Supplementary Information for:

Successive passaging of a plant-associated microbiome reveals robust habitat and host genotype-dependent selection

Morella, Norma M.[1*], Weng, Francis Cheng-Hsuan[3], Joubert, Pierre M.[1], Metcalf, C. Jessica E. [4], Lindow, Steven[1*], Koskella, Britt[2*]

Author affiliations

1 Department of Plant and Microbial Biology, UC Berkeley, Berkeley, CA, USA

2 Department of Integrative Biology, UC Berkeley, Berkeley, CA, USA

3 Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

4 Department Ecology, Evolutionary Biology & Public Affairs, Princeton University, Princeton, NJ, USA

* To whom correspondences should be addressed

Email: morella@berkeley.edu; icelab@berkeley.edu; bkoskella@berkeley.edu

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1908600116

Extended Methods

Tomato accessions

Tomato accessions were obtained from the Tomato Genetics Resource Center. Five tomato genotypes were used: *Solanum lycopersicum* money maker disease susceptible (TGRC 2706); *S. lycopersicum* money maker disease resistant (TGRC 3472); *S. lycopersicum* Rio Grande disease susceptible control for TGRC 3342 (TGRC 3343); *S. lycopersicum* Rio Grande disease resistant (TGRC 3342); and *S. pimpinellifolium* wild ancestor (2934). During the 2016 growing season, seeds for these experiments were generated by growing tomatoes in sunshine mix soil in the Jane Gray Greenhouse at UC Berkeley. Sunshine Mix #1 soil was used. Upon flowering, plants were manually pollinated by flicking flowers. Care was taken to switch gloves between plants of different genotypes. Fruits were collected and placed in plastic Ziploc bags, manually crushed, and allowed to ferment at 21°C for 2-3 weeks. After the fermentation process was complete, seeds were strained from remaining fruit material, rinsed with DI water, and allowed to dry on filter paper. Seeds were stored in the dark at 21°C until use. All genotypes were used for passages one, two, three, and p4-combined. Genotype 2934 was not used in passage four, as that genotype succumbed to fungal disease in the third generation. The community coalescence competition experiment included genotypes 2706, 3472, and 2934.

Tomato germination and growth

Seeds were surface sterilized using TGRC recommendations as follows: seeds were soaked in 2.7% bleach (sodium hypochlorite) solution for 20 minutes. Sterilized seeds were then washed with sterile ddH_2O three times to remove any excess bleach. Sterilized seeds were then transferred onto 1% water agar plates and placed in the dark at 21°C until emergence of the hypocotyl. At that point, seedling plates were moved into a growth chamber and allowed to continue germination for 1 week. Growth chamber conditions were 25°C, 65% humidity

and 16 h daylight per day. After approximately one week, seedlings were planted in sunshine mix #1 soil in seedling trays. After approximately one more week of growth, seedlings were transplanted into 8” diameter pots, making the plants approximately 2.5-3 weeks old at the first time of microbial inoculation. Age of inoculation varied slightly from experiment to experiment but was kept identical amongst genotypes within an experiment.

Inoculation preparation, first passage

Microbial inoculum for the first passage of the experiment was generated from field-grown tomato plants from the UC Davis Student Organic Farm collected in September and October of 2016. One-gallon Ziploc bags were filled with leaf, stem, and some flower material from tomato plants. One bag was collected from each of nine different sites, spread through four different fields. Plant material was collected from various genotypes of tomatoes. Other plant types, such as lettuce, eggplant, corn, and oak trees, surrounded the tomato fields. During the October collection, soil was also collected at each site. The top ~2cm of soil was brushed away, and a 50mL conical was pushed directly into the soil at the base of a plant which was in the middle of each collection site. Plant material and soil were transferred to the lab on ice and stored at 4°C briefly until processing. Sterile phosphate freezing buffer was added to the bags of leaves, and the entire bags were placed in a Branson M5800 sonicating water bath. Material was sonicated for 10 minutes. This gentle sonication washes microbes from the surfaces of the leaves but does not damage cells. The resulting leaf wash from each site was pooled. From the September collection, leaf wash was pelleted for 10 mins at 4000 x G, re-suspended in glycerol freezing buffer, and stored at -80 for approximately one month. This was then thawed, re-spun to remove the freezing buffer, and combined with the October leaf wash. At that point, the starting inoculum was divided into 6 aliquots and stored in glycerol freezing buffer. For each inoculation in the first passage, an aliquot was thawed and cells pelleted for 10 mins at 4000 X G. Cells were re-suspended in 200mL 10mM $MgCl_2$ buffer. Of

this, 40mL were and heat killed in an autoclave for a 30 minutes at 121°C. Inoculum was plated, and an absence of growth confirmed that the heat-kill was effective. To get initial concentration of inoculum, dilution plating was performed on Kings Broth agar plates (1.1 X 10 ^6 CFU/mL). Soil from each site, which had been stored at -20°C, was combined in a sterile Nalgene bucket and thoroughly mixed before inoculation.

Inoculation procedure

Soil inoculation: The top layer of every pot was supplemented with 40 grams of UC Davis Farm Soil. Soil inoculation was only performed once and only for the first passage of plants. Spray inoculation: Each plant was sprayed, using misting spray tops placed in 15mL conicals, with approximately 4.5mL of inocula. Control plants from passage 1 were inoculated with the heat-killed inocula. Control plants from subsequent experiments were inoculated with sterile 10mM $MgCl_2$. Immediately after inoculation, plants were placed in a random order in a high-humidity misting chamber for 24 hours. After 24 hours, the plants were moved to a greenhouse bench. Plants were inoculated once per week in the same manner and were placed in the misting chamber for 24 hours after every inoculation.

Plant sampling and inoculation preparation for passaging lines

Ten days after the final spray inoculation, plants were sampled. With the exception of the P4-Combined experiment, all plants were cut off at the base and immediately placed into sterile 1L bottles individually. By the end of P4-Combined, the plants had grown too large to sample the entire plant, and instead, roughly 2/3 of the plant material was sampled from each plant, with care taken to sample the same age of branches from every plant. After collection, plant material was weighed, and 200mL of sterile 10mM $MgCl_2$ were added to each bottle containing the plant material. The bottles were submerged in a sonicating water bath, sonicated for 5 minutes, vortexed, and sonicated for another 5 minutes. Half of the volume from each plant was

pelleted for 10 mins at 4200 X G, re-suspended in ~1mL of 1:1 KB Broth Glycerol, divided into aliquots, and stored at -80°C for inoculation of the subsequent passage. The other half of the volume was pelleted in the same manner and then stored as a pellet at -20°C for DNA extractions. To prepare inoculation of the next passage, microbiome glycerol stocks were thawed, briefly pelleted to remove glycerol, and re-suspended in sterile 10mM $MgCl_2$. Volume of re-suspension depended slightly on the size of the plants, but in general ranged from 5-10mL. Microbiomes were never pooled.

Inoculation preparation, combination of P4 microbiomes (Figure S7)
Frozen microbiomes from all plants from the end of passage four were thawed, and half the volume was removed from each aliquot. These aliquots were combined into one pooled meta-inoculum. This was divided into six aliquots. One was used immediately, and the rest of the aliquots were stored at -20°C in KB Glycerol and thawed by aliquot for each week of inoculation, as above.

