bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Plant neighborhood shapes diversity and reduces interspecific variation of the phyllosphere microbiome

Kyle M. Meyer[1,*], Robert Porch[2], Isabella E. Muscettola[1], Ana Luisa S. Vasconcelos[3], Julia K. Sherman[1], C. Jessica E. Metcalf[4], Steven E. Lindow[5], Britt Koskella[1]

[1] Department of Integrative Biology, University of California, Berkeley, Berkeley, California, 94720, USA

[2] Institute of Ecology and Evolution, University of Oregon, Eugene, OR, 97403, USA

[3] Department of Soil Science, College of Agriculture "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 13418-900, Brazil

[4] Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, 08544, USA

[5] Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, California, 94720, USA

* To whom correspondences should be addressed: kmmeyer@berkeley.edu

Keywords: Tomato, Bean, Pepper, Leaf-associated microbiome, Community assembly, Microbiome dispersal, Neighborhood effects, Phyllosphere

Competing Interests: the authors declare that they have no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Microbial communities associated with plant leaf surfaces (i.e. the phyllosphere) are increasingly recognized for their role in plant health. While accumulating evidence suggests a role for host filtering of its microbiota, far less is known about how community composition is shaped by dispersal, including from neighboring plants. We experimentally manipulated the local plant neighborhood within which tomato, pepper, or bean plants were grown in a three-month field trial. Focal plants were grown in the presence of con- or hetero-specific neighbors (or no neighbors) in a fully factorial combination. At 30-day intervals, focal plants were harvested and replaced with a new age- and species-matched cohort while allowing neighborhood plants to continue growing. 16S community profiling revealed that the strength of host filtering effects (i.e. interspecific differences in composition) decreased over time. In contrast, the strength of neighborhood effects increased over time, suggesting dispersal from neighboring plants becomes more important as neighboring plant biomass increases. We next implemented a cross-inoculation study in the greenhouse using inoculum generated from the field plants to directly test host filtering of microbiomes while controlling for directionality and source of dispersal. This experiment further demonstrated that focal host species, the host from which the microbiome came, and in one case the donor hosts' neighbors, contribute to variation in phyllosphere bacterial composition. Overall, our results suggest that local dispersal is a key factor in phyllosphere assembly, and that demographic factors such as nearby neighbor identity and biomass or age are important determinants of phyllosphere microbiome diversity.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Plant leaf surfaces, commonly termed the phyllosphere, harbor a wide diversity of microorganisms [1]. These endophytic and epiphytic communities can influence plant health and fitness through a variety of means, including protection against pathogens [2,3], plant growth promotion [4], primary productivity enhancement [5], protection against abiotic conditions including frost [6], and fixation of atmospheric nitrogen (N) [7]. Plant hosts can exert some control over the abundance and composition of their microbiome members by virtue of the differing chemical and physical features of resources provided on their surfaces [8], but also through immune activity, molecular signaling, and barrier formation [9–15]. This filtering effect can give rise to predictable differences in microbiome composition among hosts [10,16–18], a phenomenon referred to as species identity (or genotype) effects. Evidence for such effects comes from phylogenetic clustering of associated microbial taxa [19,20], deviation from null or neutral expectations [21,22], changes consequent to host genetic manipulation [10], or compositional differences explained by species or genotype as a factor [16–18,20,23]. While species identity effects suggest the importance of host control over microbiota, such effects are often weak or variable when tested in broader environmental or ecological contexts [24,25]. This raises the question of whether and how host effects can be swamped by environmental factors in shaping the microbiome.

The neighboring plant community constitutes a major component of a plant's environmental and ecological context. Neighborhood effects, also known as associational effects, have been extensively studied for pathogen and herbivore

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transmission [26–28], revealing patterns of transmission that relate to the nearest conspecific neighbor (i.e. conspecific negative density dependence) [29–32] as well as species frequency-dependent patterns of host fitness [33–35]. Much less effort has focused on the role of neighborhood effects for non-pathogenic plant-associated microorganisms [36]. Given the prominent role of aerial transmission in shaping phyllosphere microbial communities [23,37–39], both neighbor identity and proximity are likely to be important factors shaping epiphytic microbial communities. Moreover, it has been shown both theoretically [40] and empirically [41] that in the presence of high dispersal rates, community members can persist even in the face of strong selection against them (e.g. as a result of plant filtering effects), a phenomenon termed mass effects. As such, differences in microbiota composition that arise between species could be diminished when inter-host dispersal is high. Indeed this has been shown in zebrafish, where differences in bacterial community composition among host variants were dramatically reduced when inter-host dispersal was allowed [42].

Recent observational research in tree communities has revealed detectable neighborhood effects on epiphytic communities [38], but many open questions remain. For instance, it is unclear whether neighborhood effects are general and causative, a crucial gap in knowledge if such effects are to be incorporated into agricultural practice. It is also unclear what role neighbor or focal plant identity and age/biomass play in microbiome assembly. Lastly, although host filtering and microbial dispersal are intimately intertwined, the relative impacts of each in shaping microbiome differentiation among species has not been described. We

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

address these knowledge gaps using field- and greenhouse-based experiments involving tomato, pepper, and bean host plants. By manipulating both the focal plant species identity and the local neighborhood composition in the field, we were able to directly test the relative importance of plant host filtering versus local dispersal sources in shaping microbiome composition on leaves. We then performed a controlled cross-inoculation study in the greenhouse to directly examine the effects of host filtering and inoculum source on microbiome assembly of the three plant species involved. We hypothesized that: 1) neighborhood effects would increase as neighboring plants increase in age and biomass; 2) that neighborhood effects would depend on both neighbor and focal plant identity due to host filtering effects; 3) that host species effects would be diminished in the presence of neighboring plants; and 4) that experimental transplantation of microbiomes across hosts would result in compositional change as a result of host filtering, but that there would remain a detectable signal of past host.

Methods

Experimental Design: Neighborhood study

To test for the relative influence of host species and neighborhood effects on foliar microbial communities, we implemented a fully factorial, randomized block design at the Oxford Tract, a research farm near the University of California, Berkeley. The study included three plant species: tomato (*Solanum lycopersicum* var. Moneymaker), pepper (*Capsicum anuum* var. Early Cal Wonder), and bean (*Phaseolus vulgaris* var. Bush Blue Lake 274). Plant neighborhoods were established

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

116 in which a single 5 week old tomato or pepper plant, or a cluster of 6, 2-week old
117 beans, was planted as the focal individual in the middle of a circle of eight
118 neighborhood plants, each planted 0.61 m from the focal plant, with fully reciprocal
119 combinations of focal and neighborhood plants (Fig. 1A) in a randomized complete
120 block design with 6 replicates. Neighborhood plots were established in 3 zones (2
121 blocks per zone) spaced 1.22m apart separated by at least 0.91m from any other
122 plants at the experimental site to minimize edge effects. Each neighborhood plot
123 was 1.22 x 1.22 m, and separated by 1.22 m from adjacent plots. No neighbor
124 control plots, in which focal plants had no neighbors encircling them, were included
125 for each species. Thus, the experiment contained 9 neighborhood comparisons and
126 3 no neighbor comparisons. Two weeks prior to planting, soil was tilled for weed
127 management and drip lines were installed underneath plastic sheeting to provide
128 irrigation. The plastic sheeting prevented the growth of weeds and minimized
129 dispersal of soil onto plants. Plants were planted through small holes made through
130 the plastic sheeting. Individual tomato and pepper plants were propagated in a
131 greenhouse for 5 weeks, to a height of about 20 cm before transplantation into the
132 field. Beans, 6 seeds per pot, were grown for 20 days to a height of 20 cm before
133 transplantation. All greenhouse plants were watered using drip irrigation to
134 minimize the wetting of the leaves, and thus the development of large epiphytic
135 bacterial community sizes. All plants, including focal individuals and neighbors,
136 were transplanted in the field June 1, 2019.
137 Focal plants were harvested and replaced at 30-day intervals (3 times of
138 establishment) while the neighborhoods were retained and continued to grow

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

throughout the study (Fig. 1B, C). Focal plants were all the same age upon planting as the original cohorts to allow for direct comparisons across cohorts. Thus, bacterial community composition of the focal plants was assessed on 3 separate occasions for each of 72 focal plants in the neighborhoods (totaling 216 focal samples). Only one focal plant (pepper with no neighbors) died prior to sampling. Additionally, at each round of planting, bacterial community composition was assessed on 5 plants of each species at the time of transplantation to identify taxa that had established on plants in the greenhouse.

Neighborhood Plant Attribute Measurements

Several attributes of neighborhood plants were measured before each monthly harvest of focal plants in order to determine how neighbors might impact phyllosphere communities of the focal plant. These attributes included: average neighbor height, distance of the focal plants to the nearest neighbor, the number of neighbors touching the focal plant (if any), the total number of flowers on the neighborhood plants, and whether the neighborhood plants had signs of herbivory, infestation, or disease (yes or no). Further, the biomass of the neighborhood was estimated without harvesting the plants by fitting a linear model relating the height and weight of the focal plants, and extrapolating to that of the neighbor plant weights based on their height. Separate linear models were fit for each plant species.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sample processing

Immediately before harvesting the focal plants, their height was measured. Focal plants were then excised at their base using ethanol-sterilized scissors, transferred to 1-gallon (3.79 L) sterile plastic bags, and transported in a chilled cooler to the laboratory. Plant weight was recorded and then plants were subsampled, collected, and re-weighed. Subsampling was necessary to reduce biomass differences among samples and to enable more efficient collection of epiphytic bacteria by sonication. Foliar bacteria were collected from plant subsamples (range 3.84 – 621.14 g, median 40.84 g) by adding 180 ml of sterile 10mM $MgCl_2$ to the sample bags and sonicating for 10 minutes in a sonicating water bath (Branson model 5800). Leaf wash was then filtered through an autoclaved coffee filter and distributed to four 50 ml conical tubes, which were then centrifuged at 4000 rcf at 10° C for 10 minutes to pellet microbial cells. The supernatant was then decanted from each tube and the pellets resuspended in 1.8 ml King's broth (KB). 600 µl of the resuspended pellet was frozen at -80° C for subsequent DNA extraction, while the remaining two 600 µl aliquots were each mixed with 400 ul 1:1 KB:Glycerol and frozen at -80° C for subsequent experimentation.

Experimental Design – Follow-up transplant study

To further test the importance of host filtering, inoculum source, and dispersal history, we conducted a follow-up greenhouse study in which bacterial communities recovered from field plants at harvest time point 2 were reciprocally inoculated onto these same species under controlled conditions. Cryopreserved

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

phyllosphere communities from a single focal field plant were transferred to either the same plant species from which they were isolated, or onto the plant species that previously had surrounded that focal plant when it was in the field. For instance, the microbiome from a tomato that was surrounded by beans was applied equally to a tomato and a bean plant. This was done for all combinations. Experimental blocks from the field trial were treated as experimental blocks in the greenhouse trial, using blocks 2-6 from the field (5 replicates per treatment). We deliberately did not equalize inoculum density, as we anticipated that bacterial abundances would vary according to plant species and thereby constitute an important component of species identity effects. The biomass of every donor plant, however, was recorded for downstream statistical analysis. Additionally, for each plant species we included 5 replicate blank inoculum controls in which the same volume of sterile 10 mM $MgCl_2$ that was used to resuspend inoculum was sprayed onto plants. We further included replicate heat-killed controls, in which field-derived leaf wash was autoclaved for 40 minutes before being applied to each of three plant hosts, in the same manner as the experimental inocula.

Inocula were prepared by thawing the freezer stock, centrifuging at 4000 rcf and 10° C for 10 minutes to pellet cells, decanting the supernatant, and re-suspending cells in 7 ml 10 mM $MgCl_2$ then splitting in half to make two 3.5 mL inocula. Twenty-two samples were inoculated each day (block) such that each block contained every comparison, and this was repeated for 5 days. Inocula were sprayed onto the adaxial (top) and abaxial (bottom) sides of leaves using ethanol- and UV-sterilized misting caps. After inoculation, the moist, sprayed plants were placed in a

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

chamber maintaining ca. 100% relative humidity for 20 hours in order to maintain leaf moistness, thus encouraging microbial growth, before being transferred to a greenhouse. After 7 days, the plants were returned to the humid chamber for 20 hours immediately before harvest in order to facilitate further microbial multiplication on leaves and thus allow for maximal host filtering. Plants were then harvested and processed as in the field study.

