Video Article

Exploring the Root Microbiome: Extracting Bacterial Community Data from the Soil, Rhizosphere, and Root Endosphere

Tuesday Simmons[1], Daniel F. Caddell[1], Siwen Deng[1], Devin Coleman-Derr[1,2]

[1]Department of Plant and Microbial Biology, University of California, Berkeley

[2]Plant Gene Expression Center, USDA ARS

Correspondence to: Devin Coleman-Derr at colemanderr@gmail.com

URL: https://www.jove.com/video/57561
DOI: doi:10.3791/57561

Keywords: Genetics, Issue 135, 16S rRNA, sequencing, root endosphere, rhizosphere, soil, microbiome, phylogenetic profiling

Date Published: 5/2/2018

Citation: Simmons, T., Caddell, D.F., Deng, S., Coleman-Derr, D. Exploring the Root Microbiome: Extracting Bacterial Community Data from the Soil, Rhizosphere, and Root Endosphere. *J. Vis. Exp.* (135), e57561, doi:10.3791/57561 (2018).

Abstract

The intimate interaction between plant host and associated microorganisms is crucial in determining plant fitness, and can foster improved tolerance to abiotic stresses and diseases. As the plant microbiome can be highly complex, low-cost, high-throughput methods such as amplicon-based sequencing of the 16S rRNA gene are often preferred for characterizing its microbial composition and diversity. However, the selection of appropriate methodology when conducting such experiments is critical for reducing biases that can make analysis and comparisons between samples and studies difficult. This protocol describes in detail a standardized methodology for the collection and extraction of DNA from soil, rhizosphere, and root samples. Additionally, we highlight a well-established 16S rRNA amplicon sequencing pipeline that allows for the exploration of the composition of bacterial communities in these samples, and can easily be adapted for other marker genes. This pipeline has been validated for a variety of plant species, including sorghum, maize, wheat, strawberry, and agave, and can help overcome issues associated with the contamination from plant organelles.

Video Link

The video component of this article can be found at https://www.jove.com/video/57561/

Introduction

Plant-associated microbiomes consist of dynamic and complex microbial communities comprised of bacteria, archaea, viruses, fungi, and other eukaryotic microorganisms. In addition to their well-studied role in causing plant disease, plant-associated microbes can also positively influence plant health by improving tolerance to biotic and abiotic stresses, promoting nutrient availability, and enhancing plant growth through the production of phytohormones. For this reason, particular interest exists in characterizing the taxa that associate with plant root endospheres, rhizospheres, and the surrounding soil. While some microbes can be cultured in isolation on laboratory generated media, many cannot, in part because they may rely on symbiotic relationships with other microbes, grow very slowly, or require conditions that cannot be replicated in a lab environment. Because it circumvents the need for cultivation and is relatively inexpensive and high-throughput, sequence-based phylogenetic profiling of environmental and host-associated microbial samples has become a preferred method for assaying microbial community composition.

The selection of appropriate sequencing technologies provided by various next generation sequencing (NGS) platforms[1] is dependent on the users' needs, with important factors including: desired coverage, amplicon length, expected community diversity, as well as sequencing error-rate, read-length, and the cost-per-run/megabase. Another variable that needs to be considered in amplicon-based sequencing experiments is what gene will be amplified and what primers will be used. When designing or choosing primers, researchers are often forced to make tradeoffs between the universality of amplification and the taxonomic resolution achievable from the resulting amplicons. For this reason, these types of studies often chose primers and markers that selectively target specific subsets of the microbiome. Evaluating the composition of bacterial communities is commonly accomplished by sequencing one or more of the hypervariable regions of the bacterial 16S rRNA gene[2,3]. In this study, we describe an amplicon based sequencing protocol developed for a NGS platform that targets the 500 bp V3-V4 region of the bacterial 16S rRNA gene, which allows for broad amplification of bacterial taxa while also providing sufficient variability to distinguish between different taxa. Additionally, this protocol can easily be adapted for the use with other primer sets, such as those targeting the ITS2 marker of fungi or the 18S rRNA subunit of eukaryotes.

While other approaches such as shotgun metagenomics, metatranscriptomics, and single-cell sequencing, offer other advantages including resolved microbial genomes and more direct measurement of community function, these techniques are typically more expensive and computationally intensive than the phylogenetic profiling described here[4]. Additionally, performing shotgun metagenomics and metatranscriptomics on root samples yields a large percentage of reads belonging to the host plant genome, and methods to overcome this limitation are still being developed[5,6].

As with any experimental platform, amplicon-based profiling can introduce a number of potential biases which should be considered during the experimental design and data analysis. These include the methods of sample collection, DNA extraction, selection of PCR primers, and how library preparation is performed. Different methods can significantly impact the amount of usable data generated, and can also hinder the efforts to compare results between studies. For example, the method of removing rhizosphere bacteria[7] and the use of different extraction techniques or choice of DNA extraction kits[8,9] have been shown to significantly impact downstream analysis, which leads to different conclusions regarding which microbes are present and their relative abundances. Since amplicon-based profiling can be customized, making comparisons across studies can be challenging. The Earth Microbiome Project has suggested that researchers studying complex systems such as the plant-associated microbiome would benefit from the development of standardized protocols as a means of minimizing the variability caused by the application of different methods between studies[10,11]. Here, we discuss many of the above topics and offer suggestions as to best practices where appropriate.

The protocol demonstrates the process of collecting soil, rhizosphere, and root samples from *Sorghum bicolor* and extracting DNA using a well-established DNA isolation kit[11]. Additionally, our protocol includes a detailed amplicon sequencing workflow, using a commonly utilized NGS platform, to determine the structure of the bacterial communities[12,13,14]. This protocol has been validated for the use in a wide range of plant hosts in a recent published study of the roots, rhizosphere, and associated-soils of 18 monocot species including *Sorghum bicolor, Zea mays,* and *Triticum aestivum*[15]. This method has also been validated for use with other marker genes, as demonstrated by its successful application to studying the fungal ITS2 marker gene in studies of the agave microbiome[16,17] and strawberry microbiome[18].

