Endophytic microbiome variation at the level of a single plant seed

Authors

AF Bintarti[1], A Sulesky-Grieb2[,3], N Stopnisek[1,4], A Shade[1,2,3,4]*

Affiliations

1. Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48824
2. Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824
3. Program in Ecology, Evolution and Behavior, Michigan State University, East Lansing, MI 48824
4. The Plant Resilience Institute, Michigan State University, East Lansing, MI 48824

*Corresponding author, shadeash@msu.edu

Abstract

Like other plant compartments, the seed harbors a microbiome. The members of the seed microbiome are the first to colonize a germinating seedling, and they initiate the trajectory of microbiome assembly for the next plant generation. Therefore, the members of the seed microbiome are important for the dynamics of plant microbiome assembly and the vertical transmission of potentially beneficial symbionts. However, it remains challenging to

assess the microbiome at the individual seed level (and, therefore, for the future individual plant) due to low endophytic microbial biomass, seed exudates that can select for particular members, and high plant and plastid contamination of resulting reads. Here, we report a protocol for extracting metagenomic DNA from an individual seed (common bean, *Phaseolus vulgaris* L.) with minimal disruption of host tissue, which we expect to be generalizable to other medium- and large- seed plant species. We applied this protocol to quantify the 16S rRNA V4 and ITS2 amplicon composition and variability for individual seeds harvested from replicate common bean plants grown under standard, controlled conditions to maintain health. Using metagenomic DNA extractions from individual seeds, we compared seed-to-seed, pod-to-pod, and plant-to-plant microbiomes, and found highest microbiome variability at the plant level. This suggests that several seeds from the same plant could be pooled for microbiome assessment, given experimental designs that apply treatments at the maternal plant level. This study adds protocols and insights to the growing toolkit of approaches to understand the plant-microbiome engagements that support the health of agricultural and environmental ecosystems.

Keywords

16S rRNA, ITS, *Phaseolus vulgaris* L., legume, pod, fruit, DNA extraction, microbiome assembly

Introduction

Seed microbiomes offer a reservoir of microbiota that can be vertically passed from maternal plants to offspring (Mitter et al. 2017; Shade et al. 2017; Truyens et al. 2015) and some of these members have plant-beneficial phenotypes (Adam et al. 2018; Berg and Raaijmakers 2018; Bergna et al. 2018; López-López et al. 2010). Therefore, the seed microbiome is expected to a play a key role in plant health and fitness (Barret et al. 2015), and especially in the assembly and establishment of the developing plant's microbiome (Chesneau et al. 2020). This expected importance of the seed microbiome has fueled recent research activity to use high-throughput sequencing to characterize the seed microbiomes of various plants (e.g., Chartrel et al. 2021; Dai et al. 2020; Eyre et al. 2019; Raj et al. 2019; Rodríguez et al. 2020; Xing et al. 2018).

Seed microbiomes include microbial members that live on the seed surface as epiphytes and members that colonize inside the internal tissue of the seed as endophytes (Nelson 2018). Among these microbiome members, endophytes that closely associate with endosperm and embryo are more likely to be transmitted to the next plant generations than are seed-associated epiphytes (Barret et al. 2016; Nelson 2018). By itself, an endophytic association does not confirm that there is a functional benefit or co-evolutionary relationship between the plant and the microbiome member (Nelson 2018). However, endophytic microbes offer the first source of inoculum for the germinating seedling (as reviewed in Nelson 2018; Vujanovic and Germida 2017), and, given the potential for priority effects or pathogen exclusion, these members can have implications for the mature plant's microbial community structure. Therefore, understanding the endophytic seed microbiome is expected to provide insights into

mechanisms of seed facilitation of microbiome assembly and the vertical transmission of microbiome members over plant generations.

As is true for other plant compartments, different plant species or divergent crop lines/varieties/cultivars often have different seed microbiome composition or structure (Wassermann et al. 2019; Klaedtke et al. 2016; Johnston-Monje and Raizada 2011; López-López et al. 2010). However, many seed microbiome studies have reported generally high variability across seed samples from the same plant type and treatment (Bergna et al. 2018; López-López et al. 2010; Bintarti et al. 2020), with strong explanatory value of either seed origin/seed lot, geographic region or soil edaphic conditions (Chartrel et al. 2021; Klaedtke et al. 2016; Johnston-Monje and Raizada 2011; but see also Adam et al. 2018 for an exception). While these insights may call into question the proportion of "inherited" versus acquired seed microbiome members, the high microbiome variability may be in part due to methods applied to extract the microbial DNA from the seed compartment, and different methods applied across studies. For instance, some studies surface sterilize the seeds while others do not; some germinate the seed prior to microbiome analysis while others do not, etc. One source of microbiome variability could be the common practice of the pooling of many seeds from the same or different plants to produce a composite seed microbiome sample for DNA extraction. Because multiple seeds are investigated at once, it is unclear at what level the most microbiome variability is highest– the seed, the pod or fruit, the plant, or the field or treatment. This information is required to determine the necessary sample size in well-powered experimental designs. More importantly, the question of vertical transmission cannot directly be addressed without seed microbiome assessment of an individual.

Our study objectives were: 1) to determine the appropriate observational unit of endophytic seed microbiome assessment for common bean (*Phaseolus vulgaris* L) by quantifying seed-to-seed, pod-to-pod, and plant-to-plant variability in 16S rRNA V4 and ITS2 amplicon analyses; and 2) to develop a robust protocol for individual seed microbiome extraction that could be generally applied to other plants that have similarly medium- to large-sized seeds. We found that that plant-to-plant variability under controlled growth conditions exceeded within-plant variability and conclude that seeds can be pooled by maternal plant (but, not across different plants) in study designs that aim to compare seed microbiomes resulting from treatments applied at the plant level.

Materials and Methods

Plant growth conditions

Because we targeted the endophytic seed microbiome, surface sterilization of the bean seeds was conducted before germination and planting. To sterilize, seeds were soaked in a solution of 10% bleach with 0.1% Tween20 for 15 minutes, then rinsed four times with sterile water. The final rinse water was plated on tryptic soy agar (TSA) and potato dextrose agar (PDA) plates to test for sterilization efficacy. Sterilized seeds were placed in Petri dishes on sterile tissue paper moistened with sterile water, and allowed to germinate in in the dark for four days. After four days, the radicle had emerged and the germinated seeds were ready to be transferred to the growth chamber. The germinated seeds were planted in three 4.54 L (1-gallon) pots filled with a 50:50 v/v mixture of agricultural bean field soil and vermiculite. The pots were placed in a BioChambers model SPC-37 growth chamber with a 14-h day/10-h night

cycle at 26°C and 22°C, respectively, 260 mE light intensity, and 50% relative humidity. All plants received 300 mL of water every other day and 200 mL of half-strength Hoagland solution (Hoagland and Arnon 1950) once a week.

Study design

We planted three germinated seeds per pot and culled to one seedling per pot at the early vegetative growth stage. There were three plant replicates designated as A, B, and C, grown under the above-described conditions for normal, healthy growth. The three plants produced different numbers of pods and total seeds (plant A = 5 pods, 22 seeds; plant B = 6 pods, 29 seeds; and plant C = 7 pods, 26 seeds) with the number of seeds varying across pods (2 to 6 seeds per pod). We aimed to balance and maximize number of seeds across plants. Therefore, we extracted metagenomic DNA from 3 pods from plants A and C, and 6 pods from plant B, with 3 to 4 seeds in each pod. For the 16S V4 analysis we had 3 pods from plant A (A1, A2 and A3= 4 seeds), 6 pods from plant B (B1 through B6 = 4 seeds), and 3 pods from plant C (C5= 3 seeds, C6 and C7= 4 seeds) for a total of 47 individual seed samples analyzed. For the ITS2 analysis, we were unable to amplify fungal target DNA from pod A1 or pod B1, for a total of 45 individual seed samples analyzed.

