



# Performance of 16s rDNA Primer Pairs in the Study of Rhizosphere and Endosphere Bacterial Microbiomes in Metabarcoding Studies

Bram Beckers<sup>1</sup>, Michiel Op De Beeck<sup>1†</sup>, Sofie Thijs<sup>1</sup>, Sascha Truyens<sup>1</sup>, Nele Weyens<sup>1</sup>, Wout Boerjan<sup>2,3</sup> and Jaco Vangronsveld<sup>1\*</sup>

<sup>1</sup> Centre for Environmental Sciences, Hasselt University, Diepenbeek, Belgium, <sup>2</sup> Department of Plant Systems Biology, Flanders Institute for Biotechnology (VIB), Gent, Belgium, <sup>3</sup> Department of Plant Biotechnology and Bioinformatics, Ghent University, Gent, Belgium

## OPEN ACCESS

### Edited by:

Benjamin Gourion,  
LIPM Centre National de la Recherche  
Scientifique/Institut National de la  
Recherche Agronomique, France

### Reviewed by:

Sanguin Hervé,  
Cirad, France  
Samuel Mondy,  
Institut National de la Recherche  
Agronomique, France

### \*Correspondence:

Jaco Vangronsveld  
jaco.vangronsveld@uhasselt.be

### † Present Address:

Michiel Op De Beeck,  
Department of Biology, Lund  
University, Lund, Sweden

### Specialty section:

This article was submitted to  
Plant Biotic Interactions,  
a section of the journal  
Frontiers in Microbiology

Received: 12 November 2015

Accepted: 18 April 2016

Published: 13 May 2016

### Citation:

Beckers B, Op De Beeck M, Thijs S, Truyens S, Weyens N, Boerjan W and Vangronsveld J (2016) Performance of 16s rDNA Primer Pairs in the Study of Rhizosphere and Endosphere Bacterial Microbiomes in Metabarcoding Studies. *Front. Microbiol.* 7:650. doi: 10.3389/fmicb.2016.00650

Next-generation sequencing technologies have revolutionized the methods for studying microbial ecology by enabling high-resolution community profiling. However, the use of these technologies in unraveling the plant microbiome remains challenging. Many bacterial 16S rDNA primer pairs also exhibit high affinity for non-target DNA such as plastid (mostly chloroplast) DNA and mitochondrial DNA. Therefore, we experimentally tested a series of commonly used primers for the analysis of plant-associated bacterial communities using 454 pyrosequencing. We evaluated the performance of all selected primer pairs in the study of the bacterial microbiomes present in the rhizosphere soil, root, stem and leaf endosphere of field-grown poplar trees (*Populus tremula* × *Populus alba*) based on (a) co-amplification of non-target DNA, (b) low amplification efficiency for pure chloroplast DNA (real-time PCR), (c) high retrieval of bacterial 16S rDNA, (d) high operational taxonomic unit (OTU) richness and Inverse Simpson diversity and (e) taxonomic assignment of reads. Results indicate that experimental evaluation of primers provide valuable information that could contribute in the selection of suitable primer pairs for 16S rDNA metabarcoding studies in plant-microbiota research. Furthermore, we show that primer pair 799F-1391R outperforms all other primer pairs in our study in the elimination of non-target DNA and retrieval of bacterial OTUs.

**Keywords:** 16S rDNA metabarcoding, 454 pyrosequencing, plant microbiome, chloroplast DNA, endophytes

## INTRODUCTION

The development and implementation of next-generation sequencing technologies (NGS) and their corresponding bioinformatics tools have revolutionized the methods for studying microbial ecology by enabling high-resolution community profiling (Margulies et al., 2005; Sogin et al., 2006; Parameswaran et al., 2007; Shendure and Ji, 2008; Metzker, 2010; Caporaso et al., 2011). DNA sequencing of 16S rDNA genes or gene fragments have fuelled large microbial community studies in humans [Andersson et al., 2008; Qin et al., 2010; HMPC (Human Microbiome Project Consortium), 2012], the gut microfauna of insects (Sudakaran et al., 2012; Hansen and Moran, 2014), different natural habitats (Andrew et al., 2012; DeLeon-Rodriguez et al., 2013; Hamdan et al., 2013) as well as plant microbiome studies (Gottel et al., 2011; Mendes et al., 2011; Bulgarelli et al., 2012; Lundberg et al., 2012; Peiffer et al., 2013; Shakya et al., 2013; Edwards et al., 2015). Currently,

the most used next generation sequencing platforms in the profiling of microbiomes are the 454 pyrosequencing technology (Margulies et al., 2005) and the Illumina MiSeq and HiSeq systems (Caporaso et al., 2011, 2012). 454 pyrosequencing has continuously demonstrated its effectiveness in describing microbial communities by enabling highly multiplexed sequencing of short hypervariable 16S rDNA regions (Gottel et al., 2011; Lundberg et al., 2012; Bulgarelli et al., 2013; Shakya et al., 2013). In recent years, the Illumina HiSeq and MiSeq systems have also earned their place in high-throughput 16S rDNA sequencing and have even surpassed 454 pyrosequencing in terms of read quantity and quality (Caporaso et al., 2012).

For both technologies, several studies have evaluated their technical aspects such as error profiles (Kunin et al., 2010; Balzer et al., 2011; Gilles et al., 2011; Nakamura et al., 2011; Schirmer et al., 2015), platform-specific effects on the observed microbial communities (Claesson et al., 2010; Luo et al., 2012; Nelson et al., 2014; Tremblay et al., 2015) as well as biases introduced by primer specificity with *in silico* studies (Klindworth et al., 2013) and *in vivo* studies (Aird et al., 2011; Berry et al., 2011; Pinto and Raskin, 2012; Kennedy et al., 2014). Another important aspect within microbial ecology and 16S rDNA community profiling with NGS techniques is the occurrence of contaminating sequences, which are routinely co-extracted during DNA extraction from various biotic samples. In numerous research areas, NGS-based approaches are susceptible to contamination with undesired sequences (non-target DNA) such as in the diagnosis of symptomatic infections (microbial DNA contamination; Strong et al., 2014), in malaria clinical sequencing (human DNA contamination) (Oyola et al., 2013), and food web analysis (Pompanon et al., 2012). More specifically within microbial ecology, in the study of interorganismal association such as the human microbiome [HMPC (Human Microbiome Project Consortium), 2012], plant-microbiome associations (Lundberg et al., 2012; Bodenhausen et al., 2013; Bulgarelli et al., 2013; Ghyselinck et al., 2013) and insect-microbiome studies (Sudakaran et al., 2012; Hansen and Moran, 2014) sequences from an organellar origin (e.g., mitochondria and/or chloroplast DNA) represent a major source of contamination.

This is of particular interest in plant-microbiome research since plants house eukaryotic cells, prokaryotic cells, and eukaryotic plant organelles with a prokaryotic lineage (mitochondria and chloroplast/plastids) (Dyall et al., 2004; Raven, 1970). The number of mitochondria and chloroplasts vary depending on the plant species, cell type and age of the tissue but can be as high as 10,000 chloroplast DNA copies in tobacco leaf cells (Shaver et al., 2006). The homology between bacterial 16S rDNA, chloroplast DNA, and mitochondrial DNA leads to significant challenges in the selection of appropriate primer pairs to study plant-microbe interactions (Ghyselinck et al., 2013). Currently, three general methods exist to reduce the impact of these contaminating sequences: (a) adaptation of existing DNA extraction protocols to reduce co-extraction of organellar DNA (Lutz et al., 2011) or post-extraction separation of host DNA from microbial DNA based on differences in CpG methylation density (Feehery et al., 2013), (b) the development of blocking primers to block and/or reduce amplification of

sequences originating from a eukaryotic host such as peptide nucleic acid-mediated PCR clamping (Lundberg et al., 2013) and suicide polymerase endonuclease restriction (SuPER) (Green and Minz, 2005), and (c) the use of specific mismatch primers during PCR amplification (Chelius and Triplett, 2001; Sakai et al., 2004).

The preferred or most utilized technique is the use specific mismatch primers, which amplify bacterial 16S rDNA sequences while simultaneously avoiding the amplification of chloroplast DNA sequences. Chelius and Triplett (2001) developed the first mismatch primer (799F), with a primer design which centered around two base pair mismatches at positions 798–799 and two additional base pair mismatches at positions 783 and 784 in the chloroplast DNA. Primer 799F has been used with varying success in several plant systems (Bulgarelli et al., 2012; Bodenhausen et al., 2013; Shade et al., 2013). Further, Sakai et al. (2004) modified primer 799F into primer 783Rabc, in an attempt to access the hypervariable regions V3–V4 of the bacterial 16S rDNA genes in the study of the rhizobacterial communities of wheat and spinach. Indeed hypervariable regions V3 and V4 have been the preferred target of the 16S rDNA in studying soil and rhizosphere assemblages and databases are more exhaustive for these regions (Klindworth et al., 2013). Furthermore, Rastogi et al. (2010) used primer 783Rabc to develop a PCR-based method to determine the degree of chloroplast and mitochondrial contamination in DNA samples from plant environments.

However, the experimental performance of these mismatch primers (and their potential to reduce co-amplification of non-target DNA) in different plant compartments with low chloroplast/plastid input (rhizosphere soil) and higher chloroplast/plastid input (endosphere compartments) has not been evaluated. Although, *in silico* analyses provide valuable technical information and indicate the theoretical optimal performance of primer pairs, they fail to capture the true experimental potential and are expected to result in an incomplete picture of how primers will perform during PCR amplification (Op De Beeck et al., 2014). Therefore, experimental evaluation of the amplification efficiency and robustness of selected primer pairs in plant-bacteria interaction studies is essential to assess their behavior in these specific conditions.

For this reason, we experimentally tested a set of commonly used primers for the analysis of plant-associated bacterial communities using 454 pyrosequencing (Table 1). We tested all selected primer pairs in the study of plant-associated bacterial communities in rhizosphere, roots, stems, and leaves of hybrid poplar trees (*Populus tremula* × *P. alba*). The different amounts of plastid DNA content of these plant compartments, ranging from virtually no plastid content (rhizosphere soil) to very high plastid (chloroplast) content (leaves) allows us to evaluate the performance of the selected primer sets in specific conditions.

## MATERIALS AND METHODS

### Study Site Description and Sampling

A poplar (*Populus tremula* × *Populus alba*, cv “717-1-B4,” female clones) field trial located in Ghent, Belgium (property of VIB)

**TABLE 1 | Summary of primers used in the current study.**

Primer pairs	Primer sequence (5'–3')	A	B	C	D	References
799F	AACMGGATTAGATACCCCKG	79.7	0.29	4	V5-V6-V7	Chelius and Triplett, 2001
1391R	GACGGGCGGTGWGTRCA	84.6	1.44	0		Walker and Pace, 2007
967F	CAACGCGAAGAACCCTTACC	80.9	0.34	0	V6-V7	Sogin et al., 2006
1391R	GACGGGCGGTGWGTRCA	84.6	1.44	0		Walker and Pace, 2007
799F	AACMGGATTAGATACCCCKG	79.7	0.29	4	V5-V6-V7	Chelius and Triplett, 2001
1193R	ACGTCATCCCCACCTTCC	78.1	0.20	0		Bodenhausen et al., 2013
341F	CCTACGGGNGGCWGCAG	91.2	0.05	0	V3-V4	Klindworth et al., 2013
785R	GACTACHVGGGTATCT AATCC	86.2	0.09	0		Klindworth et al., 2013
68F	TNANACATGCAAGTCGRRRCG	72.5	0.60	0	V1-V4	McAllister et al., 2011
783Rabc	CTACC*AGGGTATCTAATCC*TG	70.9	5.05	3		Sakai et al., 2004
68F	TNANACATGCAAGTCGRRRCG	72.5	0.60	0	V1-V3	McAllister et al., 2011
518R	WTTACCGCGGCTGCTG G	87.6	0.09	0		Lee et al., 2010
341F	CCTACGGGNGGCWGCAG	91.2	0.05	0	V3-V4	Klindworth et al., 2013
783Rabc	CTACC*AGGGTATCTAATCC*TG	70.9	5.05	3		Sakai et al., 2004

Primers are indicated as forward (F) or reverse (R).

