Plant Microbiome Identification and Characterization

Sarah L. Lebeis[1]

[1]Department of Microbiology, University of Tennessee, Knoxville, Tennessee

To fully exploit the potential of plant microbiome alterations to improve plant health, reliable methods must be used to prepare and characterize microbiome samples. The power of culture-independent studies is that they allow the characterization of novel microbial community members, but only microbial members consistently represented between different research groups are likely to become broadly applicable treatments. The identification of plant microbiome members can be affected by several experimental stages, including design, sample preparation, nucleic acid extraction, sequencing, and analysis. The protocols described here therefore aim to highlight crucial steps that experimenters should consider before beginning a plant microbiome study. © 2017 by John Wiley & Sons, Inc.

Keywords: epiphyte • endophyte • plant microbiomes • phyllosphere • rhizosphere

INTRODUCTION

As with any type of scientific research, it is imperative to develop clearly defined hypotheses and experimental procedures to yield supporting or refuting data. The most common first set of studies to characterize a plant microbiome is amplicon sequencing of phylogenetically informative marker genes, such as bacterial and archaeal 16S ribosomal RNA (Caporaso et al., 2012), eukaryotic 18S rRNA (Stoeck et al., 2010), and the fungal intergenic transcribed spacer (ITS; Menkis et al., 2012). Without clear predictions, results are merely descriptive, which is especially dangerous when the phenotypes are inherently complex, such as clusters of sequences into Operational Taxonomic Units (OTUs) and their relative abundances. To avoid these pitfalls, experimental design should build off foundational studies that previously defined critical aspects of plant microbiomes, such as microbial diversity, density, and dynamics (Lebeis, 2015). The protocols presented here will ensure sample consistency and allow individual research groups to tailor their downstream sequencing and analysis methods to best suit their research questions. Although these protocols were created for working with the model plant *Arabidopsis thaliana* in the greenhouse (Bulgarelli et al., 2012; Lundberg et al., 2012), they can be amended for other plants, and even field experiments (Wagner et al., 2016).

DIFFERENTIAL HARVEST OF PLANT MICROBIOMES IN THE LABORATORY

A common theme to emerge from comparisons of plant microbiome studies is microbial communities vary between plant tissues (e.g., roots, leaves, and flowers) and between

BASIC PROTOCOL 1

Plant Microbiome Identification and Characterization

the inside (i.e., endophyte or endosphere) and outside (i.e., epiphyte in leaves/flowers, and rhizosphere in roots) of a tissue (Bodenhausen, Horton, & Bergelson, 2013; Bulgarelli et al., 2012; Lundberg et al., 2012; Shade, McManus, & Handelsman, 2013; Wagner et al., 2016). While microbes from leaves can colonize roots and vice versa (Bai et al., 2015), the presence of tissue-distinct communities in nature suggests that environmental stresses and nutrient levels play an role in microbial colonization of each tissue (Lebeis, 2015; Pieterse, de Jonge, & Berendsen, 2016; Vorholt, 2012). This protocol describes how to generate robustly distinct epiphytic and endophytic microbial communities.

Materials

Seeds
70% (v/v) ethanol with 0.01% (v/v) Triton-X-100
1% (v/v) bleach with 0.01% (v/v) Triton-X-100
Germinating plate medium (see recipe)
70% and 95% ethanol
Natural soil from field site
Sterile sand
Phosphate buffer with 0.02% silwet (see recipe)

Plant growth chamber
Petri dishes
Autoclavable pots
Miracloth (Millipore)
Aluminum foil
Flats and domes
Flame resistant pan, mallet, and mesh kitchen strainer
Gloves
Metal spatula
Metal tweezers and scissors
50-ml conical tubes
100-μm basket cell strainers (Fisher)
Squeeze bottle containing distilled water
Centrifuge
Liquid nitrogen
Bath ultrasonicator (Branson) or Biotruptor (Diagenode)

Grow plants and assemble microbiomes

1. Sterilize the seed surface and vernalize the appropriate time and temperature for the selected plant. For example, 1 min treatment in 70% (v/v) ethanol with 0.01% (v/v) Triton-X-100, 10 to 12 min in 1% (v/v) bleach with 0.01% (v/v) Triton-X-100, and several washes with sterile distilled water will sterilize most seeds.

   Some seeds do contain endophytes, which should be reduced by physically removing the seed coat or by a short heat treatment (e.g., 50° to 60°C for 1 hr) prior to step 1, to prevent founder's effects in the plant microbiome assembly.

2. Germinate sterile seeds on solid medium for 1 to 2 weeks, or until dicots have reached the two-leaf stage.

   Look for microbial contaminants on emerging seedlings as an indicator of contamination or seed born endophytes, but be sure to note the difference between root hairs and fungal contaminants. If microbial contamination is visible, start again with fresh seeds and increase the bleach sterilization time.

3. Prepare pots by placing a small piece of miracloth in the bottom of pots to prevent fine soil and particles from passing through drainage holes. Stack pots in an autoclavable bin, cover with foil, and sterilize in an autoclave on a dry cycle for at least 20 min.

    An initial study should include a minimum of 10 biological replicates in each test group to identify statistically robust differences.

