AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

EVOLUTIONARY AND GENOMIC MICROBIOLOGY

Unveiling Genomic Diversity among Members of the Species *Bifidobacterium pseudolongum*, a Widely Distributed Gut Commensal of the Animal Kingdom

Gabriele Andrea Lugli,[a] Sabrina Duranti,[a] Korin Albert,[b,c] Leonardo Mancabelli,[a] Stefania Napoli,[a] Alice Viappiani,[d] Rosaria Anzalone,[a] Giulia Longhi,[d] Christian Milani,[a] Francesca Turroni,[a,e] Giulia Alessandri,[f] David A. Sela,[b,c,g] Douwe van Sinderen,[h] Marco Ventura[a,e]

[a]Laboratory of Probiogenomics, Department of Chemistry, Life Sciences, and Environmental Sustainability, University of Parma, Parma, Italy
[b]Department of Food Science, University of Massachusetts, Amherst, Massachusetts, USA
[c]Molecular and Cellular Biology Graduate Program, University of Massachusetts, Amherst, Massachusetts, USA
[d]GenProbio srl, Parma, Italy
[e]Microbiome Research Hub, University of Parma, Parma, Italy
[f]Department of Veterinary Medical Science, University of Parma, Parma, Italy
[g]Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester, Massachusetts, USA
[h]APC Microbiome Institute and School of Microbiology, Bioscience Institute, National University of Ireland, Cork, Ireland

ABSTRACT Bifidobacteria are commensals of the animal gut and are commonly found in mammals, birds, and social insects. Specifically, strains of *Bifidobacterium adolescentis*, *Bifidobacterium bifidum*, *Bifidobacterium longum*, and *Bifidobacterium pseudolongum* are widely distributed in the mammalian gut. In this context, we investigated the genetic variability and metabolic abilities of the *B. pseudolongum* taxon, whose genomic characterization has so far not received much attention. Phylogenomic analysis of the genome sequences of 60 *B. pseudolongum* strains revealed that *B. pseudolongum* subsp. *globosum* and *B. pseudolongum* subsp. *pseudolongum* may actually represent two distinct bifidobacterial species. Furthermore, our analysis highlighted metabolic differences between members of these two subspecies. Moreover, comparative analyses of genetic strategies to prevent invasion of foreign DNA revealed that the *B. pseudolongum* subsp. *globosum* group exhibits greater genome plasticity. In fact, the obtained findings indicate that *B. pseudolongum* subsp. *globosum* is more adaptable to different ecological niches such as the mammalian and avian gut than is *B. pseudolongum* subsp. *pseudolongum*.

IMPORTANCE Currently, little information exists on the genetics of the *B. pseudolongum* taxon due to the limited number of sequenced genomes belonging to this species. In order to survey genome variability within this species and explore how members of this taxon evolved as commensals of the animal gut, we isolated and decoded the genomes of 51 newly isolated strains. Comparative genomics coupled with growth profiles on different carbohydrates has further provided insights concerning the genotype and phenotype of members of the *B. pseudolongum* taxon.

KEYWORDS *Bifidobacterium*, animal commensals, bifidobacteria, genomics, next-generation sequencing

The animal and bacterial kingdoms have coevolved and coadapted to establish interspecies relationships for hundreds of millions of years (1). The Human Microbiome Project and many other recent studies have achieved astounding progress in deciphering details on the human gut microbiota composition as well as its role in human health (2, 3). Members of the genus *Bifidobacterium* are prevalent and some-

Citation Lugli GA, Duranti S, Albert K, Mancabelli L, Napoli S, Viappiani A, Anzalone R, Longhi G, Milani C, Turroni F, Alessandri G, Sela DA, van Sinderen D, Ventura M. 2019. Unveiling genomic diversity among members of the species *Bifidobacterium pseudolongum*, a widely distributed gut commensal of the animal kingdom. Appl Environ Microbiol 85:e03065-18. https://doi.org/10.1128/AEM.03065-18.

Editor Andrew J. McBain, University of Manchester

Address correspondence to Marco Ventura, marco.ventura@unipr.it.

G.A.L. and S.D. contributed equally to this work.

Received 20 December 2018
Accepted 3 February 2019
Accepted manuscript posted online 8 February 2019
Published 4 April 2019

times very abundant among the hundreds of bacterial species that inhabit the gut of humans and other mammals (2). Bifidobacteria are Gram-positive, non-spore-forming, and nonmotile bacteria that belong to the phylum *Actinobacteria*, and they represent one of the dominant microbial colonizers of the human and animal gut (4). Currently, 72 distinct bifidobacterial (sub)species are recognized, mostly isolated from the gastrointestinal tract of various animals, the human intestine and oral cavity, and insect hindgut (5–12). Previous genome analyses clearly indicate that bifidobacteria have developed a diverse number of genetic strategies to adapt to their specific ecological niches (5–12).

Recently, Milani et al. described bifidobacterial populations present among a wide range of 291 adult animals, thereby unveiling their widespread distribution across the mammalian kingdom (13). In this context, *Bifidobacterium adolescentis*, *Bifidobacterium bifidum*, *Bifidobacterium longum*, and *Bifidobacterium pseudolongum* were shown to be the predominant bifidobacterial species present in the mammalian gut (13). To date, comprehensive comparative genomic analyses of these bifidobacterial taxa have been performed (14–16), with the notable exception of the *B. pseudolongum* species, for which currently little information exists with respect to its genomic characterization.

The *B. pseudolongum* species consists of two subspecies, *B. pseudolongum* subsp. *pseudolongum* and *B. pseudolongum* subsp. *globosum*, of which *B. pseudolongum* subsp. *pseudolongum* was identified and classified in 1969 from swine feces (17). In the same year, Scardovi et al. isolated from the bovine rumen a strain they named to be *Bifidobacterium globosum*, which was further classified in 1992 as a subspecies of the *B. pseudolongum* species, i.e., *B. pseudolongum* subsp. *globosum* (18, 19). More recent studies have assessed the level of genome relatedness among the two subspecies of this taxon, casting doubt on the correct taxonomic classification of type strains previously assigned to these species (5, 20). Thus, in order to investigate the genetic variability and the metabolic capabilities of this taxon, 51 newly isolated *B. pseudolongum* strains from feces of various animals were subjected to *de novo* sequencing. The obtained genomic data sets were further supplemented with publicly available *B. pseudolongum* chromosomal sequences and were then subjected to in-depth comparative genomic analyses. Our findings revealed genotype and phenotype differences between the two subspecies of this taxon, highlighting two evolutionary routes that may be responsible for their differential host colonization preference.

RESULTS AND DISCUSSION

Isolation and genomic characterization of the *B. pseudolongum* taxon. Recently, internal transcribed spacer (ITS) bifidobacterial profiling analysis was performed on fecal samples of various animals, revealing that *B. pseudolongum* is one of the most prevalent species (13). However, the genomic characterization of *B. pseudolongum* taxon is rather limited, and comparative genome analyses have not been reported that would characterize the genetic diversity and genomic features of members of this species. Thus, through a culture-dependent approach, we isolated 51 different *B. pseudolongum* strains from fecal samples of animals (Table 1). As shown in Table 1, we isolated *B. pseudolongum* strains from 22 mammalian species, as well as from birds, such as chicken (*Gallus gallus domesticus*), duck (*Anas platyrhynchos domesticus*), pigeon (*Columba livia domestica*), and quail (*Coturnix coturnix*). Exploring the bifidobacterial biodiversity among the mammalian gut microbiota, we were able to isolate strains from the following: (i) domesticated animals, e.g., alpaca (*Vicugna pacos*), camel (*Camelus dromedarius*), cow (*Bos taurus*), donkey (*Equus africanus asinus*), goat (*Capra aegagrus hircus*), horse (*Equus ferus caballus*), pig (*Sus scrofa*), and sheep (*Ovis aries*); and (ii) wild animals kept in captivity, such as bear (*Ursus arctos*), capybara (*Hydrochoerus hydrochaeris*), mara (*Dolichotis patagonum*), fox (*Vulpes vulpes*), hare (*Lepus europaeus*), hippopotamus (*Hippopotamus amphibius*), kangaroo (*Macropus rufus*), mouse (*Mus musculus*), porcupine (*Erethizon dorsatum*), mouflon (*Ovis musimon*), roe deer (*Capreolus capreolus*), and tapir (*Tapirus terrestris*). Furthermore, several canine breeds were inves-