P1, P4 coalescence experiment (Figure 5)
Genotypes 2706, 3472, and 2934 were used for this experiment, and four plants of each genotype received each treatment (P1, P4, and Mix). One control plant of each genotype was spray inoculated with $MgCl_2$. To prepare the inoculum, microbiomes from the end of passage one and the end of passage four were used. All aliquots (one from each plant, except for plant 4 which had exhibited disease symptoms) were thawed and combined. The same was done for all of the individual microbiomes that came off of passage 4 plants. To remove the glycerol, the samples were spun down and resuspended in 10mM $MgCl_2$. In order to generate the 50/50 mix of P1 and P4 microbiomes, live/dead PCR with PMA treatment was used, adapted from the following method [1]. Briefly, serial dilutions of P1 and P4 were performed in $MgCl_2$. Each sample then received PMA at a final concentration of 100uM and vortexed. Samples were incubated in the dark at room temp for 5 minutes. Then they were placed in ice on a tray exactly 10cm away from a 700 watt

halogen lamp. The light was turned on for 30 seconds, and turned off for 30 seconds. During the 30 seconds without light, the samples were all vortexed. This was repeated three more times. Samples were then pelleted for 10 minutes at 5000 X G. The supernatant including the excess PMA was removed, and cells were re-suspended in sterile 10mM $MgCl_2$. Droplet Digital PCR (as described below) was then utilized to quantify bacteria from each sample, and concentration was matched to 7.7 x 10^6 cells/mL. P1 and P4 were aliquotted separately and then re-combined for the mixed inoculum so that each plant received ~9 x 10^4 bacteria each week that they were inoculated. Plants were inoculated for three weeks and harvested 10 days after the final inoculation as described previously.

Bacterial quantification using ddPCR

The BioRad QX200 system was used for culture independent quantification of bacteria. Complete ddPCR methods are described elsewhere [2]. Bacterial abundance was measured directly after microbes were sonicated off plant surfaces into sterile buffer. For consistency, the same region of the 16S gene used below for amplicon sequencing was used for bacterial quantification. PNAs, were used as well to limit any background amplification of plant mitochondrial or chloroplast DNA. Five ul of sample were used in every reaction. All data were normalized to weight, in grams, and concentrations are reported as 16S copy number/gram.

DNA extractions

DNA was extracted from microbial pellets using the Qiagen PowerSoil DNA extraction kit. A buffer control extraction was included for every set of extractions in order to identify and exclude taxa present in the dataset due to buffer contamination.

16S Libraries

The 16S rRNA gene was amplified using dual-indexed primers designed for the V3- V4 region [3] using the following primers: 341F (5 -CCTACGGGNBGCASCAG-3) and 785R (5 -GACTACNVGGGTATCTAATCC-3) [4]. Additionally, we also used peptide nucleic acids, PNAs [5] to decrease amplification of plant mitochondrial and chloroplast DNA. Negative buffer controls and PCR controls were sequenced along with experimental samples. Reaction conditions were 94°C for 3 min, 94°C for 45 s, 78°C for 10 s, 50°C for 1 min, 72°C for 1.5 min, repeat steps 2–5 30 times and 72°C for 10 min. PCR mixtures were randomized in order, run in duplicate for each sample, pooled and quantified using Qubit. Amplicons from each sample were pooled in equimolar concentrations, cleaned using an AMpure bead clean-up kit. Libraries were prepared for paired 300-nucleotide reads in Illumina's MiSeq V3 platform (Illumina) at The California Institute for Quantitative Biosciences (QB3) at UC Berkeley.

ITS Libraries

Using the same DNA as above, the ITS2 region was amplified using ITS9-F: GAACGCAGCRAAIIGYGA and ITS4-R: TCCTCCGCTTATTGATATGC following a protocol published online by the Joint Genome Institute. A second PCR was performed (7 cycles) in order to anneal MiSeq Illumunia adapters and barcodes onto the amplicons. PCRs were carried out in duplicate and pooled before they were prepared for sequencing by the QB3 sequencing facility as described above.

Sequence Processing and Analysis

MiSeq sequencing files were demultiplexed by QB3 sequencing facility. Fastq Files are deposited in the NCBI BioProject database (BioProject ID: PRJNA578761). Reads were combined into contigs using VSearch [6], and the remainder of the analysis was carried out in Mothur [7] following their MiSeq SOP [8]. Data were quality-filtered, and chimeras were removed using UChime [9]. Singletons were removed using the split.abund command in Mothur after pre-

clustering of similar sequences. We used a 97% similarity cut-off for defining OTUs. The Silva reference database [10] was used for sequence alignment and taxonomic assignment. Archaeal, chloroplast, mitochondrial and unknown domain DNA sequences were removed. To account for reagent contaminants, we also sequenced two DNA extraction kit controls and PCR controls along with our samples. Contaminant OTUs from control samples that were at a similar or higher relative abundance in control samples compared to experimental samples were removed from the full OTU table. Bacterial were rarified to 8,000 reads per sample. For the fungal community, an OTU table was generated from the fungal community sequencing data using QIIME 2. Trimmed, paired reads were first denoised, without read trimming, using the DADA2 plug-in [11]. Chimeric sequences were then filtered using the uchime-denovo command of the Vsearch plug-in [12]. Reads were then clustered into OTUs at 97% identity using the cluster-features-closed-reference command in the VSEARCH plug-in and the 2017 version of the UNITE database [13]. In order to assign taxonomy to the clustered OTUs, a Naïve-Bayes classifier was first trained using the UNITE database and the feature-classifier plug-in [14]. The classify-sklearn command of the feature-classifier plug-in was finally used to assign taxonomy to the clustered OTUs. Fungal reads were rarified to 415 reads/sample. Once bacterial and fungal OTU tables were generated in Mothur and QIIME, the remainder of the analysis was performed in R using the following packages: Phyloseq [15], vegan [16], ampvis2 [17], and MicrobiomeSeq (Alfred Ssekagiri, William T. Sloan, Umer Zeeshan Ijaz). Occupancy abundance plots were generated using R code from the following source: http://rstudio-pubs-static.s3.amazonaws.com/266780_cac4994322494658904507a7606b1dd8.html. Goodness of fit for linear and quadratic models of Bray-Curtis dissimilarity to P1 communities was based on both R^2 values and calculating Akaike information criterion values. For determining the effects of specific variables on Bray-Curtis dissimilarities between samples, PERMANOVA tests were run using Vegan's Adonis and adonis2 functions.

Genotype PERMANOVAs in P1 and P2

For the genotype effect observed in P1 and P2, the data were also analyzed with the removal of the primary outlying line in P1, which was a diseased plant. The significance does not change (30% of dissimilarity explained, p=0.002). That same line had too low read depth to be analyzed at P2, and thus was excluded from this analysis at the rarefaction step. By P3, this line was included, as it did not fall outside of the 95% confidence intervals for P3 clustering. Additionally, replicate lines from accession 2934 were lost after passage 3 due to a fungal pathogen present in the original inoculum that seemingly only infected this genotype. To ensure that this genotype alone was not driving the effect of genotype on community composition, we excluded it from the dataset and re-ran statistical tests. There remains a significant overall genotype effect when all passages are analyzed together (PERMANOVA, $F_{3, 79}$= 1.9723, p= 0.034). We then subsetted the data, again ran univariate PERMANOVAs at each passage, and we found that genotype remains a significant factor explaining Bray-Curtis dissimilarity at P1 and P2, but not P3 nor P4. (P1: R^2=0.25, p=0.017; P2: R^2=0.31, p=0.009; P3: R^2=0.16, p=0.228; P4: R^2=0.10, p=0.943).