Protocol

1. Collection and Separation of Root Endosphere, Rhizosphere, and Soil Samples

1. Prior to entering the field, autoclave ultrapure water (at least 90 mL of water per sample) to sterilize. Prepare epiphyte removal buffer (at least 25 mL per sample) by adding 6.75 g of KH_2PO_4, 8.75 g of K_2HPO_4, and 1 mL of Triton X-100, to 1 L of sterile water. Sterilize the buffer using a vacuum filter with 0.2 µm pore size.
    1. For steps 1.2 to 1.5, wear clean gloves sterilized with ethanol at all times and replace the gloves between each sample to prevent contamination. Sterilize all equipment with 70% ethanol and wipe clean all equipment between samples. Before sampling, determine the optimal sampling depth for your experiment, and be consistent with all soil and root collections.

2. To collect bulk soil samples, use an ethanol-sterilized soil core collector to obtain soil that is free of plant roots by collecting a core approximately 23 to 30 cm from the base of the plant.
3. Transfer the soil to a plastic bag, homogenize the soil by gentle shaking, and transfer an aliquot of the soil sample (approximately 600 mg) to fill one 2 mL tube. Immediately place the 2 mL tube on dry ice or flash freeze the tube in liquid N_2 until ready to proceed with DNA extraction (step 2).
    1. In some environments, the surrounding soil can contain plant material. In this case, use a sterilized 2 mm sieve to separate the plant debris from the soil prior to placing it in the plastic bag.

4. To collect the root and rhizosphere, use an ethanol-sterilized shovel to dig up the plant, taking care to obtain as much of the root as possible. Depth is dependent upon the plant. While small plants such as wheat can be removed by digging several centimeters, larger plants such as sorghum may require 30 cm or more. Gently shake off excess soil from the roots until there is approximately 2 mm of soil adhering to the root surface.
   NOTE: Take care when working with small plants, with fragile roots, or in dry, high-clay content soils. Ideally, there should only be a thin layer of soil remaining on the roots after shaking. If large aggregates of soil remain, a rubber mallet can be used to dislodge the soil by gently hitting the base of the shoot. If the amount of soil remaining after this process exceeds or falls short of 2 mm, the approximate thickness should be noted.
5. For large plants, use sterile scissors and/or shears to cut a representative subsection of roots and place a minimum of 500 mg of root tissue into a 50 mL conical vial. For smaller grasses, place the entire root system into the vial. Add enough epiphyte removal buffer to cover the roots, then immediately place the sample on dry ice or flash freeze the sample in liquid N_2.
   NOTE: Take care not to overfill the 50 mL conical vial, as it will make washing step difficult. There should be enough empty space such that the epiphyte buffer is able to flow to the bottom, surround the roots throughout the vial, and cover the top. Because some grasses have more root biomass than will fit into a 50 mL conical vial, a subsection of the roots should be collected. However, it should be noted that cutting the roots could lead to endophytic bacteria being washed out into the rhizosphere fraction, so breaking roots should be minimized. If samples are not processed immediately after returning to lab, they can be stored at -80 °C.
6. To separate the rhizosphere from the roots, thaw the root sample on ice, then sonicate the root samples at 4 °C for 10 min with pulses of 160 W for 30 s, separated by 30 s. Transfer the roots into a chilled (4 °C), clean 50 mL tube using sterile forceps. Do not dispose the original tube with buffer and soil, which is the rhizosphere fraction (**Figure 1**).
7. Centrifuge the tube containing buffer and rhizosphere for 10 min at 4 °C, 4,000 x g. Decant the supernatant, flash freeze the tube containing the rhizosphere fraction in liquid N_2, and store the rhizosphere fraction at -80 °C until ready to proceed with DNA extraction (step 2).
8. To wash the roots, add approximately 20 mL of chilled (4 °C) sterile water to the root fraction. Wash the root by shaking vigorously (by hand or mixer, for 15 - 30 s), and then drain the water.
9. Repeat this step at least twice, until no soil remains on the root surface. If the DNA extraction (step 2) is not performed immediately, wrap the roots in sterile aluminum foil, flash freeze the roots in liquid N_2, and store the samples at -80 °C until ready to proceed with DNA extraction.

2. DNA Extraction

NOTE: Throughout steps 2 and 3, clean gloves sterilized with ethanol should be worn at all times and all work should be performed on a surface sterilized with ethanol.

1. Extract DNA from the soil and rhizosphere samples.
    1. Use a sterile spatula to quickly transfer 250 mg of soil and rhizosphere from steps 1.3 and 1.7 into separate collection tubes provided in a commercial DNA isolation kit designed for extraction from soil, then proceed with DNA isolation using the kit supplier's protocol.
    2. After eluting the DNA in the elution buffer supplied by the DNA isolation kit, store the DNA at -20 °C until ready to proceed with step 3.

2. Extract DNA from the root samples.
    1. Chill a sterilized mortar and pestle using liquid N_2. Measure out 600 to 700 mg of root tissue and place the tissue into the mortar. Carefully, add enough liquid N_2 to cover the roots.
    2. Grind the roots into small pieces. Continue the process of adding liquid N_2 and grinding (at least two times, be consistent between samples), until the roots are a fine powder. Ensure that the root tissue does not thaw during this step.
    Caution: Use appropriate personal protective equipment (lab coat, protective eyewear, and cryogenic gloves) when working with liquid N_2.
    NOTE: For a low-quality DNA extraction, it can be beneficial to grind excess roots into powder and store the powder at -80 °C.
    3. Quickly, before the root powder begins to thaw, use a sterile spatula to transfer the root powder into pre-weighed 1.5 mL tubes on ice. Record the weight of the tube and powder. Typically, 300 - 400 mg of powder is transferred.
    4. Use a sterile spatula to quickly transfer 150 mg of root powder to the collection tube provided in a commercial DNA isolation kit designed for extraction from soil, then proceed with DNA isolation using the kit supplier's protocol.
    NOTE: For some root samples, there can be a high concentration of organics remaining in the DNA pellet, which prevents the amplification of the DNA during PCR, especially when a different DNA extraction protocol (*e.g.*, CTAB extraction) is used. If necessary, clean the DNA by following the instructions provided in the environmental DNA clean-up kit.