Seed harvest and endophyte metagenomic DNA extraction

Once the plants reached maturity at the R9 growth stage (yellowing leaves and dry pods), the seeds were harvested for endophytic microbiome analysis. Seeds were distinguished by plant and pod. The endophytic microbiome of each seed was extracted and sequenced

132 individually. To extract the endophytic metagenomic DNA (mgDNA), a protocol was adapted
133 from Barret et al. 2015 and Rezki et al. 2018. First, the seeds were surface-sterilized as above
134 and the seed coat was carefully removed using sterilized forceps. Each seed was then soaked in
135 3 mL of PBS solution with 0.05% Tween20 (hereafter, "soaking solution") overnight at 4°C with
136 constant agitation of 170 rpm. Since low levels of microbial biomass are expected in single seed
137 extractions, positive and negative controls were included in the extraction protocol. This
138 ensures that if no extractable microbial DNA is present in a sample that it is representative of
139 the sample, rather than the extraction methods. A mock community was used as a DNA
140 extraction positive control by adding one, 75 µL aliquot of the ZymoBIOMICS™ Microbial
141 Community Standard (Zymo Research, Irvine, CA, United States) to 3 mL of the soaking solution
142 immediately prior to conducting the extraction protocol. Sterile soaking solution (3 mL) was
143 used as a negative DNA extraction control.
144 After soaking overnight, the samples were centrifuged at 4500xg for 60 minutes at 4°C
145 to pellet any material that had been released from the seed tissues. After centrifugation the
146 seed was removed, and the pelleted material was resuspended in 1-2 mL of supernatant and
147 transferred to a microcentrifuge tube for DNA extraction using the E.Z.N.A® Bacterial DNA Kit
148 (Omega Bio-tek, Inc. Norcross, GA, United States). The manufacturer's Centrifugation Protocol
149 was used with minor modifications. Specifically, the pelleted seed material was suspended in TE
150 buffer (step 4), the incubation for the lysozyme step was extended to 20 minutes, 30 µL of
151 elution buffer was used, and the elution step was extended to a 15 minute incubation. These
152 modifications were performed to maximally recover the limited amount of mgDNA expected

153 from a single seed. We detail the standard operating protocol, and provide notes on the
154 alternatives that we tested in optimizing this protocol in the Supplementary Material.
155
156 ***PCR amplification and amplicon sequencing***
157 To confirm successful DNA extraction from the seed pellet, DNA quantification and
158 target gene polymerase chain reaction (PCR) assays were performed. First, the DNA extracted
159 from the seed samples and the positive and negative controls were quantified using the
160 Qubit™dsDNA BR Assay Kit (ThermoFisher Scientific, Waltham, MA, United States). Then, PCR
161 amplification and sequencing of the V4 region of 16S rRNA bacterial/archaeal gene and the ITS2
162 region of the ITS fungal gene were performed. The V4 region of 16S rRNA gene amplification
163 was conducted using 515f (5'-GTGCCAGCMGCCGCGGTAA-3') and 806r (5'-
164 GGACTACHVGGGTWTCTAAT-3') universal primers (Caporaso et al. 2011) under the following
165 conditions: 94°C for 3 min, followed by 35 cycles of 94°C (45 s), 50°C (60 s), and 72°C (90 s), with
166 a final extension at 72°C (10 min). The amplification was performed in 25 µl mixtures containing
167 12.5 µl GoTaq®Green Master Mix (Promega, Madison, WI, United States), 0.625 µl of each
168 primer (20 µM), 2 µl of DNA template (~1 ng per µl), and 9.25 µl nuclease free water. The
169 mgDNA (concentration of ~ 1 ng per µl) was sequenced at the Research Technology Support
170 Facility (RTSF) Genomics Core, Michigan State sequencing facility using the Illumina MiSeq
171 platform v2 Standard flow cell. The sequencing was performed in a 2x250bp paired end format.
172 The PCR amplification of the ITS2 region of the ITS gene was performed using ITS86f (5'-
173 GTGAATCATCGAATCTTTGAA-3') and ITS4 (5'- TCCTCCGCTTATTGATATGC-3') primers (Op De
174 Beeck et al. 2014) with addition of index adapters by the RTSF Genomics Core. The PCR

amplification of the ITS2 was conducted under the following conditions: 95°C for 2 min, followed by 40 cycles of 95°C (30 s), 55°C for (30 s), and 72°C for (1 min), with a final extension at 72°C for 10 min. The amplification was performed in 50 µl mixture containing 20 µl GoTaq®Green Master Mix (Promega, Madison, WI, United States), 1 µl of each primer (10 µM), 1 µl of DNA template (~ 1 ng per µl), and 27 µl nuclease free water. The PCR products were purified using QIAquick® PCR Purification Kit (QIAGEN, Hilden, Germany). Purified PCR products with a concentration range 6-10 ng per µl were sequenced at the RTSF Genomics Core using Illumina MiSeq platform v2 Standard flow cell and 2x250bp paired end format.

Sequence analysis

The USEARCH pipeline (v.10.0.240) was used to merge paired-end bacterial/archaeal raw reads, filter for low-quality sequences, dereplicate, remove singletons, denoise, and check for chimeras (Edgar and Flyvbjerg 2015). An in-house open reference strategy was performed for OTU clustering (Rideout et al. 2014). First, closed-reference OTU picking was performed by clustering the quality filtered reads against the SILVA database (v.132) (Quast et al. 2013) at 97% identity using USEARCH algorithm (usearch_global command) (Edgar 2010). Then, de novo OTU picking process was performed on the reads that failed to match the reference using UPARSE-OTU algorithm (cluster_otus command) (Edgar 2013) at 97% identity. Finally, closed-reference and de novo OTUs were combined into a full set of representative sequences. The merged sequences were then mapped back to the representative sequences using the usearch_global command.

Sequence alignment, taxonomy assignment, non-bacteria/archaea filtering, and phylogenetic diversity calculation were performed using QIIME 1.9.1. The representative sequences were aligned against the SILVA database (v.132) (Caporaso, Kuczynski, et al. 2010) using PyNAST (Caporaso, Bittinger, et al. 2010). The unaligned OTUs and sequences were excluded from the OTU table and the representative sequences file, respectively. Taxonomy assignment was performed using the default classifier method (UCLUST algorithm) at a minimum confidence of 0.9 (Edgar 2010) using SILVA database (v.132) as the reference. Plant contaminants (chloroplast and mitochondria) and unassigned taxa were removed from the OTU table and the representative sequences using filter_taxa_from_otu_table.py and filter_fasta.py command. Filtering the microbial contaminants from the OTU table was conducted in R (v.3.4.2) (R Core Development Team) using the microDecon package (McKnight et al. 2019). Reads were normalized using Cumulative Sum Scaling (CSS) method in metagenomeSeq Bioconductor package on R (Paulson et al. 2013).

The fungal ITS raw reads were processed using the USEARCH (v.10.0.240) pipeline. Read processing included merging paired-end reads, removing primers using cutadapt (v.2.1) (Martin 2011), dereplication, and singleton removal. OTUs were picked and chimeras removed using de novo clustering at 97% identity threshold with the UPARSE-OTU algorithm (cluster_otus command, Edgar 2013). Then, all merged sequences were mapped to the clustered reads using usearch_global command to generate an OTU table. Fungal taxonomic classification was performed in CONSTAX (Gdanetz et al. 2017) using RDP Classifier (v.11.5) (Cole et al. 2014; Wang et al. 2007) at a minimum confidence of 0.8 and with the UNITE reference database (release 01-12-2017). Plant and microbial contaminants removal and read normalization were

218 performed in R (v.3.4.2). Plant contaminants were removed from the OTU table by filtering out
219 OTUs that were assigned into Kingdom Plantae. Microbial contaminants were removed using
220 the microDecon package (McKnight et al. 2019). The CSS method from the metagenomeSeq
221 Bioconductor package was performed to normalize the fungal reads (Paulson et al. 2013).
222
223 ***Microbial community analysis***
224 Microbiome statistical analyses were conducted in R (v.3.4.2) (R Core Development
225 Team). Microbial alpha and beta diversity were calculated on the CSS- normalized OTU table
226 using the vegan package (v.2.5-7) (Oksanen et al. 2019). Richness and Faith's phylogenetic
227 diversity were used to analyze the bacterial/archaeal alpha diversity. For fungal alpha diversity,
228 we used richness. The evenness of the seed microbiomes was visualized using rank-abundance
229 curves (Phyloseq package (v.1.28.0) in R (McMurdie and Holmes 2013)). Differences in alpha
230 diversity among plants and pods were determined by fitting the Linear Mixed-Effects Model
231 (LMM) using the lme() function of the nlme package (version 3.1-152) (Pinheiro et al. 2021). We
232 performed LMM because the study has an unbalanced nested design with pod as the random
233 factor, nested within plant as the fixed factor. Microbial composition and relative abundance
234 were analyzed using the Phyloseq package (v.1.28.0) in R (McMurdie and Holmes 2013).
235 Beta diversity was calculated using Jaccard distances and visualized using principal
236 coordinate analysis (PCoA) plot. We used the Jaccard index, which is based on presence-
237 absence, rather than a metric based on relativized abundance because we reasoned that the
238 seed microbiome members are likely to be dormant inside the seed prior to germination (Cope-
239 Selby et al. 2017), and that any differences in relative abundances are not directly attributable

to competitive fitness outcomes inside the seed. Furthermore, exponential growth would allow that any viable cell successfully packaged and passaged via the seed could, in theory, successfully colonize the new plant. Nested permutational multivariate analysis of variance (PERMANOVA) using the function nested.npmanova() from the BiodiversityR package (Kindt 2020) was performed to assess the microbial community structure among plants and pods. We performed multivariate analysis to check the homogeneity of dispersion (variance) among groups using the function betadisper() (Oksanen et al. 2019). We performed PERMADISP to test the significant differences in dispersions between groups and Tukey's HSD test to determine which groups differ in relation to the dispersions (variances).

