INTEGRATIVE GENOMICS TO STUDY THE PLANT MICROBIOME

Isai Salas-González

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum in Bioinformatics and Computational Biology

Chapel Hill
2021

Approved by:

Jeffery L. Dangl

Corbin D. Jones

Terrence Furey

Di Wu

Scott Gifford

ABSTRACT

Isai Salas-González: Integrative genomics to study the plant microbiome
(Under the direction of Jeffery L. Dangl)

Plants associate with a myriad of microorganisms on the basis of environmental and plant derived cues resulting in a broad spectrum of plant-microbe interactions ranging from mutualism to pathogenicity. These microorganisms, collectively term the plant microbiota, influence plant phenotypes and ultimately fitness. Widely supported trends have emerged based on high-throughput culture independent exploratory surveys: plant habitats exert a selective bottleneck for microbes in relation to the surrounding soil, the enrichment of microbes within the plant is phylogenetically constrained and plant-associated microbes influence host phenotypes (Chapter 1).

The work in this dissertation exploits the increasing availability of plant-associated bacterial isolates to infer the molecular mechanisms modulating plant-microbiome interactions via a combination of multi-omics analyses. First, by developing and deploying a phylogenomic and statistical framework in conjunctions with thousands of bacterial isolates, we identified novel traits associated with bacterial adaptation to plant lifestyle (Chapter 2). Afterwards, by employing a multi-omics profiled synthetic ecology approach, we explored the influence of abiotic stresses over the plant microbiome assembly (Chapter 3) and the discovered dominant microbe-microbe interactions responsible for the modulation of the plant host root growth (Chapter 4). Finally, as an epilogue (Chapter 5), I provide a perspective about the current challenges and future directions concerning plant microbiome research.

A ellas, que son mi soporte, Rosui, Tota y Borre

ACKNOWLEDGEMENTS

This dissertation was only possible thanks to the cumulative effort of an excellent team of scientists assembled by Prof. Jeff Dangl in his research group. I want to thank all the laboratory members of the Dangl group, that through their scientific insight have nourished me to develop as the scientist I am today.

I want to thank my advisor, Prof. Jeff Dangl, for his inexhaustible trust in me and his exemplary passion for discovery and basic research. As well, I want to thank my other advisor, Prof. Corbin Jones, for his always positive encouragement and advice through my PhD. I also thank the rest of my thesis committee: Prof. Terry Furey, Prof. Di Wu, and Prof. Scott Gifford, for dedicating some of their valuable time to me, and for their thoughtful comments and advice.

I also want to thank my dear friends: Dr. Omri Finkel and Dr. Gabriel Castrillo, for showing me the value of collaboration, effort, resilience, discipline, and friendship. Thank you for letting me grow along the road with you guys.

I thank my PhD program, the Bioinformatics and Computational Biology Curriculum, for the opportunity they gave me to pursue a doctoral degree.

Primarily, I want to thank my family for their infinite love, trust and encouragement that allowed me to reach this stage of my life. Muchas gracias a mi madre, María de Lourdes González Beltrán que me crío con ese infinito amor heredado de mis chamanas: Josefina Sedano Morales y Ana María Beltrán Herrera. Muchas gracias a mi tía (que es mi madre también), Ana María Ramírez Reyes, quien tantas veces ha sido mi ángel guardián y a quien le debo una niñez integral

y llena de risas. Muchas gracias a mi padre, Silvestre Salas Chávez, por haberme dado bases sólidas y por siempre haber estado ahí para escucharme. Muchas gracias a mi tía Carmen, mi tío Jaime y mis dos primos Josué y Abimael, por haberme acogido con ese calor de hogar único.

Finalmente, quiero agradecer a mi Rosui, por ser la chibi de este oh

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: THE PLANT MICROBIOME (FROM ECOLOGY TO REDUCTIONISM)[1]

Introduction

The causes and consequences of plant-associated microbial variation have been the subject of intense study for over a century. Following the discovery that atmospheric nitrogen is fixed by bacteria residing in leguminous root nodules came the understanding that plants are associated with an abundance of diverse microbes. Hypotheses that arose in that period are fundamental to the field to this day. Among them are the notions articulated by Lorenz Hiltner (Bulgarelli et al., 2012)(Hartmann et al., 2008): that plant-derived nutrients attract beneficial microbiota in a species-specific manner and that this mechanism is exposed to exploitation by pathogens.

For over a century, the field relied on culture-dependent approaches to illuminate and study the multitude of plant microbiome inhabitants, which include fungi, bacteria, protists, and viruses. However, the stunning extent and distribution of this diversity revealed by culture-independent and high-throughput molecular approaches over the past two decades have had a transformative effect on our understanding, study, and application of plant-microbe research.

1.1: Ecological and evolutionary processes shaping the plant microbiome

The era of culture-independent, high-throughput plant microbiome study began with an exploratory phase of the diversity and composition of microbial taxa and genes across plant habitats. Primarily based on marker-based census data, these studies confirmed the century-old

[1] A portion of the results presented in this chapter appeared as an article in:
Fitzpatrick CR*, Salas-González I*, Conway JM, Finkel OM, Gilbert S, Russ D, Teixeira PJPL, Dangl JL. The Plant Microbiome: From Ecology to Reductionism and Beyond. Annu Rev Microbiol. 2020 Sep 8;74:81-100.

hypothesis that plants harbor distinct microbiota, which represent a subset of those found in the ambient environment (Bulgarelli et al., 2012; Lundberg, Lebeis, Paredes, Yourstone, Gehring, Malfatti, Tremblay, Engelbrektson, Kunin, del Rio, et al., 2012). This led to the multistep model of plant microbiome assembly, whereby specific microbes in the environment are recruited to plant surfaces, followed by additional filtering as microbial taxa colonize the interior of plant organs (Bulgarelli et al., 2013)(Edwards et al., 2015). Such large effects of plant habitats on microbial communities led to the hypothesis that plant-associated microorganisms are adapted to the unique environments provided by their plant hosts.

To date studies, besides conforming to this paradigm, have revealed that additional ecological and evolutionary processes can strongly shape plant microbiota.

1.1.1: Selection

Plants are not homogenous microbial habitats. Different plant habitats such as leaves (J. H. Leveau, 2019), roots (Pascale et al., 2020), or flowers (Rebolleda-Gómez et al., 2019) typically harbor unique microbiota ((Bai et al., 2015; Bodenhausen et al., 2013); but see (Meaden et al., 2016)). This is due to variation across plant habitats in plant-derived resources as well as physical and chemical properties resulting from structural differences and exposure to different features of the environment. For example, roots and leaves impose different selection on microbiota due to both their structural differences and exposure to soil versus air, respectively (Bai et al., 2015; Grinberg et al., 2019). After plant habitat, variation in the abiotic and biotic environment can exert large direct and indirect effects on plant-associated microbiota. Climate-driven geographical variation in soil microorganisms (Bahram et al., 2018) can drive the composition of plant microbiota due to the predominance of horizontal transmission (Liu et al., 2019). Alternatively, environmental variation, both abiotic (Gallart et al., 2018) and biotic (Humphrey & Whiteman,

2020), can indirectly shape plant microbiota through plant responses. Interactions among microbes can also have large effects on community composition, in which the presence of particular microbial groups or even single taxa alter the plant microbiome via both antagonistic and beneficial interactions (Carrión et al., 2019; Durán et al., 2018; Snelders et al., 2020; Uroz et al., 2019; X. Wang et al., 2019). Finally, numerous studies demonstrate that variation (Bowsher et al., 2020; Cregger et al., 2018) and across (Fitzpatrick et al., 2018; Koyama et al., 2019; Naylor et al., 2017; Tkacz et al., 2020) host plant species can affect both the diversity and composition of plant microbiota. However, the extent to which variation across host plants shapes microbiota seems to vary across environments (Bowen et al., 2017; Bowsher et al., 2020; Veach et al., 2019) and habitats within (Massoni et al., 2020; Rochefort et al., 2019).

1.1.2: Ecological drift

Ecological drift—stochastic variation in growth and death—can be a potent driver of community composition under certain scenarios, notably when communities have few species and exhibit low overall abundance and selection is weak (B. Gilbert & Levine, 2017; Zhou & Ning, 2017). Endophytic communities found within plant organs may be particularly prone to ecological drift due to their low overall abundance compared to epiphytic communities, especially during early plant development (Edwards et al., 2015; Z. Gao et al., 2019). Stochastic changes in the relative abundance of individual species within a community can have large downstream effects on community composition when coupled with selection, such as altered interaction strengths among community members. For example, increasing the relative abundance of six randomly chosen members of a soil microbial community led to compositional changes consistent with competitive exclusion between the increased member and close relatives in the community (Zhao et al., 2019). The high levels of unexplained variance in plant microbiome composition and

diversity, as well as the abundance of individual taxa, are, at least in part, due to drift (Maignien et al., 2014).

1.1.3: Dispersal

While selection and drift alter the abundance of existing members within a community, dispersal and diversification are how new species arise in communities. Plant-colonizing microorganisms disperse from the surrounding environment (i.e., horizontal transmission), including soil (Liu et al., 2019), neighboring plants and interacting animals (Rebolleda Gómez & Ashman, 2019), and the air column (Choudoir et al., 2018), although vertical transmission via seed can also occur (Newcombe et al., 2018; Shade et al., 2017). After initial colonization, microbial dispersal to and from plants is likely to occur. Yet despite its omnipresence, dispersal is a difficult process to quantify; this is especially true for microorganisms. The importance of dispersal is inferred by correlating the compositional similarity between two communities with the physical distance separating them, although distance can be confounded by known or unknown environmental factors. A negative correlation between distance and similarity (i.e., distance-decay) suggests that microbial dispersal limitation may contribute to compositional differences between communities. Distance-decay studies of plant microbiota show that the importance of dispersal as a driver of community composition will vary according to the spatial scale, plant habitat, and microbial taxa under (Amend et al., 2019; Donald et al., 2020; Finkel et al., 2012; Hurtado-McCormick et al., 2019; Laforest-Lapointe et al., 2016; Meaden et al., 2016). Researchers are also beginning to use experimental approaches to evaluate the importance of dispersal in the plant microbiome. Experimentally reducing dispersal among floral nectar microbial communities increased compositional similarities between communities (Vannette & Fukami, 2017). Experimental studies show that bacterial dispersal in soil is restricted to a narrow taxonomic

breadth of organisms, which could have direct impacts on the colonization of plant roots (Krüger et al., 2018; Wolf et al., 2015). Dispersal can also shape plant microbiota through the introduction of priority effects, whereby the order of arrival among microorganisms can either facilitate or inhibit the success of future colonists (Carlström et al., 2019). Furthermore, changes in environmental variables such as resource availability and pH can alter the strength of priority effects (Grainger et al., 2019).

1.1.4: Evolution

Evolutionary change can shape the growth of individual populations and the interactions between species and can ultimately give rise to new species, all of which can affect community composition (Gómez et al., 2016; Pelletier et al., 2009). Direct evidence of evolutionary change occurring within plant microbiota, let alone shaping community composition, is rare outside well-known symbionts (Porter et al., 2017). Evolved resistance to phage reduced the ability of a bacterial pathogen to proliferate on its host plant, which in turn led to altered composition of plant microbiota (X. Wang et al., 2019). Indeed, the evolutionary outcomes of multitrophic interactions within the plant microbiome may play a much larger role in shaping community diversity and composition than is currently recognized (Z. Gao et al., 2019; Koskella & Taylor, 2018; Morella et al., 2018). The interaction between evolutionary dynamics and dispersal may also have large effects on plant microbiomes. Conceivably, a beneficial mutation arising in one microbial population could spread via dispersal to others, either within or between plant hosts, potentially altering microbe-microbe interactions and ultimately community composition (Miller et al., 2018). Although the above examples include evolution by natural selection, genetic drift occurring within microbial populations could also lead to evolutionary change, with consequences for ecological dynamics.

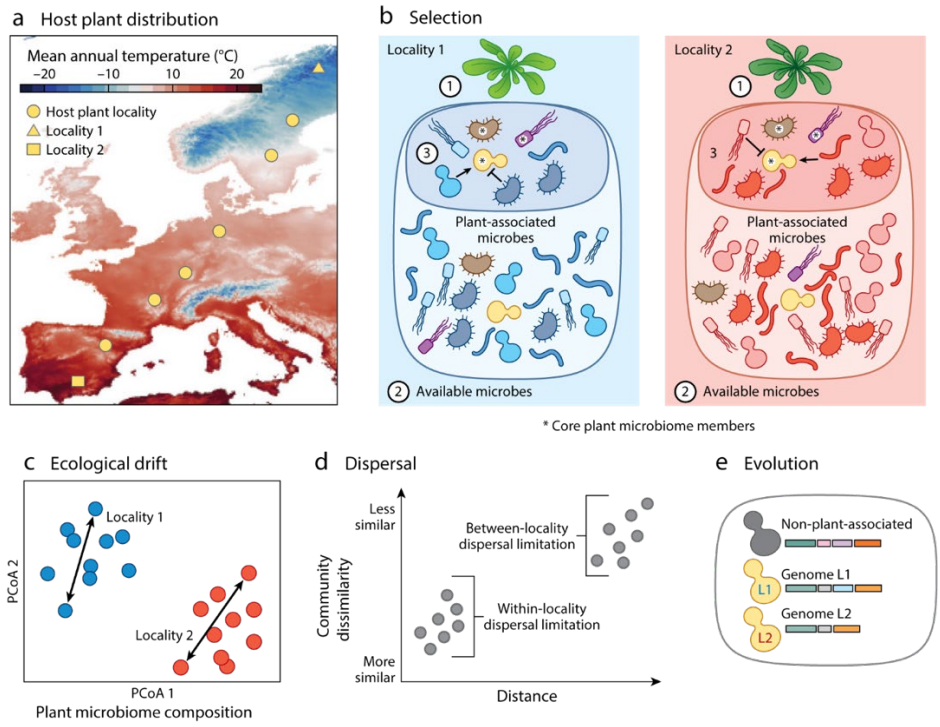


Figure 1.1 - **Ecological and evolutionary processes shaping the plant microbiome.**

An illustrative example of the ecological and evolutionary processes shaping the plant microbiota. **a)** Plants and their microbiomes occur across large geographic areas that vary in abiotic and biotic environmental factors. Here, we focus on two localities found at the extremes of a temperature gradient. **b)** Selection imposed by numerous factors (1 –3 in the figure) can shape the diversity and composition of plant microbiota. (1) Microbes found in the environment are winnowed during colonization and assemble unique plant-associated microbiomes, a fraction of which are found at high occurrence across host plants and localities (taxa denoted by asterisks in panel b). Moreover, host plant variation across localities as a result of responses to temperature or other environmental factors (indicated by different host plant color in panel b), genetically distinct populations, or different plant species can also shape plant microbiota. (2) Geographic location and corresponding environmental features determine the pool of microbes available for plant colonization. (3) Microbe-microbe interactions will likely vary across localities, in turn driving compositional differences in plant microbiota. In addition to selection, plant microbiota is also shaped by (c) drift, (d) dispersal, and (e) evolution. **c)** Drift results from stochastic differences in the growth rates of individual plant microbiome members and can cause plant microbiota to compositionally diverge despite exposure to the same environments. Each point in panel c depicts the composition of a sampled plant microbiome, and the distance between two points reflects compositional similarity. Although location has a clear effect on plant microbiota, samples within a locality still exhibit variation despite being collected from the same host plant species in the same environment. Some of this variation will be caused by ecological drift. **d)** Dispersal can also contribute to compositional differences observed among plant microbiota across spatial scales. Microbial

dispersal can occur over small and large spatial extents, giving rise to compositional differences between pairs of plant microbiota that scale with distance. **e)** Evolutionary change occurring in individual plant microbiome members can lead to altered population growth rates and species interactions (as depicted in panel b), both of which could lead to shifts in plant microbiota. In addition to elucidating its role in shaping plant microbiota, research on evolutionary change in the plant microbiome has led to important discoveries of adaptations to a plant-associated lifestyle. For example, genetic differences between plant-associated and free-living relatives reveal microbial adaptations to life with plants (as shown in panel e). However, that microbial adaptation might be occurring to other features of the local plant-associated environment, including host plant variation, the abiotic environment, and resident microbiota (as depicted in 1 –3 in panel b). Genomic analysis of core microbiome members across these factors could reveal adaptations that are otherwise hidden by the cryptic diversity within amplicon sequence variants or operational taxonomic units.

1.2: Environmental variation shapes the mechanistic interactions between plants and microbiota

Plants and their microbiomes must cope with environmental variation, including changing temperature, light, humidity, soil chemistry, and water availability. This environmental variation often leads to shifts in the diversity or composition of plant microbiota (Caddell et al., 2019; Cheng et al., 2019; Hiruma, 2019). While these environmentally induced shifts can be due to direct microbial responses, they are often indirectly caused by plant responses (Carvalhais et al., 2013), which can shift plant microbiota through the mechanisms described above. For example, in *Arabidopsis*, phosphorus starvation alters the root microbiome through the phosphate starvation response, which is integrated with the immune system through the master transcriptional regulator, phosphate starvation response 1 (*phr1*), and its effects on the jasmonic and salicylic acid pathways (Castrillo et al., 2017; Hiruma et al., 2016; Morcillo et al., 2020). Secondary metabolites produced by plants under various environmental stresses, including iron (Stringlis et al., 2018) and phosphate (Hiruma et al., 2016) limitation, can also shape the plant microbiome by either selectively enriching or inhibiting particular members. Plant responses to abiotic stress can also be linked with microbiota through shared signaling components, such as salt stress tolerance, which

is linked through the chitin receptor, chitin elicitor receptor kinase 1 (Espinoza et al., 2017). Plant exudation and bacterial uptake of the metabolite glycerol-3-phosphate was strongly associated with the enrichment of Actinobacteria in the Sorghum bicolor root under drought stress (L. Xu et al., 2018). Moreover, the cross talk between abiotic and biotic environmental responses can change over the course of plant development, in turn shifting the composition of plant microbiota (Berens et al., 2019). Lastly, insect herbivory can perturb the leaf microbiome through the induction of plant defense responses (Humphrey & Whiteman, 2020). These examples illustrate that variation in both the abiotic and biotic environment can elicit changes in the plant microbiome through plant responses (Figure 1.2).

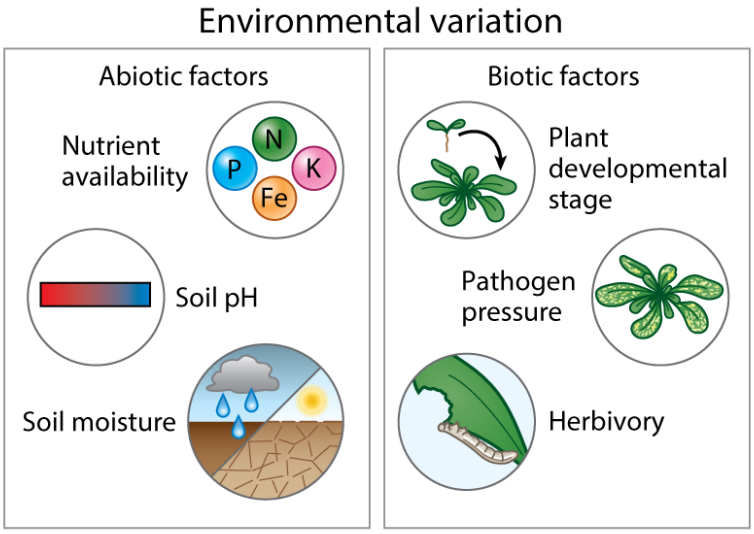


Figure 1.2 - **Exemplary abiotic and biotic factors influencing the plant microbiome**

1.3: Integrating ecological and reductionist approaches for a more mechanistic understanding of the plant microbiome.

The association between the plant microbiome and the increased tolerance of plant hosts to biotic and abiotic stresses (Berendsen et al., 2018; Carrión et al., 2019; Fitzpatrick et al., 2018; Stringlis et al., 2018; L. Xu et al., 2018) has propelled the field into a quest of comprehending the mechanisms behind the distribution of microbes within plant environments. This transition in the

research perspective, from a descriptive to a mechanistic one, has been driven by novel advances including the use of synthetic communities of genome-sequenced microbial strains, dual omics of host plants and microbiota, and increased resolution of microfluidic and microscopy techniques. Along the following chapters, I will illustrate how these new reductionists approaches have proven and will continue to be critical for our understanding of the diversity and composition observed in plant microbiota across host plant habitats and environments (Figure 1.3).

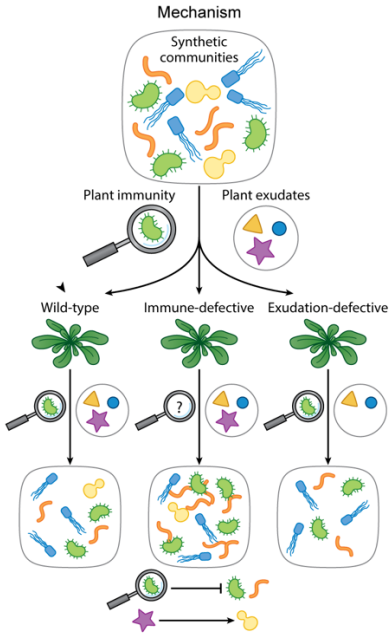


Figure 1.3 - **Scheme showing a synthetic ecology framework to elucidate plant-microbiome mechanisms.**

CHAPTER 2: GENOMIC FEATURES OF BACTERIAL ADAPTATION TO PLANTS[2]

Introduction

The microbiota of plants and animals have coevolved with their hosts for millions of years (Baumann, 2005; Ley et al., 2008; Sprent, 2008). Through photosynthesis, plants serve as a rich source of carbon for diverse bacterial communities. These include mutualists and commensals, as well as pathogens. Phytopathogens and growth-promoting bacteria have considerable effects on plant growth, health, and productivity (Chowdhury et al., 2015; Fibach-Paldi et al., 2012; Pfeilmeier et al., 2016; Santhanam et al., 2015). Except for intensively studied relationships such as root nodulation in legumes (Peters et al., 1986), T-DNA transfer by Agrobacterium (Hiei et al., 1994) , and type III secretion–mediated pathogenesis(Hueck, 1998), the molecular mechanisms that govern plant–microbe interactions are not well understood. It is therefore important to identify and characterize the bacterial genes and functions that help microbes thrive in the plant environment. Such knowledge should improve the ability to combat plant diseases and harness beneficial bacterial functions for agriculture, with direct effects on global food security, bioenergy, and carbon sequestration.

Cultivation-independent methods based on profiling of marker genes or shotgun metagenome sequencing have considerably improved the overall understanding of microbial ecology in the plant environment.

---

[2] A portion of the results presented in this chapter appeared as an article in:
Levy A*, Salas Gonzalez I*, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F, Rangel Alvarez B, Lundberg DS, Lu TY, Lebeis S, Jin Z, McDonald M, Klein AP, Feltcher ME, Rio TG, Grant SR, Doty SL, Ley RE, Zhao B, Venturi V, Pelletier DA, Vorholt JA, Tringe SG, Woyke T, Dangl JL. Genomic features of bacterial adaptation to plants. Nat Genet. 2017 Dec 18;50(1):138-150

In parallel, reduced sequencing costs have enabled the genome sequencing of plant-associated bacterial isolates at a large scale (Bai et al., 2015). Importantly, isolates enable functional validation of in silico predictions. Isolate genomes also provide genomic and evolutionary context for individual genes, as well as the potential to access genomes of rare organisms that might be missed by metagenomics because of limited sequencing depth. Although metagenome sequencing has the advantage of capturing the DNA of uncultivated organisms, multiple 16S rRNA gene surveys have reproducibly shown that the most common plant-associated bacteria are derived mainly from four phyla (Bulgarelli et al., 2013; Hardoim et al., 2015) (Proteobacteria, Actinobacteria, Bacteroidetes, and Firmicutes) that are amenable to cultivation. Thus, bacterial cultivation is not a major limitation in sampling of the abundant members of the plant microbiome (Bai et al., 2015). Our objective was to characterize the genes that contribute to bacterial adaptation to plants (plant-associated genes) and those genes that specifically aid in bacterial root colonization (root associated genes).

2.1 Expanding the plant-associated bacterial reference catalog

Here, we sequenced the genomes of 484 new bacterial isolates and single bacterial cells from the roots of Brassicaceae, maize, and poplar trees. We combined the newly sequenced genomes with existing genomes to create a dataset of 3,837 high-quality, nonredundant genomes. We then developed a computational approach to identify plant-associated genes and root-associated genes based on comparison of phylogenetically related genomes with knowledge of the origin of isolation. We experimentally validated two sets of plant-associated genes, including a previously unrecognized gene family that functions in plant-associated microbe–microbe competition. In addition, we characterized many plant-associated genes that are shared between bacteria of different phyla, and even between bacteria and plant-associated eukaryotes. This study

represents a comprehensive and unbiased effort to identify and characterize candidate genes required at the bacteria–plant interface.

To obtain a comprehensive reference set of plant-associated bacterial genomes, we isolated and sequenced 191, 135, and 51 novel bacterial strains from the roots of Brassicaceae (91% from *Arabidopsis thaliana*), poplar trees (*Populus trichocarpa* and *Populus deltoides*), and maize, respectively (Methods, Table 2.1).

Table 2.1 | Novel and previously sequenced genomes used in this analysis

Taxon	Taxonomic rank	Novel sequenced PA genomes	Scanned genomes	Genomes used in analysis	PA	NPA	Soil	RA
Alphaproteobacteria[1]	Class	126	784	610	368	199	43	169
Burkholderiales[1]	Order	85	612	433	160	209	64	86
Acinetobacteria[1]	Genus	4	926	454	7	442	5	3
Pseudomonas[1]	Genus	75	506	349	169	137	43	61
Xanthomonadaceae[1]	Family	15	264	147	110	26	11	26
Bacillales[2]	Order	54	664	454	97	185	172	54
Actinobacteria 1[3]	NA	69	504	394	164	142	88	89
Actinobacteria 2[3]	NA	19	845	587	29	526	32	18
Bacteroidetes[4]	Phylum	37	481	409	56	293	60	17
Total		484	5,586	3,837	1,160	2,159	518	523

[1]Proteobacteria [2]Firmicutes. [3]Actinobacteria phylum. PA, plant-associated bacteria; NPA, non-plant-associated bacteria; soil, soil-associated bacteria; RA, root-associated bacteria; NA, not available (an artificial taxon).

Table 2.1 - **Taxonomic distribution and environmental classification (Plant-Associated, Non-Plant-Associated, Soil or Root-Associated) of the isolates used in this analysis.**

The bacteria were specifically isolated from the interior (endophytic compartment) or surface (rhizoplane) of plant roots, or from soil attached to the root (rhizosphere). In addition, we isolated and sequenced 107 single bacterial cells from surface-sterilized roots of *A. thaliana*. All genomes were assembled, annotated, and deposited in public databases and in the following dedicated website.
(http://labs.bio.unc.edu/Dangl/Resources/gfobap_website/).

In addition to the newly sequenced genomes noted above, we collected 5,587 bacterial genomes belonging to the four most abundant phyla of plant-associated bacteria13 from public databases (Methods). We manually classified each genome as plant-associated, non-plant-associated (NPA), or soil-derived on the basis of its unambiguous isolation niche (Methods). The plant-associated genomes included organisms isolated from plants or rhizospheres. A subset of the

plant-associated bacteria was also annotated as 'root-associated' when isolated from the rhizoplane or the root endophytic compartment. Genomes from bacteria isolated from soil were considered as a separate group, as it is unknown whether these strains can actively associate with plants. Finally, the remaining genomes were labeled as NPA genomes; these were isolated from diverse sources, including humans, non-human animals, air, sediments, and aquatic environments. We carried out stringent quality control to remove low-quality or redundant genomes (Methods). This led to a final dataset of 3,837 high-quality and nonredundant genomes, including 1,160 plant-associated genomes, 523 of which were also root-associated. We grouped these 3,837 genomes into nine monophyletic taxa to allow comparative genomics analysis among phylogenetically related genomes (Figure 2.1, Methods).

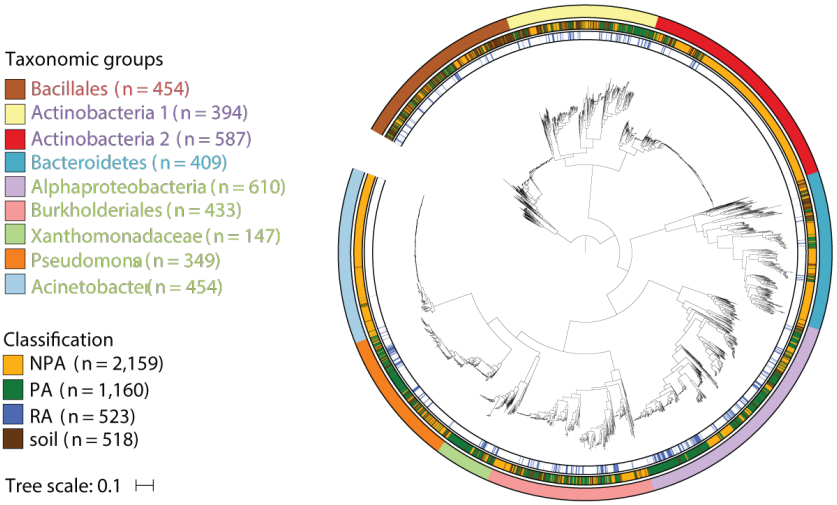


Figure 2.1 - **Maximum-likelihood phylogenetic tree of 3,837 high- quality and nonredundant bacterial genomes.**

The tree was constructed based on a super-matrix inference approach based on the concatenated alignment of 31 single-copy genes distributed across the 3,837 isolates. The outer ring shows the taxonomic group, the central ring shows the isolation source, and the inner ring shows the root-associated (RA) genomes within plant-associated (PA) genomes. Taxon names are color-coded according to phylum: green, Proteobacteria; red, Firmicutes; blue, Bacteroidetes; purple, Actinobacteria.

To determine whether our genome collection from cultured isolates was representative of plant-associated bacterial communities, we analyzed cultivation-independent 16S rDNA surveys and metagenomes from the plant environments of Arabidopsis (Bulgarelli et al., 2012; Lundberg, Lebeis, Paredes, Yourstone, Gehring, Malfatti, Tremblay, Engelbrektson, Kunin, Rio, et al., 2012), barley (Bulgarelli et al., 2015), wheat, and cucumber (Ofek-Lalzar et al., 2014) (Methods). The nine taxa analyzed here account for 33–76% (median, 41%) of the total bacterial communities found in plant-associated environments and therefore represent a substantial portion of the plant microbiota, consistent with previous reports (Bai et al., 2015; Bulgarelli et al., 2013; Hacquard et al., 2015).

2.2 Identification and validation of plant- and root-associated genes

We compared the genomes of bacteria isolated from plant environments with those from bacteria of shared ancestry that were isolated from non-plant environments. We assumed that the two groups should differ in the set of accessory genes that evolved as part of their adaptation to a specific niche. First, employing a phylogenetic aware approach (PhyloGLM) and a naïve approach (t-test) we compared genome functional content at a broad sequence scale (clusters of orthologous groups, COGs) between PA and NPA isolates. (Figure 2.2). Two gene categories demonstrated similar phylogeny-independent trends suggestive of an environment-dependent selection process. The “Carbohydrate metabolism and transport” gene category was expanded in the plant-associated organisms of six taxa (Figure 2.2). This was the most expanded category in Alphaproteobacteria, Bacteroidetes, Xanthomonadaceae, and Pseudomonas In contrast, mobile genetic elements (phages and transposons) were underrepresented in four plant-associated taxa (Fig. 2.2).

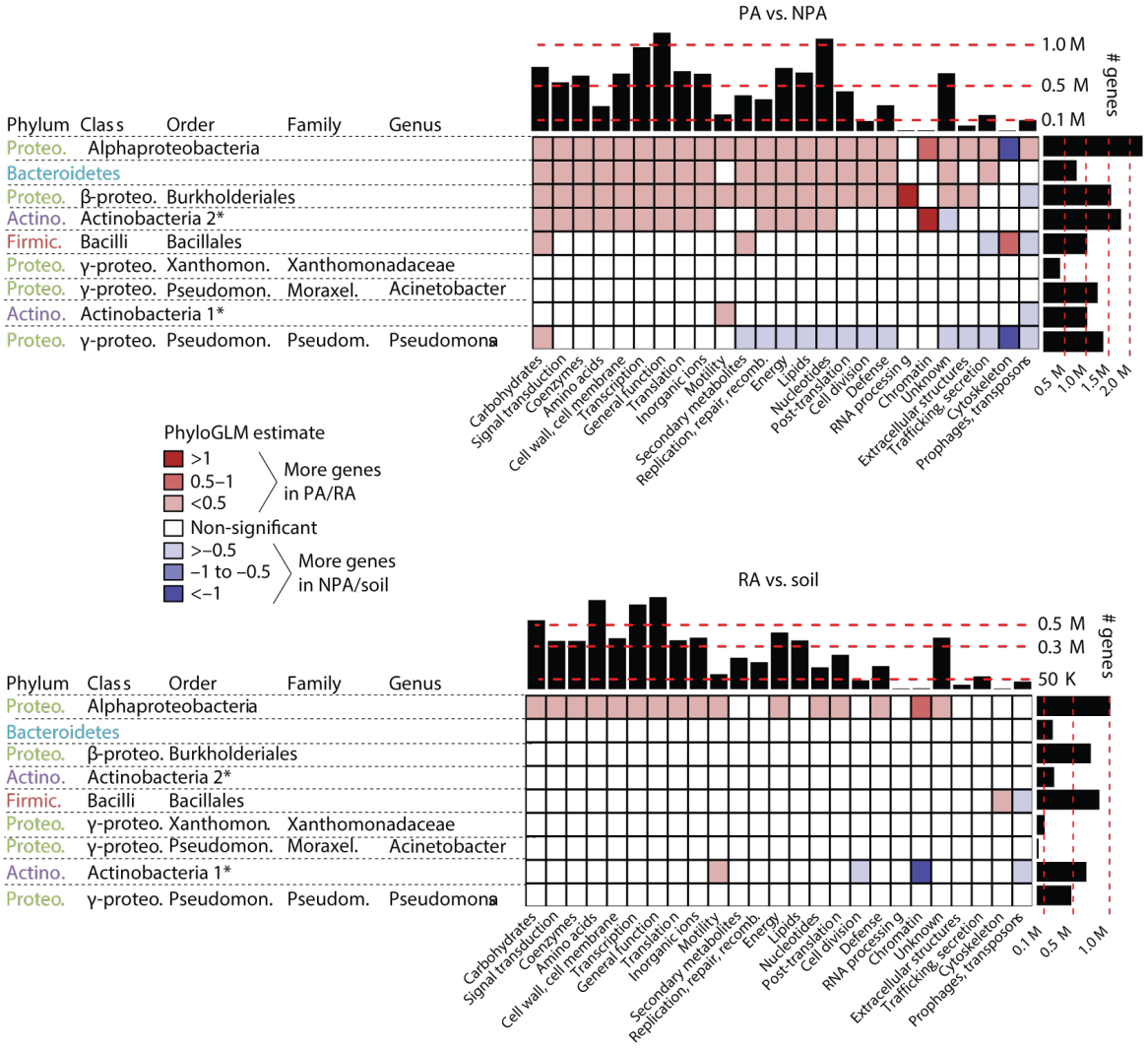


Figure 2.2 - **Differences in functional gene categories (COGs) between plant-associated and non-plant-associated bacteria.**

Both heat maps indicate the level of enrichment or depletion based on a PhyloGLM test. Significant cells (color-coded according to the key) represent P values of < 0.05 (FDR-corrected). Pink-red cells indicate significantly more genes in plant-associated and root-associated genomes in the top and bottom heat maps, respectively. Bar graphs at the top and right of each heat map represent the total number of genes compared in each column and row, respectively. Asterisks indicate non-formal class names. “Carbohydrates” denotes the carbohydrate metabolism and transport gene category. Note that cells representing high absolute estimate values (dark colors) are based on categories of few genes and are therefore more likely to be less accurate. Phylum names are color- coded as in Figure 2.1. Abbreviations: Xanthomon., Xanthomonadales; Pseudomon., Pseudomonadales; Pseudom., Pseudomonadaceae; Moraxel., Moraxellaceae.

Next, using a phylogenomic and statistical pipeline (Figure 2.3), we sought to identify specific genes enriched in plant- and root-associated genomes compared with NPA and soil-derived genomes, respectively (Methods). First, we clustered the proteins and/or protein domains of each taxon on the basis of homology, using the annotation resources COG (Tatusov et al., 2000), KEGG Orthology (Kanehisa et al., 2016), and TIGRFAM (Haft et al., 2003), which typically comprise 35–75% of all genes in bacterial genomes (Huntemann et al., 2015). To capture genes that do not have existing functional annotations, we also used OrthoFinder (Emms & Kelly, 2015) to cluster all protein sequences within each taxon into homology-based orthogroups. Finally, we clustered protein domains with Pfam (Finn et al., 2016) (Methods). We used these five protein/domain-clustering approaches in parallel comparative genomics pipelines. Each protein/domain sequence was additionally labeled as originating from either a plantassociated genome or an NPA genome. Next, we determined whether protein/domain clusters were significantly associated with a plant-associated lifestyle by using five independent statistical approaches: hypergbin, hypergcn (two versions of the hypergeometric test), phyloglmbin, phyloglmcn (two phylogenetic tests based on PhyloGLM (Ives & Garland, 2010)), and Scoary (Brynildsrud et al., 2016) (a stringent combined test) (Methods). These analyses were based on either gene presence/absence or gene copy number. We defined a gene as significantly plant-associated if at least one test showed that it belonged to a significant plant-associated gene cluster, and if it originated from a plant-associated genome. We defined significant NPA, root-associated, and soil genes in the same way.

