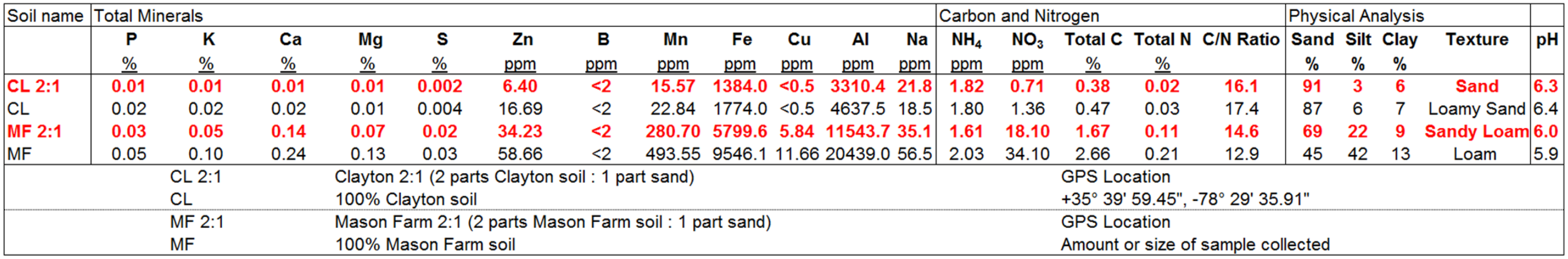
Supplementary Figures and Legends



Soil name	Total Minerals												Carbon and Nitrogen					Physical Analysis				
	P	K	Ca	Mg	S	Zn	B	Mn	Fe	Cu	Al	Na	NH_4	NO_3	Total C	Total N	C/N Ratio	Sand	Silt	Clay	Texture	pH
	%	%	%	%	%	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	%	%		%	%	%		
CL 2:1	**0.01**	**0.01**	**0.01**	**0.01**	**0.002**	**6.40**	**<2**	**15.57**	**1384.0**	**<0.5**	**3310.4**	**21.8**	**1.82**	**0.71**	**0.38**	**0.02**	**16.1**	**91**	**3**	**6**	**Sand**	**6.3**
CL	0.02	0.02	0.02	0.01	0.004	16.69	<2	22.84	1774.0	<0.5	4637.5	18.5	1.80	1.36	0.47	0.03	17.4	87	6	7	Loamy Sand	6.4
MF 2:1	**0.03**	**0.05**	**0.14**	**0.07**	**0.02**	**34.23**	**<2**	**280.70**	**5799.6**	**5.84**	**11543.7**	**35.1**	**1.61**	**18.10**	**1.67**	**0.11**	**14.6**	**69**	**22**	**9**	**Sandy Loam**	**6.0**
MF	0.05	0.10	0.24	0.13	0.03	58.66	<2	493.55	9546.1	11.66	20439.0	56.5	2.03	34.10	2.66	0.21	12.9	45	42	13	Loam	5.9

CL 2:1	Clayton 2:1 (2 parts Clayton soil : 1 part sand)	GPS Location
CL	100% Clayton soil	+35° 39' 59.45", -78° 29' 35.91"
MF 2:1	Mason Farm 2:1 (2 parts Mason Farm soil : 1 part sand)	GPS Location
MF	100% Mason Farm soil	Amount or size of sample collected

Supplementary Table ST1: Mason Farm and Clayton soil micronutrient analysis and GPS location.

a

Arabidopsis Genotypes and Seed Stocks				
Accession	**Region**	**Latitude**	**Longitude**	**Stock Center**
Col-0	USA (Germany?)	38.3	-92.3	CS22625
Ct-1	Italy	37.3	15	CS22639
Cvi-0	Cape Verde Isl.	16	-24	CS22614
Ler-1	Poland	52-53	15-16	CS22618
Mt-0	Libya	33	23	CS22642
Oy-0	Norway	60.23	6.13	CS22658
Shahdara	Tajikistan	38.35	68.48	CS22652
Tsu-0	Tsushima	34-35	136-137	CS28780

b

Table of Samples (after pooling smaller samples for normalization by rarefaction)

	Col-0		Ct-1		Cvi-0		Ler-1		Mt-0		Oy-0		Sha-0		Tsu-0		Soil	Dig date	Experiment
	R	EC	R	EC	R	EC	R	EC	R	EC	R	EC	R	EC	R	EC	S		Start
CL1 yng	11		11		9		10	1	10	4	10	4	10	3	8	5	10		
CL1 old	9	6	10	6	9	8	10	8	9	7	10	6	10	5	9	7	10	Dec-09	Feb-10
CL2 yng	5	7	10	4	5	1	3	4	4	2	7	2	7	3	7	1	8		
CL2 old	9	4	9	5	7	7	10	3	8	5	9	7	11	9	7	7	10	Dec-09	Apr-10
MF1 yng	8	3	8	4	8	4	8	8	7	6	8	2	9	8	10	6	10		
MF1 old	9	7	6	6	8	8	9	8	8	5	9	7	10	9	9	5	10	Feb-10	Mar-10
MF2 yng	10	10	7	7	10	10	10	10	8	8	10	10	10	10	8	8	10		
MF2 old	9	9	7	7	9	9	5	5	9	9	10	10	9	9	9	9	10	Apr-10	Apr-10
MF3 yng	6	6															11	Jun-10	Aug-10
MF4 yng	7	7															10	Nov-10	Jan-11

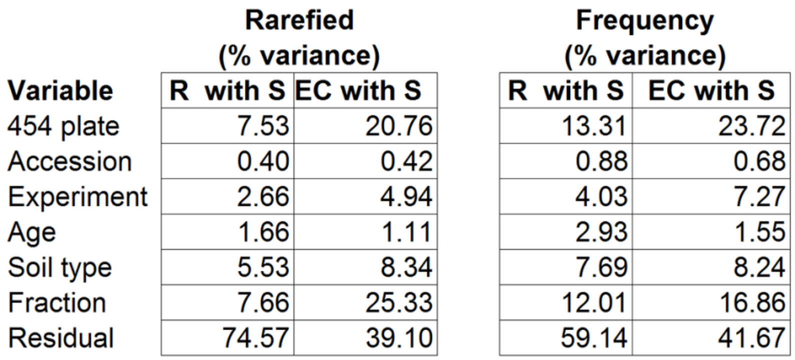
Table of Samples (Frequency)

	Col-0		Ct-1		Cvi-0		Ler-1		Mt-0		Oy-0		Sha-0		Tsu-0		Soil	Dig date	Experiment
	R	EC	R	EC	R	EC	R	EC	R	EC	R	EC	R	EC	R	EC	S		Start
CL1 yng	11		11		9		10	1	10	7	10	7	10	7	8	7	10		
CL1 old	9	8	10	8	9	9	10	8	9	7	10	7	10	8	9	8	10	Dec-09	Feb-10
CL2 yng	5	9	10	7	8	4	3	6	4	5	8	8	7	9	7	3	8		
CL2 old	9	9	9	8	7	8	10	3	8	7	9	9	11	12	7	9	10	Dec-09	Apr-10
MF1 yng	8	10	8	8	8	11	8	9	7	8	8	5	9	11	10	10	10		
MF1 old	9	8	6	8	8	10	9	8	8	5	9	8	10	9	10	5	10	Feb-10	Mar-10
MF2 yng	11	11	10	10	10	10	10	10	8	8	10	10	10	10	8	8	10		
MF2 old	9	9	7	7	9	9	5	5	9	9	10	10	9	9	9	9	10	Apr-10	Apr-10
MF3 yng	6	10															12	Jun-10	Aug-10
MF4 yng	8	8															10	Nov-10	Jan-11

Supplementary Table ST2: a) *Arabidopsis thaliana* genotypes and seed stocks used. **b)** Number of high quality samples for the frequency-normalized table (top) and the rarefaction normalized table (bottom), in which some replicate samples were pooled to make the rarefaction threshold. Does not include the four sterile seedling samples (Supplementary Figure S13).

Supplementary Table ST3: All 778 Measurable OTUs including GLMM predictions, taxonomic assignments, sequences, and location of notable OTUs within main figures. Provided as a separate Excel document with a full table legend on Sheet 1, the table based on rarefaction -normalized data on Sheet 2, and the table based on frequency-normalized data on Sheet 3.

(as separate Excel document available from Nature website)

