

Design of high specificity and sensitive sequencing primers to detect the bifidobacterial population in biological samples

Xiaoye Bai (✉ baixiaoye2012@163.com)

Inner Mongolia Agricultural University <https://orcid.org/0000-0003-0476-5492>

Lingling Shen

Inner Mongolia Agricultural University

Zhongjie Yu

Inner Mongolia Agricultural University

Xu Gao

Inner Mongolia Agricultural University

Lai-Yu Kwok

Inner Mongolia Agricultural University

Zhihong Sun

Inner Mongolia Agricultural University

Heping Zhang

Inner Mongolia Agricultural University

Research article

Keywords: Bifidobacterium, sequencing primer, rpsK gene, biological samples, microbiota

Posted Date: September 18th, 2019

DOI: <https://doi.org/10.21203/rs.2.14489/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background : Bifidobacteria were poorly represented or even totally absent in many of the currently published microbiota/metagenomics datasets. This work designed Bifidobacterium -targeting primers that were suitable for use in high-throughput sequencing to characterized the feature of Bifidobacterium .

Results : After multiple steps of in silico prediction, phylogenetic analysis, polymerase chain reaction and single-molecule real-time sequencing (SMRT) experiments, the Bif-6 primer pair was confirmed to be optimal for such purpose. Up to 99.42% and 99.32% of the detected sequences in human stool and breast milk samples represented up to 6 and 8 different Bifidobacterium species, respectively. Additionally, the Bif-6 primer pair detected the species Bifidobacterium catenulatum and Bifidobacterium kashiwanohense , which have not previously been reported in human breast milk to our knowledge. The high specificity of Bif-6 could be explained by the lower average nucleotide identity of 78.42% (versus 93.91% for the universal primers 27f/1492r). Moreover, to optimize the process of microbiota annotation, here a Bifidobacterium -specific database was constructed that largely reduced the time required for annotation compared with using the NCBI nucleotide (nt) database. The two databases generated highly consistent profiles of bifidobacteria when tested on the same samples.

Conclusion : this work designed a highly reliable primer and database specific for detecting the bifidobacterial population in complex biological samples, which will facilitate future studies of this group of bacteria.

Background

Bifidobacterium is a Gram-positive, nonmotile, non-spore-forming, catalase- and oxidase-negative, and often branched anaerobic bacterial genus [1]. Members of this genus are part of the endogenous gut microbiota and are generally considered to be health-promoting [2]. They play important roles in modulating gut mucosal physiological responses and immunity [3–5], as well as guarding against infection, inhibiting pathogenic bacteria, and alleviating inflammatory diseases [6, 7]. Therefore, *Bifidobacterium* can be regarded as gut health indicators.

The advent of sequencing technology has made it possible to comprehensively describe the microbial community in various ecosystems, but the detection of low abundant taxa remains challenging [8]. The bifidobacterial population is often undetected or underestimated due to its intrinsic low abundance and limitation in detection methodology [9–11]. Particularly, the choice of polymerase chain reaction (PCR) primers would affect the sensitivity and specificity of detection, as well as taxonomic coverage. An inadequate primer design might result in underestimation of a specific microbial population like bifidobacteria [12, 13]. Most previous studies applied universal 16S rRNA gene-specific PCR primers to detect the bifidobacterial biodiversity in the ecosystem [14, 15], although several other studies attempted to design PCR primers based on rRNA internal transcribed spacers sequences (ITS) [16, 17], some housekeeping genes and functional genes such as *HSP60*, *tuf*, and *recA* [18–20]. However, the majority of

these primers failed to distinguish between closely related bifidobacterial species [16, 21]; and most of these genes when used for primer design suffer from limitations like multicopy, non-uniform product size, and low taxonomic resolution [22, 23].

Thus, this study aimed to develop new bifidobacteria-targeting primers that were suitable for use in high-throughput sequencing. The primers were designed based on core genes universal to all known bifidobacterial (sub)species. Our work has established cost-effective tools for in-depth study of bifidobacteria in complex environments.

Results

Consensus *Bifidobacterium* sequences and deduced primers

Sixty-nine *Bifidobacterium* genome sequences were downloaded (Supplementary material 1), and seven consensus sequences were identified from these *Bifidobacterium* genomes. Four of these genes were *Bifidobacterium*-specific based on search against the NCBI nt database, from which 30 *Bifidobacterium*-specific primer pairs were deduced. Then, 17 pairs of primers were rejected due to mismatching and non-specificity after querying against the NCBI nt database. The predicted amplicon size of the remaining 13 primer pairs was 400–700 bp.

Primer specificity to *Bifidobacterium* culture

The amplification specificity of the 13 primer pairs was tested in PCR on 24 bacterial strains (Table 1). Four of the primer pairs (namely Bif-4, Bif-5, Bif-6, and Bif-11) exhibited high specificity to *Bifidobacterium*, evidenced by positive PCR on two bifidobacterial strains while absent of non-specific amplification when non-bifidobacterial strains were tested (Table S1 and Figure S1). The sequences and the optimum annealing temperature (T_m) of these four primer pairs are listed in Table 2.

Primer specificity to *Bifidobacterium* diversity in biological samples

(A) Initial screening

To select for the primer pair that was most suitable for use in targeting the bifidobacteria population in biological samples, Bif-4, Bif-5, Bif-6, and Bif-11 were tested on samples 1–3 together with the control 27f/1492r primers by SMRT sequencing. Although the number of observed OTUs did not saturate, the

Shannon diversity index reached a stable value in all cases, suggesting a representative microbial diversity had already been captured (Figure S2).

For comparison between samples, the number of sequence per sample was randomized and standardized to 1,610 sequences. The proportion of bifidobacterial sequences in each sample was calculated (Table 3). The four currently designed primer pairs showed high effectiveness in amplifying bifidobacterial sequences, representing 96.01% to 100% of total sequences (<1% using 27f/1492r primers) (Table 3). The *Bifidobacterium*-specific primers could detect between four to seven different species per sample, while the 27f/1492r primers could only detect at most two bifidobacterial species per sample (Figure 1). The distribution of the eight bifidobacterial species detected by different primer pairs is listed in Table S2. Bif-6 showed the highest numerical means in both Shannon diversity index and number of observed OTUs, though no significant differences were indeed observed in both parameters among the four tested primer pairs ($p > 0.05$) (Figure 2). Thus, Bif-6 was selected for the next round of verification.

(B) Further SMRT verification

Ten faecal and breast milk samples were sequenced with the primer pairs Bif-6 and 27f/1492r using SMRT. For comparison between samples, the number of sequence per sample was standardized to 2,024 and 3,473 sequences for the faecal and breast milk samples, respectively. Similarly, the Shannon diversity index but not the number of observed OTUs reached a stable value as sequencing depth increased, indicating most microbial diversity of each sample had already been captured although new taxa might still be found (Figure S3).