\*Primer 783Rabc is a primer mix (Sakai et al., 2004):

(a) 5'-CTACCAGGGTATCTAATCC-3',

(b) 5'-CTACCGGGGTATCTAATCCCG-3',

(c) 5'-CTACCGGGGTATCTAATCCCG-3'.

(A) Primer coverage (%) for Bacteria using *Silva* (Quast et al., 2013) and *probeBase* (Loy et al., 2007), (B) Primer Score tested with *PrimerProspector* 1.0.1 (Walters et al., 2011) (Table S1), (C) Number of mismatches with poplar chloroplast DNA, (D) Hypervariable region of the 16S rRNA operon targeted by primer pairs.

was selected to acquire samples for this study (Custers, 2009). The field trial was established in April 2009 with a density of 15,000 trees per hectare and an inter-plant distance of 0.75 m (Van Acker et al., 2014). Briefly, poplar trees were sampled after ~3.5 years of growth in October 2012. We collected samples from rhizosphere soil, roots, stems, and leaves of three biologically independent poplar individuals. Per poplar individual, we collected (a) 10 g of roots at a depth of 5–10 cm below ground in 50 mL plastic tubes, (b) one complete offshoot for the stem and leaf samples. To standardize and maximize reproducibility of stem samples, several small “cores” with bark (5–7 cores, each about 1 cm) were collected from each offshoot. Finally all leaves of each offshoot were gathered and placed in sealed plastic bags for transportation.

## Processing of Samples

Poplar root samples were shaken for 10 min on a shaking platform (100 rpm) and soil particles dislodged from the roots were collected as rhizosphere soil. Rhizosphere soil was sieved using a 2 mm sieve for homogenization and removal of residual roots and debris. Subsequently the samples were stored at  $-80^{\circ}\text{C}$  until DNA was extracted.

Epiphytes (microbes living on the plant surface) were removed from all plant samples (roots, stems, and leaves) via surface-sterilization under aseptic conditions. Samples were sequentially washed with (a) sterile Millipore water (30 s), (b) followed by immersion in 70% (v/v) ethanol (2 min), (c) sodium hypochlorite solution (2.5% active  $\text{Cl}^-$ , 5 min) supplemented with 0.1% Tween 80, (d) 70% (v/v) ethanol (30 s) and finalized by rinsing the samples five times with sterile Millipore water. Finally plant samples (~5 g of each compartment per sample) were

homogenized by (a) portioning the samples into small fragments using a sterile scalpel and (b) macerating them in sterile 10 mM phosphate saline (PBS) buffer (130 mM NaCl, 7 mM  $\text{Na}_2\text{HPO}_4$ , 3 mM  $\text{NaH}_2\text{PO}_4$ , pH 7.4) using a Polytron PR1200 mixer (Kinematica A6) in four cycles of 2 min with cooling of the mixer on ice between cycles to reduce heating of the samples. Finally, quadruplicate aliquots of each sample (1.5 mL) of the homogenized plant material (root, stem, and leaf) were stored for all poplar individuals at  $-80^{\circ}\text{C}$  until DNA was extracted.

## DNA Extraction

DNA from rhizosphere, roots, stems, and leaves (further denoted as “plant compartments”) was extracted in quadruplicate from three biologically independent poplar individuals to minimize DNA extraction bias (Feinstein et al., 2009; Op De Beeck et al., 2014). In total, DNA was extracted from 48 samples (3 poplar individuals  $\times$  4 plant compartments  $\times$  4 quadruplicate extractions per sample).

Approximately 250 mg of rhizosphere soil was used for each individual DNA extraction. DNA extraction was performed with the Power Soil DNA Isolation Kit following the protocol provided by the manufacturer (MoBio, Carlsbad, CA, USA). For the plant compartments (roots, stems, leaves), aliquots (1.5 mL) of homogenized plant material were first centrifuged (13,400 rpm, 30 min) to collect all cells. Supernatants were discarded and DNA extractions were performed on pelleted plant material. After optimization of the DNA extraction kit (Figure S1), DNA was extracted from the plant samples using the Invisorb Spin Plant Mini Kit according to the manufacturer’s protocol (Stratec Biomedical AG, Birkenfeld, Germany).

## In Silico Evaluation of Primer Pairs

To select suitable primer pairs from all available 16S rDNA primers, several parameters (Table 1):

- (A) Primer coverage and phylum spectrum of all 16S rDNA primers were queried using data from Silva (Klindworth et al., 2013; Quast et al., 2013) and probeBase (Loy et al., 2007).
- (B) To evaluate the primer-to-target 3' mismatches, primers were tested with PrimerProspector 1.0.1 (Walters et al., 2011) against the Greengenes database (gg\_13\_5.fasta) containing 1,262,986 sequences (DeSantis et al., 2006). This database was curated using mothur (version 1.34.3) to remove sequences containing ambiguous bases and sequences with homopolymers longer than eight bases. The curated database contained 946,815 16S rDNA sequences with an average read length of 1399.4 basepairs. All primer tests were performed as described by Walters et al. (2011) using standard settings. Primer scores were calculated based on the following formula:  $\text{weighted score} = \text{non-3' mismatches} \times 0.40 + \text{3' mismatches} \times 1.00 + \text{non-3' gaps} \times 1.00 + \text{3' gaps} \times 3.00$ . An additional penalty score of 3.00 was assigned if the final 3' base of a primer had a mismatch with its target sequence.
- (C) The presence of mismatches with poplar chloroplast DNA was evaluated by downloading the full chloroplast genome of *Populus alba* from NCBI and blasting each primer using Fast PCR (Kalendar et al., 2014).
- (D) Overall coverage of the mostly used hypervariable regions of the 16S rRNA operon and the resulting amplicon length.

## PCR Amplification and 454 Pyrosequencing

For the PCR amplification, we selected seven different primer pairs to evaluate their performance in metabarcoding studies for rhizospheric and endophytic bacteria (Table 1). Primer pairs covered all hypervariable regions from V1 until V7 of the 16S rDNA gene and included, amongst others, primer 799F (5'-AACMGATTAGATACCCKG-3') (Chelius and Triplett, 2001) and primer 783Rabc (5'-CTACC\*AGGGTATCTAATCC\*TG; Sakai et al., 2004), which theoretically minimize chloroplast contamination by providing considerable mismatches with the poplar plastid DNA (3–4 mismatches). All primers and their sequences are listed in Table 1. Except for primer pair 68F-518R, all forward primers were fused to the Roche 454 pyrosequencing adaptor A and a sample-specific 10 bp barcode (multiplex identifiers, MID) and all reverse primers were fused to adaptor B (Roche Applied Science, Mannheim, Germany). For primer pair 68F-518R, the reverse primer was fused to adaptor A and the forward primer was fused to adaptor B.

DNA samples ( $n = 48$ ) were individually amplified using a Techne TC-5000 thermocycler (Bibby Scientific Limited, Staffordshire, UK) with the seven different primer pairs. Since the concentration of bacterial DNA in comparison with the plant DNA is low, we chose a nested PCR strategy to amplify all samples and thereby minimize the formation of primer dimers. A first round of PCR amplification was conducted using primers

without the Roche 454 pyrosequencing adaptors and sample-specific barcode. Each 25  $\mu$ l PCR reaction contained  $\sim 10$  ng of DNA and was carried out using the FastStart High Fidelity PCR System (Roche Applied Science, Mannheim, Germany). Each reaction contained 2.75  $\mu$ l FastStart 10 x reaction buffer, 1.8 mM MgCl<sub>2</sub>, 0.2 mM dNTP mix, 0.4  $\mu$ M of each primer, and 2 U FastStart HiFi polymerase. Cycling conditions included: initial denaturation at 94°C for 3 min, followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 53°C for 1 min, and extension at 72°C during 1 min; a final extension phase was performed at 72°C during 10 min. PCR amplicon pools were cleared from residual primers and primer dimers by separating the PCR products on a 1.5% agarose gel. Bacterial amplicons were excised from the gels using the QIAquick gel extraction kit (Qiagen Benelux N.V., Venlo, The Netherlands). Mitochondrial products produced by primers 799F-1391R and 799F-1193R of respectively 1000 and 800 bp were also eliminated via the gel purification (Figure S2). Following the first round of PCR amplification and gel-purification of the PCR products, a second round of PCR amplification was carried out for all seven primer pairs with the Roche 454 pyrosequencing adaptors and the sample specific barcodes. Amplicon length of sequences produced by primer pairs 799F-1391R and 68F-783Rabc was reduced by amplifying the samples with 967F-1391R and 68F-518R in the second round. PCR cycling conditions were identical as previously described, with the exception of the number of PCR cycles, which was lowered to 25.

Subsequently, quadruplicate PCR amplicon pools from the corresponding samples were pooled together to end up with 12 samples from four different compartments (rhizosphere, root, stem, leaf) of three biologically independent poplar individuals. PCR amplicon pools were purified to remove residual PCR primers and primer dimers using the QIAquick PCR purification kit (Qiagen Benelux B.V., Venlo, The Netherlands). Following purification, the quality of the amplicon pools was evaluated using an Agilent 2100 Bioanalyzer system (Agilent Technologies, Diegem, Belgium) according to the manufacturer's protocol. Finally, purified amplicon libraries were quantified with the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA, USA) and a Fluostar Omega plate reader (BMG Labtech, Ortenberg, Germany) and pooled in equimolar concentrations. The resulting seven amplicon pools (one for each primer pair), each of them containing 12 samples, were sequenced on one eighth of a Pico Titer Plate on a Roche Genome Sequencer FLX+ using Titanium chemistry (Roche Applied Science, Mannheim, Germany) by LGC Genomics (Berlin, Germany). Total amplicon pools consisted of 26 samples, including 14 samples related to another study.

## Sequence Processing

Sequencing generated seven individual Standard Flowgram Format (SFF) files, which were analyzed separately using the software package mothur (version 1.34.3) following the Standard Operating Procedure outlined in [http://www.mothur.org/wiki/Schloss\\_SOP](http://www.mothur.org/wiki/Schloss_SOP) (Schloss et al., 2009). Briefly, sequencing errors were reduced by denoising (shhh.flows, mothur implementation of Amplicon Noise algorithm) and quality trimming, which

removed reads shorter than 200 bases, reads with homopolymers longer than eight bases and reads containing ambiguous bases. Unique sequences were identified, whilst archiving the abundance data of the unique sequences, and aligned using `align.seqs` with the SILVA reference alignment (Release 119) (Pruesse et al., 2007). Within the unique sequences, chimeric sequences were identified using the Uchime tool (*de novo* chimera detection) (Edgar et al., 2011) followed by their removal from the dataset. Taxonomic classification of the sequences were done using a cut-off of 80%. Sequences matching “Chloroplast” and “Mitochondria” were identified using `classify.seqs` and abundance data of these sequences were used to compare the performance of all primer pairs (Table 2). Subsequently these sequences were removed from the data set. Finally, genus-level OTUs (Operational Taxonomic Unit) were defined based on a 97% sequence similarity level. Complete parametrical evaluation was conducted with primer pairs 799F-1391R, 799F-1193R, and 341F-783Rabc based on low co-amplification of non-target DNA and high retrieval of bacterial reads. Because these selected primer pairs resulted in differential amounts of reads per sample, the number of reads per sample were rarefied to 417 reads per sample. Samples, for which fewer than 417 reads were obtained, were removed from the data set. Only for primer pair 341F-783Rabc, we removed 3 samples, all belonging to the stem compartment. Rarefaction curves were assembled based on 10,000 permutations and intra-sample richness, diversity, and Good's coverage estimates which were calculated in `mothur` (version 1.34.3) based on 10,000 iterations.

The standard flowgram format (SFF) files were deposited in the NCBI Sequence Read Archive (SRA) under the Bioproject number PRJNA318176 and BioSample accession numbers SAMN04633889 to SAMN04633970.