Power analysis and sample size were calculated using the pwr.t.test() function from the pwr package (v.1.3-0). We performed power analysis of two-category t-test. Because the most microbiome variability was observed across plants, we pooled individual seed sequence profiles in silico at the plant level for this analysis. We calculated Cohen's d effect size given the information of mean and standard deviation of bacterial/archaeal alpha diversity (richness and phylogenetic diversity) from three plant samples from this study: Plant A (n = 12; richness: M = 30.58, SD = 6.42, phylogenetic diversity: M = 4.17, SD = 0.89), Plant B (n = 24; richness: M = 18.21, SD = 7.35, phylogenetic diversity: M = 2.92, SD = 0.82) and Plant C (n= 11; richness: M = 19.09, SD = 10.95, phylogenetic diversity: M = 3.09, SD = 1.39). We calculated the common standard deviation (σpool of all groups) using the above information, then we calculated Cohen's d effect size for both richness and phylogenetic diversity. Cohen's d effect size was defined by calculating the difference between the largest and smallest means divided by the square root of the mean square error (or the common standard deviation). Power analysis was

run with Hedges's g effect size (corrected with Cohen's d effect size) and significant level of 0.05.

Data and code availability

The computational workflows for sequence processing and ecological statistics are available on GitHub (https://github.com/ShadeLab/Bean_seed_variability_Bintarti_2021). Raw sequence data of bacteria/archaea and fungi have been deposited in the Sequence Read Archive (SRA) NCBI database under Bioproject accession number PRJNA714251.

Results

Sequencing summary and microbiome coverage

A total of 5,056,769 16S rRNA V4 and 8,756,009 ITS2 quality reads were generated from 47 mgDNA samples purified from individual seeds for bacteria/archaea, and from 45 samples for fungi. We removed more than 90 % of reads that were plant contaminants (**Fig. S1**), resulting in 17,128 and 67,878 16S rRNA bacterial/archaeal and ITS fungal reads, respectively. After removing plant and microbial contaminants, we determined 211 bacterial/archaeal and 57 fungal operational taxonomic units (OTUs) defined at 97% sequence identity. While the majority of individual seeds from plants A and B had exhaustive to sufficient sequencing effort, some seeds from plant C did not (**Fig. 1a**). However, the fungal rarefaction curves reached asymptote and had sufficient effort (**Fig. 1b**). Both bacterial/archaeal and fungal seed microbiomes were highly uneven with few dominant and many rare taxa, as typical for microbiomes (**Fig 1c,d**).

284

285 ***Microbiome Diversity***

286 There were differences in bacterial/archaeal richness among seeds from different plants
287 (LMM; df = 2, F-value = 6.91, p-value = 0.015) (**Fig. 2a**), where plant B and C had lower seed
288 richness than plant A (Tukey's HSD post hoc test; p-value = 0.001 and 0.006, respectively).
289 However, bacterial/archaeal richness among seeds from pods collected from the same plant
290 were not different (LMM, p-value > 0.05) (**Fig. 2b**). Similarly, bacterial/archaeal phylogenetic
291 diversity were different among seeds collected from different plants (LMMs; df = 2, F-value =
292 6.56, p-value = 0.003) (**Fig. 2c**), but not among seeds from pods within the same plant (LMM, p-
293 value > 0.05) (**Fig. 2d**). Plants B and C had lower seed microbiome bacterial/archaeal
294 phylogenetic diversity compared to plant A (Tukey's HSD post hoc test, p-value = 0.001 and
295 0.013, respectively). We observed no differences in fungal richness among seeds from different
296 plants (LMM; df = 2, F-value = 1.11, p-value = 0.37) (**Fig. 2e**), and among seeds from pods within
297 the same plant (LMM, p-value > 0.05) (**Fig. 2f**). To summarize, these results suggest that seed
298 bacterial/archaeal alpha diversity, but not fungal, varied plant to plant.

299 We detected a difference in seed bacterial/archaeal structure among plants (nested
300 PERMANOVA, df =2, F-ratio = 2.94, p-value = 0.001) (**Fig. 3a**), but, again, not among pods from
301 the same plant (nested PERMANOVA, df =9, F-ratio = 0.99, p-value = 0.63). Though separation
302 among pods and plants are not obvious on the PCoA for the fungal seed microbiomes, we
303 detected modest differences in fungal community structure among seeds from different plants
304 (nested PERMANOVA, df =2, F-rati0 = 1.55, p-value = 0.02) (**Fig. 3b**), as well as among seeds
305 from pods from the same plant (nested PERMANOVA, df =9, F-rati0 = 1.27, p-value = 0.03). An

306 analysis of beta-dispersion revealed that there were differences in seed microbiome dispersion
307 across different plants for bacterial/archaeal communities (PERMADISP, df = 2, F-value = 63.9,
308 p-value = 0.001) (**Fig.3c**), but not for fungal communities (PERMADISP, df = 2, F-value = 0.22, p-
309 value = 0.798) (**Fig. 3d**). Therefore, statistical differences in the seed microbiome across plants
310 for the bacteria/archaea may be attributed to either centroid or dispersion, while fungal seed
311 communities were different by centroid.
312
313 ***Bean seed microbiome composition***
314 We identified 135 bacterial/archaeal and 49 fungal taxa at the genus level. The
315 bacterial/archaeal individual seed communities were dominated by taxa from class
316 Gammaproteobacteria (50.47%), Bacilli (24.48%), Alphaproteobacteria (8.68 %), and
317 Bacteroidia (6.59 %) (**Fig. 4a**), and include Pseudomonas (13.58 %), Bacillus (10.2 %),
318 Acinetobacter (9.5 %), Raoultella (7.09%), and Escherichia-Shigella (5.19%) as the major genera.
319 Among members of the class Alphaproteobacteria, we also found genera Bradyrhizobium and
320 Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium with relative abundance of 2.57 and
321 0.85 %, respectively. Although seed fungal community composition varied among plants and
322 also pods within plant, the fungal community was dominated by taxa belonging to classes
323 Pezizomycetes (53.44 %), Agaricomycetes (25.7 %), and Dothideomycetes (11.17 %) (**Fig. 4b**),
324 and the genera Helvella (53.44 %), Gautieria (19.65%), Acidomyces (7.29 %),
325 Capnodiales_unidentified_sp_23791 (2.52 %), and Phlebiopsis (1.82%).
326 A key objective of this research was to understand the sources of variability in the
327 individual bean seed microbiome to inform future study design. Because we found that the

328 plant-to-plant seed microbiome variability was highest when grown in control conditions, we
329 performed a power analysis to determine how many plants would be required to observe a
330 treatment effect from seed samples pooled per plant. To detect the effect of treatment to
331 bacterial/archaeal richness and phylogenetic diversity, pooled seeds from 9 and 12 plants are
332 needed, respectively, for 16S rRNA richness and phylogenetic diversity, to achieve power of 0.8;
333 and 13 and 19 plants to achieve power of 0.95 (**Fig. 5**).
334
335 **Discussion**
336 There remain gaps in our understanding of the persistence and assembly of seed
337 microbiome members, especially across plant generations, and which microbiome members are
338 beneficial and actively selected by, or even co-evolved with, the host. Here, we investigated the
339 variability of the common bean microbiome at the resolution of the individual seed, which is
340 the unit that delivers any vertically transmitted microbiome to the offspring. Because multiple
341 legume seeds within a pod develop as a result of a single flower pollination, one simple
342 hypothesis is that the individual seeds within a pod may harbor a highly similar microbiome if
343 the floral pathway of assembly is prominent. However, recent work has suggested that the
344 endophytic seed microbiome of green bean varieties of common bean likely colonize
345 predominantly via the internal vascular pathway, and not the floral pathway (Chesneau et al.
346 2020), which may result in more homogeneity among seed microbiomes of the same plant.
347 Our data support this finding, as seeds from the same plant (and therefore a common vascular
348 pathway across pods) had relatively low microbiome variability, especially as compared across
349 plants. It is expected that the vascular pathway of seed microbiome assembly is more likely to

colonize the internal seed compartments (e.g., embryo), and therefore more likely to be vertically transmitted (Barret et al. 2016). It is yet unclear whether plant species that have a stronger relative importance of the floral pathway in seed microbiome assembly may exhibit higher microbiome variability at the pod/fruit level. Such an outcome may indicate that the experimental unit should instead be the pod level rather than the plant level for plant species dominated by floral assembly pathways.