DNA extraction, PCR, Library Preparation, and Sequencing

One sixth of the total leaf surface microbial extraction per plant was used for DNA extraction with DNeasy Powersoil Kits (Qiagen). Sample order was randomized to avoid batch effects, and a blank (no sample) control was included in every round of DNA extraction. DNA concentration of each sample was quantified using the Qubit dsDNA HS Assay Kit. 10 µl of sample DNA was used as template and PCR amplified for 35 cycles at the University of California - Davis Host Microbe Systems Biology Core using the 799F (5' – AACMGGATTAGATACCCKG – 3') - 1193R (5' – ACGTCATCCCCACCTTCC – 3') primer combination, which targets the V5-V7 region of the 16S rRNA gene, and was designed to minimize chloroplast amplification [43,44]. To further minimize host mitochondrial and chloroplast amplification, peptide nucleic acid (PNA) clamps were added to each reaction [45]. Resulting amplicons were diluted 8:1 and were further amplified for 9 cycles to add sample-specific barcodes, then quantified using Qubit, pooled in equal amounts, cleaned with magnetic beads and size selected via electrophoresis on a Pippin Prep gel (Sage Science, USA). The

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

resultant library was then sequenced on the Ilumina MiSeq (paired-end 300) platform.

Sequence Processing

Amplicon sequences were processed using the DADA2 pipeline [46] implemented in the R statistical environment [47], including the packages ShortRead [48], Biostrings [49], and Phyloseq [50]. Forward and reverse reads were truncated at 260 and 160 bp, respectively, and quality filtered using the function 'filterAndTrim' with default settings (i.e. maxN=0, maxEE=c(2,2), and truncQ=2). Error rates for forward and reverse reads were determined using the 'learnErrors' function, and then applied to remove sequencing errors from reads and assign them to amplicon sequence variants (ASVs) using the 'dada' function. Filtered paired reads were merged using the function 'mergePairs' and then converted into a sequence table using the 'makeSequenceTable' function. Chimeric sequences were removed from the sequence table using the function 'removeBimeraDeNovo' (method = consensus). Taxonomy was assigned to the remaining sequences using the 'assignTaxonomy' function, which implements the RDP Naïve Bayesian Classifier algorithm with kmer size 8 and 100 bootstrap replicates [51]. We used the Silva SSU taxonomic training dataset (version 138) formatted for DADA2 [52]. Chloroplast and mitochondrial sequences were filtered from the ASV table by removing any ASVs with a taxonomic assignment of 'Chloroplast' at the Order level or 'Mitochondria' at the Family level, respectively. Lastly, we applied the 'isContaminant' function (method = prevalence) from the package 'decontam' [53] to our samples using our

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

253 blank (no sample) DNA extractions to identify and remove putative contaminants
254 introduced during DNA extraction.
255
256 *Bacterial Quantification using Droplet Digital PCR (ddPCR)*
257 In order to estimate foliar bacterial abundances of each plant sample, droplet
258 digital PCR (ddPCR) using the Bio-Rad QX200 system on bacterial DNA extracted
259 from leaf wash. Comprehensive ddPCR methods are described elsewhere [54], but
260 briefly, we targeted the V5-V7 region of the 16S rRNA gene in sample DNA using the
261 chloroplast-excluding 799F (5' – AACMGGATTAGATACCCKG – 3') – 1389R (5' –
262 ACGGGCGGTGTGTRC – 3') primer combination. 5 µl of 1:10 diluted DNA template
263 were combined with 11 µl of 2X EvaGreen Supermix (Bio-Rad, USA) and 0.22 µl of
264 each primer, and 5.56 µl of molecular grade water to a total volume of 20 µl.
265 Reaction mixes were then loaded into the QX200 droplet generator with 70 ul of
266 droplet generation oil, then transferred to a PCR plate. 39 cycles of PCR were
267 performed under the following conditions: 95°C for 10 minutes, 95°C for 30
268 seconds, 55°C for 30 seconds, 72°C for 2 minutes, with steps 2-4 repeated 39 times,
269 4°C for 5 minutes, and 90°C for 5 minutes. EvaGreen signal was measured on the
270 QX200 droplet reader, cutoff thresholds were set for each column based on
271 background fluorescence in no template controls, and concentrations were
272 determined using the associated QuantaSoft software. Abundances are reported as
273 16S rRNA copies per g plant material as well as estimates of 16S copies per
274 individual plant by taking into account the proportion of the total plant sample that
275 was used for sample processing.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

276

277 *Statistical Analysis*

278 All statistical analyses were performed using R version 4.0.3 [47]. Community
279 matrices were rarefied to 6400 counts per sample ten times and averaged in order
280 to account for differences in sampling extent across samples. Bray Curtis bacterial
281 community dissimilarities were calculated between samples using the 'vegdist'
282 function in the vegan package in R [55]. Community structure differences among host
283 species identity, neighbor species identity, and experimental block were assessed
284 using a PERMANOVA on Bray-Curtis distances using the 'adonis' function (also in
285 the vegan package), which performs a sequential test of terms and uses the
286 algorithm presented in [56]. To assess the change in the relative strength of these
287 factors through time, the PERMANOVA was performed for each of the three
288 harvesting time points separately. Since not all samples successfully sequenced,
289 generating slight differences in sample numbers among harvests, we adjusted R^2
290 values to take into account sample numbers and degrees of freedom using the
291 'RsquareAdj' function in the vegan package. Indicator taxa analysis was performed
292 using the 'multipatt' function in the indicspecies package [57].

293 In order to assess the unique contribution of plant species identity for each
294 neighborhood type, we used variation partitioning on Hellinger-transformed
295 community matrices to partition out the effects of space. A geographic distance
296 matrix was calculated for all experimental plots based on plot GPS coordinates using
297 the program Geographic Distance Matrix Generator [58], and then pairwise distances
298 were decomposed into principal coordinates using the 'pcnm' function in the vegan

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

299 package. For each combination of plants, significant principal coordinates were
300 selected using forward and backward (direction = "both") model selection with the
301 'ordistep' function in the vegan package. The unique contribution of host species
302 identity was then calculated after partitioning out the significant spatial PCNMs that
303 were selected using the 'varpart' function in vegan. We further assessed the
304 contribution of other host factors including plant height and weight by performing
305 the above-mentioned model selection and variation partitioning. The statistical
306 significance of variation fractions was then tested by performing redundancy
307 analysis ordination (RDA), and declaring the non-focal factors as conditions.
308 Neutral modeling of phyllosphere communities was performed using the
309 'fit_sncm' function in the package reltools [59]. This package fits the neutral model
310 from [60], as implemented by [21]. In order to assess phylogenetic patterns in the
311 phyllosphere communities, we constructed a phylogenetic tree of all ASVs with
312 greater than 20 counts in the community matrix, which included 7949 ASVs.
313 Sequences were aligned using the 'AlignSeqs' function in the DECIPHER package [61]
314 using default settings. Next, pairwise distances between sequences were calculated
315 using the 'dist.ml' function in the phangorn package version 2.5.5 [62]. These distances
316 were then used to construct a neighbor-joining tree using the 'NJ' function in
317 phangorn. Lastly, the neighbor-joining tree was used as a starting point to create a
318 generalized time-reversible with gamma rate variation (GTR+G+I) maximum
319 likelihood tree using the 'pml', 'update', and 'optim.pml' functions in the phangorn
320 package. Lastly, we calculated the mean pairwise distance (MPD) of taxa in each
321 sample and compared to the MPD of a null model to calculate the standardized effect

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

size (SES) using the 'ses.mpd' function in the picante package [63]. We used a community randomization null model (null.model = species.pool, iterations=999) whereby a randomized community matrix is constructed by drawing from the total species pool with equal probability. Using such a procedure, a Z-score (the SES of MPD versus the null community) below zero can be interpreted as phylogenetically clustered whereby taxa co-occurring in a sample are more closely related than the same number of taxa drawn at random from the species pool. By contrast, samples above zero are interpreted as phylogenetically overdispersed, i.e. phylogenetic distance among co-occurring taxa is greater than the above-stated null expectation.

For univariate data such as ASV-level richness, MPD SES, and ddPCR-based abundance data, a three-way ANOVA was fit to test for significant effects of host, neighbor, and harvest time point, with interactions therein. The appropriateness of this procedure was verified by checking for a normal distribution of residuals on the model.

Results

Experimental manipulation of plant neighborhood in the field

We first compared the phyllosphere microbiome of plants that were surrounded by no neighbors, conspecific (same species) neighbors, or heterospecific (different species) neighbors. Because the field trial was conducted over the course of 3 months, with focal plants being replaced with a plant of the same species but at the original age of planting each month, we were also able to compare neighborhood age/biomass effects on microbiome assembly. After processing, 175 of the 216 focal

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

plant samples from the field yielded high quality sequencing reads (Tomato n = 63, Pepper n = 52, Bean n = 60). Of these 175, 64 were from harvest 1, 58 were from harvest 2, and 53 were from harvest 3. The 41 excluded samples each had less than 24 total reads and thus failed either to amplify or to yield sequences. The greenhouse control plants that were included to assess the bacterial communities established prior to transplantation into the field yielded very few reads (28 of the 45 samples had less than 50 reads), indicating that bacterial colonization prior to transplantation was minimal. Of the 17 greenhouse control samples with detectable sequences, communities were dominated by Enterobacterales, Corynebacterales, Burkholderiales, and Pseudomonadales.

The field trial dataset contained 5,414,393 observations of 19,818 ASVs, 13,455 of which had >10 occurrences, and 2,253 of which had >100 occurrences. Within-host ASV-level richness ranged from 22 to 769 ASVs across all treatments and hosts. Richness levels varied significantly by harvest time ($F_{2,174} = 24.21$, $p < 0.001$), declining throughout the season, and varying by host identity ($F_{2,174} = 4.96$, $p < 0.01$), with beans harboring a greater richness, especially at the first harvest. Neighborhood did not impact bacterial richness, however the total number of flowers on the neighborhood plant species at the time of focal plant harvest was positively correlated with bacterial richness on the focal plants ($R^2_{adj} = 0.044$, $p = 0.01$). Similar qualitative trends were observed for Shannon diversity, except that host plant weight and height were also positively correlated with diversity ($R^2_{adj} = 0.022$, $p = 0.03$ and $R^2_{adj} = 0.044$, $p < 0.01$, respectively).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bacterial abundance per plant varied significantly by host identity ($F_{2,174}$ = 27.35, $p < 0.001$), neighborhood ($F_{3,174} = 7.96$, $p < 0.001$), and harvest time point ($F_{2,174}$= 3.85, $p = 0.051$, Fig. 2A), with no significant interactions among variables. Tomato and bean plants tended to harbor higher bacterial abundances than pepper plants ($p < 0.001$). As expected, abundance per plant was positively correlated with plant weight ($R^2_{adj} = 0.187$, $p < 0.001$) and plant height ($R^2_{adj} = 0.117$, $p < 0.001$). If we normalize bacterial abundances by the plant material weight used for processing, we see weaker effects of host identity ($F_{2,174} = 3.24$, $p = 0.041$) and neighbor ($F_{3,174} = 2.17$, $p = 0.094$), but a stronger effect of harvest time point ($F_{2,174}$ = 5.15, $p = 0.024$). Several neighborhood attributes had interesting associations with bacterial abundance on focal hosts. Specifically, estimated neighborhood biomass ($R^2_{adj} = 0.146$, $p < 0.001$), average neighbor height ($R^2_{adj} = 0.123$, $p < 0.001$), and total number of flowers ($R^2_{adj} = 0.022$, $p = 0.038$) were all negatively correlated with bacterial abundance on focal plants. Lastly, bacterial abundance per focal plant was negatively associated with community richness ($R^2_{adj} = 0.016$, $p = 0.05$) and Shannon diversity ($R^2_{adj} = 0.03$, $p < 0.01$).

Overall, phyllosphere communities were dominated by the phyla Proteobacteria, Firmicutes, and Actinobacteriota. The most abundant bacterial orders were the Bacillales, Burkholderiales, Enterobacterales, Lactobacillales, Micrococcales, Rhizobiales, Sphingomonadales, and Xanthomonadales (Fig. 2B). Bray Curtis dissimilarities among samples were driven by harvest time point (R^2 = 0.063, $p = 0.003$), host species ($R^2 = 0.055$, $p = 0.003$), neighbor ($R^2 = 0.023$, p = 0.048), and block ($R^2 = 0.036$, $p = 0.06$), with a significant interaction between host

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and harvest (R^2 = 0.034, p = 0.033), and a trending interaction between neighbor and harvest (R^2 = 0.041, p = 0.072).