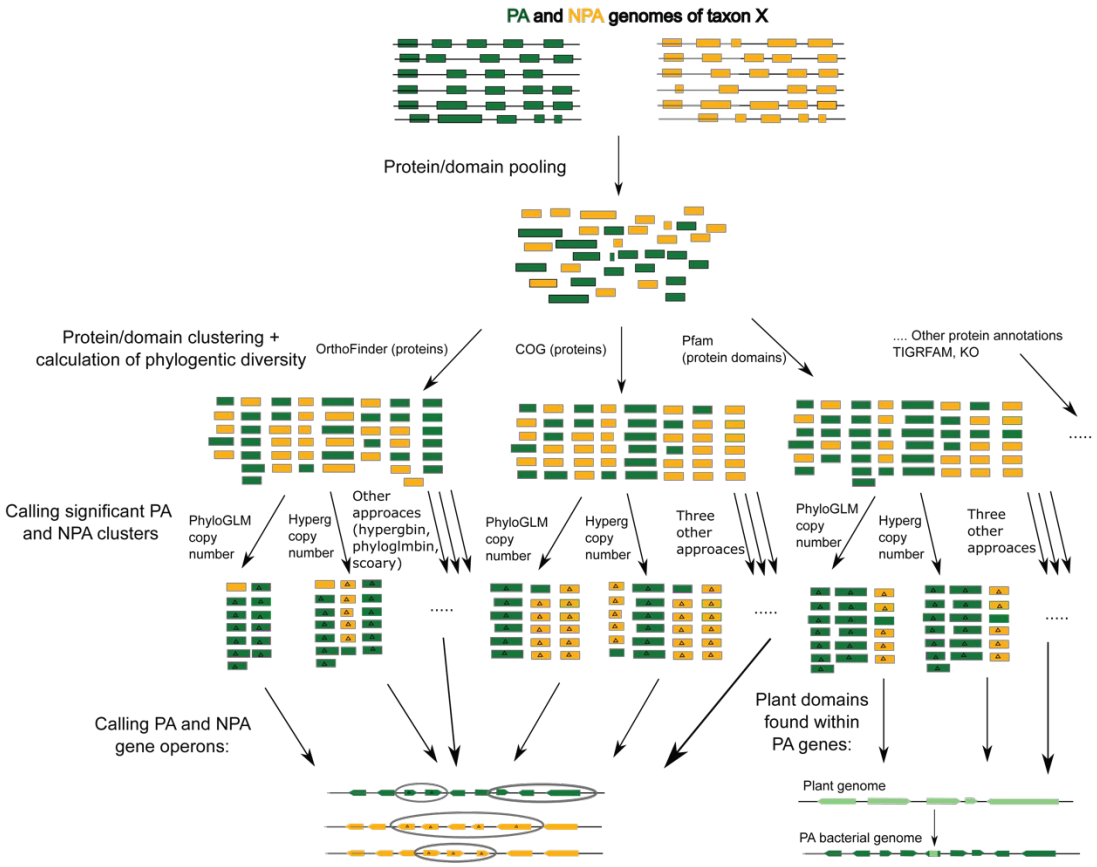


Figure 2.3 - **Overview of the algorithm used to call PA and NPA genes (proteins) and gene operons.**

High quality PA and NPA genomes were collected. All protein and protein domains were retrieved from genomes. Different protein/domain clustering approaches were used based on existing functional annotation (COG, Pfam, TIGRfam, KEGG orthology) or based on running OrthoFinder over all protein coding genes (for simplicity TIGRfam and KEGG orthology were not mentioned in the figure). Note that clusters may contain a combination of orthologous and paralogous genes. Significant PA/NPA clusters (enriched with PA/NPA proteins/domains) were called based on five tests: PhyloGLM and the Hypergeometric test, both gene copy number and gene presence/absence versions (phyloglmcn, phyloglmbin, hypergcn, hypergbin), and Scoary. Genes from PA and NPA genomes in PA and NPA clusters, respectively, are marked with a triangle. Genes from the significant protein clusters (OrthoFinder, COG) were separately used to predict PA/NPA gene operons comprised of nearly exclusively adjacent PA/NPA genes sharing the same orientation. PA Pfam domains were used to search the overlap between those and plant-like protein domains (PREPARADOs).

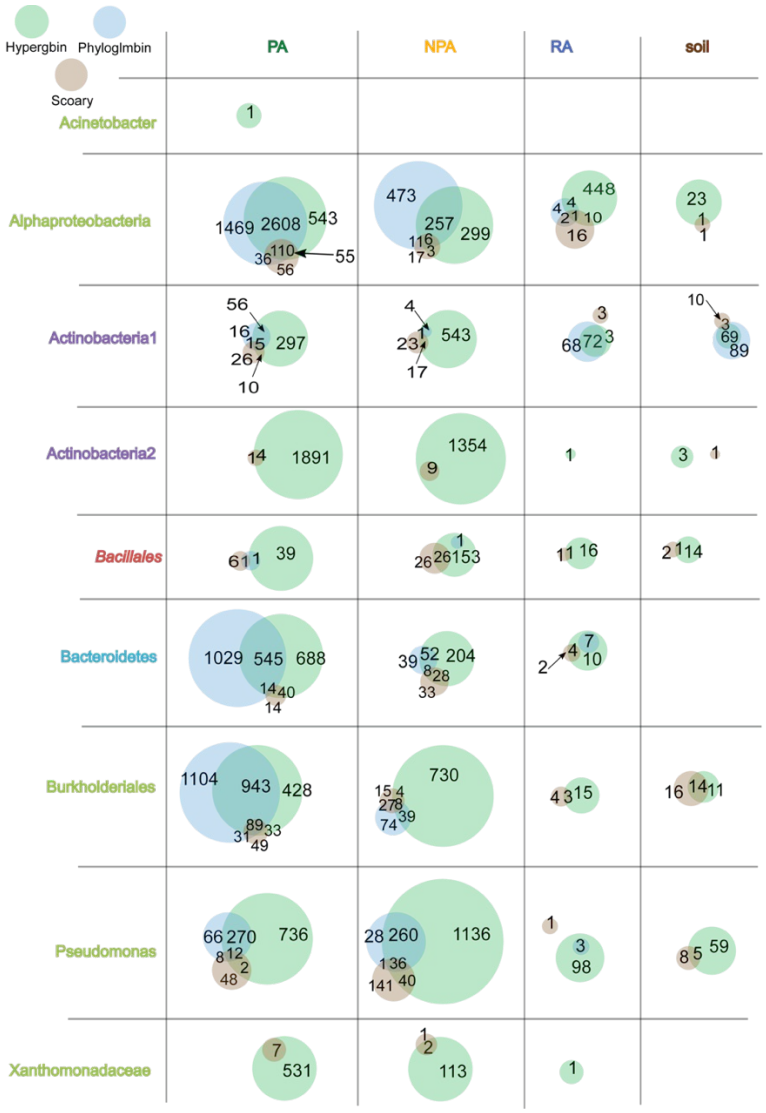


Figure 2.4 - **Overlap of statistically significant genes between the different algorithms to determine enrichment.**

Number of significant orthogroups predicted by the copy number versions of the Hypergeometric test (hypergcn) and PhyloGLM (phyloglmcn), and by Scoary. The numbers represent gene clusters found either in each group separately (where circles do not overlap) or in the overlap between the groups. Hypergcn is likely the most promiscuous yet sensitive approach as it predicts enriched genes in high numbers and does not require a phylogenetic signal (monophyletic genes can be significant in hypergbin). It may lead to many false positive predictions. Phyloglmcn is more stringent than hypergcn but it may be less sensitive than hypergcn as it cannot predict any significant gene in certain taxa that lack sufficiently strong phylogenetic signal (e.g. Actinobacteria2, Xanthomonadaceae). Scoary is probably the most stringent approach that combines a naive statistical test, a phylogenetic test, and label permutations. Therefore, it frequently yields the lowest number of significant predictions.

We provide full lists of statistically significant plant-associated, root-associated, soil-associated, and NPA proteins and domains according to the five clustering techniques and five statistical approaches for each taxon in the following website: http://labs.bio.unc.edu/Dangl/Resources/gfobap_website/enrichments.html.

To validate our predictions, we assessed the abundance patterns of plant-associated and root-associated genes in natural environments. We retrieved 38 publicly available plant-associated, NPA, root-associated, and soil-associated shotgun metagenomes, including some from plant-associated environments that were not used for isolation of the bacteria analyzed here (Hultman et al., 2015; Louca et al., 2016; Ofek-Lalzar et al., 2014). We mapped reads from these culture-independent metagenomes to plant-associated genes found with all statistical approaches (Methods). Plant- associated genes in up to seven taxa were more abundant ($p < 0.05$, t-test) in plant-associated metagenomes than in NPA metagenomes (Figure 2.5a). Root-associated, soil-associated, and NPA genes, in contrast, were not necessarily more abundant in their expected environments (Figure 2.5a).

In addition, we selected eight genes that were predicted to be plant-associated by multiple approaches for experimental validation via an in planta bacterial fitness assay (Methods). We inoculated the roots of surface-sterilized rice seedlings (n = 9–30 seedlings per experiment) with wild-type *Paraburkholderia kururiensis M130* (a rice endophyte (Coutinho et al., 2015)) or a knock- out mutant strain for each of the eight genes. We grew the plants for 11 d and then collected and quantified the bacteria that were tightly attached to the roots (Methods). Mutations in two genes led to fourfold to sixfold reductions in colonization (false discovery rate (FDR)-corrected Wilcoxon rank sum test, $q < 0.1$) relative to that by wild-type bacteria (Figure. 2.5b). These two genes encode an outer-membrane efflux transporter from the nodT family and a Tir chaperone

protein (CesT), respectively. It is plausible that the other six genes assayed function in facets of plant association not captured in this experimental context.

Functions for which co-expression of and cooperation between different proteins are needed are often encoded by gene operons in bacteria. We therefore tested whether our methods could correctly predict known plant-associated operons. We grouped plant- associated and root-associated genes into putative plant-associated and root-associated operons on the basis of their genomic proximity and orientation (Figure 2.3, Methods). This analysis yielded some well-known plant-associated functions, such as those of the *nodABCSUIJZ* and *nifHDKENXQ* operons (Figures 2.5c, d). Nod and Nif proteins are integral for biological nitrogen cycling and mediate root nodulation (Long, 1989) and nitrogen fixation (Ruvkun et al., 1982), respectively. We also identified the biosynthetic gene cluster for the precursor of the plant hormone gibberellin (Hershey et al., 2014; Nett et al., 2017) (Figure 2.5e). Other known plant-associated operons identified are related to chemotaxis (Scharf et al., 2016), secretion systems such as T3SS (Büttner & He, 2009) and T6SS (R. Gao et al., 2015), and flagellum biosynthesis (De Weert et al., 2002; De Weger et al., 1987; Weller-Stuart et al., 2017) (Figure 2.5f–i).

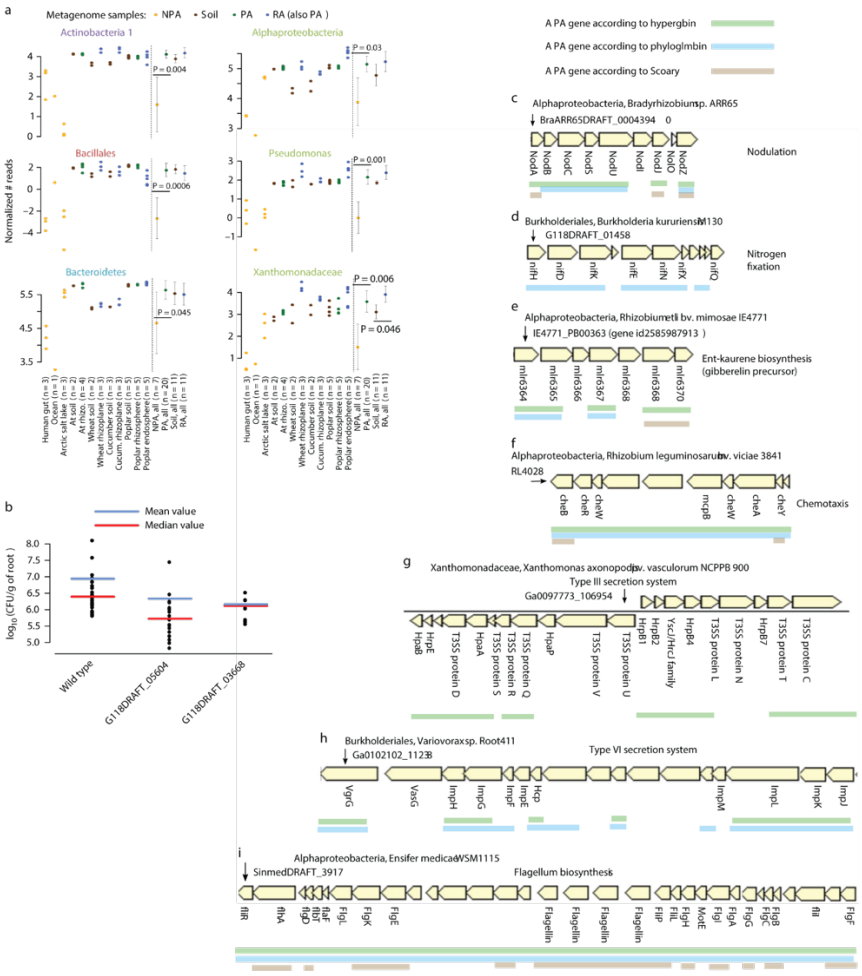


Figure 2.5 - **Validation of predicted plant-associated genes by multiple approaches.**

a) Plant-associated (PA) genes, which were predicted from isolate genomes, were more abundant in PA metagenomes than in NPA metagenomes. Reads from 38 shotgun metagenome samples were mapped to significant PA, NPA, RA, and soil-associated genes predicted by Scoary. P values are indicated for the significant differences between PA and NPA genes or RA and soil-associated genes in each taxon (two-sided t-test). **b)** Results of a rice root colonization experiment using wild-type *Paraburkholderia kururiensis M130* or knockout mutants for two predicted plant-associated genes. Two mutants showed reduced colonization compared with the wild type: G118DRAFT_05604 (*q-value* = 0.00013, Wilcoxon rank sum test), which encodes an outer membrane efflux transporter from the nodT family, and G118DRAFT_03668 (q-value = 0.0952, Wilcoxon rank sum test), a Tir chaperone protein (CesT). Each point represents the average count of a minimum of three to six plates derived from the same plantlet, expressedas colony-forming units (CFU) per gram of root. **c–i)** Examples of known functional plant-associated operons captured by different statistical approaches. The plant-associated genes are highlighted by shaded bars, colored according to the key. **c)** Nod genes. **d)** NIF genes. **e)** Ent-kaurene (gibberelin precursor). **f)** Chemotaxis proteins in bacteria from different taxa. **g)** Type III secretion system. **h)** Type VI secretion system, including the imp genes (impaired in nodulation). **i)** Flagellum biosynthesis in Alphaproteobacteria. Labels show the gene symbol or the protein name for which such information was available.

Thus, we identified thousands of plant-associated and root- associated gene clusters by using five different statistical approaches and validated them by means of computational and experimental approaches, broadening our understanding of the genetic basis of plant–microbe interactions and providing a valuable resource to drive further experimentation.

2.3 Protein domains reproducibly enriched in diverse plant-associated genomes

Plant-associated and root-associated proteins and protein domains conserved across evolutionarily diverse taxa are potentially pivotal to the interaction between bacteria and plants. We identified 767 Pfam domains as significant plant-associated domains in at least three taxa, on the basis of multiple tests. Below we elaborate on a few domains that were plant-associated or root-associated in all four phyla. Two of these domains, a DNA-binding domain (pfam00356) and a ligand-binding (pfam13377) domain, are characteristic of the LacI transcription factor (TF) family. These TFs regulate gene expression in response to different sugars (Ravcheev et al., 2014), and their copy numbers were expanded in the genomes of plant-associated and root-associated bacteria in eight of the nine taxa analyzed (Figure 2.6a). Examination of the genomic neighbors of lacI-family genes identified strong enrichment for genes involved in carbohydrate metabolism and transport in all of these taxa, consistent with their expected regulation by a LacI-family protein (Ravcheev et al., 2014). We analyzed the promoter regions of these putative regulatory targets of LacI-family TFs, and identified three AANCGNTT palindromic octamers that were statistically enriched in all but one taxon, and which may serve as the TF-binding site. These data suggest that accumulation of a large repertoire of LacI family-controlled regulons is a common strategy across bacterial lineages during adaptation to the plant environment. Another domain, the metabolic domain aldo-keto reductase (pfam00248), was enriched in the genomes of plant-associated and root-associated bacteria from eight taxa belonging to all four phyla investigated (Figure 2.6b). This

domain is involved in the metabolic conversion of a broad range of substrates, including sugars and toxic carbonyl compounds (Yamauchi et al., 2011). Thus, bacteria that inhabit plant environments may consume similar substrates.

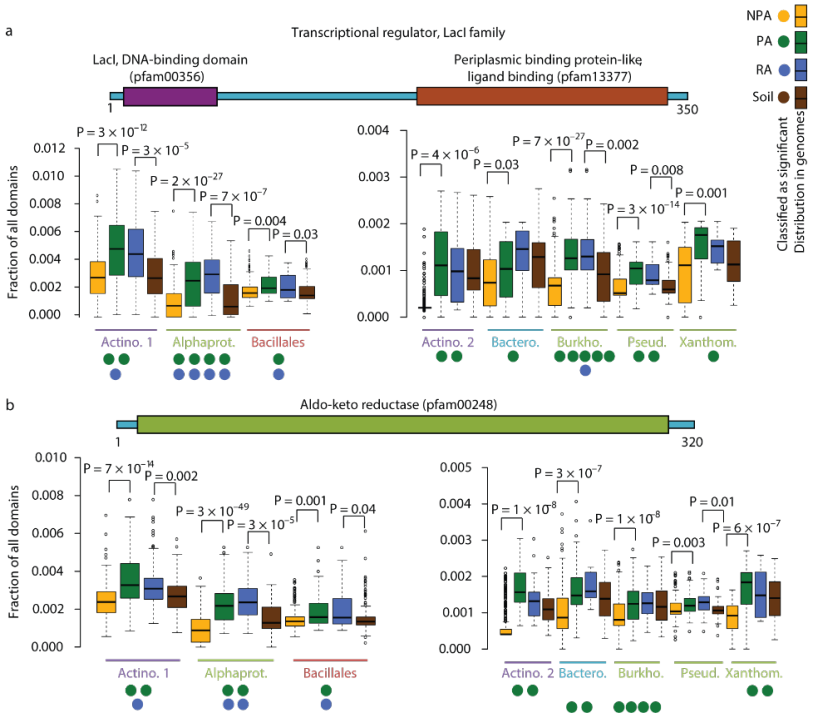


Figure 2.6 - **Proteins and protein domains that were reproducibly enriched as plant-associated or root-associated in multiple taxa.**

We compared the occurrence of protein domains (from Pfam) between plant-associated (PA) and NPA bacteria and between root-associated (RA) and soil-associated bacteria. Color-coding is as in Figure 2.1. **a)** Transcription factors with LacI (Pfam00356) and periplasmic-binding protein domains (Pfam13377). These proteins are often annotated as COG1609. **b)** Aldo-keto reductase domain (Pfam00248). Proteins with this domain are often annotated as COG0667. We used a two-sided t-test to test for the presence of the genes in a and b in genomes that shared the same label and to verify the enrichment reported by the various tests. FDR-corrected P values are shown for significant results (*q-value* < 0.05). Colored circles indicate the number of different statistical tests (≤ 5) supporting plant, non-plant, root, or soil association of a gene or domain, with each circle representing one test. Gene illustrations above each graph represent random protein models. Note that **a** and **b**, each contain two graphs because of the different scales. Abbreviations: Actino., Actinobacteria; Alphaprot., Alphaproteobacteria; Burkho., Burkholderiales; Bactero., Bacteroidetes; Pseud., Pseudomonas; Xanthom., Xanthomonadaceae. Box-and-whisker plots show the median (center lines), 25th and 75th percentiles (box edges), extreme data points within 1.5 times the interquartile range from the box edge (whiskers), and outliers (isolated data points).

2.4 Putative plant protein mimicry by plant- and root-associated proteins

Convergent evolution and horizontal transfer of protein domains from eukaryotes to bacteria have been suggested for some microbial effector proteins that are secreted into eukaryotic host cells to suppress defense and facilitate microbial proliferation (Burstein et al., 2009; Dean, 2011; Stebbins & Galaán, 2001). We searched for new candidate effectors or other functional plant-protein mimics. We retrieved a set of significant plant-associated and root-associated Pfam domains that were reproducibly predicted by multiple approaches or in multiple taxa, and we cross-referenced these with protein domains that were also more abundant in plant genomes than in bacterial genomes (Methods). This analysis yielded 64 plant-resembling plant-associated and root-associated domains (PREPARADOs) encoded by 11,916 genes (Figure 2.7).

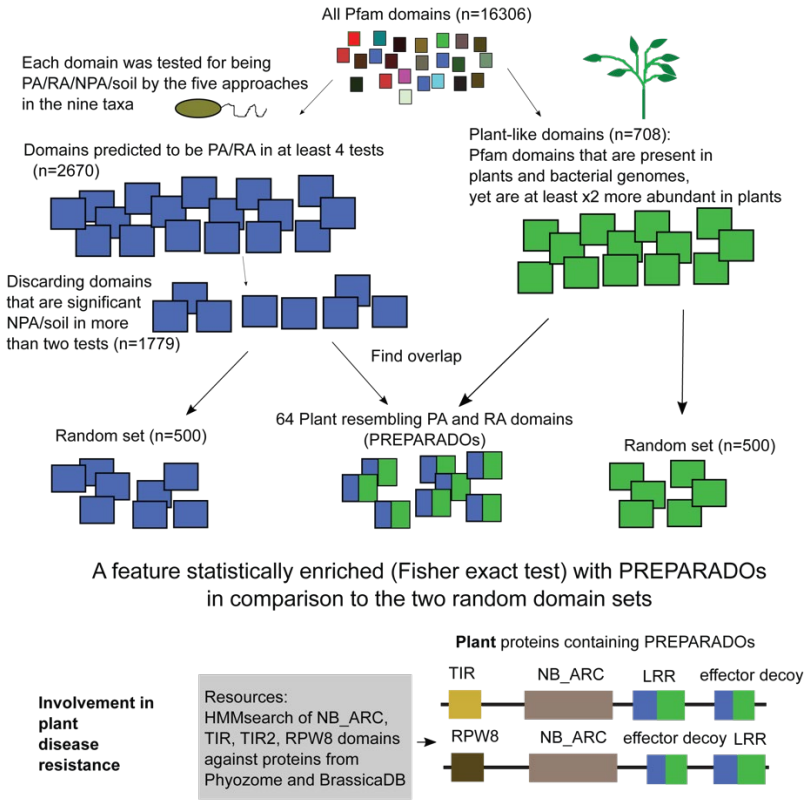


Figure 2.7 - **Algorithm to predict the plant-resembling plant-associated and root-associated domains (PREPARADOs).**

LRR is illustrated as a PREPARADO as LRR6 and LRR8 are also PREPARADOs.

The number of PREPARADOs was fourfold higher than the number of domains that overlapped reproducible NPA/soil-associated domains and plant domains (n=15). Some PREPARADOs were previously described as domains within effector proteins, such as Ankyrin repeats (C. T. Price et al., 2009), regulator of chromosome condensation repeat (RCC1) (Rothmeier et al., 2013), leucine-rich repeat (LRR) (R. Q. Xu et al., 2008), and pectate lyase (Shevchik et al., 1997). PREPARADOs from plant genomes were enriched 3–14-fold ($p < 10-5$, Fisher's exact test) as domains predicted to be 'integrated effector decoys' when fused to plant intracellular innate immune receptors of the NLR class (Cesari et al., 2014; Sarris et al., 2015, 2016) (compared with two random domain sets; Methods, Figure 2.8).

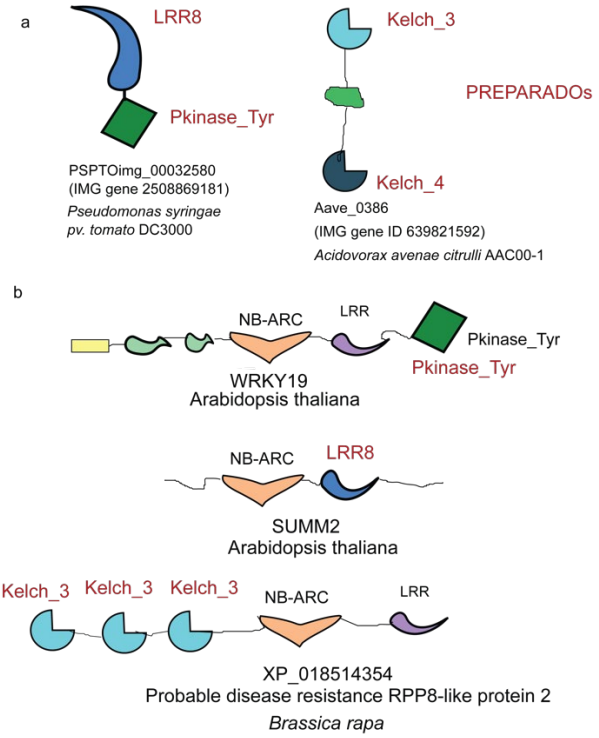


Figure 2.8 - **Exemplary PREPARADOs contained in putative effector binding or disease resistance proteins in plants.**

a) Examples of microbial proteins, each with two PREPARADOs (LRR8, Pkinase_Tyr, Kelch_3, Kelch_4). **b)** Integration ofPREPARADOs into NB-ARC domains in different plant proteins. NB-ARC is present inmany disease resistance (R) proteins. SUMM2 was suggested to act as an R gene (Zhang et al., 2012).

Seven PREPARADO-containing protein families were characterized by N-terminal eukaryotic or bacterial signal peptides followed by a PREPARADO dedicated to carbohydrate binding or metabolism. One of these domains, Jacalin, is a mannose-binding lectin domain that is found in 48 genes in the *A. thaliana* genome, compared with three genes in the human genome (Finn et al., 2016). Mannose is found on the cell wall of different bacterial and fungal pathogens and could serve as a microbial-associated molecular pattern that is recognized by the plant immune system (Gadjeva et al., 2004; Ma et al., 2010; Osborn et al., 1964; Sahly et al., 2002; Weidenbach et al., 2016; Xiang et al., 2011; Yamaji et al., 2012). We identified a family of ~430-amino-acid-long microbial proteins with a signal peptide followed by a functionally ill-defined endonuclease/exonuclease/phosphatase family domain (pfam03372), and ending with a Jacalin domain (pfam01419). This domain architecture is absent in plants but is found in diverse microorganisms, many of which are phytopathogens, including Gram-negative and Gram-positive bacteria, fungi from the Ascomycota and Basidiomycota phyla, and oomycetes (Figure 2.9). We speculate that these microbial lectins may be secreted to outcompete plant immune receptors for mannose-binding on the microbial cell wall, effectively serving as camouflage. We thus discovered a large set of protein domains that are shared between plants and the microbes that colonize them. In many cases the entire protein is conserved across evolutionarily distant plant-associated microorganisms.

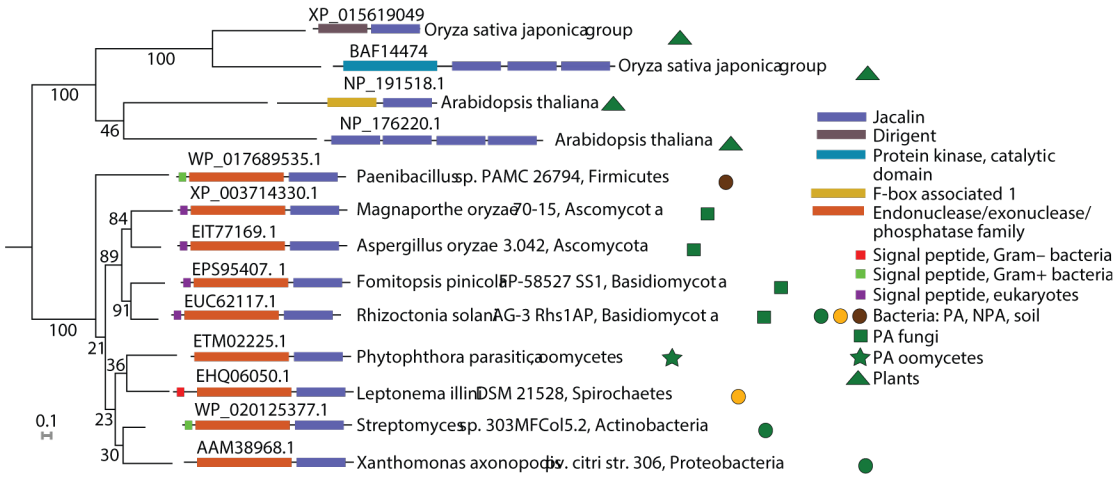


Figure 2.9 - A **protein family shared by plant-associated bacteria, fungi, and oomycetes that resemble plant proteins.**

A maximum-likelihood phylogenetic tree of representative proteins with Jacalin-like domains across plants and plant-associated (PA) organisms. Endonuclease/exonuclease/phosphatase-Jacalin proteins are present across PA eukaryotes (fungi and oomycetes) and PA bacteria. In most cases these proteins contain a signal peptide in the N terminus. The Jacalin-like domain is found in many plant proteins, often in multiple copies. The protein accession is shown above each protein illustration.

2.5 Co-occurrence of plant-associated gene clusters

We identified numerous cases of plant-associated gene clusters (orthogroups) that demonstrate high co-occurrence between genomes. When the plant-associated genes were derived by phylogeny-aware tests (i.e., PhyloGLM and Scoary), they were candidates for inter-taxon horizontal gene transfer events. For example, we identified a cluster predicted by Scoary of up to 11 co-occurring genes (mean pairwise Spearman correlation: 0.81) in a flagellum-like locus from sporadically distributed plant-associated or soil-associated genomes across 12 different genera in Burkholderiales (Figure 2.10). Two of the annotated flagellar-like proteins, FlgB (COG1815) and FliN (pfam01052), are also encoded by plant-associated genes in Actinobacteria 1 and Alphaproteobacteria taxa. Six of the remaining genes encode hypothetical proteins, all but one of which are specific to Betaproteobacteria, suggestive of a flagellar structure variant that evolved in

this class in the plant environment. Flagellum-mediated motility or flagellum-derived secretion systems (for example, T3SS) are important for plant colonization and virulence (Cole et al., 2017; De Weert et al., 2002; De Weger et al., 1987; Tans-Kersten et al., 2001) and can be horizontally transferred (Poggio et al., 2007).

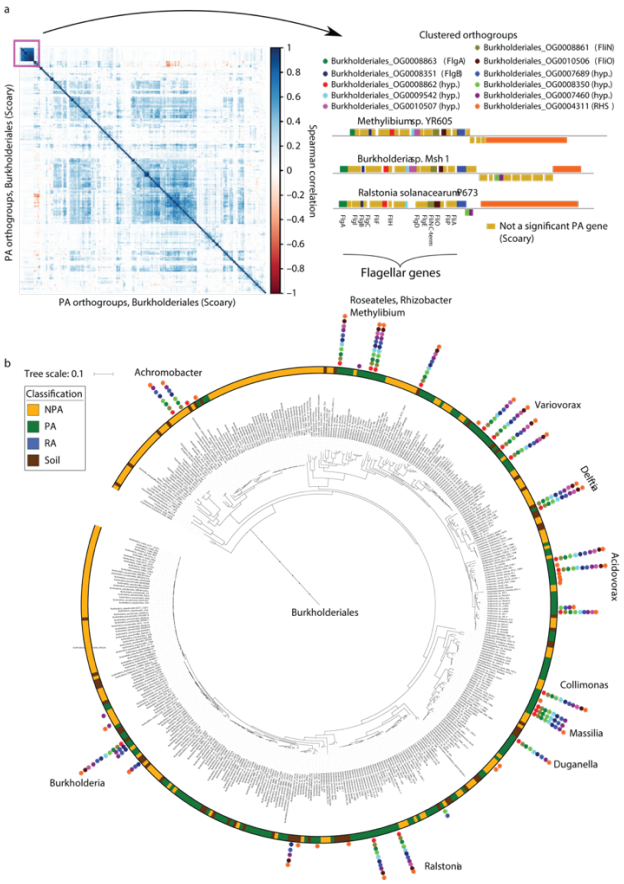


Figure 2.10 - **Co-occurring plant-associated and soil-associated flagellum-like gene clusters are sporadically distributed across Burkholderiales.**

a) Left, a hierarchically clustered correlation matrix of all 202 significant plant-associated (PA) orthogroups (gene clusters) from Burkholderiales, predicted by Scoary. Right, the orthogroups present within and adjacent to the flagellar-like locus of different genomes. Gene names based on a BLAST search are shown in parentheses. Hyp., hypothetical protein; RHS, RHS repeat protein. Genes illustrated above and below the black horizontal line for each species are located on the positive and negative strand, respectively. **b)** The Burkholderiales phylogenetic tree based on the concatenated alignment of 31 single- copy genes. Colored circles represent the 11 orthogroups presented in a, with the same color-coding as in a. Genus names are shown next to pillars of stacked circles. RA, root-associated.

2.6 Novel plant-associated gene operons

In addition to successfully capturing several known plant-associated operons (Figures 2.5-6), we also identified putative plant-associated bacterial operons. Two previously uncharacterized plant-associated gene families were conspicuous. These genes are organized in multiple loci in plant-associated genomes, each with up to five tandem gene copies. They encode short, highly divergent, high-copy-number proteins that are predicted to be secreted, as explained below. These two plant-associated protein families never co-occurred in the same genome, and their genomic presence was perfectly correlated with lifestyles of pathogenic or nonpathogenic bacteria of the genus *Acidovorax* (order Burkholderiales) (Figure 2.11). We named the gene families present in non-pathogens and pathogens Jekyll and Hyde, respectively, after the characters in Robert Louis Stevenson's classic novel. The typical Jekyll gene is 97 amino acids long, contains an N-terminal signal peptide, lacks a transmembrane domain, and, in 98.5% of cases, appears in non-pathogenic plant-associated or soil-associated *Acidovorax* isolates (Figure 2.11a). We recently isolated four *Acidovorax* strains from the leaves of naturally grown Arabidopsis (Bai et al., 2015). Even these nearly identical isolates carried hypervariable Jekyll loci that were substantially more divergent than neighboring genes and included copy-number variations and various mutations (Figure 2.11a, b). The Hyde putative operons, in contrast, are composed of two distinct gene families unrelated to Jekyll. A typical Hyde1 protein has 135 amino acids and an N-terminal transmembrane helix. Hyde1 proteins are also highly variable, as demonstrated by copy-number variation, sequence divergence, and intra-locus transposon insertions (Figure 2.11a, c). Hyde1 was found in 99% of cases in phytopathogenic *Acidovorax*. These genomes carried up to 15 Hyde1 gene copies distributed in up to ten loci (Figure 2.11a). In 70% of cases Hyde1 was located directly downstream from a more conserved ~300-amino-acid-long plant-associated protein-coding gene that we named

Hyde2 (Figure. 2.11c, d). We identified loci with Hyde2 followed by Hyde1-like genes in different members of the Proteobacteria phylum. Hyde-carrying organisms included other phytopathogens, such as *Pseudomonas syringae*, in which the Hyde1-like-Hyde2 locus was again highly variable between closely related strains (Figure 2.11d). Notably, we observed that Hyde genes often are directly preceded by genes that encode core structural T6SS proteins, such as PAAR, VgrG, and Hcp (Ho et al., 2014), or are fused to PAAR (Figure 2.11d). We therefore suggest that Hyde1 and/or Hyde2 might constitute a new T6SS effector family.

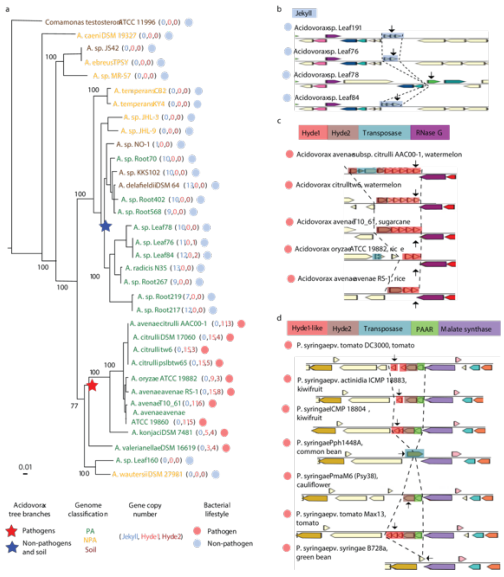


Figure 2.11 - **Rapidly diversifying, high-copy-number Jekyll and Hyde plant-associated genes.**

a) A maximum likelihood phylogenetic tree of *Acidovorax* isolates based on concatenation of 35 single-copy genes. The pathogenic and non-pathogenic branches of the tree are perfectly correlated with the presence of Hyde1 and Jekyll genes, respectively. **b)** An example of a variable Jekyll locus in highly related *Acidovorax* species isolated from leaves of wild *Arabidopsis* from Brugg, Switzerland. Arrows indicate the following locus tags (from top to bottom): Ga0102403_10161, Ga0102306_101276, Ga0102307_107159, and Ga0102310_10161. **c)** An example of a variable Hyde locus from pathogenic *Acidovorax* infecting different plants (the host plant is shown after the species name). The transposase in the first operon fragmented a Hyde2 gene. Arrows indicate the following locus tags (from top to bottom): Aave_3195, Ga0078621_123525, Ga0098809_1087148, T336DRAFT_00345, and AASARDRAFT_03920. **d)** An example of a variable Hyde locus from pathogenic *Pseudomonas syringae* infecting different plants. Arrows indicate the following locus tags (from top to bottom): PSPTOimg_00004880 (a.k.a.

PSPTO_0475), A243_06583, NZ4DRAFT_02530, Pphimg_00049570, PmaM6_0066.00000100, PsyrptM_010100007142, and Psyr_4701. Genes color- coded with the same colors in b–d are homologous, with the exception of genes colored in ivory (unannotated genes) and Hyde1 and Hyde1-like genes, which are analogous in terms of their similar size, high diversification rate, position downstream of Hyde2, and tendency to have a transmembrane domain. PAAR, proline-alanine-alanine-arginine repeat superfamily.

The high sequence diversity of Jekyll and Hyde1 genes suggests that the two plant-associated protein families encoded by these genes could be involved in molecular arms races with other organisms in the plant environment. As many Type VI effectors are used in interbacterial warfare, we tested *Acidovorax* Hyde1 proteins for antibacterial properties. Expression of two variants of the gene in Escherichia coli led to a 105 –106 -fold reduction in cell numbers (Figure 2.12a). We constructed a mutant strain of the phytopathogen *Acidovorax citrulli* AAC00-1with deletion of five Hyde1 loci (Δ5-Hyde1), encompassing 9 of 11 Hyde1 genes. Wild-type, Δ5-Hyde1, and T6SS-mutant (ΔT6SS) *Acidovorax* strains were co-incubated with an *E. coli* strain that is susceptible to T6SS killing (MacIntyre et al., 2010) and nine phylogenetically diverse *Arabidopsis* leaf bacterial isolates (Bai et al., 2015). Survival of wild-type *E. coli* and six of the leaf isolates after coincubation with wild-type *Acidovorax* was reduced 102 –106 -fold compared with that after coincubation with Δ5-Hyde1 or ΔT6SS *Acidovorax* (Figure 2.12b). Combined with the genomic association of Hyde loci with T6SS, these results suggest that the T6SS antibacterial phenotype of *Acidovorax* is mediated by Hyde proteins and that these toxins could be used in competition against other plant-associated organisms (Tian et al., 2015).