For the faecal samples, the average proportion of *Bifidobacterium* detected by the primer pairs 27f/1492r and Bif-6 were $0.02 \pm 0.03\%$ and $99.42 \pm 1.18\%$, respectively (Figure 3; Table S3). At the species level, the 27f/1492r failed to detect any bifidobacteria in most samples, except HF2, in which two bifidobacterial species were detected. In contrast, Bif-6 was able to detect three to five different bifidobacterial species per sample (Figure 3; Table S3). The prevalence and average relative abundance of bifidobacterial species detected by the Bif-6 primer pair in descending order were: *Bifidobacterium catenulatum* (prevalence: 100%, proportion: 56.88%), *Bifidobacterium longum* (prevalence: 90%, proportion: 17.11%), *Bifidobacterium adolescentis* (prevalence: 90%, proportion: 15.73%), *Bifidobacterium bifidum* (prevalence: 70%, proportion: 9.43%), *Bifidobacterium animalis* (prevalence: 60%, proportion: 0.26%), and *Bifidobacterium breve* (prevalence: 10%, proportion: 0.15%) (Table S3).

For the breast milk samples, the average proportion of *Bifidobacterium* detected by the primer pairs 27f/1492r and Bif-6 were $1.94 \pm 0.09\%$ and $99.32 \pm 0.26\%$, respectively (Figure 4; Table S4). The 27f/1492r primers failed to detect any bifidobacteria in four of the samples, while only one species was detected in each of the other four samples. On the other hand, the bifidobacterial sequences detected by the Bif-6 primer pair comprised over 99% of the total sequences in all cases (Table S4). Nine bifidobacterial

species were present across all samples as detected by using Bif-6, namely *Bifidobacterium catenulatum* (48.89%), *Bifidobacterium longum* subsp. *longum* (28.46%), *Bifidobacterium stercoris* (16.47%) and *Bifidobacterium bifidum* (3.74%), *Bifidobacterium adolescentis* (0.71%), *Bifidobacterium kashiwanohense* (0.35%), *Bifidobacterium animalis* subsp. *lactis* (0.29%), *Bifidobacterium dentium* (0.20%), and *Bifidobacterium breve* (0.20%). The species *Bifidobacterium longum* subsp. *infantis* (prevalence: 37%, proportion: 0.01%) and *Bifidobacterium longum* subsp. *suis* (prevalence: 37%, proportion: 0.01%) were only occasionally detected by the Bif-6 primer pair (Table S4).

Currently constructed *Bifidobacterium* database and its reliability

To identify the taxonomic resolution of Bif-6, here the anticipated amplicon sequences were predicted from all available *Bifidobacterium* genomes to construct a bif-6-sequence database, from which a phylogenetic tree was created. Different bifidobacterial species were distinctive in the Bif-6-sequence phylogenetic tree (Figure S4), whereas the bifidobacterial species formed a cluster in the 16S rRNA sequence-based phylogenetic tree and were not distinguishable (Figure S5). A distance of 100% identity indicated no distinction between two sequences. By pairwise comparison of sequences in the bif-6 database, 100% identity occurred only in two pairs of species, namely *Bifidobacterium gallinarum* vs *Bifidobacterium saeculare* and *Bifidobacterium kashiwanohense* vs *Bifidobacterium catenulatum* (contrasting to six pairs of species using the 16S rRNA gene) (Supplementary Material 2). Such result indicated a higher taxonomic distinction could be achieved using the bif-6 amplicon than full-length 16S rRNA sequences. Additionally, the mean ANI value of bif-6 amplicon sequences was 78.42%, contrasting to 93.91% for the 27f/1492r database (Supplementary Material 3). According to Milani et al., the cutoff level of ANI value required for distinguishing between different *Bifidobacterium* species was 98.5% [16]. Based on this ANI cutoff level, six and ten *Bifidobacterium* species could not be distinguished by using Bif-6 and 27f/1492r primer pairs, respectively (Supplementary Material 3), suggesting Bif-6 offered a higher taxonomic resolution of *Bifidobacterium*. Besides, the ten faecal and breast milk microbiota profiles generated by the Bif-6 database and the nt library were largely consistent, except for the taxonomic assignment of two species, *Bifidobacterium adolescentis* and *Bifidobacterium stercoris* (Figure 5B).

Discussion

Bifidobacterium is a bacterial genus that exists naturally in most ecosystems; however, many reports have underestimated the prevalence and biodiversity of the bifidobacterial population due to its generally low relative abundance, particularly in biological samples that contained complex microbiota [24–26]. In order to improve the detection of bifidobacteria in natural habitats, this work aimed to design PCR primers that targeted the *Bifidobacterium* genus in complex samples using high-throughput technology.

By multiple steps of *in silico* prediction, analysis, and sequencing experiments, the Bif-6 primer pair was superior over the other three tested primers pairs, as demonstrated by the apparently higher number of observed OTUs in the dataset. By using the Bif-6 primer pair, the average relative abundance of bifidobacterial sequences comprised up to 99.42% (for human faecal samples) and 99.32% (for breast milk samples) of the total sequences, contrasting to 0.15% and 1.94% using 27f/1492r that targeted to the full-length 16S rRNA gene, suggesting tremendous improvement in the effectiveness in detecting the bifidobacterial populations in these two types of samples. Some published studies designed primers that were specific to the order Bifidobacteriales and reported a proportion between 24.1% to 94.4% of specific sequences [10, 12, 14]. The high specificity toward bifidobacteria of Bif-6 could be explained by its target gene, *rpsK*, which is a single copy core gene specific to the *Bifidobacterium* genus. Another reason for the high specificity and distinction of Bif-6 toward bifidobacteria could be the large difference in the mean ANI values between amplicons of Bif-6 and 27f/1492r primers, which were 78.42% and 93.91%, respectively, as determined by phylogenetic analysis.

Apart from the specificity and relative abundance, the Bif-6 primer pair was able to detect a much wider spectrum of bifidobacterial species both in human stool and breast milk samples compared with the 27f/1492r primer pair. The Bif-6 amplicons of the faecal samples comprised predominantly with the species *Bifidobacterium adolescentis*, *Bifidobacterium catenulatum*, and *Bifidobacterium longum*, which agreed with the results published by other reports describing the biodiversity of human bifidobacteria [6, 27–29]. In addition, three other commonly reported faecal bifidobacterial species (namely *Bifidobacterium animalis*, *Bifidobacterium bifidum*, and *Bifidobacterium breve*) [16] were also detected by using the Bif-6 primer pair.

