

TECHNICAL ADVANCE

e-Xtra*

Validation of a PNA Clamping Method for Reducing Host DNA Amplification and Increasing Eukaryotic Diversity in Rhizosphere Microbiome Studies

Stephen J. Taerum,¹ Blaire Steven,² Daniel J. Gage,³ and Lindsay R. Triplett^{1,†}

¹ Department of Plant Pathology and Ecology, The Connecticut Agricultural Experiment Station, New Haven, CT 06511

² Department of Environmental Science, The Connecticut Agricultural Experiment Station, New Haven, CT 06511

³ Department of Molecular and Cell Biology, The University of Connecticut, Storrs, CT 06269

Accepted for publication 4 September 2020.

ABSTRACT

Protists and microscopic animals are important but poorly understood determinants of plant health. Plant-associated eukaryotes could be surveyed by high-throughput sequencing of 18S ribosomal RNA (rRNA) genes but the abundance of plant DNA in rhizosphere samples makes 18S rRNA gene amplification with universal primers unfeasible. Here, we applied a pipeline to generate peptide nucleic acid (PNA) clamps to suppress the amplification of maize host DNA during 18S rRNA gene library preparation. PNA clamps targeting the V4 and V9 hypervariable regions of the 18S rRNA gene of maize were designed and evaluated *in silico*, and the performance of the V9 targeting clamp PoacV9_01 was evaluated *in vitro*. PoacV9_01 suppressed the

amplification of five crop species in quantitative PCR assays. In an 18S rRNA gene sequencing survey of the rhizosphere of maize, PoacV9_01 reduced the relative abundance of plant reads from 65 to 0.6%, while drastically increasing the number and diversity of animal, fungal, and protist reads detected. Thus, PoacV9_01 can be used to facilitate the study of eukaryotes present in grass phytobiomes. In addition, the pipeline developed here can be used to develop PNA clamps that target other plant species.

Keywords: agriculture, maize, metagenomics, microbiome, microorganism, phyllosphere, phytobiome, PNA clamp, Poaceae, protist, rhizosphere, soil ecology, *Zea mays*

Phytobiomes are multikingdom communities that significantly affect the health, growth, and survival of plants (Huang et al. 2014). Plants host diverse communities of bacteria, archaea, fungi, animals, and viruses, as well as autotrophic and heterotrophic protists (Leach et al. 2017). However, sequence-based taxonomic exploration of the phytobiome has focused mostly on its bacterial and fungal components, leaving an underexplored realm of microbial “dark matter” consisting of protists,

animals, and viruses. The diversity missing from many studies limits our understanding of microbial food webs and their effects on plant health.

Eukaryotic microbes are abundant in the rhizosphere, or the zone of soil in direct contact with plant roots, and are an important component of the phytobiome (Berendsen et al. 2012; Mendes et al. 2013). In the rhizosphere, plants actively produce exudates that recruit a diverse community of microorganisms (Berendsen et al. 2012; Compant et al. 2019; Mendes et al. 2013), including non-fungal eukaryotes. A single gram of soil is estimated to contain up to 1 million protists and 100 nematodes, while each square meter can harbor thousands of microscopic arthropods (Leach et al. 2017). The best-understood of these eukaryotes are economically important plant pests such as parasitic nematodes and *Phytophthora* spp.; however, commensal eukaryotes are also highly influential to ecosystem structure and plant health. Heterotrophic protists and nematodes may shape rhizosphere communities through selective or generalist predation on bacteria and fungi (Geisen et al. 2015, 2016; Jousset et al. 2009), mineralizing nutrients and enhancing the survival and dispersal of beneficial microbes in the process (Gao et al. 2019b; Rubinstein et al. 2015). Bacterivorous eukaryotes serve as influential hubs in soil microbial networks (Jiang et al. 2017; Xiong et al. 2017), affecting plant biomass, nutritional status, and disease outcomes (Trap et al. 2016; Xiong et al. 2020).

[†]Corresponding author: L. R. Triplett; Lindsay.Triplett@ct.gov

Funding: This work was supported by an AFRI Foundational Program grant from the United States Department of Agriculture–National Institute of Food and Agriculture (USDA-NIFA) to L. R. Triplett, D. J. Gage, and B. Steven (grant number 2019-67019-29315). Salary support for D. J. Gage and funding for the collection of some preliminary data were supported through a USDA-NIFA grant to D. J. Gage (grant number 2016-67013-24412). K. Manning and C. Calderon provided technical and field assistance, and were supported through an AFRI Education and Learning Initiative grant from USDA-NIFA (grant number 2017-67032-26013).

*The e-Xtra logo stands for “electronic extra” and indicates that three supplementary figures, six supplementary tables, and supplementary materials are published online.

The author(s) declare no conflict of interest.

© 2020 The American Phytopathological Society

Predation may also enhance the beneficial activities of mycorrhizae and plant-growth-promoting rhizobacteria (Henkes et al. 2018; Weidner et al. 2017), or directly inhibit pathogen growth (Long et al. 2018). Soil arthropods are less studied in the phytobiome but

are thought to play critical roles in soil nutrient cycling (Lavelle et al. 2006; Parmelee 1995).

High-throughput amplicon sequencing of 16S or 18S ribosomal RNA (rRNA) genes is the most widely adopted approach for measuring the bacterial and fungal diversity of the phytobiome. However, there are very few studies profiling nonfungal eukaryotic communities that live in direct contact with plants such as the rhizosphere. Several factors prevent the broader inclusion of protist and animal taxa in amplicon sequencing studies. For one, many protists and microscopic animals are still uncharacterized on a taxonomic level, although availability of reference sequences for these taxa has increased greatly in recent years. Another complication is that protists comprise many high-order clades across the eukaryote Tree of Life (Adl et al. 2019; Burki et al. 2020), encompassing far greater taxonomic diversity than fungi or plants alone. Such diversity hampers the ability to design “universal” protist primers. Group-specific primers have been useful for targeted studies of major protist taxa associated with plants (Geisen et al. 2019; Ploch et al. 2016) but these are generally only capable of amplifying a fraction of the eukaryotic portion of the phytobiome.

Several sets of broad-spectrum or universal barcoding primers have been applied to profile diverse eukaryote communities in soil, including those in heavily vegetated and agricultural soils (De Gruyter et al. 2020; Geisen et al. 2019; Mahé et al. 2017). Although these tools have proven highly valuable to measure soil protist diversity, few broad-spectrum studies have focused specifically on the rhizosphere soil compartment, which may contain high concentrations of plant DNA from root hairs, border cells, and secreted mucilage. Plant-derived amplification products are more likely to dominate rhizosphere sequencing results, potentially limiting sequencing depth, the number of rare taxa that can be detected, and the number of samples that can be analyzed simultaneously (Geisen et al. 2019).

One strategy to limit host plant sequence recovery is to block amplification of host DNA using peptide nucleic acid (PNA) clamps, or synthetic oligonucleotides that bind to host targets in a thermally stable manner to block DNA elongation (Karkare and Bhatnagar 2006). This approach has been successfully applied to suppress plastid and mitochondria amplification in studies of the plant bacterial microbiome (Lundberg et al. 2013; Steven et al. 2018), although specialized PNA clamps are needed for some plant species with plastids that differ in the target site (Fitzpatrick et al. 2018). PNA clamps have also been applied to block amplification of host 18S rRNA gene sequences from animals, facilitating surveys of their eukaryotic prey and symbionts (Belda et al. 2017; Terahara et al. 2011). This suggests that amplification of plant host 18S rRNA gene products could also be blocked by PNA clamping. Recently, a study reported the use of a plant-suppressing PNA clamp during sequencing of soil eukaryotes but no studies of its efficacy, breadth, or potential bias were reported (De Gruyter et al. 2020). The true utility of a PNA clamping approach will depend on the degree of host interference that must be overcome, and whether broad-spectrum plant blocking could be achieved without introducing sequence bias.