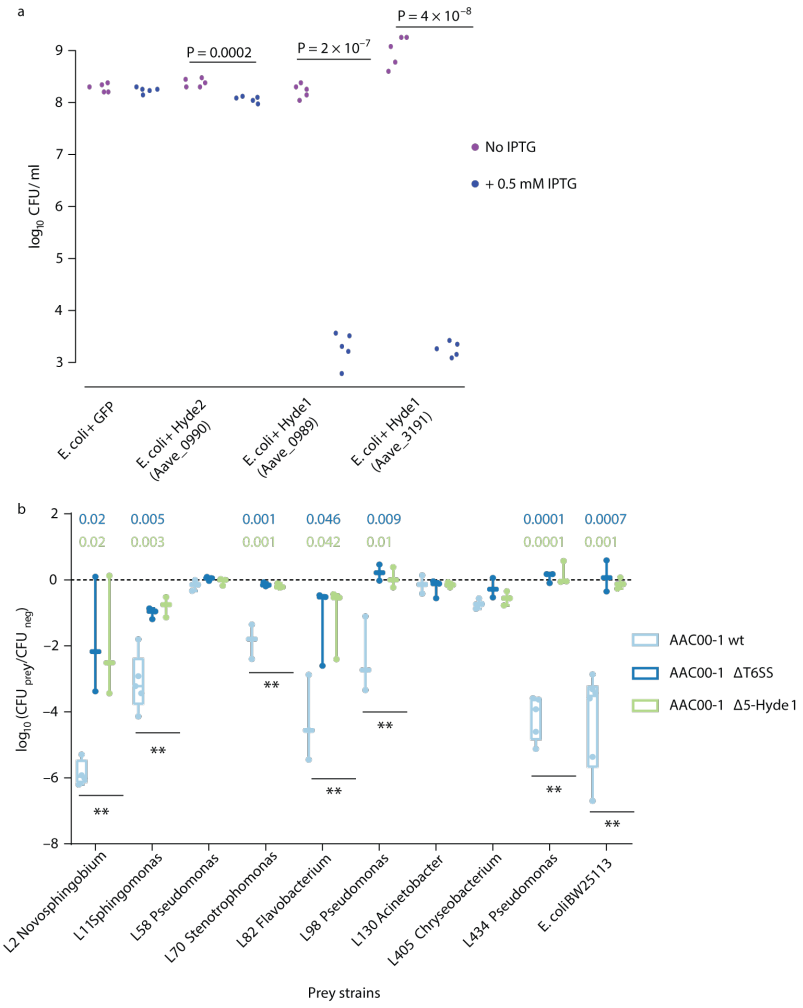


Figure 2.12 - **Hyde1 proteins of *Acidovorax citrulli* AAC00-1 are toxic to *E. coli* and various plant-associated bacterial strains.**

a) Toxicity assay of Hyde proteins expressed in *E. coli.* GFP, Hyde2-Aave_0990, and two Hyde1 genes from two loci, Aave_0989 and Aave_3191, were cloned into pET28b and transformed into *E. coli* C41 cells. Aave_0989 and Aave_3191 proteins were 53% identical. Bacterial cultures from five independent colonies were spotted on an LB plate. Gene expression of the cloned genes was induced with 0.5 mM IPTG. P values are shown for significant results (two-sided t-test). **b)** Quantification of recovered prey cells after co-incubation with *Acidovorax* aggressor strains. Antibiotic-resistant prey strains *E. coli* BW25113 and nine different *Arabidopsis* leaf isolates were mixed at equal ratios with different aggressor strains or with NB medium (negative control). Five Hyde1 loci (including 9 out of 11 Hyde1 genes) are deleted in Δ5-Hyde1. ΔT6SS contains a vasD (Aave_1470) deletion. After coincubation for 19 h on NB agar plates, mixed populations were resuspended in NB medium and spotted on selective antibiotic-containing NB agar. The box plots represent results from at least three independent experiments, with individual values superimposed as dots. The center line represents the median, the box limits represent the 25th and 75th percentiles, and the edges represent the minimal and maximal values. P values are shown at the top; double asterisks denote a significant difference (one-way ANOVA followed by Tukey's honest significant difference test) between results for wild type versus ΔT6SS and for wild type versus Δ5-Hyde1.

2.7 Discussion

Here we sequenced nearly 500 root-associated bacterial genomes isolated from different plant hosts. These new genomes were combined in a collection of 3,837 high-quality bacterial genomes for comparative analysis. We developed a systematic approach to identify plant-associated and root-associated genes and putative operons. Our method is accurate as reflected by its ability to capture numerous operons previously shown to have a plant-associated function, the enrichment of plant-associated genes in plant-associated metagenomes, the validation of Hyde1 proteins as likely type VI effectors in *Acidovorax* directed against other plant-associated bacteria, and the validation of two new genes in *P. kururiensis* that affect rice root colonization. We note that bacterial genes that are enriched in genomes from the plant environment are also likely to be involved in adaptation to the many other organisms that share the same niche, as we demonstrated for Hyde1. We used five different statistical approaches to identify genes that were significantly associated with the plant/root environment, each with its advantages and disadvantages. The phylogeny-correcting approaches (phyloglmbin, phyloglmcn, and Scoary) allow accurate identification of genes that are polyphyletic and correlate with an environment independently of ancestral state. On the basis of our metagenome validation, the hypergeometric test predicts more genes that are abundant in plant-associated communities than PhyloGLM does. It also identifies monophyletic plant-associated genes, but it yields more false positives than the phylogenetic tests, because in every plant-associated lineage many lineage-specific genes will be considered plant-associated. Scoary is the most stringent method of all and yielded the fewest predictions. Future experimental validation should prioritize genes predicted in multiple taxa and/or by multiple approaches.

2.8 Methods

2.8.1 Bacterial isolation and genome sequencing

Bacterial strains from Brassicaceae and poplar were isolated via previously described protocols (Doty et al., 2009; Weston et al., 2012). Poplar strains were cultured from root tissues collected from *Populus deltoides* and *Populus trichocarpa* trees in Tennessee, North Carolina, and Oregon. Root samples were processed as described previously (Gottel et al., 2011; Weston et al., 2012). Briefly, we isolated rhizosphere strains by plating serial dilutions of root wash, whereas for endosphere strains, we pulverized surface-sterilized roots with a sterile mortar and pestle in 10 mL of MgSO4 (10 mM) solution before plating serial dilutions. Strains were isolated on R2A agar media, and the resulting colonies were picked and re-streaked a minimum of three times to ensure isolation. Isolated strains were identified by 16S rDNA PCR followed by Sanger sequencing.

For maize isolates, we selected soils associated with Il14h and Mo17 maize genotypes grown in Lansing, NY, and Urbana, IL. The rhizosphere soil samples of each maize genotype were grown at each location and were collected at week 12 as previously described (Peiffer et al., 2013). From each rhizosphere soil sample, soil was washed and samples were plated onto Pseudomonas isolation agar (BD Diagnostic Systems). The plates were incubated at 30 °C until colonies formed, and DNA was extracted from cells.

For isolation of single cells, *A. thaliana* accessions Col-0 and Cvi-0 were grown to maturity. Roots were washed in distilled water multiple times. Root surfaces were sterilized with bleach. Surface-sterilized roots were then ground with a sterile mortar and pestle. Individual cells were isolated by flow cytometry followed by DNA amplification with MDA, and 16S rDNA screening as described previously (Rinke et al., 2013).

DNA from isolates and single cells was sequenced on next-generation sequencing platforms, mostly using Illumina HiSeq technology. Sequenced genomic DNA was assembled via different assembly methods. Genomes were sequenced annotated using the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4)23 and deposited at the IMG database, ENA, or Genbank for public use.

2.8.2 Data compilation of 3,837 isolate genomes and their isolation-site metadata.

We retrieved 5,586 bacterial genomes from the IMG system. Isolation sites were identified through a manual curation process that included scanning of IMG metadata, DSMZ, ATCC, NCBI Biosample and the scientific literature. On the basis of its isolation site, each genome was labeled as plant-associated, NPA, or soil-associated. Plant-associated organisms were also labeled as root-associated when isolated from the endophytic compartments or from the rhizoplane. We applied stringent quality control measures to ensure a high-quality and minimally biased set of genomes:

a) Known isolation site: genomes with missing isolation-site information were filtered out.
b) High genome quality and completeness: all isolate genomes passed this filter if N50 (the shortest sequence length at 50% of the genome) was more than 50,000 bp. Single amplified genomes passed the quality filter if they had at least 90% of 35 universal single-copy clusters of orthologous groups (COGs) (Beszteri et al., 2010). In addition, we used CheckM (Parks et al., 2015) to assess isolate genome completeness and contamination. Only genomes that were at least 95% complete and no more than 5% contaminated were used.
c) High-quality gene annotation: genomes that passed this filter had at least 90% genome sequence coding for genes, with an exception—in the Bartonella genus most genomes have coding base percentages below 90%.

d) Non-redundancy: we computed whole-genome average nucleotide identity and alignment fraction values for each pair of genomes (Varghese et al., 2015). When the alignment fraction exceeded 90% and the whole-genome average nucleotide identity was greater than 99.995% we considered the genome pair redundant. In such cases one genome was randomly selected, and the other genome was marked as redundant and was filtered out.

e) Consistency in the phylogenetic tree: we filtered out 14 bacterial genomes that showed discrepancy between their given taxonomy and their actual phylogenetic placement in the bacterial tree.

2.8.3 Construction of the bacterial genome tree

To generate a phylogenetic tree of the 3,837 high-quality and nonredundant bacterial genomes, we retrieved 31 universal single-copy genes from each genome with AMPHORA (Martin Wu & Eisen, 2008). For each individual marker gene, we used Muscle with default parameters to construct an alignment. We masked the 31 alignments by using Zorro (M Wu et al., 2012) and filtered the low-quality columns of the alignment. Finally, we concatenated the 31 alignments into an overall merged alignment, from which we built an approximately maximum-likelihood phylogenetic tree with the WAG model implemented in FastTree 2.1 (M. N. Price et al., 2010). Trees of each taxon are provided in the following website: http://labs.bio.unc.edu/Dangl/Resources/gfobap_website/faa_trees_metadata.html.

2.8.4 Clustering of the 3,837 genomes into nine phylogenetic bins

We divided the dataset into different taxa (taxonomic groups) to allow downstream identification of genes enriched in the plant-associated or root-associated genomes of each taxon compared with the NPA or soil-associated genomes from the same taxon, respectively. To determine the number of taxonomic groups to analyze, we converted the phylogenetic tree into a

distance matrix, using the cophenetic function implemented in the R package ape (Paradis et al., 2004). We then clustered the 3,837 genomes into nine groups using k-medoids clustering as implemented in the PAM (partitioning around medoids) algorithm from the R package fpc. The k-medoids algorithm clusters a dataset of n objects into k a priori–defined clusters. To identify the optimal k value for the dataset, we compared the silhouette coefficients for values of k ranging from 1 to 30. We selected a value of $k = 9$ because it yielded the maximal average silhouette coefficient (0.66). In addition, at $k = 9$ the taxa were monophyletic, contained hundreds of genomes, and were relatively balanced between plant-associated and NPA genomes in most taxa (Table 2.1). The resulting genome clusters generally overlapped with annotated taxonomic units. One exception was in the Actinobacteria phylum. Here our clustering divided the genomes into two taxa that we named, for simplicity, Actinobacteria 1 and Actinobacteria 2. However, our rigorous phylogenetic analysis supports previous suggestions for revisions in the taxonomy of phylum Actinobacteria (Sen et al., 2014). In addition, the tree showed very divergent bacterial taxa in the Bacteroidetes phylum that could not be separated into monophyletic groups. Specifically, the Sphingobacteriales order (from class Sphingobacteria) and the Cytophagaceae (from class Cytophagia) are paraphyletic. Therefore, we decided to unify all Bacteroidetes into one phylum-level taxon.

2.8.5 COG enrichment analysis

Protein-coding genes were retrieved and mapped to COG IDs with the program RPS-BLAST at an e-value cutoff of 1e–2 and an alignment length of at least 70% of the consensus sequence length. Each COG ID was mapped to at least one COG category. For each genome, we counted the number of genes from a given category. A t-test and PhyloGLM test were used to

compare the number of genes in the genomes that shared the same taxon and category but different labels (e.g., plant-associated versus NPA).

2.8.6 Identification of plant-associated, NPA, root-associated, and soil genes/domains

The following description applies to plant-associated, NPA, root-associated, and soil genes. For conciseness, only plant-associated genes are described here. Plant-associated genes were identified via a two-step process that included protein/domain clustering on the basis of amino acid sequence similarity and subsequent identification of the protein/domain clusters significantly enriched in proteins/domains from plant-associated bacteria. Clustering of genes and protein domains involved five independent methods: OrthoFinder (Emms & Kelly, 2015), COG (Tatusov et al., 2000), KEGG Orthology (KO) (Kanehisa et al., 2016), TIGRFAM (Haft et al., 2003), and Pfam (Finn et al., 2016). OrthoFinder was selected as the *de novo* clustering approach that included all proteins, including those that lack any functional annotation. We first compiled, for each taxon separately, a list of all proteins in the genomes. For COG, KO, TIGRFAM, and Pfam, we used the existing annotations of IMG genes that are based on BLAST alignments to the different protein/domain models (Huntemann et al., 2015). This process yielded gene/domain clusters. For each of the nine taxa, we used custom scripts to transform the orthogroup result output from Orthofinder into pan genome matrices depicting the distribution of orthogroups across the genomes of that given taxon. Additionally, we constructed tables exhibiting the distribution of orthogroups across genomes based on the CDS IDs.

The pan-genome matrices and tables described above can be downloaded from: http://labs.bio.unc.edu/Dangl/Resources/gfobap_website/matrices_df_ogs.html

The scripts used to compute the matrices and tables can be found in: https://github.com/isaisg/gfobap/tree/master/orthofinder_orthogroups_to_matrices_dataframes

Next, we determined which clusters were significantly enriched with genes derived from plant-associated genomes. These clusters were termed plant-associated clusters. In the statistical analysis, we used only clusters of more than five members. We corrected P values with Benjamini–Hochberg FDR and used $q < 0.05$ as the significance threshold, unless stated otherwise. The proteins in each cluster were categorized as either plant-associated or NPA, on the basis of the label of the encoding genome. Namely, a plant-associated gene is a gene derived from a plant-associated gene cluster and a plant-associated genome.

The three main approaches were the hypergeometric test (Hyperg), PhyloGLM, and Scoary. Hyperg looks for overall enrichment of gene copies across a group of genomes but ignores the phylogenetic structure of the dataset. PhyloGLM (Ives & Garland, 2010) takes into account phylogenetic information to eliminate apparent enrichments that can be explained by shared ancestry. The Hyperg and PhyloGLM tests were used in two versions, based on either gene presence/absence data (hypergbin, phyloglmbin) or gene copy-number data (hypergcn, phyloglmcn). We also used a stringent version of Scoary (Brynildsrud et al., 2016), a gene presence/absence approach that combines Fisher's exact test, a phylogenetic test, and a label-permutation test. The first hypergeometric test, hypergcn, used the gene copy-number data, with the cluster being the sample, the total number of plant-associated and NPA genes being the population, and the number of plant-associated genes within the cluster being considered as 'successes'. The second version, hybergbin, used gene presence/absence data. P values were corrected by Benjamini–Hochberg FDR (Benjamini & Hochberg, 1995) for clusters of COG/KO/TIGRFAM/Pfam. For the abundant OrthoFinder clusters, we used Bonferroni correction with a threshold of $p < 0.1$, as downstream validation with metagenomes showed fewer false positives with the more significant clusters. The third and fourth statistical approaches used

PhyloGLM (Ives & Garland, 2010), implemented in the phylolm (v 2.5) R package. PhyloGLM combines a Markov process of lifestyle (e.g., plant-associated versus NPA) evolution with a regularized logistic regression. This approach takes advantage of the known phylogeny to specify the residual correlation structure between genomes that share common ancestry, and so it does not need to make the incorrect assumption that observations are independent. Intuitively PhyloGLM favors genes found in multiple lineages of the same taxon. For each taxon we used the subtree from Figure 2.1 to estimate the correlation matrix between observations and used the copy number (in phyloglmcn) or presence/absence pattern (in phyloglmbin) of each gene as the only independent variable. Positive and negative estimates in phyloglmbin/phyloglmcn indicated plant-associated/root-associated and NPA/soil-associated proteins/domains, respectively.

Finally, the fifth statistical approach was Scoary (Brynildsrud et al., 2016), which uses a gene presence/absence dataset. Scoary combines Fisher's exact test, a phylogeny-aware test, and an empirical label-switching permutation analysis. A gene cluster was considered significant by Scoary only if (1) it had a *q*-value less than 0.05 for Fisher's exact test, (2) the 'worst' P value from the pairwise comparison algorithm was < 0.05, and (3) the empirical (permutation-based) *p* value was < 0.05. These are very stringent criteria that yielded relatively few significant predictions. Odds ratios greater than or less than 1 in Scoary indicated plant-associated/root-associated and NPA/soil-associated proteins/domains, respectively.

2.8.7 Validation of predicted plant-associated, NPA, root-associated, and soil-associated genes using metagenomes.

For each orthogroup in our dataset consisting of more than two CDS, we built multiple sequence alignments using MAFFT20 (v7.305b) (Katoh et al., 2002). Subsequently, we used these alignments as inputs for Hidden Markov Model (HMMs) construction using the hmmbuild command from the HMMER suite21 (v 3.1b2) (Eddy, 2011). We built nine HMM databases, one

corresponding to each of the nine taxa analyzed in this study. Also, to complement these databases, we developed a scanning pipeline that utilizes the HMMER suite to search for the HMM orthogroups in our nine databases over any genome provided to the pipeline. The MAFFT alignments, HMM profiles and HMM databases can be downloaded from: http://labs.bio.unc.edu/Dangl/Resources/gfobap_website/mafft_hmm.html

The scripts to compute the alignments and the HMM profiles plus the pipeline to scan novel genomes using the HMM databases can be downloaded from: https://github.com/isaisg/gfobap/tree/master/mafft_hmm

https://github.com/isaisg/gfobap/tree/master/scanner_orthogroups_scripts

In parallel metagenome samples (n = 38) were downloaded from NCBI and GOLD databases. The reads were translated into proteins, and proteins at least 40 amino acids long were aligned using HMMsearch (Eddy, 2011) against the different protein references. The protein references included the predicted plant-associated, root-associated, soil-associated, and NPA proteins from OrthoFinder that were found to be significant by the different approaches.

2.8.8 Validation of plant-associated genes in Paraburkholderia kururiensis M130 affecting rice root colonization.

Paraburkholderia kururiensis strain M130 was grown at 30 C in King's B medium. *E coli* strains were grown at 37° C in Luria Bertani medium. When needed, antibiotics were used in the following concentrations: ampicillin, 100 μg/mL; kanamycin, 50 μg/mL; nitrofurantoin, 50 μg/mL; rifampicin 50 μg/mL.

For the mutant construction, internal fragments of 200–900 bp from each gene of interest were PCR-amplified. Fragments were cloned in the pGem2T easy vector (Promega) and sequenced (GATC Biotech; Germany), then excised with EcoRI restriction enzyme and cloned in the corresponding site in pKNOCK-Km R94. These plasmids were then used as a suicide delivery

system to create the knockout mutants and transferred to *P. kururiensis* M130 by triparental mating. All the mutants were verified by PCR with primers specific to the pKNOCK-Km vector and to the genomic DNA sequences upstream and downstream from the targeted genes.

For the rhizosphere colonization experiments with *P. kururiensis*. Seeds of *Oryza sativa* (BALDO variety) were surface-sterilized and then left to germinate in sterile conditions at 30 °C in the dark for 7 d. Each seedling was then aseptically transferred into a 50-mL Falcon tube containing 35 mL of half-strength Hoagland solution semisolid substrate (0.4% agar). The tubes were then inoculated with 107 colony-forming units (cfu) of a *P. kururiensis* suspension. Plants were grown for 11 d at 30 °C (16/8-h light/dark cycles). For the determination of the bacterial counts, plants were washed under tap water for 1 min and then cut below the cotyledon to excise the roots. Roots were air-dried for 15 min, weighed, and then transferred to a sterile tube containing 5 mL of PBS. After vortexing, the suspension was serially diluted to 10−1 and 10−2 in PBS, and aliquots were plated on KB plates containing the appropriate antibiotic (rifampicin 50 µg/mL for the wild type, rifampicin 50 µg/mL and kanamycin 50 µg/mL for the mutants). After 3 d of incubation at 30 °C, we counted colony-forming units (CFU). Three replicates for each dilution from ten independent plantlets were used to determine the average CFU values.

2.8.9 Plant-mimicking plant-associated and root-associated proteins

Figure 2.7 summarizes the algorithm used to find plant-mimicking plant-associated and root-associated proteins. Pfam (Finn et al., 2016) version 30.0 metadata were downloaded. Protein domains that appeared in both Viridiplantae and bacteria and occurred at least two times more frequently in Viridiplantae than in bacteria were considered as plant-like domains (n = 708). In parallel, we scanned the set of significant plant-associated, root-associated, NPA, and soil-associated Pfam protein domains predicted by the five algorithms in the nine taxa. We compiled a

list of domains that were significantly plant-associated/root-associated in at least four tests, and significantly NPA/soil-associated in up to two tests (n = 1,779). The overlapping domains between the first two sets were defined as PREPARADOs (n = 64). In parallel, we created two control sets of 500 random plant-like Pfam domains and 500 random plant-associated/root-associated Pfam domains. Enrichment of PREPARADOs integrated into plant NLR proteins in comparison to the domains in the control groups was tested by Fisher's exact test. To identify domains found in plant disease-resistance proteins, we retrieved all proteins from Phytozome and BrassicaDB. To identify domains in plant disease-resistance proteins, we used hmmscan to search protein sequences for the presence of NB-ARC (PF00931.20), TIR (PF01582.18), TIR_2 (PF13676.4), or RPW8 (PF05659.9) domains. Bacterial proteins carrying the PREPARADO domains were considered as having full-length identity to fungal, oomycete, or plant proteins on the basis of LAST alignments to all Refseq proteins of plants, fungi, and protozoa. "Full-length" is defined here as an alignment length of at least 90% of the length of both query and reference proteins. The threshold used for considering a high amino acid identity was 40%.

We annotated PREPARADO-containing proteins as secreted by Sec if they had a predicted signal peptide (Deng et al., 2016) and lacked a transmembrane domain according to IMG annotation. A protein was marked as being secreted by T3SS if it had a score > 0.999 according to EffectiveT3 as implemented by the EffectiveDB server (Eichinger et al., 2016). A domain was predicted to be associated with secretion by Sec or T3SS if over 50% of the proteins carrying the domain were predicted to be secreted by these secretion systems.

2.8.10 Prediction of plant-associated, NPA, root-associated, and soil-associated operons and their annotation as biosynthetic gene clusters

Significant plant-associated, NPA, root-associated, and soil-associated genes of each genome were clustered on the basis of genomic distance: genes sharing the same scaffold and

strand that were up to 200 bp apart were clustered into the same predicted operon. We allowed up to one spacer gene, which is a non-significant gene, between each pair of significant genes within an operon. Operons were predicted for the genes in COG and OrthoFinder clusters using all five approaches. Operons were annotated as biosynthetic gene clusters if at least one of the constituent genes was part of a biosynthetic gene cluster from the IMG-ABC database (Hadjithomas et al., 2015).

2.8.11 Jekyll and Hyde analyses

To find all homologs and paralogs of Jekyll and Hyde genes, we used IMG BLAST search with an e-value threshold of 1e–5 against all IMG isolates. We searched Hyde1 homologs of *Acidovorax*, Hyde1 homologs of *Pseudomonas*, Hyde2, and Jekyll genes using proteins of genes Aave_1071, A243_06583, Ga0078621_123530, and Ga0102403_10160 as the query sequence, respectively. Multiple sequence alignments were done with Mafft (Katoh et al., 2002). A phylogenetic tree of *Acidovorax* species was produced with RaxML (Stamatakis, 2014), based on concatenation of 35 single-copy genes (Martin Wu & Eisen, 2008).

To verify the toxicity of Hyde1 and Hyde2 proteins to *E. coli*, we cloned genes encoding proteins Aave_0990 (Hyde2), Aave_0989 (Hyde1), and Aave_3191 (Hyde1), or GFP as a control, to the inducible pET28b expression vector via the LR reaction. The recombinant vectors were transformed into *E. coli* C41 competent cells by electroporation after sequencing validation. Five colonies were selected and cultured in LB liquid media supplemented with kanamycin with shaking overnight. The OD600 of the bacterial culture was adjusted to 1.0, and then the culture was diluted by 102, 104, 106, and 108 times successively. Bacteria culture gradients were spotted (5 μL) on LB plates with or without 0.5 mM IPTG to induce gene expression.

A. citrulli strain AAC00-1 and its derived mutants were grown on nutrient agar medium supplemented with rifampicin (100 μg/ml). To delete a cluster of five Hyde1 genes (Aave_3191–3195), we carried out a marker-exchange mutagenesis as previously described (Traore et al., 2014). The marker-free mutant was designated as Δ1-Hyde1, and its genotype was confirmed by PCR amplification and sequencing. The marker-exchange mutagenesis procedure was repeated to delete four other Hyde1 loci. The final mutant, with deletion of 9 out of 11 Hyde1 genes (in five loci), was designated as Δ5-Hyde1 and was used for competition assay. The ΔT6SS mutant was provided by Ron Walcott's lab.

For the competition assay of *Acidovorax citrulli* AAC00-1 against different strains. *E. coli* BW25113 pSEVA381 was grown aerobically in LB broth (5 g/L NaCl) at 37 °C in the presence of chloramphenicol. Naturally antibiotic-resistant bacterial leaf isolates (Bai et al., 2015) and *Acidovorax* strains were grown aerobically in NB medium (5 g/L NaCl) at 28 °C in the presence of the appropriate antibiotic.

Competition assays were conducted similarly as described elsewhere (Basler et al., 2013; MacIntyre et al., 2010). Briefly, bacterial overnight cultures were harvested and washed in PBS (pH 7.4) to remove excess antibiotics, and resuspended in fresh NB medium to an optical density of 10. Predator and prey strains were mixed at a 1:1 ratio, and 5 μL of the mixture was spotted onto dry NB agar plates and incubated at 28 °C. As a negative control, the same volume of NB medium was mixed with prey cells instead of the predator strain. After 19 h of coincubation, bacterial spots were excised from the agar and resuspended in 500 μL of NB medium and then spotted on NB agar containing antibiotic selective for the prey strains. CFUs of recovered prey cells were determined after incubation at 28 °C. All assays were performed in at least three biological replicates

CHAPTER 3: THE EFFECTS OF SOIL PHOSPHORUS CONTENT ON PLANT MIOCROBIOTA ARE DRIVEN BY THE PLANT PHOSPHATE STARVATION RESPONSE[3]

Introduction

Phosphate starvation response (PSR) in nonmycorrhizal plants comprises transcriptional reprogramming resulting in severe physiological changes to the roots and shoots and repression of plant immunity. Thus, plant-colonizing microorganisms—the plant microbiota—are exposed to direct influence by the soil's phosphorus (P) content itself as well as to the indirect effects of soil P on the microbial niches shaped by the plant. The individual contribution of these factors to plant microbiota assembly remains unknown. To disentangle these direct and indirect effects, we planted PSR-deficient *Arabidopsis* mutants in a long-term managed soil P gradient and compared the composition of their shoot and root microbiota to wild-type plants across different P concentrations. PSR-deficiency had a larger effect on the composition of both bacterial and fungal plant-associated microbiota than soil P concentrations in both roots and shoots.

To dissect plant–microbe interactions under variable P conditions, we conducted a microbiota reconstitution experiment. Using a 185-member bacterial synthetic community (SynCom) across a wide P concentration gradient in an agar matrix, we demonstrated a shift in the effect of bacteria on the plant from a neutral or positive interaction to a negative one, as measured by rosette size. This phenotypic shift was accompanied by changes in microbiota composition: the

[3] A portion of the results presented in this chapter appeared as an article in:
Finkel OM*, Salas-González I*, Castrillo G*, Spaepen S, Law TF, Teixeira PJPL, Jones CD, Dangl JL. The effects of soil phosphorus content on plant microbiota are driven by the plant phosphate starvation response. PLoS Biology. 2019 Nov 13;17(11):e3000534.

genus *Burkholderia* was specifically enriched in plant tissue under P starvation. Through a community drop-out experiment, we demonstrated that in the absence of *Burkholderia* from the SynCom, plant shoots accumulated higher ortophosphate (Pi) levels than shoots colonized with the full SynCom but only under Pi starvation conditions. Therefore, Pi-stressed plants are susceptible to colonization by latent opportunistic competitors found within their microbiome, thus exacerbating the plant's Pi starvation.

Plants provide the primary energy source for terrestrial heterotrophs, most of which are microbial. The interaction of these microbial heterotrophs with plants ranges between the extremes of mutualistic symbiosis (Desbrosses & Stougaard, 2011) and pathogenesis (Dangl & Jones, 2001; Hacquard et al., 2016). However, the vast majority of plant-associated microbial diversity, the plant microbiota, lies between these 2 extremes, inducing more subtle, context-dependent effects on plant health (Castrillo et al., 2017; Hacquard et al., 2017; Haney et al., 2015). The microbiota consumes plant photosynthate (Eissfeller et al., 2013; Hernández et al., 2015; Pickles et al., 2017), and it provides benefits via protection from pathogens (Cha et al., 2016; Hu et al., 2016; Mendes et al., 2011; Santhanam et al., 2015; Shen et al., 2015) or abiotic stress (Yuan et al., 2016)(Vargas et al., 2014) or by increasing nutrient bioavailability (Castrillo et al., 2017; Hacquard et al., 2016; Hiruma et al., 2016; Vessey, 2003).

The plant microbiota is derived from the microbial community composition in soil (Edwards et al., 2015; Lundberg, Lebeis, Paredes, Yourstone, Gehring, Malfatti, Tremblay, Engelbrektson, Kunin, del Rio, et al., 2012; Yeoh et al., 2016) which is governed by its own set of ecological processes (Fierer & Jackson, 2006). Correlations with soil microbial diversity, and by derivation, with plant microbiota composition and diversity, were observed for soil abiotic factors, such as pH (Fierer & Jackson, 2006; Griffiths et al., 2011; Lauber et al., 2009; Zhalnina et al.,

2014), drought (Angel et al., 2010; Fitzpatrick et al., 2018; Griffiths et al., 2011; Naylor et al., 2017; Santos-Medellín et al., 2017; L. Xu et al., 2018), and nutrient concentrations (Fierer et al., 2012; Fierer & Jackson, 2006; Griffiths et al., 2011; Leff et al., 2015; Pantigoso et al., 2018; Ramirez et al., 2012; P. Yu et al., 2018). Soil nutrient concentrations, in particular orthophosphate (Pi)—the only form of phosphorus (P) that can be taken up by plants—produce comparatively modest changes in microbial community composition (Fabiańska et al., 2019; Leff et al., 2015). Nevertheless, available soil Pi concentrations influences where a plant–microbe interaction lies along the mutualism–pathogenicity continuum (Hiruma et al., 2016).

Nonmycorrhizal plants respond to phosphate limitation by employing a range of PSR mechanisms. These manifest as severe physiological and morphological changes to the root and shoot, such as lateral root growth prioritization, depletion of shoot Pi stores (Raghothama, 1999), and changes to root exudate profiles (Herrera Paredes et al., 2018; Ziegler et al., 2016); these changes can potentially affect both root and shoot microbiota. In *Arabidopsis*, most of the transcriptional PSR driving these physiological responses is controlled by the 2 partially redundant transcription factors PHOSPHATE STARVATION RESPONSE 1 (PHR1) and PHR1-LIKE (PHL1) (Bustos et al., 2010). As a result, the double mutant *phr1 phl1* has an impaired PSR and accumulates a low level of Pi. Pi transport into roots relies on the PHOSPHATE TRANSPORTER TRAFFIC FACILITATOR1 (PHF1) gene, which is required for membrane localization of high-affinity phosphate transporters (González et al., 2005). In axenic conditions, *phf1* mutants constitutively express PSR and accumulate low levels of Pi (González et al., 2005). In addition to inducing physiological changes, the plant's response to its nutrient status is also linked to its immune system. PHR1 negatively regulates components of the plant immune system, which can lead to enhanced pathogen susceptibility but also to the alteration of the plant's microbiota under

phosphate starvation (Castrillo et al., 2017). *Arabidopsis* microbiota are altered in *phr1 phl1* and *phf1* mutants (Castrillo et al., 2017; Fabiańska et al., 2019) in experiments using both natural and synthetic microbial communities (Castrillo et al., 2017).

Here, we examined (i) the effect of soil P content on plant microbiota composition, (ii) how PSR modulates the plant microbiota, and (iii) the interplay between PSR and soil P content in shaping the plant microbiota composition. We used a combination of greenhouse experiments with differentially P-fertilized soils, *Arabidopsis* PSR mutants and laboratory microcosms utilizing tractable synthetic bacterial communities. Using PSR mutants planted in P-amended soil, we demonstrate that the plant PSR regulators have a profound effect on the composition of root and shoot microbiota, overshadowing the effect of the soil P content. We constructed an ecologically tractable system utilizing a complex bacterial synthetic community (SynCom) as a model of the plant root microbiome and used this system to study the interactions between microbiota assembly and abiotic stress. We demonstrate deterministic responses of the SynCom members to changes in Pi concentrations, and we identify Pi-dependent shifts in community composition along the mutualist–pathogen continuum.

3.1 The plant phosphate starvation response is activated in soil

To better understand the effect of PSR genes on the plant microbiome under both Pi-limiting and Pi-replete conditions, we investigated how microbiota adapted to varying soil P levels interact with the plant's PSR. We grew wild-type (wt) *Arabidopsis* and the PSR mutants *phf1* and *phr1 phl1* in soils collected from the "Halle long-term soil fertilization experiment," ongoing since 1949 (Gransee & Merbach, 2000). Throughout this long-term experiment, each transect of soil has received 1 of 3 P fertilization treatments: 0 (low), 15 (medium) and 45 (high) Kg[P].Ha^{-1}.$Year^{-1}$, resulting in a 3- to 5-fold difference in bioavailable P

between the low and high treatments (Robbins et al., 2018). To differentiate the long-term adaptive effect of P limitation on the microbial community from the effect of short-term changes in P availability, we also fertilized a subset of the low P soil at the time of planting and designated this condition low+P.

We examined whether PSR, defined and typically studied in axenic conditions, is active in our soil-based experimental system. We harvested 8-week-old plants grown in the different soils and quantified developmental and molecular phenotypes typically associated with PSR in both wt plants and mutants. We found a strong positive correlation among all developmental features analyzed: shoot area, shoot fresh weight, and shoot Pi accumulation across all soil conditions (Figures 3.1a and 3.7a-c). In wt plants, shoot Pi accumulation reflected soil P conditions (Figure 3.1a). As expected (González et al., 2005), *phr1 phl1* showed a dramatic reduction in all phenotypic parameters (Figures 3.1a and 3.7b,c) and *phf1* accumulated less shoot Pi than wt but did not display any obvious morphological effect (Figures 3.1a and 3.7b-d).

To identify the transcriptomic signature of PSR in a low P soil, we compared the root transcriptomes of the 3 genotypes from the low P samples with those of the low+P samples. Using a likelihood ratio test, we identified 210 genes that were differentially expressed across genotypes and P conditions ($q < 0.1$). After hierarchical clustering, 123 (59%) of these genes fall into a single cluster (Cluster 1) of co-expressed genes that are exclusively highly expressed in wt under low P and not in either of the PSR mutants (Figure 3.1b). Thus, these genes represent a PSR under our experimental conditions. A gene ontology (GO) enrichment analysis (Figure 3.1d) illustrates that these genes are involved in processes such as ion homeostasis, detoxification, and response to oxidative stress. Interestingly, few PSR genes defined from in vitro experiments were significantly differentially expressed in our soil experiment. From a previously defined set of 193 PSR marker

genes defined using 7-day-old seedlings exposed in vitro to P limitation for up to 2 days (Castrillo et al., 2017), only 7 were called as significant in our experiment using 8-week-old plants. Nevertheless, all 7 of these genes were enriched in wt in low P soil (Figure 3.1b). Surprisingly, despite the fact that *phf1* and *phr1 phl1* have contrasting transcriptional responses to Pi limitation in vitro (González et al., 2005), the 123 genes in Cluster 1 were not up-regulated in both mutants. To corroborate that the canonical in vitro defined PSR is also induced in wt plants, we compared the median expression of the set of 193 PSR marker genes (Castrillo et al., 2017) across the different soils (Figures 3.1c and 3.7e). As expected, shoot Pi content was significantly correlated with the induction of PSR marker genes (Figures 3.7c, f), with the highest median expression level in the low P conditions. We conclude that although the response of 8-week-old plants to low P conditions in natural soil is markedly different from in vitro defined PSR, wt plants indeed respond to low P conditions in the soils tested in a Pi concentration- and PHR1/PHL1-dependent manner.