Human breast milk is another natural source of *Bifidobacterium* [30]; thus, this study tested the Bif-6 primer pair on human breast milk samples by SMRT sequencing. By using the Bif-6 primer pair, eleven different bifidobacterial species/subspecies were detected in the eight breast milk samples; nine of them were common to all tested samples, and six of them were previously reported in breast milk [31, 32]. The most prevalent bifidobacterial species detected in the current breast milk samples were *Bifidobacterium catenulatum* (48.88%), *Bifidobacterium longum* subsp. *longum* (28.46%), and *Bifidobacterium stercoris* (16.47%), while each of the other species comprised over >0.1% of the total sequences.

Chen *et al.* reported that *Bifidobacterium longum* (prevalence: 62.4%, abundance: 0.31%) was a predominant bifidobacterial species in human breast milk, whereas several reports found a low relative abundance of *Bifidobacterium* in human milk (<0.01%) [31–34], which could be due to inefficient detection methods used and difficulty in cultivation. Additionally, to our knowledge, this is the first report that observed *Bifidobacterium catenulatum* and *Bifidobacterium kashiwanohense* in human breast milk; both species were often reported in healthy infant feces [35–37]. Our data support that a rich array of bifidobacteria exists in human breast milk.

Finally, the bif-6-sequence database constructed here facilitates rapid taxonomic assignment of bifidobacteria. However, since there are over 70 species and 10 subspecies of bifidobacteria [38], the

current database (contained only 69 taxa) would need to be updated regularly subject to the availability of the *rpsK* gene sequences.

Conclusions

By combining *in silico* prediction, phylogenetic analysis, and empirical verification, we demonstrated that theBif-6 primer pair, targeting to the single copy *rpsK* gene that is unique to the *Bifidobacterium* genus, could effectively detect the bifidobacterial diversity in complex microbiota communities using the high-throughput sequencing approach. Our data support that a wide bifidobacterial diversity exists in human breast milk. The newly designed primer pair and bifidobacteria-specific database will serve as valuable tools for future in-depth studies of bifidobacteria in natural ecological niches.

Experimental Procedures

Experimental procedures

Ethics statement

This study was approved by the Ethical Committee of the Inner Mongolia Agricultural University (Hohhot, China). Written informed consent was obtained from each participant prior to the study.

Bacterial strains

Twenty-four bacterial strains, including four bifidobacteria and twenty non-bifidobacteria strains (Table 1), were used for testing primer specificity. All strains were obtained from the Lactic Acid Bacteria Culture Collection of Inner Mongolia Agricultural University (Huhhot, China).

Sample information

Three faecal samples (Samples 1, 2, and 3) were collected from healthy donors who visited the Inner Mongolia People's Hospital, and the other ten faecal samples (HF1 to HF10) were collected from healthy individuals from Hainan province of China. The breast milk samples (BM1-BM8) were donated by eight healthy mothers who visited the Inner Mongolia Health Hospital for Woman and Children, China.

DNA extraction

Bacterial genomic DNA of each strain was extracted from 2 ml overnight culture using the TIANamp Bacteria DNA Extraction Kit (Tiangen, Beijing, China). The faecal genomic DNA was extracted from 0.5 g

of sample using the QIAamp Stool Mini Kit (Qiagen, Hilden, Germany), while the breast milk DNA was extracted from 5 ml breast milk using the OMEGA Soil DNA Kit (Solarbio, Shanghai, China) according to the manufacturers' instructions. The extracted DNA was stored at -20°C until use.

Identification of consensus *Bifidobacterium* sequences and primer design

All accessible genome sequences of *Bifidobacterium* included in the List of Prokaryotic names with Standing in Nomenclature (LPSN) were downloaded from the NCBI genome database (<http://www.ncbi.nlm.nih.gov/genome/>), and the consensus sequences were obtained using in-house script. The identified consensus sequences were queried against the NCBI nucleotide (nt) library using the BLAST software to ensure that they were specific to the *Bifidobacterium* genus. Based on the results of the BLAST search, *Bifidobacterium*-specific primers were designed using the Primer 5.0 software (Premier Biosoft, Palo Alto, CA, USA). The length of each deduced primer was between 17 and 24 bp with no more than two degenerate bases. The primers were synthesized by Shanghai Sunny Biotechnology (Shanghai, China).

Determination of primer specificity

The deduced primers were firstly tested by PCR using genomic DNA of the four bifidobacterial strains as templates (Table 1). The size and specificity of generated PCR products were assessed by agarose gel electrophoresis. Primers that could amplify clean single DNA fragments of predicted sizes for all four strains were selected for a second round of PCR verification. In the second round, primer pairs were selected based on positive reactions for two selected bifidobacterial strains (*Bifidobacterium animalis* subsp. *animalis* and *Bifidobacterium longum* subsp. *infantis*) and negative reactions for the twenty non-bifidobacterial strains (Table 1). The PCR was performed in a final volume of 50 μl containing 2 μl extracted DNA (20 ng/ μL) on a thermocycler (LeikangHengtai Trading Co., Ltd., Beijing). The PCR cycling conditions were as follows: 95°C for 10 min, 30 cycles of 95°C for 30 s, annealing temperature (T_m) for 1 min, 72°C for 1 min, and 72°C for 7 min.

In the third round, the genomic DNA of three faecal samples (namely Samples 1, 2, and 3) were used as PCR templates. These three samples were known to contain a rich consortium of bifidobacteria as determined previously (data not shown). Based on the obtained results, four PCR primer pairs were further tested in single-molecule real-time (SMRT) sequencing (Table 2). Additionally, the universal primer pair that could amplify the full-length 16S rRNA gene, 27f/1492r (5'-AGAGTTTGATCCTGGCTCAG-3'/5'-CTACGGCTACCTTGTACGA-3') [39], were included as control.

The PCR for SMRT sequencing was performed in a final volume of 50 μl containing 1 μl of extracted DNA (20 ng/ μl), 25 μl $2 \times$ KAPA HiFi HotStart ReadyMix, 1.2 μL forward primer (10 μM), 1.2 μl reverse primer

(10 μ M), ddH₂O to a total volume of 50 ml [40]. The PCR primers were barcoded to distinguish between samples during SMRT sequencing. The reagents were purchased from Pacific Biosciences (MenloPark, CA, United States). The amplification program was as follows: 95°C for 10 min, 30 cycles of 95°C for 30 s, annealing temperature (T_m) for 1 min, 72°C for 1 min, and 72°C for 7 min. The DNA library construction and PacBio SMRT II sequencing (MenloPark, CA, United States) were performed with commercial kits purchased from Pacific Biosciences (MenloPark, CA, United States) according to the manufacturer's instructions.

