Extraction and 16S rRNA Sequence Analysis of Microbiomes Associated with Rice Roots

Joseph Edwards[1, #, $], Christian Santos-Medellín[1, #] and Venkatesan Sundaresan[1, 2, *]

[1]Department of Plant Biology, University of California, Davis, USA; [2]Department of Plant Sciences, University of California, Davis, USA; [$]Current address: Department of Integrative Biology, University of Texas, Austin, USA

*For correspondence: sundar@ucdavis.edu

[#]Contributed equally to this work

[Abstract] Plant roots associate with a wide diversity of bacteria and archaea across the root-soil spectrum. The rhizosphere microbiota, the communities of microbes in the soil adjacent to the root, can contain up to 10 billion bacterial cells per gram of soil (Raynaud and Nunan, 2014) and can play important roles for the fitness of the host plant. Subsets of the rhizospheric microbiota can colonize the root surface (rhizoplane) and the root interior (endosphere), forming an intimate relationship with the host plant. Compositional analysis of these communities is important to develop tools in order to manipulate root-associated microbiota for increased crop productivity. Due to the reduced cost and increasing throughput of next-generation sequencing, major advances in deciphering these communities have recently been achieved, mainly through the use of amplicon sequencing of the 16S rRNA gene. Here we first present a protocol for dissecting the microbiota from various root compartments, developed using rice as a model. We next present a method for amplifying fragments of the 16S rRNA gene using a dual index approach. Finally, we present a simple workflow for analyzing the resulting sequencing data to make ecological inferences.

Keywords: Root microbiome, Amplicon sequencing, Rhizosphere, Rhizoplane, Endosphere

[Background] Various plant root niches host different microbial communities (microbiota) originating from the soil (Bulgarelli *et al.*, 2012; Lundberg *et al.*, 2012; Edwards *et al.*, 2015; Zarraonaindia *et al.*, 2015; Wagner *et al.*, 2016). Distinct microbiota acquired by each root niche likely have varying metabolic potential and may therefore impact the health of the host plant in different ways (Finkel *et al.*, 2017). Bacterial and archaeal community composition in root-associated microbiota can be inferred through the use of 16S rRNA gene sequencing (Caporaso *et al.*, 2012). The relatively low cost of sequencing now allows for comparative studies across plant species using datasets gathered by different research groups; however, small aberrations in specimen collection, sequencing, and analysis protocols may lead to large differences in the inferred microbial communities (Duvallet *et al.*, 2017). We present this protocol detailing how to collect and analyze root microbiota from rice in an attempt to promote reproducibility across the plant microbiome field.

Materials and Reagents

1. Falcon 50 ml conical centrifuge tubes (Corning, Falcon®, catalog number: 352070)
2. 1.5 ml microfuge tubes (E&K Scientific Products, catalog number: 280150)
3. 1.5 ml non-stick microfuge tubes (Thermo Fisher Scientific, Ambion™, catalog number: AM12450)
4. 0.2 ml PCR tubes (GeneMate, catalog number: 3235-00-210IS)
5. Filtered pipette tips (10, 200, 1,000 µl) (VWR, catalog numbers: 89168-750, 89140-936, 89168-754)
   Manufacturer: Biotix, catalog number: BT10XLS3, BT200, BT1250.
6. Gloves (Medline Industries, catalog numbers: large, MDS192086; medium, MDS192085; small, MDS192084)
7. Qubit 0.5 ml assay tubes (Thermo Fisher Scientific, catalog number: Q32856)
8. Single edge razor blade (Personna, catalog number: 94-115-71)
9. Nuclease-free water (Thermo Fisher Scientific, Ambion™, catalog number: AM9939)
10. DNeasy PowerSoil kit (QIAGEN, catalog number: 12888-100)
11. Primer 515F (GTGCCAGCMGCCGCGGTAA)
12. Primer 806R (GGACTACHVGGGTWTCTAAT)
13. HotStar High Fidelity DNA polymerase kit (QIAGEN, catalog number: 202602)
14. Agencourt Ampure XP beads (Beckman Coulter, catalog number: A63880)
15. Qubit dsDNA HS assay kit (Thermo Fisher Scientific, catalog number: Q32851)
16. Ethanol 200 proof (Sigma-Aldrich, catalog number: E7023-500ML)
17. Agarose (Biotech Sources, catalog number: G01PD-500)
18. DNA gel loading dye
19. NucleoSpin gel and PCR clean-up kit (MACHEREY-NAGEL, catalog number: 740609.250)
20. NaCl (Fisher Scientific, catalog number: S271-1)
21. KCl (Fisher Scientific, catalog number: P217-500)
22. Na_2HPO_4 (Fisher Scientific, catalog number: S374-500)
23. KH_2PO_4 (Fisher Scientific, catalog number: P285-500)
24. Autoclaved phosphate buffered saline (PBS) solution (~100 ml/plant) (see Recipes)