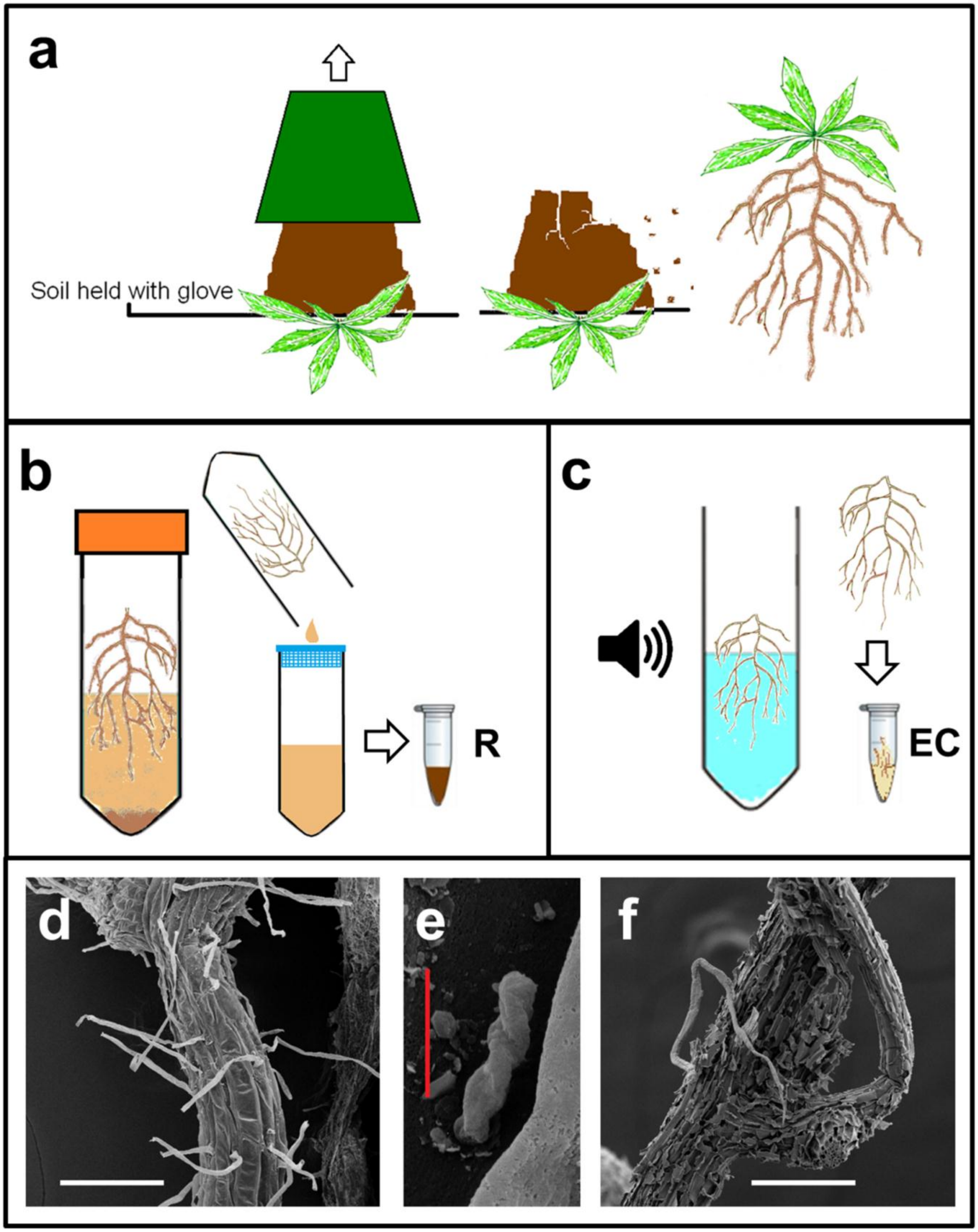


	Rarefied (% variance)		Frequency (% variance)	
Variable	**R with S**	**EC with S**	**R with S**	**EC with S**
454 plate	7.53	20.76	13.31	23.72
Accession	0.40	0.42	0.88	0.68
Experiment	2.66	4.94	4.03	7.27
Age	1.66	1.11	2.93	1.55
Soil type	5.53	8.34	7.69	8.24
Fraction	7.66	25.33	12.01	16.86
Residual	74.57	39.10	59.14	41.67

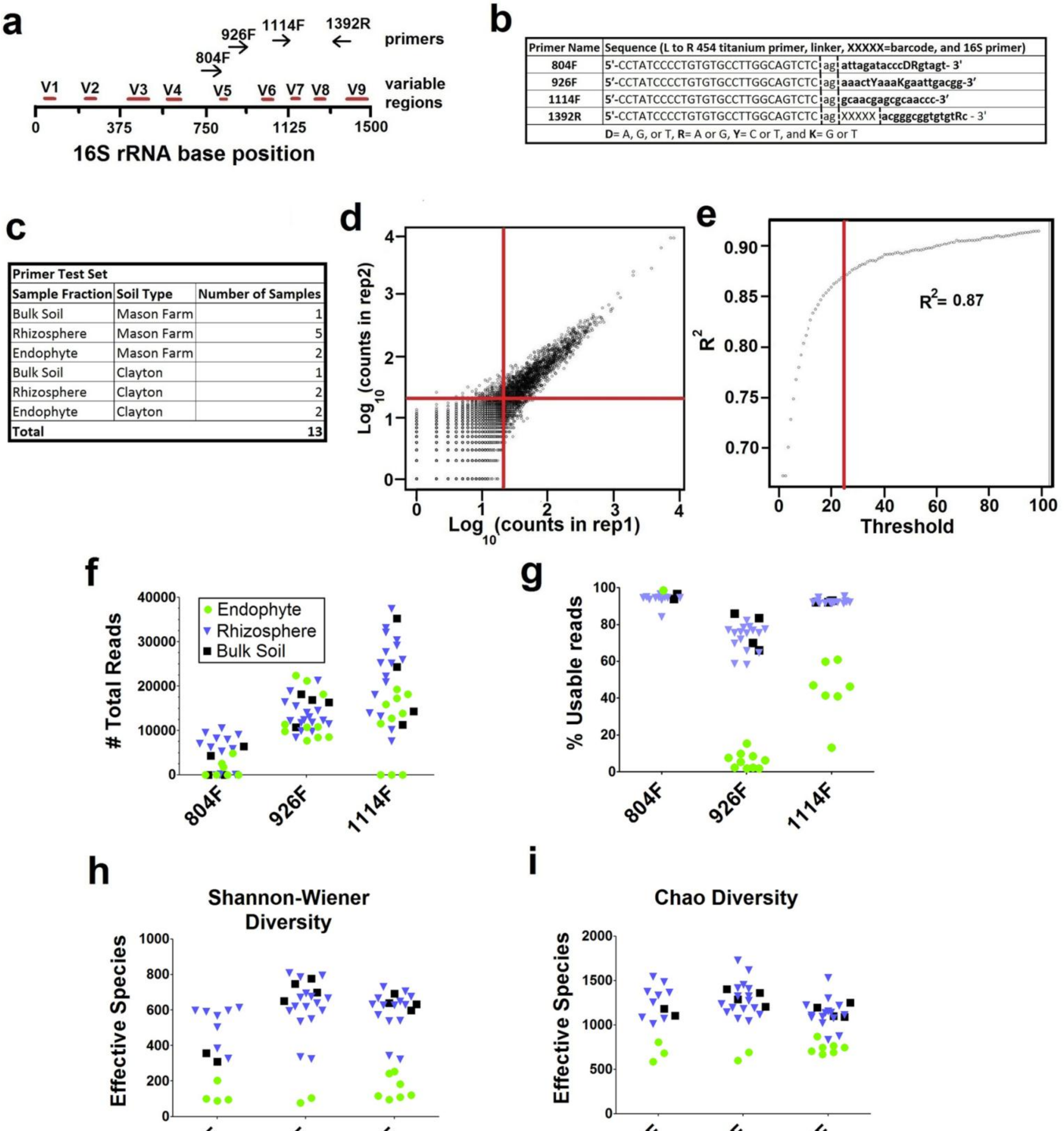
Supplementary Table ST4: Percent variance explained by each variable in the Full GLMM.

Supplementary Table ST5: ANOVA statistics comparing Shannon Diversity and taxonomic distributions across the S, R, and EC fractions. This table accompanies Figure 2, Supplementary Figure S7, and Supplementary Figure S15. It is provided as a separate Excel document with statistics on rarefaction-normalized data on Sheet 1, and statistics on frequency-normalized data on Sheet 2. Further explanation is given in the table.

(as separate Excel document available from Nature website)



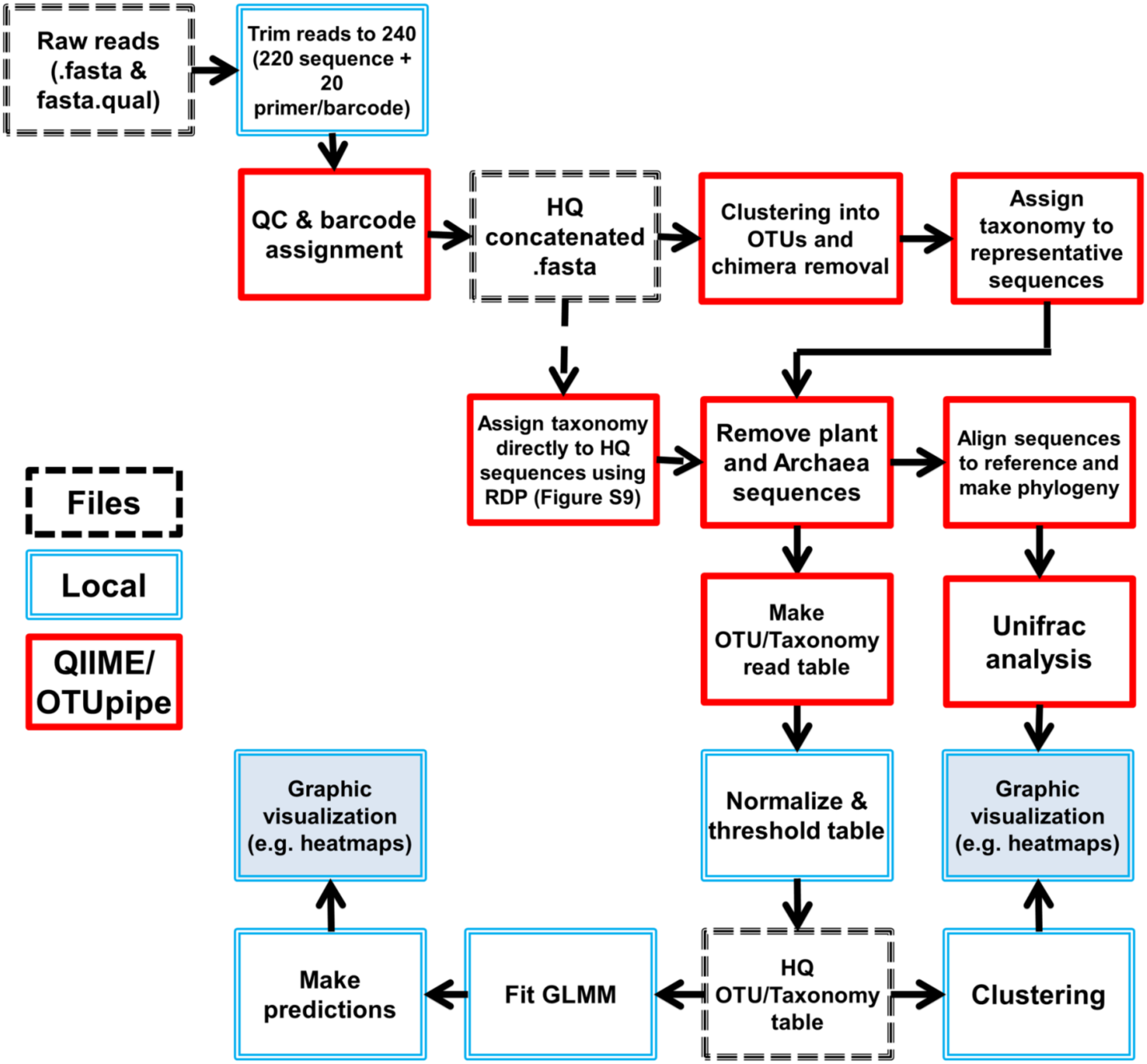
Supplementary Figure S1: Harvesting scheme. a) Using gloves and a flame-sterilized work surface, plants are overturned, pots are removed, and soil is crumbled/brushed away leaving ≤1 mm rhizosphere soil on roots. **b)** The above-ground parts are cut away and rhizosphere soil is harvested from roots by shaking them in sterile phosphate buffer with Silwet L-77; the rinse is pelleted and becomes the rhizosphere R fraction. **c)** Roots are placed in a new tube with sterile phosphate buffer and sonicated for five 30 second bursts at low intensity (see Supplementary Methods). The surface-cleaned roots are then snap frozen and lyophilized to become the EC fraction. **d)** SEM showing intact root surface after rhizosphere soil has been removed, but prior to sonication. Scale = 100 microns. **e)** SEM showing a root-surface bacterium on root shown in **d**. Scale = 1 micron. **f)** SEM showing the disruptive clearing of nearly the entire root surface after sonication. Scale = 100 microns.



