Quantitative and Qualitative Characterization of Plant Endo-bacteriome by Plant Dna-free Sequencing Method

Lili Zhang (✉ zhangll@im.ac.cn)

Institute of Microbiology Chinese Academy of Sciences <https://orcid.org/0000-0002-4321-0521>

Liying Chen

Institute of Microbiology Chinese Academy of Sciences

Mengting Zhang

Institute of Microbiology CAS

Da Liu

Institute of Microbiology CAS

Hongbo Sun

Institute of Microbiology CAS

Jianxiang Wu

Institute of Biotechnology Zhejiang University

Yan Huo

Institute of Microbiology CAS

Xiaoying Chen

Institute of Microbiology CAS

Rongxiang Fang (✉ fangrx@im.ac.cn)

Institute of Microbiology CAS

Research

Keywords: Plant endo-bacteriome, plant DNA-free amplification, 16S rRNA gene, next generation sequencing, absolute quantification of plant endophytic bacteria

Posted Date: June 9th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-33863/v1>

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

[Read Full License](#)

Abstract

Background: Plant endophytic bacteria colonize plants' internal tissues and interact with plants more closely than do epiphytic or environmental bacteria. However, the community structure of such endophytic bacteria could not be efficiently deciphered, and the microbiota abundance could not be quantified through absolute quantification. Application of 16S rRNA gene sequencing to characterize the plant endophytic community is greatly hindered by the high sequence identities among bacterial 16S rRNA, plant mitochondrial 18S rRNA and chloroplast 16S rRNA genes. This makes it difficult to identify bacterial sequences among total DNAs extracted from plant material.

Results: We designed PCR primer sets that are able to specifically amplify bacterial DNAs, even when there are very few bacteria colonizing the plant material. We computationally and experimentally evaluated the specificity, coverage, and accuracy of the newly designed primer sets (322F/796R and 799F/1107R) and two widely used primer sets (338F/806R and 799F/1193R). When applied to a same planting-soil community through next generation sequencing (NGS), the four different primer sets revealed similar high-abundant taxa composition with variation in taxa abundance. Primer sets amplifying the same 16S variable regions generated more comparable sequencing results. When applied to a rice endo-bacteriome, both 799F/1107R and modified 322F-Dr/796Rs (Primer pair 322F/796R with a penultimate-base substitution in 322F) produced plant DNA-free bacterial amplicon libraries. Primer 322F-A/796R was then used through NGS to decipher the rice endo-bacteriomes. The rice root and leaf endo-bacteriomes shared 66.36% OTU identity but enriched different bacterial species. Within the same host genotype and soil type, the root endo-bacteriome was more stable than the leaf endo-bacteriome across individual plants. 322F-A/796R was used through absolute quantitative PCR to quantitate the population size of leaf or root endophytes, which revealed 10^6 – 10^7 and 10^9 – 10^{10} bacteria per gram fresh weight, respectively.

Conclusions: This is the first study to develop plant DNA-free bacterial 16S amplification methods. The newly designed primer sets combined with NGS deciphered the rice endo-bacteriome structure, and absolute quantitative PCR quantitated the size of the endobacterial population. The protocols developed here are suitable for various plants, will significantly advance studies on plant endo-bacteriomes.

Background

Endophytes are microbial symbionts that live inside the plant for the majority of their life cycle without any detrimental effects on their host [1–3]. Endophytes are generally isolated from plant materials after surface sterilization. Inside plant tissues, endophytes mainly colonize the intercellular spaces [4]. A few endophytes, such as *Rhizobium*, live inside the plant cells. Compared with bacteria living on plants' surfaces and in the environment, endophytes that live inside plant tissues are less affected by environmental conditions. Therefore, when endophytes carry a genetic character that benefits the plant, the beneficial effects can be directly and efficiently exerted on the host.

To obtain the full structure of the endophytic bacterial community (hereafter, the “endo-bacteriome”) is important for study of the endo-bacteriome function under a clear background. Two methods have been used to determine bacterial composition and abundance in endo-bacteriomes: culture methods and non-culture methods. However, because many plant endophytic bacteria are not culturable for isolation and quantification, culture methods cannot give an accurate overview of plants’ endo-bacteriomes. High-throughput sequencing and bioinformatics tools have the potential to reveal the full structure of plant endo-bacteriomes. Next-generation sequencing (NGS) of the 16S rRNA gene (16S) has been widely used to decipher microbiomes associated with humans and other animals, soil, plants, and various environmental niches [5–11]. By amplifying all 16S sequences from a sample and sequencing the acquired 16S amplicon library, the complete bacterial composition and the relative abundance of each bacterium can be determined [12–15]. However, this non-culture method is still not suitable for analyses of the plant endo-bacteriome because of the high sequence identity among bacterial 16S, plant Mt18S and Ct16S sequences. In a few previous studies, bacterial cells have been enriched from plant tissues by washing and vortexing [16, 17] or the host DNA has been filtered out from the sequencing results [12, 18–21]. However, the former approach inevitably leads to an incomplete collection of endophytes, and the latter results in limited sequencing depth and low coverage of bacterial communities. For example, when the primer pair S-D-Bact-0341-b-S-17/S-D-Bact-0785-a-A-21 was used to amplify rice endophytes, only 310 tags from a stem sample and 1,771 tags from a shoot sample were acquired [18]. The design of plant organellar DNA-free and bacterial 16S-specific PCR primers with high bacterial coverage is one of the keys to resolve this problem.

The bacterial 16S contains nine hypervariable regions (V_1 – V_9) along with highly conserved sequences between these regions. The overall identities among bacterial 16S, plant Mt18S and Ct16S exceed 70% and are even higher in the conserved regions. The commonly used PCR primers, for example, primers 564F/785R [6, 22] and 515F/806R [12, 23] that amplify the 16S V_4 region, 27F/338R [24, 25] amplifies the V_1 – V_2 , 341F/785R [21] and 338F/806R [26] amplify the V_3 – V_4 regions, and 515F/907R [27] amplifies the V_4 – V_5 regions, have the primer sequences 100% identity to the plant Mt18S or Ct16S. As a result, PCR amplification with these primers resulted in co-amplification of high-proportion plant DNAs. The most commonly used primer pair for plant-associated microbiota is 799F/1193R that amplifies the 16S V_5 – V_7 regions [15, 19, 28, 29]. This primer pair is efficient in producing Ct16S-free amplicons and excludes Mt18S via removal of the ~ 350 bp-longer amplicons produced from most plants [15, 28, 30]. However, this primer pair is not suitable for rice endophytes because there is only a ~ 50-bp difference in length within this amplification region between rice Mt18S and bacterial 16S. Moreover, when the endophytes are much less abundant than the plant mitochondria, the bacterial 16S cannot be efficiently amplified [31, 32]. At present, it is still not possible to efficiently access the rice endophytes by using the plant whole-genomic DNA as PCR templates.

In this study, using rice as a model, we designed new PCR primer sets to specifically amplify bacterial 16S sequences from total DNA extracted from rice, thus providing clean bacterial amplicon libraries for next-generation sequencing. Even when only a few bacteria were present in a plant material, the bacterial

community composition could be determined. Using these newly designed primer sets, we determined the composition of rice leaf and root endo-bacteriomes. The rice leaf and root endo-bacteriomes shared more than 60% identity at the operational taxonomic unit (OTU) level, but their structures were very different. Compared with the leaf endo-bacteriome, the root endo-bacteriome was more stable among individual plants of the same genotype growing in the same environment. We quantitated the endobacteria of rice leaf and root. In greenhouse-grown rice plants, the size of the endophyte population in the root was 10^9 – 10^{10} bacteria per gram fresh weight, which was much larger than that in the leaf (10^6 – 10^7 bacteria per gram fresh weight).