For comparison between samples, the number of OTUs sequence of each sample was normalized independently and randomly basing on the sample having the fewest sequence count. Then, the sequencing coverage was estimated by two alpha diversity indexes, namely OTUs rarefaction and Shannon diversity index.

Construction of a bifidobacteria-specific database

Since Bif-6 outperformed all other tested primer pairs, sixty-nine bifidobacteria (sub)species genome sequences (Supplementary Material 1) were downloaded from the NCBI website (<http://www.ncbi.nlm.nih.gov/genome/>), from which sequences between the forward and reverse Bif-6 primers were extracted using in-house script (Supplementary Material 4). Then these sequences were queried against the nt library to verify the constructed bifidobacteria-specific database. Moreover, the *Bifidobacterium* sequences between the Bif-6 primer pair were aligned to construct a phylogenetic tree and calculate the distance between any two strains using the Mega 6.0 software [41, 42]. The phylogenetic tree was visualized with the FigTree software (version 1.4.3). To further determine how well the constructed database could reflect the phylogenetic relationship among closely related bifidobacteria, the average nucleotide identity (ANI) values were calculated based on pairwise alignment of the extracted bifidobacterial sequences (Supplementary Material 4) using in-house script. The same procedures were performed on the bifidobacterial 16S rRNA gene sequences for comparison. Data were plotted in a heatmap.

Verification of the *Bifidobacterium*-specific database

Ten faecal samples (HF1 to HF10) and eight breast milk samples (BM1 to BM8) were sequenced with the primer pairs Bif-6 and 27f/1492r using SMRT. The taxonomic assignment of the high-quality sequences was based on the currently constructed *Bifidobacterium* Bif-6 sequence database (Supplementary Material 4) and NCBI nucleotide (nt) database using in-house script.

Bioinformatic analyses

Raw data were generated using the protocol RS_ReadsOfinsert.1, followed by the extraction of high-quality sequences using the QIIME package. The minimum full passes, minimum predicted accuracy, minimum read lengths of inserts, and maximum read lengths were set to 5, 90, 1,400, and 1,800, respectively [43]. The barcode and primer sequences were removed prior to operational taxonomic units (OTUs) assignment using UCLUST [44, 45] with a 100% pairwise identity threshold. The taxonomic assignment of sequences generated by the 27f/1492r primer pair was based on the Greengenes (version 13.5) [45], RDP (release 11.5) [46], and Silva (version 128) [47] databases.

Statistical analyses

The relative abundances of bifidobacterial species were calculated and represented by heatmaps and pie charts. The graphic presentations were generated using the R software “pheatmap” package (R Foundation for Statistical Computing, Vienna, Austria) and Excel 2013 (Microsoft, Redmond, WA, USA), respectively. Data are presented as mean \pm standard deviation (SD). Significant differences in the Shannon diversity index and the number of observed *Bifidobacterium* species between datasets were identified by one-way ANOVA followed by Turkey’s post hoc test using SPSS statistical software (version 20) (IBM, Armonk, NY, USA). A p-value of below 0.05 was considered statistically significant.

Data deposition in public database

Raw sequences reported in this article have been deposited in the metagenomics Rapid Annotation using Subsystem Technology (under the subject number of mgp88725 and mgp86051 for the faecal and breast milk samples, respectively).

Declarations

Abbreviations

PCR: Polymerase chain reaction with agarose gel electrophoresis; SMRT: Single-molecule real-time sequencing; OUT: Operational taxonomic units; ITS: Internally transcribed spacer.

Acknowledgements

The authors would like to acknowledge Jie Zhao and Haiyan Xu for their advice on this work.

Funding

This research was supported by the National Natural Science Foundation of China (Projects: 31720103911, 31622043), China Agriculture Research System (Grant no. CARS-36), and the Inner Mongolia Autonomous Region Major Basic Research Open Project.

Authors' Contributions

HZ and ZS designed the experiments; LS, HZ, and XG performed the experiments; ZY, QH, and XB analyzed the data; XB and LYK wrote the manuscript; all authors read and approved the final manuscript.

Conflict Of Interest

The authors declare that they have no conflicts of interest.

References

1. Biavati B, Mattarelli P: The Family Bifidobacteriaceae. *The Prokaryotes* 2006:322–82.
2. Sun Z, Zhang W, Guo C, Yang X, Liu W, Wu Y, Song Y, Kwok LY, Cui Y, Menghe B *et al*: Comparative genomic analysis of 45 type strains of the genus Bifidobacterium: a snapshot of its genetic diversity and evolution. *Plos One* 2015, 10(2):e0117912.
3. Duranti S, Gaiani F, Mancabelli L, Milani C, Grandi A, Bolchi A, Santoni A, Lugli GA, Ferrario C, Mangifesta M: Elucidating the gut microbiome of ulcerative colitis: bifidobacteria as novel microbial biomarkers. *Fems Microbiology Ecology* 2016, 92(12):fw191.
4. Round JL, Mazmanian SK: The gut microbiota shapes intestinal immune responses during health and disease (vol 9, pg 313, 2009). *Nature Reviews Immunology* 2009, 9(8):600-.
5. Turrone F, Taverniti V, Ruas-Madiedo P, Duranti S, Guglielmetti S, Lugli GA, Gioiosa L, Palanza P, Margolles A, van Sinderen D *et al*: Bifidobacterium bifidum PRL2010 Modulates the Host Innate Immune Response. *Appl Environ Microb* 2014, 80(2):730–40.
6. Arbolea S, Stanton C, Ryan CA, Dempsey E, Ross PR: Bosom Buddies: The Symbiotic Relationship Between Infants and Bifidobacterium longum ssp. longum and ssp. infantis. Genetic and Probiotic Features. *Annu Rev Food Sci Technol* 2016, 7(1):1–21.
7. Guarner F, Malagelada JR: Gut flora in health and disease. *Lancet*. *Lancet* 2003, 361(9356):512–9.
8. Mande SS, Mohammed MH, Ghosh TS: Classification of metagenomic sequences: methods and challenges. *Brief Bioinform* 2012, 13(6):669–81.