3. Measure the concentration of all DNA samples using a high-sensitivity benchtop fluorometer.
    1. Add 1 - 20 µL of each eluted DNA sample into tubes provided in the dsDNA high-sensitivity assay kit. Add fluorometer working solution (1:200 dye:buffer) up to 200 µL.
    2. Prepare two additional tubes containing 10 µL of DNA standard 1 (0 ng/µL DNA) or 10 µL of standard 2 (100 ng/µL), and add 190 µL of fluorometer working solution to each standard.
    3. Measure the concentration of the standards and each sample. If it is not done automatically, calculate the DNA concentration from the absorbance output by a linear regression of the two standards.

3. Amplicon Library Preparation and Submission

1. Set up materials for the amplification reaction.
    1. Thaw DNA samples at 4 °C and keep them on ice throughout step 3. Randomize the order of DNA samples to minimize bias due to the location on the PCR plate (**Table 2**.)
    2. In a 96-well PCR plate, dilute DNA from each sample in molecular-grade water to 5 ng/µL in a total volume of 20 µL. Add 20 µL of molecular-grade water to the four corner wells as negative controls for amplification (blanks) (**Table 2**).
    3. Arrange the barcoded primers (10 µM) in either PCR strip tubes or a 96-well plate such that they can be added with a multi-channel pipette (**Figure 2**).
    4. Prepare sufficient PCR master mix to amplify each DNA sample in triplicate. Prepare 1.5 µL of BSA (20 mg/mL), 37.5 µL of pre-made 2x master mix (composed of PCR buffer, $MgCl_2$, dNTPs, and *Taq* DNA polymerase), 0.57 µL of chloroplast PNA (100 µM), 0.57 µL of mitochondrial PNA (100 µM), and 25.86 µL of molecular grade water.
    5. Pour the master mix into a sterile 25 mL of multichannel pipette reservoir and distribute 66 µL of master mix into each well of a new 96-well PCR plate using a multichannel pipette.
    NOTE: When calculating reagent volumes for the master mix, make sure to also include the 4 blank wells per plate.
    6. Using a multi-channel pipette, add 6 µL of 5 ng/µL DNA (from the normalized DNA plate) to the master mix plate. Then add to the master plate 1.5 µL of 10 µM forward primer such that each column has a different forward barcode, and 1.5 µL of 10 µM reverse primer such that each row has a different reverse barcode (**Figure 2**).
    NOTE: Prior to adding primers, the randomized plates and master mix could be used to amplify the ITS or ITS2 fungal genes if different primers were added. If this is the case, a similar primer design can be used.
    7. Spin down the plate briefly at 3,000 x g. Use a multi-channel pipette to mix gently, then divide into three plates with 25 µL of reaction mixture.
    NOTE: Although three replicates are not strictly necessary, it decreases the impact of technical variability.

2. Amplify the DNA in each plate using a thermocycler set to the following conditions: 180 s at 98 °C, 30 cycles of: 98 °C for 45 s (denaturing), 78 °C for 10 s (PNA annealing), 55 °C for 60 s (primer annealing), and 72 °C for 90 s (extension), then 600 s at 72 °C followed by a 4 °C hold step. After the amplification, pool the three replicate plates into one single 96-well plate.
3. Quantify the DNA using high-sensitivity fluorometer reagents in a 96-well plate reader.
    1. Add 2 µL of each PCR product to a 96-well microplate, along with 98 µL of fluorometer working solution (1:200 dye:buffer). Include 4 wells as standards: 5 µL of DNA standard 1 (0 ng/µL DNA), 1 µL of standard 2 (10 ng/µL), 2 µL of standard 2 (20 ng/µL), and 5 µL of standard 2 (50 ng/µL). Then add fluorometer working solution for a final volume of 100 µL.
    NOTE: Each sample can be measured using a benchtop fluorometer as described in step 2.3 if a plate reader is not available.
    2. Calculate the DNA concentration from the absorbance output by a linear regression of the four standards.
    3. For the successfully amplified barcoded products (those that have a concentration greater than 15 ng/µL), pool 100 ng of each sample into a single 1.5 mL tube (**Table 2**).
    4. Calculate the average volume of samples added to the pool by using the =AVERAGE() function in a spreadsheet program. Add the volume of the "blank" PCR products to the pooled samples.
    NOTE: Since the "blank" PCR products have their own unique barcode combinations, they can be sequenced to check for any laboratory contaminants.

4. Measure the concentration of the pooled product using a benchtop fluorometer as described in step 2.3, and take 600 ng of DNA and dilute in molecular-grade water to a final volume of 100 µL in a 1.5 mL tube. Store the remaining pooled product at -20 °C.
5. Wash the 600 ng DNA aliquot by following the established PCR purification process with paramagnetic purification beads in a 96-well format with a few exceptions.
    1. Make a fresh 600 µL aliquot of 70% ethanol. Shake the bottle of magnetic beads to re-suspend beads that settle to the bottom.
    2. Add 1x volume (100 µL) of bead solution to the 600 ng aliquot of DNA. Mix thoroughly by pipetting 10 times. Incubate for 5 min at room temperature.
    3. Place the tube onto the magnetic stand for 2 min (or until solution is clear) to separate beads from solution. While the tube is still in the magnetic stand, aspirate the clear supernatant carefully without touching the magnetic beads, and discard the clear supernatant.
    NOTE: At this point, the amplicon products are bound to the magnetic beads. Any beads that are disturbed or lost during aspiration will result in a loss of DNA.
    4. Leave the tube in the magnetic stand and add 300 µL of 70% ethanol to the tube; incubate at room temperature for 30 s. Aspirate out the ethanol and discard. Repeat this process, and remove all ethanol after the second wash. Remove the tube from the magnetic stand, and air dry for 5 min.
    5. Add 30 µL of molecular-grade water to the dried beads and mix by pipetting 10 times. Incubate at room temperature for 2 min. Return the tube to the magnetic stand for 1 min to separate the beads from solution. Transfer the eluate to a new tube.
    NOTE: Magnetic beads will not affect downstream reactions.
6. Measure the final concentration of cleaned, pooled DNA using a benchtop fluorometer as described in step 2.3. Dilute an aliquot to 10 nM in a final volume of 30 µL, or to the concentration and volume preferred by the sequencing facility.
7. Utilize the services of a sequencing facility to sequence the DNA on a NGS platform, 2 x 300 bp paired-end sequencing.