Incorporation of repeated measures into statistical models: In the serial passaging experiment, each microbiome line was independently passaged across four cohorts of tomato plants, and each microbiome line was sampled at the end of each passage. Although the microbiomes were never sampled multiple times from the same tomato plant, the data structure is similar to what one would find in time series experiment. Thus, wherever possible, “Line ID” was incorporated into models to take this into account. The following linear mixed effects model was utilized for determining significant changes in diversity over time: lmer(Values ~ Passage + (1|LineID). In the case of PERMANOVA tests, the strata term was used to limit permutations within Line IDs to test for the main effect of Passage. Furthermore, the strata term cannot be utilized when determining significance of terms by=”margin” (Type III tests). Instead, significance is assigned to each term sequentially from first to last. Thus, order of

terms in the model may impact significance. In the data presented in this manuscript, all iterations of term order were tested in each model. Statistics are presented using the following models with the use of strata: adonis(bray.matrix ~ Passage + Genotype, permutations=999, strata= LineID). Importantly, although changing the order of terms sometimes slightly impacted the R^2 values, none of the differences had any impact on a variable's significance level. Changing the order of terms did not impact the interpretation of the importance of any variable tested in this dataset. The adonis2 test with the by="margin" term was used whenever the strata term was not included in the model. The following model was utilized in these cases: adonis(bray.matrix ~ Passage + SampleType, by="margin", permutations=999).

Community Cohesion Metrics

The estimations of positive and negative cohesion values follows the cohesion metrics approach proposed by Herren *et al*. [18]. Herren *et al*. multiplied the connectedness metrics determined by relative abundance profile by the same relative abundance profile to estimate cohesion values. We modified their method to estimate cohesion values by using two relative abundance profiles of a training set and test set. Relative abundance profile of the training set was obtained by randomly selecting half of the samples in each microbiome passage. The test set consists of the other half of the samples. Using the training set and following the same procedure as Herren *et al*., connectedness metrics were calculated. The estimated connectedness metrics subtracts a null model. The objective of the null model was to calculate the strength of pairwise correlations that would be observed if there were no true relationship between OTUs. The obtained connectedness metrics are multiplied by relative abundance profile of test set to estimate positive and negative cohesion values. Two hundred iterations of sampling randomization in each microbiome passage were carried out at OTU level to obtain training set and test set for P1, P2, P3, and P4.

Neutral model

The neutral model was proposed by Sloan *et al*. to describe both microbial diversity and taxa-abundance distribution of a community [19]. Burns *et al*. [20] have developed a R package based on Sloan's neutral model to determine the potential importance of neutral process to a community assembly. In brief, the neutral model creates a potential neutral community by a single free parameter describing the migration rate, *m*, based on two sets of abundance profiles – a local community and metacommunities. The local community describes the observed relative abundance of OTUs, while the metacommunity is estimated by the mean relative abundance across all local communities. The estimated migration rate is the probability of OTU dispersal from the metacommunity to replace a randomly lost individual in the local community. The migration rate can be interpreted as dispersal limitation. In each microbiome passage, half of the samples were randomly selected and the relative abundance profile at the OTU level was used. The neutral model fit and migration rate were estimated in the resolution results of 200 iterations for P1, P2, P3, P4, and P4 Combined.

Null model predictions

We applied a null model approach on the serial passaging data P1-P4 to characterize the changes of stochastic process driving the assembly of plant microbiome over time. Lines that had high quality sequencing data at every time point (thirteen in total), were used for this analysis. The null scenario for each line at each passage was generated using the data for that same line at the previous passage. The null scenario of P1 was generated using the original field inoculum sample. The null model approach was based on community pairwise dissimilarity proposed by Chase and Myers (Chase and Myers 2011) and extended by Stegen *et al.* to incorporate species abundance (Stegen, Lin et al. 2013). Chase and Myers proposed a degree of species turnover by a randomization procedure where species probabilistically occur at each local community until observed local richness is reached. However, the estimated degree of turnover does not include species abundance. To take full advantage of our dataset, we also

incorporated species relative abundance into the procedure proposed by Stegen *et al*. Zinger *et al*. has developed R code for the null model and applied the null model approach on the soil microbiome [21]. This approach does not require *a priori* knowledge of the local community condition and determines if each plant microbiome at the current passage deviates from a null scenario generated by that same microbiome at the previous passage. In brief, the null scenario of each was generated by random resampling of OTUs and remained the same richness and number of reads with the original sample. Total OTUs observed in the sample and the corresponding relative abundance were used as probabilities of selecting an OTU and its associated number of reads, respectively. The Bray-Curtis metric is used to calculate dissimilarities across null communities with 1,000 permutations. The average of dissimilarities among permutations represents null expectations of community dissimilarities. The null deviation shows the differences between average null expectation and the observed microbiome of the same line.

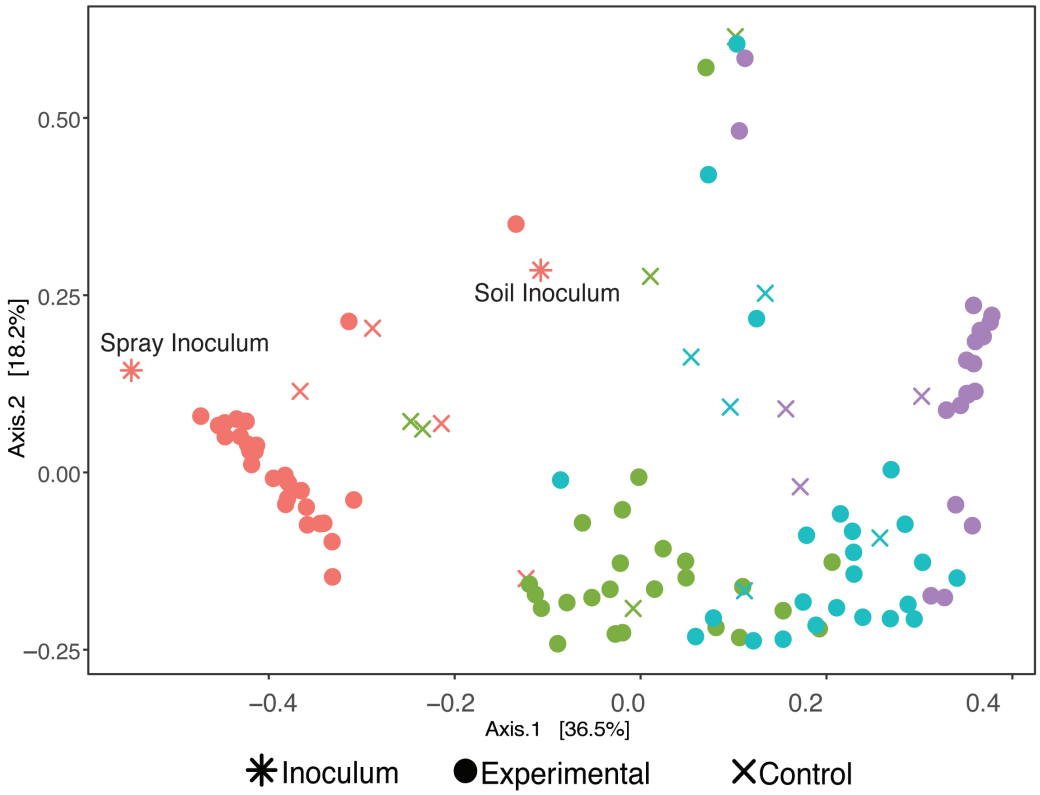


Fig. S1. **Bacterial community differences between sample type and passage number**