In this study, we evaluated whether a plant host-suppressing PNA clamp could facilitate surveys of the eukaryotic communities in the phytobiome. A comparative pipeline was applied to identify plant-specific target sequences within universal 18S rRNA gene amplicons, and to design PNA clamps that targeted multiple plant species. A PNA clamp specific to the family Poaceae was selected for validation, and it suppressed amplification of the 18S rRNA gene in maize, rice, wheat, barley, and sorghum. In a sequencing study of rhizosphere soil samples from field-grown maize, addition of the PNA clamp drastically reduced the presence of plant reads and

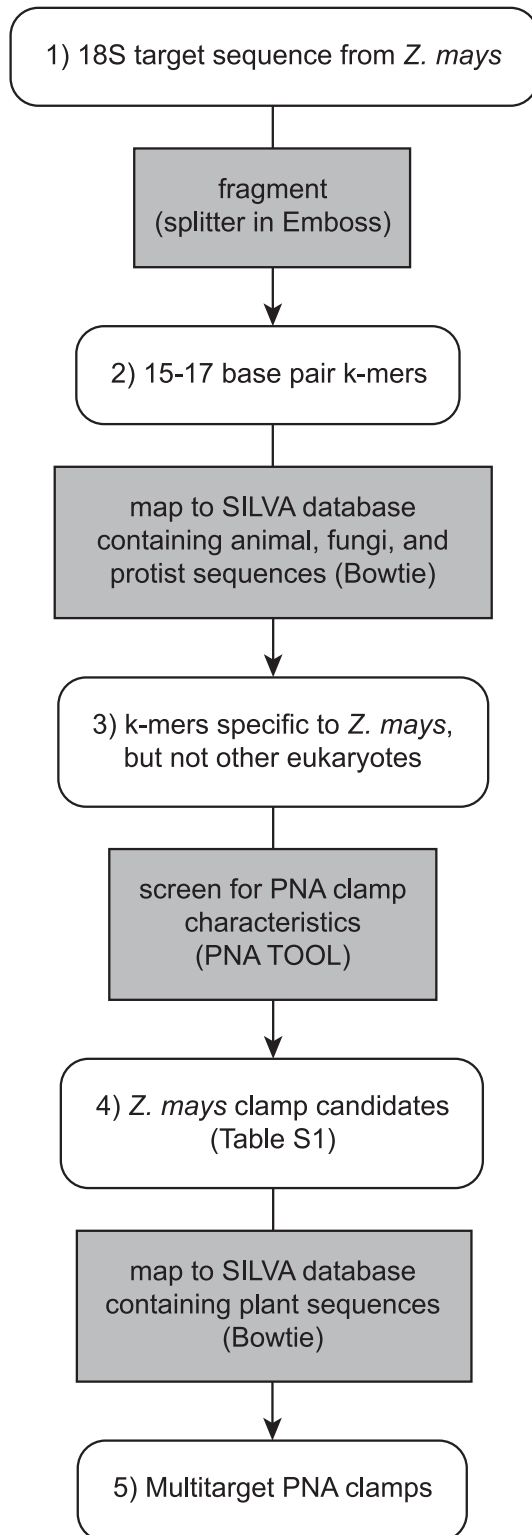


Fig. 1. Flowchart depicting the computational pipeline utilized in this study, which can be adapted for future peptide nucleic acid (PNA) clamp design. Gray square boxes denote processes, rounded white boxes indicate inputs and outputs, and *Z. mays* = *Zea mays*.

substantially increased the measured number and diversity of nonhost eukaryotic sequence variants, including those of fungi, heterotrophic protists, algae, nematodes, and arthropods. The PNA clamp did not introduce any measurable compositional bias. This study demonstrates that the use of a PNA clamp is a highly promising approach for characterizing the eukaryotic phytobiome.

MATERIALS AND METHODS

PNA clamp development. PNA clamps were developed to target the V4 and V9 hypervariable regions of the 18S rRNA gene, which are the most commonly utilized markers for high-throughput studies on eukaryotes. The clamps were designed following a protocol modified from Belda et al. (2017) and Lundberg et al. (2013). The PNA design pipeline is shown in Figure 1, and specific command-line arguments are provided in Supplementary File S1.

A 1,801-bp sequence of the 18S rRNA gene from maize (*Zea mays*) was downloaded from the online SILVA rRNA gene database (accession number LPUQ01000139) (Quast et al. 2013). The V4 and V9 regions were isolated after alignment in MEGA v. 7.0.26 (Kumar et al. 2016). Target regions were fragmented in silico into 15-, 16-, and 17-bp k-mers using the splitter application in EMBOSS (Rice et al. 2000), which was implemented in Jembooss (Carver and Bleasby 2003). Fragments were mapped onto an index of nonplant eukaryote 18S rRNA gene sequences available on the SILVA SSU r132 database using Bowtie v. 1.2.3 (Langmead et al. 2009). The number of times a k-mer matched a sequence in the index was recorded, allowing up to one mismatch in the alignments. Any k-mers that aligned with one or more sequences from the index were removed from consideration.

Candidate k-mers were screened using the online PNA TOOL (https://www.pnabio.com/support/PNA_Tool.htm). Clamps were designed to consist of fewer than 35% guanines and fewer than 50% purines, and a temperature between 76 and 82°C. K-mers that passed screening were mapped to an index of all 26,564 Streptophyta (plant) sequences from the SILVA SSU r132 database, a curated collection of 18S genes from all major plant lineages, using Bowtie. Two clamp sequences (k-mers) were selected for further testing (see Results): PoacV4_01 (V4 region, sequence 5'-TCGGTCTCGCCGTGA-3') and PoacV9_01 (V9 region, sequence 5'-GCCGCCCGGACGTC-3'). The PNA clamps were synthesized at PNA Bio, Inc. (Newbury Park, CA, U.S.A.), and resuspended in nuclease-free water to a concentration of 100 μ M.

Plant lines, protist isolates, and growth conditions. Seed of maize B73, rice (*Oryza sativa* L. 'Kitaake'), wheat (*Triticum aestivum* L. 'Byrd'), barley (*Hordeum vulgare* L. 'Morex'), sorghum (*Sorghum bicolor* L. 'BT623'), *Arabidopsis thaliana* ecotype Col-0, and *Nicotiana benthamiana* were germinated in wetted Turface (Turface Athletics, Buffalo Grove, IL, U.S.A.) in a greenhouse (28°C and 14-h days). After 1 week, the stems and leaves of the seedlings were collected.

Protist DNA was isolated from two laboratory lines (Daniel Gage Lab, University of Connecticut) isolated from soybean rhizospheres: UC1 (*Colpoda* sp.) was isolated in Mansfield, CT, in 2013, while UC5 (*Cercomonas* sp.) was isolated in Columbia, CT, in 2015. These species were selected because they are commonly associated with plant rhizospheres, easily culturable, and represent two distinct lineages of protists: *Colpoda* is a genus in the supergroup Alveolata, while *Cercomonas* is a genus in the supergroup Rhizaria. Both cultures were initiated from single cells into soil extract medium (20 mg of KH_2PO_4 , 20 mg of $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 200 mg of KNO_3 , and 100 ml of soil extract per liter; Culture Collection of Algae and Protozoa, Dunstaffnage Marine Laboratory, U.K.). Cultures were maintained using heat-killed *Escherichia coli* (DH5 α , optical density at 600 nm = 0.005) as a food source.

Initial PCR tests of clamp function. DNA was extracted from plant tissue using the GeneJET Plant Genomic DNA Purification kit (Thermo Scientific, Waltham, MA, U.S.A.). Protist DNA was extracted using the GeneJET Genomic DNA Purification kit (Thermo Scientific). DNA concentrations were measured using the Qubit dsDNA HS kit (Thermo Scientific), after which the samples were each diluted to 1 ng/ μ l. An initial set of PCR assays was conducted on the plant and UC1 templates. The V4 region was amplified using the primers TAREuk454FWD1 (5'-CCAGCAS CYGCGGTAATTCC-3') and TAREukREV3 (5'-ACTTTCGTTCTT GATYRA-3') (Stoeck et al. 2010), while the V9 region was amplified with the Earth Microbiome Project primers Euk1391F (5'-GTACA CACCGCCCGTC-3') (Lane 1991) and EukBr (5'-TGATCCTCTG CAGGTTACCTAC-3') (Medlin et al. 1988). These are primer pairs that were designed to universally amplify eukaryote DNA. Reactions consisted of 1 U of Invitrogen Taq DNA polymerase (Thermo Scientific), 1 $\times$ PCR buffer, 0.5 μ M forward and reverse primers, 0.2 mM dNTPs, 1.5 mM MgCl_2 , and template at 0.2 ng/ μ l in a 25- μ l reaction volume. Clamps PoacV4_01 or PoacV9_01 were added to reach the following concentrations: 0, 1.5, 7.5, and 15 μ M. Cycling conditions were 3 min at 94°C; followed by 30 cycles of 94°C for 45 s, 55°C for 30 s, and 72°C for 90 s; followed by a final extension of 72°C for 10 min. No PNA annealing step was used. The PCR amplicons were separated on a 1% agarose gel and visualized after staining with ethidium bromide.

Quantitative PCR test of PNA clamps. Quantitative PCR (qPCR) assays of the V4 and V9 regions were conducted using the SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, U.S.A.) and the same primers used for PCR assays. Reactions consisted of 1 $\times$ Supermix, 0.5 μ M forward and reverse primers, and template at 0.2 ng/ μ l in a 10- μ l reaction volume. For initial qPCR tests, the PNA clamps were added to a final concentration of 1.5 μ M. Cycling conditions for amplifying V4 were 2 min at 95°C followed by 40 cycles of 95°C for 10 s and 60°C for 60 s, while cycling conditions for V9 were 2 min at 95°C followed by 40 cycles of 95°C for 10 s and 60°C for 15 s. Reactions were conducted on a CFX96 Touch Real-Time PCR machine (Bio-Rad).

Further studies focused on optimizing and testing the V9 clamp, PoacV9_01. To determine the optimal concentration of this clamp, qPCR assays were conducted as described above, except that PoacV9_01 was added to reach the following concentrations: 0, 0.75, 1.5, 3.75, and 7.5 μ M. Each PNA concentration was tested on maize DNA in triplicate. The optimal concentration of PoacV9_01 was found to be 3.75 μ M, and this concentration was used in all subsequent experiments.