4. To prepare soil inoculum, dry soil should be placed in a 95% (v/v) ethanol flame-sterilized large metal container and crushed using a flame-sterilized metal mallet. The resulting soil, which will likely contain dead organic material and rocks, should be sifted through a flame-sterilized stainless steel mesh kitchen strainer, resulting in a fine homogenous soil. Using gloved hands sterilized with 70% ethanol, mix 2 parts homogenized soil with 1 part sterile sand in a large sterile container to improve water drainage.

5. Fill sterile pots with the amended soil, and arrange in a flat.

6. Aseptically transfer sterile seedlings into these pots, leaving some pots unplanted as "soil-only" controls.

7. Water each pot with distilled water, and then water every 2 to 3 days.

    Decide if pots will be watered from the top or the bottom, and water them the same way throughout the experiment to prevent unwanted variance between samples. If pots are watered from the top with a squeeze bottle, air contaminants could be introduced. Watering from below—by filling the flat—will allow exchange of microbes, and differences between samples (particularly below ground) that develop over time, may be lost.

8. Cover flats with plastic domes for the first 3 to 5 days of the experiment, and grow in a greenhouse or plant growth chamber.

Harvest microbiome

9. Stop watering plants 2 to 3 days prior to harvest.

10. To harvest plants, invert pots onto a flame-sterilized glass or metal pan. Remove loose soil from roots using a sterile metal spatula or gloved hands wiped with 70% (v/v ethanol).

    If soil is too wet, the weight of the soil will break the root systems into pieces, decreasing root tissue yield. To determine if the soil is an appropriate harvesting consistency, squeeze the pot. Soil should move, but only separate from the pot at the edges. If the soil cracks down the middle of the pot without crumbling, the soil is not an appropriate moisture level for harvesting.

11. Separate leaves from roots using steriled tweezers and scissors, and place the tissue of interest for each plant into a 50-ml conical tube containing ~25 ml phosphate buffer with 0.02% silwet. Sterilize the pan and tools with fire between each sample.

    In this "microbiome harvest" protocol, keep each plant as an individual sample. If there is insufficient material for later protocols, samples may be pooled during nucleic acid extraction, library preparation, or during analysis (see Troubleshooting).

    If only DNA extraction is planned, the protocol can be performed at room temperature. For corresponding plant transcriptome or meta-transcriptome studies, keep the samples on ice and consider using an RNA stabilizing product.

    For field experiments, follow the microbiome harvest protocol as closely as possible for steps 9 to 11. Then store conical tubes on wet ice and transport samples back to the lab to perform steps 12 to 17 in the lab (Agler et al., 2016; Wagner et al., 2016).

12. Vortex conical tubes containing the samples and phosphate buffer for at least 30 sec to separate the rosettes or root systems from a subset of loosely attached external microbes.

13. Filter the resulting liquid slurry through a 100-μm cell strainer basket to remove the plant tissue, soil debris, and sand. The liquid that flows through can be considered the "leaf epiphyte" or "rhizosphere" samples.

    This is an operation definition for the leaf epiphyte and rhizosphere fractions. Thus, not all microbes will be able to be separated from their tissue. If the leaf epiphyte or rhizosphere composition varies between samples or cannot be differentiated from endosphere composition, increase the stringency of this step by changing the harvesting buffer or by vortexing for a longer time.

14. Centrifuge this liquid until a tight pellet forms (minimum 3200 × *g* for 12 min). Remove the supernatant and freeze pellets at −80°C until nucleic acid extraction (see Basic Protocol 3, step 6).

15. To remove additional debris from tissues, aseptically transfer the leaves or roots from the cell strainer basket (step 13) into 20 ml sterile water in a conical tube and vortex for 15 to 30 sec at least once. Use sterile tweezers to pick out visible debris that remains tangled in the roots. Place clean roots in a conical tube 1/3 full of phosphate buffer without silwet.

    Reserve tissue if starting a collection of microbial isolates (see Basic Protocol 2).

16. Remove tightly attached, external microbes from plant tissues by physical (Step 16a) or chemical treatment (Step 16b).

    a. Physical removal of tightly bound microbes: Place conical tube containing tissue in phosphate buffer into a bath sonicator on the lowest setting for 5 to 10 min.
    b. Chemical elimination of external microbes: Place tissue in bleach solution for 10 to 12 min, and wash several times with sterile distilled water.

    To choose between the two methods, consider the root structure of your plant. Thus, although the sonication method is less common, it may be preferred for root systems with fine roots (e.g., A. thaliana) that tangle during harvesting and leave tissue protected from bleach treatment.

    For RNA sequencing studies, consider skipping this step to get samples into liquid nitrogen as soon as possible.

17. Place tissue into a sterile tube, snap freeze by immersing the tube in liquid nitrogen, and store the sample at −80°C until nucleic acid extraction. Microbes associated with the tissue in the resulting samples are considered the "endophytic compartment" or "endosphere."