## Isolation of Intact Chloroplasts to Extract Pure Chloroplast DNA

Intact chloroplasts were isolated from (*Populus tremula* × *P. alba*) leaves following a method described by Cortleven et al. (2011). Briefly, fresh leaves (~10 g) were harvested and homogenized in 100 mL ice-cold grinding buffer (2.0 mM NaEDTA; 1.0 mM MgCl<sub>2</sub>; 1.0 mM MnCl<sub>2</sub>; 50.0 mM Hepes/KOH, pH 7.5; 0.33 M sorbitol; 5.0 mM sodium ascorbate) using a Braun MX-32 mixer. The resulting homogenate was filtered through four layers of Miracloth (pore size: 22–25 μm) and centrifuged (1400 g, 5 min). The pellet was resuspended in 1 mL of grinding buffer whereafter the suspension was loaded on a continuous 10–80% Percoll gradient (3% PEG 6000; 1% Ficoll; 1% BSA) and centrifuged (8000 g, 20 min). Finally, intact chloroplasts were collected after centrifugation (lower band), washed twice with five volumes of grinding buffer and stored at –70°C until chloroplast DNA was extracted. DNA was extracted from intact chloroplasts using the Invisorb Spin Plant Mini Kit following the manufacturer's instructions.

## Quantitative Real-Time PCR

To evaluate the primer efficiency of the selected primer pairs amplifying pure chloroplast DNA (*Populus tremula* × *P. alba*), we tested all primer pairs in a qPCR set-up. From five chloroplast

DNA samples, we made a two-fold dilution series (1:2 up to 1:64). qPCR cycling conditions included: initial denaturation at 94°C for 3 min, followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 53°C for 1 min and extension at 72°C during 1 min; a final extension phase was performed at 72°C during 10 min. Finally, a dissociation curve was generated to verify amplification efficiency. Each reaction contained 2 μl of template DNA, 5 μl 2 × Fast SYBR Green Master Mix (Applied Biosystems, Foster City, CA, USA), 0.3 μl forward and reverse primer (0.3 μM each) and 2.4 μl nuclease-free water in a total volume of 10 μl. PCR efficiencies (*E*) were calculated as  $E = (10^{-1/\text{slope}} - 1) \times 100$ .

## Statistical Analysis

Statistical analyses were performed in R 2.15.1 (The R Foundation for Statistical Computing, Vienna, Austria). Normal distributions of the data were checked with the Shapiro–Wilk test and homoscedasticity of variances was analyzed using either Bartlett's or the Fligner–Killeens test. Significant differences in the variance of parameters were evaluated, depending on the distribution of the estimated parameters, either with ANOVA or the Kruskal Wallis Rank Sum Test. *Post hoc* comparisons were conducted by either the Tukey's Honest Significant Differences tests or Pairwise Wilcoxon Rank Sum tests. Poisson corrections were used for abundance data and distributions of ratios were compared with Pearson's Chi-squared tests. Statistical analysis of multivariate data was performed according to the recommendations of Anderson and Willis (2003). Non-metric multi-dimensional scaling (NMDS) was performed using the `Vegan 2.0–8` package in R (Oksanen et al., 2013) with 10,000 iterations. The OTU abundance data were square-root transformed (to downweight quantitatively abundant OTUs) and similarities in the bacterial community structures were displayed with non-metric multi-dimensional scaling (NMDS) with Bray–Curtis distances (Bray and Curtis, 1957). Differences between the *a priori* defined groups were evaluated with permutation-based hypothesis tests, namely analysis of similarities (ANOSIM), an analog of univariate ANOVA. Community richness estimators (number of OTUs, Chao1 estimator, ACE estimator, and Bootstrap) and community diversity estimators (Berger–Parker, Shannon, non-parametric Shannon, QStat, Simpson, and Inverse Simpson indices) were calculated in `Mothur` using 10,000 iterations.

## RESULTS AND DISCUSSION

After the *in silico* analyses, our final selection included primers 799F and 783Rabc (Chelius and Triplett, 2001; Sakai et al., 2004), containing several mismatches with the chloroplast DNA. Based on primer-to-target 3' mismatches, overall coverage and taking into account amplicon length, both primers were matched with two primers to produce four primer pairs, respectively 799F-1391R, 799F-1193R, 341F-783Rabc, and 68F-783Rabc. Further, we included primer pair 341F-785R, as described by Klindworth et al. (2013), to evaluate the performance of the mismatch primer sets with the ideal primer pair for 16S rDNA metabarcoding studies with 454 applications Primer sequences and full results of the *in silico* analysis are displayed in Table 1 and Table S1.

**TABLE 2 | Quality metrics of pyrosequencing analysis, co-amplification of non-target DNA, and amplification of bacterial rDNA reads.**

<b>A. Total reads</b>	<b>799F-1391R</b>	<b>967F-1391R</b>	<b>799F-1193R</b>	<b>341F-785R</b>	<b>68F-783Rabc</b>	<b>68F-518R</b>	<b>341F-783Rabc</b>
Rhizosphere soil	2235 ± 165	2550 ± 673	956 ± 285	1961 ± 119	3346 ± 454	1519 ± 217	2196 ± 317
Root	2728 ± 74	2577 ± 56	1943 ± 129	2916 ± 438	2484 ± 155	3488 ± 532	2548 ± 403
Stem	2811 ± 117	2502 ± 159	2456 ± 486	2199 ± 350	2068 ± 384	3412 ± 632	1386 ± 18
Leaf	2665 ± 100	2402 ± 231	2410 ± 197	2621 ± 134	1961 ± 64	3257 ± 367	1678 ± 81
Read length before QC	405 ± 96	401 ± 101	364 ± 105	392 ± 105	348 ± 139	349 ± 105	361 ± 129
Read length after QC	207 ± 4	208 ± 4	217 ± 5	233 ± 5	222 ± 5	200 ± 4	205 ± 4
<b>NORMALIZATION TO 1000 READS</b>							
<b>B. Chloroplast DNA</b>	<b>799F-1391R</b>	<b>967F-1391R</b>	<b>799F-1193R</b>	<b>341F-785R</b>	<b>68F-783Rabc</b>	<b>68F-518R</b>	<b>341F-783Rabc</b>
Rhizosphere soil	0 <sup>a</sup>	0.2 ± 0.3 (< 0.1) <sup>a</sup>	0 <sup>a</sup>	1 ± 2 (0.1) <sup>a</sup>	0 <sup>a</sup>	0 <sup>a</sup>	0.2 ± 0.3 (< 0.1) <sup>a</sup>
Root	0 <sup>a</sup>	786 ± 79 (79) <sup>b</sup>	0 <sup>a</sup>	863 ± 54 (86) <sup>b</sup>	736 ± 90 (74) <sup>b</sup>	975 ± 8 (97) <sup>c</sup>	270 ± 87 (26) <sup>d</sup>
Stem	2 ± 3 (0.2) <sup>a</sup>	997 ± 3 (99) <sup>b</sup>	0 <sup>a</sup>	962 ± 1 (96) <sup>b</sup>	993 ± 4 (99) <sup>b</sup>	998 ± 1 (99) <sup>b</sup>	804 ± 36 (80) <sup>c</sup>
Leaf	0 <sup>a</sup>	907 ± 35 (91) <sup>b</sup>	0 <sup>a</sup>	910 ± 29 (91) <sup>b</sup>	894 ± 12 (89) <sup>b</sup>	985 ± 4 (98) <sup>c</sup>	518 ± 71 (52) <sup>d</sup>
<b>C. Mitochondrial DNA</b>	<b>799F-1391R</b>	<b>967F-1391R</b>	<b>799F-1193R</b>	<b>341F-785R</b>	<b>68F-783Rabc</b>	<b>68F-518R</b>	<b>341F-783Rabc</b>
Rhizosphere soil	0 <sup>a</sup>	0 <sup>a</sup>	0.5 ± 0.5 (< 0.1) <sup>a</sup>	0 <sup>a</sup>	0 <sup>a</sup>	0 <sup>a</sup>	0 <sup>a</sup>
Root	0 <sup>a</sup>	0 <sup>a</sup>	9 ± 1 (1) <sup>b</sup>	45 ± 17 (5) <sup>c</sup>	15 ± 5 (1) <sup>b</sup>	4 ± 1 (0.5) <sup>b</sup>	136 ± 17 (14) <sup>d</sup>
Stem	0 <sup>a</sup>	0 <sup>a</sup>	19 ± 11 (2) <sup>b</sup>	35 ± 1 (4) <sup>b</sup>	6 ± 3 (0.5) <sup>a</sup>	1 ± 1 (0.1) <sup>a</sup>	173 ± 25 (17) <sup>c</sup>
Leaf	0 <sup>a</sup>	0 <sup>a</sup>	11 ± 2.5 (1) <sup>b</sup>	69 ± 16 (7) <sup>c</sup>	20 ± 13 (2) <sup>b</sup>	6 ± 3 (0.5) <sup>b</sup>	196 ± 53 (20) <sup>d</sup>
<b>D. Bacterial rDNA</b>	<b>799F-1391R</b>	<b>967F-1391R</b>	<b>799F-1193R</b>	<b>341F-785R</b>	<b>68F-783Rabc</b>	<b>68F-518R</b>	<b>341F-783Rabc</b>
Rhizosphere soil	1000 ± 0 (100) <sup>a</sup>	999 ± 0.26 (99) <sup>a</sup>	999 ± 0.3 (99) <sup>a</sup>	998 ± 3 (99) <sup>a</sup>	1000 ± 0 (100) <sup>a</sup>	1000 ± 0 (100) <sup>a</sup>	999 ± 0.52 (99) <sup>a</sup>
Root	1000 ± 0 (100) <sup>a</sup>	414 ± 79 (21) <sup>b</sup>	992 ± 1 (99) <sup>a</sup>	92 ± 41 (9) <sup>b</sup>	250 ± 88 (25) <sup>b</sup>	22 ± 7 (2) <sup>c</sup>	594 ± 72 (60) <sup>d</sup>
Stem	997 ± 3 (99) <sup>a</sup>	2 ± 3 (0.2) <sup>b</sup>	982 ± 11 (98) <sup>a</sup>	4 ± 2 (0.3) <sup>b</sup>	1 ± 1 (0.1) <sup>b</sup>	1 ± 2 (< 0.1) <sup>b</sup>	25 ± 12 (3) <sup>b</sup>
Leaf	1000 ± 0 (100) <sup>a</sup>	93 ± 35 (9) <sup>b</sup>	989 ± 3 (98) <sup>a</sup>	22 ± 15 (2) <sup>b</sup>	85 ± 37 (9) <sup>b</sup>	10 ± 6 (1) <sup>b</sup>	278 ± 25 (28) <sup>c</sup>

(A) Total number of reads ( $\pm$  standard deviation) obtained per plant compartment for each primer pair and average read length ( $\pm$  standard deviation) before and after quality control (QC). Average number of chloroplast (B) and mitochondrial (C) sequences (non-target DNA) obtained from each plant compartment by the selected primer pairs. (D) Amplification of bacterial rDNA reads. Values were normalized to 1000 reads and are averages of three biologically independent replicates  $\pm$  standard deviation. Values between brackets represent the average percentage (%) of reads. Sequence counts were statistically analyzed using a one-way ANOVA within each plant compartment to compare primer pairs. Differences at the 95% significance level are indicated with lower case letters ( $P < 0.05$ ).

## 454 Pyrosequencing

We analyzed 12 samples derived from four different plant compartments (rhizosphere, roots, stems, and leaves) of poplar (*Populus tremula*  $\times$  *P. alba*) with seven selected bacterial 16S rDNA primer pairs (Table 1). Sequencing of the amplicon libraries generated a total of 799,429 reads with an average of ( $\pm$  standard deviation) 114,204 ( $\pm$  12013) reads. Quality metrics are displayed in Table 2 with the total amount of reads obtained per plant compartment and the average read length before and after quality checking and trimming (Table 2A).

