A new Illumina MiSeq high-throughput sequencing-based method for evaluating the composition of the *Bacteroides* community in the intestine using the *rpsD* gene sequence

Chen Wang,1,2 Saisai Feng,1,2 Yue Xiao,1,2 Mingluo Pan,1,2 Jianxin Zhao,1,2 Hao Zhang,1,2,3,4,5 Qixiao Zhai1,2,6 and Wei Chen1,2,3,7

1State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi, Jiangsu 214122, China.  
2School of Food Science and Technology, Jiangnan University, Wuxi, Jiangsu 214122, China.  
3National Engineering Research Center for Functional Food, Jiangnan University, Wuxi, Jiangsu 214122, China.  
4Wuxi Translational Medicine Research Center, Jiangsu Translational Medicine, Research Institute Wuxi Branch, Wuxi, China.  
5(Yangzhou) Institute of Food Biotechnology, Jiangnan University, Yangzhou 225004, China.  
6International Joint Research Laboratory for Probiotics, Jiangnan University, Wuxi, Jiangsu 214122, China.  
7Beijing Innovation Centre of Food Nutrition and Human Health, Beijing Technology and Business University (BTBU), Beijing 100048, China.

Summary

*Bacteroides* is a bacterial genus that is known to closely interact with the host. The potential role of this genus is associated with its ecological status and distribution in the intestine. However, the current 16S V3–V4 region sequencing method can only detect the abundance of this genus, revealing a need for a novel sequencing method that can elucidate the composition of *Bacteroides* in the human gut microbiota. In this study, a core gene, *rpsD*, was selected as a template for the design of a *Bacteroides*-specific primer set. We used this primer set to develop a novel assay based on the Illumina MiSeq sequencing platform that enabled an accurate assessment of the *Bacteroides* compositions in complex samples. Known amounts of genomic DNA from 10 *Bacteroides* species were mixed with a complex sample and used to evaluate the performance and detection limit of our assay. The results were highly consistent with those of direct sequencing with a low *Bacteroides* DNA detection threshold (0.01 ng), supporting the reliability of our assay. In addition, the assay could detect all the known *Bacteroides* species within the faecal sample. In summary, we provide a sensitive and specific approach to determining the *Bacteroides* species in complex samples.

Introduction

*Bacteroides* is among the most abundant gram-negative bacterial genera in the human gut, accounting for up to 25% of the total intestinal microbiota (Ochoa-Repáraz et al., 2010). According to data from the National Center for Biotechnology Information (NCBI) and the European Molecular Biology Laboratory (EMBL), the genomic data of 42 *Bacteroides* species can be collected from NCBI and EMBL. It has been reported as a relatively complex genus and includes diverse species (Ley et al., 2008). As commensal and mutualist bacteria, *Bacteroides* species can establish stable relationships with hosts (Faith et al., 2013) and play potential probiotic roles, such as resolving diseases (Mazmanian et al., 2008; Ochoa-Repáraz et al., 2010; Hsiao et al., 2013), aiding digestion (Xu and Gordon, 2003) and enhancing immunity (Mazmanian et al., 2005). Interestingly, some of these beneficial functions are associated with the ecological status and distribution of *Bacteroides* species in the intestine (Ley et al., 2005, 2006). For example, studies have shown that a low abundance of *Bacteroides* uniformis in the intestine of a formula-fed infant is associated with a...
high risk of obesity (Owen et al., 2005; Sanchez et al., 2011). The abundance of *Bacteroides acidifaciens* may be associated with metabolic diseases, such as diabetes and obesity (Yang et al., 2017). Similarly, a significantly increased abundance of other *Bacteroides* species, such as *Bacteroides vulgatus*, was observed in subjects with type 2 diabetes (Remely et al., 2016). Given the importance of *Bacteroides* spp. in human health, ongoing research has focused on assessments of the diversity and composition of this genus in the human intestinal tract.

The development of molecular biological methods, particularly next-generation sequencing (NGS) based on the 16S rRNA gene, has enabled the thorough examination of a variety of samples. Consequently, a variety of methods based on this approach have been proposed. One previous study described a two-step multiplex PCR assay based on the 16S rRNA gene, the 16S-23S rRNA ISR and a variable region of the 23S rDNA that could be used to identify 10 *Bacteroides fragilis* group species (Liu et al., 2003). However, based on the current classification of *B. fragilis* group species, this method may be inaccurate since two species (*Bacteroides distasonis* and *Bacteroides merdae*) reported in that study have been classified into *Parabacteroides* species by a following study (Sakamoto and Benno, 2006). Another study designed 19 oligonucleotide primers based on the 16S rRNA genes and detected 14 *Bacteroides* spp. at different hierarchical levels using a method called hierarchical oligonucleotide primer extension (HOPE) (Hong et al., 2008). However, the sensitivity of the HOPE method may be affected by the excessively long poly(A) tail (Wu and Liu, 2007) and the presence of nontarget DNA in a complex environmental sample (Hong et al., 2008). The number of *Bacteroides* species (14) that can be detected by HOPE is also limited. Besides, one study has reported a gyrB-based real-time PCR system that can detect *B. fragilis* as a human-specific marker of faecal contamination (Lee and Lee, 2010). But this method lack efficacy and can only be used to identify one species of *Bacteroides* (*B. fragilis*). Consequently, an improved dual-indexing amplification and sequencing approach based on the V3–V4 region of the 16S rRNA gene and the Illumina MiSeq platform was developed to assess the composition of microbial communities in clinical samples (Fadrosh et al., 2014). However, an analysis of the V3–V4 region can only identify microbial flora at the genus level, which has limited the progress of *Bacteroides*-specific research. Therefore, there is a need to develop a simple and direct method to elucidate the relative abundances of different *Bacteroides* spp. present in complex intestinal microbial communities.