There are many challenges in analyzing the microbiome of seeds generally and of a single seed in particular, which may be why cultivation-independent studies of single seeds are few (Abdelfattah et al. 2021). Previous studies showed that seeds have low microbial biomass and diversity (Adam et al. 2018; Chesneau et al. 2020; Rezki et al. 2016), especially relative to other plant compartments or soil. Therefore, many studies pool seeds to analyze the aggregated microbiome of many seeds and to get enough microbial biomass for metagenomic DNA extraction (Latz et al. 2021; Bergna et al. 2018; Wassermann et al. 2019; Adam et al. 2018; Johnston-Monje and Raizada 2011; Klaedtke et al. 2016). Generally, microbiome samples that have low biomass have numerous challenges in sequence-based analysis, as discussed elsewhere (Eisenhofer et al. 2019; Bender et al. 2018). First, unknown contaminants, either from nucleic acid kits or from mishandling of the samples, can have relatively high impact on the observed community composition, and so extraction and PCR controls are needed for assessment of contaminants and subtraction of suspected contaminants from the resulting community (Davis et al. 2018). Second, the sparse datasets (e.g., many zero observations for many taxa in many samples) generated from low biomass samples often require special statistical consideration and data normalization (Weiss et al. 2017; Anderson et al. 2011).

372 Plant host contamination of the microbiome sequence data is another consideration
373 expected with analysis of the seed, and this challenge also applies to other plant compartments
374 (Fitzpatrick et al. 2018; Song and Xie 2020). For 16S rRNA amplicon sequencing, the
375 contaminant reads typically derive from host mitochondria and chloroplasts, but ITS or 18S
376 amplicon analysis may also have reads annotated as Plantae. Therefore, nucleic acid extractions
377 may attempt minimal disturbance of the plant tissue that is that target of microbiome
378 investigation; for example, grinding tissues to include in the extraction will result in higher plant
379 DNA contamination than separating microbial biomass from intact tissue. For seeds in
380 particular, it is known that seeds can exude both antimicrobials and attractants to select for
381 particular microbial members early in microbiome assembly of the germinated seed and
382 emerging seedling (Chesneau et al. 2020; Meldau et al. 2012), and there is an active zone of
383 plant and microbiome activity at the seed-soil-interface of a germinating seed (the
384 spermosphere, e.g., Schiltz et al. 2015). Therefore, to target the native endophytic seed
385 microbiome without also allowing for the plant's potential selection for or filtering against
386 particular members, it is important to use dormant seeds and also to minimally disrupt the seed
387 compartment during extraction. Notably, many protocols have opted to first germinate seeds
388 and, therefore, study the outcome of any plant selection prior to analyzing the seed
389 microbiome (Wassermann et al. 2021; Bergna et al. 2018; López-López et al. 2010).
390 Taking all of these methodological aspects into consideration, this study presents a
391 protocol and analysis pipeline for endophyte microbiome DNA extraction from a single dormant
392 seed that experiences minimal tissue disruption in the extraction process, includes both
393 positive and negative sequencing controls, and includes bioinformatic steps to identify

394 contamination and remove host signal from the marker gene amplification. Notably, we chose
395 to perform microbiome analysis based on a presence/absence taxon table rather than a table
396 with relativized taxon abundances. This was done in consideration of the ecology of the seed
397 endophyte microbiome members to likely be dormant until germination (Cope-Selby et al.
398 2017), and therefore the differences in relativized abundances do not reflect differences in
399 fitness outcomes inside the dormant seed. We acknowledge that relative abundances could
400 reflect differential microbiome member recruitment by the host plant, but this is not the
401 objective of the study and would be best addressed with a different design to determine the
402 multi-generation consistency and transmission rates of any observed enrichments, which would
403 be supported by assessment of the seed microbiome within individual seeds, and across plant
404 generations.

405 In conclusion, individual seed microbiome assessment provides improved precision in
406 our understanding of plant microbiome assembly and sets the stage for studies of vertical
407 transmission. We found that seeds produced by an individual bean plant can be considered as
408 a unit (for comparative treatment study designs), and that seeds produced by different plants
409 are expected to have slightly different microbiomes, even if grown under the same, controlled
410 conditions and in the same soil source. Future work may consider whether functional
411 redundancy in plant beneficial phenotypes across seed microbiome members may provide one
412 mechanism for consistent outcomes in beneficial plant microbiome establishment.

413

414 **Acknowledgments**

415 This research was supported by USDA 2019-67019-29305. AS acknowledges support
416 from the USDA National Institute of Food and Agriculture and Michigan State University
417 AgBioResearch (Hatch). AFB acknowledges a doctoral fellowship from the Fulbright Foundation
418 and from Michigan State University. NS acknowledges support from the Michigan State
419 University Plant Resilience Institute.

420

References

Abdelfattah, A., Wisniewski, M., Schena, L., and Tack, A. J. M. 2021. Experimental evidence of microbial inheritance in plants and transmission routes from seed to phyllosphere and root. Environ. Microbiol.

Adam, E., Bernhart, M., Müller, H., Winkler, J., and Berg, G. 2018. The Cucurbita pepo seed microbiome: genotype-specific composition and implications for breeding. Plant Soil. 422:35–49.

Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L., et al. 2011. Navigating the multiple meanings of β diversity: A roadmap for the practicing ecologist. Ecol. Lett. 14:19–28.

Barret, M., Briand, M., Bonneau, S., Préveaux, A., Valière, S., Bouchez, O., et al. 2015. Emergence shapes the structure of the seed microbiota. Appl. Environ. Microbiol. 81:1257–1266.

Barret, M., Guimbaud, J. F., Darrasse, A., and Jacques, M. A. 2016. Plant microbiota affects seed transmission of phytopathogenic microorganisms. Mol. Plant Pathol. 17:791–795.

Op De Beeck, M., Lievens, B., Busschaert, P., Declerck, S., Vangronsveld, J., and Colpaert, J. V. 2014. Comparison and validation of some ITS primer pairs useful for fungal metabarcoding studies. PLoS One. 9:e97629.

Bender, J. M., Li, F., Adisetiyo, H., Lee, D., Zabih, S., Hung, L., et al. 2018. Quantification of variation and the impact of biomass in targeted 16S rRNA gene sequencing studies. Microbiome. 6:155.

Berg, G., and Raaijmakers, J. M. 2018. Saving seed microbiomes. ISME J. 12:1167–1170.

443 Bergna, A., Cernava, T., Rändler, M., Grosch, R., Zachow, C., and Berg, G. 2018. Tomato seeds
444 preferably transmit plant beneficial endophytes. Phytobiomes J. 2:183–193.
445 Bintarti, A. F., Kearns, P. J., Sulesky, A., and Shade, A. 2020. Abiotic treatment to common bean
446 plants results in an altered seed microbiome. bioRxiv. Available at:
447 https://doi.org/10.1101/2020.06.05.134445.
448 Caporaso, J. G., Bittinger, K., Bushman, F. D., Desantis, T. Z., Andersen, G. L., and Knight, R.
449 2010. PyNAST: A flexible tool for aligning sequences to a template alignment.
450 Bioinformatics. 26:266–267.
451 Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al.
452 2010. QIIME allows analysis of high-throughput community sequencing data. Nat.
453 Methods. 7:335–336.
454 Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et
455 al. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per
456 sample. Proc. Natl. Acad. Sci. U. S. A. 108:4516–4522.
457 Chartrel, V., Dugat-Bony, E., Sarthou, A. S., Huchette, S., Bonnarme, P., and Irlinger, F. 2021. The
458 microbial community associated with pea seeds (Pisum sativum) of different
459 geographical origins. Plant Soil.
460 Chesneau, G., Torres-Cortes, G., Briand, M., Darrasse, A., Preveaux, A., Marais, C., et al. 2020.
461 Temporal dynamics of bacterial communities during seed development and maturation.
462 FEMS Microbiol. Ecol. 96.
463 Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. 2014. Ribosomal
464 Database Project: Data and tools for high throughput rRNA analysis. Nucleic Acids Res.