A PCoA plot of Bray-Curtis dissimilarity amongst samples illustrates a significant effect (determined by a multivariate PERMANOVA test) of both passage (colors) and sample type (shapes).

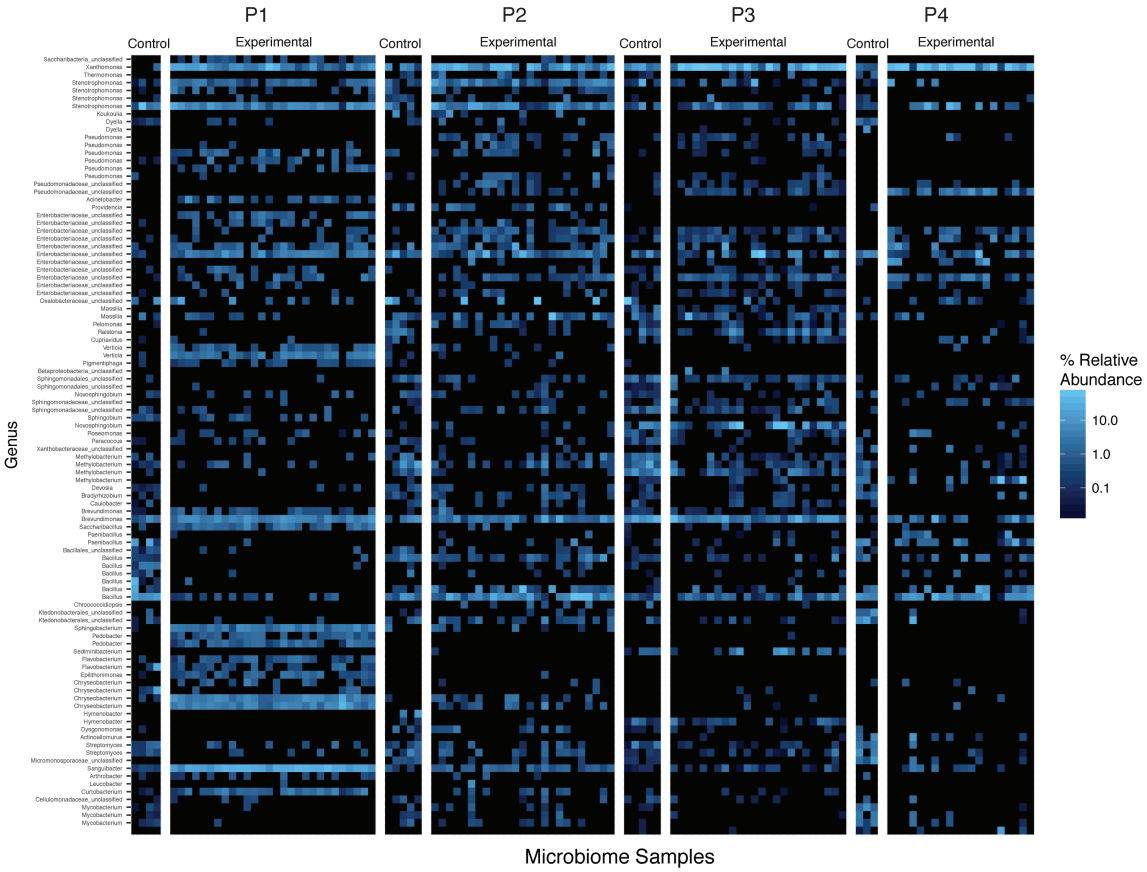


Fig. S2. Changing relative abundance of top 100 non-inoculum OTUs

Only OTUs that were not identifiable in the spray inoculum sample were included in this analysis. Relative abundance top 100 non-inoculum OTUs are plotted as a heatmap. Each OTU is shown with its Genus-level taxonomic identification. Importantly, some OTUs classified as “non-inoculum” may have in fact been present in the initial spray inoculum but at too low of a relative abundance to recover sufficient high-quality sequences.

