Title

Multiple forces shape the phyllosphere microbiome: The importance of vertical transmission, environmental selection, and bacteriophages

Author

Morella, Norma Marie

Publication Date

2019

Peer reviewed|Thesis/dissertation

Multiple forces shape the phyllosphere microbiome: The importance of vertical transmission, environmental selection, and bacteriophages

By

Norma Marie Morella

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Britt Koskella, Co-Chair
Professor Steven Lindow, Co-Chair
Professor Devin Coleman-Derr
Professor Sheng Luan

Spring 2019

Abstract

Multiple forces shape the phyllosphere microbiome: The importance of vertical transmission, environmental selection, and bacteriophages

by

Norma Marie Morella

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Britt Koskella, Co-Chair

Professor Steven Lindow, Co-Chair

As our understanding of host-associated microbial communities (microbiomes) deepens, there is a simultaneous revelation of key gaps in our understanding of these systems. Among these is knowledge of the forces underlying the assembly of, selection within, and dynamics among microbiota. These questions relate to broad principles that are shared across host species, and synthesis across these systems will identify conserved principles in the larger field of microbiome research. The work presented here seeks to identify and explore the relative importance of multiple forces simultaneously shaping the microbiome, specifically, that of the phyllosphere (above-ground surfaces of plants).

This research begins by investigating the importance of vertically transmitted (parent to offspring) microbes in seedling health. Currently, there is little understanding of the ecological importance of commensal or mutualistic bacteria that are transmitted on or within the seeds. Using a tomato study system and combination of classic microbiological techniques and next generation sequencing, we found that the vertically transmitted seed microbiome is capable of protecting seedlings against a common plant pathogen. This work provides evidence that the seed microbiome plays an important ecological role in early seedling life, and very likely shapes the development of the microbiome both directly through priority effects and indirectly through interactions with the plant host.

This work then explores the importance of host genotype and environmental selection in shaping the phyllosphere microbiome in tomato plants. A successive passaging experiment was used to address this question by selecting upon the phyllosphere microbiome. Beginning with a diverse microbial community generated from field-grown tomato plants, replicate plants across five plant genotypes were inoculated for four eight-week long passages, and the microbial community was sequenced at each passage. We observed consistent shifts in both the bacterial (16S amplicon sequencing) and fungal (ITS amplicon sequencing) communities across replicate lines over time, as well as a general loss of diversity over the course of the experiment suggesting that much of the naturally observed microbial community in the phyllosphere in an open environment in outdoor setting is likely transient or poorly adapted. We found that both host genotype and environment shape microbial composition, but the relative importance of

genotype declines through time. Furthermore, using a community coalescence experiment, we found that the bacterial community from the end of the experiment was robust to invasion by the starting community. These results highlight that selecting for a stable microbiome that is well adapted to a particular host environment is indeed possible, emphasizing the great potential of this approach in agriculture and other systems.

In the final chapters of this work, the importance of bacteriophages (viruses that infect bacteria) in the phyllosphere is examined. This question was addressed by transferring microbial communities from field-grown tomato plants to juvenile plants grown under mostly sterile conditions in either the presence or absence of their associated bacteriophage (phage) community. In three separate experiments, we found that the presence of phages affects overall bacterial abundance during colonization of new host plants. Furthermore, bacterial community analysis (16S amplicon sequencing) shows that phages significantly alter the relative abundance of dominant community members and can influence both within- and among-host diversity. This is then extended through the measuring of the impact of phages on bacterial communities weeks after initial inoculation. This work describes the impact of disrupting bacteria-phage temporal dynamics on bacterial communities. Together, these results underscore the importance of both lytic and lysogenic phages in host-associated microbiomes but how they can have fluctuating impacts over relatively short timescales.

Together, these results contribute to fundamental knowledge gaps by demonstrating the ecological importance of vertically transmitted microbes, determining that microbiomes can be adapted to their host and environment, and uncovering the temporal variability of key driving forces underlying microbiome structure.

Table of Contents

Acknowledgements

I would first like to thank my advisor, Dr. Britt Koskella, for her mentorship during my PhD. Britt is a superb PI, and her approach to science has helped me build my own scientific identity and place in Microbial Ecology over the past several years. In addition to providing immense scientific support, Britt is truly a mentor to her students: she fosters an inclusive and supportive work environment, and I have benefitted greatly from conversations with Britt regarding the challenges and highlights of life in academia. Britt is a role model in many ways, and I will always be immensely grateful that I had the opportunity to conduct my PhD research under her guidance and mentorship.

I would next like to thank Dr. Steve Lindow, my co-advisor. Steve has provided invaluable insight on my research that can only come from decades of experience and knowledge. He provides this insight in the form of inquisitive questioning that has fostered my ability to think critically about my own research and the work of others. All the while, Steve plans camping trips that bring the lab to beautiful outdoor locations. I feel very fortunate to have been a member of the Lindow lab, and I will miss our conversations around the campfire.

I would like to thank my committee members, Dr. Devin Coleman-Derr and Dr. Sheng Luan, for their feedback on my work, including my written dissertation.

Thank you to the Oxford Greenhouse staff for help in caring for tomato plants. I also thank Shana McDevitt and Dylan Smith at the QB3 sequencing facility for their continued support of my sequencing efforts through the years.

I would like to acknowledge current and former members of the Koskella Lab. They not only make the days more enjoyable, but their intelligence helps create a collaborative scientific environment that elevates the quality of everybody's research. I would especially like to thank all of the undergraduate students that I have worked with for their positive attitudes and exceptional work ethic: Annika Gomez, Michelle Leung, Grant Wang, Chris Yang, Merissa Washalaski, Shirley Zhang, and Alina Lee. I would also like to acknowledge current and former members of the Lindow lab for welcoming me into their lab and providing feedback on my research.

Thank you to my friends. From care boxes, to good luck messages, to celebrating small victories throughout the years, I will always feel overwhelmed by the love and encouragement that my friends have shown. And thank you to Willow for making every day better in so many ways.

Lastly, I would like to thank my family: Mom, Dad, Jake, Jenni, Anna, Carmen, and all of my extended family- of whom there are too many to list. They have made the tough parts of graduate school easier and the successes sweeter. The last five years would not have been possible without for their unending support and love.

Chapter 1 Introduction

Parts of this chapter have been adapted from the following with permission:
Morella, Norma M, and Britt Koskella. "The Value of a Comparative Approach to Understand the Complex Interplay between Microbiota and Host Immunity." *Frontiers in immunology* vol. 8 1114. 14 Sep. 2017

1.1 Motivation

As the demand for agricultural production grows and additional challenges arise as our climate changes, there is increasing interest in harnessing beneficial plant-microbe interactions to improve promote plant growth and health. Furthermore, though plants and microbes have been co-evolving for millions of years, human agriculture has only been in practice for 12,000 years [1]. Plant breeding and agricultural processes such as chemical usage, monoculture, and annual tillage are far removed from how plants naturally grow and thrive, and as such, there is a need to understand how agricultural and breeding practices impact plant-microbe interactions. There remain key open questions regarding the effects of plant breeding for agriculturally beneficial traits, transmission of plant microbiota, and overall importance of microbial community living on the aboveground surfaces of plants, a habitat known as the phyllosphere.

1.2 *Solanum lycopersicum* as a model system

The plant host study system in this work is the modern tomato plant, *Solanum lycopersicum*. Tomatoes have an extensive and interesting history of domestication. Their wild ancestors, *Solanum pimpinellifolium*, originate from the Andean region of South America. They are thought to have been domesticated initially in Mexico and then brought to Europe in the 1500s [2]. Though wild tomatoes have a large genetic diversity, domesticated cultivars are estimated to contain less than 5% of this genetic diversity due to a history of inbreeding and backcrossing of domesticated lines [3]. It was not until the early 1900s that breeders began to cross *S. lycopersicum* with wild lines to reintroduce genetic diversity and select for desirable traits such as disease resistance.

The tomato plant is an ideal host system for various reasons. First, researchers and breeders have access to thousands of tomato accessions through the UC Davis Tomato Genetics Resource Center (TGRC), allowing for consistent and reliable control over host genetics. Furthermore, tomato plants can be easily propagated in controlled-temperature growth chambers, the greenhouse, or the field. Seedlings can be grown for up to three weeks in entirely aseptic conditions on water agar plates. Finally, there is an applied interest in work such as this, as the tomato industry in the US produces 2.6 billion dollars worth of fresh and market tomatoes (USDA-AMS, 2017), and locally, is one of California's top ten most valued agricultural products (CDFA).

1.3 Host Immune system- Microbe interactions

Many of the ways in which microbial communities interact with their host is through the host immune system. As such, a cursory understanding of the plant immune system, and how this compares to immunity in other systems, is necessary to fully understand how microbes influence host health.

The adaptive immune system is thought to have arisen in jawed fish ≈500 million years ago [4], whereas the innate immune system likely dates back to early eukaryotic cells themselves

[5, 6]. As microbial communities greatly pre-date the existence of multicellular eukaryotes, both branches of the immune system evolved in the presence of microbes, and it follows that tolerance for commensal microbiota (those associated with hosts, but which do not cause disease) must have been a key factor in shaping the evolution of immunity. Innate immunity, found across all kingdoms of life, is non-specific and responds broadly to 'non-self' invaders. Its hallmarks include protective physical barriers and general pattern recognition receptors (often referred to as microbe-associated or pathogen-associated molecular patterns) that sense non-self signals and elicit host responses. Adaptive immunity is unique to vertebrates and responds to specific pathogens through detection of antigens via somatically-generated receptors [7] and specialized white blood cells (B and T cells), resulting in immunological memory, in which the organism is protected from future infection (the basis of vaccination). Broadly speaking, the adaptive immune response is highly specific to particular pathogens and can change over the course of a host's lifetime whereas the innate immune response is a general resistance that can only respond to selection across host generations.

Plant innate immunity consists of two primary responses to microbes [8]. The first branch of the immune system recognizes microbial- or pathogen-associated molecular patterns (MAMPs or PAMPs, such as flagellin and lipopolysaccharides) through the use of transmembrane pattern recognition receptors and results in pattern-triggered immunity. However, many plant pathogens have evolved to overcome these defenses through the use of effectors. Plants with resistance genes for specific pathogens can detect the effectors through NB-LRR proteins, which represent the second response to microbes: effector triggered immunity. Additionally, plants have physical barriers to infection such as cell wall defenses [9], and they can also secrete antimicrobial peptides to ward off infection [10]. Although vertebrate adaptive immunity is typically considered more specific, plant immunity is also the result of specific recognition, either of pathogen effectors or molecular patterns, and immune 'priming' in plants has been shown repeatedly [11–14].

As is becoming increasingly evident, in addition to its well-studied role in preventing pathogen establishment, the immune system also influences both the composition and abundance of non-pathogenic microbiota. In mammals, this is best studied in the gut microbiome, where differentiating between these diverse commensals and colonizing pathogens is clearly a highly complex problem. The human immune system maintains a homeostatic relationship with commensals through numerous mechanisms, including stratification and compartmentalization of the intestine, production of a mucous layer and antimicrobial proteins, and limiting epithelial exposure and immune response [15]. The plant immune system is also important in shaping the non-pathogenic microbiome [16]. Two studies in *Arabidopsis thaliana* demonstrate that disrupting components of the plant immune system, such as the signaling molecules salicylic and jasmonic acid, influences microbial communication composition: the first shows evidence for altered root microbiome communities in plant hosts lacking genes controlling production of SA compared to control plants [17]; and the second shows altered microbial communities in plants with mutations in genes controlling ethylene response (another signaling molecule) and cuticle formation [18]. Recent work in wheat also demonstrated a role for jasmonic acid in shaping composition of the microbiome, and again in this case, activation of JA signaling pathways altered microbial diversity and composition of root endophytes [19].

In mammals, microbiota are critical in development and function of components of adaptive immunity, such as B and T cell diversity and differentiation [20]. In plants, commensal bacteria influence host immunity by priming the plant for future exposure to pathogens through

the induction of a systemic response, causing broad-range basal levels of protection. A primed plant may respond more rapidly and strongly to pathogen invasion through a variety of mechanisms including: quicker closing of stomata, less sensitivity to bacterial manipulation of defenses, up regulation of defense-related genes, and stronger salicylic acid related immune responses [21]. In some cases, the effects of priming can even be trans-generational through chromatin and histone modification, where the subsequent generation of primed plants exhibits enhanced resistance to bacterial, fungal, and herbivorous pathogens [22–25]. Host-associated microbiota can also directly influence host resistance against invading pathogens. Common across most systems, the microbiome can serve a protective role that is independent of the host immune system through antagonism, competitive exclusion, or physical exclusion of pathogens, collectively referred to as defensive symbiosis [26, 27]. Recently, the phyllosphere microbiome, discussed in this work, has been found to protect its plant host against pathogens [28–30].

In mammals it is clear that early exposure to microbes is crucial to the development of both branches of the immune system [31], influencing not only immune development and response against pathogens, but also tolerance to commensal microbiota [32]. The role of early exposure to microbiota suggests it would be advantageous for a community of beneficial microbes to be transmitted vertically from parent to offspring (e.g. through birth or seeds) from generation to generation. Transmission of microbiota in plants can occur vertically through the seeds, or horizontally from the soil and surrounding environment. Plants ranging from trees to grasses are known to harbor bacteria in their seeds, many of which are reported to promote plant health [33–35]. Despite this, there is no evidence that plants actively select for transmission of specific microbial communities, and there are no clear examples of adaptations to ensure seed-mediated transmission. My work in Chapter 2 explores this topic through uncovering the importance of some seed-transmitted microbes in early seedling health.

1.4 The phyllosphere

The plant phyllosphere is defined as the aerial surfaces of plants, or, all plant tissues growing aboveground. This work primarily focuses on microbial epiphytes of the phyllosphere: the bacteria, viruses, and fungi that are found on the surfaces of leaves. The phyllosphere is a massive habitat estimated to exceed 10^8 km^2 of plant surface area worldwide [36]. It is, in general, a nutrient poor environment that undergoes fluctuations in temperature, UV, and moisture [37]. The phyllosphere microbiome is known to harbor primarily four phyla of bacteria, and they reach an abundance of ~10^6 cells/cm^2 [18, 38, 39]. Microbes from surrounding plant species, dust, soil, and other sources are thought to be the primary colonization source for the phyllosphere. In particular, neighboring plants have been shown to contribute to both the density and composition of local airborne microbes [40]. However, as demonstrated in Chapter 3, the microbes frequently described as members of the phyllosphere microbiome may in fact be transient visitors and not well-adapted colonizers of the environment.

Although there is a trend in phyllosphere research to focus on the bacterial portion of the microbiome, there have been some studies describing the fungal community as well [39, 41, 41, 42]. There is even less work on the viral community, although from culture-based work, we know that bacteriophage viruses do indeed inhabit the phyllosphere and predate upon the bacterial community [43–45]. There are many technical limitations that impede the field's ability to fully describe the phyllosphere phage community. Nevertheless, my work in Chapters 4 and 5 contribute to our understanding of the importance of bacteriophages in this system.

Compared to the field's understanding of the belowground microbial habitat, the rhizosphere, the phyllosphere has been relatively understudied. Despite this, there are many

advantages to the system. Specifically, the phyllosphere has a naturally distinct spatial structure, it is relatively easy to sample, and the microbes are highly culturable. Through inoculation using a fairly simple spray technique, the environment can be evenly saturated with diverse microbial inoculum, and it is possible to sample the successfully colonized community in its entirety. It is also easy to visualize, and spatial patterns of colonization and survival can be easily ascertained. Moreover, bacterial abundance and growth can be tracked using droplet digital PCR, (ddPCR, Appendix 1), and the bacterial and fungal communities can be described using next generation sequencing. Overall, the phyllosphere is an ideal system in which to study topics such as the relative importance of transmission events, host characteristics, the environment and microbe-microbe interactions in shaping the microbiome.

1.5 Concluding introductory remarks

Symbiotic associations between plants and microbes span from pathogenic to beneficial, and these interactions have been studied from many angles of science- from evolution to agriculture. My dissertation research seeks to address fundamental questions about microbial community ecology and host-microbiome interactions. It is motivated by the belief that rational design or manipulation of complex microbial communities has the potential to shape the future of medicine and agriculture, but this success will largely depend on our basic understanding of the systems at hand. The findings described in this dissertation make a small but significant step in this direction.

Chapter 2. Tomato seed-associated bacteria confer protection of seedlings against foliar speck caused by *Pseudomonas syringae*

Parts of this chapter have been adapted from the following with permission:
Norma M. Morella[a,+], Xuening Zhang[b,+], and Britt Koskella[c,] Tomato seed-associated bacteria confer protection of seedlings against foliar speck caused by *Pseudomonas syringae.* *Phytobiomes.* Accepted May 6, 2019.
[+] These authors contributed equally

2.1 Introduction

Plant associated microbiomes are capable of enhancing host fitness through a number of mechanisms. They can promote growth through production of phytohormones [46] and fixation of nutrients from the environment [47, 48], confer both drought and stress tolerance [49, 50], and even influence flowering time of their hosts [51]. Perhaps one of the most influential ways that microbial organisms affect host fitness is through their impact on host immunity and disease resistance. In plants, microbes can confer disease resistance through both indirect and direct mechanisms and can indirectly protect against disease via the plant immune system. Plants are able to detect microbial-associated molecular patterns such as lipopolysaccharides (LPS) in the environment, activating a generalized anti-microbial defense mechanism, and effectively priming the plant to respond more effectively when subsequently exposed to a pathogen [8, 52, 53].

Experimental studies using isolated strains of bacteria have demonstrated that many can protect plants against pathogen colonization through direct inhibition of the pathogen's growth, either through competition for resources [29] or production of antimicrobials [54]. Furthermore, there is long-standing understanding that plants can be 'primed' against pathogen colonization by colonization of non-pathogenic bacteria [12]. It is now becoming clear that the microbiome as a whole might act collectively to confer disease resistance, although it is more difficult to pinpoint mechanisms underlying the effects of whole consortia compared to studying individual strains using culture-dependent methods. Both the rhizosphere [55] and phyllosphere microbiomes [28, 30, 56] have recently been shown to provide protection against pathogens. But even as we begin to understand microbiome-mediated protection against disease, it is unclear how a naturally protective community might assemble on a plant, and, once assembled, if it can be stably maintained over time. A broader consideration of how plant-associated microbiomes are acquired and transmitted among hosts is required to better understand how a generally beneficial community might persist across generations.

The two dominant sources for assembly of the plant microbiome are horizontal transmission from the environment and unrelated plants and vertical transmission from parental plants. Local plant populations are important contributors to the airborne microorganism community, and thus that movement of microbes among neighboring plants can readily occur through aerial dispersal [40]. Unlike horizontal transmission, however, vertical transmission holds the potential to connect, extend, and reinforce beneficial symbioses across temporal and spatial scales. In plants, vertical transmission of microbial communities is observed in both vegetative [57] and sexual reproduction [58]. Parental microbiota can be transmitted through the foliar (e.g. via the stigma) and vascular pathways onto seeds [59], though the most likely route across plant species remains unclear . Once on or within the seeds, they can act as the incipient members of a mature plant microbiome, critically shaping growth, development, and susceptibility to pathogens of newly emerging seedlings. Such transmission would allow plant

lineages to maintain beneficial symbioses across multiple generations and pave the way for coevolution of the partners, as has been well-characterized in other systems (e.g. aphids and their vertically transmitted symbiont, *Buchnera aphidocola*; [60]. Indeed, recent work investigating tomato seed endophytes suggests plants may preferentially transmit plant growth promoting bacteria to the next generation [61].

Overall, studies of the biological significance of vertically transmitted seed microbiota remain fragmented, with a large focus on seed-mediated transmission of pathogens. High-throughput community profiling has offered important insight to seed-associated microbiomes [33] but often lacks support from experimental manipulation, limiting the scope of conclusions on their ecological implications (although see [62] for an elegant counter-example). Moreover, studies on seed-associated microbes have focused primarily on endophytes from surface-sterilized seeds, despite the fact that the seed surface is the most immediate interface between the embryo and parental tissues. As a result, endogenous seed epiphytes remain a relatively unexplored group, despite their potential importance in early colonization of plants. Here we present a study in which we examine whether endogenous seed epiphytic microbes, both as a community and in isolation, protect seedlings of various tomato types against a common plant pathogen, *Pseudomonas syringae pv* tomato strain DC3000 (*Pst*). By transferring naturally occurring seed-associated microbial communities back onto surface-sterilized seeds of either the original cultivar or different genotypes, and comparing pathogen colonization and disease susceptibility against un-inoculated control seedlings, we were able to test the impact of multiple seed-associated communities and bacterial isolates on disease progression, and examine the dose-dependence of protection conferred.

2.2 Results

2.2.1 Fruit Collection (See methods for complete details)

Tomato fruits were collected from UC Davis Student Organic Farm in September 2017. We collected mature, intact fruits from a total of four different tomato types based on distinct fruit morphologies and field locations including: orange cherry tomatoes, red cherry tomatoes, medium-orange-sized tomatoes, and an unidentified heirloom variety. Fruit of the same tomato type/cultivar were collected from multiple plants planted in the same row, resulting in four tomato types (designated as Tomato Type 1-4). Tomato Type 1-3 were collected from non-neighboring lanes from one field, and the heirloom variety (Tomato Type 4) was collected from a neighboring field. Tomatoes were brought into the lab, pooled within tomato type (Figure 2-1a), sterilized, and then fermented to collect seeds.

2.2.2 Initial screening of protective effects of natural seed microbiota

To first determine if endogenous epiphytic microbiota of field tomato seeds are protective against pathogen infection, we germinated seeds with or without their endogenous epiphytic community from each of the four tomato types before challenging with *Pst* and disease severity recorded daily (Figure 2-1b). We found that the application of a endogenous seed microbiome significantly reduces disease symptoms at the seedling stage (Figure 2-2a; repeated measures GLM, $F_{1,42}$=5.182 , p= 0.02327), and there is no significant effect of host tomato type on the protective effects of their microbiome ($F_{3,42}$= 0.597, p= 0.6212). There is also no significant interaction between disease severity and tomato type ($F_{3,42}$= 0.807, p= 0.4988).

As an alternative measure of disease, we also calculated the area under the disease progression curve AUDPC as a cumulative measure of disease symptoms over time (Figure 2-

2b). Seedlings with their original seed microbiome had significantly lower AUDPC values (ANOVA $F_{1, 34}$= 11.310, p= 0.0019), and alleviation of disease severity is again largely independent of the tomato type ($F_{3, 34}$= 0.612, p= 0.612). There was no significant interaction between the two variables ($F_{3, 34}$= 0.758, p= 0.5252).

2.2.3 Application of protective TT4 seed microbiota across multiple tomato types

We next sought to test the protective effects of the seed microbiome independent of the host genetic background. We chose to use TT4 because, based on our initial results, it appeared to have the strongest protective effect. We harvested the TT4 seed microbiome from TT4 seeds and applied it onto surface sterilized seeds of both itself and the other two tomato types (with the exception of tomato type 3, as an insufficient number of tomatoes were collected from the field to conduct the experiment). We found that TT4 seed microbiota was also capable of protecting tomato types 1 and 2 from disease. The application of TT4 microbiome is significantly protective against *Pst* infection when the disease progression data were compared to controls ($F_{1,21}$= 16.531, p= 0.0006; Figure 2-3a). There was also a significant effect of recipient tomato type ($F_{2,21}$= 4.852, p= 0.018), but no interaction between tomato type and treatment was observed ($F_{2,21}$= 0.481, p= 0.625). When analyzing the same data using AUDPC value (Figure 2b) we similarly observed that seedlings which received TT4 microbiota are significantly less diseased ($F_{1,21}$= 16.234, p= 0.0006; Figure 2-3b). Again, there is a significant effect of recipient genotype ($F_{2,21}$= 4.355, p= 0.0261), but with no significant interaction between the two variables ($F_{2,21}$= 0.278, p= 0.7604).

In addition to analyzing disease severity, we used ddPCR to quantify total *Pst* density from the seedlings at the end of the experiment. We found that seedlings which did not receive TT4 microbiota carried significantly more *Pst* than those which did ($F_{1,17=}$ 11.237, p= 0.0038; Figure 2-3c). Unlike disease progression, we did not detect any effect of recipient tomato type on *Pst* density ($F_{2,17}$= 0.966, p= 0.4007) and there was not an interaction amongst tomato type and treatment ($F_{2,17}$= 0.784, p= 0.472).

2.2.4 Taxonomic characterization of seedling microbiota

Intrigued as to what made TT4 seed microbiota protective on not only its own tomato genotype but also others, we used 16S rRNA community profiling to sequence the bacterial communities of two week old seedlings whose seeds had been inoculated with TT4 microbiota. We found that these seedlings were strikingly dominated by OTUs in the genus *Pantoea* (Figure 2-4). Knowing that *Pantoea* is highly culturable, and also that many species are already used as biocontrol strains [63], we next sought to culture isolates from TT4 seeds to determine the exact species of *Pantoea* that were endogenously found on these seeds. We were able to culture six bacterial isolates from TT4 seeds. We also tried to culture bacterial isolates from the other three tomato types, and we were only able to culture one bacterial isolate from TT2, which we identified as a *Bacillus* species (G1A). Using Sanger sequencing, we sequenced the 16S genes of our isolates and identified them as species of *Pantoea* (Table 2-1). Because *Pantoea* spp. are notoriously difficult to differentiate using 16S sequences [64, 65], we chose three isolates (ZM1, ZM2, and ZM3) based on distinct colony morphology and different 16S sequences, and sequenced their *gyrB* and *rpoB* genes as well. We were able to further confirm their identities and place them within a phylogenetic tree of *Pantoea spp.* (Figure 3b) [66]. To our knowledge, our isolates have not been previously identified nor used as biocontrol strains, although some related strains have been developed (Figure 3b, noted as “Biocontrol”). Interestingly, ZM3 and

ZM2 appear to be similar based on DNA sequencing, with their partial 16S sequences aligning 99% to one another and their partial *gyrB* sequences aligning 100%. However, when grown on nutrient agar, their colonies are distinctly different colors; yellow and white, respectively. Whole genome sequences will further elucidate genetic differences between the isolates and are currently underway. Lastly, we aligned our ~420bp of amplicon sequencing data to near full length reverse Sanger sequencing reads of the isolates and observed 100% match of some of these OTUs with our isolate sequences (Figure 2-5).

Table 2-1 Isolate Identities using Sanger Sequencing

Isolate	Gene	Top Match
ZM1		
	16S (27F-1492R)	99% Pantoea dispersa strain AS18
	rpoB A	NA
	rpoB B	Plautia stali symbiont
	gyrB	98-99% Plautia stali symbiont; Pseudomonas amygdali pv. sesami strain BC3275; Pantoea dispersa strain ATCC 14589
ZM2		
	16S (27F-1492R)	99%: Pantoea sp. strain 33
	rpoB A	99%: Pantoea agglomerans strain B025670
	rpoB B	99%: Pantoea agglomerans strain C410P1
	gyrB	99%: Pantoea agglomerans strain C410P1
ZM3		
	16S (27F-1492R)	99%: Pantoea sp. strain CA15-11 16S; Pantoea agglomerans strain CZ-BHG003; Pantoea vagans strain Y-4
	rpoB A	99%: Pantoea agglomerans strain B025670
	rpoB B	93% Pantoea agglomerans strain C410P1
	gyrB	99%: Pantoea agglomerans strain C410P1

2.2.5 Protective ability of seed microbiota isolates and known biocontrol strains

Next, we sought to determine if the isolates alone, in the absence of the remaining microbiome, were sufficient in protecting seedlings against disease. We evaluated the protective ability of ZM2, ZM3, two commercially available biocontrol agents of the same genus, and the putative *Bacillus* spp. (GIA) that we cultured from the less protective TT2 seed microbiome. Treatments were categorized as the following treatments: single-strain, consortia, or control. When comparing AUDPC values after ten days of disease scoring (Figure 2-6a), we found a significant effect of treatment ($F_{2,30}$= 12.01, $p < 0.0001$). Post-hoc Tukey tests show both single-strain inoculated seedlings and consortia inoculated seedlings were significantly less diseased than seedlings which did not receive any bacteria ($p= 0.0001$ and $p< 0.0001$, respectively). There was no significant difference between seedlings inoculated with single strains and consortia ($p= 0.89$). We also uncovered a significant effect of strain ($F_{5,12}$= 5.07, $p = 0.01$). Post-hoc Tukey

tests revealed that the only significant differences amongst strains are between strains ZM1 and G1A (p=0.03), and strains ZM2 and G1A (p=0.007).

When analyzing *Pst* density (Figure 2-6b), we again found a significant effect of treatment ($F_{2,30}$= 32.29, p < 0 .0001). Post-hoc Tukey tests show both single-strain inoculated seedlings and consortia inoculated seedlings were significantly less diseased than seedlings which did not receive any bacteria (p< 0.0001). There was no significant difference between seedlings inoculated with single strains and consortia (p= 0.73).
Unlike AUDPC, there was no significant effect of strain in terms of total pathogen load ($F_{9,22}$= 1.094, p= 0.412).

We next compared the protective ability of live isolates compared to their UV-killed counterparts (Figure 2-6c, d). When analyzing the effect on disease severity via AUDPC values, we found a significant effect of both UV-treatment ($F_{1,26}$= 69.3786, p<0.00001) and isolate treatment ($F_{6,26}$= 6.9200 p= 0.00018). There was no significant interaction of the two ($F_{6,26}$= 1.1786, p= 0.34644). When analyzing *Pst* density, we observed a significant effect of isolate treatment ($F_{6,26}$=10.151, p<0.00001), but we did not observe an overall significant effect of UV-treatment ($F_{1,26}$= 0.443, p= 0.512). Again, there was no significant interaction between isolate and UV-treatment ($F_{6,26}$=0.838, p=0.535).

In order to determine if pair-wise differences existed between seedlings treated with live and UV-killed isolates and the no-treatment control seedlings, we performed Welch's t-tests and controlled for multiple comparisons using the Bonferroni correction. For live bacteria, all isolates were capable of lowering disease severity symptoms compared to controls (p<0.026). In contrast, we found that all *Pantoea* isolates were capable of lowering *Pst* density (p<0.048), but *E. coli* was not (p=0.123). None of the UV-killed isolates lowered AUDPC values as compared to controls (p>0.068). In contrast, we find that ZM1, ZM2, ZM3, and E325 significantly lower *Pst* density (p=0.04104, p=0.03783), p=0.03246, p=0.04104 respectively), and C9-1 and *E. coli* do not (p=0.08417, p=0.08242). In conclusion, when comparing to no-treatment control plants, all strains of live bacteria are capable of reducing disease severity, and all strains except for *E. coli* also lower *Pst* density. In comparison, when seeds are treated with UV-killed bacterial strains, none lower disease severity, but all except for one *Pantoea* isolate and *E. coli* can still lower *Pst* density.

2.2.6 Pathogen dose response curves

We next determined the dose-dependence of protection by varying *Pst* density at inoculation from an OD_{600} of $2x10^{-5}$ to $2x10^{-2}$ in an effort to capture both the low end and high end of infection. We tested our three isolates in addition to the two commercially available *Pantoea* strains. As previously, we measured both disease severity (Figure 2-7a) and overall *Pst* density (Figure 2-7b). We found that disease severity and infection dosage are positively correlated (GLM, adjusted R^2= 0.7, $F_{11,48}$= 13.31, p < 0 .0001). Both the seed inoculum treatment ($F_{5,48}$= 20.08, p< 0.0001) and infection dosage ($F_{1,48}$= 38.58, p< 0.0001) had significant effects on disease severity, with no significant interaction between the two ($F_{5,48}$= 2.281, p< 0.0612). Post-hoc Tukey tests reveal that all treatments are significantly less diseased than the no treatment controls (p< 0.006), and ZM1 is significantly more protective than both E325 (p= 0.02) and ZM3 (p= 0.04).

Pst densities on seedlings were also linearly correlated with infection dose, although weakly so (GLM, adjusted R^2= 0.4, p= 0.0003). Like AUDPC values, bacterial isolate ($F_{5,48}$= 6.5743, p< 0.001) and dose ($F_{1,48}$= 17.882, p= 0.001) had significant effects on *Pst* densities.

Here, we observe a significant interaction between the two ($F_{5,48}$= 7.486, p< 0.001), with negative controls having significantly more disease than every treatment (p<0.01) except for E325 (p=0.7383).

2.2.7 Protective isolate dose response curves

We varied the dose of *Pantoea* strains from $4x10^{-1}$ to $4x10^{6}$ CFU/seed. Given that they were cultured from TT4 seeds at an average of 40 CFU/seed (dashed vertical line, Figure 2-8b), this range captured five orders of magnitude above and 2 orders of magnitude below the biologically relevant abundance. After inoculation, we recorded disease severity of seedlings for 9 days, and calculated AUDPC (Figure 2-8a) and overall *Pst* density (Figure 2-8b). Seedlings that were inoculated with sterile buffer had an average AUDPC of 29+/- 2.3 and overall PST density of 9.96×10^{9} +/- 1.07×10^{9} copies/gram (gray dashed horizontal lines). Treatment at all doses, with the exception very low density treatments, were effective at reducing both disease severity and *Pst* density compared to seedlings which did not receive any protective bacteria. With regard to disease severity, AUDPC values negatively correlated with inoculation densities significantly (adjusted R^2= 0.7999, p< 0.0001), and there was a significant effect of treatment type on disease severity outcomes (ANOVA, $F_{4,65}$= 32.973, p< 0.0001). However, there was no linear relationship between final *Pst* density and inoculation densities (adjusted R^2= 0.0948, $F_{9,65}$= 1.861, p= 0.0740). In this case, a cubic model was the best fit to our data (adjusted R^2= 0.6567, $F_{19,55}$= 8.449, p< 0.0001). Under this model, we found that isolate type had a marginally significant effect on final *Pst* density (ANOVA, $F_{4,55}$= 2.4836, p= 0.05414).

In order to determine which *Pantoea* density resulted in the lowest *Pst* density, we used a SPLINE model to: 1) fit a cubic model to the data (curved line); 2) identify the minimum value on the curve (horizontal line); and 3) identify the distribution of the minimum value (shaded area). We found that for all three of our isolates, the distribution around the calculated minimum *Pst* density encompasses the concentration at which isolates were originally cultured for TT4 seedlings (diamond data points).

2.3 Discussion

Through a combination of culture-dependent and independent methods, we were able to directly test the protective effects of naturally occurring seed-associated microbiota, both in consortia and as single isolates. We found one particularly protective seed-associated microbial community that was able to significantly decrease the density of *P. syringae pv tomato* DC3000 (*Pst*) growth on seedlings and reduce disease symptoms across multiple tomato types.

Community profiling uncovered that *Pantoea spp.* dominated this seed microbiome, regardless of which seedling type it was applied to, and we were able to culture specific *Pantoea* strains directly from the surface of these seeds. When we applied these culturable isolates to seeds, we found that individual strains were as protective when applied in isolation as when combined. In order to understand how application density impacts protection, we varied the dose of isolates ZM1, ZM2, and ZM3, and we found a non-linear pattern of inoculation density correlation with pathogen density.

The seed surface is the primary site of contact between seed and fruits, and it is known to harbor a diversity of microbes across plant species. Despite this, few studies have included seed epiphytes when investigating seed-associated microbes, focusing primarily on seed endophytes (reviewed, [35]. Studies on seed-associated endophytes in maize [67], cucurbits [68], and rice [69] have revealed both high microbial diversity and the presence of conserved genera across

geographical, ecological, and host genotypic variations. In maize, seed-associated bacterial communities vary in accordance with host phylogeny, suggesting a close evolutionary partnership between the conserved microbiota and host plants [67]. Moreover, seed-associated microbiota can successfully colonize seedlings, although their community structure has been shown to change as seedlings develop [34]. Pairwise assays have also shown that certain individual isolates from seed endophytic communities can interact antagonistically to pathogens in eucalyptus [70], rice [71] and cucurbit seedlings [72], acting as key disease-resisting members. More recently, a study on the seed microbiota of *Brassica napus* revealed that inoculating seeds with bacterial strains can lead to resistance against a fungal pathogen in seedlings in a cultivar-dependent manner [73], further suggesting that specific seed endophytes can yield disease-protection effect in early development of the host.

Here, we used seed collection methods that allowed us to preserve the microbiota naturally occurring on the seed surface of four field-grown tomato types. Across four biological replicates (TT1-TT4), endogenous seed-microbiota were found to suppress disease symptoms in juvenile seedlings of their natural hosts when challenged with a common tomato pathogen, *Pst* (Figure 2-2). When TT4 microbiota was inoculated onto two other field tomato types, it was able to significantly reduce disease symptoms and decrease the density of *Pst* by 10 to 100-fold (Figure 2-3). Although the tomato types themselves differed in their overall susceptibility to disease, we did not observe that any single tomato type was more protected by the TT4 microbiome than another. This may suggest that the pathogen- suppressive effects of TT4, whether attributable to microbiome members with antagonistic activities against *Pst* or immune system priming, are capable of acting independently of their host genotypic context. Due to the way in which tomatoes were collected (which was based on plant location and fruit morphology rather than specific genotype of tomatoes), we were not able to point to the specific differences amongst host genotypes, but this would be useful in future studies. Furthermore, as we did not sequence the microbiome of adult plants from which seeds were collected, future work should explore if differences amongst seed microbiomes are driven by differences in the microbiome composition of the adult plants themselves. These microbiome differences may be a result of field location, host genotype, or other unknown factors. Our data suggest that it may be possible to breed plants to specifically recruit or harbor beneficial seed microbiomes that may ensure a more disease resistant crop in subsequent generations.

To better understand the protective effects observed, we sequenced the bacterial communities associated with seedlings inoculated with the TT4 microbiome and found the communities to be dominated by *Pantoea spp* (Figure 2-4). This is in line with community profiling results from the seed surface of *Triticum* and *Brassica* [62]. We then isolated culturable bacteria from seeds, and again found primarily *Pantoea spp*. Inoculation of seeds with our *Pantoea* isolates showed that they are highly protective against *Pst,* both in terms of colonization and disease (Figure 2-6). *Pantoea spp.* is a known antagonist of many bacterial [54, 74, 75] as well as fungal [62, 76] pathogens, and they are common biocontrol strains. *Pantoea dispersa* strain ZM1 appears to be novel and not previously described as a biocontrol species, but provides protection that is on par with, if not better than, currently commercially available strains. Genome sequencing will reveal if *P. agglomerans* strains ZM2 and ZM3 are novel biocontrol strains.

Our work also helps to disentangle the link between diversity and disease protection. Although there exists a speculative relationship between taxonomic diversity and the strength of a microbiome's disease-resistance effect, little empirical evidence exists to support or disprove

this. A recent study on the protective effect of a constructed community against *Pst* shows that variation in inoculum diversity affects disease-resistant effects in a significantly non-linear manner [28], demonstrating that increasing taxonomic diversity can have no impact, or even decrease, the protective effect of the community. In this study, seedling bacterial communities are low in richness and diversity, dominated primarily by one genus: *Pantoea*, although there are multiple species and strains of *Pantoea*. It is possible that a greater diversity of bacteria existed on the seeds, but we still find that inoculation of individual strains of *Pantoea* is sufficient in seedlings for protection against the pathogen used in this study. We are also aware that the fermentation step used to collect seeds might have enriched certain members of the seed microbiome that are able to survive acidic conditions. However, this may be a biologically relevant filtering step for epiphytic seed microbes, as seeds are likely to experience acidic conditions both during fermentation of fruit in the field or through the digestive track of animals.

Although we only test the protective ability of isolates against a bacterial pathogen, the *Pantoea* isolates and the *Bacillus* isolate may have other growth promoting capabilities as well, as a recent paper describes various growth promotion traits of tomato seed endophytes (in addition to disease protection) [61]. They may also have protective effects against fungal pathogens, as has been previously demonstrated in *Pantoea* species [62, 76]. Practically, seed-associated bacteria are an excellent target for probiotic/biocontrol application, and it may even be possible to apply the protective strain to the flowers of the previous generous, as was demonstrated with a plant growth promoting endophyte [77]. Taken together, these studies and our results suggest that the common agricultural practice of seed sterilization may be disrupting persistent mutualisms between plants and microbes across generations. While seed sterilization is an agriculturally important procedure to purge seed-transmitted pathogens, we and other groups have shown that it may also be removing beneficial symbionts. How the simultaneous disruption of pathogenic and mutualistic symbioses would impact host health over ecological time scales, and how agricultural practices should preserve the beneficial traits conferred by the vertically transmitted microbiome while still preventing the spread of pathogens, are outstanding questions in need of future research.

The focus of this work was to examine the potential protective effects of seed epiphytic communities rather than describe the mechanisms underlying protection. Previous work has demonstrated that some *Pantoea spp.* are protective through antibiosis activity [54, 74], or it may also be mediated through competition for resources [29]. In addition to direct interactions between microbes, application of *Pantoea spp*. to seeds ensures that germinating seedlings are in immediate contact with microbes, and this may prime the plant's immune system so that it is better able to mount a response against *Pst*, thus indirectly protecting against disease. In our experiments, the data suggest that both direct and indirect mechanisms are mediating protection. We find that all strains of live bacteria, including a non-plant associated strain of *E. coli*, are capable of decreasing disease severity symptoms when compared to non-treated controls. When seeds are treated with UV-killed bacteria, we find that none of the strains are capable of decreasing disease severity. Our results suggest that all bacteria included in our experiment can protect seedlings against *Pst* through direct interactions, as UV-killed bacteria were unable to decrease disease severity. Additionally, we found that all live isolates except for *E.coli* lowered *Pst* densities in seedlings, suggesting that this characteristic may be unique to our *Pantoea* isolates. When seedlings were treated with UV-killed bacteria, we again found that ZM1, ZM2, and ZM3 were capable of lowering *Pst* densities, but *E.coli* was not. This suggests that some of the protective capability we are observing is conferred through indirect mechanisms, perhaps

through immune activation by UV-resistant membrane-bound antigens. The inability of UV-killed *E. coli* to decrease *Pst* density suggests that these *Pantoea* strains may have plant host or pathogen specific protective traits. Future work will explore the protective ability of these isolates in adult plants and will further dissect direct versus indirect mechanisms of protection.

By varying the concentration of *Pst* inoculated onto seedlings, we observed that increasing *Pst* increases AUDPC, as expected (Figure 2-7). Interestingly, we also observed a decoupling of plant disease symptoms and pathogen density. This was similarly observed when we tested for protective effects of each isolate (Figure 2-6). Here, we saw a linear increase of disease severity as *Pst* inoculation was increased in inoculation density, but saw a much weaker linear correlation between dose and *Pst* densities. We posit that this non-linear increase of *Pst* density might either be due to 1) a carrying capacity of *Pst* density that is reached on the seedling leaves, or 2) the possibility that the detection of dead or inactive *Pst* cells disguises a linear pattern. Furthermore, disease severity was calculated based on foliar symptoms, but it is very likely that *Pst* also asymptomatically colonizes the seedling root tissue. The entire seedlings, including roots, were homogenized prior to *Pst* quantification; this may have obscured differences in foliar *Pst* densities.

By varying the dose of the protective strain (Figure 2-8), we were able to find that an increased dose does not necessarily correlate with decreased pathogen density, as was recently uncovered in a study investigating the protective effects of the phyllosphere community in adult tomato plants [28]. We vary the dosage of protective *Pantoea* strains from less than one CFU/seed to 10^8 CFU/seed, the highest of which is five orders of magnitude higher than the concentration at which we originally recovered bacteria on the seeds (Figure 5). When analyzing *Pst* density seven days after inoculation, all culturable isolates' ability to *Pst* suppress growth resulted in a non-linear pattern of pathogen density, whereby increasing *Pantoea* does did not linearly correlate with decreasing *Pst* density. The same is true for the two commercially available biocontrol strains. Most notably, all three TT4 isolates exhibit optimal suppression of *Pst* at densities close to that found in naturally occurring seeds (40 CFU/seed). At isolate densities above 10^4 CFU/seed, *Pst* density as detected by ddPCR, reached a similarly high level for all strains, suggesting a maximum density beyond which additional cells of the protective strains do not result in further protection. In light of our results that UV-killed *Pantoea* are capable of decreasing *Pst* density through presumed plant-immune activation, we posit that this activation, or priming, may be dependent on bacterial density on the seeds. In such a model, induction of resistance responses in the plant would be fully activated when such a threshold of signal is achieved. This is further supported by the result that UV-killed isolate C9-1 was the only *Pantoea* isolate unable to decrease *Pst* density (Figure 2-6d), and its dose response curve (Figure 2-8b) was also the only one that did not follow the cubic pattern observed in the other *Pantoea* isolates.

To rule out the possibility that higher densities of *Pantoea* resulted in the killing of *Pst,* a scenario that would be undistinguishable because of the use of ddPCR to quantify the pahtogen, we treated samples with a PMAxx ™ and repeated the ddPCR. PMA treatment blocks the amplification of dead cells and allows the measurement of only live cells. The data are quantitatively similar (data not shown), indicating that even when only live cells are quantified, *Pst* densities reach an asymptote. Our results are suggestive of the possibility that plants may not only preferentially passage beneficial symbionts, (also recently shown in [61], but it is possible that they are being passaged at an optimal density as well. This density may be driven by nutrient or physical characteristics of the seeds, or it may reflect a concentration that is most effective at

inducing a host immune response in seedlings. More empirical data are needed to more robustly test these hypotheses.

Our work also demonstrates an important practical use for dose-response curves. When a biocontrol agent is applied in the field, its goal is usually to decrease disease severity symptoms in the host. However, in the case of food-borne human pathogens, the goal is likely to decrease total numbers of bacterial cells. The same is true for decreasing likelihood of disease transmission, or secondary effects caused by pathogens such as frost damage. We found that the optimal dose for decreasing disease severity may not be the same as that needed to decrease overall bacterial abundance. Indeed, increasing the titer of biocontrol strains may indirectly increase the number of pathogenic cells when it exceeds a biologically relevant dosage.

In conclusion, this study explores the epiphytic seed microbiome from both an ecological and applied perspective. Our fruit sterilization method is a simple yet effective modification of previous studies that allowed us to study only biologically relevant microbes on the surfaces of seeds. Through both culture-dependent and independent approaches, we described a complete microbiome, dominated by key species, which confer protection in seedlings against a common and agriculturally relevant pathogen. Furthermore, our finding that increasing the dose of a protective strain does not necessarily imply a decreased pathogen load has critical importance in the field of biocontrol and probiotics. As a whole, this work highlights the importance of approaching questions from multiple perspectives, as often these different approaches are likely to inform one another and increase both the validity and applicability of the results. Overall, this work not only contributes to our understanding of the ecological importance of seed-associated microbes, but it also provides novel insight as to how agricultural processes can be informed by natural patterns and used to enhance plant health and fitness.

2.4 Materials and Methods

2.4.1 Tomato fruit and seed collection

Tomato fruits were collected from UC Davis Student Organic Farm in September 2017. We collected mature, intact fruits from a total of four different tomato types based on distinct fruit morphologies and field locations including: orange cherry tomatoes, red cherry tomatoes, medium-orange-sized tomatoes, and an unidentified heirloom variety. Fruit of the same tomato type/cultivar were collected from multiple plants planted in the same row, resulting in four tomato types (designated as Tomato Type 1-4). Tomato Type 1-3 were collected from non-neighboring lanes from one field, and the heirloom variety (Tomato Type 4; generated by the Student Collaborative Organic Plant breeding Education (SCOPE) program; USDA-NIFA award # 2015-51300-24157) was collected from a neighboring field. Fruits were transported to UC Berkeley on ice and immediately stored in 4°C until processing. Intact tomato fruits from the same type were pooled in a sterile 1L beaker until they reached roughly the 500 mL line (approximately 10-20 tomatoes). To ensure that no additional microbes other than those found naturally were introduced to the seed surface, we surface sterilized the tomato fruits themselves before processing of seeds. Tomatoes were submerged in 75% ethanol for 20 minutes. They were then washed with sterile double-distilled H_2O (ddH_2O) three times. The last wash was plated onto Kings Broth (KB) agar [78], and no colony forming units (CFUs) were detected. Sterilized tomatoes were then pooled into another sterile one-liter bottle, crushed with sterile forceps and spatula until becoming a thick fruit mixture, and allowed to ferment at room temperature (approximately 20°C) for seven days. We employed this as a common seed collection method for removal of seeds from the fruit endocarp. After fermentation, seeds were then strained out from

the fermented liquid with a sterilized metal strainer, minimally washed with sterile ddH_2O to remove any excess fruit, and dried on filter paper within sterile petri dishes. All procedures were carried out sterilely in a Biological Safety Cabinet. Harvested seeds were stored in sterile petri dishes in darkness at 21°C, and these same seed stocks were used for all experiments. The first experiment was performed less than one month after collection, and the last experiment was done six months after collection.

2.4.2 Germinating TT1-TT4 seeds with/without their endogenous microbiota

For each experimental replicate (Figure 2-1a) within an experiment (six per treatment), 6 seeds from each field-collected tomato type were placed into sterile 1.5 mL Eppendorf tubes and submerged in 400µl of sterile 10 mM $MgCl_2$ solution and sonicated for 15 mins in a Branson M5800 sonicator. This sonicating water bath is different from laboratory sonicators used to disrupt cells; instead, these baths will dislodge bacteria with minimal disruption of their cell integrity. The liquid was then transferred into new sterile tubes and used as seed microbiome inocula. Prior to inoculation, seeds were surface sterilized using the following procedure: Seeds were first soaked in 2.7% bleach (sodium hypochlorite) solution for 20 minutes, then washed with sterile ddH_2O three times to remove any excess bleach. The last washes were plated on KB agar plates and incubated at 28 °C for 24 hours. No culturable bacteria were present in any post-sterilization seed wash from any experiments. After sterilization, 40µl of the original seed wash (bacterial inoculum) was pipetted directly on top of each individual seed. We did this so that each seed would receive roughly the same number of microbes that was removed during the sonication step. The removal and re-addition process was done, in general, so that every seed used in the experiment would undergo the exact same procedure, and the only difference would be receiving microbiota or not. Negative control seeds were each inoculated with 40µl of 10 mM $MgCl_2$.

2.4.3 Germinating TT1 and TT2 with/without TT4 microbiota

In this experiment, microbiota was removed from TT1, TT2, and TT4 seeds via surface sterilization, as described above. In this case, for every experimental replicate of TT1 and TT2 (each with six seeds), an equivalent number of TT4 seeds from our original seed stock were used to generate microbial seed inoculum. We had a total of five experimental replicates per tomato type, per treatment. After sterilization, TT1, TT2, and TT4 seeds were all inoculated with either TT4 microbiota or 40µl of 10 mM $MgCl_2$. As above, 40µl of seed wash (or buffer) was pipetted directly on top of each seed within an experimental replicate.

2.4.4 Culturing and growth of isolates

In order to culture bacteria from the seeds used in this experiment, seeds were sonicated into sterile buffer, as above. Next, the seed wash was diluted 1:10 in sterile 10 mM $MgCl_2$ solution and plated onto KB agar and Lysogeny Broth [79] agar. They were incubated for 48 hours at 28 °C. We were only able to culture bacteria from tomato types 4 and 2. On average, we cultured 40 (+/- standard error 0.53) colony- forming units (CFUs) from each TT4 seed. To isolate individual strains from the microbial community, we picked morphologically distinct colonies, based on color and surface, and streaked them on new nutrient agar where they were grown for 24 hours at 28 °C. Liquid cultures were attained by inoculation into liquid KB and grown on an orbital shaker at 28 °C overnight.

2.4.5 Germinating Money Maker seeds with/without live bacterial isolates

For consistency amongst tomato plant hosts, Money Maker seeds were used for all further experiments measuring the impact of particular seed-associated microbiota. Seeds were sterilized as described above. In addition to testing our own bacterial isolates ZM1, ZM2, and ZM3, we also included Biological Control strains, kindly provided by Dr. V. Stockwell, Oregon State. These two strains are *Pantoea agglomerans* strain E325A and *Pantoea vagans* strain C9-1. Bacterial inoculum was prepared as follows: isolates were grown overnight on an orbital shaker in LB at 28 °C. We measured the optical density at 600 nm (OD600) of the overnight culture and plated the culture on LB agar, incubated overnight at 28 °C to obtain their CFU counts. The remainder of the liquid culture was stored at 4 °C overnight. The next day, we calculated a CFU to OD ratio, and re-measured the OD to account for any growth that occurred of the liquid culture overnight. We pelleted the bacteria at 4000 X G for 5 minutes, re-suspended in sterile 10 mM $MgCl_2$ solution, and diluted to the appropriate concentration. Each seed was inoculated by pipetting the bacterial culture directly on top of each individual seed. In Figure 2-6, we inoculated seeds with pure cultures at a final inoculum density of 40 CFU/seed to approximately match the observed natural densities. Each experimental replicate held four seedlings, and we had three experimental replicates per isolate per treatment. Disease severity was monitored for 10 days after plate flooding (see below). For dose response curves, the density of bacteria applied to the seeds ranged from $4x10^{-1}$ to $4x10^{6}$ CFU/seed with control replicates not receiving any, and disease was monitored for nine days. We did not replicate at the plate level for dose response curves.