A previous study demonstrated that because each species harbours a unique DNA sequence, the choice of the correct probe sequence and use of sufficiently stringent assay conditions can enable a very selective DNA sequence-based detection method (Kreader, 1995). Recently, *Lactobacillus*-specific and *Bifidobacterium*-specific primer pairs based on a hypervariable core gene have been developed for the precise taxonomic identification and detection of intestinal *Lactobacillus* (Xie et al., 2019) and *Bifidobacterium* (Hu et al., 2017) species, respectively. Thus, the selection of a unique DNA sequence suggests a potential approach to the rapid detection and measurement of the relative abundances of other bacteria at species levels.

In this study, we developed a *Bacteroides*-specific primer pair using the core gene *rpsD*, which is present in the genomic sequences of all *Bacteroides* species, as a discriminative marker. We then assessed the precision, quantification limit, detection limit and detection efficiency of this primer pair and developed a novel method to quantify *Bacteroides* spp. in human and mouse faecal samples using high-throughput sequencing.

**Results**

Selection and phylogenetic analysis of the core gene

Thirty-one core genes were identified as alternative target genes by Roary. Only the *rpsD* gene was demonstrated to have a higher discriminating power for *Bacteroides* than the 16S rRNA V3-V4 region gene. As shown in the neighbour-joining tree based on the hypervariable sequence region within the *rpsD* gene in Figure S1, different *Bacteroides* species were placed into different clades, while the same species were placed into a single clade. In contrast, the phylogenetic tree based on the V3–V4 region gene did not allow distinctions between closely related *Bacteroides* species. As shown in the Figure S2, *B. uniformis* dntKV2 01149 and *Bacteroides fluxus* YIT12057 were resolved into one clade, as were *Bacteroides faecis* MAJ27 and *Bacteroides thetaiotaomicron* 7330. Thus, the *rpsD* gene was selected as the template because of its high resolving power for species discrimination in the *Bacteroides* genus.

Comparative analysis of the *rpsD* and 16S rRNA gene resolving power

The comparative analysis of the *rpsD* of 42 *Bacteroides* species showed that the per cent identity of the *rpsD* gene ranged from 76.38% to 100%, and the average value was 87.22%. The minimum and maximum per cent identity of the 16S rRNA gene were 84.77% and 100%, respectively, with an average value of 91.96%. This finding demonstrates that the *rpsD* gene had higher levels of taxonomic and phylogenetic resolving power at the
Bacteroides species level than the 16S rRNA gene. Therefore, the rpsD gene was used as the template for Bacteroides-specific primer design.

Design of the species-specific primer
The full-length rpsD gene is approximately 606-bp long, which is unsuitable for a short-read sequencing platform. Therefore, a partial Bacteroides rpsD gene was used for sequencing and designing a novel primer set as per the primer design criteria. Based on the results of a multiple sequence alignment, the 24–508-bp region of the rpsD gene was targeted for PCR amplification. All of these rpsD genes have been deposited in NCBI database (GenBank accession number: MT152002–MT152101). A potential Bacteroides group-specific primer pair, Bif-rpsD-F (5′-AWCDAGAATHGCMCGTAA-3′)/Bif-rpsD-R (5′-YRTCCCAYTCCAACCA-3′), was selected and manually designed based on the hypervariable sequence region within the rpsD gene using MEGA 5 and Primer Premier 6.0. The total volume per reaction was 50 μl, and each reaction contained 1 μl of DNA template, 25 μl of 2×Taq PCR MasterMix (Sangon, Shanghai, China), 1 μl of each primer at 10 μM (Sangon, Shanghai, China) and 22 μl of double-distilled water (ddH2O). The following PCR conditions were used to amplify the 485-bp rpsD target sequence: initial denaturation at 95°C for 8 min; 30 cycles of 95°C for 40 s, 50°C for 40 s and 72°C for 40 s; and a final extension at 72°C for 8 min.

Database construction based on rpsD genes
To construct a DNA database for sequencing on the Illumina MiSeq platform, all rpsD genes from reported Bacteroides species should be considered. Finally, 520 rpsD genes from 42 different species of Bacteroides were collected from NCBI and EMBL. These genes were used for the construction of the database and the identification of amplified sequences. Notably, sequence similarity of 97% could be clustered into one operational taxonomic unit (OTU). OTUs based on the rpsD gene composition of Bacteroides were comparable with those in database.