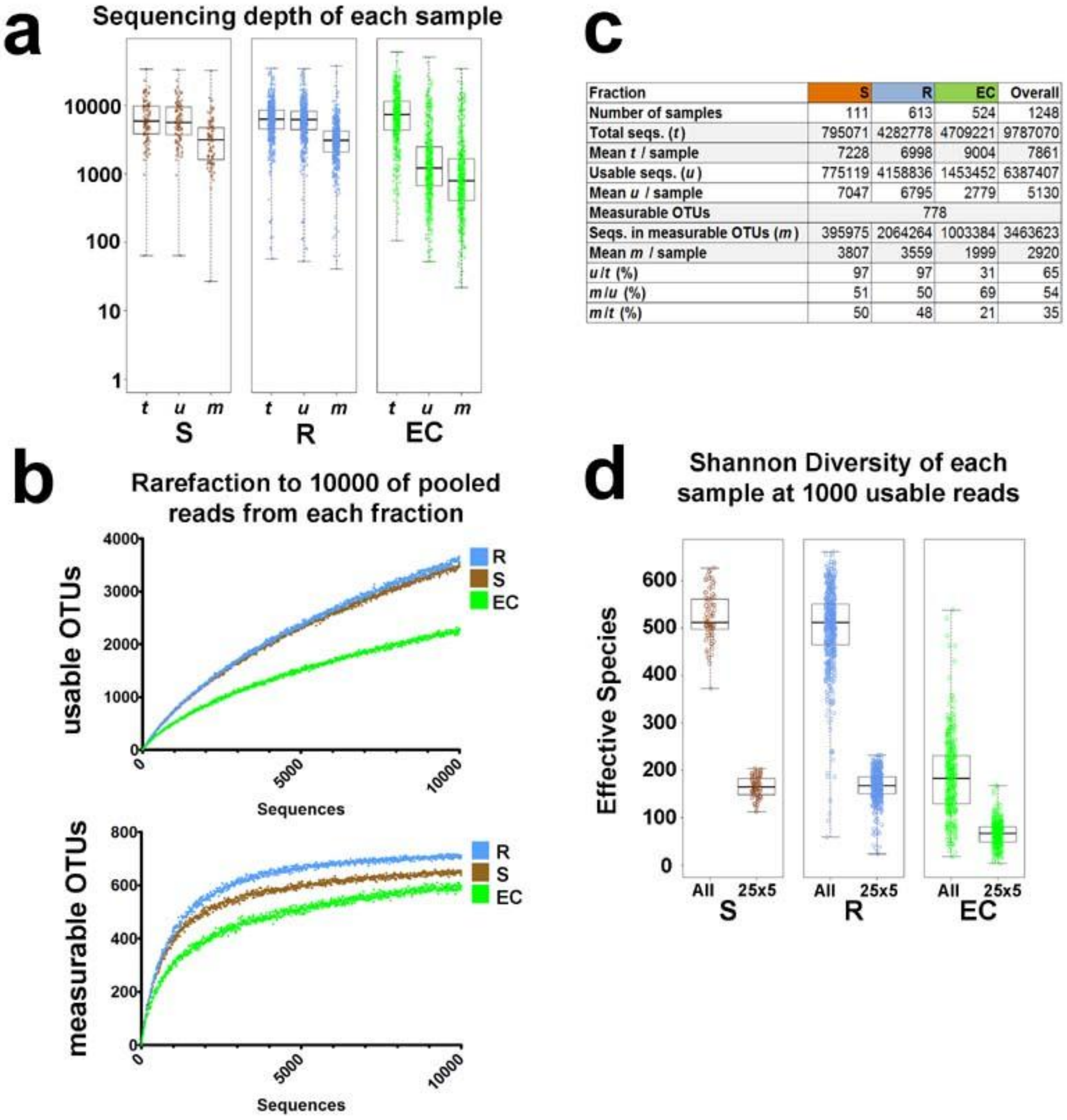
Primer Name	Sequence (L to R 454 titanium primer, linker, XXXXX=barcode, and 16S primer)
804F	5'-CCTATCCCCTGTGTGCCTTGGCAGTCTC ag attagatacccDRgtagt- 3'
926F	5'-CCTATCCCCTGTGTGCCTTGGCAGTCTC ag aaactYaaaKgaattgacgg-3'
1114F	5'-CCTATCCCCTGTGTGCCTTGGCAGTCTC ag gcaacgagcgcaaccc-3'
1392R	5'-CCTATCCCCTGTGTGCCTTGGCAGTCTC ag XXXXX acgggcggtgtgtRc - 3'
	D= A, G, or T, **R**= A or G, **Y**= C or T, and **K**= G or T

Primer Test Set		
Sample Fraction	Soil Type	Number of Samples
Bulk Soil	Mason Farm	1
Rhizosphere	Mason Farm	5
Endophyte	Mason Farm	2
Bulk Soil	Clayton	1
Rhizosphere	Clayton	2
Endophyte	Clayton	2
Total		13

Supplementary Figure S2: Primer test and technical reproducibility. a) Position on the 16S gene of each of the primers tested. **b)** Sequence of each primer used. **c)** Composition of the 13 samples tested. **d)** Log10 transformation of raw reads per OTU for one independent replicate (x-axis) vs. the other (y-axis), where both replicates were PCR-amplified and sequenced from the same sample (axes labels are transformed and cover a range of 0-10,000 reads). The intersection of the red lines shows where an OTU with 25 reads in both replicates would lie. **e)** Progressive drop-out analysis displaying the R^2 correlation of the data in **d** as OTUs with low read numbers are discarded. When only OTUs with ≥25 reads are considered (red line) the R^2 is acceptable at 0.87, a balance between reproducibility and data loss for low-abundance OTUs. In **f-i,** green circles are EC samples, blue triangles are R samples, and black squares are bulk soil samples. **f)** Total reads obtained from amplicons made with 804F, 926F, or 1114F paired with bar-coded 1392R. **g)** Percent of the 'usable' reads from **f** which are not identified as plant or chimeric OTUs. **h)** Shannon-Weiner species diversity of 1000 usable reads (for each sample with ≥1000 reads). **i)** Chao1 diversity of 1000 usable reads from each sample (for each sample with ≥1000).

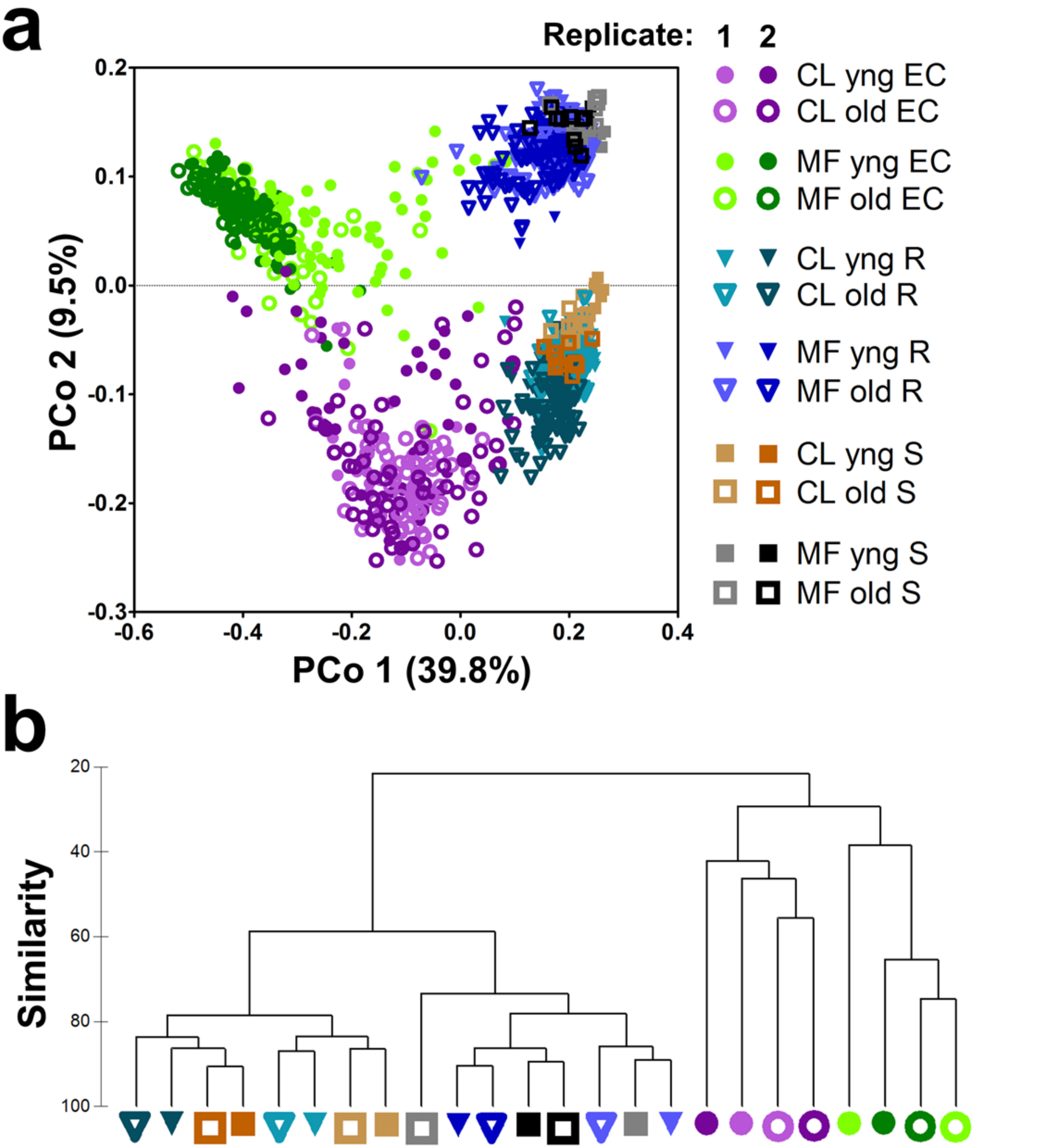

Raw reads (.fasta & fasta.qual)
Trim reads to 240 (220 sequence + 20 primer/barcode)
QC & barcode assignment
HQ concatenated .fasta
Clustering into OTUs and chimera removal
Assign taxonomy to representative sequences
Assign taxonomy directly to HQ sequences using RDP (Figure S9)
Remove plant and Archaea sequences
Align sequences to reference and make phylogeny
Files
Local
QIIME/ OTUpipe
Make OTU/Taxonomy read table
Unifrac analysis
Graphic visualization (e.g. heatmaps)
Normalize & threshold table
Graphic visualization (e.g. heatmaps)
Make predictions
Fit GLMM
HQ OTU/Taxonomy table
Clustering

Supplementary Figure S3: Informatics pipeline. Order of events. Broken-line black-line boxes represent files. Blue double-line boxes describe events that occur locally using custom scripts. Red boxes describe events that are implemented through QIIME/OTUpipe.