2.4.6 Effects of UV-Killed bacterial isolates

Liquid bacterial cultures were prepared as above. In addition to *Pantoea* strains, we included a non-environmental bacterium: *E. coli* BW25113 [80]. To UV kill bacteria, 500µl of diluted cultures were exposed to UV radiation (intensity of 15 mW / cm^2 / sec) in 24-well plates without a lid for a total of two hours. A portion of the culture was plated onto KB Agar to ensure absence of any viable CFU. Plates were incubated at 28 °C for the first 48 hours and then room temperature for the subsequent 5 days. No growth was detected from UV killed isolates. Bacteria were applied to seeds as above. After inoculation of *Pst*, disease was monitored for 8 days.

2.4.7 Tomato seedling growth

In all experiments, seedlings were grown on 1% agar plates. Inoculated seeds were kept in the dark at 21°C for six days and transferred into growth chamber at 25°C, 65% humidity and 16 h daylight per day for 2 days until challenged with *P. syringae* pv. *tomato* strain DC3000 (see below: flooding assay).

2.4.8 Pathogen flooding assay

The pathogen used in this work is *Pseudomonas syringae* pathovar tomato strain DC3000 (*Pst*). To challenge the seedlings with the pathogen, we modified a previous *P. syringae* flooding assay [81]. *Pst* was grown overnight in KB at 28 °C, and OD_{600} was measured of the overnight culture. The culture was diluted to OD_{600}=0.0002 in sterile 10 mM $MgCl_2$ solution, representing approximately 5 x 10^3 CFU/mL. We further added Silwet at 0.015% by volume to the diluted culture to ensure even dispersal of the pathogen. We flooded each 1% water agar plate of seedlings with 6 ml of the final *Pst* culture. The *Pst* culture was deposited directly on top of cotyledon leaves to ensure contact between *Pst* and the leaf surface. Flooded plates were shaken

gently on an orbital shaker for 4 minutes; the liquid was then poured off of the plates. For negative control seedlings, sterile 10 mM $MgCl_2$ solution was used in place of *Pst* culture. Infected tomato seedlings were kept vertically with cotyledons up in growth chamber at 25°C, 65% humidity and 16 h daylight per day until final collection. For the pathogen dose response curve (Figure 2-7), seedlings were flooded with *Pst* density ranging from OD_{600} of $2x10^{-5}$ to $2x10^{-2}$.

2.4.9 Disease severity scoring

Each experimental replicate plate held four to six seedlings (Figure 2-9). The disease symptoms of every individual seedling were scored daily after flooding until the final collection. All seedlings were randomly assigned numbers and scored blindly for every experiment. The duration of experiments ranged from 8 to 14 days (see X axis of figures, and methods above, for details). As has been done in previous plant disease assays [82, 83], we established a series of disease symptom scores that correspond to different levels of disease severity, with chlorosis and speckling as the two primary symptoms. To optimize both the descriptive power and the accuracy of our scoring, we set the following scoring indices: 1 = mildly diseased, 2 = moderately diseased, 3 = severely diseased, and 4 = loss of leaf or death (See Figure 2-1b for representative photographs of disease severity levels). Individual seedlings were scored blind daily for the duration of the experiments. Seedlings within one experimental replicate (Figure 2-1a, Figure 2-9) were then pooled into one 15mL conical and processed as described below. Disease severities of individual seedlings within an experimental replicate were averaged, and these data are displayed as a single data point to avoid problems of pseudo-replication. We analyzed these data in two ways. First, we used repeated measures Generalized Linear Model to analyze disease progression of individual experimental replicates over time. We then calculated the area under the disease progression curve (AUDPC;[84] as a cumulative measure of disease symptoms over time, and analyzed treatment effects using ANOVA.

2.4.10 Seedling collection and processing

Seedlings from each experimental replicate were collected with sterile forceps and combined into 15 mL conical tubes and weighed. Seedlings were submerged in 5 mL of sterile 10 mM $MgCl_2$ solution, homogenized with round ceramic beads for 40 seconds at 4m/sec for 35s using a Fastprep-24^{TM}5G. The final homogenates had little or no large visible plant material.

2.4.11 ddPCR quantification of bacteria

Pseudomonas syringae density was quantified from each experimental replicate using droplet digital PCR using a fluorescent probe targeting the *Pseudomonas* 16S gene [85] as fully described elsewhere [86]. Briefly, seedling homogenates were diluted 1:10, and 2µl of homogenate was used as template in the BioRad QX200 ddPCR reaction. In analyzing positive droplets, all thresholds were set using negative, no template controls and positive pure *Pst* DNA controls. As with analysis of AUDPC, *Pst* densities are a measure of each plate experimental replicate, as described above. Bacterial abundances are normalized to total seedling weight within a plate and reported as copy of 16S rRNA gene per gram of plant material. For negative ddPCR controls, we always attempted to measure *Pst* in the $MgCl_2$ inoculated plant controls for all experiments as well as *Pantoea* DNA. Although the *Pst* probe was designed to be specific to Pseudomonads, we did this to ensure our probe was only amplifying *Pst* and not *Pantoea* nor any plant material. The signal amplitudes for sterile plant-only and *Pantoea* isolates-only controls

were the same as those of no-template, sterile ddH_20 controls, indicating that indeed, there was no detectable background amplification of *Pantoea* species when using the *Pseudomonas* probe, and no *Pseudomonas* was present in the negative controls.

2.4.12 DNA extraction of isolates and community

Homogenized seedling material was used for DNA extractions for community profiling using the Qiagen PowerSoil DNA extraction kit. For bacterial isolates, DNA was extracted using the Qiagen Blood and Tissue kit protocol for gram-negative bacteria.
DNA extractions of all bacterial isolates from the seed wash were performed using the QIAGEN DNeasy Blood & Tissue Kit. Samples were taken with sterile pipet tips from axenic colonies. We followed the kit's accompanying protocol for DNA extractions of gram-negative bacteria. Success of extraction was verified by NanoDrop nucleotide quantification.

2.4.13 16S Amplicon sequencing of the seedling communities

The 16S rRNA gene was amplified using dual-indexed primers designed for the V3- V4 region [87] using the following primers: 341F (5 -CCTACGGGNBGCASCAG-3) and 785R (5 -GACTACNVGGGTATCTAATCC-3) [88]. Additionally, we used peptide nucleic acids, PNAs [89] to decrease amplification of plant mitochondrial and chloroplast DNA. For community profiling, we also sequenced a DNA extraction of the reagents from the DNA extraction kit and our negative PCR control along with experimental samples. Reaction conditions were 94°C for 3 min, 94°C for 45 s, 78°C for 10 s, 50°C for 1 min, 72°C for 1.5 min, repeat steps 2–5 30 times and 72°C for 10 min. PCR mixtures were randomized in order, run in triplicate for each sample, pooled and quantified using Qubit. Amplicons from each sample were pooled in equimolar concentrations, cleaned using an AMpure bead clean-up kit. Libraries were prepared for paired 300-nucleotide reads in Illumina's MiSeq V3 platform (Illumina) at The California Institute for Quantitative Biosciences (QB3) at UC Berkeley and run in 1 lane.

2.4.14 Identification of bacterial isolates

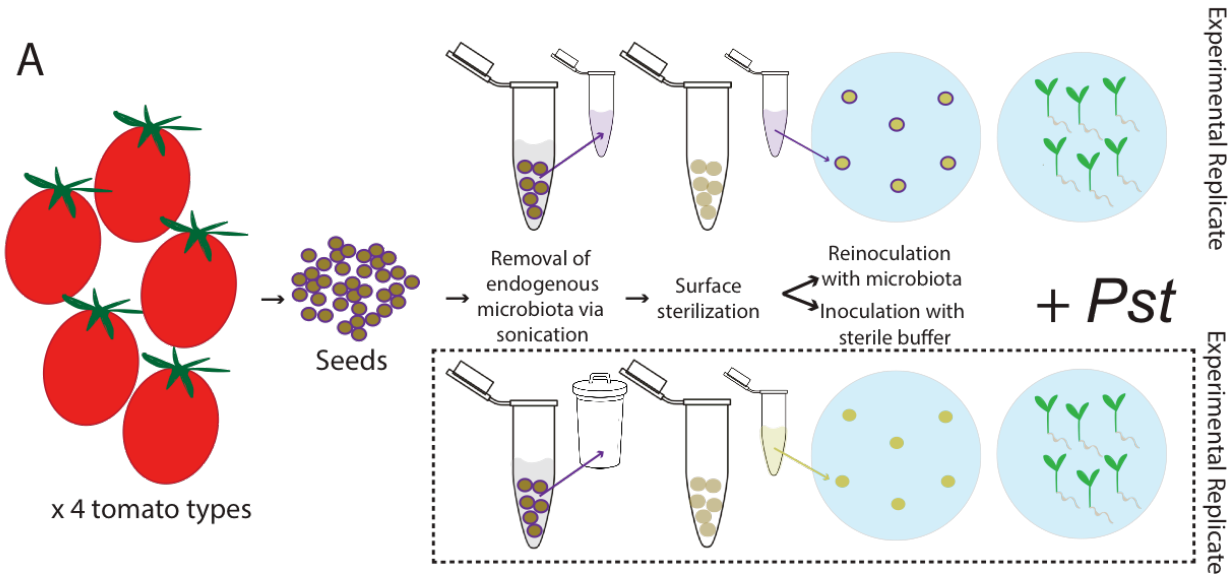
Using pure DNA of individual bacterial isolates, we amplified and sequenced 16S genes using 27R and 1492R primers (Lane 1991). In addition to 16S, we sought to further discriminate against our potential strains, and so we amplified the *gyrB* gene and *rpoB* genes with previously published primers and PCR protocols [64, 65]. We performed a BLASTn search of all isolate sequences and recorded the top hits with the highest identity (Table 2-1). Phylogenetic tree of isolates and neighbors were built using *gyrB* sequences. We placed our isolates within a subset of samples previously mapped in a *Pantoea* phylogenetic tree by Rezzonico et al. [66]. Dr. T. Smits (personal communication) kindly provided the E325 *gyrB* sequence. The evolutionary history of our isolates and other strains was inferred by using the Maximum Likelihood method based on the Tamura-Nei model [90]. The tree with the highest log likelihood is shown. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 13 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 316 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 [91].

2.4.15 Amplicon sequencing analysis

MiSeq sequencing files were demultiplexed by QB3 sequencing facility. Reads were combined into contigs using VSearch [92], and the remainder of the analysis was carried out in Mothur version 1.41.3 [93] following their MiSeq SOP [94]. Data were quality-filtered by length, ambiguous bases, and homopolymer length using the recommended Mothur parameters. Chimeras were removed using UChime [95]. We used a 99% similarity cut-off for defining OTUs. The Silva reference database [96] was used for sequence alignment and taxonomic assignment. Archaeal, chloroplast, mitochondrial and unknown domain DNA sequences were removed. Once an OTU table was generated in Mothur, the remainder of the analysis was performed in R using the Phyloseq package version 1.19.1 [97] and Vegan package version 2.4-5 [98]. To account for reagent contaminants, we also sequenced DNA extraction kit controls and PCR controls along with our samples. Contaminant OTUs from control samples that were at a similar or higher relative abundance in control samples compared to experimental samples were removed from the full OTU table. Data were rarified to 50,000 reads per sample and singletons were removed.

2.4.16 Statistical analysis

All statistical tests were done using R (version 3.4.4) and SPSS 24.0 (SPSS, Chicago, IL, USA). Details of statistical tests and results are reported in the text. Graphs were plotted in R using the ggplot2 package version 2.2.1 [99]. Individual data points represent the value averaged across seedlings within an experimental replicate.



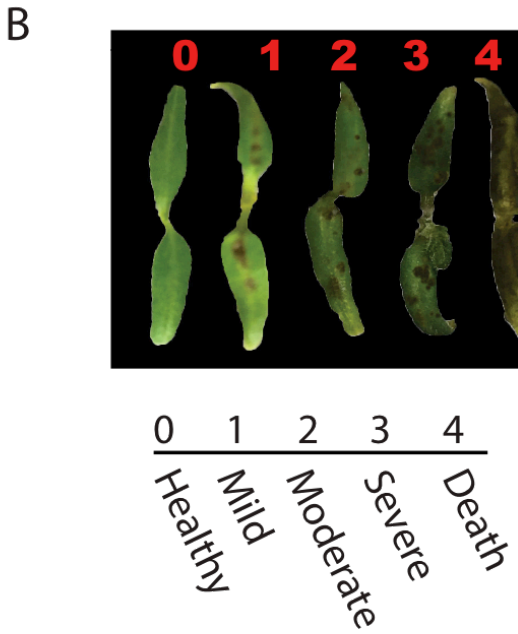


Figure 2-1 Experimental Design
Seeds from each tomato type, or biological replicate, were depleted of their epiphytic microbiome, sterilized, and then restored with their microbiome or not. Experimental replicates consisted of 6 seeds (Figure a). After germination and challenge with the pathogen, a disease severity index (b) was used to rate seedlings on a scale of 0 to 4 for four different seed microbiome sources.

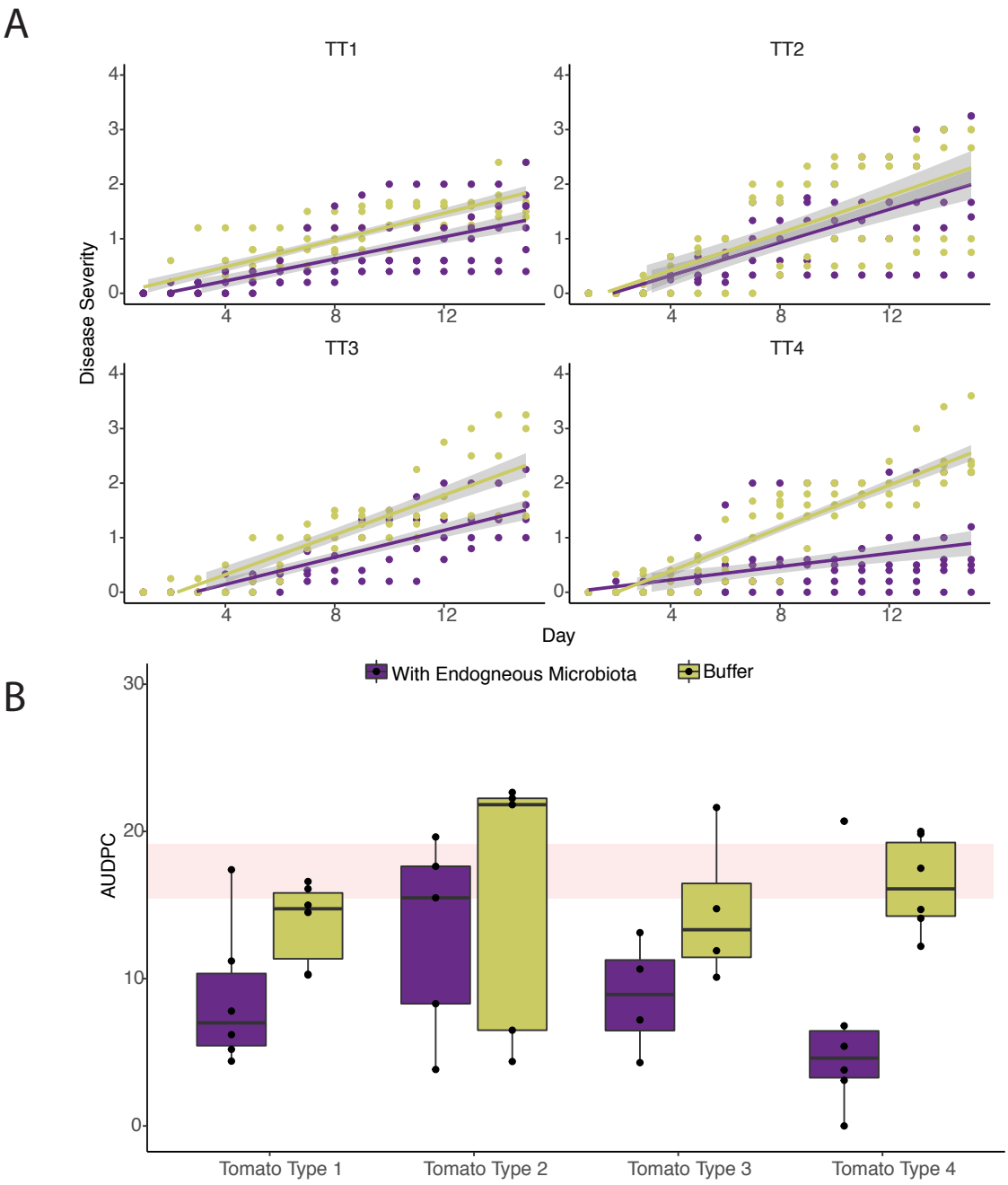


Figure 2-2 Protective ability of epiphytic seed microbiota
Disease severity was tracked for 15 days and averaged across seedlings within an experimental replicate, fit to a linear regression and plotted with confidence intervals (shading) (a). Total area under the disease progression curve was calculated for replicates in each treatment and graphed as boxplots (b) where black line indicates median, the upper portion of the box is third quartile, lower portion is first quartile, whiskers are maximum and minimum, and outliers are black dots. The red shaded area indicates an AUDPC at which 50% or more of seedlings on a plate were dead.

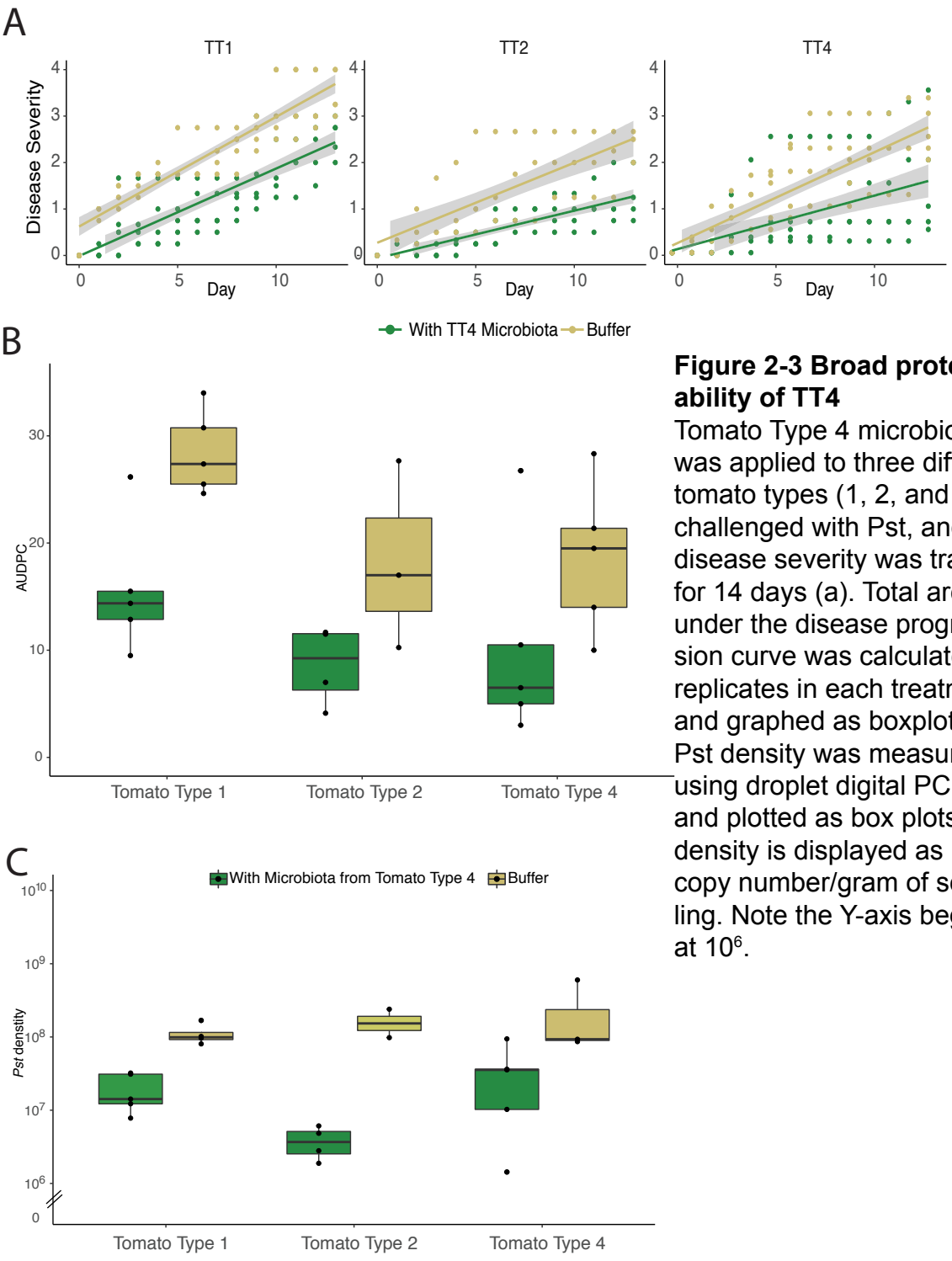


Figure 2-3 Broad protective ability of TT4
Tomato Type 4 microbiota was applied to three different tomato types (1, 2, and 4), challenged with Pst, and disease severity was tracked for 14 days (a). Total area under the disease progression curve was calculated for replicates in each treatment and graphed as boxplots (b). Pst density was measured using droplet digital PCR (c) and plotted as box plots. Pst density is displayed as 16S copy number/gram of seedling. Note the Y-axis begins at 10^6.

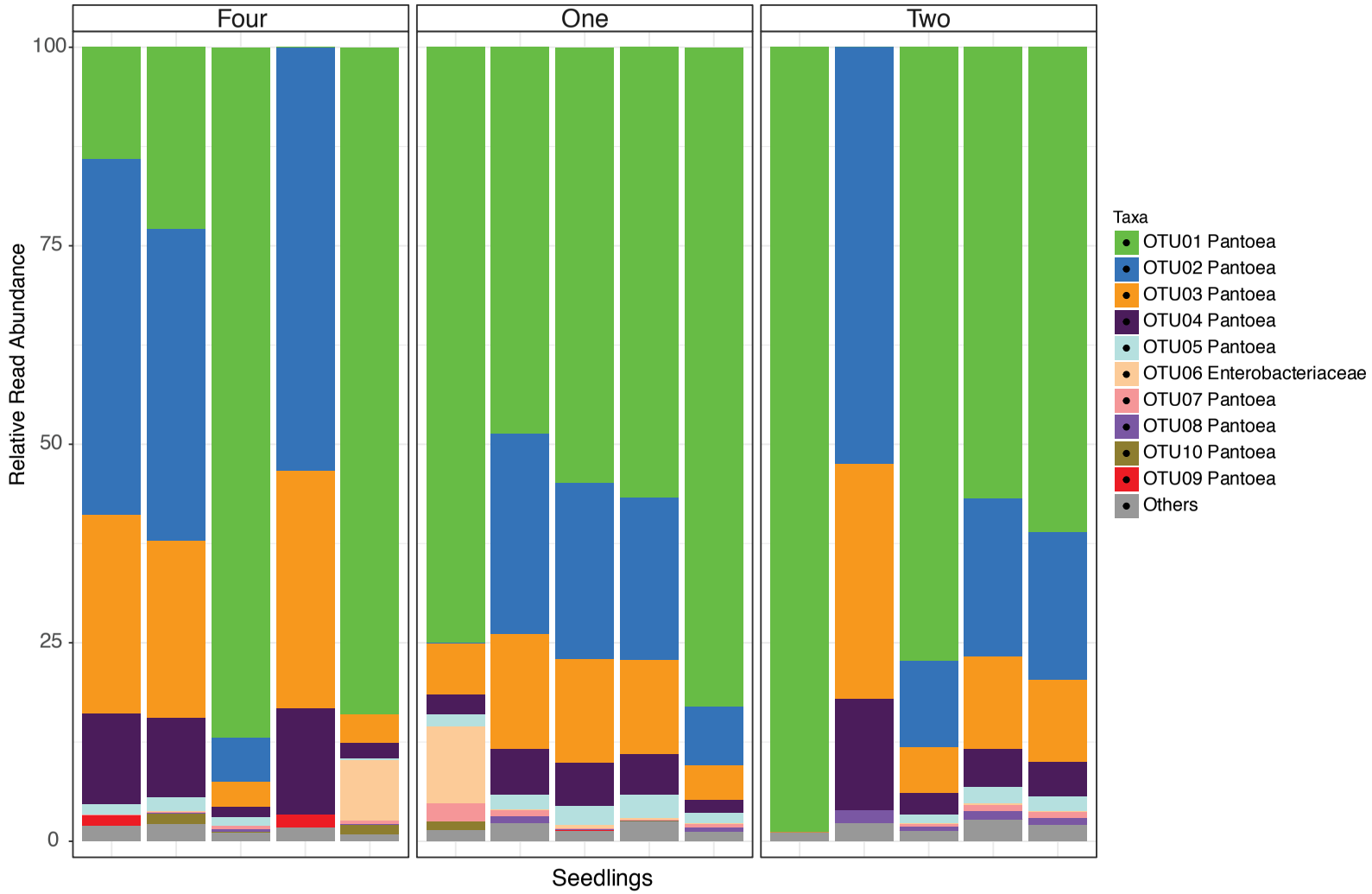


Figure 2-4 Bacterial communities of seedlings innoculated with TT4
Relative abundance of top OTUs in seedlings that were inoculated with TT4 complete microbiomes are displayed with different colors representing different 99% cutoff OTUs

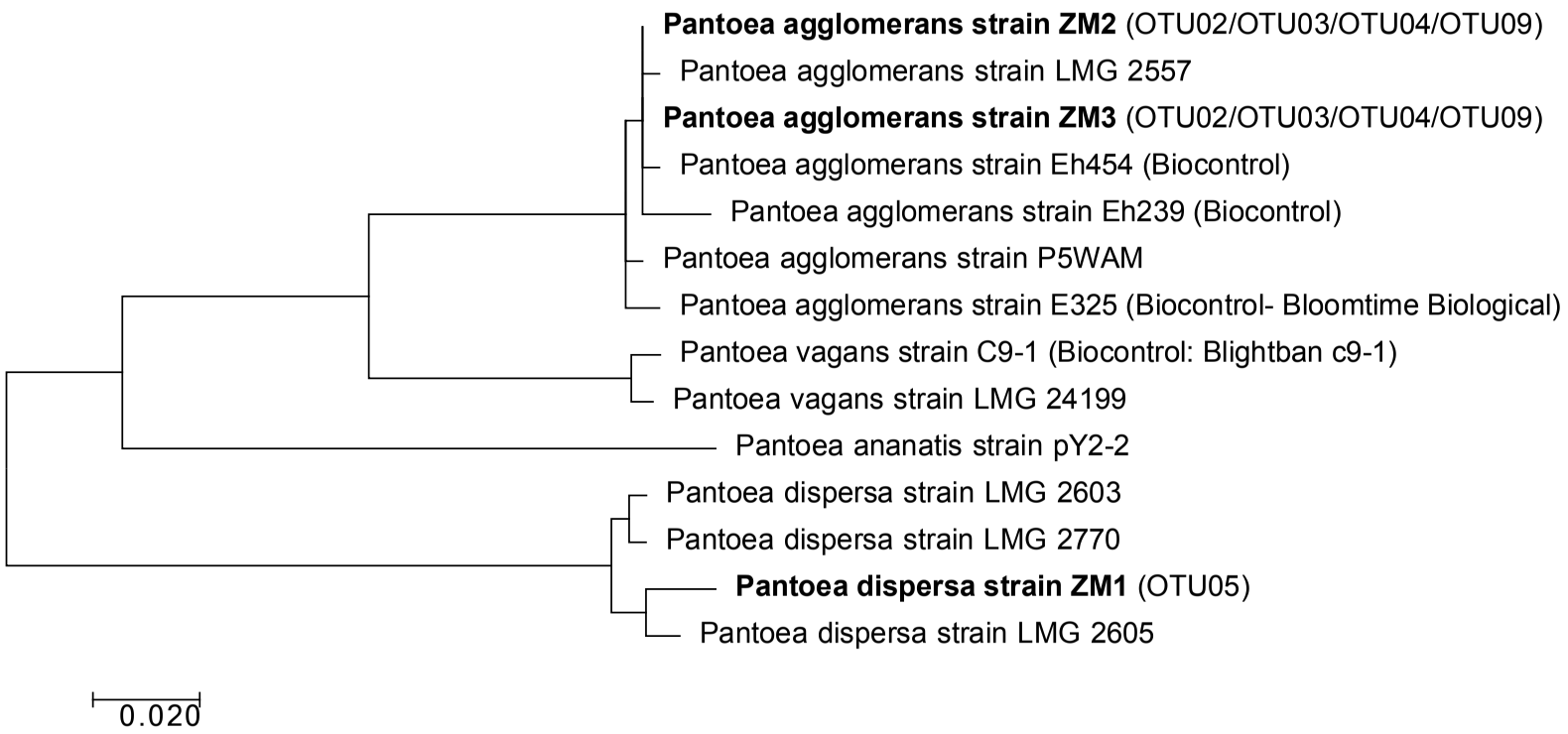


Figure 2-5 Taxonomic identity of *Pantoea* isolates
Isolated Pantoea strains are placed within a phylogenetic tree using the maximum likelihood method based on the Tamura-Nei model (b). Pantoea isolated in this study are bolded and known biocontrol strains are labeled as “Biocontrol”. Details of tree building are included in methods. The OTUs whose sequences aligned 100% to the 16S Sanger sequences are indicated after the isolate name

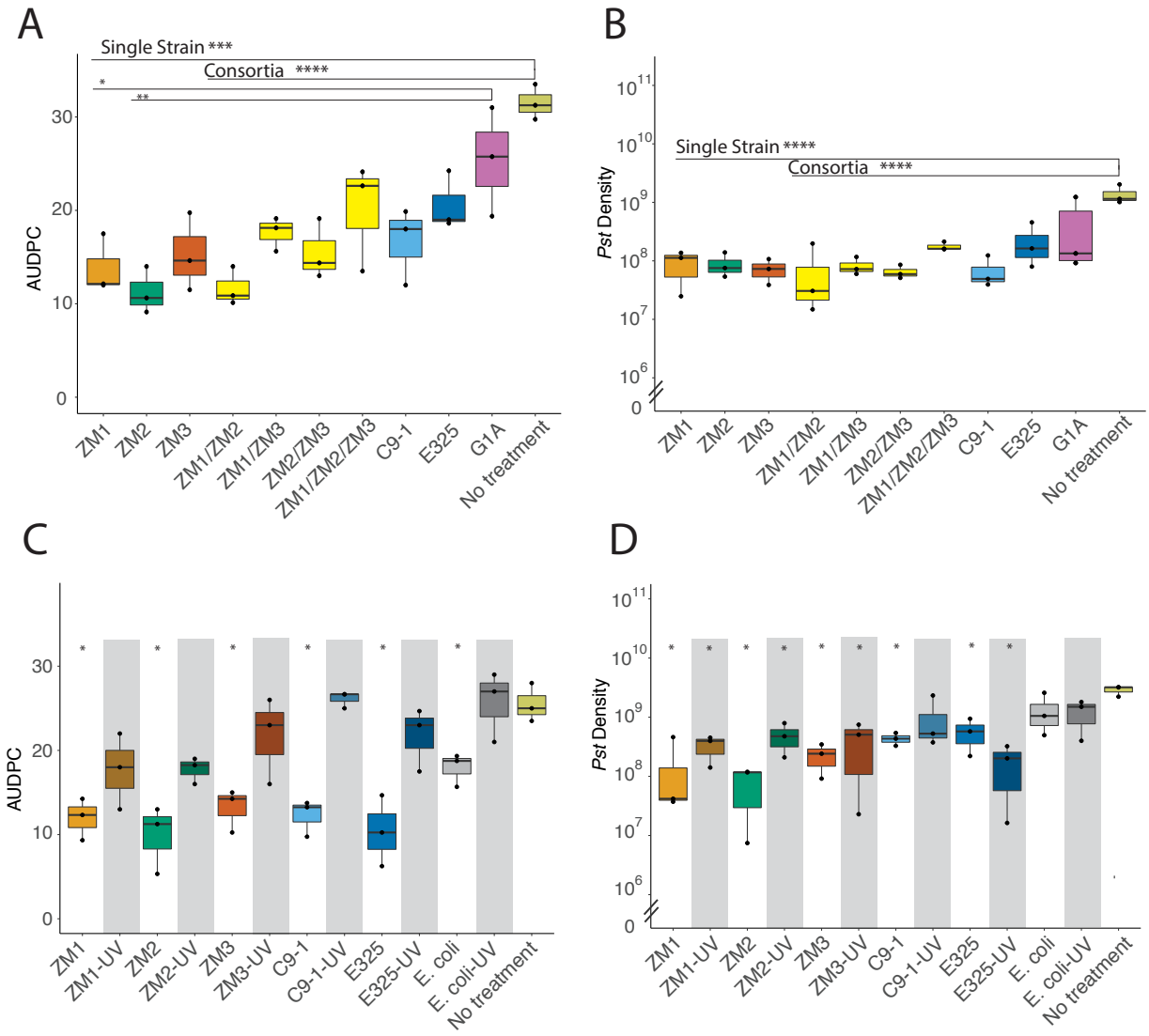


Figure 2-6 Protective ability of bacterial isolates

Money Maker seeds were inoculated with all isolates, combination of isolates, or sterile buffer “no treatment” control and challenged with Pst. and disease severity was tracked. AUDPC was calculated for each replicated and is compared across isolates (a). Pst densities for all treatments were also measured (b). The protective ability of UV-killed strains was also tested. AUDPC was calculated (c), and Pst density was measured (d). Note the Y-axis begins at 106. In panels c and d, significance values indicate results of a Welch’s t-test comparison to no-treatment controls with a correction for multiple comparisons. * p≤0.05; ** p≤0.01; *** p≤0.001; ****p≤0.0001.

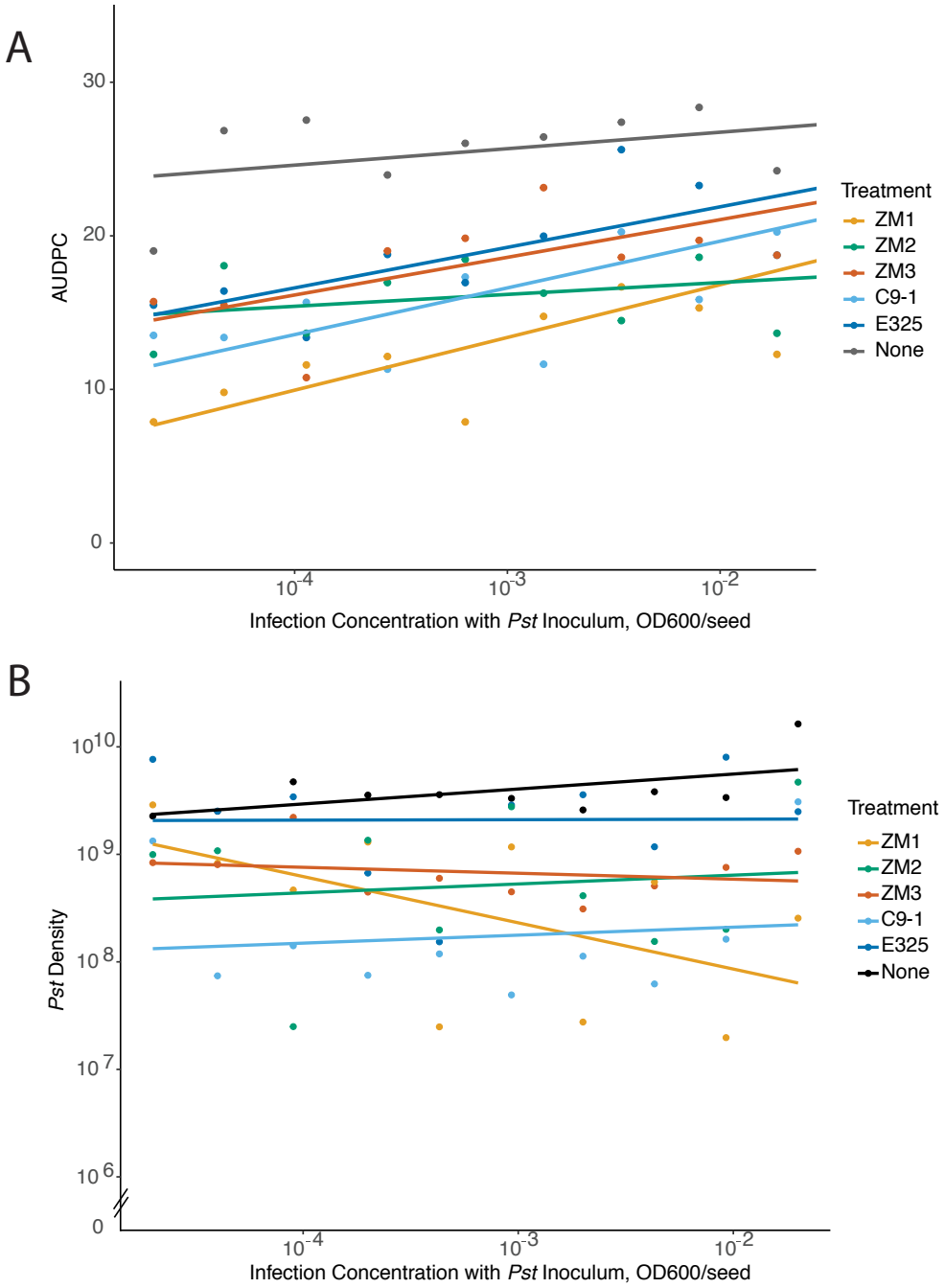


Figure 2-7 Pst dose response curve
Pathogen density inoculation was varied, and AUDPC was calculated for each treatment (a). Additionally, total Pst was quantified measured (b). Note the Y-axis begins at 10^6.

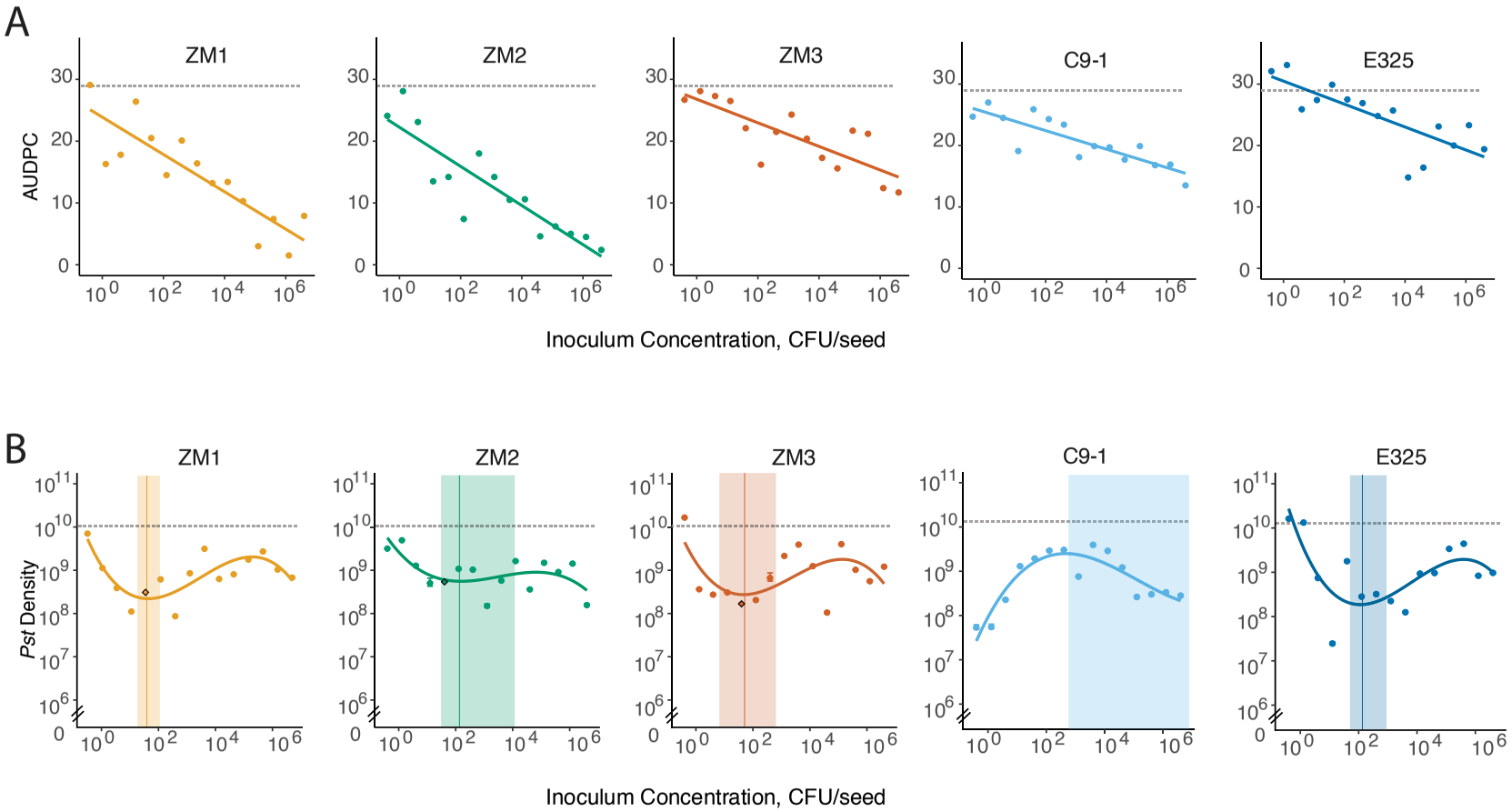


Figure 2-8 *Pantoea* Dose response curve

Seeds were inoculated with varying densities of protective strains, challenged with *Pst* and disease severity was tracked for 9 days. AUDPC was calculated (a), and a linear model was fit to the data. The dashed horizontal line indicates the mean AUDPC of control seedlings that did not receive any treatment of protective isolates. *Pst* densities were also measured (b). A cubic model was fit to the data (curved line). Vertical dashed line indicates the predicted minimum value from a SPLINE model used on a cubic fit to the data; the shaded area indicates the distribution of the minimum range. The diamond data point outlined in black indicates the concentration at which Pantoea was originally cultured off the seedlings. Error bars represent the Poisson Error of the ddPCR measurement of *Pst* density. Pst density is displayed as *Pseudomonas* 16S copy number/gram of seedling.

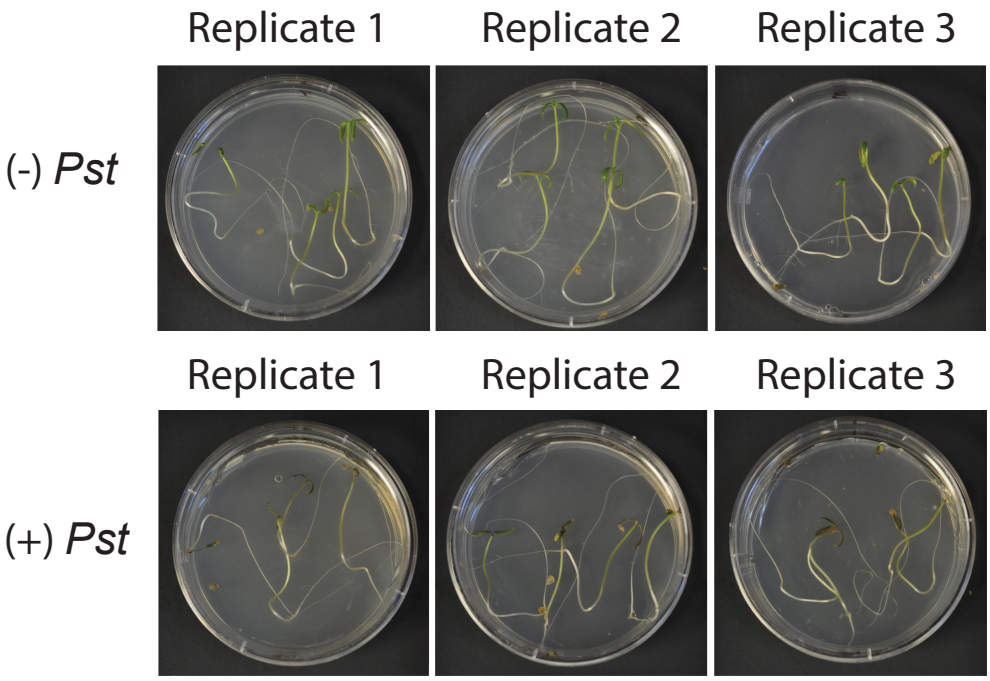


Figure 2-9 Disease severity across replicates within treatment
Three experimental replicates with and without Pst treatment in which disease severity can be observed at the individual seedling level.

Chapter 3. Successive passaging of a plant-associated microbiome reveals robust habitat and host genotype-dependent selection

Parts of this chapter have been adapted from the following with permission:
Norma M. Morella, Francis Cheng-Hsuan Weng, Pierre M. Joubert, C. Jessica E.Metcalf, Steven Lindow, Britt Koskella, Successive passaging of a plant-associated microbiome reveals robust habitat and host genotype-dependent selection. *bioRxiv* 627794; doi: https://doi.org/10.1101/627794

3.1 Introduction

The field of microbiome science spans both basic and applied research in human health, agriculture, and environmental change. As our understanding of the ability of the microbiome (diverse microbial communities and their collective genomes) to influence host health and shape host traits deepens, there is increasing interest in selecting and/or designing microbiomes for specific traits or functions. Such trait-based selection of microbiomes has the potential to shape the future of agriculture and medicine [100][101]. In agriculture, below-ground microbiota have already proven capable of shifting the flowering time of plant hosts [102], enhancing drought resistance [103, 104], and even altering above-ground herbivory [105]. However, long-term, repeatable success of future efforts will rely on a fundamental understanding of the assembly of, selection within, and co-evolution among microbiota within these communities. One of the challenges facing successful, rational microbiome manipulation and assembly is disentangling the forces naturally shaping the communities, including both host characteristics and constant microbial immigration on community stability. For example, in both humans and plants, there is contrasting evidence for the relative importance of the environment versus host genotype in shaping the microbiome [18, 106–113], and dispersal has been shown to override host genetics in an experimental zebra fish system [114].

One powerful but under-utilized approach to understand and experimentally control for the factors shaping microbiome composition and diversity is experimental evolution. Measuring changes of populations or communities over time under controlled settings in response to a known selection pressure has proved a powerful force in gaining fundamental understanding of both host-pathogen (co)evolution [115] and microbial evolution [116]. Here, we harness an experimental evolution approach in order to study how an entire microbial community can be selected upon in a plant host environment that varies across disease resistance-associated genotypes. We employ a microbiome passaging approach using the phyllosphere microbiome of tomato as a model system to determine if the microbial community could become adapted to the plant host environment. The phyllosphere, defined as the aerial surfaces of the plant, is a globally important microbial habitat [36]. Microbial communities in this habitat can shape important plant traits such as protection against foliar disease [28, 29] and growth [47, 117]. Successful trait-based selection on the phyllosphere could therefore allow for enhancement of plant health, but this critically depends on the ability to select for a well-adapted microbial community that is relatively stable against invasion.

We collected a diverse phyllosphere microbiome from tomatoes grown in an agricultural setting and transplanted it onto green-house grown plants using a transplantation method previously shown to be effective for lettuce [118]. We serially passaged this diverse microbiome on each of four cohorts of tomato plants (six lines per cohort) of five different genotypes (pairs of near isogenic genotypes that differed at known disease resistance loci, as well as a wild

tomato accession) for a total of 30 weeks. We then measured adaptation of the community both computationally by fitting community structure to neutral models, and empirically using community coalescence experiments [119] in which communities from different passaged lines are combined together and re-inoculated onto host plants in a common garden experiment. Overall, we were able to measure and characterize the response of the phyllosphere microbiome to selection in the plant host environment under greenhouse conditions, and select for a stable and well-adapted plant-associated microbiome

3.2 Results

3.2.1 Serial passaging experiment

A diverse starting inoculum was collected from field grown, mature tomato plants. This field-microbiome was spray inoculated onto 30 tomato plants of 5 different genotypes, with six replicates each. Two-week old tomato plants were spray-inoculated once per week for five weeks, and then sampled in their entirety ten days after the final inoculation (Figure 3-1b). The phyllosphere microbiome of each plant was then individually passaged on these genetically distinct hosts over the course of four eight-week long passages; P1, P2, P3, and P4 (Figure 3-1a; see methods for details). Microbiomes were not pooled across plants within a given plant genotype, resulting in 30 independent selection lines. Control plants were inoculated with an equal volume of either heat killed inoculum (P1) or sterile buffer (subsequent passages) every week. At the end of each passage, bacterial density was measured and normalized to the weight of each plant (Figure 3-1c), and communities were sequenced using 16S rRNA amplicon sequencing.

We first measured the impact of host genotype on bacterial community structure (Figure 3-1d). Using Bray-Curtis dissimilarity measures, we performed an ANOSIM test (a nonparametric multivariate analysis of variance) and found that plant genotype could explain 29% of dissimilarity between microbiomes in P1 ($p=0.003$). In P2, plant genotype similarly explains 28% of the variation in bacterial community dissimilarity ($p=0.004$). However, genotype becomes an insignificant driver of community composition in both P3 (18%, $p=0.378$) and P4 (9%, $p=0.937$). The genotype effect observed in P1 was robust to removal of the primary outlying line (30% $p=0.002$), and that same line had too low read depth to be analyzed at P2, and thus was excluded from this analysis at the rarefaction step. By P3, this line was included, as it did not fall outside of the 95% confidence intervals for P3 clustering.

We also sought to determine if there were more subtle influences of host genotype on the community that were not uncovered through analyzing Bray-Curtis distances alone. From the original inoculum sample, we identified ten Operational Taxonomic Units (OTUs) using linear discriminant analysis effect-size (LEfSe) analysis [120] that were significantly associated with particular genotypes in P1 and P2. We compared their presence/absence at the end of P4 to those OTUs that were not found to be associated with genotype. Interestingly, those OTUs that were significantly associated with particular genotypes at the start of the experiment were significantly more likely to be present at the end of the experiment than those not associated with genotype (Fisher's exact test, $p=0.013$).

In addition to genotype effects, we were interested in what other factors were driving our observed change in community composition. We found that the number of passages on tomato plants strongly shaped microbial community diversity. Bray-Curtis distances across all samples uncovered a significant effect of both passage number and sample type (i.e. experimental, control, or inoculum) on bacterial communities (Figure 3-1e; ANOSIM test;

effect of Passage $F_{3, 114}$= 27.8895, p= 0.001; Sample Type $F_{3, 114}$= 3.0075, p=0.001). As this was an open system, we next sought to determine if there was a high degree of dispersal amongst plants within the greenhouse by directly comparing the communities of experimental and control plants. At every passage, control and experimental plants are found to host significantly different communities (Bray-Curtis distances; p-values <0.04), suggesting minimal effects of dispersal within the greenhouse relative to our inoculations. When inoculum and control samples are removed from analysis, there remains a significant effect of passage number ($F_{3, 89}$= 33.3023p=0.001) and a significant overall effect of plant genotype on community composition ($F_{4, 89}$= 1.9991, p=0.016). When variance is partitioned, passage can explain 51% of dissimilarity, whereas genotype explains only 4%. Replicate lines from accession 2934 were lost after P3 due to a stem rot fungal pathogen present in the original inoculum that seemingly only infected this genotype. However, the observed overall genotype effect was not driven by this accession, as there remains a significant effect of genotype after its removal ($F_{3, 79}$= 1.9723, p= 0.034), and passage number remains highly significant ($F_{3, 79}$= 31.9804, p= 0.001).