Detection of the specificity, accuracy and sensitivity of the novel primer set
In silico PCR using PRIMER-BLAST generated only a single amplicon in the Bacteroides pan-genome. We also performed PCR using genomic DNA extracted from known bacterial species, including 14 Bacteroides strains and 8 non-Bacteroides strains. As shown in Figure 1, a PCR product was obtained only when the genomic DNA of Bacteroides species was used as the template DNA. In addition, all reads generated for the artificial sample (with known amounts of genomic DNA extracted from 10 known Bacteroides species) were compatible with the 10 known species, and a strong correlation was observed between the normalized relative abundance predicted for Bacteroides species and the relative abundance observed by an rpsD-profiling analysis (Fig. 2A). Furthermore, the lowest detectable amount of Bacteroides DNA (amplified using the designed primer pair) was 0.01 ng, corresponding to a detection limit of 10³ CFU (Fig. 2B).

Comparison of the robustness of the rpsD gene and the 16S rRNA V3-V4 region
PCR using our designed primer set generated 832 342 and 669 429 copies of the high-quality 16S rRNA gene and 528 618 and 951 729 copies of the rpsD gene from the 20 human and 20 mouse faecal samples, respectively (Table 1). In addition, approximately 18.91% and 1.56% of the reads generated from the human and mice faecal samples, respectively, could be assigned to the Bacteroides genus when using the primer set targeting the V3-V4 region gene (Fig. 3A and B). In contrast, almost all the sequences could be assigned to the Bacteroides genus when using the novel primer pair Bif-rpsD-F/Bif-rpsD-R (Fig. 3C and D). Furthermore, the primer pair designed to target the partial rpsD gene could identify Bacteroides at the species level, whereas the...
universal primer set that targeted the V3-V4 region of 16S rRNA could identify Bacteroides only at the genus level.

Discussion

The Illumina MiSeq platform provides a scalable, high-throughput and streamlined sequencing platform for analysing the community compositions of complex samples (Fadrosh et al., 2014). Based on this sequencing technology, some approaches based on a high throughput, long sequence read length and high level of accuracy, including single- (Caporaso et al., 2012) and dual- (Kozich et al., 2013) indexing methods targeting the hypervariable region of the 16S rRNA gene, have been developed and are used widely (Fadrosh et al., 2014). These approaches allow the in-depth analysis of complex samples (Tringe and Hugenholtz, 2008). However,

Figure 2. Detection accuracy and detection limit of the novel designed primer set.
A. Relationship between the normalized relative abundance predicted for Bacteroides species and the relative abundance determined by the rpsD-profiling analysis.
B. The detection limit of the novel designed primer set based on the selected partial rpsD gene sequence. CFU: colony-forming units.

Table 1. Overview of sequencing results for each sample.

| Sample ID | Sequence numbera (16S) | OTU numberb (16S) | Sequence number (rpsD) | OTU number (rpsD) | Sample ID | Sequence numbera (16S) | OTU numberb (16S) | Sequence number (rpsD) | OTU number (rpsD) |
|-----------|------------------------|------------------|------------------------|-------------------|-----------|------------------------|------------------|------------------------|-------------------|
| M-1       | 36 855                 | 4069             | 92 624                 | 3999              | Hu-1      | 29 441                 | 3464             | 11 891                 | 827               |
| M-2       | 26 708                 | 3049             | 57 356                 | 4589              | Hu-2      | 62 019                 | 5151             | 12 130                 | 484               |
| M-3       | 29 059                 | 3476             | 63 827                 | 4791              | Hu-3      | 56 297                 | 4766             | 25 825                 | 668               |
| M-4       | 34 669                 | 3727             | 70 629                 | 4744              | Hu-4      | 51 110                 | 5530             | 30 837                 | 1591              |
| M-5       | 48 116                 | 4774             | 80 354                 | 5887              | Hu-5      | 20 584                 | 2174             | 41 508                 | 2379              |
| M-6       | 35 925                 | 3616             | 49 640                 | 3685              | Hu-6      | 30 888                 | 2233             | 15 648                 | 1002              |
| M-7       | 48 782                 | 5805             | 26 034                 | 2059              | Hu-7      | 25 166                 | 2699             | 28 256                 | 1286              |
| M-8       | 27 033                 | 3123             | 29 453                 | 2664              | Hu-8      | 31 504                 | 2558             | 37 582                 | 1345              |
| M-9       | 40 273                 | 4827             | 41 948                 | 3413              | Hu-9      | 23 045                 | 2539             | 59 035                 | 2935              |
| M-10      | 45 963                 | 4850             | 21 980                 | 2055              | Hu-10     | 51 727                 | 3657             | 15 959                 | 881               |
| M-11      | 23 407                 | 2757             | 33 698                 | 2927              | Hu-11     | 43 126                 | 3344             | 29 815                 | 1200              |
| M-12      | 27 307                 | 3201             | 45 208                 | 3911              | Hu-12     | 48 765                 | 5278             | 16 268                 | 1115              |
| M-13      | 26 085                 | 3146             | 41 987                 | 3876              | Hu-13     | 35 439                 | 3516             | 19 954                 | 1124              |
| M-14      | 33 210                 | 3596             | 43 417                 | 3510              | Hu-14     | 60 166                 | 4545             | 21 169                 | 1292              |
| M-15      | 26 735                 | 3171             | 46 157                 | 3755              | Hu-15     | 40 807                 | 2814             | 37 444                 | 1534              |
| M-16      | 25 529                 | 2902             | 55 517                 | 4307              | Hu-16     | 45 431                 | 5279             | 42 246                 | 1920              |
| M-17      | 38 377                 | 4488             | 44 553                 | 3697              | Hu-17     | 32 437                 | 4289             | 15 167                 | 852               |
| M-18      | 24 024                 | 3024             | 37 192                 | 3273              | Hu-18     | 49 466                 | 4178             | 13 247                 | 800               |
| M-19      | 28 566                 | 3091             | 27 451                 | 2550              | Hu-19     | 42 768                 | 4020             | 37 892                 | 1666              |
| M-20      | 42 806                 | 5028             | 42 704                 | 3424              | Hu-20     | 52 156                 | 4446             | 16 745                 | 1115              |