Methods

Sampling of planting soil and rice tissues

Soil from a rice field in Shunyi, Beijing, China, was collected in September 2017 using shovels to gather 50-ml samples at a depth of approximately 10 cm below the ground. The sampling sites of the three soil samples were 3 m apart. The soil samples were immediately placed in liquid nitrogen and then transported to the lab for storage at -80°C until DNA was extracted within one week. *Oryza sativa* spp. *Japonica* cv. Nipponbare (referred to as 'Nipponbare' in this work) was used as the model plant in this work. Nipponbare plants grown in greenhouse for five weeks were collected for endo-bacteriome sequencing. The attached soil was removed from the plants by washing with ddH₂O three times. Leaf or root samples from each rice plant were placed into a 50-ml Falcon tube, washed in 75% ethanol for 5 min and then in 0.1% Tween20 (mixed in ddH₂O) on a shaking platform for 5 min at 180 rpm, followed by washing four times in ddH₂O. The obtained leaves or roots were used as leaf- or root-endosphere fractions for analysis of the microbial commensals.

Scanning electron microscopy

We conducted scanning electron microscope (SEM) analyses to determine whether bacteria were still present on the surface of rice root and leaf samples after surface sterilization. The surface-sterilized roots and leaves were fixed in 4% paraformaldehyde at 4 °C for 3–7 days and then washed with ddH₂O three times (for 6, 7, and 8 min, respectively). The fixed samples were dehydrated in a graded ethanol series (50%, 70%, 85%, 95%, and 100%; 15 min immersion at each concentration). The dehydration in 100% ethanol was repeated three times. After dehydration, the specimens were critical-point dried with liquid CO₂ (Leica EM CPD300; Leica, Hanau, Germany) and sputter coated with gold-palladium (E-1045 ion sputter; Hitachi, Tokyo, Japan). The samples were viewed under a Quanta200 SEM (FEI, Hillsboro, OR, USA).

DNA extraction from planting soil or plant tissues

Planting soil DNA was extracted using Power Lyzer™ PowerSoil® DNA Isolation Kit (MoBio) (MoBio Laboratories, Carlsbad, CA, USA) in accordance with the protocols provided by the manufacturer. The leaf and root samples were ground to a fine powder in liquid nitrogen and then 50 mg of the ground sample was added to a 2.0 ml centrifuge tube. The samples were suspended in 567 µl of 1×TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and the suspensions were thoroughly mixed by vortexing. Plant and bacterial cells in the mixture were disrupted by incubation with 30 µl of 10% SDS, 20 µl of proteinase K (100 µg/ml AMRESCO Solon USA) and 2 µl RNase (100 mg/ml; Tiangen Biotechnology Co., Ltd., Beijing, China) for 1 h at 37°C, followed by incubation with 100 µl of 5 M NaCl and 80 µl of CTAB/NaCl solution (700 mM NaCl and 10% CTAB in ddH₂O) for 10 min at 37°C. Genomic DNA was purified by phenol:chloroform:isoamyl alcohol extraction and chloroform:isoamyl alcohol extraction. DNA was precipitated by adding 0.7 volumes of isopropanol and washed with 70% ethanol. After drying, the DNA pellet was dissolved in ddH₂O.

Design of plant DNA-free bacterial 16S amplification primers

The bacterial 16S sequences were obtained from the EZBioCloud database. The bacterial strains were searched for in the “Taxonomy” (<https://www.ezbiocloud.net/taxonomy>) column of the Dashboard of Prokaryotic Diversity. In total, 33 16S sequences representing 33 species from four phyla, Proteobacteria, Actinobacteria, Bacteroidetes, and Firmicutes, were obtained. The Ct16S and Mt18S of several plants, including *O. sativa*, *Z. mays*, and *A. thaliana*, were obtained from NCBI databases. Multiple sequence alignment of the 33 bacterial 16S and the Ct16S and Mt18S was performed using the CLUSTALW program. After alignment, the bacterial 16S-specific and plant organellar DNA-free PCR primers were designed in accordance with the following primer design criteria:

- The primer sequence is in the highly conserved regions of bacterial 16S;
- The primer length is between 17 and 25 bp, and the annealing temperature is between 50°C and 60°C. Annealing temperatures were calculated with the Oligo calculator;
- The 3' end of each primer should be an exact match to the bacterial 16S but mismatch to the plant Mt18S or Ct16S;
- Fewer than two mismatches between a primer and its bacterial 16S template, but more than three mismatches between a primer and the Mt18S or Ct16S template.
- No significant sequence homology between the designed primer and the plant Mt18S or Ct16S at positions beyond the designed regions.

After primer sequences had been designed, the forward and the reverse primers were respectively named in accordance with the position of the first and the last nucleotide in the 16S rRNA gene of *Escherichia coli* DSM 30083 (GenBank accession number: JMST01000030). The primer 799F is not designed in this study, and its naming method is not the same.

Design of plant DNA-free bacterial 16S amplification primer pairs suitable for the Illumina next-generation sequencing platform

The primers designed as described above were paired according to amplicon length, specificity for domain bacteria, overall coverage and capacity for being free of plant organellar DNA.

- Amplicon length. According to the Illumina next-generation sequencing platform, the amplicon length is less than 500 bp;
- Specificity for domain Bacteria. The primer pairs should be specific to the domain Bacteria;
- Overall coverage. At two-mismatch stringency, the overall coverage for the domain Bacteria should more than 75%;
- Plant organellar DNA-free amplification. For one-step PCR, the primer pairs should avoid both Mt18S and Ct16S co-amplification. For nested PCR, the primer pair for the first round of PCR should avoid Mt18S or Ct16S co-amplification, while that for the second round of PCR should avoid Ct16S or Mt18S.

After the suitable primer pairs had been established, the amplification efficiency was assayed by PCR using plant whole-genomic DNA (containing bacterial DNA) as a template.

In silico evaluation of primers and primer pairs

Primer evaluation was based on the nonredundant SILVA SSU (small subunit rRNA gene) reference database (release 132) (referred to here as 'SSU r132') containing 675,860 SSU sequences, including 575,268 Bacteria, 75,716 Eukaryota and 24,876 Archaea sequences. For SSU Ref, bacterial and eukaryotic sequences must be at least 1,200 bases in length, while archaeal sequences must be at least 900. The Ref NR (non-redundant) dataset is created by clustering at 99% (up to SILVA 108) and 98% (SILVA 111) sequence identity using UCLUST [33]. For each cluster, only the longest sequence is kept. Sequences from cultivated species including type strains and multiple operons are preserved in all cases to serve as an anchor for taxonomy. It was recommended that the resulting SSU Ref NR dataset be used as the standard SILVA reference dataset.

SILVA Probe Match and Evaluation Tool, TestProbe 3.0 (<https://www.arb-silva.de/search/testprobe/>), was used to detect all occurrences of a given primer sequence in the SILVA databases. The maximum number of mismatches should be set as zero, one or two. Primer pairs were evaluated using TestPrimer 1.0 (<https://www.arb-silva.de/search/testprime/>) by running an *in silico* PCR using the 'ARB PT server' [34] on the SILVA databases, the same as described above. The stringency for the primer pair was set at three levels: zero, one or two mismatches. For the 3' end of the primer, a perfect match was required.

For each primer or primer pair and stringency level, the database entries were sorted into three groups: match, mismatch, and no_data. The last group contained all sequences for which no clear decision could be made. Typically, this is the case when the sequence in question does not cover the primer match position. The coverage values are then computed for each taxonomic unit by dividing the number of matched sequences in that group by the number of matched or mismatched sequences.

16S rRNA gene amplification

Platinum Taq DNA polymerase (Invitrogen, USA), which has no proofreading function (no 3'-5' exonuclease activity), was used in all of the 16S rDNA amplification experiments to decrease the probability of DNA elongation from an incorrect base pair located at the 3' end of the primer. Primer pairs used for 16S V_3-V_4 amplification included 322F/796R, 322F-A/796R, 322F-C/796R, 322F-T/796R, and 322F-H/796R. Primer pair 799F/1107R was used for 16S V_5-V_6 amplification. PCR reaction mixtures were made in a volume of 50 μ l containing 200 μ M dNTPs, 0.2 μ M of each primer, 2U Platinum Taq DNA polymerase, 1.5 mM $MgCl_2$ and 1 $\times$ PCR buffer. The concentration of template DNA depends on the sample type, including 200 ng/50 μ l for plant leaf DNA or rice callus DNA, 30ng/ μ l for root DNA, and 10 ng/50 μ l for soil DNA. The following PCR conditions were used: initial denaturation at 94°C for 2 min, followed by 26 (soil) or 34 (rice leaf, root or callus) cycles consisting of denaturation (94 °C for 30 s), annealing (30 s, for temperatures, see Table 1) and extension (72 °C for 30 s) and a final extension step at 72 °C for 5 min. The reactions were run on a 1% agarose gel to ensure that the amplification was successful.