TABLE 1 General genome features of *B. pseudolongum* strains

Strain	Origin	Cluster	No. of bases	Coverage (fold)	No. of contigs	No. of ORFs	No. of rRNAs	No. of tRNAs	No. of TUGs	HGT %	CRISPR system	RM system type	Prophages	Accession no.
Newly sequenced genomes														
22506[a]	Goat	glob-A	1,913,074	177	27	1,569	3	52	37	9.8	I-E	2	None	RYUB00000000
22511[a]	Goat	glob-A	2,018,875	233	35	1,645	3	52	62	11.9	I-E	3	None	RYUA00000000
102015[a]	Cow	glob-A	1,918,182	497	29	1,605	1	45	51	11	I-U	1	None	RYTZ00000000
102017[a]	Cow	glob-A	1,884,108	914	21	1,558	1	46	33	8.5	II	3	None	RYTY00000000
112206[a]	Goat	glob-A	1,947,489	851	29	1,613	1	47	54	10.8	None	5	None	RYTX00000000
1511B[b]	Chicken	glob-B	1,971,056	152	11	1,604	3	51	37	11.2	I-E	1	None	RYVN00000000
1513B[b]	Piglet	pseudo	1,855,765	124	6	1,532	3	52	34	8.3	I-C	3	None	RYVM00000000
1546B[b]	Chicken	glob-B	1,955,522	68	15	1,635	2	53	13	10.8	I-E	1	1	RYVL00000000
1550B[b]	Pigeon	glob-B	2,000,909	204	5	1,659	4	53	32	11.6	II	None	1	RYVK00000000
1565B[b]	Donkey	glob-B	1,956,592	266	15	1,638	2	53	12	11.2	I-E	1	1	RYVJ00000000
1577B[b]	Horse	glob-B	1,950,120	219	13	1,590	3	52	14	9.2	None	1	None	SBKZ00000000
1578B[b]	Duck	glob-B	1,950,074	155	13	1,590	3	52	12	9.4	None	1	None	RYVI00000000
1604B[b]	Piglet	pseudo	1,882,038	123	6	1,541	2	52	23	8.3	I-E	1	None	RYVH00000000
1612B[b]	Pig	pseudo	1,891,497	155	5	1,583	3	53	38	7.8	None	1	1	RYVG00000000
1616B[b]	Camel	glob-A	2,031,794	246	39	1,676	2	52	49	10.1	I-E	None	1	RYVF00000000
1629B[b]	Tapir	pseudo	1,852,430	156	6	1,542	1	52	30	9.5	None	5	None	RYVE00000000
1655B[b]	Alpaca	glob-A	2,032,959	234	31	1,680	3	52	16	11	None	1	None	SBKY00000000
1678B[b]	Hippopotamus	glob-A	2,037,960	84	34	1,680	3	52	21	11.1	None	2	None	SBKX00000000
1770B[b]	Fox	glob-B	2,114,575	118	21	1,768	4	65	59	12.3	I-E	1	None	RYVD00000000
1780B[b]	Goat	glob-A	1,981,485	177	27	1,618	2	52	39	11	None	4	None	RYVC00000000
1791B[b]	Roe deer	glob-A	2,130,549	148	34	1,773	2	58	82	12.4	I-E	5	None	RYVB00000000
1805B[b]	Hare	glob-B	2,102,705	113	12	1,729	3	52	58	12.6	I-U	None	2	RYVA00000000
2000B[b]	Dog	glob-B	2,112,310	177	34	1,799	2	53	80	11.7	None	2	1	RYUZ00000000
2001B[b]	Dog	glob-B	2,095,316	137	21	1,736	3	52	62	11.8	II	2	1	RYUY00000000
2002B[b]	Dog	glob-B	2,057,959	271	20	1,722	2	54	43	9.7	I-C	2	1	RYUX00000000
2003B[b]	Dog	glob-B	2,148,604	145	37	1,816	3	59	77	12.7	I-C	1	2	RYUW00000000
2004B[b]	Cow milk	glob-A	1,910,716	305	17	1,571	2	53	32	9.4	None	4	None	RYUV00000000
2009B[b]	Dog	glob-B	2,107,529	141	42	1,738	4	58	62	12.5	I-C	3	None	RYUU00000000
2012B[b]	Dog	glob-B	2,070,775	215	35	1,753	3	53	82	13.2	None	3	2	SBKW00000000
2017B[b]	Quail	glob-B	2,189,937	178	12	1,891	3	60	20	14.7	II	2	2	RYUT00000000
2019B[b]	Sheep	glob-A	1,943,135	179	30	1,602	1	52	46	10.5	I-E	3	None	RYUS00000000
2023B[b]	Bear	glob-B	1,983,829	214	19	1,636	2	53	23	10.6	None	None	1	SBKV00000000
2029B[b]	Hippopotamus	glob-B	2,132,640	232	13	1,710	2	52	13	12.5	I-E	3	1	RYUR00000000
2032B[b]	Capybara	glob-B	2,132,192	250	14	1,709	2	52	12	12.4	I-E	3	1	RYUQ00000000
2048B[b]	Pigeon	glob-B	2,030,037	272	11	1,720	2	55	53	13.7	None	1	2	RYUP00000000
2049B[b]	Quail	glob-B	2,189,952	177	13	1,893	3	60	20	14.9	II	1	2	RYUO00000000
2054B[b]	Dog	pseudo	1,974,289	161	19	1,611	2	52	54	9.7	None	3	None	RYUN00000000
2071B[b]	Dog	glob-B	2,084,649	247	36	1,767	3	60	73	12.8	I-C	None	3	RYUM00000000
2072B[b]	Hippopotamus	glob-B	2,083,912	174	21	1,749	3	53	62	12.1	None	1	3	SBKU00000000
2083B[b]	Sheep	glob-A	2,059,818	131	20	1,679	3	53	45	9.8	II	None	None	RYUL00000000
2086B[b]	Hippopotamus	glob-B	1,996,821	293	20	1,633	2	52	14	11.9	I-U	2	None	RYUK00000000
2088B[b]	Hippopotamus	glob-B	1,991,771	295	17	1,633	1	52	14	11.4	I-U	2	None	RYUJ00000000
2089B[b]	Mara	glob-B	2,047,482	322	9	1,633	2	52	22	10.8	II	None	None	RYUI00000000
2093B[b]	Snake	glob-B	2,132,964	248	23	1,736	2	53	96	13	II	1	1	RYUH00000000
2098B[b]	Mouflon	glob-B	1,998,347	215	13	1,633	2	52	12	11.9	I-U	2	None	RYUG00000000
2103B[b]	Mouse	glob-B	2,080,594	246	33	1,722	2	52	80	12.5	None	None	1	SBKT00000000
2105B[b]	Mouse	glob-B	1,964,892	249	21	1,608	2	52	31	10.1	II	2	None	RYUF00000000
2109B[b]	Kangaroo	glob-B	2,039,410	208	6	1,659	2	53	34	11.8	I-E	None	1	SBKS00000000
2113B[b]	Hippopotamus	glob-B	1,996,348	218	15	1,631	2	52	13	11.8	I-U	2	None	RYUE00000000
2114B[b]	Porcupine	glob-B	2,047,198	273	19	1,646	2	52	22	11.7	II	None	None	RYUD00000000
2115B[b]	Porcupine	glob-B	2,016,784	329	12	1,627	3	52	17	11.5	II	None	None	RYUC00000000
Public genomes														
DSM20092	Cow	glob-A	1,978,157	NA[c]	1	1,598	4	52	21	9.1	None	1	None	CP017695
LMG11569	Cow	glob-A	1,935,255	NA	26	1,565	3	52	29	9.2	None	3	None	JGZG00000000
LMG11571	Pig	pseudo	1,898,684	NA	11	1,494	3	52	61	8.8	I-E	1	None	JGZH00000000
PV8-2	Human infant	glob-B	2,032,698	NA	1	1,685	4	52	47	11.8	I-U	None	1	CP007457
UMB-MBP-01	Mouse	glob-B	2,008,102	NA	1	1,645	4	52	56	10.5	II	1	1	CP022544
1370B	Pig	pseudo	1,902,036	NA	17	1,582	5	52	54	10.8	None	2	None	PCHI00000000
1520B	Hamster	glob-B	2,008,481	NA	19	1,630	4	52	59	10.6	None	1	None	PCHH00000000
1549B	Brahma chicken	glob-B	1,959,452	NA	19	1,647	2	53	16	11.4	I-E	1	1	PCHG00000000
1595B	Pig	pseudo	1,936,418	NA	16	1,603	4	53	67	10.4	None	2	None	PCHF00000000

[a]Genome sequences decoded using a NextSeq platform (Illumina, UK).
[b]Genome sequences decoded using a MiSeq platform (Illumina, UK).
[c]NA, not applicable.

tigated, allowing the isolation of eight *B. pseudolongum* strains, while a single strain was collected from a snake (*Boa constrictor*).

Accordingly, the genomes of these 51 newly isolated *B. pseudolongum* strains were decoded by means of a next-generation sequencing (NGS) approach, and the obtained