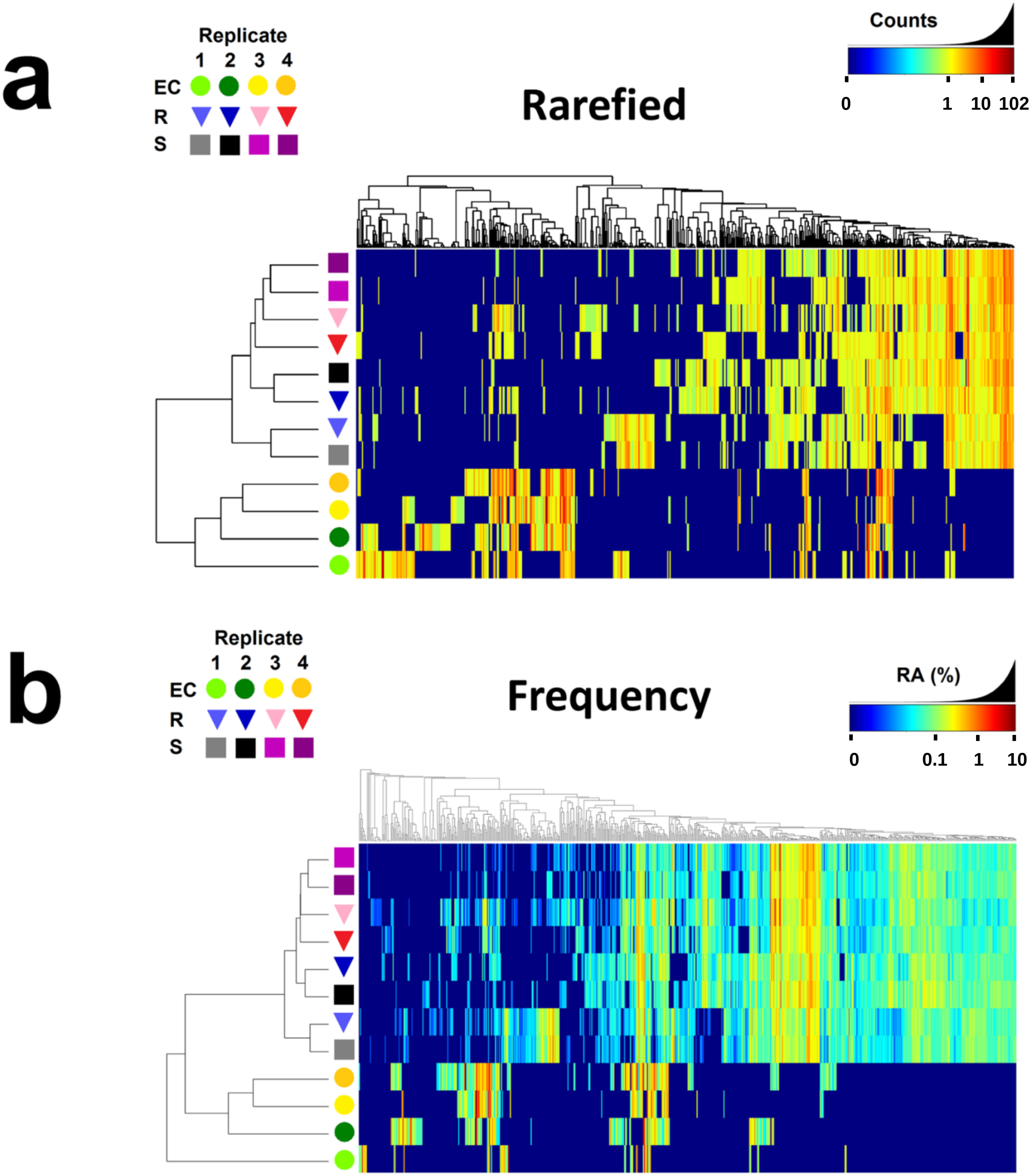


Fraction	S	R	EC	Overall
Number of samples	111	613	524	1248
Total seqs. (*t*)	795071	4282778	4709221	9787070
Mean *t* / sample	7228	6998	9004	7861
Usable seqs. (*u*)	775119	4158836	1453452	6387407
Mean *u* / sample	7047	6795	2779	5130
Measurable OTUs	778			
Seqs. in measurable OTUs (*m*)	395975	2064264	1003384	3463623
Mean *m* / sample	3807	3559	1999	2920
u/*t* (%)	97	97	31	65
m/*u* (%)	51	50	69	54
m/*t* (%)	50	48	21	35

Supplementary Figure S4: Sequencing statistics and quality. a) Sequencing depth per sample in reads for the three sample fractions S, R, and EC. Each dot represents a single plant or soil sample. Within each fraction, the total (*t*), usable *(u)*, and measurable *(m)* read counts are shown for all samples. The box plots contain the 1st and 3rd quartiles, split by the median; whiskers extend to include the farthest outliers. **b)** Rarefaction curves to 10,000 sequences for cumulative reads from S, R, and EC fractions considering all usable OTUs (top) and only measurable OTUs (bottom) **c)** Table, split by sample fraction, summarizing: cumulative numbers of total high quality reads, 'usable' (non-plant & non-chimera) reads, number of OTUs after the technical reproducibility '25x5' threshold is applied, 'measurable' reads (reads contained in OTUs that pass the 25x5 threshold). **d)** Shannon diversity of individual samples from each fraction, calculated from the rarefaction-normalized table, before (left) and after (right) applying the 25x5 measurable OTU threshold.



Supplementary Figure S5: Sample fraction and soil type drive the microbial composition of root-associated endophyte communities. a) Principal Coordinate Analysis (PCoA) of pairwise normalized weighted Unifrac distances between the samples considering relative abundance of all (*unthresholded*) OTUs. **b)** The median RAs for the *25x5 thresholded* 'measurable' OTUs from each of 24 soil/stage/fraction groups were $\log_2$ transformed (see methods) to make 24 representative samples (branch labels) and the pairwise Bray Curtis Similarity was used to hierarchically cluster these representatives (group average linkage).



Supplementary Figure S6: OTUs identified from four independent biological replicates are reproducible. Heat map displaying the reproducibility between four independent replicates at the yng developmental stage of bulk soil (squares), Col-0 R samples (triangles), and Col-0 EC samples (circles). Each symbol represents the median of six or more samples. All data were $\log_2$ transformed for visualization, but for ease of interpretation the quantities shown in the color key represent the original (untransformed) counts (in panel a) and frequencies (in panel b) for each color. Although all 778 measurable OTUs were included, some OTUs had a median of 0 in all Col-0 and soil groups shown and were removed from the display.