Illumina next-generation sequencing

The primer barcode sequences used in all high-throughput sequencing were synthesized by Majorbio. The PCR amplification was performed in our lab according to the reaction mixtures and PCR conditions described above. The annealing temperature remained the same as for the corresponding primer pair without a barcoding sequence. The amplification efficiency and specificity were analyzed by subjecting 3 μ l of PCR products to agarose gel electrophoresis (2% agarose gel, 5 V/cm). The quantity and quality of DNA were analyzed using BioTek Synergy2 (BioTek Instruments, Inc., Winooski, VT, USA). PCR products with a single band on agarose gel electrophoresis, the correct size and a concentration >1 ng/ μ l were used in the following sequencing experiments.

Purification of the PCR products and construction of the amplicon libraries were performed at Majorbio Bio-pharm Technology Co., Ltd. (Shanghai, China) (referred to as 'Majorbio' hereafter) and the amplicon libraries were sequenced using the Illumina MiSeq PE300 platform. Using the 300-bp paired-end (PE) MiSeq protocol generates 300-bp PE reads and the paired reads would be merged into a sequence according to the overlap between PE reads. The minimum overlap length is 10 bp. The obtained sequence

length would be up to 600 bp containing up to 550 bp of 16S/18S-rDNA sequence. A maximum mismatch ratio of 0.2 was allowed in the overlapping area of the spliced sequence.

Sequences with a quality score lower than 20 were discarded [35]. Short reads were merged into tags using Fast Length Adjustment of Short Reads (FLASH, v. 1.2.11) [36]. Effective tags were grouped into OTUs by clustering them based on a sequence similarity score of 97% using the Usearch program v. 7.0.1090 [37]. Single OTUs and chimeras were removed during clustering. Taxonomic classification of the representative sequence for each OTU was performed using QIIME's version (http://qiime.org/scripts/assign_taxonomy.html) of the Ribosomal Database Project's classifier (version 2.2 <http://sourceforge.net/projects/rdp-classifier/>) [38] against the SILVA database (Release 128, <http://www.arb-silva.de>) using the default parameters. All OTUs identified as belonging to chloroplasts and mitochondria were included in the dataset. The community composition of each sample was annotated at each classification level, namely, domain, phylum, class, order, family, genus, and species. OTUs of all samples of each project were normalized by the minimum number of sample sequences (Table S1) and were then used for subsequent analysis. All data analysis was performed on the free online platform, Majorbio I-Sanger Cloud Platform (www.i-sanger.com).

Quantitative PCR to determine bacterial abundance in rice tissues

Relative quantification was performed to estimate the number of bacterial endophyte cells relative to the number of rice cells. We extracted DNA from surface-sterilized plant tissues and used the primer pair 322F-A/796R to amplify the bacterial 16S sequences. The *OsActin* gene was used as an internal reference gene to determine the number of plant cells and was amplified using the primer pair OsActin-F/OsActin-R. We conducted SYBR-Green-based quantitative PCR (qPCR) according to the manufacturer's instructions (Bio-Rad, Hercules, CA, USA).

We also quantified bacterial cells in rice plant materials using a standard curve method. First, a standard curve was created as follows: total DNA was extracted from rice callus containing no or very few bacteria, and then the DNA concentration was adjusted to 100 ng/μl with ddH₂O. Fresh cultured *E. coli* DH5α cells were washed with ddH₂O and suspended in ddH₂O (final concentration, 10¹⁰ cells/ml). The bacterial solution was boiled in ddH₂O for 10 min to release DNA into the supernatant. After centrifugation at 13,000 × *g* for 1 min, the supernatant was collected. One microliter of this DNA solution corresponded to the total genomic DNA from 10⁷ *E. coli* cells. This DNA solution was serially diluted ten-fold so that 1 μl of each DNA solution corresponded to genomic DNA from 10⁶, 10⁵, 10⁴, 10³, and 10² *E. coli* cells. Each of the *E. coli* DNA dilutions in 1 μl ddH₂O was then mixed with 1 μl 100 ng/μl rice callus DNA. The six DNA dilutions were PCR-amplified using the primer pair 322F-A/796R to construct the standard curve. A standard curve was also constructed for the endogenous reference gene. The *OsActin* gene, which was amplified using the primers OsActin-F/OsActin-R, was used as the internal control to standardize the

amount of plant DNA added to the reaction. Each concentration gradient was established with an internal reference reaction and three repeated reactions. The 25- μ l SYBR-Green-based qPCR mixture comprised 12.5 μ l 2 $\times$ iQTM SYBR Green Mix, 2 μ l DNA template (1 μ l rice callus DNA + 1 μ l bacterial DNA), 0.5 μ l primer pair, and 9.5 μ l ddH₂O. After running each qPCR, the standard curve of the Log value of bacterial cell numbers and the corresponding Ct value was generated.

Next, we determined the DNA extraction loss rate. Rice callus (50–100 mg) was placed in a 2 ml sterile Eppendorf tube, and then sterilized steel balls were added to the tube to break the callus. Freshly cultured *E. coli* DH5 α cells were washed with TE buffer and then suspended in TE buffer to a final concentration of 10⁹ cells/ml. The bacterial solution was serially diluted ten-fold to obtain bacterial solutions at concentrations of 10⁸ – 10 cells/ml. Then, 250 μ l of each bacterial solution was added to the broken callus cells, and then TE buffer was added to obtain a total volume of 567 μ l. The DNA was extracted as described above and then dissolved in 50 μ l ddH₂O. The above experiments were repeated three times to obtain three sets of gradient DNA dilutions. Each DNA sample was used as one PCR template for amplification with either 322F-A/796R or OsActin-F/OsActin-R. The 25- μ l SYBR-Green-based qPCR mixture comprised 12.5 μ l 2 $\times$ iQTM SYBR Green Mix, 1 μ l DNA template, 0.5 μ l primer pair, and 10.5 μ l ddH₂O. After running the q-PCR reaction, the Ct value was substituted into the standard curve to calculate the Log value of the bacterial number, and then it was compared with the known number of added bacteria to calculate the DNA extraction loss rate.

Results

Primer design and in silico evaluation of primer coverage

To design PCR primers that specifically amplify the bacterial 16S, sequence alignment was performed using plant Mt18S and Ct16S and bacterial 16S. Because plant bacterial microbiota is mainly composed of Proteobacteria, Actinobacteria, Bacteroidetes and Firmicutes [19, 29, 39, 40], we selected 33 species from these four phyla (Fig. S1A) and aligned the corresponding 16S with Mt18S and Ct16S of several plants, including *O. sativa*, *A. thaliana*, and *Zea mays*. The primer selection criteria included the following: the 3' end should be an exact match to bacterial 16S but mismatch to Mt18S or Ct16S, with fewer than two mismatches between a primer and its bacterial 16S template. Four primers with a mismatch(es) at the 3' end to either Mt18S or Ct16S were selected, including 322F with one mismatch to Mt18S, 796R with two mismatches to Ct16S, 799F [30] with two mismatches to Ct16S and 1107R with three mismatches to Mt18S (Fig. 1A and S1B).