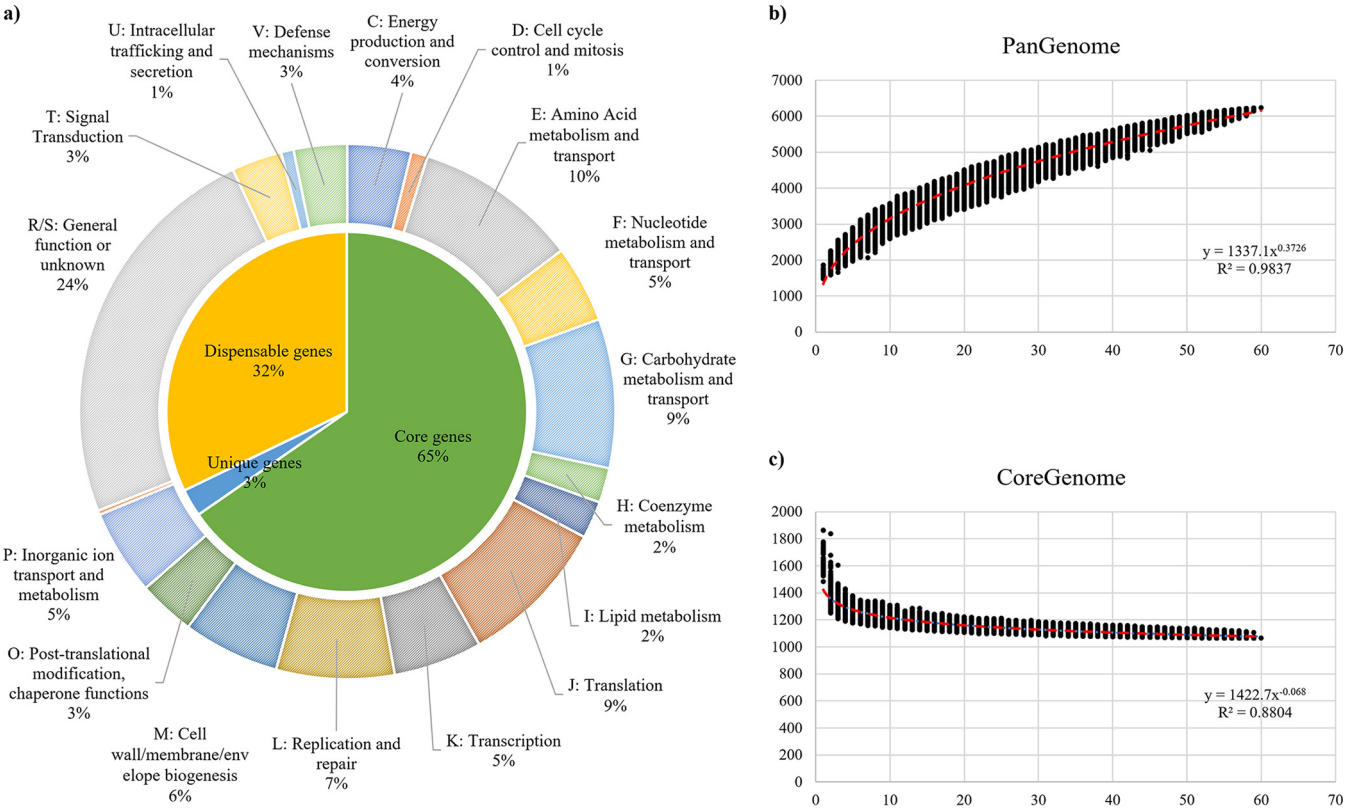

FIG 1 *B. pseudolongum* pangenome and core genome. (a) Number of core genes (green), unique genes (blue), and dispensable genes (yellow) identified in the pangenome analysis (internal pie chart). COG classifications of the whole *B. pseudolongum* pangenome are highlighted in different colors (external pie chart). (b) Pangenome size based on sequential addition of the 60 *B. pseudolongum* genomes. (c) Core genome size based on sequential addition of the *B. pseudolongum* genomes.

data were analyzed together with nine publicly available *B. pseudolongum* sp. genomes (Table 1). All *B. pseudolongum* genomes of newly isolated strains were sequenced to a coverage depth that ranged from 68-fold to 914-fold, which upon assembly resulted in five to 42 contigs (Table 1). We were able to predict the contig orientation and the order for each draft genome using the complete genome of the type strain *B. pseudolongum* DSM 20092 (NCBI accession no. CP017695) as a reference sequence. The individual genome length of each *B. pseudolongum* strain was retrieved by the assembly, resulting in genomes whose sizes were shown to range from 1,852,430 to 2,189,952 bp (Table 1). As outlined in Table 1, the number of predicted open reading frames (ORFs) in each genome ranged from 1,494 ORFs for *B. pseudolongum* subsp. *pseudolongum* LMG 11571 to 1,893 ORFs for *B. pseudolongum* subsp. *globosum* 2049B, displaying a higher average number of ORFs in genomes of isolates that belong to the *B. pseudolongum* subsp. *globosum* taxon (Table 1). Furthermore, 81% of the predicted *B. pseudolongum* ORFeome was functionally classified on the basis of the eggNOG database (21). A large proportion of classified genes was predicted to be involved in housekeeping functions, amino acid and carbohydrate metabolism, and associated transport activities (Fig. 1a), resembling the functional content in genomes of other members of the *Bifidobacterium* genus (6, 20, 22, 23).

Pangenome and core genome of the *B. pseudolongum* species. All identified genes for each *B. pseudolongum* isolate were employed for a comparative genome analysis of this species. Thus, pangenome and core-genome analyses of this taxon were undertaken following a previously described method based on Clusters of Orthologous Groups (COGs) (24). This analysis resulted in the identification of 6,179 COGs, representing the pangenome of the *B. pseudolongum* species. Notably, 1,069 COGs of this pangenome were shared among the 60 *B. pseudolongum* genomes, thus representing

the so-called core genome of this taxon (Fig. 1c). Furthermore, dispensable genes present in two or more (but not all) strains, and truly unique genes (TUGs) which are present in just one of the analyzed strains were also identified. TUGs of each *B. pseudolongum* strain, ranging from 12 for *B. pseudolongum* 1565B to 96 for *B. pseudolongum* 2093B, were detected, with an average of 41 TUGs per genome. When this number is compared to other previously analyzed bifidobacterial pangenomes, such as *Bifidobacterium adolescentis* and *Bifidobacterium bifidum* (14, 15), it appears that *B. pseudolongum* ORFeome contains a relatively small number of TUGs, indicating that we have sufficiently explored the variability within this taxon.

The pangenome size, when plotted versus the number of included bifidobacterial genomes, shows that the power trendline tends to reach a plateau, where genomic data from the last strain added to the analysis do not substantially expand the total gene pool (Fig. 1b). Therefore, according to these data, the resulting pangenome curve suggests a "closed" *B. pseudolongum* pangenome. This means that after addition of the 60th *B. pseudolongum* genome, any further genome additions will result in only minor increases in the pangenome. In the same fashion, a pangenome was previously disclosed for *Bifidobacterium animalis*, *Bifidobacterium breve*, and *Bifidobacterium longum* species (16, 25, 26).

Phylogenomic analyses and evolutionary development of *B. pseudolongum* taxon. The comparative genomic analysis also allows the assessment of phylogeny of the *B. pseudolongum* taxon, applying a previously described methodology (20, 22, 27). To perform such a phylogenomic analysis, the genomes of the 60 *B. pseudolongum* strains were used for this exercise, as well as the gene sequence of *B. animalis* subsp. *animalis* ATCC 25527, which served as an outgroup. Furthermore, paralogs were excluded from the 1,069 *B. pseudolongum* core genes identified and described above by means of PGAP analysis (see Materials and Methods), resulting in 1,038 genes of which the concatenated amino acid sequences were employed to build a supertree (Fig. 2). The resulting *B. pseudolongum*-based phylogenomic supertree clearly shows the presence of two major groups which were assumed to correspond to the subspecies of this taxon. Within the supertree, the type strain *B. pseudolongum* subsp. *pseudolongum* LMG 11571 shared the same phylogenetic branch as that of seven *B. pseudolongum* strains, while the type strain *B. pseudolongum* subsp. *globosum* LMG 11569 was positioned in the second branch together with the remaining 52 strains (Fig. 2). Thus, our isolation attempts from stool samples of animals indicates that strains belonging to the *B. pseudolongum* subsp. *globosum* group outnumber members of the *B. pseudolongum* subsp. *pseudolongum* group.

Interestingly, strains of the smaller cluster, represented by members of the *B. pseudolongum* subsp. *pseudolongum* group, were all isolated from fecal samples of pigs, with the exception of the 1629B and 2054B strains that were retrieved from tapir and dog fecal samples, respectively (Fig. 2). Furthermore, none of the 52 strains belonging to the *B. pseudolongum* subsp. *globosum* group were of porcine origin, a finding that suggests that members of the *B. pseudolongum* subsp. *pseudolongum* taxon reside in a specific ecological niche. The *B. pseudolongum* subsp. *globosum* group also appeared to include a conserved subcluster composed of 15 strains (Fig. 2). Remarkably, and similar to the *B. pseudolongum* subsp. *pseudolongum* group, members of this subcluster were isolated from specific animals in which the other 37 strains of the *B. pseudolongum* subsp. *globosum* group had not been detected. In this context, the hosts of this subcluster were represented by several ruminant species, such as cow, goat, sheep, camel, and alpaca, perhaps highlighting a genomic adaptation of these *B. pseudolongum* strains to a diet rich in grass-based fiber content. In contrast, the remaining strains of the *B. pseudolongum* subsp. *globosum* group were isolated from a wide selection of mammals, both herbivorous and carnivorous, and including birds, such as chicken, pigeon, quail, and duck (Fig. 2). Thus, members of the *B. pseudolongum* species appear to be widespread among the animal kingdom, but it appears that different animals host a specific cluster of members of this taxon.