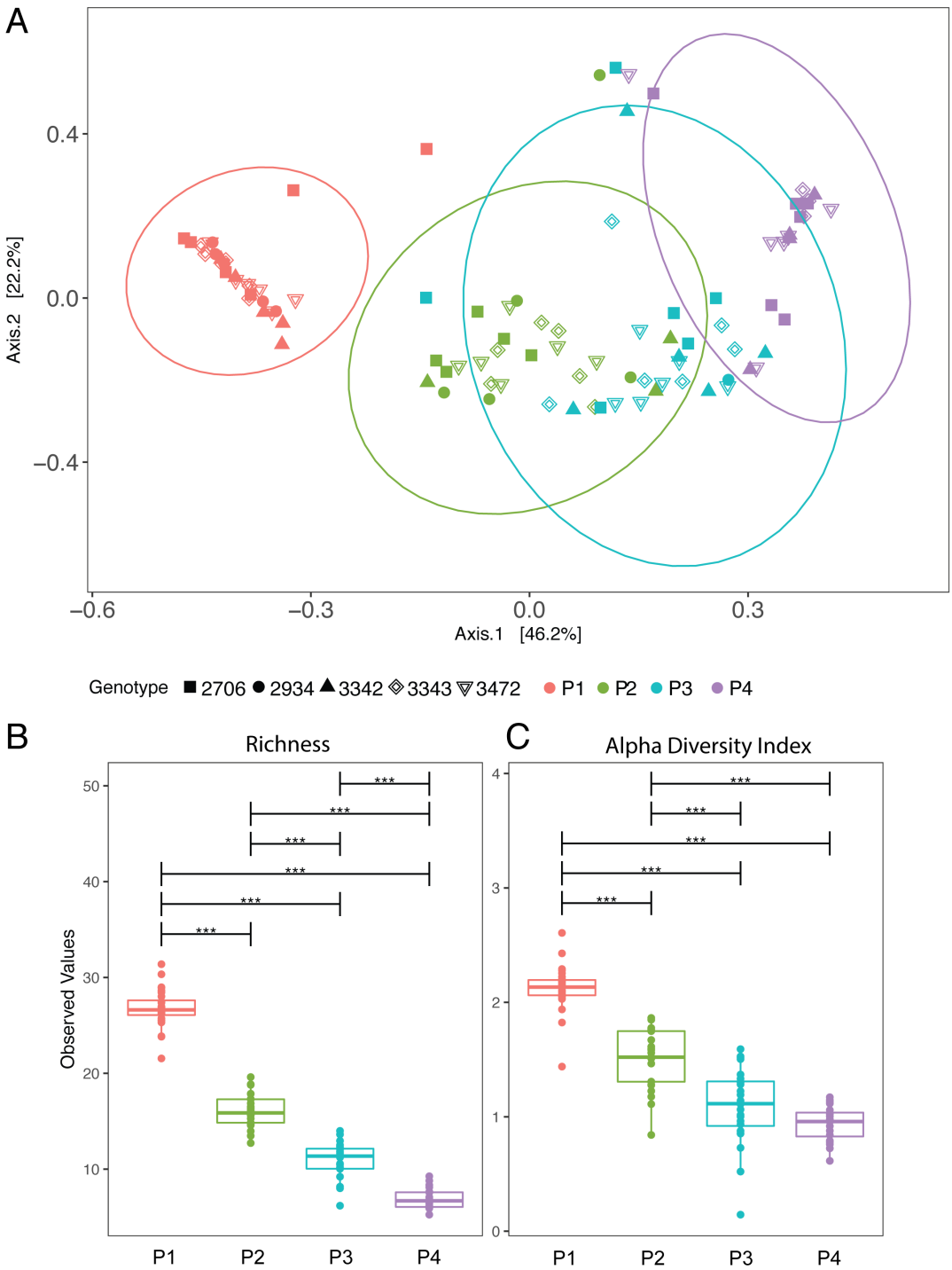
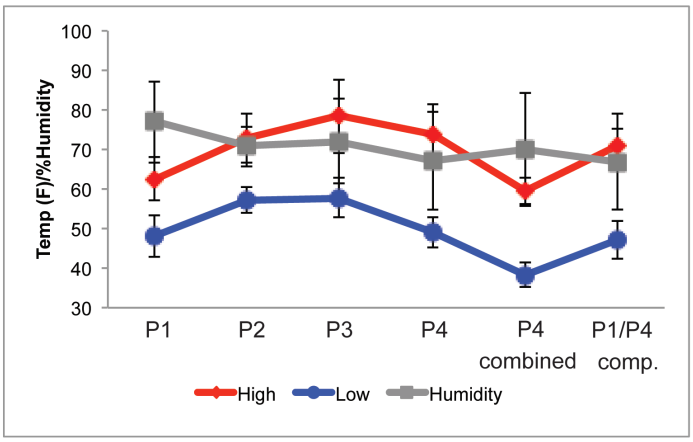


Fig. S3. Serial passaging of the phyllosphere microbiome: Inoculum only taxa

The dataset was subsampled to only contain OTUs that were present in the initial spray inoculum. A PCoA plot of Bray-Curtis dissimilarity of experimental plants shows a significant effect of passage (colors) and genotype (shapes) (a). Richness (b) and Shannon's alpha diversity index (c) of each experimental plant are plotted at each passage and show a significant decrease over time. Corrected p values from multiple pairwise comparisons are illustrated on the graph * p≤0.05; ** p≤0.01; *** p≤0.001.

A



B

Passage/Exp	Start	End
One	11/8/16	12/15/16
Two	6/1/17	7/4/17
Three	8/25/17	9/26/17
Four	10/3/17	11/1/17
P4 Combined	11/28/17	1/3/18
P1/P4 Comp.	10/18/18	11/12/18

Fig. S4. Climatic variables

Although all experiments were performed in the greenhouse, outside climatic variables varied for the duration of the six experiments. Humidity, high temp, and low temp are plotted (a). The date at which the plants were first inoculated and the date at which they were harvested are shown (b).

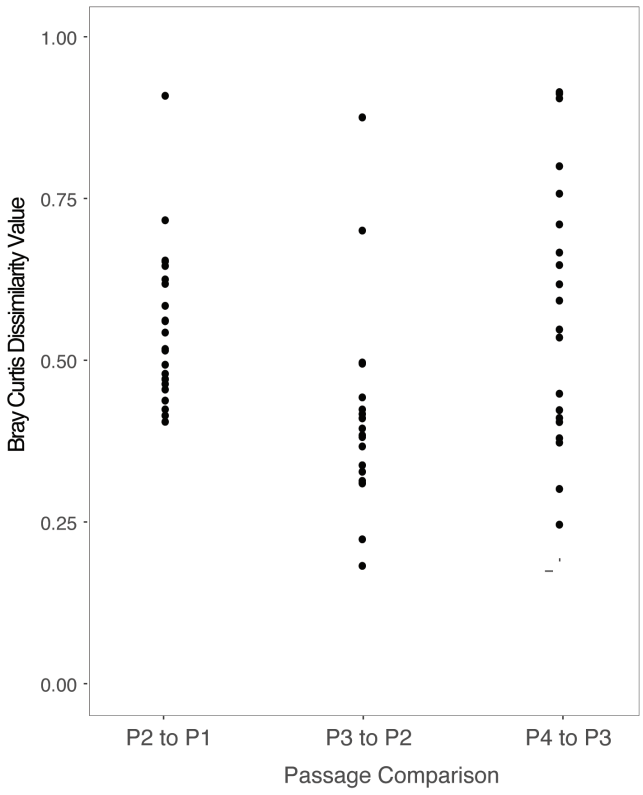


Fig. S5. Bray-Curtis dissimilarity within lines over time

Bray-Curtis dissimilarities between (n) and (n-1) are displayed for each passaged microbiome line. There is no observable significant effect of “comparison” on Bray-Curtis dissimilarity, but we did uncover a moderately significant effect of “Line ID” (ANOVA: see main text for statistical values).