ISOLATION OF MICROBES FROM SURFACE STERILIZED PLANT TISSUE

To start a culture collection of endophytic bacteria or fungi, harvested tissue (from Basic Protocol 1, step 15) will be surface sterilized and plated on a variety of dilute medium. After microbes are isolated, they can be identified by 16S rRNA (for bacteria) or ITS (for fungi) gene sequencing, and preserved for long-term storage. These microbial strains can be used for future molecular mechanism experiments and genome sequencing.

Materials

Plant tissue (see Basic Protocol 1, step 15)
1% (v/v) bleach with 0.01% (v/v) Triton-X100
2.5% (w/v) sodium thiosulfate

Dilute bacterial agar media (e.g., PDA, R2A, 10% LB, 2% TSA, and 20% KB)
DNeasy UltraClean Microbial DNA Isolation Kit (Qiagen)
40% to 80% (v/v) glycerol
Oligonucleotides (e.g., 27F and 1492R for bacteria or ITS9F and ITS4R for fungi)

1. Sterilize the tissue (from Basic Protocol 1, step 15) in 1% (v/v) bleach with 0.01% (v/v) Triton-X100 for 10 to 12 min.

2. Rinse tissue in sterile distilled water.

3. Neutralize bleach with 2.5% (w/v) sodium thiosulfate for 5 min.

4. Rinse in 2 to 3 additional sterile distilled water washes.

5. Homogenize tissue and dilute 1:10 in sterile distilled water as described in Basic Protocol 3, steps 3 to 5, substituting fresh tissue for freeze-dried tissue.

6. Plate 50 to 100 μl of the tissue slurry to various dilute media, e.g., PDA, R2A, 10% LB, 2% TSA, and 20% KB.

   Dilute medium prevents overgrowth of faster growing microbes and enables the isolation of a larger diversity of microbes. Potato Dextrose Agar (PDA) is a common fungal medium. Reasoner's 2A agar (R2A) is a nutrient-poor medium often used to isolate environmental microbes. 10% Luria-Bertani (LB) medium, 2% Tryptic Soy Agar (TSA), and 20% King's B (KB) agar use the indicated percent nutrients with the standard amount of agar.

7. As colonies appear, streak them to rich medium (e.g., LB).

8. Pick isolated colonies and inoculate individual liquid LB cultures.

9. Once liquid cultures have grown turbid, culture may be used to prepare glycerol stocks for long-term storage (i.e., culture collection) or to extract DNA. For glycerol stocks, combine equal volumes of liquid culture and sterile glycerol in a freezer tube, mix thoroughly, snap freeze in liquid nitrogen, and store at −80°C. For DNA extraction, follow manufacturer's instructions (e.g., Qiagen's DNeasy UltraClean Microbial DNA Isolation Kit).

10. Use Sanger sequencing to obtain the full 16S rRNA gene sequence by using the 27F and 1492R primer pair for bacteria. An ITS primer pair (e.g., ITS9F and ITS4R) could be used to identify fungi.

TISSUE PREPARATION AND NUCLEIC ACID EXTRACTION

BASIC PROTOCOL 3

While neither culture-dependent nor culture-independent approaches are completely representative of the actual plant microbiome, is it important to minimize avoidable inconsistencies or diversity underestimations by completely homogenizing tissue and using the same nucleic acid extraction protocols.

Materials

Plant tissue and pellets (from Basic Protocol 1)
Nucleic acid extraction kit, e.g., DNeasy PowerSoil Kit (Qiagen)
DNA quantification kit, e.g., Qubit or Pico Green (Thermo Fisher Scientific)
PCR primers (see Table 1)
PCR Master Mix, e.g., HiFi HotStart Ready Mix (Kapa Biosystems)

Conical tubes (appropriate size for the volume of samples)
Dissecting probe, needle, or other flame-resistant sharp object
Liquid nitrogen

Table 1 Primer Pairs for Phylogenetic Analysis of Bacterial Components of Plant Microbiomes

Primer Pair	Variable region(s)	Uses	Strengths	Weaknesses
27F-1492R	V1-V9	Full length Sanger sequencing of isolates	Accurate	Bad for sample multiplexing
515F-806R	V4	454 and MiSeq (Caporaso et al., 2012; Lebeis, 2015; Lundberg et al., 2013; Tremblay et al., 2015)	High microbial accuracy	Few primer biases
779F-1192R	V5-V7	454 and MiSeq (Agler et al., 2016; Bulgarelli et al., 2012)	Good plant DNA exclusion	Primer biases
926F-1392R	V6-V8	454 and MiSeq (Tremblay et al., 2015)	Captures most sequences	Universal
1114F-1392R	V8-V9	454 and MiSeq (Lundberg et al., 2012)	Some plant DNA exclusion	Primer biases against archaea

Bell jar and lyophilizer/freeze drier
Beads of various sizes, glass, metal, ceramic, or garnet (Corning, Qiagen)
Mechanical disruptor, e.g., a grinding mill, Geno/Grinder (SPEX SamplePrep), or FastPrep24 (MP Biomedicals)
Vortex

Additional reagents and equipment for PCR (Kramer and Coen, 2000) and agarose gel electrophoresis (Voytas, 2000)

1. Aseptically place the plant tissue (e.g., leaf or roots from the step 17 of Basic Protocol 1) in a sterile plastic tube, and melt a hole in the top of the plastic tube using a flame-sterilized dissecting probe while it is still hot.