To better understand how the original, diverse, field inoculum changed over four passages on plants in the greenhouse, we calculated the percentage of OTUs in the original inoculum that were detectable over the course of the experiment (Figure 3-1f, green diamonds). At the end of P1, 92% of the field inoculum OTUs were still present on the plants, but by P4, this was reduced to 29%. We then calculated if the decrease in original community member diversity was the result of replacement by non-inoculum taxa (i.e. those that colonized plants over the course of the experiment). In this case, we observed that the proportion of sequencing reads (divided by total reads) representing the original inoculum OTUs remains above 78% (Figure 3-1f, box plots). This suggests that a relatively small percentage of the community was made up of OTUs that colonized plants from the greenhouse environment. Of note, some OTUs considered “non-inoculum” were likely present in the initial inoculum, but in too low of abundance to detect. In particular, there were 27 OTUs with reads in the spray inoculum sample in the non-rarefied dataset, but this was number was reduced to zero after data rarefaction. To account for the impact of the small percentage of arriving species on community composition, we re-analyzed the dataset using only those OTUs that were observed to be present in the initial inoculum (Figure 3-2a). In this case, passage number remains a significant driver of community dissimilarity ($F_{3\ 89}$= 37.6813, p=0.001), as does genotype ($F_{4, 89}$= 2.0393, p=0.015).

We next measured changes in community diversity over the course of passaging and across lines. We found a significant decrease in both OTU richness and alpha diversity over time across all plant genotypes (Figure 3-3a and 3-3b), including when only original spray inoculum OTUs are considered (Figure 3-2b). Importantly, this drop in diversity from the start of the experiment does not correspond to a decrease in overall bacterial abundance on plants (Figure 3-1b). Note that our measures of bacterial growth likely largely overestimate the starting densities (which were based on what was known to be sprayed onto each plant rather than what survived there) and do not account for population turnover (as a result of cell death and replacement within a passage), and are therefore highly conservative. In P1, we also estimated fold change of bacterial abundance on control plants that were sprayed with heat-killed inoculum, and found an average change of 0.76, which is significantly lower than the averaged 11-fold change for experimental plants which received live inoculum (Welch’s Two sample T-Test, p<0.0001). Finally, although passaging was performed in a control

temperature greenhouse, outside high and low temperatures and humidity all varied significantly across passages (Figure 3-4; ANOVA $P<0.001$ for all measures), which may have impacted the observed differences in both abundance and growth across passages.

With the knowledge that communities were drastically changing over time, we sought to determine if the rate at which the communities were changing was consistent. To do this, we calculated Bray-Curtis distances of microbiomes in each passage to P1 microbiomes (Figure 3-3c). As we similarly observed through ordination plots in Figure 1, the communities become less similar to P1 over time. We then fit both a linear and quadratic regression to these data, and we found a better fit of a quadratic model than linear. We determined this through comparing both R^2 values and calculating Akaike information criterion values (Linear R^2 0.774, AIC -3563.231; Quadratic R^2 0.8379, AIC: -4414.637). Both models were highly significant ($p<0.001$). Taken together, this suggests that community change may be slowing down, although it appears to have not entirely stopped.

We next observed changes in relative abundance of specific taxa within lines over time (Figure 3-5, top 100 OTUs plotted). At each passage, there are numerous taxa that are differentially abundant compared to other passages. In some cases, there was evidence for replacement of OTUs within taxonomic groups. Specifically, in the top 10 most differentially abundant taxa as determined by using a Kruskal-Wallis test [121] (Figure 3-6), three of them are in the family *Pseudomonadaceae*. OTUs 0010 and 0004 are in significantly higher relative abundance in P1 than in P4 ($p<0.0001$), and gradually decreased in relative abundance, whereas OTU0002, an unclassified *Pseudomonadaceae*, is significantly more abundant in P4 as compared to other passages ($p<0.0001$). All three OTUs are present in the initial spray inoculum, although OTU0002 represents only 0.03% of rarified spray inoculum reads whereas OTU0004 represents 27% and OTU0010 represents 21%.

To better understand how bacterial community dynamics were changing over the course of the four passages, we utilized a recently developed cohesion metric to quantify connectivity of microbial community [122]. In brief, community cohesion is a computational method used to predict within-microbiome dynamics by quantifying connectivity of microbial communities based on pairwise correlations and relative abundance of taxa. Changes in community cohesion over time are suggestive of biotic interactions, where connectivity can arise from either, or both, positive and negative interactions resulting from cross-feeding (positive) or competition (negative) as well as environmental co-filtering. When applied to our dataset (Figure 3-7), we find a mild but significant increase in positive cohesion values (among 200 permutations) from P1 to P4 ($R^2=0.19$, $p=1.4 \times 10^{-38}$). Consistent with positive cohesion values showing increased biotic interactions, there are also increasingly negative cohesion values from P1 to P4, which again is mild but significant ($R^2=0.257$, $p=1 \times 10^{-53}$). To test if bacterial communities were changing due to neutral processes alone, we first applied the Sloan neutral community model [123] and found that a neutral model is less correlated with observed communities on the plants over time (Figure 3-8a). However our data violate a key assumption of the neutral model in that dispersal was experimentally constrained within lines, and thus we took the approach of generating a null prediction based on the known community composition of inocula applied at each passage and comparing our observed communities to the predicted neutral community using a recently developed approached [124] (see methods for complete details). We found that Bray Curtis distances between predicted (null) and observed communities moderately increases over time ($R^2=0.261$, $p<0.0001$) (Figure 3-8b), suggesting that community change over the course of the passaging experiment

is likely the result of deterministic rather than neutral processes. Further evidence for a shift away from neutrality can be observed using occupancy- abundance curves in which the occupancy, or proportion of individuals in which an OTU is found, is plotted against its relative abundance. A positive correlation between the two is expected to occur by chance, as in a neutrally assembled community, but a change in distribution of individuals may indicate a community shaped by deterministic processes [125, 126]. When our data are visualized in this manner (Figure 3-9), we see that in P1, the most abundant taxa also occupy the highest proportion of plants, as you would expect in a neutral community not undergoing niche selection. However, this trend collapses by P4 with many abundant taxa occupying far fewer individuals than would be expected under an assumption of neutrality.

We next designed an experiment to which we could apply Sloan's model of neutral theory (Figure 3-10a). All lines from the end of P4 were pooled together and re-inoculated onto tomato plants, mimicking the inoculation procedure from the first passage. Plants that received the P4-pooled inoculum had significantly different bacterial community composition than the P4 plants themselves (48% of variation explained, P=0.001; Figure 3-10b). We did not observe an effect of genotype on the communities assembled from this combined inoculum (ANOSIM, p=0.565). We also found that the majority of the variation between samples (ANOSIM; 76%, p=0.001) was driven by an exceptional situation of introduction of a greenhouse taxon (OTU0003) to the plants (Figure 3-10c). To test if neutral processes were driving community structure in this experiment, we again examined fit to a neutral model using the Sloan model approach. In this case, as with P1, the assumption of equal dispersal potential among plants is met. In 200 iterative predictions, the fit of the neutral model is significantly higher in P1 (0.87 ± 0.01) than P4 Combined (0.52 ± 0.05; Student's *t*-test, *p*-value < 0.01), suggesting that neutral processes are dictating the community structure after the first passage, but not in the P4 Combined experiment (Figure 3-10d). We also see the occupancy-abundance relationship breakdown in P4-Combined when compared to P1 directly (Figure 3-10e).

3.2.3. Mycobiome

For P1 and P4, we also used ITS amplicon sequencing to describe the fungal communities across lines, and observe patterns that are similar to the bacterial communities. We again found a significant effect of passage number on fungal communities (Figure 3-11a; Bray-Curtis distances for all samples, ADONIS, 43%, p=0.001). The significant effect of passage number remained after inoculum, control samples, and accession 2934 were removed (Figure 3-11b; 47%, p=0.001). However, unlike in the bacterial community analysis, we found no significant differences in community composition between control and experimental plants at P1 (p=0.117), P4 (p=0.649) or in both passages combined (p=0.588). Additionally, we found no effect of host genotype at either passage (p=0.612, p=0.576) or overall (p=0.997). We also measured a significant decrease in both OTU richness (t-test, p=0.013) and Shannon's diversity (p=0.0005) between P1 and P4 across all genotypes (Figure 3-11c). Finally, analysis of the 5 most common taxa overall identified a single OTU, identified as *Rhodosporidiobolus nylandii*, which was not detectable in the inoculum or P1 but dominated the fungal community in P4 (Figure 3-11d).

3.2.4 Measure of microbiome adaptation using community coalescence approach

The similarity of changes in community structure both across replicates and genotypes over the course of the passaging experiment led us to predict that these microbiomes were adapting to the local plant and greenhouse environment.

To further determine if the community changes we observed from P1 to P4 were due to habitat selection rather than neutral processes, we employed a community coalescence competition experiment. In this experiment (Figure 3-12a), phyllosphere communities from the end of P1 (pooled across all lines) and the end of P4 (again, pooled across lines) were inoculated onto a new cohort of plants, either on their own or in an approximately 50:50 mixture of live cells (as determined using live/dead PMA treatment followed by ddPCR; see methods for complete details).

To ensure that our method for the mixed inoculum was effective, we sequenced multiple replicates of the P1, P4, and Mix inoculums and found that source of inoculum explains 88% of dissimilarity amongst samples (ANOSIM, $p=0.002$). To ensure that the Mix inoculum was significantly different than both P1 and P4 separately, we compared P1 and Mix inocula directly and found that 75% of difference between samples can be explained by this variable ($p=0.02$). Similarly, when P4 and Mix are compared directly, 74% of variation in the community is explained ($p=0.02$). This consistent difference among the three inocula allowed us to compare the communities colonizing plants from each treatment.

We first measured final bacterial abundance and found that colonization was lower on these plants than in previous experiments, but does not significantly differ among treatments ($p=0.419$), apart from control plants, where bacterial colonization was greatly reduced (Figure 3-12b). We then compared bacterial communities again using 16S amplicon sequencing and ordinated samples on a PCoA based on Bray-Curtis distances. Plants that received P1 inoculum had distinctly different communities than those that received either P4 or the Mixed inoculum. Plants that received the Mixed inoculum clustered together with those receiving P4 and were relatively indistinguishable. Using ADONIS tests, we determined that inoculum source can explain 45% of Bray-Curtis dissimilarity amongst samples (Figure 3-12c; $p=0.001$), and there was no effect of plant genotype ($p=0.743$; although note that only three genotypes were used in this experiment). In a pairwise analysis between P1 and Mixed, inoculum source explains 31% of the community dissimilarity ($p=0.001$). In contrast, inoculum source does not explain any significant variation in dissimilarity amongst P4 and Mixed inoculum plants ($p=0.103$). Together, these results suggest that the plants receiving the 50:50 mixed inoculum were indistinguishable in community composition from those receiving the pooled, P4 adapted microbiomes, and that these selected communities were not invadable by the microbial communities from the start of the experiment. Consistent with our results from the passaging experiment itself, alpha diversity was highest in P1 plants compared to both P4 and Mixed plants (Figure 3-12d). Alpha diversity did not differ amongst communities colonizing plants from the P4 and Mixed inoculums, despite being different between the two inocula themselves. We also examined compositional makeup of the communities (Figure 3-12e), and consistent with P1 to P4 passaging results, we see differentially abundant taxa between groups (Figure 3-13). Again, two *Pseudomonas* OTUs are more abundant in P1 plants as compared to P4 and Mix, in which there was an unclassified *Pseudomonaceae* that was higher in relative abundance.

3.3 Discussion

The impact of a microbiome on host health and fitness depends not only on which microbial organisms are present in the community, but also on how they interact with one another within the microbiome [127]. Unlocking the great potential of microbiome manipulation and pre/probiotic treatment in reshaping host health will therefore depend on our ability to understand and predict these interactions. We took a microbiome passaging approach, inspired by classic experimental evolution, to test how selection for growth in the tomato phyllosphere under greenhouse conditions would impact microbiome diversity and adaptation across genotypes that differ in disease resistance genes.

Across independently selected lines passaged on five tomato genotypes, we observed a dramatic shift in community structure and composition, accompanied by a loss of alpha diversity (Figures 3-1 to 3-3). We also found that host genotype shapes community composition early in passaging (P1 and P2), explaining over 24% of variation amongst samples, but diminishes over time. The relative importance of host genotype and environment in shaping microbiome composition remains highly debated. Our results suggest that the relative importance of genotype versus other factors, such as the growth environment or strength of within-microbiome interactions, changes over the course of passaging on a constant host background. We did observe that even in the absence of a strong genotype effect, there remains a legacy of genotype effect, in that OTUs found to be significantly associated with particular genotypes early on are more likely to be present at the end of passaging than those that did not exhibit any host preference.

In order to test if the phyllosphere microbiome undergoes habitat filtering, we chose to begin the experiment with a diverse inoculum. This starting community generated from field grown tomato plants likely contained microbes from other surrounding plant species, dust, soil, and other sources. In particular, neighboring plants have been shown to contribute to both the density and composition of local airborne microbes [40]. We found that although the total number of these field inoculum OTUs decreased over the course of the experiment, the taxa that remained consistently made up 78-95% of the community. This provides evidence that the original spray inoculum underwent strong niche selection over the course of the experiment. To test the alternative hypothesis that community changes were due to neutral processes such as bottle necking or random dispersal, we first fit our data to neutral and null models, finding a poorer fit over time. We then tested this experimentally by conducting a community coalescence experiment to measure fitness of passaged microbiomes as compared to those from the start of the experiment. The results of this experiment strongly support the idea that these phyllosphere microbiomes adapted to the plant host environment over the course of four passages (Figure 3-12). Independent of overall bacterial abundance, P4 microbiomes were able to dramatically outcompete the less-adapted P1 microbiomes. This community coalescence approach [119] allowed us to demonstrate non-neutral adaption of a bacterial community that is independent of host genotype and resistant to invasion by a more diverse, less-adapted community. This community coalescence approach was used by others in a study conducted on methanogenic bacterial communities [128]. The authors found that when multiple methanogenic communities were combined, a single dominant community emerged from the mix. This emergent dominant community resembles the single community with the highest methane production that went into the combination, suggesting that the most-fit community is capable of reassembly, even in the presence of other bacteria.

While adaptation to both the local host environment (tomato plants, host genotype) and the larger environment (the greenhouse) were likely driving the increasingly non-neutral selection over time, the strength of within microbiome biotic interactions likely also increased over the course of the experiment. We see evidence for this through both increasing positive and negative community cohesion values, and our finding that a greenhouse-acquired taxon contributed to shaping community structure in one experiment (Figures 3-7 and 3-10). Though we are not able to determine what drove certain plants to be more colonized by this taxon than others, we did observe strong shifts in community composition associated with its relative abundance that may be due to spatial organization of plants in the greenhouse and/or stochastic initial colonization events. In a greenhouse study conducted on *Arabidopsis thaliana* phyllosphere communities, the authors found that abundance of certain dominant taxa could be tied to spatial organization of the plants that was likely driven by early stochastic events [111].

Although we focus primarily on the bacterial portion of the microbiome, the mycobiome was observed to change over the course of passaging as well (Figure 3-11). Previous work in *A. thaliana* demonstrated that "hub" fungal taxa strongly influence both bacterial alpha and beta diversity [129]. Although it is possible that multi-kingdom interactions played a role in shaping community composition, our experimental methods, especially the process of sonicating epiphytic microbiota and freezing in between passages, likely biased passaging towards bacterial taxa and epiphytes. Similarly, pelleting of the community and removal of the supernatant at each passage would have selected against any free lytic bacteriophages. Previously, we found that the phage fraction of the microbiome is capable of altering both abundance and composition in the tomato phyllosphere [130]. Moreover, by reducing selection for dispersal ability across the phyllosphere environment by evenly spraying microbes onto leaves in a high humidity environment, we may have tipped the balance in favor of bacterial species that are better competitors within the microbiome. A dispersal-competition tradeoff was recently demonstrated using functional traits of soil microbial communities along a marine-to-land gradient, where bacterial communities from more disturbed habitats were found to be dominated by cell chemosensory and motility behaviors whereas those from more stable environments were dominated by traits for competition and chemical defense [131]. Future work is required to disentangle both the selective impacts of the plant versus environment versus multi-kingdom interactions in shaping microbiome adaptation, and the change in microbial function as a result of this response to selection.

Overall, we were able to show rapid and robust habitat selection of these communities over relatively short time scales. The results uncover great promise of this approach and system for answering fundamental questions about the forces shaping microbiome assembly over time, and also paves the way for selecting stable, uninvadable host-associated microbiomes, which may inform rational microbiome manipulation and probiotic design. Experiments such as these are crucial if we are to understand general principles governing microbiome assembly and adaptation and use this knowledge for transformative applications in both medicine and agriculture.

3.4 Materials and Methods

3.4.1 Tomato accessions

Tomato accessions were obtained from the Tomato Genetics Resource Center. Five tomato genotypes were used: *Solanum lycopersicum* money maker disease susceptible (TGRC 2706); *S. lycopersicum* money maker disease resistant (TGRC 3472); *S. lycopersicum* Rio Grande

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

3.4.2 Tomato germination and growth

Seeds were surface sterilized using TGRC recommendations as follows: seeds were soaked in 2.7% bleach (sodium hypochlorite) solution for 20 minutes. Sterilized seeds were then washed with sterile ddH_2O three times to remove any excess bleach. Sterilized seeds were then transferred onto 1% water agar plates and placed in the dark at 21°C until emergence of the hypocotyl. At that point, seedling plates were moved into a growth chamber and allowed to continue germination for 1 week. Growth chamber conditions were 25°C, 65% humidity and 16 h daylight per day. After approximately one week, seedlings were transferred planted in sunshine mix #1 soil in seedling trays. After approximately one more week of growth, seedlings were transplanted into 8" diameter pots, making the plants approximately 2.5-3 weeks old at the first time of microbial inoculation. Age of inoculation varied slightly from experiment to experiment but was kept identical amongst genotypes within an experiment.

3.4.3 Inoculation preparation, first passage

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

pelleted for 10 mins at 4000 X G. Cells were re-suspended in 200mL 10mM $MgCl_2$ buffer. Of this, 40mL were and heat killed in an autoclave for a 30 minutes at 121°C. Inoculum was plated, and an absence of growth confirmed that the heat-kill was effective. To get initial concentration of inoculum, dilution plating was performed on Kings Broth agar plates (1.1 X 10 ^6 CFU/mL). Soil from each site, which had been stored at -20°C, was combined in a sterile Nalgene bucket and thoroughly mixed before inoculation.

3.4.4 Inoculation procedure

Soil inoculation: The top layer of every pot was supplemented with 40 grams of UC Davis Farm Soil. Soil inoculation was only performed once and only for the first passage of plants. Spray inoculation: Each plant was sprayed, using misting spray tops placed in 15mL conicals, with approximately 4.5mL of inocula. Control plants from passage 1 were inoculated with the heat-killed inocula. Control plants from subsequent experiments were inoculated with sterile 10mM $MgCl_2$. Immediately after inoculation, plants were placed in a random order in a high-humidity misting chamber for 24 hours. After 24 hours, the plants were moved to a greenhouse bench. Plants were inoculated once per week in the same manner and were placed in the misting chamber for 24 hours after every inoculation. Passage one plants received 5 weeks of inoculation, P2-P4: four weeks, and the fifth cohort: five weeks.

3.4.5 Plant sampling and inoculation preparation for P2-4

Ten days after the final spray inoculation, plants were sampled. With the exception for plant cohort 5, all plants were cut off at the base and immediately placed into sterile 1L bottles individually. By the end of cohort 5, the plants had grown too large to sample the entire plant, and instead, roughly 2/3 of the plant material was sampled from each plant, with care taken to sample the same age of branches from every plant. After collection, plant material was weighed, and 200mL of sterile 10mM $MgCl_2$ were added to each bottle containing the plant material. The bottles were submerged in a sonicating water bath, sonicated for 5 minutes, vortexed, and sonicated for another 5 minutes. Half of the volume from each plant was pelleted for 10 mins at 4200 X G, re-suspended in ~1mL of 1:1 KB Broth Glycerol, divided into aliquots, and stored at -80°C for inoculation of the subsequent passage. The other half of the volume was pelleted in the same manner and then stored as a pellet at -20°C for DNA extractions. To prepare inoculation of the next passage, microbiome glycerol stocks were thawed, briefly pelleted to remove glycerol, and re-suspended in sterile 10mM $MgCl_2$. Volume of re-suspension depended slightly on the size of the plants, but in general ranged from 5-10mL. Microbiomes were never pooled.

3.4.6 Inoculation preparation, combination of P4 microbiomes

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

3.4.7 P1, P4 coalescence experiment

Genotypes 2706, 3472, and 2934 were used for this experiment, and four plants of each genotype received each treatment (P1, P4, and Mix). One control plant of each genotype was

spray inoculated with $MgCl_2$ as a control. To prepare the inoculum, microbiomes from the end of passage one and the end of passage four were used. All aliquots (one from each plant, except for plant 4 which had exhibited disease symptoms) were thawed and combined. The same was done for all of the individual microbiomes that came off of passage 4 plants. To remove the glycerol, the samples were spun down and re-resuspended in 10mM $MgCl_2$. In order to generate the 50/50 mix of P1 and P4 microbiomes, live/dead PCR with PMA treatment was used, adapted from the following method [132]. Briefly, serial dilutions of P1 and P4 were performed in $MgCl_2$. Each sample then received PMA at a final concentration of 100uM and vortexed. Samples were incubated in the dark at room temp for 5 minutes. Then they were placed in ice on a tray exactly 10cm away from a 700 watt halogen lamp. The light was turned on for 30 seconds, and turned off for 30 seconds. During the 30 seconds without light, the samples were all vortexed. This was repeated three more times. Samples were then pelleted for 10 minutes at 5000 X G. The supernatant including the excess PMA was removed, and cells were re-suspended in sterile 10mM $MgCl_2$. Droplet Digital PCR (as described below) was then utilized to quantify bacteria from each sample, and concentration was matched to 7.7×10^6 cells/mL. P1 and P4 were aliquotted separately and then re-combined for the mixed inoculum so that each plant received ~9×10^4 bacteria each week that they were inoculated. Plants were inoculated for three weeks and harvested 10 days after the final inoculation as described previously.

3.4.8 Bacterial quantification using ddPCR

The BioRad QX200 system was used for culture independent quantification of bacteria. Complete ddPCR methods are described elsewhere [130]. Bacterial abundance was measured directly after microbes were sonicated off plant surfaces into sterile buffer. For consistency, the same region of the 16S gene used below for amplicon sequencing was used for bacterial quantification. PNAs were used as well to limit any background amplification of plant mitochondrial or chloroplast DNA. Five ul of sample were used in every reaction. All data were normalized to weight, in grams, and concentrations are reported as 16S copy number/gram.

3.4.9 DNA extractions

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

3.4.10 16S Libraries

The 16S rRNA gene was amplified using dual-indexed primers designed for the V3- V4 region [87] using the following primers: 341F (5 -CCTACGGGNBGCASCAG-3) and 785R (5 -GACTACNVGGGTATCTAATCC-3) [88]. Additionally, we also used peptide nucleic acids, PNAs [89] to decrease amplification of plant mitochondrial and chloroplast DNA. Negative buffer controls and PCR controls were sequenced along with experimental samples. Reaction conditions were 94°C for 3 min, 94°C for 45 s, 78°C for 10 s, 50°C for 1 min, 72°C for 1.5 min, repeat steps 2–5 30 times and 72°C for 10 min. PCR mixtures were randomized in order, run in duplicate for each sample, pooled and quantified using Qubit. Amplicons from each sample were pooled in equimolar concentrations, cleaned using an AMpure bead clean-up kit. Libraries were

prepared for paired 300-nucleotide reads in Illumina's MiSeq V3 platform (Illumina) at The California Institute for Quantitative Biosciences (QB3) at UC Berkeley.

3.4.11 ITS Libraries

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

3.4.12 Data Analysis

MiSeq sequencing files were demultiplexed by QB3 sequencing facility. Reads were combined into contigs using VSearch [92], and the remainder of the analysis was carried out in Mothur [93] following their MiSeq SOP [94]. Data were quality-filtered, and chimeras were removed using UChime [95]. Singletons were removed using the split.abund command in Mothur after pre-clustering of similar sequences. We used a 97% similarity cut-off for defining OTUs. The Silva reference database [96] was used for sequence alignment and taxonomic assignment. Archaeal, chloroplast, mitochondrial and unknown domain DNA sequences were removed. To account for reagent contaminants, we also sequenced two DNA extraction kit controls and PCR controls along with our samples. Contaminant OTUs from control samples that were at a similar or higher relative abundance in control samples compared to experimental samples were removed from the full OTU table. Bacterial were rarified to 8,000 reads per sample. For the fungal community, an OTU table was generated from the fungal community sequencing data using QIIME 2 (https://qiime2.org, version 2018.8). Trimmed, paired reads were first denoised, without read trimming, using the DADA2 plug-in [133]. Chimeric sequences were then filtered using the uchime-denovo command of the Vsearch plug-in [134]. Reads were then clustered into OTUs at 97% identity using the cluster-features-closed-reference command in the VSEARCH plug-in and the 01/12/2017 version of the UNITE database [135]. In order to assign taxonomy to the clustered OTUs, a Naïve-Bayes classifier was first trained using the UNITE database and the feature-classifier plug-in [136]. The classify-sklearn command of the feature-classifier plug-in was finally used to assign taxonomy to the clustered OTUs. Once bacterial and fungal OTU tables were generated in Mothur and QIIME, the remainder of the analysis was performed in R using the following packages: Phyloseq [97], vegan [98], ampvis2 [137], and MicrobiomeSeq (Alfred Ssekagiri, William T. Sloan, Umer Zeeshan Ijaz).

3.4.13 Community Cohesion Metrics

The estimations of positive and negative cohesion values follows the cohesion metrics approach proposed by Herren *et al*. [122]. Herren *et al*. multiplied the connectedness metrics determined by relative abundance profile by the same relative abundance profile to estimate cohesion values. We modified their method to estimate cohesion values by using two relative abundance profiles of a training set and test set. Relative abundance profile of the training set was obtained by randomly selecting half of the samples in each microbiome passage. The test set consists of the other half of the samples. Using the training set and following the same procedure as Herren *et al*., connectedness metrics were calculated. The estimated connectedness metrics subtracts a null model. The objective of the null model was to calculate the strength of pairwise correlations that

would be observed if there were no true relationship between OTUs. The obtained connectedness metrics are multiplied by relative abundance profile of test set to estimate positive and negative cohesion values. Two hundred iterations of sampling randomization in each microbiome passage were carried out at OTU level to obtain training set and test set for P1, P2, P3, and P4.

3.4.14 Neutral model

The neutral model was proposed by Sloan *et al*. to describe both microbial diversity and taxa-abundance distribution of a community [123]. Burns *et al*. [114] have developed a R package based on Sloan's neutral model to determine the potential importance of neutral process to a community assembly. In brief, the neutral model creates a potential neutral community by a single free parameter describing the migration rate, *m*, based on two sets of abundance profiles – a local community and metacommunities. The local community describes the observed relative abundance of OTUs, while the metacommunity is estimated by the mean relative abundance across all local communities. The estimated migration rate is the probability of OTU dispersal from the metacommunity to replace a randomly lost individual in the local community. The migration rate can be interpreted as dispersal limitation. In each microbiome passage, half of the samples were randomly selected and the relative abundance profile at the OTU level was used. The neutral model fit and migration rate were estimated in the resolution results of 200 iterations for P1, P2, P3, P4, and P4 Combined.

3.4.15 Null model predictions

We applied a null model approach on the serial passaging data P1-P4 to characterize the changes of stochastic process driving the assembly of plant microbiome over time. Lines that had high quality sequencing data at every time point (thirteen in total), were used for this analysis. The null scenario for each line at each passage was generated using the data for that same line at the previous passage. The null scenario of P1 was generated using the original field inoculum sample. The null model approach was based on community pairwise dissimilarity proposed by Chase and Myers (Chase and Myers 2011) and extended by Stegen *et al.* to incorporate species abundance (Stegen, Lin et al. 2013). Chase and Myers proposed a degree of species turnover by a randomization procedure where species probabilistically occur at each local community until observed local richness is reached. However, the estimated degree of turnover does not include species abundance. To take full advantage of our dataset, we also incorporated species relative abundance into the procedure proposed by Stegen *et al*. Zinger *et al*. has developed R code for the null model and applied the null model approach on the soil microbiome [124]. This approach does not require *a priori* knowledge of the local community condition and determines if each plant microbiome at the current passage deviates from a null scenario generated by that same microbiome at the previous passage. In brief, the null scenario of each was generated by random resampling of OTUs and remained the same richness and number of reads with the original sample. Total OTUs observed in the sample and the corresponding relative abundance were used as probabilities of selecting an OTU and its associated number of reads, respectively. The Bray-Curtis distance is used to calculate dissimilarities across null communities with 1,000 permutations. The average of dissimilarities among permutations represents null expectations of community dissimilarities. The null deviation shows the differences between average null expectation and the observed microbiome of the same line.

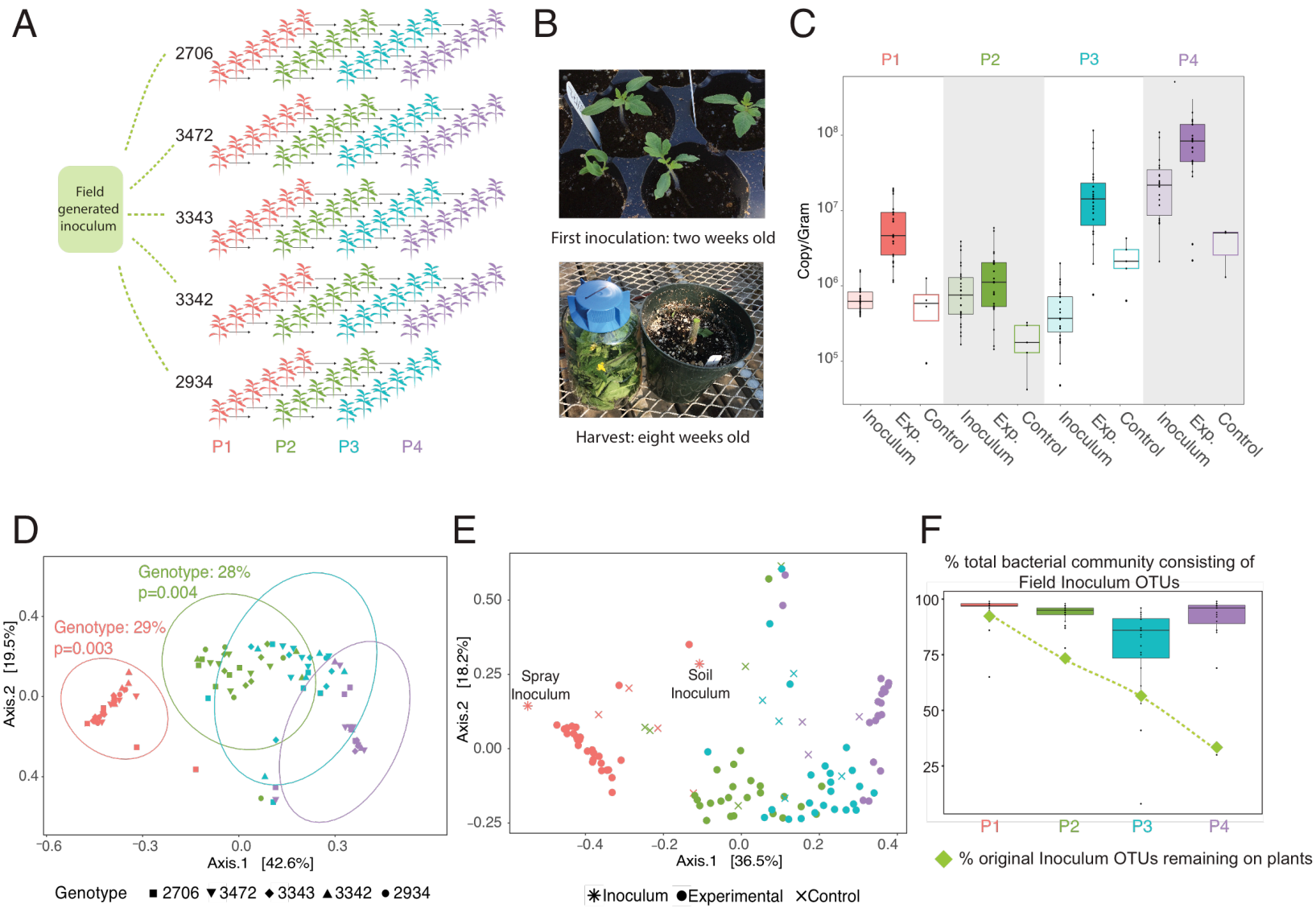


Figure 3-1 Serial passaging of the phyllosphere microbiome

Figure 3-1 (continued) Serial passaging of the phyllosphere microbiome Experimental design of serial passaging experiment in which microbial inoculum from an agricultural tomato field was inoculated onto replicates of five genotypes and passaged for four passages (a). Plants were first inoculated when they were 2 weeks old, and the entire plant was sampled at 8 weeks old (b). Bacterial abundance was measured at the end of each passage from experimental and control plants using ddPCR and normalized to the weight of each plant. Inoculum density was calculated as well (c). PCoA plots of Bray-Curtis distances show a significant effect of genotype in P1 and P2 (d) and passage (colors) and sample type (shapes) (e). Ellipses indicate 95% confidence around the clustering.The percent of original inoculum OTUs present at each passage was calculated (green diamonds), and the reads/sample of inoculum OTUs out of total reads was calculated for each plant at every passage and displayed on a box plot (f).

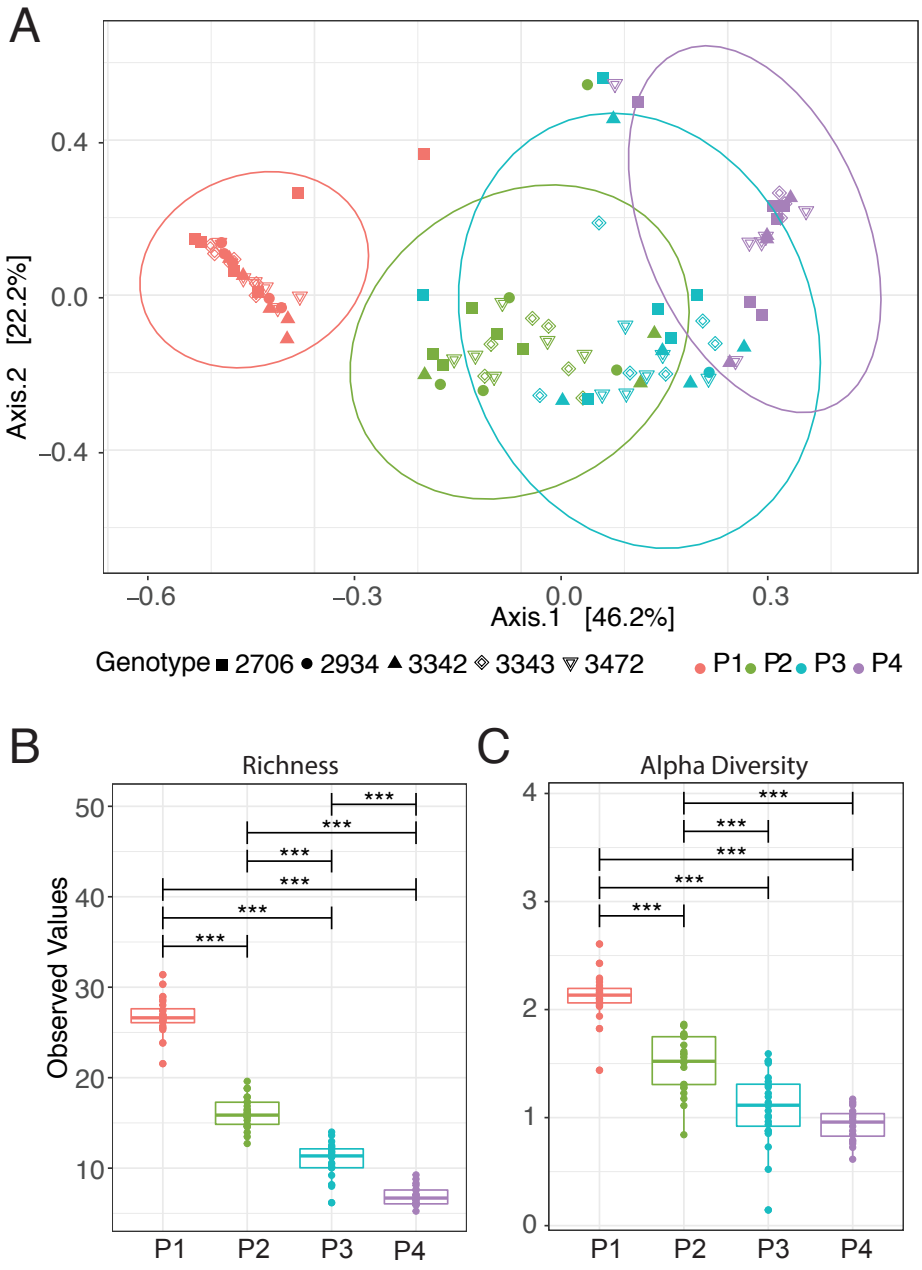


Figure 3-2 Serial passaging of the phyllosphere microbiome: Inoculum only taxa

The dataset was subsampled to only contain OTUs that were present in the initial spray inoculum. Experimental plants are ordinated on a PCoA plot based on Bray-Curtis distances (a), and an ADONIS test indicates a significant effect of passage (colors) and genotype (shapes). Richness (b) and Shannon's alpha diversity (c) are plotted of each experimental plant at each passage and show a significant decrease over time. Significance values of pairwise comparisons are illustrated on the graph * p≤0.05; ** p≤0.01; *** p≤0.001; ****p≤0.0001.

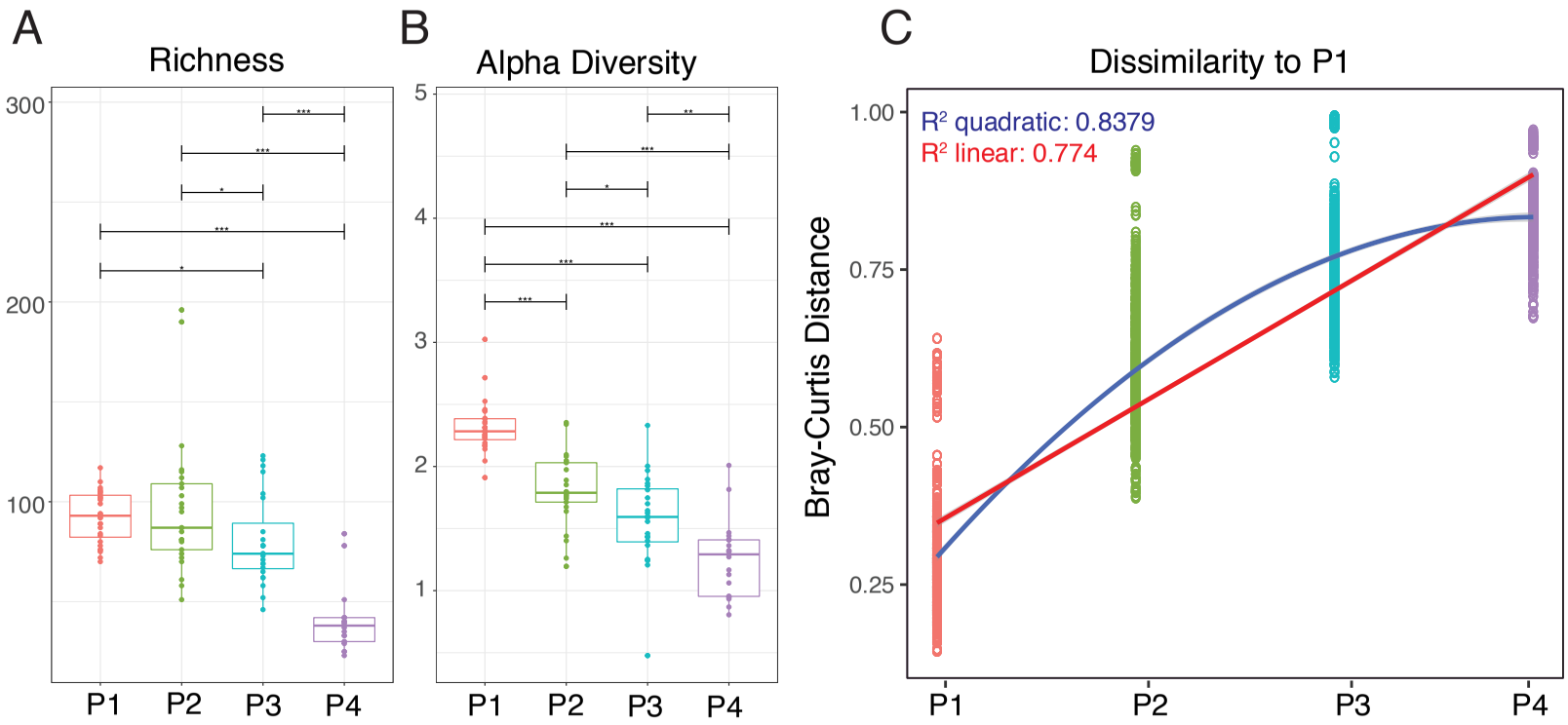
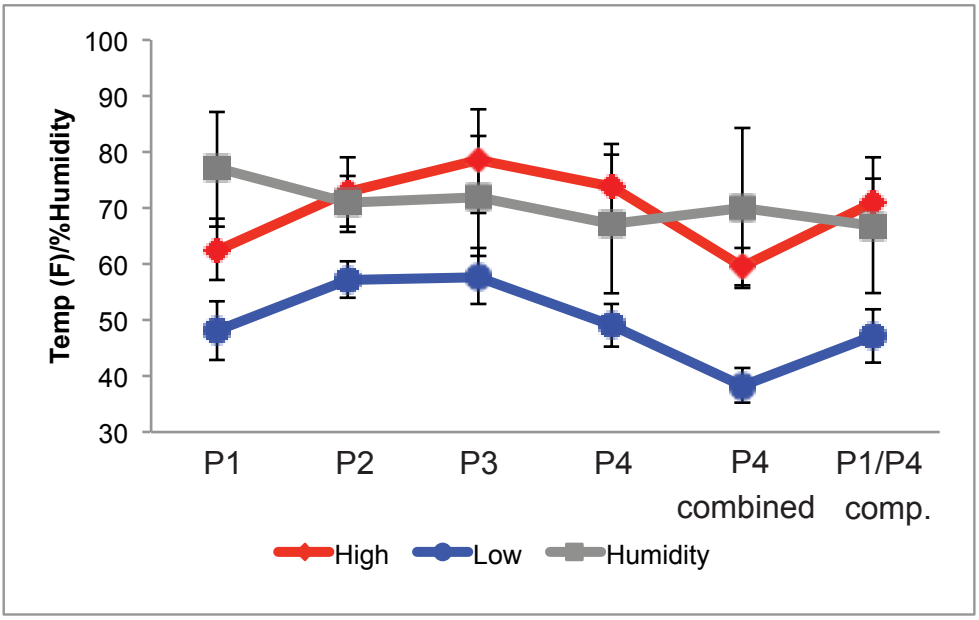


Figure 3-3 Changes in diversity from P1 to P4
Plots of richness (a) and Shannon's alpha diversity (b) at each passage show a significant decrease over time. Bray-Curtis distances between microbiomes in P1 were compared to those in P1, P2, P3, and P4, and linear and quadratic models were fit to the data (c). Significance values of pairwise comparisons in (a) and (b) are illustrated on the graph * p≤0.05; ** p≤0.01; *** p≤0.001; ****p≤0.0001.

A



B

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

Figure 3-4 Climatic Variables
Although all experiments were performed in the greenhouse, outside climatic variables were plotted for the duration of the six experiments. Humidity, high temp, and low temp are plotted (a). The date at which the plants were first inoculated and the date at which they were harvested are shown (b).

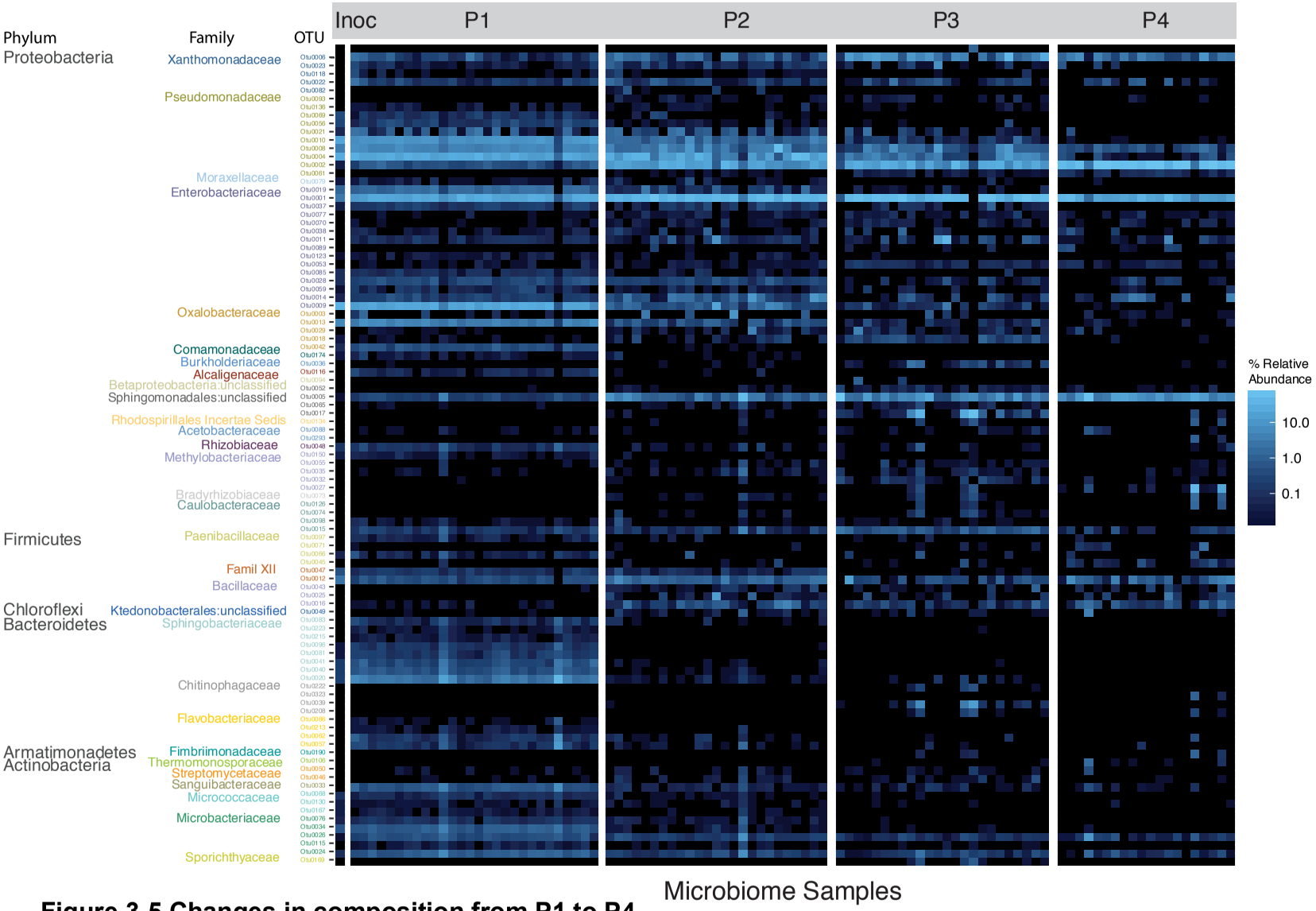
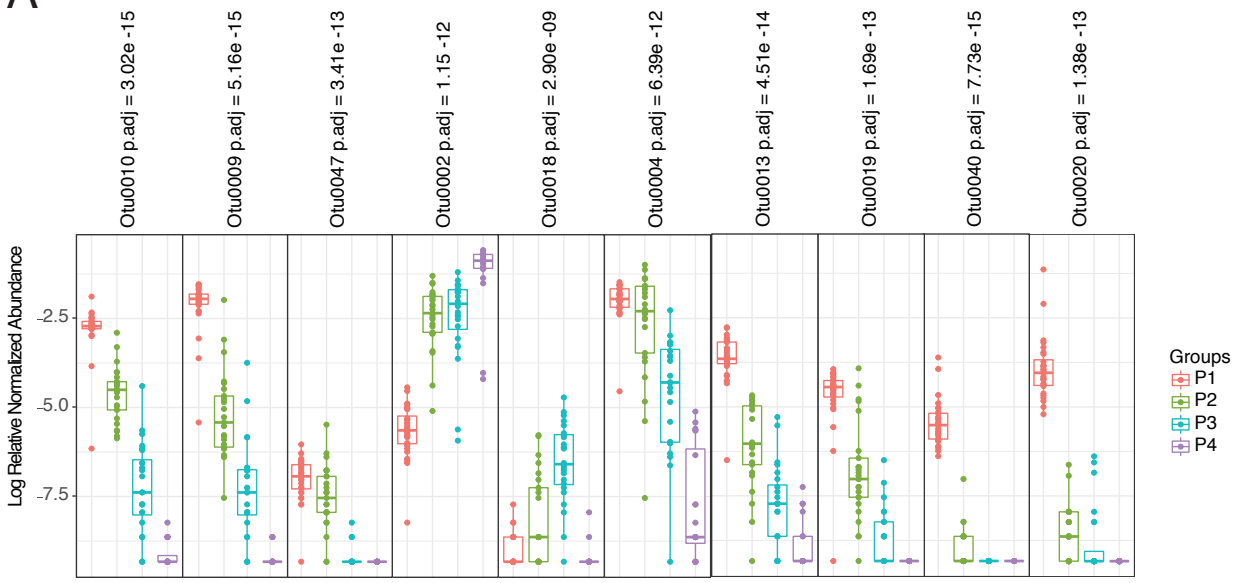


Figure 3-5 Changes in composition from P1 to P4

A heat map showing relative abundance of the top 100 OTUs illustrates the changing community composition at multiple taxonomic levels (d). Full taxonomy of OTUs is found in Appendix B.

A



B

	Phylum	Class	Order	Family	Genus
Otu0010	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Otu0009	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Enterobacteriaceae_unclassified
Otu0047	Firmicutes	Bacilli	Bacillales	Family_XII	Exiguobacterium
Otu0002	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonadaceae_unclassified
Otu0018	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Massilia
Otu0004	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Otu0013	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Massilia
Otu0019	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Rahnella
Otu0040	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Pedobacter
Otu0020	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Pedobacter

Figure 3-6 Differentially abundant taxa amongst passaged lines
We performed a Kruskal-Wallis test on log-relative transformed OTU abundance at different passages using the MicrobiomeSeq package (a). This is a non-parametric method, and it tests whether samples originate from the same distribution. P-values are corrected for multiple testing using family wise error rates. Significant OTU rankings 1-10 are assigned importance using random forest classifier. Identities of OTUs are displayed as well (b).

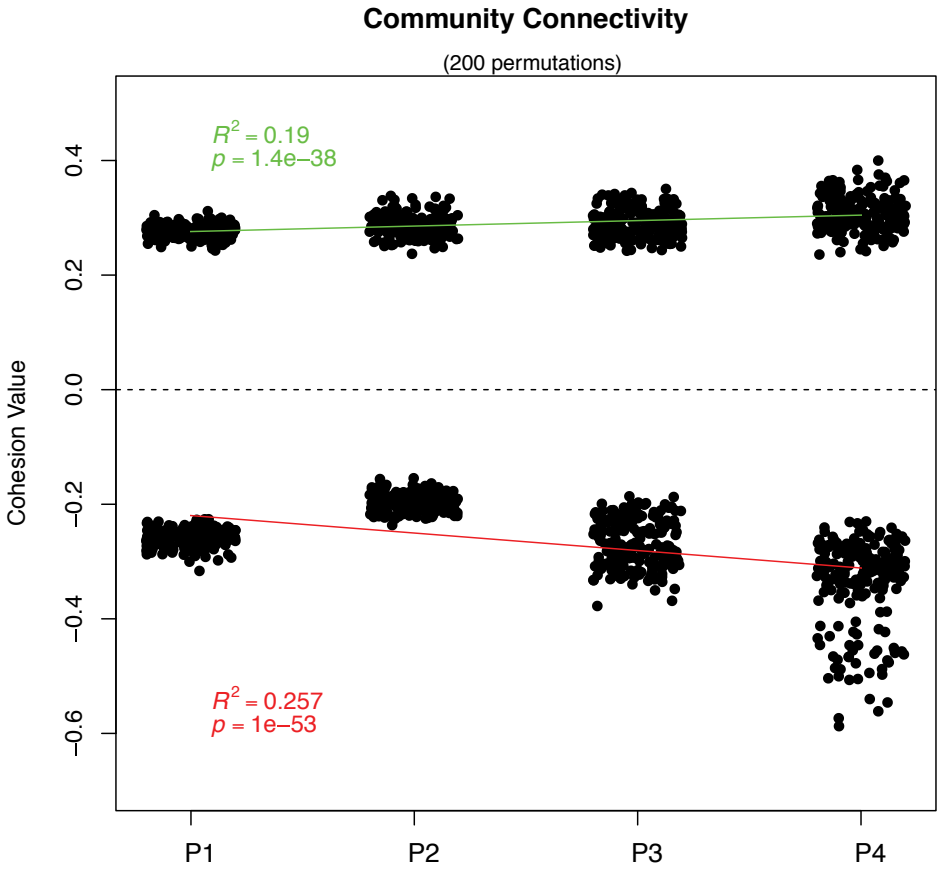


Figure 3-7 Community Cohesion from P1 to P4
We applied cohesion metrics on tomato microbiome data set to describe microbial dynamics in P1, P2, P3, and P4. We calculated both positive and negative cohesion values and found a moderate but significant increase in positive and negative cohesion values from P1 to P4.