Equipment

1. 96 well magnetic plate (Alpaqua Engineering, catalog number: A001219R)
2. 1.5 ml tube magnetic rack (Thermo Fisher Scientific, catalog number: MR01)
3. Pipettes (2.5, 10, 200, 1,000 µl) (Thermo Fisher Scientific, Finnpipette™, catalog numbers: 4641010N, 4641030N, 4641080N, 4641100N)
4. Ultrasonic cleaning bath, 40 kHz (Branson, model: Branson 1800, catalog number: CPX-952-116R)

5. Dissection tools (scissors and forceps) (scissors: Bioseal, catalog number: KI011/50; forceps: Integra LifeSciences, Miltex, catalog number: 6-184)

6. -80 °C freezer

7. Microcentrifuge (Eppendorf, model: 5417C)

8. Mini-Beadbeater-96 high-throughput cell disrupter (Bio Spec Products, catalog number: 1001)

9. Electrophoresis gel unit (Bio-Rad Laboratories, catalog number: 1704468)

10. Qubit fluorometer (Thermo Fisher Scientific, catalog number: Q33226)

11. PCR thermal cycler (Bio-Rad Laboratories, model: T100™ Thermal Cycler, catalog number: 1861096)

Software

1. Python2 version 2.7.12
2. R version 3.4.3

Procedure

The procedure outlined below is generally applicable across a wide array of conditions and has been successfully used to survey bacterial and archaeal community composition in the greenhouse and the field (Edwards *et al.*, 2015), as well as across environmental perturbations (Santos-Medellín *et al.*, 2017) and plant developmental stages (Edwards *et al.*, 2018).

Notes:

1. *Contamination control*

    Because introduced contaminants can obscure the compositional data generated from this protocol, it is important to observe the following measures to minimize contamination:

    a. *Always wear gloves.*

    b. *Before starting each main step, wipe your hands and work surface with 70% ethanol.*

    c. *Avoid leaving bottles and tubes open to the environment.*

    d. *Use filtered pipette tips for Procedure A (compartment separation), B (DNA extraction), and C (PCR amplification). After PCR, you can switch to sterile non-filtered tips.*

2. *For sample storage*

    Preprocessed roots should be stored for no longer than 24 h at 4 °C before compartment separation. Additionally, individual compartments should be separated before storing them for longer periods of time at -80 °C. Freezing and thawing preprocessed roots will cause microbial cells to lyse and the DNA to possibly diffuse across compartments, therefore reducing resolution.

    a. *For the rhizosphere compartment, pipette 500 µl of the soil suspension generated in Step A4 to a 1.5 ml microfuge tube, spin down (10,000 x g for 1 min), remove the supernatant, and store at -80 °C. When ready to perform DNA extractions, thaw samples at room temperature (~23 °C) and resuspend the rhizosphere in 500 µl of PBS.*

b. *For the rhizoplane compartment, store the 500 µl of concentrated microbial suspension generated in Step A6 at -80 °C. When ready to perform DNA extractions, thaw samples at room temperature (~23 °C).*

c. *For the endosphere compartment, use fire-sterilized forceps to transfer 0.25 g of the thrice-sonicated roots from Step A7 to a 1.5 ml microfuge tube and store at -80 °C. When ready to perform DNA extractions, thaw samples at room temperature (~23 °C).*

A. Compartment separation of root-associated microbiota

The following protocol uses a combination of washing and sonicating steps to separate the rhizosphere, rhizoplane, and endosphere fractions of the root-associated microbiota. This approach has been successfully employed to harvest compositionally distinct communities that harbor microorganisms enriched in each of these spatial compartments (Edwards *et al.*, 2015). The protocol for separation of the rhizoplane is based on the method for endospheric bacteria isolation developed by Lundberg *et al.* (2012). The method utilizes a bath sonicator to remove the microbiota in the rhizoplane, and avoids hypochlorite treatment for the reasons detailed in Lundberg *et al*. (2012). Because the DNA yields from rhizoplane samples are low, the sequences can exhibit higher variability following PCR amplification, and may require additional replicates to draw statistically significant conclusions. It should be noted that the compartment dissection protocol cannot ensure complete purity of samples free of contamination from adjacent compartments, especially where they overlap spatially, but despite these limitations the protocol has been proven to be efficient and reproducible for studies of overall compositional profiles (Edwards *et al.*, 2015; Santos-Medellin *et al.*, 2017; Edwards *et al.*, 2018). All steps are visually detailed in Video 1.
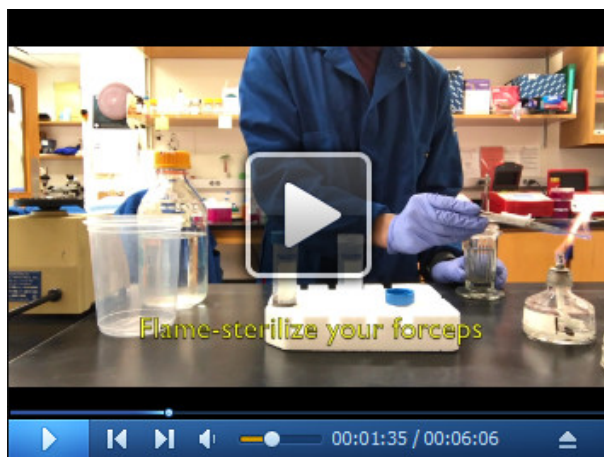
Video 1. Collection of rice roots and separation of compartments