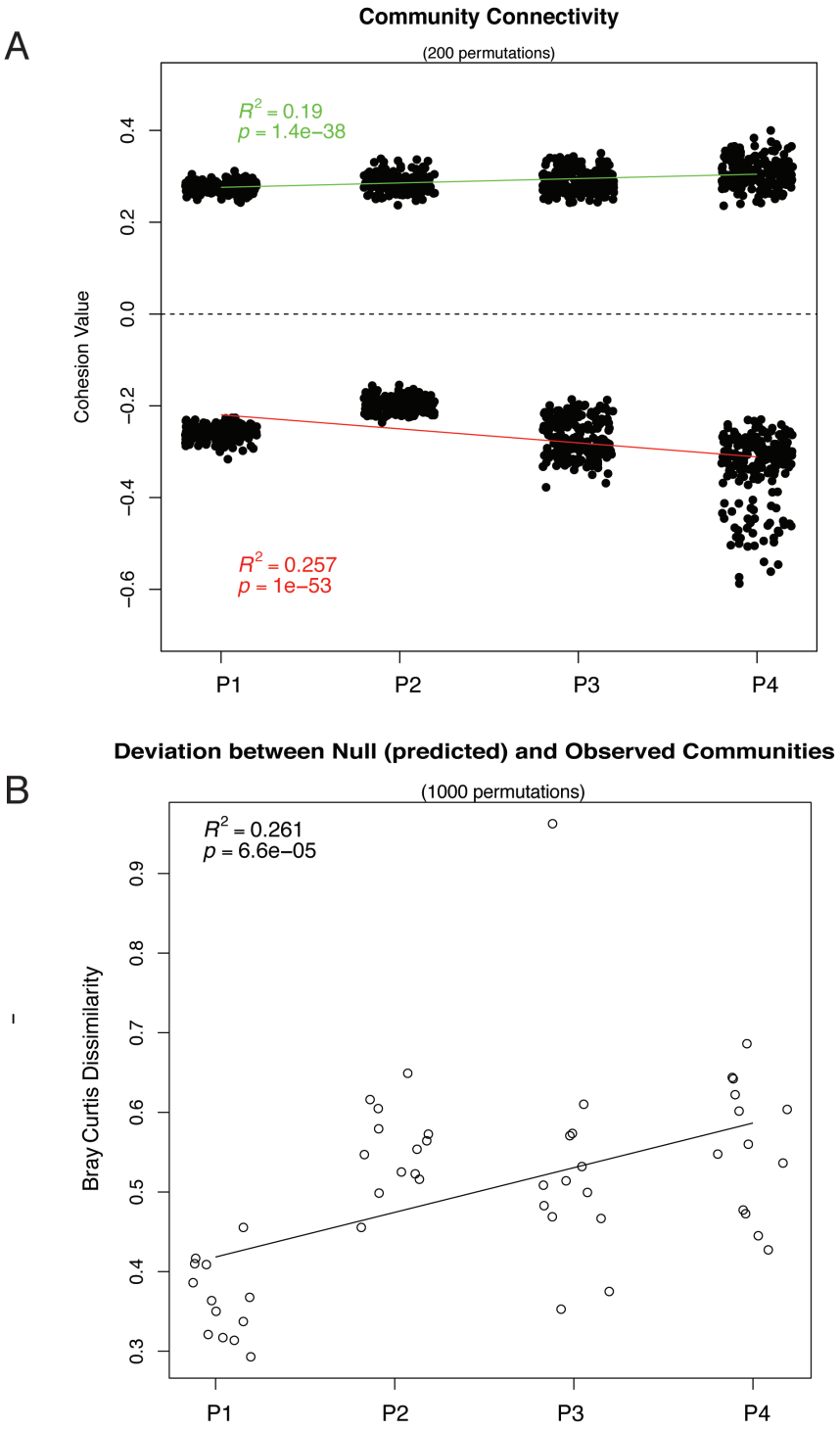


Fig. S6. Modeling bacterial community dynamics

We applied community cohesion metrics on our dataset to describe microbial dynamics in P1, P2, P3, and P4 (a). We calculated both positive and negative cohesion values and found a mild but significant increase in positive and negative cohesion values from P1 to P4 (values indicated on graph). We next compared Bray-Curtis dissimilarities between a predicted null model from the (n-1) passage with observed communities from passage n over 1,000 iterations and we found a mild but significant increase in deviation from the null prediction from P1 to P4 (b).

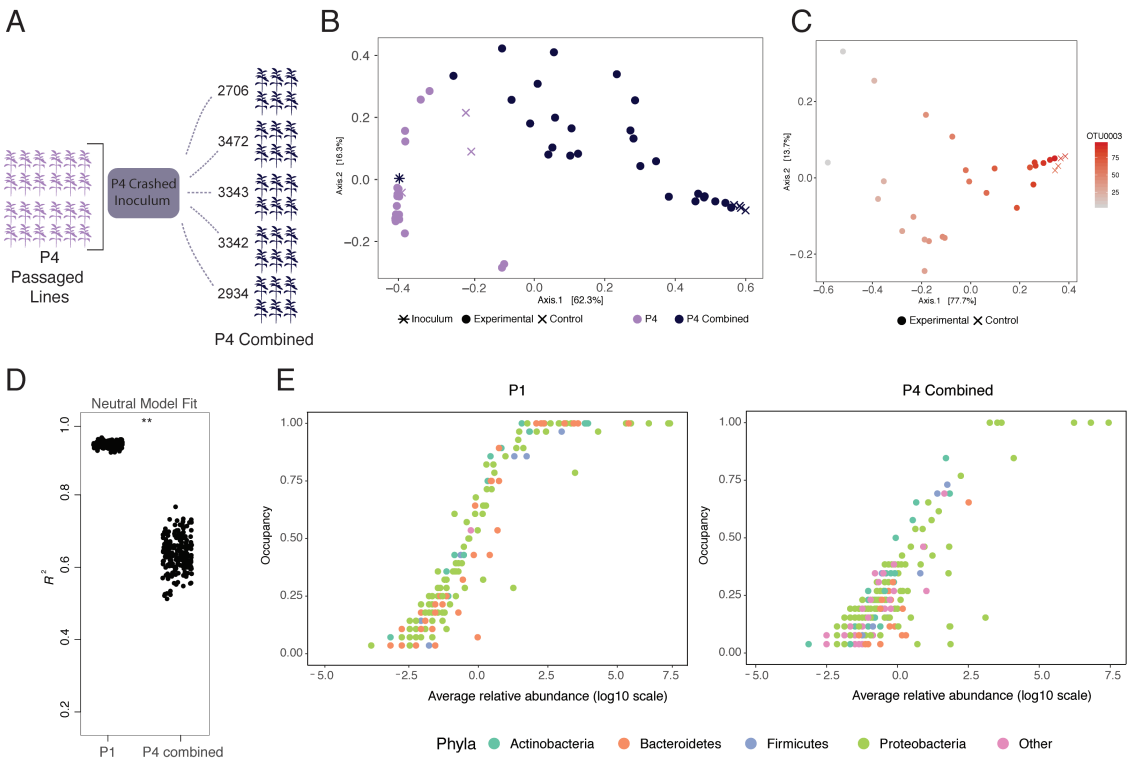


Fig. S7. Combination of passaged lines and re-inoculation

Passaged microbiomes from the end of P4 were pooled and then sprayed onto tomato plants (six replicates of five genotypes) (a). P4 plants, inoculum generated from P4 plants, P4-Combined, and controls are plotted on a PCoA plot of Bray-Curtis dissimilarities, and P4-Combined plants cluster apart from P4 plants (b). We compared the observations of community structure and predicted community composition by a neutral model for P1 and P4-Combined (c), and in 200 iteratively predictions, the fit of the neutral model is significantly higher in P1 than P4-Combined (Student's *t*-test, *p*-value < 0.01). Visualized on a PCoA plot (d), relative abundance of OTU0003 can be seen driving Bray-Curtis dissimilarity amongst plants, both experimental and control. Occupancy-Abundance curves for P1 and P4-combined are shown side-by-side (e).