## Co-Amplification of Non-Target DNA

Firstly, we evaluated the co-amplification of non-target DNA, e.g., chloroplast and mitochondrial DNA, by all primer pairs in the different plant compartments after normalization to 1000 reads (Table 2B,C). In the rhizosphere, some of the selected primer pairs retrieved minute fractions of chloroplast (967F-1391R: <0.1%, 341F-785R: 0.1%, 341F-783Rabc: <0.1%) and mitochondrial (799F-1193R; <0.1%) sequences. Retrieval of plastid and mitochondrial DNA from rhizosphere soil samples is most likely attributable to trace amounts of decaying root,

stem or leaf tissue, and the presence of eukaryotic organisms (mitochondrial sequences) in the soil. Furthermore, rhizosphere soil is inevitably contaminated with live and dead root cap border cells which have been shown to remain alive after desquamation from the root corpus (Vermeer and McCully, 1982; Hawes et al., 2000; Bulgarelli et al., 2013).

In the plant compartments, we found significant differences in performance of the primer pairs for co-amplification of non-target DNA (Table 2B,C). As expected, interference of non-target DNA was significantly reduced by the primer pairs containing a forward or reverse primer with incorporated mismatches with chloroplast DNA, e.g., primers 799F and 783Rabc (only in combination with 341F). Primer pairs 799F-1391R and 799F-1193R completely eliminated the co-amplification of plastid DNA sequences in the root samples and plastid/chloroplast sequences in the stem and leaf samples. Indeed, primer 799F and variations thereof have been used to minimize chloroplast contamination in plant samples with varying success and mostly resulting in some co-amplification of chloroplast DNA (Sun et al., 2008; Sagaram et al., 2009; Redford et al., 2010; Trivedi et al., 2010; Bragina et al., 2012; Bulgarelli et al., 2012;

Bodenhausen et al., 2013; Santhanam et al., 2014; Schlaeppi et al., 2014). However, although primer 783Rabc (Sakai et al., 2004) displayed 3 mismatches with poplar DNA during *in silico* analyses (Table 1), in an experimental set-up it failed to efficiently eliminate chloroplast DNA amplification (Table 2). Primer pair combination 341F-783Rabc performed reasonably well in the root samples (26% of plastid DNA) but higher chloroplast content from stems and leaf samples resulted in significant co-amplification of chloroplast DNA (80% in the stems and 52% in the leaves) (Table 2). Although primer pair 341F-783Rabc performed better than the other primer pairs (without chloroplast mismatches) (Table 2) in reducing the co-amplification of chloroplast DNA, *in silico* analyses portrayed an incorrect image of the primer potential and endorse that the position of the mismatches are crucial for their effectiveness in PCR amplification (Ayyadevara et al., 2000; Klindworth et al., 2013; Lefever et al., 2013). Complete sequence counts for the chloroplast and mitochondrial sequences and statistical differences are presented in Table 2B,C.

Interestingly, we consistently retrieved more chloroplast sequences from the stem samples as compared to the leaf samples for all primer pairs (except 799F-1193R), although not statistically significant for all primer pairs. Although absolute chloroplast DNA concentration is clearly higher in the leaf samples (being the major photosynthetic organ) than the stem samples, the balance between endophytic bacterial DNA and chloroplast DNA seems to be a more crucial factor in the co-amplification of chloroplast DNA. Poplar stems are highly lignified and consist of a high proportion of dead cells (xylem vessels) (Boerjan et al., 2003; Vanholme et al., 2010) with low nutrient content (Siebrecht et al., 2003; Danielsen et al., 2013; Morhart et al., 2013) and therefore most likely harbor fewer total bacterial cells than the leaf samples thereby skewing the balance toward the chloroplast DNA.

Finally, we also calculated the number of bacterial 16S rDNA sequences for each primer pair within each plant compartment (Table 2D). These sequence counts of course correlate with the sequence counts of the retrieved non-target DNA from organellar source.

## Primer Efficiency for Pure Chloroplast DNA (Poplar)

To assess whether or not the observed differences in the amplification of chloroplast sequences in the plant-bacteria DNA extracts during 16S rDNA metabarcoding would correlate with PCR amplification efficiency of the selected primer pairs for pure chloroplast DNA, a qPCR experiment was conducted. To this end, we isolated pure chloroplast DNA from the intact chloroplasts (Cortleven et al., 2011) of 5 poplar leaf samples and amplified the pure chloroplast DNA with the selected primer pairs (Figure 1). We observed a strong correlation between the qPCR set-up and the pyrosequencing results. Primer pairs 799F-1391R and 799F-1193R showed very low affinity for pure poplar chloroplast DNA (efficiency of respectively, 9.2 and 17.4%) resulting in virtually no amplification of chloroplast DNA in the pyrosequencing set-up. Medium affinity for chloroplast

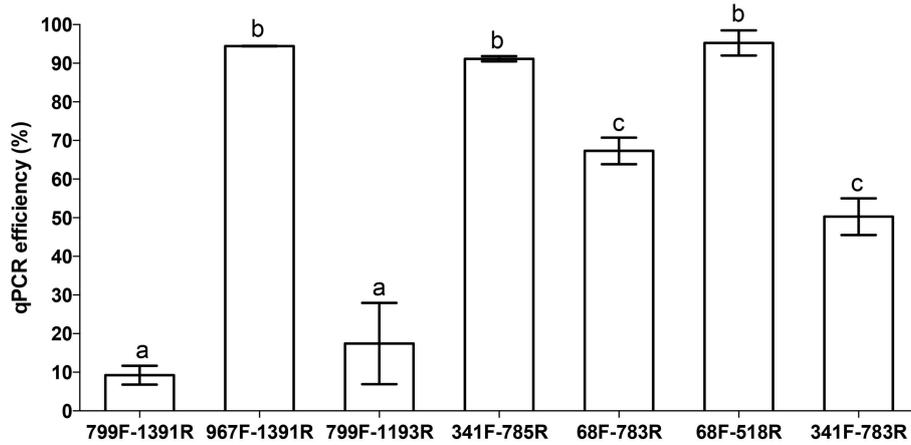
16S DNA was observed for primer pairs 68F-783Rabc (67.3%) and 341F-783Rabc (50.3%) resulting in differential amplification of chloroplast DNA (specifically for 341F-783Rabc) depending on the plastid/chloroplast content (root vs. stem and leaf) in the pyrosequencing set-up. The other primer pairs 967F-1391 (94.5%), 341F-785R (91.1%) and 68F-518R (95.3%) displayed very high PCR amplification efficiencies and indeed resulted in high co-amplification of chloroplast DNA during metabarcoding (Table 2B,C).

## Parametrical Comparison of Selected Primer Pairs

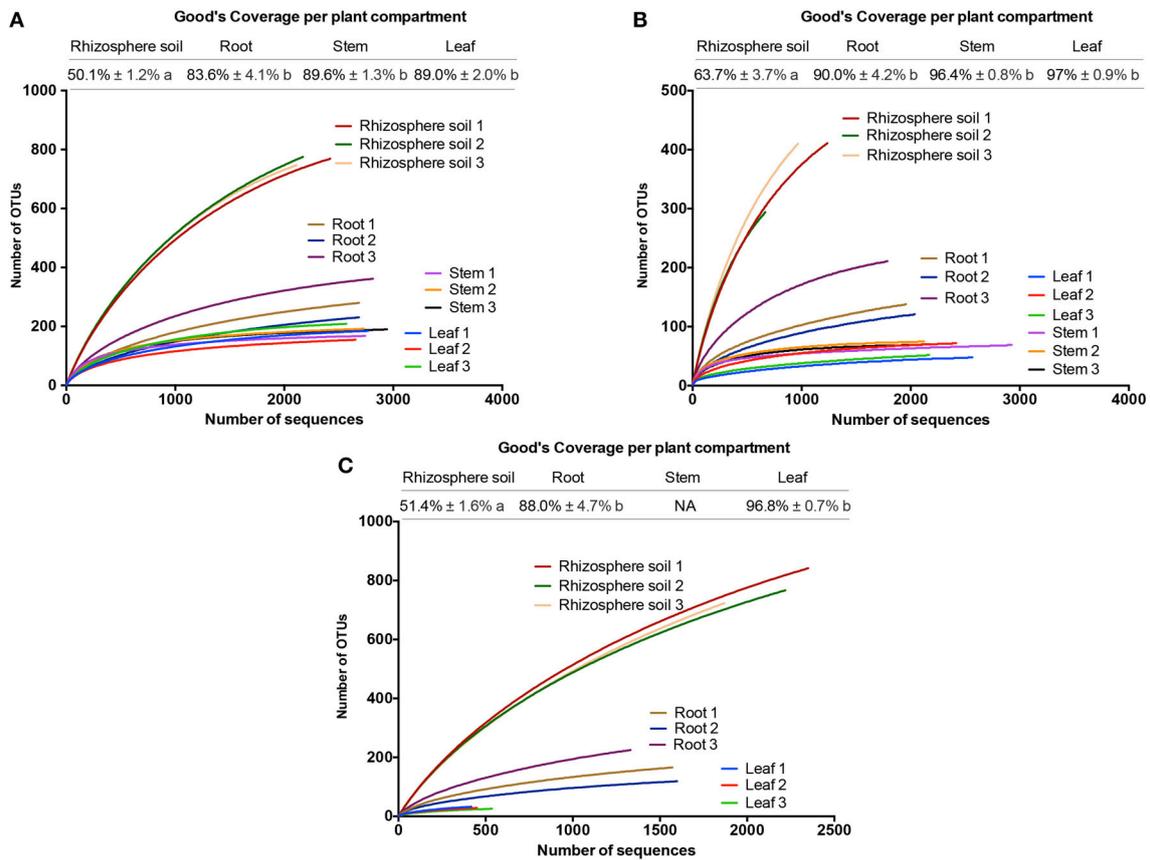
Based on low co-amplification levels of chloroplast and mitochondrial sequences and consequently high retrieval of bacterial rDNA reads (Table 2D), we selected primer pairs 799F-1391R, 799F-1193R, and 341F-783Rabc for further parametrical analysis (Figures 2, 3).

Rarefaction curves and Good's coverage estimates indicated that our sampling effort (regardless of primer pair) for the rhizospheric samples was inadequate (ranging from 50.1 to 63.7%) to fully capture the bacterial communities (Figure 2). Indeed, rarefaction curves from other studies using rhizosphere soil samples only tend toward saturation after 5000–6000 sequenced reads (Gottel et al., 2011). Observed low Good's coverage estimates were partly the result of our low level of subsampling (417 sequences) used to include as much samples as possible in the parametrical comparison of the primer pairs. Higher subsampling levels (1500 sequences) revealed significantly higher Good's coverage estimates in the rhizosphere soil samples for all primer pairs (ranging from 76 to 91%; data not shown). However, the comparison of the selected primer pairs revealed the highest OTU richness for primer pairs 799F-1391R and 341F-783Rabc, which yielded on average the same amount of OTUs ( $P = 0.80$ ) with 277 OTUs (min = 271; max = 281) and 270 OTUs (min = 265; max = 276), per sample respectively. For primer pair 799F-1193R, OTU richness was significantly lower ( $P < 0.01$ ) with on average 236 OTUs (min = 227; max = 252) per sample (Figure 3A). Inverse Simpson diversity estimates were comparable for all primers (Figure 3A).