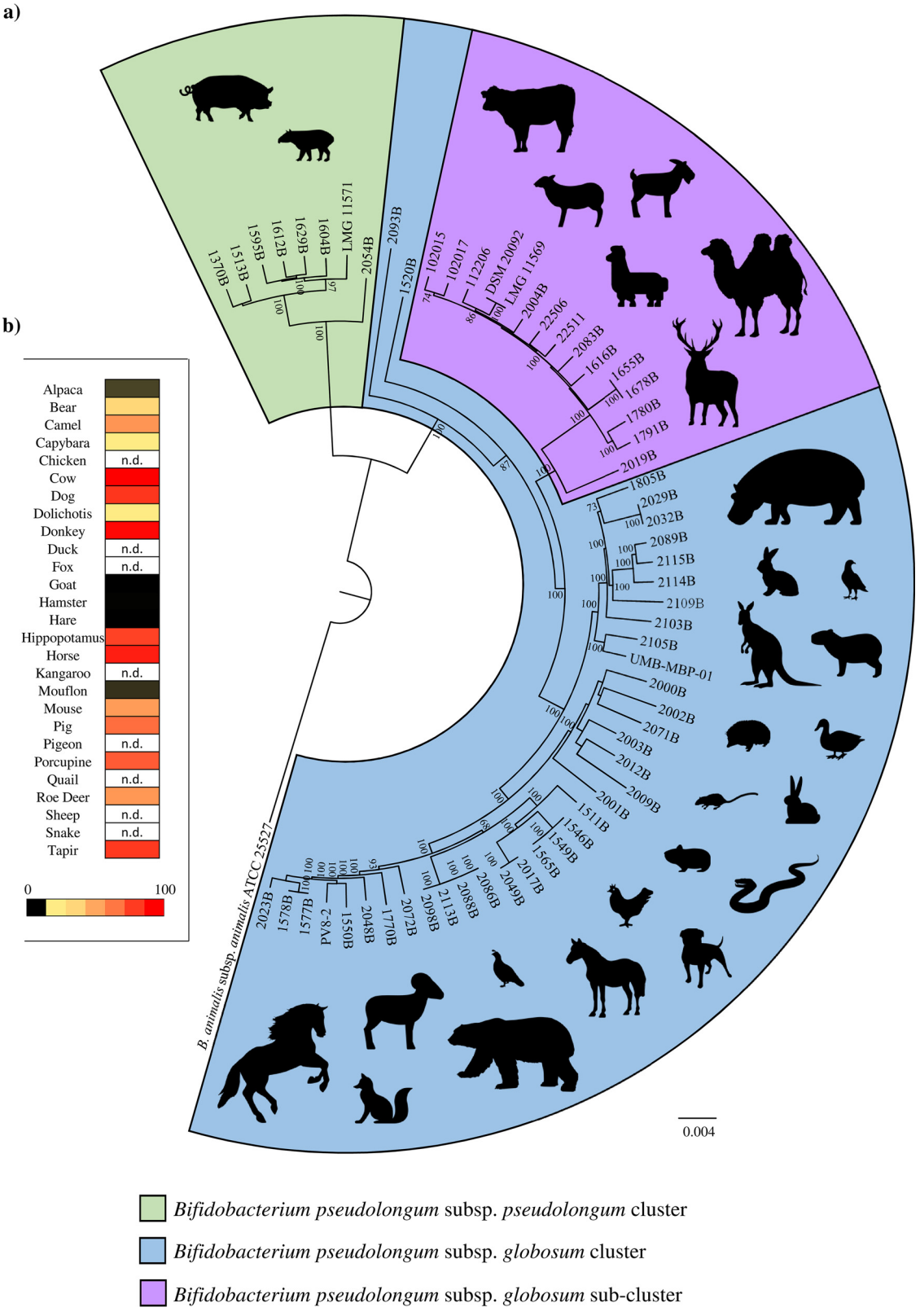

FIG 2 Phylogenomic tree of the *B. pseudolongum* taxon. (a) Proteomic tree based on the concatenation of 1,038 core genes identified in the pangenome analysis of the 60 *B. pseudolongum* strains. Phylogenetic groups are highlighted in different colors. The tree was constructed by the neighbor-joining method, and the genome sequence of *B. animalis* subsp. *animalis* ATCC 25527 was used as an outgroup. Bootstrap percentages above 50 are shown at node points, based on 1,000 replicates. (b) Relative abundance of *B. pseudolongum* within analyzed animals based on ITS bifidobacterial profiling. n.d., not detected.

Using the average nucleotide identity (ANI) approach, between all collected *B. pseudolongum* pairs, we highlighted the genome synteny among members of this species, with associated ANI values ranging from 92.7% to 99.9% (see Table S1 in the supplemental material). Notably, two strains that display an ANI value of <95% may be considered to belong to two distinct species (28). In this regard, in our previous phylogenomic studies concerning the *Bifidobacterium* genus, we applied an ANI threshold of about 94% to discriminate between bifidobacterial species (6, 20, 27). Assessment of ANI values among members of a given *B. pseudolongum* subspecies showed that *B. pseudolongum* subsp. *pseudolongum* members exhibit higher values (greater than 98.4%) than those of the *B. pseudolongum* subsp. *globosum* group (greater than 94.8%) (Table S1). Thus, despite the absence of a subspecies-specific core genome within the species, phylogenomic analyses highlight extensive variability between genomes of members of the two groups. Altogether, our findings suggest that *B. pseudolongum* subsp. *pseudolongum* and *B. pseudolongum* subsp. *globosum* represent distinct bifidobacterial species rather than two separate subspecies.

Genome plasticity of the *B. pseudolongum* species. In order to identify genes that may have been acquired by horizontal gene transfer (HGT), the genomes of the 60 *B. pseudolongum* strains were screened using the software tool COLOMBO (see Materials and Methods), resulting in the identification of HGT genes that make up 7.8% of the total gene pool in the case of *B. pseudolongum* 1612B and up to 14.9% of that of *B. pseudolongum* 2049B (Table 1). The functional classification based on the eggNOG database revealed that just 45% of the predicted alien genes could be functionally assigned (Fig. S2). Furthermore, unlike the whole proteome of the *B. pseudolongum* taxon, 44% of these genes are of unknown function, followed by genes predicted to encode replication and repair systems and cell wall and membrane proteins (Fig. S2). Interestingly, members of the *B. pseudolongum* subsp. *pseudolongum* group exhibit a lower average percentage of predicted HGT events (9.2%), followed by members of the ruminant subcluster of *B. pseudolongum* subsp. *globosum* group (10.4%) and the remaining strains (11.8%). Thus, *B. pseudolongum* subsp. *globosum* strains that were isolated from a wider number of animals seem to possess an enhanced propensity to acquire alien genes. In a similar fashion, the length of the analyzed genomes and their corresponding number of dispensable genes reflect the same trend. In fact, the average length of the *B. pseudolongum* subsp. *pseudolongum* genomes is significantly lower than that of the *B. pseudolongum* subsp. *globosum* ($P < 0.001$), i.e., 1.89 and 2.03 Mb, respectively (Table 1), while the number of dispensable genes increases to 21% in members of the *B. pseudolongum* subsp. *globosum* group. Thus, a correlation seems to exist between genome size, number of dispensable genes, and ecological niches of the analyzed *B. pseudolongum* strains.

It has already been shown that prophage sequences represent a large part of the bifidobacterial genetic repertoire acquired by HGT (22). We therefore performed a prophage profiling of the pangenome of the *B. pseudolongum* by means of previous bifidophage sequences classified by Lugli et al. (29). This genomic screen revealed the presence of 35 complete prophage sequences harbored by 25 *B. pseudolongum* strains (Table 1). In this context, the retrieved prophage sequences were predicted to belong to the *Siphoviridae* family and were shown to include modules that putatively encode functions involved in lysogeny, DNA replication, DNA packaging, head and tail morphogenesis, and host lysis (30). Furthermore, 10 apparently incomplete prophage sequences, i.e., prophages that exhibit extensive genome degeneration, were identified in just as many strains. Overall, this analysis showed that complete prophages were retrieved from *B. pseudolongum* subsp. *globosum* strains only, with the exception of one prophage sequence contained in the *B. pseudolongum* subsp. *pseudolongum* 1612B genome (Table 1). Moreover, among members of the *B. pseudolongum* subsp. *globosum* taxon, 33 out of 34 prophage sequences were identified in strains belonging to the wider phylogenetic cluster (Fig. 2).

Defense mechanisms of the *B. pseudolongum* species. In order to investigate the ability of members of the *B. pseudolongum* group to defend themselves against invasion by foreign DNA, we investigated the presence of active CRISPR-Cas systems, as previously reported for the type strains of the genus *Bifidobacterium* by Briner et al. (31). Among the 60 *B. pseudolongum* genomes analyzed, we observed a high percentage of CRISPR-Cas system occurrence (67%) (Tables 1 and S2). According to the *cas* gene content and CRISPR length, we identified 28 type I systems and 12 type II systems, while type III systems appear to be absent in our assessed strain collection (Fig. S2). In this context, the lower number of type II systems, compared to the type I systems, reflects the previously stated notion that type II systems are fairly rare in nature, occurring in just 5% of currently known bacteria (32). When scrutinizing the CRISPR sequences, 14 different repeats were identified, of which each sequence corresponds to a specific (sub)type system, e.g., I-E, I-C, and I-U. Interestingly, strains 2009B and 2093B possess two different CRISPR sequences, yet only one CRISPR-Cas system appears to be complete, being type I-C and type II, respectively (Table S2). Type I systems, including the I-E, I-C, and I-U subtypes, were the most commonly encountered CRISPR-Cas systems and the only ones identified in the *B. pseudolongum* subsp. *pseudolongum* cluster. In contrast, type II systems were identified only in members of the *B. pseudolongum* subsp. *globosum* cluster. Furthermore, the type II system distribution among the *B. pseudolongum* strains, based on the phylogenomic tree of this taxon, highlighted a more recent acquisition of this system with respect to type I systems (Fig. 2). Moreover, genome sequencing of the *B. pseudolongum* strains allowed the identification of prophages matching spacer sequences of CRISPR loci identified in the study. Strains possessing a CRISPR-Cas system displayed at least one match with prophage sequences with an identity value ranging from 97% to 100% (Table S3). Interestingly, both subspecies seem to have acquired immunity through these CRISPR-Cas systems against *B. pseudolongum* subsp. *globosum*-derived prophages retrieved from strains of the wide cluster (94% of the hits) (Table S3).