a. The sequence number refers to the count of assembled sequences after quality filtering.
b. The OTU (Operational Taxonomic Units) number is presented for all sequences without rarefaction. M, mice sample; Hu, human sample.
the targeted region of the 16S rRNA gene has a limited resolving power at the *Bacteroides* species level. Therefore, a novel molecular marker with a high resolving power is needed to identify *Bacteroides* species. Notably, previous studies have reported that an appropriate target gene for species-specific primers should meet the following criteria: (i) the target gene region should be prevalent in the genus at a high resolving power; (ii) the target gene region should encompass a hypervariable region and two constant regions at both ends; and (iii) the PCR amplification region in the target gene should not be longer than 500 bp (Dieffenbach *et al.*, 1993; Hu *et al.*, 2017; Xie *et al.*, 2019).

In this study, the core gene *rpsD*, which encodes the 30S ribosomal protein S4 and exists in all *Bacteroides* species, was identified by Roary. The *rpsD* gene was reported to exist in *Bacillus subtilis* and is co-transcribed with the genes for initiation factor 1 and ribosomal proteins B, S13, S11 and L17 (Boylan *et al.*, 1989). A previous study also revealed that this gene is monocistronic (Grundy and Henkin, 1990). In our study, we observed that the partial sequences of both ends of the 485-bp (< 500-bp) *rpsD* gene (24–508 bp) were highly conserved, whereas the other sequences were more variable. These results suggest that the *rpsD* gene fulfills all prerequisites and should be considered as a reliable alternative phylogenetic marker for *Bacteroides* species.

Based on the partial *rpsD* gene, we designed a pair of *Bacteroides*-specific primers, *Bif-rpsD*-F/*Bif-rpsD*-R,
which produce a 485-bp (< 500-bp) amplicon and enable the rapid discrimination of all known Bacteroides species from non-Bacteroides species. We then developed a novel method based on the illumina MiSeq sequencing platform that allowed an accurate assessment of the Bacteroides composition in complex samples. A complex sample containing 10 genomic DNA samples was evaluated, and the results were highly consistent with those of direct sequencing, thus supporting the reliability of our assay. Moreover, the minimum Bacteroides DNA detection threshold was 0.01 ng, indicating that the designed primer set possessed a higher sensitivity than those previously reported for Bifidobacterium-specific primers (0.05 ng) (Hu et al., 2017) and Lactobacillus-specific primers (0.05 ng) (Xie et al., 2019).

Compared with previously reported methods that could only detect 10–14 B. fragilis group species (Liu et al., 2003; Hong et al., 2008), our newly developed rpsD-based sequencing method was more 'broad-spectrum,' as it could identify 29 Bacteroides species in human and mouse faecal samples. Besides, only one pair of Bacteroides-specific primers were used in the present method, which improved the convenience of the assay. However, our method did not detect Bacteroides ihuae and Bacteroides timonensis in the 20 Chinese human faecal samples. These two Bacteroides species have been reported only in the sputum of healthy Frenchwomen living in Marseille (Fonkou et al., 2017) and in the faecal sample of a 21-year-old French Caucasian woman with severe anorexia nervosa (Ramasamy et al., 2014), respectively, indicating that they may not exist in the faecal samples of the Chinese studies in our study. Notably, B. acidificiens and Bacteroides caecimuris were detected in our human faecal samples, although to date, these species have been detected only in mouse samples (Miyamoto and Itoh, 2000; Lagkouvardos et al., 2016). In addition, some Bacteroides species, such as Bacteroides barnesiae, Bacteroides gallinarum, B. salanitronis and Bacteroides paurosacharolyticus, were not detected in our human and mouse samples, consistent with the findings of previous studies that identified the first three species in chicken caecal samples (Lan et al., 2006) and the latter in rice-straw residue from a methanogenic reactor that treated waste from cattle farms (Ueki et al., 2011). However, our method has some limitations. One possible drawback of the rpsD gene is the high level of similarity between some Bacteroides species, as observed between Bacteroides dorei DSM 17855 and B. vulgatus ATCC 8482 (99.5%); Bacteroides cellulosityllicus DSM 14838 and B. timonensis AP1 (99.34%); and B. faecis MAJ27 and B. thetaiotaomcron VPI-5482 (99.5%). Similar low levels of discriminatory power have also been reported for the groEL gene in Bifidobacterial species (Hu et al., 2017) and Lactobacillus species (Xie et al., 2019). Therefore, caution should be exercised when using the rpsD gene as a marker for the identification of certain Bacteroides species. In addition, the complete genomic information of some Bacteroides species, such as Bacteroides kribii and Bacteroides koreensis, is not available in the current database. Consequently, these species would be labelled as unassigned Bacteroides when annotated in the database. Regular updates of the Bacteroides database are warranted to enable the identification of all Bacteroides species.