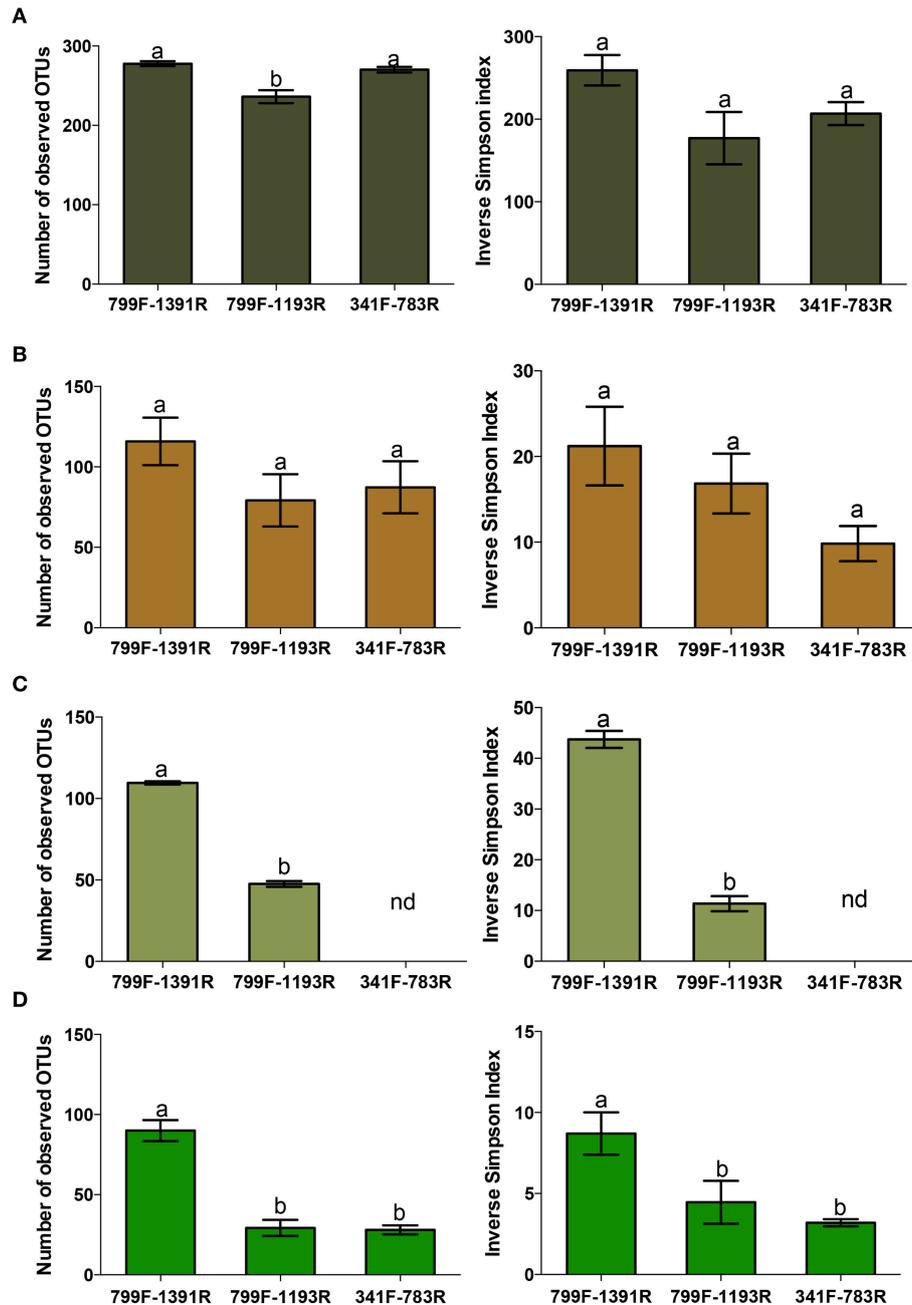
For the root, stem and leaf samples, rarefaction curves of all primer pairs tended toward saturation demonstrating that the sequencing effort was sufficient to obtain the most abundant bacterial OTUs. Good's coverage estimates ranged from 83.6% to up to 97% (Figure 2). In the roots, the comparison of the selected primer pairs again revealed the highest OTU richness for primer pair 799F-1391R which obtained on average 115 OTUs (min = 93; max = 143) per sample. This in comparison with primer set 799F-1193R which retrieved 79 OTUs (min = 58; max = 111) per sample and primer pair 341F-783Rabc which yielded 87 OTUs (min = 62; max = 117) (Figure 3B). Gottel et al. (2011) reported high variability in the OTU retrieval isolated from the roots of poplar trees in mature, natural ecosystems (83 OTUs per sample  $\pm$  78). However, important to mention, is that we rarefied our samples to 417 sequences per sample, thereby reducing the amounts of OTUs retrieved. Indeed much higher OTUs richness counts were observed in



**FIGURE 1 | Average PCR amplification efficiency of selected 16S rDNA primers for pure chloroplast DNA (poplar) using quantitative real-time PCR.** Values are averages of five biologically independent replicates  $\pm$  standard error. PCR efficiencies were compared using a one-way ANOVA. Differences at the 95% significance level are indicated with lower case letters ( $P < 0.05$ ).



**FIGURE 2 | Good's coverage estimates and rarefaction curves of the different replicates from each plant compartment (rhizosphere soil, root, stem, and leaf) for each primer pair including (A) 799F-1391R, (B) 799F-1193R, and (C) 341F-783Rabc.** Good's coverage estimates were calculated in mothur based on 10,000 iterations. Differences at the 95% significance level between the plant compartments are indicated with lower case letters ( $P < 0.05$ ). Rarefaction curves were assembled showing the number of observed OTUs, defined at a 97% sequence similarity cut-off, relative to the total number of identified bacterial rDNA sequences. To calculate the community estimators, samples were rarefied to 417 reads. NA = Not available due to low retrieval of bacterial rDNA reads.



**FIGURE 3 | Comparison of parametrical alpha diversity between selected 16S rDNA primer pairs (799F-1391R, 799F-1193R, and 341F-783Rabc) for all sampled plant compartments [(A) Rhizosphere soil, (B) Root, (C) Stem, (D) Leaf] after subsampling to 417 reads.** All averages were calculated across three biologically independent poplar individuals for each primer pair. Left panels: average number of operational taxonomic units (OTUs) observed based on a 97% sequence similarity cutoff (richness) and right panels: Inverse Simpson diversity indices. OTU counts and Inverse Simpson indices were statistically analyzed using a one-way ANOVA per plant compartment. Differences at the 95% significance level are indicated with lower case letters ( $P < 0.05$ ). nd = not determined as a result of low bacterial rDNA reads.

the roots and leaves of *Arabidopsis thaliana* (Bulgarelli et al., 2012; Lundberg et al., 2012; Bodenhausen et al., 2013). Inverse Simpson diversity estimates displayed a clear trend toward higher diversity estimates in primer 799F-1391R as compared to primer pairs 799F-1193R and 341F-783Rabc although

not statistically significant for the rhizosphere soil and root samples (Figure 3B).

Finally, in the stem and leaves (Figures 3C,D), we consistently observed the highest OTU richness for primer pair 799F-1391R ( $P < 0.01$ ). In the stem (Figure 3C), primer pair 799F-1391R

retrieved on average 109 OTUs (min = 107; max = 110). In contrast, primer set 799F-1193R only obtained 47 OTUs (min = 45; max = 51) per sample in the stems. For primer pair 341F-783Rabc, OTU richness could not be determined in the stem samples (nd) as a consequence of very low amplification of bacterial rDNA reads (Table 2 and Figure 3C). In the leaves (Figure 3D), primer pair 799F-1391R yielded on average 90 OTUs (min = 80; max = 102) whereas primer sets 799F-1193R and 341F-783Rabc retrieved significantly less ( $P < 0.01$ ) OTUs with respectively, 29 OTUs (min = 22; max = 39) and 28 OTUs (min = 23; max = 33) per sample. For the Inverse Simpson index in the stems and leaves, primer pair 799F-1391R consistently showed a higher diversity as compared to primer pairs 799F-1193R and 341F-783Rabc ( $P < 0.01$ ). Diversity indices could not be calculated for primer pair 341F-783Rabc in the stem samples due to low amplification of bacterial reads.

To exclude bias in the community richness and diversity estimators, we included several other alternative estimators (Table S2). Most estimators re-enforce the results obtained from the number of observed OTUs and the inverse Simpson index in Figure 3 and resulted in similar trends, although not all statistically significant.

## Community Similarity between Primer Pairs and Plant Compartments

To compare the bacterial communities retrieved by each selected primer pair (799F-1391R, 799F-1193R, and 341F-783Rabc) at the phylum and genus level, relative frequency distributions of the obtained genus-level OTUs and phyla were analyzed with chi-squared tests for the three primer pairs, based on average abundances across replicate samples. At the phylum and genus-level, differences were observed ( $P < 0.05$ ) for all three primer pairs within all plant compartments. Further, we compared phylum and genus-level OTU abundances of samples for each primer pair within every plant compartment using non-metric multi-dimensional scaling (NMDS) with Bray-Curtis dissimilarities (Figure S3). The observed dissimilarities in the recovery of OTUs, according to the selected primer pair, at phylum-level (Figure S3a, ANOSIM:  $P < 0.05$ ) and genus-level (Figure S3b, ANOSIM:  $P < 0.05$ ) demonstrate the major bias primer selection can introduce in metabarcoding studies (Aird et al., 2011; Berry et al., 2011; Pinto and Raskin, 2012; Ghyselinck et al., 2013; Klindworth et al., 2013; Op De Beeck et al., 2014; Tremblay et al., 2015).

Finally, we compared the bacterial communities (NMDS) retrieved per plant compartment within each primer pair (Figure S4). In general, irrespective of the primer pair, significantly different OTUs (ANOSIM:  $P < 0.05$ ) were detected in each plant compartment at the phylum level (Figure S4a) and genus-level (Figure S4b) clearly illustrating the specific niche differentiation of plant-associated bacteria. Lower intravariability of the samples (mainly at the phylum level) was observed for primer pair 341F-783R resulting in slightly lower stress values (and slightly higher significance levels) of the NMDS fit as compared to the other primer pairs. However, primer pair 341F-783R also displayed significantly lower OTU richness and diversity index estimates (Figure 3) indicating that lower intravariability of the samples is instigated by low (and less diverse) retrieval of OTUs.