Representative Results

Performing the recommended protocol should result in a dataset of indexed paired-end reads that can be matched back to each sample and assigned to either a bacterial operational taxonomic units (OTU) or exact sequence variant (ESV, also referred to as amplicon sequence variant (ASV) and sub-operational taxonomic unit (sOTU)), depending on downstream analysis. In order to obtain high-quality sequence data, care must be taken at each step to maintain consistency between samples and minimize the introduction of any potential bias during the sample processing or library preparation. After collecting, processing, and extracting DNA from samples (steps 1 and 2), the resulting eluate should appear clear and free of organics that would inhibit amplification. While purity can be verified by measuring each DNA sample via a microvolume spectrophotometer, we have found that the soil DNA extraction kit reliably removes all contaminants. As a result of the predictable DNA quality, quantification methods that rely on fluorescence-based dyes that specifically bind DNA are more appropriate than those based on UV absorbance[19,20,21]. Prior to PCR amplification, soil and rhizosphere samples average around 10 ng/µL DNA, while root samples typically have a mean concentration of approximately 30 ng/µL (**Table 2**).

Following the amplification of the environmental DNA (step 3), success or failure can be determined by measuring the concentration of the PCR product via benchtop fluorometer reagents on a plate reader, if available, or manually (**Table 2**). In our experience, successful amplifications that result in high-quality amplicon data yield greater than 15 ng/µL PCR products. If there are multiple failures on a plate, the positional arrangement within the plant and the sample type of failed samples may help determine the problem. For instance, if they are all adjacent on the plate, it may indicate pipette error, whereas if they are all in the same row or column, it could suggest issues with a specific primer. If they all belong to the same sample type, it might suggest problems with sample processing or DNA extraction.

It is important to check the compatibility of the universal PNAs with your specific plant system bioinformatically during experimental design in order to verify that they will block amplification of chloroplast and mitochondrial 16S genes. Following the amplification step, it is not clear whether the PNAs successfully bound to mitochondrial and chloroplast templates; this is only revealed after sequencing (**Figure 3**). To help ensure that the PNAs will effectively block contaminant amplification, an alignment of the PNA sequence to each chloroplast and mitochondrial 16S rRNA gene (there may be multiple copies) for the plant host being investigated should not reveal any mismatches. Even a single mismatch to the 13 bp PNA sequence, especially in the middle of the PNA clamp, can drastically reduce the effectiveness, as in the case of the provided chloroplast PNA sequence and the chloroplast 16S rRNA gene of *Lactuca sativa* (lettuce) (**Figure 3**).

Since an equal amount of amplified DNA is pooled per sample, there should be an approximately even number of reads obtained per sample after sequencing and sorting reads based on their barcoded index (**Figure 4**). The majority of these reads should match to bacterial taxa. Any eukaryotic, mitochondrial, or chloroplast matches should be discarded. Depending on the analysis pipeline and taxonomic database chosen, chloroplast and mitochondrial reads can mistakenly be assigned to bacterial lineages, often *Cyanobacteria* and *Rickesttia*, respectively (**Figure 3**). A degree of manual curation is often prudent to check for these common mis-assignments. Specific details will depend on the choice of analysis, but relative abundance profiles should generally be similar (no significant difference) among biological replicates and significantly different between soil, rhizosphere, and root samples (**Figure 5**). It is important to note, however, that while there may be no significant difference between biological replicates, it is important to collect at least three replicates per.

Methods for interpreting the data obtained in these experiments are hotly debated amongst microbial ecologists. Until recently, amplicon sequence analysis has been dependent upon grouping reads into OTUs. However, these are problematic because: 1) they are based on a somewhat arbitrary threshold of 97% similarity, 2) diversity is often underestimated, and 3) there can be low taxonomic resolution. Recently developed tools such as DADA2, Deblur, and UNOISE2[22,23,24] are able to sort reads into ESVs, which solves some problems presented when using OTUs. Caveats to using ESVs include: 1) artificial increases in diversity due to the differences in rRNA copies within a species, and 2) increased sensitivity to PCR and sequencing errors[25,26].

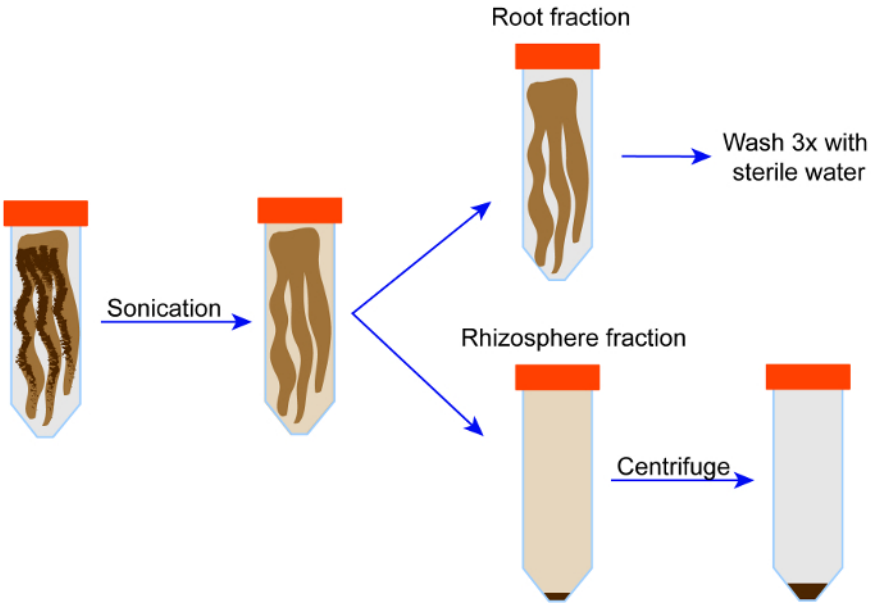

Figure 1: Separation of root and rhizosphere fractions. Flowchart displaying the steps for separating the rhizosphere from the root samples, followed by washing the roots with sterile water to remove any remaining rhizoplane organisms. Please click here to view a larger version of this figure.