The effects of host identity on bacterial community composition decrease through time while neighborhood effects increase through time and vary by host identity

We next examined the relative influence of host species identity and neighborhood on focal plant microbiome structure at each time point during the field experiment by performing a PERMANOVA on Bray Curtis dissimilarities using host species identity (i.e. tomato, pepper, or bean), neighborhood (i.e. tomato, pepper, bean, or no neighbor), and experimental block (1-6) as independent variables. The effect of host identity was significant, but diminished in size over the three time points (Harvest 1: Adj. R^2 = 0.096, p < 0.001; Harvest 2: Adj. R^2 = 0.068, p < 0.001; Harvest 3: Adj. R^2 = 0.027, p < 0.001, Fig. 3A, see Table 1 for pre-adjusted R^2 values). In contrast, the effect of neighborhood status was initially not statistically significant, but increased in size over the three time points (Harvest 1: Adj. R^2 = -0.001, p = 0.242; Harvest 2: Adj. R^2 = 0.017, p < 0.01; Harvest 3: Adj. R^2 = 0.032, p < 0.001, Table 1, Fig 3A). Block effects tended to decrease throughout the experiment (Harvest 1: Adj. R^2 = 0.011, p = 0.021, Harvest 2: Adj. R^2 = 0.009, p = 0.028, Harvest 3: Adj. R^2 = 0.009, p = 0.07, Fig. 3A). No significant interactions among variables were observed at individual harvests.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: Results of a PERMANOVA on phyllosphere bacterial community Bray Curtis dissimilarities for harvest time points 1, 2, and 3 in the field trial. Variables tested include: host species identity (tomato, pepper, or bean), neighborhood (tomato, pepper, bean, no neighbor), and experimental block (1 through 6). R^2 values represent the fit of the model and adjusted R^2 values have been adjusted based on sample numbers of degrees of freedom to render values comparable across harvest time points.

Harvest 1

	df	*Pseudo-f*	*R2*	*Adj. R2*	*P*	*Significance*
Host Species	2	4.49	**0.125**	**0.096**	**0.001**	***
Neighborhood	3	1.11	0.046	-0.001	0.242	NS
Block	5	1.28	0.089	**0.011**	**0.021**	*

Harvest 2

	df	*Pseudo-f*	*R2*	*Adj. R2*	*P*	*Significance*
Host Species	2	3.22	**0.1**	**0.068**	**0.001**	***
Neighborhood	3	1.46	**0.069**	**0.017**	**0.001**	***
Block	5	1.23	**0.096**	**0.009**	**0.028**	*

Harvest 3

	df	*Pseudo-f*	*R2*	*Adj. R2*	*P*	*Significance*
Host Species	2	1.81	**0.064**	**0.027**	**0.001**	***
Neighborhood	3	1.66	**0.088**	**0.032**	**0.001**	***
Block	5	1.18	0.104	0.009	0.07	.

By excluding the 'no neighbor' controls, we then tested for an effect of neighbor identity by treating neighbor type (i.e. tomato, bean, or pepper) as an independent variable. On this subset of plants we see similar trends through time: host identity effects diminish (Harvest 1: Adj. $R^2 = 0.114$, $p < 0.001$; Harvest 2: Adj. $R^2 = 0.050$, $p < 0.001$; Harvest 3: Adj. $R^2 = 0.024$, $p = 0.007$) and neighbor identity effects increase (Harvest 1: Adj. $R^2 = -0.003$, $p = 0.325$; Harvest 2: Adj. $R^2 = 0.010$, $p = 0.031$; Harvest 3: Adj. $R^2 = 0.018$, $p = 0.013$). In this case a significant block effect

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

was only observed at time point 3 (Adj. $R^2 = 0.032$, $p = 0.013$), suggesting the block effect trend described above is influenced by the no neighbor controls.

We further asked whether a closer approximation of bacterial taxon absolute abundances might impact our conclusions. The relative abundance of each taxon in each sample was multiplied by the total 16S rRNA copies per 10 µl DNA (the same volume used for sequencing library preparation) to yield an estimate of each taxon's absolute abundance (quasi-absolute abundance). In this new dataset, sample dissimilarities were modeled using the same PERMANOVA procedure as above. This generated the same qualitative findings as the relative abundance data, but with slightly stronger effect sizes for influence of neighborhood plant species (Supp. Table 1). One new result revealed by this approach, however, was a host by neighborhood interaction, which increased from harvest 2 (Adj. $R^2 = 0.008$ $p = 0.002$) to harvest 3 (Adj. $R^2 = 0.024$, $p = 0.003$, see Supp. Table 1 for pre-adjusted R^2 values). In other words, the effect of neighborhood depended on the host's species identity, and this effect became stronger over time.

To further assess whether the three host plants species differed in their susceptibility to neighborhood effects, we subset the data by plant species and assessed the strength of neighborhood effects separately over the 3 time points. Similar to the combined data, no plant species exhibited a detectable neighborhood effect of microbiome composition at harvest 1. The tomato focal plants only exhibited a detectable neighborhood effect at harvest 3 ($R^2_{adj} = 0.086$, $P = 0.009$, Fig. 3B), and no block effects at any harvest. For the pepper focal plants, there was only a neighborhood effect at harvest 2 ($R^2_{adj} = 0.108$, $P = 0.013$, Fig. 3B). Lastly for the

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bean focal plants, a neighborhood effect was only detected at harvest 3 (R^2_{adj} = 0.069, P = 0.012, Fig. 3B). No significant block effects were observed for pepper or bean at any harvest.

Neighbor status and identity diminish effects of host species identity on phyllosphere bacterial communities

To test the hypothesis that focal plants with experimental neighbors would experience higher rates of inter-host dispersal than focal plants without nearby neighbors, we performed model selection on hosts that were distinguished by their neighbor status to identify the most explanatory host variables. We repeated this procedure for spatial variables generated by principle coordinates of neighborhood matrix analysis and then performed variation partitioning to partition out the unique contribution of host factors explaining diversity. We find supporting evidence for our hypothesis in harvests 2 and 3, but not harvest 1 (Fig. 4), indicating a dependency on the age structure of neighbors. For harvests 2 and 3, the unique contribution of host identity was stronger for the plants having no neighbors than the plants having either tomato, pepper, or bean as neighbors. In fact, at harvest 3 focal plants that were surrounded by bean or tomato neighbors had no detectable effect of host species identity after separating out the effect of space. Interestingly at harvest 1, focal plants with neighbors had higher effects of host species identity than the no neighbor controls, and this was especially the case for plants with tomato or bean neighbors. Interestingly in the focal plants surrounded by peppers, host species effects followed a hump-shaped relationship (increasing at first, then

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

474 decreasing), rather than the monotonic decrease observed for tomato- or bean-
475 surrounded plants. It thus appears that host-mediated selection of epiphytic
476 bacterial communities becomes subservient to the effect of immigrant inoculum
477 from neighboring plants as the biomass of neighboring plants increases.
478 Additionally, in certain cases, host identity combined with host height,
479 weight, or both height and weight in a way that increased explanatory power of host
480 factors. While in several instances this boosted the explanatory power of host
481 factors (e.g. at harvest 1), our qualitative conclusions remain the same (Supp. Fig. 1).
482 In other words, the effects of host identity are weaker for all plants that have
483 neighbors at both harvests 2 and 3.
484
485
486 *Phylogenetic clustering and neutral model fit vary by host and through time*
487 To better understand the predominance of deterministic processes in
488 shaping phyllosphere community membership and determine whether the three
489 plant species might be influenced by different assembly processes, we tested for
490 patterns of phylogenetic clustering. Evidence of phylogenetic clustering within a
491 host species would suggest that phylogenetically-conserved traits are being selected
492 for in a host-specific way [19,20]. We tested this idea using the standardized effect size
493 (SES) of the mean pairwise distance (MPD) of bacterial ASVs in each sample. MPD
494 SES was significantly influenced by host species identity ($F_{2,175} = 219.86$, $p < 0.001$),
495 harvest time point ($F_{2,175} = 12.33$, $p < 0.001$), a host by harvest interaction ($F_{4,175} =$
496 43.84, $p < 0.001$), and neighborhood ($F_{3,175} = 3.38$, $p < 0.05$, Fig. 5A). Tomato- and

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

497 pepper-associated communities were more phylogenetically clustered than would
498 be expected by chance (as indicated by negative SES values), and this was the case
499 for all three time points for both plants. In contrast, bean-associated communities
500 showed evidence of phylogenetic overdispersion (as indicated by highly positive
501 SES values). Bean SES values tended to decrease (i.e. tend towards clustering) over
502 time, but were highly variable. One exception was at harvest 3, where beans with no
503 neighbors had high variability in MPD SES, beans with conspecific neighbors were
504 phylogenetically clustered, and beans with tomato neighbors were overdispersed
505 (Tukey's HSD conspecific neighbor vs tomato neighbor $p = 0.02$).

506 We next asked how well the occupancy-abundance relationships within each
507 host species could be fit by a neutral model, whereby passive dispersal and
508 ecological drift are the primary drivers of establishment, and then asked whether
509 the fit to that model changed over time. Of the three hosts, bean-associated
510 communities had the highest goodness-of-fit values followed by tomato-associated
511 communities, suggesting differences among hosts in the role of neutral processes in
512 shaping community structure (Fig 5B). Both bean and tomato hosts showed a
513 decline in the fit of a neutral model from harvest 1 to 2 (Bean: harvest 1 R^2=0.483,
514 harvest 2 R^2=0.074; Tomato: harvest 1 R^2=0.214, harvest 2 R^2=0.052), and the
515 neutral model failed to fit either set of plants for harvest 3 (as indicated by a
516 negative goodness-of-fit). At all three time points, pepper hosts were never fit by a
517 neutral model (indicated by negative goodness-of-fit values).
518

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

519 *Experimental greenhouse transplantation of field study-derived inocula replicates host*
520 *filtering and reveals effects of inoculum source*

521 The subsequent greenhouse study allowed us to more closely examine the
522 effects of inter-host dispersal of bacterial taxa, as phyllosphere bacterial
523 communities that were recovered from the field were transplanted onto either the
524 same plant species from which they were collected, or onto the plant species that
525 had previously surrounded the source plant. From the 105 total such reciprocal
526 inoculations, 103 samples yielded sufficient high quality sequences for analysis. The
527 resultant dataset contained 2,640,588 observations of 1734 ASVs, 1379 of which
528 had greater than 10 observations, and 621 of which had over 100 observations.

529 We observed a linear relationship between the ASV-level richness of the
530 sample from which the inoculum was derived and the number of inoculum ASVs
531 that were detectable in the greenhouse samples (R^2_{adj} = 0.105, p = 0.004, Supp. Fig
532 3). The number of overlapping ASVs between the inoculum and the experimental
533 plants was significantly related to the host species identity ($F_{2,88}$ = 8.503, p < 0.001),
534 the previous host species from which the inoculum was sampled ($F_{2,88}$ =3.871, p
535 =0.028), and interactions between the host and previous host ($F_{4,88}$ = 2.598,
536 p=0.049) as well as between the host and previous neighbor ($F_{4,88}$ =2.968 p=0.03).
537 Similar to the field study, phyllosphere communities were dominated by the phyla
538 Proteobacteria and Firmicutes. The most abundant bacterial orders were the
539 Bacillales, Burkholderiales, Enterobacterales, and Pseudomonadales (Supp. Fig. 2).
540 The bacterial community structure on treated plants differed significantly from that

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of control plants to which only sterile buffer had been applied (PERMANOVA R^2 = 0.0156, p = 0.048).

We used the heat-killed inoculum as a control to gain insights into the level of host selection and subsequent inoculum establishment across our experimental plants. To do so, we asked how strongly phyllosphere microbiomes were differentiated by host species identity. Plants that received a heat-killed inoculum were less differentiable than plants that received a live inoculum (PERMANOVA Adj. R^2 = 0.100, p = 0.04 versus Adj. R^2 = 0.103, p < 0.001, respectively). The plants that received the sterile buffer as a control were even less differentiable by host species identity (Adj. R^2 = 0.056, p = 0.07). Interestingly, if we subset samples based on whether experimental plants received an inoculum from heterospecific (different species) or conspecific (same species) hosts, we see that heterospecific transplants resulted in more differentiable hosts than conspecific transplants (Adj. R^2 = 0.123, p < 0.001 vs Adj. R^2 = 0.109, p < 0.001, respectively). It thus appears that the treatment plants receiving live cells more efficiently filtered communities, driving differentiation of host species, and that heterospecific inoculum sources further bolstered host differentiation.