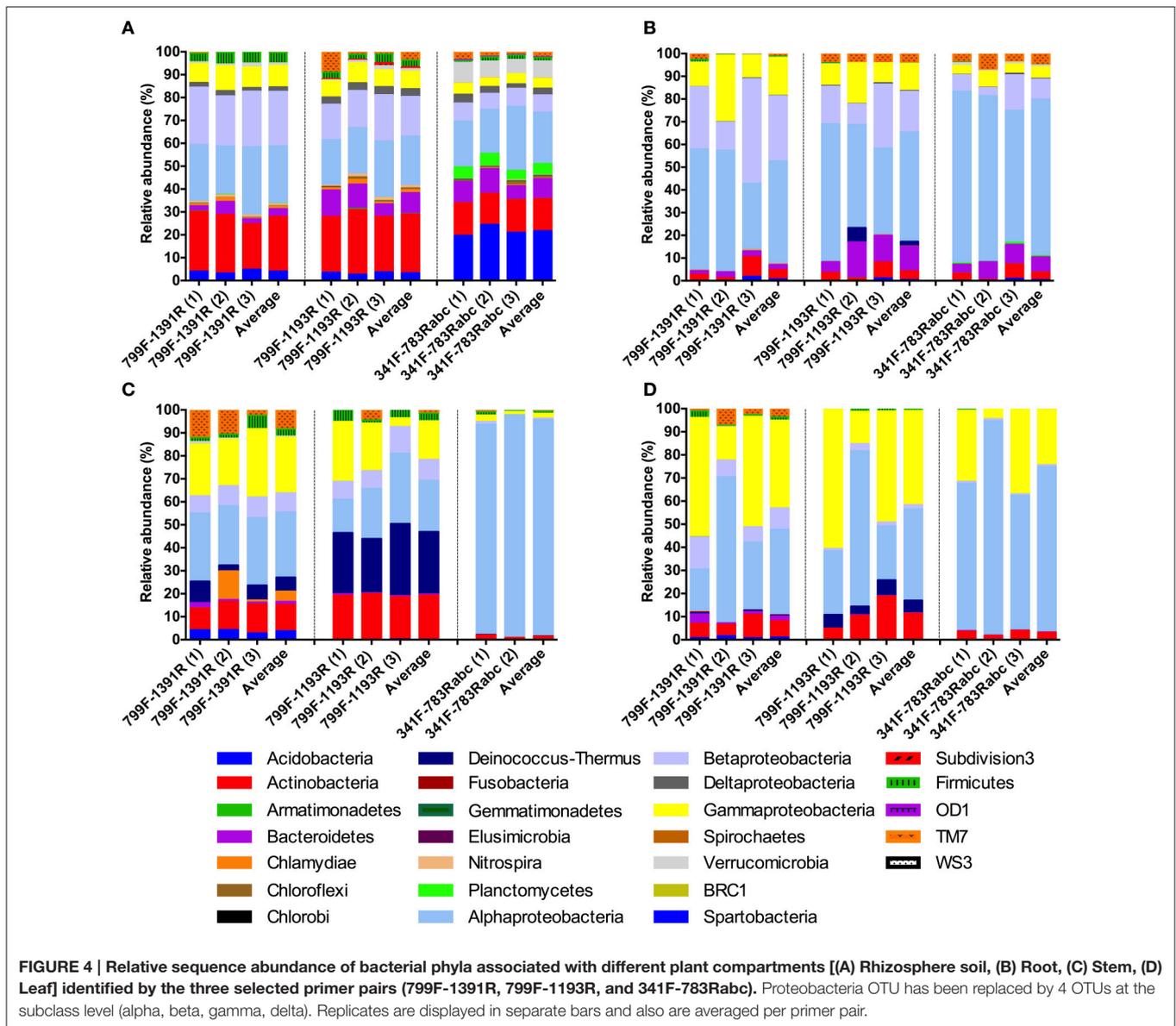
We previously observed the same niche differentiation in the isolation of cultivable bacteria from poplar trees in the same field study (Beckers et al., 2016). Niche differentiation between the rhizosphere and root endophyte microbiome has been described for mature poplar trees growing in natural ecosystems (*Populus deltoides*) (Gottel et al., 2011; Shakya et al., 2013), for *Arabidopsis thaliana* ecotypes (Bulgarelli et al., 2012; Lundberg et al., 2012; Schlaeppi et al., 2014; Bai et al., 2015) and other plant species (Inceoglu et al., 2010; Weinert et al., 2011; Ofek-Lalzar et al., 2014; Edwards et al., 2015). Recently, Bulgarelli et al. (2013) proposed a two-step selection model for root microbiota differentiation from the rhizosphere where rhizodeposition and host genotype-dependent fine-tuning converge to select specific endophytic assemblages. Although we used a limited amount of biological replicates (3), our data indicate additional fine-tuning and niche differentiation of the microbiota in the aerial plant organs, with the stem and leaf bacterial communities being remarkably dissimilar from the root and rhizosphere (Figure S4). This somewhat in contrast with the findings of Bai et al. (2015) who reported an extensive overlap in taxonomy and genome-encoded functional competences between the leaf and root microbiota of *Arabidopsis thaliana*. Although, some of their evidence from recolonization experiments also pointed to microbiome habitat speciation to their respective ecological niche (Bai et al., 2015).

## Core Microbiome Members Identified by Each Primer Pair

Finally, we also obtained a first look at the rhizospheric and endophytic bacterial communities associated with the different plant compartments of poplar trees (*Populus tremula* × *P. alba*) in the field trial under investigation.

Remarkably, we found a clear difference in the numbers of reads that could not be unambiguously classified at the phylum level in the V6-V7 region (799F-1391R:  $46\% \pm 0.5$ , 799F-1193R:  $49\% \pm 2.6$ ) as compared to the V3-V4 region (341F-783Rabc:  $21\% \pm 0.5$ ) ( $P < 0.05$ ) in the rhizosphere samples. This is indicative of an insufficient database representation of the biodiversity of soil-borne bacteria and an underrepresentation of the hypervariable V6-V7 region (Gans et al., 2005; Bulgarelli et al., 2012). Indeed, V3-V4 has been the preferred region for next-generation studies (Klindworth et al., 2013). Therefore, at least for the time being, for the study of plant-associated bacteria a trade-off enforces itself to choose for the V6-V7, with the availability of primers to avoid co-amplification of organellar DNA but with an underrepresentation of sequences in this region in the databases.

At the phylum level (Figure 4), the majority of the OTUs in the rhizosphere soil (Figure 4A) identified by all primer pairs were assigned to Proteobacteria (37 to 60%), Actinobacteria (14 to 25%), Acidobacteria (3 to 21%), and Bacteroidetes (3 to 9%). In the roots (Figure 4B), for all primer pairs, we observed a strong dominance of Proteobacteria (78 to 91%) with a minority of the identified OTUs belonging to Bacteroidetes (2 to 11%), Actinobacteria (2 to 4%), and TM7 (1 to 4%). In the stems (Figure 4C), the dominance of phylum Proteobacteria (48 to 97%) persisted although it was slightly less pronounced for the analyses on primer pairs 799F-1391R (61%) and 799F-1193R



(48%) compared to primer pair 341F-783Rabc (97%). Most likely, this observation is directly related to the very reduced number of bacterial sequences obtained from the stem samples by primer pair 341F-783Rabc. For the analyses based on primer pairs 799F-1391R and 799F-1193R, the rest of the identified OTUs mainly belonged to Actinobacteria (respectively 11 and 19%) and Deinococcus-Thermus (respectively 6 and 27%). Finally in the leaves, the majority of the OTUs were also identified as Proteobacteria (82 to 96%). A minority of the OTUs found in the leaves belonged to Actinobacteria lineages (3 to 11%) (Figure 4D).

Taking a closer look at the phylum level, in the rhizosphere we predominantly identified *Proteobacteria*, *Acidobacteria*, and *Actinobacteria*, irrespective of the selected primer pair. The ratio of *Proteobacteria* to *Acidobacteria* in rhizosphere bacterial communities has previously been shown to be an indicator of soil nutrient content where *Proteobacteria* were linked to

nutrient-rich soils and *Acidobacteria* to nutrient-poor soils (Smit et al., 2001; Castro et al., 2010; Gottel et al., 2011). Endophytic communities were, for the most part dominated by *Proteobacteria* suggesting substantial overlap in key community members across host species (Gottel et al., 2011; Bulgarelli et al., 2012; Lundberg et al., 2012; Bodenhausen et al., 2013; Shakya et al., 2013; Romero et al., 2014).

To give an idea of the bacterial communities at the genus level, we defined the core bacterial community, described by each primer pair, as the 10 most abundant genus-level OTUs per compartment. This resulted in 21 OTUs for the rhizosphere, 18 OTUs for the roots, and 23 OTUs for the stem and leaf samples. The percentage of sequences represented by these 10 most abundant OTUs for each primer pair and plant compartment are listed in Tables S3–S6. For all plant compartments, and in particular the rhizosphere soil, we observed long-tailed rank-abundance curves characteristic of

microbial communities (Hartmann et al., 2012). Worth noting here is that for all compartments, except for the leaves, the most abundantly identified OTU was different for each primer pair. This demonstrates the major effect of primer choice on the observed bacterial communities. In the rhizosphere, *Actinomycetales* (9.9%, 799F-1391R), *Rhizobiales* (11.3%, 799F-1193R), and *Acidobacteria\_Gp6* (15.7, 341F-783Rabc) were the dominant OTUs. In the roots, *Pseudomonas* (11.9%, 799F-1391R), *Rhizobium* (15.9%, 799F-1193R), and *Rhizobiales* (38.9%, 341F-783Rabc) constituted the major identified OTUs. In the stems, *Pseudomonas* (12.9%, 799F-1391R), *Deinococcaceae* (18.2%, 799F-1193R), and *Sphingomonadaceae* (34.2%, 341F-783Rabc) were the most observed OTUs. Finally, in the leaves, *Pseudomonas* dominated bacterial communities regardless of the primer pair used.

At genus-level, most remarkable is the efficiency of the endophytic colonization of the OTU *Pseudomonas* as indicated by all studied primer pairs (Tables S3–S6). A low relative abundance of *Pseudomonas* in the rhizosphere soil (0.3 to 1.9%) is contrasted by its dominance in the endosphere samples (3.67 to 40.13%). Similarly, comparative studies revealed high relative abundance of a *Pseudomonas*-like OTU in the root microbiome of *Populus deltoides* (Gottel et al., 2011), although a new study revealed similar dominance in the root endosphere by a *Streptomyces*-like OTU (Shakya et al., 2013). In contrast, root and leaf microbiota of *Arabidopsis* genotypes appear to be not heavily populated by *Pseudomonas*-like OTUs (Bulgarelli et al., 2012; Lundberg et al., 2012; Schlaeppi et al., 2014; Bai et al., 2015). Endophytic colonization of *Pseudomonas* may occur via the rhizosphere where most endophytic bacteria should originate from and/or via leaf stomatal colonization since aerosol samples were found to harbor abundant *Pseudomonas* sequences (Fahlgren et al., 2010). Finally, important to mention with the interpretation of community studies with next generation sequencing platforms is the 16S rRNA operon copy number. This copy number may vary, depending on species, from 1 to 15, introduces significant bias and distort views on bacterial communities (Crosby and Criddle, 2003; Lee et al., 2009). Furthermore, to fully evaluate the reliability and robustness of the primers concerning bias against and/or in favor of specific taxonomic groups and community richness and diversity estimators, the use of mock communities would provide more insight into their experimental behavior during metabarcoding studies (Caporaso et al., 2011; Pinto and Raskin, 2012).

## CONCLUDING REMARKS

We experimentally evaluated the performance of seven 16S rDNA primer pairs in 16S rDNA metabarcoding studies of endophytic and rhizospheric bacterial communities. Our results show that different primer pairs display different efficiencies in the elimination of non-target DNA. In this study, the primer pair 799F-1391R, which amplifies the V5-V7 hypervariable regions of the 16S rDNA gene, displayed very low amplification of non-target DNA across all sampled plant compartments. And retrieved the highest number of OTUs

as well as exhibited the highest Inverse Simpson diversity, especially in the plant compartments with high chloroplast content (stem and leaf samples). Therefore, we propose primer pair 799F-1391R as best suited for 16S rDNA metabarcoding studies which simultaneously investigate rhizosphere and endosphere microbiomes. Specifically, in experimental setups where direct comparisons are made between different endosphere microbiomes such as the microbiomes of wild-type plants and genetically modified plants. However, to perform an in-depth characterization of the rhizosphere and/or endosphere microbiomes and uncover the true bacterial diversity, the use of multiple primer pairs is highly advisable.

Recently, the use of other 16S rDNA metabarcoding applications such as the HiSeq2000, MiSeq Illumina and Ion Torrent platforms have come to the foreground (Claesson et al., 2010; Metzker, 2010; Caporaso et al., 2012; Logares et al., 2013; Kennedy et al., 2014). However, in plant-microbiota research, the application of these platforms has been mainly limited, up till now, to a couple of studies of the rhizosphere microbiome (Jiang et al., 2013; Panke-Buisse et al., 2014; Sun et al., 2014) and root and leaf microbiota (Ofek-Lalzar et al., 2014; Bai et al., 2015; Coleman-Derr et al., 2016). Evaluating the potential of our optimized approach, with platform-specific modifications (e.g., amplicon length), in combination with HiSeq2000 and MiSeq Illumina could further contribute to the high-resolution 16S rDNA-based community profiling of plant-associated bacterial communities.

## AUTHOR CONTRIBUTIONS

Conceptualization: BB, JV, and WB; Methodology: BB, MODB, SoT, and SaT; Investigation: BB; Writing—original draft: BB; Writing—Review and editing: MODB, NW, WB, and JV; Funding Acquisition: JV and WB.

## FUNDING

This work was funded by the Fund for Scientific Research Flanders (FWO-Vlaanderen), project number G032912N, Ph.D. grants for Michiel Op De Beeck, Sofie Thijs and Sascha Truyens and a post-doc grant for Nele Weyens. Furthermore, this work has been financially supported by the UHasselt Methusalem project 08M03VGRJ.