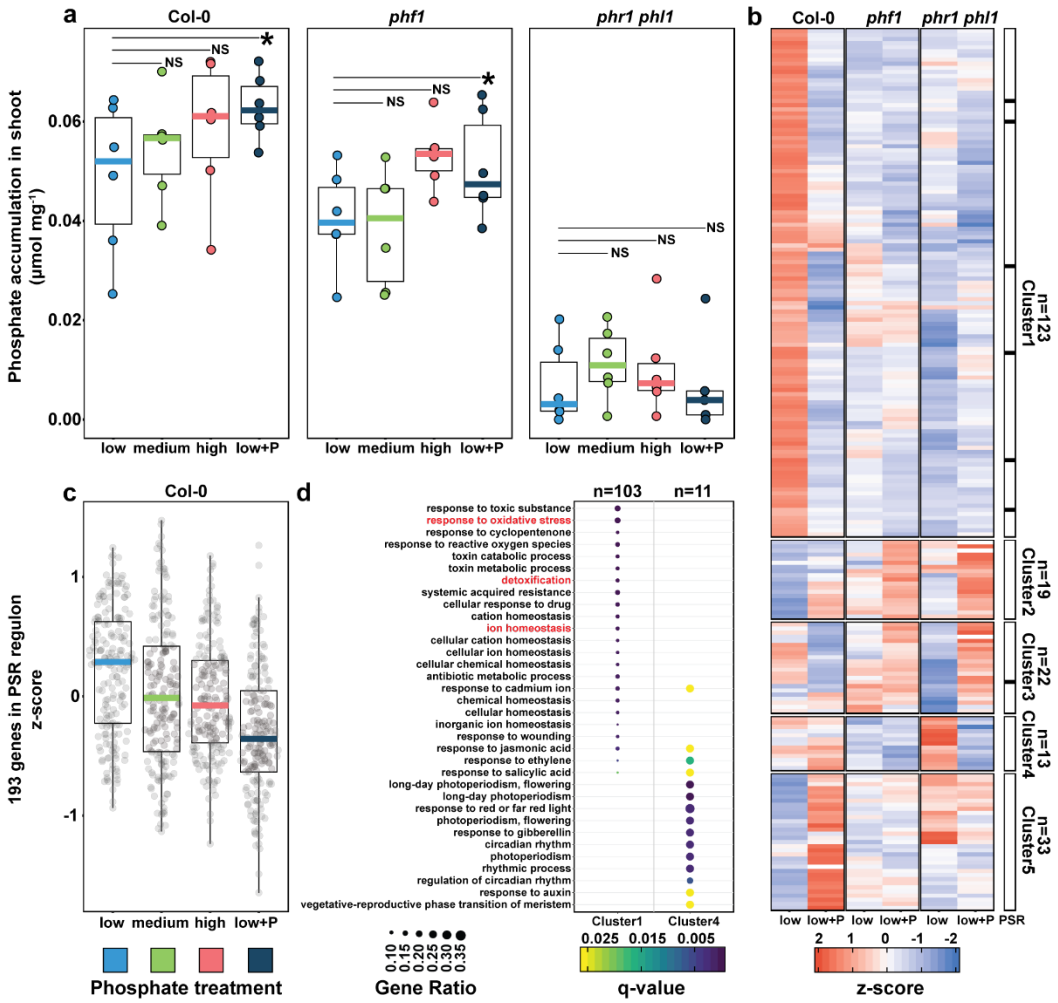


Figure 3.1 - **Plants respond to differential P conditions in soil.**

a) Free phosphate content normalized by shoot fresh weight (mmol·mg−1) across wt Col-0 plants and 2 PSR mutants, phf1 and phr1 phl1. Statistical significance between low P and low+P treatments was determined across each genotype independently by a paired t test ($p < 0.05$). **b)** Heat map showing the average standardized expression of 210 DEGs across the low P and low+P samples in the Col-0, phf1 and phr1 phl1 genotypes. The black bar to the right highlights the distribution of 7 genes belonging to the in vitro defined PSR marker genes across the 5 clusters in the heat map. **c)** Average expression of 193 PSR marker genes across the 4 phosphorus regimes in the Col-0 genotype. **d)** GO enrichment for Clusters 1 and 4. Clusters 2, 3, and 6 did not show any statistically significant GO enrichment. The gene ratio is the proportion of genes per cluster that belong to a GO category. DEG, differentially expressed gene; GO, gene ontology; P, phosphorus; PSR, phosphate starvation response; wt, wild type.

3.2 Bacterial and fungal plant microbiota differ in plant recruitment patterns

We studied the relationship between PSR and the plant microbiome in wt plants and the 2 PSR mutants grown in all 4 soils. Total DNA was extracted from shoots, roots, and soil, and the 16S rRNA (V3-V4) and ITS1 regions were amplified and sequenced to obtain bacterial and fungal community profiles, respectively. Sequences were collapsed into amplicon sequence variants (ASVs. Bacterial and fungal alpha- and beta-diversity measures conform to previously published data (Fabiańska et al., 2019; Lundberg, Lebeis, Paredes, Yourstone, Gehring, Malfatti, Tremblay, Engelbrektson, Kunin, del Rio, et al., 2012): Microbial diversity decreased from the soil to the root and shoot compartments (Figures 3.2a, d and 3.8a, b), and roots and shoots harbor bacterial and fungal communities distinct from the surrounding soil community and from each other (Figures 3.2b, e and 3.8c-f). Plant-derived samples were primarily enriched in comparison to soil with members of the phyla Proteobacteria, Bacteroidetes, and Actinobacteria and depleted in members of Acidobacteria and Gemmatimonadetes (Figures 3.2c and 3.8e). Plant-enriched fungal ASVs belonged mainly to the phyla Ascomycota (orders Hypocreales and Pleosporales) and Basidiomycota (order Agaricales). Plant-depleted fungal ASVs belonged mainly to Filobasidiales (Basidiomycota) and Mortierellales (Zygomycota) (Figures 3.2f and 3.8f).

To quantify the effect of soil community composition on the composition of root and shoot microbiota, we used Mantel tests to detect correlation between dissimilarity matrices of the 3 fractions (root, shoot, and soil). For bacteria, both root and shoot community dissimilarities were strongly correlated with soil community dissimilarity (Figures 3.9a, b), whereas for fungi, no correlation was measured between root and soil (Figure 3.9d), and only a weak correlation was measured between shoot and soil (Figure 3.9e). This observation indicates that both root and shoot bacterial communities are strongly dependent on soil community composition, despite the fact that

bacterial microbiota is distinct from the soil community (Figure 3.2b). By contrast, the fungal microbiota composition both above and below ground is independent of relative abundances within the soil inoculum. This difference implies that the plant's microbiota filtering mechanisms are fundamentally different for fungi and bacteria.

3.3 Shoot and root microbiota are both correlated and distinct

Shoot and root microbiomes are linked, and substantial crosstalk is expected to occur between these 2 niches (Bai et al., 2015; Knief et al., 2012). We show here that roots and shoots harbor distinct communities from each other (Figures 3.2b, e and 3.8c-f). To further explore organ specificity in the plant microbiome composition, we compared root and shoot samples at the ASV level. Shoots were mainly enriched with the bacterial phyla Cyanobacteria and Patescibacteria compared to the root, whereas roots were enriched with Proteobacteria, Chloroflexi, and Bacteroidetes (Figures 3.2c and 3.8e). With regard to fungal orders, shoots were enriched with Capnodiales, Glomerellales, Pleosporales, and Hypocreales, whereas roots were enriched with Pezizales, Helotiales, and Mucorales (Figures 3.2f and 3.8f). The shoot enrichment of Cyanobacteria suggests that the availability of light is an important factor in niche differentiation within the plant (Atamna-Ismaeel, Finkel, Glaser, Sharon, et al., 2012; Atamna-Ismaeel, Finkel, Glaser, von Mering, et al., 2012; Finkel et al., 2016). We used Mantel tests to detect correlation between dissimilarity matrices of root and shoot samples. Despite the fact that they harbor distinct communities, roots and shoots were correlated with each other for both bacteria and fungi (Figures 3.9c, f). Thus, although roots and shoots form distinct bacterial and fungal niches, shifts in microbiota in both of these niches are correlated, suggesting that either one or both of these fractions serves as an inoculum source for the other.

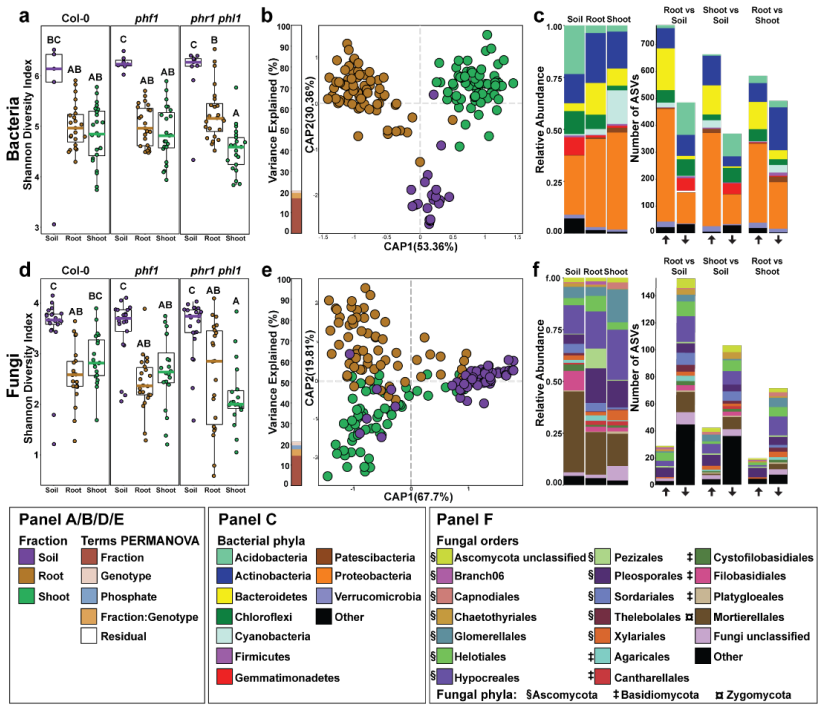


Figure 3.2 - **Plant recruitment patterns of bacteria and fungi.**

a, d) Bacterial (a) and fungal (d) alpha diversity estimated using the Shannon Diversity Index. Letters represent post hoc test results, based on a full factorial ANOVA model. **b, e)** CAP based on Bray-Curtis dissimilarities between bacterial (b) and fungal (e) communities across the soil, root, and shoot. The bar graph to the left of the CAP depicts the percentage of variance explained by statistically significant ($p < 0.05$) terms in a PERMANOVA model. **c)** Left panel: Relative abundance profiles of the main bacterial phyla across the soil, root, and shoot fractions. Right panel: Number of statistically significant ASVs enriched in specific fractions. The arrows on the bottom of the panel denote the direction of the enrichment relative to the name of the contrast tested; the up arrow means enrichment in the left fraction of the contrast, whereas the down arrow means enrichment in the right fraction of the contrast (e.g., RootvsSoil, up arrow enriched in root relative to soil, bottom arrow enriched in soil relative to root). A detailed interactive visualization of the bacterial enrichment patterns across the multiple taxonomic levels can be found at https://itol.embl.de/tree/152231625417470155198725**3. f)** Left panel: Relative abundance profiles of the main fungal orders across soil, root, and shoot fractions. Right Panel: Number of statistically significant ASVs enriched in specific fractions. The arrows on the bottom of the panel denote the direction of the enrichment relative to the name of the contrast tested; the up arrow signifies enrichment in the left fraction of the contrast, whereas the down arrow signifies enrichment in the right fraction of the contrast (e.g., RootvsSoil, up arrow enriched in root relative to soil, bottom arrow enriched in soil relative to root). Plot is colored by order. The symbols besides the colors in the legend denote phylum. A detailed interactive visualization of the fungal enrichment patterns across the multiple taxonomic levels can be found at https://itol.embl.de/tree/136561721374648315710970**84**. ASV, amplicon sequence variant; CAP, canonical analysis of principal coordinates; PERMANOVA, Permutational Multivariate Analysis of Variance.

3.4 The plant microbiome composition is driven by the plant PSR status

We investigated the influences of plant PSR signaling and the different soil P concentrations on microbial community composition. Constrained ordination showed significant differences, explaining a similar proportion of variance, in both bacterial and fungal community compositions. These differences persist across the P accumulation gradients represented by both the difference in soil P and PSR mutants (Figures 3.3a, b and 3.10a, b). This effect is maintained also when considering 13 soil edaphic factors that were measured in for the same soil samples (Robbins et al., 2018) (Figures 3.11a-d). For both bacteria and fungi, the PSR genotype effect in roots was more consistent than the soil P effect (Figures 3.3a, b), which was mainly driven by the P-amended low+P samples. In shoots, both bacteria and fungi responded to PSR genotype but in this case did not respond to soil P (Figures 3.10c, d). We did not observe a significant soil P:genotype interaction effect for either bacteria or fungi (Figures 3.10a, b), confirming that *phf1* and *phr1 phl1* both had atypical bacterial microbiomes regardless of Pi status. As expected, we did not observe a PSR effect in the soil samples (Figures 3.10e, f). The notable genotype effect illustrates that the plant niche filtering (Figures 3b, e) is partly shaped by PSR.

To define which taxa at the ASV-levels were influenced by soil P and/or plant PSR, we applied a generalized linear model (GLM) to the count datasets. Contrasting the low P samples against the low+P samples, we detected 769 bacterial and 140 fungal ASVs, accounting for 23% and 33% of the bacterial and fungal abundance in the root, respectively, that were differentially abundant in at least 1 genotype (Figures 3.3c-f). Of these, most (568 bacterial and 85 fungal ASVs) were genotype specific, suggesting that the Pi response of these taxa is not direct but is rather driven by Pi responses in the plant. Taken together, these results indicate that plant microbiota are relatively robust to differences in soil P content but are sensitive to the plant PSR status. Responses

to soil P concentration are contingent on PSR regulatory elements under both low and high P conditions.

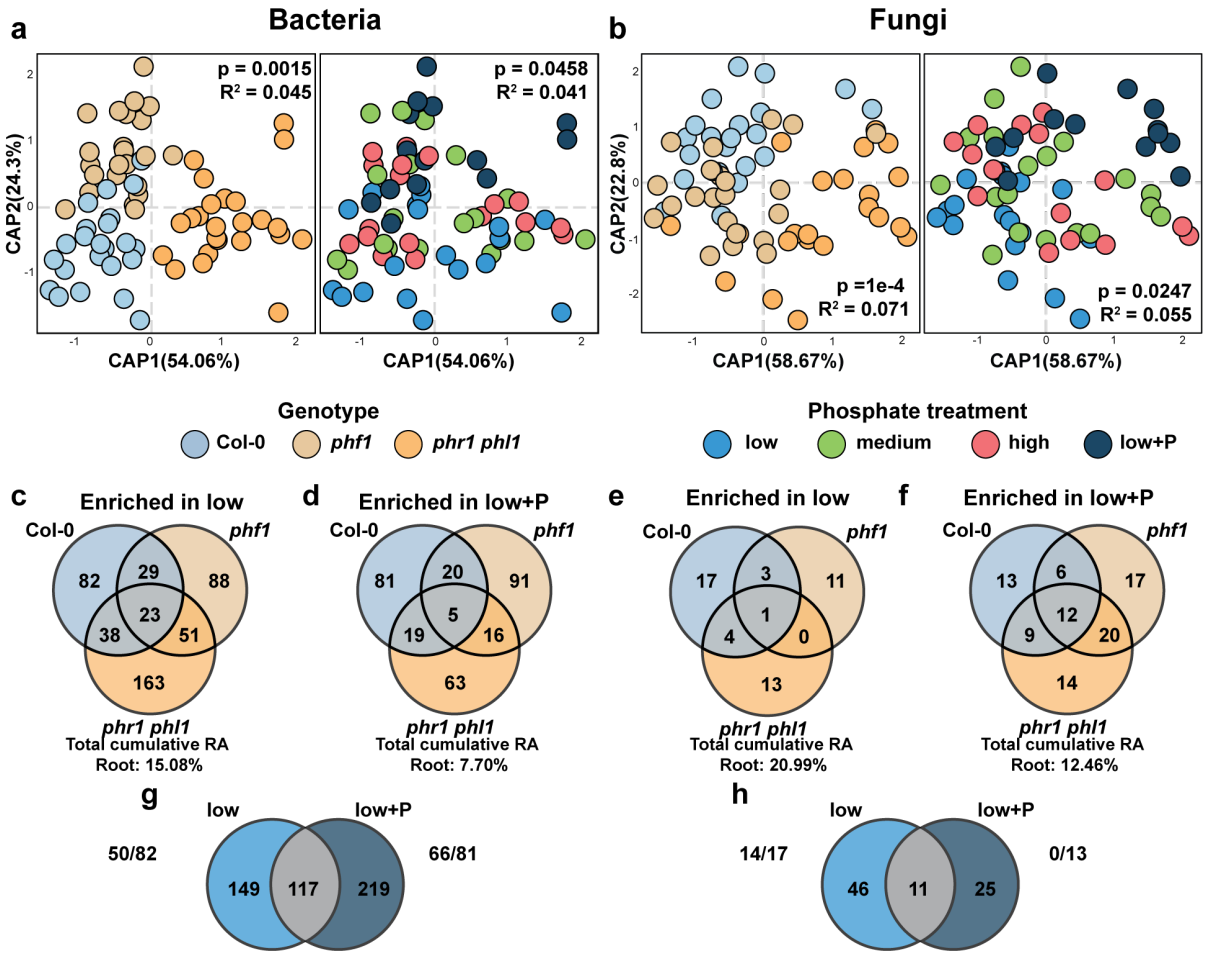


Figure 3.3 - **Plant PSR controls the assembly of the plant microbiome.**

a, b) Canonical analysis of principal coordinates showing the influence of plant genotypes and soil P content over the (a) bacterial and (b) fungal communities in the root. The p-value and R2 values inside each plot are derived from a PERMANOVA model and correspond to the genotype and P term, respectively. **c, e)** Venn diagrams showing the distribution of (c) bacterial and (e) fungal ASVs with statistically significant ($q < 0.1$) higher abundance in the low P treatment in comparison to the low+P treatment in the Col-0, phf1 and phr1 phl1 roots. **d, f)** Venn diagrams showing the distribution of (d) bacterial and (f) fungal ASVs with statistically significant ($q < 0.1$) higher abundance in the low+P treatment in comparison to the low P treatment across the Col-0, phf1 and phr1 phl1 roots. ASV, amplicon sequence variant; P, phosphorus; PERMANOVA, Permutational Multivariate Analysis of Variance; PSR, phosphate starvation response; RA, relative abundance.

3.5 Bacterial synthetic communities modulate the plant PSR

The results obtained from the soil experiment suggest that the community structure of the plant microbiome is not only determined by first-order interactions (plant–microbe, microbe–microbe, microbe–environment) but also by higher-order interactions, such as the effect of abiotic conditions on plant–microbe interactions. This is evident in the large proportion of ASVs that respond to soil P in a genotype-specific manner (Figures 3.3c-f). To establish a system in which interactions of different orders of complexity can be studied reproducibly, we constructed a plant–microbe microcosm that can be deconstructed to its individual components while retaining a complexity that is comparable to natural ecological communities. We designed a representative bacterial SynCom from a culture collection composed of isolates derived from surface-sterilized *Arabidopsis* roots (Levy et al., 2018). We selected 185 genome-sequenced isolates representing a typical plant-associated taxonomic distribution (Figures 3.4a, b). We grew each isolate separately and mixed the grown cultures to equal optical densities. We grew 7-day-old *Arabidopsis* seedlings in a Pi concentration gradient (0, 10, 30, 50, 100, 1000 μM KH_2PO_4) and concomitantly exposed them to the SynCom on vertical agar plates for 12 days.

First, we investigated whether PSR is induced in our experimental system. Similar to the natural soil-based experiment, we quantified developmental and transcriptional phenotypes associated with PSR in plants grown in different concentrations of Pi. In line with the work by Castrillo and colleagues (Castrillo et al., 2017), the presence of the SynCom consistently decreased primary root elongation across all Pi concentrations compared with the uninoculated control, but the Pi gradient did not affect this parameter (Figure 3.12a). Shoot size increased with Pi concentration, and the slope of this trend was affected by the presence of the SynCom: At high Pi, the SynCom tended to increase shoot size, whereas at low Pi the SynCom decreased it (Figure

3.5a), suggesting that the microbiome plays a role in shaping the plant's response to different Pi concentrations.

We performed RNA sequencing (RNA-Seq) on inoculated and uninoculated seedlings exposed to high (1,000 μM) and low (50 μM) Pi. To confirm that our low Pi treatments induce PSR, we examined the expression of the 193 PSR markers defined in the work by Castrillo and colleagues (Castrillo et al., 2017). We found that 168 of the 193 PSR markers genes were significantly induced in uninoculated plants at low Pi compared with high Pi conditions. In the presence of the SynCom, 184 out of 193 PSR marker genes were significantly induced, and the average fold change increased from 4.7 in uninoculated conditions to 11 in the presence of the SynCom (Figure 3.12b). We further examined whether the 123 low Pi-responsive genes from the soil experiment (Cluster 1 in Figure 3.1b) are overexpressed in the agar system as well. We found that 59 of the 123 genes (47.2%) were low Pi-enriched in uninoculated plants and 72 (58.5%) were low Pi-enriched in the presence of the SynCom. The average fold change for this set of 123 genes was 1.6 in uninoculated conditions and 2.0 in the presence of the SynCom (Figure 3.12c). These results confirm that (i) in both our systems, wild soils and axenic conditions, PSR is induced at low Pi, and (ii) the SynCom enhances this induction, similar to the results reported in the work by Castrillo and colleagues (Castrillo et al., 2017).

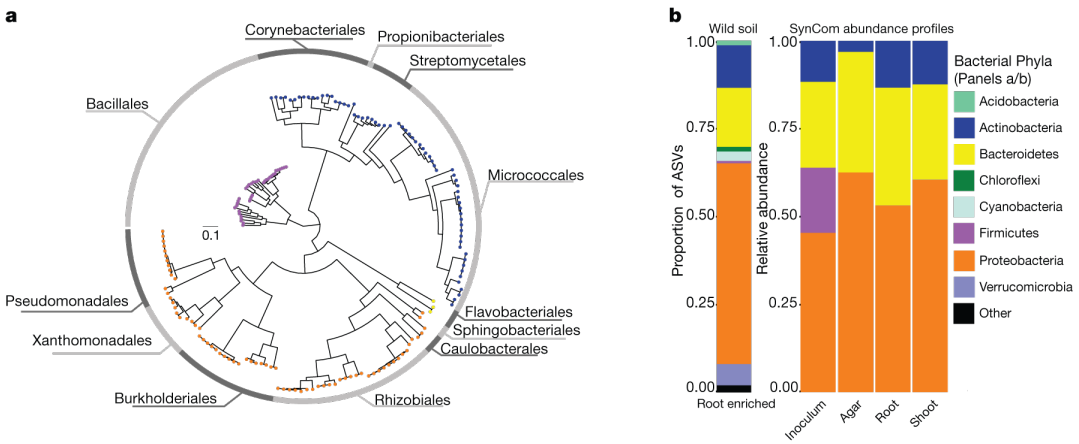


Figure 3.4 - **Bacterial SynCom reproduces the typical plant-associated taxonomic distribution found in soil.**

a) Phylogenetic tree of 185 bacterial genomes included in the SynCom. The tree tips are colored according to the phylum classification of the genome in panel B; the outer ring shows the distribution of the 12 distinct bacterial orders present in the SynCom. **b)** Left Panel: Proportion of ASVs enriched in the root in comparison to the natural soil across all treatments and genotypes based on a fitted GLM ($q < 0.1$). Each ASV is colored according to its phylum-level classification. Right Panel: Relative abundance profiles of bacterial isolates across the initial bacterial inoculum, planted agar, root, and shoot fractions. Each isolate is colored according to its phylum-level classification based on the genome-derived taxonomy. ASV, amplicon sequence variant; GLM, generalized linear model; SynCom, synthetic community.

3.6 Bacterial SynComs display deterministic community assembly in plants

To quantify the establishment of the SynCom in the plants, we determined bacterial community composition after 12 days of co-inoculation in roots, shoots, and agar via 16S rRNA gene amplicon sequencing, mapping reads to 97 unique sequences (USeqs) representing the 185-strain SynCom. We found that plant roots and shoots sustained a higher bacterial alpha diversity than the surrounding agar (Figure 3.5b), an aspect in which our experimental system differs from a natural environment where species richness is higher in the surrounding soil than in the plant (Figure 3.2a). As in natural soil experimental systems, agar, roots, and shoots assembled distinct bacterial communities, and this difference among these 3 fractions explained most of the variance in community composition despite the different Pi concentrations (Figures 3.5c and 3.12d).

To study which strains are enriched in the roots and shoots under the different Pi concentrations, we utilized a generalized linear model (GLM). Noticeably, plant (root and shoot) enrichment is strongly linked to phylogeny (Figure 3.5d) and is robust across the phosphate gradient assayed. In contrast, the root versus shoot comparison did not exhibit a significant phylogenetic signal, highlighting the fact that the ability to differentially colonize the shoot from the root under these conditions is phylogenetically scattered across the SynCom. As in the soil census, shoot, root, and agar beta diversities were significantly correlated (Figures 3.12e-g).

We hypothesized that by establishing a standardized protocol for producing the inoculum and controlling the growth conditions, we will have created a reproducible system in which most of the variance can be accounted for. To test this, we compared the number of ASVs and total relative abundance captured by the fraction (root/shoot versus soil/agar) in a GLM in the natural community experiment versus the SynCom experiment. Supporting our hypothesis, only 1,518 out of 3,874 measurable ASVs (32% of the total ASVs), accounting for 72% of the relative abundance in plant tissue, shift significantly between root and soil in the natural community survey, whereas 58 out of 97 USeqs (59%), accounting for 99% relative abundance in plant tissue, were significantly enriched or depleted in plant tissue in the SynCom experiment (Figure 3.5d). These results indicate that plant colonization is largely deterministic in our SynCom system, in comparison to microbiomes in nature. The reproducibility of this system, coupled with our ability to edit it as a tool for hypothesis-testing, is crucial to bridge ecological observation with mechanistic understanding of plant–microbiota interactions.

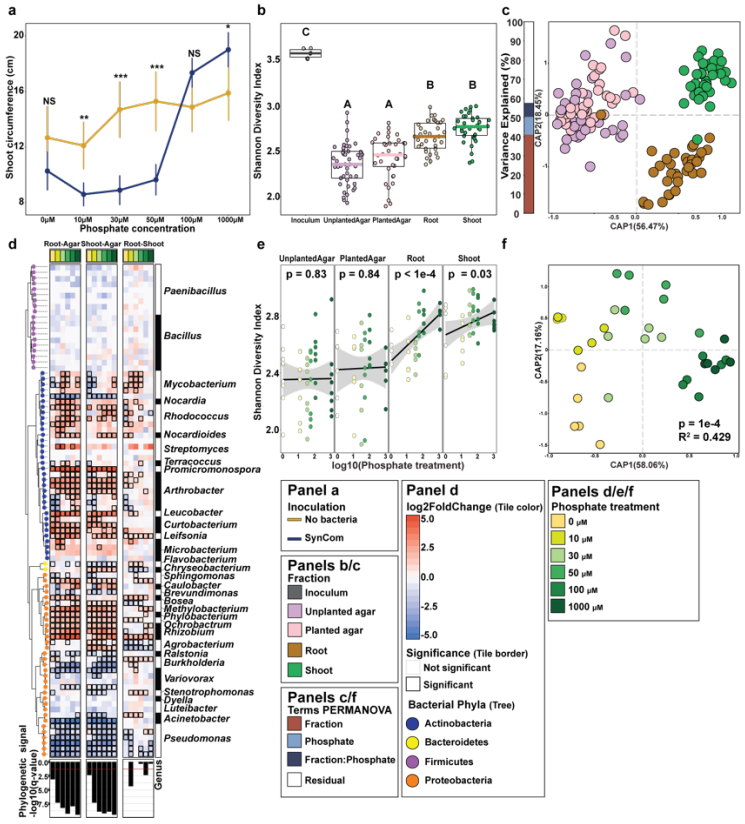


Figure 3.5 - **Synthetic bacterial communities display deterministic community assembly in plants.**

a) Strip chart displaying the average shoot size of Col-0 Arabidopsis grown across a Pi gradient either in sterile conditions or with the SynCom. Each dot in the scatter plot represents the mean value for that particular treatment; the range crossing each dot represents the 95% confidence interval calculated. The lines are drawn to connect the means. **b)** Alpha diversity across the fractions sampled was estimated using the Shannon Diversity index. An ANOVA model followed up by a Tukey HSD test were applied to estimate differences between inoculum, unplanted agar, planted agar, root, and shoot fractions. Letters represent the results of the post hoc test. **c)** CAP based on Bray-Curtis dissimilarities between bacterial communities across the 4 fractions sampled. The bar graph to the left of the CAP depicts the percentage of variability explained by statistically significant ($p < 0.05$) terms in the PERMANOVA model. **d)** Enrichment patterns of the SynCom. Each row along the different panels of the figure represents a USeq: a USeq encompasses a set of indistinguishable V3-V4 16S rRNA sequences present in the 185-member SynCom. Phylogenetic tree (on the left) is colored based on the phylum-level classification of the corresponding USeq. Each column in the heat maps represents a specific contrast in the enrichment model. We calculated root versus agar (left heat map), shoot versus agar (middle heat map), and root versus shoot (right heat map) enrichments within each Pi treatment (e.g., Root_0Pi versus Agar_0Pi). The heat maps are colored based on log2 fold changes derived from the fitted GLM. Positive fold changes (colored in red gradient) represent enrichments on the left side of the name of the contrast (e.g., Root-Agar, enriched in root in comparison to agar), whereas negative fold changes (colored

in blue gradient) represent enrichments on the right side of the name of the contrast (e.g., Root-Agar, enriched in agar in comparison to agar). Boxed cells represent statistically significant enrichment/depletion. The bottom panel depicts the transformed (−log10) q-value derived from a phylogenetic signal Pagel's λ test. Tests were performed per column in the heat map (e.g., Root0μM Pi versus Agar0μM Pi). **e)** Bacterial alpha diversity estimated using the Shannon Diversity index. p-values derived from a linear model are shown for each fraction. Linear regression line is shown in black and the 95% confidence interval is shaded in gray. **f)** CAP showing the influence of phosphate on the bacterial communities in the root. The bar graphs to the left of the CAP depict the percentage of variability explained by statistically significant ($p < 0.05$) variables based on a PERMANOVA model. CAP, canonical analysis of principal coordinates; GLM, generalized linear model; HSD, Honestly Significant difference; PERMANOVA, Permutational Multivariate Analysis of Variance; Pi, orthophosphate; SynCom, synthetic community; USeq, unique sequence.

3.7 P stress-induced changes in the root microbiome

The shifting role of the SynCom from increasing shoot size under replete Pi to decreasing shoot size and PSR induction under Pi limitation (Figures 3.5a and 3.12b) can be explained by either a shift in the lifestyle of individual bacteria along the mutualist–pathogen continuum or by changes in the microbiota composition along the Pi gradient. The latter would favor the proliferation of mutualist bacteria only when sufficient nutritional requirements are met. To measure the effect of the Pi concentration in the media on the SynCom composition in wt plants, we measured alpha and beta diversity along our Pi gradient (0, 10, 30, 50, 100, 1000 μM KH_2PO_4) in roots, shoots, and agar. We observed a positive correlation between alpha diversity and Pi concentrations, resembling a partial ecological diversity–productivity relationship—the prediction/observation of a bell-shaped response of ecological diversity to environmental productivity (Kassen et al., 2000)—in roots and shoots but not in the surrounding agar (Figure 3.5e). As for beta diversity, the composition of the SynCom shifted significantly along the Pi concentration gradient (Figures 3.5f and 3.13a-e). Pi concentration therefore alters the plant

microbiome, shifting from a net-positive outcome for the plant to a net-negative one as measured by shoot size (Figure 3.5a).

3.8 Burkholderia respond to Pi stress-induced changes in the plant

In a previous publication (Castrillo et al., 2017), we demonstrated that PHR1 negatively regulates defense-related genes under low Pi conditions. Suppression of plant defense and consequent alterations in colonization could account for some of the shift we observed from a beneficial to a detrimental community. We thus aimed to identify bacteria that respond to Pi stress-induced changes in the plant, rather than the Pi concentration itself. To do so, we searched for USeqs that displayed a strong Pi:fraction (shoot, root, agar) interaction in our GLM (Figure 3.14a). Two of the three USeqs displaying the strongest Pi:fraction interaction belonged to *Burkholderiaceae*, representing all 5 *Burkholderia* strains used in this experiment. The relative abundance of these USeqs is positively correlated with Pi concentration in the agar but is negatively correlated with Pi concentration in the root and shoot (Figures 3.6a and 3.14a-c). This pattern suggests that these strains are responding to physiological changes in the plant.

To measure the physiological effect of the specific recruitment of *Burkholderia* under Pi stress on the plant, we conducted a drop-out experiment in which we compared plants inoculated with the full SynCom to plants inoculated with a SynCom excluding all 5 *Burkholderia* isolates. We also included a SynCom excluding all members of the neighboring *Ralstonia* clade (Figure 3.4a), which did not display any discernible Pi response. We measured shoot size and Pi concentrations in the shoots (a proxy for PSR) of plants grown in high (1,000 μM) and low (50 μM) Pi with the different SynComs. In addition, we measured shoot size and Pi content in a refeeding treatment with SynCom-inoculated plants grown in low (50 μM) Pi and then transferred to high Pi (1,000 μM) conditions. As seen before (Figure 3.5a), the SynCom decreased shoot sizes

in low Pi and increased them in high Pi. However, the different taxon drop-outs did not affect shoot size compared with the full SynCom, except for a slight decrease in post-refeeding shoot size in the *Ralstonia* drop-out treatment (Figure 3.15). All SynCom treatments decreased shoot Pi content in the low Pi conditions compared with the uninoculated plants but recovered to a higher shoot Pi level than the uninoculated treatments upon transferring to high Pi conditions, reproducing our previous report (Castrillo et al., 2017) (Figure 3.6b). Among inoculated treatments, plants colonized with the *Burkholderia* drop-out treatment (SynCom excluding all *Burkholderia*) had a higher Pi content than either plants colonized with the full SynCom or with the *Ralstonia* drop-out SynCom only in the low Pi conditions. There was no difference in shoot Pi among the SynCom treatments in either the high Pi treatment or following the refeeding treatment. This finding suggests that the enrichment of *Burkholderia* in plant tissue under Pi starvation can be considered a shift in the effect of bacteria on the plant from a positive interaction to a negative one.

To test whether *Burkholderia* are recruited to the plant under low Pi via a PSR-dependent mechanism, we inoculated 7-day-old wt, *phf1*, and *phr1 phl1* seedlings with the SynCom and profiled the community composition after 12 days of growth in 3 Pi concentrations (0, 50, and 1,000 μM). In accordance with our soil experiment, and with the work by Castrillo and colleagues (Castrillo et al., 2017), community composition assembled in the roots of the 3 genotypes differed significantly (Figure 3.16). However, *Burkholderia* sequences were enriched in low Pi in all 3 genotypes (Figure 3.6c), indicating that their recruitment to the root under low Pi is independent of PSR activation and of the immune dampening that accompanies it.

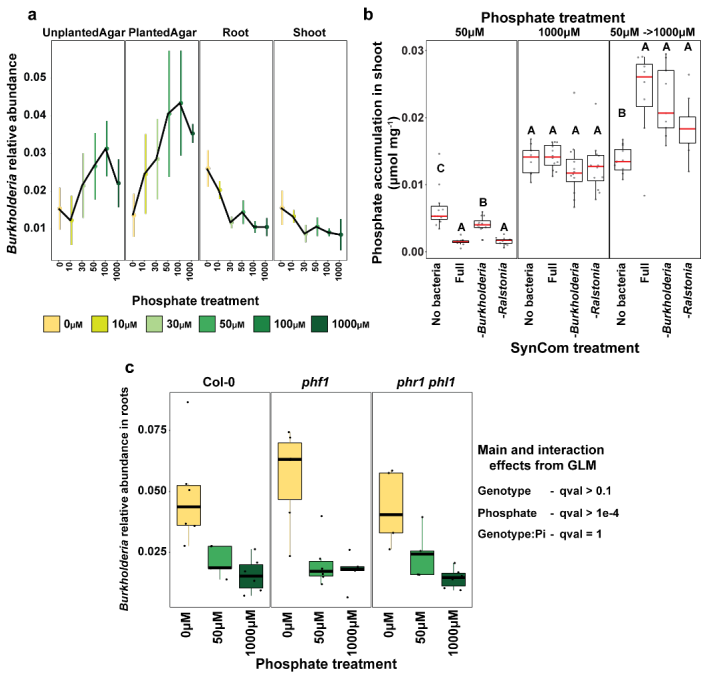


Figure 3.6 - **Bacterial strains respond to Pi-stress-induced physiological changes in the wt plants.**

a) Relative abundance of Burkholderia USeqs, both of which exhibit a statistically significant ($q < 0.1$) Pi-enrichment between the plant fractions and the agar fraction. The middle dot of each strip bar corresponds to the mean of that particular condition, the range of the strip bar corresponds to the 95% confidence interval of the mean. The lines are drawn connecting the means for each Pi concentration. **b)** Box plots showing the phosphate accumulation in plants exposed to different SynComs across 3 phosphate treatments. Statistically significant differences among SynCom treatments were computed within each phosphate treatment separately using an ANOVA model. Letters represent the results of the post hoc test. **c)** Box plot showing relative abundance of Burkholderia USeqs across 3 Pi concentrations and 3 plant genotypes. Summary of the NB-GLM for Burkholderia is shown on the right. GLM, generalized linear model; NB, no bacteria; Pi, orthophosphate; SynCom, synthetic community; USeq, unique sequence; wt, wild type.

3.9 Discussion

This study shows that despite 60 years of differential fertilization, the plant's PSR and accompanying changes to its microbiome composition between the low P and high P soils are subtle, possibly because Pi status of the plant is highly buffered by the plant ionomic regulatory network (Baxter et al., 2008). Only when comparing the low P versus the P-supplemented low+P

samples is there a discernible difference in both plant transcriptome and shoot Pi accumulation, which correlates to a stronger effect on microbiota composition. This suggests that bioavailable Pi added to the soil is quickly consumed, and short-term amendments are needed in order to detect changes. A similar short-term fertilization approach was successfully applied in the work by Fabiańska and colleagues (Fabiańska et al., 2019), using soils from a different long-term experimental site.

As opposed to in vitro defined transcriptomes (González et al., 2005), here, *phf1* and *phr1 phl1* plants, although each having a distinct transcriptional profile, were impaired in the response of a core set of genes (Cluster 1 in Figure 3.1b) to low P. This is despite the fact that *phf1* plants were not impaired in growth (Figures 3.7b, c) and were only slightly impaired in shoot Pi accumulation (Figure 3.7d). A paralogous discrepancy was also observed in the work by Hiruma and colleagues (Hiruma et al., 2016), in which both *phf1* and *phr1 phl1* were impaired in their response to a plant growth promoting fungus, whereas only *phr1 phl1* but not *phf1* was impaired in P translocation to the shoot, raising the speculation that PHF1 is needed for fine, rather than bulk, P translocation within the plant. Our results support this idea and raise a similar speculation that PHF1-guided transporters are involved in fine grained phosphate translocation within the plant toward Pi sensing tissues.

An additional apparent departure from the axenically defined PSR transcriptional signature is that only 7 of the genes that were induced in wt plants under low Pi conditions overlapped with the list of 193 axenically defined PSR genes we used as a reference. It is notable, however, that despite the difference in plant age and environmental conditions between our soil and axenic agar systems, the agar-defined transcriptomic signature is detectable in the soil system (Figure 3.1c) and vice versa: the signature of the 123 genes identified in our soil system is detectable on agar

(Figure 3.12c). Thus, despite the apparent differences, agar-defined PSR signatures are also maintained under complex real-life conditions.