1. Using gloves, harvest the rice plant by firmly holding the shoot and slowly pulling the root system out of the ground (Figure 1A). In the case of seedlings, carefully scoop the roots to avoid ripping the tissue.

2. Vigorously shake the roots to remove loose soil, leaving only the soil layer firmly attached to the root. This layer constitutes the rhizosphere compartment (Figure 1B).

3. Using flame-sterilized scissors, cut ~5 cm of root immediately below the root-shoot junction (red box Figures 1B and 1C) and place the tissue in a sterile 50 ml Falcon tube with 15 ml of autoclaved PBS solution. For potted plants, avoid collecting roots immediately adjacent to the inner walls.

4. Vortex the roots for 15 sec to mix the rhizosphere fraction in the PBS solution (Figure S1). Save the resulting soil suspension for DNA extraction (Step B1).

5. Using flame-sterilized forceps, transfer the roots to a new 50 ml Falcon tube. Wash the roots thoroughly by adding 20 ml of fresh PBS, vortexing for 15 sec at maximum speed, and discarding the PBS. Repeat these steps for a total of three washes (Figure 1D). If any soil remains in the bottom of the tube, perform additional washing steps until no soil is visible.
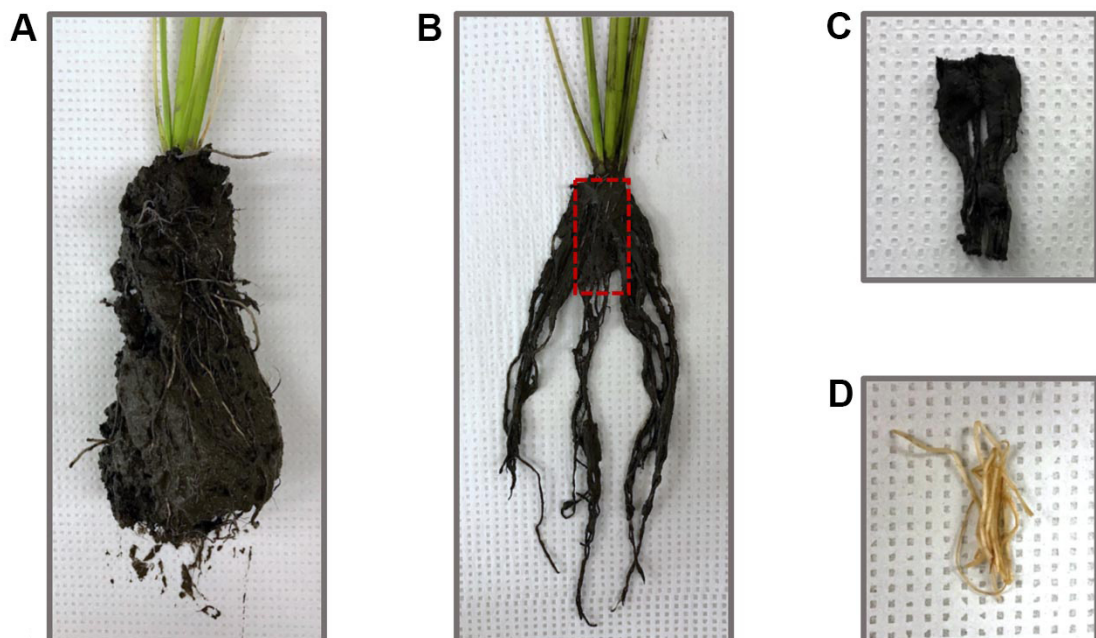

Figure 1. Root harvesting and processing for microbiome studies. A. Rice roots pulled out of the soil. B. Rice roots after being vigorously shaken to remove loose soil. The red box indicates the ~5 cm of root cut with flame-sterilized scissors. C. Rice root section collected into 50 ml Falcon tubes for compartment separation. The soil layer firmly attached to the roots constitutes the rhizosphere. D. Rice root section after being thoroughly washed with sterile PBS solution.

6. Separate the rhizoplane compartment by sonicating the roots for 30 sec at 50-60 Hz and transferring the 10 ml of PBS with the sonicated microbes to a new tube (Figure S1). Add 1.5 ml of the microbial suspension to a 1.5 ml microfuge tube, spin down at 10,000 x g for 1 min and discard 1 ml of supernatant. Add 1 ml of microbial suspension, spin down (10,000 x g for 1 min), and discard 1 ml. Repeat these steps once more for a total of three centrifugations.