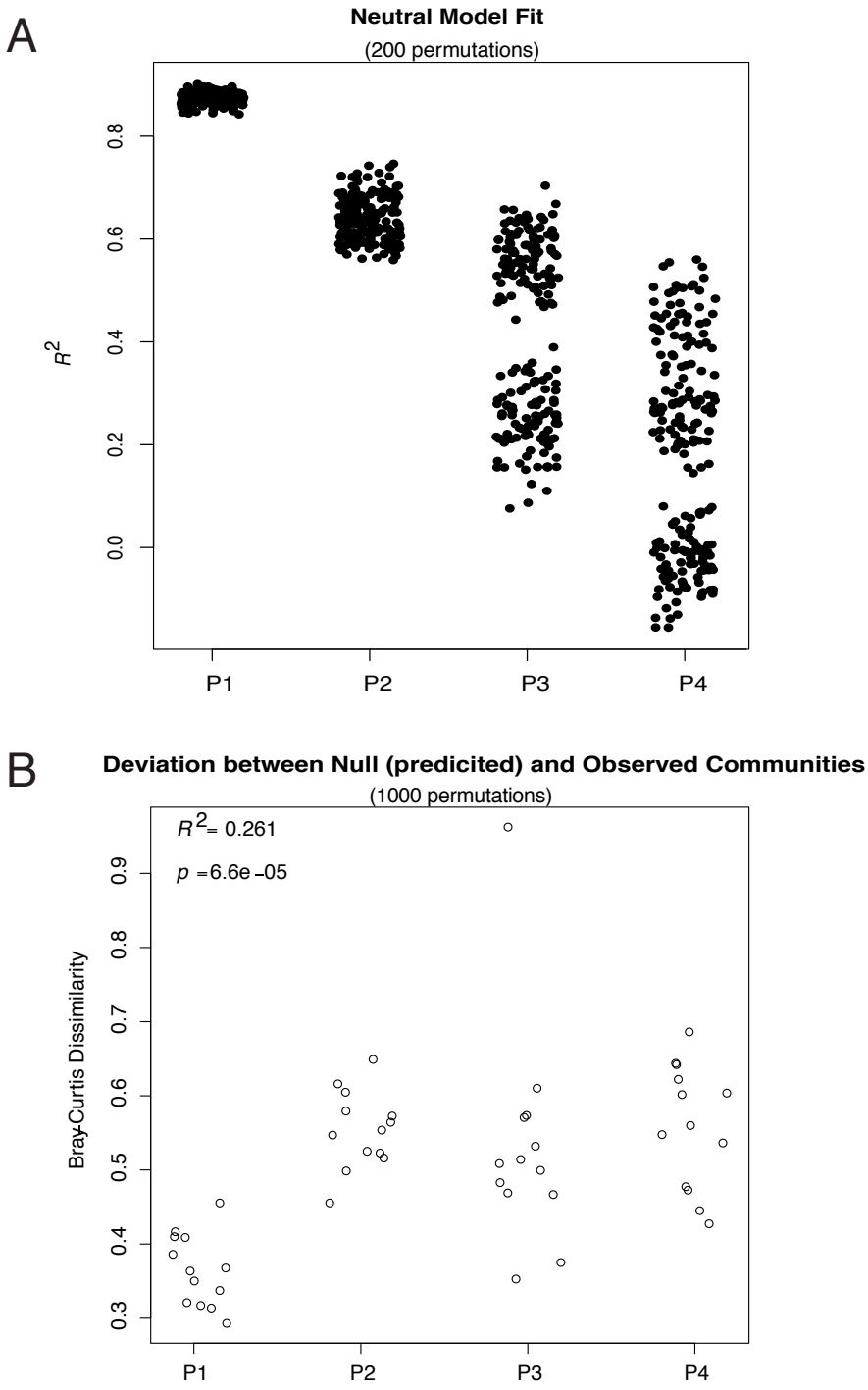


Figure 3-8 Neutral model fit and comparison of null to observed communities We compared the observations of community structure and predicted community composition using a neutral model and found a poorer fit over time (a). We compared Bray-Curtis distances between a predicted null model from the n-1 passage with observed communities from passage n and found a moderate but significant increase in deviation from the null prediction from P1 to P4 (b).

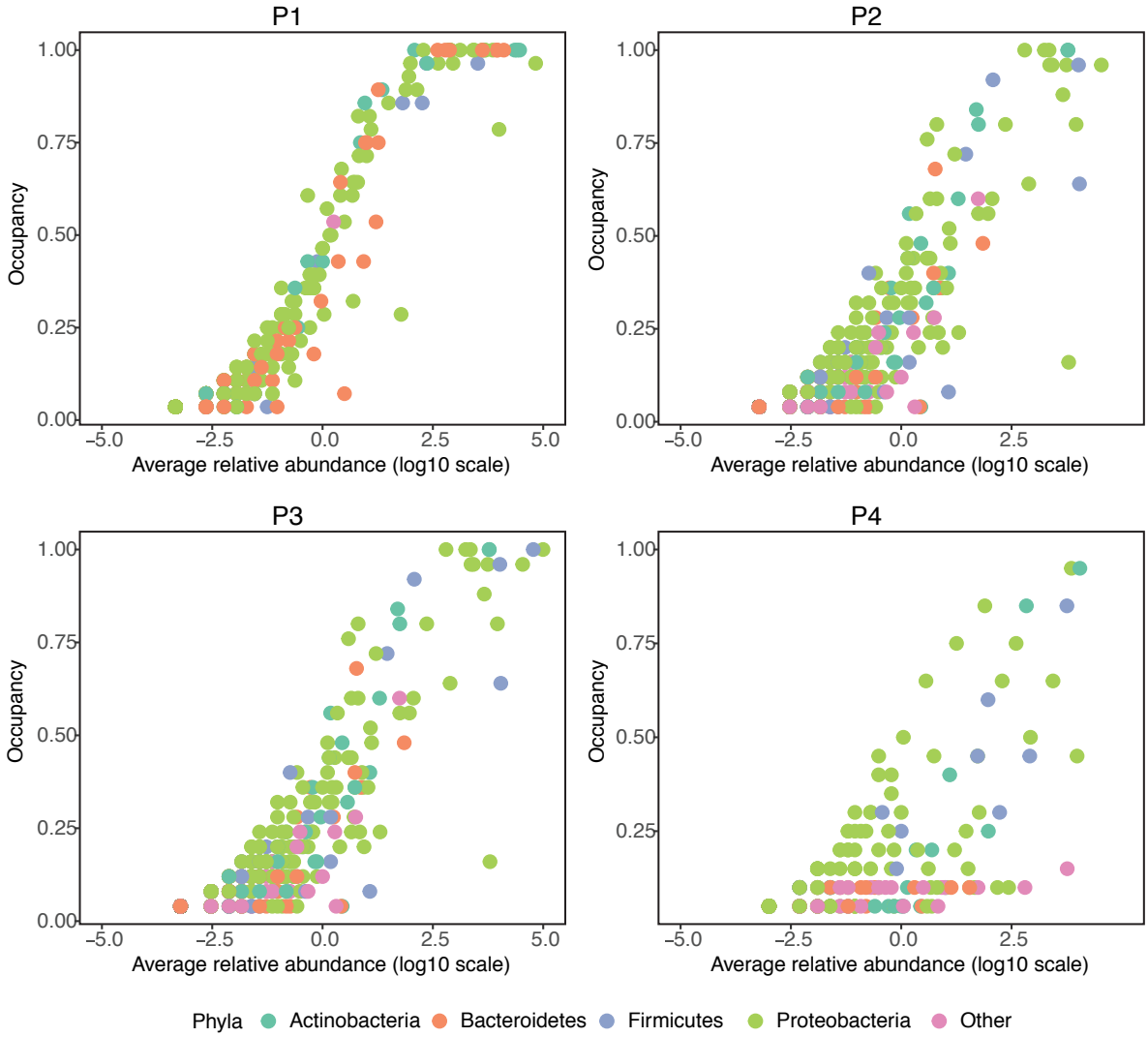


Figure 3-9 Occupancy- Abundance of taxa from P1 to P4
For each OTU, its occupancy (or, proportion of plant hosts in which it was found) is plotted against the log 10 of its relative abundance. OTUs belonging phyla other than those in the top 4 Phyla are classified as “other”

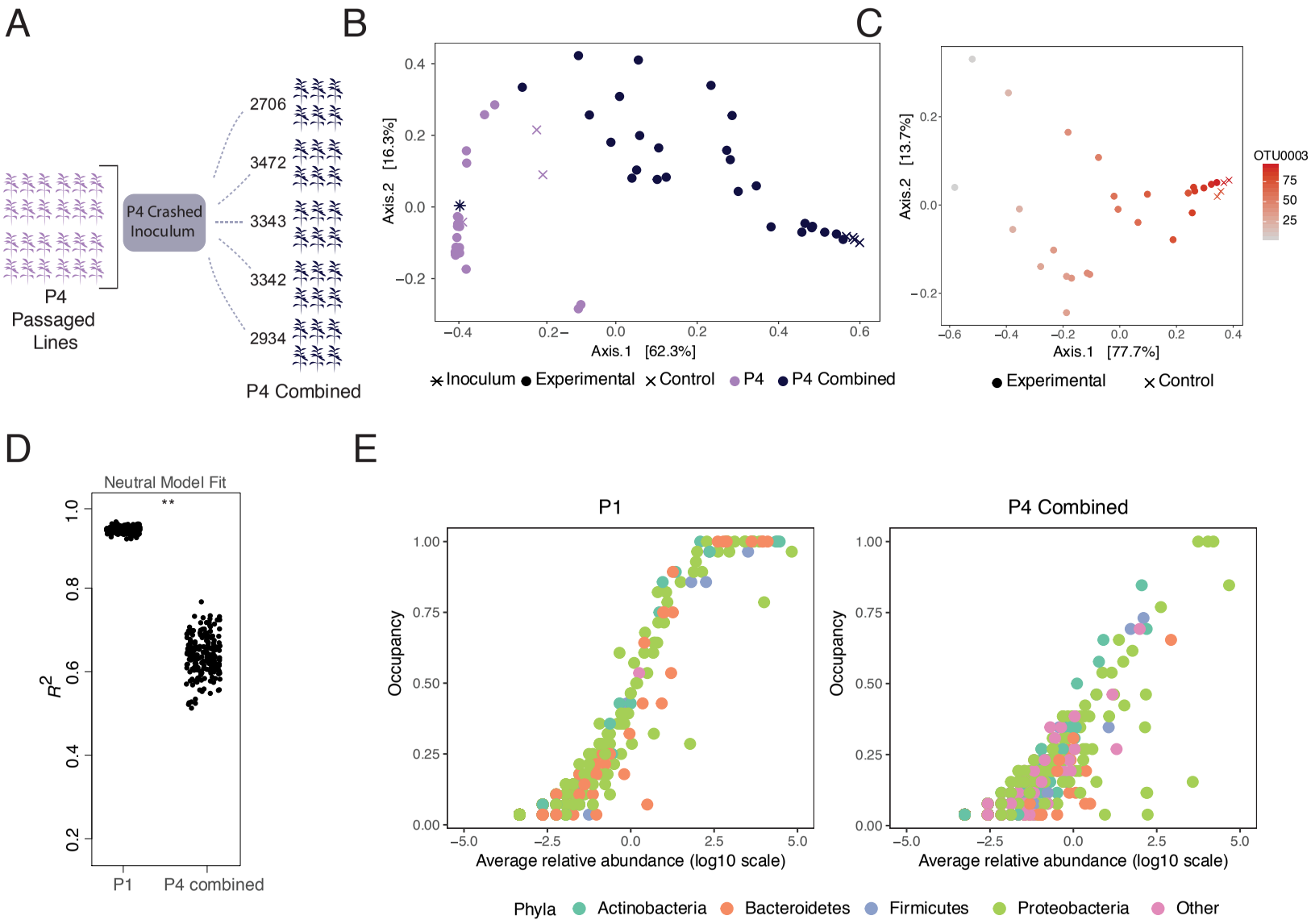


Figure 3-10 Combination and reincoulation of serial passaged lines

Figure 3-10 (continued) Combination of serial passaged lines
Passaged microbiomes from the end of P4 were pooled and then sprayed onto a fifth cohort of tomato plants, and there six replicates of five genotypes (a). P4 plants, inoculum generated from P4 plants, and controls are plotted on a PCoA plot based on Bray-Curtis distances (b), and P4-Crashed plants can be seen clustering away from P4 plants. We compared the observations of community structure and predicted community composition by a neutral model for P1 and P4-Crashed (c), and in 200 iteratively predictions, the fit of the neutral model is significantly higher in P1 than P4-Crashed (Student's t-test, p-value < 0.01). Visualized on a PCoA plot (d), relative abundance of OTU0003 can be seen driving Bray-Curtis distances amongst plants, both experimental and control. Occupancy-Abundance curves for P1 and P4-combined are shown side-by-side in order to visualize the shift away from neutrality (e).

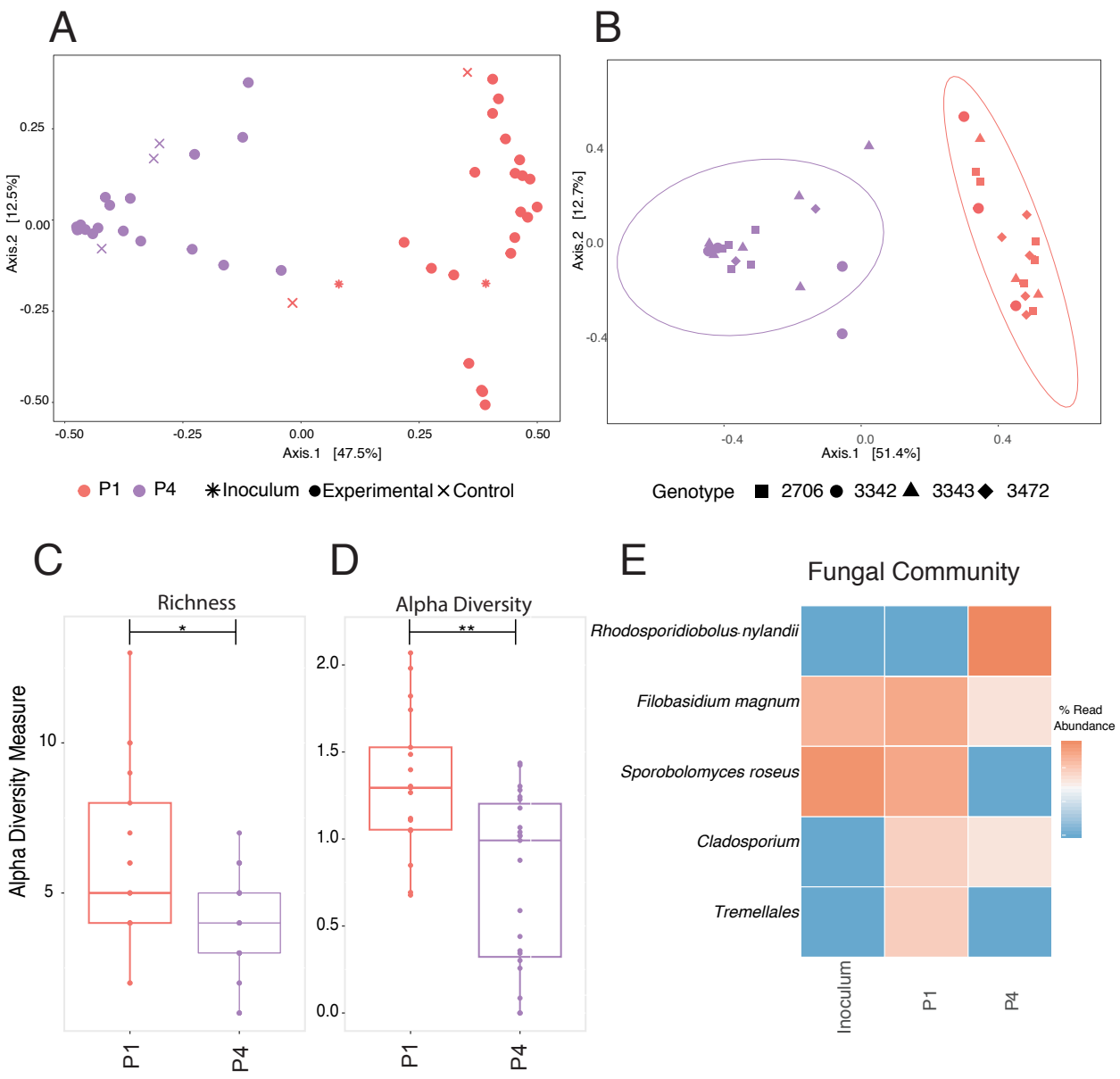


Figure 3-11 The fungal community

A PCoA plot of Bray-Curtis distances show a significant change in the community from P1 to P4, as determined by a PERMANOVA test (a). There is no effect of genotype (shapes) on the fungal community (b). Ellipses indicate 95% confidence around the clustering. Both richness (c) and Shannon's alpha diversity (d) significantly decrease between P1 and P4. Relative abundance of the top five fungal taxa is plotted for the original inoculum, P1 and P4 (e). Significance values of pairwise comparisons in (c) and (d) are illustrated on the graph
* p≤0.05; ** p≤0.01; *** p≤0.001; ****p≤0.0001.

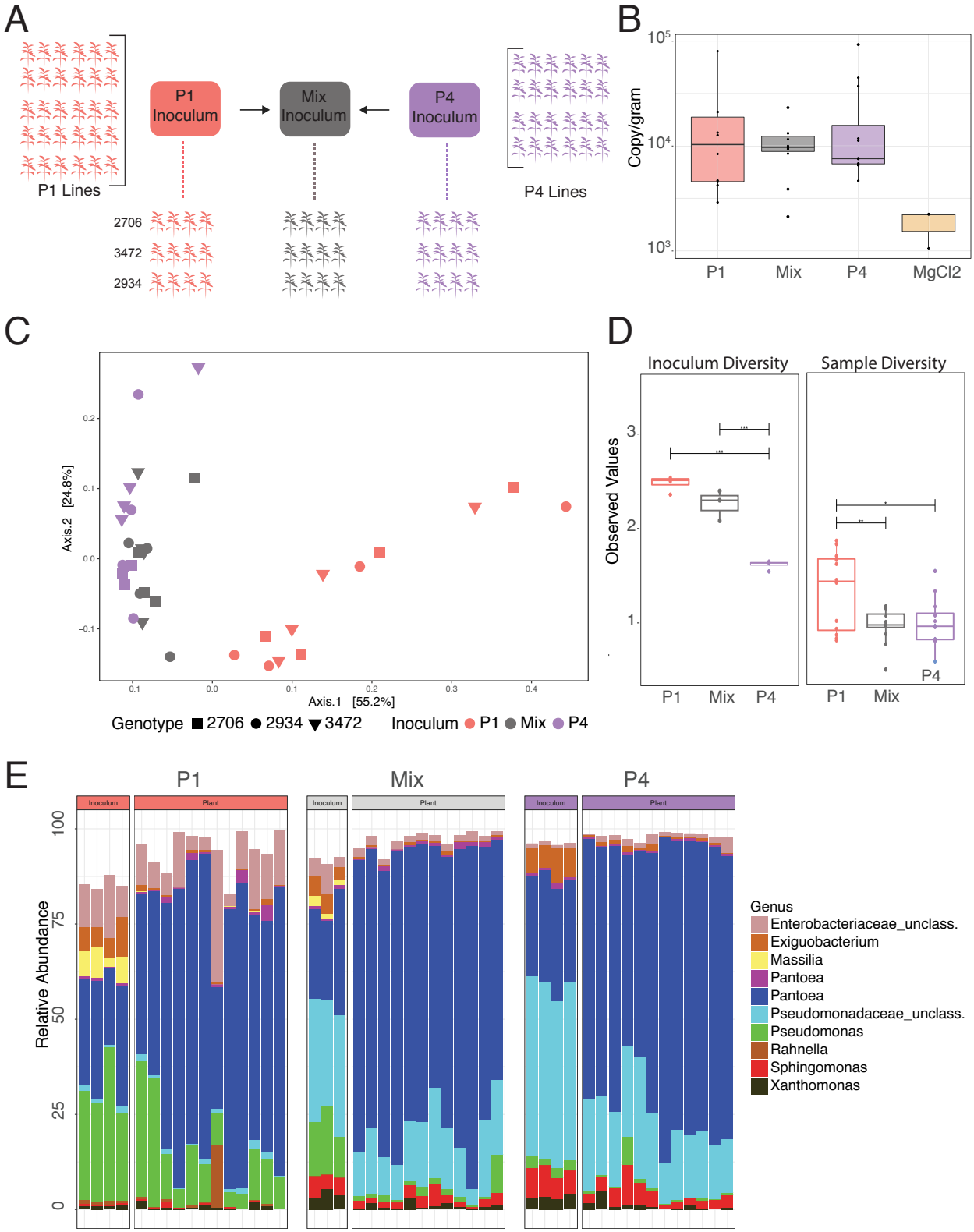
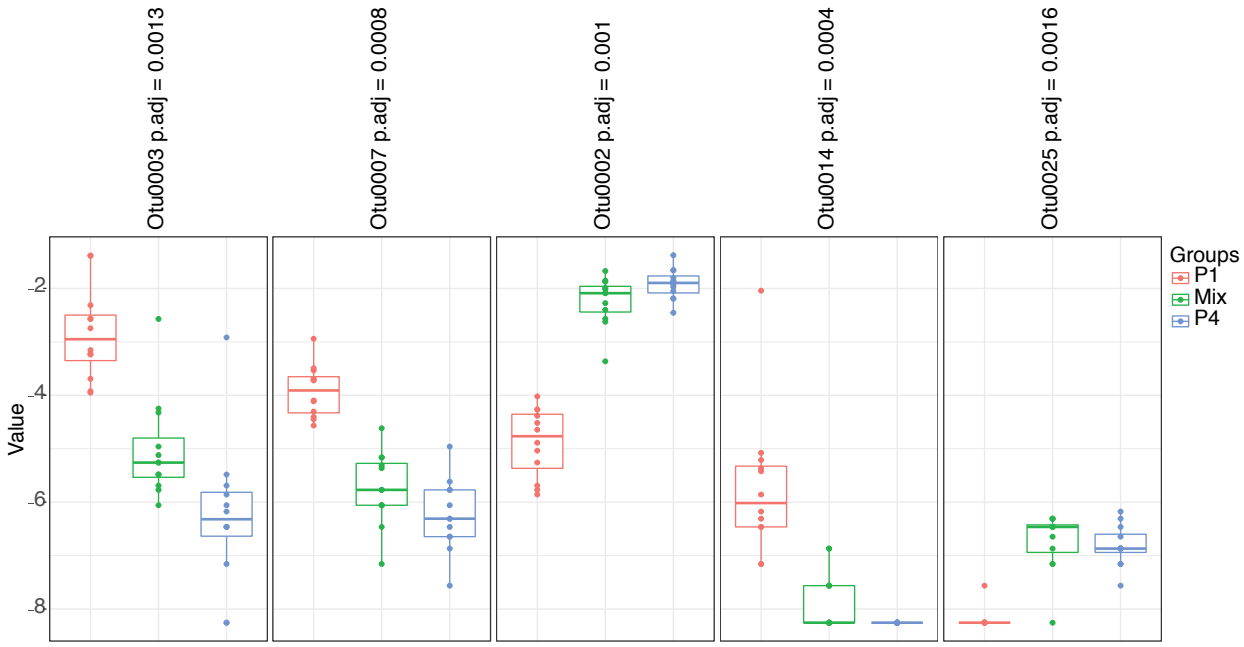


Figure 3-12 Testing microbiome adaptation

Figure 3-12 (continued) Testing microbiome adaptation
Plants were inoculated with pooled, passaged microbiomes from the end of P1, P4, or a 50:50 Mix of the two (a). Bacterial abundance was measured using ddPCR (b). A PCoA plot of Bray-Curtis distances (colored by inoculum source) shows that P1 plants have bacterial communities that are significantly different from P4 and Mix plants, which are indistinguishable (c). Shannon's alpha diversity of the inoculum and experimental plants (d) show significant differences between samples. A bar graph illustrating composition of the top 10 OTUs shows differences in taxa amongst both the inoculum and experimental plants (e).

A



B

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

Figure 3-13 Differentially abundant taxa
We performed a Kruskal-Wallis test on log-relative transformed OTU abundance at different passages using the MicrobiomeSeq package (a). This is a non-parametric method, and it tests whether samples originate from the same distribution. P-values are corrected for multiple testing using family wise error rates. Significant OTU rankings 1-5 are assigned importance using random forest classifier. Identities of OTUs are displayed as well (b).

Chapter 4. The impact of bacteriophages on phyllosphere bacterial abundance and composition

Parts of this chapter have been adapted from the following with permission:
Morella, N. M., Gomez, A. L., Wang, G., Leung, M. S., & Koskella, B. (2018). The impact of bacteriophages on phyllosphere bacterial abundance and composition. *Molecular ecology*, *27*(8), 2025-2038.

4.1 Introduction

Bacteriophage viruses (phages) that infect bacteria are both ubiquitous and abundant, so much so that they are estimated to largely outnumber bacterial cells in the environment [138–140]. Their role in controlling bacterial populations has been studied since their discovery as "bacteria-killers" in the early 1900s, and there are countless studies investigating bacteria–phage pairwise dynamics. Even the fundamental Luria–Delbruck fluctuation experiment that demonstrated mutations arise in the absence of selection was conducted using phage as the selective pressure [141]. In addition to fundamental research, lytic phages (i.e., phages that lyse their bacterial hosts in order to transmit) have been widely studied for their use as biological control and therapeutic agents, and have been successfully used in control of plant pathogens[142], including in tobacco [143], tomato [144–146], detached flowers of *Rosaceae* trees [147], and even some ready-to-eat foods such as hot dogs and lettuce leaves [148]. Beyond just controlling bacterial abundance, phage predation in the ocean is thought to impact biogeochemical cycling and food web processes through bacterial lysis, converting biomass into dis- solved organic matter and contributing to the dissolved organic carbon pool [140, 149]. Recently, increased interest in the human microbiome is also beginning to acknowledge a key role of lytic phages in these communities [150, 151], but in contrast to free-living microbiota, little empirical work has been done to uncover the role of phages in these systems. Specifically lacking is an understanding of when and how phages shape the abundance and composition of host-associated bacteria.

Phages can have important impacts on the competitive dynamics and structure of bacterial communities [152–155], and they are predicted to maintain bacterial diversity through a variety of mechanisms [156, 157]. As bacteria evolve to escape phage predation and phages counter-adapt to these resistances, evolutionary and co-evolutionary dynamics can drive important phenotypic and genotypic variation within the bacterial community (e.g., [153, 158]). One idea that is frequently put forth in marine systems, known as "kill-the- winner" dynamics [159, 160], suggests the most abundant bacteria should also be the most susceptible to phage predation. In this model, an increase in bacterial abundance is followed by an increase in the associated phage population and a subsequent decrease in bacterial abundance, effectively preventing one type of bacteria from ever dominating the community. Antagonistic co-evolution between bacteria and phage has recently been put forth as a driver of bacterial diversification and variation within the human gut microbiome [161], with potential impacts on microbiome function and human health through phage-mediated homeostasis or dysbiosis [162, 163].

The impact of lytic phages on bacterial cell density and community diversity may in part be the result of cell lysis, which not only has a direct effect on the population of cells, but also has an indirect effect on competition among bacterial strains and species within a community. Phages may also increase bacterial density and diversity by releasing nutrients into the environment via lysis (increasing availability for other members of the community [164]) by

conferring metabolic or morphological traits to bacteria upon integration into the genome (such as the gain of virulence factors in *Staphylococcus aureus* MRSA [165] and the cholera-causing toxin in *Vibrio cholera* [166], or by facilitating horizontal gene transfer among bacteria [155]. As such, understanding and predicting the link between phage and bacterial diversity could, in theory, be harnessed to increase the overall productivity of a system, or at least studied to better understand how diversity contributes to whole-system function and productivity [167].

Technical limitations have also impeded our ability to under- stand the role of phages in host-associated microbiomes. Specifically, there is no universal gene marker similar to 16S rRNA for bacteria or 18S/ITS for fungi, making culture-independent identification of phages difficult and dependent on more costly and computationally challenging metagenomics sequencing (although taking this approach can be very powerful and has recently proven fruitful in describing a "core phageome" in the human gut [162]). Furthermore, culturing phages can be difficult, as it is reliant on having an isolate of a suitable bacterial host. Despite these difficulties, studying phages in host-associated microbiomes may reveal interactions between bacteria and phages that are different from those occurring in free-living microbial populations. For example, recent work has suggested that microbiomes may be dominated by temperate phages that integrate into the host genome as opposed to lytic phages that lyse their host cells (summarized [168]). In this case, changes in bacterial density or composition due to lysis would likely represent only one small way in which phages impact their microbiome. Furthermore, host factors such as age, immunity and health are likely to change the dynamics between bacteria and phages within the host environment (discussed [163]). Another key difference in host-associated microbiomes compared to free- living bacterial communities is the process of colonization of a new host. Phages may play a key role in early microbiome establishment, the importance of which will be specific to the mode of microbiome transmission and the diversity and composition of colonizing bacteria.

In this work, we sought to investigate the role of lytic phages in shaping bacterial abundance and community composition during colonization of the phyllosphere (the aerial surfaces of plants). We used a filtration method to deplete phages from the microbial community associated with tomato leaves, modified from an approach that has previously proven effective for separation of phage and bacteria in seawater [169–173]. Through inoculation of a size-fractionated field-grown tomato microbiome onto juvenile, growth chamber-grown plants, we are able to test whether the lytic phage fraction of the phyllosphere has an impact on bacterial abundance, composition and diversity during microbiome establishment. We find an impact of phages on overall bacterial abundance and relative abundance of specific taxa that is measurable after 24 hr, but not at 7 days post-inoculation. We also find evidence for slightly higher alpha diversity after 7 days in those communities in which phages were initially present in the inoculum relative to those in which they were depleted. Our results emphasize that lytic phages are likely to be an important component of the microbiome and are capable of influencing both bacterial abundance and diversity over short timescales.

Table 4-1: Experimental Summary

Title	Description	Treatments	Data	Figure
Proof of Concept	Proof of Concept for phage depletion methods	Bacteria, Bacteria + known phage	Abundance using ddPCR, plating for live bacterial cells in various fractions	3-1
Field microbiome Exp. 1	6 tomato field site communities	Bacteria only, Bacteria + 0.22 µm filtrate containing free phage particles	16S General abundance (ddPCR)	3-2
Field microbiome Exp. 2	1 tomato field community	Bacteria only, Bacteria + phage fraction in buffer, Bacteria + 100K MWCO leaf wash filtrate	16S General abundance (ddPCR), 16S Illumina sequencing	3-3, 3-4, 3-5, 3-6
Constructed Community Exp.	Constructed Community (CC) of phyllosphere bacterial isolates	CC, CC + phage fraction from field community, CC + 100K MWCO leaf wash filtrate from field community	16S General abundance (ddPCR), phage fishing on bacterial isolates	Figure 3-7

4.2 Results

4.2.1 Phage Depletion Proof of Concept

In our first of multiple experiments (Table 4-1), we conducted a proof of concept experiment. We used ddPCR to measure quantities of known phage and bacterial host in size fractions of our mock community (two phages and *Pseudomonas syringae* pv. tomato), and we determined that our fractionation method effectively concentrates phages from the leaf wash, allowing us to deplete them from both the "bacteria only" and 100K MWCO filtrate fractions of the leaf microbiome (Figure 4-1a, note the log scale). FRS and SHL bacteriophages were effectively depleted, although the ddPCR signal was not entirely eliminated in the 0.22- µm filter bacterial recovery fraction (representing the bacteria-only inoculum). Phage levels were concentrated from the 0.22- µm flow-through fractions in the 100K MWCO centrifugation unit, representing bacteria plus phage treatment. Lastly, we also measured decreased levels of phage in the 100K MWCO flow-through fractions, representing the additional phage-depleted inoculum: bacteria plus filtrate. FRS and SHL phages are approximately 60 and 80 nm in size, respectively, and we thus presume that most phages in the environmental samples that are that size or larger should be retained in the upper portion of the 100K MWCO centrifugation unit. Membrane pore size for the unit we used is 10 nm; therefore, smaller phage particles should have been retained in the upper fraction as well. Overall, we therefore consider both the bacterial/fungal fraction and the 100K MWCO flow-through fraction phage-depleted, but not necessarily absent of all phage. Lastly, levels of *P. syringae* pv. tomato abundance was measured in all fractions (Figure 4-1b), and signal was also detected in the nonbacterial fractions. However, this is likely due to the detection of DNA and not the presence of live cells, as bacteria could not be cultured from those filtered fractions (Figure 4-1c). As seen in Figure 4-1d,

infectious phage particles were present in the initial leaf wash, and they were also sufficiently high in concentration to completely lyse the bacterial lawn in the 0.22 μm flow through (not shown) and 100K MWCO concentrate fractions, as little to no bacterial growth is observed. By comparison, a solid bacterial lawn is seen in the 0.22- μm filter recovery sample, where most phages appear to be depleted. As evidenced by a small number of plaques, a few bacteriophages are present in the 100K MWCO filtrate. This further supports the possibility that the third treatment, bacteria plus filtrate, was phage-depleted, but not completely free of phages, in our subsequent field experiments.

4.2.2 Bacterial Abundance

In field-generated inoculum experiment 1, with 6 different leaf wash sources (Figure 4-2), a general linear model uncovered an overall significant effect of treatment (phage-depleted versus phage re-suspended; $F_{1,18}$= 8.847, p= 0.008) as well as day of sampling ($F_{1,18}$= 26.476, p<0.001) on bacterial abundance. Additionally, there was a significant interaction effect between treatment and day ($F_{1, 18}$=10.168, p=0.005). For analysis of data separated by day, we used a Mann-Whitney nonparametric test used to account for unequal variances. We found significantly higher bacteria abundance of bacteria-only samples at day 1, (Z= -2.567, p= 0.01) but not at day 7 (Z= -0.640, p= 0.522). Two samples from the bacteria plus phage plants at day 7 were excluded from this analysis due to insufficient levels of bacteria to measure.

In field-generated inoculum experiment 2, which consisted of a single microbial inoculum with six experimental replicates for each treatment (bacteria only, bacteria plus phage, and bacteria plus filtrate; Figure 4-3), we also found an overall significant effect of day, but in this case not treatment, on abundance (GLM, effect of day, $F_{1,30}$= 27.366, p< 0.001, and effect of treatment, $F_{2,30}$= 1.840, p=0.176) and no interaction effect. When data were analyzed by day, we observed only a marginally significant effect of treatment on bacterial density at day 1 (Kruskal-Wallis test, x^2_2= 4.994, p= 0.082) and no effect at day 7 (x^2_2= 4.538, p= 0.103). As can be seen in Figure 4-3 and Figure 4-7, we consistently observe high variation amongst experimental replicates in the combined bacteria plus filtrate samples, which we cannot explain given the complexity and possible heterogeneity of molecules likely to be found in this fraction. When this treatment is removed from the analysis to test our *a priori* predictions regarding the impact of phage depletion on abundance, we observe significantly lower bacteria abundance in the bacteria plus phage samples than bacteria only at day 1 (Z= -2.402, p= 0.016). In agreement with the first experiment, there are no significant differences in abundance by day 7 between bacteria and bacteria plus phage (Z =-0.641, p= 0.522).

4.2.3 Bacterial Composition

After rarefaction and filtering, there were a total of 200 OTUs present in the spray inoculum from field experiment 2 representing taxa from the four top phyla commonly found in the phyllosphere: *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Firmicutes* [38]. As expected, the bacterial composition of inoculum from the three different treatments, sampled after resuspension with/without phage but before growth, has similar rank order of relative abundance for the top OTUs (Figure 4-4a). Observed differences in relative abundance of specific taxa may be due in part to concentrated free bacterial DNA in the 100K MWCO fraction. Given the way in which inocula was prepared (suspension of bacteria from the filter in a concentrated volume, vortexing thoroughly, and equally aliquotting into three treatments), it is unlikely that the bacterial communities differed substantially between treatments at inoculation.

After 1 day on plants, OTUs in the family *Pseudomonadaceae* appear to be higher in relative abundance in plants from the bacteria-only treatment, while *Enterobacteriaceae* are more abundant in the other two treatments (Figure 4-4b). To test this, we used linear discriminant analysis effect size (LEfSe) analysis and confirmed the enrichment of OTU0002, family *Enterobacteria*, in bacteria plus phage treatment ($p< 0.05$) and OTU0005, family *Pseudomonadaceae*, in bacteria only ($p< 0.05$). Although there were a small number of additional significantly different OTUs observed across treatments, these were not present at high enough abundance for detection in the initial inoculum, and therefore it is unclear whether they were part of the inoculum or already present on the juvenile plants that were sprayed. By day 7, there was only one significantly different OTU among treatments, which was from the family *Micrococcaceae* and enriched in the bacteria plus filtrate plants ($p< 0.01$).

4.2.4 Community Dissimilarity

Using Bray-Curtis dissimilarity measures, we performed an ANOSIM test (a nonparametric multivariate analysis of variance) on all samples at both days from field experiment 2 to determine the effect size of variables day and inoculum treatment, or the percentage of variation amongst samples explained by each variable. When all samples from all treatments are analyzed together, we found that day explains 11% of variation ($p= 0.001$), and treatment accounts for 13% of variation ($p= 0.009$). Then, to directly compare bacteria only and bacteria plus phage samples, we removed bacteria plus filtrate samples, re-ran the ADONIS, and found similar results: day explains 12% of variation ($p=0.006$) and treatment explains 13% ($p= 0.002$). Samples are visualized on a Principle Coordinates Analysis plot (PCoA) based on Bray-Curtis distances (Figure 4-5a). We then split samples by day to analyze them separately (Figure 4-5b, c). Here, an ADONIS test indicates that treatment accounts for 31% of variation between bacteria and bacteria plus phage samples at day 1 ($p= 0.003$), but that by day 7, treatment does not significantly account for any variation amongst samples.

4.2.5 Richness and Diversity

At day 1, mean richness of OTUs for treatments (min, max) are as follows: bacteria only 68 (22, 143), bacteria and phage 74 (27, 101), and bacteria and filtrate 64 (19, 104), and at day 7: bacteria only 57 (11, 130), bacteria and phage 100 (78, 120), and bacteria and filtrate 76 (15, 122) (Figure 4-6a). Alpha diversity (Figure 4-6b) was calculated using Shannon's diversity index, which is a measure of richness and evenness of OTUs. A GLM uncovered an overall significant effect of day on Shannon's Diversity ($F_{1,28}= 8.53$, $p= 0.007$). There was no significant effect of treatment when days were analyzed together ($F_{2,28}=0.893$, $p=0.421$), nor when days were analyzed separately and bacteria plus filtrate samples were included in the model (Kruskal Wallis; Day 1: $x^2_2= 0.129$ $p= 0.937$; Day 7: $x^2_2= 3.315$, $p= 0.191$). There was, however, a marginally significant effect of treatment on diversity at day 7 when only the bacteria and bacteria plus phage treatments were compared ($F_{1,9}= 5.374$, $p= 0.046$).

To compare community dissimilarity among replicate plants across treatments, we calculated a mean Bray-Curtis distance for each plant (compared to all other plants within a given treatment at each day). This allowed us to determine whether communities were more or less dissimilar across replicate plants in one treatment versus another (Figure 4-6c). A GLM showed a significant effect of day ($F_{1,28} = 51.431$, $p< 0.001$), with no overall effect of treatment ($F_{2,28}= 1.389$, $p= 0.266$), but an interaction effect between day and treatment was also observed ($F_{2,28}= 3.463$, $p= 0.045$). When bacteria plus filtrate samples were excluded from this analysis,

there was no significant difference among treatments at day 1 ($F_{1,9}$= 2.961, p= 0.119), but by day 7 plants from the bacteria plus phage treatment were more similar to one another than were plants from the bacteria only treatment ($F_{1,9}$ = 6.231, p= 0.034).

4.2.6 Constructed Community Experiment

In the constructed community experiment using the phage and leaf wash 100K MWCO filtrate fractions from organic garden outdoor tomato plants with phyllosphere bacterial isolates, we were able to qualitatively replicate the outcome on bacterial abundance after 24 hours, as determined by ddPCR (Figure 4-7). We detected only low-levels of bacterial signal from plants inoculated with sterile buffer and were not able to culture any bacteria from these plants, and they were subsequently excluded from statistical analysis. We observed a significant difference in bacterial abundance among treatments overall (x^2_2= 8.346, p= 0.015), and a significant difference between the bacteria and bacteria plus phage treatments when the filtrate treatment was removed (Z= -2.309, p= 0.021). Furthermore, as determined by a large zone of clearance on the soft agar overlay plate, we found a bacteriophage from the phage-fraction that could lyse an isolate from the family Enterobacteriaceae (96% identity match of 16S rRNA gene amplified with primers 27F/1492R to *Pantoea agglomerans*).

4.3 Discussion

Using a community-level phage depletion approach, we found that the phage fraction of the phyllosphere microbiome from field-grown tomato plants impacted bacterial abundance and composition during microbiome establishment on a new host. When microbial communities were sprayed onto juvenile tomato plants after either phage depletion or resuspension with the depleted phage-fraction, we observed decreased abundance in the latter treatment after 24 hours across three different experiments (Table 1): first with six independent leaf wash sources (Figure 4-2), then with one leaf wash source and six plant replicates per treatment (Figure 4-3), and finally with a constructed bacterial community and natural phage fraction (Figure 4-7). Using 16S rRNA Illumina MiSeq data from field experiment 2, we were able to further show that the phage-fraction of the phyllosphere affects microbiome composition, including relative abundance of specific OTUs (Figure 4-4). We observed an effect of phage depletion treatment on community dissimilarity between treatments after 24 hours, but not after 7 days (Figure 4-5). We also found some evidence for differences in both alpha and beta diversity between phage depleted and phage re-suspended communities after 7 days (Figure 6). Overall, these results support the idea that lytic phages can mediate bacterial dynamics within host-associated bacterial communities, as they have been found to do in free-living communities [169, 171, 174].

Across these experiments we observed a decrease in overall bacterial abundance 24 hours after inoculation, suggesting that phages affected growth of the most common and/or fastest growing bacterial strains during colonization of a new plant host. However, it is important to note that decreased overall bacterial abundance is not necessarily an expected outcome of lytic phage action within a microbiome. This is both because phage-mediated lysis has been shown in some cases to increase population growth due to release of nutrients [164] but also because other strains that are not being targeted by phages should be able to offset any decreased growth of susceptible bacteria. That the impact of phages on abundance in our experiments was short-lived (i.e. no longer evident at day 7) suggests either that phages are particularly impactful during initial colonization, as bacterial population are rapidly growing, or that resistant bacterial strains/species increased in density over time to utilize existing resources. Indeed, the Kill the

Winner hypothesis predicts that phages should most commonly prey upon highly competitive bacterial species [159, 160]. Results of our sequencing efforts supports this model, as we found different relative abundances of the two dominant families when the phage fraction was versus was not present in the initial inoculum. After 24 hours, the bacteria plus phage treatment plants were observed to have lower abundances of *Pseudomonads*, but when the phage-fraction was depleted there was an overabundance of an OTU within the family *Enterobacteriaceae*. However, after seven days the differences in relative abundance of these two OTUs were no longer observed to differ among treatments.

Although only marginally significant, the presence of phage in the inoculum also led to an increase in alpha diversity at seven days post-inoculation. Again, this result may have been driven by a decrease of *Pseudomonads* after the first 24 hours, perhaps allowing a richer community to develop after the first week. Interestingly, when comparing beta diversity among treatments using averaged Bray-Curtis distances between samples within a treatment, we found an interaction effect between day sampled and inoculum treatment. This suggests that the phage fraction of the microbiome may also be having an effect on among-host microbiome diversity, initially driving divergence among communities as the empty niches are filled, (as has been observed previously at the within-population level; [175]), but eventually leading to more synchronous community structure. It is important to note that the patterns we observed were based on the depletion of lytic phages from the microbiome at the point of inoculation, but there were almost certainly many temperate phages remaining (i.e. phages embedded within the host genomes) and possibly some lytic phages contained within bacterial cells (*i.e.* pre-lysis) at the time of collection/filtration. As such, it is possible that differences in treatment effect observed between 24 hours and 7 days were due to the resurgence of phages in the phage-depleted communities rather than loss of phages in the bacteria plus phage treatment. The observed transience of phage-mediated impacts on abundance and diversity is intriguing, and longer-term studies with more time points are needed to better understand temporal effects of phages on bacterial communities.

One question we were not able to directly address in this series of experiments is the constituents of the leaf wash filtrate (i.e. the flow through from the 100K MWCO filter). The molecules and small proteins found in this filtrate had a surprisingly large and variable impact on the phyllosphere microbiome, impacting both abundance and community composition and causing high variation among biological replicates. In future experiments, additional size fractionation of the leaf wash filtrate and/or mass spectrometry analysis of these fractions may help address this question. As observed in our proof of concept experiment, it is also possible that some bacteriophages made it through the filtration step and were present in this treatment. We decided to eliminate this treatment from many of our analyses due not to the effect of the treatment itself (which was often in the same general direction as the phage treatment) but rather due to the high variances observed across replicate plants. In most cases, plants within this treatment spanned the variation observed in both the bacteria alone treatment and the bacteria plus phage treatment. It was therefore unclear to us how to interpret this treatment and what biological significance it might have, but further study is certainly warranted. Another limitation of this work is that we have not identified the specific phages in the phage-fraction of the experiment. We have taken measures to ensure that the method used for separation of microbiome fractions is effective at separating phage from bacteria, but in order to fully describe the diversity of phage, as we have done for the bacterial community, one would need to take a metagenomics approach. Furthermore, there may be other entities that are phage-sized in that

fraction of the microbiome, such as extracellular vesicles [176, 177] or spores of bacteria such as *Bacillus* [178] that impact upon microbiome colonization. However, given that the current estimates of phages largely outnumber bacteria in the environment, we expect non-phage particles to be far less abundant than phages in this size fraction. This was recently shown for outer-membrane vesicles, where they were estimated to represent less than 0.01-1% of SYRB DNA-stained phage-sized particles quantified in seawater [179]. Furthermore, we cannot rule out the possibility that the presence of phage, but not their predation on specific taxa, is causing the effects we are observing. However, by recapitulating the results of decreased abundance in bacteria after 24 hours when a phage fraction was present in our constructed community, we were able to lend some insight to this question. In this case, our detection of a phage capable of lysing a member of the constructed community suggested that the phage fraction was most likely driving the observed decrease in abundance. This is further supported by the fact that the phage was found to lyse *Pantoea agglomerans*, a member of the family *Enterobacteriaceae*, which we have found to be in high relative abundance in 16S rRNA community data in both this experiment and other unpublished work. Another important note is that the ddPCR protocol used here relies on lysis of bacteria cells through a hot-start step in the PCR. Because of this, it is possible that our abundance measures do not take into account hard-to-lyse bacteria. Finally, we did not include any analyses of the fungal communities in these microbiomes, as it was outside the scope of the current work. However, it is possible that our filtration methods also impacted any fungal viruses that might have been present in this study. How fungal communities are influenced by viruses within the microbiome is certainly an open question in the field that warrants further study.

Given the building evidence that the phyllsophere microbiome is a key component of plant fitness, influencing key functional traits [180] and likely protecting host plants against disease [28, 56, 181], the idea that lytic phages impact these communities is of direct relevance to plant health. A better understanding of bacteria-phage dynamics within these systems may present opportunities for manipulating the plant microbiome and ultimately increasing plant health. These ideas can be extended to the human microbiome, where the role of phages is proving to be appreciable [161, 168, 182]. With regard to using phages in therapeutics, their role in controlling bacterial community dynamics and local adaptation is an important consideration for both phage-therapy to target specific pathogens and full-microbiome perturbations or replacements via fecal transplants. Overall, our results make a significant contribution towards the empirically demonstration of the role that phages play in shaping bacterial community structure in natural systems. This may be through, but is not limited to, impacts on bacterial abundance, composition, competitive-dynamics, and/or diversity. These effects are ultimately likely to affect the overall stability and function of the microbiome, and consequently, host fitness. In conclusion, it is becoming increasingly clear that phages should be considered when seeking to understand the diversity, evolution, and ecology of any microbiome.

4.4 Materials and Methods

4.4.1 Phage Depletion

We first verified a method to deplete phages from the phyllosphere microbiome by creating a mock bacterial community with known bacteria (*Pseudomonas syringae* pv. tomato strain DC3000*)* and two tailed DNA phages (FRS and SHL, provided by Omnilytics Inc). The mock community was sprayed onto tomato leaves (n=3) and, after leaves no longer appeared to be dripping (approximately 10 minutes), plants were placed in 15mL conical tubes in sterile

10mM $MgCl_2$ into a sonicating water bath (Branson) for 10 minutes. The leaf wash was passed through a 0.22µm filter, and bacteria were recovered by vortexing the removable filter for 5 minutes in 10mM $MgCl_2$ at medium speed. Then, to separate phage away from other small molecules and proteins in the leaf wash, the 0.22µm flow through was further fractionated using a 100 kilodalton Molecular Weight Cutoff (100K MWCO) centrifugation unit (Pall Macrosep, 10nm pore size) for 25 minutes at 4,000 X G, as per manufacturer's instructions. We then used droplet digital PCR (ddPCR; BioRad QX200^{TM} System) to measure levels of bacteria and phages in each fraction. Our ddPCR probes target the capsid gene of each of the two phages or the 16S rRNA gene of *P. syringae* pv. tomato (Bergmark et al., 2012). We measured levels of bacteria and phages from original leaf wash, the filter recovery, the 0.22µm flow-through fractions, and the upper and low collection fractions from the 100K MWCO unit. Additionally, we plated 100µl of each fraction on Kings Broth Agar plates to verify presence or absence of live *P. syringae* pv. tomato cells, as DNA from lysed cells would be expected to be present in each fraction after filtration. To verify the presence of infectious bacteriophages (plaque forming units) in the same samples, 100µl of each fraction was combined with 200 µl of an overnight culture of *Pseudomonas syringae* pv. tomato strain DC3000 and 7mL of KB soft agar, briefly vortexed, and plated onto KB hard agar plates. All plates were incubated overnight at 28°C.

4.4.2 Plant Collection and Inoculum Generation

With the goal of using a 'natural' microbiome for subsequent studies, we sampled tomato leaves from the UC Davis Student Farm between the months of August and October. For field experiment 1, (Table 1), inoculum was generated from each of six different sites from across three different fields (n=1 tomato plant per site). For the subsequent experiment with sequencing data, field experiment 2, leaves were pooled across fields into a single diverse inoculum source (n=6 tomato plants). We sought to create as diverse of a starting inoculum as possible, and thus we sampled from multiple fields that had different tomato genotypes and different surrounding plants, as microbes from surrounding plants are known to contribute to colonization of the phyllosphere [40]. Tomato genotypes sampled for inoculum preparation included Beef-eaters, Romas, Lucky Tiger, Red Cherry, Yellow Cherry, and Blush Tomato. Surrounding plants included oak trees, corn, and watermelon. Leaves were collected from individual plants at multiple canopy levels. Inoculum for each leaf wash was then generated by submerging 100-200 grams of leaf material in 10mM $MgCl_2$ and sonicating it in a Branson 550 sonicating water bath for 10 minutes.