42.

Cope-Selby, N., Cookson, A., Squance, M., Donnison, I., Flavell, R., and Farrar, K. 2017. Endophytic bacteria in Miscanthus seed: implications for germination, vertical inheritance of endophytes, plant evolution and breeding. GCB Bioenergy. 9:57–77.

Dai, Y., Li, X. Y., Wang, Y., Li, C. X., He, Y., Lin, H. H., et al. 2020. The differences and overlaps in the seed-resident microbiome of four Leguminous and three Gramineous forages. Microb. Biotechnol. 13:1461–1476.

Davis, N. M., Proctor, Di. M., Holmes, S. P., Relman, D. A., and Callahan, B. J. 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 6.

Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 26:2460–2461.

Edgar, R. C. 2013. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. Nat. Methods. 10:996–998.

Edgar, R. C., and Flyvbjerg, H. 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics. 31:3476–3482.

Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., and Weyrich, L. S. 2019. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. Trends Microbiol. 27:105–117.

Eyre, A. W., Wang, M., Oh, Y., and Dean, R. A. 2019. Identification and characterization of the core rice seed microbiome. Phytobiomes J. 3:148–157.

Fitzpatrick, C. R., Lu-Irving, P., Copeland, J., Guttman, D. S., Wang, P. W., Baltrus, D. A., et al.

2018. Chloroplast sequence variation and the efficacy of peptide nucleic acids for blocking host amplification in plant microbiome studies. Microbiome. 6.

Gdanetz, K., Benucci, G. M. N., Vande Pol, N., and Bonito, G. 2017. CONSTAX: A tool for improved taxonomic resolution of environmental fungal ITS sequences. BMC Bioinformatics. 18.

Hoagland, D. R., and Arnon, D. I. 1950. The water-culture method for growing plants without soil. Circ. Calif. Agric. Exp. Stn. 347:32.

Johnston-Monje, D., and Raizada, M. N. 2011. Conservation and Diversity of Seed Associated Endophytes in Zea across Boundaries of Evolution, Ethnography and Ecology. PLoS One. 6:e20396.

Kindt, R. 2020. BiodiversityR: Package for Community Ecology and Suitability Analysis. R package version 2.13-1. https://cran.r-project.org/web/packages/BiodiversityR/index.html.

Klaedtke, S., Jacques, M. A., Raggi, L., Préveaux, A., Bonneau, S., Negri, V., et al. 2016. Terroir is a key driver of seed-associated microbial assemblages. Environ. Microbiol. 18:1792–1804.

Latz, M. A. C., Kerrn, M. H., Sørensen, H., Collinge, D. B., Jensen, B., Brown, J. K. M., et al. 2021. Succession of the fungal endophytic microbiome of wheat is dependent on tissue-specific interactions between host genotype and environment. Sci. Total Environ. 759:143804.

López-López, A., Rogel, M. A., Ormeño-Orrillo, E., Martínez-Romero, J., and Martínez-Romero, E. 2010. Phaseolus vulgaris seed-borne endophytic community with novel bacterial

509 species such as Rhizobium endophyticum sp. nov. Syst. Appl. Microbiol. 33:322–327.
510 Martin, M. 2011. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing
511 Reads. EMBnet.journal. 17:10–12.
512 McKnight, D. T., Huerlimann, R., Bower, D. S., Schwarzkopf, L., Alford, R. A., and Zenger, K. R.
513 2019. microDecon: A highly accurate read-subtraction tool for the post-sequencing
514 removal of contamination in metabarcoding studies. Environ. DNA. 1:14–25.
515 McMurdie, P. J., and Holmes, S. 2013. Phyloseq: An R Package for Reproducible Interactive
516 Analysis and Graphics of Microbiome Census Data. PLoS One. 8.
517 Meldau, D. G., Long, H. H., and Baldwin, L. T. 2012. A native plant growth promoting bacterium,
518 Bacillus sp. B55, rescues growth performance of an ethylene-insensitive plant genotype
519 in nature. Front. Plant Sci. 3.
520 Mitter, B., Pfaffenbichler, N., Flavell, R., Compant, S., Antonielli, L., Petric, A., et al. 2017. A new
521 approach to modify plant microbiomes and traits by introducing beneficial bacteria at
522 flowering into progeny seeds. Front. Microbiol. 8:1–10.
523 Nelson, E. B. 2018. The seed microbiome: Origins, interactions, and impacts. Plant Soil. 422:7–
524 34.
525 Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. 2019. vegan:
526 Community Ecology Package. R package version 2.5-7. https://CRAN.R-
527 project.org/package=vegan.
528 Paulson, J. N., Colin Stine, O., Bravo, H. C., and Pop, M. 2013. Differential abundance analysis for
529 microbial marker-gene surveys. Nat. Methods. 10:1200–1202.
530 Pinheiro, J., Bates, D., DebRoy, S., and Sarkar, D. 2021. nlme: Linear and Nonlinear Mixed Effects

Models. R package version 3.1-152. https://CRAN.R-project.org/package=nlme.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 41:D590–D596.

Raj, G., Shadab, M., Deka, S., Das, M., Baruah, J., Bharali, R., et al. 2019. Seed interior microbiome of rice genotypes indigenous to three agroecosystems of Indo-Burma biodiversity hotspot. BMC Genomics. 20:924.

Rezki, S., Campion, C., Iacomi-Vasilescu, B., Preveaux, A., Toualbia, Y., Bonneau, S., et al. 2016. Differences in stability of seed-associated microbial assemblages in response to invasion by phytopathogenic microorganisms. PeerJ. 4:e1923.

Rezki, S., Campion, C., Simoneau, P., Jacques, M.-A., Shade, A., and Barret, M. 2018. Assembly of seed-associated microbial communities within and across successive plant generations. Plant Soil. 422:67–79.

Rideout, J. R., He, Y., Navas-Molina, J. A., Walters, W. A., Ursell, L. K., Gibbons, S. M., et al. 2014. Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. PeerJ. 2:e545.

Rodríguez, C. E., Antonielli, L., Mitter, B., Trognitz, F., and Sessitsch, A. 2020. Heritability and functional importance of the setaria viridis bacterial seed microbiome. Phytobiomes J. 4:40–52.

Schiltz, S., Gaillard, I., Pawlicki-Jullian, N., Thiombiano, B., Mesnard, F., and Gontier, E. 2015. A review: What is the spermosphere and how can it be studied? J. Appl. Microbiol. 119:1467–1481.

553 Shade, A., Jacques, M.-A. M. A. M.-A., and Barret, M. 2017. *Ecological patterns of seed*
554 *microbiome diversity, transmission, and assembly*.
555 Song, L., and Xie, K. 2020. Engineering CRISPR/Cas9 to mitigate abundant host contamination
556 for 16S rRNA gene-based amplicon sequencing. Microbiome. 8.
557 Truyens, S., Weyens, N., Cuypers, A., and Vangronsveld, J. 2015. Bacterial seed endophytes:
558 Genera, vertical transmission and interaction with plants. Environ. Microbiol. Rep. 7:40–
559 50.
560 Vujanovic, V., and Germida, J. J. 2017. Seed endosymbiosis: A vital relationship in providing
561 prenatal care to plants. Can. J. Plant Sci. 97:971–981.
562 Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. 2007. Naïve Bayesian classifier for rapid
563 assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ.
564 Microbiol. 73:5261–5267.
565 Wassermann, B., Cernava, T., Müller, H., Berg, C., and Berg, G. 2019. Seeds of native alpine
566 plants host unique microbial communities embedded in cross-kingdom networks.
567 Microbiome. 7:108.
568 Wassermann, B., Rybakova, D., Adam, E., Zachow, C., Bernhard, M., Müller, M., et al. 2021.
569 Studying seed microbiomes. In *The Plant Microbiome: Methods and Protocols, Methods*
570 *in Molecular Biology*, eds. L C Carvalhais and P G Dennis. , p. 1–21.
571 Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. 2017. Normalization
572 and microbial differential abundance strategies depend upon data characteristics.
573 Microbiome. 5.
574 Xing, H. Q., Ma, J. C., Xu, B. L., Zhang, S. W., Wang, J., Cao, L., et al. 2018. Mycobiota of maize

575 seeds revealed by rDNA-ITS sequence analysis of samples with varying storage times.