According to the up to 500-bp read-length of the amplicons by the Illumina Miseq PE300 platform, the designed primers could be combined into two sets of PCR primers, which include 322F/796R which amplifies the bacterial 16S V₃–V₄ regions generating an amplicon length of about 454 bp and 799F/1107R which amplifies the V₅–V₆ regions of about 308 bp (Fig. 1B). We then evaluated the overall coverage of the primer pairs and their specificity to domain Bacteria. Probe sequences were blasted

against the nonredundant SSU r132 using TestProbe 3.0 with the default settings, while sequences of the primer pairs were blasted against the SILVA Database SSU r132 using TestPrimer 1.0 with the default settings. As shown in Table 1, all the four primer pairs are specific to domain Bacteria. At zero or one-mismatch stringency, primer pair 322F/796R showed overall coverage higher than 799F/1107R.

Experimental evaluation of designed primer pairs using soil samples

Soil microbiota was used as the material to compare the overall coverage and accuracy of each primer pair in deciphering a microbiota structure. Soil represents the most diverse ecosystem on earth, with exceptionally high bacterial species diversity. Planting soils are colonized by bacterial communities with more abundant bacteria and wider species diversity than those of bacterial communities on and inside plants. It is generally believed that the bacteria inside plant roots and leaves are mainly derived from the soil through root absorption [12, 32, 41, 42]. Meanwhile, planting soil does not contain any plant DNA that would affect bacterial-16S amplification. Using planting soil as the material, the designed primer sets were compared experimentally with the widely used primer pairs 338F/806R and 799F/1193R.

We extracted DNA from soil samples and conducted PCR using the protocol described above. The amplicon libraries were sequenced using the Illumina Miseq PE300 platform. In total, the four primer sets revealed 7,656 OTUs. Because the four primer sets amplified different 16S rDNA regions, the four 16S amplicon libraries did not share any identity at the OTU level (Fig. 2A). At the genus and phylum levels, the four 16S amplicon libraries shared 48.06% and 75.96% identities, respectively (Fig. 2B and C). When the same or similar 16S regions were amplified, the amplicon libraries shared higher identities. At the OTU, genus, and phylum levels, the amplicon libraries obtained using 322F/796R and 338F/806R shared 70.27%, 82.28%, and 90.48% identity, respectively, and those obtained using 799F/1107R and 799F/1193R shared 63.87%, 77.85%, and 90.48% identity, respectively. These results indicated a better comparability when the same 16S variable regions were amplified.

Community β -diversity was analyzed by principal coordinate analysis (PCoA) based on Bray–Curtis distances. At the OTU level, the first principal coordinate accounted for 70.44% of the variability, and the samples with different 16S variable-region amplified were separated along that axis (Fig. 2D). Samples primed with the V_3 – V_4 primer sets, including 338F/806R (Fig. 2D, red circle) and 322F/796R (Fig. 2D, blue triangle), clustered together. Those samples were separated from those primed with the V_5 – V_7 and V_5 – V_6 primer sets. The clustering pattern was similar at the species and genus levels (Fig. 2D). Because host compartment, environmental factors, and host genotype determine the diversity of microbiota composition at lower taxonomic ranks [12, 43, 44], that is, at the genus and species levels, it is better to use primers targeting the same 16S variable regions to produce more comparable sequencing results.

At various taxonomic levels, the four primer sets revealed similar community composition, especially the predominant taxa. However, some taxa exhibited differences in relative abundance among the different

amplicon libraries, and a few low-abundant taxa were undetectable using some of the primer sets (Fig. 2E–F). At the phylum level, amplification with each primer set revealed the same four predominant phyla: Proteobacteria, Firmicutes, Actinobacteria, and Bacteroidetes (Fig. 2E). These four phyla constituted 85.68% (322F/796R), 81.61% (799F/1107R), 80.30% (799F/1193R), and 71.80% (338F/806R) of the 16S amplicons. Among the four main phyla, Proteobacteria had the highest relative abundance (>30%), followed by Firmicutes (17%–20%) (Fig. 2E). The four primer sets revealed six other phyla with relative abundance of >1%: Acidobacteria, Chloroflexi, Nitrospirae, Verrucomicrobia, Gemmatimonadetes, and Patescibacteria. Among them, Chloroflexi exhibited significantly higher relative abundance in the 338F/806R 16S amplicon library (11.61%) than in the amplicon libraries obtained using the other three primer sets (less than 5%). A few phyla were amplified by only some primer sets, for example, Verrucomicrobia was not detected by 322F/796R and constituted 1.62%–3.53% relative abundance in the other three amplicon libraries (Fig. 2E).

At the class level, amplification using each primer set revealed seven highly abundant classes: Gammaproteobacteria, Deltaproteobacteria, Actinobacteria, Clostridia, Bacteroidia, Alphaproteobacteria, and Bacilli (Fig. 2F). The relative abundance of each class in the soil microbiota resolved by different primer sets is close to each other and the total abundances of the seven classes constituted 82.91% (322F/796R), 79.12% (799F/1107R), 76.51% (799F/1193R), and 69.77% (338F/806R) of the 16S amplicons. There were eight other classes whose relative abundance is over 1%. Amplification preference was observed among these classes, and it corresponded to that observed at the phylum level. For example, the class Anaerolineae (phylum Chloroflexi) exhibited significantly higher relative abundance in the 338F/806R 16S amplicon library (9.998%) than in the other three amplicon libraries (<2%). Class Subgroup_6 (phylum Gemmatimonadetes) was only detected using the primer pair 338F/806R. Class Verrucomicrobiae (phylum Verrucomicrobia) was not detected using the primer pair 322F/796R (Fig. 2F). These differences are summarized in Supplementary file 1: Fig. S2.

Because Chloroflexi exhibited high relative abundance in the 16S amplicon library obtained using 338F/806R (Fig. 2E and Fig. S2A), we analyzed this phylum at various lower taxonomic levels to explore its composition (Fig. S2B–D). Anaerolineae was the main detected class in Chloroflexi (Fig. S2B). At the order level, SBR1031, which belongs to Anaerolineae, was amplified only by the primer pair 338F/806R, while Anaerolineales was amplified by 338F/806R (40.65%) and 799F/1193R (1.10%) at very different relative abundance. The high abundance of SBR1031 and Anaerolineales was the reason for the high abundance of Chloroflexi detected by 338F/806R. Dehalococcoidales, which belongs to Dehalococcoidia, was much more abundant in the libraries obtained using 799F/1193R (14.63%) and 322F/796R (6.05%) than in those obtained using 338F/806R (<1%) and 799F/1107R (<1%) (Fig. S2C).

Definition of rice root and leaf endophytic bacteria

We observed root and leaf surfaces by SEM to determine whether any surface bacteria remained after surface sterilization. The results showed that it was easy to obtain bacteria-free leaf surfaces using a

standard surface sterilization method. In the SEM analyses, there were no bacteria attached to the leaf surface or in and around the stomata (Fig. 3A–C). In contrast, many bacteria remained tightly attached to the root surface after surface sterilization and were mainly distributed between adjacent cells or in grooves on the cell surface (Fig. 3D–F). This tight attachment of bacteria suggests that this may be one invasion route for endophytes. On the basis of these results, in this study, the leaf endophytic bacteria refer to all bacteria colonizing the internal tissues, while the root endophytes include both the bacteria colonizing the internal tissues and those unremovable by surface sterilization.

Experimental evaluation of designed primer sets using plant materials

The designed primer sets were applied to the rice genomic DNA template to evaluate the efficiency in excluding plant DNA contamination. Leaf and root samples were collected from 5-week-old greenhouse-grown rice plants (cultivar Nipponbare), and DNA was extracted from the surface-sterilized plant materials. PCR was performed in accordance with the protocols described above, and the amplicon libraries were sequenced using the Illumina Miseq PE300 platform. The sequencing results indicated that from all the plant materials, including leaf and root samples, there was no Ct16S sequence co-amplification (Fig. S3A and 3B). Except for the leaf sample amplified using 322F/796R, all the other three amplicon libraries had little (<1.8%) or no MtDNA contamination. Rice leaf samples amplified using 322F/796R had 87.8% MtDNA contamination on average (Fig. 4A&B), indicating that primer 322F was not very efficient in avoiding Mt18S contamination. Considering the decreased efficiency of bacterial-DNA amplification when endophytes are much less abundant than plant mitochondria [31, 32], the surface-sterilized rice leaf might contain bacterial amounts significantly fewer than the root. The large proportion of Mt18S in the plant genomic DNA samples might be the reason for the Mt18S contamination.