   1.5-ml tubes are usually appropriate for small root systems (e.g., A. thaliana), whereas 15-ml or 50-ml tubes may be necessary for larger root systems.

2. Snap freeze samples by immersing the tubes in liquid nitrogen, and place samples in a bell jar compatible with the lyophilizer. Completely dry tissue samples, usually overnight is sufficient.

3. Transfer dried tissue to a sterile tube containing enough beads of different sizes and shapes to fill the conical tip of the tube.

4. Freeze the sample by immersing the tube in liquid nitrogen, so the tissue is dry and brittle.

5. Homogenize the tissue samples for 5 min on a mechanical disruptor. Visually check samples to make sure they are pulverized. Repeat steps 4 and 5 until the tissue is a powder.

6. Extract nucleic acids from resulting pulverized tissue or from pellets (see Basic Protocol 1, step 14) using a chosen nucleic acid extraction kit (e.g., DNeasy PowerSoil kit) according to the manufacturer's instructions (see Critical Parameters).

   If using a bead tube supplied with a kit for the first step of the protocol, use the liquid from the manufacturer kit to transfer the pulverized tissue from step 5.

7. Quantify double-stranded DNA using a Qubit or PicoGreen fluorescent dye stain.

   Amplicon library protocols typically recommend DNA concentrations of 10 ng/μl.

8. Prior to preparing library amplicons, perform test PCRs using the selected primers and a PCR Master Mix to ensure that the extracted DNA is present.

   Use basic versions of the selected amplicon primers without spacers, barcodes, sequencing adaptors, etc. (see Critical Parameters).

9. Fractionate and visualize PCR products on an agarose gel.

10. Samples are ready for library amplicon preparation and sequencing.

    While the actual protocol use for amplicon library preparation can vary greatly, it is critical to include steps that: a) amplify the desired sequence for taxonomic identification; b) attach the sequencing adaptor to the targeted sequence; and c) allow samples to be multiplexed. A single round of PCR with long primers can accomplish these goals, but most protocols use at least two PCR reactions (Caporaso et al., 2012; Lundberg, Yourstone, Mieczkowski, Jones, & Dangl, 2013; Kircher, Sawyer, & Meyer, 2012). Although amplicon sequencing is often performed at sequencing facilities, the Commentary section discusses the Critical Parameters to consider.

REAGENTS AND SOLUTIONS

Use deionized, distilled water in all recipes and protocol steps.

Germinating plate medium

Dissolve 2.22 mg of MS salts with Gambourg's vitamins, 1% sucrose, and 9.5 g Phytoagar in 1 liter of distilled water. Autoclave medium on a liquid cycle for at least 20 min at 121°C. When the medium has cooled enough to touch the container, pour into sterile petri dishes and allow to solidify. Store plates at 4°C for up to 1 month.

Phosphate buffer

Dissolve 6.33 mg NaH_2PO_4 and 16.5 g Na_2HPO_4 in 1 liter of distilled water. Store autoclaved buffer at room temperature for up to 3 months.

Phosphate buffer with 0.02% silwet

Prepare 0.02% (v/v) silwet (Lehle Seeds) in phosphate buffer by transferring 5 µl of sterile silwet into 25 ml of phosphate buffer. Store at 4°C for up to 1 week.

COMMENTARY

Background Information

Although the procedures described here focus on preparation of samples for amplicon sequencing, if sufficient sample is produced, the nucleic acids generated from these protocols can be used for a number of downstream applications, including metagenomics, plant transcriptomics, and metatranscriptomics. Whereas amplicon sequencing only requires 150 ng of material per sample, metagenomics sequencing requires at least 500 ng, and transcriptome sequencing requires 1 to 3 µg of rRNA-depleted RNA. There are unique challenges associated with each type of plant material. Material for metagenomic or metatranscriptomic sequencing can be readily extracted from soil, but soil samples are extremely diverse, so producing useful data can be a computational challenge (Lebeis, 2015). In contrast, plant tissue is microbially less complex, but the microbial signal is low compared to the host signal.

Critical Parameters

Whether an investigator will prepare and perform their own amplicon sequencing or send the DNA samples to a sequencing facility, there are several critical parameters each investigator should understand.

Selection of nucleic acid extraction kit

It is imperative to choose a nucleic acid extraction kit or protocol that will be appropriate for all of the samples. For example, a study investigating both soil and root tissue requires a common extraction kit or protocol to neutralize humic acid, a soil component that interferes with DNA polymerases and other enzymes required for amplicon preparation. Since different nucleic acid kits and protocols yield

statistically distinct communities (Lundberg et al., 2012), it is necessary to use a humic acid neutralizing kit or protocol for all samples in the study, such as the DNeasy PowerSoil kit.