9. De Vuyst L, Leroy F: Cross-feeding between bifidobacteria and butyrate-producing colon bacteria explains bifidobacterial competitiveness, butyrate production, and gas production. *Int J Food Microbiol* 2011, 149(1):73–80.
10. Zheng H, Liang H, Wang Y, Miao M, Shi T, Yang F, Liu E, Yuan W, Ji ZS, Li DK: Altered Gut Microbiota Composition Associated with Eczema in Infants. *Plos One* 2016, 11(11):e0166026.
11. Dai Z, Coker OO, Nakatsu G, Wu W, Zhao L, Chen Z, Chan F, Kristiansen K, Sung J, Wong SH: Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. *Microbiome* 2018, 6(1):70.
12. Milani C, Hevia A, Foroni E, Duranti S, Turrone F, Lugli GA, Sanchez B, Martín R, Gueimonde M, Van SD: Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. *Plos One* 2013, 8(7):e68739.
13. Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ, Scott KP: 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. *Microbiome* 2015, 3.
14. Turrone F, Peano C, Pass DA, Foroni E, Severgnini M, Claesson MJ, Kerr C, Hourihane J, Murray D, Fuligni F: Diversity of Bifidobacteria within the Infant Gut Microbiota. *Plos One* 2012, 7(5):e36957.
15. Canfora EE, van der Beek CM, Hermes GDA, Goossens GH, Jocken JWE, Holst JJ, van Eijk HM, Venema K, Smidt H, Zoetendal EG *et al*: Supplementation of Diet With Galacto-oligosaccharides Increases Bifidobacteria, but Not Insulin Sensitivity, in Obese Prediabetic Individuals(Article). *Gastroenterology* 2017(No.1):87–97(e3).
16. Milani C, Lugli GA, Turrone F, Mancabelli L, Duranti S, Viappiani A, Mangifesta M, Segata N, Van SD, Ventura M: Evaluation of bifidobacterial community composition in the human gut by means of a targeted amplicon sequencing (ITS) protocol. *Fems Microbiology Ecology* 2015, 90(2):493–503.
17. Duranti S, Lugli GA, Mancabelli L, Armanini F, Turrone F, James K, Ferretti P, Gorfer V, Ferrario C, Milani C: Maternal inheritance of bifidobacterial communities and bifidophages in infants through vertical transmission. *Microbiome* 2017, 5(1):66.
18. Ventura M, Meylan V, Zink R: Identification and tracing of Bifidobacterium species by use of enterobacterial repetitive intergenic consensus sequences. *Appl Environ Microb* 2003, 69(7):4296–301.
19. Ventura M, Zink R: Comparative sequence analysis of the tuf and recA genes and restriction fragment length polymorphism of the internal transcribed spacer region sequences supply additional tools for discriminating Bifidobacterium lactis from Bifidobacterium animalis. *Appl Environ Microb* 2003, 69(12):7517–22.

20. Jian W, Zhu L, Dong X: New approach to phylogenetic analysis of the genus *Bifidobacterium* based on partial HSP60 gene sequences. *Int J Syst Evol Microbiol* 2001, 51(Pt 5):1633–8.
21. Ventura M, Zink R: Rapid identification, differentiation, and proposed new taxonomic classification of *Bifidobacterium lactis*. *Appl Environ Microb* 2002, 68(12):6429–34.
22. Ventura M, Canchaya C, Del CA, Dellaglio F, Neviani E, Fitzgerald GF, Van SD: Analysis of bifidobacterial evolution using a multilocus approach. *International Journal of Systematic and Evolutionary Microbiology* 2006, 56(Pt 12):2783.
23. Ludwig W, Schleifer KH: Phylogeny of Bacteria beyond the 16S rRNA Standard. *Asm News* 1999, 65(11):752–7.
24. Opstelten JL, Jonathan P, Van MSWC, Emna A, Matthieu P, Siersema PD, Bas O, Cervino ACL: Gut Microbial Diversity Is Reduced in Smokers with Crohn's Disease. *Inflammatory Bowel Diseases* 2016, 22(9):2070–7.
25. Zhang Z, Li D: Thermal processing of food reduces gut microbiota diversity of the host and triggers adaptation of the microbiota: evidence from two vertebrates. *Microbiome* 2018, 6(1):99.
26. Chu DM, Antony KM, Ma J, Prince AL, Showalter L, Moller M, Aagaard KM: The early infant gut microbiome varies in association with a maternal high-fat diet. *Genome Medicine* 2016, 8(1):77.
27. Scardovi V, Crociani F: *Bifidobacterium catenulatum*, *Bifidobacterium dentium*, and *Bifidobacterium angulatum*: Three New Species and Their Deoxyribonucleic Acid Homology Relationships. *International Journal of Systematic Bacteriology* 1974, 24(2):310-.
28. Firman J, Liu L, Zhang L, Arango AG, Wang M, Tomasula P, Kobori M, Pontious S, Xiao W: The effect of quercetin on genetic expression of the commensal gut microbes *Bifidobacterium catenulatum*, *Enterococcus caccae* and *Ruminococcus gauvreauii*. *Anaerobe* 2016, 42:130–41.
29. Garrido D, Barile D, Mills DA: A molecular basis for bifidobacterial enrichment in the infant gastrointestinal tract. *Advances in Nutrition* 2012, 3(3):415S.
30. Soto A, Martin V, Jimenez E, Mader I, Rodriguez JM, Fernandez L: Lactobacilli and Bifidobacteria in Human Breast Milk: Influence of Antibiotherapy and Other Host and Clinical Factors. *J Pediatr Gastr Nutr* 2014, 59(1):78–88.
31. Chen PW, Lin YL, Huang MS: Profiles of commensal and opportunistic bacteria in human milk from healthy donors in Taiwan. *Journal of Food and Drug Analysis* 2018, 26(4):1235–44.
32. Murphy K, Curley D, O'Callaghan TF, O'Shea CA, Dempsey EM, O'Toole PW, Ross RP, Ryan CA, Stanton C: The Composition of Human Milk and Infant Faecal Microbiota Over the First Three Months of Life: A Pilot Study. *Scientific Reports* 2017, 7.