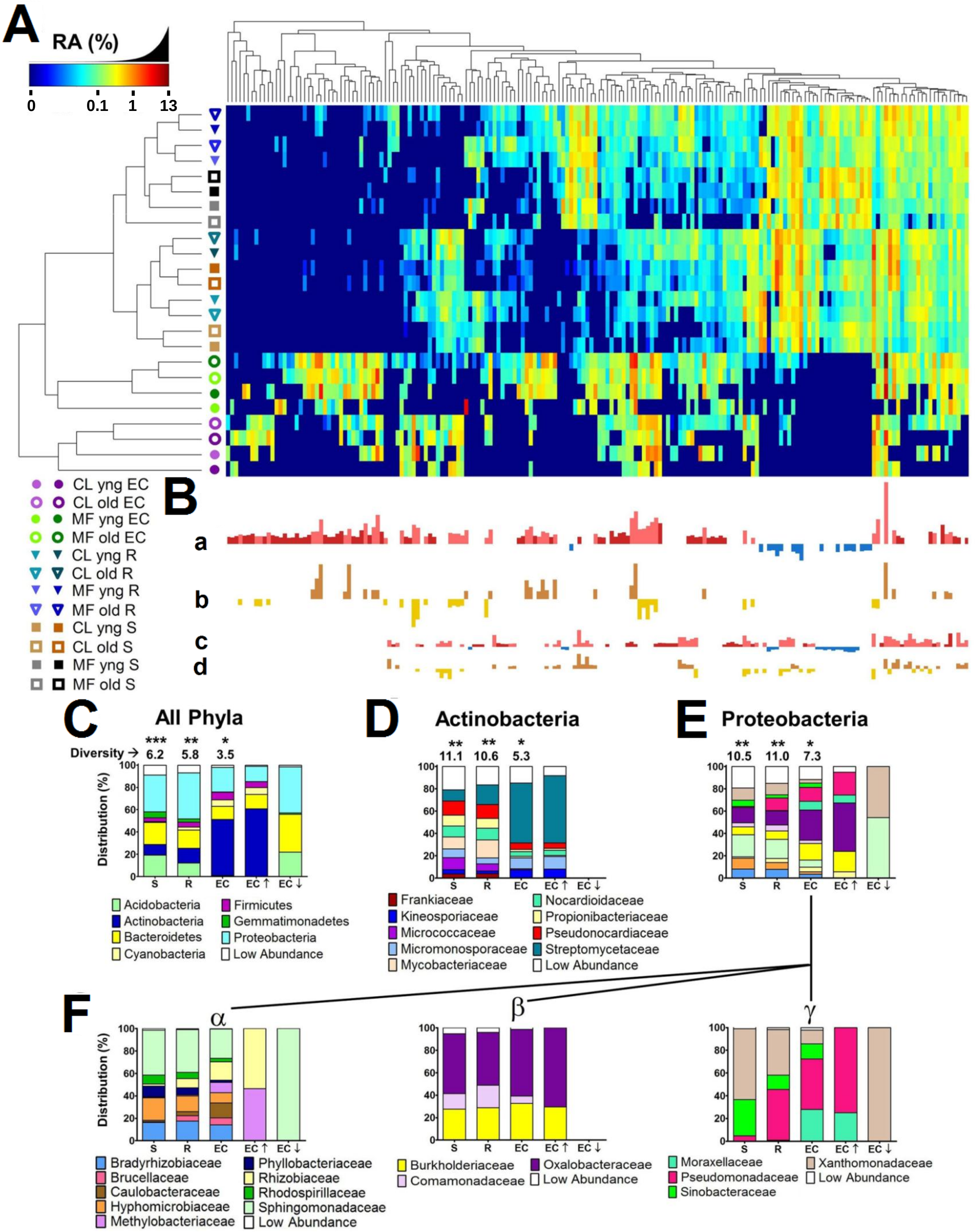
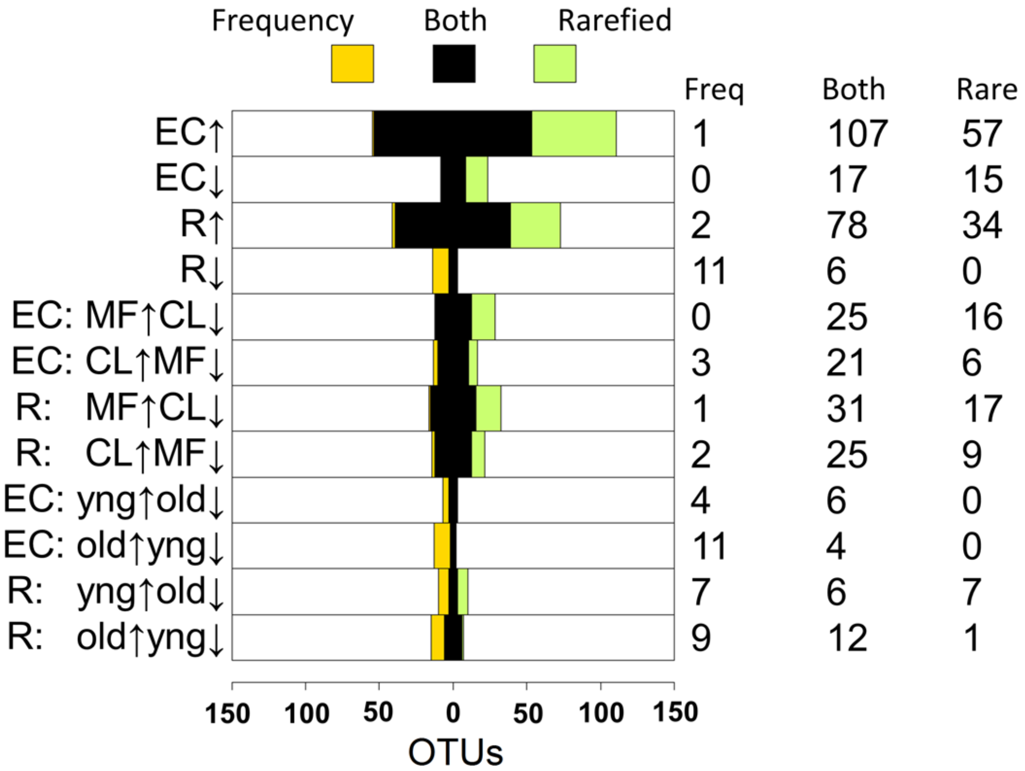
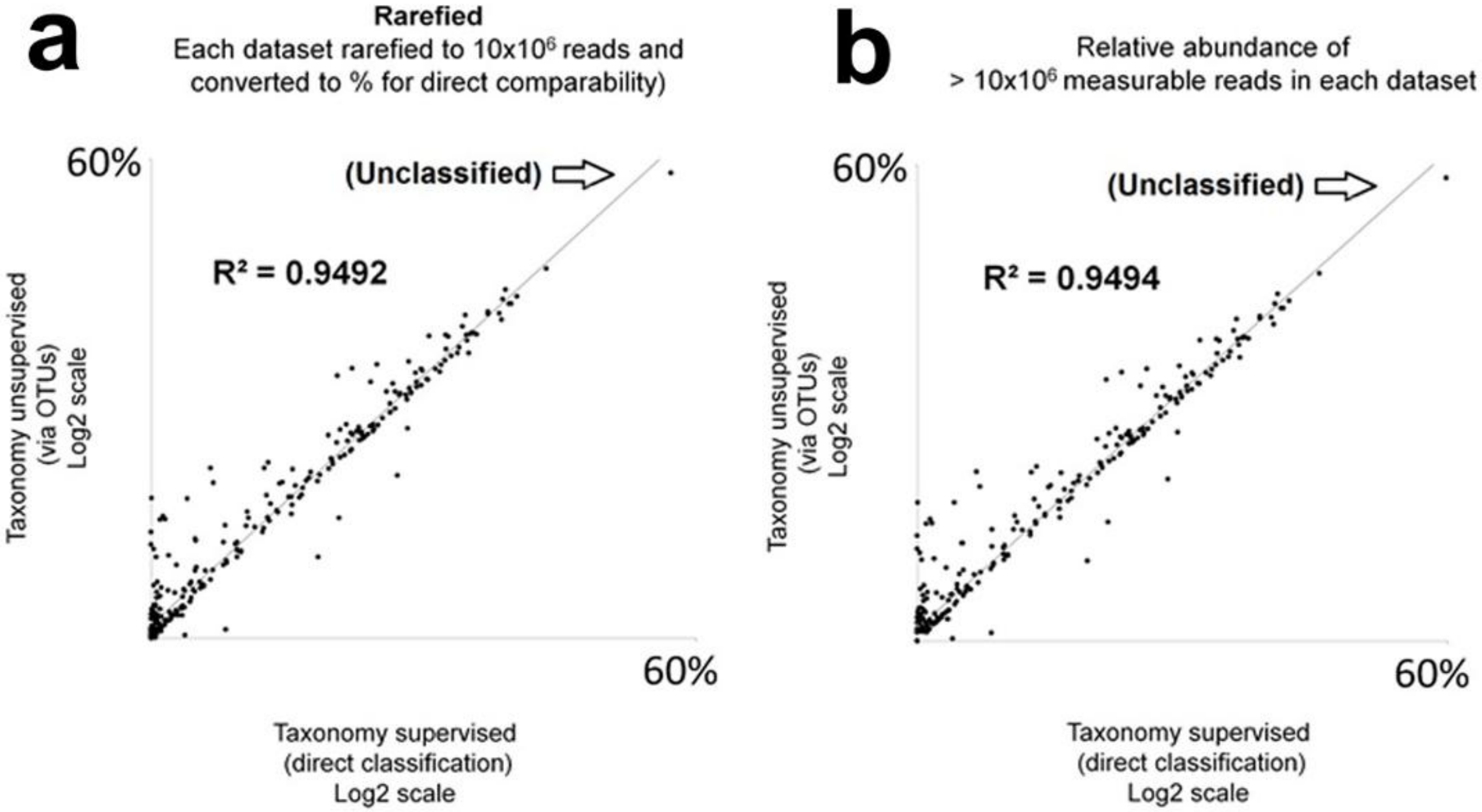

A
RA (%)
0
0.1
1
13
CL yng EC
CL old EC
MF yng EC
MF old EC
CL yng R
CL old R
MF yng R
MF old R
CL yng S
CL old S
MF yng S
MF old S
B
a
b
c
d
C
All Phyla
Diversity →
*** 6.2
** 5.8
* 3.5
Distribution (%)
100
80
60
40
20
0
S
R
EC
EC ↑
EC ↓
Acidobacteria
Actinobacteria
Bacteroidetes
Cyanobacteria
Firmicutes
Gemmatimonadetes
Proteobacteria
Low Abundance
D
Actinobacteria
** 11.1
** 10.6
* 5.3
Frankiaceae
Kineosporiaceae
Micrococcaceae
Micromonosporaceae
Mycobacteriaceae
Nocardioidaceae
Propionibacteriaceae
Pseudonocardiaceae
Streptomycetaceae
Low Abundance
E
Proteobacteria
** 10.5
** 11.0
* 7.3
F
α
β
γ
Bradyrhizobiaceae
Brucellaceae
Caulobacteraceae
Hyphomicrobiaceae
Methylobacteriaceae
Phyllobacteriaceae
Rhizobiaceae
Rhodospirillaceae
Sphingomonadaceae
Low Abundance
Burkholderiaceae
Comamonadaceae
Oxalobacteraceae
Low Abundance
Moraxellaceae
Pseudomonadaceae
Sinobacteraceae
Xanthomonadaceae
Low Abundance