Several studies link host physiological response to the soil phosphate status with the bacterial (Castrillo et al., 2017; Herrera Paredes et al., 2018) and fungal (Fabiańska et al., 2019; P. Yu et al., 2018) microbiome. A recent report of *Arabidopsis* planted in a 60-year-long annual P fertilization gradient (the same soil used in the current study) showed a modest P effect on plant microbiome composition (Robbins et al., 2018). Previously, we showed that PSR mutants in *Arabidopsis* have different bacterial microbiomes in Pi replete conditions (Castrillo et al., 2017), and a recent publication showed that PSR mutants had a slightly altered fungal microbiome in Pi-replete but not in Pi-depleted conditions (Fabiańska et al., 2019). Here, we analyzed fungi and bacteria side by side and demonstrated a pronounced effect of PSR impairment on both bacterial and fungal components of the plant microbiota. We noted an intriguing difference that emerged in the patterns of community assembly between bacteria and fungi (Figures 3.9a, b, d, e). The bacterial microbiota composition is strongly dependent on the soil bacterial community composition, whereas changes to the fungal microbiota are uncoupled from changes to the soil fungal community composition. This indicates that the plant is markedly more selective as to the fungi allowed to proliferate in its tissue than it is with bacteria. The observation that much of the modulation of the plant's fungal microbiota is mediated by the bacterial microbiota itself (Durán et al., 2018) may also contribute to this complex pattern. Our results show that impairment of PSR genes profoundly affects the composition of the plant microbiota, independently of P conditions, and that observed shifts in root-derived microbial communities may not be a result of sensitivity to P concentrations but rather a response to PSR regulation in the hosts. The mechanism by which PSR regulation affects microbiota assembly is not fully understood. On one hand, PSR and plant

immunity have been shown to be transcriptionally linked (Castrillo et al., 2017). On the other, Pi depletion drastically changes the root's exudate profiles (Cha et al., 2016; Herrera Paredes et al., 2018; Pant et al., 2015; Ziegler et al., 2016), which have been shown to play a critical role in plant microbiota assembly (Huang et al., 2019; Stringlis et al., 2018). It is likely that both factors contribute to shifts in microbiota composition. Evidently, however, the *Burkholderia* plant-enrichment in our SynCom occurs via a PSR-independent mechanism.

Our SynCom, comprising 185 genome-sequenced endophytic bacterial isolates, was designed to resemble a natural bacterial community (Figure 3.4b). The community assembly patterns shown for this system are highly reproducible, demonstrating that microbiome assembly is largely a deterministic process. The reproducibility and editability of this system is attractive for detailed mechanistic study of the processes that determine community assembly and its influence on plant phenotype and fitness. From within this SynCom, the genus *Burkholderia* emerges as a PSR-independent, low Pi-responsive taxon. We compared the effect of *Burkholderia* on shoot Pi accumulation from within a full SynCom (a realistic proxy for the bacterial community) to that of the full SynCom lacking *Burkholderia*, a strategy akin to knocking-out a gene of interest, also recently applied in the work by Durán and colleagues (Durán et al., 2018). The control treatment for this type of approach is the full SynCom, whereas in a plant–bacterium binary association experiment it would typically be sterile conditions. Because both sterile conditions and binary association are strong deviations from conditions that may be encountered in the field, the results of binary association experiments may be correspondingly distorted. Using the drop-out approach, we expect to see more subtle differences, because the microbial load on the plant does not change much, but also that these differences be more relevant to the field—an expectation that is yet to be empirically tested. Our observation that dropping *Burkholederia* out of the SynCom increased

shoot Pi in Pi-limiting conditions (50 μM Pi) but not in Pi-replete conditions (1,000 μM Pi) suggests that strains in this genus shift their relationship with the plant from a seeming commensal to a competitor/pathogen. Having ruled out PSR-dependent processes, another plausible explanation for *Burkholderia* enrichment patterns is that when Pi is limiting in the media, the plant becomes a source of Pi for the bacteria and strains with an enhanced ability to utilize plant-derived organic Pi or polyphosphate have an advantage under these conditions. Strains belonging to the genus *Burkholderia* have been shown to be particularly efficient at polyphosphate accumulation at pH 5.5, which is similar to the pH we used in our media (Mullan et al., 2002).

Shifts in microbiota composition that accompany PSR are either adaptive to the plant or reflect opportunistic strategies by bacteria (Castrillo et al., 2017)(Hiruma et al., 2016). Under the former hypothesis, microbes recruited by the plant under Pi stress provide the plants with an advantage vis-a-vis coping with this stress, whereas under the latter, opportunistic microbes might be making a bad situation worse for the plant. In the case of *Burkholderia* in our SynCom, results support the latter hypothesis. *Burkholderia* contribute to depletion of shoot Pi stores, only under Pi-limiting conditions. However, plant-adaptive microbial recruitment under low Pi has been shown to occur as well (Hiruma et al., 2016). The fact that bacteria responding to PSR genes are not a monophyletic group in soil indicates that multiple mechanisms are involved in community assembly. It is likely that these mechanisms encompass both plant-adaptive and opportunistic strategies.

3.10 Supporting figures

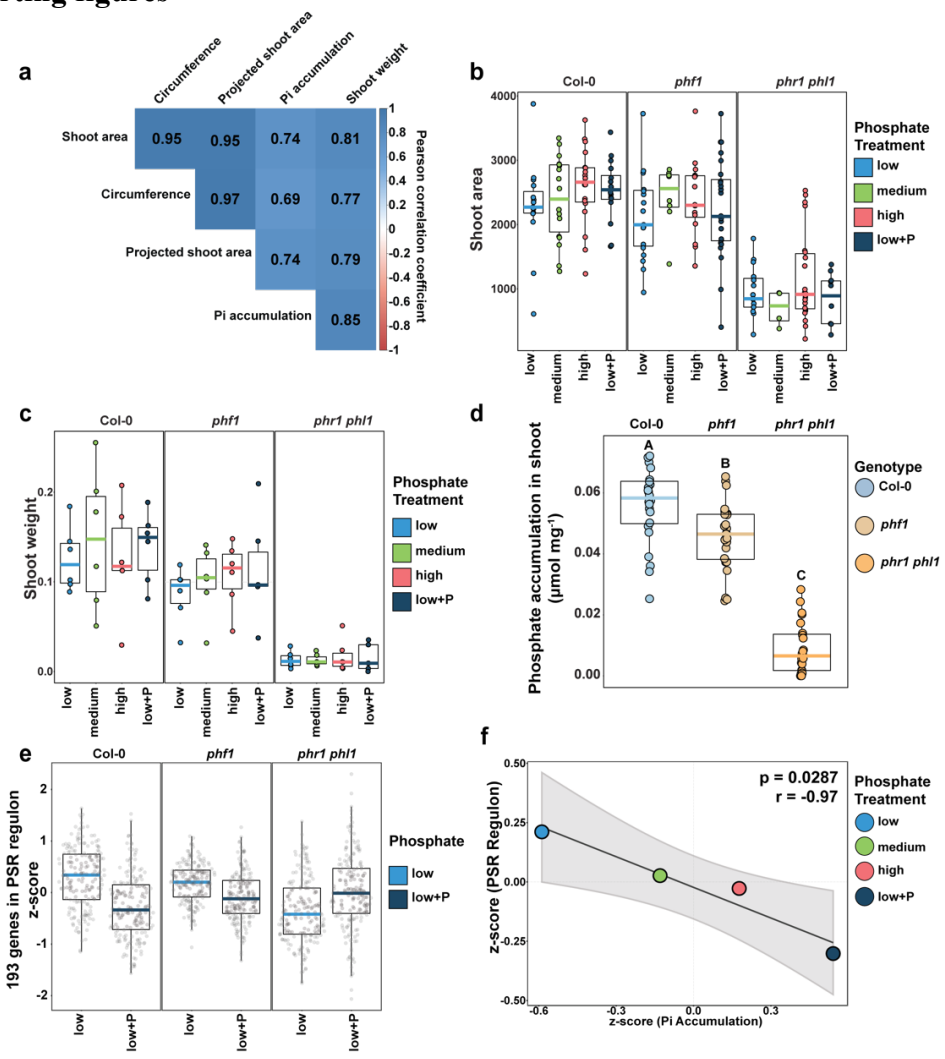


Figure 3.7 - **PSR in soil.**

a) Heat map showing the all versus all pairwise Pearson correlation coefficient calculated between the quantified phenotypes associated with the PSR: shoot area, shoot fresh weight, and shoot free Pi accumulation. **b)** Box plot showing the distribution of the shoot area measured across the P gradient within each of the 3 genotypes. **c)** Boxplot showing the distribution of shoot fresh weight measured across the P gradient within each of the 3 genotypes. **d)** Box plot showing the shoot Pi accumulation across the 3 genotypes. Letters represent the results of the post hoc test. **e)** Box plots displaying the average expression of 193 PSR marker genes across the low and low+P samples in each of the 3 genotypes tested. **f)** Scatter plot showing the relationship between the standardized average phosphate accumulation in leaves (x-axis) and the average standardized expression of 193 PSR marker genes (y-axis). The p-value and R value were calculated according to Pearson's product moment correlation coefficient. P, phosphorus; Pi, orthophosphate; PSR, phosphate starvation response.

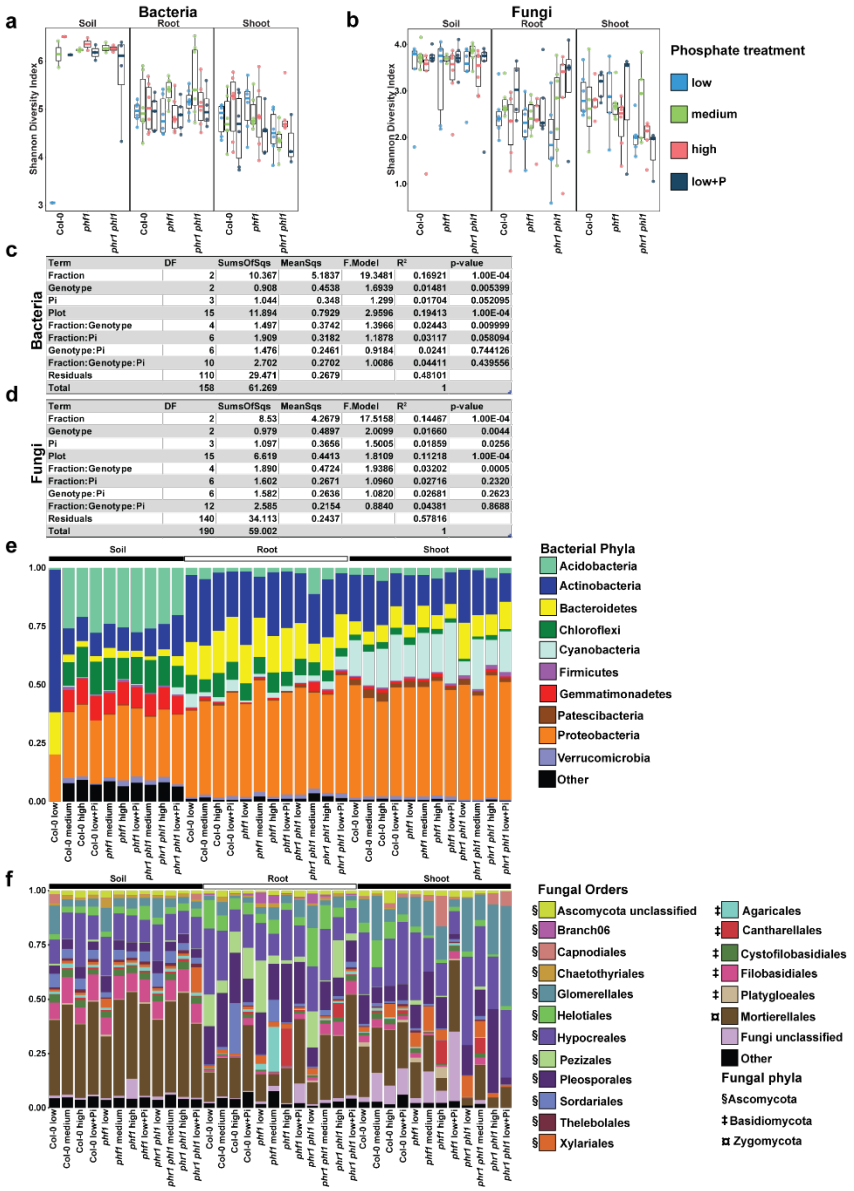


Figure 3.8 - **Characterization of the soil and plant microbiota in soils exposed to different level of P fertilization.**

a, b) Box plots showing the distribution of the alpha diversity (Shannon diversity index) across all levels of P in the soil for bacteria (a) and fungi (b). **(c, d)** PERMANOVA results in which the effect of the 3 variables (fraction, genotype, and soil P) and their interaction on the assembly of the bacterial (c) and fungal (d) communities were tested. **e)** Relative abundance profiles of the main bacterial phyla in the 3 variables (fraction, genotype, and soil P) across all levels of P in the soils. **f)** Relative abundance profiles of the main fungal orders in the 3 variables (fraction, genotype, and soil P) across all the levels of P in the soils. P, phosphorus; PERMANOVA, Permutational Multivariate Analysis of Variance.

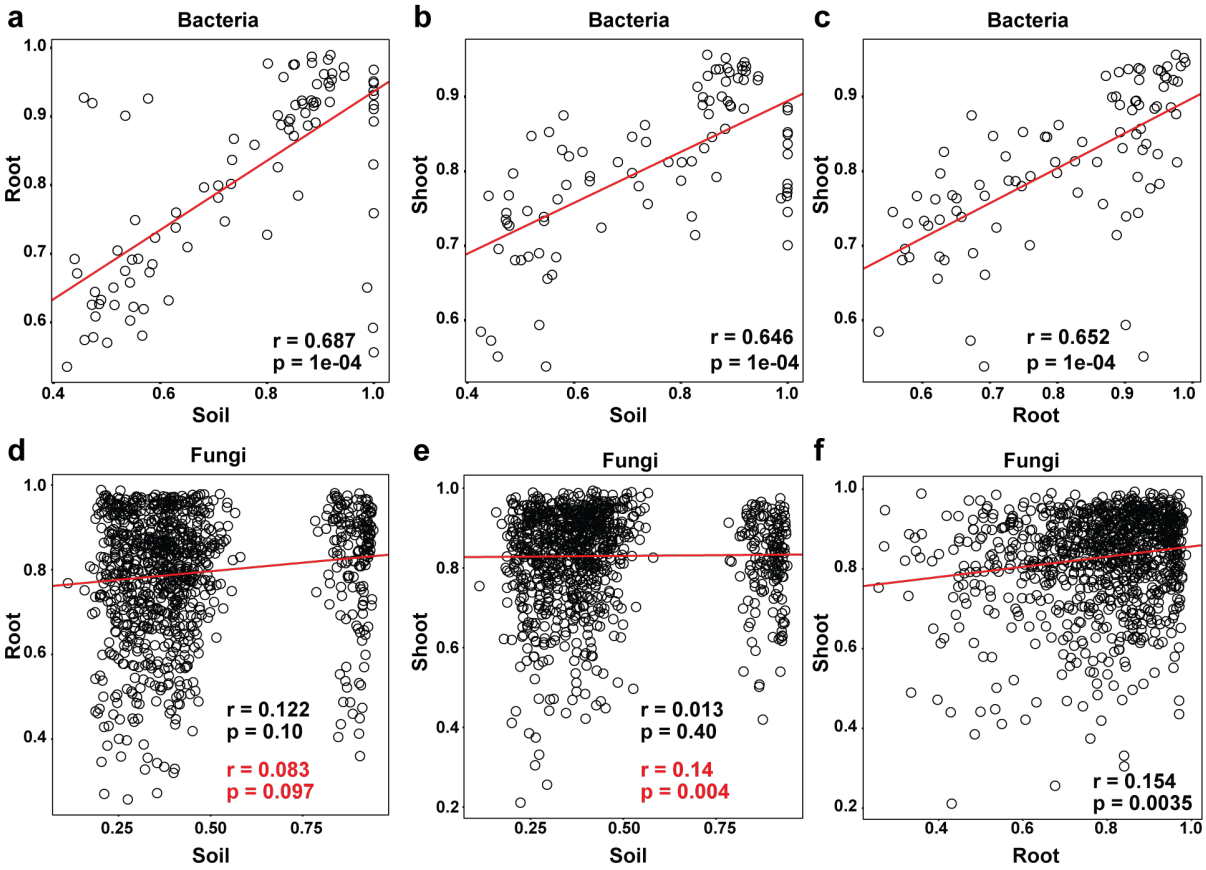


Figure 3.9 - **Bacterial but not fungal plant microbiota composition is strongly dependent on soil inoculum.**

a, b, c) Correlation plots between Bray-Curtis distance matrices calculated for bacteria within soil treatments, root, and shoot fractions. The R and p-values were calculated using Mantel tests. a) Correlation plot of soil versus root. a) Correlation plot of soil versus shoot. c) Correlation plot of root versus shoot. **d, e, f)** Correlation plots between Bray-Curtis distance matrices calculated for fungi within soil treatments, root, and shoot fractions. The R and p-values were calculated using Mantel tests. R and p-values colored in red were calculated excluding the cloud of large distances appearing in graphs panels d and e. d) Correlation plot of soil versus root. e) Correlation plot of soil versus shoot. **f)** Correlation plot of root versus shoot.

a Bacteria

Term	DF	SumsOfSqs	MeanSqs	F.Model	R^2	p-value
Genotype	2	1.0282	0.51412	2.3503	0.04558	0.0015
Pi	3	0.9467	0.31556	1.4426	0.04197	0.0458
Plot	15	9.8092	0.65395	2.9895	0.43484	1.00E-04
Genotype:Pi	6	1.368	0.22801	1.0423	0.06064	0.3461
Residuals	43	9.4062	0.21875		0.41697	
Total	69	22.5584			1	

b Fungi

Term	DF	SumsOfSqs	MeanSqs	F.Model	R^2	p-value
Genotype	2	1.5003	0.75013	2.7429	0.07058	1.00E-04
Pi	3	1.1674	0.38914	1.4229	0.05492	0.0259
Plot	15	6.4217	0.42812	1.5654	0.30209	1.00E-04
Genotype:Pi	6	1.5025	0.25042	0.9157	0.07068	0.7197
Residuals	39	10.665	0.27347		0.50173	
Total	65	21.2575			1	

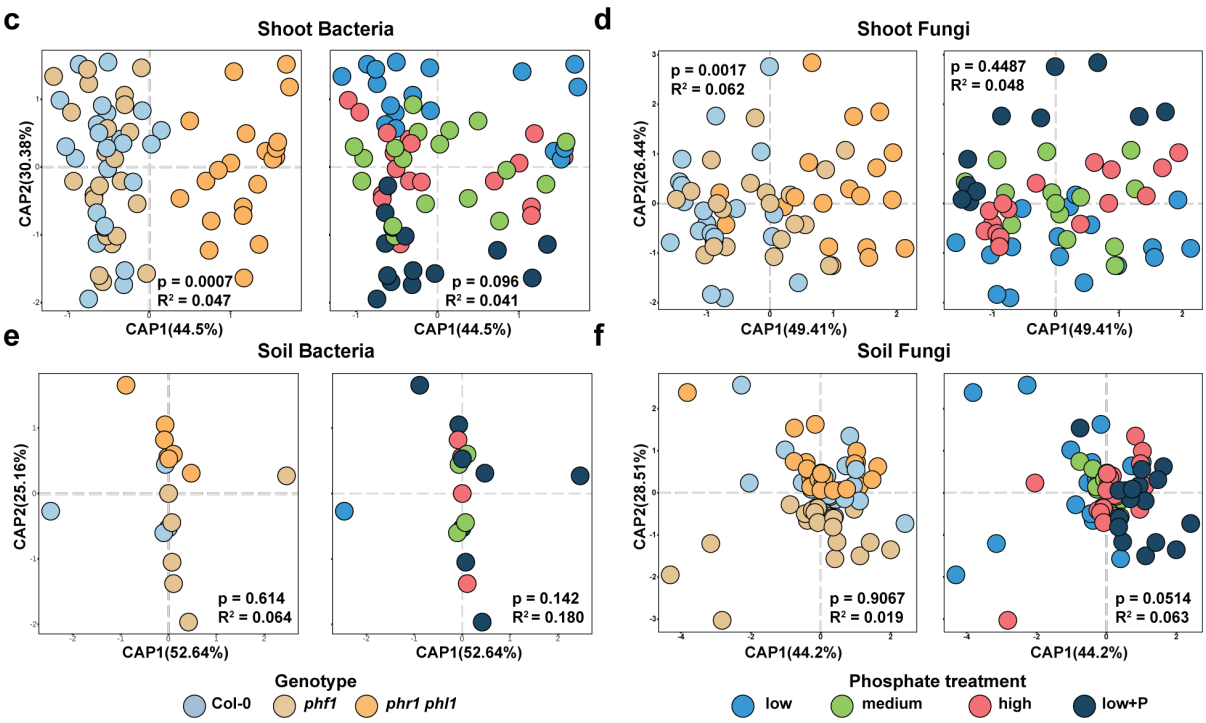


Figure 3.10 - **Plant genotypes and soil P concentrations influence the composition of the plant microbiota.**

a, b) PERMANOVA results showing the influence of the plant genotype and soil P concentration and their interaction on the assembly of the root (a) bacterial and (b) fungal communities. **c, d)** CAP showing the effect of plant genotype and P content in the soil over the shoot (c) bacterial and (d) fungal communities. The p-value and R2 values in each plot are derived from a PERMANOVA model and correspond to the genotype and soil P term, respectively. **e, f)** CAP showing the influence of genotype and P on the soil (e) bacterial and (f) fungal communities. Note smaller number of points in bacterial soil samples. The p-value and R2 values in each plot are derived from a PERMANOVA model and correspond to the genotype and soil P term, respectively. CAP, canonical analysis of principal coordinates; P, phosphorus; PERMANOVA, Permutational Multivariate Analysis of Variance.

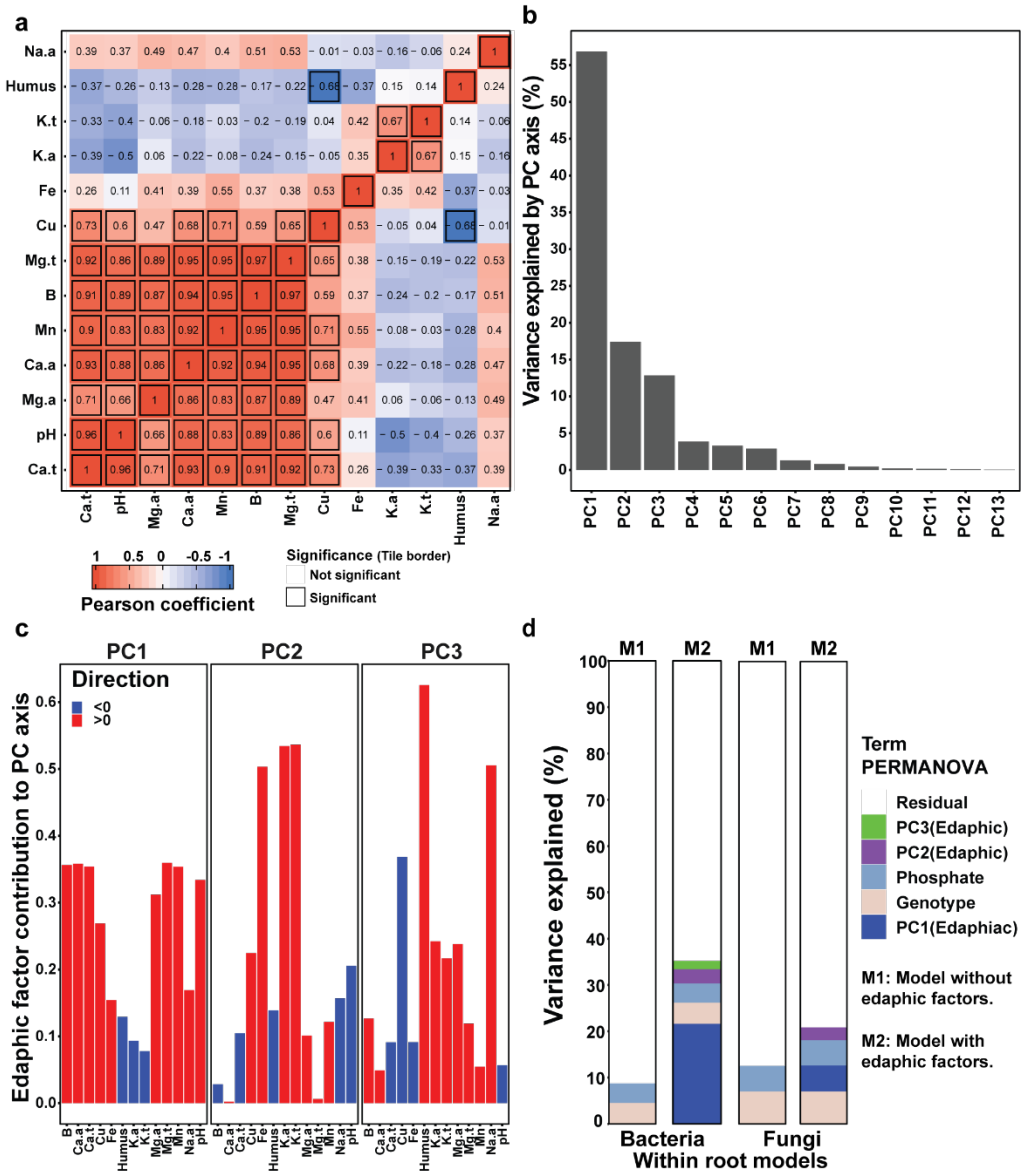
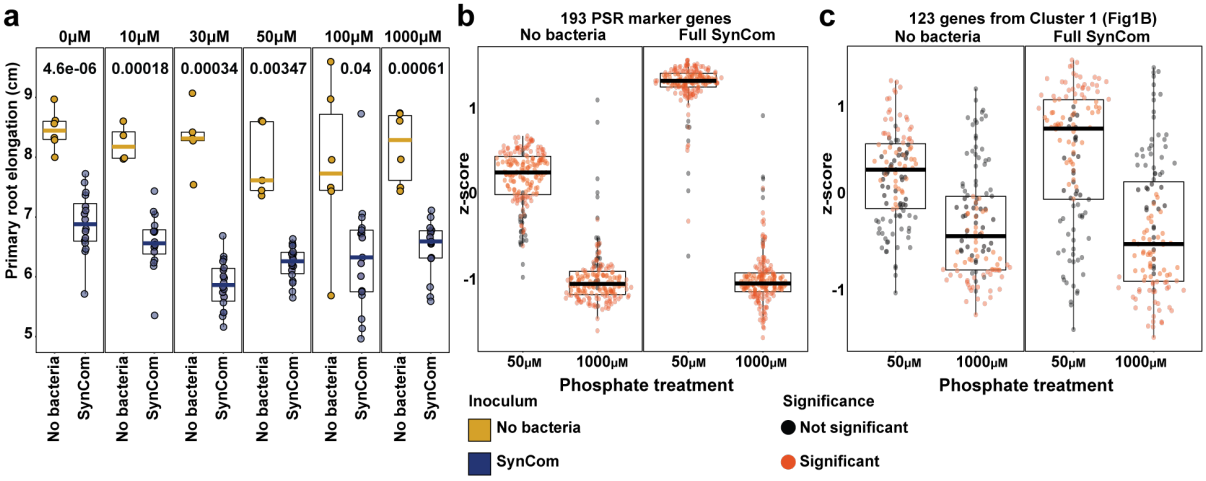


Figure 3.11 - **Variation in soil edaphic factors does not confound soil P effect.**

a) Correlation heat map of the 13 edaphic factors reported in the work by Robbins and colleagues **b)** Bar plot displaying the amount of variance in the edaphic factor matrix explained by PC. **c)** Bar plot depicting the contribution of the different edaphic factors to the first 3 PC. Colors denote the direction of the variable in PCA space. **d)** Proportion of the variance explained for the different variables in models including (M2) and excluding (M1) edaphic factors for bacteria (left) and fungi (right). Only variables with a statistically significant effect are shown. P, phosphorus; PC, principal component; PCA, principal component analysis.



d

Term	DF	SumsOfSqs	MeanSqs	F.Model	R^2	p-value
Fraction	3	3.5223	1.1741	41.384	0.41082	1.00E-04
Pi	5	0.7945	0.1589	5.601	0.09267	1.00E-04
Fraction:Pi	15	0.6256	0.04171	1.47	0.07297	0.0039
Residuals	128	3.6314	0.02837		0.42355	
Total	151	8.5738			1	

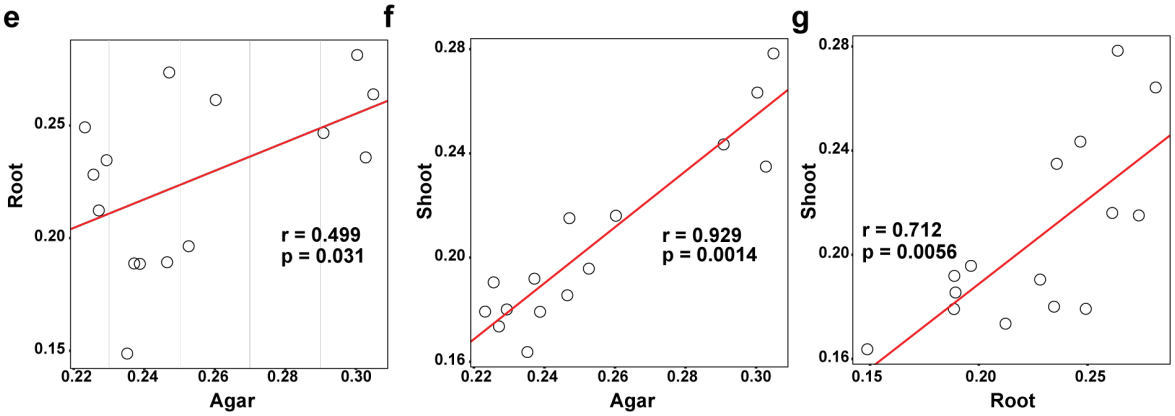
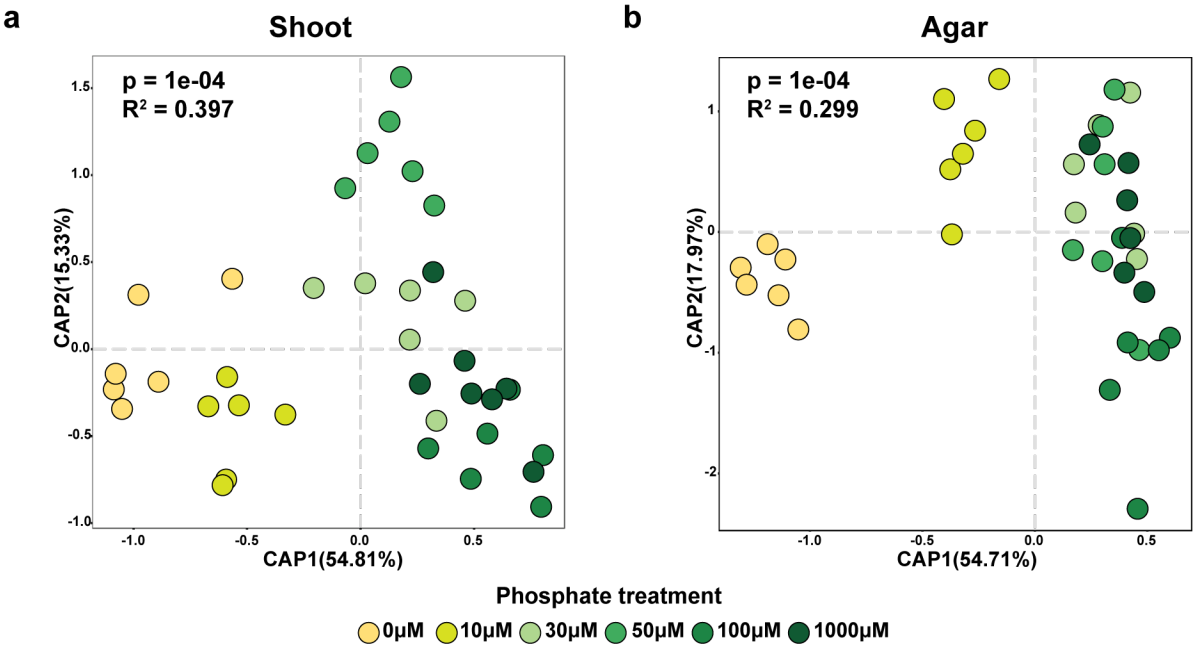


Figure 3.12 - **A bacterial synthetic community modifies the plant PSR.**

a) Box plots displaying the primary root elongation of plants grown in a gradient of Pi concentrations in sterile conditions or with the SynCom. A t test was used for each Pi treatment to estimate differences between SynCom-treated and uninoculated plants. **b)** Average expression of the 193 PSR markers genes in low (50 µM) and high (1,000 µM) Pi conditions within SynCom-treated and uninoculated plants. (C) Average expression of the 123 genes from Cluster 1 in low (50 µM) and high (1,000 µM) Pi conditions within SynCom-treated and uninoculated plants. **d)** PERMANOVA model results showing the influence of the 2 variables (fraction and Pi concentration) and their interaction on the assembly of the bacterial community in the plant**. e, f, g)** Correlation plots between Bray-Curtis distance matrices calculated for bacterial profile within agar, root, and shoot fractions. The R and p-values were calculated using Mantel tests. e) Correlation plot of agar versus root. f) Correlation plot of agar versus shoot. g) Correlation plot of root versus shoot. PERMANOVA, Permutational Multivariate Analysis of Variance; Pi, orthophosphate; PSR, phosphate starvation response; SynCom, synthetic community.



c

Root

Term	DF	SumsOfSqs	MeanSqs	F.Model	R^2	p-value
Pi	5	0.3609	0.072179	4.0715	0.42098	1.00E-04
Residuals	28	0.49638	0.017728		0.57902	
Total	33	0.85728			1	

D

Shoot

Term	DF	SumsOfSqs	MeanSqs	F.Model	R^2	p-value
Pi	5	0.3082	0.061641	4.0894	0.39744	1.00E-04
Residuals	31	0.46728	0.015073		0.60256	
Total	36	0.77548			1	

E

Agar

Term	DF	SumsOfSqs	MeanSqs	F.Model	R^2	p-value
Pi	5	0.3634	0.072679	2.5647	0.29945	1.00E-04
Residuals	30	0.85014	0.028338		0.70055	
Total	35	1.21353			1	

Figure 3.13 - **Bacterial synthetic community responds to the phosphate concentration in the media.**

a, b) CAP showing the influence of Pi concentration in the media on the bacterial communities in the (a) plant shoot and (b) agar. The bar graphs to the left of each plot depict the percentage of variability explained by statistically significant ($p < 0.05$) variables based on a PERMANOVA model. **c, d, e)** PERMANOVA model results showing the influence of Pi concentration on the assembly of the bacterial community in (c) roots, (d) shoot, and (e) agar. CAP, canonical analysis of principal coordinates; PERMANOVA, Permutational Multivariate Analysis of Variance; Pi, orthophosphate.

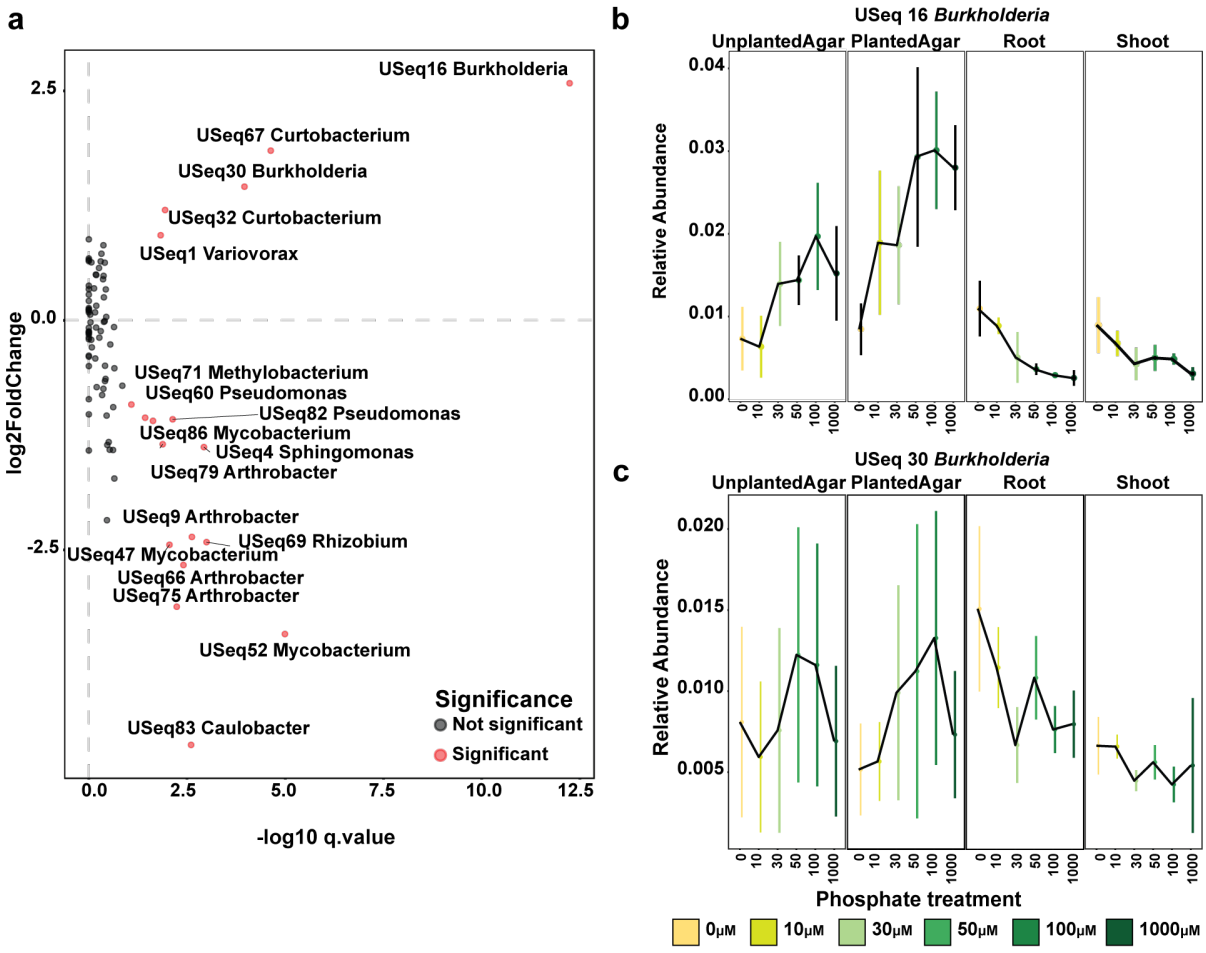


Figure 3.14 - **USeqs in the bacterial synthetic community displayed a strong Pi:fraction (shoot, root, agar) interaction.**

a) Scatter plot (volcano plot) showing the results of the GLM interaction model between fraction and Pi concentration. The axes of the plot represent the output of the statistical test. The x-axis is the transformed q-value and the y-axis the log2 fold change. Each dot in the scatter plot represents a USeq. USeqs that showed a statistically significant fraction:Pi interaction are colored in red. USeqs genus and ID are displayed. The top right quadrant represents USeqs that are enriched in the plant tissues under low Pi conditions. **b, c)** Relative abundance of Burkholderia Useqs 16 (b) and 30 (c) that exhibits a statistically significant ($q < 0.1$) Pi enrichment between the plant fractions and the agar fraction. The middle dot of each strip bar corresponds to the mean of that particular condition; the range of the strip bar corresponds to the 95% confidence interval of the mean. GLM, generalized linear model; Pi, orthophosphate; USeq, unique sequence.