Other multiprotein complexes that prevent acquisition of foreign DNA are represented by restriction-modification (RM) systems (33). Screening the predicted *B. pseudolongum* proteome for RM systems revealed that type I RM systems represent the predominant gene cluster, being present in 35 strains, followed by type II RM systems identified on the genomes of 25 *B. pseudolongum* strains (Tables 1 and S4). Besides, type III and IV RM systems were identified at a lower frequency, i.e., in six and 16 strains, respectively (Fig. S2). Remarkably, members of the *B. pseudolongum* subsp. *pseudolongum* group encoded an average of 2.3 complete RM systems per genome, while *B. pseudolongum* subsp. *globosum* group encoded an average of only 1.5 systems. Thus, it appears that the genomes of *B. pseudolongum* species isolated from pigs were equipped with a wider genomic arsenal to prevent the invasion by foreign DNA sequences.

Taken together, these findings indicate a higher occurrence of prophage sequences in members of the *B. pseudolongum* group, which possess a lower number of genetic clusters encoding proteins aimed to defend themselves against foreign DNA invasion (Table 1). Thus, our analyses revealed a correlation between a lower number or RM/CRISPR systems and an increase of dispensable genes of the *B. pseudolongum* subsp. *globosum* taxon, similar to what was recently shown for the *B. breve* taxon (34). Accordingly, the higher genome plasticity of members of the *B. pseudolongum* subsp. *globosum* group, perhaps as a result of the lower abundance of RM/CRISPR systems, may explain an enhanced ability to adapt to different ecological niches such as mammals and birds.

***B. pseudolongum* carbohydrate-active enzymes and growth profiles on different carbohydrates.** To identify the carbohydrate-active enzyme repertoire of each analyzed *B. pseudolongum* genome, we investigated the presence of genes predicted to encode glycosyl hydrolases (GHs). This analysis identified members of 28 GH families, highlighting a predominance of genes encoding GHs belonging to the GH13, GH43,

and GH36 families, predicted to be responsible for the breakdown of plant-derived polysaccharides, such as starch, and a wide range of carbohydrates, e.g., maltodextrin, melibiose, and raffinose (Fig. 3). Together with GH2, GH3, GH25, GH30, GH51, and GH77 families, which were also identified in the glycobiome of each *B. pseudolongum* strain, these GH families represent the core glycobiome of the *B. pseudolongum* taxon. The *B. pseudolongum* subsp. *pseudolongum* group was shown to encode a higher number of GH112 than did the *B. pseudolongum* subsp. *globosum* group, encoding lacto-*N*-biose phosphorylase. In contrast, the glycobiome of members of the *B. pseudolongum* subsp. *globosum* group unveiled a higher number of GH29, GH31, and GH42 families, encoding putative α-L-fucosidase, α-glucosidase, and β-galactosidase activity, respectively (Fig. 3). Furthermore, analysis of the glycobiomes of members of the *B. pseudolongum* subsp. *globosum* group isolated from ruminant species revealed that GH1 and GH94 members are rare compared to other strains belonging to *B. pseudolongum* subsp. *globosum*. Altogether, members of the *B. pseudolongum* subsp. *pseudolongum* group and the ruminant subcluster of the *B. pseudolongum* subsp. *globosum* taxon possessed a lower number of genes encoding GHs than did the remainder of the *B. pseudolongum* strains analyzed here. Accordingly, *B. pseudolongum* subsp. *globosum* strains that were not isolated from ruminants revealed an extensive variability in GH enzymes which would be consistent with their ability to colonize a wider number of different animal hosts (Fig. 3).

In order to validate the above-described genomic-based analyses, we carried out growth experiments of *B. pseudolongum* species strains on 27 carbohydrates including host- and plant-derived glycans as the sole carbon source (Fig. 3). As displayed in Fig. 3, all *B. pseudolongum* subsp. *pseudolongum* strains were able to grow on several sugars, such as glucose, glycogen, lactose, maltodextrin, maltose, melibiose and raffinose. In contrast, fermentation capabilities of carbohydrates were shown to be quite varied for *B. pseudolongum* subsp. *globosum*. However, statistical analyses were performed to corroborate the observed growth differences between *B. pseudolongum* subsp. *pseudolongum* and *B. pseudolongum* subsp. *globosum* strains on different sugars. As shown in Fig. 4, the comparison of metabolic capabilities between *B. pseudolongum* subsp. *globosum* and *B. pseudolongum* subsp. *pseudolongum* strains showed a significant growth difference ($P < 0.05$) for 18 carbohydrates. Specifically, *B. pseudolongum* subsp. *globosum* strains displayed higher growth performances when the growth medium was supplemented with cellobiose, rhamnose, starch, trehalose, *N*-acetyl-D-galactosamine, or *N*-acetyl-D-glucosamine. On the other hand, *B. pseudolongum* subsp. *pseudolongum* strains were shown to grow significantly better in medium supplemented with glycogen, glucose, pullulan, maltose, ribose, lactose, sucrose, maltodextrin, melibiose, turanose, fructose, mannitol, or lacto-*N*-tetraose (Fig. 4). As expected, the higher growth performance of *B. pseudolongum* subsp. *pseudolongum* grown in the presence of these carbohydrates is consistent with the abundance of GH112 as unveiled by the *in silico* glycobiome analysis (Fig. 3). Interestingly, in contrast to *B. pseudolongum* subsp. *pseudolongum* strains, members of the *B. pseudolongum* subsp. *globosum* subcluster display limited growth on glucose and lactose. Strains of the *B. pseudolongum* subsp. *globosum* subcluster have been isolated from ruminants, which are likely to have a rather low abundance of simple sugars, such as glucose and lactose, in the large intestine since most of these carbohydrates are metabolized in the rumen (35). Thus, members of the *B. pseudolongum* subsp. *globosum* subcluster may have evolved genetic capabilities toward the metabolism of complex dietary carbohydrates that end up in the large intestine of ruminants without being hydrolyzed (35).

Conclusions. The current study provides insights regarding the genotype and phenotype of members of the *B. pseudolongum* taxon. The genome sequencing of 51 newly isolated strains from mammals and birds allowed us to perform a comparative genomic analysis unveiling the genetic makeup of members of the *B. pseudolongum* subsp. *pseudolongum* and *B. pseudolongum* subsp. *globosum* subspecies. Phylogenomic analyses highlighted a phylogenetic cluster consisting of eight strains and representing