576 Microbiologyopen. 7.

577

578

579 **Figure Legends**

580 **Figure 1**. Rarefaction curves of bacteria/archaea (a) and fungi (b) from seed samples (marked)
581 at 97 % of clustering threshold were constructed by plotting the OTU number after
582 decontamination (microbial contaminants removal) to the sequence (read) number. The
583 rarefaction curves were constructed using vegan package (v2.5-4). Rank abundance curve of
584 decontaminated and normalized bacterial/archaeal (c) and fungal (d) OTU tables. Samples
585 (n=47 and n=45 for bacteria/archaea and fungi, respectively) were grouped by plant.

586

587 **Figure 2**. Bacterial/archaeal richness among plants were different (linear mixed-effects model,
588 LMM; df = 2, F-value = 6.91, p-value = 0.015) (a), but not among pods within plant (p-value >
589 0.05) (b). Specifically, plant B and C displayed lower bacterial/archaeal richness compared to
590 the plant A (Tukey's HSD post hoc test; p-value = 0.001 and 0.006, respectively).
591 Bacterial/archaeal phylogenetic diversity among plants were different (linear mixed-effects
592 model, LMM; df = 2, F-value = 6.56, p-value = 0.003) (c), but not among pods within plant (p-
593 value > 0.05) (d). Specifically, plant B and C displayed lower bacterial/archaeal phylogenetic
594 diversity compared to the plant A (Tukey's HSD post hoc test, p-value = 0.001 and 0.013,
595 respectively). Fungal richness was not different among plants (linear mixed-effects model,
596 LMM; df = 2, F-value = 1.11, p-value = 0.37) (e) and among pods within plant (p-value = 0.55) (f).

597

598 **Figure 3**. Principal coordinate analysis (PCoA) plot based on Jaccard dissimilarities of
599 bacterial/archaeal (a) and fungal (b) OTUs. The samples were plotted and grouped based on
600 plant as illustrated different colors. Each point was labelled by pod. Statistical analysis showed

601 that seed bacterial/archaeal community structure differ among plants (nested PERMANOVA, df
602 =2, F-ratio = 2.94, p-value = 0.002) but not pods (nested PERMANOVA, df =9, F-ratio = 0.99, p-
603 value = 0.63). Statistical analysis also showed that seed fungal community structure differs
604 among plants (nested PERMANOVA, df =2, F-rati0 = 1.55, p-value = 0.023) and pods (nested
605 PERMANOVA, df =9, F-rati0 = 1.27, p-value = 0.03). Distance to centroid analysis using
606 betadisper function from the vegan package revealed that there is variation in
607 bacterial/archaeal Beta diversity among plant (PERMADISP, df = 2, F-value = 63.9, p-value =
608 9.6e-14) (c). In contrast, there were no variation in fungal Beta diversity among plant
609 (PERMADISP, df = 2, F-value = 0.22, p-value = 0.802) (d).
610
611 **Figure 4**. Bar plot represents mean relative abundances of bacterial/archaeal (a) and fungal (b)
612 classes detected across plants. For bacteria/archaea, each pod consisted of 4 seeds (except for
613 C5; 3 seeds); and for fungi, each pod consisted of 4 seeds (except for A1, B1 and C5; 3 seeds).
614 The endophyte microbiome was assessed from the DNA extracted from single seed collected
615 from each pod. Bacterial/archaeal and fungal classes with mean relative abundances of less
616 than 10 % were grouped into the 'Other' classification, which includes many lineages (not
617 monophyletic).
618
619 **Figure 5**. Analysis of power using pwr.t.test() function from the pwr package revealed that an
620 effect of treatment on the 16S rRNA bacterial/archaeal alpha diversity (richness (a) and
621 phylogenetic diversity (b)) would be detectable 12 plants at a power of 0.8. Because the
622 highest seed microbiome variability was at the maternal plant level, individual seed microbiome

623 sequence profiles were pooled in silico by plant to perform this power analysis at the individual
624 plant level.

625

626 **Figure S1**. The proportion of plant reads of the total bacterial/archaeal (a) and fungal (b) reads
627 showed that more than 90 % reads obtained were plant contaminants.

628

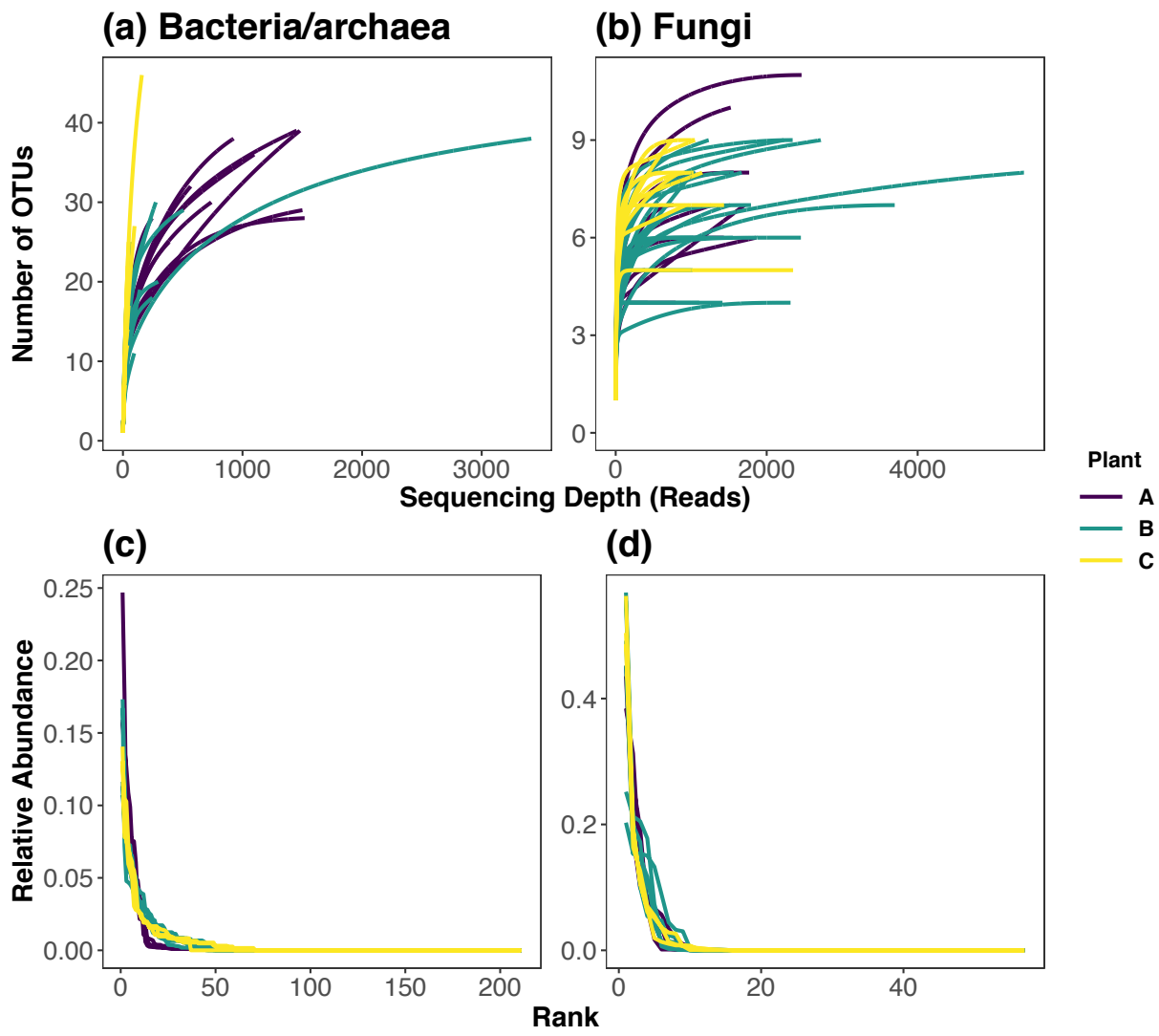

Fig 1. Rarefaction curves of bacteria/archaea (a) and fungi (b) from seed samples (marked) at 97 % of clustering threshold were constructed by plotting the OTU number after decontamination (microbial contaminants removal) to the sequence (read) number. The rarefaction curves were constructed using vegan package (v2.5-4)**.** Rank abundance curve of decontaminated and normalized bacterial/archaeal (c) and fungal (d) OTU tables. Samples (n=47 and n=45 for bacteria/archaea and fungi, respectively) were grouped by plant.