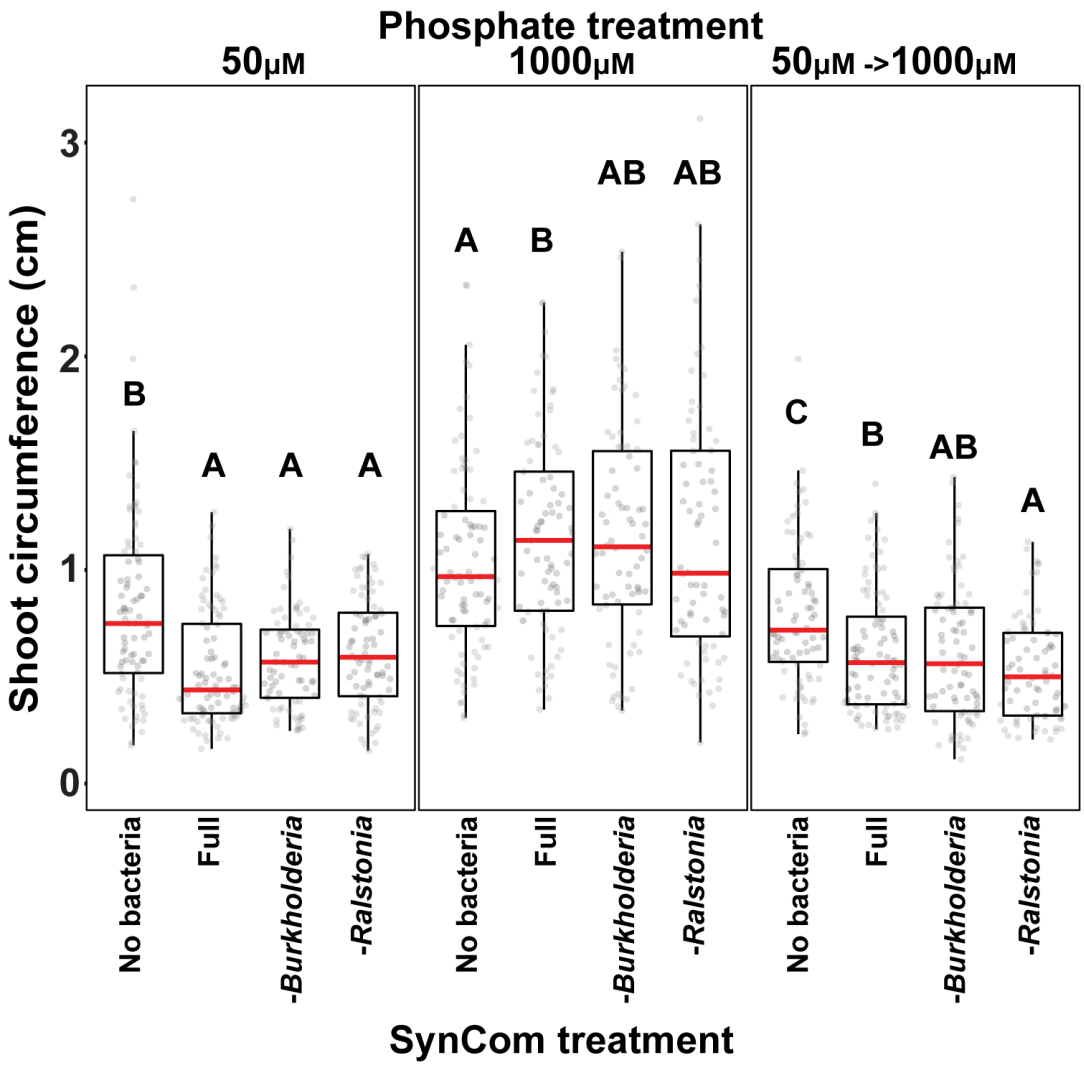


Figure 3.15 - **Shoot size is not affected by Burkholderia drop-out.**

Box plots showing shoot circumference in plants exposed to different SynComs across 3 phosphate treatments. Statistically significant differences among SynCom treatments were computed within each phosphate treatment separately using an ANOVA model. Letters represent the results of the post hoc test. SynCom, synthetic community.

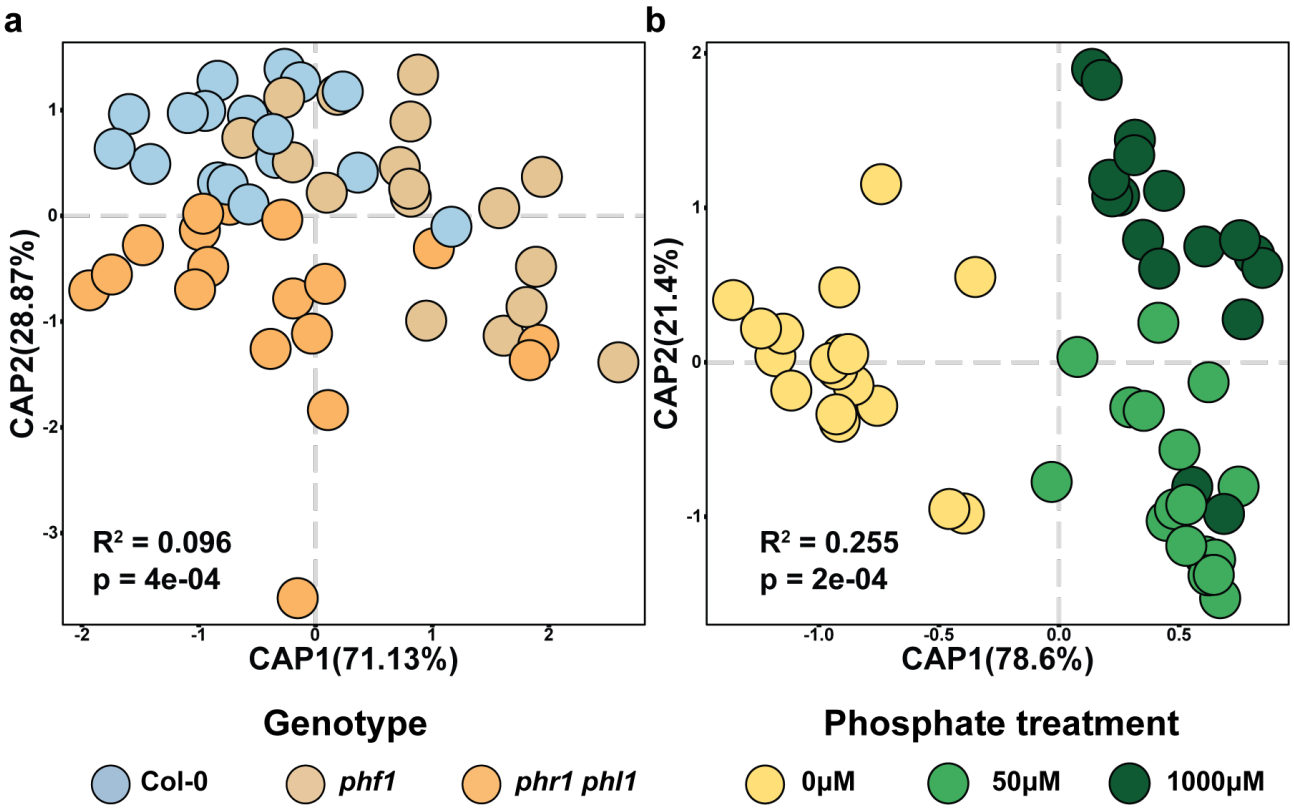


Figure 3.16 - **Plant genotype and Pi concentration affect root community composition in the agar system.**

CAP showing the influence of plant genotypes (a) and agar Pi concentration (b) over the bacterial SynCom in the root. The p-value and R2 values inside each plot are derived from a PERMANOVA model and correspond to the genotype and Pi terms, respectively. CAP, canonical analysis of principal coordinates; PERMANOVA, Permutational Multivariate Analysis of Variance; Pi, orthophosphate; SynCom, synthetic community.

3.11 Methods

3.11.1 Soil P gradient experiment

3.11.1.1 Collection of soil from field site

Soil used in this experiment was collected from the long-term Pi fertilization field ("Field D") trial at the Julius Kühn Experimental Station at Martin Luther University of Halle-Wittenberg (51°29′45.6″N, 11°59′33.3″E) (Gransee & Merbach, 2000; Merbach et al., 2000). Soil cores (10 cm diameter × 15 cm depth) were taken from 18 6 × 5 m unplanted plots belonging to 2 strips.

These plots represent 3 P fertilization regimens: low, medium, and high P (0, 15, and 45 kg P ha−1 year−1, respectively). Differences in soil mineral content between strips and P fertilization regimens are reported in the work by Robbins and colleagues (Robbins et al., 2018), showing that the different Pi regiments significantly differed only in P content. Soil cores were harvested in the middle of March 2016 (strip 1) and beginning of April 2016 (strip 2). Approximately 2 cm of the topsoil was discarded, and the remaining lower 13 cm of soil was stored at 4°C until use. Soils from each core were homogenized separately with a mesh sieve wire (5 × 5 m2). The sieved soil cores were stored at 4°C until use. About 300 g of soil were added to each pot (7 × 7 × 7 cm3).

3.11.1.2 Experimental design

Each of the 3 Arabidopsis genotypes was grown in soil from all 18 plots (6 plots per P treatment). In addition, a fourth P regimen designated "Low+P" was created by adding additional P to a set of pots with low P. The amount of P added to these pots is based on the difference in total P between low and high P plots. The average difference between low and high P over all the plots is 42 mg P per kg soil (Robbins et al., 2018). Per pot, this is 12.6 mg P (accounting for 300 g soil per pot). Thus, a 10 ml solution consisting of 4.2 mg P in the form of 20% K2HPO4 (MilliporeSigma, St. Louis, MO) and 80% KH2PO4 (MilliporeSigma, St. Louis, MO) was added to the pots in 3 applications (Weeks 2, 4, and 6) before watering (in order to distribute the P through the soil).

Thus, the experiment included 2 variables: soil treatment (low P, medium P, high P, low+P) and genotypes (Col-0, *phf1*, and *phr1/phl1*) with 6 independent replicates, amounting to 72 pots. Pot positions in the greenhouse were randomized.

3.11.1.3 Plant growth conditions

Arabidopsis thaliana ecotype Col-0 and mutants *phf1* and *phr1 phl1* (both in the Col-0 background) were used. Seeds were surface sterilized (20 min 70% EtOH (MilliporeSigma, St. Louis, MO), 10 s 100% EtOH) and planted directly onto moist soil. Sown seeds were stratified for 3 days at 4°C before being placed in a greenhouse under short-day conditions (6/18 day-night cycle; 19 to 21°C) for 8 weeks. Germinating seedlings were thinned to 4 plants per pot.

3.11.1.4 Sample harvest

After 8 weeks of growth, pots were photographed, and shoot size was quantified using WinRhizo software (Regent instruments Inc., Québec, Canada). Samples were harvested in random order to avoid any confounding circadian effect on the results. For DNA extraction, 2 roots, 2 shoots, and soil from each pot were harvested separately. Roots and shoots were rinsed in sterile water to remove soil particles, placed in 2 ml Eppendorf tubes (Eppendorf, Hamburg, Germany) with 3 sterile glass beads (MilliporeSigma, St. Louis, MO), then washed 3 times with sterile distilled water to remove soil particles and weakly associated microbes. Root and shoot tissue were then pulverized using a tissue homogenizer (TissueLyser II; Qiagen, Hilden, Germany) and stored at −80˚C until processing. Five ml of soil from each pot was suspended in 20 ml of sterile distilled water. The resulting slurry was sieved through a 100 μm sterile cell strainer (Fisher Scientific, Hampton, NH) and the flow-through was centrifuged twice at maximum speed for 20 minutes, removing the supernatant both times. The resulting pellet was stored at −80˚C until processing. For RNA extraction, one root system and one shoot were taken from 3 replicates of each treatment, washed lightly to remove soil particles, placed in 2 ml Eppendorf tubes with 3 glass beads and flash frozen with liquid nitrogen. Tubes were stored at −80˚C until processing. For shoot Pi measurement, 2 to 3 leaves from the remaining shoot in each pot were placed in an

Eppendorf tube and weighed; 1% acetic acid (MilliporeSigma, St. Louis, MO) was then added, and samples were flash frozen and stored at −80°C until processing. The Ames method (Ames, 1966) was used to determine the phosphate concentration in these samples.

3.11.1.5 DNA extraction

DNA extractions were carried out on ground root and shoot tissue and soil pellets, using the 96-well-format MoBio PowerSoil Kit (MoBio Laboratories; Qiagen, Hilden, Germany) following the manufacturer's instruction. Sample position in the DNA extraction plates was randomized, and this randomized distribution was maintained throughout library preparation and sequencing.

3.11.1.6 RNA extraction

RNA was purified from plant tissue using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and stored at −80˚C.

3.11.2 Bacterial SynCom Experiment

3.11.2.1 Bacterial isolation and culture

The 185-member bacterial SynCom contained genome-sequenced isolates obtained from Brassicaceae roots, nearly all Arabidopsis, planted in 2 North Carolina, US, soils. Because both bacteria and fungi responded similarly to PSR in our soil experiments, we only included bacteria, which are more compatible with our experimental system in our SynCom. A detailed description of this collection and isolation procedures can be found in the work by Levy and colleagues (Levy et al., 2018). One week prior to each experiment, bacteria were inoculated from glycerol (MilliporeSigma, St. Louis, MO) stocks into 600 µL KB medium in a 96 deep well plate. Bacterial cultures were grown at 28°C, shaking at 250 rpm. After 5 days of growth, cultures were inoculated into fresh media and returned to the incubator for an additional 48 hours, resulting in 2 copies of

each culture, 7 days old and 48 hours old. We adopted this procedure to account for variable growth rates of different SynCom members and to ensure that nonstationary cells from each strain were included in the inoculum. After growth, 48-hour-old and 7-day-old plates were combined and optical density (OD) of the culture was measured at 600 nm using an Infinite M200 Pro plate reader (TECAN, Männedorf, Switzerland). All cultures were then pooled while normalizing the volume of each culture according to the OD (we took a proportionally higher volume of culture from cultures with low OD). The mixed culture was then washed twice with 10 mM MgCl2 (MilliporeSigma, St. Louis, MO) to remove spent media and cell debris and vortexed vigorously with sterile glass beads to break up aggregates. OD of the mixed, washed culture was then measured and normalized to OD = 0.2. A total of 100 μL of this SynCom inoculum was spread on each agar plate prior to transferring seedlings.

3.11.2.2 Experimental design of agar experiments

We performed the Pi gradient experiment in 2 independent replicas (experiments performed at different times, with fresh bacterial inoculum and batch of plants), each containing 3 internal replications, amounting to 6 samples for each treatment. We had 2 SynCom treatments: no bacteria (NB) and SynCom; 6 Pi concentrations: 0, 10, 30, 50, 100, or 1,000 μM KH2PO4(henceforth, Pi); and 2 plant treatments: planted plates and unplanted plates (NP).

For the drop-out experiment, the entire SynCom, excluding all 5 *Burkholderia* and both *Ralstonia* isolates, was grown and prepared as described above. The Burkholderia and *Ralstonia* isolates were grown in separate tubes, washed, and added to the rest of the SynCom to a final OD of 0.001 (the calculated OD of each individual strain in a 185-Member SynCom at an OD of 0.2) to form the following 4 mixtures: (1) Full community—all *Burkholderia* and *Ralstonia* isolates added to the SynCom; (2) *Burkholderia* drop-out—only *Ralstonia* isolates added to the SynCom;

(3) *Ralstonia* drop-out—only *Burkholderia* isolates added to the SynCom; (4) uninoculated plants—no SynCom. The experiment had 3 Pi conditions: low Pi (50 μM Pi), high Pi (1,000 μM Pi), and low→high Pi. Twelve days post-inoculation the low Pi and high Pi samples were harvested, and the low→high plants were transferred from 50 μM Pi plates to 1,000 μM Pi plates for an additional 3 days. The experiment was performed twice, and each rep consisted of 6 plates per SynCom mixture and Pi treatment, amounting to 72 samples. Upon harvest, shoot Pi accumulation was measured using the Ames method (Ames, 1966).

For the drop-out experiment with PSR mutants, the entire SynCom, excluding all 5 *Burkholderia*, was grown and prepared as described above. The *Burkholderia* isolates were grown in separate tubes, washed, and added to the SynCom to a final OD of 0.001 (the calculated OD of each individual strain in a 185-Member SynCom at an OD of 0.2) to form the following 2 mixtures: (1) Full community—all *Burkholderia* isolates added to the SynCom; (2) Burkholderia drop-out—no isolates added to the SynCom. For each SynCom, we inoculated 6 agar plates for each of 3 Pi conditions: 0, 50, and 1,000 μM Pi. Three 7-day-old seedlings from each of the 3 genotypes (wt Col-0, phf1, and phr1 phl1) were transferred to each plate. Roots were harvested 12 days post-inoculation, and bacterial DNA was extracted.

3.11.2.3 *In vitro* plant growth conditions

Arabidopsis thaliana accession Col-0 was used. All seeds were surface-sterilized with 70% bleach (Clorox, Oakland, CA), 0.2% Tween-20 (MilliporeSigma, St. Louis, MO) for 8 minutes, and rinsed 3 times with sterile distilled water to eliminate any seed-borne microbes on the seed surface. Seeds were stratified at 4°C in the dark for 2 days. Plants were germinated on vertical square 12 X 12 cm agar plates (Fisher Scientific, Hampton, NH) with Johnson medium (JM; [4]) containing 0.5% sucrose (MilliporeSigma, St. Louis, MO) and 1,000 μM Pi, for 7 days. Then, 10

plants were transferred to each vertical agar plate with amended JM lacking sucrose at one of the following experimental Pi concentrations: 0, 10, 30, 50, 100, or 1,000 μM Pi. The SynCom was spread on the agar prior to transferring plants. Each experiment included unplanted agar plates with SynCom for each media type (designated NP) and uninoculated plates with plants for each media type (designated NB). Plants were placed in randomized order in growth chambers and grown under a 16-hour dark/8-hour light regime at 21°C day/18°C night for 12 days (the period of time it takes roots to reach the bottom of the plate).

3.11.2.4 Sample harvest

Twelve days post-transferring, plates were imaged using a document scanner. For DNA extraction, roots, shoots, and agar were harvested separately, pooling 6 plants for each sample. Roots and shoots were placed in 2.0 ml Eppendorf tubes with 3 sterile glass beads. Samples were washed 3 times with sterile distilled water to remove agar particles and weakly associated microbes. Tubes were stored at −80˚C until processing. For RNA, samples were collected from a separate set of 2 independent experiments, using the same SynCom and conditions as above but with just 2 Pi concentrations: 1,000 μM Pi (high) and 50 μM Pi (low). Four seedlings were harvested from each sample, and samples were flash frozen and stored at −80˚C until processing.

3.11.2.5 DNA extraction

Root and shoot samples were lyophilized for 48 hours using a Freezone 6 freeze dry system (Labconco, Fisher Scientific, Hampton, NH) and pulverized using a tissue homogenizer (MP Biomedicals, Solon, OH). Agar from each plate was stored in a 30 ml syringe (Fisher Scientific, Hampton, NH) with a square of sterilized Miracloth (Millipore) at the bottom and kept at −20°C for a week. Syringes were then thawed at room temperature, and samples were squeezed gently into 50 ml tubes. Samples were centrifuged at maximum speed for 20 minutes, and most of the

supernatant was discarded. The remaining 1 to 2 ml of supernatant containing the pellet was transferred into clean microfuge tubes. Samples were centrifuged again, supernatant was removed, and pellets were stored at −80°C until DNA extraction.

DNA extractions were carried out on ground root and shoot tissue and agar pellets using 96-well-format MoBio PowerSoil Kit (MOBIO Laboratories; Qiagen, Hilden, Germany) following the manufacturer's instruction. Sample position in the DNA extraction plates was randomized, and this randomized distribution was maintained throughout library preparation and sequencing.

3.11.2.6 RNA extraction

RNA was extracted from *Arabidopsis* seedlings following the work by Logemann (Logemann et al., 1987). Frozen seedlings were ground in liquid nitrogen, then homogenized in a buffer containing 400 µl of Z6-buffer; 8 M guanidinium-HCl (MilliporeSigma, St. Louis, MO), 20 mM MES, (MilliporeSigma, St. Louis, MO)20 mM EDTA (MilliporeSigma, St. Louis, MO) at pH 7.0; 400 µL phenol:chloroform:isoamylalcohol (25:24:1) (MilliporeSigma, St. Louis, MO) was added, and samples were vortexed and centrifuged (20,000g, 10 minutes) for phase separation. The aqueous phase was transferred to a new 1.5 ml tube, and 0.05 volumes of 1 N acetic acid (MilliporeSigma, St. Louis, MO) and 0.7 volumes 96% ethanol were added. The RNA was precipitated at −20°C overnight. Following centrifugation (20,000g, 10 minutes, 4°C), the pellet was washed with 200 µl sodium acetate (pH 5.2) (MilliporeSigma, St. Louis, MO) and 70% ethanol. The RNA was dried and dissolved in 30 µL of ultrapure water and stored at −80°C until use.

3.11.2.7 Quantification of plant phenotypes

The Ames method (Ames, 1966) was used to determine the phosphate concentration in the shoots of plants grown on different Pi regimens and treatments. Primary root length elongation was measured using ImageJ (Schindelin et al., 2012), and for shoot area and total root network measurement, WinRhizo software (Regent Instruments Inc., Quebec, Canada) was used.

3.11.3 DNA and RNA sequencing

3.11.3.1 Bacterial 16s sequencing

We amplified the V3-V4 regions of the bacterial 16S rRNA gene using primers 338F (5′-ACTCCTACGGGAGGCAGCA-3′) and 806R (5′-GGACTACHVGGGTWTCTAAT-3′). Two barcodes and 6 frames hifts were added to the 5' end of 338F, and 6 frameshifts were added to the 806R primers, based on the protocol in the work by Lundberg and colleagues (Lundberg, Lebeis, Paredes, Yourstone, Gehring, Malfatti, Tremblay, Engelbrektson, Kunin, del Rio, et al., 2012). Each PCR reaction was performed in triplicate and included a unique mixture of 3 frameshifted primer combinations for each plate. PCR conditions were as follows: 5 µl Kapa Enhancer (Kapa Biosystems, Wilmington, MA), 5 µl Kapa Buffer A (Kapa Biosystems, Wilmington, MA), 1.25 µl of 5 µM 338F, 1.25 µl of 5 µM 806R, 0.375 µl mixed rRNA gene-blocking peptide nucleic acids (PNAs; 1:1 mix of 100 µM plastid PNA and 100 µM mitochondrial PNA; PNA Bio (Kapa Biosystems, Wilmington, MA), 0.5 µl Kapa dNTPs (Kapa Biosystems, Wilmington, MA), 0.2 µl Kapa Robust Taq (Kapa Biosystems, Wilmington, MA), 8 µl dH2O, 5 µl DNA; temperature cycling: 95°C for 60 seconds, 24 cycles of 95°C for 15 seconds, 78°C (PNA) for 10 seconds, 50°C for 30 seconds, 72°C for 30 seconds, 4°C until use. Following PCR cleanup, the PCR product was indexed using 96 indexed 806R primers with the same reaction mix as above and 9 cycles of the cycling conditions described in the work by Lundberg and colleagues (Lundberg, Lebeis, Paredes,

Yourstone, Gehring, Malfatti, Tremblay, Engelbrektson, Kunin, del Rio, et al., 2012). PCR products were purified using AMPure XP magnetic beads (Beckman Coulter, Brea, CA) and quantified with a Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA). Amplicons were pooled in equal amounts and then diluted to 10 pM for sequencing. Sequencing was performed on an Illumina MiSeq instrument (Illumina, San Diego, CA) using a 600-cycle V3 chemistry kit. The raw data for the natural soil experiment is available in the NCBI SRA Sequence Read Archive (accession PRJNA531340). The raw data for the SynCom experiment is available in the NCBI SRA Sequence Read Archive (accession PRJNA531340).

3.11.3.2 Fungal/Oomycete ITS sequencing

We amplified the ITS1 region using primers ITS1-F (5′-CTTGGTCATTTAGAGGAAGTAA-3′; (GARDES & BRUNS, 1993)) and ITS2 (5′-GCTGCGTTCTTCATCGATGC-3′;). Samples were diluted to concentrations of 3.5 ng µl−1 of DNA with nuclease-free water for the first PCR reaction to amplify the ITS1 region. Reactions were prepared in triplicate in 25 µl volumes consisting of 10 ng of DNA template, 1× incomplete buffer, 0.3% bovine serum albumin, 2 mM MgCl2, 200 µM dNTPs, 300 nM of each primer, and 2 U of DFS-Taq DNA polymerase (Bioron, Ludwigshafen, Germany); temperature cycling: 2 minutes at 94°C, 25 cycles: 30 seconds at 94°C, 30 seconds at 55°C, and 30 seconds at 72°C; and termination: 10 minutes at 72°C. PCR products were cleaned using an enzymatic cleanup (24.44 µl: 20 µl of template, 20 U of exonuclease I, 5 U of Antarctic phosphatase, 1× Antarctic phosphatase buffer; New England Biolabs, Frankfurt, Germany); incubation conditions were 30 minutes at 37°C, 15 minutes at 85°C; centrifuge 10 minutes at 4,000 rpm. A second PCR was then performed (2 minutes at 94°C; 10 cycles: 30 seconds at 94°C, 30 seconds at 55°C, and 30 seconds at 72°C; and termination: 10 minutes at 72°C), in triplicate using 3 µl of cleaned PCR product and

sample-specific barcoded primers (5′-AATGATACGGCGACCACCGAGATCTACACTCACGCGCAGG-ITS1F-3′; 5′-CAAGCAGAAGACGGCATACGAGAT-BARCODE(12-NT)-CGTACTGTGGAGA-ITS2-3′). PCR reactions were purified using with Agencourt AMPure XP purification kit (Beckman Coulter, Krefeld, Germany). Amplicons were pooled in equal amounts and then diluted to 10 pM for sequencing. Sequencing was performed on an Illumina MiSeq instrument using a 600-cycle V3 chemistry kit. The raw data are available in the NCBI SRA Sequence Read Archive (Project Number PRJNA531340).

3.113.3 Plant RNA Sequencing

Illumina-based mRNA-Seq libraries were prepared from 1 μg RNA following the work by Herrera Paredes and colleagues (Herrera Paredes et al., 2018). mRNA was purified from total RNA using Sera-mag oligo(dT) magnetic beads (GE Healthcare Life Sciences, Chicago, IL) and then fragmented in the presence of divalent cations (Mg2+) at 94°C for 6 minutes. The resulting fragmented mRNA was used for first-strand cDNA synthesis using random hexamers and reverse transcriptase (Enzymatics, Qiagen, Beverly, MA), followed by second-strand cDNA synthesis using DNA Polymerase I (Enzymatics, Qiagen, Beverly, MA) and RNAseH (Enzymatics, Qiagen, Beverly, MA). Double-stranded cDNA was end-repaired using T4 DNA polymerase (Enzymatics, Qiagen, Beverly, MA), T4 polynucleotide kinase (Enzymatics, Qiagen, Beverly, MA), and Klenow polymerase (Enzymatics, Qiagen, Beverly, MA). The DNA fragments were then adenylated using Klenow exo-polymerase (Enzymatics, Qiagen, Beverly, MA) to allow the ligation of Illumina Truseq HT adapters (D501–D508 and D701–D712; Illumina, San Diego, CA). Following library preparation, quality control and quantification were performed using a 2100 Bioanalyzer instrument (Agilent Technologies, Santa Clara, CA) and the Quant-iT PicoGreen

dsDNA Reagent (Invitrogen, Carlsbad, CA), respectively. Libraries were sequenced using HiSeq4000 sequencers (Illumina, San Diego, CA) to generate 50-bp single-end reads.

3.11.4 Data processing and statistical analyses

3.11.4.1 Quantification of plant phenotypes—Soil experiment

To measure correlation between all measured plant phenotypes (shoot Pi, shoot weight, shoot size) we applied hierarchical clustering based on a matrix of Pearson correlation coefficients between all pairs of phenotypes. We used the R package corrplot version 0.84 to visualize correlations. To compare shoot Pi accumulation, we treated the low P sample as the control, because this soil did not receive any treatment. We performed paired t tests between the different P-treated samples and the low P samples within each plant genotype independently ($\alpha < 0.05$).

3.11.4.2 Amplicon sequence data processing—Soil experiments

Bacterial sequencing data were processed with MT-Toolbox (Yourstone et al., 2014). Usable read output from MT-Toolbox (i.e., reads with 100% correct primer and primer sequences that successfully merged with their pair) were quality filtered using Sickle by not allowing any window with a Q score under 20. After quality filtering, samples with low total reads recruited (<3,000 reads), amounting to 51 soil samples were discarded. The resulting sequences were collapsed into ASVs using the R package DADA2 version 1.8.1 (Callahan et al., 2016). Taxonomic assignment of each ASV was performed using the naïve Bayes kmer method implemented in the DADA2 package using the Silva 132 database as training reference (Quast et al., 2013).

Fungal ITS sequence data were processed using DADA2 (Callahan et al., 2016) with default parameters using only the forward reads. Taxonomic assignment of each ASV was performed using the naïve Bayes kmer method implemented in the MOTHUR package (Schloss

et al., 2009) using the UNITE database (Nilsson et al., 2019) as training reference. The resulting bacterial and fungal count tables were deposited at https://github.com/isaisg/hallepi.

3.11.4.3 Community analyses—Soil experiments

The resulting bacterial and fungal count tables were processed and analyzed with functions from the ohchibi package. Both tables were rarefied to 3,000 reads per sample. An alpha diversity metric (Shannon diversity) was calculated using the diversity function from the vegan package version 2.5–3 (Dixon, 2003). We used ANOVA to test for differences in Shannon Diversity indices between groups. Tukey's HSD post hoc tests here and elsewhere were performed using the cld function from the emmeans R package (Searle et al., 1980). Beta-diversity analyses (Principal coordinate analysis and canonical analysis of principal coordinates [CAP]) were based on Bray-Curtis dissimilarity calculated from the rarefied abundance tables. We utilized the capscale function from the vegan R package v.2.5–3 (Dixon, 2003) to compute a CAP. To analyze the full data set (all fraction, all genotypes, all phosphorus treatments), we constrained by fraction, plant genotype, and phosphorus fertilization treatment, while conditioning for the plot effect. We performed the genotype:phosphorus interaction analysis over each fraction independently, constraining for the plant genotype and phosphorus fertilization treatment while conditioning for the plot effect. In addition to CAP, we performed Permutational Multivariate Analysis of Variance (PERMANOVA) over the 2 data sets described above using the adonis function from the vegan package version 2.5–3 (Dixon, 2003). Finally, we used the function chibi.permanova from the ohchibi package to plot the R2 values for each significant term in the PERMANOVA model tested.

The relative abundance of bacterial phyla and fungal taxa were depicted using the stacked bar representation encoded in the function chibi.phylogram from the ohchibi package.

We used the R package DESeq2 version 1.22.1 (Love et al., 2014) to compute the enrichment profiles for both bacterial and fungal ASVs. For the full data set model, we estimated main effects for each variable tested (Fraction, Plant genotype, and phosphorus fertilization) using the following design:

Abundance ~ Fraction + Genotype + Phosphorous treatment

We delimited ASV fraction enrichments using the following contrasts: soil versus root, soil versus shoot, and root versus shoot. An ASV was considered statistically significant if it had $q < 0.1$.

We implemented a second statistical model in order to identify ASVs that exhibited statistically significant differential abundances depending on genotype. For this analysis, we utilized only root-derived low P and P-supplemented low P (low+P) treatments. We utilized a group design framework to facilitate the construction of specific contrasts. In the group variable we created, we merged the genotype and phosphate levels per sample (e.g., Col-0_lowP, phf1_low+P, or phr1 phl1_lowP). We controlled the paired structure of our design by adding a plot variable, resulting in the following model design:

Abundance ~ Plot + group

We delimited 6 sets (S1, S2, S3, S4, S5, S6) of statistically significant ($q < 0.1$) ASVs from our model using the following contrasts:

S1 = {Samples from Col-0, higher abundance in low treatment in comparison to low+P treatment}

S2 = {Samples from phf1, higher abundance in low treatment in comparison to low+P treatment}

S3 = {Samples from phr1 phl1, higher abundance in low treatment in comparison to low+P treatment}

S4 = {Samples from Col-0, higher abundance in low+P treatment in comparison to low treatment}

S5 = {Samples from phf1, higher abundance in low+P treatment in comparison to low treatment}

S6 = {Samples from phr1 phl1, higher abundance in low+P treatment in comparison to low treatment}

The 6 sets described above were used to populate Figures S3.3c-f.

The interactive visualization of the enrichment profiles was performed by converting the taxonomic assignment of each ASV into a cladogram with equidistant branch lengths using R. We used the interactive tree of life (iTOL) interface (Letunic & Bork, 2016) to visualize this tree jointly with metadata files derived from the output of the statistical models described above. The cladograms for both bacteria and fungi can be downloaded from the links described above or via the iTOL user hallepi.

In order to compare beta-diversity patterns across samples, we only used samples coming from pots in which sequence data from all 3 fractions (soil, root, and shoot) passed quality filtering. Then, for each fraction, we estimated a distance structure between samples inside that fraction using the Bray-Curtis dissimilarity metric. Finally, we computed Mantel (Mantel, 1967)correlations between pairs of distance objects (e.g., samples from root or samples from shoot) using the vegan package version 2.5–3 (Dixon, 2003)implementation of the Mantel test. All scripts and data sets required to reproduce the soil experiment analyses are deposited in the following GitHub repository: https://github.com/isaisg/hallepi/.

3.11.4.4 Inspection of other edaphic factors in the soil

To inspect whether the genotype or P effects that we observed are confounded by another edaphic factor in the soil, we cross-referenced our data set with the edaphic factors reported for the same soil plots in the work by Robbins and colleagues (Robbins et al., 2018). Because most of the edaphic factors are correlated (Figure 3.11a), we considered the first 3 principal components

(PCs) derived from these edaphic factors. These 3 PCs encompass 87% of the cumulative variance in the edaphic factor matrix (Figure 3.11b). A PERMANOVA model of the root community composition that considers these 3 PCs assigns 21% of explained variance to the first PC, which is composed of 8 edaphic factors (Figure 3.5c and 3.11b). Nonetheless, the P and genotype variables explain a similar proportion of variance as in a model that did not account for the other edaphic factors, indicating that they are orthogonal to the other variables that can be accounted for and are not confounded by them.

3.11.4.5 Amplicon sequence data processing—SynCom experiments

SynCom sequencing data were processed with MT-Toolbox (Yourstone et al., 2014). Usable read output from MT-Toolbox (i.e., reads with 100% correct primer and primer sequences that successfully merged with their pair) were quality filtered using Sickle by not allowing any window with Q-score under 20. The resulting sequences were globally aligned to a reference set of 16S rRNA gene sequences extracted from genome assemblies of SynCom member strains. For strains that did not have an intact 16S rRNA gene sequence in their assembly, we generated the 16S rRNA gene using Sanger sequencing. The reference database also included sequences from known bacterial contaminants and Arabidopsis organellar 16S sequences. Sequence alignment was performed with USEARCH version 7.1090 (Edgar, 2010) with the option 'usearch_global' at a 98% identity threshold. On average, 85% of sequences matched an expected isolate. Our 185 isolates could not all be distinguished from each other based on the V3-V4 sequence and were thus classified into 97 USeqs. A USeq encompasses a set of identical (clustered at 100%) 16S rRNA V3-V4 sequences coming from a single or multiple isolates.

Sequence mapping results were used to produce an isolate abundance table. The remaining unmapped sequences were clustered into Operational Taxonomic Units (OTUs) using UPARSE

(Edgar, 2013) implemented with USEARCH version 7.1090 at 97% identity. Representative OTU sequences were taxonomically annotated with the RDP classifier (Q. Wang et al., 2007) trained on the Greengenes database (DeSantis et al., 2006) (4 February 2011). Matches to Arabidopsis organelles were discarded. The vast majority of the remaining unassigned OTUs belonged to the same families as isolates in the SynCom. We combined the assigned USeq and unassigned OTU count tables into a single table.

The resulting count table was processed and analyzed with functions from the ohchibi package. Samples were rarefied to 1,000 reads per sample. An alpha diversity metric (Shannon diversity) was calculated using the diversity function from the vegan package version 2.5–3 (Dixon, 2003). We used ANOVA to test for differences in alpha diversity between groups. Beta-diversity analyses (Principal coordinate analysis and CAP) were based on were based on Bray-Curtis dissimilarity calculated from the rarefied abundance tables. We used the capscale function from the vegan R package version 2.5–3 (Dixon, 2003) to compute the CAP. To analyze the full data set (all fraction, all phosphate treatments), we constrained by fraction and phosphate concentration while conditioning for the replicate effect. We performed the Fraction:Phosphate interaction analysis within each fraction independently, constraining for the phosphate concentration while conditioning for the rep effect. In addition to CAP, we used PERMANOVA analysis over the 2 data sets described above using the adonis function from the vegan package version 2.5–3 (Dixon, 2003). Finally, we used the function chibi.permanova from the ohchibi package to plot the R2 values for each significant term in the PERMANOVA model tested.

We visualized the relative abundance of the bacterial phyla present in the SynCom using the stacked bar representation encoded in the chibi.phylogram from the ohchibi package.

We used the package DESeq2 version 1.22.1 (Love et al., 2014) to compute the enrichment profiles for USeqs and OTUs present in the count table. For the full data set model, we estimated main effects for each variable tested (fraction and phosphate concentration) using the following model specification:

Abundance ~ Fraction + Phosphate treatment + Replicate

We calculated the USeqs/OTUs fraction enrichments using the following contrasts: agar versus root, agar versus shoot, and root versus shoot. A USeq/OTU was considered statistically significant if it had $q < 0.1$. In order to populate the heat maps shown in Figure 3.5c, we grouped the fraction and phosphate treatment variable into a new group variable that allowed us to fit the following model:

Abundance ~ Replicate + group

We used the fitted model to estimate the fraction effect inside each particular phosphate level (e.g., Root versus agar at 0Pi, or shoot versus agar at 1,000Pi).