Resuspend the pellet by vortexing (15 sec) and save the concentrated microbial suspension for DNA extraction.

Note: Depending on how much root material was sampled, a pellet may or may not be visible.

7. Add enough fresh PBS to fully cover the roots and sonicate for 30 sec at 50-60 Hz. Discard the PBS and repeat this step once more. The thrice-sonicated roots constitute the endosphere compartment (Figure S1).

B. DNA Extraction

Use the DNeasy PowerSoil kit to isolate the genomic DNA from the root-associated communities. The input for each of the compartments is as follows:

1. For the rhizosphere compartment, add 500 µl of the soil suspension generated in Step A4 to a PowerBead tube.

   Note: Large particles which inhibit the uptake of the soil suspension into the pipette tip should be avoided. If the pipette tip is too clogged to continue, retrieve a new pipette tip. Repeat the process until successful.

2. For the rhizoplane compartment, transfer the 500 µl of concentrated microbial suspension generated in Step A6 to a PowerBead tube.

3. For the endosphere compartment, use fire-sterilized forceps to transfer 0.25 g of the thrice-sonicated roots from Step A7 to a PowerBead tube. Pre-homogenize the endosphere by bead-beating the roots in the PowerBead tube with the included garnet particles for 1 min.

After adding the samples to the PowerBead tubes, follow the PowerSoil kit protocol with these adjustments:

4. After adding Solution C1, the PowerBead tubes can be homogenized using a beadbeater for 2 min instead of vortexing them for 10 min.

5. Elute the final product in 30 µl of Solution C6 instead of 100 µl.

C. 16S rRNA amplification

For library construction, this protocol uses primer 515F (GTGCCAGCMGCCGCGGTAA) and 806R (GGACTACHVGGGTWTCTAAT) to amplify the V4 region of the 16S rRNA gene. Primer design for Illumina sequencing follows the one described in Caporaso *et al.* (2012), except both forward and reverse primers are barcoded (Figure 2). By using a unique combination of barcodes for each sample, this dual-indexing strategy allows us to multiplex a large number of libraries with a limited amount of primers. Full sequences for the forward primers can be found on this GitHub page and the reverse primers can be found on this GitHub page. We recommend working with sets of 24 samples, in which all reactions share the same barcoded 515F primer but have a unique 806R barcode. Additionally, it is important to run a negative control for each individual reaction to detect any potential contamination.
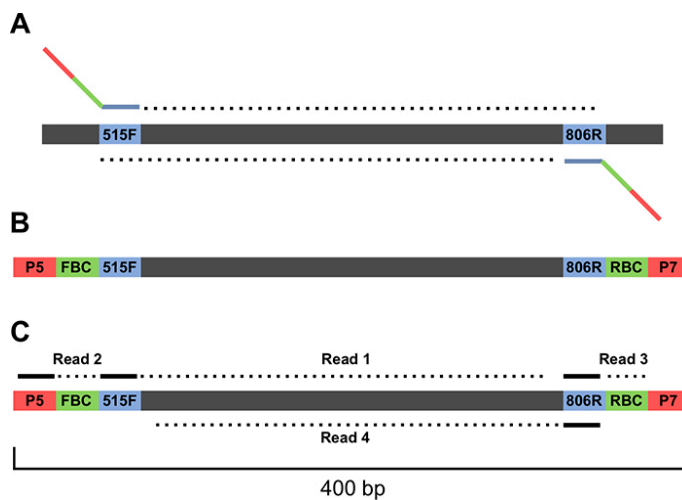
Figure 2. Schematic of 16S V4 region amplicon. A. Genomic regions before amplification. The primer binding sites are blue and the number corresponds to the position within the 16S rRNA gene where the primers bind. B. Amplicon after PCR amplification. FBC stands for forward barcode and RBC stands for reverse barcode. C. The sequencing strategy for the amplicons. Note that three custom primers are used in the sequencing: a primer for the forward read starting at position 515, a primer for the reverse sequencing read starting at position 806, and a primer for the reverse barcode. A list of forward primer sequences and reverse primer sequences can be found on GitHub.