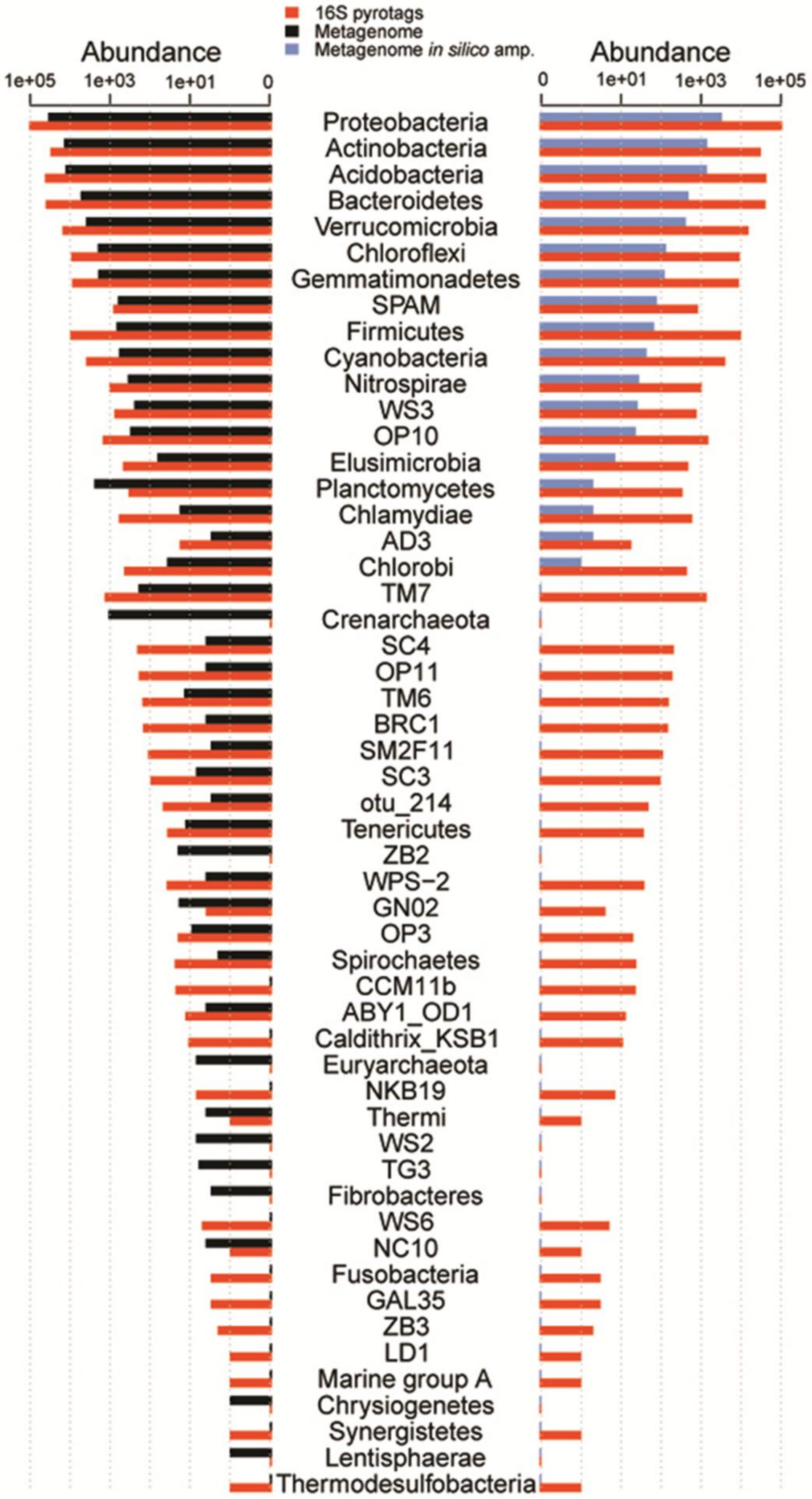
Figure S7: OTUs that differentiate the endophyte compartment and rhizosphere from soil. A, Heat map displaying the median RA (log_2 transformed) of each of 108 ‘R and EC-differentiating OTUs’ present across experimental replicates, where samples and OTUs are clustered on their Bray Curtis Similarity (group average linkage). The color key relates the colors to the untransformed RAs. **B,** The strength of the GLMM predictions (Best Linear Unbiased Predictors or BLUPs) is represented by the height of the bars. **a,** shows OTUs predicted as EC–enriched (red, up) or EC depleted (blue, down). **b,** shows OTUs found higher in the EC in MF soil than CL (brown, up) or higher in CL than MF (gold, down). OTUs in **a** that are not differentially affected by soil type as are shown in darker hues in **a**. **c,** OTUs predicted as R-enriched (as in **a** above). **d** OTUs higher in R in one soil type (as in **b**). **C)** Histogram displaying the distribution of the phyla present in the 778 measurable OTUs in soil (S), rhizosphere (R) and endophytic compartments (EC) compared to phyla present in the subset of EC OTUs enriched (EC-Up), or depleted (EC-Down) compared to soil. Shannon Diversity (considering phyla as individuals) is shown above. A differential number of asterisks above the Shannon Diversity values represents a significant difference ($p<0.05$, weighted ANOVA, Supplementary Methods, Supplementary Table ST5) **D)** Distribution of families present among the OTUs of the phylum Actinobacteria. **E)** Distribution of families present among the OTUs of the phylum Proteobacteria. **F)** Distribution of families present among the OTUs of three classes of the phylum Proteobacteria – Alpha (left), Beta (center), Gamma (right). Statistical evidence for presence, enrichment in, or depletion from EC is detailed in Supplementary Table S6. Data in **(D-F)** are from both soil types, pooled (see Supplementary Figure S15 for each soil separately).



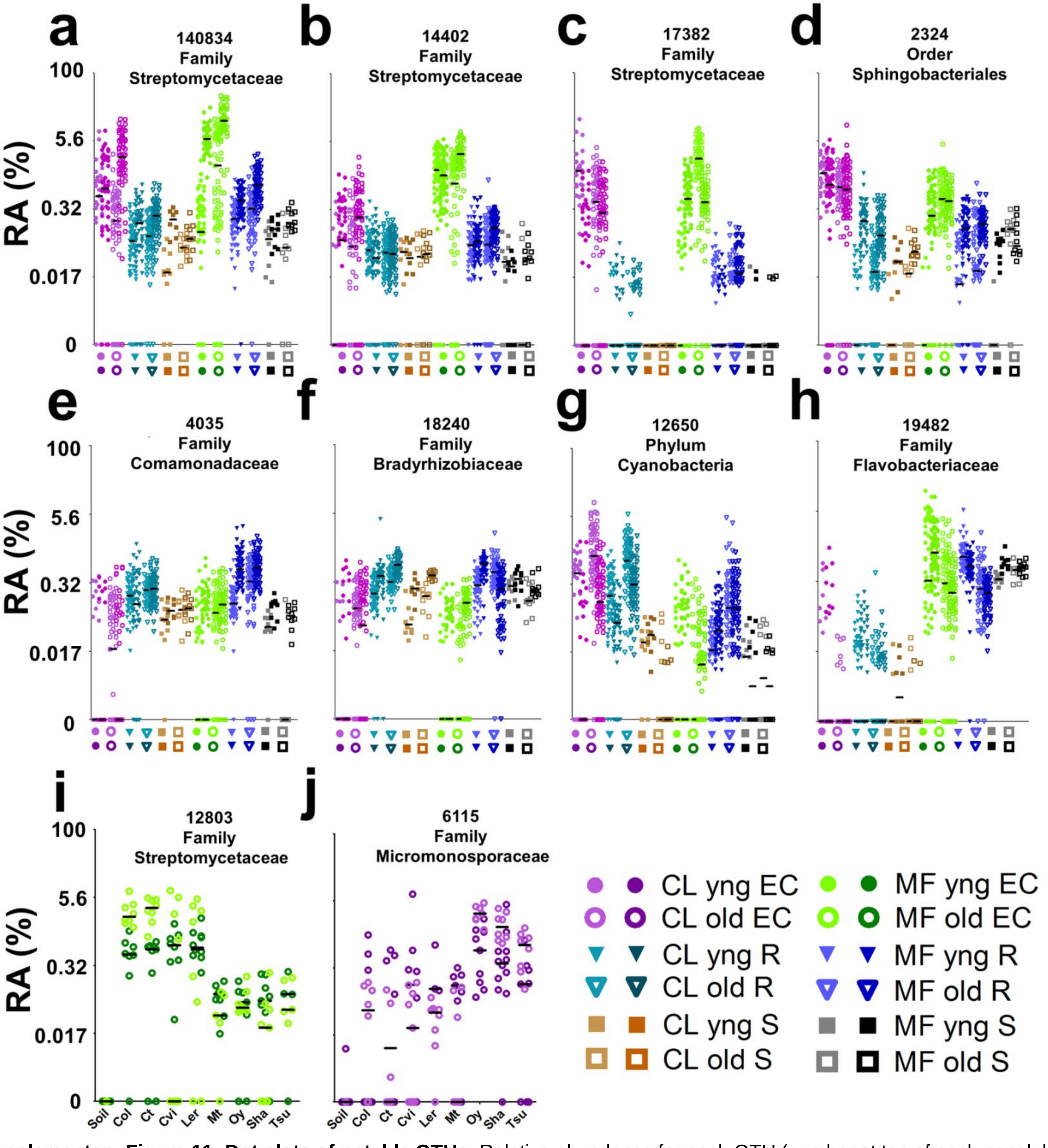
Supplementary Figure S8: Overlap of GLMM predictions between rarefaction-normalized and frequency-normalized OTU tables. The number of OTUs predicted by the full GLMM in each category that are unique to the frequency table is shown in orange. The number of OTUs predicted by the full GLMM in each category that are unique to the rarefied table are shown in green. The number of OTUs that were shared predictions in the two tables is shown in black.



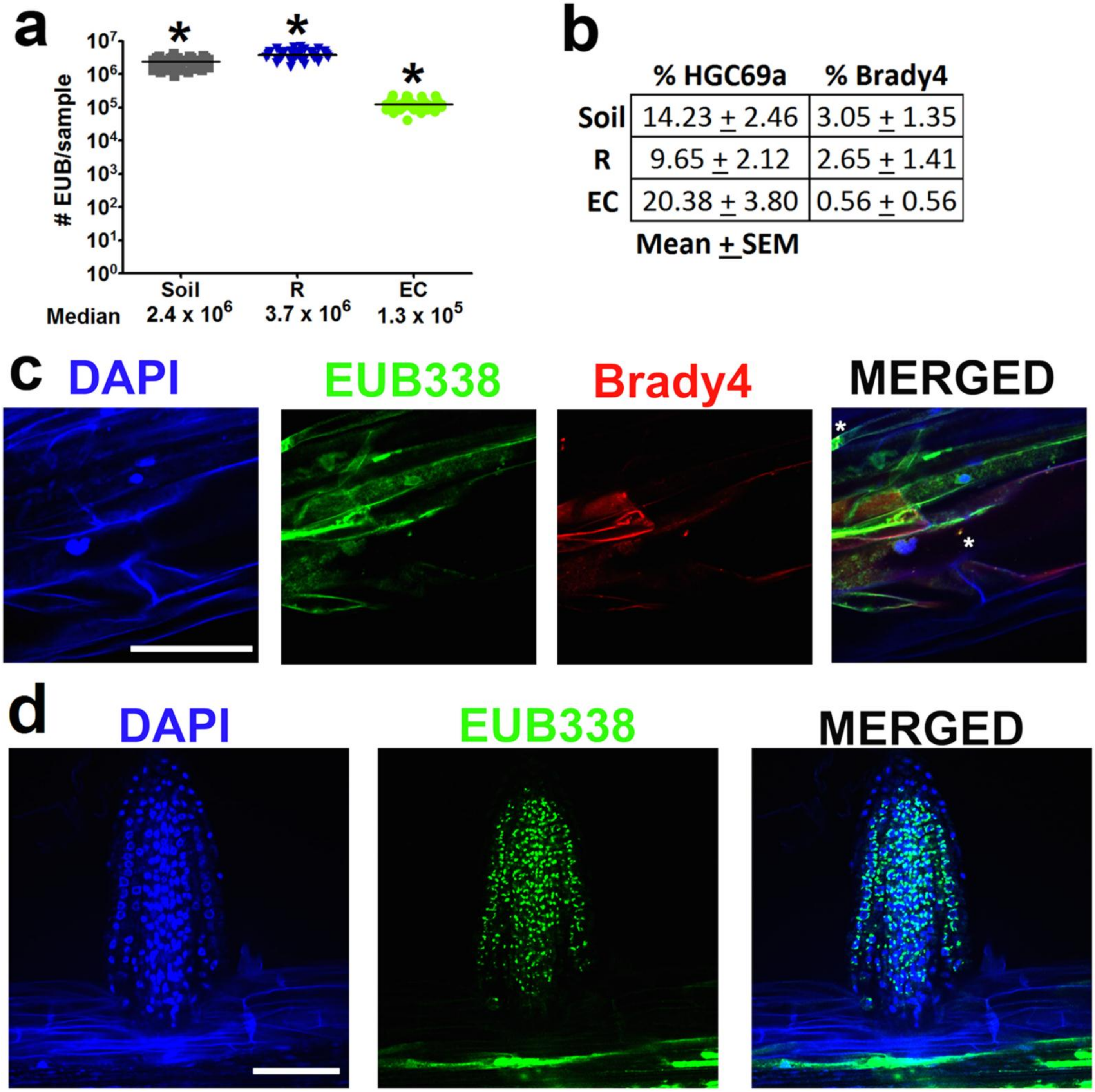
Supplementary Figure S9: 16S taxonomy classification at the family level is robust to method. For taxonomy-supervised classification, reads that passed default QIIME quality thresholds (but that were not clustered into OTUs) were trimmed to 220bp and were classified via RDP against Greengenes (Feb. 4 2011 version) training set to get family-level taxonomy. The abundance of each family was compared to the abundance of that family when the family assignments were assigned *after* the taxonomy-unsupervised grouping of reads into OTUs. In **a)** The total reads from non-chloroplast families from both taxonomy-supervised and taxonomy-unsupervised methods were rarefied to 10,000,000 reads, and the reads per family are shown as the $\log_2$ transformed relative abundance of the total reads, whereas **b)** shows the relative abundance of each family using all non-chloroplast reads, omitting the rarefaction step. The scatterplots thus show the high correlation at the family level for supervised and unsupervised taxonomy assignment. The dataset used for this figure included extra samples not described here, and was clustered as a single .fasta using the default QIIME implementation of Uclust [28].