As described above, we used both 0.22µm vacuum filtration and the 100K MWCO centrifugation unit to separate environmental phage particles from the rest of the microbiome. The leaf wash was first passed through a 0.22µm filter, and bacteria were recovered by vortexing the removable filter for 20 minutes in 10mM $MgCl_2$ at medium speed and then centrifuged at 4,000 x G for 10 minutes and resuspended in 3mL sterile $MgCl_2$. The 0.22µm leaf wash flow-through was passed through a second vacuum filtration unit to ensure removal of bacteria. The bacteria were collected from the first filter unit and divided evenly into three aliquots and recombined with either: (i) sterile 10mM $MgCl_2$ buffer (**bacteria only**); (ii) 10mM $MgCl_2$ containing the phage fraction from the 100K MWCO unit (**bacteria plus phage**); or (iii) an equal volume of the leaf wash filtrate from the 100K MWCO unit (**bacteria plus filtrate**), which should represent another phage-depleted inoculum, but includes any molecules or particles smaller than 100 kDa in the sample.

For field experiment 1, we prepared the bacteria only and bacteria plus 0.22µm flow-through fractions, but we did not perform the third separation of the phage fraction from the rest of the leaf wash using the 100K MWCO centrifugation unit (see Table 1). After observing the significant effect that the 0.22µm fraction had on bacterial abundance, we then sought to dissect the effects of phage particles away from that of other small molecules and chemicals that were also removed from the leaved by sonication. In field experiment 2 and the constructed community experiment, therefore, we tested the effect of the leaf wash 100K MWCO filtrate treatment separately from that of the phage fraction. For field experiment 1 we verified that starting inoculum across the six different sources was within the same order of magnitude using OD_{600} measurements, and for the second experiment, we quantified starting bacterial density using CFU plating on Kings Medium, which was determined to be $\approx 5*10^5$ CFU/mL. Size fractions and bacteria were combined to create each inoculum immediately before application onto tomato plants to avoid any impact of phages as a result of interactions outside of the host.

4.4.3 Plant Inoculation

Money Maker tomato seeds were surface sterilized by soaking 20 minutes in 2.4% Sodium Hypochlorite and then rinsed twice in H_20. Seeds were germinated at 21°C on sterile 1% water agar in the dark. Upon emergence of cotyledon leaves, tomato plants were transplanted into Sunshine Mix potting soil and grown for three additional weeks in a growth chamber at 80% relative humidity, 28°C, and 15-hour days. At four weeks of age, plants were spray inoculated with inoculum (n=6 plants per treatment), each receiving approximately 3.5mL of inoculum, enough to thoroughly wet all leaves. Plants were returned to the growth chamber and spatially randomized across treatments.

4.4.4 Sample Collection

After approximately 24 hours and again after 7 days, plants were sampled in the following manner. For field experiment 1, four hole-punches from two separate leaves were taken from each plant and placed in 1.5mL tubes. For field experiment 2, two complete leaves, one from each side of the plant, were sampled from each replicate plant and placed into 15mL conical tubes (Falcon brand). Whole-leaf samples were weighed, and we then added 500µl of sterile $MgCl_2$ to tubes from the first experiment and 10mL of sterile $MgCl_2$ to tubes from experiment two, before sonicating as described above for 10 minutes. Bacterial density was measured from this leaf wash for both experiments, and in the case of the second, the remainder was pelleted for DNA extractions.

4.4.5 Measuring Bacterial Abundance

Bacterial abundance was measured directly from leaf wash using ddPCR and primers designed to decrease amplification of plant chloroplast and mitochondrial DNA, 799F/1389R [183, 184]. Three µl of leaf wash was used as template with BioRad QX200™ EvaGreen Supermix and primers at a final concentration of 20nm. Twenty µl PCR reactions were loaded into the droplet generator with 70 µl of droplet generation oil, droplets were generated as per manufacturer instructions, and resulting droplets were transferred to a PCR plate. PCR cycling conditions were: 95°C for 3 minutes, 94°C for 20 seconds, 53°C for 40 seconds, 72°C for 40 seconds, repeat steps 2-4 35 times, 72°C for 7 minutes, 4°C 5 mins, and 90°C 5 minutes. The PCR plate was then transferred to the BioRad droplet reader where EvaGreen signal was measured by the machine. Cutoff thresholds were set for each row based on background

fluorescence in no template controls. Bacterial abundance values were normalized to amount of plant tissue collected for each sample (though care was taken to sample the same type and approximate age of leaves of each plant), and bacterial abundance is reported as rRNA copy number per mL of buffer per gram of plant material for field experiment 2. Field experiment 1 did not require plant weight normalization due to the sampling method described above

4.4.6 DNA extractions, 16S rRNA Amplification and Sequencing

DNA was extracted using MP Biomedicals FastDNA™ Spin Kit and cleaned using MoBio PowerClean DNA Clean-Up Kit. The 16S rRNA gene was amplified using dual-indexed primers designed for the V3-V4 region [87] using the following primers: (341F (5'-CCTACGGGNBGCASCAG-3') and 785R (5'-GACTACNVGGGTATCTAATCC-3') [88]. In order to decrease amplification of plant mitochondrial and chloroplast DNA, we also used peptide nucleic acids, PNAs [89]. PCR reactions contained 20µg BSA, PNAs at a final concentration of 0.5uM, 2 µl of template, primers at a final concentration of 20nm, and ThermoFisher Platinum™ Taq DNA Polymerase High Fidelity. Negative buffer controls, PCR controls, and original inoculum were also sequenced. Reaction conditions were: 94°C for 3 minutes, 94°C for 45 seconds, 78°C for 10 seconds, 50°C for 1 minute, 72°C for 1.5 minutes, repeat steps 2-5 30 times, and 72°C for 10 seconds. PCR reactions were randomized in order, run in triplicate for each sample, pooled, and quantified using Qubit. Amplicons from each sample were then pooled in equimolar concentrations, cleaned using an AMpure bead cleanup kit. Libraries were prepared for paired 300-nucleotide reads in Illumina's MiSeq V3 platform (Illumina) at The California Institute for Quantitative Biosciences (QB3) at UC Berkeley and run in 1 lane.

4.4.7 Data Analysis

For culture-independent community analyses, sequence files were demultiplexed by QB3 sequencing facility. Reads were combined into contigs using VSearch [134] and the remainder of the analysis was carried out in Mothur [93] following their MiSeq SOP [94]. Data were quality filtered and chimeras were removed using UChime [95]. The Mothur pipeline uses *de novo* cluster-based OTU picking that does not rely on taxonomic assignment [185, 186], and we used a 97% similarity cutoff for defining OTUs. The Silva reference database [96] was used for sequence alignment and taxonomic assignment. Archaeal, chloroplast, mitochondrial, and unknown domain DNA sequences were removed. Once an OTU table was generated in Mothur (the remainder of the analysis was done in R using the Phyloseq [97] and vegan packages [98]. In order to account for reagent contaminants, we also sequenced two DNA extraction kit controls and two PCR controls along with our samples. Contaminant OTUs from control samples that were at a similar or higher relative abundance in control samples compared to experimental samples were removed from the full OTU table. Reads were then rarified to 8,000 reads/sample. Cleaned, but unfiltered data, rarified to 500 reads/sample, were used for alpha diversity measures, though patterns remain consistent when analyses were run using the 8,000 read/sample rarified dataset. We chose to analyze data at both rarefaction levels because rarefaction to 8000 reads/sample resulted in the loss of two samples due to insufficient read depth. The 8000 reads/sample data were further filtered and the bottom 0.01% of OTUs were removed and significant differences in abundance of OTUs between samples were determined using the LDA effect size (LEfSe) via Galaxy web application [120]. The same filtered data were used to calculate Bray-Curtis distances, which were used for beta diversity and Principle Coordinates

analysis. Graphs were plotted using the GG2 plot package in R [99]. All other statistical tests were performed in SPSS version 24. For field experiment 2, statistical analyses were first run with all treatments, including the 100K MWCO filtrate treatment. However, the *a priori* predictions we sought to test with this experiment were comparing phage-depleted and phage co-inoculated treatments. We had no expectations about what would happen in the 100K MWCO filtrate treatment. Therefore, in some cases (clearly indicated when this was the case), the bacteria plus filtrate samples were removed from statistical analysis in order to allow direct testing of our predictions. This was especially important given the surprisingly large amount of among-replicate variation in both abundance and composition observed in the bacteria plus filtrate treatment.

4.4.8 Constructed Community

A final experiment was conducted using a constructed bacterial community of tomato phyllosphere isolates. We sought to include representative isolates of the most abundant families based on bacterial community sequencing data, and sixteen isolates were chosen (Appendix C). Isolates were grown separately overnight in Kings Broth, and the cultures were normalized to an OD_{600} of 0.0013, approximately $1x10^6$ CFU/mL. To generate the non-bacterial fractions of inoculum, tomato leaves were collected from the UC Berkeley Student Organic Garden, and leaf wash and fractionation was performed as previously described. The bacterial portion of the leaf wash was discarded, and the phage fraction and leaf wash fraction were combined with the constructed community of phyllosphere isolates to generate bacteria plus phage and bacteria plus filtrate treatments. Tomato plants were inoculated and sampled as described above (n=4 for each treatment) with the addition of four plants inoculated with sterile buffer only. Additionally, given that in this case we had a panel of culturable hosts, we sought to isolate phage from the phage-fraction used in this experiment with the following method. Ten µl of phage fraction was combined with 80 µl of bacterial culture from each of the 16 isolates and 7mL of Kings Broth and grown overnight at 28°C to enrich for phage. The following day, cultures were filtered using a 0.22µm filter unit. Filtrate was spotted on soft agar overlays of each of the 16 bacterial isolates from which it was generated. Additionally, filtrate from each of the 16 enrichments was pooled and spotted onto all 16 soft agar overlays. Plates were incubated at 28°C overnight and presence of lytic phage was determined by presence of zones of clearing, *i.e.* plaques, on the soft-agar overlay plates.

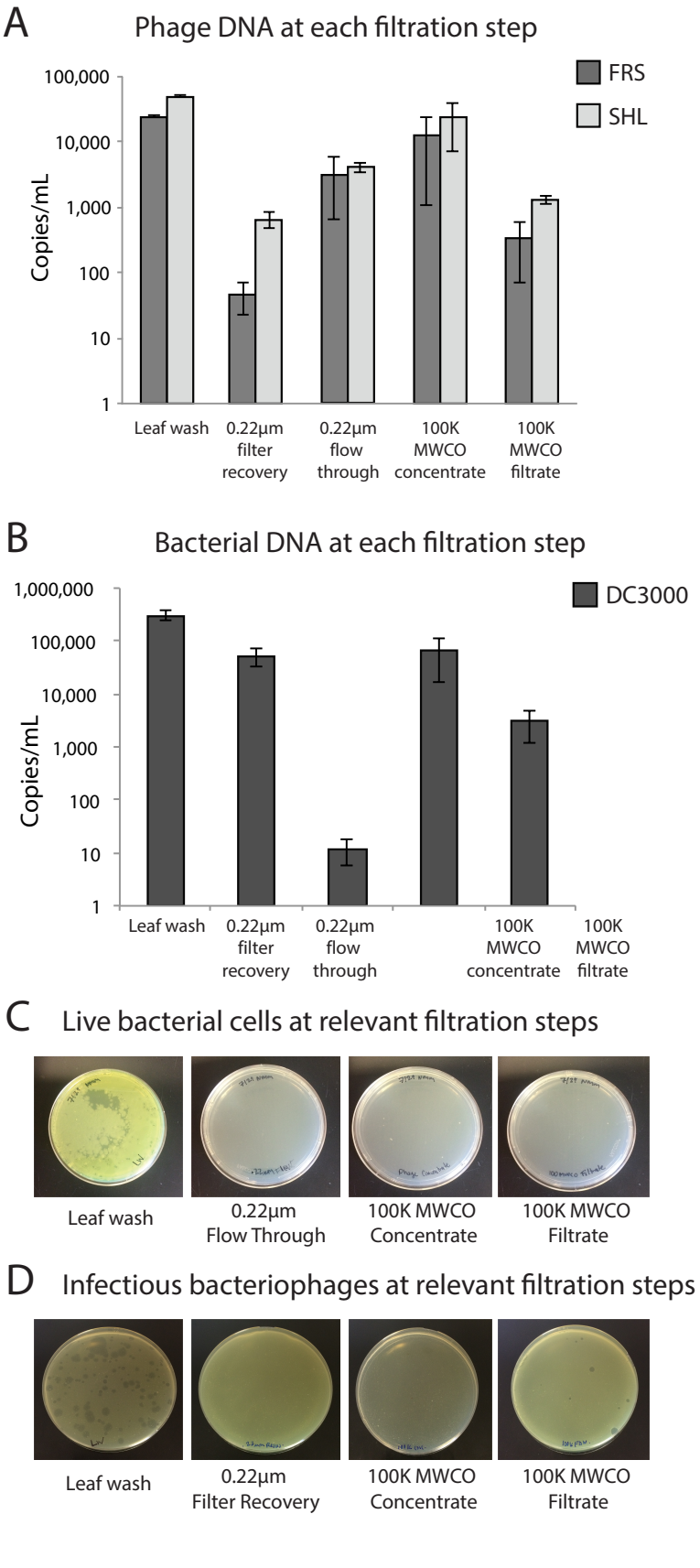


Figure 4-1 Phage depletion proof of concept
Size fractionation depletion of environmental phage was verified using purified phage particles, FRS and SHL, pure bacterial culture (DC3000) and DNA probes designed for each phage and bacteria. (a) Measurements of phage levels indicate depletion of phage from both the 0.22-µm filter recovery and 100K MWCO filtrate fractions. (b) Bacterial signals indicate that bacteria were successfully recovered from the 0.22-µm filter. (c) Although there was bacterial signal in both fractions of the 100K MWCO filtration step, bacteria could not be cultured from the 0.22 µm filtrate, 100K MWCO concentrate nor the final 100K MWCO leaf filtrate—as compared to the initial leaf wash. (d) Infectious bacteriophages were detected in the initial leaf wash (seen as individual plaques) and the 100K MWCO concentrate (little to no bacterial growth was observed). Phages were not detected in the 0.22-µm filter recovery fraction (solid bacterial lawn) and were detected in low concentration in the 100K MWCO leaf filtrate (minimal plaque formation). Error bars are ±1 SE of experimental replicates, and all y-axis are log scale

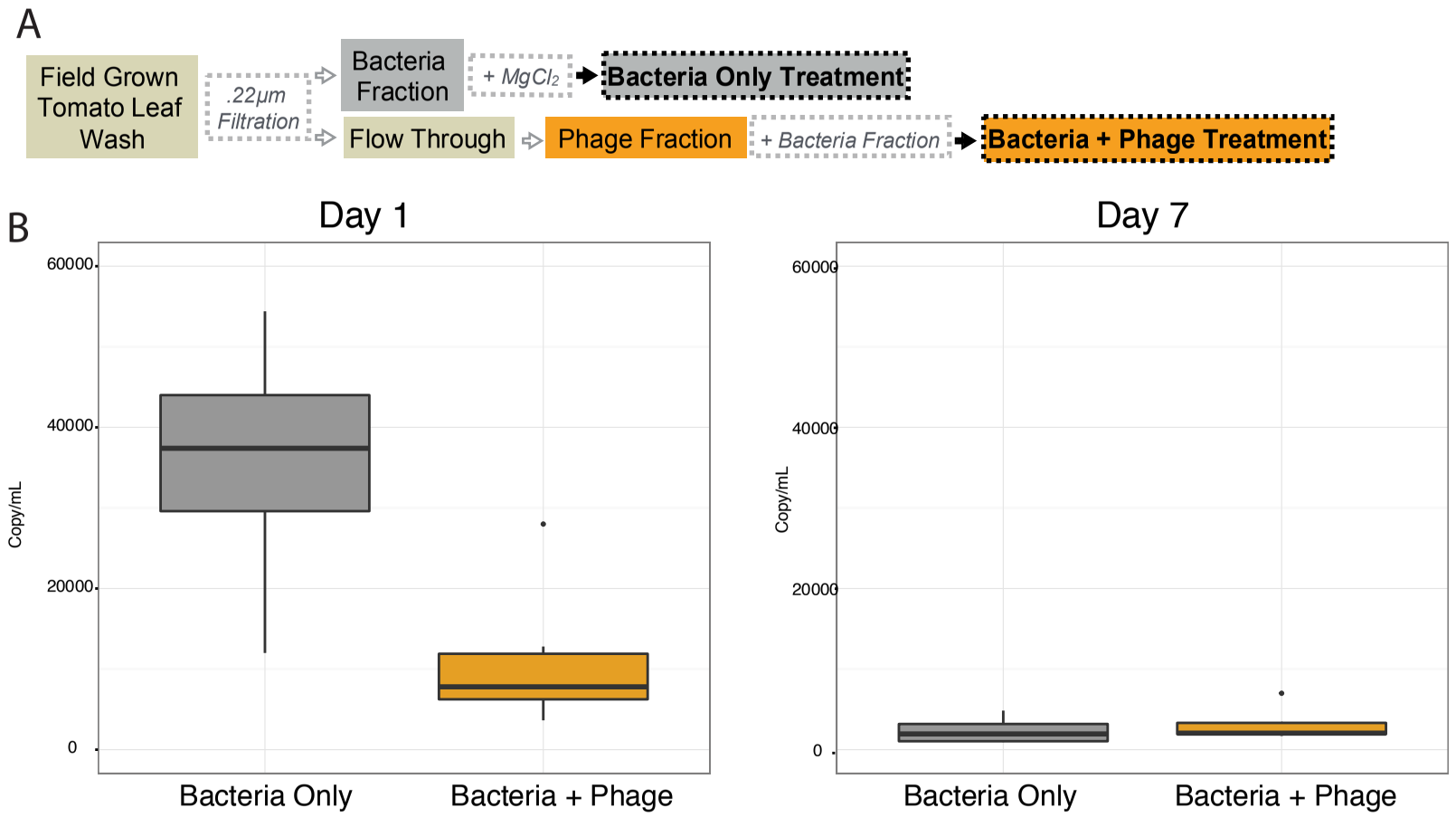


Figure 4-2 Comparison of bacterial abundance in phage-depleted vs. phage-resuspended phyllosphere communities grown on tomato

For field microbiome experiment 1, bacteria and bacteria plus phage treatments were generated from six different leaf wash sources (a). Bacterial abundance of plants (n = 1 per leaf wash per treatment) was measured at day 1 and day 7 using 16S rRNA primers and ddPCR (b) and displayed on boxplots where black line indicates median, the upper portion of the box is third quartile, lower portion is first quartile, whiskers are maximum and minimum, and outliers (any value more than 1.5× length of the box) are black dots. Y values indicate 16S copy number per ml of leaf wash volume (for a fixed size of leaf tissue sampled). **represents p value of <.001

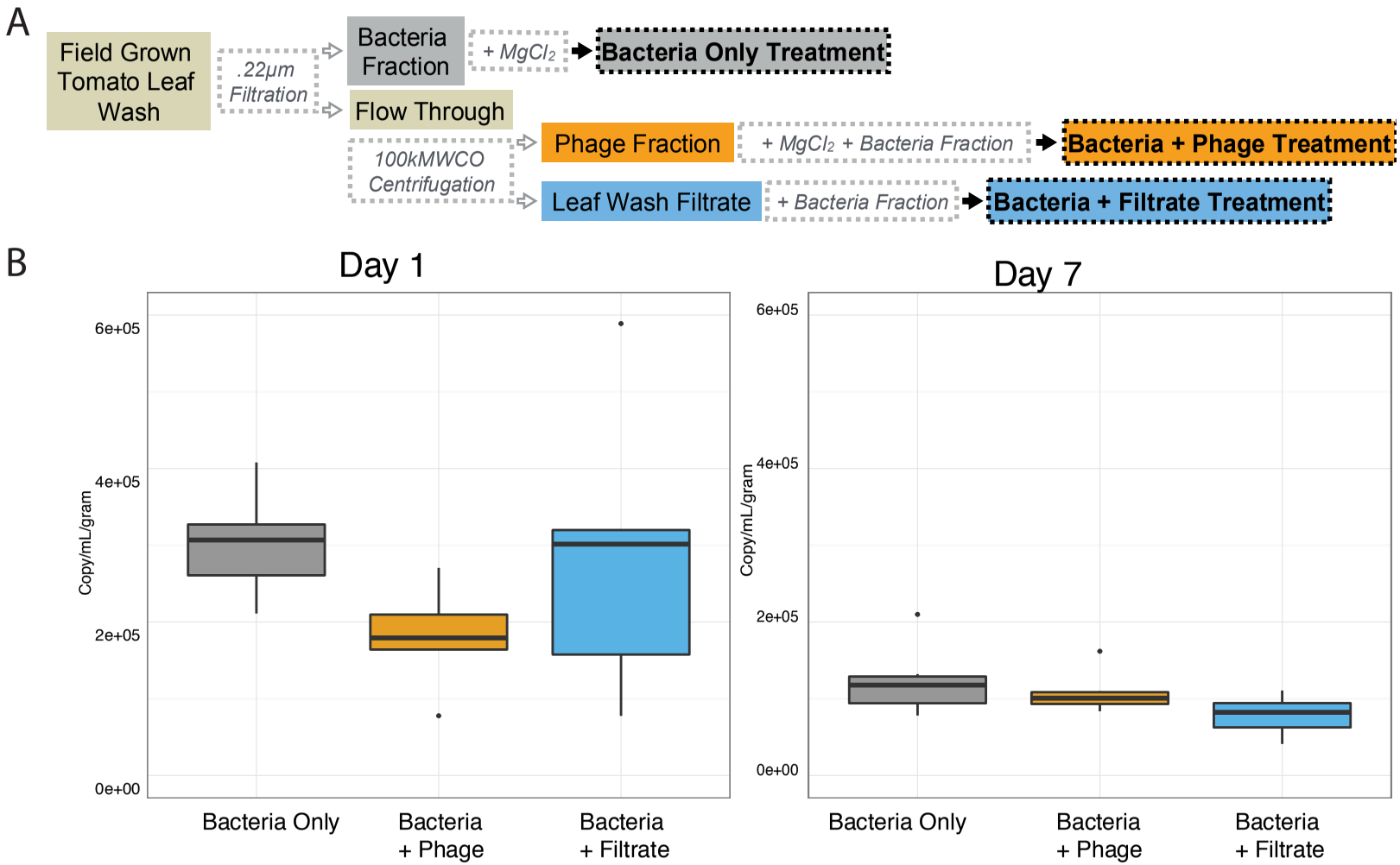


Figure 4-3 Comparison of bacterial abundance in phage-depleted vs. phage-resuspended or 100K MWCO leaf wash filtrate treatments

For field microbiome experiment 2, bacteria, bacteria plus phage and bacteria plus filtrate treatments were generated from one leaf wash source (a). Bacterial abundance of plants (n = 6 per treatment) was measured at day 1 and day 7 (b) and displayed on boxplots Y values indicate 16S copy number per ml of leaf wash volume, normalized per gram of plant tissue collected. * represents p value <.05.

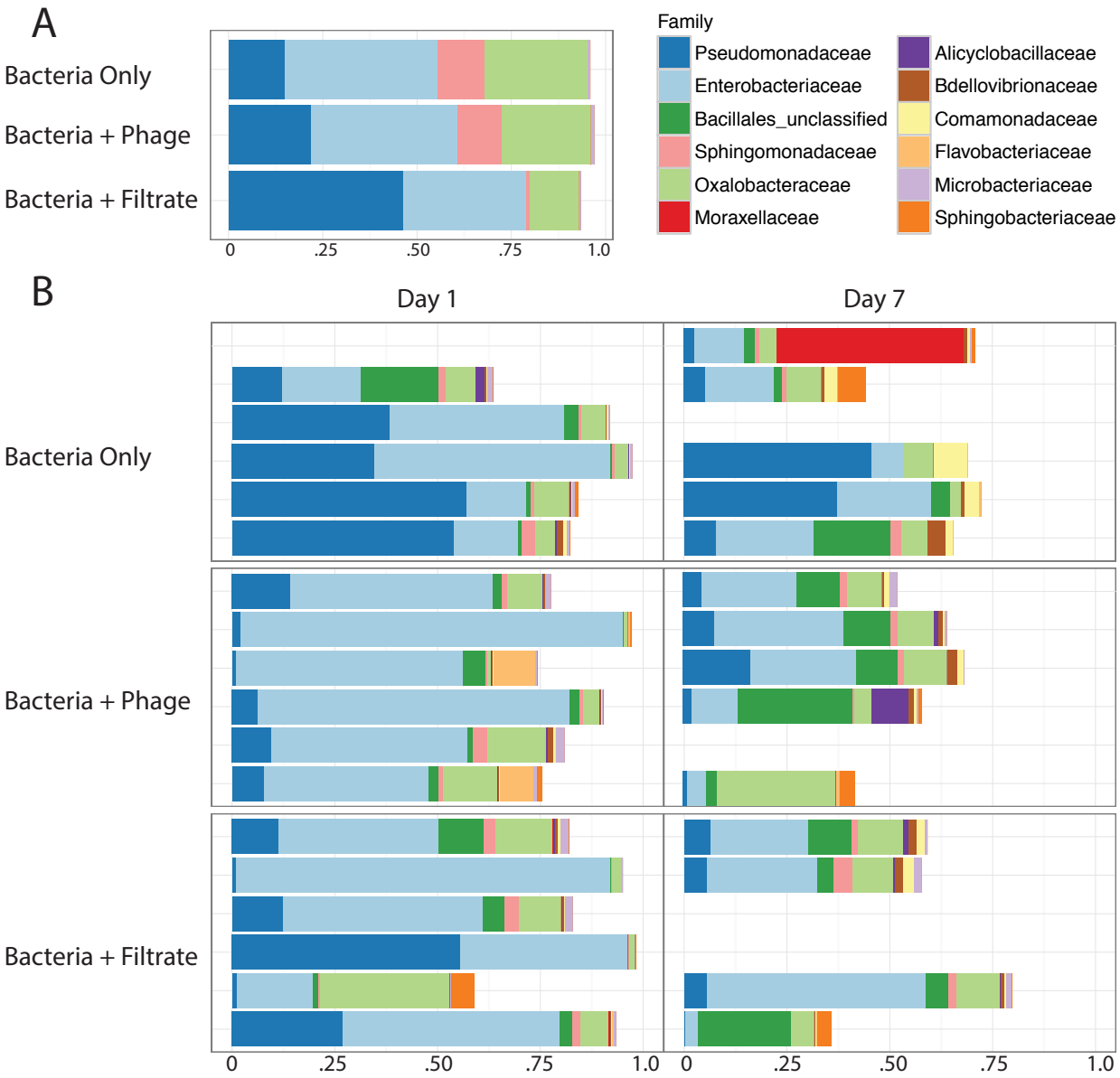


Figure 4-4 : Representation of the top bacterial families in inocula and samples Bacterial community compositions of the three different treatments were very similar in terms of relative abundance of members, as expected (a). Bar graphs show relative abundance of top 12 bacterial families at day 1 and day 7 (b), where an OTU from family Pseudomonadaceae is statistically overabundant after 1 day in the bacteria-only samples and an OTU from family Enterobacteriaceae is overabundant after 1 day in bacteria plus phage samples. Missing bars are where sequencing depth was insufficient for analysis of sample, and these samples were therefore also not included in statistical analysis. White space up to 1.0 (for illustration purposes only) represents the remaining bacterial composition of that sample that is not made up of members in the top 12 families across all samples from both days

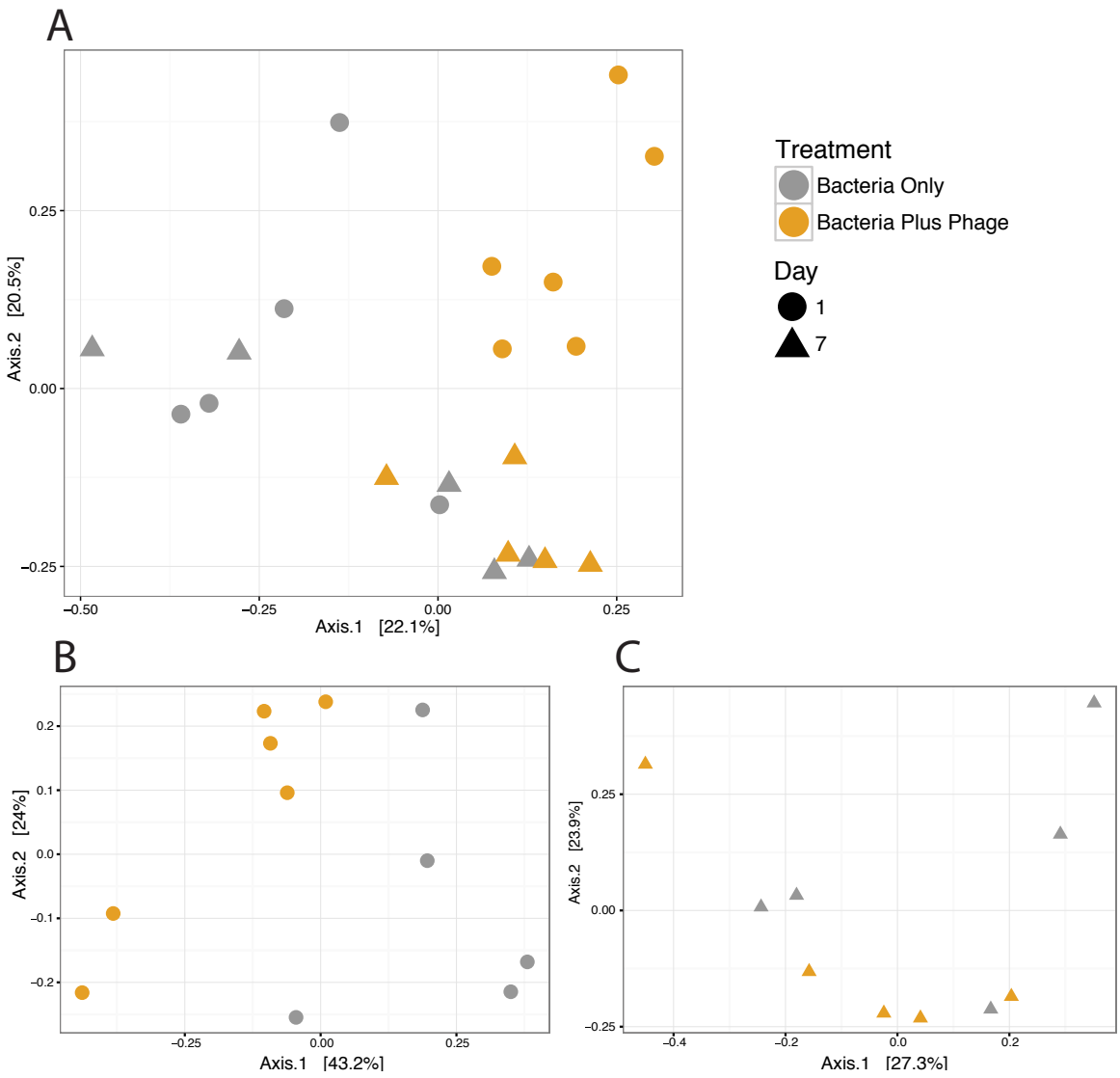


Figure 4-5 Observed differences in compositional dissimilarity among bacterial communities

Principal coordinates analysis plot of samples, based on Bray–Curtis values, where day explains 12% of variation (p = .006) and treatment explains 13% (p = .002) (a). When bacteria and bacteria plus phage samples are analysed by day and axes 1 and 2 are graphed on a principal coordinates analysis plot, treatment accounts for 31% of variation at day 1 (p = .003) (b), but treatment has no significant effect on sample dissimilarity samples at day 7 (c). Filtrate samples were excluded from visualization for simplicity

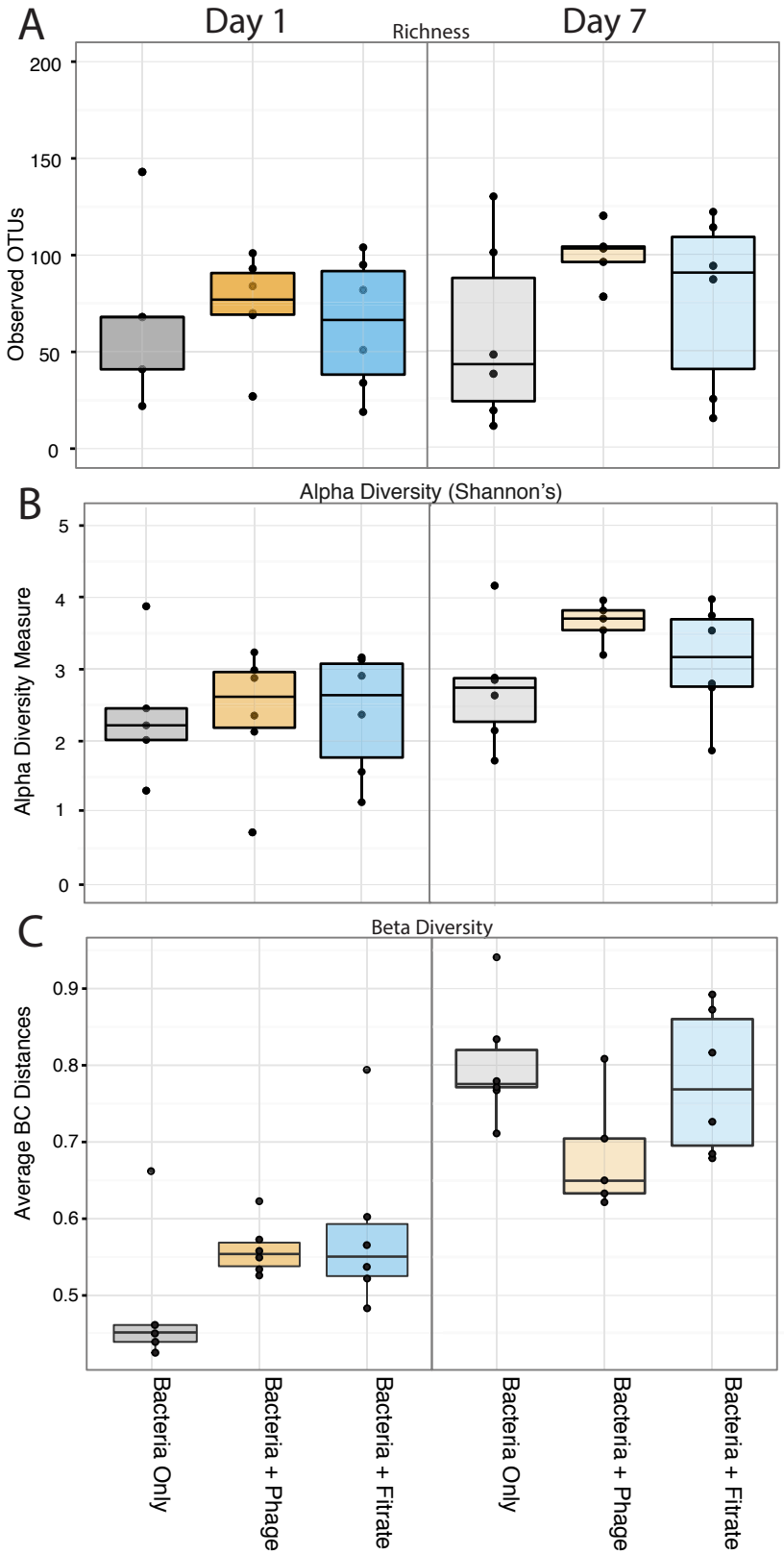


Figure 4-6 OTU richness and diversity across treatments and day

OTU richness (a) and Shannon's alpha diversity (b) were measured at day 1 and day 7 and beta diversity (c) was calculated using Bray–Curtis distances between bacterial communities of replicates within each treatment for each day. All data are displayed on boxplots where black line indicates median, the upper portion of the box is third quartile, lower portion is first quartile, whiskers are maximum and minimum, and outliers (any value more than 1.5× length of the box) are black dots. * represents p value <.05 (bacteria + filtrate treatment removed from statistical analysis)

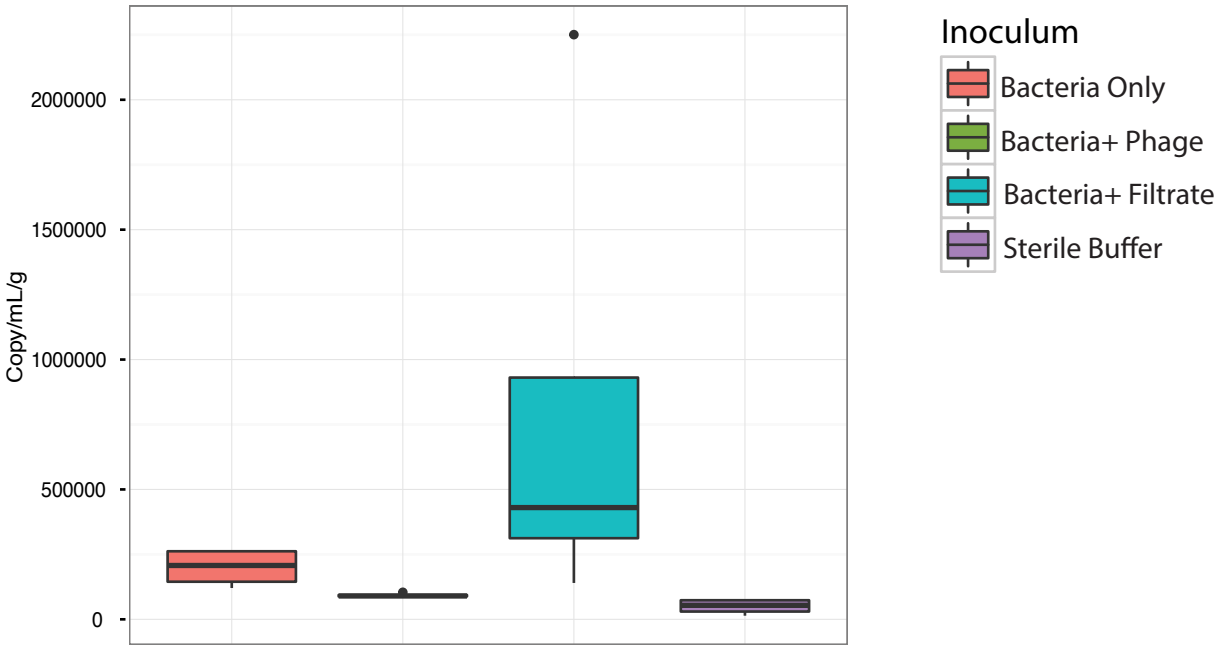


Figure 4-7 Constructed Community
Bacterial density was measured after 24 hours of plants inoculated with a community constructed of phyllosphere isolates and phage-fraction and leaf wash-filtrate from garden-grown tomatoes (n=4 per inoculum) and displayed on boxplots where black line indicates median, the upper portion of the box is third quartile, lower portion is first quartile, whiskers are maximum and minimum, and outliers (any value more than 1.5X length of the box) are black dots. *p value <0.05 (bacteria + filtrate treatment removed from statistical analysis)

Chapter 5. The impact of bacteriophages on phyllosphere bacterial communities after progressive growth on leaves

5.1 Introduction

Until recently, very little was known about the role of bacteriophages (viruses that infect bacteria) in the phyllosphere community. Work from our lab on tomato leaves (Chapter 4) shows that the phage fraction of the phyllosphere microbiome is capable of decreasing bacterial abundance within the first day of colonization of a new host, and this phage fraction also impacts bacterial composition and diversity. Other work conducted in the horse chestnut tree phyllosphere [45] demonstrated that bacteria evolve resistance to phages relatively quickly over time in a natural system. The bacterial evolution of resistance to phage predation is thought to be one mechanism by which phages can contribute to, or maintain their diversity [155]. Phage can also increase bacterial density and diversity by releasing nutrients into the environment via cell lysis, thus increasing its availability to other members of the community; [164]. One of the more frequently discussed mechanisms by which they maintain diversity in the community is through "kill-the- winner" dynamics [159, 160]. This theory suggests that the most abundant bacteria are also the most susceptible to phage predation. In such a system, an increase in abundance of a particular bacterial taxon is followed by an increase in abundance of the associated phage population and therefore a subsequent decrease in abundance of the susceptible bacterial strain, effectively preventing one type of bacterium from ever dominating the community

In the phyllosphere studies described above, only the effects of lytic phage on the bacterial communities were studied. Lytic phages are those that infect a bacterial host, and upon replication, destroy the host cell. There is also increasing evidence for the prevalence and importance of temperate phages that are integrate into the hosts' genome [168], also commonly referred to as lysogenic phages. In humans, these temperate phages are thought to dominate the phageome and outnumber lytic phages. The relative abundance of lytic to lysogenic phages may depend on factors such as host density and nutrient characteristics of the environment [187, 188]. Like lytic phages, temperate phages are also thought to increase bacterial diversity in communities through mechanisms such as facilitating horizontal gene transfer and conferring novel traits to bacteria upon lysogeny. For example, phage transduction mediates the acquisition of virulence factors in *Staphylococcus aureus* MRSA [165]. They can also lysogenize and revert to a lytic lifecycle and contribute to bacterial diversity via the killing mechanisms described above. Virtually nothing is known about the prevalence and importance of temperate phages in plant-associated microbial communities.

Here, I use a community passaging approach to disentangle the effects of lytic and lysogenic bacteriophages on epiphytic bacterial communities in the phyllosphere. Phyllosphere microbiomes were allowed to colonize plants in succession in the presence and absence of their "free" phage fraction (i.e. those lytic phages found outside of host cells), and under conditions that did or did not allow co-evolution of the bacterial and phage in these communities. Specifically, the experiment consisted of the following passage lines: 1) bacteria only, 2) both bacteria and phage, 3) bacteria passaged between plants but only in the presence of ancestral phage, and 4) ancestral bacteria introduced at each plant passage but with phage that had evolved during previous passages (Figure 5-1a). Passaging occurred weekly for three weeks on a total of three cohorts of plant hosts; the communities were passaged from one cohort of plants to the next every week for a total of three passages. At the end of three weeks, microbiomes were collected, and 16S rRNA amplicon sequencing was used to describe the bacterial communities.

Additionally, bacteria were cultured from each plant cohort at the final time point, and attempts were made to culture lytic phages. Attempts were almost made to induce lysogenic phages from the bacterial isolates.

Table 5-1: Abbreviations for each treatment

Treatments	Abbreviation
1) Passaged bacteria	B
2) Passaged bacteria and passaged phage	BP
3) Passaged bacteria; ancestral phage	BPa
4) Ancestral bacteria; passaged phage	BaP

5.2 Results

5.2.1 Bacterial community differences

By analyzing bacterial communities via 16S amplicon sequencing, it was apparent that by three weeks after initial inoculation that the treatments resulted in distinct bacterial communities. After calculating Bray-Curtis distances between all samples, a measure of bacterial community dissimilarity, samples are plotted on a PCoA plot (Figure 5-1b). The plot indicates separation amongst treatment. Furthermore, the original inoculum sample that was sequenced was dissimilar to all experimental samples, indicating that the community that was originally sprayed onto the plants changed over the course of the three weeks. Indeed, sample type (inoculum versus experimental plant-derived cells) explains 14% of the dissimilarity amongst all samples (ANOSIM, $p=0.04$). When the inoculum sample is removed and the effect of treatment is analyzed using the same ANOSIM test, 42% of variation amongst the remaining samples can be explained by treatment ($p=0.001$). Because treatment BaP seems distinct from the rest, those samples were removed and the test was re-run. In this case, treatment remains significant ($p=0.011$), but with only 26% of variation explained (Figure 5-1c).

The data can also be analyzed by distinguishing between which types of phages were passaged during the experiment (Figure 5-2a). In B only lines, only lysogenic phages would have been passaged. In the BP line, both lytic and lysogenic would have been passaged. In the BaP line (ancestral bacteria), only lytic phages would have been passaged. Lastly, in the BPa line (ancestral phage), only lysogenic phages would have been passaged: although ancestral lytic phages were applied at each passage, only those that were already in or had been incorporated into the bacterial genome were passaged onward. Therefore, lines can be classified as "lytic passage (Lyt)"; "lytic and lysogenic passaged (LytLys)"; and "lysogenic passaged (Lys)" respectively. When visualized in this manner (Figure 5-2b), samples do appear to be distinguished by that of the phages that were passaged. Statistically, 22% of the dissimilarity in bacterial community composition is explained by this variable ($p=0.008$), although this is, of course, reflective of the overall treatment effect observed in Figure 5-1. Visually, it appears as though there may be a continuum in bacterial community composition in which lytic and lysogenic only lines are the least similar, while communities associated with lytic/lysogenic show intermediate changes between these two extremes. Statistically however, this is not the case. Lyt/Lys samples are just as similar to Lyt samples as they are to Lys samples ($p>0.63$ for both comparisons) (Figure 5-2c). A similar analysis can be conducted in which the bacteria-only

lines are excluded, and only treatments in which lytic phages were passaged and/or applied are compared. In this analysis, treatment explains 39% of variation amongst samples ($p=0.001$).

5.2.2 Bacterial alpha diversity

Next, differences in both richness (number of OTUs) and alpha diversity (considering both richness and evenness) were calculated for treatments (Figure 5-3a, b). There were no significant differences in richness amongst the 4 treatments. Treatment BPa, or passaged bacteria and ancestral phage, had a lower alpha diversity than both passaged bacteria only as well as passaged bacteria and phage ($p<0.05$). The treatment in which only bacteria were passaged had higher alpha diversity than BaP ($p<0.05$).

5.2.3 Bacterial beta-diversity

The effect of beta diversity, the differences amongst individuals within a treatment, was then determined. Bray-Curtis distances were calculated between each individual sample and every other sample within the same treatment. This was done for every treatment, and the distances are graphed on a box plot and analyzed using an ANOVA and post-hoc Tukey tests (Figure 5-3c). There is a significant overall effect of treatment on beta diversity ($F_{3, 54} =10.86$, $p<0.0001$). Post-hoc tests indicate that the BPa treatment had significantly lower beta diversity than all other treatments ($p<0.007$).

5.2.3 Composition of taxa in treatment groups

The relative abundance of the top bacterial families was assessed in each treatment (Figure 5-4). There are apparent differences in bacterial family level representation between treatments. The family *Pseudomonaceae* was significantly more abundant in the BPa treatment ($p=0.003$) than in other treatments. *Enterobacteraceae*, on the other hand, was differentially abundant amongst treatments ($p<0.0001$) with the highest relative abundance in BaP and the lowest in B. This was reflective of the initial inoculum, which was dominated by *Enterobacteraceae*. In most treatments, 25% of families were not found in the top 20 families.

5.2.4 Evidence for presence of lytic and lysogenic phages

From each of the six replicates of the 4 treatments, bacteria were isolated, totaling 80 bacterial strains. These bacteria isolates were used to "fish" for lytic phages in both the resultant phage fractions obtained after passaging in each treatment and also from the phage fraction from the original inoculum. Then, for the B and BP lines, lysogeny of temperate phages was induced via UV treatment. Plaques, a zone of clearing on the bacterial lawn, were construed as evidence for the presence of phage. Overall, we found very little evidence for the presence of either lytic or lysogenic phages that attacked any of our bacterial isolates. We did find some evidence for lysogens: 20% of bacterial isolates showed evidence for the presence of lysogenic phages in the B only lines. There was evidence for lysogeny in only 7% of isolates in the BP lines where both types of phages were passaged. With regard to lytic phages in the original phage fraction, we also did not find evidence for phage predation of any bacterial isolate using this starting inoculum. The results are reported in Table 2.

Table 5-2: Bacterial Isolates and evidence for phages

	Average Bacterial Isolate / replicate	Min[a]	Max[a]	Total Bacterial Isolates	Evidence for Lytic Phages from endpoint phage fraction	Evidence for Lysogenic Phages
B	2.5	1	4	15	None	Three isolates
BP	3.4	2	6	17	None	One Isolate
BPa	4.5	3	6	27	One isolate	Three isolates
BaP	3.5	2	5	21	None	Three isolates

[a]Min and Max indicate the minimum and maximum number of bacterial isolates cultured from lines (6) of each treatment.

5.3 Discussion

Overall, disruption of co-passaging of bacteria and phage on leaves over time and between plants was found to have an impact on both the composition and diversity of the resultant epiphytic bacterial communities. The bacterial communities resulting from passaging bacteria with their ancestral phage (BPa) appeared to be the most dissimilar as compared to the other passaging treatments. This microbiome also had lower alpha diversity than that of both only bacteria passaged only (B) and bacteria passaged along with any potentially evolving phage (BP). BPa lines also have lower beta diversity than all other treatments, and *Pseudomonaceae* dominated the communities. All of this taken together suggests that the original phage present in the inoculum is capable of having the largest impact on the community, even to a bacterial community that may have changed in composition by passaging on plants. The most parsimonious explanation for this finding is that the phages present in the initial inoculum have low persistence in the phyllosphere, at least in a growth chamber, and as such, there were both more, and perhaps more diverse phage present in the initial inoculum than what remained after attrition on the leaf surface during passaging in the growth chamber. This is, however, not fully supported by our phage-isolation attempts, as we were unable to isolate phages from the original phage-fraction on any bacterial isolates. However, this may be explained by the apparent decay of phages during refrigeration as observed by others [189], and it does not preclude the possibility that there were lytic phages present at the time of the passaging experiments (which was performed 10 months prior to phage-culturing attempts).

In treatments in which lytic phages were passaged for three weeks, it is likely that the phage fraction contained very little, if any, active phage particles by the end of the experiment. This is supported by the fact that we were unable to recover any phage isolates from the final time point of the experiment, except for one isolate from plants exposed to a BPa microbiome. Poor persistence of phages in the phyllosphere is a finding supported also supported by the work of others. Phages, in general, are found in very low incidence on the surface of leaves compared to that in endophytic compartments [190]. This may be due to phage's sensitivity to UV on the surface of leaves [191], or may be due to low replication of bacteria in the phyllosphere, which is a nutrient-deplete environment that may limit bacterial growth [37], and hence lytic phage replication.

The strong effect of the ancestral phage fraction on the bacterial community is consistent, in some regards, to the findings described in Chapter 4. There, we found that after one week, plants that received bacteria and phage together (BP) had lower beta diversity than plants receiving only bacteria (B). The beta diversity results from this experiment support that finding.

Plants receiving passaged bacteria and ancestral phage (BPa) have significantly lower beta diversity than both B and BP treatments. In both cases, it may be that the impact the original phage fraction had on the bacterial community shaped it in a way that made all the microbiomes similar to one another. Interestingly, alpha diversity was also the lowest in this treatment, yet there are no differences in alpha diversity between the B and BP treatments. The alpha diversity finding may be explained by the taxonomy of the bacteria that was most impacted by the phage fraction. Here, *Pseudomonaceae* was in highest relative abundance in the BPa treatment- the treatment with the lowest alpha diversity. In previous work, the Bacteria-only treatment had a significantly higher relative abundance of *Pseudomonaceae* as well. Again, this treatment had the lowest alpha diversity. Thus, it is possible that the phage fraction has an initial impact on a dominant bacterial family in the phyllosphere, and this lethality subsequently has a ripple effect on alpha diversity of the rest of the community. Future work using rationally designed synthetic communities of bacteria and phages could address with hypothesis with more clarity.

One of the primary goals of this work was to disentangle the effects of lytic and lysogenic phages on bacterial communities in the phyllosphere. Through our experimental design, we attempted to include a variety of lytic versus lysogenic phage challenges to the bacteria. Visually, it appears that the bacterial communities in which only lysogenic phages were passaged are different from those in which lytic phages were passaged. Furthermore, 20% of bacterial isolates showed evidence for the presence of lysogenic phages in the B only lines in which only lysogenic phages were passaged. In contrast, there was evidence for lysogeny in only 7% of isolates in the BP lines where both types of phages were passaged. However, these findings are not statistically significant, and thus no conclusions can be drawn.

This work highlights the importance of time-scale when studying the effects of phages on the phyllosphere bacterial community. Predictions about alpha diversity based on results after only a one week experiment are not entirely consistent with our findings from the study of communities that were passaged between plants for three weeks. Specifically, we did not find an increase in alpha diversity in the lines in which both bacteria and phage were passaged together compared to bacteria-only lines. This may indicate that temperate phages are able to mediate long-term bacterial diversity as well as lytic phage, or it may be unrelated to the presence of phages and reflect other microbial dynamics occurring in the community. Future work that involves more rigorous identification of lysogenic phages, such as bacterial genome sequencing, may help address the question of their importance in maintaining diversity. We also found that the ancestral phage fraction has the strongest impact on both composition and diversity of the bacterial communities – probably because it was more abundant. Finally, our lack of ability to culture phages after passaging on plants suggests strongly that lytic phage particles do not persist well in the phyllosphere of plants grown in the growth chamber This supports work by others indicating that the feasibility of using phages as biocontrol agents in agriculture may be largely dependent on their ability to persist in the phyllosphere. Overall, these findings are an important extension of previous work (Chapter 4), and they underscore many of the unanswered questions that remain regarding the abundance, persistence, and importance of bacteriophages in the phyllosphere.