1. For a set of 24 reactions and 24 negative controls, prepare a master mix using the following recipe:

	1 reaction	50 reactions
RNase-free water	6.25 µl	312.5 µl
HotStar HiFidelity PCR Buffer 5x (contains dNTPs and MgCl$_2$)	2.5 µl	125 µl
10 µM barcoded 515F Primer	1.25 µl	62.5 µl
HotStar HiFidelity DNA Polymerase	0.25 µl	12.5 µl

2. Aliquot 20.5 µl of the master mix into each PCR tube.
3. Add 2.5 µl of the corresponding 10 µM barcoded 806R primer into each tube and mix well by pipetting.
4. Aliquot 11.5 µl of each reaction into new PCR tubes to run as negative controls.
5. Add 1 µl of the corresponding template to the remaining 11.5 µl.
6. Cap the tubes and spin down.
7. Run the following touchdown PCR program:

Initial denaturation:

95 °C, 5 min

7 cycles, decreasing the annealing temperature 2 °C each cycle:

95 °C, 45 sec

65 °C, 1 min (-2 °C/cycle)

72 °C, 1:30

30 cycles:

95 °C, 45 sec

50 °C, 30 sec

72 °C, 1:30

Final extension:

72 °C, 10 min

Pause:

4 °C, ∞

Note: If the user is experiencing a high sample to sample variation, there may be noise generated during the PCR step. The Earth Microbiome Project has established a protocol where 3 separate PCRs are performed between samples and subsequently pooled in order to minimize variation.

D. Gel

1. Add 1 µl of PCR product to 5 µl of gel loading dye (1x).
2. Run the samples on a 1% agarose gel at 120 V for 20 min.
3. Verify proper amplification (expected band size is ~400 bp long) and absence of contamination in the negative controls.

E. PCR Cleanup

1. Remove beads from 4 °C and allow them to reach room temperature (~23 °C).
2. Prepare a fresh batch of 70% ethanol solution (500 µl/reaction).
3. Aliquot 9 µl of the PCR product into a new 0.2 ml tube.
4. Add 5.4 µl (0.6 volume) of Ampure XP beads, mix by pipetting, and let incubate at room temperature (~23 °C) for 5 min. We have found that 0.6x volume of AMPure beads to PCR product is the correct ratio to remove primer dimers and unused primers while leaving PCR product intact. Users of this protocol may need to experiment to ensure that this ratio also works for their experiments.
5. Transfer the tubes to a magnet plate and let stand for 2 min.
6. Carefully remove the cleared solution without disturbing the beads.
7. Keeping the tubes on the magnet plate, add 200 µl of 70% ethanol, incubate for 30 sec, and remove with a pipette. Repeat this step once more for a total of two ethanol washes. For the final wash, remove all ethanol from the bottom.
8. With the tubes still on the magnet plate, air-dry the beads for 2 min. Take care not to over-dry the beads as this will prevent the beads from being resuspended in Step E9.

9. Take the plates off the magnet plate, add 50 µl of nuclease-free water, and resuspend the beads by pipetting up and down.

10. Place the tube back to the magnet plate and incubate for 1 min.

11. Transfer 35 µl of the eluant to a new tube.

F. Quantification of amplicon libraries

Use the Qubit dsDNA HS assay to measure the concentrations of the purified libraries:

1. Set up two Assay Tubes for the standards and one for each sample to be quantified. Make sure not to label the side of the tube as this could interfere with the sample read.

2. For each tube, prepare 200 µl of Qubit Working Solution by mixing 199 µl of Qubit buffer and 1 µl of Qubit Working Solution. It is important not to mix the working solution in a glass container.

3. For the standards, aliquot 190 µl of working solution to 0.5 ml Assay Tubes, add 10 µl of the corresponding standard, and mix by vortexing.

4. For the samples, aliquot 195 µl of working solution to 0.5 ml Assay Tubes, add 5 µl of the corresponding sample, and mix by vortexing.

5. Incubate tubes for 2 min at room temperature.

6. Select the dsDNA High Sensitivity Assay on the Qubit Fluorometer, read the standards, and run each sample. Typical DNA concentrations range from 0.5 to 2.0 ng/µl.

G. Pooling of amplicon libraries

1. Multiplex ~150 libraries per MiSeq sequencing run. If multiple sequencing runs are needed, randomize the libraries across sequencing runs to avoid batch effects.

2. Based on the range of concentrations obtained, determine a target amount of DNA to be pooled. Aim for at least 5 ng of DNA per sample.

3. For each library, calculate the volume need by dividing the target amount of DNA by the sample concentration.

4. Pool the libraries in a non-stick RNase-free 1.5 ml microfuge tube.

5. Avoid pipetting volumes lower than 1 µl. If some samples are too concentrated, predilute them before pooling.

H. Library concentration

1. Remove AmpureXP beads from 4 °C and allow them to reach room temperature.

2. Prepare a fresh batch of 70% ethanol solution.

3. Add 1.8 volumes of Ampure XP beads to the pooled libraries, mix by pipetting, and let incubate at room temperature for 5 min. 1.8x volume ensures that there is no loss of product due to saturation of beads.

4. Transfer the tube to a magnet rack and let stand for 2 min.

5. Carefully remove the cleared solution without disturbing the beads.

6. Keeping the tubes on the magnet, add 1.5 ml of 70% ethanol, incubate for 30 sec, and remove with a pipette. Repeat this step once more for a total of two ethanol washes. For the final wash, remove all ethanol from the bottom.