Supplementary Figure S10: Test for PCR bias in pyrotagging. a) Relative abundance of 16S metagenomics and pyrotag reads. To assess possible bias introduced by amplification for pyrotagging, we compared the taxonomic distribution of a metagenome library created without amplification with a corresponding pyrotag dataset. Both datasets are from Col-0 Mason Farm young samples. 16S rDNA reads from this metagenome library (One HiSeq lane; more than 400 million 150 bp paired-end reads) were extracted by alignment against the 16S Silva database (release 106). Aligned reads were then assigned a taxonomy using an RDP training set built with the Greengenes reference database (version: May 9th 2011). This allowed classification of 57,663 16S reads from the metagenome sample using a bootstrap threshold >=0.50. There is an excellent overall correlation between the relative abundance of pyrotags and metagenome 16S rDNA reads across the major phyla represented in the datasets. Only two major classes, Thaumarchaeota and Planctomycea, were not amplified by the 1114F-1392R primers. Slightly higher abundance of Actinobacteria and Betaproteobacteria was observed in pyrotag data than in metagenome 16S reads. This was investigated further. **b)** For those classes in which underrepresentation in the pyrotag data are observed (red class names in Supplemental Figure S10a), we used *in silico* PCR analyses using the Greengenes database as template and our pyrotags primer pair, allowing a maximum of 2 mismatches, to investigate at which taxonomic level the under-representation would be discerned (Supplemental Figure S10b). We show that Thaumarchaeota (class) and Planctomycea (class) may be misrepresented in our pyrotag data. Since the Greengenes database contains many sequences amplified with the 1392R primer and therefore lacks this primer's sequence, we removed all sequences shorter than 6449 (in absolute position) in our reference database to minimize false negative rate (*i.e.* sequences not amplifying because they are not long enough to match the 1392R primer sequence).



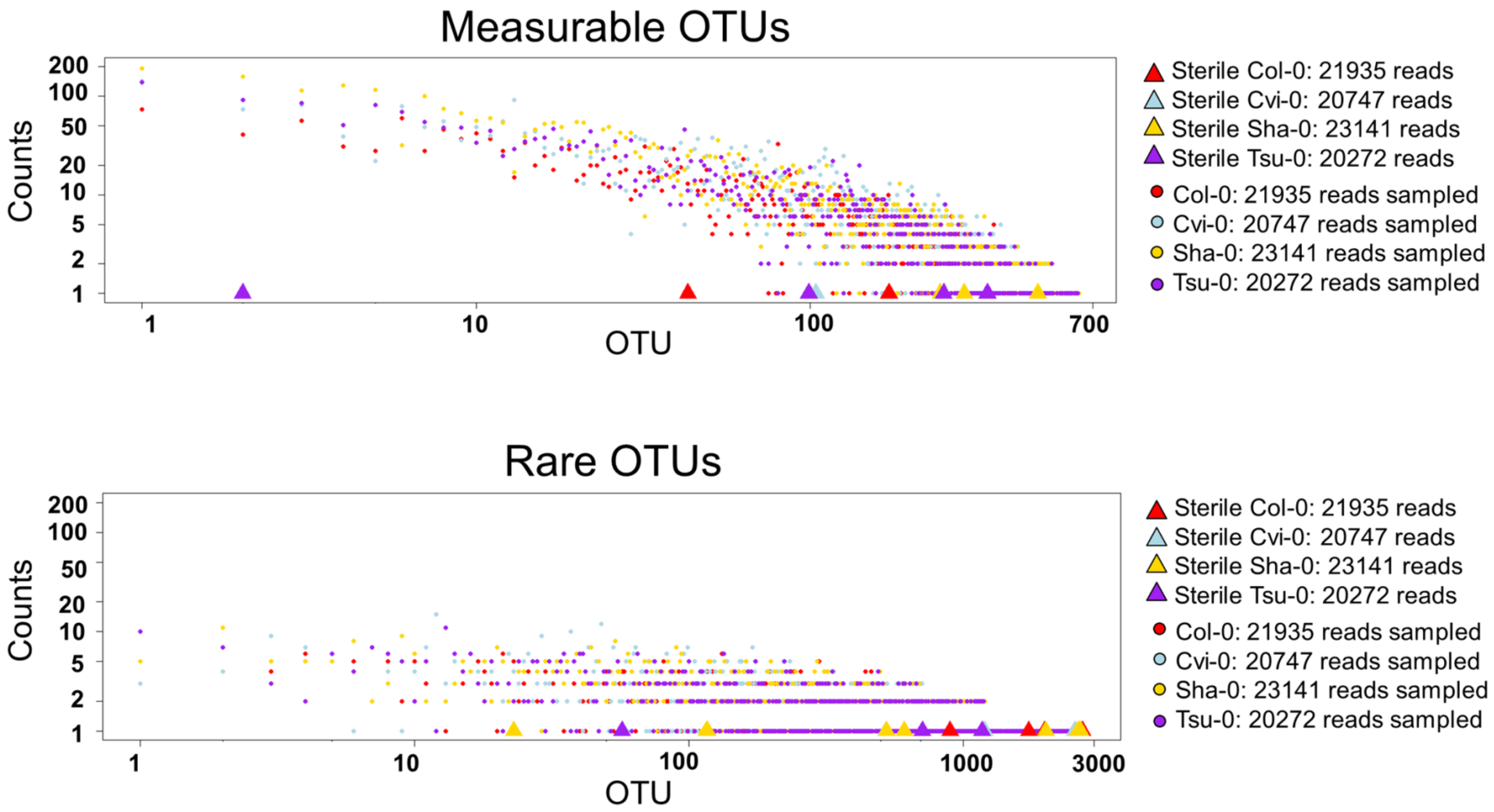
Supplementary Figure 11: Dot plots of notable OTUs. Relative abundance for each OTU (number at top of each panel; keyed to Supplementary Table ST3) from the frequency-normalized table was $\log_2$ transformed and the abundance for each sample (y-axis) plotted as an individual symbol. The y-axis is labeled with the actual (untransformed) relative abundance values. In **a-h**, each position on the x-axis is labeled with a symbol to represent the sample group (legend, lower right), and samples from that group are plotted column-wise directly above. Biological replicates are shown in the same column with different hues. The median of each biological replicate is shown with a horizontal black bar; some may not be visible because they are at 0. In **i** and **j**, sample color is according to the legend, and each position on the x-axis is labeled by Arabidopsis accession, with samples from that accession plotted above each label. Each OTU in the figure has model predictions in several categories (Supplemental table ST3).



	% HGC69a	% Brady4
Soil	14.23 + 2.46	3.05 + 1.35
R	9.65 + 2.12	2.65 + 1.41
EC	20.38 + 3.80	0.56 + 0.56