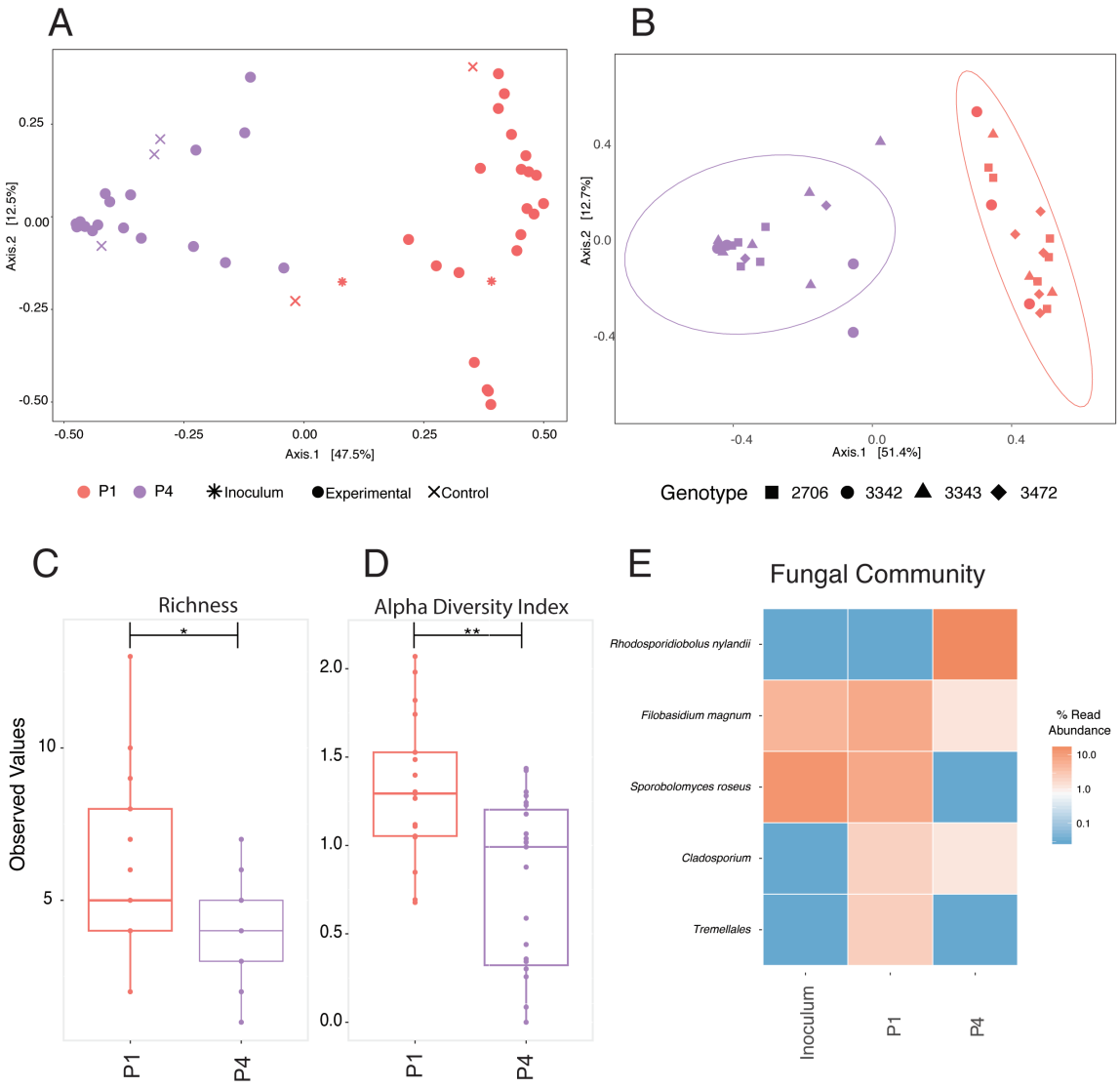
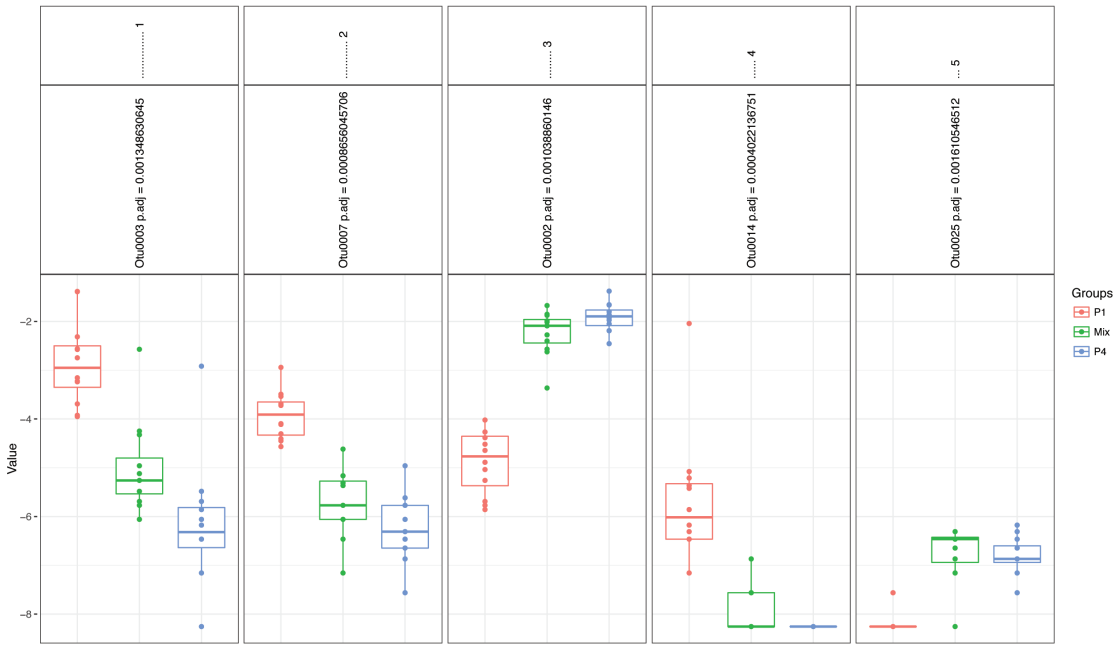


Fig. S8. The fungal community

A PCoA plot of Bray-Curtis dissimilarity amongst samples shows a significant change in the community from P1 to P4, as determined by a PERMANOVA test (a). There is no effect of genotype (shapes) on the fungal community at either passage (b). Ellipses indicate 95% confidence around the clustering. Both richness (c) and Shannon's alpha diversity (d) significantly decrease between P1 and P4. Relative abundance of the top five fungal taxa is plotted for the original inoculum, P1 and P4 (e). Corrected p values from multiple pairwise comparisons are illustrated on the graph * p≤0.05; ** p≤0.01.



	Phylum	Class	Order	Family	Genus
Otu0003	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Otu0007	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Otu0002	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonadaceae_unclassified
Otu0014	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Rahnella
Otu0025	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Enterobacteriaceae_unclassified

Fig. S9. Differentially abundant taxa in community coalescence experiment

We performed a Kruskal-Wallis test on log-relative transformed OTU abundance at different passages using the MicrobiomeSeq package (a). This is a non-parametric method, and it tests whether samples originate from the same distribution. P-values are corrected for multiple testing using family wise error rates. Significant OTU rankings 1-5 are assigned importance using random forest classifier. Taxonomies of OTUs are displayed as well (b).