In this study, it is theoretically possible to apply the Bacteroides species composition from the novel primers to break down the overall relative abundance of the Bacteroides genus as assessed by the V3-V4 data. We have tried to verify the feasibility of this method. The result showed that B. fragilis and B. vulgatus account for 1.33% and 5.05% of the total gut microbial population in the 20 human faecal samples, respectively. Interestingly, one previous study (Ruseler-van Embden and Both-Patoir, 1983) has reported that B. vulgatus accounts for 6% of the total gut microbiota of healthy humans. Besides, the abundance of B. fragilis accounts for only up to 1% of the total gut microbial population (Rocha and Smith, 2013). These results reinforced that our method is feasible. Notably, caution should be exercised when apply the primers. More well-conducted experiments are needed to propose the efficiency and specificity of the novel primers to assess the species composition of the Bacteroides genus as assessed by the V3-V4 data.

In summary, we developed a powerful Illumina MiSeq sequencing platform-based method for the accurate, sensitive and rapid identification of different Bacteroides species. This method enabled the determination of the relative abundances of different Bacteroides species at a low detection limit of 10^3 CFU ml\(^{-1}\). It also yielded a high resolving power for discriminating between Bacteroides species from complex samples such as human and mouse faeces. This method can enable the elucidation of Bacteroides diversity in different ecological systems, as well as the potential roles of different Bacteroides species in host health.

**Experimental procedures**

**Bacterial strains, culture media and DNA extraction**

All bacterial strains were obtained from the Culture Collection of Food Microorganisms of Jiangnan University (Wuxi, China) and cultured at 37 °C in an anaerobic workstation (N\(_2\), 85%; H\(_2\), 10%; CO\(_2\), 5%) in different culture media (Table 2). Genomic DNA was extracted...
from these bacteria using the TIANamp Bacteria DNA Kit (TianGen, Beijing, China) as per the manufacturer’s instructions.

Faecal sample collection and genomic DNA extraction

The faecal samples of 20 mice (male C57BL/6 mice, 6 weeks old) and humans (20 healthy people from China) were collected rapidly after defecation in faecal collection tubes under aseptic conditions and stored at −80°C until genomic DNA extraction. Genomic DNA was extracted from these samples using the method described in the Fast DNA SPIN Kit for Feces (MP Biomedicals; Carlsbad, CA, USA), with the following modifications: 0.1 g faeces sample was used to extract genomic DNA extraction and 60 μl DNA eluent were added in the clean catch tube to collect purified DNA (this step was performed twice to increase the DNA concentration of collected samples).

Selection and phylogenetic analysis of the core gene

As shown in the Table S1, one hundred genomic sequences from 42 Bacteroides species were collected from NCBI and EMBL. The pan-genome of the Bacteroides genus was identified using Roary software, and the selected core gene data were aligned using the CLUSTAL_X program (Thompson et al., 1997). A tree of the homologous genes was then constructed and used to select the hypervariable sequence region that would allow the precise taxonomic identification and detection of all Bacteroides spp.

Design of Bacteroides-specific primers

The selected partial homologous gene sequences from the pan-genome of Bacteroides species were amplified by PCR using the barcoded fusion primers (341F/806R) designed in this study. Meanwhile, the PCR conditions for the region covered by this primer pair and for the 16S rRNA V3-V4 regions were set as described in a previous report (Jia et al., 2016). The major PCR products of the selected (i.e. primer-covered) region and V3-V4 region gene sequences were electrophoresed on a 2.0% agarose gel in TBE buffer, stained with SYBR SAFE (Invitrogen, Eugene, OR, USA), purified and quantified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) as per the manufacturer’s instructions. A further quantification step was performed using the Quant-IT PicoGreen dsDNA Assay Kit (Life Technologies, Carlsbad, CA, USA). The selected gene regions from all known Bacteroides species, based on NCBI and EMBL, were used to construct a DNA amplicon sequence library. This library was then sequenced on the Illumina MiSeq platform as described in a previous study (Hu et al., 2017).