Mean + SEM

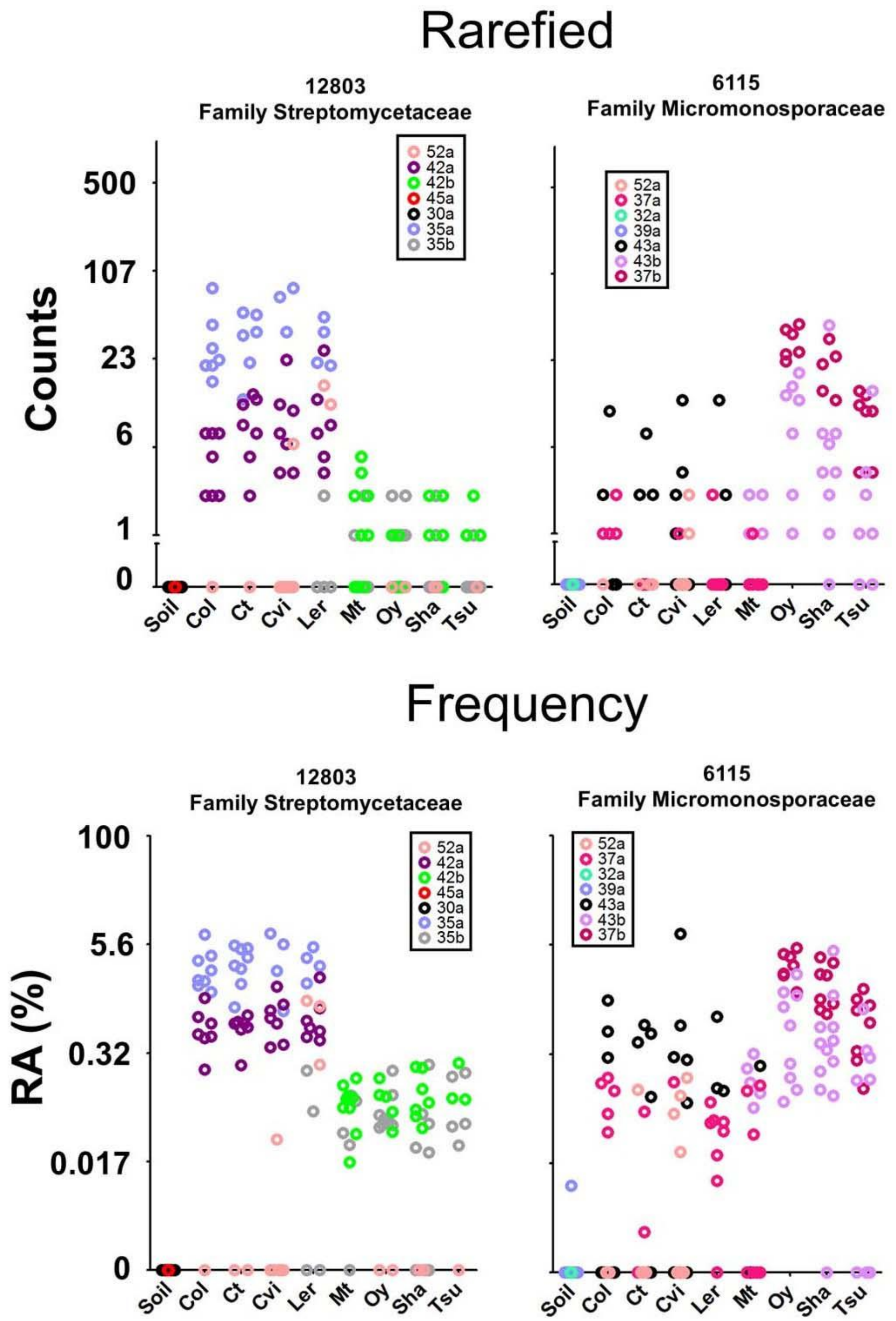
Supplementary Figure S12: Quantification of microbes in the three sample fractions using CARD-FISH. Four sets of Col-0 roots were pooled, processed, diluted, and put onto filters. **(a)** CARD-FISH using the EUB338, eubacterial probe, was applied and counterstained with DAPI. The number of EUB positive signals co-localizing with a DAPI signal was counted and the number of EUB positive signals per sample was calculated. This is an estimate for the number of bacteria present in each of our samples that DNA was extracted from with bulk soil (n=40), rhizosphere (n=39), and endophytic compartment (n=40). * indicates statistical significance at $p<1x10^{16}$ (ANOVA with post-hoc TukeyHSD) between each of the sample groups **(b)** Using double CARD-FISH on filters made from equal concentration of the 3 sample fractions, we determined the % of DAPI positive eubacteria that are also co-localize with either the HGC69a (Actinobacteria) or Brady4 (Bradyrhizobiaceae) probes on filters made from bulk soil (n=10), rhizosphere (n=10), and endophytic compartment (n=10) samples. Actinobacteria was in higher abundance in EC samples and Bradyrhizobiaceae was in lower abundance in EC samples compared to soil and R samples as expected from our pyrotag sequencing data. **(c)** Double CARD-FISH was applied using the EUB338, eubacterial probe (green) and the Brady4, *Bradyrhizobiaceae* probe (red), counterstained with DAPI (the asterisks indicate signals that are positive in all 3 channels). **(d)** Newly forming lateral roots and root tips were found commonly to be heavily colonized. Scale bars represent 50 microns.



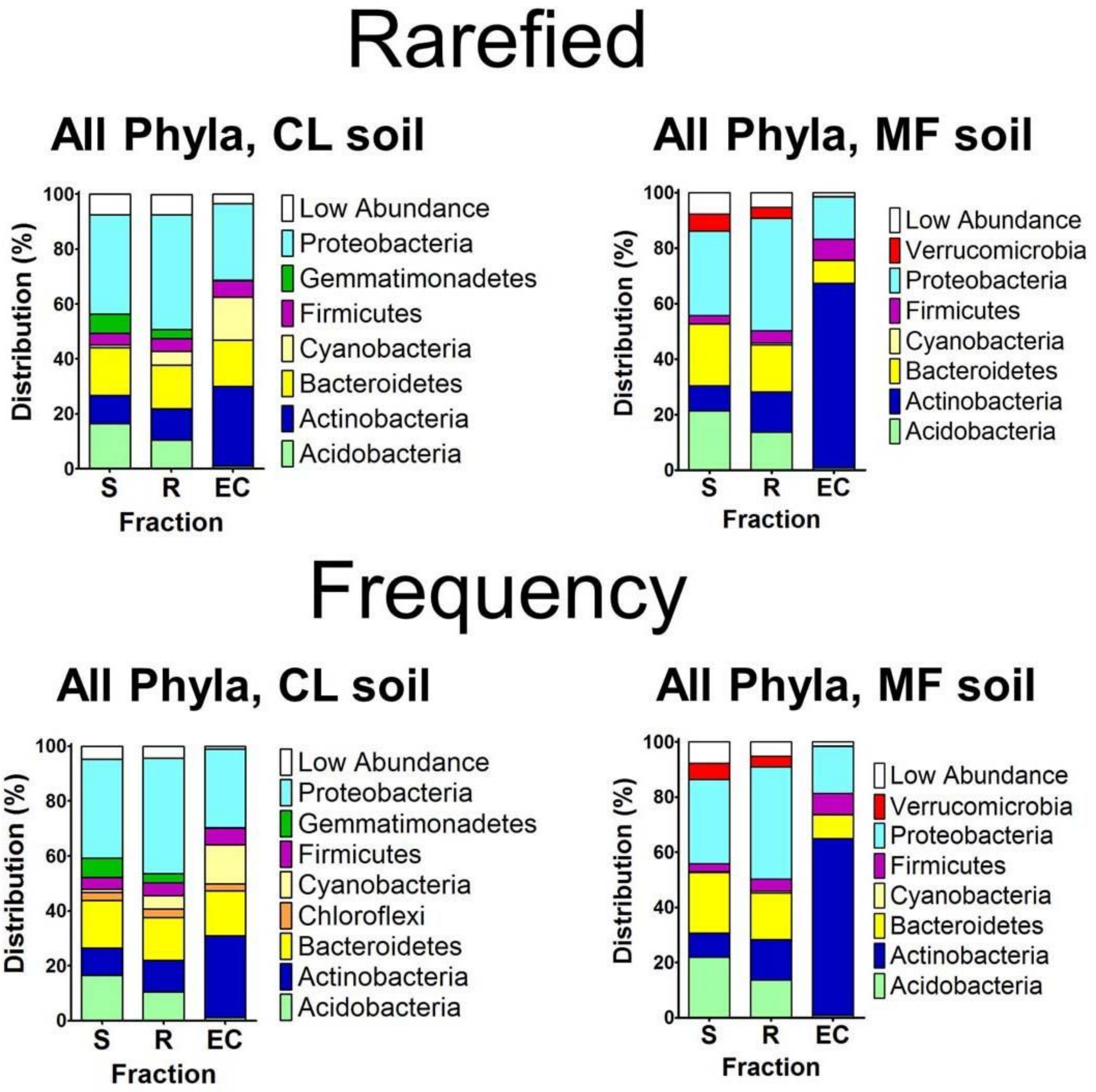
OTUs detected in >1 sterile sample

Rare OTU_9990: Unclassified (1 read each in sterile Cvi-0, Sha-0, and Tsu-0)
Rare OTU_17330: Phylum Cyanobacteria (1 read each in sterile Cvi-0, and Tsu-0)

Supplementary Figure S13: Pyrosequencing of sterile seedlings as compared to vs. non-sterile EC samples. DNA was extracted from homogenates from gnotobiotic seedlings of the genotypes Col-0, Cvi-0, Sha-0, and Tsu-0 (from which no culturable microbes were found), using bacteriolytic DNA preps, and these were pyrosequenced and clustered into OTUs as part of our full dataset. 21935, 20747, 23141, and 20272 high quality reads were obtained from each gnotobiotic genotype, respectively (triangles). The same total number of total reads was sampled from using pooled EC data from the full dataset for these accessions (circles). Each position on the X axis represents an OTU in the full dataset (measurable OTUs on top, rare OTUs on bottom) and the position on the Y axis represents the number of sequence reads found in that OTU. Both axes are shown in log scale. Of the 86095 HQ reads obtained from both sterile plants and non-sterile plants, the majority were from chloroplast OTUs (not shown). Far more non-plant reads were obtained from the non-sterile plants (19093 of 86095, or 22%) vs. sterile plants (34 of 86095, or 0.04%), a difference approaching three orders of magnitude. The 34 reads from non-sterile plants were members of 31 OTUs (triangles – some overlap on the log-scale axis). No OTU in a sterile plant sample was represented by more than one read, and only two OTUs were shared by more than one of the accessions - both of these shared OTUs were not in the measurable set, and had poor taxonomic classification. 11 of these 31 OTUs were not represented in the non-sterile samples. Furthermore, by including extra unused barcodes in our mapping files, or by sequencing sterile water in excess, we have been able to occasionally 'detect' single representatives of OTUs in our dataset, demonstrating that technical noise can cause singletons (data not shown). While we cannot rule out that unculturable microbes survive surface sterilization and exist at extremely low abundance, we have no evidence that such microbes exist in *A. thaliana* roots.



Supplementary Figure S14: Genotype-variable OTUs colored by sequence plate. Displays the data from **Fig. 3i** (MF old EC, left) and **Fig. 3j** (CL old EC right), colored by sequence plate (instead of biological replicate as in Figure 3) according to the legend within each plot. The top panel is based on rarefied data, as in Figure 3, and the bottom panel is based on the relative abundance, as in Supplementary Figure S11. (Note: ‘a’ and ‘b’ in our plate naming scheme do not represent different regions of the same plate. All 454 regions were modeled independently in the Full GLMM).



Supplementary Figure S15: Phyla in each sample fraction by soil type. Histogram displaying the distribution of the phyla present in the 778 measurable OTUs in soil (S), rhizosphere (R) and endophytic compartments (EC) with each soil type, MF and CL, considered independently. Rarefaction-normalized on top; frequency-normalized on bottom. Accompanying statistics on the distributions are in Supplementary Table ST5.