To determine the efficacy of PoacV9_01 in blocking predicted target and nontarget organisms, qPCR assays were then conducted as described above using the DNA isolated from a selection of plants and protists. Water was used as the negative control. Reactions were conducted in triplicate using a PNA clamp concentration of 3.75 μ M. Cycle threshold values were statistically compared between clamp and no-clamp reactions using a Student's *t* test.

Preparation and high-throughput sequencing of maize rhizosphere libraries. We next studied the impact of the clamp on high-throughput sequencing of the maize rhizosphere. B73 maize was planted on 15 May 2019 at Griswold Farm in Griswold, CT and 17 May 2019 at Lockwood Farm in Hamden, CT. Root crowns of four individual plants, or two from each farm, were harvested at 8 weeks after planting (V6 stage) using a drain spade. Root crowns were shaken vigorously for 1 min, and the top 10 cm from five assorted roots was collected using sterile scissors and forceps and placed on ice. Rhizosphere soil samples were collected within 30 min of root collection by vortexing 2 min in 35 ml of phosphate-buffered

saline following the protocol of McPherson et al. (2018) and stored at -80°C until DNA was extracted. The two samples collected at the Hamden, CT site were labeled L1 and L2 and the samples from Griswold site were G1 and G2.

Rhizosphere DNA was extracted using the DNeasy PowerSoil Kit (Qiagen, Germantown, MD, U.S.A.). Two separate V9 amplicon libraries were generated for each sample, one with the PNA clamp and one without. Preparation of the eight libraries for high-throughput sequencing involved two separate PCR steps using Invitrogen Platinum *Pfx* DNA polymerase (Thermo Fisher). In the first set of reactions, PCR amplification was conducted using Euk1391F and EukBr primers with the Illumina sequencing adaptors (Euk1391F: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTACACACCGCCCGTC and EukBr: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGA TCCTTCTGCAGGTTCACTAC) at a final concentration of $0.5\ \mu\text{M}$. Clamp reactions were conducted with the PNA clamp at a final concentration of $3.75\ \mu\text{M}$. The thermal cycler protocol consisted of a denaturation step of 3 min at 95°C ; followed by 25 cycles of a denaturation step of 30 s at 95°C , an annealing step of 30 s at 55°C , and an extension step of 30 s at 72°C ; followed by a final extension step of 5 min at 72°C . After amplification, the libraries were cleaned using the GeneJET Gel Extraction and DNA Cleanup Micro kit (Thermo Scientific). In the second set of reactions, the samples were barcoded on both ends using Nextera DNA CD Indexes (Illumina Inc., San Diego, CA, U.S.A.). The thermal cycler protocol consisted of a denaturation step of 3 min at 95°C ; followed by eight cycles of a denaturation step of 30 s at 95°C , an annealing step of 30 s at 55°C , and an extension step of 30 s at 72°C ; followed by a final extension step of 5 min at 72°C . The barcoded libraries were cleaned one final time using the GeneJET Gel Extraction and DNA Cleanup Micro kit.

The eight libraries were quantified using the Qubit dsDNA HS kit, then diluted to the same concentration. The libraries were pooled and sequenced on the Illumina iSeq 100 platform. The samples were sequenced using $2 \times 150\text{-bp}$ chemistry. The raw sequence data have been submitted to the NCBI Short-Read Archive database under accession number NCBI PRJNA630266.

Bioinformatic analyses. *Quality control and filtering.* Read filtering and assembly were conducted using the *mothur* v. 1.44.0 package (Schloss et al. 2009). Paired-end demultiplexed sequences were assembled into contigs. Contigs were screened to retain sequences between 100 and 200 bp in length, with no ambiguous bases, and a maximum homopolymer length of eight. Filtered contigs were aligned to the SILVA v. 132 SSU reference library. Chimeric sequences were identified using VSEARCH (Rognes et al. 2016) as implemented in *mothur*, and all potentially chimeric sequences were removed.

Taxonomic classification. Sequences were assigned to amplicon sequence variants (ASVs) (sequences sharing 100% sequence similarity) using *mothur*. We examined ASVs instead of operational taxonomic units because ASV methods have a potentially higher sensitivity and resolution (Callahan et al. 2017), which would allow us to better assess the impact of PoacV9_01 on the detection of taxonomic diversity. Representative ASVs were classified using the naïve Bayesian classifier in *mothur* (Wang et al. 2007) against the SILVA v. 132 SSU reference database (Quast et al. 2013). ASVs were also separately classified against the Protist Ribosomal Reference (PR²) database (Guillou et al. 2013). Classifications with a bootstrap of 80% or higher were retained.

Analysis of abundance and diversity. ASV abundance data were merged with the SILVA and PR² taxonomic classification files using the *phyloseq* package (McMurdie and Holmes 2013)

implemented in R. The SILVA and PR² ASV classifications were used to compare ASV abundance between corresponding clamped and unclamped libraries. Rank abundance curves were generated by sorting nonplant ASVs by abundance in R and plotting in *ggplot2*. After removing the plant sequences and filtering out ASVs that had a lower than 0.01% abundance in both libraries, relative abundances of each taxon in the clamp and no-clamp libraries were plotted against each other in *ggplot2*. Similarities were calculated between the clamp and no-clamp libraries using the Bray-Curtis metric, followed by a permutational multivariate analysis of variance (PERMANOVA) to test for significant differences in composition between the treatments. Furthermore, differences between the community composition of clamp and no-clamp libraries were examined for each sample using the Kolmogorov-Smirnov test. α -Diversity statistics were calculated using *mothur*. Rarefaction curves of the ASVs in each library were generated using the *vegan* package in R (Oksanen et al. 2019).

RESULTS

Implementation of a pipeline for design of PNA clamps. To facilitate sequencing of the eukaryotic component of the rhizosphere microbiome, we sought to design a PNA clamp that would bind specifically to plant sequences but not nonplant sequences. We targeted two commonly used 18S rRNA gene universal taxonomic primer sets for the V4 and V9 hypervariable regions, respectively. To design such a clamp, we implemented a k-mer comparative alignment pipeline adapted from previous work (Fig. 1) (Lundberg et al. 2013).

Selection and in silico analysis of clamps. Nineteen V4 k-mers and one V9 k-mer met the design criteria for effective PNA clamps (Supplementary Table S1). No maize-specific k-mers from either region mapped to all or most plant sequences in the index but three V4 k-mers mapped to 3 species in the family Poaceae and 1 species in the family Asparagaceae, while the one V9 k-mer mapped to 40 species in the family Poaceae, 1 species in the family Asparagaceae, and 1 species in the family Xyridaceae. One k-mer from each variable region was selected as a PNA clamp sequence (Supplementary Fig. S1); these were named PoacV4_01 (V4 region, sequence 5'-TCGGTTCTCGCCGTGA-3') and PoacV9_01 (V9 region, sequence 5'-GCCGCCCCGACGTC-3'). These were selected because they were predicted in silico to bind multiple plant species (Supplementary Table S2). PoacV9_01 had a 100% match with several other model grain species (Fig. 2), and only differed by one to two bases from the target region of some nongrass plants that were examined, suggesting that small modifications to the sequence could make it applicable for studies in other host plants.

Preliminary qualitative PCR evaluation of clamp function. Initial conventional PCR tests using 18S rRNA gene universal primers in clamp and no-clamp reactions demonstrated that both clamps suppressed amplification of DNA from maize, as seen by a reduced band intensity compared with a no-clamp reaction (Supplementary Fig. S2). Neither clamp suppressed amplification from the protist UC1. Both clamps also suppressed amplification of rice DNA, and PoacV9_01 appeared to suppress amplification of sorghum and *A. thaliana* as well. Although these results were only visualized at the end of 30 cycles and were not quantitative, they suggested that both PoacV4_01 and PoacV9_01 could effectively block 18S rRNA gene amplification from maize and rice.

Optimization and validation of the V9 PNA clamp PoacV9_01. To quantify the efficiency of blocking by the PNA clamps, we followed up these initial tests with qPCR assays. Over numerous attempts, we were unable to obtain consistent qPCR amplification of the maize V4 region using the primer set

TAREuk454FWD1 and TAREukREV3, which has a longer amplicon than recommended for qPCR using fluorescent staining methods. Because of the broader potential species range of the PoacV9_01 clamp, we decided to focus on PoacV9_01 for in-depth analyses.