Indicator taxon analysis allowed us to examine the taxa enriched on each of the three host species in the field trial (at harvest 2) and the greenhouse trial. Of the 34 taxa that distinguished pepper plants in the field, 6 were found in the greenhouse dataset (Supp. Table 2). The collective relative abundance of these taxa was significantly higher on pepper plants in the greenhouse than on tomatoes (Tukey's HSD p < 0.001) or beans (Tukey's HSD p < 0.001). Of the 65 taxa that distinguished

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

564 tomato hosts in the field, 25 were detected in the greenhouse dataset (Supp. Table
565 2). The collective relative abundance of these taxa was significantly higher on
566 tomato and pepper plants relative to beans (Tukey's HSD $p < 0.05$ and $p < 0.01$,
567 respectively). Of the 300 taxa that distinguished bean, 27 were found in the
568 greenhouse (Supp. Table 2), but their collective relative abundance was not
569 significantly different among hosts ($p > 0.05$).
570
571 *The effect of donor plant biomass on recipient plant phyllosphere richness in the*
572 *greenhouse depends on the origin of field inoculum*
573 Within-host ASV-level richness of treatment plants ranged from 24 to 231
574 ASVs and varied significantly by host species identity ($F_{2,74} = 12.41$, $p < 0.001$), an
575 interaction between host and previous host identities ($F_{4,74} = 3.50$, $p < 0.05$), and
576 experimental block ($F_{4,74} = 2.43$, $p = 0.058$). Of the three plant species, peppers
577 harbored significantly higher richness than tomatoes or beans ($p < 0.001$), which
578 were indistinguishable from each other ($p > 0.05$). When the inoculation was a
579 conspecific transfer (i.e. moving between two plants of the same species), a negative
580 but weak relationship was observed between donor plant biomass and recipient
581 plant richness (Adj. $R^2 = 0.09$, $p < 0.05$, Fig. 6A). However, when the inoculation was
582 a heterospecific transfer (i.e. between two different plant species), a positive and
583 stronger relationship was observed between donor plant biomass and recipient
584 plant richness (Adj. $R^2 = 0.21$, $p < 0.01$, Fig. 6B). No significant differences in richness
585 were observed between conspecific transplants and heterospecific transplants ($p >$
586 0.05).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

587

588 *Experimental transplantation of phyllosphere communities reveals an influence of*
589 *current and previous host identity, as well as donor plant neighbor*

590 PERMANOVA on Bray Curtis dissimilarities revealed that host species
591 identity, previous host identity (i.e. the species from which inoculum was derived),
592 and experimental block all significantly contributed to phyllosphere community
593 structure differences among experimental greenhouse plants (Host species R^2 =
594 0.127, p <0.001, Previous host species R^2 = 0.0468, p <0.001, Block R^2 = 0.156, p
595 <0.001, Table 2). We also observed an effect of donor plant biomass on recipient
596 plant phyllosphere community structure (R^2 = 0.018, p = 0.05, Table 2). Moreover, if
597 we interrogate the dataset by plant species, we see a "grandparent effect" in the
598 phyllosphere community structure of greenhouse-grown pepper plants. That is,
599 pepper phyllosphere communities were most strongly shaped by the previous
600 host's neighbor (R^2 = 0.194, p < 0.001), i.e. the plant species that previously
601 surrounded the donor plant. The previous host's species identity was also
602 significantly associated with pepper plant phyllosphere community structure, albeit
603 to a lesser extent (R^2 = 0.107, p = 0.061). In contrast to what was observed on
604 pepper, neither previous neighbor effects nor previous host effects were observed
605 for tomato or bean subsets.

606

607

608

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

609 **Table 2:** Results of a PERMANOVA on phyllosphere bacterial community Bray
610 Curtis dissimilarities for the greenhouse experiment. Variables tested include: host
611 species identity (tomato, pepper, or bean), previous host species identity (tomato,
612 pepper, bean), previous neighbor (tomato, pepper, bean), previous host biomass (g),
613 and experimental block (1 through 5). R^2 values represent the fit of the model.

Greenhouse Trial

	df	*Pseudo-f*	*R2*	*P*	*Significance*
Host Species	2	6.21	**0.127**	**0.001**	***
Previous Host Species	2	2.28	**0.046**	**0.001**	***
Previous Neighbor Species	2	0.98	0.02	0.478	NS
Previous Host Biomass	1	1.72	**0.018**	**0.05**	*
Block	4	3.92	**0.156**	**0.001**	***

614

615

616 ***Discussion***

617 Plant-microbe associations form in part through host filtering of microbiota
618 that arrive via dispersal [64]. Our study experimentally manipulated neighbor
619 presence, identity, and age in order to understand how these factors influence host
620 filtering of phyllosphere communities. Over the course of the experiment, we found
621 that host species identity effects on focal plant microbiomes decreased, while the
622 effects of neighborhood increased (Fig 3A). This finding builds on past studies
623 showing that host species- or genotype-level differences in microbiota change over
624 the growing season (e.g. Wagner et al. 2016; Chaparro, Badri, and Vivanco 2014;
625 İnceoğlu et al. 2011). However, an important distinction is that we experimentally
626 held host developmental stage constant throughout the experiment, thereby
627 demonstrating that changes in host species identity effects over time are not simply

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

due to host ontogeny, but hinge on characteristics of the neighboring plant community such as neighbor identity and biomass.

The increasing strength of neighborhood effects throughout the field experiment suggests that as neighboring plants grow and enrich for host species-specific microbes, they become larger sources of microbial propagules to their surrounding neighborhood, and thus alter the outcome of host filtering through compositional changes to the local species pool. Two other lines of evidence from our study further underscore the importance of neighbor identity for phyllosphere community assembly. First, when we directly control the directionality of dispersal in our greenhouse microbial transplant study, we see that the source of inoculum (i.e. the species identity of the donor plant) significantly contributed to the microbiome composition of recipient plants (Table 2). Second, the field experiment uncovered strong differences in host species identity effects depending on the identity of neighbors (Fig. 4). For instance at harvests 2 and 3, hosts that were surrounded by tomato or bean neighbors were substantially less differentiable in their phyllosphere community structure than plants surrounded by pepper or that had no neighbors. This suggests firstly that having a neighbor impacts the differentiation of hosts, but crucially that the outcome of neighborhood effects depends on neighbor identity. Interestingly, it has also been reported that inter-host dispersal among zebrafish greatly diminished genotype-level microbiome differences [42]. Our results not only reinforce this concept in plants, but suggest that this effect depends on neighbor identity.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

650 One of the key differences among experimental neighborhoods was plant
651 biomass, which varied by species (Fig. 1C). This is likely an important driver of the
652 neighborhood effects observed in this study. Larger biomass plants could harbor
653 higher abundances of microorganisms by virtue of having more microbial habitat,
654 thereby increasing the load of propagules dispersing onto focal plants. While we
655 chose not to sample the epiphytic communities of the neighborhood plants in order
656 to leave them undisturbed for the duration of the experiment, we observed a
657 positive correlation between focal host biomass and epiphytic bacterial abundance.
658 Thus, larger biomass neighborhood could have diminished the strength of host
659 filtering through mass effects, i.e. rescuing via dispersal the taxa that went locally
660 extinct due to host selection. We see evidence for the importance of neighbor
661 biomass in several results. First, as the higher-biomass tomato and bean
662 neighborhoods grew, we see that focal plant species identity effects became weaker,
663 so much so that by harvest 3 hosts were indistinguishable by their species identity if
664 they were surrounded by tomato or bean (Fig. 4). Interestingly, at harvest 1, relative
665 to no neighbor controls, focal hosts surrounded by tomatoes or beans exhibited
666 higher species identity effects, suggesting that at an early stage, neighbor plants may
667 bolster host filtering by providing higher abundance and/or diversity of propagules.
668 Moreover, the effects of the smaller pepper neighborhoods followed a similar but
669 lagged trend, whereby host differentiation was highest at harvest time point 2,
670 followed by more diminished host species effects at harvest 3. This could be driven
671 by the observed slower growth of peppers relative to beans or tomatoes. In the
672 greenhouse study, we see that the biomass of the donor plants significantly

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

673 contributed to variation in both community composition (Table 2) and diversity
674 (Fig. 6) of recipient plants. Here, the relationship between recipient plant
675 phyllosphere richness and donor plant biomass was negative and weak if the
676 transplant was conspecific (same species) but stronger and positive if it was
677 heterospecific (between different species). Together these results underscore that
678 biomass is not only an important component of neighborhood effects, but that such
679 influences of biomass may depend at least partly on the identity of the neighbors.
680 Our study also makes clear that differences in the strength of host filtering
681 across species may impact susceptibility to neighborhood effects. We found that
682 plant species exhibited neighborhood effects differently through time (Fig. 3B).
683 Together with the observation of a host-by-neighborhood interaction effect, these
684 results suggest that the relative impact of local neighborhood differs among focal
685 host species, perhaps due to differences in the degree to which dispersal versus host
686 filtering influence phyllosphere assembly. For instance, host-specific carrying
687 capacities could be driving the observed differences in bacterial abundances across
688 species (Fig. 2A). This could mean that species with lower abundances of bacteria
689 (e.g. peppers) are more invasible, and hence taxa that are selected for may more
690 quickly become outnumbered by immigrating taxa. This may explain why pepper
691 plants exhibited neighborhood effects at harvest 2, while tomato and beans did not
692 do so until harvest 3. Several lines of evidence also suggest that the plants may differ
693 in their selective abilities. For instance, tomato- and pepper-associated communities
694 consistently exhibited phylogenetic clustering (Fig. 5A), indicating that closely
695 related taxa were often observed to co-occur on a single plant, perhaps due to host

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

696 selection of shared traits. In contrast, bean microbiomes tended to be
697 phylogenetically overdispersed, and hence exhibited much less clustering than one
698 would expect by chance. Overdispersion could indicate that traits under selection
699 are not phylogenetically conserved, that competition is strong (e.g. between
700 ecologically similar taxa), or that selection is relatively weak [19]. That we also see a
701 strong fit of bean microbiomes to neutral models at harvests 1 and 2 (Fig. 5B)
702 suggests that the bean phyllosphere may be less selective relative to the tomato and
703 pepper phyllosphere, and therefore more susceptible to dispersal effects.
704 Interestingly in the greenhouse experiment, only pepper plant microbiomes
705 exhibited a "grandparent effect" of inoculum, i.e. an effect of the donor plants'
706 previous neighbor. This result demonstrates that for certain species, microbiome
707 composition not only reflects its contemporary host and its source history, but also
708 its previous dispersal history. This is analogous to a child inheriting a parent's
709 microbiome that carries with it traces of the parent's former house, pet, or domestic
710 partner. Results from the field trial may help shed light on why only pepper
711 microbiomes contained detectable traces of dispersal history. The aforementioned
712 patterns of phylogenetic clustering in pepper plant communities and the
713 observation that neutral models failed to fit pepper plants (Fig. 5A,B) together
714 suggest that the pepper phyllosphere may impose particularly strong selection on
715 microbial communities. This strong filtering ability may not only select against taxa,
716 but it could act to amplify taxa that previously dispersed from pepper neighbors,
717 thus giving rise to effects of previous neighbor. In other words, selective plants such

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

718 as pepper may bolster pepper-associated microorganisms upon arrival, even if they
719 have become rare through multiple dispersal events.

720 One major implication of our work is that it highlights the potential
721 importance of local host frequencies in recruiting microbiota. At harvest 3, bean
722 phyllosphere communities exhibited phylogenetic clustering only when the plants
723 were surrounded by conspecific (same species) neighbors. As phylogenetic
724 clustering could be interpreted as host filtering for phylogenetically-conserved
725 traits, this result suggests that for beans in particular, the frequency of conspecifics
726 in the plant metacommunity may be an important determinant of host filtering
727 efficacy. Interestingly, for pepper- or tomato-associated communities, the con- or
728 hetero-specific status of neighborhoods had little influence over phylogenetic
729 clustering, suggesting that this effect may depend on the strength of host filters.
730 Similar findings have recently been reported for *Acer saccharum* trees [38], where the
731 abundance of conspecific trees in the local metacommunity was positively
732 correlated with the degree of host specialization in the phyllosphere, and
733 importantly, this was not the case for all tree species surveyed. Moreover in cacao
734 trees, leaf litter of healthy conspecific hosts was shown to protect against pathogen
735 damage [67]. While we did not test the fitness effects of host filtering for specialized
736 microbial taxa, our results alongside several others may challenge the conspecific
737 negative density-dependence (i.e. Janzen-Connell) hypothesis that posits that higher
738 local densities of conspecifics may be disadvantageous due to the possibility of
739 shared pests or pathogens [29,30]. While there remain many examples of conspecific
740 negative density-dependence [68], particularly in the tropics, meta-analyses seeking

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

741 general trends have generated mixed results [69,70]. Our work emphasizes that
742 recruitment of specialized taxa from nearby conspecific hosts could outweigh the
743 negative effects of pest pressure in certain contexts, or for certain host species.
744 Overall, our work makes clear that local neighborhood identity and biomass
745 are key components that shape assembly of the phyllosphere microbiome. In both
746 the field trial and greenhouse experiment, we find that although plants are able to
747 select upon their microbial communities, the outcome of this selection is shaped by
748 both neighbor identity and local biomass. Moving forward, this work has opened a
749 number of critical questions regarding how neighborhood effects on the plant
750 microbiome might shape plant health, fitness, and – in agricultural settings
751 especially – yield. The work also raises questions about how invasive plant species
752 might alter microbial dispersal within their communities, and potentially negatively
753 feedback on native plant species' fitness by reducing their ability to filter the
754 optimal microbiome. In sum, our work demonstrates that host filtering and local
755 dispersal are intimately intertwined and represent crucial considerations for the
756 study of host-microbe associations.
757

758 ***Acknowledgements***

759 We acknowledge that this work was conducted on the territory of xučyun
760 (Huichin), the ancestral and unceded land of the Chochenyo speaking Ohlone
761 people; therefore this work would not have been possible were it not for their past
762 land stewardship. We thank the staff of UC-Berkeley's Oxford Tract Greenhouse for
763 their assistance in maintaining the experiment. Thank you to R. Koutsoukis, T. Caro,

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

X. Zhang, R. Debray, C. Hernandez, and E. Mehlferber for assistance with initial planting, and to the Koskella lab for invaluable feedback throughout.