5.4 Materials and Methods

Experimental tomato plant germination and growth

Solanum lycopersicum cultivar Money Maker was used as a host plant in this experiment. Seeds were sourced from Park Seeds and surface sterilized as follows. Seeds were first soaked in 2.7%

bleach (sodium hypochlorite) solution for 20 minutes, then washed with sterile ddH_2O three times to remove any excess bleach. Sterile seeds were then germinated on 1% sterile water agar in the dark at 21°C for roughly seven days, or until the emergence of cotyledon leaves. At that point, they were transplanted into Sunshine Mix #1 potting soil and transferred into growth chamber at 25°C, 65% humidity and 16 h daylight per day. Seedlings were germinated once per week for three weeks in order to age-match plants for each passage of the experiment.

Plant material collection and inoculum preparation

A diverse field inoculum was generated using field-grown tomato plants. Above ground tomato plant material was collected from two fields from the UC Davis Student Organic Farm in June 2018. The material was stored on ice for transportation to the lab. One hundred grams of plant material was submerged in 1.5L of 10mM $MgCl_2$ and sonicated for 5 minutes, vortexed for 30 seconds, and sonicated again. This was repeated with an additional 500 grams of plant material, 100 grams at a time. The leaf wash from above was filtered using 8 µm filter paper to remove large pieces of plant debris. The flow-through containing all microbes was then filtered using .22µm filter units. Microbes collected on the filters, which should be most bacteria, were sonicated off the filter paper into sterile buffer for 10 minutes. To concentrate the phage fraction of the microbiome, .22 µm flow-through was then concentrated using 100Kda MWCO Millipore filter units. Both the bacterial and phage fractions were split into 8 aliquots to account for the 3 weeks of inoculation and the need for ancestral phage and ancestral bacteria for weeks 2 and 3. Bacterial fractions were stored at -80°C in 1:1 KB glycerol, and phage fractions were stored in the dark at 4°C. On each day of inoculation, bacterial aliquots were re-suspended in 3mL of $MgCl_2$ without the addition of any other fractions for the "B only" treatment. A bacterial aliquot was combined with a phage fraction aliquot for "B and P". For the first week, three treatment groups received B and P. Inoculum was spray inoculated onto each plant individually. After 1 week, entire plants were harvested individually. Bacterial and phage fractions were recovered from the plants as described previously (chapter 40 (filtering and centrifugation). The bacterial and phage fractions were re-combined and inoculated onto the plants. For evolved bacteria and ancestral phage, the bacterial fraction from the end of the passage was combined with an ancestral phage aliquot from frozen storage. For the ancestral bacteria and evolved phage treatment, the phage fraction from the end of the passage was combined with an aliquot of the ancestral bacterial treatment from frozen storage. The experiment was continued for three passages in total, each consisting of one week.

Collection and processing of endpoint microbial communities

At the end of the third week, bacterial and phage fractions were collected as above. The bacterial fraction was pelleted at 4,000 x G for 10mins. Half of the pellet was used for DNA extractions. The other half was re-suspended in 1:1 KB glycerol and stored at -80°C for future use. The entire phage fraction from each microbiome was stored at 4°C for future use.

DNA extractions and 16S amplicon library preparation

DNA was extracted from bacterial pellets using the Qiagen PowerSoil DNA Extraction Kit following manufacturer's instructions. The 16S rRNA gene was amplified using dual-indexed primers designed for the V3-V4 region [87] using the following primers: (341F (5'-CCTACGGGNBGCASCAG-3') and 785R (5'-GACTACNVGGGTATCTAATCC-3') [88]. In order to decrease amplification of plant mitochondrial and chloroplast DNA, we also used

peptide nucleic acids, PNAs [89]. PCR reactions contained 20µg BSA, PNAs at a final concentration of 0.5uM, 2 µl of template, primers at a final concentration of 20nm, and ThermoFisher Platinum™ Taq DNA Polymerase High Fidelity. Negative buffer controls, PCR controls, and original inoculum were also sequenced. Reaction conditions were: 94°C for 3 minutes, 94°C for 45 seconds, 78°C for 10 seconds, 50°C for 1 minute, 72°C for 1.5 minutes, repeat steps 2-5 30 times, and 72°C for 10 seconds. PCR reactions were randomized in order, run in triplicate for each sample, pooled, and quantified using Qubit. Amplicons from each sample were then pooled in equimolar concentrations, cleaned using an AMpure bead cleanup kit. Libraries were prepared for paired 300-nucleotide reads in Illumina's MiSeq V3 platform (Illumina) at The California Institute for Quantitative Biosciences (QB3) at UC Berkeley and run in 1 lane.

Amplicon sequencing analysis

MiSeq sequencing files were demultiplexed by the QB3 sequencing facility. Reads were combined into contigs using VSearch [92], and the remainder of the analysis was performed using Mothur version 1.41.3 [93] following their MiSeq SOP [94]. Data were quality-filtered by length, ambiguous bases, and homopolymer length using the recommended Mothur parameters. Singletons were removed using the split.abund command in Mothur after pre-clustering of similar sequences. Chimeras were removed using UChime [95]. We used a 97% similarity cut-off for defining OTUs. The Silva reference database [96] was used for sequence alignment and taxonomic assignment. Archaeal, chloroplast, mitochondrial and unknown domain DNA sequences were removed. Once an OTU table was generated in Mothur, the remainder of the analysis was performed in R using the Phyloseq package version 1.19.1 [97] and Vegan package version 2.4-5 [98] and MicrobiomeSeq (Alfred Ssekagiri, William T. Sloan, Umer Zeeshan Ijaz). Data were rarified to 5,000 reads per sample.

Culturing bacteria from endpoint samples

Bacterial fractions from the endpoint of the experiment were thawed on ice. Approximately 100ul was removed, pelleted to remove the glycerol, re-suspended in 10mM $MgCl_2$, and serially diluted. Dilutions were plated on Kings medium B hard agar and incubated at 28 °C for at least 24 hours, or until colonies were large enough to isolate. At that point, colonies having distinct morphologies (color, size, texture) were re-streaked onto KB agar plates in order to isolate individual bacterial strains. Once individual strains were isolated, liquid cultures were generated by inoculation into liquid KB and grown on an orbital shaker at 28 °C overnight.

Lytic phage isolation

In order to detect lytic phages, bacterial enrichments were generated using the potential phage source (both the original phage source, and the endpoint phage fractions from the experiment). Three mL of KB broth, 50 µL of the bacterial isolate, and 10 µL of potential phage stock was combined in a 15 mL conical tube. This culture was incubated overnight at 28 °C on an orbital shaker. Then, using a 5 mL syringe the culture was filtered through a 0.22 µm filter. The filtrate was spotted onto KB soft agar overlays, in which 8 mL of soft agar was mixed with 50 µL of bacterial isolate which was then placed on top of a base of hard agar. Ten µL of each enrichment were spotted onto the soft agar overlay and incubated at 28 °C for an additional 12 hours. After removing the petri dishes from the incubator, zones of clearing in the bacterial growth in the soft agar were noted, which are evidence for the presence of lytic phages.

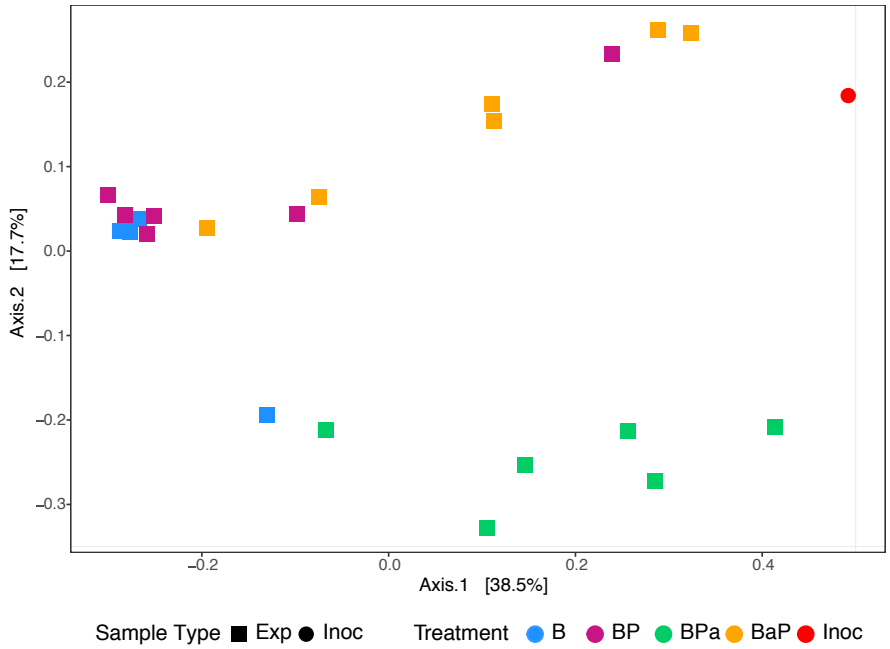
Lysogenic phage isolation

Two methods to induce the lysogenic phages from the bacterial genomes were used: UV light treatment and mitomycin C treatment. For the UV treatment, agar petri dishes with soft agar overlays as described above were used. Directly after pouring the soft agar onto plates, the petri dish was exposed under a UV light for 1 minute (intensity of 15 mW / cm^2 / sec.) These petri dishes were incubated at 28 °C for 12 hours, and then checked for plaques as signs of lysogenic phages that were induced to be lytic. For the Mitomycin C treatment, a 96-well plate approach was used to search for evidence of lysogens. Wells received 100 μL of KB broth and 10 μL of the bacterial isolate. For the Mitomycin C-treated wells, Mitomycin C was added at a final concentration of 500um/mL. OD_{600} measurements were taken every 5 minutes for 24 hours. Growth reader conditions were 28 °C with shaking every 5 minutes.

A

Treatments	Abbreviation
Passaged bacteria	B
Passaged bacteria and passaged phage	BP
Passaged bacteria; ancestral phage	BPa
Ancestral bacteria; passaged phage	BaP

B



C

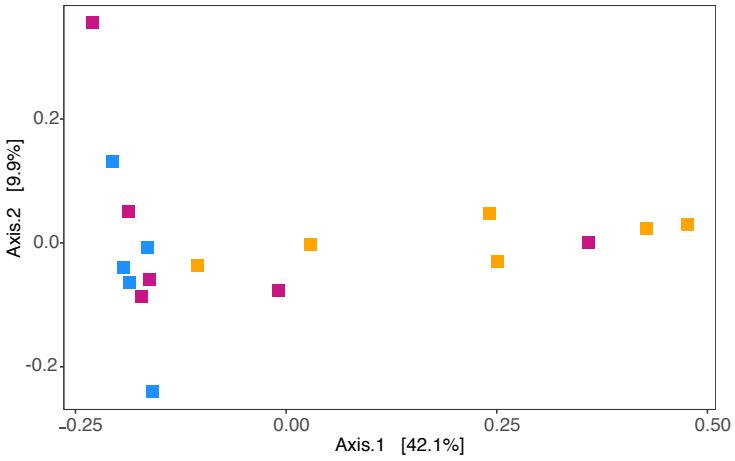
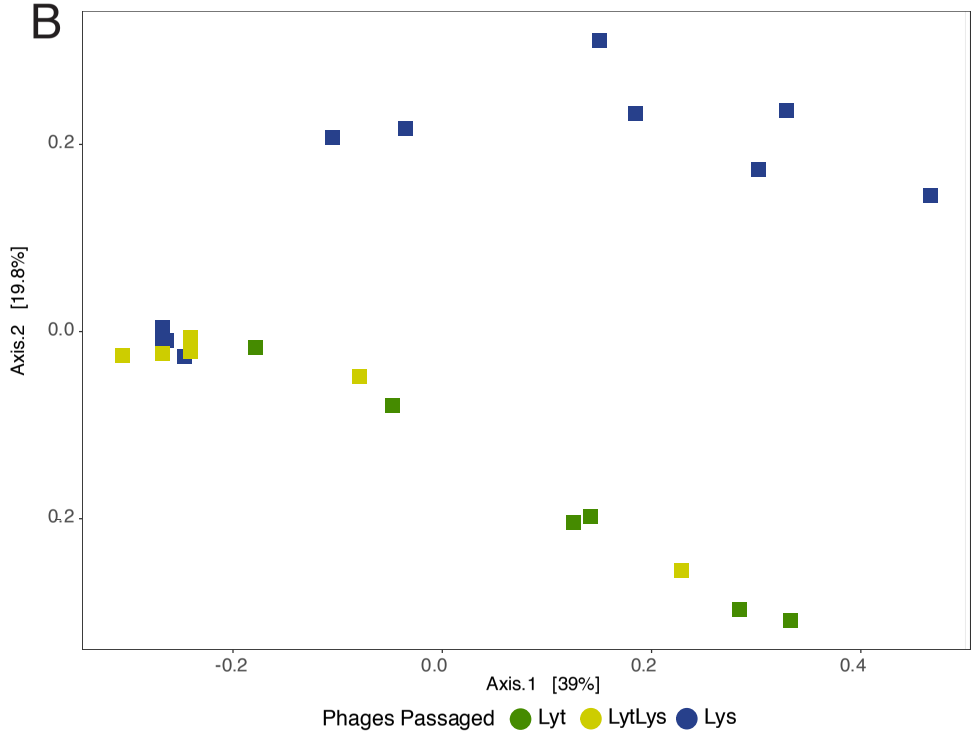


Figure 5-1 Disruption of bacteria-phage temporal dynamics
Four treatments were designed in which phyllosphere microbiomes were passaged with or without their phage fraction, and with or without co-passaging of the bacterial and phage fractions of the community (a). All samples are ordinated on a PCoA plot based on Bray-Curtis distances, and an ADONIS test indicates a significant effect of sample type and treatment (b). When the least similar group is removed (BPa), there is still a significant effect of treatment (c).

A

Treatments	Abbreviation
Lytic phages passaged	Lyt
Lytic and lysogenic phages passaged	Lyt/Lys
Lysogenic phages passaged	Lys



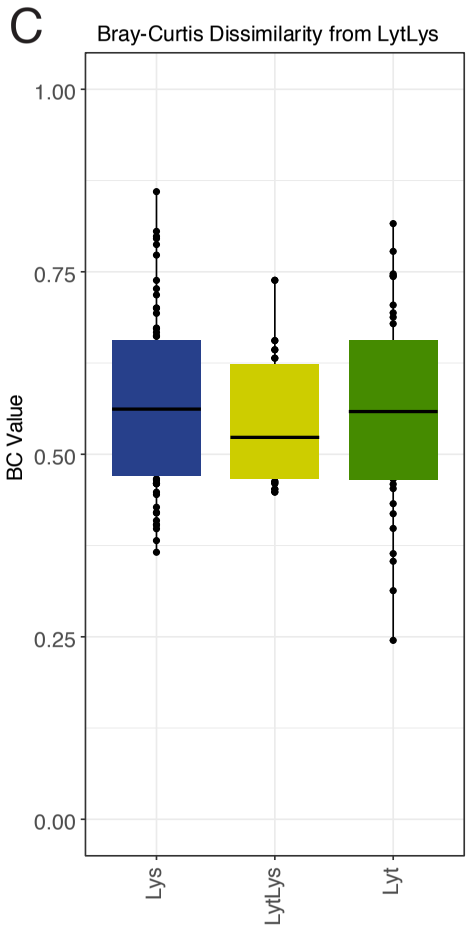


Figure 5-2 Lytic versus lysogenic phage passaging
Treatments were classified by which phage fraction was passaged for the duration of the experiment (a). All samples are ordinated on a PCoA plot based on Bray-Curtis distances, and an ADONIS test indicates a significant effect of treatment (b). Bray-Curtis distances are displayed on a box plot in order to directly compare community dissimilarity of LytLys to both Lys, Lyt, and itself.

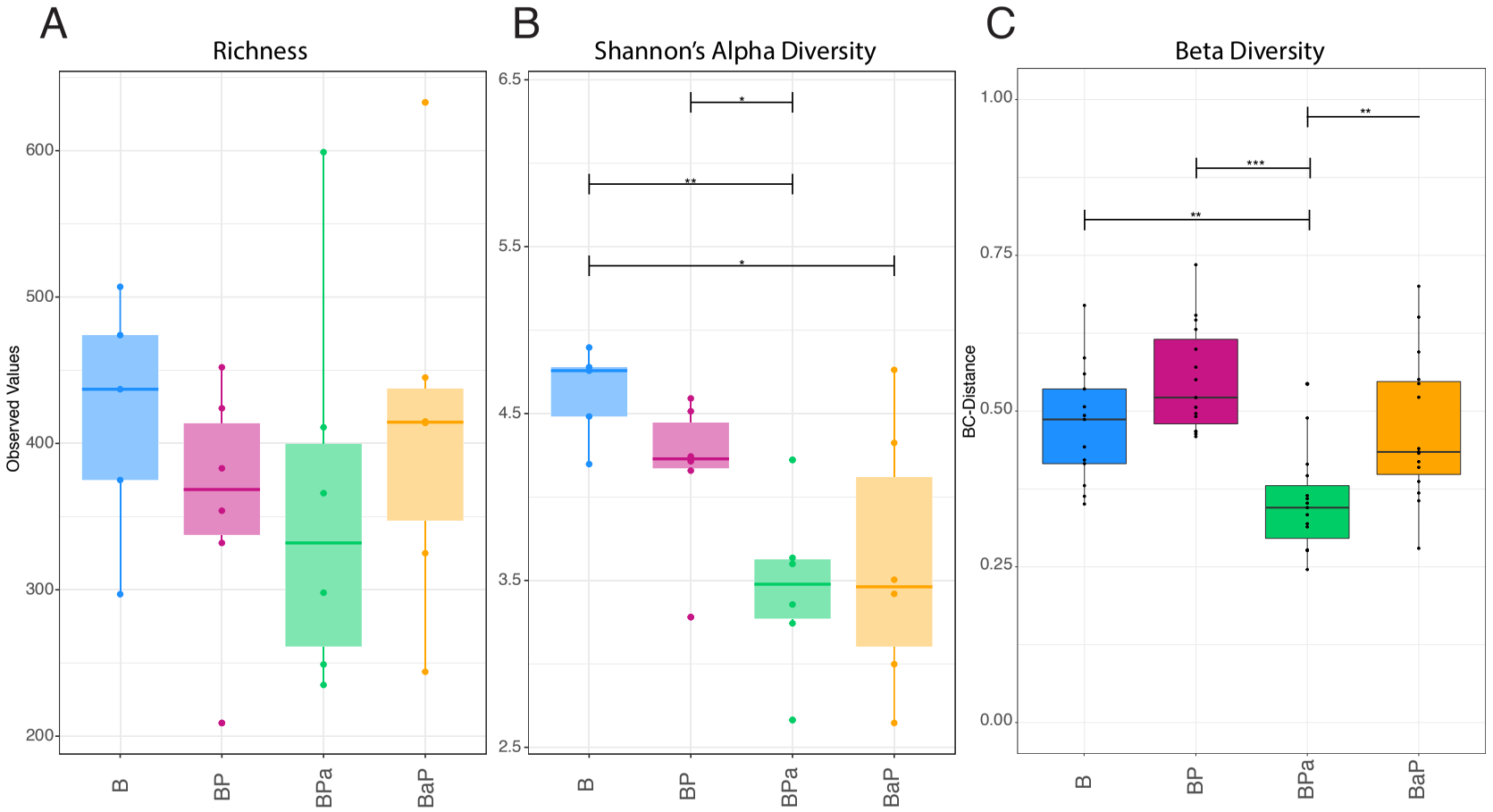


Figure 5-3 Diversity amongst treatments
Richness (a), Shannon's alpha diversity (b), and beta diversity (c) are plotted of each experimental line at each passage. Significance values of pairwise comparisons are illustrated on the graph * p≤0.05; ** p≤0.01; *** p≤0.001.

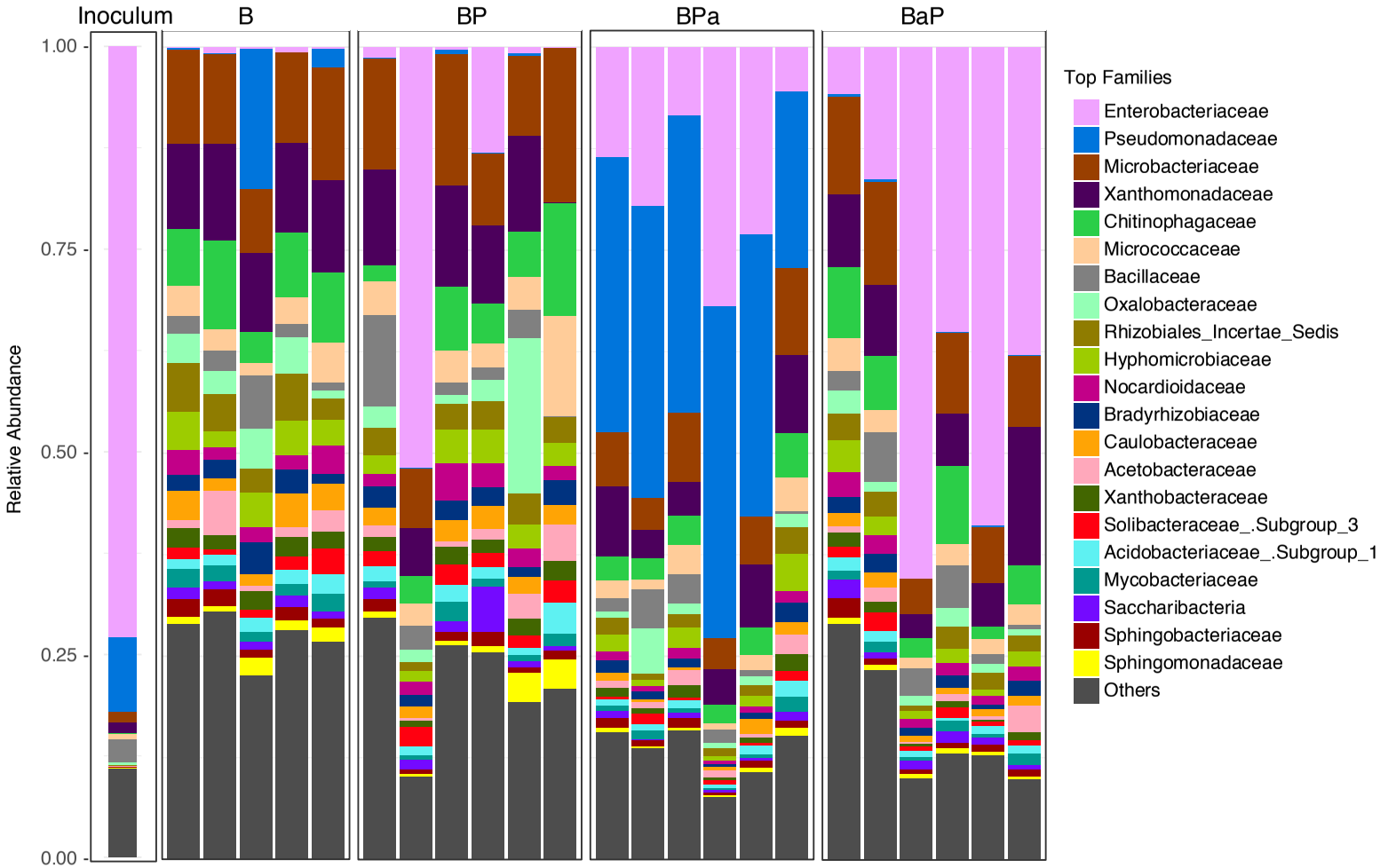


Figure 5-4 Bacterial composition of original inoculum and samples
A bar graph illustrating the relative abundance of the top families shows differences in composition amongst treatment groups.

Chapter 6 Conclusions

6.1 The importance of seed-transmitted bacteria

My thesis work began by demonstrating that vertically transmitted bacteria on the surface of tomato seeds are capable of protecting seedlings against a common bacterial pathogen. Vertical transmission is a well-studied process in other systems such as termites and aphids [192]. However in plants, it has been primarily limited to the transmission of pathogens and endophytes. Both are important fields of study- especially for the prevention of plant diseases. However, I felt that there was a significant lack of knowledge as to the functional importance of vertically transmitted commensals or mutualists in plants. Intuitively, vertical transmission of a microbiome or symbiont would allow for maintenance of key members of the microbial community across generations. Beneficial microbes would have primary access to both spatial niches and nutrients provided by seedlings. Interestingly, plants have a differential onset of resistance to pathogens throughout their life-stages, something described as age-related resistance (ARR) or developmental resistance [193–195]. However, much of the work on ARR investigates exposure and resistance to specific pathogens throughout the developmental stage of the plant and does not address if there is a crucial window of exposure to commensals, and whether these commensals ae contributing to ARR, as observed in other systems. To my knowledge, there are no studies to date that test the importance of timing of commensal microbial exposure on microbiome establishment or immune function in plants, although there is a wealth of literature on establishing biological control agents [54, 75, 196, 197]. For example, would a seedling exposed to beneficial microbes mount as strong of a response as an older plant? And would exposure of otherwise sterile adults result in the same successional dynamics of microbiome establishment as often observed in seedlings? Given that we know that resistance to pathogens can change throughout the life cycle of a plant, research focused on age-related tolerance and recruitment of commensals and plant-growth promoting bacteria has large implications in seed treatment and agricultural practices. Future work on this topic will address if this critical window of exposure exists for plants.

One fundamental question that I was unable to answer is: are the types of bacteria transmitted on seeds merely a reflection of the parental plant from which the progeny was generated? If so, do transmitted microbes vary based on adult-plant microbiomes? If so, which anatomical portion of the plant is most influential in shaping the seed microbiome? In this work- we found that some seed microbiomes were more protective than others. This suggests that they may have differed in their composition, but we never confirmed that this was the case. Furthermore in my work on adult plants, I found that plant host genotype influences the phyllosphere microbiome. Are these differences between microbiomes heritable through vertical transmission? In order to address these outstanding questions, I would conduct a common garden experiment wherein multiple host genotypes would be planted in replicated field sites. Upon fruiting, I would not only collect tomato fruits, but I would also collect leaf, flower, and soil from each site. I would collect seeds from fruits using the same approach as described in Chapter 2. I would carry out *Gyrase B* amplicon sequencing (which provides finer taxonomic resolution than 16S) in order to describe the bacterial communities of the seeds in addition to the adult plants from which they came. The results of this proposed work would help us to determine if 1) there is variability in vertically transmitted seed microbiomes across location and host genotype and 2) what these differences arise from. By fully describing the bacterial community of the leaves and fruits as well, we could determine seed colonists' origin and address if transmission of

leaf epiphytes is possible via the seeds. A similar culture-based approached could be used to determine the protective ability of vertically transmitted microbes, as it is possible that taxonomic identity may differ- but function may be the same.

In conclusion, the work described in Chapter 2 makes an important step in filling in the gap in the literature as to the existence and importance of vertically transmitted symbionts in this system. Future work will help us understand the origin of these symbionts and if genotype-specific microbes are physically heritable through seed transmission.

6.2 Community passaging is a valuable tool in microbiome research

Breeding for agriculturally beneficial traits, such as disease resistance, is an area of ongoing research in the tomato industry, and agriculture in general. Many resistance genes in modern tomatoes originate from their wild Peruvian relative, *S. pimpinellifolium*, which has a much larger genetic diversity than modern cultivars. Whether or not there are differences in tomato microbiota due to domestication and the presence/absence of resistance genes is a relatively unexplored topic. My work on adult phyllosphere microbiomes originally began as a way to test the hypothesis that breeding for disease resistance genes in tomatoes would ultimately impact the microbiome of those plants. I was, indeed, able to show that host genotype shaped the phyllosphere community for the first two passages of the experiment. When genotypes are classified "resistant" or "susceptible" (e.g. three genotypes have specific resistance loci and two do not), there is no overall effect of resistance in microbiomes that had been fully selected on various tomato varieties. This lack of an overall effect of resistance leads me to conclude that it may not be the resistance genes themselves, but rather other genetic differences that existed between the genotypes that drove genotype specificity of the bacterial communities. Although I chose two pairs of near-isogenic lines and one out-group, it is likely that other genetic differences existed between near isogenic lines due to the introgression breeding technique that was used to generate the lines. Furthermore, the pairs of near isogenic lines are different tomato cultivars, further contributing to genetic differences amongst hosts. Future work by others will further test the degree to which differences in host genetics impacts the phyllosphere community, and indeed, some evidence for heritable taxa already has been produced using genome wide association studies in corn and *Arabidopsis* [198, 199]. Moving forward, if there are taxonomic differences between microbiomes, whole genome metagenomics sequencing will help us determine if the functional capability of the microbiomes differed as well.

One of the puzzling results from the microbiome passaging work that we were not able to fully explain is that host genotype had a significant impact on bacterial community at the beginning of the experiment, but this declined over time. Through identifying the specific taxa that were significantly associated with the five genotypes in P1 and P2 it's clear that it is not only the rare taxa associated with particular genotypes that drove such a genotype effect on the microbiomes, and thus the decline in genotype effect cannot be fully explained simply by an overall decrease in diversity. It seems likely that the microbiome underwent environmental selection driven by three factors: 1) the greenhouse, 2) the tomato phyllosphere, and 3) specific tomato genotypes. It seems reasonable that the relative strength of each selection pressure would change over time, whereby host genotype is important early on, but the community experiences progressively more time in the tomato phyllosphere in the greenhouse, the pressure of those environments overshadows a genotype effect. Even with well-designed and well-controlled experiments, it is difficult to disentangle the selection pressures at play.

The work in Chapter 3 also highlights the advantages of the phyllosphere as a system in microbiome research. Through our microbiome transplant and passaging technique that is biologically relevant to how the phyllosphere is naturally colonized, we were not only able to select upon entire host-associated microbial communities, but we could also experimentally test hypotheses regarding microbiome adaption in subsequent experiments. Again, this is due to the physical accessibility of the phyllosphere community and the ease at which it can be inoculated onto hosts.

These findings also shed light on a notable challenge in microbiome research. Our data suggest that when describing the microbiome of an open environment, such as plant surfaces, many of the taxa found there may be transient visitors. In the case of the phyllosphere, there are microbes on leaf surfaces that may have emigrated from air, soil, surrounding plants, or other non-plant habitats and do not necessarily represent an adapted community that is capable of growth and persistence. Passaging of microbiomes on a particular host seems to be a powerful way of differentiating those taxa that are, or can become, well adapted to a plant host environment and those that were present upon sampling, but are not well adapted to the environment. Across all systems, much of the work in microbial ecology is highly descriptive: the community associated with a particular host or ecosystem at a given time is described to be its microbiome, implying strong selection for a particular interactive community- rather than a context-dependent assemblage with many recent immigrants, for example. Our findings raise the question as to if a microbiome should be defined as the community that is merely found there upon sampling, or alternatively, if a true microbiome is only one that is adapted to its host or environment. The latter definition might prove hard to establish in many habitats, but fortunately can be readily addressed in the phyllosphere. Thus we expect that our phyllosphere studies will provide important conceptual contributions to the field as a whole.

6.3 Bacteriophages are a challenging but crucial component of the microbiome

The final chapters of my dissertation explore the importance of bacteriophages (phages) in the phyllosphere community. There are many challenges facing phage research, and studying environmental phages is an especially difficult field due to lack of a universal marker gene for sequencing, a lack of cultivability, and our nascent understanding of phage genetic diversity. Thus, we are unable to describe the abundance and diversity of phages within our samples without shotgun metagenome sequencing. Some of these challenges I was able to overcome, and others I was not. My work primarily depended on the assumption that there were phages on the leaves used to generate the initial inoculum. I took a "black box approach" in which I isolated the size fraction of the microbiome that should contain most lytic phage particles. I treated this as the "phage fraction". I then looked for an effect of this fraction on bacterial abundance, composition, and diversity. This approach allowed me to overcome the difficulty of identifying and quantifying phages. My findings show that there is, indeed, an important effect of the phage fraction on the microbiome as a whole. This work also provides empirical support for the theory that phages mediate prokaryotic diversity and contribute to temporal population size dynamics.

In order to measure phage abundance in starting samples, I attempted to use both fluorescence microscopy and transmission electron microscopy. Although both approaches yielded images of "phage-like particles", it was impossible to quantify these particles, primarily due to the amount of background fluorescence that interfered with microscopy and the sheer difficulty of identifying low-abundance phages using electron microscopy. Another alternative for quantifying phage particles is flow cytometry, but this method suffers from the same

limitations due to the presence of intrinsically fluorescent contaminating particles. In Appendix 1, I describe a method that I was able to develop for quantification of known phages. If I were to continue the work described in Chapter 4, I would do so with a defined synthetic community of bacteria and phages, from which I could sensitively measure both bacterial and phage abundance throughout the course of the experiment.

There remains much to learn about lytic and temperate phages. My findings in Chapter 5 attempt to disentangle the effect of lytic versus temperate phages on the bacterial community on leaves. This work is an important extension from Chapter 4 in a number of ways. First, I wanted to test if the patterns that I observed of the effect of lytic phages on the bacterial community after only a short time together on leaves (one week) were persistent over time. I found that patterns of the effect of phages on bacterial communities differed when examined after 3 weeks compared to 1 week. This has important implications in how we think about the issue of timescale in microbial community interactions. It also begs the question: how many lytic phages were present at the start of the experiments, and how long did they persist on the surface of the plants? If lytic phages do not persist, can selection for lysogenic phages produce some of the same patterns in the bacterial community as well? Interestingly, I did observe that treatment consisting only of bacteria that were passaged on leaves for several weeks, i.e. the treatment in which lysogenic phages would have been selected, had a qualitatively higher alpha bacterial diversity than the other treatments. This may suggest that lysogenic phages are capable of promoting bacterial diversity over longer time scales (for example, through facilitating horizontal gene transfer upon transduction or increasing genetic diversity through conferring resistance against other phages) We were not, however, able to find conclusive evidence for an increased presence of lysogenic phage in the communities of bacteria passaged on plants in the absence of lytic phage.

The questions that I attempted to address in Chapters 4 and 5 are difficult to answer with current, common bacteriophage techniques. Moving forward, the best way to uncover both lytic and temperate phage abundance and diversity in the phyllosphere is likely through shotgun metagenomic sequencing. Sequencing the phage fraction (viral metagenomics) would better describe the lytic phages in the system. Sequencing the bacterial fraction should reveal the prevalence of temperate phages integrated into the bacterial genomes. This approach is not devoid of challenges, however, as it can be difficult to sequence microbiomes associated with plants because of the presence of abundant contaminating plant genetic material. Improvements in high throughput sequencing are allowing us to overcome this limitation, if only by the sheer amount of sequences that can be obtained in environmental samples. I see this approach as the most promising way to comprehensively understand the abundance, diversity, and importance of bacteriophages in the phyllosphere.

6.4 Current limitations in microbiome research

That the microbiome is an entity that fundamentally influences host health and function has caught the attention of researchers, medical doctors, nutritionists, and every other person interested in the microbial world that exists around and within them. Next generation sequencing and other “omics” approaches have enabled us to address the diversity and complexity of various microbial communities, but there are limitations to these approaches. For example, many labs frequently use 16S rRNA amplicon sequencing to describe bacterial communities. This is the most accessible approach due to reasonable cost, accessibility to protocols, and ease of use of sequence analysis pipelines. However, 16S amplicon sequencing only gives us coarse taxonomic resolution of the community, and it does not give any idea of function (other than predicted

function). The movement of the field away from the division of bacterial diversity within Operational Taxon Units (OTUs) to Exact Sequence Variants (ASV) should provide more resolution of bacterial diversity, but still does not provide much insight into the functions of these taxa. It seems likely that as the cost of sequencing continues to fall and more labs have access to both the sequencing and analysis technology required, shotgun metagenomics will surpass 16S amplicon sequencing in popularity. Even so, these advanced sequencing approaches must be coupled with hypothesis-driven experiments and highly controlled experimental design. This, in addition to culture-based approaches and the use of synthetic communities when possible, will enable us to move the field beyond observational and correlational findings.

6.5 Concluding remarks

I have studied the importance of vertical transmission of seed epiphytes, host genotype, environmental selection, and bacteriophages in shaping the phyllosphere community of *Solanum lycopersicum*. The results from this work are directly applicable to agricultural processes such as seed treatment, plant probiotic design, and breeding practices. They also suggest that both evolutionary and ecological forces have shaped, and continue to shape, a longstanding partnership between microbiomes and plants. To truly understand the microbiome is a monumental task. It will not be achieved in one or two labs or in a single system. Instead, the findings from multiple systems from researchers seeking to answer a multitude of questions may one day converge to reveal universal patterns that exist. This will not only enable transformative capabilities in both medicine and agriculture, but it will also illuminate the remarkable complexity underlying host-microbiome symbioses.

References

1. Doebley JF, Gaut BS, Smith BD (2006) The Molecular Genetics of Crop Domestication. Cell 127:1309–1321. https://doi.org/10.1016/j.cell.2006.12.006

2. Jenkins JA (1948) The origin of the cultivated tomato. Econ Bot 2:379–392. https://doi.org/10.1007/BF02859492

3. Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. Theor Appl Genet 80:437–448. https://doi.org/10.1007/BF00226743

4. Flajnik MF, Kasahara M (2010) Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet 11:47–59. https://doi.org/10.1038/nrg2703

5. Buchmann K (2014) Evolution of Innate Immunity: Clues from Invertebrates via Fish to Mammals. Front Immunol 5:. https://doi.org/10.3389/fimmu.2014.00459

6. Charles A Janeway J, Travers P, Walport M, Shlomchik MJ (2001) Evolution of the innate immune system

7. Medzhitov R, Janeway CJr (2000) Innate Immunity. N Engl J Med 343:338–344. https://doi.org/10.1056/NEJM200008033430506

8. Jones JDG, Dangl JL (2006) The plant immune system. Nature 444:323–329. https://doi.org/10.1038/nature05286

9. Underwood W (2012) The Plant Cell Wall: A Dynamic Barrier Against Pathogen Invasion. Front Plant Sci 3:. https://doi.org/10.3389/fpls.2012.00085

10. Mansfield JW (2000) Antimicrobial Compounds and Resistance. In: Slusarenko AJ, Fraser RSS, Loon LC van (eds) Mechanisms of Resistance to Plant Diseases. Springer Netherlands, pp 325–370

11. Jung HW, Tschaplinski TJ, Wang L, et al (2009) Priming in Systemic Plant Immunity. Science 324:89–91. https://doi.org/10.1126/science.1170025

12. Conrath U (2011) Molecular aspects of defence priming. Trends Plant Sci 16:524–531. https://doi.org/10.1016/j.tplants.2011.06.004

13. Conrath U, Beckers GJM, Flors V, et al (2006) Priming: Getting ready for battle. Mol Plant Microbe Interact 19:1062–1071. https://doi.org/10.1094/MPMI-19-1062

14. Ahmad S, Gordon-Weeks R, Pickett J, Ton J (2010) Natural variation in priming of basal resistance: From evolutionary origin to agricultural exploitation. Mol Plant Pathol 11:817–827. https://doi.org/10.1111/j.1364-3703.2010.00645.x

15. Hooper LV, Littman DR, Macpherson AJ (2012) Interactions Between the Microbiota and the Immune System. Science 336:1268–1273. https://doi.org/10.1126/science.1223490

16. Zipfel C, Oldroyd GED (2017) Plant signalling in symbiosis and immunity. Nature 543:328–336. https://doi.org/10.1038/nature22009

17. Lebeis SL, Paredes SH, Lundberg DS, et al (2015) Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. Science 349:860–864. https://doi.org/10.1126/science.aaa8764

18. Bodenhausen N, Bortfeld-Miller M, Ackermann M, Vorholt JA (2014) A Synthetic Community Approach Reveals Plant Genotypes Affecting the Phyllosphere Microbiota. PLOS Genet 10:e1004283. https://doi.org/10.1371/journal.pgen.1004283

19. Liu H, Carvalhais LC, Schenk PM, Dennis PG (2017) Effects of jasmonic acid signalling on the wheat microbiome differ between body sites. Sci Rep 7:. https://doi.org/10.1038/srep41766

20. Honda K, Littman DR (2016) The microbiota in adaptive immune homeostasis and disease. Nature 535:75–84. https://doi.org/10.1038/nature18848

21. Pastor V, Luna E, Mauch-Mani B, et al (2013) Primed plants do not forget. Environ Exp Bot 94:46–56. https://doi.org/10.1016/j.envexpbot.2012.02.013

22. Cerf-Bensussan N, Gaboriau-Routhiau V (2010) The immune system and the gut microbiota: friends or foes? Nat Rev Immunol 10:735–744. https://doi.org/10.1038/nri2850

23. Luna E, Bruce TJA, Roberts MR, et al (2012) Next-generation systemic acquired resistance. Plant Physiol 158:844–853. https://doi.org/10.1104/pp.111.187468

24. Rasmann S, De Vos M, Casteel CL, et al (2012) Herbivory in the previous generation primes plants for enhanced insect resistance. Plant Physiol 158:854–863. https://doi.org/10.1104/pp.111.187831

25. Slaughter A, Daniel X, Flors V, et al (2012) Descendants of Primed Arabidopsis Plants Exhibit Resistance to Biotic Stress. Plant Physiol 158:835–843. https://doi.org/10.1104/pp.111.191593

26. Clay K (2014) Defensive symbiosis: a microbial perspective. Funct Ecol 28:293–298. https://doi.org/10.1111/1365-2435.12258

27. Hussa EA, Goodrich-Blair H (2013) It takes a village: ecological and fitness impacts of multipartite mutualism. Annu Rev Microbiol 67:161–178. https://doi.org/10.1146/annurev-micro-092412-155723

28. Berg M, Koskella B (2018) Nutrient-and Dose-Dependent Microbiome-Mediated Protection against a Plant Pathogen. Curr Biol

29. Innerebner G, Knief C, Vorholt JA (2011) Protection of Arabidopsis thaliana against leaf-pathogenic Pseudomonas syringae by Sphingomonas strains in a controlled model system. Appl Environ Microbiol 77:3202–3210. https://doi.org/10.1128/AEM.00133-11

30. Christian N, Herre EA, Mejia LC, Clay K (2017) Exposure to the leaf litter microbiome of healthy adults protects seedlings from pathogen damage. Proc R Soc B 284:20170641. https://doi.org/10.1098/rspb.2017.0641

31. Kaplan JL, Shi HN, Walker WA (2011) The Role of Microbes in Developmental Immunologic Programming. Pediatr Res 69:465–472. https://doi.org/10.1203/PDR.0b013e318217638a

32. Kelly D, King T, Aminov R (2007) Importance of microbial colonization of the gut in early life to the development of immunity. Mutat Res Mol Mech Mutagen 622:58–69. https://doi.org/10.1016/j.mrfmmm.2007.03.011

33. Shade A, Jacques M-A, Barret M (2017) Ecological patterns of seed microbiome diversity, transmission, and assembly. Curr Opin Microbiol 37:15–22. https://doi.org/10.1016/j.mib.2017.03.010

34. Barret M, Briand M, Bonneau S, et al (2015) Emergence Shapes the Structure of the Seed Microbiota. Appl Environ Microbiol 81:1257–1266. https://doi.org/10.1128/AEM.03722-14

35. Truyens S, Weyens N, Cuypers A, Vangronsveld J (2015) Bacterial seed endophytes: genera, vertical transmission and interaction with plants. Environ Microbiol Rep 7:40–50. https://doi.org/10.1111/1758-2229.12181

36. MORRIS CE (2002) Fifty years of phyllosphere microbiology : significant contributions to research in related fields. Phyllosphere Microbiol

37. Lindow SE, Brandl MT (2003) Microbiology of the Phyllosphere. Appl Environ Microbiol 69:1875–1883. https://doi.org/10.1128/AEM.69.4.1875-1883.2003

38. Bulgarelli D, Schlaeppi K, Spaepen S, et al (2013) Structure and Functions of the Bacterial Microbiota of Plants. Annu Rev Plant Biol 64:807–838. https://doi.org/10.1146/annurev-arplant-050312-120106

39. Ottesen AR, González Peña A, White JR, et al (2013) Baseline survey of the anatomical microbial ecology of an important food plant: Solanum lycopersicum (tomato). BMC Microbiol 13:114. https://doi.org/10.1186/1471-2180-13-114

40. Lymperopoulou DS, Adams RI, Lindow SE (2016) Contribution of vegetation to the microbial composition of nearby outdoor air. Appl Environ Microbiol AEM.00610-16. https://doi.org/10.1128/AEM.00610-16

41. Cordier T, Robin C, Capdevielle X, et al (2012) The composition of phyllosphere fungal assemblages of European beech (Fagus sylvatica) varies significantly along an elevation gradient. New Phytol 196:510–519. https://doi.org/10.1111/j.1469-8137.2012.04284.x

42. Coleman-Derr D, Desgarennes D, Fonseca-Garcia C, et al (2016) Plant compartment and biogeography affect microbiome composition in cultivated and native Agave species. New Phytol 209:798–811. https://doi.org/10.1111/nph.13697

43. Koskella B, Thompson JN, Preston GM, Buckling A (2011) Local biotic environment shapes the spatial scale of bacteriophage adaptation to bacteria. Am Nat 177:440–451. https://doi.org/10.1086/658991

44. Koskella B (2013) Phage-Mediated Selection on Microbiota of a Long-Lived Host. Curr Biol 23:1256–1260. https://doi.org/10.1016/j.cub.2013.05.038

45. Koskella B, Parr N (2015) The evolution of bacterial resistance against bacteriophages in the horse chestnut phyllosphere is general across both space and time. Philos Trans R Soc Lond B Biol Sci 370:. https://doi.org/10.1098/rstb.2014.0297

46. Egamberdieva D, Wirth SJ, Alqarawi AA, et al (2017) Phytohormones and Beneficial Microbes: Essential Components for Plants to Balance Stress and Fitness. Front Microbiol 8:. https://doi.org/10.3389/fmicb.2017.02104

47. Fürnkranz M, Wanek W, Richter A, et al (2008) Nitrogen fixation by phyllosphere bacteria associated with higher plants and their colonizing epiphytes of a tropical lowland rainforest of Costa Rica. ISME J 2:561–570. https://doi.org/10.1038/ismej.2008.14

48. Zahran HH (1999) Rhizobium-Legume Symbiosis and Nitrogen Fixation under Severe Conditions and in an Arid Climate. Microbiol Mol Biol Rev 63:968–989

49. Lau JA, Lennon JT (2012) Rapid responses of soil microorganisms improve plant fitness in novel environments. Proc Natl Acad Sci 109:14058–14062. https://doi.org/10.1073/pnas.1202319109

50. Badri DV, Zolla G, Bakker MG, et al (2013) Potential impact of soil microbiomes on the leaf metabolome and on herbivore feeding behavior. New Phytol 198:264–273. https://doi.org/10.1111/nph.12124

51. Wagner MR, Lundberg DS, Coleman-Derr D, et al (2014) Natural soil microbes alter flowering phenology and the intensity of selection on flowering time in a wild Arabidopsis relative. Ecol Lett 17:717–726. https://doi.org/10.1111/ele.12276

52. Newman M-A, Dow JM, Molinaro A, Parrilli M (2007) Priming, induction and modulation of plant defence responses by bacterial lipopolysaccharides. J Endotoxin Res 13:69–84. https://doi.org/10.1177/0968051907079399

53. Zamioudis C, Korteland J, Van Pelt JA, et al (2015) Rhizobacterial volatiles and photosynthesis-related signals coordinate MYB72 expression in Arabidopsis roots during onset of induced systemic resistance and iron-deficiency responses. Plant J 84:309–322. https://doi.org/10.1111/tpj.12995

54. Pusey PL, Stockwell VO, Reardon CL, et al (2011) Antibiosis activity of Pantoea agglomerans biocontrol strain E325 against Erwinia amylovora on apple flower stigmas. Phytopathology 101:1234–1241. https://doi.org/10.1094/PHYTO-09-10-0253

55. Santhanam R, Luu VT, Weinhold A, et al (2015) Native root-associated bacteria rescue a plant from a sudden-wilt disease that emerged during continuous cropping. Proc Natl Acad Sci 112:E5013–E5020. https://doi.org/10.1073/pnas.1505765112

56. Ritpitakphong U, Falquet L, Vimoltust A, et al (2016) The microbiome of the leaf surface of Arabidopsis protects against a fungal pathogen. New Phytol 210:1033–1043. https://doi.org/10.1111/nph.13808

57. Vannier N, Mony C, Bittebiere A-K, et al (2018) A microorganisms' journey between plant generations. Microbiome 6:. https://doi.org/10.1186/s40168-018-0459-7

58. Rezki S, Campion C, Simoneau P, et al (2018) Assembly of seed-associated microbial communities within and across successive plant generations. Plant Soil 422:67–79. https://doi.org/10.1007/s11104-017-3451-2

59. Shahzad R, Khan AL, Bilal S, et al (2018) What Is There in Seeds? Vertically Transmitted Endophytic Resources for Sustainable Improvement in Plant Growth. Front Plant Sci 9:. https://doi.org/10.3389/fpls.2018.00024

60. Douglas AE (1998) Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria Buchnera. Annu Rev Entomol 43:17–37. https://doi.org/10.1146/annurev.ento.43.1.17

61. Bergna A, Cernava T, Rändler M, et al (2018) Tomato seeds preferably transmit plant beneficial endophytes. Phytobiomes J. https://doi.org/10.1094/PBIOMES-06-18-0029-R

62. Links MG, Demeke T, Gräfenhan T, et al (2014) Simultaneous profiling of seed-associated bacteria and fungi reveals antagonistic interactions between microorganisms within a shared epiphytic microbiome on Triticum and Brassica seeds. New Phytol 202:542–553. https://doi.org/10.1111/nph.12693

63. Walterson AM, Stavrinides J (2015) Pantoea: insights into a highly versatile and diverse genus within the Enterobacteriaceae. FEMS Microbiol Rev 39:968–984. https://doi.org/10.1093/femsre/fuv027

64. Brady C, Cleenwerck I, Venter S, et al (2008) Phylogeny and identification of Pantoea species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). Syst Appl Microbiol 31:447–460. https://doi.org/10.1016/j.syapm.2008.09.004

65. Delétoile A, Decré D, Courant S, et al (2009) Phylogeny and Identification of Pantoea Species and Typing of Pantoea agglomerans Strains by Multilocus Gene Sequencing. J Clin Microbiol 47:300–310. https://doi.org/10.1128/JCM.01916-08

66. Rezzonico F, Smits TH, Montesinos E, et al (2009) Genotypic comparison of Pantoea agglomerans plant and clinical strains. BMC Microbiol 9:204. https://doi.org/10.1186/1471-2180-9-204

67. Johnston-Monje D, Raizada MN (2011) Conservation and Diversity of Seed Associated Endophytes in Zea across Boundaries of Evolution, Ethnography and Ecology. PLoS ONE 6:e20396. https://doi.org/10.1371/journal.pone.0020396

68. Khalaf EM, Raizada MN (2016) Taxonomic and functional diversity of cultured seed associated microbes of the cucurbit family. BMC Microbiol 16:. https://doi.org/10.1186/s12866-016-0743-2

69. Hardoim PR, Hardoim CCP, van Overbeek LS, van Elsas JD (2012) Dynamics of Seed-Borne Rice Endophytes on Early Plant Growth Stages. PLoS ONE 7:e30438. https://doi.org/10.1371/journal.pone.0030438

70. Ferreira A, Quecine MC, Lacava PT, et al (2008) Diversity of endophytic bacteria from Eucalyptus species seeds and colonization of seedlings by Pantoea agglomerans. FEMS Microbiol Lett 287:8–14. https://doi.org/10.1111/j.1574-6968.2008.01258.x

71. Cottyn B, Debode J, Regalado E, et al (2009) Phenotypic and genetic diversity of rice seed-associated bacteria and their role in pathogenicity and biological control. J Appl Microbiol 107:885–897. https://doi.org/10.1111/j.1365-2672.2009.04268.x

72. Khalaf EM, Raizada MN (2018) Bacterial Seed Endophytes of Domesticated Cucurbits Antagonize Fungal and Oomycete Pathogens Including Powdery Mildew. Front Microbiol 9:42. https://doi.org/10.3389/fmicb.2018.00042