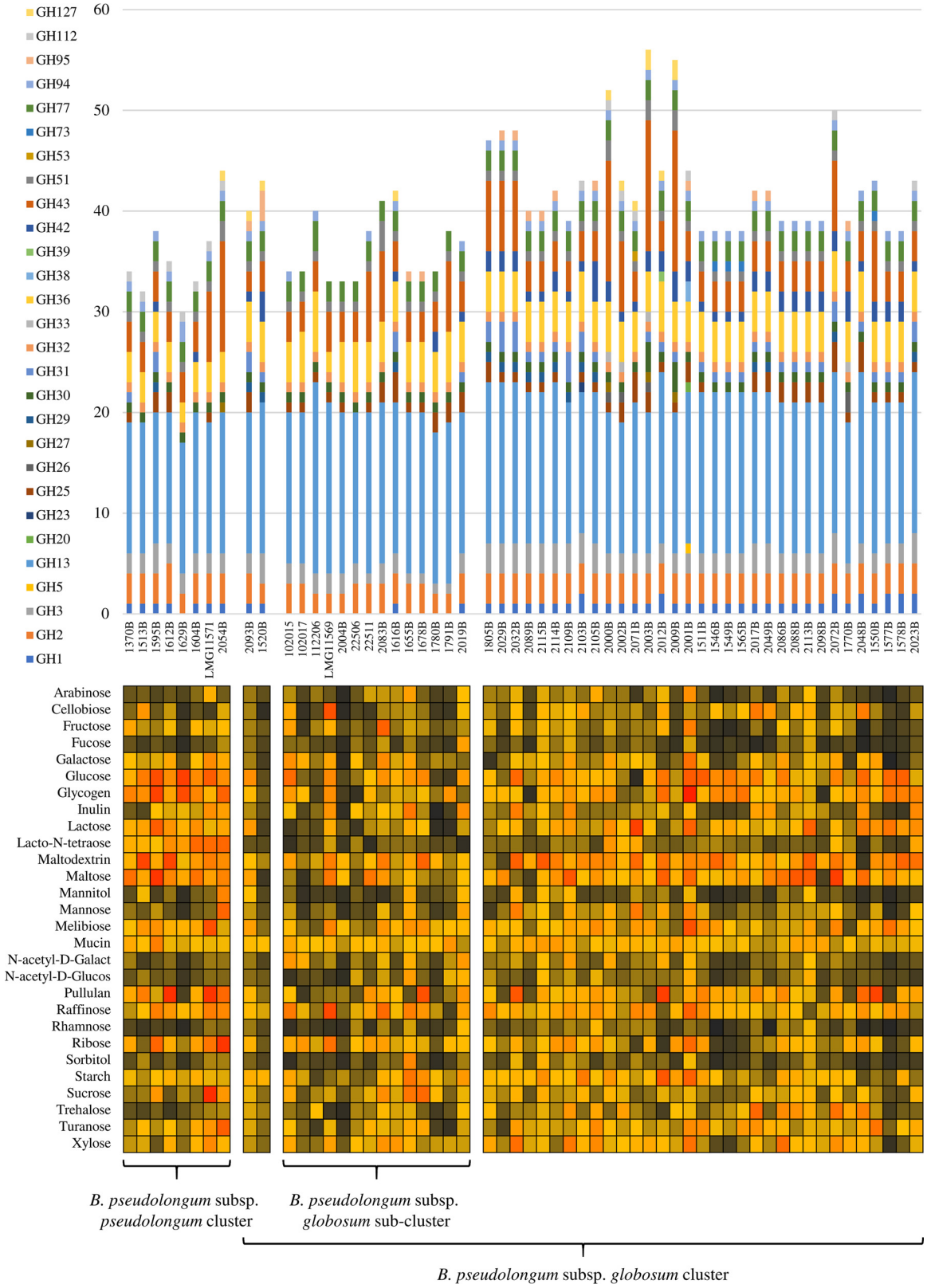

FIG 3 Predicted glycobiome of the *B. pseudolongum* species and relative growth performances. The top of the image represents the predicted glycobiome of the 60 *B. pseudolongum* strains, while the heat map at the bottom of the image depicts growth performances of *B. pseudolongum* strains on different sugars. *N*-acetyl-D-Galact, *N*-acetyl-D-galactosamine; *N*-acetyl-D-Glucos, *N*-acetyl-D-glucosamine.

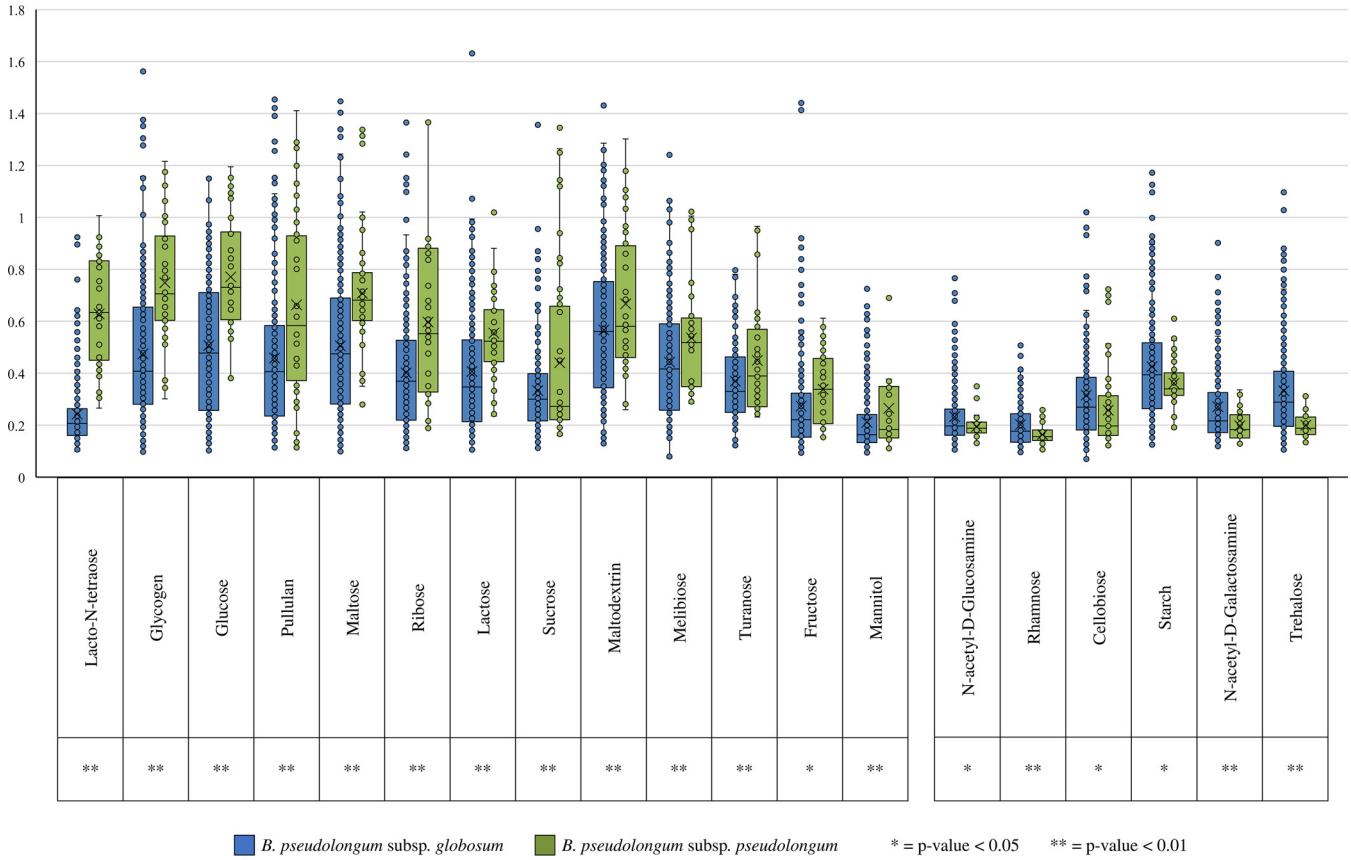

FIG 4 Statistical analysis based on the carbohydrate growth assays. Whisker plot based on optical density values of those sugars that results in a *P* value of <0.05 between subspecies (Student's *t* test). The *x* axis represents the sole carbon source used for the growth experiments, while the *y* axis shows the optical density values obtained for *B. pseudolongum* subsp. *pseudolongum* strains (green) and *B. pseudolongum* subsp. *globosum* strains (blue). Dots reflect the distribution of a data set, while the boxes represent 50% of the data set, distributed between the 1st and 3rd 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.

members of the *B. pseudolongum* subsp. *pseudolongum* group that were exclusively isolated from the gut of pigs. Conversely, members of the *B. pseudolongum* subsp. *globosum* cluster were isolated from a plethora of animals, reflecting their relatively (compared to *B. pseudolongum* subsp. *pseudolongum*) high genome variability and providing support for a possible taxonomic reclassification of these subspecies into two separate species. Furthermore, *in silico* analyses revealed an apparently higher genome plasticity in members of the *B. pseudolongum* subsp. *globosum* group. *B. pseudolongum* subsp. *pseudolongum* strains, on the other hand, showed a higher number of defense systems to prevent foreign DNA invasion, perhaps also explaining their smaller genomes. Moreover, differences between subspecies were highlighted through the prediction of their glycobiome and growth profiles on different carbohydrates. Altogether, these results highlight that *B. pseudolongum* subsp. *pseudolongum* strains evolved as commensals of a specific ecological niche, trying to minimize the acquisition of alien DNA, while members of the *B. pseudolongum* subsp. *globosum* group evolved in an opposite direction in order to colonize the gastrointestinal tract of a wide range of animals.

MATERIALS AND METHODS

Bifidobacterial genome sequences. We retrieved the complete genome sequences of three *B. pseudolongum* strains from the National Center for Biotechnology Information (NCBI) public database. In the same fashion, partial genome sequences of six *B. pseudolongum* strains were retrieved from the NCBI in order to include the type strain of both subspecies and assembled genomes that consisted of <20 contigs.

Recovery of bifidobacterial strains and growth conditions. One gram of a given fecal sample was mixed with 9 ml of phosphate-buffered saline (PBS; pH 6.5). Serial dilutions and subsequent platings were performed using de Man-Rogosa-Sharpe (MRS) agar (Scharlau Chemie, Barcelona, Spain) supplemented with 0.05% (wt/col) L-cysteine hydrochloride and 50 μg/ml mupirocin (Delchimica, Italy). The agar plates were incubated in an anaerobic atmosphere (2.99% H_2, 17.01% CO_2, and 80% N_2) in a chamber (Concept 400; Ruskin) at 37°C for 48 h. Approximately 3,500 colonies were selected and were subcultivated in MRS broth supplemented with 0.05% (wt/col) L-cysteine hydrochloride in anaerobic chamber at 37°C for 16 h. DNA was extracted using GenElute bacterial genomic DNA kits (Sigma-Aldrich) following the manufacturer's guide and then subjected to a *B. pseudolongum*-specific species identification PCR using primers Blong1 (5′-TTCCAGTTGATCGCATGGTC-3′) and Blong2 (5′-GGGAAGCCGTATCTCTACGA-3′), which had been designed on the 16S rRNA gene sequence of this species. PCR products were detected with SYBR safe DNA gel stain after amplification according to the following protocol: one cycle of 94°C for 5 min, followed by 30 cycles of 94°C for 30 s, 54°C for 30 s, and 72°C for 50 s, and finally one cycle of 72°C for 5 min. The isolated strains employed in this study are listed in Table 1. Furthermore, in order to discriminate between subspecies, PCR analysis were performed using DNA extracted from fecal samples of animals, as previously described by Milani et al. (13). The specific primers for the *B. pseudolongum* subsp. *pseudolongum* taxon are Bpseudolo_F (5′-CAGGCGTTCCTGTGGTTC-3′) and Bpseudolo_R (5′-GCGATGATGGCGAATGAC-3′), while primers B.p.glob_F (5′-GCAAGTCTCCAATGTTGAGG-3′) and B.p.glob_R (5′-CTGTGCGGACGAGACGTAG-3′) were used for the *B. pseudolongum* subsp. *globosum* taxon. Amplicons were detected with SYBR safe DNA gel stain after amplification according to the following protocol: one cycle of 94°C for 5 min, followed by 30 cycles of 94°C for 20 s, 54°C or 58°C for 20 s, and 72°C for 30 s, and finally one cycle of 72°C for 5 min.