- 33.Chen PW, Tseng SY, Huang MS: Antibiotic Susceptibility of Commensal Bacteria from Human Milk. *Curr Microbiol* 2015, 72(2):1–7.
- 34.Cabrera-Rubio R, Mira-Pascual L, Mira A, Collado MC: Impact of mode of delivery on the milk microbiota composition of healthy women. *Journal of Developmental Origins of Health & Disease* 2016, 7(1):54–60.
- 35.Morita H, Nakano A, Onoda H, Toh H, Oshima K, Takami H, Murakami M, Fukuda S, Takizawa T, Kuwahara T: Bifidobacterium kashiwanohense sp. nov., isolated from healthy infant faeces. *International Journal of Systematic and Evolutionary Microbiology* 2011, 61(11):2610–5.
- 36.Vazquez-Gutierrez P, Lacroix C, Jaeggi T, Zeder C, Zimmerman M, Chassard C: Bifidobacteria strains isolated from stools of iron deficient infants can efficiently sequester iron. *Bmc Microbiol* 2015, 15(1):3.
- 37.Bunesova V, Lacroix C, Schwab C: Fucosyllactose and L-fucose utilization of infant Bifidobacterium longum and Bifidobacterium kashiwanohense. *Bmc Microbiol* 2016, 16(1):248.
- 38.Parte AC: LPSN - List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *International Journal Of Systematic And Evolutionary Microbiology* 2018, 68(6):1825–9.
- 39.Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ: Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microb* 2008, 74(8):2461–70.
- 40.Hou Q, Bai X, Li W, Gao X, Zhang F, Sun Z, Zhang H: Design of Primers for Evaluation of Lactic Acid Bacteria Populations in Complex Biological Samples. *Frontiers in Microbiology* 2018, 9(2045).
- 41.Calixto CPG, Waugh R, Brown JWS: Evolutionary Relationships Among Barley and Arabidopsis Core Circadian Clock and Clock-Associated Genes. *Journal of Molecular Evolution* 2015, 80(2):108–19.
- 42.Kumar S, Stecher G, Tamura K: MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution* 2016, 33(7):1870–4.
- 43.Hou QC, Xu HY, Zheng Y, Xi XX, Kwok LY, Sun ZH, Zhang HP, Zhang WY: Evaluation of bacterial contamination in raw milk, ultra-high temperature milk and infant formula using single molecule, real-time sequencing technology. *Journal of Dairy Science* 2015, 98(12):8464–72.
- 44.Edgar RC: Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010, 26(19):2460.
- 45.Desantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL: Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microb* 2006, 72(7):5069–72.

46.Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM, Tiedje JM: The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Research* 2007, 35:D169-D72.

47.Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO: SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* 2007, 35(21):7188–96.

Figures

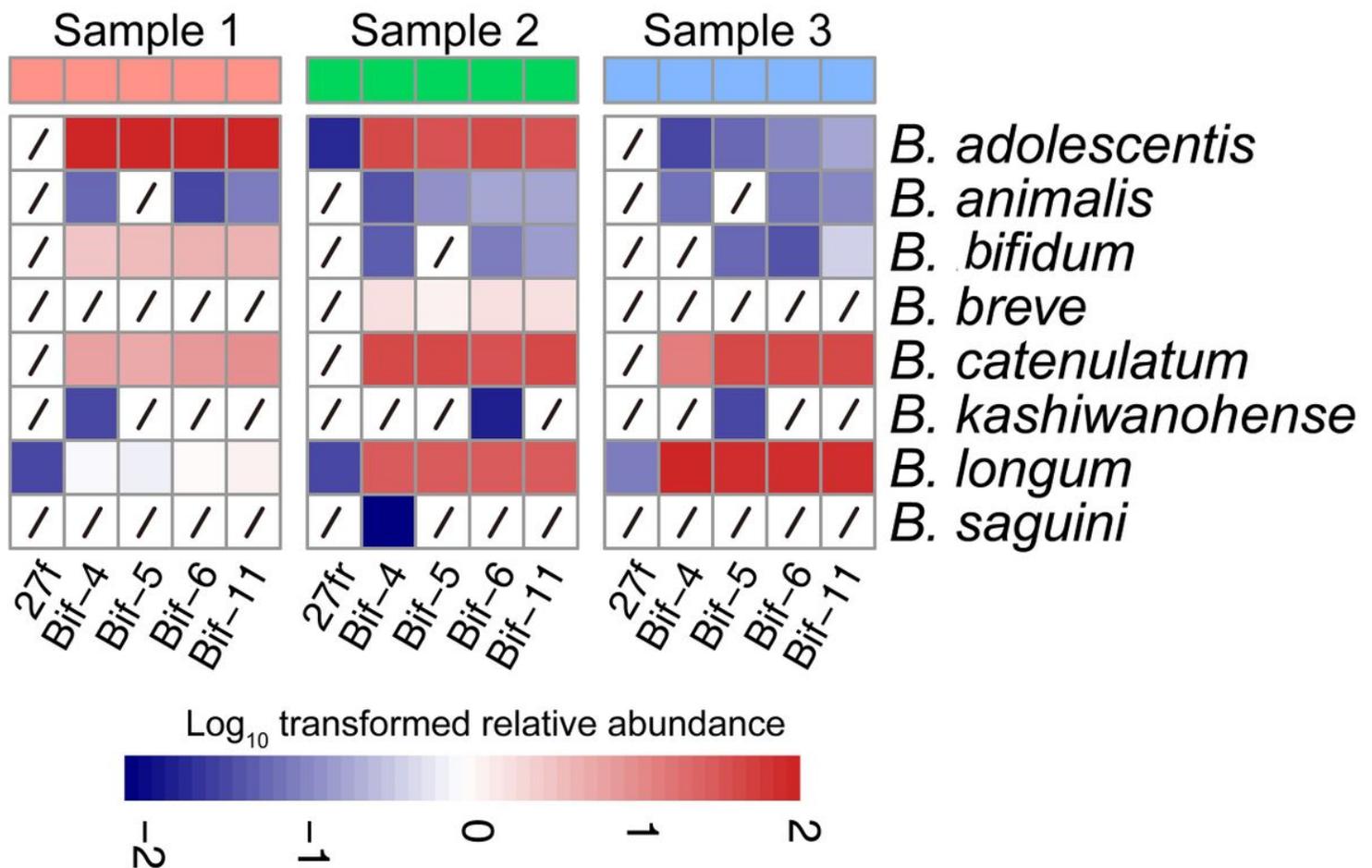


Figure 1

Bifidobacterial diversity of three faecal samples detected by five primer pairs. Primer pairs: 27f/1492r, Bif-4, Bif-5, Bif-6, and Bif-11. 'B.' represents Bifidobacterium. The heatmap represents the relative abundance of different Bifidobacterium species, as illustrated by the colour scale; a larger number represents a higher relative abundance.