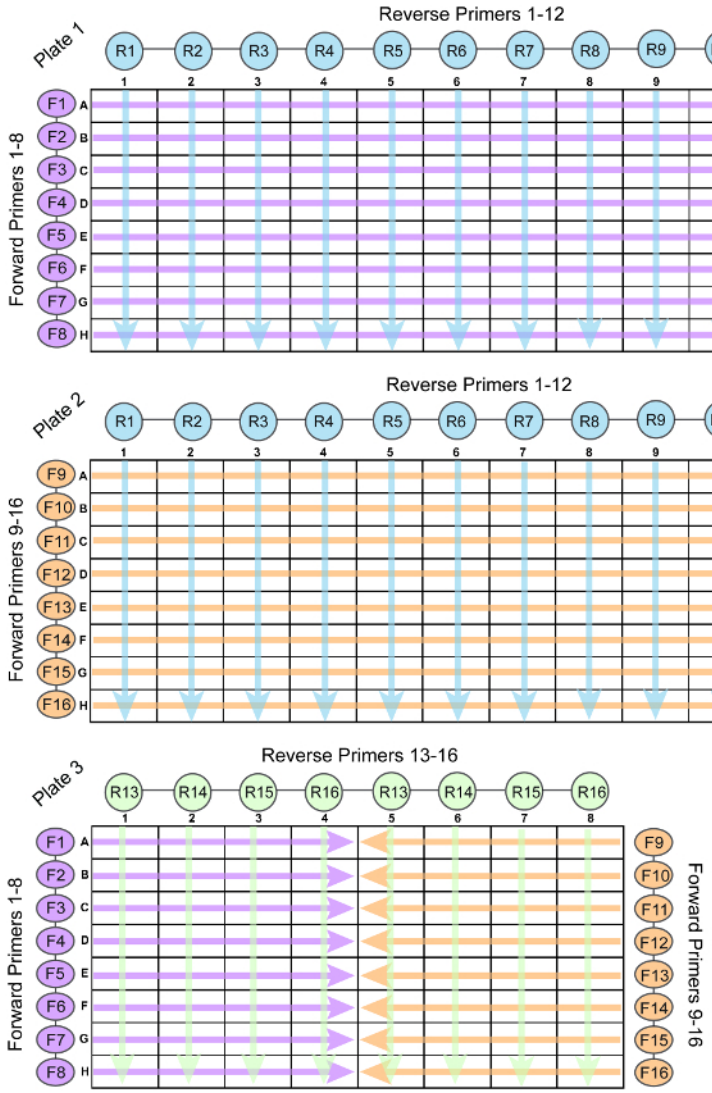

Figure 2: Example of stock primer layout for amplification and distribution within plates. Stock primers (**Table 1**) can be prepared in strip tubes for optimal distribution within 96-well plates (each strip of primers is represented by a different color; purple for forward primers 1 - 8, orange for forward primers 9 - 16, blue for reverse primers 1 - 12, and green for reverse primers 13 - 16.) In this case, 16 forward and 16 reverse primers can be distributed efficiently with a multi-channel pipette such that each well has a unique barcode combination. Please click here to view a larger version of this figure.

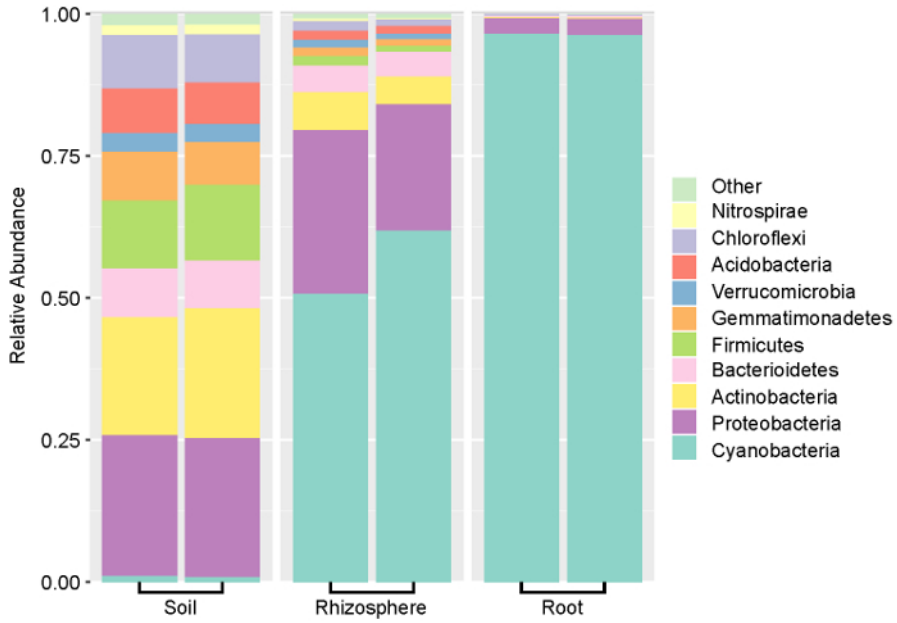

Figure 3: Results that suggest chloroplast PNA is ineffective. Representative result from rhizosphere ("Rhizo"), root, and soil samples from lettuce that were non-treated (NT) or treated (VT) with a biological soil amendment. The PNA sequence used to block chloroplast contamination of most plants is GGCTCAACCCTGGACAG[27]. However, lettuce contains a mismatch in the chloroplast 16S ribosomal RNA gene (GGCTCAAC**T**CTGGACAG). This renders the PNA ineffective, resulting in a high relative abundance of reads that match to Cyanobacteria in rhizosphere and root samples. Please click here to view a larger version of this figure.