7. With the tube still on the magnet, air-dry the beads for 2 min.

8. Take the tube off the magnet, add 50 µl of nuclease-free water, and resuspend the beads by pipetting.

9. Place the tube back to the magnet and incubate for 1 min.

10. Transfer the cleared eluant to a new tube.

I. Gel cleanup

1. Run the concentrated pool of libraries on a 1.8% agarose gel at 120 V for 40 min.

2. Using a new blade, excise the ~400 bp band.

3. Purify the libraries using the NucleoSpin Gel and PCR Clean-up kit.

J. Sequencing

1. Submit the pooled libraries for 2 x 250 MiSeq sequencing. Use the following custom sequencing primers:

    Read 1 Primer: TATGGTAATTGTGTGCCAGCMGCCGCGGTAA

    Read 2 Primer: AGTCAGTCAGCCGGACTACHVGGGTWTCTAAT

    3' Index Primer: ATTAGAWACCCBDGTAGTCCGGCTGACTGACT

2. Include a PhiX control for low diversity samples.

K. Sequence analysis

A more detailed version of the sequence analysis pipeline including code for running the analysis can be found on GitHub.

1. Compile metadata in a spreadsheet

    a. Assign each sample a unique identifier. This identifier should be unique not just across a single experiment, but across all previous experiments as well. The identifier should only contain alphanumeric and period (".") characters.

    b. Create column(s) for barcodes. Each sample should have a unique barcode for each run.

    c. Create additional columns for the experimental variables associated with each sample. Some typical examples of experimental variables are root compartment, plant developmental stage, plant genotype, plot location, date, year, and collector.

        Note: A detailed protocol for designing a metadata table can be found at the Earth Microbiome Projects website .

2. Download sequencing files from sequencing facility

    Downloading of individual files can be accomplished either through an FTP client or by using command line tools (such as Wget).

3. Demultiplex sequences

There should be 4 fastq files provided by the sequencing facility: two read files (R1 and R2) and two index files (I1 and I2).

4. Construct full-length contiguous sequences

Full-length sequences can be assembled using PANDAseq (Masella *et al.*, 2012). Note that full-length contigs are not necessary for clustering OTUs using DADA2 (see below).

5. Cluster Sequences and build OTU table

a. If clustering with QIIME (Caporaso *et al.*, 2010), several options and algorithms are available. The user may want to conduct reference based clustering against a database of 16S rRNA genes (closed reference clustering) or a user may want to perform *de novo* OTU clustering. This method is known as closed reference clustering. Alternatively, a user may prefer to perform a hybrid between these two methods where sequences are first referenced against a database. Reads that do not have a match within the database are then clustered *de novo*. This method is known as open reference clustering. The user can define the similarity threshold for one read to be considered a match with an entry in the database. Historically, > 97% sequence identity has been used as the standard for clustering sequences into operational taxonomic units (OTUs). One relatively new and extremely fast method for closed reference clustering of sequences into OTUs is through using the NINJA-OPS pipeline (Al-Ghalith *et al.*, 2016). NINJA-OPS leverages the speed and memory efficiency of Bowtie (Langmead and Salzberg, 2012), mapping reads back to a synthetic genome of concatenated 16S genes. This method can be performed on a laptop computer.

b. Alternatively, users may prefer to bin sequences based upon exact matches using DADA2 (Callahan *et al.*, 2016a). If using this method, there is no need to construct full-length contiguous sequences before clustering.

6. Assign taxonomies to OTUs

Multiple algorithms exist to assign taxonomies to the OTU sequences. If using closed-reference OTU clustering, there is no need to perform this step because the database sequences have already been classified (DeSantis *et al.*, 2006). QIIME defaults to using the UCLUST (Edgar, 2010) algorithm for taxonomic assignment, while DADA2 uses the RDP naive Bayes method (Wang *et al.*, 2007) for assigning taxonomies to sequences.

Data analysis

A detailed description of how to perform the data analysis including R code can be found on GitHub. In this tutorial we use data from Santos-Medellín *et al.* (2017) to illustrate analytical techniques.

1. Remove plastidial and mitochondrial sequences from the dataset

a. Mitochondria and plastids are a result of an ancient endosymbiosis event. The mitochondria and plastids have retained their own ribosomal machinery, therefore a fraction of the resulting sequences will belong to these organelles. These reads are not part of the root microbiota and should be removed from the dataset before further analysis. This is not to

say that organellar reads are not useful–these reads can be used for quantification purposes (Edwards *et al.*, 2015), but they should not be considered part of the microbiota.

b. Mitochondrial and plastidial OTUs can be identified via their associated taxonomies by searching for 'mitochondria' in the Family column and 'Chloroplast' under the Class column.