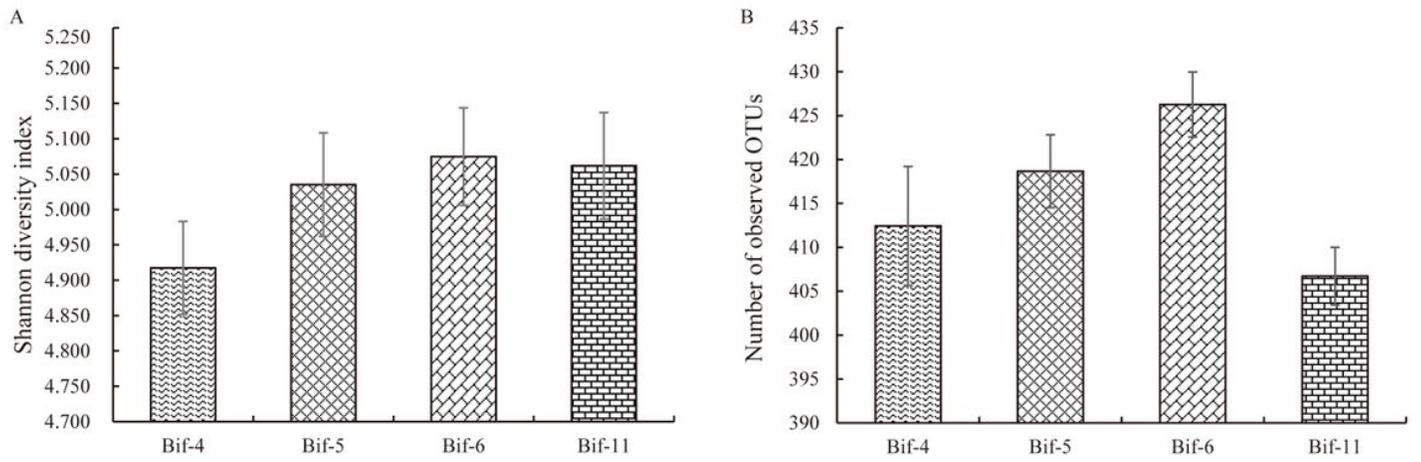


Figure 2

Shannon diversity index (A) and number of observed operational taxonomic units (OTUs) (B) of faecal sample microbiota detected by four bifidobacterial primer pairs. Primer pairs: 27f/1492r, Bif-4, Bif-5, and Bif-6. Error bars represent SD.

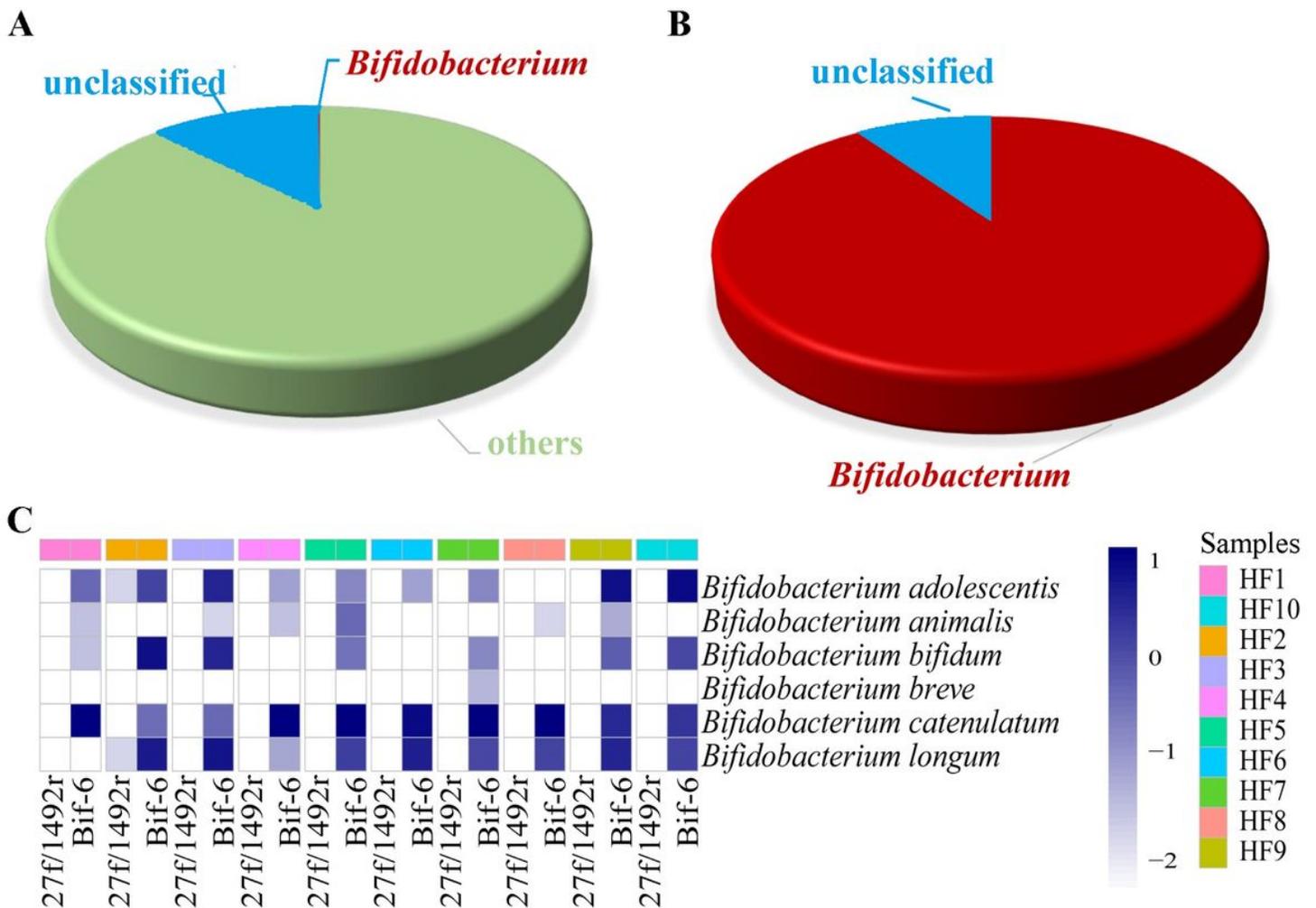


Figure 3

Bifidobacterial composition of ten samples from Hainan province. Composition of microbiota at the genus level detected by the primer pairs 27f/1492r (A) and Bif-6 (B); heatmap representing the relative abundance of different *Bifidobacterium* species (C). Samples were encoded as HF1-HF10. The colour scale represents the relative abundance; a larger number represents a higher relative abundance.