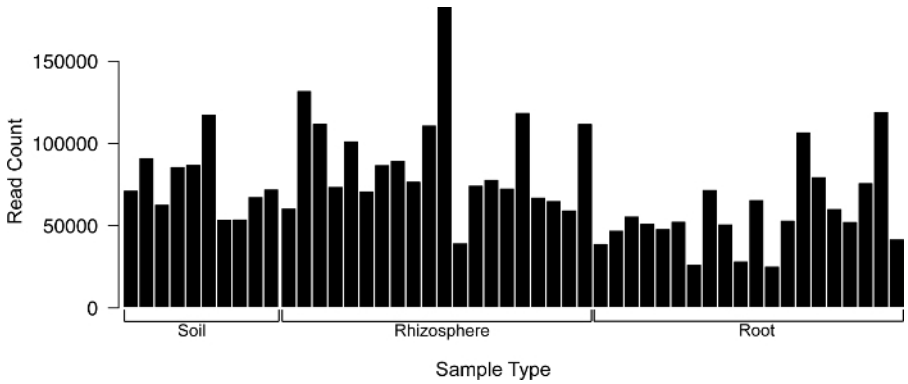

Figure 4: Distribution of read counts among samples in a library. Bar chart showing number of read counts (y-axis) from different samples (bars, x-axis), matched by the barcode combination in the read. The number of reads per sample can vary based on how many samples are in the library; this subset was sequenced in a library of 192 samples. Please click here to view a larger version of this figure.

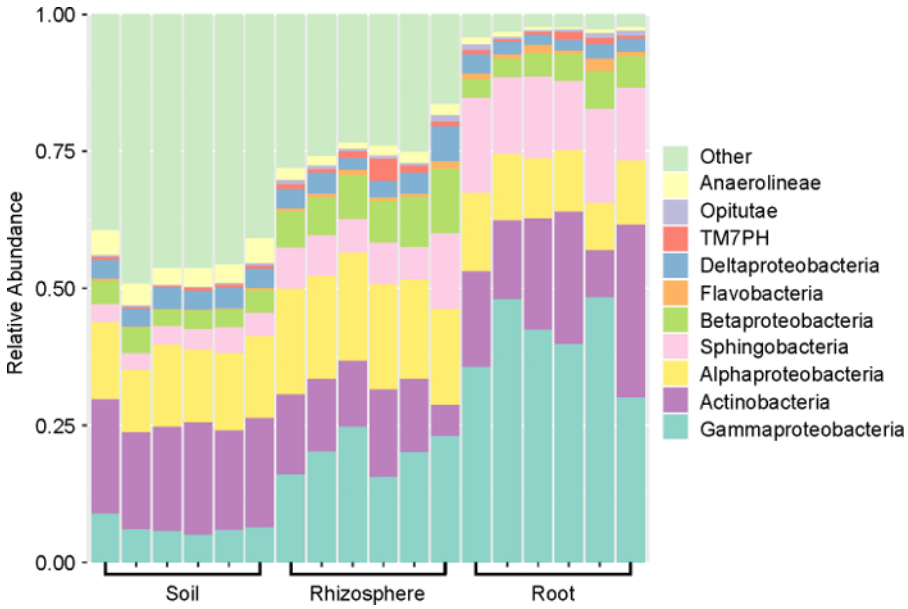

Figure 5: Relative abundance of the top 12 classes in root, rhizosphere, and soil communities. Stacked bar chart showing relative abundance of classes present in a representative 16S dataset containing 6 replicates for each sample type (bulk soil, rhizosphere, and root endosphere). Please click here to view a larger version of this figure.

Table 1: Primers for amplifying the V3-V4 region of the 16S rRNA gene. Primers are composed of, sequentially: an adapter for a common NGS platform, a unique barcode, the primer for NGS, a spacer region of variable length to shift the frame for sequencing, and a universal PCR primer that amplifies either 341F or 785R of the 16S rRNA gene. The number of primers needed is dependent upon how many samples are sequenced per library; a combination of 16 forward and 16 reverse primers is sufficient for 244 samples (256 primer combinations with 12 used for blank wells during PCR (**Figure 2**)). Please click here to download this file.

Table 2: Normalization of randomized DNA samples prior to and following amplification. Example worksheet listing samples in a randomized order and indicating their location on a single 96-well plate, which also determines the primer combination assigned to it. Formulas in the bottom row describe calculations for adding 100 ng of each sample to the normalized plate, plus the volume of water to reach 20 µL. Following amplification, the volume of 100 ng of each successful product is calculated and added to a final pool. The volume of "blank" PCR product to add to the final pool is the average of the other samples. Please click here to download this file.

Supplemental Figure 1: Approximation of minimum root biomass during sample collection. When collecting roots, try to collect at least 500 mg of tissue. Here, roots collected from a young sorghum plant (left, in both A and B) and a young rice plant (right) are shown next to (A) and inside (B) 50 mL conical tubes. Both samples weigh approximately 1 g, however, it is important to note that this weight includes rhizosphere and root, and the rhizosphere weight is, in this case, approximately half the total weight. Please click here to download this file.

Discussion

This protocol demonstrates an established pipeline for exploring root endosphere, rhizosphere, and soil microbial community compositions, from field sampling to sample processing and downstream sequencing. Studying root-associated microbiomes presents unique challenges, due in part to the inherent difficulties in sampling from soil. Soils are highly variable in terms of physical and chemical properties, and different soil conditions can be separated by as little as a few millimeters[28,29]. This can lead to the samples which are collected from adjacent sampling sites having considerably different microbial community compositions and activities[30,31]. Thus, using soil core collectors and shovels to maintain consistent sampling depths and homogenization prior to DNA extraction are essential to the reproducibility within root microbiome studies.
It is also essential to efficiently separate the rhizosphere and root fractions; using a harsh method of root surface sterilization can potentially lyse endophytes within roots prior to DNA extraction, while a more conservative wash may not remove all microbes from the root surface[7].
Another key factor that can negatively impact or disrupt sequencing results is bacterial contamination, which can come from many sources and is sometimes impossible to distinguish from the sampled environmental bacteria[32,33]. For this reason, careful sterilization of sampling tools, experimental materials, and working environments are vital in order to avoid contamination.