References

1 Lindow SE, Brandl MT. Microbiology of the Phyllosphere. *Appl Environ Microbiol* 2003; **69**: 1875 LP-1883.

2 Morella NM, Zhang X, Koskella B. Tomato Seed-Associated Bacteria Confer Protection of Seedlings Against Foliar Disease Caused by Pseudomonas syringae. *Phytobiomes J* 2019; **3**: 177–190.

3 Innerebner G, Knief C, Vorholt JA. Protection of Arabidopsis thaliana against leaf-pathogenic Pseudomonas syringae by Sphingomonas strains in a sontrolled model system. *Appl Environ Microbiol* 2011; **77**: 3202–3210.

4 Fu S-F, Sun P-F, Lu H-Y, Wei J-Y, Xiao H-S, Fang W-T *et al.* Plant growth-promoting traits of yeasts isolated from the phyllosphere and rhizosphere of Drosera spatulata Lab. *Fungal Biol* 2016; **120**: 433–448.

5 Laforest-Lapointe I, Paquette A, Messier C, Kembel SW. Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. *Nature* 2017; **546**: 145–147.

6 Lindow SE, Leveau JHJ. Phyllosphere microbiology. *Curr Opin Biotechnol* 2002; **13**: 238–243.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

787 7 Fürnkranz M, Wanek W, Richter A, Abell G, Rasche F, Sessitsch A. Nitrogen
788 fixation by phyllosphere bacteria associated with higher plants and their
789 colonizing epiphytes of a tropical lowland rainforest of Costa Rica. *ISME J*
790 2008; **2**: 561–570.
791 8 Ottesen AR, Gorham S, Reed E, Newell MJ, Ramachandran P, Canida T *et al.*
792 Using a Control to Better Understand Phyllosphere Microbiota. *PLoS One*
793 2016; **11**: e0163482.
794 9 Jones JDG, Dangl JL. The plant immune system. *Nature* 2006; **444**: 323–329.
795 10 Bodenhausen N, Bortfeld-miller M, Ackermann M, Vorholt JA. A Synthetic
796 Community Approach Reveals Plant Genotypes Affecting the Phyllosphere
797 Microbiota. *PLOS Biol* 2014; **10**. doi:10.1371/journal.pgen.1004283.
798 11 Horton MW, Bodenhausen N, Beilsmith K, Meng D, Muegge BD, Subramanian S
799 *et al.* Genome-wide association study of Arabidopsis thaliana leaf microbial
800 community. *Nat Commun* 2014; **5**: 5320.
801 12 Hacquard S, Spaepen S, Garrido-Oter R, Schulze-Lefert P. Interplay Between
802 Innate Immunity and the Plant Microbiota. *Annu Rev Phytopathol* 2017; **55**:
803 565–589.
804 13 Zhalnina K, Louie KB, Hao Z, Mansoori N, Nunes U, Shi S *et al.* Dynamic root
805 exudate chemistry and substrate preferences drive patterns in rhizosphere
806 microbial community assembly. *Nat Microbiol* 2018. doi:10.1038/s41564-
807 018-0129-3.
808 14 Humphrey PT, Whiteman NK. Insect herbivory reshapes a native leaf
809 microbiome. *Nat Ecol Evol* 2020; **4**: 221–229.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

15 Yadav RKP, Karamanoli K, Vokou D. Bacterial Colonization of the Phyllosphere of Mediterranean Perennial Species as Influenced by Leaf Structural and Chemical Features. *Microb Ecol* 2005; **50**: 185–196.

16 Morella NM, Weng FCH, Joubert PM, Metcalf CJE, Lindow S, Koskella B. Successive passaging of a plant-associated microbiome reveals robust habitat and host genotype-dependent selection. *Proc Natl Acad Sci U S A* 2020; **117**: 1148–1159.

17 Wagner MR, Busby PE, Balint-Kurti P. Analysis of leaf microbiome composition of near-isogenic maize lines differing in broad-spectrum disease resistance. *New Phytol* 2019; **225**: 2152–2165.

18 Wagner MR, Lundberg DS, Del Rio TG, Tringe SG, Dangl JL, Mitchell-Olds T. Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nat Commun* 2016; **7**: 1–15.

19 Horner-Devine MC, Bohannan BJM. Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* 2006; **87**: S100-8.

20 Kembel SW, O'Connor TK, Arnold HK, Hubbell SP, Wright SJ, Green JL. Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. *Proc Natl Acad Sci* 2014; : 1–6.

21 Burns AR, Stephens WZ, Stagaman K, Wong S, Rawls JF, Guillemin K *et al.* Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J* 2016; **10**: 655–664.

22 Sloan WT, Woodcock S, Lunn M, Head IM, Curtis TP. Modeling Taxa-

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abundance Distributions in Microbial Communities using Environmental Sequence Data. *Microb Ecol* 2007; **53**: 443–455.

23 Laforest-Lapointe I, Messier C, Kembel SW. Host species identity, site and time drive temperate tree phyllosphere bacterial community structure. *Microbiome* 2016; : 1–10.

24 Schlaeppi K, Dombrowski N, Oter RG, Ver Loren van Themaat E, Schulze-Lefert P. Quantitative divergence of the bacterial root microbiota in Arabidopsis thaliana relatives. *Proc Natl Acad Sci* 2014; **111**: 585 LP-592.

25 Gallart M, Adair KL, Love J, Meason DF, Clinton PW, Xue J *et al.* Host Genotype and Nitrogen Form Shape the Root Microbiome of Pinus radiata. *Microb Ecol* 2018; **75**: 419–433.

26 Hambäck PA, Inouye BD, Andersson P, Underwood N. Effects of plant neighborhoods on plant–herbivore interactions: resource dilution and associational effects. *Ecology* 2014; **95**: 1370–1383.

27 Underwood N, Inouye BD, Hambäck PA. A Conceptual Framework for Associational Effects: When Do Neighbors Matter and How Would We Know? *Q Rev Biol* 2014; **89**: 1–19.

28 Barbosa P, Hines J, Kaplan I, Martinson H, Szczepaniec A, Szendrei Z. Associational Resistance and Associational Susceptibility: Having Right or Wrong Neighbors. *Annu Rev Ecol Evol Syst* 2009; **40**: 1–20.

29 Janzen DH. Herbivores and the Number of Tree Species in Tropical Forests. *Am Nat* 1970; **104**: 501–528.

30 Connell JH. On the role of natural enemies in preventing competitive exclusion

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in some marine animals and in rain forest trees. In: Den Boer PJ, Gradwell G (eds). *Dynamics of populations*. PUDOC, 1971, pp 298–312.

31 Mangan SA, Schnitzer SA, Herre EA, Mack KML, Valencia MC, Sanchez EI *et al.* Negative plant–soil feedback predicts tree-species relative abundance in a tropical forest. *Nature* 2010; **466**: 752–755.

32 Miller EC, Perron GG, Collins CD. Plant-driven changes in soil microbial communities influence seed germination through negative feedbacks. *Ecol Evol* 2019; **0**: 1–14.

33 Antonovics J, Ellstrand NC. Experimental studies of the evolutionary significance of sexual reproduction. I. A test of the frequency-dependent selection hypothesis. *Evolution (N Y)* 1984; **38**: 103–115.

34 Ellstrand NC, Antonovics J. Experimental studies of the evolutionary significance of sexual reproduction II. A test of the density-dependent selection hypothesis. *Evolution (N Y)* 1985; **39**: 657–666.

35 Naeem S, Tjossem SF, Byers D, Bristow C, Li S. Plant neighborhood diversity and production. *Ecoscience* 1999; **6**: 355–365.

36 Worrich A, Musat N. Associational effects in the microbial neighborhood. *ISME J* 2019; : 2143–2149.

37 Copeland JK, Yuan L, Layeghifard M, Wang PW, Guttman DS. Seasonal Community Succession of the Phyllosphere Microbiome. *Mol Plant-Microbe Interact* 2015; **28**: 274–285.

38 Lajoie G, Kembel SW. Host neighborhood shapes bacterial community assembly and specialization on tree species across a latitudinal gradient. *Ecol*

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Monogr 2021; **0**: 1–18.

39 Lymperopoulou D, Adams R, Lindow SE, Löffler F. Contribution of Vegetation to the Microbial Composition of Nearby Outdoor Air. *Appl Environ Microbiol* 2016; **82**: 3822–3833.

40 Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF *et al.* The metacommunity concept: a framework for multi-scale community ecology. *Ecol Lett* 2004; **7**: 601–613.

41 Fodelianakis S, Lorz A, Valenzuela-cuevas A, Barozzi A, Booth JM, Daffonchio D. Dispersal homogenizes communities via immigration even at low rates in a simplified synthetic bacterial metacommunity. *Nat Commun* 2019; **10**: 1–12.

42 Burns AR, Miller E, Agarwal M, Rolig AS, Milligan-Myhre K, Seredick S *et al.* Interhost dispersal alters microbiome assembly and can overwhelm host innate immunity in an experimental zebrafish model. *Proc Natl Acad Sci* 2017; **114**. doi:10.1073/pnas.1702511114.

43 Chelius MK, Triplett EW. The Diversity of Archaea and Bacteria in Association with the Roots of Zea mays L. *Microb Ecol* 2001; **41**: 252–263.

44 Bodenhausen N, Horton MW, Bergelson J. Bacterial Communities Associated with the Leaves and the Roots of Arabidopsis thaliana. *PLoS One* 2013; **8**: e56329.

45 Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangl JL. Practical innovations for high-throughput amplicon sequencing. *Nat Methods* 2013; **10**: 999–1002.

46 Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

902 DADA2: High-resolution sample inference from Illumina amplicon data. *Nat*
903 *Methods* 2016; **13**: 581.
904 47 R Core Team. R: A language and environment for statistical computing.
905 2020.http://cran.r-project.org.
906 48 Morgan M, Anders S, Lawrence M, Aboyoun P, Pagès H, Gentleman R.
907 ShortRead: a bioconductor package for input, quality assessment and
908 exploration of high-throughput sequence data. *Bioinformatics* 2009; **25**:
909 2607–2608.
910 49 Pagès H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: Efficient
911 manipulation of biological strings. 2020.
912 50 McMurdie P, Holmes S. phyloseq: An R Package for Reproducible Interactive
913 Analysis and Graphics of Microbiome Census Data. *PLoS One* 2013;
914 **8(4)**.https://doi.org/10.1371/journal.pone.0061217tle.
915 51 Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian Classifier for Rapid
916 Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl*
917 *Environ Microbiol* 2007; **73**: 5261–5267.
918 52 Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P *et al.* The SILVA
919 ribosomal RNA gene database project: improved data processing and web-
920 based tools. *Nucleic Acids Res* 2013; **41**: D590–D596.
921 53 Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical
922 identification and removal of contaminant sequences in marker-gene and
923 metagenomics data. *Microbiome* 2018; **6**: 226.
924 54 Morella NM, Gomez AL, Wang G, Leung MS, Koskella B. The impact of

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bacteriophages on phyllosphere bacterial abundance and composition. *Mol Ecol* 2018; **27**: 2025–2038.