PNA clamp concentration may affect its efficacy (Belda et al. 2017; Lundberg et al. 2013). In a test of several clamp

concentrations, PoacV9_01 suppressed the amplification of maize DNA by approximately five cycles at concentrations as low as 0.75 μ M (Supplementary Table S3). A concentration of 3.75 μ M increased the suppression by approximately one more cycle, and with little to no increased efficiency at higher concentrations (Supplementary Table S3). Therefore, 3.75 μ M was selected as the optimal PNA concentration for further analyses. PoacV9_01 was

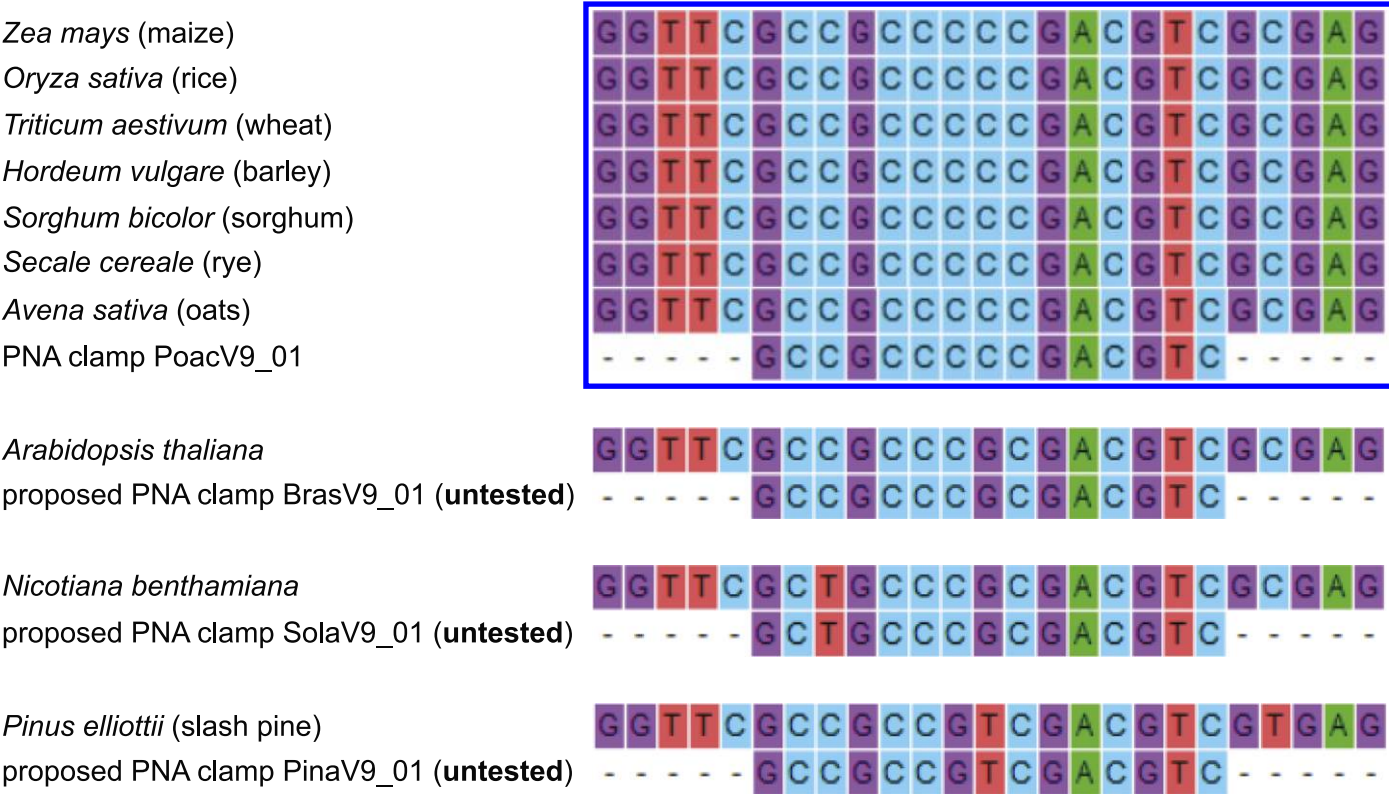


Fig. 2. 18S ribosomal RNA gene fragments of select plant species that align with the peptide nucleic acid (PNA) clamp developed in this study (PoacV9_01). The clamp was experimentally demonstrated to effectively block the amplification of DNA from maize, rice, wheat, barley, and sorghum, and should be effective for other grass species such as rye and oat. Small modifications in the clamp sequence should make it effective for other plant species such as *Arabidopsis thaliana*, *Nicotiana benthamiana*, and *Pinus elliotii*. The proposed clamps for *A. thaliana*, *N. benthamiana*, and *P. elliotii* have not been tested on plant material but have been tested in silico following the protocol outlined in this study (see Materials and Methods).

TABLE 1 Mean cycle thresholds and ranges observed in 18S V9 amplification of selected species, with and without addition of the peptide nucleic acid clamp PoacV9_01 at 3.75 μ M					
Plant species	No clamp ^a	Clamp ^a	Difference	<i>t</i> value (df = 4)	<i>P</i> value ^b
Maize	17.27 (17.17 to 17.45)	25.16 (24.93 to 25.58)	7.89	33.97	<0.001
Rice	18.86 (18.77 to 18.91)	24.21 (24.16 to 24.29)	5.35	91.34	<0.001
Wheat	19.43 (19.20 to 19.57)	26.53 (26.15 to 27.07)	7.10	23.70	<0.001
Barley	17.00 (16.87 to 17.14)	24.97 (24.88 to 25.10)	7.97	79.15	<0.001
Sorghum	13.47 (13.30 to 13.69)	22.59 (22.19 to 23.12)	9.12	30.48	<0.001
<i>Arabidopsis thaliana</i>	12.75 (12.62 to 12.91)	13.75 (13.42 to 14.26)	1.00	3.72	0.021
<i>Nicotiana benthamiana</i>	16.69 (16.50 to 16.81)	16.63 (16.53 to 16.75)	0.06	−0.49	0.650
<i>Colpoda</i> sp.	20.43 (20.39 to 20.46)	20.45 (20.37 to 20.56)	0.02	0.44	0.681
<i>Cercomonas</i> sp.	17.67 (17.53 to 17.80)	17.58 (17.54 to 17.60)	0.09	−1.14	0.319
Negative control	36.06 (34.06 to 37.45)	34.29 (33.80 to 34.56)	1.77	−1.93	0.126

^a Represents the mean value of three independent reactions. Range of the three values is shown in parentheses.
^b Numbers in bold indicate that the difference is statistically significant.

priced at roughly \$450 per 100 nmol at the time of this study, or 530 50- μ l reactions at the optimal concentration (\$0.84/reaction).

Next, we tested the efficacy of the PoacV9_01 clamp in suppressing amplification from five Poaceae crops that have an identical match to the V9 target sequence: maize, rice, wheat, barley, and sorghum. In addition, we evaluated suppression of the model plant *A. thaliana*, predicted to have only one mismatch in the eighth nucleotide of the target sequence (Fig. 2), and *N. benthamiana*, predicted to have two mismatches in the third and eighth nucleotides. DNA from protist isolates UC1 and UC5 were also tested to evaluate potential off-target suppression of nonplant eukaryotes. PoacV9_01 was highly effective in suppressing the amplification of all five Poaceae crops (Table 1). Clamp addition delayed fluorescence detection in these five species by 5.35 to 9.12 cycles compared with the corresponding unclamped reactions. There was slight suppression of amplification of the *A. thaliana* V9 region, although it was much less efficient than that observed with the grass species (Table 1). The reduction in suppression efficiency with one mismatch has also been observed for PNA clamps developed for plant plastids (Fitzpatrick et al. 2018). In contrast, PoacV9_01 had no significant effect on amplification of protist or *N. benthamiana* DNA. These results indicate that the PNA clamp PoacV9_01 effectively suppresses amplification from grass species with matching V9 target sequences, and may reduce amplification from species with a single central mismatch, but does not suppress amplification from species with two or more mismatches.

Clamp PoacV9_01 reduced amplification of host DNA and increased ASV abundance and diversity. Although PoacV9_01 suppressed amplification of purified maize DNA in a qPCR assay, it

was still uncertain how effective the clamp would be in the context of sequencing a complex rhizosphere sample, or whether clamp addition would introduce a bias to amplification and recovery of sequences. To answer these questions, we applied the PoacV9_01 clamp to high-throughput sequencing of rhizosphere soil samples collected from four field-grown maize plants. Clamp and no-clamp libraries were simultaneously prepared and sequenced for each plant in order to determine the effect of PoacV9_01 on sequence recovery within each sample.

After quality filtering, we recovered 2,291,527 sequences (286,441 average per sample). These sequences could be clustered into an average of 44,286 ASVs per sample.

PoacV9_01 drastically reduced the sequencing of host reads in rhizosphere samples. Representative ASV sequences were classified according to two taxonomic databases commonly used in eukaryote profiling studies, SILVA (Quast et al. 2013) and PR² (Guillou et al. 2013). The PR² database includes a more comprehensive catalog of eukaryotic taxa, especially protists (Guillou et al. 2013); therefore, we hypothesized that this database would allow us to classify a higher proportion of eukaryotic sequence variants or produce classifications to deeper taxonomic ranks. Use of the SILVA database would allow us to determine whether the conclusions would be affected by the classification database employed, and to determine the extent of noneukaryote sequence recovery.