Chromosomal DNA extraction. In order to perform chromosomal DNA extraction, *B. pseudolongum* strains were inoculated in de Man-Rogosa-Sharpe (MRS; Scharlau Chemie) medium supplemented with 0.05% (wt/vol) L-cysteine hydrochloride and incubated at 37°C in an anaerobic atmosphere (2.99% [vol/vol] H_2, 17.01% [vol/vol] CO_2, and 80% [vol/vol] N_2) using an anaerobic chamber (Concept 400; Ruskin). Cells from 10 ml of an overnight culture were harvested by centrifugation at 6,000 rpm for 8 min, and the obtained cell pellet was used for DNA extraction using the GenElute bacterial genomic DNA kit (Sigma-Aldrich) following the manufacturer's guidelines. Internal transcribed spacer (ITS) sequences were amplified from extracted DNA using primer pair Probio-bif_Uni/Probio-bif_Rev (36) and sequenced to avoid decoding of clonal strains.

Genome sequencing and assemblies. The genome sequences of 46 *B. pseudolongum* strains were determined by GenProbio srl (Parma, Italy) using a MiSeq platform (Illumina, UK). A genome library was generated using the TruSeq Nano DNA kit following a specified protocol (part no. 15041110 rev. D). The generated library samples were then loaded into a 600-cycle flow cell version 3 (Illumina). The remaining five *B. pseudolongum* strains were sequenced on the Illumina NextSeq platform with NextSeq V2 reagents using a Nextera XT 150-bp paired-end library preparation kit (catalog no. FC-131-1096), as per the manufacturer's instructions. Fastq files of paired-end reads obtained from each individual genome sequencing effort were used as input for the genome assembly through the MEGAnnotator pipeline (37). The SPAdes program (version 3.12.0) was used for *de novo* assembly of each bifidobacterial genome sequence with the pipeline option "–carefull" and a list of k-mer sizes of 21,33,55,77,99,127 (38). Contigs greater than 1,000 bp were then employed by MEGAnnotator for the prediction of protein-encoding open reading frames (ORFs) using Prodigal (39). Predicted ORFs were functionally annotated by means of RAPSearch2 (Reduced Alphabet based Protein similarity Search) (cutoff E value, 1×10^{-5}; minimum alignment length, 20 amino acids) performed against the NCBI nr database (40) coupled with hidden Markov model profile (HMM) searches (http://hmmer.org/) performed against the manually curated Pfam-A database (cutoff E value, 1×10^{-10}). Furthermore, tRNA genes were identified using tRNAscan-SE version 1.4 (41), while rRNA genes were detected using RNAmmer version 1.2 (42). In order to ensure the consistency of the genomic analyses, *B. pseudolongum* chromosomes retrieved from public databases were reannotated using the same bioinformatics pipeline applied for the 51 *B. pseudolongum* strains isolated in the current study.

Comparative genomics. The genomes of the decoded *B. pseudolongum* strains (Table 1) were subjected to a pangenome calculation using the PGAP (43). Each predicted proteome of a given *B. pseudolongum* strain was screened for orthologues against the proteome of every collected *B. pseudolongum* strain by means of BLAST analysis (44) (cutoff E value, $<1 \times 10^{-5}$; 50% identity over at least 80% of both protein sequences). The resulting output was then clustered into protein families named as Clusters of Orthologous Groups (COGs) by means of MCL (graph theory-based Markov clustering algorithm) (45), using the gene family (GF) method. Pangenome profiles were built using an optimized algorithm incorporated in the PGAP software, based on a presence/absence matrix that included all identified COGs in the 60 analyzed genomes. Protein families that are shared among all analyzed genomes allowed us to formulate the core genome of the *B. pseudolongum* species. In addition, unique protein families encoded by the analyzed *B. pseudolongum* genomes that are not present in other genomes were also identified.

Phylogenomic analyses. The concatenated core genome sequences of each *B. pseudolongum* strain were aligned using MAFFT software (46), and corresponding phylogenetic trees were constructed using the neighbor-joining method in ClustalW version 2.1 (47). Accordingly, a *B. pseudolongum* supertree was built using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). For each genome pair, an ANI value was calculated using the program JSpecies, version 1.2.1 (28).

Genomic analyses. Carbohydrate-active enzymes predicted to be encoded by each of the 60 *B. pseudolongum* genomes were identified based on similarity to genes incorporated within the

Carbohydrate-Active enZyme (CAZy) database (48). For this purpose, we used GH data of 12,039 bacterial genomes available in the CAZy database, retrieving functional annotation by means of RAPSearch2 (cutoff E value, 1×10^{-30}). The prediction of genes encoding restriction enzymes was performed by means of the REBASE database (49). Identification of clustered regularly interspaced short palindromic repeats (CRISPRs) was achieved through CRISPRfinder (50), while related Cas-encoding genes were manually evaluated in each *B. pseudolongum* genome. Identification of genes that are predicted to be acquired by horizontal gene transfer (HGT) events was performed using COLOMBO version 4.0 (51). Prediction of prophage sequences was evaluated by means of a custom database composed of bifidobacterial genes previously classified as prophage genes (29). Additional analyses were performed to detect similarity between CRISPR spacers and prophage sequences retrieved within *B. pseudolongum* genomes. Functional classification of the *B. pseudolongum* proteome was performed on the basis of the eggNOG database (21) by means of RAPSearch2 (cutoff E value, 1×10^{-30}).

Carbohydrate growth assays. *B. pseudolongum* growth on semisynthetic MRS medium supplemented with 1% (wt/vol) of a specific sugar was monitored by optical density at 600 nm using a plate reader (BioTek, Winooski, VT, USA). The plate reader was run for 24 h and 48 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. SPSS software (IBM, Italy) was used to perform statistical analysis between strains of the *B. pseudolongum* subsp. *pseudolongum* group and *B. pseudolongum* subsp. *globosum* group by Student's *t* test.

Data availability. Fifty-one newly isolated *B. pseudolongum* genomes were sequenced and deposited at DDBJ/ENA/GenBank under the accession numbers reported in Table 1 (BioProject no. PRJNA510800), together representing a collection of 60 *B. pseudolongum* genomes.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.03065-18.

SUPPLEMENTAL FILE 1, PDF file, 0.5 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

We thank GenProbio srl for financial support of the Laboratory of Probiogenomics. This work was funded by the EU Joint Programming Initiative "A healthy diet for a healthy life" (JPI HDHL; http://www.healthydietforhealthylife.eu/) to D.V.S. (in conjunction with Science Foundation Ireland [SFI], grant 15/JP-HDHL/3280) and to M.V. (in conjunction with MIUR, Italy). D.V.S. is a member of APC Microbiome Ireland funded by the Science Foundation Ireland (SFI), through the Irish Government's National Development Plan (grant SFI/12/RC/2273).

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

We declare no conflicts of interest.