Table 2. Bacteria were used in this study.

| Number | Species | Isolation source | Strain | Medium and reference |
|--------|---------|------------------|--------|---------------------|
| 1      | Bilidobacterium longum | Human faecal | FGDZ58M1 | MRS (with 1% L(+)-Cysteine) |
| 2      | Lactobacillus brevis | Human faecal | X3 | |
| 3      | Lactobacillus plantarum | Human faecal | Z2 | |
| 4      | Lactobacillus fermentum | Human faecal | B7 | |
| 5      | Pediococcus acidilactici | Human faecal | B18 | TSB broth |
| 6      | Enterococcus faecalis | Human faecal | CCFM596 | BHI (with 1% L(+)-Cysteine) |
| 7      | Escherichia coli | Human faecal | CCFM21 | BHI (with 1% L(+)-Cysteine) |
| 8      | Akkermansia muciniphila | Human faecal | ATCC BAA-835 | Anaerobic medium (gastric mucin as the sole carbon and nitrogen source) (Derrien et al., 2004) |
| 9      | B. caccae | Human faecal | FSDTA-ELH-2.5MIC-3 | Modified BHI (Tan et al., 2019) |
| 10     | B. dorei | Human faecal | FSDTA-HCK-B-6 | |
| 11     | B. eggerthii | Human faecal | FSDTA-HCK-B-9 | |
| 12     | B. faecis | Human faecal | FNMH1BE10K3 | |
| 13     | B. salyersiae | Human faecal | FSDTA-ELI-BHI-9 | |
| 14     | B. uniformis | Human faecal | FSDTA-HCK-B1 | |
| 15     | B. ovatus | Human faecal | FSDTA-HCK-B4 | |
| 16     | B. fragilis | Human faecal | FSDTA-HCK-B8 | |
| 17     | B. vulgatus | Human faecal | FSDTA-HCM-XY-14 | |
| 18     | B. stercolis | Human faecal | FFJLY21K3 | |
| 19     | B. xylanisolvens | Human faecal | FFJLY22K22 | |
| 20     | B. thetaiotaomicron | Human faecal | FGSZY48K9 | |
| 21     | B. kribii | Human faecal | FOHXXNK2 | |
| 22     | B. koreensis | Human faecal | FYNLJ19K1 | |

B., Bacteroides; MRS, de Man, Rogosa and Sharp broth; LB, Luria-Bertani broth; BHI, Brian Heart Infusion broth; TSB: Tryptic Soy Broth.
Detection of primer specificity

Four tests were performed to determine the specificities of the primer set. (i) The tree generated from the V3-V4 region of the 16S rRNA gene and the tree of selected genes constructed from the alignment of Bacteroides sequences were used to compare the efficacy of the novel primer set via MEGA 5.02 (Tamura et al., 2011). (ii) In silico PCR was performed using PRIMER-BLAST, with the NCBI nonredundant database as the template (Ye et al., 2012). (iii) The genomic DNA of 14 Bacteroides strains and 8 non-Bacteroides strains (Table 2) were extracted and PCR amplified using the novel primer set. (iv) The genomic DNA extracted from 20 healthy human and 20 mouse faecal samples were PCR amplified using the novel primer set and the V3-V4 region gene primer set. These sequence data have been submitted to the GenBank databases under accession number SRR11212985 (Fig. 3A); SRR11213067 (Fig. 3B); SRR11213107 (Fig. 3C); SRR11213137 (Fig. 3D). Microbiota analyses were performed using 16S rRNA gene amplicon sequencing and the Quantitative Insights Into Microbial Ecology (QIIME) version 1 software.

Evaluation of primer sensitivity and detection limit

Known amounts (0.001–50 ng) of genomic DNA extracted from 10 known Bacteroides species were mixed to be an artificial sample and used to evaluate the detection sensitivity and detection limit of the designed primer set. The genomic DNA served as the template for PCR amplification with this primer set, and the obtained amplicons were then sequenced on the Illumina MiSeq sequencing platform. These sequence data have been submitted to the GenBank databases under accession number SRR11212976. Among these amplicons, the lowest detectable concentration of gene copies and the colony-forming units (CFU) of the corresponding strain was identified as the PCR detection limit. The copy numbers of the selected gene and the CFU of the corresponding strain were estimated using a dsDNA copy number calculator (Calculator for Determining the Number of Copies of a Template, URI Genomics & Sequencing Center).

Statistical analysis

Data analysis was performed using GRAPHPAD PRISM 8. The data corresponding to each treatment were reported as the mean ± standard error of the mean (SEM). The statistical significance of the data was determined at the P < 0.05 level. Mean values were subjected to an analysis of variance, and statistically significant results were compared using Tukey’s test.

Author contributions

CW, JXZ, HZ, QXZ and WC conceptualized and designed the study; YX, SSF and MLP organized the database; CW performed the statistical analysis and wrote the first draft of the manuscript; SSF wrote sections of the manuscript. All authors contributed to the manuscript revision process and read and approved the submitted version. QXZ takes primary responsibility for communication with the journal and editorial office during the submission process, throughout the peer review and during publication. All authors read and approved the final manuscript.