After removing the Mt18S reads from the sequencing results, we compared the obtained plant endophytic bacterial structures. At the genus level, the two 16S amplicon libraries of rice leaf endophytes shared 47.9% identity, and those of rice root symbionts shared 52.4% identity (Fig. 4C&D). At each taxonomic level, the two primer sets revealed similar bacterial composition; however, various taxa exhibited differences in relative abundance (Fig. S4). A few taxa were amplified by only one primer set. At the family level, both Verrucomicrobia and Pedosphaeraceae (phylum Verrucomicrobia), were only detected by 799F/1107R at a low relative abundance (<0.4% in leaf samples and <2% in root samples). Both Amoebohilaceae and Flavobacteriaceae were only detected by 322F/796R. All these primer-specific taxa had low relative abundance (Fig. S4C).

Optimization of primer pair 322F/796R

To specifically amplify bacterial 16S from the plant materials, especially those containing very few endophytes, we modified the primer 322F to reduce co-amplification of Mt18S, thus increasing the

efficiency of 322F/796R in producing plant-DNA free bacterial amplicon library from the plant materials that contain low-abundant endobacteria. The penultimate base was changed from G (guanine) to A (adenosine), C (cytosine), T (thymine), or H (a proportional mixture of A, C, and T); these modified primers were named 322F-A, 322F-C, 322F-T, and 322F-H, respectively. These 322F-derivatives are hereafter referred to as “322F-Dr” (Table 1). All of these primers contained two adjacent mismatches to Mt18S at the 3' end. The primer sets 322F-A/796R, 322F-C/796R, 322F-T/796R and 322F-H/796R (collectively, “322F-Dr/796Rs”) were then used to amplify bacterial 16S from rice leaf genomic DNA. As expected, amplification with each of the 322F-Dr/796Rs resulted in very little MtDNA contamination (<1.8% at the genus level; Fig. 4E).

After removing the Mt18S sequences, we compared the taxonomic classifications and the α -diversity and β -diversity revealed by 322F/796R and each of the 322F-Dr/796Rs. At the OTU level, both Ace and Shannon's index analyses showed no significant difference among the samples primed with different primer sets (Fig. 4F&G). These results indicated that all the 322F-Dr/796Rs, when compared with 322F/796R, did not change the ability in decipher the richness and diversity of the bacterial communities. We then compared the β -diversity using principal component analysis (PCA) analysis at the OTU level. As shown in Fig. 4H, the bacterial communities deciphered using 322F/796R and each of the 322F-Dr/796Rs were clustered together, indicating the high similarity of the four primer pairs in revealing the bacterial community structure of the same rice leaf sample.

We then compared the primer sets 322F/796R and 322F-Dr/796Rs in deciphering the community structure of bacteria in an environmental material that is free of plant DNA interference. Total DNA extracted from planting soil was used as template. At the OTU level, the five 16S amplicon libraries shared 60.08% identity, with only 4.09% of the 16S amplicons belonging to a single amplification group (Fig. S5A). At the genus and phylum levels, the proportions of shared amplicons were 77.21% and 86.84%, respectively, while the proportions of single-group specific amplicons were 3.24% and 5.27%, respectively (Fig. S5B&C). All the 322F-Dr/796Rs, as compared with 322F/796R, had similar abilities to reveal the richness and diversity of the bacterial communities (Fig. S5D&E). In a PCA of β -diversity, samples with the same template rather than those amplified with the same primer pair grouped closer together (Fig. S5F), indicating a weak influence, if any, of primer pairs on the definition of soil bacterial structure.

Rice leaf and root contain different endo-bacteriomes

Using the newly designed primer set 322F-A/796R, we performed 16S NGS to compare the composition of the rice leaf and root endo-bacteriomes. The coverage index of all samples was >0.97. Sequencing results indicated that rice leaf and root shared 66.36% identity at the OTU level; however, the structures of bacterial communities were very different (Fig. 5A and B). At the phylum level, the root and leaf exhibited similar bacterial community composition but differences in the relative abundance of phyla (Fig. 5C). Proteobacteria had the highest relative abundance in all leaf and root samples. In leaf samples, more

than 70% 16S amplicons belonged to Proteobacteria. The phyla Bacteroidetes, Actinobacteria, Fibrobacteres, and Firmicutes had relative abundance ranging between 2% and 5%. There were low-abundant (<2%) Chloroflexi and Deinococcus-Thermus. In root samples, there were three predominant phyla: Proteobacteria (44.6%), Fibrobacteres (35.7%), and Bacteroidetes (16.9%). The phyla Actinobacteria (1.9%) and Chloroflexi (0.5%) were at low abundance. Compared with the root, leaf enriched Proteobacteria, Firmicutes, and Deinococcus-Thermus (Fig. 5D). At the family level, the relative abundance of many taxa differed significantly between the leaf and root endo-bacteriomes (Fig. 5E).

Compared with the root endo-bacteriome, the leaf endo-bacteriome had significantly lower species richness but similar species diversity (Fig. 5F&G), suggesting that there might be large differences among individual leaf samples. As illustrated in a Venn diagram (Fig. 5H), the six leaf 16S amplicon libraries shared 20% identity at the OTU level, with more than 10% of the 16S amplicons belonging to a single amplification group. In contrast, the six root samples shared 50% identity at the OTU level and less than 5% of the OTUs belonged to a single amplification group (Fig. 5I). PCoA indicated that the leaf samples were more separated than the root samples, providing further evidence that the leaf endo-bacteriome was more variable among samples (Fig. 5B).

Using PICRUSt, we explored the function of the endo-bacteriomes in the leaf and root. The rice leaf and root endo-bacteriomes enriched very different KEGG pathways, indicative of different microbial functional features (Fig. 5J). The level 3 KEGG pathway data indicated that the secretion system, bacterial secretion system, two-component system, pentose phosphate pathway, bacterial motility proteins, transporters, and flagellar assembly were more enriched in leaf samples, while the DNA repair and recombination proteins, transcription machinery, pyrimidine metabolism, peptidases, alanine, aspartate and glutamate metabolism, cysteine and methionine metabolism, and amino acid related enzymes were more enriched in root samples. The level 1 KEGG pathway data indicated that environmental information processing and cellular processes were more enriched in the leaf endo-bacteriome, while genetic information processing and metabolism were more enriched in the root endo-bacteriome.

Together, these results suggest that the interactions between the leaf endo-bacteriome and the plant would be more readily affected by environmental factors, whereas root endophytes have stronger interactions with the plant and would participate more in plants' physiological functions.

Quantification of rice endophytic bacteria

Using the designed primer sets, qPCR was performed to quantify the endobacteria in the rice leaf and root. Leaf and root samples were collected from 5-week-old greenhouse-grown rice plants (Nipponbare) and then surface-sterilized before extracting total genomic DNA. The primer 322F-A/796R was used to amplify the endophyte DNA. Within the rice leaf, the number of bacteria was 1/100 that of plant cells. In contrast, the number of bacteria within root and tightly attaching onto the root surface was 10-fold that of plant cells (Fig. 6A). It remained unclear what proportion of these root bacteria were localized inside the root.

Next, we used the standard curve method to quantitate the endobacteria. A qPCR standard curve was created by using gradient dilutions of *E. coli* genomic DNA in 100 ng rice callus DNA as the PCR templates. Rice callus was considered to contain no or very few bacteria. The primer pair 322F-A/796R was used to amplify the bacterial 16S rDNA gene to create a standard curve for the bacterial number (Fig. 6B). The PCR efficiency of 322F-A/796R was 90.55%. The DNA loss rate during extraction was 66.7%–90%, as estimated by adding dilutions of *E. coli* cells to 50–100 mg plant materials before DNA extraction and PCR amplification. We calculated that the rice leaf contained endophytic bacteria at a concentration of 10^6 – 10^7 cells per gram plant fresh weight (Fig. 6C). The surface-sterilized rice root contained much higher bacterial densities (10^9 – 10^{10} cells per gram plant fresh weight; Fig. 6C). The low bacterial abundance in the rice leaf might be the reason for high Mt18S DNA contamination in the amplicon library obtained using 322F/796R.