If the designed studies require both RNA and DNA isolation, consider using an extraction kit that isolates RNA and subsequently elutes DNA, such as the RNeasy PowerSoil Total RNA kit (Qiagen) followed by the RNeasy PowerSoil DNA Elution kit (Qiagen). Because extraction of environmental RNA is almost always more challenging than DNA, and requires much more material to acquire similar quantities, the user should always use the RNA material and extraction requirements to guide the necessary amount of sample material (Wang, Hayatsu, & Fujii, 2012). Difficulties in obtain sufficient RNA are caused by a number of factors including: RNA stability, the activity of environmental RNases, sample contamination with humic acid, and adsorption to chemicals present in the soil, such as Andosols (Wang et al., 2012). If the studies will investigate expression levels, the user must also consider that the amount of mRNA will be significantly lower after depletion of ribosomal RNA following treatment with a protocol or commercially available kit (e.g., ThermoFisher's RiboMinus or Illumina's Ribo-Zero; Wang et al., 2012).

Selection of amplicon primers and library preparation

Primer biases are well documented in amplicon sequencing (Ghyselinck, Pfeiffer, Heylen, Sessitsch, & De Vos, 2013; Tremblay et al., 2015; Walters et al., 2016), highlighting the importance of selecting primers known to amplify the microbes of interest, or if unknown, the primers most commonly used by the field. Standard census primers include: 515F and 806F for amplifying the V4 region of bacterial and archaeal 16S rRNA gene (Caporaso et al., 2012), 565F and 948R for amplifying the V4 region of the eukaryotic 18S rRNA gene (Stoeck et al., 2010), and ITS9F and ITS4R for amplifying the fungal ITS2 region between 5.8S and 26S rRNA (Menkis et al., 2012).

By far the most commonly used marker gene is the 16S rRNA gene, which—in its ~1500-bp sequence—contains constant regions required for ribosome function and phylogenetically informative variable regions (V1 to V9). Primers are named for their location in the 16S rRNA sequence and tend bind in the constant regions and allow sequencing across the variable regions. For example, Basic Protocol 2 uses low-throughput, long-read Sanger sequencing of nearly the entire 16S rRNA gene to identify isolated microbes. Ultra-high-throughput sequencing platforms (e.g., Illumina) cannot achieve the long read lengths of Sanger sequencing, so primer pairs that generate shorter amplicons have been investigated (Table 1). A recent study compared the relative strengths and weaknesses of three primer pairs (515F-806R, 926F-1392R, and 1114F-1392R) with a common set of samples. This study clearly showed that primer pair affects alpha diversity (Tremblay et al., 2015). Importantly, the 515F-806R primer pair yielded the most consistent metagenome shotgun-sequencing results, suggesting that it's the most accurate primer pair in the study (Tremblay et al., 2015). The original 515F primer sequence missed *Crenarcheota* and *Thaumarchaetoa*, and the original 806R sequence missed SAR11 (Caporaso et al., 2012), but both primers have recently been improved to capture these taxa (Walters et al., 2016). Such primer improvements have also occurred with the ITS1 region primers (ITS1F-ITS2) to minimize biases (Walters et al., 2016).

While the 515F and 1114F primers were designed to be bacterial specific, 926F, 806R, and 1392R were designed to be "universal" primers that bind to prokaryotic and eukaryotic 16S rRNA genes. For a universal primer pair, such as 926F-1392R, it will be difficult to survey the microbial community in plant tissues (Lundberg et al., 2013). Versions of the 799F-1192R primer pair offer the best exclusion of plant DNA, but they also contain biases against microbes (Bulgarelli et al., 2012, Table 1). If using the 515F-806R primer pair to examine microbial communities within plant tissues, consider adding a peptide nucleic acid that perfectly matches the plant 16S sequence during the PCR reaction to prevent DNA amplification. Such approaches can increase bacterial reads in root tissue from ~5% to ~60% (Lundberg et al., 2013). Because many ultra-high-throughput sequencing technologies rely on images of the flow cell, a lack in nucleotide diversity increases the difficulty of distinguishing between nucleotides on the flow cell (Krueger, Andrews, & Osborne, 2011). Nucleotide diversity, and thus accuracy in the resulting sequence data, can be introduced by frameshifting nucleotides in the primers in addition to or instead of adding PhiX bacteriophage DNA in the sequencing reaction (Caporaso et al., 2012; Krueger et al., 2011).

Plant microbiome census amplicon primers also include molecular barcodes, which allows many samples to be multiplexed into a single sequencing run. Even for diverse soil and rhizosphere samples, 192 samples can easily be multiplexed with sufficient read depth (Lebeis et al., 2015). Lastly, adaptor sequences vary between sequencing platforms, so the sequencing platform is a critical aspect of primer selection (Tremblay et al., 2015).

Aside from biological variance, experiment variation and sequencing errors are often introduced by amplicon preparation or sequencing. By tagging each DNA molecule with a unique molecular barcode in an initial round of PCR (Burke & Darling, 2016; Lundberg et al., 2013), it is possible to identify and eliminate any sequences that arise from PCR errors that occur early during amplicon preparation (Lundberg et al., 2013). These errors must be minimized to achieve non-inflated estimates of sample diversity (Kunin, Engelbrektson, Ochman, & Hugenholtz, 2010). One way to minimize the influence of technical factors is to reduce the number of sequencing runs for each experiment by multiplexing as many samples as possible. Because leaf samples and root endophyte samples tend to have significantly lower diversity than soil or rhizosphere samples (Lebeis, 2015), it is possible to multiplex more samples and still capture the diversity.