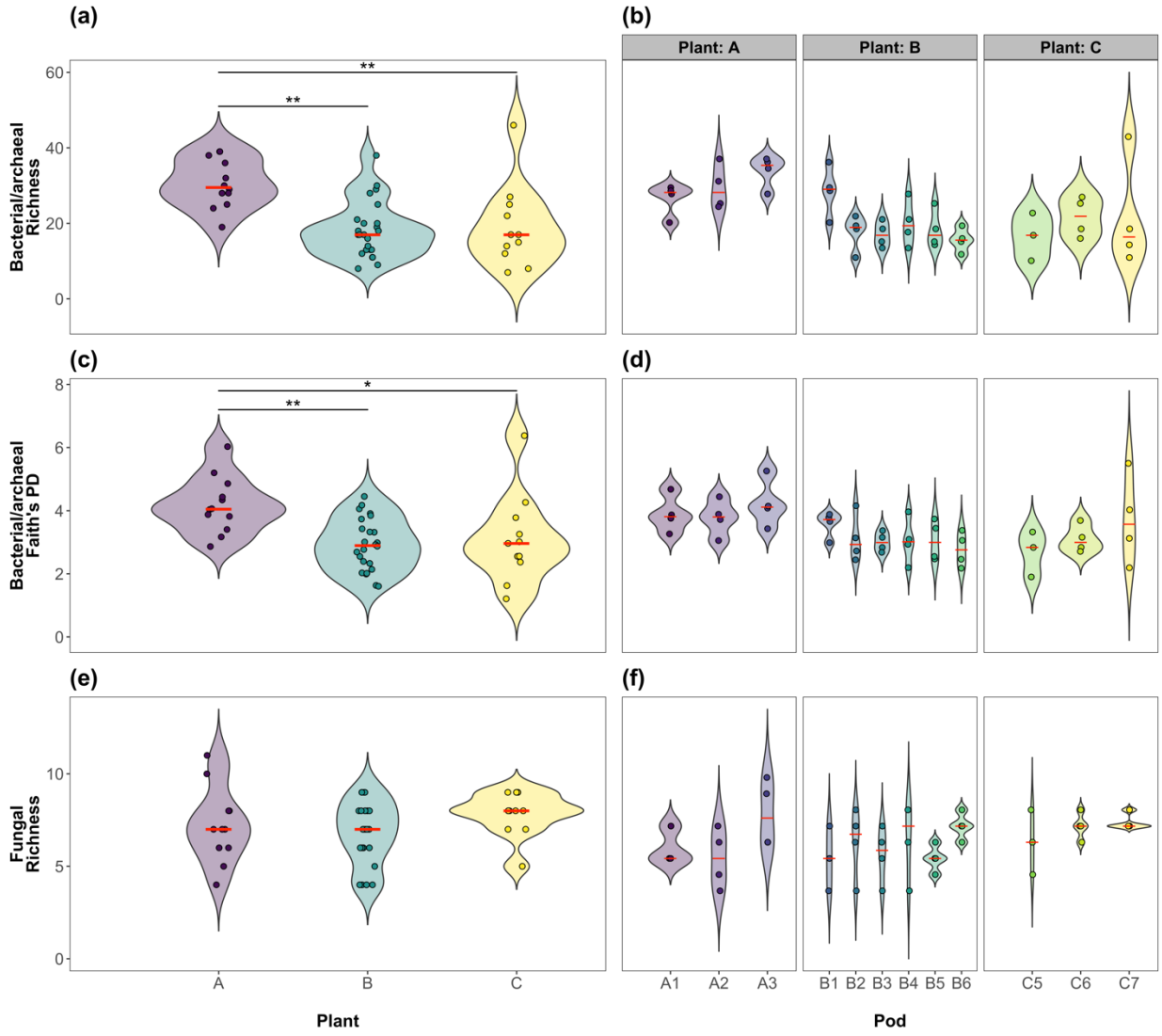

Fig 2. Bacterial/archaeal richness among plants were different (linear mixed-effects model, LMM; df = 2, F-value = 6.91, p-value = 0.015) (a), but not among pods within plant (p-value > 0.05) (b). Specifically, plant B and C displayed lower bacterial/archaeal richness compared to the plant A (Tukey's HSD post hoc test; p-value = 0.001 and 0.006, respectively). Bacterial/archaeal phylogenetic diversity among plants were different (linear mixed-effects model, LMM; df = 2, F-value = 6.56, p-value = 0.003) (c), but not among pods within plant (p-value > 0.05) (d). Specifically, plant B and C displayed lower bacterial/archaeal phylogenetic diversity compared to the plant A (Tukey's HSD post hoc test, p-value = 0.001 and 0.013, respectively). Fungal richness was not different among plants (linear mixed-effects model, LMM; df = 2, F-value = 1.11, p-value = 0.37) (e) and among pods within plant (p-value = 0.55) (f).

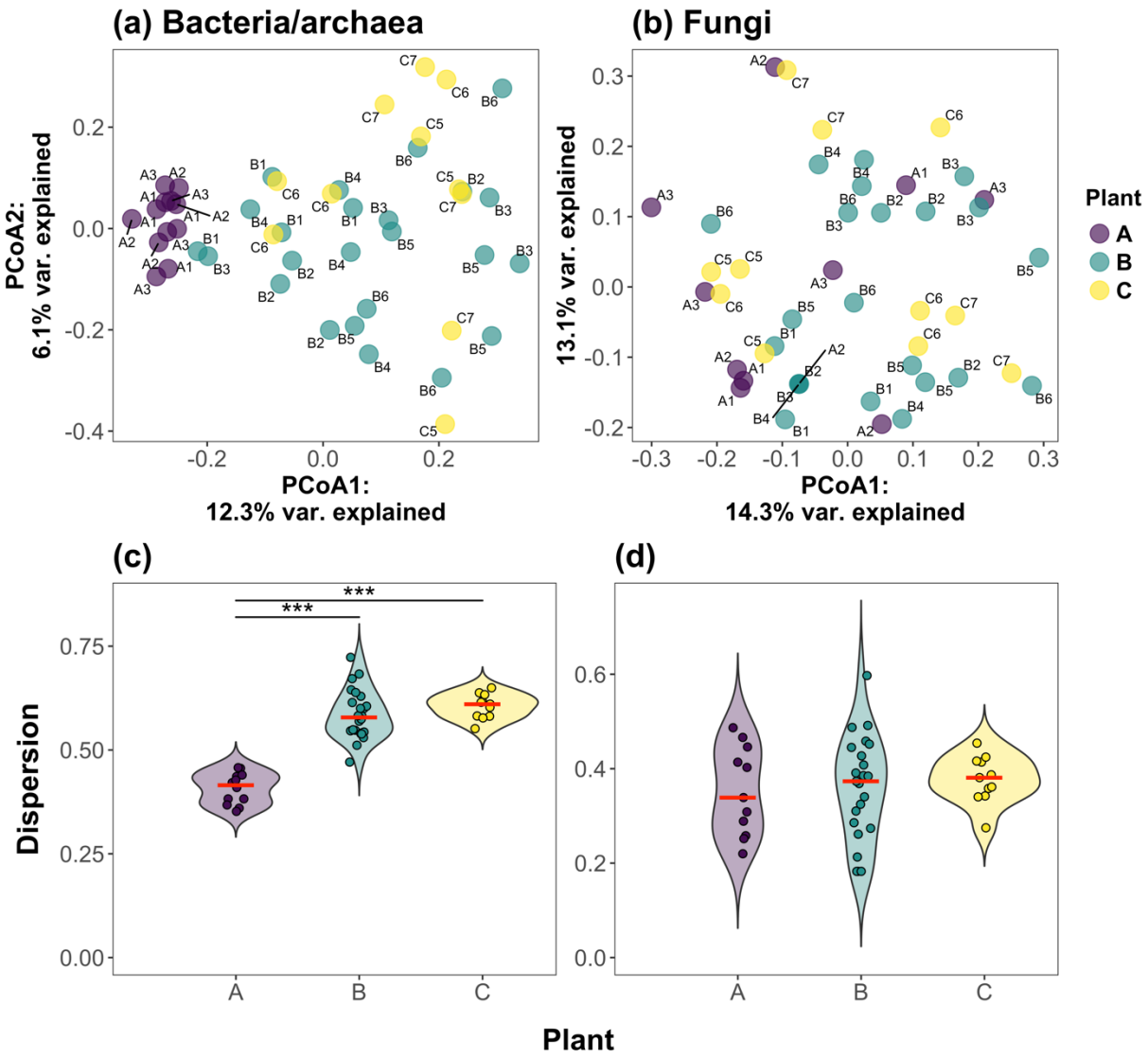

Fig 3. Principal coordinate analysis (PCoA) plot based on Jaccard dissimilarities of bacterial/archaeal (a) and fungal (b) OTUs. The samples were plotted and grouped based on plant as illustrated different colors. Each point was labelled by pod. Statistical analysis showed that seed bacterial/archaeal community structure differ among plants (nested PERMANOVA, df =2, F-ratio = 2.94, p-value = 0.002) but not pods (nested PERMANOVA, df =9, F-ratio = 0.99, p-value = 0.63). Statistical analysis also showed that seed fungal community structure differs among plants (nested PERMANOVA, df =2, F-rati0 = 1.55, p-value = 0.023) and pods (nested PERMANOVA, df =9, F-rati0 = 1.27, p-value = 0.03). Distance to centroid analysis using betadisper function from the vegan package revealed that there is variation in bacterial/archaeal Beta diversity among plant (PERMADISP, df = 2, F-value = 63.9, p-value = 9.6e-14) (c). In contrast, there were no variation in fungal Beta diversity among plant (PERMADISP, df = 2, F-value = 0.22, p-value = 0.802) (d).