REFERENCES

1. Kostic AD, Howitt MR, Garrett WS. 2013. Exploring host-microbiota interactions in animal models and humans. Genes Dev 27:701–718. https://doi.org/10.1101/gad.212522.112.
2. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The human microbiome project. Nature 449:804–810. https://doi.org/10.1038/nature06244.
3. Donovan SM. 2017. Introduction to the special focus issue on the impact of diet on gut microbiota composition and function and future opportunities for nutritional modulation of the gut microbiome to improve human health. Gut Microbes 8:75–81. https://doi.org/10.1080/19490976.2017.1299309.
4. Milani C, Duranti S, Bottacini F, Casey E, Turroni F, Mahony J, Belzer C, Delgado Palacio S, Arboleya Montes S, Mancabelli L, Lugli GA, Rodriguez JM, Bode L, de Vos W, Gueimonde M, Margolles A, van Sinderen D, Ventura M. 2017. 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. https://doi.org/10.1128/MMBR.00036-17.
5. Lugli GA, Mangifesta M, Duranti S, Anzalone R, Milani C, Mancabelli L, Alessandri G, Turroni 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 41:173–183. https://doi.org/10.1016/j.syapm.2018.01.002.
6. Lugli GA, Milani C, Duranti S, Mancabelli L, Mangifesta M, Turroni F, Viappiani A, van Sinderen D, Ventura M. 2018. Tracking the taxonomy of the genus Bifidobacterium based on a phylogenomic approach. Appl Environ Microbiol 84:e02249-17. https://doi.org/10.1128/AEM.02249-17.
7. Duranti S, Mangifesta M, Lugli GA, Turroni F, Anzalone R, Milani C, Mancabelli L, Ossiprandi MC, Ventura M. 2017. Bifidobacterium vansinderenii sp. nov., isolated from faeces of emperor tamarin (Saguinus imperator). Int J Syst Evol Microbiol 67:3987–3995. https://doi.org/10.1099/ijsem.0.002243.
8. Modesto M, Michelini S, Oki K, Biavati B, Watanabe K, Mattarelli P. 2018. 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. https://doi.org/10.1099/ijsem.0.002545.
9. Modesto M, Puglisi E, Bonetti A, Michelini S, Spiezio C, Sandri C, Sgorbati B, Morelli L, Mattarelli P. 2018. 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. https://doi.org/10.1016/j.syapm.2018.07.005.
10. Modesto M, Michelini S, Sansosti MC, De Filippo C, Cavalieri D, Qvirist L, Andlid T, Spiezio C, Sandri C, Pascarelli S, Sgorbati B, Mattarelli P. 2018. Bifidobacterium callitrichidarum sp. nov. from the faeces of the emperor tamarin (Saguinus imperator). Int J Syst Evol Microbiol 68:141–148. https://doi.org/10.1099/ijsem.0.002472.
11. Michelini S, Modesto M, Filippini G, Spiezio C, Sandri C, Biavati B, Pisi A, Mattarelli P. 2018. Corrigendum To "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]. Syst Appl Microbiol 41:528. https://doi.org/10.1016/j.syapm.2018.05.002.
12. Pechar R, Killer J, Salmonova H, Geigerova M, Svejstil R, Svec P, Sedlacek I, Rada V, Benada O. 2017. Bifidobacterium apri sp. nov., a thermophilic actinobacterium isolated from the digestive tract of wild pigs (Sus scrofa). Int J Syst Evol Microbiol 67:2349–2356. https://doi.org/10.1099/ijsem.0.001956.
13. Milani C, Mangifesta M, Mancabelli L, Lugli GA, James K, Duranti S, Turroni F, Ferrario C, Ossiprandi MC, van Sinderen D, Ventura M. 2017. Unveiling bifidobacterial biogeography across the mammalian branch of the tree of life. ISME J 11:2834–2847. https://doi.org/10.1038/ismej.2017.138.
14. Duranti S, Milani C, Lugli GA, Turroni 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 17:2515–2531. https://doi.org/10.1111/1462-2920.12743.
15. Duranti S, Milani C, Lugli GA, Mancabelli L, Turroni 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. https://doi.org/10.1038/srep23971.
16. O'Callaghan A, Bottacini F, O'Connell Motherway M, van Sinderen D. 2015. Pangenome analysis of Bifidobacterium longum and site-directed mutagenesis through by-pass of restriction-modification systems. BMC Genomics 16:832. https://doi.org/10.1186/s12864-015-1968-4.
17. Mitsuoka T. 1969. Comparative studies on bifidobacteria isolated from the alimentary tract of man and animals (including descriptions of bifidobacterium thermophilum nov. spec. and bifidobacterium pseudolongum nov. spec). Zentralbl Bakteriol Orig 210:52–64. (In German.)
18. Scardovi V, Trovatelli LD, Crociani F, Sgorbati B. 1969. Bifid bacteria in bovine rumen. New species of the genus Bifidobacterium: B. *globosum* n.sp. and B. ruminale n.sp. Arch Mikrobiol 68:278–294. https://doi.org/10.1007/BF00409919.
19. Yaeshima T, Fujisawa T, Mitsuoka T. 1992. *Bifidobacterium globosum*, Subjective Synonym of *Bifidobacterium pseudolongum*, and Description of *Bifidobacterium pseudolongum* subsp. *pseudolongum* comb. nov. and *Bifidobacterium pseudolongum* subsp. *globosum* comb. nov. Syst Appl Microbiol 15:380–385. https://doi.org/10.1016/S0723-2020(11)80211-0.
20. Lugli GA, Milani C, Turroni 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. https://doi.org/10.1128/AEM.02004-14.
21. Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, Gabaldon T, Rattei T, Creevey C, Kuhn M, Jensen LJ, von Mering C, Bork P. 2014. eggNOG v4.0: nested orthology inference across 3686 organisms. Nucleic Acids Res 42:D231–D239. https://doi.org/10.1093/nar/gkt1253.
22. Milani C, Lugli GA, Duranti S, Turroni F, Bottacini F, Mangifesta M, Sanchez B, Viappiani A, Mancabelli L, Taminiau 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. https://doi.org/10.1128/AEM.02308-14.
23. Milani C, Andrea Lugli G, Durati S, Turroni 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. https://doi.org/10.1038/srep15782.
24. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome." Proc Natl Acad Sci U S A 102:13950–13955. https://doi.org/10.1073/pnas.0506758102.
25. Milani C, Duranti S, Lugli GA, Bottacini F, Strati F, Arioli S, Foroni E, Turroni F, van Sinderen D, Ventura M. 2013. Comparative genomics of Bifidobacterium animalis subsp. lactis reveals a strict monophyletic bifidobacterial taxon. Appl Environ Microbiol 79:4304–4315. https://doi.org/10.1128/AEM.00984-13.
26. Bottacini F, Morrissey R, Esteban-Torres M, James K, van Breen J, Dikareva E, Egan M, Lambert J, van Limpt K, Knol J, O'Connell Motherway M, van Sinderen D. 2018. Comparative genomics and genotype-phenotype associations in Bifidobacterium breve. Sci Rep 8:10633. https://doi.org/10.1038/s41598-018-28919-4.
27. Lugli GA, Milani C, Turroni F, Duranti S, Mancabelli L, Mangifesta M, Ferrario C, Modesto M, Mattarelli P, Jiří K, van Sinderen D, Ventura M. 2017. Comparative genomic and phylogenomic analyses of the Bifidobacteriaceae family. BMC Genomics 18:568. https://doi.org/10.1186/s12864-017-3955-4.
28. 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. https://doi.org/10.1073/pnas.0906412106.
29. Lugli GA, Milani C, Turroni F, Tremblay D, Ferrario C, Mancabelli L, Duranti S, Ward DV, Ossiprandi MC, Moineau S, van Sinderen D, Ventura M. 2016. Prophages of the genus Bifidobacterium as modulating agents of the infant gut microbiota. Environ Microbiol 18:2196–2213. https://doi.org/10.1111/1462-2920.13154.
30. Botstein D. 1980. A theory of modular evolution for bacteriophages. Ann N Y Acad Sci 354:484–490. https://doi.org/10.1111/j.1749-6632.1980.tb27987.x.
31. Briner AE, Lugli GA, Milani C, Duranti S, Turroni F, Gueimonde M, Margolles A, van Sinderen D, Ventura M, Barrangou R. 2015. Occurrence and diversity of CRISPR-Cas systems in the genus Bifidobacterium. PLoS One 10:e0133661. https://doi.org/10.1371/journal.pone.0133661.
32. Chylinski K, Makarova KS, Charpentier E, Koonin EV. 2014. Classification and evolution of type II CRISPR-Cas systems. Nucleic Acids Res 42:6091–6105. https://doi.org/10.1093/nar/gku241.
33. Ershova AS, Rusinov IS, Spirin SA, Karyagina AS, Alexeevski AV. 2015. Role of restriction-modification systems in prokaryotic evolution and ecology. Biochemistry (Mosc) 80:1373–1386. https://doi.org/10.1134/S0006297915100193.
34. Bottacini F, Morrissey R, Roberts RJ, James K, van Breen J, Egan M, Lambert J, van Limpt K, Knol J, Motherway MO, van Sinderen D. 2018. Comparative genome and methylome analysis reveals restriction/modification system diversity in the gut commensal Bifidobacterium breve. Nucleic Acids Res 46:1860–1877. https://doi.org/10.1093/nar/gkx1289.
35. Orskov ER. 1986. Starch digestion and utilization in ruminants. J Anim Sci 63:1624–1633.
36. Milani C, Lugli GA, Turroni 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. https://doi.org/10.1111/1574-6941.12410.
37. 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:fnw049. https://doi.org/10.1093/femsle/fnw049.
38. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
39. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
40. Zhao Y, Tang H, Ye Y. 2012. RAPSearch2: a fast and memory-efficient

protein similarity search tool for next-generation sequencing data. Bioinformatics 28:125–126. https://doi.org/10.1093/bioinformatics/btr595.
41. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964. https://doi.org/10.1093/nar/25.5.955.
42. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108. https://doi.org/10.1093/nar/gkm160.
43. Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J. 2012. PGAP: pan-genomes analysis pipeline. Bioinformatics 28:416–418. https://doi.org/10.1093/bioinformatics/btr655.
44. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
45. Vlietstra WJ, Zielman R, van Dongen RM, Schultes EA, Wiesman F, Vos R, van Mulligen EM, Kors JA. 2017. Automated extraction of potential migraine biomarkers using a semantic graph. J Biomed Inform 71:178–189. https://doi.org/10.1016/j.jbi.2017.05.018.
46. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. https://doi.org/10.1093/nar/gkf436.
47. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31:3497–3500. https://doi.org/10.1093/nar/gkg500.
48. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178.
49. Roberts RJ, Vincze T, Posfai J, Macelis D. 2015. REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 43:D298–D299. https://doi.org/10.1093/nar/gku1046.
50. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 35:W52–W57. https://doi.org/10.1093/nar/gkm360.
51. Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF, Surovcik K, Meinicke P, Merkl R. 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC Bioinformatics 7:142. https://doi.org/10.1186/1471-2105-7-142.