73. Rybakova D, Mancinelli R, Wikström M, et al (2017) The structure of the Brassica napus seed microbiome is cultivar-dependent and affects the interactions of symbionts and pathogens. Microbiome 5:104. https://doi.org/10.1186/s40168-017-0310-6

74. Stockwell VO, Johnson KB, Sugar D, Loper JE (2002) Antibiosis Contributes to Biological Control of Fire Blight by Pantoea agglomerans Strain Eh252 in Orchards. Phytopathology 92:1202–1209. https://doi.org/10.1094/PHYTO.2002.92.11.1202

75. Wright SAI, Zumoff CH, Schneider L, Beer SV (2001) Pantoea agglomerans Strain EH318 Produces Two Antibiotics That Inhibit Erwinia amylovoraIn Vitro. Appl Env Microbiol 67:284–292. https://doi.org/10.1128/AEM.67.1.284-292.2001

76. Enya J, Koitabashi M, Shinohara H, et al (2007) Phylogenetic Diversities of Dominant Culturable Bacillus, Pseudomonas and Pantoea Species on Tomato Leaves and Their Possibility as Biological Control Agents. J Phytopathol 155:446–453. https://doi.org/10.1111/j.1439-0434.2007.01256.x

77. Mitter B, Pfaffenbichler N, Flavell R, et al (2017) A New Approach to Modify Plant Microbiomes and Traits by Introducing Beneficial Bacteria at Flowering into Progeny Seeds. Front Microbiol 8:. https://doi.org/10.3389/fmicb.2017.00011

78. King EO, Ward MK, Raney DE (1954) Two simple media for the demonstration of pyocyanin and fluorescin. J Lab Clin Med 44:301–307

79. Bertani G (1951) STUDIES ON LYSOGENESIS I. J Bacteriol 62:293–300

80. Baba T, Ara T, Hasegawa M, et al (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:2006.0008. https://doi.org/10.1038/msb4100050

81. Hassan JA, Zhou Y-MJ, Lewis JD (2017) A rapid seedling resistance assay identifies wild tomato lines that are resistant to Pseudomonas syringae pv. tomato race 1. Mol Plant Microbe Interact. https://doi.org/10.1094/MPMI-11-16-0247-R

82. Song GC, Choi HK, Kim YS, et al (2017) Seed defense biopriming with bacterial cyclodipeptides triggers immunity in cucumber and pepper. Sci Rep 7:14209. https://doi.org/10.1038/s41598-017-14155-9

83. Rajendran DK, Park E, Nagendran R, et al (2016) Visual Analysis for Detection and Quantification of Pseudomonas cichorii Disease Severity in Tomato Plants. Plant Pathol J 32:300–310. https://doi.org/10.5423/PPJ.OA.01.2016.0032

84. Madden LV, Hughes G, Bosch F van den (2007) The study of plant disease epidemics. Study Plant Dis Epidemics

85. Bergmark L, Poulsen PHB, Al-Soud WA, et al (2012) Assessment of the specificity of Burkholderia and Pseudomonas qPCR assays for detection of these genera in soil using 454 pyrosequencing. FEMS Microbiol Lett 333:77–84. https://doi.org/10.1111/j.1574-6968.2012.02601.x

86. Morella NM, Yang SC, Hernandez CA, Koskella B (2018) Rapid quantification of bacteriophages and their bacterial hosts in vitro and in vivo using droplet digital PCR. J Virol Methods 259:18–24. https://doi.org/10.1016/j.jviromet.2018.05.007

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

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

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

90. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023

91. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054

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

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

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

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

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

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

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

99. Wickham H (2009) Ggplot2: elegant graphics for data analysis. Springer, New York

100. Gopal M, Gupta A (2016) Microbiome Selection Could Spur Next-Generation Plant Breeding Strategies. Front Microbiol 7:. https://doi.org/10.3389/fmicb.2016.01971

101. Mimee M, Citorik RJ, Lu TK (2016) Microbiome Therapeutics – Advances and Challenges. Adv Drug Deliv Rev 105:44–54. https://doi.org/10.1016/j.addr.2016.04.032

102. Panke-Buisse K, Poole AC, Goodrich JK, et al (2015) Selection on soil microbiomes reveals reproducible impacts on plant function. ISME J 9:980–989. https://doi.org/10.1038/ismej.2014.196

103. Marasco R, Rolli E, Ettoumi B, et al (2012) A drought resistance-promoting microbiome is selected by root system under desert farming. PloS One 7:e48479. https://doi.org/10.1371/journal.pone.0048479

104. Rolli E, Marasco R, Vigani G, et al (2015) Improved plant resistance to drought is promoted by the root-associated microbiome as a water stress-dependent trait. Environ Microbiol 17:316–331. https://doi.org/10.1111/1462-2920.12439

105. Pineda A, Kaplan I, Bezemer TM (2017) Steering Soil Microbiomes to Suppress Aboveground Insect Pests. Trends Plant Sci 22:770–778. https://doi.org/10.1016/j.tplants.2017.07.002

106. Wagner MR, Lundberg DS, del Rio TG, et al (2016) Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. Nat Commun 7:12151. https://doi.org/10.1038/ncomms12151

107. Costello EK, Lauber CL, Hamady M, et al (2009) Bacterial Community Variation in Human Body Habitats Across Space and Time. Science 326:1694–1697. https://doi.org/10.1126/science.1177486

108. Benson AK, Kelly SA, Legge R, et al (2010) Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. Proc Natl Acad Sci U S A 107:18933–18938. https://doi.org/10.1073/pnas.1007028107

109. Spor A, Koren O, Ley R (2011) Unravelling the effects of the environment and host genotype on the gut microbiome. Nat Rev Microbiol 9:279–290. https://doi.org/10.1038/nrmicro2540

110. Micallef SA, Channer S, Shiaris MP, Colón-Carmona A (2009) Plant age and genotype impact the progression of bacterial community succession in the Arabidopsis rhizosphere. Plant Signal Behav 4:777–780. https://doi.org/10.4161/psb.4.8.9229

111. Maignien L, DeForce EA, Chafee ME, et al (2014) Ecological Succession and Stochastic Variation in the Assembly of Arabidopsis thaliana Phyllosphere Communities. Mbio 5:e00682-13. https://doi.org/10.1128/mBio.00682-13

112. Rothschild D, Weissbrod O, Barkan E, et al (2018) Environment dominates over host genetics in shaping human gut microbiota. Nature 555:210–215. https://doi.org/10.1038/nature25973

113. Laforest-Lapointe I, Messier C, Kembel SW (2016) Host species identity, site and time drive temperate tree phyllosphere bacterial community structure. Microbiome 4:27. https://doi.org/10.1186/s40168-016-0174-1

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

115. Ebert D (1998) Experimental evolution of parasites. Science 282:1432–1435

116. Buckling A, Craig Maclean R, Brockhurst MA, Colegrave N (2009) The Beagle in a bottle. Nature 457:824–829. https://doi.org/10.1038/nature07892

117. Stone BWG, Weingarten EA, Jackson CR (2018) The Role of the Phyllosphere Microbiome in Plant Health and Function. In: Annual Plant Reviews online. American Cancer Society, pp 1–24

118. Williams TR, Marco ML (2014) Phyllosphere Microbiota Composition and Microbial Community Transplantation on Lettuce Plants Grown Indoors. mBio 5:e01564-14. https://doi.org/10.1128/mBio.01564-14

119. Rillig MC, Antonovics J, Caruso T, et al (2015) Interchange of entire communities: microbial community coalescence. Trends Ecol Evol 30:470–476. https://doi.org/10.1016/j.tree.2015.06.004

120. Segata N, Izard J, Waldron L, et al (2011) Metagenomic biomarker discovery and explanation. Genome Biol 12:R60. https://doi.org/10.1186/gb-2011-12-6-r60

121. Kruskal WH, Wallis WA (1952) Use of Ranks in One-Criterion Variance Analysis. J Am Stat Assoc 47:583–621. https://doi.org/10.2307/2280779

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

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

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

125. Wright DH (1991) Correlations Between Incidence and Abundance are Expected by Chance. J Biogeogr 18:463–466. https://doi.org/10.2307/2845487

126. Gonzalez A, Lawton JH, Gilbert FS, et al (1998) Metapopulation Dynamics, Abundance, and Distribution in a Microecosystem. Science 281:2045–2047. https://doi.org/10.1126/science.281.5385.2045

127. Gould AL, Zhang V, Lamberti L, et al (2018) Microbiome interactions shape host fitness. Proc Natl Acad Sci 115:E11951–E11960. https://doi.org/10.1073/pnas.1809349115

128. Sierocinski P, Milferstedt K, Bayer F, et al (2017) A Single Community Dominates Structure and Function of a Mixture of Multiple Methanogenic Communities. Curr Biol 27:3390-3395.e4. https://doi.org/10.1016/j.cub.2017.09.056

129. Agler MT, Ruhe J, Kroll S, et al (2016) Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. PLOS Biol 14:e1002352. https://doi.org/10.1371/journal.pbio.1002352

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

131. Dini-Andreote F, Elsas JD van, Olff H, Salles JF (2018) Dispersal-competition tradeoff in microbiomes in the quest for land colonization. Sci Rep 8:9451. https://doi.org/10.1038/s41598-018-27783-6

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

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

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

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

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

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

138. Bergh Ø, BØrsheim KY, Bratbak G, Heldal M (1989) High abundance of viruses found in aquatic environments. Nature 340:467–468. https://doi.org/10.1038/340467a0

139. Clokie MR, Millard AD, Letarov AV, Heaphy S (2011) Phages in nature. Bacteriophage 1:31–45. https://doi.org/10.4161/bact.1.1.14942

140. Weinbauer MG (2004) Ecology of prokaryotic viruses. FEMS Microbiol Rev 28:127–181. https://doi.org/10.1016/j.femsre.2003.08.001

141. Luria SE, Delbrück M (1943) Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics 28:491–511

142. Álvarez B, Biosca EG (2017) Bacteriophage-Based Bacterial Wilt Biocontrol for an Environmentally Sustainable Agriculture. Front Plant Sci 8:. https://doi.org/10.3389/fpls.2017.01218

143. Tanaka H, Negishi H, Maeda H (1990) Control of Tobacco Bacterial Wilt by an Avirulent Strain of *Pseudomonas solanacearum* M4S and Its Bacteriophage. Jpn J Phytopathol 56:243–246. https://doi.org/10.3186/jjphytopath.56.243

144. Bae JY, Wu J, Lee HJ, et al (2012) Biocontrol potential of a lytic bacteriophage PE204 against bacterial wilt of tomato. J Microbiol Biotechnol 22:1613–1620

145. Bhunchoth A, Phironrit N, Leksomboon C, et al (2015) Isolation of Ralstonia solanacearum-infecting bacteriophages from tomato fields in Chiang Mai, Thailand, and their experimental use as biocontrol agents. J Appl Microbiol 118:1023–1033. https://doi.org/10.1111/jam.12763

146. Fujiwara A, Fujisawa M, Hamasaki R, et al (2011) Biocontrol of Ralstonia solanacearum by Treatment with Lytic Bacteriophages. Appl Environ Microbiol 77:4155–4162. https://doi.org/10.1128/AEM.02847-10

147. Born Y, Fieseler L, Thöny V, et al (2017) Engineering of Bacteriophages Y2::dpoL1-C and Y2::luxAB for Efficient Control and Rapid Detection of the Fire Blight Pathogen, Erwinia amylovora. Appl Environ Microbiol 83:e00341-17. https://doi.org/10.1128/AEM.00341-17

148. Guenther S, Huwyler D, Richard S, Loessner MJ (2009) Virulent Bacteriophage for Efficient Biocontrol of Listeria monocytogenes in Ready-To-Eat Foods. Appl Environ Microbiol 75:93–100. https://doi.org/10.1128/AEM.01711-08

149. Suttle CA (2007) Marine viruses -- major players in the global ecosystem. Nat Rev Microbiol Lond 5:801–12. http://dx.doi.org/10.1038/nrmicro1750

150. Barr JJ, Auro R, Furlan M, et al (2013) Bacteriophage adhering to mucus provide a non-host-derived immunity. Proc Natl Acad Sci U S A 110:10771–10776. https://doi.org/10.1073/pnas.1305923110

151. Scanlan PD (2017) Bacteria-Bacteriophage Coevolution in the Human Gut: Implications for Microbial Diversity and Functionality. Trends Microbiol. https://doi.org/10.1016/j.tim.2017.02.012

152. Bohannan B j. m., Lenski R e. (2000) Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. Ecol Lett 3:362–377. https://doi.org/10.1046/j.1461-0248.2000.00161.x

153. Koskella B, Brockhurst MA (2014) Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. Fems Microbiol Rev 38:916–931. https://doi.org/10.1111/1574-6976.12072

154. Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, et al (2009) Explaining microbial population genomics through phage predation. Nat Rev Microbiol 7:828–836. https://doi.org/10.1038/nrmicro2235

155. Weinbauer MG, Rassoulzadegan F (2004) Are viruses driving microbial diversification and diversity? Environ Microbiol 6:1–11. https://doi.org/10.1046/j.1462-2920.2003.00539.x

156. Korytowski DA, Smith H (2017) Permanence and Stability of a Kill the Winner Model in Marine Ecology. Bull Math Biol 79:995–1004. https://doi.org/10.1007/s11538-017-0265-6

157. Maslov S, Sneppen K (2017) Population cycles and species diversity in dynamic Kill-the-Winner model of microbial ecosystems. Sci Rep 7:39642. https://doi.org/10.1038/srep39642

158. Gómez P, Buckling A (2011) Bacteria-phage antagonistic coevolution in soil. Science 332:106–109. https://doi.org/10.1126/science.1198767

159. Thingstad T, Lignell R (1997) Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. Aquat Microb Ecol 13:19–27. https://doi.org/10.3354/ame013019

160. Thingstad TF (2000) Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol Oceanogr 45:1320–1328. https://doi.org/10.4319/lo.2000.45.6.1320

161. Scanlan PD (2017) Bacteria–Bacteriophage Coevolution in the Human Gut: Implications for Microbial Diversity and Functionality. Trends Microbiol 25:614–623. https://doi.org/10.1016/j.tim.2017.02.012

162. Manrique P, Bolduc B, Walk ST, et al (2016) Healthy human gut phageome. Proc Natl Acad Sci U S A 113:10400–10405. https://doi.org/10.1073/pnas.1601060113

163. Mirzaei MK, Maurice CF (2017) Menage a trois in the human gut: interactions between host, bacteria and phages. Nat Rev Microbiol 15:397–408. https://doi.org/10.1038/nrmicro.2017.30

164. Weitz JS, Wilhelm SW (2012) Ocean viruses and their effects on microbial communities and biogeochemical cycles. F1000 Biol Rep 4:17. https://doi.org/10.3410/B4-17

165. Kraushaar B, Hammerl JA, Kienoel M, et al (2017) Acquisition of virulence factors in livestock-associated MRSA: Lysogenic conversion of CC398 strains by virulence gene-containing phages. Sci Rep 7:2004. https://doi.org/10.1038/s41598-017-02175-4

166. Waldor MK, Mekalanos JJ (1996) Lysogenic Conversion by a Filamentous Phage Encoding Cholera Toxin. Science 272:1910–1914. https://doi.org/10.1126/science.272.5270.1910

167. Zhang J, Gao Q, Zhang Q, et al (2017) Bacteriophage-prokaryote dynamics and interaction within anaerobic digestion processes across time and space. Microbiome 5:57. https://doi.org/10.1186/s40168-017-0272-8

168. Manrique P, Dills M, Young MJ (2017) The Human Gut Phage Community and Its Implications for Health and Disease. Viruses-Basel 9:141. https://doi.org/10.3390/v9060141

169. Bouvier T, del Giorgio PA (2007) Key role of selective viral-induced mortality in determining marine bacterial community composition. Environ Microbiol 9:287–297. https://doi.org/10.1111/j.1462-2920.2006.01137.x

170. Fuhrman JA, Schwalbach M (2003) Viral Influence on Aquatic Bacterial Communities. Biol Bull 204:192–195. https://doi.org/10.2307/1543557

171. Schwalbach MS, Hewson I, Fuhrman JA (2004) Viral effects on bacterial community composition in marine plankton microcosms. Aquat Microb Ecol 34:117–127. https://doi.org/10.3354/ame034117

172. Staniewski MA, Short CM, Short SM (2012) Contrasting Community versus Population-Based Estimates of Grazing and Virus-Induced Mortality of Phytoplankton. Microb Ecol 64:25–38. https://doi.org/10.1007/s00248-012-0019-9

173. Wilcox RM, Fuhrman JA (1994) Bacterial viruses in coastal seawater: lytic rather than lysogenic production. Mar Ecol Prog Ser 114:35–45

174. Rodriguez-Brito B, Li L, Wegley L, et al (2010) Viral and microbial community dynamics in four aquatic environments. ISME J 4:739. https://doi.org/10.1038/ismej.2010.1

175. Buckling A, Rainey PB (2002) The role of parasites in sympatric and allopatric host diversification. Nature 420:496–499. https://doi.org/10.1038/nature01164

176. Beveridge TJ (1999) Structures of Gram-Negative Cell Walls and Their Derived Membrane Vesicles. J Bacteriol 181:4725–4733

177. Brown L, Wolf JM, Prados-Rosales R, Casadevall A (2015) Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. Nat Rev Microbiol 13:620–630. https://doi.org/10.1038/nrmicro3480

178. Carrera M, Zandomeni RO, Fitzgibbon J, Sagripanti J-L (2007) Difference between the spore sizes of Bacillus anthracis and other Bacillus species. J Appl Microbiol 102:303–312. https://doi.org/10.1111/j.1365-2672.2006.03111.x

179. Biller SJ, McDaniel LD, Breitbart M, et al (2017) Membrane vesicles in sea water: heterogeneous DNA content and implications for viral abundance estimates. ISME J 11:394–404. https://doi.org/10.1038/ismej.2016.134

180. Kembel SW, O'Connor TK, Arnold HK, et al (2014) Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. Proc Natl Acad Sci 111:13715–13720. https://doi.org/10.1073/pnas.1216057111

181. Koskella B, Meaden S, Crowther WJ, et al (2017) A signature of tree health? Shifts in the microbiome and the ecological drivers of horse chestnut bleeding canker disease. New Phytol 215:737–746. https://doi.org/10.1111/nph.14560

182. Reyes A, Blanton LV, Cao S, et al (2015) Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. Proc Natl Acad Sci 112:11941–11946. https://doi.org/10.1073/pnas.1514285112

183. Chelius MK, Triplett EW (2001) The Diversity of Archaea and Bacteria in Association with the Roots of Zea mays L. Microb Ecol 41:252–263. https://doi.org/10.1007/s002480000087

184. Johnston-Monje D, Lundberg DS, Lazarovits G, et al (2016) Bacterial populations in juvenile maize rhizospheres originate from both seed and soil. Plant Soil 405:337–355. https://doi.org/10.1007/s11104-016-2826-0

185. Schloss PD, Westcott SL (2011) Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Appl Environ Microbiol 77:3219–3226. https://doi.org/10.1128/AEM.02810-10

186. Westcott SL, Schloss PD (2015) De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. PeerJ 3:e1487. https://doi.org/10.7717/peerj.1487

187. Ortmann AC, Lawrence JE, Suttle CA (2002) Lysogeny and Lytic Viral Production during a Bloom of the Cyanobacterium Synechococcus spp. Microb Ecol 43:225–231. https://doi.org/10.1007/s00248-001-1058-9

188. Knowles B, Silveira CB, Bailey BA, et al (2016) Lytic to temperate switching of viral communities. Nature 531:466. https://doi.org/10.1038/nature17193

189. Petrie KL, Palmer ND, Johnson DT, et al (2018) Destabilizing mutations encode nongenetic variation that drives evolutionary innovation. Science 359:1542–1545. https://doi.org/10.1126/science.aar1954

190. Koskella B, Thompson JN, Preston GM, Buckling A (2011) Local biotic environment shapes the spatial scale of bacteriophage adaptation to bacteria. Am Nat 177:440–451. https://doi.org/10.1086/658991

191. Iriarte FB, Balogh B, Momol MT, et al (2007) Factors Affecting Survival of Bacteriophage on Tomato Leaf Surfaces. Appl Environ Microbiol 73:1704–1711. https://doi.org/10.1128/AEM.02118-06

192. Morella NM, Koskella B (2017) The Value of a Comparative Approach to Understand the Complex Interplay between Microbiota and Host Immunity. Front Immunol 8:. https://doi.org/10.3389/fimmu.2017.01114

193. Develey-Rivière M-P, Galiana E (2007) Resistance to pathogens and host developmental stage: a multifaceted relationship within the plant kingdom. New Phytol 175:405–416. https://doi.org/10.1111/j.1469-8137.2007.02130.x

194. Whalen MC (2005) Host defence in a developmental context. Mol Plant Pathol 6:347–360. https://doi.org/10.1111/j.1364-3703.2005.00286.x

195. Panter SN, Jones DA (2002) Age-related resistance to plant pathogens. Adv Bot Res 38:251–280. https://doi.org/10.1016/S0065-2296(02)38032-7

196. Huang Y, Kuang Z, Wang W, Cao L Exploring potential bacterial and fungal biocontrol agents transmitted from seeds to sprouts of wheat. Biol Control. https://doi.org/10.1016/j.biocontrol.2016.02.013

197. Lindow SE, Suslow TV (2003) Temporal Dynamics of the Biocontrol Agent Pseudomonas fluorescens Strain A506 in Flowers in Inoculated Pear Trees. Phytopathology 93:727–737. https://doi.org/10.1094/PHYTO.2003.93.6.727

198. Peiffer JA, Spor A, Koren O, et al (2013) Diversity and heritability of the maize rhizosphere microbiome under field conditions. Proc Natl Acad Sci 110:6548–6553. https://doi.org/10.1073/pnas.1302837110

199. Horton MW, Bodenhausen N, Beilsmith K, et al (2014) Genome-wide association study of Arabidopsis thaliana's leaf microbial community. Nat Commun 5:5320. https://doi.org/10.1038/ncomms6320

200. Koskella B, Meaden S (2013) Understanding Bacteriophage Specificity in Natural Microbial Communities. Viruses 5:806–823. https://doi.org/10.3390/v5030806

201. Anderson TF (1948) The Influence of Temperature and Nutrients on Plaque Formation by Bacteriophages Active on Escherichia coli Strain B 1. J Bacteriol 55:659–665

202. Howard-Varona C, Hargreaves KR, Solonenko NE, et al (2018) Multiple mechanisms drive phage infection efficiency in nearly identical hosts. ISME J 1. https://doi.org/10.1038/s41396-018-0099-8

203. Hindson CM, Chevillet JR, Briggs HA, et al (2013) Absolute quantification by droplet digital PCR versus analog real-time PCR. Nat Methods 10:1003–1005. https://doi.org/10.1038/nmeth.2633

204. Peng X, Nguyen A, Ghosh D (2018) Quantification of M13 and T7 bacteriophages by TaqMan and SYBR green qPCR. J Virol Methods 252:100–107. https://doi.org/10.1016/j.jviromet.2017.11.012

205. Hayden RT, Gu Z, Ingersoll J, et al (2013) Comparison of Droplet Digital PCR to Real-Time PCR for Quantitative Detection of Cytomegalovirus. J Clin Microbiol 51:540–546. https://doi.org/10.1128/JCM.02620-12

206. Yan Y, Jia X-J, Wang H-H, et al (2016) Dynamic quantification of avian influenza H7N9(A) virus in a human infection during clinical treatment using droplet digital PCR. J Virol Methods 234:22–27. https://doi.org/10.1016/j.jviromet.2016.04.001

207. Rački N, Morisset D, Gutierrez-Aguirre I, Ravnikar M (2014) One-step RT-droplet digital PCR: a breakthrough in the quantification of waterborne RNA viruses. Anal Bioanal Chem 406:661–667. https://doi.org/10.1007/s00216-013-7476-y

208. Tadmor AD, Ottesen EA, Leadbetter JR, Phillips R (2011) Probing Individual Environmental Bacteria for Viruses by Using Microfluidic Digital PCR. Science 333:58–62. https://doi.org/10.1126/science.1200758

209. Lim SW, Lance ST, Stedman KM, Abate AR (2017) PCR-activated cell sorting as a general, cultivation-free method for high-throughput identification and enrichment of virus hosts. J Virol Methods 242:14–21. https://doi.org/10.1016/j.jviromet.2016.12.009

210. Ma L, Zhu S, Tian Y, et al (2016) Label-Free Analysis of Single Viruses with a Resolution Comparable to That of Electron Microscopy and the Throughput of Flow Cytometry. Angew Chem Int Ed 55:10239–10243. https://doi.org/10.1002/anie.201603007

211. Allers E, Moraru C, Duhaime MB, et al (2013) Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. Environ Microbiol 15:2306–2318. https://doi.org/10.1111/1462-2920.12100

212. Gaudin R, Barteneva NS (2015) Sorting of small infectious virus particles by flow virometry reveals distinct infectivity profiles. Nat Commun 6:6022. https://doi.org/10.1038/ncomms7022

213. Rajamanickam V, Wurm D, Slouka C, et al (2017) A novel toolbox for E-coli lysis monitoring. Anal Bioanal Chem 409:667–671. https://doi.org/10.1007/s00216-016-9907-z

214. Verthé K, Verstraete W (2006) Use of flow cytometry for analysis of phage-mediated killing of Enterobacter aerogenes. Res Microbiol 157:613–618. https://doi.org/10.1016/j.resmic.2006.02.007

215. Zemb O, Manefield M, Thomas F, Jacquet S (2013) Phage adsorption to bacteria in the light of the electrostatics: A case study using E. coli, T2 and flow cytometry. J Virol Methods 189:283–289. https://doi.org/10.1016/j.jviromet.2013.02.007

216. Brussaard CPD, Marie D, Bratbak G (2000) Flow cytometric detection of viruses. J Virol Methods 85:175–182. https://doi.org/10.1016/S0166-0934(99)00167-6

217. Bacteriophages - Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions | Martha R. J. Clokie | Springer. http://www.springer.com/gp/book/9781588296825. Accessed 2 Apr 2018

218. Rohwer F, Edwards R (2002) The Phage Proteomic Tree: a Genome-Based Taxonomy for Phage. J Bacteriol 184:4529–4535. https://doi.org/10.1128/JB.184.16.4529-4535.2002

219. d'Herelle On an invisible microbe antagonistic to dysentery bacilli.

220. Panec M, Katz DS (2006) Plaque Assay Protocols. http://www.asmscience.org/content/education/protocol/protocol.3073. Accessed 17 May 2017

221. Mazzocco A, Waddell TE, Lingohr E, Johnson RP (2009) Enumeration of bacteriophages using the small drop plaque assay system. Methods Mol Biol Clifton NJ 501:81–85. https://doi.org/10.1007/978-1-60327-164-6_9

222. Macken C (1999) Design and analysis of serial limiting dilution assays with small sample sizes. J Immunol Methods 222:13–29. https://doi.org/10.1016/S0022-1759(98)00133-1

223. Otawa K., Satoh H., Kanai Y., et al (2008) Rapid quantification of total viral DNA in the supernatants of activated sludge samples with the fluorescent dye PicoGreen®. Lett Appl Microbiol 46:434–438. https://doi.org/10.1111/j.1472-765X.2008.02335.x

224. DNase I Treatment. https://www.protocols.io/view/DNase-I-Treatment-c3myk5. Accessed 2 Apr 2018

225. Ellis EL, Delbrück M (1939) THE GROWTH OF BACTERIOPHAGE. J Gen Physiol 22:365–384

226. Kelln RA, Warren RAJ (1971) Isolation and properties of a bacteriophage lytic for a wide range of pseudomonads. Can J Microbiol 17:677–682. https://doi.org/10.1139/m71-109

227. Hyman P, Abedon S Practical Methods for Determining Phage Growth Parameters. In: ResearchGate. https://www.researchgate.net/publication/23642166_Practical_Methods_for_Determining_Phage_Growth_Parameters. Accessed 6 Apr 2018

228. You L, Suthers PF, Yin J (2002) Effects of Escherichia coli Physiology on Growth of Phage T7 In Vivo and In Silico. J Bacteriol 184:1888–1894. https://doi.org/10.1128/JB.184.7.1888-1894.2002

229. Rački N, Dreo T, Gutierrez-Aguirre I, et al (2014) Reverse transcriptase droplet digital PCR shows high resilience to PCR inhibitors from plant, soil and water samples. Plant Methods 10:42. https://doi.org/10.1186/s13007-014-0042-6

230. Maruyama A, Oda M, Higashihara T (1993) Abundance of Virus-Sized Non-DNase-Digestible DNA (Coated DNA) in Eutrophic Seawater. Appl Environ Microbiol 59:712–717

Appendix A: Quantification of Bacteriophages and Their Bacterial Hosts *in Vitro* and *in Vivo* using Droplet Digital PCR

Parts of this chapter have been adapted from the following with permission:
Morella, Norma M., Shangyang Christopher Yang, Catherine A. Hernandez, and Britt Koskella. "Rapid quantification of bacteriophages and their bacterial hosts in vitro and in vivo using droplet digital PCR." *Journal of virological methods* 259 (2018): 18-24.

Scientific interest in bacteriophages, viruses that infect bacteria, spans multiple fields, including biotechnology, pathology, and evolutionary biology. Despite this widespread interest, current methods in phage research, such as plaque assays, are often laborious and dependent upon both the host bacterium being examined and the environmental conditions under which the assay is run [200–202]. Though plaque assays remain an essential gold standard for quantification of infectious particles in phage research, there is a need for more high-throughput and sensitive methods to study bacteria-phage dynamics. To that end, we set out to optimize droplet digital polymerase chain reaction (ddPCR) as a method for studying phage dynamics in pairwise interactions with host bacteria, in phage-phage competition experiments, and within a eukaryotic host.

Droplet digital PCR, thoroughly described elsewhere [203], is a digital PCR technique for absolute quantification of DNA or RNA targets. Using primers or fluorescent probes designed for the gene or region of interest, PCR reactions are partitioned into approximately 20,000 droplets using water-oil emulsion, and the reaction is then carried out independently within each individual droplet. A droplet reader then measures the amount of amplified target in each droplet using the fluorescent signal it produces. Then, using a Poisson distribution, optimized software calculates the absolute copy number of target molecules in the reaction mixture based on this fluorescent signal. Multiple gene targets can be quantified simultaneously in a sample with different fluorescent signals or even further multiplexed byvarying concentrations of primers and probes. Another advantage is that unlike quantitative PCR [204], ddPCR does not require standard curves, and it has also been shown to be more precise and accurate than quantitative PCR when quantifying low-copy- number genes [205, 206]. This technology has shown great promise for studying viral dynamics in other systems, such as quantification of cytomegalovirus in human plasma [205], waterborne rotavirus in water samples [207], and quantification of avian influenza throughout treatment for human infection [206]. For phage, it has been used to explore phylogenetic patterns between viruses and bacteria residing in termite guts [208], and it has also been used to sort infected cells in a culture-independent manner, a method called PCR activated cell sorting, PACS [209]. Despite these advances for using ddPCR to study phage, to our knowledge, it has not been used as a method to track phage dynamics over time, and it promises to have many advantages over traditional methods for studying bacteria-phage interactions such as being high-throughput, repeatable, and sensitive (Table 1).

Table 1: Current methods in bacteriophage research

Attribute/Measurement	ddPCR	qPCR	Plaque Assays	Electron Microscopy	Flow Cytometry	Phage FISH
Resolution Scale for enumeration	Gene copy number/ul from plaques or DNA	Gene copy number/ul from plaques or DNA	Plaque forming unit	Visible viral particles	Unlabeled: 27nm- sized particles; 4nm resolution (must be pure) [210]	Gene copy [211]
High Throughput	Yes	Yes	No	No	Yes	No
Specialized equipment required	Yes	Yes	No	Yes	Yes	Yes
Genome or partial genome sequence required	Yes	Yes	No	No	No	Yes
Discernment between infectious and non-infectious particles	No	No	Yes	Possible	Sometimes [212]	Possible if probe detects non-infectious particles
Time-to-Lysis	Yes	Not yet demonstrated	Yes	Not easily	Yes [213, 214]	Yes
Adsorption Rate	Possible	Not yet demonstrated	Yes	Not easily	Yes [215]	Yes
Discernment between taxonomically different phage in competition	Yes	Yes [204]	No	Not easily	Sometimes: different families can take up fluorescent dye differently [216]	Possible

Table 2: Primer and probe sequences

Organism	Gene Target	Forward Primer 5' to 3'	Reverse Primer 5' to 3'	Probe	Fluorophore
FRS	Capsid	GCAACTGGCGGTCAACAAAT	ATGCAGTCAGAACTTCGCCA	AGGGCCAATCGGCTGCTGA	HEX
SHL	Capsid	CGATGTTCAGAACGATGGCAG	GTGCTGAATACTCGGCGCAA	CTACGACATCGAAGACGCGA	FAM
Pseudomonas Syringae pv Tomato Strain DC3000	16S	ACTTTAAGTTGGGAGGAAGGG	ACACAGGAAATTCCACCACCC	TGCCAGCAGCCGCGG	FAM

Bacteriophages are typically classified morphologically using electron microscopy and/or taxonomically by genome sequencing, and further described using biological characteristics such as host range, time-to-lysis, and burst size [217, 218]. Perhaps the most widely used method to enumerate phages and determine infectivity characteristics are plaque assays, such as soft-agar overlays[219, 220]. Briefly, to perform these assays, a layer of soft agar mixed with host bacteria is poured on top of a layer of harder agar. Purified phage can either be mixed in with the soft agar and bacteria or diluted and plated separately in small quantities onto the soft agar [221], and plates are incubated to produce a bacterial lawn. Susceptible bacteria will be lysed by infectious phage particles, causing the formation of "plaques" (absence of bacterial growth) within these lawns that can be enumerated, and the number of plaque forming units (PFUs) can be calculated to quantify infectious bacteriophages. In addition to being time consuming and low-throughput, plaque assays are also variable across growth conditions, host strain being tested, and bacterial growth phase. For example, the results of plaque assays are influenced by host bacterial strain/genotype [200] and environmental variables such as temperature and nutrients [201]. They also require serial dilutions, a large source of experimental error [222]. Though plaque assays have been commonly used to track host-phage dynamics, the field of phage research would greatly benefit from a more precise and higher throughput method such as ddPCR.

We sought to further develop ddPCR as a tool to **enumerate phages** (thus avoiding many of the drawbacks of soft-agar overlays), **track bacteria-phage dynamics** over time, **compare phage time-to-lysis** to predict risk of infection, and **explore phage-phage competition** dynamics. In our work, we used BioRad's QX200™ Droplet Digital™ PCR System. For verification of ddPCR in the bacteria/bacteriophage system, we used two lytic phages, FRS and SHL (in the Podoviridae family and Myoviridae family, respectively; provided by Omnilytics) that infect the tomato plant pathogen, *Pseudomonas syringae* pathovar tomato strain DC3000 (hereafter referred to only as *P. syringae*). Primers were designed to target the phage capsid gene (Table 2), and we used previously described primers for quantification of *P. syringae* [85] . Probes were synthesized by BioRad based on primer sequences. We next validated specificity of our DNA probes (Figure 1) and tested their ability to detect a range of gene-target quantities in both pure DNA and phage stocks. For purified DNA, we began with a starting concentration of 10 ng/ul and found that our probes measured copy number accurately between 10-1 and 1x10-5 ng/µl (Figure 2). For phage stocks (viral particles), we began with phage at a concentration of 1.5x105 PFU/mL and 8.5x105 PFU/mL of FRS and SHL, respectively. Phage titer was too low to be accurately quantified using soft agar drop plating below a concentration of approximately 100 PFU/mL (1PFU/10µl). Using ddPCR to measure the same dilutions of phage stock, we were able to detect phage concentrations in samples 100 times more dilute, corresponding to approximately 200 DNA copies/mL (Figure 3a and 3b, solid lines), comparable to the sensitivity of recently developed qPCR methods [204].

For FRS and SHL, we frequently do not observe a 1:1 ratio between PFU/mL and DNA copies/mL, and ddPCR values are typically 2 to 3 orders of magnitude higher than PFU values. It is most likely that we do not observe a 1:1 ratio either because of the presence of free DNA in phage stocks or due to non-infectious phage particles (at least in the context of the host being used in plaque assays). In order to eliminate free DNA in stocks, we treated each dilution of phage stock from Figure 3a and 3b with DNase and Proteinase K/EDTA (Figure 3a and Figure 3b, dashed lines) based on previously published protocols for DNase treatment and quantification of phage DNA within particles [223, 224]. DNase eliminates free DNA, and the additional Proteinase K/EDTA treatment will subsequently break open the capsids allowing us to

quantify only those copies that were packaged within a capsid at the time of collection. On average, phage concentrations are 3 to 4 times higher prior to enzyme treatment than post-treatment, but the copies/mL values are still 2-3 orders of magnitude higher than PFU/mL for the same samples. This supports the hypothesis that a large portion of copies being detected by ddPCR represents non-infectious particles in our samples and not free DNA. Ratios of PFU/mL to copy number/mL may vary from system to system and should be validated by the researcher. It is important to note that we do not detect higher levels of phage signal after enzyme treatment, indicating that the hot start step of our PCR protocol is sufficient to break open capsids and release DNA to be amplified. If phage signal were higher after enzyme treatment, it would suggest that the additional Proteinase K treatment is necessary to break open capsids to quantify DNA. Again, this may vary from phage to phage but, when true, allows for very high-throughput characterization of phage dynamics.

In addition to static measures of DNA or RNA abundance, ddPCR can also be used to track phage and bacterial levels simultaneously through different stages of phage infection such as adsorption and time-to-lysis (typically referred to as the latent-period [225]). Phages must first recognize bacterial cells, then attach and inject their DNA (or RNA) into the cell, where they replicate and package their genomic content and finally assemble into progeny phage particles. After phage assembly is complete, the bacterial cells burst and phage particles are released into the environment. We sought to measure both adsorption and time-to-lysis using ddPCR and classic PFU plating methods. First, to measure **adsorption** for FRS phage, we infected a high concentration of *P. syringae* (OD600=0.70) with 700 µl of 1x 105 PFU/mL of FRS in 12mL of Kings Broth. We collected samples every 10 minutes for 60 minutes. At each time point, half of the sample was filtered through a 0.22µm filter (to quantify 'free' phages that are not associated with a bacterial cell), and the other half was frozen on dry ice. The free phage was then either plated in triplicate on soft agar overlays for PFU quantification, or analyzed in triplicate using ddPCR (Figure 4a). We found that for FRS, PFU plating is a better way to determine adsorption rate, as the decrease in free phage particles within the first 30 minutes is more apparent (Figure 4b). We next treated samples with DNase and Proteinase K before running ddPCR, which again lowered overall signal (Figure 4c), but did not look more similar to PFU plating than untreated samples. Taken together with the enzyme treatment results from Figure 1a and b, this supports the idea that a large population of non-infectious phage particles exists in our samples, despite numerous single plaque purification steps, as is common in the field. There may also be a significant portion of the phage population that either does not adsorb, or adsorbs at a slower rate than the other, as has been seen previously in a *Pseudomonas* phage [226]. In such cases, although PFU plating should be more accurate for characteristics such as adsorption rate, ddPCR offers an interesting and exciting way to explore the production of non-infectious particles. Furthermore, ddPCR will most likely still be useful in characterizing adsorption of phages for which the majority of particles are infectious and there is complete adsorption.

We next sought to measure **time to lysis** using ddPCR for each of our two phages. To do this we infected *P. syringae* with each phage separately at a multiplicity of infection of 1 PFU to 100 bacterial cells, as determined by PFU plating and bacterial optical density. Here, samples were collected from the coculture over the course of 24 hours, and each sample was filtered through a 0.22µm filter (Millipore) to measure only free phage at each time point. Though ddPCR would pick up non-encapsidated phage DNA as well, we expect this to be only a small minority of phage signal based on previous enzyme treatment and quantification of phage stocks (Figure 3a and b). We observe a sharp increase in free phage particles, indicative of **time-to-**

lysis, after approximately 3 hours (Figure 3c FRS, 1d SHL). We refer to this drastic increase in phages and complete lysis of bacterial cells as "time-to-lysis" throughout the manuscript; however, it is possible that smaller bursts are occurring within the three hours. When required, a definitive biological characterization of time-to-lysis would require stringent synchronization of phage infection, MOI, and bacterial growth phase [227]. To validate ddPCR as a technique comparable to that of the gold standard PFU plating, samples were also plated on soft agar overlays to determine the infectious particle count at each time point. Using a Pearson's correlation test, we found that there was a positive correlation between ddPCR results and PFU plating for both FRS (r^2=0.92, p=0.001) and SHL (r^2=0.76, p=0.03).

In light of the clear ability of ddPCR to enumerate phage over time, we then explored how it can be used to study phage-bacteria dynamics both *in vitro* and *in* vivo. To do this, we simultaneously measured bacterial and phage density in a six-hour time course samples of co-cultures using two probes within the same reaction for FRS and *P. syringae* (Fam-probe, Hex-probe) and duplicate reactions on the same plate for SHL and *P. syringae*, as both are Fam-probes. To determine the effect of phages on bacterial density, we compared bacterial levels to a culture in which phages were not introduced. Time-to-lysis is again clearly observed after 3 hours, when phage particles increase rapidly as bacterial density subsequently decreases (Figure 5a and b). This decrease in bacterial density was confirmed by plating for colony forming units from the various time points (Figure 6). Bacteria could not be cultured beyond 3 hours for SHL and beyond 4 hours for FRS, as the populations had been driven to extinction by the phage. Detectable bacteria using ddPCR at those time points is likely DNA from lysed bacterial cells that has not yet been degraded.

A unique advantage of using ddPCR to measure both adsorption and time-to-lysis is the ability to compare the number of droplets that are positive for both bacteria and phage signals relative to the total positive droplets. This gives insight to how often phages are found to be attached to or infecting bacterial cells, and it could be used to measure the risk of infection in a given treatment or environment. To demonstrate this approach, we again measured signal of samples collected from an *in vitro* co-culture of *P. syringae* and FRS over the course of 24 hours. We observed an increase in co-occurrence of signals until roughly 5 hours as phage adsorption and replication begins to take place (Figure 5c) Between hours 5 and 6, we observed a clear decrease in co-localization of signal coinciding with a decrease in bacterial signal and sharp increase in phage signal, likely due to release of viral particles from lysed bacterial hosts. This supports an observable **time-to-lysis** in this co-culture of roughly 4 hours. Though this is slightly longer than time-to-lysis observed in Figure 3 for FRS, and longer than might be deduced from our adsorption assay (Figure 4), this may due to the infection time course being started with bacterial culture in a slightly different growth phase and at different MOIs. Care should be taken to standardize host concentration and growth rate between experiments, as these are known to have an observable effect on the outcome [228].

Next, we sought to determine whether ddPCR could be used to detect replication of phage and bacteria *in vivo*. Using a tomato plant as a eukaryotic host, we inoculated leaves with either *P. syringae* alone or with both *P. syringae* and FRS, and measured bacteria and phage concentrations at 0 and 72 hours post-inoculation. Tissue samples were homogenized, plated for colony forming units (CFUs), and measured on the ddPCR. Phage concentration was estimated using number of visible plaques in bacterial dilution plates (the small drop plaque assay system [221]), which was only possible at 72 hours due to low bacterial density at time 0. As expected, bacterial concentration increased over 72 hours as bacteria replicated in the apoplast of each leaf

(significant effect of time in a repeated measures ANOVA after normalization of data via log-transformation, $p<0.01$ for both ddPCR and CFUs; Figure 7A). There was no significant effect of phage on the bacterial concentrations measured via either method (non-significant time by treatment interaction, $p=0.252$ for ddPCR, $p= 0.211$ for CFUs; Figure 7A). However, in plants inoculated with both bacteria and phage we detected a modest yet significant increase in the ddPCR phage signal after 72 hours (paired t-test, $p=0.032$; Figure 7B). Phage signal in plants inoculated with bacteria only was zero for 5 out of 6 samples from both time points, and negligible for one sample from each time point (1.5 and 3 copies/µl, likely post-sampling contamination; data not shown). On average, the phage signal in co-inoculated plants doubled over the 72 hours, which is a much lower level of replication than typically occurs by this time *in vitro*. Thus, while phages were replicating and necessarily lysing bacterial cells within the leaves, this was undetectable by comparing bacterial densities across treatments over time and was not possible to determine using traditional plating methods. Through use of the ddPCR method, we were therefore able to sensitively detect *in vivo* phage replication despite a statistically undetectable impact on bacterial density. Importantly, phage and bacteria levels were also measured directly from homogenized host plant tissue with virtually no background amplification of plant tissue from magnesium chloride-treated plants (data not shown). This has been observed before, where ddPCR was found to be superior to qPCR in its tolerance of PCR inhibitors from the environment [229]. Overall, we have demonstrated the utility of this method to sensitively quantify bacteria and phage interaction within host tissue without the need for DNA extraction.

Perhaps one of the most exciting applications of ddPCR in studying bacteriophage dynamics is the ability to **compete** multiple phage types against one another on a single bacterial host. Using differently labeled probes, we were able to precisely quantify levels of taxonomically distinct phages over the course of infection. There are currently very few effective and sensitive ways to compare individual phage types in competition. Flow cytometry can be used but has important drawbacks, including size limitations and necessary sample purity. Recent work demonstrates qPCR as a viable method to measure different phage types [204], and here, we present an additional molecular method to do so. We infected *P. syringae* with both FRS and SHL at equivalent MOIs, as determined by PFU/mL, and tracked the dynamics of each in a mixed-infection culture over a 24-hour time course. At an MOI of 1 PFU to 100 bacterial cells, we observed that the FRS population performed poorly in the presence of SHL as compared to its growth alone (Figure 8a). We then tested the idea that a prolonged time-to-lysis for FRS may have allowed SHL to outcompete FRS. To do this we sampled co- culture at two time points and centrifuged each to separate the bacterial cells (and any attached or replicating phages) from the free phage. We measured phage densities in the pellet and supernatant at 0 and 5 hours and found a higher proportion of FRS to be in the pelleted fraction as compared with SHL at five hours (40% and 1% respectively; Figure 8b). This supports our hypothesis that SHL has a competitive advantage over FRS, perhaps due to more successful attachment, adsorption, or lysis. We plan to fully describe competitive dynamics between these two phages using ddPCR in future work.

Finally, we sought to determine whether we could detect infection of bacterial cells by more than one type of phage simultaneously. To do this, we compared proportion of droplets positive for both phage signals from the 1:100 MOI infection culture. We found that only 5% of all positive droplets in the pelleted cells were positive for both FRS and SHL (Figure 9), lower than the 29% of double-positive droplets randomly co-occurring in the supernatant of the same samples, suggesting there was little if any co-infection of host cells, in line with our observation

that FRS was almost entirely out-competed by SHL in this competition. Having a high-throughput molecular approach to determine if bacterial hosts are being infected by more than one phage in competition with one another offers a powerful method to address questions about phage-phage-bacteria dynamics such as superinfection, and there are currently only a limited number ways to test this.

To summarize, we have shown that ddPCR can be used to enumerate phage particles, determine time-to-lysis, sensitively track infection dynamics of both bacteriophage and host both *in vivo* and *in vitro*, quantify percent of phage signal located in the bacterial pellet, and determine the outcome of competition between phages. The latter can be used for competition between phage species or ancestral and evolved phage. Certainly, there are drawbacks to the ddPCR system, such as cost and requirement of a genome sequence for primer and probe design. Furthermore, the utility of ddPCR may vary from system to system, and it should first be verified in one's study system using classic methods. Nevertheless, we see great promise in this approach to studying bacteriophages. We find it well worth the cost, and useful for accurate quantification of both bacteria and known phages *in vitro* and *in vivo* [130]. Methods such as these are greatly needed in bacteriophage research. In addition to being used to address fundamental questions about competition under different conditions and in different environments, this method will enhance our ability to design and implement phage therapy agents, which often contain cocktails of multiple phages. The fact that these data can be generated in such a high-throughput fashion, even in the absence of any DNA extraction methods, suggests that great headway can rapidly be made in increasing our understanding of bacteria-phage dynamics under both natural and experimental settings.

Complete Methods

Phage Stock Amplification

FRS and SHL stocks were generated by amplification on bacterial host *Pseudomonas syringae pv* tomato strain DC3000. Seven mL of Kings Broth was combined with 200ul of bacterial culture and 50ul of long-term phage stock. Cultures were incubated at 28°C overnight. The following day, cultures were filtered with a 0.22µm filter. Phage stocks were quantified by spotting serial dilutions of stock onto soft agar overlays (overlay method described below).

Phage DNA Extraction

Purified DNA was used for verification of specificity of probes (Figure 1) and determining ddPCR limit of detection for pure DNA (Figure 1b). To generate lysates, FRS and SHL were amplified as described above. Next, 10 mL of each of the filter-sterile phage lysates were transferred into individual sterile Oak Ridge tubes, and 40 µl of Nuclease Mix (150 mM NaCl, 0.25 mg/mL DNaseI, 0.25 mg/ml RNase A, 50% glycerol, 4.25 ml ddH2O) was added and mixed with gentle inversions. The oak ridge tubes were then incubated at 37°C for 30 minutes, then left at room temperature for 1 hour. 4 ml of phage precipitation solution (30% PEG 8000, 3.3 M NaCl, ddH2O to 100 ml total volume, filter sterilized) was then added to each tube and incubated overnight at 4°C. After overnight incubation, the tubes were spun at 10,000xg for 20 minutes to precipitate the phages. The supernatant was decanted and the tubes left inverted on a paper towel to drain any excess supernatant for 2-3 minutes. 500 µl of sterile ddH2O was then added to the pellet and resuspended by pipetting up and down. The tubes then sat at room temperature for 5-10 minutes. DNA was then extracted using the Wizard PCR Preps DNA Purification system (Promega #7170) kit, and eluted in 10 mM Tris.

Phage Primer and Probe Design and Validation
We sought to target an essential gene from the phage genome, and so we chose to design primers and probes for the capsid gene of both FRS and SHL. Using their annotated genomes, we found unique primers that amplified 100-150bp regions, as per BioRad's recommendations for designing probes. We used primers described in Bergmark et al. 2012 for amplification of *Pseudomonas syringae* pv tomato strain DC3000. Primers and their specificity for intended targets were validated through standard PCR and amplification of either purified DNA or live viral particles. PCR products were run on 1% agarose gels to ensure amplification of intended target. Once primers were validated, BioRad synthesized probes with signals for either HEX or FAM, which were used in all subsequent experiments. Probes were then further validated using the ddPCR system as described below and illustrated in Figure 1.

Plaque Forming Unit (PFU) Plating
In all cases, Kings Broth media and agar were used. Hard agar was used at 1.2% agar and soft agar at 0.6% agar. Seven mL of soft agar was combined with 100µl of a stationary phase overnight DC300 bacterial culture and poured onto a hard agar plate beneath. Once it had solidified, a serial dilution of the phages in sterile KB broth was spotted on top of the soft agar in triplicate. Plates were incubated at 28°C overnight, and plaques were enumerated the following day.

Enzyme Treatment
Based on previous protocols, [223, 224, 230], we used the following enzyme treatment for samples in Figures 3 and 4. DNase (NEB) was diluted in reaction buffer to 1000 U/mL, or 1U/ul. A 10X solution of EDTA and Proteinase K (Thermo Scientific) was made from stocks at 200mM EDTA and 500µg/mL Proteinse K. DNase was added in a 1:10 dilution in phage sample and incubated for 30 mins at 37°C. Then, the EDTA/Proteinase K solution was diluted into sample 1:10 for final concentrations of 20mM of EDTA and 50g/mL of Proteinase K. Samples were incubated at 65°C for 60 mins and then transferred to 4°C until ddPCR analysis.

Time Courses
For the time courses, phages were added to bacterial cultures at an MOI of either 1:10 or 1:100, as determined by OD600 measurements of bacterial culture (we use a conversion of 1.0 OD_{600}= 2*10^8 cells/mL) and PFU counts of phage stocks. Cultures were sampled at time zero and each subsequent hour thereafter by vortexing at medium speed, removing 500µl of culture, and filtering it through a 0.22µm filter unit using a 1mL syringe to remove the bacterial cells. In pelleting experiments, bacterial cells were removed by pelleting at 10,000 x g for five minutes. Samples were frozen on dry ice immediately after sampling and stored at -20°C until ddPCR or PFU/mL analysis.