After sampling, obtaining high quality DNA is a high priority for successful downstream analyses. In our experience, DNA extraction from field grown root samples through alternative methods, such as through CTAB-based extraction, often contain substantially greater quantities of humic acids and other compounds compared to rhizosphere and soil samples. These compounds can prevent the enzymatic activity of the DNA polymerase during PCR amplification, even at low concentrations[34,35]. Using DNA extraction kits designed for soils on root samples, as opposed to a CTAB extraction followed by a phenol chloroform clean-up, can effectively rid samples of humic acids and will result in high quality DNA[36,37,38,39]. Accordingly, we recommend using a commercially available DNA extraction kit for root samples as well. It should be noted that the goal is to obtain microbial genomic DNA from plant roots. Thus, thorough and consistent root grinding is important to break down the plant tissue and lyse the microbial cells to release microbial DNA without introducing bias between samples due to the variation in grinding pressure and time.

Following careful extraction of DNA from samples, there are two main sources for problems during amplification: 1) contamination of plant tissues with plant endosymbionts (chloroplast and mitochondria) and 2) selection of 16S rRNA region to amplify. The amplification from chloroplast or mitochondria 16S rRNA sequences can generate >80% of the sequences in root samples[40], and more in leaf tissues, though the amount of contamination is dependent on the choice of primers. Thus, PNA clamps are necessary during the PCR step to suppress plant host chloroplast and mitochondrial 16S contamination[27,41]. However, different plant species can have variation in the chloroplast and mitochondrial 16S sequence[27]; therefore, it is essential to confirm the sequence of the chloroplast and mitochondrial 16S genes of the plant being studied prior to library sequencing, in order to determine if alternate PNA oligos are needed (**Figure 3**). Additionally, the 16S rRNA gene consists of nine hypervariable regions flanked by nine conservative regions; different results can be obtained from the same community depending upon which hypervariable region is amplified[42]. Previous studies have found the V4 region to be one of the most reliable for assigning taxonomy[43] and it has been used for other extensive microbiome surveys[11]. Lengthening the target to the V3-V4 region is suggested here to increase variability and improve taxonomic resolution.

In this protocol, we demonstrated a pipeline to perform 16S rRNA amplicon sequencing via next generation sequencing (NGS) for studying microbial community compositions of environmental samples[12]. We recommend using amplicon sequencing as a tool for phylogenetic profiling, because it is relatively inexpensive, high-throughput, and does not require extensive computational expertise or resources to analyze. While our method focuses on analyzing the bacterial fraction of the microbiome, it can easily be adapted to investigate fungi. The protocol is identical through step 2, and the only difference in step 3 is what primers would be used during the amplification. However, it is worth nothing that amplicon based profiling is not without limitations. By sequencing a single marker gene, no information is obtained regarding the functional capacity of the community. Additionally, the taxonomic resolution can be quite low, especially when sequencing from environments with a high percentage of uncharacterized microbes. However, sequencing technologies are rapidly evolving, and we anticipate the potential to deal with some of these shortcomings by adapting this protocol for use with other sequencing platforms. Finally, as mentioned in the introduction, shotgun metagenomics and metatranscriptomics can easily be performed on soil and rhizosphere samples, and methods to eliminate plant contamination from plant tissues are currently being explored. Experimental designs which pair amplicon-based approaches and other metagenomic techniques can be particularly effective in complex communities where high species diversity and uneven representation of taxa can prevent shotgun data from accurately characterizing the less dominant members.

Disclosures

The authors have nothing to disclose.

Acknowledgements

This work was funded by the USDA-ARS (CRIS 2030-21430-008-00D). TS is supported by the NSF Graduate Research Fellowship Program.

References

1. Goodwin, S., McPherson, J.D., McCombie, W.R. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* **17** (6), 333-351 (2016).
2. Soergel, D.A.W., Dey, N., Knight, R., Brenner, S.E. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J.* **6** (7), 1440-1444 (2012).
3. Takahashi, S., Tomita, J., Nishioka, K., Hisada, T., Nishijima, M. Development of a Prokaryotic Universal Primer for Simultaneous Analysis of Bacteria and Archaea Using Next-Generation Sequencing. *PLoS One.* **9** (8), e105592 (2014).
4. Poretsky, R., Rodriguez-R, L.M., Luo, C., Tsementzi, D., Konstantinidis, K.T. Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. *PLoS One.* **9** (4), e93827 (2014).
5. Sharpton, T.J. An introduction to the analysis of shotgun metagenomic data. *Front Plant Sci.* **5**, 209 (2014).
6. Jiao, J.-Y., Wang, H.-X., Zeng, Y., Shen, Y.-M. Enrichment for microbes living in association with plant tissues. *J Appl Microbiol.* **100** (4), 830-837 (2006).
7. Richter-Heitmann, T., Eickhorst, T., Knauth, S., Friedrich, M.W., Schmidt, H. Evaluation of Strategies to Separate Root-Associated Microbial Communities: A Crucial Choice in Rhizobiome Research. *Front Microbiol.* **7**, 773 (2016).
8. Mahmoudi, N., Slater, G.F., Fulthorpe, R.R., Mahmoudi, N., Slater, G.F., Fulthorpe, R.R. Comparison of commercial DNA extraction kits for isolation and purification of bacterial and eukaryotic DNA from PAH-contaminated soils. *Can J Microbiol.* **5709**, 623-628 (2011).
9. Vishnivetskaya, T.A. *et al.* Commercial DNA extraction kits impact observed microbial community composition in permafrost samples. *FEMS Microbiol Ecol.* **87** (1), 217-230 (2014).
10. Busby, P.E. *et al.* Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLoS Biol.* **15** (3), e2001793 (2017).
11. Thompson, L.R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature.* (2017).
12. Caporaso, J.G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6** (8), 1621-1624 (2012).
13. Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Appl Environ Microb.* **79** (17), 5112-5120 (2013).
14. Degnan, P.H., Ochman, H. Illumina-based analysis of microbial community diversity. *ISME J.* **6** (1), 183-194 (2012).
15. Naylor, D., DeGraaf, S., Purdom, E., Coleman-Derr, D. Drought and host selection influence bacterial community dynamics in the grass root microbiome. *ISME J.* (2017).
16. Desgarennes, D., Garrido, E., Torres-Gomez, M.J., Peña-Cabriales, J.J., Partida-Martinez, L.P. Diazotrophic potential among bacterial communities associated with wild and cultivated *Agave* species. *FEMS Microbiol Ecol.* **90** (3), 844-857 (2014).
17. Coleman-Derr, D. *et al.* Plant compartment and biogeography affect microbiome composition in cultivated and native *Agave* species. *New Phytol.* **209** (2), 798-811 (2016).