55 Oksanen J, Blanchet FG, Roeland K, Legendre P, Minchin P, O'Hara RB *et al.* vegan: Community ecology package. 2015.http://cran.r-project.org.

56 Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 2001; **26**: 32–46.

57 Cáceres M De, Legendre P. Associations between species and groups of sites: indices and statistical inference. *Ecology* 2009; **90**: 3566–3574.

58 Ersts PJ. Geographic Distance Matrix Generator. http://biodiversityinformatics.amnh.org/open_source/gdmg.

59 Sprockett D. reltools: Microbiome Amplicon Analysis and Visualization. 2021.

60 Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* 2006; **8**: 732–740.

61 Wright ES. Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *R J* 2016; **8**: 352–359.

62 Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics* 2011; **27**: 592–593.

63 Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD *et al.* Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 2010; **26**: 1463–1464.

64 Koskella B. The phyllosphere. *Curr Biol* 2020; **30**: R1143–R1146.

65 Chaparro JM, Badri D V, Vivanco JM. Rhizosphere microbiome assemblage is

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

948 affected by plant development. *ISME J* 2014; **8**: 790–803.

949 66 İnceoğlu Ö, Al-Soud WA, Salles JF, Semenov A V, van Elsas JD. Comparative
950 Analysis of Bacterial Communities in a Potato Field as Determined by
951 Pyrosequencing. *PLoS One* 2011; **6**: e23321.

952 67 Christian N, Herre EA, Mejia LC, Clay K. Exposure to the leaf litter microbiome
953 of healthy adults protects seedlings from pathogen damage. *Proc R Soc B Biol*
954 *Sci* 2017; **284**: 20170641.

955 68 Leigh EG, Davidar P, Dick CW, Terborgh J, Puyravaud J-P, ter Steege H *et al.*
956 Why Do Some Tropical Forests Have So Many Species of Trees? *Biotropica*
957 2004; **36**: 447–473.

958 69 Hyatt LA, Rosenberg MS, Howard TG, Bole G, Fang W, Anastasia J *et al.* The
959 distance dependence prediction of the Janzen-Connell hypothesis: a meta-
960 analysis. *Oikos* 2003; **103**: 590–602.

961 70 Carson W, Anderson J, Leigh E, Schnitzer S. Challenges Associated With
962 Testing And Falsifying The Janzen_Connell Hypothesis: A Review and Critique.
963 In: Carson W, Schnitzer SA (eds). *Tropical Forest Community Ecology*. Wiley
964 Blackwell, 2008, pp 210–241.

965

966

967

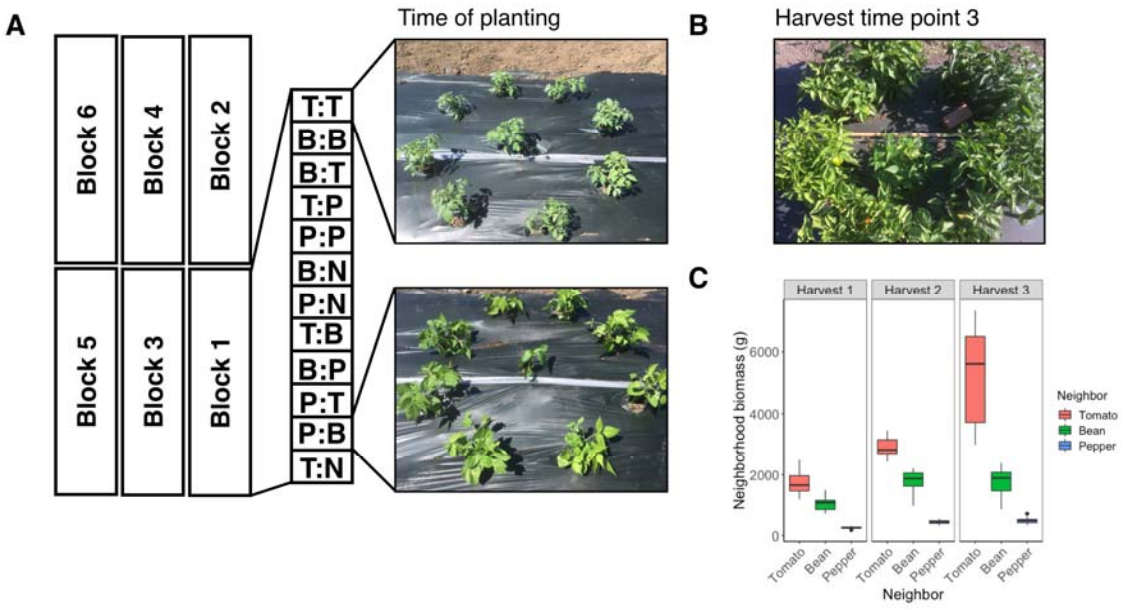
968

969

970

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

971 **Figure legends**



972

973 **Fig. 1**: Experimental design of the field trial. A) Experimental neighborhoods were
974 constructed by planting a focal plant in the center of a ring of neighbor plants, with
975 fully factorial combinations of focal and neighbor plant species. Each block
976 contained 12 comparisons, with focal and neighbor plant abbreviated (T=tomato,
977 P=pepper, B=bean, and N=no neighbor), respectively. Focal plants were harvested
978 and replaced each month, while the neighborhoods were left to continue growing.
979 Shown are a tomato focal plant with tomato neighbors (top) and a pepper focal
980 plant with bean neighbors (bottom) at the time of initial planting. B) A pepper
981 neighborhood surrounding a bean focal plant at the time of harvest number 3. C)
982 The biomass of neighborhood plants (g) increased to varying degrees with time for
983 each plant species by harvests 1, 2, and 3.

984

985

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

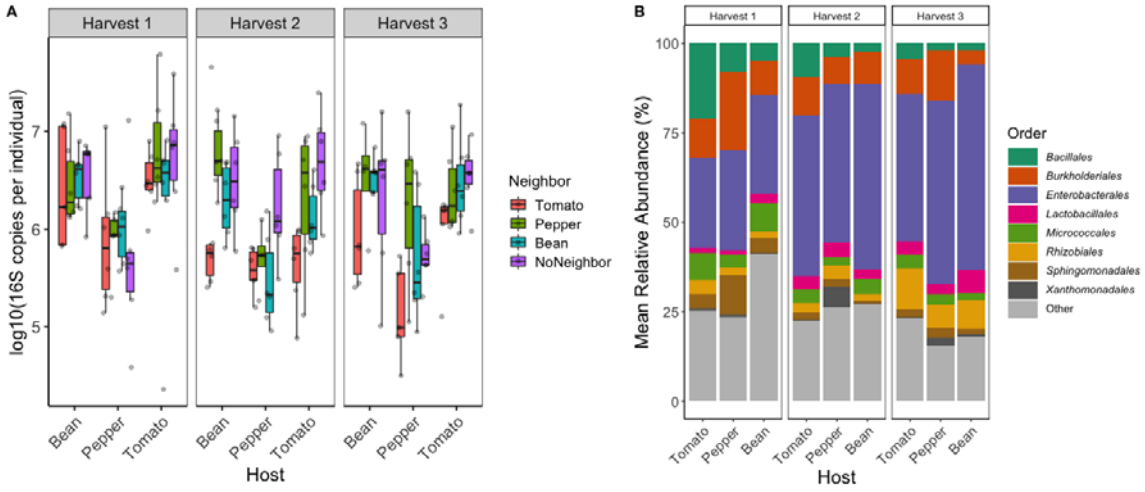


Fig. 2: Bacterial abundance and composition vary across host species and harvest time. A) The abundance (log_{10} 16S rRNA gene copies measured using ddPCR of leaf washes, y-axis) for individual focal plant species (x-axis) surrounded by different neighbor plant species (box color) and at different successive harvest times (panels 1, 2, and 3). B) Relative abundance of the 9 most abundant bacterial orders distinguished by host species and harvest time. All other less abundant or ambiguously assigned orders are grouped under 'Other'.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

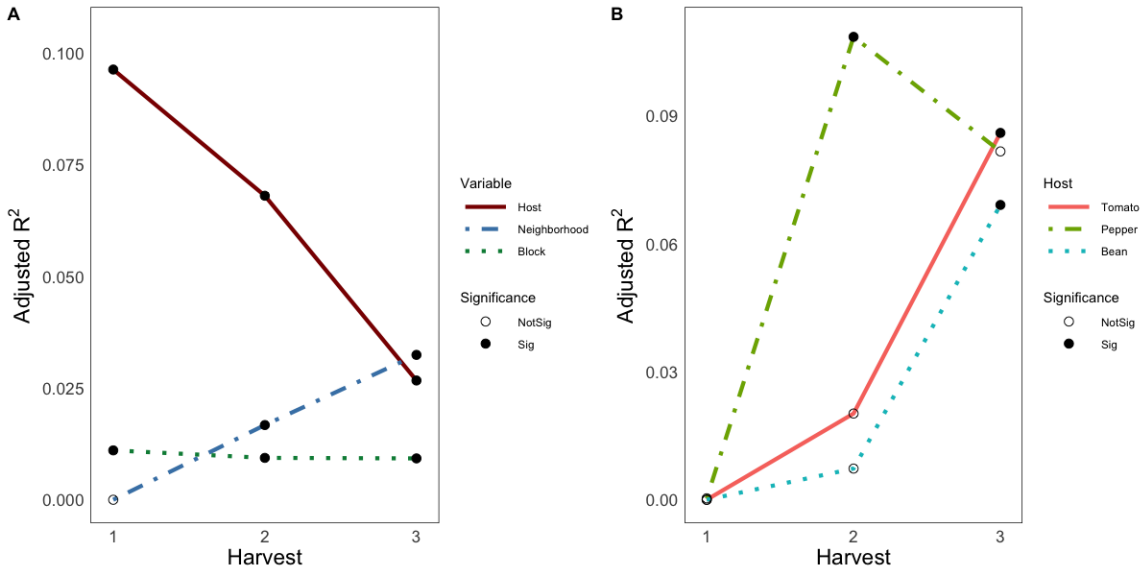
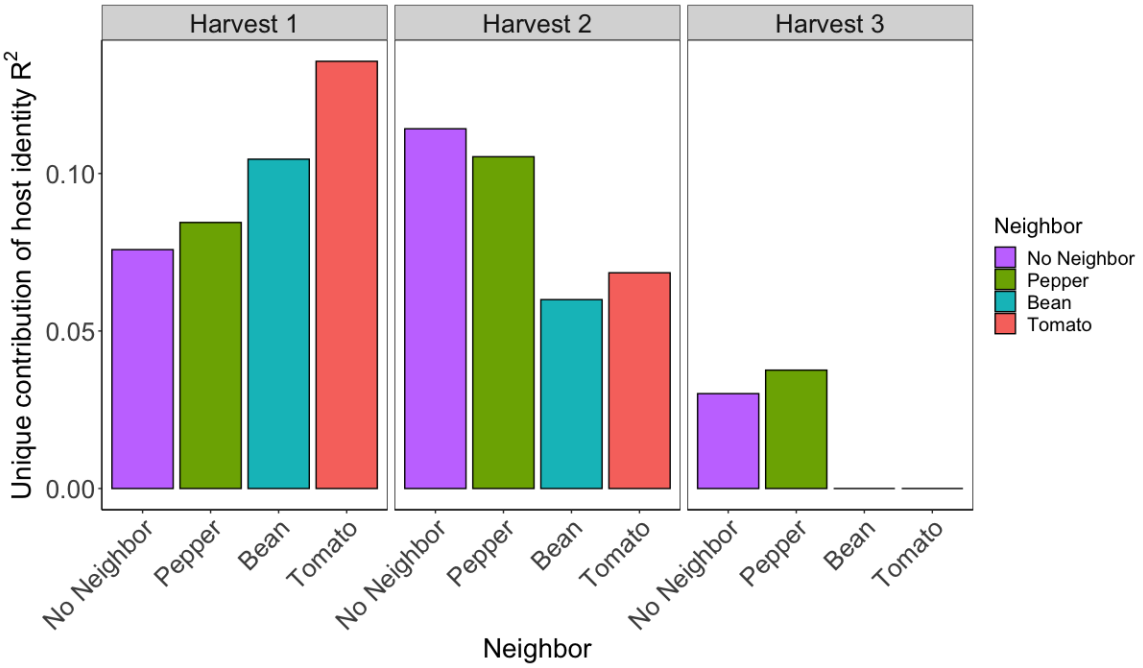