Discussion

Bacterial endophytes are non-pathogenic bacteria that live inside plant tissues for at least part of their life cycle [45]. Bacterial endophytes offer several benefits to the host plant, particularly immune priming and growth promotion. When living inside plant, under diverse environmental conditions, bacterial endophytes interact with the plant more efficiently than do rhizospheric bacteria [46–47]. Endophytes have to overcome a series of barriers before obtaining the ability to stably colonize the plant internal niches, and this co-evolution with their hosts may mean that they are more readily selected by the host plant from the external environment. The use of endophytic bacteria to promote plant growth or pathogen defense is a relatively novel but promising area for the development of sustainable agriculture.

The common definition of endophytes is derived from a practical description given in 1997 as: those (bacteria) that can be isolated from the surface-disinfested plant tissue or extracted from within the plant and that do not visibly harm the plant [2]. This definition is more suitable for culturable endophytes. New developments in high-throughput technologies, such as next-generation sequencing, have revealed that microbiomes are complex and contain large proportions of unculturable bacteria. Bioinformatic tools have made it possible to characterize the genetic and metabolic elements involved in host–microbe interactions. The combination of culture-independent sequencing, improved culturing technologies, and powerful image analysis have changed, and will continue to change, our concepts of endophytes.

The SEM analyses of surface-sterilized rice leaves and roots revealed many bacteria tightly attached to the root surface that could not be removed by surface sterilization. This phenomenon has also been observed in other studies [48]. The interactions between these bacteria and the host plant might be as strong as those between endophytic bacteria and their host, and some of these tightly attached bacteria may enter the plant and become endophytes. Because it is very hard or not possible to obtain root samples free of surface bacteria for endophyte analysis, the root bacteriome analyzed in this study included bacteria within the root and those tightly attached to the root surface after sterilization.

At present, for most plants, the endo-bacteriome cannot be well characterized by 16S sequencing using existing methods because of large-scale co-amplification of plant organellar DNAs. For example, because of co-amplification of plant DNA, the microbial abundance in the rice endosphere was analyzed based on the proportion of microbial 16S rRNA gene reads relative to organellar (plastidial and mitochondrial) reads. The proportions ranged from 0.4–40.5%, indicating a high rate of co-amplification of plant organellar genes [12]. In another study, the endophytic bacteria in a rice root sample were partially separated from large plant and fungal cells before genomic DNA was extracted; however, there were still 42.69% of the total sequenced reads corresponded to plant DNA contamination [49]. Therefore, designing specific bacterial primers with high bacterial coverage is one of the keys to resolving this problem.

In this study, we designed plant organellar DNA-free and bacterial 16S-specific PCR primers. Three primers, together with the primer 799F, constituted two primer sets. Both 322F and 1107R are specific to bacterial 16S and have the potential to avoid the amplification of plant Mt18S. Both 799F and 796R have the potential to avoid co-amplification of plant Ct16S (Fig. 1A and Fig. S1B). Thus, the two primer pairs 322F/796R and 799F/1107R have the potential to produce plant DNA-free bacterial DNA amplicons (Fig. 1). Experimentally, the primer pair 799F/1107R did not co-amplify plant DNA. The primer pair 322F/796R did not amplify plant Ct16S, and its efficiency in avoiding Mt18S co-amplification was related to the richness and diversity of the bacterial commensals. When the plant material contained bacteria at high abundance with high diversity, i.e., the rice root samples, very few (< 2%) or no Mt18S sequences were co-amplified. However, when the plant material contained bacterial communities with low richness, i.e., the rice leaf samples, plant Mt18S constituted the majority (up to 82%) of amplicons (Fig. 2C). When we introduced a substitution into primer 322F at the penultimate base, the new primers 322F-Drs completely avoided the amplification of Mt18S sequences. Even from rice leaf samples, the bacterial 16S could be specifically amplified (Fig. 3).

Based on the microbiota structures revealed by the different PCR primer sets, we found that amplification using different primers provided different results for microbiota structure. Amplification of different 16S variable regions enlarged such differences. For example, when the soil microbiota was analyzed by 338F/806R, 322F/796R, 799F/1107R, and 799F/1193R, there was only 48.06% identity at the genus level. When the same 16S variable regions were amplified by different primers, the identities were higher. At the genus level, the amplicon libraries obtained with 338F/806R and 322F/796R shared 82.28% identity, while those obtained with 799F/1107R and 799F/1193R shared 77.85% identity. The main differences among libraries were in the relative abundance of taxonomic groups, rather than bacterial community composition. At each taxonomic level, the main bacteria were the same, regardless of the primers used to obtain the library. Because bacteriomes associated with animal and environmental samples are commonly sequenced using primers targeting the 16S V₃–V₄ regions [26, 50], the primer pair 322F-A/796R was selected for the NGS and q-PCR analysis of rice endo-bacteriomes.

The results of the qPCR analyses showed that the endophytic bacterial community in the rice leaf was much smaller than that in the root. The bacterial cell density within the rice leaf was 1/1000 of that in the root, however, the two bacteriomes shared a large proportion of taxa and had similar species diversity.

Compared with the root endo-bacteriome, the leaf endo-bacteriome was much more variable among individual plants. That is, the leaf samples contained as many bacterial species as the root samples, but a large proportion of the leaf endophytes were present at low abundance or only in a single leaf sample. Under the same host genotype growing in the same soil type, the composition of root symbionts was similar among different individual plants. The differences in endo-bacteriome composition among root or leaf samples might be related to the bacterial source and density. Root symbionts are mainly derived from soil microbiota and are present at a high density. In contrast, the leaf endophytic bacteria may originate from the root, air, stomata, hydathodes, and/or feeding insects. Some endophytic bacteria are present at very low densities within leaf tissues and may be opportunistic symbionts or pathogens that have not established a successful colonization within the plant tissue.

This is the first study to quantify rice leaf endophytic bacteria using a non-culture method. In the past, neither culture nor non-culture methods have been effective for quantifying plant endophytic bacteria. A recent study quantified endophytes in leaves of *Arabidopsis thaliana* by separation, culturing, and colony counting and estimated the bacterial density at about 10^5 CFU/g fresh weight. The density is at least 100-fold lower than that of the epiphytes. The study have also shown that plants can control phyllosphere microbiota to ensure plant health. For example, the *Arabidopsis* quadruple mutant *min7fls2efrcerk1*, which is simultaneously defective in pattern-triggered immunity and the MIN7 vesicle-trafficking pathway, showed altered endophytic phyllosphere microbiota and a 100-fold increase in the density of endophytic bacteria. The mutant displayed leaf-tissue damage associated with dysbiosis. In this study, we estimated the cell density of rice phyllosphere endophytic bacteria at 10^6 – 10^7 CFU/g fresh weight, higher than that estimated for the endobacteria within *Arabidopsis* leaves. The presence of unculturable bacteria might be one reason for the higher endophytic bacterial density detected in our study. Our study did not quantify the epiphytic phyllosphere microbiota. We found that when compared to the epiphytic rhizosphere microbiota, the epiphytic phyllosphere microbiota were more easily removed by surface sterilization treatment.

Conclusions

We designed gene-specific primer sets to amplify bacterial 16S rRNA without co-amplification of plant organellar 16S or 18S rRNA genes. This allowed us to determine the composition of plant endophytic bacterial communities using plant whole-genomic DNA as the PCR template. The PCR amplicons were suitable for next-generation sequencing using the Illumina platform. These methods will be especially useful for analyzing plant materials that contain very few bacteria or an endo-bacteriome with a simple structure, which are difficult to analyze using other methods. Using the newly designed primer sets and sequencing methods, we determined the structure and composition of the endo-bacteriome in rice leaves and roots and quantified the endophytes' population size. These protocols are suitable for analyses of various plants and will significantly promote research on plant endo-bacteriomes.