18. De Tender, C. *et al.* Dynamics in the Strawberry Rhizosphere Microbiome in Response to Biochar and Botrytis cinerea Leaf Infection. *Front Microbiol.* **7**, 2062 (2016).
19. Kapp, J.R. *et al.* Variation in pre-PCR processing of FFPE samples leads to discrepancies in BRAF and EGFR mutation detection: a diagnostic RING trial. *J Clin Pathol.* **68** (2), 111-118 (2015).
20. Simbolo, M. *et al.* DNA qualification workflow for next generation sequencing of histopathological samples. *PLoS One.* **8** (6), e62692 (2013).
21. O'Neill, M., McPartlin, J., Arthure, K., Riedel, S., McMillan, N. Comparison of the TLDA with the Nanodrop and the reference Qubit system. *J Phys Conf Ser.* **307** (1), 012047 (2011).
22. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods.* **13** (7), 581-583 (2016).
23. Amir, A. *et al.* Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems.* **2** (2), e00191-16 (2017).
24. Edgar, R.C. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv.* 081257 (2016).
25. Nguyen, N.-P., Warnow, T., Pop, M., White, B. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *NPJ Biofilms Microbiomes.* **2**, 16004 (2016).
26. Callahan, B.J., McMurdie, P.J., Holmes, S.P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* (2017).
27. Lundberg, D.S., Yourstone, S., Mieczkowski, P., Jones, C.D., Dangl, J.L. Practical innovations for high-throughput amplicon sequencing. *Nat Methods.* **10** (10), 999-1002 (2013).
28. O'Brien, S.L. *et al.* Spatial scale drives patterns in soil bacterial diversity. *Environ Microbiol.* **18** (6), 2039-2051 (2016).
29. Fierer, N., Lennon, J.T. The generation and maintenance of diversity in microbial communities. *Am J Bot.* **98** (3), 439-448 (2011).
30. Fierer, N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat Rev Microbiol.* **15** (10), 579-590 (2017).
31. Buckley, D.H., Schmidt, T.M. Diversity and dynamics of microbial communities in soils from agro-ecosystems. *Environ Microbiol.* **5** (6), 441-452 (2003).
32. Salter, S.J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87 (2014).
33. Weiss, S., Amir, A., Hyde, E.R., Metcalf, J.L., Song, S.J., Knight, R. Tracking down the sources of experimental contamination in microbiome studies. *Genome Biol.* **15** (12), 564 (2014).
34. Sutlović, D., Definis Gojanović, M., Andelinović, S., Gugić, D., Primorac, D. Taq polymerase reverses inhibition of quantitative real time polymerase chain reaction by humic acid. *Croat Med J.* **46** (4), 556-562 (2005).
35. Sutlovic, D., Gamulin, S., Definis-Gojanovic, M., Gugic, D., Andjelinovic, S. Interaction of humic acids with human DNA: proposed mechanisms and kinetics. *Electrophoresis.* **29** (7), 1467-1472 (2008).
36. Aleklett, K., Leff, J.W., Fierer, N., Hart, M. Wild plant species growing closely connected in a subalpine meadow host distinct root-associated bacterial communities. *PeerJ.* **3**, e804 (2015).
37. Bogas, A.C. *et al.* Endophytic bacterial diversity in the phyllosphere of Amazon Paullinia cupana associated with asymptomatic and symptomatic anthracnose. *Springerplus.* **4**, 258 (2015).
38. Hiscox, J., Savoury, M., Müller, C.T., Lindahl, B.D., Rogers, H.J., Boddy, L. Priority effects during fungal community establishment in beech wood. *ISME J.* **9** (10), 2246-2260 (2015).
39. Zhang, T., Yao, Y.-F. Endophytic Fungal Communities Associated with Vascular Plants in the High Arctic Zone Are Highly Diverse and Host-Plant Specific. *PLoS One.* **10** (6), e0130051 (2015).
40. Ghyselinck, J., Pfeiffer, S., Heylen, K., Sessitsch, A., De Vos, P. The effect of primer choice and short read sequences on the outcome of 16S rRNA gene based diversity studies. *PLoS One.* **8** (8), e71360 (2013).
41. Wintzingerode, F., Landt, O., Ehrlich, A., Göbel, U.B. Peptide nucleic acid-mediated PCR clamping as a useful supplement in the determination of microbial diversity. *Appl Environ Microb.* **66** (2), 549-557 (2000).
42. Cruaud, P., Vigneron, A., Lucchetti-Miganeh, C., Ciron, P.E., Godfroy, A., Cambon-Bonavita, M.-A. Influence of DNA extraction method, 16S rRNA targeted hypervariable regions, and sample origin on microbial diversity detected by 454 pyrosequencing in marine chemosynthetic ecosystems. *Appl Environ Microb.* **80** (15), 4626-4639 (2014).
43. Yang, B., Wang, Y., Qian, P.-Y. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics.* **17**, 135 (2016).