Fig. 3: The effects of host identity on bacterial community composition decrease through time while neighborhood effects increase and vary by host plant species. A) Adjusted R^2 values (y-axis) are the result of PERMANOVA analyses on Bray Curtis dissimilarities for each harvest, accounting for sample number and degrees of freedom from slight differences in sample number. See Table 1 for pre-adjusted R^2 values. The effect of host identity (solid maroon line), the effect of neighborhood (blue dot-dash line), and the effect of experimental block (green dotted line) are shown. Harvest time point is shown on x-axis. Filled circles indicate statistical significance ($p < 0.05$), while open circles represent statistically insignificant effects ($p > 0.05$). B) Host plant species experience neighborhood effects on phyllosphere bacterial communities differently through time. Adjusted R^2 values (y-axis) and harvest time point (x-axis) are as described for plot A. Tomato hosts (solid red line), pepper hosts (light green dot-dash line), and bean hosts (light blue dotted line) are shown. Filled circles indicate statistical significance ($p < 0.05$), while open circles represent statistically insignificant effects ($p > 0.05$).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



1012

1013 **Fig. 4**: The unique contribution (adjusted R^2) of host plant identity to phyllosphere

1014 bacterial structure decreases over time for plants with neighbors. Unique

1015 contribution was calculated by partitioning out spatial principle coordinates using

1016 RDA-based variation partitioning. The order of depiction of the neighbor plant

1017 species is by estimated neighborhood biomass from lowest to highest. Boxes 1, 2,

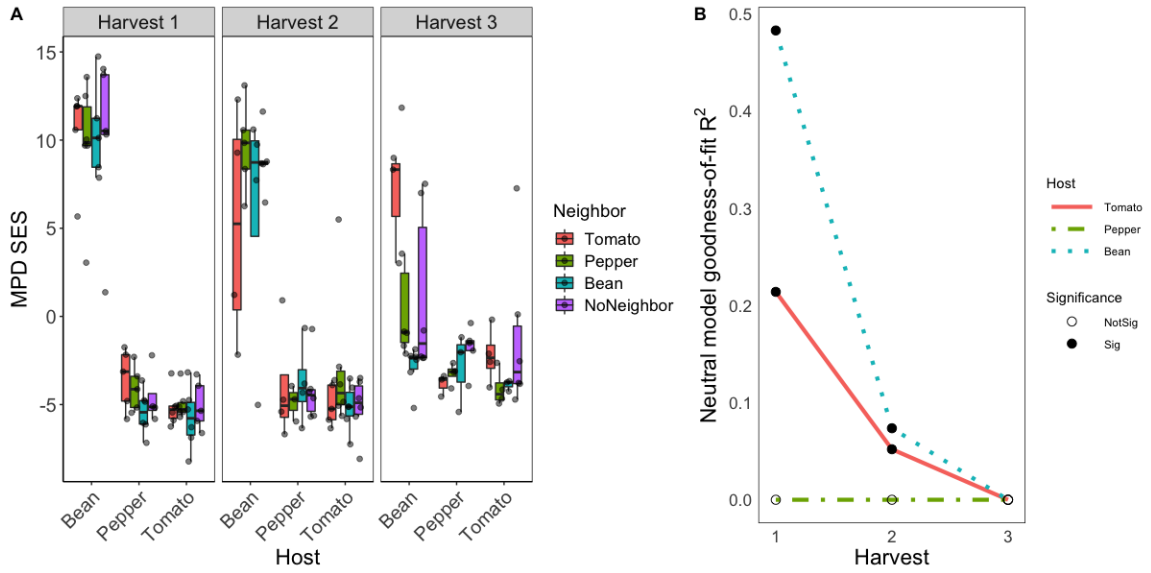
1018 and 3 represent different harvest time points. In cases where $R_{adj} = 0$, host species

1019 identity did not significantly explain variation in phyllosphere bacterial

1020 composition.

1021

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



1022

1023 **Fig 5**: Bacterial leaf surface community assembly processes differ between plant

1024 hosts, suggesting differences in host filtering. A) Standardized effect size (SES) of

1025 mean pairwise distance (MPD) of phyllosphere communities split by host (x-axis),

1026 neighbor (box color), and harvest time point (panels 1, 2, or 3). SES = (MPD_{obs} –

1027 MPD_{null})/SD(MPD_{null}), whereby values below 0 suggest phylogenetic clustering. B)

1028 The fit of a neutral model declines through time, but differs strongly by host

1029 identity. Neutral model goodness-of-fit values (R^2, y-axis) at each harvest (x-axis)

1030 for tomato (solid red line), pepper (green dot-dashed line), and bean (blue dotted

1031 line). Filled circles indicate statistical significance, open circles indicate not

1032 significant (negative or 0 goodness-of-fit values).

1033

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

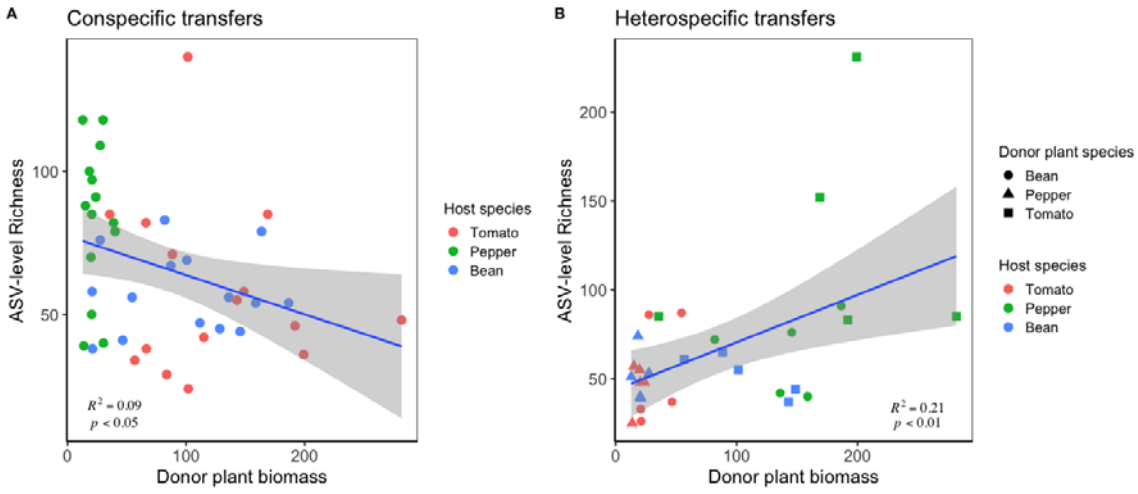
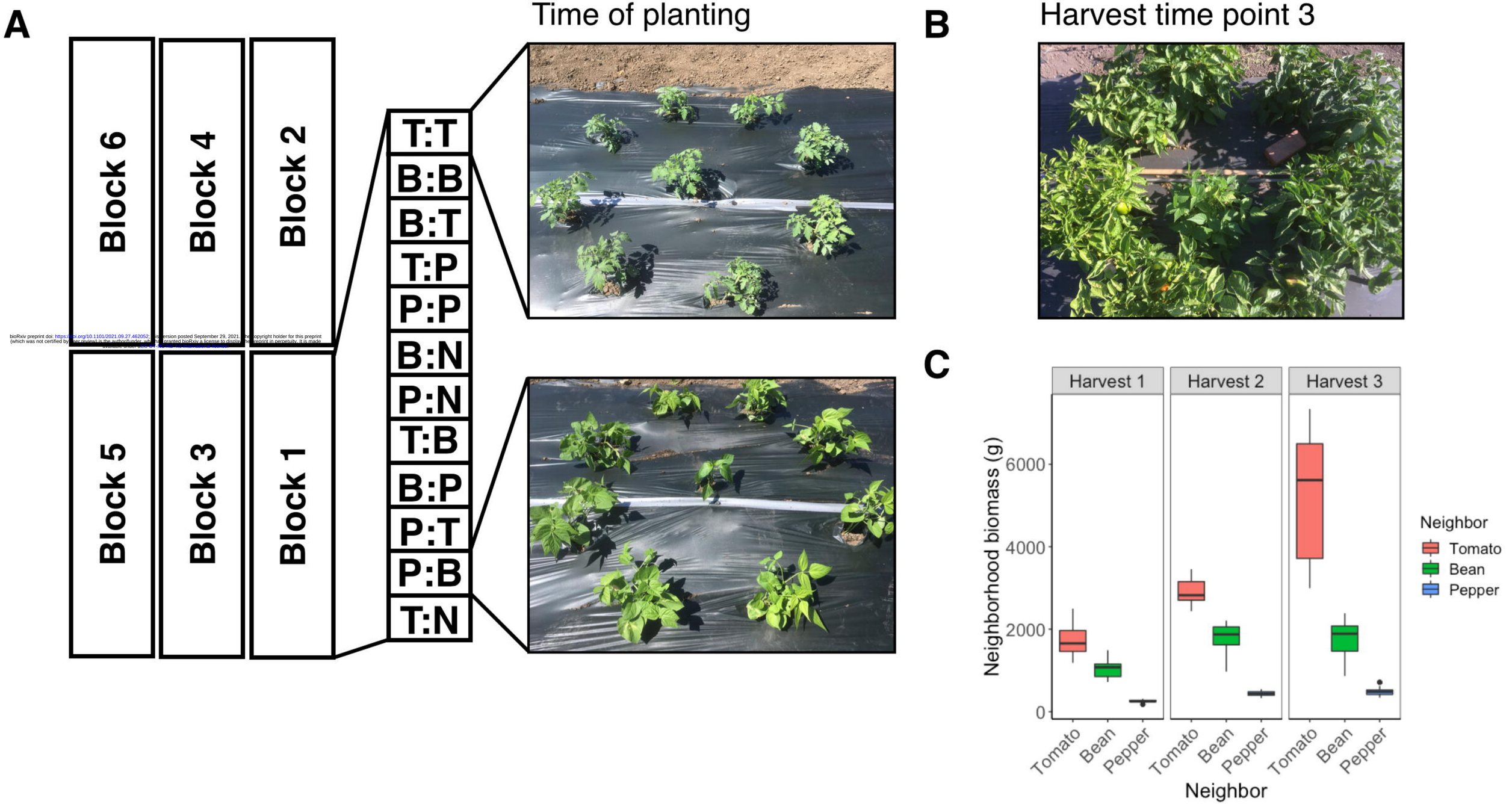
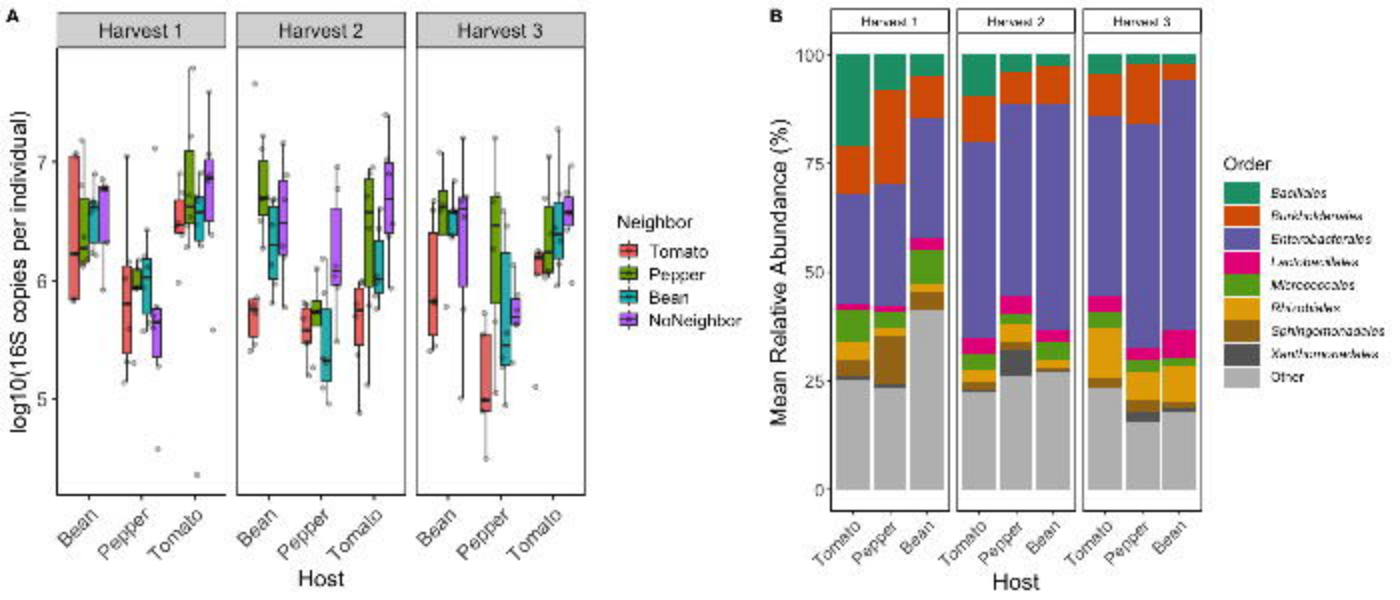
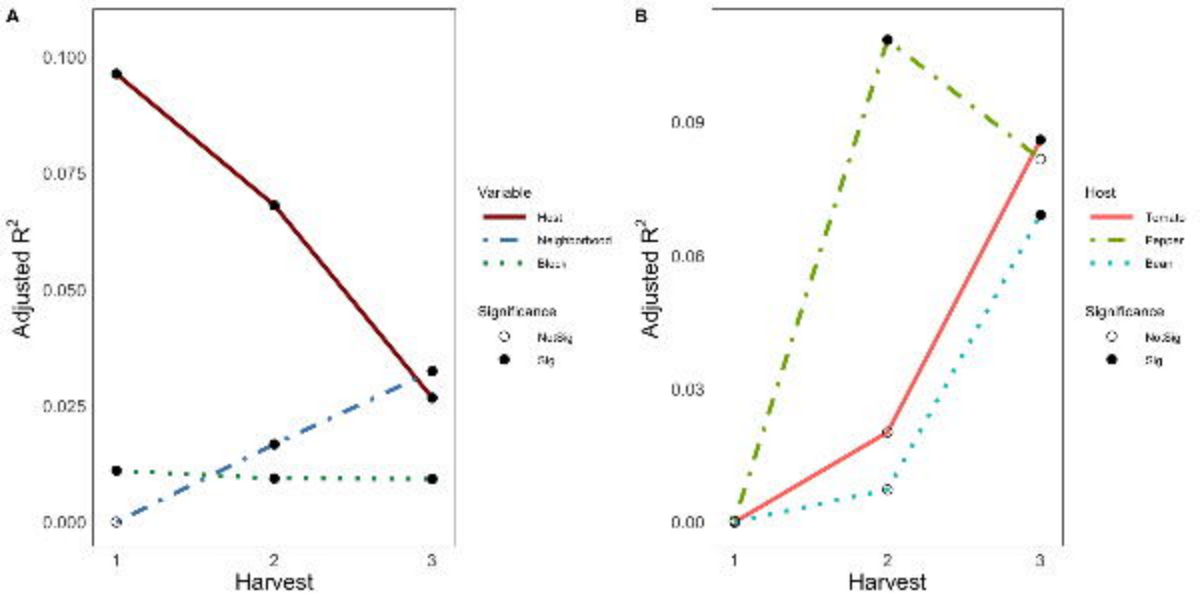