Additionally, we utilized a third model for the identification of USeqs/OTUs that exhibited a significant Fraction:Phosphate interaction between the planted agar samples and the plant fractions (root and shoot). Based on the beta-diversity and alpha-diversity results, we only used samples that were treated with 0, 10, 100, and 1,000 μM of phosphate. We grouped the samples into 2 categories based on their phosphate concentration, low (0 μM and 10 μM) and high (100 μM and 1,000 μM). Then we used the following model specification to derive the desired interaction effect:

Abundance ~ Fraction + Category + Fraction:Category + Replicate

Finally, we subset USeqs that exhibited a significant interaction (Fraction:Category, $q < 0.1$) in the following 2 contrasts (planted agar versus root) and (planted agar versus shoot).

In order to compare beta-diversity patterns across samples, we only used samples coming from pots in which sequence data from all 3 fractions (soil root and shoot) passed quality filtering. Then, for each fraction, we estimated a distance structure between samples inside that fraction using the Bray-Curtis dissimilarity metric. Finally, we computed Mantel (Mantel, 1967) correlations between pairs of distance objects (e.g., samples from root or samples from shoot) using the vegan package version 2.5–3 (Dixon, 2003) implementation of the Mantel test.

For the drop-out experiment, we ran an ANOVA model inside each of the phosphate treatments tested (50 μM Pi, 1,000 μM Pi, and 50→1,000 μM Pi). We visualized the results of the ANOVA models using the compact letter display encoded in the CLD function from the emmeans package (Searle et al., 1980). All scripts necessary to reproduce the synthetic community analyses are deposited in the following GitHub repository: https://github.com/isaisg/hallepi.

3.10.4.6 Phylogenetic inference of the SynCom isolates

To build the phylogenetic tree of the SynCom isolates, we utilized the supermatrix approach previously described in the work by Levy and colleagues (Levy et al., 2018). Briefly, we scanned 120 previously defined marker genes across the 185 isolate genomes from the SynCom utilizing the hmmsearch tool from the hmmer version 3.1b2 (Eddy, 2011). Then, we selected 47 markers that were present as single copy genes in 100% of our isolates. Next, we aligned each individual marker using MAFFT (Katoh et al., 2002) and filtered low quality columns in the alignment using trimAl (Capella-Gutiérrez et al., 2009). Afterward, we concatenated all filtered alignments into a super-alignment. Finally, FastTree version 2.1(M. N. Price et al., 2010) was used to infer the phylogeny utilizing the WAG model of evolution.

We utilized the inferred phylogeny along with the fraction fold change results of the main effect model to compute the phylogenetic signal (Pagel's λ) for each contrast (planted agar versus

root, planted agar versus shoot, and root versus shoot) along each concentration of the phosphate gradient. The function phylosig from the R package phytools (Revell, 2012) was used to test for significance of the phylogenetic signal measured. Multiple panel figures were constructed using the egg R package.

3.11.4.7 RNA-Seq read processing

Initial quality assessment of the Illumina RNA-Seq reads was performed using FastQC version 0.11.7. Trimmomatic version 0.36 (Bolger et al., 2014) was used to identify and discard reads containing the Illumina adaptor sequence. The resulting high-quality reads were then mapped against the TAIR10 (Berardini et al., 2015) Arabidopsis reference genome using HISAT2 version 2.1.0 (Kim et al., 2015) with default parameters. The featureCounts function from the Subread package (Liao et al., 2013) was then used to count reads that mapped to each one of the 27,206 nuclear protein-coding genes. Evaluation of the results of each step of the analysis was done with MultiQC version 1.1 (Ewels et al., 2016). Raw sequencing data and read counts are available at the NCBI Gene Expression Omnibus accession number GSE129396.

3.11.4.8 RNA-Seq statistical analysis—Soil experiment

To measure the transcriptional response to Pi limitation in soil, we used the package DESeq2 version 1.22.1 (Love et al., 2014) to define differentially expressed genes (DEGs) using the raw count table described above. We used only samples from low P and P-supplemented low P (low+P) treatments along the 3 genotypes tested (Col-0, phf1, and phr1 phl1). We combined the genotype and P treatment variables into a new group variable (e.g., Col-0_lowP or phf1_low+P). Because we were interested in identifying DEGs among any pair of levels (6 levels) of the group variable (e.g., Col-0_lowP versus Col-0_low+P) we performed a likelihood ratio test (LRT)

between a model containing the group variable and a reduced model containing only the intercept. Next, we defined DEGs as genes that had a $q < 0.1$.

For visualization purposes, we applied a variance stabilizing transformation to the raw count gene matrix. We then standardized (z-score) each gene along the samples. We subset DEGs from this standardized matrix and for each gene calculated the mean z-score expression value in a particular level of the group variable (e.g., Col-0_lowP); this resulted in a matrix of DEGs across the 6 levels in our design. Next, we created a dendrogram of DEGs by applying hierarchical clustering (method ward.D2, hclust R-base) to a distance object based on the correlation (dissimilarity) of the expression profiles of the genes across the 6 levels in our design. Finally, we delimited the cluster of DEGs by cutting the output dendogram into 5 groups using the R-base cutree function. GO enrichment was performed for each cluster of DEGs using the R package clusterProfiler (G. Yu et al., 2012).

For the PSR marker gene analysis, we downloaded the ID of 193 genes defined in the work by Castrillo and colleagues (Castrillo et al., 2017). Then, we subset these genes from our standardized matrix and computed for each gene the mean z-score expression value in a particular level of the group variable. Finally, we visualized the average expression of this PSR regulon across our groups of interest utilizing the function chibi.boxplot from the ohchibi package. All scripts necessary to reproduce the RNA-Seq analyses are deposited in the following GitHub repository: https://github.com/isaisg/hallepi.

3.11.4.9 RNA-Seq statistical analysis—SynCom experiment

To measure the transcriptional response to Pi limitation in the SynCom microcosm, we used the package DESeq2 version 1.22.1 (Love et al., 2014) to define DEGs using the raw count gene table. We combined the bacteria (NB, Full SynCom) and P treatment variables into a new

group variable (e.g., NB_50Pi or Full_1000Pi). Afterward we fitted the following model to our gene matrix:

$$\text{Abundance gene} \sim \text{Rep} + \text{Group}$$

Finally, utilizing the model fitted, we contrasted the phosphate treatment inside each level of the bacteria variable (e.g., NB_1000Pi versus NB_50Pi). Any gene with $q < 0.1$ was defined as differentially expressed.

For the PSR marker gene analysis, we downloaded the ID of 193 genes defined in the work by Castrillo and colleagues (Castrillo et al., 2017). Then, we subset these genes from our standardized matrix and computed for each gene the mean z-score expression value in a particular level of the group variable. Finally, we visualized the average expression of the PSR regulon across our groups of interest utilizing the function chibi.boxplot from the ohchibi package. All scripts necessary to reproduce the RNA-Seq analyses are deposited in the following GitHub repository: https://github.com/isaisg/hallepi.

CHAPTER 4: A SINGLE BACTERIAL GENUS MAINTAINS ROOT GROWTH IN A COMPLEX MICROBIOME[4]

Introduction

Plant phenotypes, and ultimately fitness, are influenced by the microorganisms that live in close association with them (Castrillo et al., 2017; Durán et al., 2018; Herrera Paredes et al., 2018). These microorganisms—collectively termed the plant microbiota—assemble on the basis of plant and environmentally derived cues (Castrillo et al., 2017; Finkel et al., 2019; Fitzpatrick et al., 2018; Thiergart et al., 2020), resulting in myriad interactions between plants and microorganisms. Beneficial and detrimental microbial effects on plants can be either direct (Herrera Paredes et al., 2018; Hogenhout et al., 2009; Ludwig-Müller, 2015; Mylona et al., 1995), or an indirect consequence of microorganism–microorganism interactions (Carlström et al., 2019; Durán et al., 2018). Although antagonistic interactions between microorganisms are known to have an important role in shaping plant microbiota and protecting plants from pathogens(Durán et al., 2018), another potentially important class of interactions is metabolic signal interference (Faure et al., 2009; Leadbetter & Greenberg, 2000): rather than direct antagonism, microorganisms interfere with the delivery of chemical signals produced by other microorganisms, which alters plant–microorganism signaling (Leadbetter & Greenberg, 2000; J. H. J. Leveau & Lindow, 2005; Zúñiga et al., 2013).

[4] A portion of the results presented in this chapter appeared as an article in:
Finkel OM*, Salas-González I*, Castrillo G*, Conway JM*, Law TF, Teixeira PJPL, Wilson ED, Fitzpatrick CR, Jones CD, Dangl JL. A single bacterial genus maintains root growth in a complex microbiome. Nature. 2020 Nov;587(7832):103-108.

Plant hormones—in particular, auxins—are both produced and degraded by an abundance of plant-associated (Donoso et al., 2017; S. Gilbert et al., 2018; J. H. J. Leveau & Gerards, 2008; Sun et al., 2018). Microbially derived auxins can have effects on plants that range from growth promotion to the induction of disease, depending on context and concentration(Ludwig-Müller, 2015). The intrinsic root developmental patterns of the plant are dependent on finely calibrated auxin and ethylene concentration gradients with fine differences across tissues and cell types (Brumos et al., 2018); it is not known how the plant integrates exogenous, microbially derived auxin fluxes into its developmental plan.

Here we apply a synthetic community to axenic plants as a proxy for root-associated microbiomes in natural soils to investigate how interactions between micro-organisms shape plant growth. We establish plant colonization patterns across 16 abiotic conditions to guide stepwise deconstruction of the synthetic community, which led to the identification of multiple levels of microorganism–microorganism interaction that interfere with the additivity of bacterial effects on root growth. We demonstrate that a single bacterial genus (*Variovorax*) is required for maintaining the intrinsically controlled developmental programme of the root, by tuning its chemical landscape. We establish *Variovorax* as a core taxon in the root microbiota of diverse plants grown in diverse soils. Finally, we identify a locus conserved across *Variovorax* strains that is responsible for this phenotype.

4.1 Microbial interactions control root growth

To model plant–microbiota interactions in a fully controlled setting, we established a plant–microbiota microcosm that represents the native bacterial root-derived microbiota on agar plates. We inoculated 7-day-old seedlings with a defined 185-member bacterial synthetic community (Finkel et al., 2019) composed of the major root-associated phyla (Castrillo et al., 2017; Finkel et

al., 2019; Fitzpatrick et al., 2018; Levy et al., 2018; Thiergart et al., 2020) (Figure 4.5a). We exposed this microcosm to each of 16 abiotic contexts by manipulating one of four variables (salinity, temperature, a previously reported phosphate concentration gradient (Finkel et al., 2019) and pH). We measured the composition of the synthetic community in root, shoot and agar fractions 12 days after inoculation using 16S rRNA amplicon sequencing.

The composition of the resulting root and shoot microbiota recapitulated phylum-level plant-enrichment patterns seen in soil-planted *Arabidopsis* (Castrillo et al., 2017) (Figure 4.5b). We validated the patterns observed in the agar-based system using seedlings grown in sterilized potting soil (Kremer et al., 2018) that were inoculated with the same synthetic community. Both relative abundance and plant-enrichment patterns at the unique sequence level were significantly correlated between the agar- and soil-based systems, which confirms the applicability of our relatively high-throughput agar-based system as a model for the assembly of plant microbiota (Figure 4.5c). Within the agar system, both fraction (substrate, root or shoot) and abiotic conditions significantly affected α- and β-diversity (Figure 4.5d-f).

To guide the deconstruction of the synthetic community into modules, we calculated pairwise correlations in relative abundance across all samples, and identified four well-defined modules of co-occurring strains that we termed modules A, B, C and D (Figure 4.1a). These modules formed distinct phylogenetically structured guilds in association with the plant. Module A contained mainly Gammaproteobacteria and was predominantly more abundant in the substrate than in the seedling; module B contained mainly low-abundance Firmicutes, with no significant enrichment trend; and modules C and D were composed mainly of Alphaproteobacteria and Actinobacteria, respectively, and showed plant enrichment across all abiotic conditions. Both

Alphaproteobacteria (module C) and Actinobacteria (module D) are consistently plant enriched across plant species (Fitzpatrick et al., 2018), which suggests that these clades contain plant-association traits that are deeply rooted in their evolutionary histories.

We next asked whether the different modules of co-occurring strains have different roles in determining plant phenotype. We inoculated seedlings with synthetic communities composed of modules A, B, C and D singly or in all six possible pairwise combinations, and imaged the seedlings 12 days after inoculation. We observed strong primary root growth inhibition (RGI) in seedlings inoculated with plant-enriched module C or D (Figure 4.1b,c). RGI did not occur in seedlings inoculated with module A or B, which do not contain plant-enriched strains (Figure 4.1b). To test whether the root phenotype derived from each module is an additive outcome of its individual constituents, we inoculated seedlings in mono-association with each of the 185 members of the synthetic community. We observed that 34 taxonomically diverse strains, distributed across all 4 modules, induced RGI (Figure 4.6a-c). However, neither the full synthetic community nor derived synthetic communities that consisted of module A or B exhibited RGI (Figure 4.1b). Thus, binary plant–microorganism interactions were not predictive of interactions in this complex-community context.

In seedlings inoculated with module pairs, we observed epistatic interactions: in the presence of module A, the RGI caused by modules C and D was reverted (Figure 4.1b, c). Thus, by deconstructing the synthetic community into four modules, we found that bacterial effects on root growth are governed by multiple levels of microorganism–microorganism interaction. This is exemplified by at least four instances: within module A or B and between module A and module C

or D. Because three of these interactions involve module A, we predicted that this module contains strains that strongly attenuate RGI and preserve stereotypic root development.

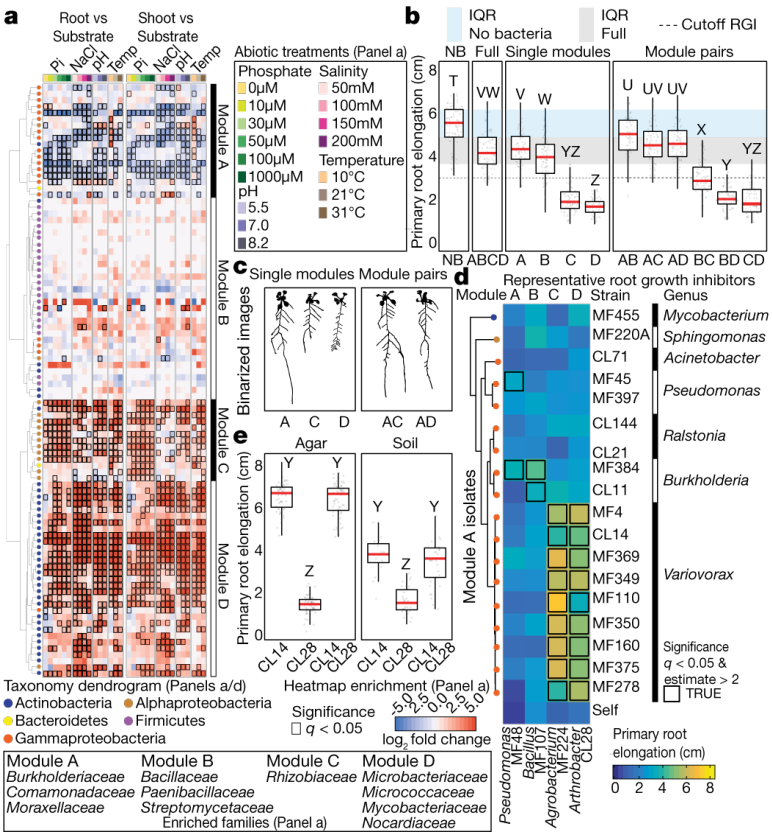


Figure 4.1 - **Arabidopsis root length is governed by bacteria–bacteria interactions within a community.**

a) Fraction enrichment patterns of the synthetic community across abiotic gradients. Each row represents a unique sequence. The four modules of co-occurring strains (A, B, C and D) are represented. Dendrogram tips are coloured by taxonomy. Heat maps are coloured by log2-transformed fold changes derived from a fitted generalized linear model and represent enrichments in plant tissue (root or shoot) compared with substrate. Comparisons with false-discovery-rate (FDR)-corrected q-value < 0.05 are contoured in black. Enriched families within each module are listed below the heat map. n = 6 biological replicates across 2 independent experiments. Pi, inorganic phosphate; temp, temperature. **b)** Primary root elongation of seedlings grown axenically (no bacteria, NB), with the full synthetic community (ABCD) or its subsets: modules A, B, C and D alone (single modules), and all pairwise combination of modules (module pairs). Significance was determined via analysis of variance (ANOVA); letters correspond to a Tukey post hoc test. n = 75, 89, 68, 94, 87, 77, 76, 96, 82, 84, 89 and 77 (from left to right) biological replicates across 2 independent experiments. **c)** Binarized image of representative seedlings inoculated with modules

A, C and D, and with module combinations A–C and A–D. **d)** Heat map coloured by average primary root elongation of seedlings inoculated with four representative RGI-inducing strains from each module (columns A–D) in combination with isolates from module A (rows) or alone (self). Significance was determined via ANOVA. **e)** Primary root elongation of seedlings inoculated with *Arthrobacter* CL28 and *Variovorax* CL14 individually or jointly across two substrates. Significance was determined via ANOVA, letters are the results of a Tukey post hoc test. n = 64, 64, 63, 17, 36 and 33 (from left to right) biological replicates across 2 independent experiments. In all box plots, the centre line represents the median, box edges show the 25th and 75th percentiles, and whiskers extend to 1.5× the interquartile range (IQR).

4.2 *Variovorax* maintain stereotypical root growth

To identify strains within module A that are responsible for intra- and inter-module attenuation of RGI, we reduced our system to a tripartite plant–microorganism–microorganism system. We individually screened the 18 non-RGI strains from module A for their ability to attenuate RGI caused by representative strains from all 4 modules. We found that all the tested strains from the genus *Variovorax* (family Comamonadaceae) suppressed the RGI caused by representative RGI-inducing strains from module C (*Agrobacterium* MF224) and module D (*Arthrobacter* CL28) (Figure 4.1d). The strains from modules A (*Pseudomonas* MF48) and B (*Bacillus* MF107) were not suppressed by *Variovorax*, but rather by two closely related *Burkholderia* strains (CL11 and MF384) (Figure 4.1d). A similar pattern was observed when we screened two selected RGI-suppressing *Variovorax* strains (CL14 and MF160) and *Burkholderia* CL11 against a diverse set of RGI-inducing strains. *Variovorax* attenuated 13 of the 18 RGI-inducing strains that we tested (Figure 4.7a).

To test whether the RGI induction and suppression we observed on agar occur in soil as well, we germinated *Arabidopsis* on sterile soil inoculated with an RGI-suppressing and -inducing pair of strains: the RGI-inducing *Arthrobacter* CL28 and the RGI-suppressing *Variovorax* CL14. As expected, *Arthrobacter* CL28 induced RGI, which was reverted by *Variovorax* CL14 in soil

(Figure 4.1e). We generalized this observation by showing that *Variovorax*-mediated attenuation of RGI extended to tomato seedlings, in which *Variovorax* CL14 reverted *Arthrobacter* CL28-mediated RGI (Figure 4.7b). Finally, we tested whether the RGI-suppressing strains maintain their capacity to attenuate RGI in the context of the full 185-member community. We compared the root phenotype of seedlings exposed to either the full synthetic community or to the same community after dropping out all ten *Variovorax* strains and/or all six *Burkholderia* strains present in the synthetic community (Figure 4.2a). We found that *Variovorax* is necessary and sufficient to revert RGI within the full community (Figures 4.2b, c and 4.8a). This result was robust across a range of substrates (including soil), and under various biotic and abiotic contexts (Figures 4.2d-f and 4.8b). Further, the presence of *Variovorax* in the synthetic community increases both the total length of root network of the plant and its shoot size (Figure 4.8c, d). Importantly, the latter is considered a reliable proxy for relative plant fitness (Abreu & Munne-Bosch, 2009; Clauss & Aarssen, 1994), which suggests that *Variovorax*-mediated suppression of RGI is adaptive.

To ascertain the breadth of the ability of *Variovorax* to attenuate RGI, we tested additional *Variovorax* strains from across the phylogeny of this genus (Figure 4.9a). All 19 of the *Variovorax* strains we tested reverted the RGI induced by *Arthrobacter* CL28. A strain from the nearest plant-associated outgroup to this genus (*Acidovorax* root219 (Bai et al., 2015)) did not revert RGI (Figure 4.9a, b). Thus, all tested strains—representing the broad phylogeny of *Variovorax*—interact with a wide diversity of bacteria to enforce stereotypic root development within complex communities, independent of biotic or abiotic contexts. Importantly, we found no evidence that this phenotype is achieved by outcompeting or antagonizing RGI-inducing strains (Figures 4.2g, h and 4.10).

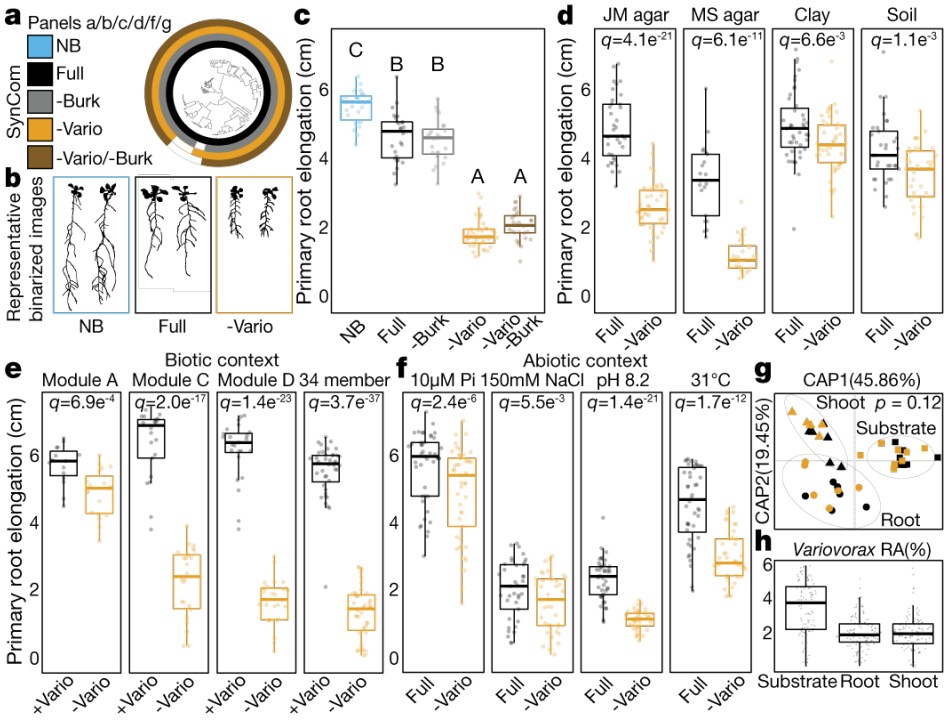


Figure 4.2 - ***Variovorax* maintains stereotypic root development.**

a) Phylogenetic tree of 185 bacterial isolates. Concentric rings represent isolate composition of each synthetic community treatment. NB, uninoculated (no bacteria); full, full synthetic community; −Burk, *Burkholderia* drop-out; −Vario, *Variovorax* drop-out; −Vario, −Burk, *Burkholderia* and *Variovorax* drop-out. **b)** Binarized image of representative seedlings inoculated or not with the full synthetic community or the *Variovorax* drop-out synthetic community. **c)** Primary root elongation of seedlings grown axenically (NB) or inoculated with the different synthetic community treatments outlined in a. Significance was determined via ANOVA, letters correspond to a Tukey post hoc test. n = 26, 26, 24, 37 and 29 (from left to right) biological replicates. **d)** Primary root elongation of seedlings inoculated with the full synthetic community or with the *Variovorax* drop-out synthetic community across different substrates: Johnson medium (JM) agar, Murashige and Skoog (MS) agar, sterilized clay or potting soil. n = 36, 47, 21, 24, 43, 48, 33 and 36 (from left to right) biological replicates across 2 independent experiments. **e)** Primary root elongation of seedlings inoculated with four subsets of the full synthetic community (module A, C, D or a previously described 35-member synthetic community with its single *Variovorax* strain removed (34-member)1), with (+Vario) or without (−Vario) Variovorax isolates. n = 40, 40, 15, 19, 22, 26, 25 and 29 (from left to right) biological replicates across 2 independent experiments. **f)** Primary root elongation of seedlings inoculated with the full synthetic community or with the *Variovorax* drop-out synthetic community across different abiotic stresses: low phosphate, high salt, high pH and high temperature. n = 43, 45, 37, 37, 43, 44, 44 and 43 (from left to right) biological replicates across 2 independent experiments. Significance was determined within each

condition via ANOVA in d, f, and with a two-sided t-test in e. FDR-adjusted P values are displayed. **g)** Canonical analysis of principal coordinates (CAP) scatter plots comparing the compositions of the full synthetic community and *Variovorax* drop-out synthetic community (colour code as in a) across fractions (substrate (squares), root (circles) and shoot (triangles)). Permutational multivariate ANOVA (PERMANOVA) P value is shown. n = 7 (substrate + full), 8 (substrate + −Vario), 6 (root + full), 6 (root + −Vario), 5 (shoot + full) or 5 (shoot + −Vario). **h)** Relative abundance (RA) of the *Variovorax* genus within the full synthetic community across the agar, root and shoot fractions. n = 127, 119 and 127 biological replicates for agar, root and shoot, respectively, across 16 independent experiments.

4.3 *Variovorax* manipulates auxin and ethylene

To study the mechanisms that underlie bacterial effects on root growth, we analyzed the transcriptomes of seedlings colonized for 12 days with the RGI-inducing strain *Arthrobacter* CL28 and the RGI-suppressing strain *Variovorax* CL14, either in mono-association with the seedling or in a tripartite combination (Figure 4.1e). We also performed RNA sequencing (RNA-seq) on seedlings colonized with the full synthetic community (no RGI) or the *Variovorax* drop-out synthetic community (RGI) (Figure 4.11a). Eighteen genes were significantly induced only under RGI conditions (RGI-induced) across both experiments (Figure 4.11a, b). Seventeen of these genes are co-expressed with genes that have proposed functions related to the root apex (Klepikova et al., 2016) (Figure 4.11b, c). The remaining gene, *GH3.2*, encodes indole-3-acetic acid-amido synthetase, which conjugates excess amounts of the plant hormone auxin and is a robust marker for late auxin responses (Takase et al., 2004; uchida et al., 2018) (Figure 4.11b). The production of auxins is a well-documented mechanism by which bacteria modulate plant root development (J. H. J. Leveau & Lindow, 2005). Indeed, the top 12 auxin-responsive genes from a previous RNA-seq study examining acute auxin response in *Arabidopsis* (uchida et al., 2018) exhibited an average transcript increase in seedlings exposed to our RGI-inducing conditions (Figure 4.11d). We hypothesized that the suppression of RGI by *Variovorax* is probably mediated by interference with bacterially produced auxin signalling.

We asked whether the suppression of RGI by *Variovorax* is directly and exclusively related to auxin signalling. Besides auxin, other small molecules cause RGI. These include the plant hormones ethylene (Chen et al., 2013) and cytokinin (Cary et al., 1995), as well as microbial-associated molecular patterns such as the flagellin-derived peptide flg22 (Gómez-Gómez et al., 1999). We tested the ability of diverse *Variovorax* strains and of the *Burkholderia* strain CL11 to revert the RGI induced by auxins (indole-3-acetic acid (IAA) and the auxin analogue 2,4-dichlorophenoxyacetic acid), ethylene (the ethylene precursor 1-aminocyclopropane-1-carboxylic acid (ACC)), cytokinins (zeatin and 6-benzylaminopurine) and the flg22 peptide. All of the *Variovorax* strains we tested suppressed the RGI induced by IAA or ACC (Figure 4.3a) —with the exception of *Variovorax* YR216, which did not suppress ACC-induced RGI and does not contain an ACC deaminase gene, a plant-growth-promoting feature that is associated with this genus (Chen et al., 2013) (Figure 4.9a). *Burkholderia* CL11 only partially reverted ACC-induced RGI (Figure 4.3a). None of the *Variovorax* strains attenuated the RGI induced by 2,4-dichlorophenoxyacetic acid, flg22 or cytokinins (Figures 4.3a and 4.12a). Importantly, this function is mediated by recognition of auxin by *Variovorax* and not by the plant auxin response per se, as the auxin response (RGI) induced by 2,4-dichlorophenoxyacetic acid is not reverted. Indeed, we found that *Variovorax* CL14 degrades IAA in culture (Figure 4.12b) and quenches fluorescence of the *Arabidopsis* auxin reporter line *DR5::GFP* caused by the RGI-inducing *Arthrobacter* CL28 (Figure 4.12c, d).

Auxin and ethylene are known to act synergistically to inhibit root growth (Robles et al., 2013). To ascertain the roles of both auxin and ethylene perception by the plant in responding to RGI-inducing strains, we used the auxin-insensitive *axr2-1* mutant (Nagpal et al., 2000) combined with a competitive inhibitor of ethylene receptors, 1-methylcyclopropene (1-MCP) (Hall et al.,

2000). We inoculated wild-type seedlings and *axr2-1* mutants, treated or not with 1-MCP, with the RGI-inducing *Arthrobacter* CL28 strain or the *Variovorax* drop-out synthetic community. We observed in both cases that bacterial RGI is reduced in *axr2-1* and 1-MCP-treated wild-type seedlings, and is further reduced in doubly insensitive 1-MCP-treated *axr2-1* seedlings; this demonstrates that both auxin and ethylene perception in the plant contribute additively to bacterially induced RGI (Figure 4.3b). Thus, in the absence of *Variovorax*, a complex synthetic community can induce severe morphological changes in root phenotypes via both auxin- and ethylene-dependent pathways, but both are reverted when *Variovorax* is present.

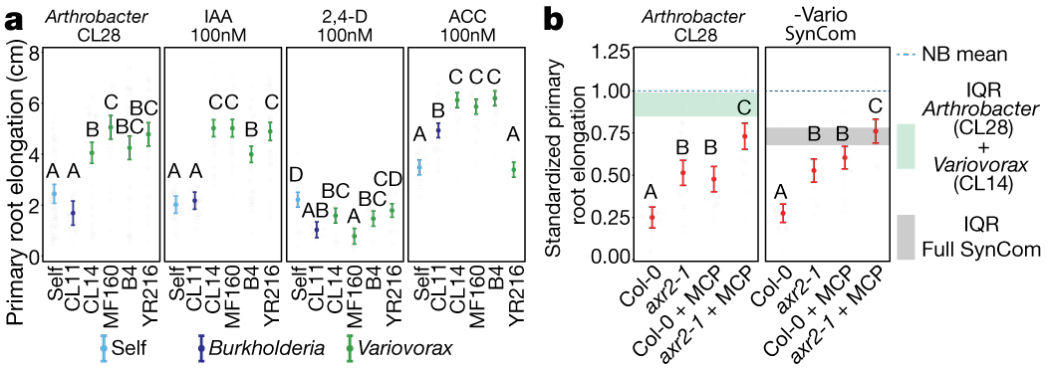


Figure 4.3 - ***Variovorax* suppression of RGI is related to auxin and ethylene signalling.**

a) Primary root elongation of seedlings grown with RGI-inducing *Arthrobacter* CL28 or three hormonal treatments (100 nM IAA, 2,4-dichlorophenoxyacetic acid (2,4-D) or ACC), individually (self) or with *Burkholderia* CL11 or one of four *Variovorax* isolates (CL14, MF160, B4 or YR216). Significance was determined within each treatment via ANOVA; letters correspond to a Tukey post hoc test. n = 74, 46, 61, 48, 49, 49, 45, 44, 46, 43, 49, 40, 22, 19, 22, 19, 20, 25, 28, 30, 29, 29, 29 and 29 (from left to right) biological replicates across 2 independent experiments. **b)** Primary root elongation, standardized to axenic conditions, of wild-type (Col-0), auxin-unresponsive (*axr2-1*), ethylene-unresponsive (Col-0 + MCP) or auxin- and ethylene-unresponsive (*axr2-1* + MCP) seedlings inoculated with RGI-inducing *Arthrobacter* CL28 or the *Variovorax* drop-out synthetic community (−Vario). The blue dotted line marks the relative mean length of uninoculated seedlings. Horizontal shading is the IQR of seedlings grown with *Arthrobacter* CL28 +*Variovorax* CL14 (aqua) or the full synthetic community (grey). Significance was determined via ANOVA; letters correspond to a Tukey post hoc test. n = 37, 25, 24, 23, 35, 21, 22 and 20 (from left to right) biological replicates across 2 independent experiments. In a, b, data represent the mean ± 95% confidence interval.

To identify the bacterial mechanism or mechanisms that are involved in the attenuation of RGI, we compared the genomes of the 10 *Variovorax* strains in the synthetic community to the genomes of the other 175 members of the synthetic community. Using de novo orthologue clustering across all 185 genomes, we identified 947 genes unique to *Variovorax*, with <5% prevalence across the 175 non-*Variovorax* members of the synthetic community and 100% prevalence among all 10 *Variovorax* strains. We grouped these genes into regions of physically contiguous genes (genomic hotspots) and focused on the 12 hotspots that contained at least 10 genes unique to *Variovorax* (Figure 4.13a). One of these hotspots (designated hotspot 33) contains weak homologues (average of about 30% identity) to the genes *iacC*, *iacD*, *iacE*, *iacF* and *iacR* of the IAA-degrading *iac* operon of *Paraburkholderia phytofirmans* strain PsJN (Donoso et al., 2017; J. H. J. Leveau & Gerards, 2008), but lacks *iacA*, *iacB* and *iacI*—which are known to be necessary for *Paraburkholderia* growth on IAA (Donoso et al., 2017) (Figures 4.4a and 4.13b). To test whether the hotspots we identified are responsive to RGI-inducing bacteria, we analysed the transcriptome of *Variovorax* CL14 in monoculture and in coculture with the RGI-inducing *Arthrobacter* CL28. We observed extensive transcriptional reprogramming of *Variovorax* CL14 when cocultured with *Arthrobacter* CL28. Among the 12 hotspots we identified, the genes in hotspot 33 were the most highly upregulated (Figures 4.4a and 4.13b, c). We thus hypothesized that hotspot 33 contains an uncharacterized auxin-degradation operon.

In parallel, we constructed a *Variovorax* CL14 genomic library in *Escherichia coli* with >12.5-kb inserts in a broad host-range vector, and screened the resulting *E. coli* clones for auxin degradation. Two clones from the approximately 3,500 that we screened degraded IAA (denoted V1 and V2). The *Variovorax* CL14 genomic inserts in both of these clones contained portions of hotspot 33 (Figures 4.4a and 4.13b). The overlap common to both of these clones contained nine

genes, among them the weak homologues to *Paraburkholedria iacC*, *iacD* and *iacE*. To test whether this genomic region is sufficient to revert RGI in plants, we transformed *Acidovorax* root219 (Bai et al., 2015), a relative of *Variovorax* that does not cause or revert RGI (Figure 4.9a, b), with the shorter functional insert (V2) (Figures 4.4a and 4.13b) or with an empty vector (EV). The resulting gain-of-function strain *Acidovorax* root219::V2 gained the ability to degrade IAA in culture (Figure 4.4b). We inoculated *Acidovorax* root219::V2 or the control *Acidovorax* root219::EV onto plants treated with IAA or inoculated with the RGI-inducing *Arthrobacter* CL28. *Acidovorax* root219::V2 fully reverted IAA-induced RGI (Figure 4.4c) and partially reverted *Arthrobacter* CL28-induced RGI, despite colonizing roots at significantly lower levels than *Variovorax* CL14 (Figures 4.4d and 4.13d). In addition, we deleted hotspot 33 from *Variovorax* CL14 (Figure 4.4a) to test whether this putative operon is necessary for the reversion of RGI. The resulting strain *Variovorax* CL14 ΔHS33—which is not impaired in plant colonization (Figure 4.13e)—did not degrade IAA in culture (Figure 4.4b), and did not revert IAA-induced (Figure 4.4c) or *Arthrobacter* CL28-induced RGI (Figure 4.4d). Thus, this *Variovorax*-specific gene cluster is necessary for the suppression of RGI and for auxin degradation. It is thus the critical genetic locus required by *Variovorax* to maintain stereotypic root development in the context of a phylogenetically diverse microbiome.

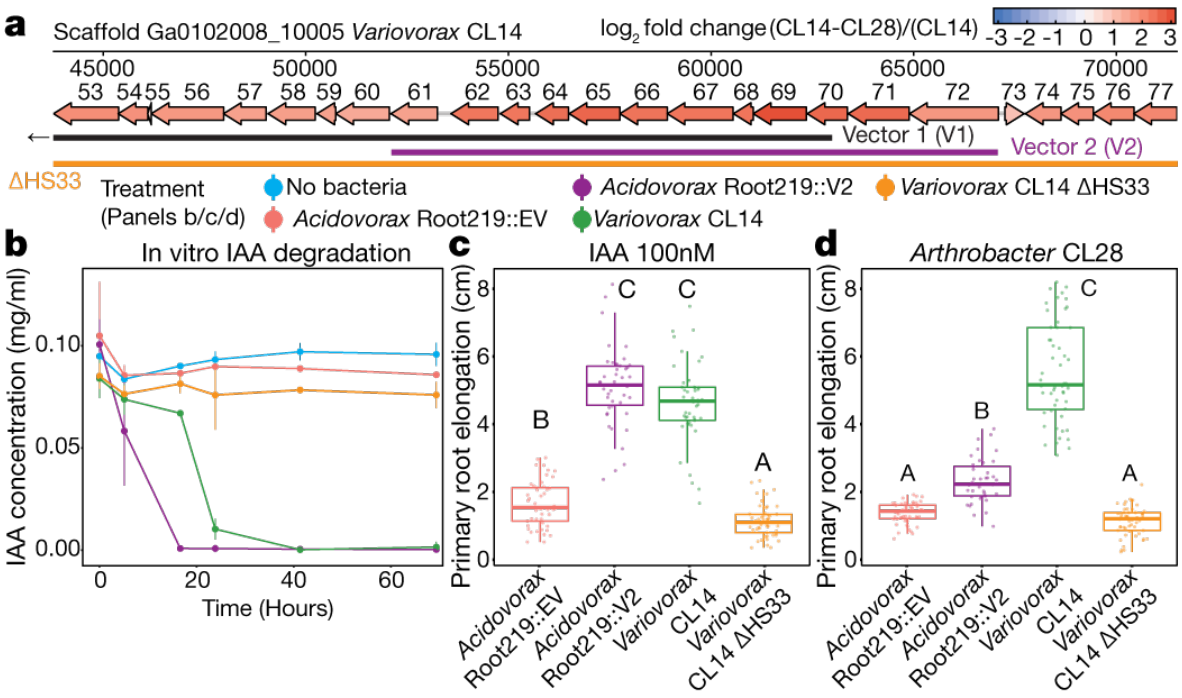


Figure 4.4 - **An auxin-degrading operon in *Variovorax* is required for root development.**

a) A map of the auxin-degrading hotspot 33. Genes are annotated with the last two digits of their IMG gene identifier (26436136XX) (Figure 4.13b), and are coloured by the log2-transformed fold change in their transcript abundance in Variovorax CL14 cocultured with *Arthrobacter* CL28 (CL14–CL28) relative to *Variovorax* CL14 (CL14) in monoculture (measured by RNA-seq). Genes contained in vector 1 (V1) and vector 2 (V2), and the region knocked-out in *Variovorax* CL14 ΔHS33, are shown below the map. **b)** In vitro degradation of IAA by *Acidovorax* root219::EV, *Acidovorax* Root219::V2, *Variovorax* CL14 and *Variovorax* CL14 ΔHS33. n = 3 biological replicates. **c) d)** Primary root elongation of seedlings treated with IAA (c) or co-inoculated with *Arthrobacter* CL28 combined with *Acidovorax* root219::EV, *Acidovorax* root219::V2, *Variovorax* CL14 and *Variovorax* CL14 ΔHS33 (d). Significance was determined via ANOVA; letters correspond to a Tukey post hoc test. c, n = 49, 46, 46 and 49 (from left to right); d, n = 51, 41, 52 and 53 (from left to right) biological replicates across 2 independent experiments.