Abbreviations

PCR: polymerase chain reaction

OTU: operational taxonomic unit

PICRUSt: Phylogenetic Investigation of Communities by Reconstruction of Unobserved States

KEGG: Kyoto Encyclopedia of Genes and Genomes

SEM: Scanning electron microscopy

PCA: principal component analysis

PCA: principal coordinate analysis

NCBI: National Center for Biotechnology Information

SSU: small subunit rRNA gene

FLASH: Fast Length Adjustment of Short Reads

CTAB: Cetyltrimethylammonium bromide

SDS: Sodium dodecyl sulfate

NGS: next generation sequencing

KO: KEGG Orthology

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of supporting data

The datasets used and/or analyzed in this study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from the National Transgenic Major Project of China (2019ZX08010-004) to LZ, the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDB11040800) to LZ, the National Natural Science Foundation of China (No. 31871932) to LZ, and the Institute Research Fund for Microbiome (Institute of Microbiology, Chinese Academy of Sciences) (No. Y854FJ2012) to LZ.

Authors' contributions

L.C. designed primers, conducted next-generation sequencing of planting soil, and conducted qPCR and SEM analyses. L.C., M.Z., D.L., and H.S. performed NGS of rice plants. M.Z. designed and performed most work on single-clonal sequencing. D.L. contributed to sampling of planting soil and rice plants. J.W. and X.C. guided the experimental design and planning. L.Z. and R.F. conceptualized, administered, and supervised the study and wrote the manuscript. All authors discussed the results and commented on the manuscript.

References

1. Wilson D. Endophyte - the evolution of a term, and clarification of its use and definition. *OIKOS*. 1995;73:274-276.
2. Hallmann J, Quadt-Hallmann A, Mahaffee WF, Kloepper JW. Bacterial endophytes in agricultural crops. *Can J Microbiol*. 1997;43:895-914.
3. Azevedo JL, Araújo WL. Diversity and applications of endophytic fungi isolated from tropical plants, in *Diversity and applications of endophytic fungi isolated from tropical plants*. Boca Raton: CRC Press; 2007. p. 189–207.
4. Hardoim PR, van Overbeek LS, Berg G, Pirttilä AM, Compant S, Campisano A, et al. The Hidden World within Plants: Ecological and Evolutionary Considerations for Defining Functioning of Microbial Endophytes. *Microbiol Mol Biol Rev*. 2015;79:293-320.
5. Kearney SM, Gibbons SM, Poyet M, Gurry T, Bullock K, Allegretti JR, et al. Endospores and other lysis-resistant bacteria comprise a widely shared core community within the human microbiota. *ISME J*. 2018;12:2403-2416.
6. Forbes JD, Chen CY, Knox NC, Marrie RA, El-Gabalawy H, de Kievit T, et al. A comparative study of the gut microbiota in immune-mediated inflammatory diseases-does a common dysbiosis exist? *Microbiome*. 2018;6:221.

7. Geva-Zatorsky N, Sefik E, Kua L, Pasman L, Tan TG, Ortiz-Lopez A, et al. Mining the Human Gut Microbiota for Immunomodulatory Organisms. *Cell*. 2017;168:928-943 e11.
8. Keller A, Brandel A, Becker MC, Balles R, Abdelmohsen UR, Ankenbrand MJ, et al. Wild bees and their nests host *Paenibacillus* bacteria with functional potential of avail. *Microbiome*. 2018;6:229.
9. McGovern E, Waters SM, Blackshields G, McCabe MS. Evaluating Established Methods for Rumen 16S rRNA Amplicon Sequencing With Mock Microbial Populations. *Front Microbiol*. 2018;9:1365.
10. Ma B, Wang H, Dsouza M, Lou J, He Y, Dai Z, et al. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J*. 2016;10:1891-901.
11. Su JQ, An XL, Li B, Chen QL, Gillings MR, Chen H, et al. Metagenomics of urban sewage identifies an extensively shared antibiotic resistome in China. *Microbiome*. 2017;5:84.
12. Edwards J, Johnson C, Santos-Medellin C, Lurie E, Podishetty NK, Bhatnagar S, et al. Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc Natl Acad Sci U S A*. 2015;112:E911-20.
13. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. *Nature*. 2009;457:480-U487.
14. Piewngam P, Zheng Y, Nguyen TH, Dickey SW, Joo H-S, Villaruz AE, et al. Pathogen elimination by probiotic *Bacillus* via signalling interference. *Nature*. 2018;562:532-537.
15. Bai Y, Muller DB, Srinivas G, Garrido-Oter R, Potthoff E, Rott M, et al. Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature*. 2015;528:364-369.
16. Knief C, Delmotte N, Chaffron S, Stark M, Innerebner G, Wassmann R, et al. Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *ISME J*. 2012;6:1378.
17. Ueda Y, Frindte K, Knief C, Ashrafuzzaman M, Frei M. Effects of elevated tropospheric ozone concentration on the bacterial community in the phyllosphere and rhizoplane of rice. *PLoS One*. 2016;11:e0163178.
18. Wang W, Zhai Y, Cao L, Tan H, Zhang R. Endophytic bacterial and fungal microbiota in sprouts, roots and stems of rice (*Oryza sativa* L.). *Microbiol Res*. 2016;188-189:1-8.
19. Chen T, Nomura K, Wang X, Sohrabi R, Xu J, Yao L, et al. A plant genetic network for preventing dysbiosis in the phyllosphere. *Nature*. 2020;580:653-657.
20. Cobo-Diaz JF, Baroncelli R, Le Floch G, Picot A. Combined metabarcoding and co-occurrence network analysis to profile the bacterial, fungal and fusarium communities and their interactions in maize stalks. *Front Microbiol*. 2019;10:261.
21. Sharaf H, Rodrigues RR, Moon J, Zhang B, Mills K, Williams MA. Unprecedented bacterial community richness in soybean nodules vary with cultivar and water status. *Microbiome*. 2019;7:63.
22. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res*. 2013;41:e1.

23. Edwards JA, Santos-Medellin CM, Liechty ZS, Nguyen B, Lurie E, Eason S, et al. Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life cycle in field-grown rice. *PLoS Biol.* 2018;16:e2003862.
24. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, Mcculle SL, et al. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A.* 2011;108 Suppl 1:4680.
25. Tkacz A, Cheema J, Chandra G, Grant A, Poole PS. Stability and succession of the rhizosphere microbiota depends upon plant type and soil composition. *ISME J.* 2015;9:2349-2359.
26. Xu N, Tan G, Wang H, Gai X. Effect of biochar additions to soil on nitrogen leaching, microbial biomass and bacterial community structure. *Eur J Soil Biol.* 2016;74:1-8.
27. Xiong J, Liu Y, Lin X, Zhang H, Zeng J, Hou J, et al. Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. *Environ Microbiol.* 2012;14:2457-2466.
28. Bulgarelli D, Garrido-Oter R, Münch PC, Weiman A, Dröge J, Pan Y, et al. Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host Microbe.* 2015;17:392-403.
29. Bulgarelli D, Rott M, Schlaeppi K, van Themaat EVL, Ahmadinejad N, Assenza F, et al. Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature.* 2012;488:91-95.
30. Chelius MK, Triplett EW. The Diversity of Archaea and Bacteria in Association with the Roots of *Zea mays* L. *Microb Ecol.* 2001;41:252-263.
31. Ghyselinck J, Pfeiffer S, Heylen K, Sessitsch A, De Vos P. The effect of primer choice and short read sequences on the outcome of 16S rRNA gene based diversity studies. *PLoS One.* 2013;8:e71360.
32. Bulgarelli D, Schlaeppi K, Spaepen S, Ver Loren van Themaat E, Schulze-Lefert P. Structure and functions of the bacterial microbiota of plants. *Annu Rev Plant Biol.* 2013;64:807-838.
33. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 2010;26:2460-2461.
34. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, et al. ARB: a software environment for sequence data. *Nucleic Acids Res.* 2004;32:1363-1371.
35. Fadrosch DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome.* 2014;2.
36. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* 2011;27:2957-2963.
37. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods.* 2013;10:996-8.
38. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 2009;37:D141-D145.