2. Normalize the sequencing depth for each library

Although the libraries were pooled in an equimolar concentration, sequencing depth can vary a few orders of magnitude between each library. It is therefore necessary to normalize data to ensure that each sample is equally represented in the analysis.

A few methods exist for normalization. Rarefaction is the process of randomly sampling from the pool of OTUs until a desired depth is achieved. It is of note that this method removes much of the data the user has acquired. For example, if the user has two libraries *a* and *b* and the depth, *d*, of each library is d_a = 100,000 and d_b = 5,000, the user may choose to rarefy to 5,000 sequences. This does not discard any sequences from d_b, but it removes 95% of the data from d_a. OTUs with low representation may be discarded using this method. Relative abundance is a method which divides the count of each OTU by the sequencing depth such that the user is left with proportional representation of each OTU in each library. This method makes full use of all the data the user has acquired. Depending on the particular analysis, the user may prefer to use alternative methods implemented in high throughput sequencing statistical libraries such as edgeR (Robinson *et al.*, 2010) or DESeq (Anders and Huber, 2010).

3. Ensure that the order of samples in the metadata file (also known as a mapping file) matches the order of samples in the OTU count table.

4. Remove low prevalence OTUs from the data

Low prevalence or non-reproducible OTUs may add unnecessary noise to the dataset. There is no specific rule of thumb for removing low abundance OTUs, but one metric that has been previously used is to remove OTUs that are not present in at least 5% of the samples (Callahan *et al.*, 2016b; Edwards *et al.*, 2018).

5. Beta diversity plots (Figure 3A)

Beta diversity measures the differences in microbiota composition between the samples.

a. Calculate pairwise dissimilarities between each sample. There are ecologically appropriate metrics for this task such as Bray-Curtis, Jaccard, and UniFrac (Lozupone and Knight, 2005) dissimilarity metrics.

b. Using the calculated dissimilarities, perform principal coordinate analysis (PCoA).

c. Plot the resulting axes and color the points based upon the factor of interest.

6. Alpha diversity plots (Figure 3B)

Alpha diversity measures the diversity within each sample.

a. Calculate alpha diversity metric for each sample. Popular metrics are the Shannon index, the Simpson Index, species richness, and Faith's phylogenetic distance.

b. Plot resulting calculations, comparing the factors of interest.

7. Phylum level analysis (Figure 3C)

a. Summarize the mean representation of each phylum in each sample type.

b. Plot the results. There are many ways to plot these results. Here we have chosen to display the data using a stacked bar plot. We have also only retained the 10 most highly represented phyla.

8. OTU differential abundance (Figure 3D)

Note: The statistical distributions of bacterial and archaeal OTU abundances do not follow a Gaussian distribution and typically cannot be log-transformed to fit a normal distribution. Therefore, methods assuming normal distributions are not recommended for performing differential abundance tests. Available statistical packages (such as edgeR or DESeq) are recommended in order to properly model OTU distributions (McMurdie and Holmes, 2014).

a. Load non-normalized count data into the statistical package of choice.

b. Normalize for sequencing depth.

c. Model sample level and OTU level dispersions.

d. Fit the model using a design matrix.

e. Perform differential abundance tests.

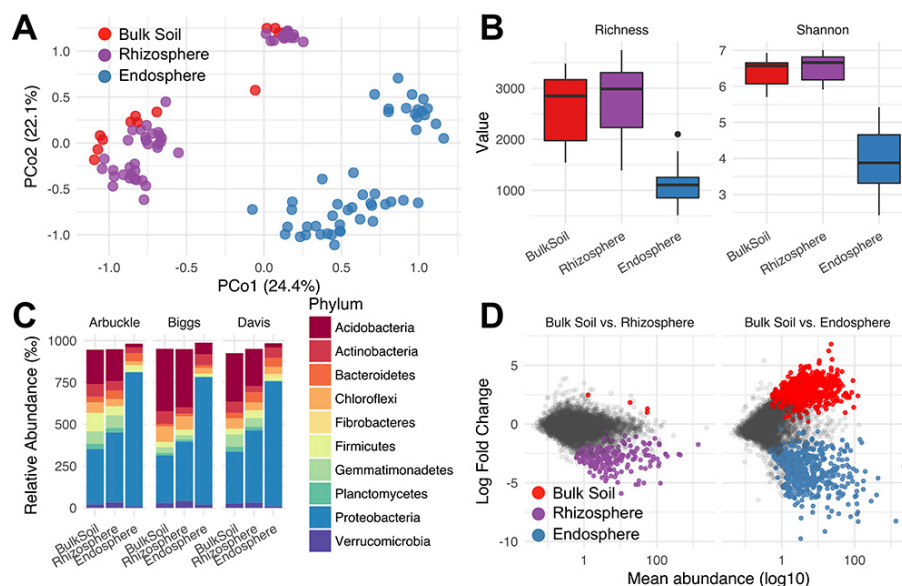f. Plot results with the average abundance on the x-axis and the fold change between the sample types on the y-axis.