Regardless of the classification database, addition of PoacV9_01 reduced the mean proportion of plant-derived reads from 66.5 to 0.6% (Fig. 3A). The sequencing recovery of other eukaryotes was increased correspondingly (Supplementary Table S4), yielding a much higher proportion of animal, fungal, and protist reads in the

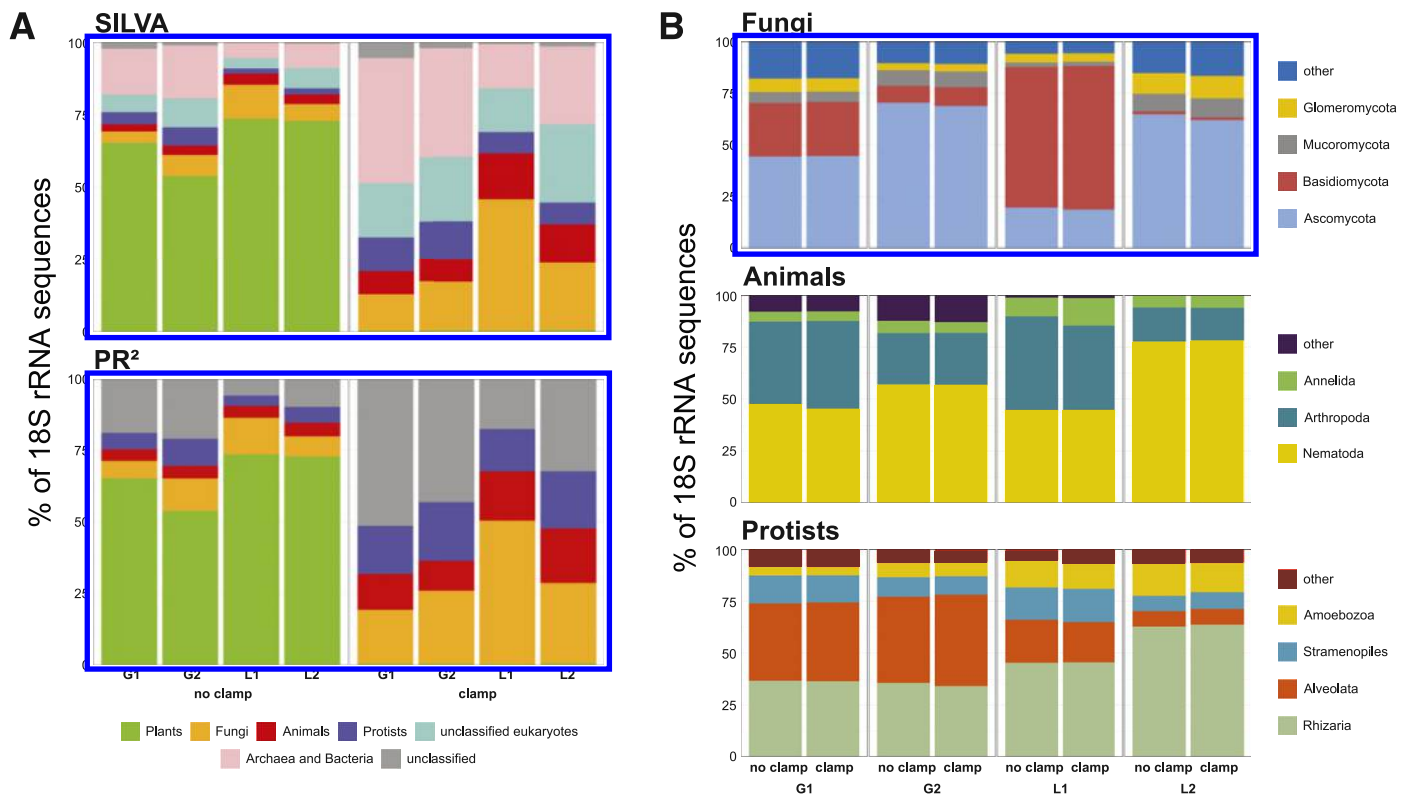


Fig. 3. Bar charts showing the relative abundance of taxonomic bins in the sequence datasets. **A**, Relative abundances of kingdom-level groups found in each rhizosphere sample, using either the SILVA or Protist Ribosomal Reference (PR²) database for classification. Bars, each of which represent a sequenced library, are organized by clamp treatment. rRNA = ribosomal RNA. **B**, Relative abundances of kingdom-level groups of nonplant eukaryotes found in this study, as identified using the PR² database. Bars are organized by sample identification. The category “other” is the sum total of the taxonomic groups not displayed.

clamped libraries than the unclamped libraries. The largest difference between reads classified by the two databases was in the bacteria and archaea (Fig. 3A). Because these organisms are not included in the PR² database, they are presumably represented the bin “unclassified” (Fig. 3A). Thus, these data suggest that there is some amplification of bacteria and archaea from the rhizosphere samples.

A higher proportion of sequences were classified as animals (14.9%), fungi (30.7%), and protists (17.9%) by the PR² database than by the SILVA database (11.1% to the animals, 24.5% to the fungi, and 10.0% to the protists), suggesting that PR² may be more discriminating in the classification of eukaryotes in the phytobiome. Despite these differences, the patterns between datasets were maintained. For example, both databases identified that the relative abundance of fungi was highest in sample L1. Taken together, these data suggest that the differences between classification databases are relatively small; however, potential biases in how the databases assign taxa should be taken into consideration before selecting one database over the other (Dupont et al. 2016).

Using the PR² database, which classified more sequences to eukaryotic groups than SILVA, we examined the major nonplant groups at deeper taxonomic levels. There were only very minor differences in the proportional abundances of animal, fungal, and protist sequences between the clamp and no-clamp libraries (Fig. 3B; Supplementary Table S5), suggesting that PoacV9_01 is not a source of bias. However, there were large differences among the four samples, with the dominant fungal, animal, and protist group varying among the samples. Studies with more replicates and deeper sequencing are needed to detect and verify any ecological patterns between plants and sites.

PoacV9_01 increased the abundance and diversity of 18S rRNA gene sequences recovered from rhizosphere soils. Next, we asked whether addition of the clamp increased the number and diversity of taxa detected by sequencing. ASV-based analysis demonstrated that PoacV9_01 increased the number of ASVs in each library by 15,161 to 38,330 compared with the corresponding unclamped library (Table 2), or an increase of 50 to 100%. Across samples, we detected a mean of 8,832 additional ASVs from fungi, 7,432 from protists, and 4,604 from animals that were not detected in the absence of PoacV9-01 (Supplementary Table S6). The increases in detected diversity were reflected in the Shannon diversity indices for each sample, which were doubled in the clamp libraries relative to the no-clamp libraries (Table 2). Good’s coverage estimator, an estimate of the proportion of total species represented in a sample, was consistently lower in the clamp libraries than in the no-clamp libraries, suggesting that deeper sequencing of the clamp libraries would result in the detection of many more ASVs

compared with the no-clamp libraries. Good’s coverage for the clamped samples ranged from approximately 86 to 89%, suggesting that the majority of the expected genetic diversity of the eukaryotic community was recovered with this sampling depth. The higher number of ASVs detected in the clamp samples was supported by rarefaction curves (Fig. 4), which confirmed that ASV counts were consistently higher for clamp samples than for no-clamp samples across the samples. However, the rarefaction curves did not plateau, even in the clamped samples, suggesting that, even with the increased recovery of ASVs with the PNA clamps, the total diversity of the eukaryotic community has only been partially sampled. This result held up even when the libraries were subsampled to the same size (Fig. 4), indicating that the increased diversity recovered in the clamp samples was due to sequences being drawn from a more diverse species pool rather than an artifact of deeper sampling after plant reads were removed.

PoacV9-01 did not bias ASV recovery. We next asked whether the PoacV9-01 clamp might introduce bias into sequencing results, potentially through off-target binding, increasing amplification bias, or other mechanisms. After bioinformatically removing plant reads from the libraries, rank abundance curves were plotted for each sample (Fig. 5). ASVs consistently had the same rank abundance relationship between clamped and nonclamped samples (Fig. 5). In other words, the most abundant ASVs in the nonclamped samples were consistently identified as the most abundant ASVs in the clamped sample. Furthermore, the ASVs were generally present in similar relative abundances, suggesting that the community profile for the dominant eukaryotic members was unaffected by including a PNA clamp in the PCR amplification. This was supported by plotting ASV relative abundances between the clamp and no-clamp libraries, which demonstrated a linear relationship between the taxa relative abundances in the clamp and no-clamp libraries, suggesting that rarer taxa were present in similar relative abundances as well (Supplementary Fig. S3).

Statistical analyses were performed to further test whether the clamp is a source of community bias. Bray-Curtis similarity measures paired with PERMANOVA statistical comparisons indicated that there was no significant difference in composition between the clamp and no-clamp libraries (Table 3). The communities were even more similar when the libraries were subsampled to the same size, demonstrating that much of the variation between the clamp and no-clamp libraries was due to there being greater sequencing depth in the clamp libraries versus the no-clamp libraries. A Kolmogorov-Smirnov test further supported that there were no significant differences in the ASV composition between the clamp and no-clamp libraries (Fig. 5).