Conflict of interest

None declared.

Ethical approval

The human participants and animal experiments in this study were approved by the Ethics Committee in Jiangnan University, China. All the faecal samples from healthy persons were for public health purposes and these were the only human materials used in present study. Written informed consent for the use of their faecal samples was obtained from the participants or their legal guardians. No human experiments were involved. The collection of faecal sample had no risk of predictable harm or discomfort to the participants.

References

Boylan, S.A., Suh, J.W., Thomas, S.M., and Price, C.W. (1989) Gene encoding the alpha core subunit of Bacillus subtilis RNA polymerase is cotranscribed with the genes for initiation factor 1 and ribosomal proteins B, S13, Sll and L17. J Bacteriol 171: 2553–2562.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6: 1621–1624.

Derrien, M., Vaughan, E.E., Plugge, C.M., and de Vos, W.M. (2004) Akkermansia muciniphila gen. nov. sp. nov. a human intestinal mucin-degrading bacterium. Int J Syst Evol Microbiol 54: 1469–1476.

Dieffenbach, C.W., Lowe, T.M.J., and Dveksler, G.S. (1993) General concepts for PCR primer design. PCR Methods Appl 3: S30–S37.

Fadrosh, D.W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R.M., and Ravel, J. (2014) An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome 2: 6.

Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., et al. (2013) The long-
term stability of the human gut microbiota. Science 341: 1237439.

Fonkou, M.D.M., Morand, A., Fournier, P.E., Raoult, D., and Dubourg, G. (2017) ‘Bacteroides ihuae’, sp. nov. a new bacterial species isolated from the human respiratory microbiome. New Microbes New Infect 17: 52–53.

Grundy, F.J., and Henkin, T.M. (1990) Cloning and analysis of the Bacillus subtilis rpsD gene, encoding ribosomal protein S4. J Bacteriol 172: 6372–6379.

Hong, P.Y., Wu, J.H., and Liu, W.T. (2008) Relative abundance of Bacteroides spp. in stools and wastewaters as determined by hierarchical oligonucleotide primer extension. Appl Environ Microbiol 74: 2882–2893.

Hsiao, E.Y., McBride, S.W., Hsien, S., Sharon, G., Hyde, E.R., McCue, T., et al. (2013) Microbiota modulate behavioral and physiological abnormalities associated with Neurodevelopmental Disorders. Cell 155: 1451–1463.

Hu, L.J., Lu, W.W., Wang, L.L., Pan, M.L., Zhang, H., Zhao, J.X., and Chen, W. (2017) Assessment of Blidobacterium species using groEL gene on the basis of Illumina MiSeq High-Throughput sequencing. Genes 8: 336.

Jia, H.R., Geng, L.L., Li, Y.H., Wang, Q., Diao, Q.Y., Zhou, T., and Dai, P.L. (2016) The effects of Bt Cry1le toxin on bacterial diversity in the midgut of Apis mellifera ligustica (Hymenoptera: Apidae). Sci Rep 6: 24664.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing microbial community structure at high taxonomic resolution. ISME J 7: 1494–1501.

Kreader, C.A. (1995) Design and evaluation of Bacteroides DNA probes for the specific detection of human fecal pollution. Appl Environ Microbiol 61: 1171–1179.

Lagkouvardos, I., Pukall, R., Abt, B., Foesel, B.U., Meier-Kolthoff, J.P., Kumar, N., et al. (2016) The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. Nat Microbiol 1: 1–15.

Lan, P.T.N., Sakamoto, M., Sakata, S., and Benno, Y. (2006) Bacteroides barnesiae sp. nov. Bacteroides salantrios sp. nov. and Bacteroides gallinarum sp. nov. isolated from chicken caecum. Int J Syst Evol Microbiol 56: 2853–2859.

Lee, C.S., and Lee, J. (2010) Evaluation of new gyrB-based real-time PCR system for the detection of B. fragilis as an indicator of human-specific fecal contamination. J Microbiol Methods 82: 311–318.

Ley, R.E., Bäckhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., and Gordon, J.I. (2005) Obesity alters gut microbial ecology. Proc Natl Acad Sci USA 102: 11070–11075.

Ley, R.E., Turnbaugh, P.J., Klein, S., and Gordon, J.I. (2008) Microbial ecology: human gut microbes associated with obesity. Nature 444: 1022–1023.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P., Ramey, R.R., Bircher, J.S., et al. (2008) Evolution of mammals and their gut microbes. Science 320: 1647–1651.

Li, C.X., Song, Y.L., Mcgeague, M., Vu, A.W., Wexler, H., and Finegold, S.M. (2003) Rapid identification of the species of the Bacteroides fragilis group by multiplex PCR assays using group- and species-specific primers. FEMS Microbiol Lett 222: 9–16.

Mazmanian, S.K., Liu, C.H., Tzianabos, A.O., and Kasper, D.L. (2005) An immunomodulatory molecule of symbiotic Bacteria directs maturation of the host immune system. Cell 122: 107–118.