## DATA ACCESSIBILITY

The standard flowgram format (SFF) files were deposited in the NCBI Sequence Read Archive (SRA) under the Bioproject number PRJNA318176 and BioSample accession numbers SAMN04633889 to SAMN04633970.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00650>

## REFERENCES

- Aird, D., Ross, M. G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., et al. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12:R18. doi: 10.1186/gb-2011-12-2-r18
- Anderson, M. J., and Willis, T. J. (2003). Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* 84, 511–525. doi: 10.1890/0012-9658(2003)084[0511:CAOPCA]2.0.CO;2
- Andersson, A. F., Lindberg, M., Jakobsson, H., Bäckhed, F., Nyrén, P., and Engstrand, L. (2008). Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PLoS ONE* 3:e2836. doi: 10.1371/journal.pone.0002836
- Andrew, D. R., Fitak, R. R., Munguia-Vega, A., Racolta, A., Martinson, V. G., and Dontsova, K. (2012). Abiotic factors shape microbial diversity in Sonoran desert soils. *Appl. Environ. Microbiol.* 78, 7527–7537. doi: 10.1128/AEM.01459-12
- Ayyadevara, S., Thaden, J. J., and Shmookler Reis, R. J. (2000). Discrimination of primer 3'-nucleotide mismatch by taq DNA polymerase during polymerase chain reaction. *Anal. Biochem.* 284, 11–18. doi: 10.1006/abio.2000.4635
- Bai, Y., Muller, D. B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., et al. (2015). Functional overlap of the *Arabidopsis thaliana* leaf and root microbiota. *Nature* 528, 364–369. doi: 10.1038/nature16192
- Balzer, S., Malde, K., and Jonassen, I. (2011). Systematic exploration of error sources in pyrosequencing flowgram data. *Bioinformatics* 27, i304–i309. doi: 10.1093/bioinformatics/btr251
- Beckers, B., Op De Beeck, M., Weyens, N., Van Acker, R., Van Montagu, M., Boerjan, W., et al. (2016). Lignin engineering in field-grown poplar trees affects the endosphere bacterial microbiome. *Proc. Natl. Acad. Sci. U.S.A.* 113, 2312–2317. doi: 10.1073/pnas.1523264113
- Berry, D., Ben Mahfoudh, K., Wagner, M., and Loy, A. (2011). Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Appl. Environ. Microbiol.* 77, 7846–7849. doi: 10.1128/AEM.05220-11
- Bodenhausen, N., Horton, M. W., and Bergelson, J. (2013). Bacterial communities associated with the leaves and the roots of *Arabidopsis thaliana*. *PLoS ONE* 8:e56329. doi: 10.1371/journal.pone.0056329
- Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. *Annu. Rev. Plant Biol.* 54, 519–546. doi: 10.1146/annurev.arplant.54.031902.134938
- Bragina, A., Berg, C., Cardinale, M., Shcherbakov, A., Chebotar, V., and Berg, G. (2012). Sphagnum mosses harbour highly specific bacterial diversity during their whole lifecycle. *ISME J.* 6, 802–813. doi: 10.1038/ismej.2011.151
- Bray, J. R., and Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* 27, 325–349.
- Bulgarelli, D., Rott, M., Schlaeppli, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F., et al. (2012). Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature* 488, 91–95. doi: 10.1038/nature11336
- Bulgarelli, D., Schlaeppli, K., Spaepen, S., Ver Loren van Themaat, E., and Schulze-Lefert, P. (2013). Structure and functions of the bacterial microbiota of plants. *Annu. Rev. Plant Biol.* 64, 807–838. doi: 10.1146/annurev-arplant-050312-120106
- 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. doi: 10.1038/ismej.2012.8
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4516–4522. doi: 10.1073/pnas.1000080107
- Castro, H. F., Classen, A. T., Austin, E. E., Norby, R. J., and Schadt, C. W. (2010). Soil microbial community responses to multiple experimental climate change drivers. *Appl. Environ. Microbiol.* 76, 999–1007. doi: 10.1128/AEM.02874-09
- Chelius, M. K., and Triplett, E. W. (2001). The diversity of archaea and bacteria in association with the roots of *Zea mays* L. *Microb. Ecol.* 41, 252–263. doi: 10.1007/s002480000087
- Claesson, M. J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J. R., Ross, R. P., et al. (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* 38, e200. doi: 10.1093/nar/kq873
- Coleman-Derr, D., Desgarenes, D., Fonseca-Garcia, C., Gross, S., Clingenpeel, S., Woyke, T., et al. (2016). Plant compartment and biogeography affect microbiome composition in cultivated and native *Agave* species. *New Phytol.* 209, 798–811. doi: 10.1111/nph.13697
- Cortleven, A., Noben, J. P., and Valcke, R. (2011). Analysis of the photosynthetic apparatus in transgenic tobacco plants with altered endogenous cytokinin content: a proteomic study. *Proteome Sci.* 9, 1–14. doi: 10.1186/1477-5956-9-33
- Crosby, L. D., and Criddle, C. S. (2003). Understanding bias in microbial community analysis techniques due to rrn operon copy number heterogeneity. *Biotechniques* 34, 790–794.
- Custers, R. (2009). First GM trial in Belgium since 2002. *Nat. Biotechnol.* 27:506. doi: 10.1038/nbt0609-506
- Danielsen, L., Lohaus, G., Sirrenberg, A., Karlovsky, P., Bastien, C., Pilate, G., et al. (2013). Ectomycorrhizal colonization and diversity in relation to tree biomass and nutrition in a plantation of transgenic poplars with modified lignin biosynthesis. *PLoS ONE* 8:e59207. doi: 10.1371/journal.pone.0059207
- DeLeon-Rodriguez, N., Latham, T. L., Rodriguez, R. L. M., Barazesh, J. M., Anderson, B. E., Beyersdorf, A. J., et al. (2013). Microbiome of the upper troposphere: species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2575–2580. doi: 10.1073/pnas.1212089110
- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi: 10.1128/AEM.03006-05
- Dyall, S. D., Brown, M. T., and Johnson, P. J. (2004). Ancient invasions: from endosymbionts to organelles. *Science* 304, 253–257. doi: 10.1126/science.1094884
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381
- Edwards, J., Johnson, C., Santos-Medellin, C., Lurie, E., Podishetty, N. K., Bhatnagar, S., et al. (2015). Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc. Natl. Acad. Sci. U.S.A.* 112, E911–E920. doi: 10.1073/pnas.1414592112
- Fahlgren, C., Hagström, A., Nilsson, D., and Zweifel, U. L. (2010). Annual variations in the diversity, viability, and origin of airborne bacteria. *Appl. Environ. Microbiol.* 76, 3015–3025. doi: 10.1128/AEM.02092-09
- Feehery, G. R., Yigit, E., Oyola, S. O., Langhorst, B. W., Schmidt, V. T., Stewart, F. J., et al. (2013). A Method for selectively enriching microbial DNA from contaminating vertebrate host DNA Highlander, SK (ed). *PLoS ONE* 8:e76096. doi: 10.1371/journal.pone.0076096
- Feinstein, L. M., Sul, W. J., and Blackwood, C. B. (2009). Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Appl. Environ. Microbiol.* 75, 5428–5433. doi: 10.1128/AEM.00120-09
- Gans, J., Wolinsky, M., and Dunbar, J. (2005). Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* 309, 1387–1390. doi: 10.1126/science.1112665
- Ghyselinck, J., Pfeiffer, S., Heylen, K., Sessitsch, A., and De Vos, P. (2013). The effect of primer choice and short read sequences on the outcome of 16S rRNA gene based diversity studies Ravel, J (ed). *PLoS ONE* 8:e71360. doi: 10.1371/journal.pone.0071360
- Gilles, A., Meglécz, E., Pech, N., Ferreira, S., Malausa, T., and Martin, J.-F. (2011). Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* 12:245. doi: 10.1186/1471-2164-12-245
- Gottel, N. R., Castro, H. F., Kerley, M., Yang, Z., Pelletier, D. A., Podar, M., et al. (2011). Distinct microbial communities within the Endosphere and Rhizosphere of *Populus deltoides* roots across contrasting soil types. *Appl. Environ. Microbiol.* 77, 5934–5944. doi: 10.1128/AEM.05255-11
- Green, S. J., and Minz, D. (2005). Suicide polymerase endonuclease restriction, a novel technique for enhancing PCR amplification of minor DNA templates. *Appl. Environ. Microbiol.* 71, 4721–4727. doi: 10.1128/AEM.71.8.4721-4727.2005
- Hamdan, L. J., Coffin, R. B., Sikaroodi, M., Greinert, J., Treude, T., and Gillevet, P. M. (2013). Ocean currents shape the microbiome of Arctic marine sediments. *ISME J.* 7, 685–696. doi: 10.1038/ismej.2012.143