TABLE 2
α-Diversity statistics for the no-clamp and clamp libraries^a

Sample	Total reads ^b		ASVs detected		ΔASVs	Shannon’s (H')		ΔH'	Good’s coverage	
	c–	c+	c–	c+		c–	c+		c–	c+
G1	260,261	383,530	35,799	74,129	+38,330	6.18	8.82	+2.64	0.90	0.86
G2	220,984	324,270	34,218	62,424	+28,206	6.62	8.46	+1.84	0.89	0.86
L1	234,097	312,264	22,946	46,103	+23,157	5.03	7.11	+2.08	0.93	0.89
L2	294,186	261,935	31,754	46,915	+15,161	5.48	7.93	+2.45	0.92	0.87
Mean	252,382	320,500	31,179	57,393	+26,214	5.83	8.08	+2.25	0.91	0.87

^a ASVs = amplicon sequence variants, and c– and c+ denote unclamped and clamped samples, respectively.

^b Total number of reads after length filtering and assembly into contigs.

DISCUSSION

The majority of plant microbiome studies have focused on bacteria and fungi, in part because the abundance of host DNA in plant-associated samples limits the sequencing from other eukaryotes. Here, we optimized a pipeline (Belda et al. 2017;

Lundberg et al. 2013) for the design of PNA clamps to suppress the amplification of plant 18S rRNA gene sequences during high-throughput sequencing. We developed and validated the PNA clamp PoacV9_01, which suppressed the amplification of 18S rRNA genes from five species of model grain crops. In high-throughput sequencing analysis of maize rhizosphere DNA, the

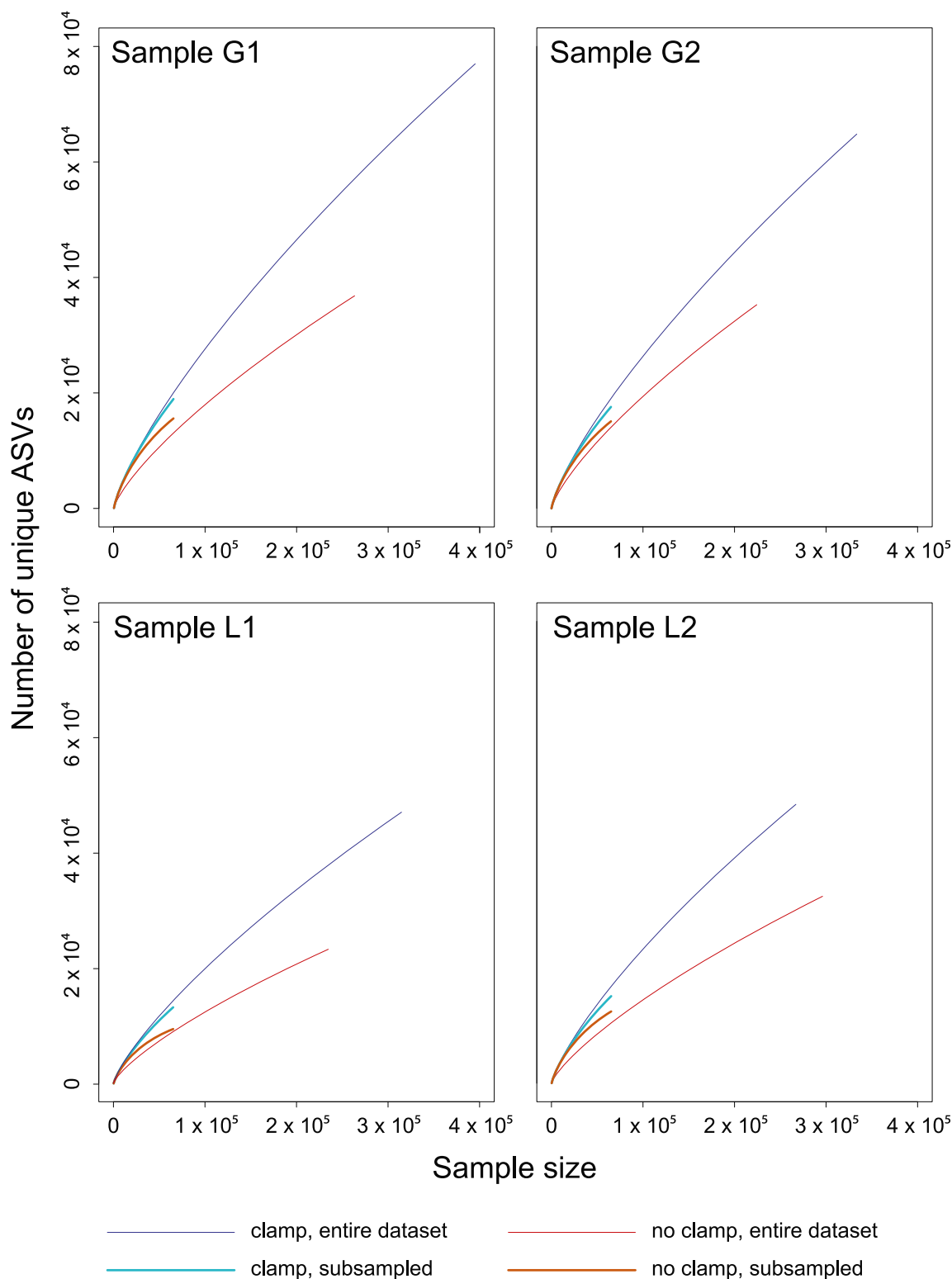


Fig. 4. Rarefaction curves of the libraries generated from the four rhizosphere samples. Curves represent the number of amplicon sequence variants (ASVs) occurring in arbitrary samples of increasing size in clamp and no-clamp libraries for each sample. Thin-lined curves represent the entire datasets, while bold-lined curves represent datasets which were subsampled to the same size after removal of plant sequences.

clamp eliminated approximately 99% of host sequence interference, reducing the proportion of plant reads from 66.5 to 0.6%. Incorporation of PoacV9_01 increased the measured diversity of other eukaryotes, including animals, fungi, and protists, increasing the number of sequence variants without biasing the composition of eukaryote groups. This study establishes PNA clamping as a promising method to facilitate profiling of eukaryotes in the phytobiome.

Addition of the PoacV9_01 clamp confers several advantages to high-throughput sequencing of plant-associated eukaryotes. Clamp

addition substantially increased the number of nonplant reads and ASVs in a small sequencing study. PoacV9_01 or other PNA clamps could allow researchers to study a greater number of samples simultaneously, obtain more informative data from inexpensive sequencing platforms, or potentially use universal 18S rRNA gene primers instead of group-specific primers, including those commonly used in studies of phytobiome fungi (Gao et al. 2019a; Hannula et al. 2017). Studies are underway to determine whether the approach described here compares favorably to fungal-specific primers in measuring community composition.