Figure 3. Example analysis of amplicon microbiome data. A. Principal coordinates analysis showing microbial community structure between root compartments using Bray-Curtis dissimilarities. Each point represents the microbial community in one particular sample. B. Alpha diversity within each community using two commonly used metrics. Richness measures how many unique OTUs were detected in each sample while Shannon-entropy measures the randomness or uncertainty in a community. C. The distribution of the 10 most abundant phyla in the dataset. We show the similar compartments between sites host similar distributions of microbes when analyzing at the phylum level. D. Differentially abundant microbes in the

rhizosphere and endosphere compartments compared to bulk soil. Each point represents a single microbial OTU. The colored points represent OTUs that were significantly differentially abundant in one of the comparisons. The color of the point represents the direction of enrichment. Differential abundance analyses were carried out using DESeq2.

Notes

It should be noted that this protocol has been optimized for rice roots and there is no guarantee that the above sonication procedure will sufficiently remove rhizoplane microbes for other plant species. Similarly, this sonication procedure may be overly invasive for other plant species. Care should be taken to optimize the procedure for specific plant taxa.

Recipes

1. Phosphate buffered saline solution (1 L)
    8 g NaCl (Fisher Scientific)
    0.2 g KCl (Fisher Scientific)
    1.44 g Na_2HPO_4 (Fisher Scientific)
    0.24 g KH_2PO_4 (Fisher Scientific)
    Adjust pH to 7.4

Acknowledgments

Competing interests

The authors declare no competing interests.

References

1. Al-Ghalith, G. A., Montassier, E., Ward, H. N. and Knights, D. (2016). NINJA-OPS: fast accurate marker gene alignment using concatenated ribosomes. *PLoS Comput Biol* 12(1): e1004658.
2. Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol* 11(10): R106.
3. Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F., Rauf, P., Huettel, B., Reinhardt, R., Schmelzer, E., Peplies, J., Gloeckner, F. O., Amann, R., Eickhorst, T. and Schulze-Lefert, P. (2012). Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* 488(7409): 91-95.
4. Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. and Holmes, S. P. (2016a). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13(7): 581-583.
5. Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J. and Holmes, S. P. (2016b). Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. *F1000Res* 5: 1492.
6. Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. and Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5): 335-336.
7. Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G. and Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6(8): 1621-1624.
8. DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72(7): 5069-5072.
9. Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A. and Alm, E. J. (2017). Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun* 8(1): 1784.
10. Edwards, J. A., Santos-Medellin, C. M., Liechty, Z. S., Nguyen, B., Lurie, E., Eason, S., Phillips, G. and Sundaresan, V. (2018). Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life cycle in field-grown rice. *PLoS Biol* 16(2): e2003862.

11. Edwards, J., Johnson, C., Santos-Medellin, C., Lurie, E., Podishetty, N. K., Bhatnagar, S., Eisen, J. A. and Sundaresan, V. (2015). Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc Natl Acad Sci U S A* 112(8): E911-920.

12. Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19): 2460-2461.

13. Finkel, O. M., Castrillo, G., Herrera Paredes, S., Salas Gonzalez, I. and Dangl, J. L. (2017). Understanding and exploiting plant beneficial microbes. *Curr Opin Plant Biol* 38: 155-163.

14. Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4): 357-359.

15. Lozupone, C. and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71(12): 8228-8235.

16. Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., Tremblay, J., Engelbrektson, A., Kunin, V., Del Rio, T. G., Edgar, R. C., Eickhorst, T., Ley, R. E., Hugenholtz, P., Tringe, S. G. and Dangl, J. L. (2012). Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 488(7409): 86-90.

17. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. and Neufeld, J. D. (2012). PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* 13: 31.

18. McMurdie, P. J. and Holmes, S. (2014). Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Comput Biol* 10: e1003531.

19. Raynaud, X. and Nunan, N. (2014). Spatial ecology of bacteria at the microscale in soil. *PLoS One* 9(1): e87217.

20. Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1): 139-140.

21. Santos-Medellín, C., Edwards, J., Liechty, Z., Nguyen, B. and Sundaresan, V. (2017). Drought stress results in a compartment-specific restructuring of the rice root-associated microbiomes. *MBio* 8(4).

22. Wagner, M. R., Lundberg, D. S., Del Rio, T. G., Tringe, S. G., Dangl, J. L. and Mitchell-Olds, T. (2016). Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nat Commun* 7: 12151.

23. Wang, Q., Garrity, G. M., Tiedje, J. M. and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73(16): 5261-5267.

24. Zarraonaindia, I., Owens, S. M., Weisenhorn, P., West, K., Hampton-Marcell, J., Lax, S., Bokulich, N. A., Mills, D. A., Martin, G., Taghavi, S., van der Lelie, D. and Gilbert, J. A. (2015). The soil microbiome influences grapevine-associated microbiota. *MBio* 6(2).