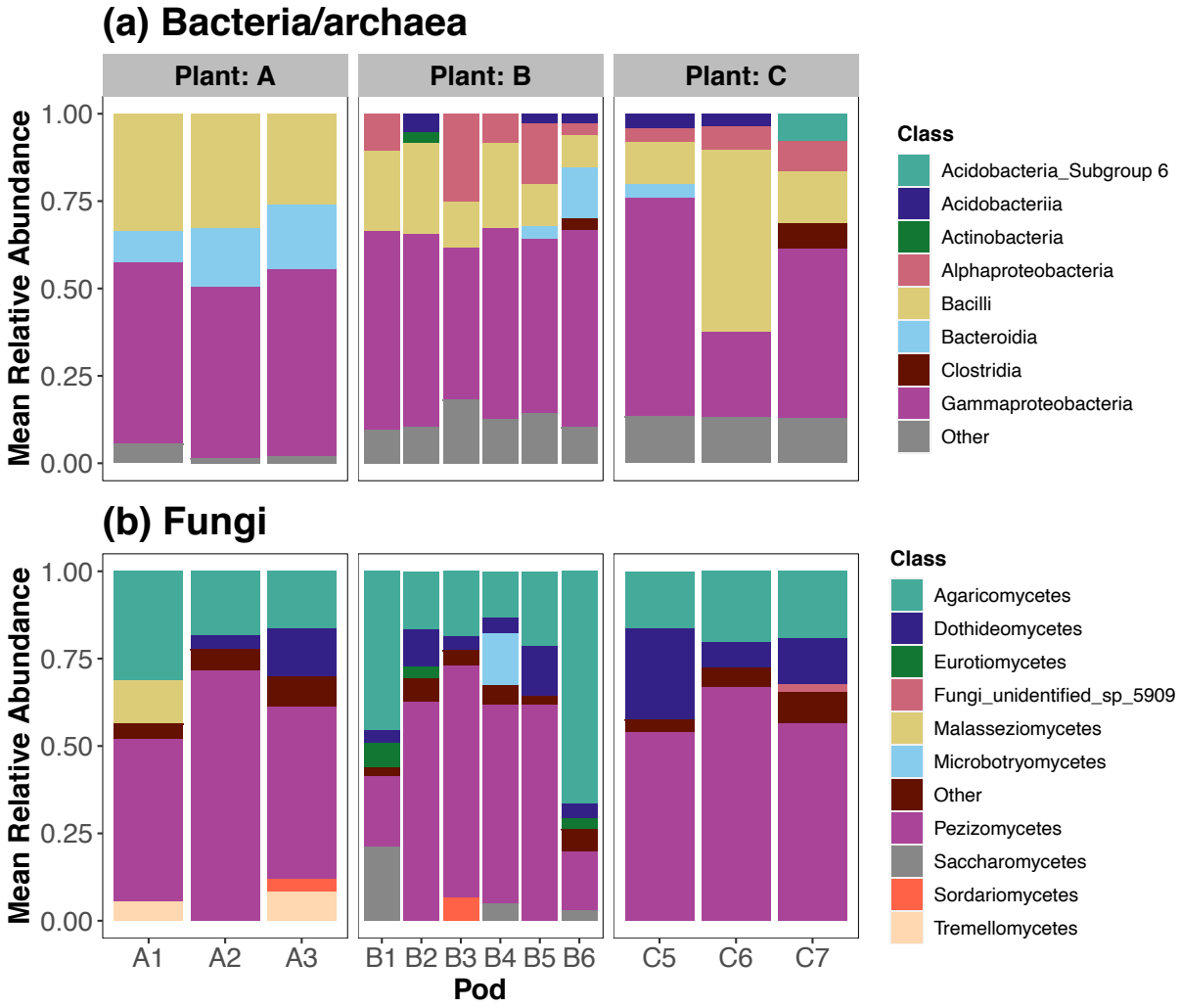

Fig 4. Bar plot represents mean relative abundances of bacterial/archaeal (a) and fungal (b) classes detected across plants. For bacteria/archaea, each pod consisted of 4 seeds (except for C5; 3 seeds); and for fungi, each pod consisted of 4 seeds (except for A1, B1 and C5; 3 seeds). The endophyte microbiome was assessed from the DNA extracted from single seed collected from each pod. Bacterial/archaeal and fungal classes with mean relative abundances of less than 10 % were grouped into the 'Other' classification, which includes many lineages (not monophyletic).

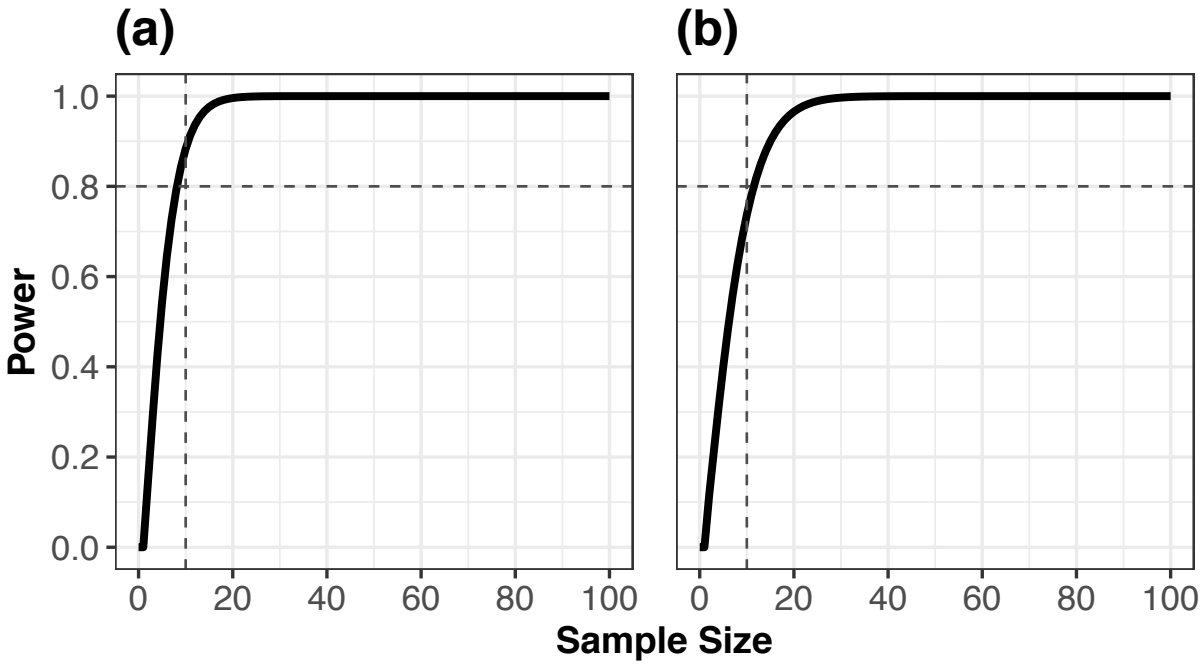

Fig 5. Analysis of power using *pwr.t.test* () function from the pwr package revealed that an effect of treatment on the 16S rRNA bacterial/archaeal alpha diversity (richness (a) and phylogenetic diversity (b)) would be detectable 12 plants at a power of 0.8. Because the highest seed microbiome variability was at the maternal plant level, individual seed microbiome sequence profiles were pooled *in silico* by plant to perform this power analysis at the individual plant level.