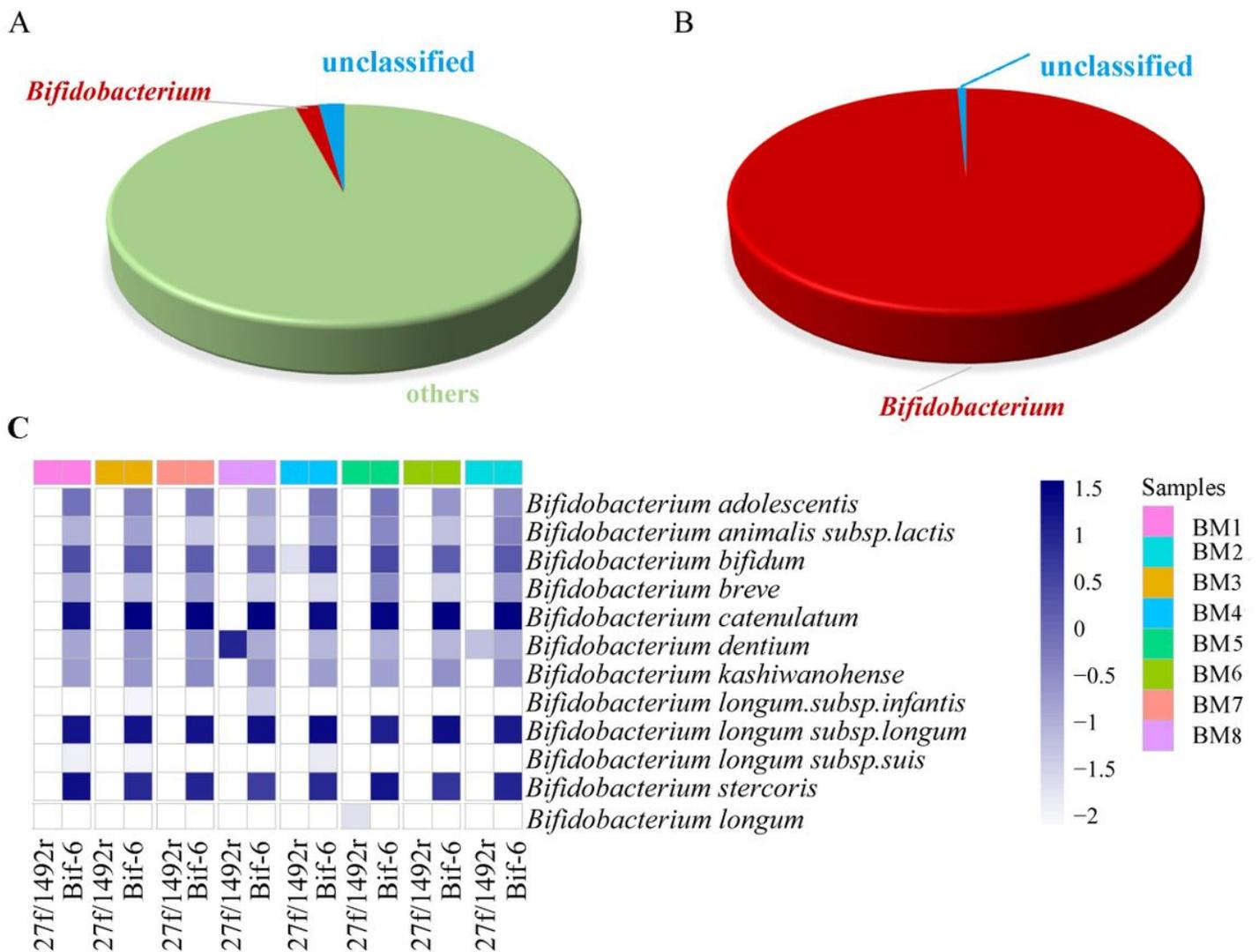


Figure 4

Bifidobacterial composition of eight breast milk samples. Composition of microbiota at the genus level detected by the primer pairs 27f/1492r (A) and Bif-6 (B); heatmap representing the relative abundance of different *Bifidobacterium* species (C). 'BM' represents breast milk. The colour scale represents the relative abundance; a large number represents a high relative abundance.

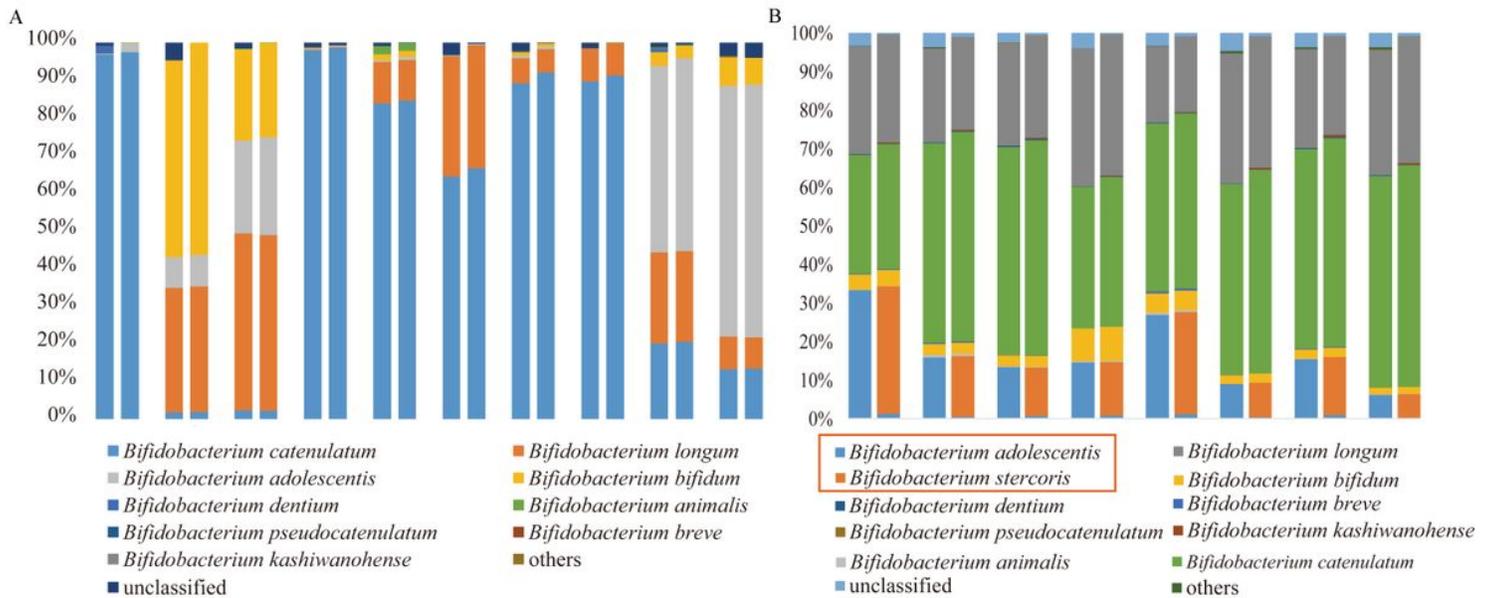


Figure 5

Annotation of bifidobacterial composition of faecal microbiota (A) and breast milk samples (B) by two databases. Bif: Bif-6-sequence database constructed in this work; nt: the NCBI nucleotide database.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [figureS3.tif](#)
- [SupplementaryMaterial1.xlsx](#)
- [SupplementaryMaterial4.txt](#)
- [Supplementarymaterial3.xls](#)
- [figureS5.pdf](#)
- [Supplementarymaterial2.xls](#)
- [figureS4.pdf](#)
- [figureS2.tif](#)
- [figureS1.tif](#)