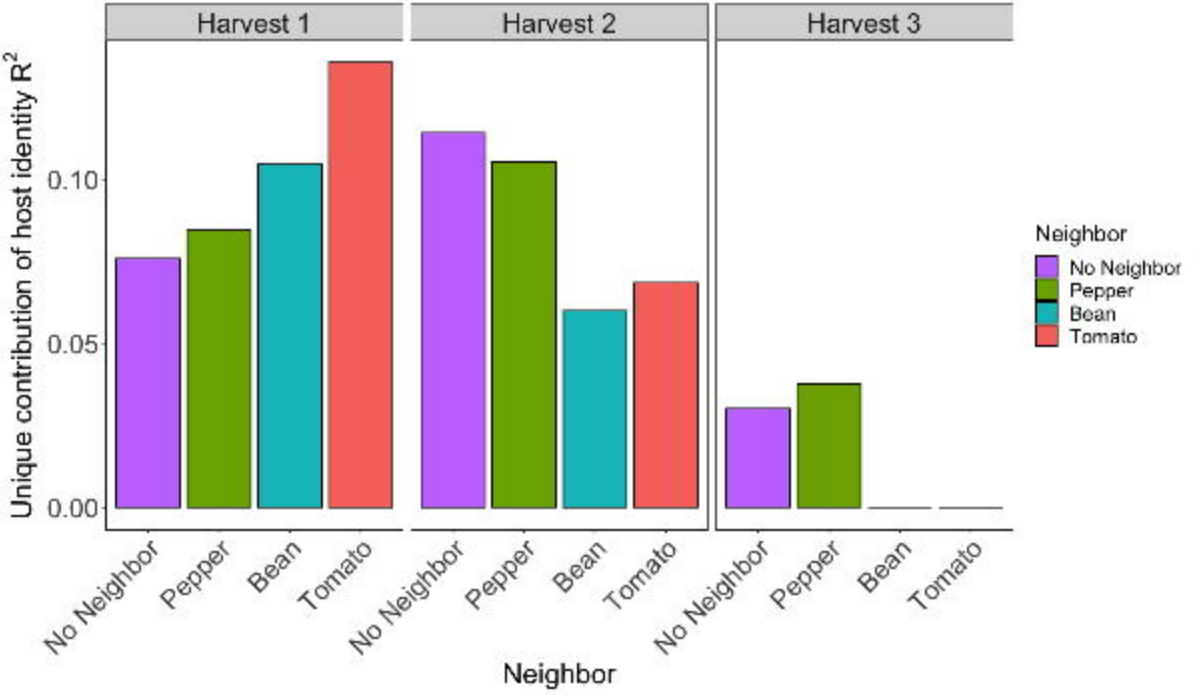
Fig 6: The effect of donor plant biomass (g) on phyllosphere community richness depends on the origin of inoculum. A) Conspecific (within the same species) transfers, where ASV-level richness (y-axis) is negatively, but weakly, correlated with the donor plant biomass (g, x-axis). Points are colored according to the recipient plant species. B) Heterospecific (across different species) transfers, where ASV-level richness (y-axis) is positively correlated with the donor plant biomass (g, x-axis). Point colors correspond to host species and point shapes correspond to the donor plant species. For both plots, R^2 and *p* values are derived from linear models.

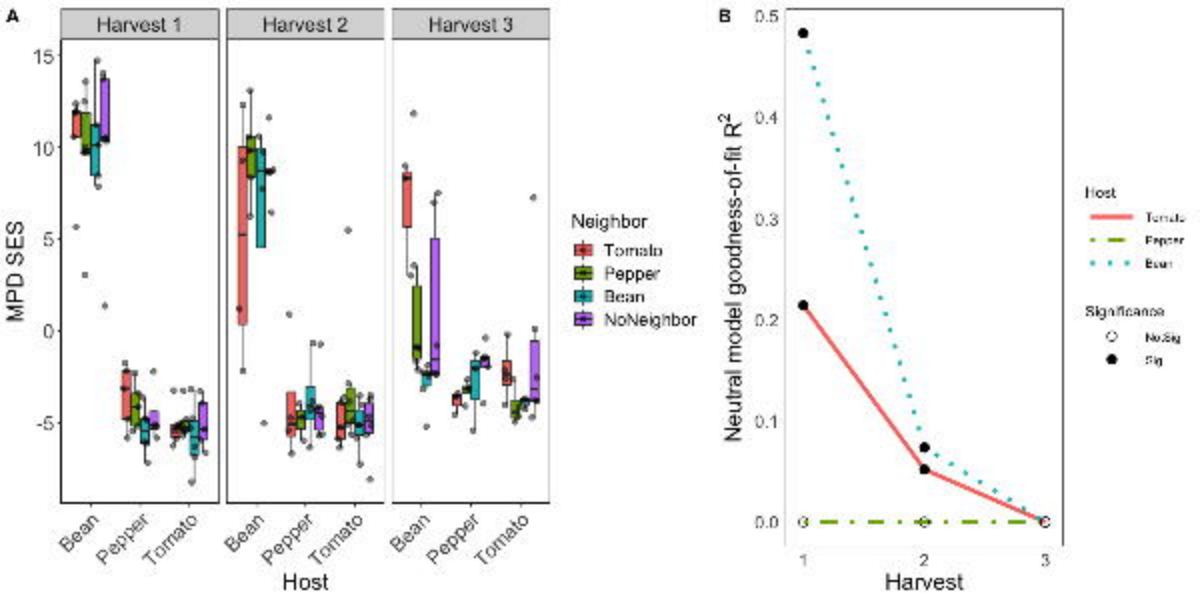
bioRxiv preprint doi: https://doi.org/10.1101/2021.09.27.462052; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

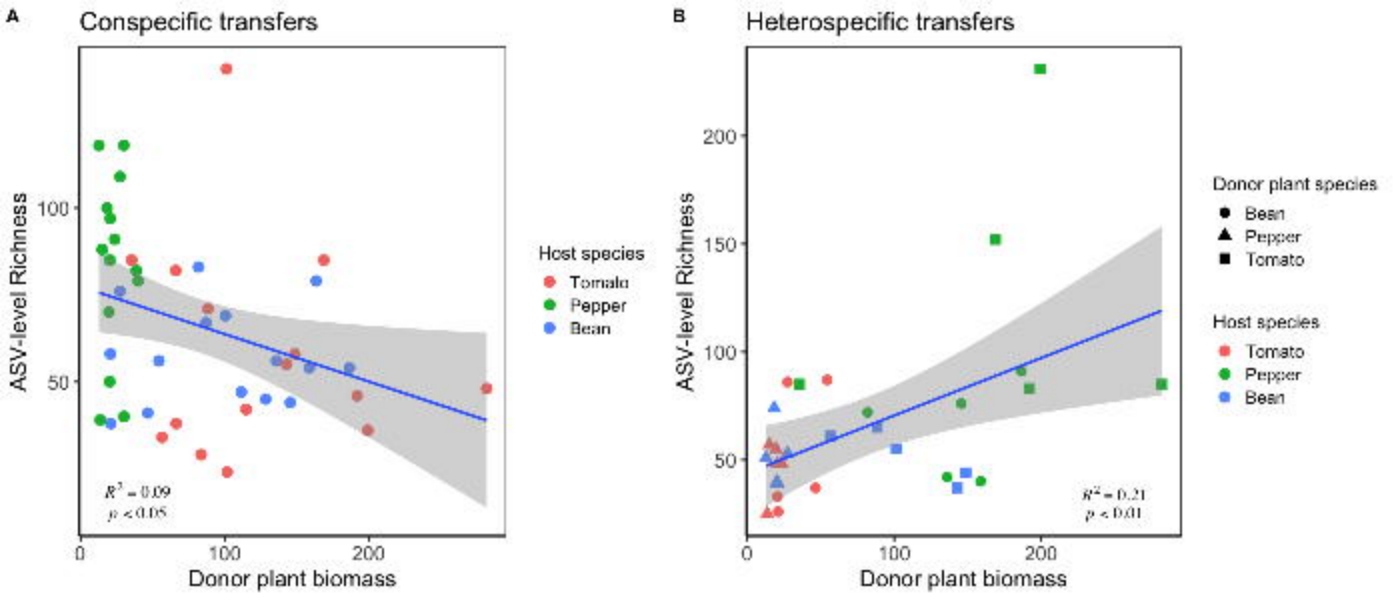



A
Harvest 1
Harvest 2
Harvest 3
log10(16S copies per individual)
7
6
5
Bean
Pepper
Tomato
Host
Neighbor
Tomato
Pepper
Bean
NoNeighbor
B
Harvest 1
Harvest 2
Harvest 3
Mean Relative Abundance (%)
100
75
50
25
0
Tomato
Pepper
Bean
Host
Order
Other


A
0.100
0.075
0.050
0.025
0.000
Adjusted R2
1
2
3
Harvest
Variable
Host
Neighborhood
Block
Significance
NoSig
Sig
B
0.09
0.06
0.03
0.00
Adjusted R2
1
2
3
Harvest
Host
Tomato
Pepper
Bean
Significance
NoSig
Sig


Harvest 1
Harvest 2
Harvest 3
Unique contribution of host identity R^2
0.10
0.05
0.00
No Neighbor
Pepper
Bean
Tomato
Neighbor
No Neighbor
Pepper
Bean
Tomato


A
Harvest 1
Harvest 2
Harvest 3
MPD SES
15
10
5
0
-5
Bean
Pepper
Tomato
Host
Neighbor
Tomato
Pepper
Bean
NoNeighbor
B
Neutral model goodness-of-fit R^2
0.5
0.4
0.3
0.2
0.1
0.0
1
2
3
Harvest
Host
Tomato
Pepper
Bean
Significance
NoSig
Sig


A
Conspecific transfers
ASV-level Richness
Donor plant biomass
$R^2 = 0.09$
$p < 0.05$
Host species
Tomato
Pepper
Bean
B
Heterospecific transfers
ASV-level Richness
Donor plant biomass
$R^2 = 0.21$
$p < 0.01$
Donor plant species
Bean
Pepper
Tomato
Host species
Tomato
Pepper
Bean