- Hansen, A. K., and Moran, N. A. (2014). The impact of microbial symbionts on host plant utilization by herbivorous insects. *Mol. Ecol.* 23, 1473–1496. doi: 10.1111/mec.12421
- Hartmann, M., Howes, C. G., VanInsberghe, D., Yu, H., Bachar, D., Christen, R., et al. (2012). Significant and persistent impact of timber harvesting on soil microbial communities in Northern coniferous forests. *ISME J.* 6, 2199–2218. doi: 10.1038/ismej.2012.84
- Hawes, M. C., Gunawardena, U., Miyasaka, S., and Zhao, X. (2000). The role of root border cells in plant defense. *Trends Plant Sci.* 5, 128–133. doi: 10.1016/S1360-1385(00)01556-9
- Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. doi: 10.1038/nature11234
- Inceoglu, O., Salles, J. F., van Overbeek, L., and van Elsas, J. D. (2010). Effects of plant genotype and growth stage on the betaproteobacterial communities associated with different potato cultivars in two fields. *Appl. Environ. Microbiol.* 76, 3675–3684. doi: 10.1128/AEM.00040-10
- Jiang, X. T., Peng, X., Deng, G. H., Sheng, H. F., Wang, Y., Zhou, H. W., et al. (2013). Illumina sequencing of 16S rRNA tag revealed spatial variations of bacterial communities in a mangrove wetland. *Microb. Ecol.* 66, 96–104. doi: 10.1007/s00248-013-0238-8
- Kalendar, R., Lee, D., and Schulman, A. H. (2014). “FastPCR software for PCR, in silico PCR, and oligonucleotide assembly and analysis,” in *DNA Cloning and Assembly Methods, Methods in Molecular Biology*, eds S. Valla and R. Lale (Humana Press), 271–302.
- Kennedy, K., Hall, M. W., Lynch, M. D. J., Moreno-Hagelsieb, G., and Neufeld, J. D. (2014). Evaluating bias of illumina-based bacterial 16S rRNA gene profiles. *Appl. Environ. Microbiol.* 80, 5717–5722. doi: 10.1128/AEM.01451-14
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41, 1–11. doi: 10.1093/nar/gks808
- Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* 12, 118–123. doi: 10.1111/j.1462-2920.2009.02051.x
- Lee, T. K., Van Doan, T., Yoo, K., Choi, S., Kim, C., Park, J., et al. (2010). Discovery of commonly existing anode biofilm microbes in two different wastewater treatment MFCs using FLX Titanium pyrosequencing. *Appl. Microbiol. Biotechnol.* 87, 2335–2343. doi: 10.1007/s00253-010-2680-6
- Lee, Z. M. P., Bussema, C., and Schmidt, T. M. (2009). rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Res.* 37, 489–493. doi: 10.1093/nar/gkn689
- Lefever, S., Pattyn, F., Hellems, J., and Vandesompele, J. (2013). Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays. *Clin. Chem.* 59, 1470–1480. doi: 10.1373/clinchem.2013.203653
- Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F. M., Ferrera, I., Sarmento, H., et al. (2013). Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ. Microbiol.* 16, 2659–2671. doi: 10.1111/1462-2920.12250
- Loy, A., Maixner, F., Wagner, M., and Horn, M. (2007). probeBase—an online resource for rRNA-targeted oligonucleotide probes: new features 2007. *Nucleic Acids Res.* 35, D800–D804. doi: 10.1093/nar/gkl856
- Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., et al. (2012). Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 488, 86–90. doi: 10.1038/nature11237
- Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D., and Dangl, J. L. (2013). Practical innovations for high-throughput amplicon sequencing. *Nat. Methods* 10, 999–1002. doi: 10.1038/nmeth.2634
- Luo, C., Tsementzi, D., Kyripides, N., Read, T., and Konstantinidis, K. T. (2012). Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample Rodriguez-Valera, F (ed). *PLoS ONE* 7:e30087. doi: 10.1371/journal.pone.0030087
- Lutz, K. A., Wang, W., Zdepski, A., and Michael, T. P. (2011). Isolation and analysis of high quality nuclear DNA with reduced organellar DNA for plant genome sequencing and resequencing. *BMC Biotechnol.* 11:54. doi: 10.1186/1472-6750-11-54
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380. doi: 10.1038/nature03959
- McAllister, S. M., Davis, R. E., McBeth, J. M., Tebo, B. M., Emerson, D., and Moyer, C. L. (2011). Biodiversity and emerging biogeography of the neutrophilic iron-oxidizing zetaproteobacteria. *Appl. Environ. Microbiol.* 77, 5445–5457. doi: 10.1128/AEM.00533-11
- Mendes, R., Kruijt, M., de Bruijn, I., Dekkers, E., van der Voort, M., Schneider, J. H. M., et al. (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* 332, 1097–1100. doi: 10.1126/science.1203980
- Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat. Rev. Genet.* 11, 31–46. doi: 10.1038/nrg2626
- Morhart, C., Sheppard, J., and Spiecker, H. (2013). Above ground leafless woody biomass and nutrient content within different compartments of a *P. maximowiczii* × *P. trichocarpa* poplar clone. *Forests* 4, 471–487. doi: 10.3390/f4020471
- Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., et al. (2011). Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res.* 39, e90. doi: 10.1093/nar/gkr344
- Nelson, M. C., Morrison, H. G., Benjamino, J., Grim, S. L., and Graf, J. (2014). Analysis, optimization and verification of illumina-generated 16S rRNA gene amplicon surveys Heimesaat, MM (ed). *PLoS ONE* 9:e94249. doi: 10.1371/journal.pone.0094249
- Ofek-Lazar, M., Sela, N., Goldman-Voronov, M., Green, S. J., Hadar, Y., and Minz, D. (2014). Niche and host-associated functional signatures of the root surface microbiome. *Nat. Commun.* 5, 4950. doi: 10.1038/ncomms5950
- Oksanen, J., Blanchet, G. F., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., et al. (2013). *vegan: Community Ecology Package. R package version 2.15.1*. Available online at: <http://CRAN.R-project.org/package=vegan>
- Op De Beek, M., Lievens, B., Busschaert, P., Declerck, S., Vangronsveld, J., and Colpaert, J. V. (2014). Comparison and validation of some ITS primer pairs useful for fungal metabarcoding studies. *PLoS ONE* 9:e97629. doi: 10.1371/journal.pone.0097629
- Oyola, S. O., Gu, Y., Manske, M., Otto, T. D., O'Brien, J., Alcock, D., et al. (2013). Efficient depletion of host DNA contamination in malaria clinical sequencing. *J. Clin. Microbiol.* 51, 745–751. doi: 10.1371/journal.pone.0076096
- Panke-Buisse, K., Poole, A. C., Goodrich, J. K., Ley, R. E., and Kao-Kniffin, J. (2014). Selection on soil microbiomes reveals reproducible impacts on plant function. *ISME J.* 9, 980–989. doi: 10.1038/ismej.2014.196
- Parameswaran, P., Jalili, R., Tao, L., Shokralla, S., Gharizadeh, B., Ronaghi, M., et al. (2007). A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res.* 35, e130. doi: 10.1093/nar/gkm760
- Peiffer, J. A., Spor, A., Koren, O., Jin, Z., Tringe, S. G., Dangl, J. L., et al. (2013). Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6548–6553. doi: 10.1073/pnas.1302837110
- Pinto, A. J., and Raskin, L. (2012). PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets Bertilsson, S (ed). *PLoS ONE* 7:e43093. doi: 10.1371/journal.pone.0043093
- Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman, S. N., and Taberlet, P. (2012). Who is eating what: diet assessment using next generation sequencing. *Mol. Ecol.* 21, 1931–1950. doi: 10.1111/j.1365-294X.2011.05403.x
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196. doi: 10.1093/nar/gkm864
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65. doi: 10.1038/nature08821
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219
- Rastogi, G., Tech, J. J., Coaker, G. L., and Leveau, J. H. J. (2010). A PCR-based toolbox for the culture-independent quantification of total bacterial

- abundances in plant environments. *J. Microb. Methods* 83, 127–132. doi: 10.1016/j.mimet.2010.08.006
- Raven, P. H. (1970). A multiple origin for plastids and mitochondria: many independent symbiotic events may have been involved in the origin of these cellular organelles. *Science* 169, 641–646. doi: 10.1126/science.169.3946.641
- Redford, A. J., Bowers, R. M., Knight, R., Linhart, Y., and Fierer, N. (2010). The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environ. Microbiol.* 12, 2885–2893. doi: 10.1111/j.1462-2920.2010.02258.x
- Romero, F. M., Marina, M., and Pieckenstein, F. L. (2014). The communities of tomato (*Solanum lycopersicum* L.) leaf endophytic bacteria, analyzed by 16S-ribosomal RNA gene pyrosequencing. *FEMS Microbiol. Lett.* 351, 187–194. doi: 10.1111/1574-6968.12377
- Sagaram, U. S., DeAngelis, K. M., Trivedi, P., Andersen, G. L., Lu, S. E., and Wang, N. (2009). Bacterial diversity analysis of Huanglongbing pathogen-infected citrus, using PhyloChip arrays and 16S rRNA gene clone library sequencing. *Appl. Environ. Microbiol.* 75, 1566–1574. doi: 10.1128/AEM.02404-08
- Sakai, M., Matsuka, A., Komura, T., and Kanazawa, S. (2004). Application of a new PCR primer for terminal restriction fragment length polymorphism analysis of the bacterial communities in plant roots. *J. Microbiol. Methods* 59, 81–89. doi: 10.1016/j.mimet.2004.06.005
- Santhanam, R., Groten, K., Meldau, D. G., and Baldwin, I. T. (2014). Analysis of plant-bacteria interactions in their native habitat: bacterial communities associated with wild tobacco are independent of endogenous jasmonic acid levels and developmental stages Simon van Overbeek, L (ed). *PLoS ONE* 9:e94710. doi: 10.1371/journal.pone.0094710
- Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T., and Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* 43, e37. doi: 10.1093/nar/gku1341
- Schlaeppli, K., Dombrowski, N., Oter, R. G., Ver Loren van Themaat, E., and Schulze-Lefert, P. (2014). Quantitative divergence of the bacterial root microbiota in *Arabidopsis thaliana* relatives. *Proc. Natl. Acad. Sci. U.S.A.* 111, 585–592. doi: 10.1073/pnas.1321597111
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09
- Shade, A., McManus, P. S., and Handelsman, J. (2013). Unexpected diversity during community succession in the apple flower microbiome. *MBio* 4, 1–12. doi: 10.1128/mBio.00602-12.Editor
- Shakya, M., Gottel, N., Castro, H., Yang, Z. K., Gunter, L., Labbé J., et al. (2013). A multifactor analysis of fungal and bacterial community structure in the root microbiome of mature *Populus deltoides* trees. *PLoS ONE* 8:e76382. doi: 10.1371/journal.pone.0076382
- Shaver, J. M., Oldenburg, D. J., and Bendich, A. J. (2006). Changes in chloroplast DNA during development in tobacco, *Medicago truncatula*, pea, and maize. *Planta* 224, 72–82. doi: 10.1007/s00425-005-0195-7
- Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. *Nat. Biotechnol.* 26, 1135–1145. doi: 10.1038/nbt1486
- Siebrecht, S., Herdel, K., Schurr, U., and Tischner, R. (2003). Nutrient translocation in the xylem of poplar-Diurnal variations and spatial distribution along the shoot axis. *Planta* 217, 783–793. doi: 10.1007/s00425-003-1041-4
- Smit, E., Leeflang, P., Gommans, S., van Mil, S., and Wernars, K. (2001). Diversity and seasonal fluctuations of the dominant members of the bacterial soil community in a wheat field as determined by cultivation and molecular methods. *Appl. Environ. Microbiol.* 67, 2284–2291. doi: 10.1128/AEM.67.5.2284
- Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., Huse, S. M., Neal, P. R., et al. (2006). Microbial diversity in the deep sea and the underexplored “rare biosphere.” *Proc. Natl. Acad. Sci. U.S.A.* 103, 12115–12120. doi: 10.1073/pnas.0605127103
- Strong, M. J., Xu, G., Morici, L., Splinter Bon-Durant, S., Baddoo, M., Lin, Z., et al. (2014). Microbial contamination in next generation sequencing: implications for sequence-based analysis of clinical samples Rall, GF (ed). *PLoS Pathog.* 10:e1004437. doi: 10.1371/journal.ppat.1004437
- Sudakaran, S., Salem, H., Kost, C., and Kaltenpoth, M. (2012). Geographical and ecological stability of the symbiotic mid-gut microbiota in European firebugs, *Pyrrhocoris apterus* (Hemiptera, Pyrrhocoridae). *Mol. Ecol.* 21, 6134–6151. doi: 10.1111/mec.12027
- Sun, J., Zhang, Q., Zhou, J., and Wei, Q. (2014). Illumina amplicon sequencing of 16S rRNA tag reveals bacterial community development in the rhizosphere of apple nurseries at a replant disease site and a new planting site. *PLoS ONE* 9:e111744. doi: 10.1371/journal.pone.0111744
- Sun, L., Qiu, F., Zhang, X., Dai, X., Dong, X., and Song, W. (2008). Endophytic bacterial diversity in rice (*Oryza sativa* L.) roots estimated by 16S rDNA sequence analysis. *Microb. Ecol.* 55, 415–424. doi: 10.1007/s00248-007-9287-1
- Tremblay, J., Singh, K., Fern, A., Kirton, E. S., He, S., Woyke, T., et al. (2015). Primer and platform effects on 16S rRNA tag sequencing. *Front. Microbiol.* 6:771. doi: 10.3389/fmicb.2015.00771
- Trivedi, P., Duan, Y., and Wang, N. (2010). Huanglongbing, a systemic disease, restructures the bacterial community associated with citrus roots. *Appl. Environ. Microbiol.* 76, 3427–3436. doi: 10.1128/AEM.02901-09
- Van Acker, R., Leplé, J. C., Aerts, D., Storme, V., Goeminne, G., Ivens, B., et al. (2014). Improved saccharification and ethanol yield from field-grown transgenic poplar deficient in cinnamoyl-CoA reductase. *Proc. Natl. Acad. Sci. U.S.A.* 111, 845–850. doi: 10.1073/pnas.1321673111
- Vanholme, R., Demedts, B., Morreel, K., Ralph, J., and Boerjan, W. (2010). Lignin biosynthesis and structure. *Plant Physiol.* 153, 895–905. doi: 10.1104/pp.110.155119
- Vermeer, J., and McCully, M. E. (1982). The rhizosphere in *Zea*: new insight into its structure and development. *Planta* 156, 45–61. doi: 10.1007/BF00393442
- Walker, J. J., and Pace, N. R. (2007). Phylogenetic composition of rocky mountain endolithic microbial ecosystems. *Appl. Environ. Microbiol.* 73, 3497–3504. doi: 10.1128/AEM.02656-06
- Walters, W. A., Caporaso, J. G., Lauber, C. L., Berg-Lyons, D., Fierer, N., and Knight R. (2011). PrimerProspector: *de novo* design and taxonomic analysis of PCR primers. *Bioinformatics* 27, 1159–1161. doi: 10.1093/bioinformatics/btr087
- Weinert, N., Piceno, Y., Ding, G.-C., Meincke, R., Heuer, H., Berg, G., et al. (2011). PhyloChip hybridization uncovered an enormous bacterial diversity in the rhizosphere of different potato cultivars: many common and few cultivar-dependent taxa. *FEMS Microbiol. Ecol.* 75, 497–506. doi: 10.1111/j.1574-6941.2010.01025.x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Beckers, Op De Beeck, Thijs, Truyens, Weyens, Boerjan and Vangronsveld. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.