Competition
When FRS and SHL were competed, the time-course was carried about as described above, except that phages were added to bacterial cultures at equal concentrations, as determined by PFU plating of the phage stocks. Phages were combined together in KB broth before bacterial culture was added.

ddPCR reactions

All steps of ddPCR were carried out on the BioRad QX200 system. PCR reactions were set up as follows: 11µl of Probe Mastermix, 1.1µl of fluorescent probes for each target, 2µl of template, and the remainder made up to 22µl with PCR-grade water. Droplets were formed in sets of 8 using the BioRad Droplet maker with 70µl of droplet generating probe oil. Droplets were transferred to a BioRad 96-well PCR plate, and the PCR reaction is as follows: 95°C for 10 minutes, 94°C for 30 seconds, 60°C for 1 minute, 72°C for 1 minute, 40 cycles of steps 2-4, and 98°C for 10 minutes. Droplets were then read on the BioRad droplet reader. Thresholds were set between positive and negative droplets based on no template controls. BioRad QuantaSoft software reports concentration in copy per µl of reaction mixture. We use this value to calculate copy per µl of initial sample, and we report values in either copy/µl or copy/mL. Samples were run in technical duplicates or triplicates, and the average of the replicates was plotted with standard deviation illustrated as error bars. Technical replication may not always be necessary for ddPCR, as the Poisson Error calculated by the BioRad software can be used as an estimate of experimental error (BioRad ddPCR Applications Guide). Users should reference BioRad software and determine the appropriate replication scale for their experiments. For bacterial signal *in vivo*, we compared replication within and between plates (Figure 10), the details of which are described below.

In vivo bacteria and phage inoculation experiment

Preparation of bacteria, phage, and spent media: An overnight culture of *P. syringae* pv. tomato strain PT23 was grown in liquid KB. After 24 hours, the culture was prepared for inoculation in plants by centrifugation of the cells at 2500xg for 5 min followed by removal of the supernatant and addition of 30 ml sterile 1 mM $MgCl_2$. This process was repeated three times, and OD600 of the final resuspended cells was measured in triplicate using a VersaMax ^{TM} microplate reader (Molecular Devices) to determine concentration. A phage lysate was prepared by filtering (with a 0.45 µm filter) a co-culture of PT23 that had been incubated overnight with FRS phage in liquid KB, and phage titer was enumerated using the soft agar overlay method described previously. An overnight culture of PT23 in liquid KB was filtered using a 0.45 µm filter to generate spent media.

Inoculations and sampling: Four week old tomato plants (*Solanum lycopersicum* cultivar Moneymaker) were inoculated with either bacteria and phage (approximately $3.5x10^4$ cells/ml of bacteria and $7x10^5$ phage particles/ml suspended in 2 ml total volume with 1 mM $MgCl_2$; an MOI of 20:1), bacteria only (approx. $3.5x10^4$ cells/ml of bacteria, spent media in the same volume as phage in the previous inoculum, suspended in 2 ml total volume with 1 mM $MgCl_2$), or a magnesium chloride control (2 ml of 1 mM $MgCl_2$). Plants were randomly assigned to a treatment, with six bacteria and phage plants, six bacteria only plants, and one magnesium chloride control. Using a sterile syringe, inoculum was injected into the underside of six randomly chosen leaves from each plant. Plants were placed randomly in a growth chamber set to 24°C with a 15h day:9h night light cycle. One hole punch (0.2 cm^2) from each of the six leaves was sampled at each time point (0 and 72 hours) from each plant using a flame-sterilized metal hole punch. The samples were homogenized in 1 ml sterile MilliQ water with two ¼" spherical ceramic beads using a FastPrep-24^{TM} 5G benchtop homogenizer (MP Biomedicals) set to 4.0 motions/second for 40 seconds.

Bacteria and phage quantification using CFU/PFU plating and ddPCR

Plant homogenates were plated for bacterial concentration directly after sampling. A dilution series of each sample was plated in duplicate on KB hard agar, and CFUs were recorded after 48 hours of incubation at 28°C. In droplets full of bacteria, samples that had phage also had visible phage plaques. We recorded this number of phage plaques in the droplets as estimated phage concentration ([221]. Homogenates were also filtered using a 0.45 µm filter and plated for phage concentration by mixing with bacteria in soft agar, but too much phage was lost in the filtration process to accurately quantify post-filtration. After plating, samples were frozen at -20°C for storage until measuring concentrations using ddPCR.

Two µl of thawed plant homogenate was used as template for ddPCR. ddPCR was performed as described above, using probes for both *P. syringae* and FRS phage in the reaction of all wells. In some wells, droplet formation was inhibited by the plant homogenate, and any samples with insufficient droplets were run on a new plate. To test whether running samples on two different plates would impact quantification, we ran 2 µl of plant homogenates with bacterial signal from a previous experiment on a single plate with no replicates, and on a second plate in duplicate. We then compared the correlation of bacterial concentration for the same samples within and between plates, and found that between-plate concentrations were no less similar than within-plate duplicates (within-plate $r=0.91$, between plates with replicate 1 $r=0.9$, between plates with replicate 2 $r=0.91$; all p-values<0.001, Figure 10)

Data analysis and statistics

All data analysis and statistics were conducted in R, and figures were made using the ggplot2 package [99].

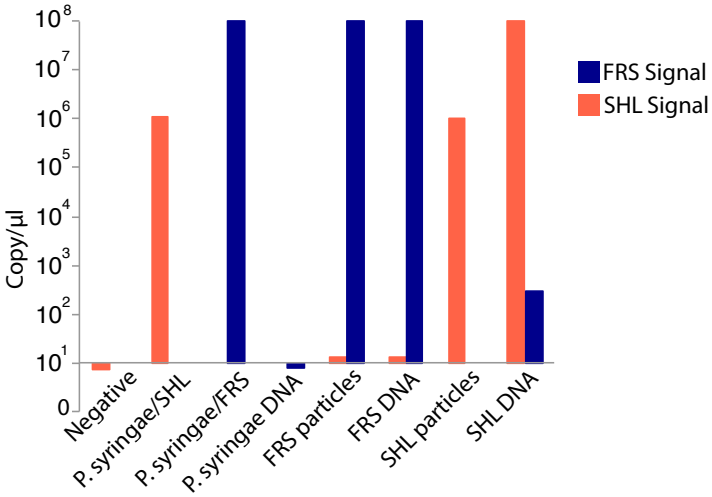


Figure 1 (left) Specificity of probes to FRS, SHL, and *Pst*
SHL and FRS probes amplify their intended target without indirect amplification of bacterial DNA or non-target phage. All probes were used with all samples, and values are reported in copy of DNA target/µl sample. Note log10 scale for y-axis.

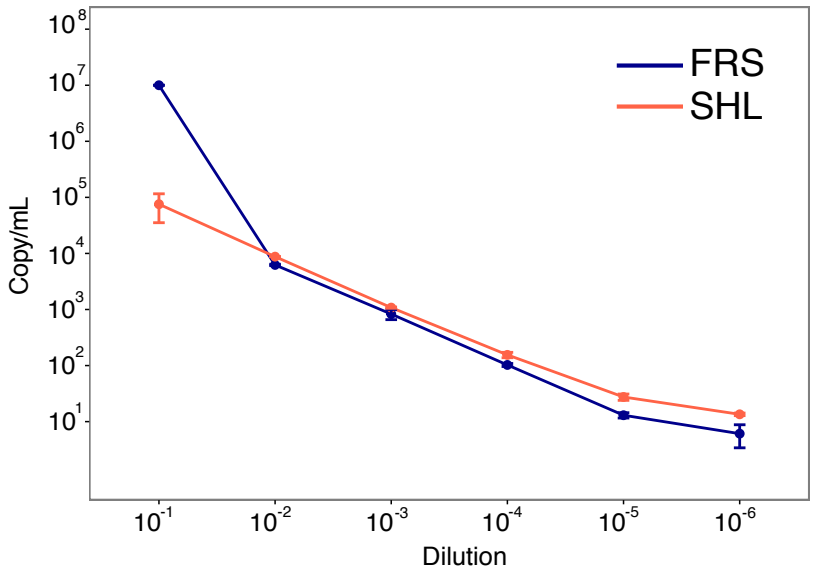


Figure 2 (above) Quantification of phage DNA
Serial dilution of FRS and SHL purified DNA stocks, with a starting concentration of 10ng/ µl. DNA could be accurately quantified between 0.1ng/ µl and 0.0001ng/ µl. Note log10 scale on y-axis. Error bars represent standard deviation of technical replicates.

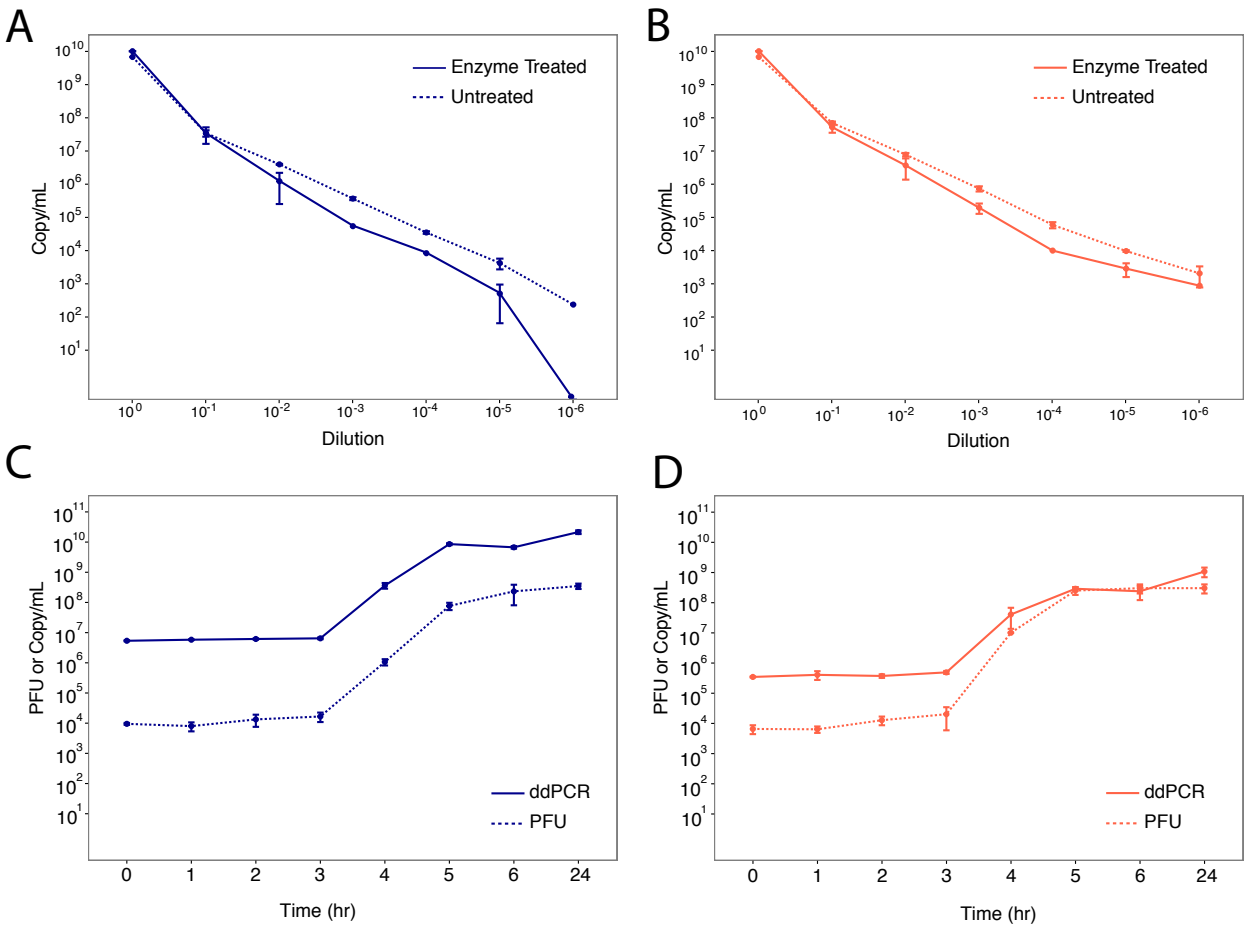
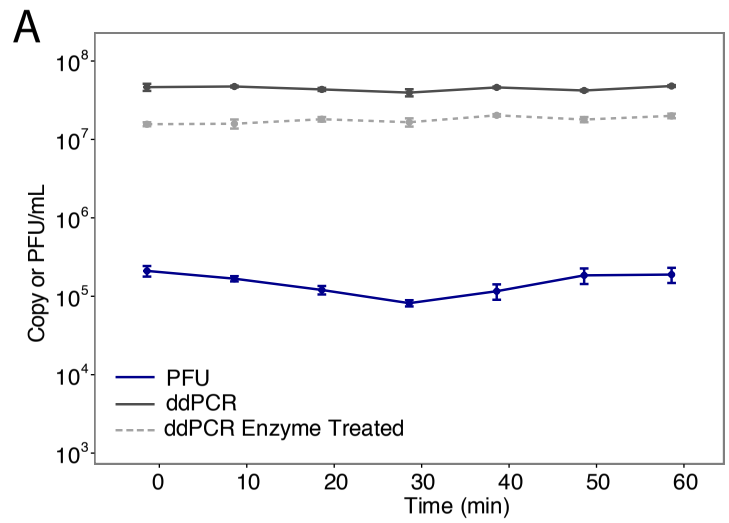
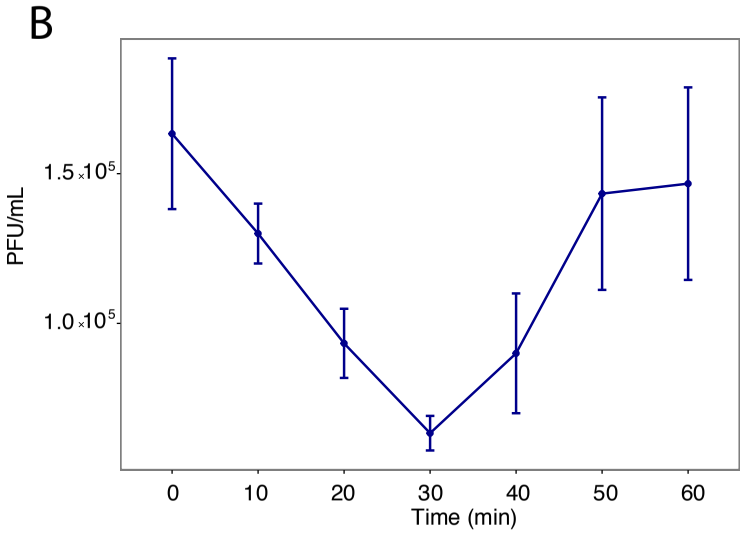


Figure 3 Enumeration of bacteriophage lysate using ddPCR
Dilution series and enumeration of viral particles from FRS (a) and SHL (b) lysates using phage- specific labeled probes. Treatment with DNase and Proteinase K/EDTA (a and b, dashed lines) results in a slight overall decrease in signal. Enumeration of phages over time using ddPCR compared to traditional plaque assays (c and d) is significantly correlated. Error bars represent standard deviation of technical replicates. Note log10 scale for y-axis.





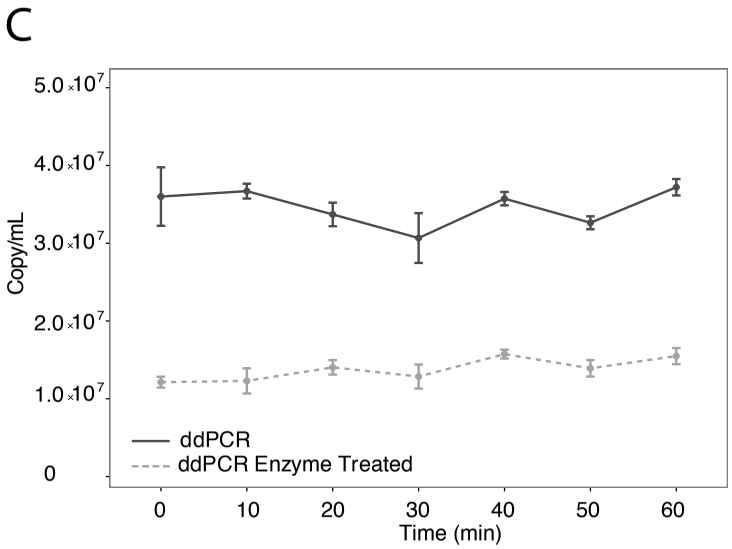


Figure 4 FRS Adsorption
Free phage from adsorption assay was measured over time via PFU plating and ddPCR (a), note the log10 scale on y-axis. PFU plating (b) shows a thirty minute time to adsorption, but the pattern as measured by ddPCR was not as evident. ddPCR samples were also treated with enzyme and compared with untreated samples (c). Error bars represent standard deviation of technical replicates.

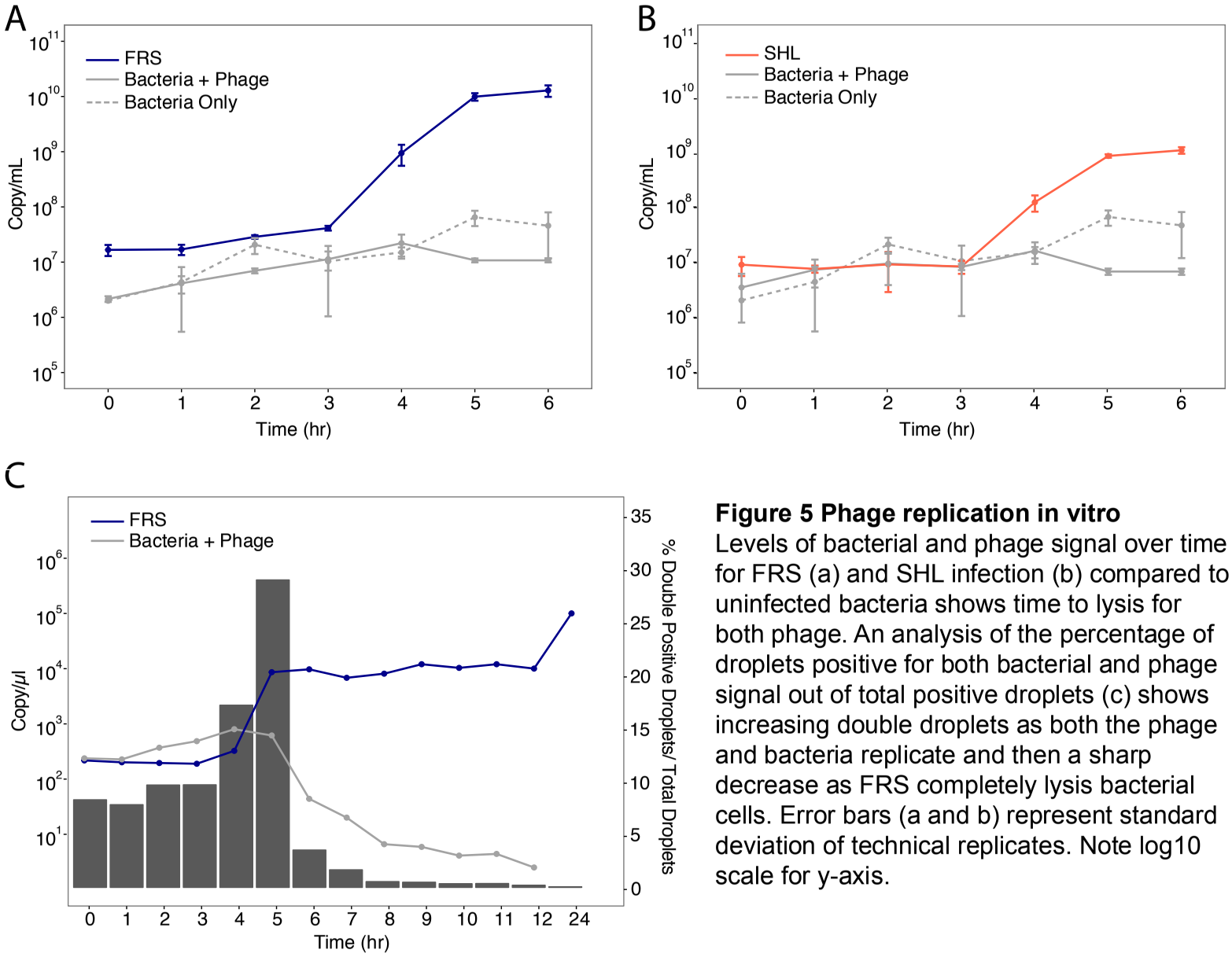


Figure 5 Phage replication in vitro
Levels of bacterial and phage signal over time for FRS (a) and SHL infection (b) compared to uninfected bacteria shows time to lysis for both phage. An analysis of the percentage of droplets positive for both bacterial and phage signal out of total positive droplets (c) shows increasing double droplets as both the phage and bacteria replicate and then a sharp decrease as FRS completely lysis bacterial cells. Error bars (a and b) represent standard deviation of technical replicates. Note log10 scale for y-axis.

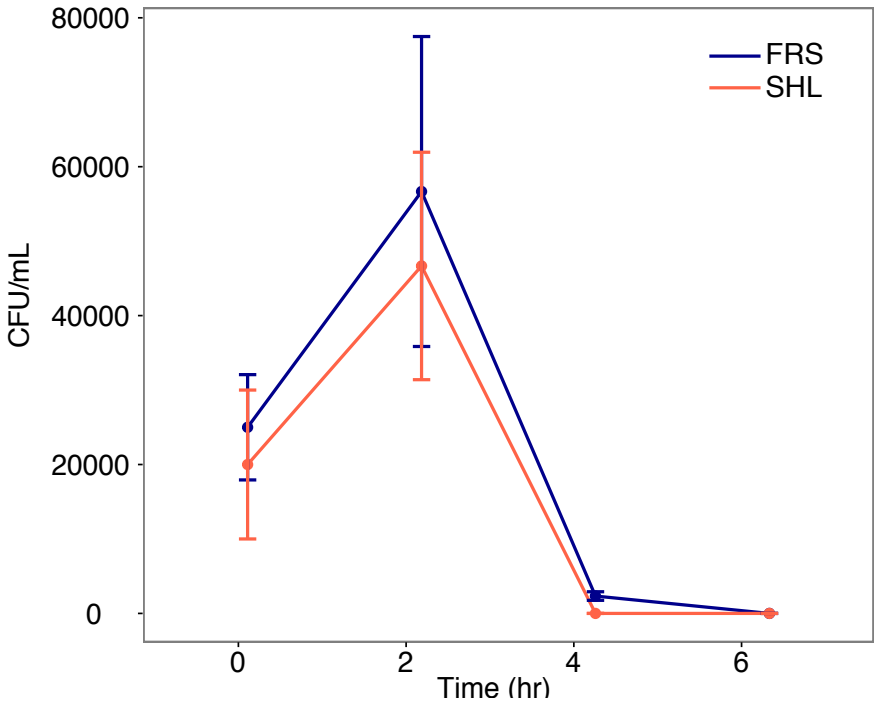


Figure 6 Viable bacterial cells over time
Unfiltered samples from time-to-lysis assays (Figure 5 and b) were diluted in serial dilutions, and 10µl were spotted onto Kings Broth hard agar plates in triplicate. Bacterial growth decreases sharply after 2 hours and is undetectable beyond three hours for FRS and four hours for SHL, confirming lysis of bacterial cells. Error bars represent standard deviation of technical triplicates.

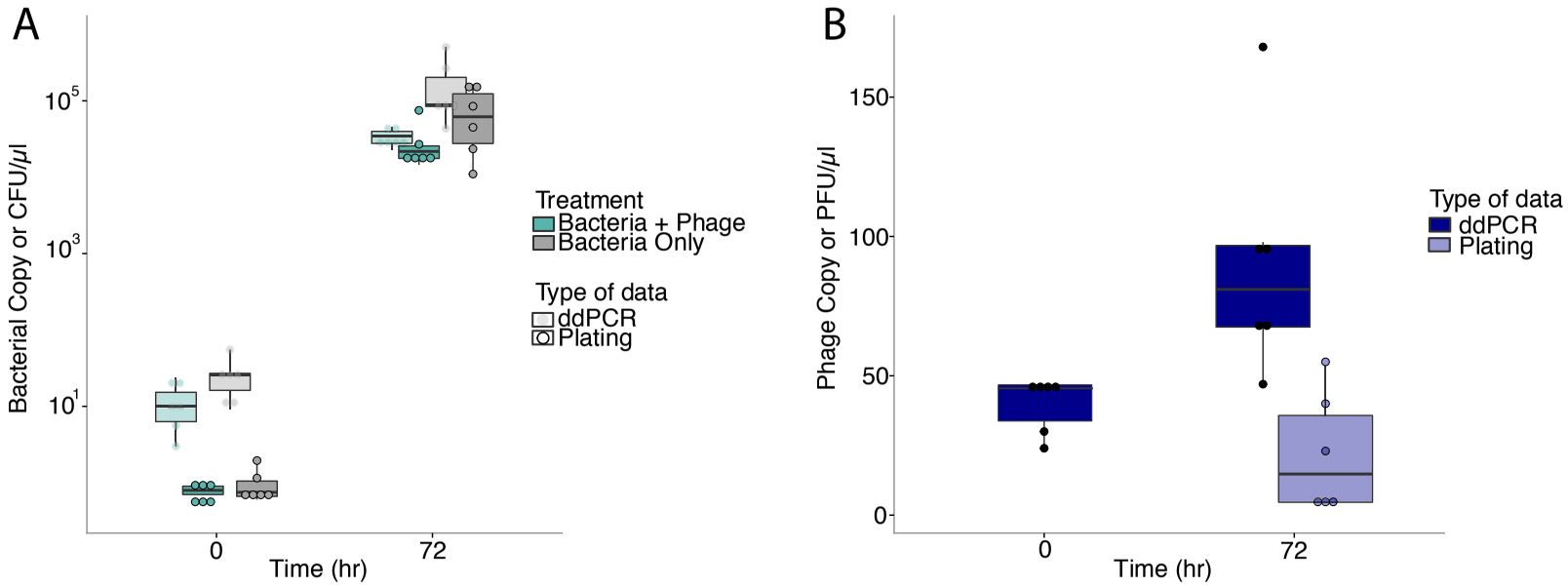


Figure 7 Phage replication in vivo

Tissue homogenates from inoculated tomato leaves at 0 and 72 hours post-inoculation were plated for bacterial colony density and measured for bacteria and phage copy concentration using ddPCR. Infective phage concentration was estimated using number of plaques visible in bacterial dilution plates. Plants were inoculated with bacteria only or both bacteria and phage. Bacterial concentration (a) significantly increased over 72 hours (note log10 scale), but there was no significant effect of treatment over time (repeated measures ANOVA, effect of time $p<0.01$, time by treatment interaction $p=0.252$ for ddPCR, $p=0.211$ for CFUs). Phage concentration (b) significantly increased in plants inoculated with both bacteria and phage (paired t-test, $p=0.032$). Determining zero hour phage concentration in bacterial dilution plates was not possible. Data are displayed on boxplots where black line indicates median, the upper portion of the box is third quartile, lower portion is first quartile, whiskers are maximum and minimum, and outliers (any value more than 1.5x length of the box) are black dots

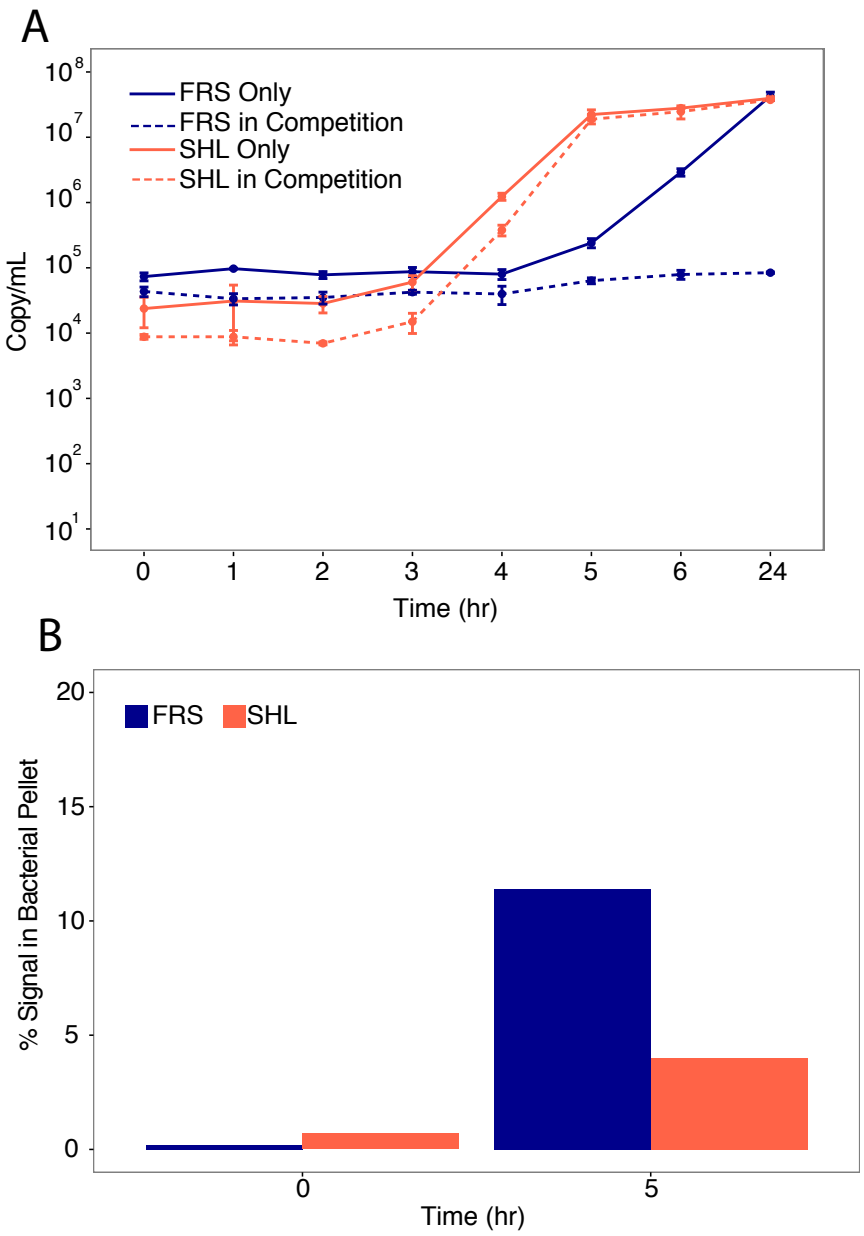


Figure 8 Phage Competition
FRS and SHL abundance in the filtered phage only fraction of phage-phage competition at an MOI of 1:100 demonstrate FRS replication is nearly absent in the presence of SHL (a) (note log10 scale for y-axis). Percent of phage signal remaining in pellet at 0 and 5 hours of FRS and SHL in competition (b) suggest SHL may have had a more efficient or quicker time to lysis than FRS, leading to its competitive dominance over FRS. Error bars represent standard deviation of technical replicates.

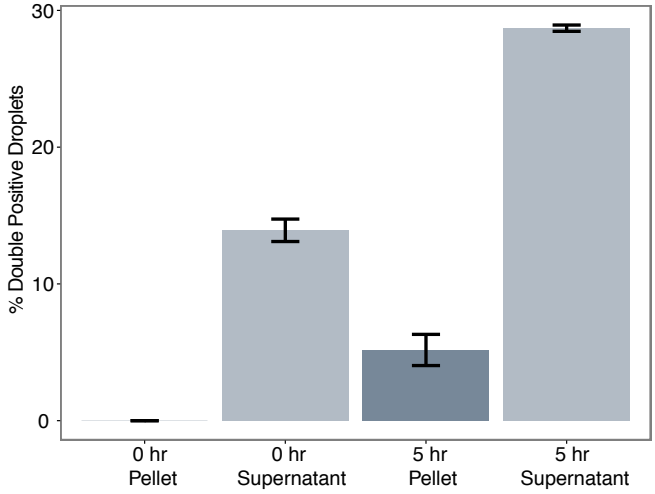


Figure 9 (left) Double positive droplets in phage-phage competition Droplets that are positive for both FRS and SHL signal in the bacterial pellet and supernatant provide little evidence for co-infection of bacterial cells with both phage types, as the percentage of double positive droplets in the supernatants, which would occur by chance, is higher than that of the double positive droplets in the pellet. Values were calculated by dividing the number of double positive droplets by total positive droplets from experimental replicates and taking the average of each sample. Error bars represent standard deviation of replicates.

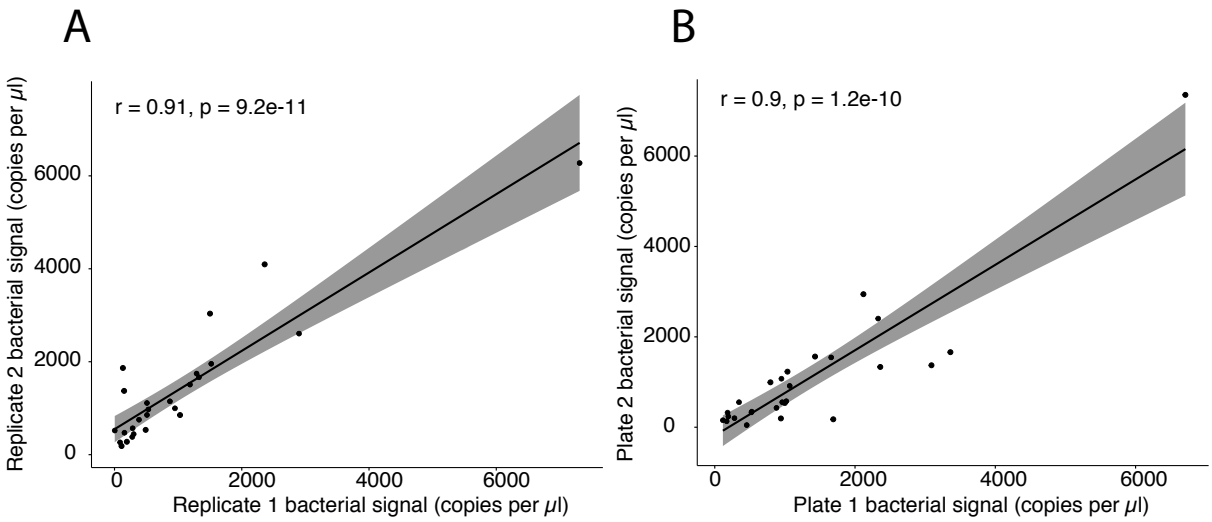


Figure 10 (above) Correlation of within and between plate ddPCR replicates Correlation of bacterial signal in sample replicates on the same plate (a), and across different plates (b). Pearson correlation coefficients were calculated for each pair of replicates (duplicates on the same plate, and the single measurement on a different plate with each of the duplicates from the first plate). The within-plate correlation coefficient was 0.91, and the between-plate correlation coefficients were 0.9 and 0.91 (a single representative of the between-plate comparisons was shown here). The shaded region represents the 95% confidence interval and all p-values<0.001.

Appendix B: Taxonomic identity of top 100 taxa in Chapter 3, Figure 3-5

OTU	Phylum	Class	Order
1	Proteobacteria(100)	Gammaproteobacteria(100)	Enterobacteriales(100)
2	Proteobacteria(100)	Gammaproteobacteria(100)	Pseudomonadales(100)
3	Proteobacteria(100)	Betaproteobacteria(100)	Burkholderiales(100)
4	Proteobacteria(100)	Gammaproteobacteria(100)	Pseudomonadales(100)
5	Proteobacteria(100)	Alphaproteobacteria(100)	Sphingomonadales(100)
6	Proteobacteria(100)	Gammaproteobacteria(100)	Xanthomonadales(100)
8	Proteobacteria(100)	Gammaproteobacteria(100)	Pseudomonadales(100)
9	Proteobacteria(100)	Gammaproteobacteria(100)	Enterobacteriales(100)
10	Proteobacteria(100)	Gammaproteobacteria(100)	Pseudomonadales(100)
11	Proteobacteria(100)	Gammaproteobacteria(100)	Enterobacteriales(100)
12	Firmicutes(100)	Bacilli(100)	Bacillales(100)
13	Proteobacteria(100)	Betaproteobacteria(100)	Burkholderiales(100)
14	Proteobacteria(100)	Gammaproteobacteria(100)	Enterobacteriales(100)
15	Proteobacteria(100)	Alphaproteobacteria(100)	Caulobacterales(100)
16	Firmicutes(100)	Bacilli(100)	Bacillales(100)
17	Proteobacteria(100)	Alphaproteobacteria(100)	Sphingomonadales(100)
18	Proteobacteria(100)	Betaproteobacteria(100)	Burkholderiales(100)
19	Proteobacteria(100)	Gammaproteobacteria(100)	Enterobacteriales(100)
20	Bacteroidetes(100)	Sphingobacteriia(100)	Sphingobacteriales(100)
21	Proteobacteria(100)	Gammaproteobacteria(100)	Pseudomonadales(100)
22	Proteobacteria(100)	Gammaproteobacteria(100)	Xanthomonadales(100)
23	Proteobacteria(100)	Gammaproteobacteria(100)	Xanthomonadales(100)
24	Actinobacteria(100)	Actinobacteria(100)	Micrococcales(100)
25	Firmicutes(100)	Bacilli(100)	Bacillales(100)
26	Actinobacteria(100)	Actinobacteria(100)	Micrococcales(100)
27	Proteobacteria(100)	Alphaproteobacteria(100)	Rhizobiales(100)
28	Proteobacteria(100)	Gammaproteobacteria(100)	Enterobacteriales(100)
29	Proteobacteria(100)	Betaproteobacteria(100)	Burkholderiales(100)
32	Proteobacteria(100)	Alphaproteobacteria(100)	Rhizobiales(100)
33	Actinobacteria(100)	Actinobacteria(100)	Micrococcales(100)
34	Actinobacteria(100)	Actinobacteria(100)	Micrococcales(100)
35	Proteobacteria(100)	Alphaproteobacteria(100)	Rhizobiales(100)
36	Proteobacteria(100)	Betaproteobacteria(100)	Burkholderiales(100)
37	Proteobacteria(100)	Gammaproteobacteria(100)	Enterobacteriales(100)

38	Proteobacteria(100)	Gammaproteobacteria(100)	Enterobacteriales(100)
39	Bacteroidetes(100)	Sphingobacteriia(100)	Sphingobacteriales(100)
40	Bacteroidetes(100)	Sphingobacteriia(100)	Sphingobacteriales(100)
41	Bacteroidetes(100)	Sphingobacteriia(100)	Sphingobacteriales(100)
42	Proteobacteria(100)	Betaproteobacteria(100)	Burkholderiales(100)
43	Firmicutes(100)	Bacilli(100)	Bacillales(100)
45	Firmicutes(100)	Bacilli(100)	Bacillales(100)
46	Actinobacteria(100)	Actinobacteria(100)	Streptomycetales(100)
47	Firmicutes(100)	Bacilli(100)	Bacillales(100)
48	Proteobacteria(100)	Alphaproteobacteria(100)	Rhizobiales(100)
49	Chloroflexi(100)	Ktedonobacteria(100)	Ktedonobacterales(100)
50	Actinobacteria(100)	Actinobacteria(100)	Streptomycetales(100)
52	Proteobacteria(100)	Alphaproteobacteria(100)	Sphingomonadales(100)
53	Proteobacteria(100)	Gammaproteobacteria(100)	Enterobacteriales(100)
55	Proteobacteria(100)	Alphaproteobacteria(100)	Rhizobiales(100)
56	Proteobacteria(100)	Gammaproteobacteria(100)	Pseudomonadales(100)
57	Bacteroidetes(100)	Flavobacteriia(100)	Flavobacteriales(100)
59	Proteobacteria(100)	Gammaproteobacteria(100)	Enterobacteriales(100)
61	Proteobacteria(100)	Gammaproteobacteria(100)	Pseudomonadales(100)
62	Bacteroidetes(100)	Flavobacteriia(100)	Flavobacteriales(100)
65	Proteobacteria(100)	Alphaproteobacteria(100)	Sphingomonadales(100)
66	Firmicutes(100)	Bacilli(100)	Bacillales(100)
68	Actinobacteria(100)	Actinobacteria(100)	Micrococcales(100)
69	Proteobacteria(100)	Gammaproteobacteria(100)	Pseudomonadales(100)
70	Proteobacteria(100)	Gammaproteobacteria(100)	Enterobacteriales(100)
71	Firmicutes(100)	Bacilli(100)	Bacillales(100)
73	Proteobacteria(100)	Alphaproteobacteria(100)	Rhizobiales(100)
74	Proteobacteria(100)	Alphaproteobacteria(100)	Caulobacterales(100)
76	Actinobacteria(100)	Actinobacteria(100)	Micrococcales(100)
77	Proteobacteria(100)	Gammaproteobacteria(100)	Enterobacteriales(100)
79	Proteobacteria(100)	Gammaproteobacteria(100)	Pseudomonadales(100)
81	Bacteroidetes(100)	Sphingobacteriia(100)	Sphingobacteriales(100)
82	Proteobacteria(100)	Gammaproteobacteria(100)	Xanthomonadales(100)
83	Bacteroidetes(100)	Sphingobacteriia(100)	Sphingobacteriales(100)
85	Proteobacteria(100)	Gammaproteobacteria(100)	Enterobacteriales(100)
86	Bacteroidetes(100)	Flavobacteriia(100)	Flavobacteriales(100)
88	Proteobacteria(100)	Alphaproteobacteria(100)	Rhodospirillales(100)
89	Proteobacteria(100)	Gammaproteobacteria(100)	Enterobacteriales(100)

93	Proteobacteria(100)	Gammaproteobacteria(100)	Pseudomonadales(100)
94	Proteobacteria(100)	Betaproteobacteria(100)	unclassified(99)
96	Bacteroidetes(100)	Sphingobacteriia(100)	Sphingobacteriales(100)
97	Firmicutes(100)	Bacilli(100)	Bacillales(100)
98	Proteobacteria(100)	Alphaproteobacteria(100)	Caulobacterales(100)
106	Actinobacteria(100)	Actinobacteria(100)	Streptosporangiales(100)
115	Actinobacteria(100)	Actinobacteria(100)	Micrococcales(100)
116	Proteobacteria(100)	Betaproteobacteria(100)	Burkholderiales(100)
118	Proteobacteria(100)	Gammaproteobacteria(100)	Xanthomonadales(100)
123	Proteobacteria(100)	Gammaproteobacteria(100)	Enterobacteriales(100)
126	Proteobacteria(100)	Alphaproteobacteria(100)	Caulobacterales(100)
130	Actinobacteria(100)	Actinobacteria(100)	Micrococcales(100)
134	Proteobacteria(100)	Alphaproteobacteria(100)	Rhodospirillales(100)
136	Proteobacteria(100)	Gammaproteobacteria(100)	Pseudomonadales(100)
150	Proteobacteria(100)	Alphaproteobacteria(100)	Rhizobiales(100)
167	Actinobacteria(100)	Actinobacteria(100)	Micrococcales(100)
169	Actinobacteria(100)	Actinobacteria(100)	Frankiales(100)
174	Proteobacteria(100)	Betaproteobacteria(100)	Burkholderiales(100)
189	Proteobacteria(100)	Gammaproteobacteria(65)	Pseudomonadales(65)
190	Armatimonadetes(100)	Fimbriimonadia(100)	Fimbriimonadales(100)
208	Bacteroidetes(100)	Sphingobacteriia(100)	Sphingobacteriales(100)
213	Bacteroidetes(100)	Flavobacteriia(100)	Flavobacteriales(100)
215	Bacteroidetes(100)	Sphingobacteriia(100)	Sphingobacteriales(100)
222	Bacteroidetes(100)	Sphingobacteriia(100)	Sphingobacteriales(100)
223	Bacteroidetes(100)	Sphingobacteriia(100)	Sphingobacteriales(100)
293	Proteobacteria(100)	Alphaproteobacteria(100)	Rhodospirillales(100)
323	Bacteroidetes(100)	Sphingobacteriia(100)	Sphingobacteriales(100)

OTU	Family	Genus
1	Enterobacteriaceae(100)	Pantoea(98)
2	Pseudomonadaceae(100)	Unclassified(93)
3	Oxalobacteraceae(100)	Unclassified(97)
4	Pseudomonadaceae(100)	Pseudomonas(100)
5	Sphingomonadaceae(99)	Sphingomonas(99)
6	Xanthomonadaceae(100)	Xanthomonas(100)
8	Pseudomonadaceae(100)	Pseudomonas(100)
9	Enterobacteriaceae(100)	Unclassified(100)
10	Pseudomonadaceae(100)	Pseudomonas(100)
11	Enterobacteriaceae(100)	Unclassified(98)
12	Family_XII(100)	Exiguobacterium(100)
13	Oxalobacteraceae(100)	Massilia(99)
14	Enterobacteriaceae(100)	Unclassified(100)
15	Caulobacteraceae(100)	Brevundimonas(100)
16	Bacillaceae(100)	Bacillus(100)
17	Sphingomonadaceae(100)	Novosphingobium(100)
18	Oxalobacteraceae(100)	Massilia(100)
19	Enterobacteriaceae(100)	Rahnella(56)
20	Sphingobacteriaceae(100)	Pedobacter(100)
21	Pseudomonadaceae(100)	Pseudomonas(100)
22	Xanthomonadaceae(100)	Stenotrophomonas(100)
23	Xanthomonadaceae(100)	Stenotrophomonas(99)
24	Microbacteriaceae(100)	Curtobacterium(100)
25	Bacillaceae(100)	Bacillus(100)
26	Microbacteriaceae(100)	Unclassified(100)
27	Methylobacteriaceae(100)	Methylobacterium(100)
28	Enterobacteriaceae(100)	Unclassified(67)
29	Oxalobacteraceae(100)	Massilia(100)
32	Methylobacteriaceae(100)	Methylobacterium(100)
33	Sanguibacteraceae(100)	Sanguibacter(100)
34	Microbacteriaceae(100)	Unclassified(91)
35	Methylobacteriaceae(100)	Methylobacterium(100)
36	Burkholderiaceae(100)	Ralstonia(100)
37	Enterobacteriaceae(100)	Pantoea(97)
38	Enterobacteriaceae(100)	Unclassified(98)
39	Chitinophagaceae(100)	Sediminibacterium(100)

40	Sphingobacteriaceae(100)	Pedobacter(100)
41	Sphingobacteriaceae(100)	Pedobacter(100)
42	Oxalobacteraceae(100)	Duganella(100)
43	Bacillaceae(100)	Bacillus(99)
45	Paenibacillaceae(100)	Paenibacillus(100)
46	Streptomycetaceae(100)	Streptomyces(100)
47	Family_XII(100)	Exiguobacterium(100)
48	Rhizobiaceae(100)	Rhizobium(59)
49	Unclassified(79)	Unclassified(79)
50	Streptomycetaceae(100)	Streptomyces(76)
52	Unclassified(98)	Unclassified(98)
53	Enterobacteriaceae(100)	Unclassified(90)
55	Methylobacteriaceae(100)	Methylobacterium(100)
56	Pseudomonadaceae(100)	Pseudomonas(100)
57	Flavobacteriaceae(100)	Chryseobacterium(100)
59	Enterobacteriaceae(100)	Unclassified(100)
61	Pseudomonadaceae(100)	Unclassified(83)
62	Flavobacteriaceae(100)	Chryseobacterium(100)
65	Sphingomonadaceae(100)	Unclassified(94)
66	Paenibacillaceae(100)	Paenibacillus(100)
68	Micrococcaceae(100)	Pseudarthrobacter(94)
69	Pseudomonadaceae(100)	Pseudomonas(100)
70	Enterobacteriaceae(100)	Unclassified(98)
71	Paenibacillaceae(100)	Paenibacillus(100)
73	Bradyrhizobiaceae(100)	Bradyrhizobium(68)
74	Caulobacteraceae(100)	Caulobacter(100)
76	Microbacteriaceae(100)	Rathayibacter(98)
77	Enterobacteriaceae(100)	Unclassified(98)
79	Moraxellaceae(100)	Acinetobacter(100)
81	Sphingobacteriaceae(100)	Pedobacter(100)
82	Xanthomonadaceae(100)	Koukoulia(100)
83	Sphingobacteriaceae(100)	Sphingobacterium(100)
85	Enterobacteriaceae(100)	Unclassified(75)
86	Flavobacteriaceae(100)	Flavobacterium(100)
88	Acetobacteraceae(100)	Roseomonas(100)
89	Enterobacteriaceae(100)	Unclassified(96)
93	Pseudomonadaceae(100)	Pseudomonas(86)
94	Unclassified(99)	unclassified(99)

96	Sphingobacteriaceae(100)	Pedobacter(100)
97	Paenibacillaceae(100)	Saccharibacillus(100)
98	Caulobacteraceae(100)	Brevundimonas(100)
106	Thermomonosporaceae(100)	Actinoallomurus(100)
115	Microbacteriaceae(100)	Curtobacterium(92)
116	Alcaligenaceae(100)	Verticia(100)
118	Xanthomonadaceae(100)	Stenotrophomonas(100)
123	Enterobacteriaceae(100)	Unclassified(94)
126	Caulobacteraceae(100)	Unclassified(82)
130	Micrococcaceae(100)	Paenarthrobacter(67)
134	Incertae Sedis(100)	Reyranella(100)
136	Pseudomonadaceae(100)	Pseudomonas(80)
150	Methylobacteriaceae(100)	Methylobacterium(100)
167	Micrococcaceae(100)	Arthrobacter(95)
169	Sporichthyaceae(100)	Unclassified(100)
174	Comamonadaceae(100)	Acidovorax(96)
189	Pseudomonadaceae(65)	Unclassified(65)
190	Fimbriimonadaceae(100)	Unclassified(99)
208	Chitinophagaceae(100)	Heliimonas(100)
213	Flavobacteriaceae(100)	Epilithonimonas(99)
215	Sphingobacteriaceae(100)	Pedobacter(100)
222	Chitinophagaceae(100)	Vibrionimonas(100)
223	Sphingobacteriaceae(100)	Pedobacter(100)
293	Acetobacteraceae(100)	Unclassified(100)
323	Chitinophagaceae(100)	Sediminibacterium(100)

Appendix C: Taxonomic identity of constructed community in Chapter 4

Isolates were identified through Sanger Sequencing of 16S rRNA gene amplified using universal 27F/1492R primers and NCBI's BLAST tool.

Family	BLAST Description
Micrococcaceae	Nesterenkonia sp. MCCC 1A09847 16S ribosomal RNA gene, partial sequence
Pseudomonadaceae	Pseudomonas oryzihabitans strain USDA-ARS-USMARC-56511, complete genome
Brevibacteriaceae	Brevibacterium frigoritolerans strain GMXG-12 16S ribosomal RNA gene, partial sequence
Bacillaceae	Bacillus sp. strain nenu DS-R06 16S ribosomal RNA gene, partial sequence
Microbacteriaceae	Microbacterium sp. HBUM178885 16S ribosomal RNA gene, partial sequence
Nocardiaceae	Rhodococcus erythropolis strain FS49 16S ribosomal RNA gene, partial sequence
Enterobacteriaceae	Pantoea agglomerans strain C410P1, complete genome
Micrococcaceae	Arthrobacter sp. S101 16S ribosomal RNA gene, partial sequence
Micrococcaceae	Pseudarthrobacter defluvii strain 52-OD12 16S ribosomal RNA gene, partial sequence
Microbacteriaceae	Pseudoclavibacter sp. JSM 2175001 16S ribosomal RNA gene, partial sequence
Microbacteriaceae	Rathayibacter festucae strain VKM Ac-2595 16S ribosomal RNA gene, partial sequence
Micrococcaceae	Paenarthrobacter nitroguajacolicus strain SIIA_Pb_R4 16S ribosomal RNA gene, partial sequence
Bacillaceae	Bacillus sp. strain nenu DS-R05 16S ribosomal RNA gene, partial sequence
Brevibacteriaceae	Brevibacterium sp. strain sz10 16S ribosomal RNA gene, partial sequence
Pseudomonadaceae	Pseudomonas koreensis strain YR-25 16S ribosomal RNA gene, partial sequence
Pseudomonadaceae	Pseudomonas fluorescens strain A528 16S ribosomal RNA gene, partial sequence
Pseudomonadaceae	Pseudomonas moraviensis strain WW1 16S ribosomal RNA gene, partial sequence
Bacillaceae	Bacillus sp. MNPK-3 16S ribosomal RNA gene, partial sequence
Microbacteriaceae	Curtobacterium flaccumfaciens strain ZSR5 16S ribosomal RNA gene, partial sequence
Nocardiaceae	Rhodococcus erythropolis strain 5WK 16S ribosomal RNA gene, partial sequence
Bacillaceae	Exiguobacterium undae gene for 16S rRNA, partial sequence, strain: Su-1