Selection of sequencing platform

Although sequencing technologies are rapidly changing, paired-end sequencing on the Illumina MiSeq platform is currently the most widespread selection for microbiome census experiments (Caporaso et al., 2012), because MiSeq combines highly accurate Illumina sequencing, longer read lengths than other Illumina platforms, and shorter run times. The high accuracy allows for more stringent quality control parameters on the reads than 454 pyrosequencing (Tremblay et al., 2015). Longer amplicons result in more accurate taxonomic information, and this platform even allows full 16S rRNA sequencing (Burke & Darling, 2016). While this approach decreases the number of samples that can be included on a sequencing run, they would certainly help with the accuracy of taxonomic assignments.

Analysis considerations

In many cases, amplicon census studies reveal abundances of taxa, rather than simple presence or absence in a community. Thus, robust and appropriate statistics and mathematical analyses will be necessary to systematically correlate phenotypes with the experimental variables. Freely available programs that process the resulting files into usable data for analysis include QIIME, UCHIME, UPARSE, and iTagger (Caporaso et al., 2010; Edgar, 2010; Edgar, 2013; Edgar, Haas, Clemente, Quince, & Knight, 2011; Tremblay et al., 2015). These programs generate tables with the abundance of clusters of sequences >97% identical into Operational Taxonomic Units (OTUs). These OTU tables list the abundance of each group of sequences for each sample, and allow the basic description of the community composition. The programs listed above also provide the tools to perform basic alpha and beta diversity analyses to begin to describe the overall community structures.

One goal of plant microbiome research is to use the composition of plant microbiomes as complex phenotype to be correlated with a number of variables, both abiotic (nutrient levels, temperature, etc.) and biotic (host species, age, genotype, etc.). Due to the inherent complexity of the phenotype, mathematical models are often employed to predict features of the plant microbiome composition that correlate with phenotypes of interest (Song, Cannon, Beliaev, & Konopka, 2014). For example, previous studies used variations of linear mixed models (Bulgarelli et al., 2012; Lundberg et al., 2012) or regression models (Lebeis et al., 2015) to identify particular OTUs or bacterial families that correlate with biological variables such as age or genotype (Wagner et al., 2016). Linear mixed models perform more accurately if the experimental design is balanced (e.g., equal number of biological replicates in each experimental group). Furthermore, more biological replicates allows more robust examination of potential differences. An initial study should aim for a minimum of 10 biological replicates in each test group. Importantly, the number of biological replicates may need to be adjusted, depending on the ability of plants to survive in natural soil. For example, immune-compromised plants have decreased survival in natural soils (Lebeis et al., 2015-2), necessitating increased replicates for these genotypes at the experiment onset. If these biological limitations cannot be overcome, and the resulting data are unbalanced (e.g., unequal number of biological replicates in each experimental group), regression models such as zero-inflated negative binomial may provide more

consistent predictions (Lebeis, 2015). While these are two major groups of mathematical modeling that can aid in plant microbiome data analysis, there are many other appropriate tools to use (Song et al., 2014). Beyond merely identifying the microbes present within a plant microbiome, predictive programs such as phylogenetic investigation of communities by reconstruction of unobserved states (PICRUSt) can be employed to assign functions to plant microbiome members (Langille et al., 2013). Although there is a danger to over-interpret such predictions, as with the results from linear mixed models and regression models, these powerful tools allow the formation of novel hypotheses regarding molecular mechanisms occurring within the plant microbiome that can then be experimentally tested.

Troubleshooting

A common problem encountered with plant microbiome studies is insufficient material to perform an intended comparison. This may improved at planting, growth, harvest, nucleic acid extraction, or data analysis. To ensure sufficient samples for statistical analyses, plant extra pots in case of mortality. Planting multiple seedlings within the same pot, however, could result in unintended plant-plant interactions. To increase the biomass of each replicate, set the growth chamber to a longer day length (e.g., 16 hr of light) to delay development and increase rosette and root system biomass (Lundberg et al., 2012). If harvesting root tissue from a small plant (e.g., *Arabidopsis thaliana*), fine lateral roots often break off of the main taproot, even when the soil in the pot is a crumbling consistency at the time of harvest. Because the harvesting pan and tweezers are sterilized between samples, additional root pieces can be harvested by raking the loose soil that has fallen into the pan. Finally, if a single sample lacks sufficient material, samples can be pooled before extraction, or reads can be pooled *in silico* after the OTU table has been generated. Though the optimal approach is to obtain sufficient material to keep as many biological replicates as possible, these represent several options for obtaining meaningful data from the experiments.