Table S1. Top 100 OTU taxonomic assignments

OTU	Phylum	Class	Order	Family	Genus
0001	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Pantoea
0002	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Unclassified
0003	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Unclassified
0004	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
0005	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
0006	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Xanthomonas
0008	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
0009	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Unclassified
0010	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
0011	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Unclassified
0012	Firmicutes	Bacilli	Bacillales	Family_XII	Exiguobacterium
0013	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Massilia
0014	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Unclassified
0015	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Brevundimonas
0016	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus
0017	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Novosphingobium
0018	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Massilia
0019	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Rahnella
0020	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Pedobacter
0021	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
0022	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas
0023	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas
0024	Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	Curtobacterium
0025	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus
0026	Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	Unclassified
0027	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae	Methylobacterium
0028	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Unclassified
0029	Proteobacteria	Betaproteobacteria	Burkhoderiales	Oxalobacteraceae	Massilia
0032	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae	Methylobacterium
0033	Actinobacteria	Actinobacteria	Micrococcales	Sanguibacteraceae	Sanguibacter
0034	Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	Unclassified
0035	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae	Methylobacterium
0036	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Ralstonia
0037	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Pantoea
0038	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Unclassified
0039	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Chitinophagaceae	Sediminibacterium
0040	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Pedobacter
0041	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Pedobacter
0042	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Duganella

0043	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus
0045	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Paenibacillus
0046	Actinobacteria	Actinobacteria	Streptomycetales	Streptomycetaceae	Streptomyces
0047	Firmicutes	Bacilli	Bacillales	Family_XII	Exiguobacterium
0048	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Rhizobium
0049	Chloroflexi	Ktedonobacteria	Ktedonobacterales	Unclassified	Unclassified
0050	Actinobacteria	Actinobacteria	Streptomycetales	Streptomycetaceae	Streptomyces
0052	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Unclassified	Unclassified
0053	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Unclassified
0055	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae	Methylobacterium
0056	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
0057	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Chryseobacterium
0059	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Unclassified
0061	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Unclassified
0062	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Chryseobacterium
0065	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Unclassified
0066	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Paenibacillus
0068	Actinobacteria	Actinobacteria	Micrococcales	Micrococcaceae	Pseudarthrobacter
0069	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
0070	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Unclassified
0071	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Paenibacillus
0073	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Bradyrhizobium
0074	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Caulobacter
0076	Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	Rathayibacter
0077	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Unclassified
0079	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter
0081	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Pedobacter
0082	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Koukoulia
0083	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Sphingobacterium
0085	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Unclassified
0086	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium
0088	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Roseomonas
0089	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Unclassified
0093	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
0094	Proteobacteria	Betaproteobacteria	Unclassified	Unclassified	Unclassified
0096	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Pedobacter
0097	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Saccharibacillus
0098	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Brevundimonas
0106	Actinobacteria	Actinobacteria	Streptosporangiales	Thermomonosporaceae	Actinoallomurus
0115	Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	Curtobacterium
0116	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	Verticia

0118	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas
0123	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Unclassified
0126	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Unclassified
0130	Actinobacteria	Actinobacteria	Micrococcales	Micrococcaceae	Paenarthrobacter
0134	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Incertae Sedis	Reyranella
0136	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
0150	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae	Methylobacterium
0167	Actinobacteria	Actinobacteria	Micrococcales	Micrococcaceae	Arthrobacter
0169	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	Unclassified
0174	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Acidovorax
0189	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Unclassified
0190	Armatimonadetes	Fimbriimonadia	Fimbriimonadales	Fimbriimonadaceae	Unclassified
0208	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Chitinophagaceae	Heliimonas
0213	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Epilithonimonas
0215	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Pedobacter
0222	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Chitinophagaceae	Vibrionimonas
0223	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Pedobacter
0293	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Unclassified
0323	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Chitinophagaceae	Sediminibacterium

References

1. Carini P, Marsden PJ, Leff JW, et al (2017) Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. Nat Microbiol 2:16242. https://doi.org/10.1038/nmicrobiol.2016.242

2. Morella NM, Gomez AL, Wang G, et al (2018) The impact of bacteriophages on phyllosphere bacterial abundance and composition. Mol Ecol. https://doi.org/10.1111/mec.14542

3. Naylor D, DeGraaf S, Purdom E, Coleman-Derr D (2017) Drought and host selection influence bacterial community dynamics in the grass root microbiome. ISME J 11:2691. https://doi.org/10.1038/ismej.2017.118

4. Takahashi S, Tomita J, Nishioka K, et al (2014) Development of a Prokaryotic Universal Primer for Simultaneous Analysis of Bacteria and Archaea Using Next-Generation Sequencing. PLOS ONE 9:e105592. https://doi.org/10.1371/journal.pone.0105592

5. Lundberg DS, Yourstone S, Mieczkowski P, et al (2013) Practical innovations for high-throughput amplicon sequencing. Nat Methods 10:999–1002. https://doi.org/10.1038/nmeth.2634

6. Rognes T, Flouri T, Nichols B, et al (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ Preprints

7. Schloss PD, Westcott SL, Ryabin T, et al (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541. https://doi.org/10.1128/AEM.01541-09

8. Kozich JJ, Westcott SL, Baxter NT, et al (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79:5112–5120. https://doi.org/10.1128/AEM.01043-13

9. Edgar RC, Haas BJ, Clemente JC, et al (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194–2200. https://doi.org/10.1093/bioinformatics/btr381

10. Quast C, Pruesse E, Yilmaz P, et al (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596. https://doi.org/10.1093/nar/gks1219

11. Callahan BJ, McMurdie PJ, Rosen MJ, et al (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13:581

12. Rognes T, Flouri T, Nichols B, et al (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584. https://doi.org/10.7717/peerj.2584

13. Nilsson RH, Larsson K-H, Taylor AFS, et al (2019) The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Res 47:D259–D264. https://doi.org/10.1093/nar/gky1022

14. Bokulich NA, Kaehler BD, Rideout JR, et al (2018) Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6:90. https://doi.org/10.1186/s40168-018-0470-z

15. McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8:e61217. https://doi.org/10.1371/journal.pone.0061217

16. Dixon P, Palmer MW (2003) VEGAN, a package of R functions for community ecology. J Veg Sci 14:927–930. https://doi.org/10.1658/1100-9233(2003)014[0927:VAPORF]2.0.CO;2

17. Skytte Andersen KS, Kirkegaard RH, Karst SM, Albertsen M (2018) ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. bioRxiv 299537. https://doi.org/10.1101/299537

18. Herren CM, McMahon KD (2017) Cohesion: a method for quantifying the connectivity of microbial communities. ISME J 11:2426–2438. https://doi.org/10.1038/ismej.2017.91

19. Sloan WT, Woodcock S, Lunn M, et al (2007) Modeling Taxa-Abundance Distributions in Microbial Communities using Environmental Sequence Data. Microb Ecol 53:443–455. https://doi.org/10.1007/s00248-006-9141-x

20. Burns AR, Miller E, Agarwal M, et al (2017) Interhost dispersal alters microbiome assembly and can overwhelm host innate immunity in an experimental zebrafish model. Proc Natl Acad Sci 114:11181–11186. https://doi.org/10.1073/pnas.1702511114

21. Zinger L, Taberlet P, Schimann H, et al (2019) Body size determines soil community assembly in a tropical forest. Mol Ecol 28:528–543. https://doi.org/10.1111/mec.14919