Mazmanian, S.K., Round, J.L., and Kasper, D.L. (2008) A microbial symbiosis factor prevents intestinal inflammatory disease. Nature 453: 620–625.

 Miyamoto, Y., and Itoh, K. (2000) Bacteroides acidifaciens sp. nov. isolated from the caecum of mice. Int J Syst Evol Microbiol 50: 145–148.

Ochoa-Reparaz, J., Melcarz, D.W., Wang, Y., Begum-Haque, S., Dasgupta, S., Kasper, D.L., and Kasper, L.H. (2010) A polysaccharide from the human commensal Bacteroides fragilis protects against CNS demyelinating disease. Mucosal Immunol 3: 487–495.

Owen, C.G., Martin, R.M., Whincup, P.H., Smith, G.D., and Cook, D.G. (2005) Effect of infant feeding on the risk of obesity across the life course: a quantitative review of published evidence. Pediatrics 115: 1367–1377.

Ramasamy, D., Lagier, J.C., Rossi-Tamisier, M., Pfleiderer, A., Michelle, C., Couderc, C., et al. (2014) Genome sequence and description of Bacteroides timonensis sp. nov. Stand Genomic Sci 9: 1181–1197.

Remely, M., Hippe, B., Zanner, J., Aumueuler, E., Brath, H., and Hasberger, A.G. (2016) Gut microbiota of obese, type 2 diabetic individuals is enriched in Faecalibacterium prausnitzii, Akkermansia muciniphila and Peptostreptococcus anaerobius after weight loss. Endocr Metab Immune Disord Drug Targets 16: 99–106.

Rocha, E.R., and Smith, C.J. (2013) Ferritin-like family proteins in the anaerobe Bacteroides fragilis: when an oxygen storm is coming, take your iron to the shelter. Biometals 26: 577–591.

Ruseler-van Embden, J.G., and Both-Patoir, H.C. (1983) Anaerobic gram-negative faecal flora in patients with Crohn’s Disease and healthy subjects. Antonie Van Leeuwenhoek 49: 125–132.

Sakamoto, M., and Benno, Y. (2006) Reclassification of Bacteroides distasonis, Bacteroides goldsteinii and Bacteroides merdae as Parabacteroides distasonis gen. nov., comb. nov., Parabacteroides goldsteinii comb. nov. and Parabacteroides merdae comb. nov. Int J Syst Evol Microbiol 56: 1599–1605.

Sanchez, E., De Palma, G., Capilla, A., Nova, E., Pozo, T., Castillejo, G., et al. (2011) Influence of environmental and genetic factors linked to celiac disease risk on infant gut colonization by Bacteroides species. Appl Environ Microbiol 77: 5316–5323.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGAS: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.

Tan, H.Z., Zhao, J.X., Zhang, H., Zhai, Q.X., and Chen, W. (2019) Novel strains of Bacteroides fragilis and Bacteroides ovatus alleviate the LPS-induced inflammation in mice. Appl Microbiol Biotechnol 103: 2353–2365.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997) The CLUSTAL_X windows
interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882.

Tringe, S.G., and Hugenholtz, P. (2008) A renaissance for the pioneering 16S rRNA gene. *Curr Opin Microbiol* **11**: 442–446.

Ueki, A., Abe, K., Ohtaki, Y., Kaku, N., Watanabe, K., and Ueki, K. (2011) *Bacteroides paurosaccharolyticus* sp. nov. isolated from a methanogenic reactor treating waste from cattle farms. *Int J Syst Evol Microbiol* **61**: 448–453.

Wu, J.H., and Liu, W.T. (2007) Quantitative multiplexing analysis of PCR-amplified ribosomal RNA genes by hierarchical oligonucleotide primer extension reaction. *Nucleic Acids Res* **35**: e82.

Xie, M.Q., Pan, M.L., Jiang, Y., Liu, X.M., Lu, W.W., Zhao, J.X., *et al*. (2019) *groEL* gene-based phylogenetic analysis of *Lactobacillus* species by High-Throughput Sequencing. *Genes* **10**: 530.

Xu, J., and Gordon, J.I. (2003) Inaugural article by a recently elected academy member. *Proc Natl Acad Sci USA* **100**: 10452.

Yang, J.Y., Lee, Y.S., Kim, Y., Lee, S.H., Ryu, S., Fukuda, S., *et al*. (2017) Gut commensal *Bacteroides acidifaciens* prevents obesity and improves insulin sensitivity in mice. *Mucosal Immunol* **10**: 104–116.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012) Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**: 134.

**Supporting information**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Phylogenetic tree derived from a neighbour-joining analysis of the *rpsD* gene region sequences. Note: Bootstrap values (%) based on 1000 replications are presented on each node. The bar indicates 2% sequence divergence.

Fig. S2. Phylogenetic tree derived from a neighbour-joining analysis of the 16S V3-V4 gene region sequences. Note: Bootstrap values (%) based on 1000 replications are presented on each node. The bar indicates 2% sequence divergence.

Table S1. Basic genomic information on *Bacteroides* species used for comparative analysis in this study.