Anticipated Results

Plant microbiome composition is typically influenced by the type of sample (e.g., soil, rhizosphere, root endosphere, leaf epiphyte, or leaf endophyte; Bodenhausen et al., 2013; Bulgarelli et al., 2012; Lebeis, 2015; Lundberg et al., 2012; Shade et al., 2013; Wagner

et al., 2016). Soils are one of the most microbially diverse environments on earth (Tringe et al., 2005), and the rhizosphere appears to be equally taxonomically diverse in many plant systems (Lebeis, 2015). Root endosphere communities tend to be less diverse subsets of the rhizosphere and soil communities (Pieterse et al., 2016) that begin to form within one week of inoculation and continue to assemble during plant growth (Bulgarelli et al., 2012; Edwards et al., 2015; Lundberg et al., 2012). Individual-to-individual variation tends to be quite low, which allows for significant grouping when necessary with relatively few biological replicates (Lebeis, 2015). Although phyllosphere microbiomes also tend to have lower community diversity than soil (Bodenhausen et al., 2013; Wagner et al., 2016), they are prone to rapid and dramatic changes in composition (Bodenhausen et al., 2013; Shade et al., 2013), with inoculation events, limited nutrients, and physical disturbances shaping alterations in community structure (Vorholt, 2012). For these reasons, future phyllosphere studies might require more time points or biological replicates. Despite these limitations, consistent influence over plant microbiome composition is attributed to developmental stage and inoculum, with plant genotype subtly affecting external communities and significantly affecting internal communities (Bodenhausen, Bortfeld-Miller, Ackermann, & Vorhold, 2014; Lebeis et al., 2015; Wagner et al., 2016).

The protocols provided here aim to allow investigators to obtain robustly differentiated sample fractions (Basic Protocol 1), representative microbial isolates (Basic Protocol 2), and sufficient nucleic acid to characterize plant microbiomes (Basic Protocol 3). Even for a small plant, such as *A. thaliana*, grown in a 2.5 × 2.5-in. pot for 6 to 8 weeks, expect to obtain 0.25 to 0.5 g rhizosphere and 0.1 to 0.25 g root tissue from each plant. Although it is similarly easy to obtain grams of leaf tissue, the mass of leaf epiphyte will be significantly lower, and you may need to pool samples prior to extraction—the MoBio PowerSoil kit recommends 0.25 g of material for each sample. If sufficient material is obtained, expect to obtain 50 µl of 10 to 50 ng/µl DNA. If less DNA is obtained, please consult Troubleshooting for solutions.

Although using dilute medium in Basic Protocol 2 will allow a greater diversity of microbes to be isolated, it will also result in slow growth. While some colonies will appear within a couple of days of plating,

investigators should continue to incubate plates for several weeks to capture slower growing microbes. Users should pick colonies as they appear, and streak them to rich medium, such as full strength LB. Expect to isolate Actinobacteria, Firmicutes (particularly *Bacillus* and *Paenibacillus* strains), and Protetobacteria, which are abundant in these communities and relatively easy to culture. Together these approaches aim to provide the investigator with the samples and microbial exemplars to test the mechanistic hypotheses that will be generated from a first glimpse into the plant microbiome system.

Time Considerations

Although the samples can be stored at −80°C between protocols, harvesting samples—especially roots—takes a significant amount of time. Thus, depending on the plant size and soil texture, one sample could take a novice up to 30 minutes to prepare. Please consider having teams of lab members working together, or harvest over several days, depending on the size of the experiment. If harvesting over several days, also consider the time resolution in the experimental design.

Acknowledgements

Conflict of Interest

The author has no conflicts of interest.

Literature Cited

Agler, M. T., Ruhe, J., Kroll, S., Morhenn, C., Kim, S. T., Weigel, D., & Kemen, E. M. (2016). Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biology*, *14*(1), e1002352. doi: 10.1371/journal.pbio.1002352.

Bai, Y., Muller, D. B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., . . . Schulze-Lefert, P. (2015). Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature*, *528*, 364–369. doi: 10.1038/nature16192.

Bodenhausen, N., Bortfeld-Miller, M., Ackermann, M., & Vorhold, J. A. (2014). A synthetic community approach reveals plant genotypes affecting the phyllosphere microbiota. *PLoS Genetics*, *10*, e1004283. doi: 10.1371/journal.pgen.1004283.

Bodenhausen, N., Horton, M. W., & Bergelson, J. (2013). Bacterial communities associated with the leaves and the roots of *Arabidopsis thaliana*. *PLoS ONE*, *8*, e56329. doi: 10.1371/journal.pone.0056329.

Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F.,

. . . Schulze-Lefert, P. (2012). Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature*, *488*, 91–95. doi: 10.1038/nature11336.

Burke, C. M., & Darling, A. E. (2016). A method for high precision sequencing of near full-length 16S rRNA genes on an Illumina MiSeq. *PeerJ*, *4*, e2492. doi: 10.7717/peerj.2492.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., . . . Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, *6*, 1621–1624. doi: 10.1038/ismej.2012.8.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., . . . Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*, 335–336. doi: 10.1038/nmeth.f.303.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*, 2460–2461. doi: 10.1093/bioinformatics/btq461.

Edgar, R. C. (2013). UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, *10*, 996–998. doi: 10.1038/nmeth.2604.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, *27*, 2194–2220. doi: 10.1093/bioinformatics/btr381.

Edwards, J., Johnson, C., Santos-Medellin, C., Lurie, E., Kumar Podishetty, N., Bhatnagar, S., . . . Sundaresan, V. (2015). Structure, variation, and assembly of the root-associated microbiomes of rice. *Proceedings of the National Academy of Sciences U.S.A.*, *112*, E911–E920. doi: 10.1073/pnas.1414592112.