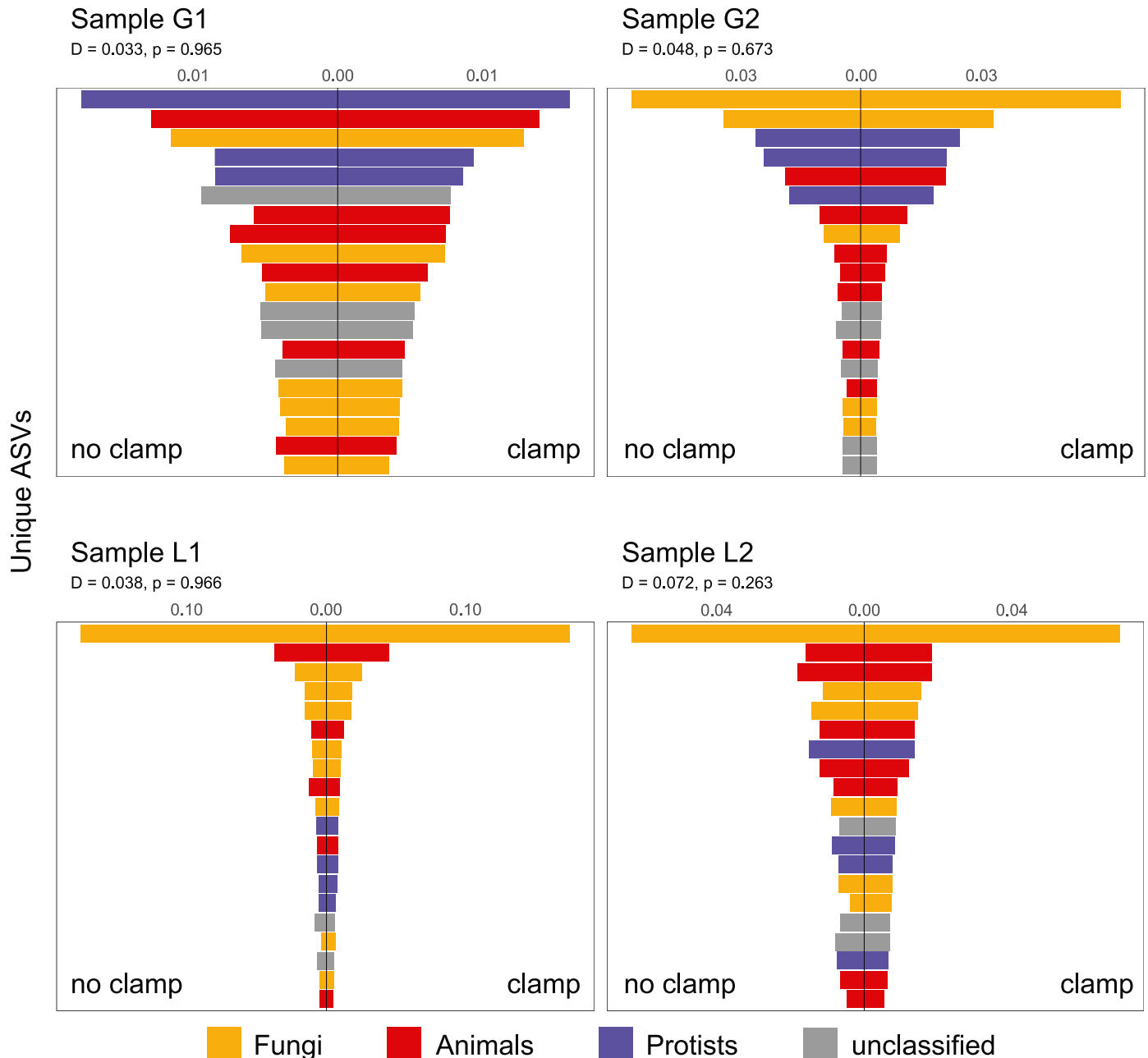


Fig. 5. Rank abundance curves of the libraries generated from the four rhizosphere samples, after plant sequences were removed. Each bar represents a single amplicon sequence variant (ASV). The 20 most abundant ASVs in each sample are shown in order of their ranking in the clamp library. In this respect, the absence of an ASV in a sample means that it was not among the 20 most abundant ASVs, rather than absent from the sample. Bar lengths represent ASV relative abundance with the clamp (right x-axis) and without the clamp (left x-axis). ASVs are colored according to their kingdom-level taxonomic classification based on the Protist Ribosomal Reference (PR²) database. The Kolmogorov-Smirnov statistical test was performed on the full datasets and the test statistics are shown underneath each sample label. *P* values > 0.05 indicate that there is no significant difference in the distribution between the curves.

TABLE 3
Table of permutational multivariate analysis of variance statistics showing the dissimilarity between the clamp and no-clamp samples, with the whole dataset after plant reads were removed, as well as the subsampled dataset

Plants	df	Sum of squares	Mean sum squares	F stat	R ²	P value
Plants removed						
Treatment	1	0.391	0.391	1.395	0.189	0.136
Residuals	6	1.679	0.280	...	0.811	...
Total	7	2.070	...	...	1.000	...
Subsampled ^a						
Treatment	1	0.061	0.061	0.212	0.034	0.780
Residuals	6	1.737	0.289	...	0.966	...
Total	7	1.798	...	...	1.000	...

^a Samples were randomly subsampled without replacement to the size of the smallest dataset.

Although we initially sought to design a PNA clamp that could block all plants, we did not identify a universal plant clamp in the V4 or V9 regions that would not also block some potential intended targets. However, in silico analysis showed that the PoacV9-01 clamp is modifiable; correcting the sequence to match the same V9 site in other plant families results in candidate clamps for other model species (Fig. 2). Work is underway in our group to validate the efficacy and breadth of these proposed clamps in other plant families. It is possible that, as more becomes known about the eukaryotic phytobiome, it will be considered acceptable to design a clamp that is more universal to plants but which also suppresses some protist groups of lesser interest.

Despite its targeted nature, PoacV9_01 effectively suppressed qPCR amplification from five major grain crops: maize, rice, wheat, sorghum, and barley. These species represent five genera from three distinct subfamilies in the Poaceae family, and the target region was confirmed in silico to be conserved in other Poaceae species (Fig. 1). Therefore, we anticipate that PoacV9_01 will be useful across varieties within these species, as well as with other grasses such as rye or oat. However, we recommend that the clamp should be optimized for each new host species prior to a sequencing run. PoacV9_01 also suppressed the amplification of *A. thaliana* to a small but significant degree. This supports prior findings that PNA clamps can still affect amplification of targets with a single mismatch (Terahara et al. 2011) but also demonstrates that a central mismatch greatly reduces suppression efficiency.

This study sought to design clamps that would work for common universal primers targeting both the V4 and V9 hypervariable regions of the maize 18S rRNA gene. Both the V4 and V9 regions are widely used in eukaryotic community studies, although they carry different biases and may result in highly distinct results when applied to the same sample (Giner et al. 2016; Hiraoka et al. 2019; Tragin et al. 2018). The V9 region is shorter and more consistent in length than the V4 region, limiting amplification bias (Geisen et al. 2018) and allowing full sequence coverage from paired ends. V9 universal primers have been adopted as a standard tool of the global cataloging effort the Earth Microbiome Project (Gilbert et al. 2014). On the other hand, the V4 region is longer and, thus, more phylogenetically informative, and is better represented in the SILVA and PR² databases (Geisen et al. 2018). We focused here on validating the V9 clamp PoacV9_01 because it matched a much wider predicted species range than any clamp designed for the V4 region, making this region more promising for comparative studies. Our proposed V4 clamp, PoacV4_01, also reduced apparent amplification from rice and maize in conventional PCR, and is a

promising candidate for sequencing studies in those species. However, PoacV4_01 has not been validated in a quantitative manner, and we would caution potential users to perform optimization studies before implementing this clamp in a sequencing reaction.

Despite its importance to plant health, the eukaryotic phytobiome is still a relatively unexplored realm, and there are several technical barriers preventing its widespread exploration. Sequencing studies can be complicated by the vast differences in 18S rRNA gene copy number among protists (Gong and Marchetti 2019), or the fact that even the most “universal” of primers introduce bias and miss some taxa (Adl et al. 2014). Functional studies can be hampered by the difficulty in isolating and culturing many soil eukaryotes. In this study, we developed a tool to overcome one important hurdle in profiling the eukaryotic phytobiome, and used a pipeline that others can implement to develop their own similar tools. We hope this will encourage the phytobiome research community to incorporate a broader spectrum of rhizosphere and phyllosphere organisms into routine studies, as well as into discussions about what standardized sequencing methods to adopt.

ACKNOWLEDGMENTS

We thank J. Leach, C. Jahn, S. Dellaporta, and E. Luna for contributing seed of maize, wheat, barley, and sorghum.

LITERATURE CITED

- Adl, S. M., Bass, D., Lane, C. E., Lukeš, J., Schoch, C. L., Smirnov, A., Agatha, S., Berney, C., et al. 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. *J. Eukaryot. Microbiol.* 66:4-119.
- Adl, S. M., Habura, A., and Eglit, Y. 2014. Amplification primers of SSU rDNA for soil protists. *Soil Biol. Biochem.* 69:328-342.
- Belda, E., Coulibaly, B., Fofana, A., Beavogui, A. H., Traore, S. F., Gohl, D. M., Vernick, K. D., and Riehle, M. M. 2017. Preferential suppression of *Anopheles gambiae* host sequences allows detection of the mosquito eukaryotic microbiome. *Sci. Rep.* 7:3241.
- Berendsen, R. L., Pieterse, C. M. J., and Bakker, P. A. H. M. 2012. The rhizosphere microbiome and plant health. *Trends Plant Sci.* 17:478-486.
- Burki, F., Roger, A. J., Brown, M. W., and Simpson, A. G. B. 2020. The new tree of eukaryotes. *Trends Ecol. Evol.* 35:43-55.
- Callahan, B. J., McMurdie, P. J., and Holmes, S. P. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* 11:2639-2643.
- Carver, T., and Bleasby, A. 2003. The design of Jemboss: A graphical user interface to EMBOSS. *Bioinformatics* 19:1837-1843.