39. Muller DB, Vogel C, Bai Y, Vorholt JA. The plant microbiota: systems level insights and perspectives. *Annu Rev Genet.* 2016;50:211-234.
40. Koiv V, Arbo K, Maivali U, Kisand V, Roosaare M, Remm M, et al. Endophytic bacterial communities in peels and pulp of five root vegetables. *Plos One.* 2019;14.
41. Cordovez V, Dini-Andreote F, Carrion VJ, Raaijmakers JM. Ecology and evolution of plant microbiomes. *Annu Rev Microbiol.* 2019;73:69-88.
42. Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, et al. Defining the core *Arabidopsis thaliana* root microbiome. *Nature.* 2012;488:86-90.
43. Duran P, Thiergart T, Garrido-Oter R, Agler M, Kemen E, Schulze-Lefert P, et al. Microbial interkingdom interactions in roots promote arabidopsis survival. *Cell.* 2018;175:973-983 e914.
44. Fitzpatrick CR, Copeland J, Wang PW, Guttman DS, Kotanen PM, Johnson MTJ. Assembly and ecological function of the root microbiome across angiosperm plant species. *Proc Natl Acad Sci U S A.* 2018;115:E1157-E1165.
45. Rosenblueth M, Martinez-Romero E. Bacterial endophytes and their interactions with hosts. *Mol Plant Microbe Interact.* 2006;19:827-37.
46. Santoyo G, Moreno-Hagelsieb G, Orozco-Mosqueda Mdel CGlick BR. Plant growth-promoting bacterial endophytes. *Microbiol Res.* 2016;183:92-9.
47. Ali S, Charles TC, Glick BR. Delay of flower senescence by bacterial endophytes expressing 1-aminocyclopropane-1-carboxylate deaminase. 2012;113:1139-1144.
48. Carrion VJ, Perez-Jaramillo J, Cordovez V, Tracanna V, de Hollander M, Ruiz-Buck D, et al. Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. *Science.* 2019;366:606-612.
49. Sessitsch A, Hardoim P, Doring J, Weilharter A, Krause A, Woyke T, et al. Functional characteristics of an endophyte community colonizing rice roots as revealed by metagenomic analysis. *Mol Plant Microbe Interact.* 2012;25:28-36.
50. Adair KL, Wilson M, Bost A, Douglas AE. Microbial community assembly in wild populations of the fruit fly *Drosophila melanogaster*. *ISME J.* 2018;12:959-972.

Figures

Figure 1

Figure 1

Primer design for plant DNA-free amplification of bacterial 16S. A. Four primers were designed to contain mismatch(es) at the 3' end to either plant Mt18S or Ct16S. Only rice Mt18S or Ct16S sequences are shown; other plants have similar sequences within aligned regions (see Figure S1B). Arrows indicate bases that mismatch to plant Mt18S or Ct16S. Arrow(s) in red indicate mismatched base or adjacent mismatches at 3' end. H=A/T/C, V=G/A/C, W=A/T, R=A/G, M=A/C, K=G/T, Y=C/T. Except for primer 799F,

other forward (322F) and reverse (796R and 1107R) primers were named according to position of first and last nucleotide in 16S rRNA gene of *Escherichia coli* DSM 30083 (GenBank accession number: JMWST01000030). B. Primer sets and amplification regions for bacterial 16S. a) Widely used primer sets. b) Primer pair 322F/796R designed to amplify bacterial 16S V3–V4 regions while avoiding plant Mt18S (322F, mtDNA-free) and Ct16S (796R, ctDNA-free). c) Primer pair 799F/1107R designed to amplify bacterial 16S V5–V6 regions while avoiding plant Mt18S (1107R, mtDNA-free) and Ct16S (799F, ctDNA-free).

Figure 2

Figure 2

Experimental comparison of newly designed and widely used primer sets using soil microbiota. A–C. Venn diagrams at OTU (A), genus (B) and phylum (C) levels. Different colors represent different primer set groups. Overlaps represent common taxa between groups; non-overlapping portions represent unique taxa in each group. D. Principal coordinate analysis (PCoA) based on Bray–Curtis distances indicating that largest separation among generated microbial communities is 16S amplification region (V3-V4 or V5-V7) and second largest separation is among samples (three different soil samples). E–F. Relative abundances deciphered using four different primer sets. E. Results at phylum level. F. Results at class level. Taxa with relative abundance >1% are shown. Arrows indicate taxa that were differentially amplified by the four primer sets.

Figure 3

Figure 3

Scanning electron micrographs of bacteria on surface of sterilized rice leaf (A–C) or root (D–F). (A–C) Sterilized rice leaf surfaces with no bacteria visible. (D–F) Sterilized rice root surfaces with abundant bacteria colonizing between adjacent cells or in grooves on cell surface.

Figure 4

Figure 4

Experimental evaluation of designed primer sets in plant DNA-free amplification. A–B. Proportion of Mt18S in each amplicon library generated using 322F/796R and 799F/1107R. C–D. Venn diagram showing community compositions at the genus level. Different colors represent different primer set groups. E. Proportion of Mt18S in each amplicon library generated using 322F/796R and 322F-Dr/796Rs. F–G. Alpha diversity indices of different groups (322F-796R and 322F-Dr/796Rs). H. Principal component analysis (PCA) based on Bray–Curtis distances showing that all five different primer sets, when applied to same leaf sample (three samples were sequenced), generated similar results for microbial community structure. R and L, rice root and leaf, respectively. G, 322F/796R; A, 322F-A/796R; T, 322F-T/796R; C, 322F-C/796R; H, 322F-H/796R.

Figure 5

Figure 5

Differences in endo-bacteriomes between rice leaf and root. A. Venn diagram showing number of OTUs shared by rice leaf and root endo-bacteriome and those unique to either. B. PCoA based on Bray–Curtis distances at OTU level at 97% identity. Each leaf or root sample is represented by a dot. Circle (leaf) or triangle (root) represent different groups. C. Relative abundance at phylum level. Taxa with relative abundance >0.5% are shown. D–E. Wilcoxon rank-sum test bar plot at phylum (D) and family (E) levels. *;

p<0.05, **, p<0.01. F–G. Alpha diversity indices of different groups at OTU level. ***, p<0.001. NS, not significantly different. H–I. Venn diagram showing number of OTUs shared among individual rice leaf samples (H) or root samples (I). J. PICRUSt prediction of functional composition of endo-bacteriome within different rice tissues. KEGG Orthology (KO) functional categories and pathways that were significantly enriched in root are shown in orange; those significantly enriched in leaf are shown in green. R1-R6 and L1-L6, six root and leaf samples, respectively. R and L, rice root and leaf endo-bacteriome, respectively.

Figure 6

Figure 6

Quantification of rice endophytic bacteria by qPCR. A. Relative quantification of leaf and root endophytic microbiota. Bacterial numbers were calculated from copy numbers of 16S rRNA gene amplified using primer 322F-A/796R. Plant cell numbers were calculated from copy numbers of rice Actin gene amplified using primers OsActin-F/OsActin-R. B. Standard curve generated by plotting mean Ct values versus ten-fold serial dilutions of standard bacterial numbers over a range of concentrations from 10^3 to 10^7 CFU/ml. C. Absolute quantities of endophytic leaf and root microbiota. Bacterial density was calculated as number of bacterial cells per mg fresh weight of rice leaf or root. Rate of DNA loss during extraction, which was 1/10 to 1/3, was not included in these calculations of bacterial density. Mean and SD were calculated from three independent experiments with 5 plants used in each experiment.

Supplementary Files

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

- [supplement1.xlsx](#)
- [supplement2.docx](#)
- [supplement3.pdf](#)
- [Table1.pdf](#)
- [Table2.pdf](#)