Ghyselinck, J., Pfeiffer, S., Heylen, K., Sessitsch, A., & De Vos, P. (2013). The effect of primer choice and short read sequences on the outcome of 16S rRNA gene based diversity studies. *PLoS ONE*, *8*, e71360. doi: 10.1371/journal.pone.0071360.

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*, e3. doi: 10.1093/nar/gkr771.

Kramer, M. F., & Coen, D. M. (2000). Enzymatic amplification of DNA by PCR: Standard procedures and optimization. *Current Protocols in Molecular Biology*, *56*, 15.1.1–15.1.14. doi: 10.1002/0471142727.mb1501s56.

Krueger, F., Andrews, S. R., & Osborne, C. S. (2011). Large scale loss of data in low-diversity Illumina sequencing libraries can be recovered by deferred cluster calling. *PLoS ONE*, *6*, e16607. doi: 10.1371/journal.pone.0016607.

Kunin, V., Engelbrektson, A., Ochman, H., & Hugenholtz, P. (2010). Wrinkles in the rare

Plant Microbiome Identification and Characterization

145

biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology*, *12*, 118–123. doi: 10.1111/j.1462-2920.2009.02051.x.

Langille, M. G. I., Zanevald, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., . . . Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, *31*, 814–821. doi: 10.1038/nbt.2676.

Lebeis, S. L. (2015). Greater than the sum of their parts: Characterizing plant microbiomes at the community-level. *Current Opinion in Plant Biology*, *24*, 82–86. doi: 10.1016/j.pbi.2015.02.004.

Lebeis, S. L., Paredes, S. H., Lundberg, D. S., Breakfield, N., Gehring, J., McDonald, M., . . . Dangl, J. L. (2015). Salicyclic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science*, *349*, 860–864. doi: 10.1126/science.aaa8764.

Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., . . . Dangl, J. L. (2012). Defining the core *Arabidopsis thaliana* root microbiome. *Nature*, *488*, 86–90. doi: 10.1038/nature11337.

Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D., & Dangl, J. L. (2013). Practical innovations for high-throughput amplicon sequencing. *Nature Methods*, *10*, 999–1002. doi: 10.1038/nmeth.2634.

Menkis, A., Burokiene, D., Gaitnieks, T., Uotila, A., Johannesson, H., Rosling, A., . . . Vasaitis, R. (2012). Occurrence and impact of the root-rot biocontrol agent *Phlebiopsis gigantea* on soil fungal communities in *Picea abies* forests of northern Europe. *FEMS Microbiology Ecology*, *81*, 438–445. doi: 10.1111/j.1574-6941.2012.01366.

Pieterse, C. M. J., de Jonge, R., & Berendsen, R. L. (2016). The soil-borne supremacy. *Trends in Plant Science*, *21*, 171–173. doi: 10.1016/j.tplants.2016.01.018.

Shade, A., McManus, P. S., & Handelsman, J. (2013). Unexpected diversity during community succession in the apple flower microbiome. *mBio*, *4*, e00602–12. doi: 10.1128/mBio.00602-12.

Stoeck, T., Bass, D., Nebel, M., Jones, M. D., Breiner, H. W., & Richards, T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*, *19*(s1), 21–31. doi: 10.1111/j.1365-294X.2009.04480.x.

Song, H. S., Cannon, W. R., Beliaev, A. S., & Konopka, A. (2014). Mathematical modeling of microbial community dynamics: A methodical review. *Processes*, *2*(4), 711–752. doi: 10.3390/pr2040711.

Tremblay, J., Singh, K., Fern, A., Kirton, E. S., He, S., Woyke, T., . . . Tringe, S. G. (2015). Primer and platform effects on 16S rRNA tag sequencing. *Frontiers in Microbiology*, *6*, 771. doi: 10.3389/fmicb.2015.00771.

Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., . . . Rubin, E. M. (2005). Comparative metagenomics of microbial communities. *Science*, *308*, 554–557. doi: 10.1126/science.1107851.

Vorholt, J. A. (2012). Microbial life in the phyllosphere. *Nature Reviews Microbiology*, *10*, 828–840. doi: 10.1038/nrmmicro2910.

Voytas, D. (2000). Agarose gel electrophoresis. *Current Protocols in Molecular Biology*, *51*, 2.5A.1–2.5A.9. doi: 10.1002/0471142727.mb0205as51.

Wagner, M. R., Lundberg, D. S., del Rio, T. G., Tringe, S. G., Dangl, J. L., & Mitchell-Olds, T. (2016). Host genotype and age shape leaf and root microbiomes of a wild perennial plant. *Nature Communications*, *7*, 12151. doi: 10.1038/ncomms12151.

Walters, W., Hyde, E. R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., . . . Knight, R. (2016). Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems*, *1*, 915. doi: 10.1128/mSystems.00009-15.

Wang, Y., Hayatsu, M., & Fujii, T. (2012). Extraction of bacterial RNA from soil: Challenges and solutions. *Microbes and Environment*, *27*(2), 111–121. doi: 10.1264/jsme2.ME11304.