- Compant, S., Samad, A., Faist, H., and Sessitsch, A. 2019. A review on the plant microbiome: Ecology, functions, and emerging trends in microbial application. *J. Adv. Res.* 19:29-37.
- De Gruyter, J., Weedon, J. T., Bazot, S., Dauwe, S., Fernandez-Garberí, P.-R., Geisen, S., Gourlez De La Motte, L., Heinesch, B., et al. 2020. Patterns of local, intercontinental and interseasonal variation of soil bacterial and eukaryotic microbial. *FEMS Microbiol. Ecol.* 96:fiia018.
- Dupont, A. Ö. C., Griffiths, R. I., Bell, T., and Bass, D. 2016. Differences in soil micro-eukaryotic communities over soil pH gradients are strongly driven by parasites and saprotrophs. *Environ. Microbiol.* 18:2010-2024.
- Fitzpatrick, C. R., Lu-Irving, P., Copeland, J., Guttman, D. S., Wang, P. W., Baltrus, D. A., Dlugosch, K. M., and Johnson, M. T. J. 2018. Chloroplast sequence variation and the efficacy of peptide nucleic acids for blocking host amplification in plant microbiome studies. *Microbiome* 6:144.
- Gao, Z., Han, M., Hu, Y., Li, Z., Liu, C., Wang, X., Tian, Q., Jiao, W., et al. 2019a. Effects of continuous cropping of sweet potato on the fungal community structure in rhizospheric soil. *Front. Microbiol.* 10:2269.
- Gao, Z., Karlsson, I., Geisen, S., Kowalchuk, G., and Jousset, A. 2019b. Protists: Puppet masters of the rhizosphere microbiome. *Trends Plant Sci.* 24:165-176.
- Geisen, S., Koller, R., Hünninghaus, M., Dumack, K., Urich, T., and Bonkowski, M. 2016. The soil food web revisited: Diverse and widespread mycophagous soil protists. *Soil Biol. Biochem.* 94:10-18.
- Geisen, S., Mitchell, E. A. D., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., Fernández, L. D., Jousset, A., et al. 2018. Soil protists: A fertile frontier in soil biology research. *FEMS Microbiol. Rev.* 42:293-323.
- Geisen, S., Rosengarten, J., Koller, R., Mulder, C., Urich, T., and Bonkowski, M. 2015. Pack hunting by a common soil amoeba on nematodes. *Environ. Microbiol.* 17:4538-4546.
- Geisen, S., Vault, D., Mahé, F., Lara, E., de Vargas, C., and Bass, D. 2019. A user guide to environmental protistology: Primers, metabarcoding, sequencing, and analyses. *bioRxiv*. doi:10.1101/850610
- Gilbert, J. A., Jansson, J. K., and Knight, R. 2014. The Earth Microbiome project: Successes and aspirations. *BMC Biol.* 12:69.
- Giner, C. R., Fom, I., Romac, S., Logares, R., de Vargas, C., and Massana, R. 2016. Environmental sequencing provides reasonable estimates of the relative abundance of specific picoeukaryotes. *Appl. Environ. Microbiol.* 82:4757-4766.
- Gong, W., and Marchetti, A. 2019. Estimation of 18S gene copy number in marine eukaryotic plankton using a next-generation sequencing approach. *Front. Mar. Sci.* 6:219.
- Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., et al. 2013. The Protist Ribosomal Reference database (PR²): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* 41:D597-D604.
- Hannula, S. E., Morriën, E., De Hollander, M., Van Der Putten, W. H., Van Veen, J. A., and De Boer, W. 2017. Shifts in rhizosphere fungal community during secondary succession following abandonment from agriculture. *ISME J.* 11:2294-2304.
- Henkes, G. J., Kandeler, E., Marhan, S., Scheu, S., and Bonkowski, M. 2018. Interactions of mycorrhiza and protists in the rhizosphere systemically alter microbial community composition, plant shoot-to-root ratio and within-root system nitrogen allocation. *Front. Environ. Sci.* 6:117.
- Hirakata, Y., Hatamoto, M., Oshiki, M., Watari, T., Kuroda, K., Araki, N., and Yamaguchi, T. 2019. Temporal variation of eukaryotic community structures in UASB reactor treating domestic sewage as revealed by 18S rRNA gene sequencing. *Sci. Rep.* 9:12783.
- Huang, X. F., Chaparro, J. M., Reardon, K. F., Zhang, R., Shen, Q., and Vivanco, J. M. 2014. Rhizosphere interactions: Root exudates, microbes, and microbial communities I. *Botany* 92:267-275.
- Jiang, Y., Liu, M., Zhang, J., Chen, Y., Chen, X., Chen, L., Li, H., Zhang, X.-X., and Sun, B. 2017. Nematode grazing promotes bacterial community dynamics in soil at the aggregate level. *ISME J.* 11:2705-2717.
- Jousset, A., Rochat, L., Péchy-Tarr, M., Keel, C., Scheu, S., and Bonkowski, M. 2009. Predators promote defence of rhizosphere bacterial populations by selective feeding on non-toxic cheaters. *ISME J.* 3:666-674.
- Karkare, S., and Bhatnagar, D. 2006. Promising nucleic acid analogs and mimics: Characteristic features and applications of PNA, LNA, and morpholino. *Appl. Microbiol. Biotechnol.* 71:575-586.
- Kumar, S., Stecher, G., and Tamura, K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33:1870-1874.
- Lane, D. J. 1991. 16S/23S sequencing. Pages 115-175 in: *Nucleic Acid Technologies in Bacterial Systematics*. E. Stackebrand and M. Goodfellow, eds. Wiley, New York, NY, U.S.A.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.
- Lavelle, P., Decaëns, T., Aubert, M., Barot, S., Blouin, M., Bureau, F., Margerie, P., Mora, P., and Rossi, J. P. 2006. Soil invertebrates and ecosystem services. *Eur. J. Soil Biol.* 42:S3-S15.
- Leach, J. E., Triplett, L. R., Argueso, C. T., and Trivedi, P. 2017. Communication in the phytobiome. *Cell* 169:587-596.
- Long, J. J., Jahn, C. E., Sánchez-Hidalgo, A., Wheat, W., Jackson, M., Gonzalez-Juarrero, M., and Leach, J. E. 2018. Interactions of free-living amoebae with rice bacterial pathogens *Xanthomonas oryzae* pathovars *oryzae* and *oryzicola*. *PLoS One* 13:e0202941.
- 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.
- Mahé, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E., Singer, D., Mayor, J., et al. 2017. Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat. Ecol. Evol.* 1:0091.
- McMurdie, P. J., and Holmes, S. 2013. Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8: e61217.
- McPherson, M. R., Wang, P., Marsh, E. L., Mitchell, R. B., and Schachtman, D. P. 2018. Isolation and Analysis of microbial communities in soil, rhizosphere, and roots in perennial grass experiments. *J. Vis. Exp.* 137:e57932.
- Medlin, L., Elwood, H. J., Stickel, S., and Sogin, M. L. 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* 71:491-499.
- Mendes, R., Garbeva, P., and Raaijmakers, J. M. 2013. The rhizosphere microbiome: Significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol. Rev.* 37:634-663.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., et al. 2019. *vegan: Community ecology package*. R Package Version 2.5-6. <https://cran.r-project.org/web/packages/vegan/vegan.pdf>
- Paramelee, R. W. 1995. Soil fauna: Linking different levels of the ecological hierarchy. Pages 107-116 in: *Linking Species and Ecosystems*. G. A. Polis and K. O. Winemiller, eds. Chapman and Hall, New York, NY, U.S.A.
- Ploch, S., Rose, L. E., Bass, D., and Bonkowski, M. 2016. High diversity revealed in leaf-associated protists (Rhizaria: Cercozoa) of Brassicaceae. *J. Eukaryot. Microbiol.* 63:635-641.
- Quast, C., Priesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F. O. 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* 41: D590-D596.
- Rice, P., Longden, L., and Bleasby, A. 2000. EMBOSS: The European molecular biology open software suite. *Trends Genet.* 16:276-277.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. 2016. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* 4:e2584.
- Rubinstein, R. L., Kadilak, A. L., Cousens, V. C., Gage, D. J., and Shor, L. M. 2015. Protist-facilitated particle transport using emulated soil micromodels. *Environ. Sci. Technol.* 49:1384-1391.
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. 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.
- Steven, B., Huntley, R. B., and Zeng, Q. 2018. The influence of flower anatomy and apple cultivar on the apple flower phytobiome. *Phytobiomes J.* 2:171-179.
- Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H.-W., and Richards, T. A. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* 19:21-31.
- Terahara, T., Chow, S., Kurogi, H., Lee, S., Tsukamoto, K., Tanaka, H., and Takeyama, H. 2011. Efficiency of peptide nucleic acid-directed PCR clamping and its application in the investigation of natural diets of the Japanese eel *Leptocephali*. *PLoS One* 6:e25715.
- Tragin, M., Zingone, A., and Vault, D. 2018. Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta. *Environ. Microbiol.* 20:506-520.
- Trap, J., Bonkowski, M., Plassard, C., Villenave, C., and Blanchart, E. 2016. Ecological importance of soil bacterivores for ecosystem functions. *Plant Soil* 398:1-24.

- Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73:5261-5267.
- Weidner, S., Latz, E., Agaras, B., Valverde, C., and Jousset, A. 2017. Protozoa stimulate the plant beneficial activity of rhizospheric pseudomonads. *Plant Soil* 410:509-515.
- Xiong, W., Jousset, A., Guo, S., Karlsson, I., Zhao, Q., Wu, H., Kowalchuk, G. A., Shen, Q., Li, R., and Geisen, S. 2017. Soil protist communities form a dynamic hub in the soil microbiome. *ISME J.* 12:634-638.
- Xiong, W., Song, Y., Yang, K., Gu, Y., Wei, Z., Kowalchuk, G. A., Xu, Y., Jousset, A., Shen, Q., and Geisen, S. 2020. Rhizosphere protists are key determinants of plant health. *Microbiome* 8:27.