4.4 Discussion

Signalling molecules and other secondary metabolites are products of adaptations that allow microorganisms to survive competition for primary metabolites. Our results illuminate the importance of a trophic layer of microorganisms that use these secondary metabolites for their own benefit, while potentially providing the unselected exaptation (Gould & Vrba, 1982) of interfering

with signalling between the bacterial microbiota and the plant host. Such metabolic signal interference has previously been demonstrated in the case of quorum quenching (Leadbetter & Greenberg, 2000), the degradation of microbial-associated molecular patterns (Bardoel et al., 2011) and the degradation of bacterially produced auxin, including among *Variovorax* (J. H. J. Leveau & Lindow, 2005)(Zúñiga et al., 2013)(Sun et al., 2018). Plant development relies on tightly regulated auxin concentration gradients (Brumos et al., 2018), which can be distorted by auxin fluxes emanating from the microbiota. Some *Variovorax* strains have the capacity to both produce and degrade auxin, which suggests a capacity to fine-tune auxin concentrations in the rhizosphere (Chen et al., 2013; Sun et al., 2018).

We have shown here that the chemical homeostasis enforced by the presence of *Variovorax* in a phylogenetically diverse, realistic synthetic community allows the plant to maintain its developmental programme within a chemically complex matrix. *Variovorax* was recently found to have the rare property of improved plant colonization upon late arrival to an established community (Carlström et al., 2019), which suggests that they use bacterially produced or induced—rather than plant-derived—substrates. Furthermore, after re-analysing a recent large-scale time and spatially resolved survey of the *Arabidopsis* root microbiome (Thiergart et al., 2020) and a common garden experiment including 30 plant species (Fitzpatrick et al., 2018), we noted that *Variovorax* is among a limited group of core bacterial genera found in 100% of the sampled sites and plant species (Figure 4.14a, b. These ecological observations, together with our results using a reductionist microcosm, reinforce the importance of *Variovorax* as a key player in the bacteria–bacteria–plant communication networks that are required to maintain root growth within a complex biochemical ecosystem.

4.5 Supporting figures

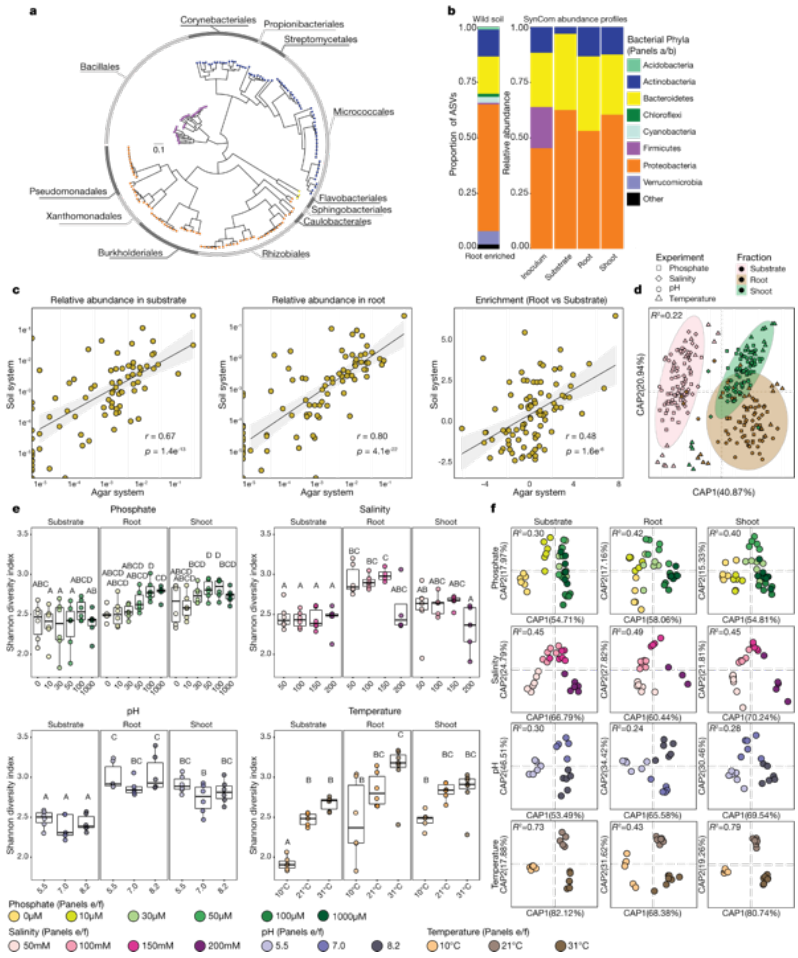


Figure 4.5 - **Synthetic community resembles the taxonomic make-up of natural communities.**

a) Phylogenetic tree of 185 genome-sequenced isolates obtained from surface-sterilized Arabidopsis roots, included in the synthetic community. The composition of this synthetic community captures the diversity of Actinobacteria, Proteobacteria and Bacteroidetes; the three major root-enriched phyla (Castrillo et al., 2017; Finkel et al., 2019; Fitzpatrick et al., 2018; Levy et al., 2018; Thiergart et al., 2020); and Firmicutes, which are abundant in plant-associated culture collections (Levy et al., 2018). Tree tips are coloured according to phylum. The outer ring shows the distribution of the 12 bacterial orders present in the synthetic community. **b)** Comparison of proportions of Proteobacteria, Bacteroidetes and Actinobacteria in synthetic-community (SynCom)-inoculated roots to root microbiota derived from plants grown in natural soil (Castrillo et al., 2017). Firmicutes, which are not plant-enriched, were reduced to <0.1% of the relative abundance (Figure 4.1a). The left panel (wild soil) shows the proportion of ASVs enriched (*q* value < 0.1) in the plant root in comparison to soil in a microbiota profiling study from the same soil from which the synthetic community strains were isolated. ASVs are coloured according to

phylum and proteobacterial ASVs are coloured by class. The right panel (synthetic community) represents the relative abundance profiles of bacterial isolates across the initial inoculum, planted agar, and root and shoot in plants inoculated with the full synthetic community. **c)** Comparison of synthetic community composition in agar versus the soil-based microcosms. Left, relative abundance in the substrate. Middle, relative abundance in root. Right, enrichment in root versus substrate. Each dot represents a single unique sequence. Pearson correlation line, 95% confidence intervals, r value and P value are shown for each comparison. n = 24 (soil system) and 8 (agar system) biological replicates. **d)** Canonical analysis of principal coordinates (CAP) showing the influence of the fraction (planted agar, root or shoot) on the assembly of the bacterial synthetic community across the four gradients used in this Article (phosphate, salinity, pH and temperature). Different colours differentiate between the fractions, and different shapes differentiate between experiments. Ellipses denote the 95% confidence interval of each fraction. Fraction (substrate, root or shoot) explains most (22%) of the variance across all abiotic variables. n = 94 (substrate), 90 (root) and 95 (shoot) biological replicates across 8 independent experiments. **e)** Abiotic conditions displayed reproducible effects on α-diversity. Each panel represents bacterial α-diversity across the different abiotic gradients (phosphate, salinity, pH and temperature) and fractions (substrate, root and shoot) used in this Article. Bacterial α-diversity was estimated using Shannon diversity index. Letters represent the results of the post hoc test of an ANOVA model testing the interaction between fraction and abiotic condition. **f)** Canonical analysis of principal coordinates scatter plots showing the effect on the composition of the synthetic community of each of the four abiotic gradients (phosphate, salinity, pH and temperature) within the substrate, root and shoot fractions. PERMANOVA R2 values are shown within each plot. e, f, Phosphate, n = 6, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 5 and 6; salinity n = 6, 6, 6, 6, 6, 6, 6, 4, 5, 6, 4 and 5; pH, n = 6, 6, 6, 5, 5, 6, 6, 6 and 6; temperature n = 6, 6, 6, 6, 6, 6, 6, 6 and 6. All samples are biologically independent and represent two independent experiments.

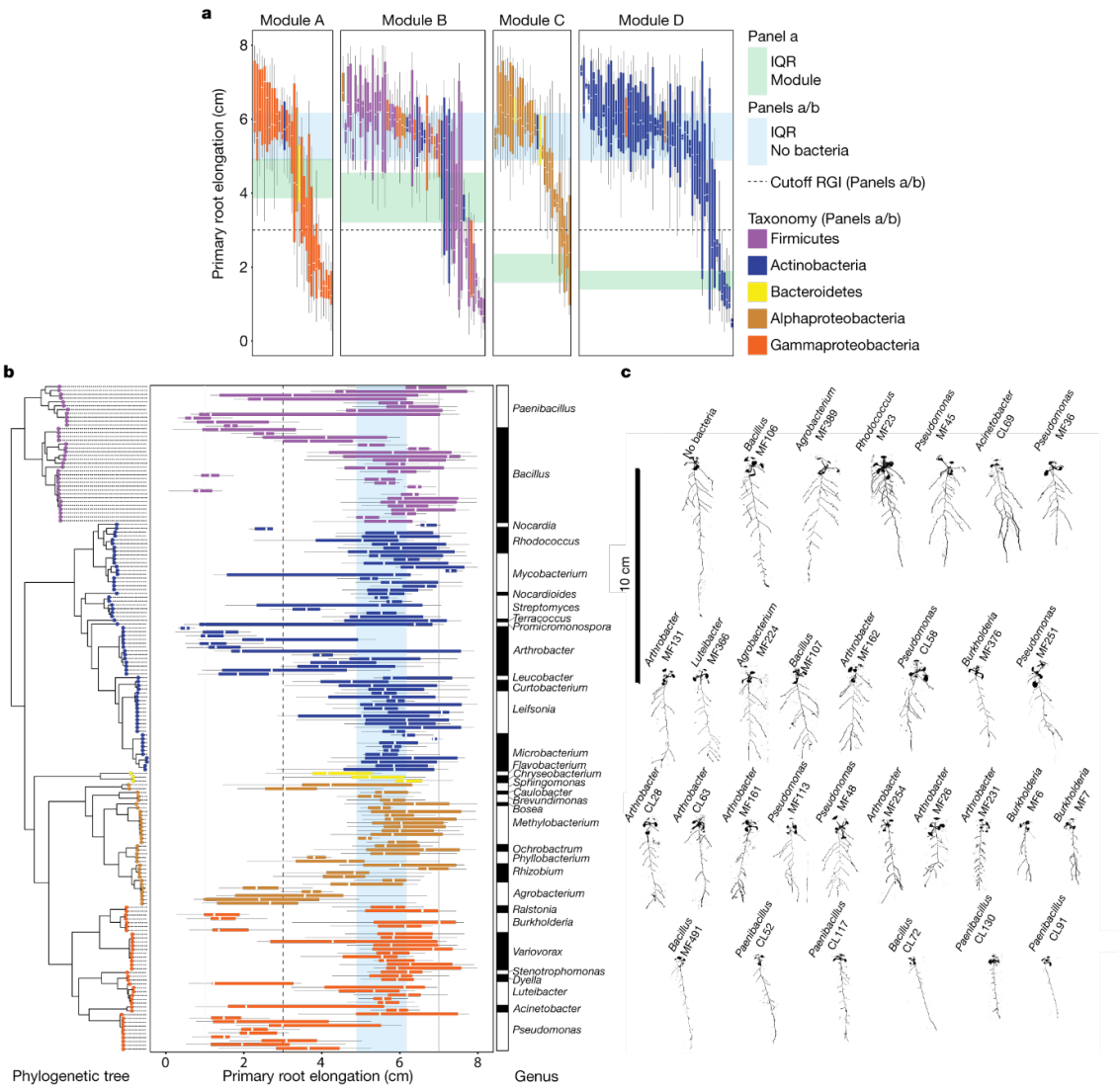


Figure 4.6 - **RGI trait is distributed across bacterial phylogeny.**

a, b) Primary root elongation of seedlings inoculated with single bacterial isolates (one box plot per isolate). Isolates are coloured by taxonomy. a, Isolates are grouped by module membership. The strips across the panels correspond to the interquartile range (IQR), as noted at the far right. The dotted line represents the cut-off used to classify isolates as root-growth inhibiting (cutoff RGI). b, Isolates are ordered according to the phylogenetic tree on the left, and coloured on the basis of their genome-based taxonomy. The vertical blue stripes across the panel correspond to the IQR of plants grown in sterile conditions. The vertical dotted line represents the 3-cm cut-off used to classify strains as RGI strains. The bar on the right denotes the genus classification of each isolate. **c)** Binarized image of representative seedlings grown axenically (no bacteria) or with 34 RGI strains individually.

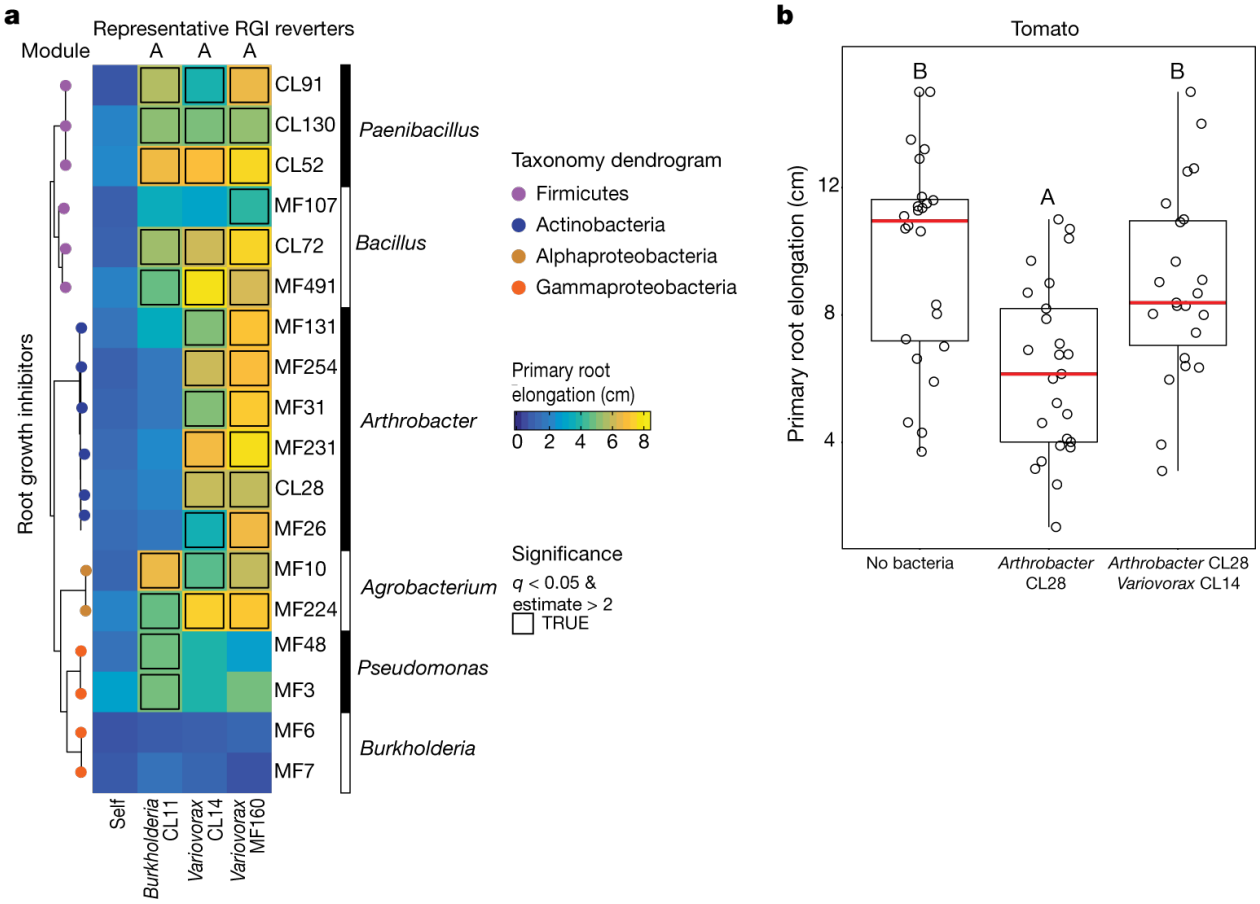


Figure 4.7 - ***Variovorax*-mediated reversion of RGI.**

a) Heat map coloured by average primary root elongation of seedlings inoculated with eighteen RGI-inducing strains (rows) alone (self) or in combination with *Burkholderia* CL11, *Variovorax* CL14 or *Variovorax* MF160 (columns). Statistically significant RGI reversions were determined via ANOVA and are outlined in black. **b)** *Variovorax*-mediated reversion of RGI is maintained in a second plant species. Primary root elongation of uninoculated tomato seedlings (no bacteria) or seedlings inoculated with the RGI-inducer *Arthrobacter* CL28 individually or along with *Variovorax* CL14 grown on vertical agar plates. Significance was determined via ANOVA while controlling for experiment; letters correspond to a Tukey post hoc test. n = 24, 25 and 23 biological replicates across 2 independent experiments.

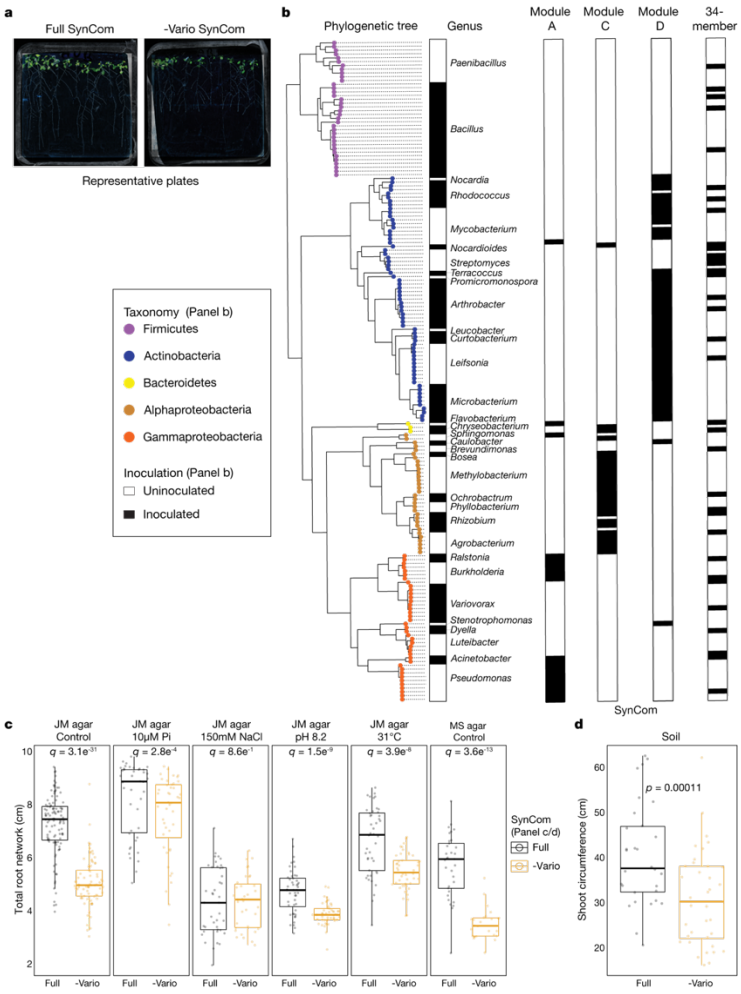


Figure 4.8 - ***Variovorax*** **maintain stereotypic plant growth.**

a) Representative plate images of plants grown with the full 185-member synthetic community (full SynCom) or the *Variovorax* drop-out community (−Vario SynCom) for 12 days. **b)** Bar graphs showing the isolate composition of synthetic communities composed by module A, module C, module D and a previously described1 34-member synthetic community (34-member). Isolates are ordered according to the phylogenetic tree on the left. The tips of the phylogenetic tree are coloured on the basis of the genome-based taxonomy of each isolate. Presence of an isolate across the different synthetic communities is denoted by a black filled rectangle (labelled 'inoculated'). **c)** Total root network quantification of *Arabidopsis* seedlings grown with the full synthetic community (full), or with the full synthetic community excluding *Variovorax* (−Vario), across different abiotic conditions: JM agar control, low phosphate (JM agar 10 μM Pi), high salt (JM agar 150 mM NaCl), high pH (JM agar pH 8.2) and high temperature (JM agar 31 °C), as well as half Murashige and Skoog medium (MS agar control). Significance was determined within each condition via ANOVA while controlling for experiment. n = 110, 77, 42, 44, 41, 35, 48, 49, 44, 45, 33 and 26 biological replicates across 2 independent experiments**. d)** Shoot circumference of plants grown in pots with potting soil with the full synthetic community or with the full synthetic community excluding *Variovorax*. Significance was determined within each condition via ANOVA while controlling for experiment. n = 30 and 36 biological replicates across 2 independent experiments.

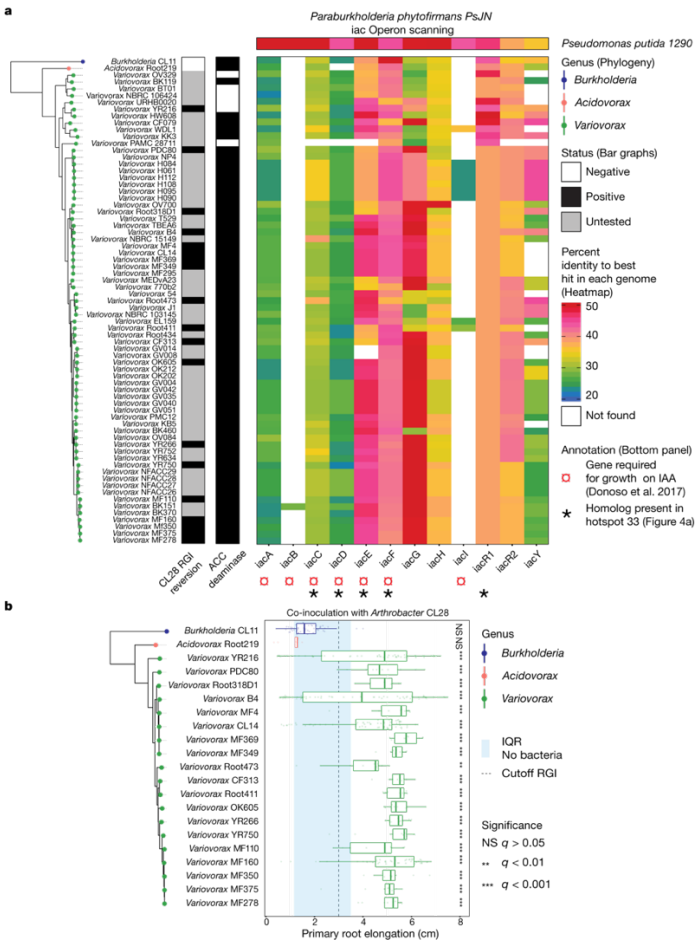


Figure 4.9 - **Reversion of RGI is prevalent across the *Variovorax* phylogeny.**

a) Phylogenetic tree of 69 publically available *Variovorax* genomes and 2 outgroup isolates, *Acidovorax* root 219 and *Burkholderia* CL11. The CL28 RGI reversion bar categorizes (positive, negative or untested) the ability of each isolate in the phylogeny to revert the RGI caused by *Arthrobacter* CL28. The ACC deaminase bar denotes the presence of the KEGG Orthology (KO) term KO1505 (1-aminocyclopropane-1-carboxylate deaminase) in each of the genomes. The heat map denotes the per cent identity of BLASTp hits in the genomes to the genes from the auxin-degrading iac operon in *Paraburkholderia phytophirmans* (Donoso et al., 2017). Synteny is not necessarily conserved, as these BLAST hits may be spread throughout the genomes. **b)** All tested *Variovorax* isolates reverted RGI. Phylogenetic tree of 19 *Variovorax* genomes and 2 outgroup isolates (*Acidovorax* root 219 and *Burkholderia* CL11) that were tested for their ability to revert the RGI imposed by *Arthrobacter* CL28. The blue vertical stripe denotes the IQR of plants treated solely with *Arthrobacter* CL28. The dotted vertical line denotes the 3-cm cut-off used to classify a treatment as an RGI. Each box plot is coloured according to the genus classification of each isolate. Significance was determined via ANOVA while controlling for experiment, letters correspond to a Tukey post hoc test. n = 59, 9, 55, 71, 10, 10, 10, 9, 10, 10, 57, 10, 10, 48, 10, 10, 10, 9, 10, 10 and 9 biological replicates across 2 independent experiments.

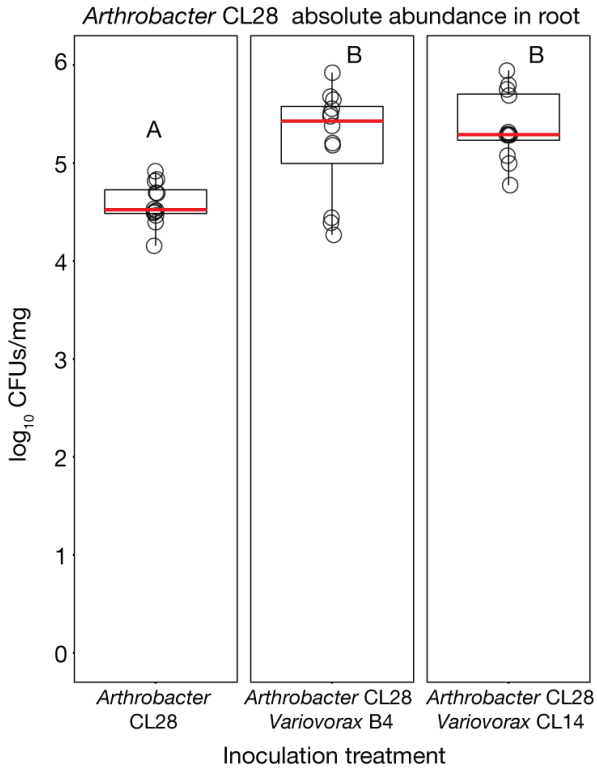


Figure 4.10 - ***Variovorax* does not compete with or antagonize RGI strains.**

To test whether *Variovorax* attenuates RGI by inhibiting the growth of RGI-inducing strains, we compared the bacterial relative abundance profiles in seedlings colonized with the full synthetic community to that of seedlings colonized with the *Variovorax* drop-out community. We found no changes in the abundances of RGI-inducing strains in response to the *Variovorax* drop-out (Figure 4.2g). In addition, we measured in planta absolute abundance of *Arthrobacter* CL28 when inoculated alone or with two *Variovorax* representatives: *Variovorax* B4 and *Variovorax* CL14. Here we show log-transformed CFUs of *Arthrobacter* CL28 normalized to root weight. To selectively count *Arthrobacter* CL28, CFUs were counted on Luria Bertani (LB) agar plates containing 50 µg ml−1 of apramycin, on which neither *Variovorax* B4 nor *Variovorax* CL14 grow. *Arthrobacter* CL28 CFUs are not reduced in the presence of *Variovorax*. Notably, Variovorax account for only about 1.5% of the root community (Figure 4.2h). These results rule out the possibility that *Variovorax* enforces stereotypic root growth by antagonizing or outcompeting RGI inducers. Significance was determined via ANOVA, letters correspond to a Tukey post hoc test. n = 12 biologically independent samples.

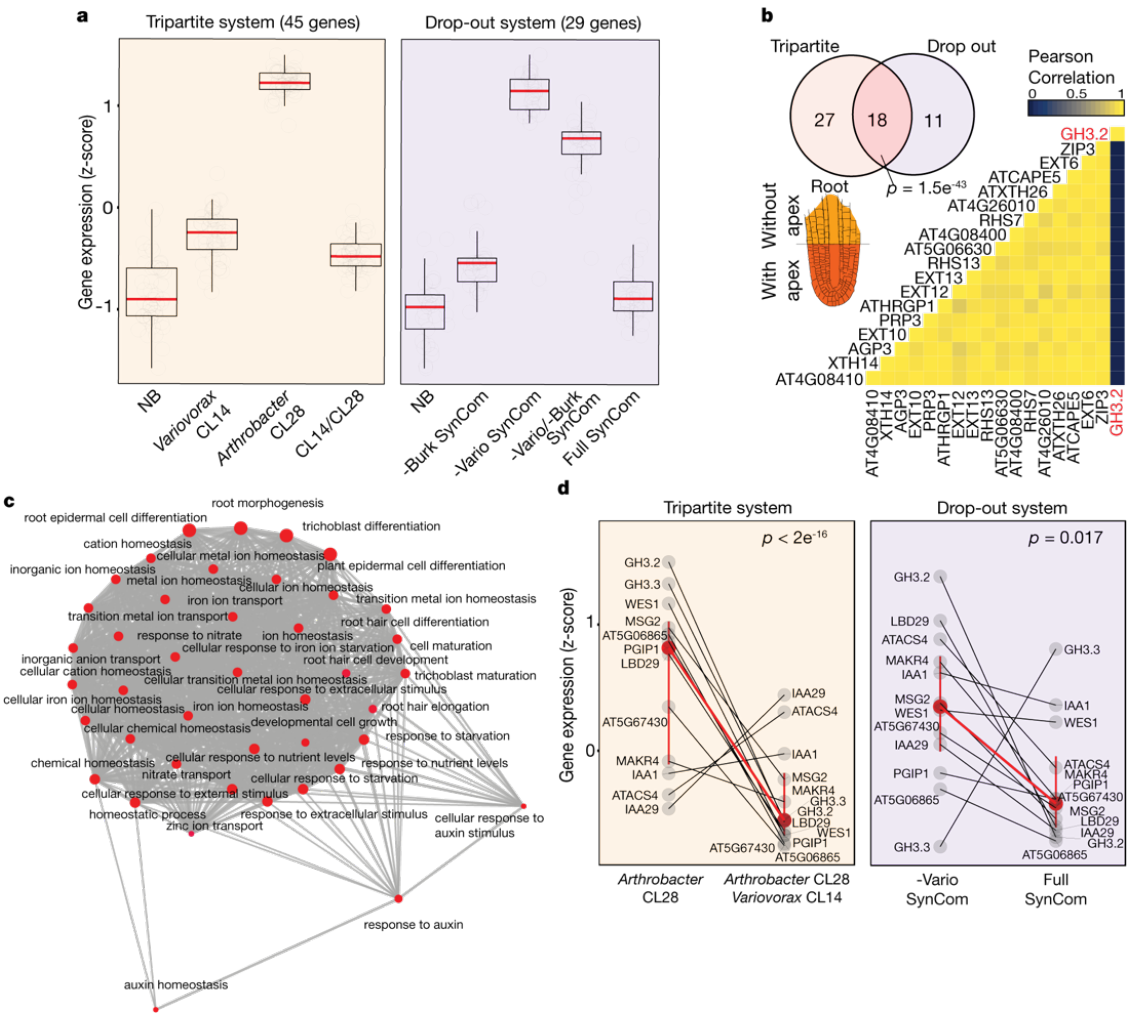


Figure 4.11 - **Auxin-responsive genes are induced in response to RGI strains.**

a) Box plots showing the average standardized expression of genes significantly induced only under RGI conditions (RGI-induced), across the following treatments. Left (tripartite system), uninoculated seedlings (NB) or seedlings inoculated with *Variovorax* CL14, *Arthrobacter* CL28 or both *Arthrobacter* CL28 and *Variovorax* CL14 (CL14/CL28). Right (drop-out system), uninoculated seedlings (NB), *Burkholderia* drop-out synthetic community (−Burk SynCom), *Variovorax* drop-out synthetic community (−Vario SynCom), *Variovorax* and *Burkholderia* drop-out synthetic community (−Vario −Burk SynCom) or the full synthetic community (full SynCom). RGI-induced genes are defined as genes that are significantly overexpressed in RGI treatments. Left, genes overexpressed in *Arthrobacter* CL28-inoculated seedling versus NB and in *Arthrobacter* CL28-inoculated seedlings versus CL14/CL28. Right, genes overexpressed in −Vario versus NB and in −Vario versus full. n = 3 (left); 5 (right). **b)** Venn diagram showing the overlap of enriched genes between the tripartite and drop-out systems. The heat map shows the pairwise correlation in expression of these 18 genes across tissues on the basis of the Klepikova Atlas (Klepikova et al., 2016). Seventeen of the 18 genes show high correlation across the atlas,

with the exception of the auxin-conjugating gene GH3.2. A root apex diagram from the Arabidopsis eFP Browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi) is shown, illustrating the spatial distribution of transcripts from the 17 highly correlated root apex-associated genes in the Klepikova Atlas (Klepikova et al., 2016). Significance of the overlap of enriched genes was determined via hypergeometric test. **c)** RGI-related genes share gene ontologies. Network of statistically significant gene ontology terms contained within the 18 overlapping RGI-induced genes (a, b). The network was computed using the emapplot function from the package clusterProfiler (G. Yu et al., 2012) in R. A P value for terms across the gene ontology was computed using a hypergeometric test (only significant ontologies are shown). Point size (gene ontology term) denotes the number of genes mapped to that particular term. **d)** Standardized expression of 12 late-responsive auxin genes (uchida et al., 2018) across the tripartite and drop-out systems. Each dot represents a gene. Identical genes are connected between bacterial treatments with a black line. Mean expression (95% confidence intervals) of the aggregated 12 genes in each treatment is highlighted in red and connected between bacterial treatments with a red line. Significance was calculated using a resampling approach, in which we compared the calculated difference between means across groups. After 10,000 resamplings, we calculated the P value by comparing the distribution of means calculated against the real observed differences between means.

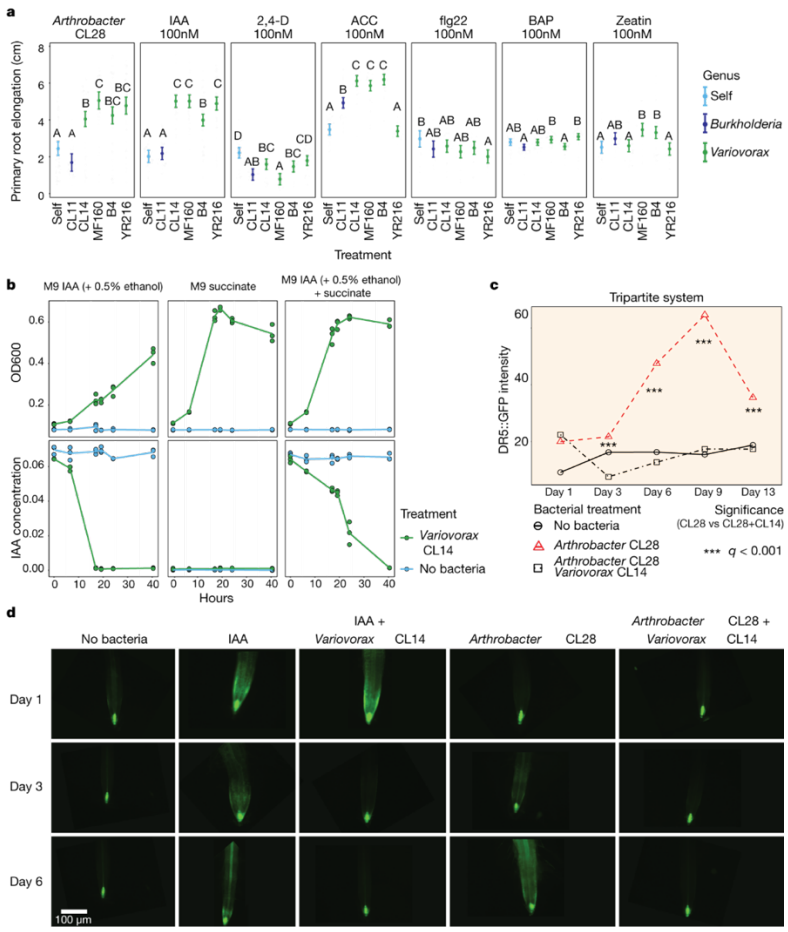


Figure 4.12 - ***Variovorax*** **degrades auxin and quenches auxin perception by the plant.**

a) Primary root elongation of seedlings grown with six hormonal or microbial-associated molecular pattern RGI treatments (panels) individually (self) or with either *Burkholderia* CL11 or each of four *Variovorax* isolates (CL14, MF160, B4 and YR216). Significance was determined via ANOVA within each panel; letters correspond to a Tukey post hoc test. n = 74, 46, 61, 48, 49, 49, 45, 44, 46, 43, 49, 40, 22, 19, 22, 19, 20, 25, 28, 30, 29, 29, 29, 29, 12, 12, 21, 20, 19, 18, 26, 30, 30, 29, 30, 30, 29, 30, 30, 26, 29 and 28 biological replicates across 2 independent experiments. **b)** *Variovorax* degrades auxin. Growth curves showing optical density at OD600 (top) and IAA concentrations (mg ml−1) (bottom) in *Variovorax* CL14 cultures grown in M9 medium with different carbon sources. Left, IAA (+ 0.5% ethanol as solvent); middle, succinate; right, succinate and IAA (+ 0.5% ethanol as solvent). **c)** *Variovorax* quenches induction of the auxin bioreporter DR5::GFP. Quantification of GFP intensity in DR5::GFP Arabidopsis seedlings grown with no bacteria, *Arthrobacter* CL28 and *Arthrobacter* CL28 + *Variovorax* CL14. GFP fluorescence was imaged 1, 3, 6, 9 and 13 days after inoculation, and quantified in the root elongation zone. Significance was determined via ANOVA within each time point, while controlling for experiment, and denoted with asterisks. n = 20 biological replicates across 2 independent experiments. **d)** Representative primary root images of DR5::GFP plants quantified in c, showing roots from 1, 3 and 6 d after inoculation. In addition to the bacterial treatments shown in c, an exogenous IAA control is shown (IAA, second column), as well as IAA-treated plants inoculated with Variovorax CL14, illustrating that IAA-induced fluorescence is quenched in the presence of Variovorax CL14 within 3 d.

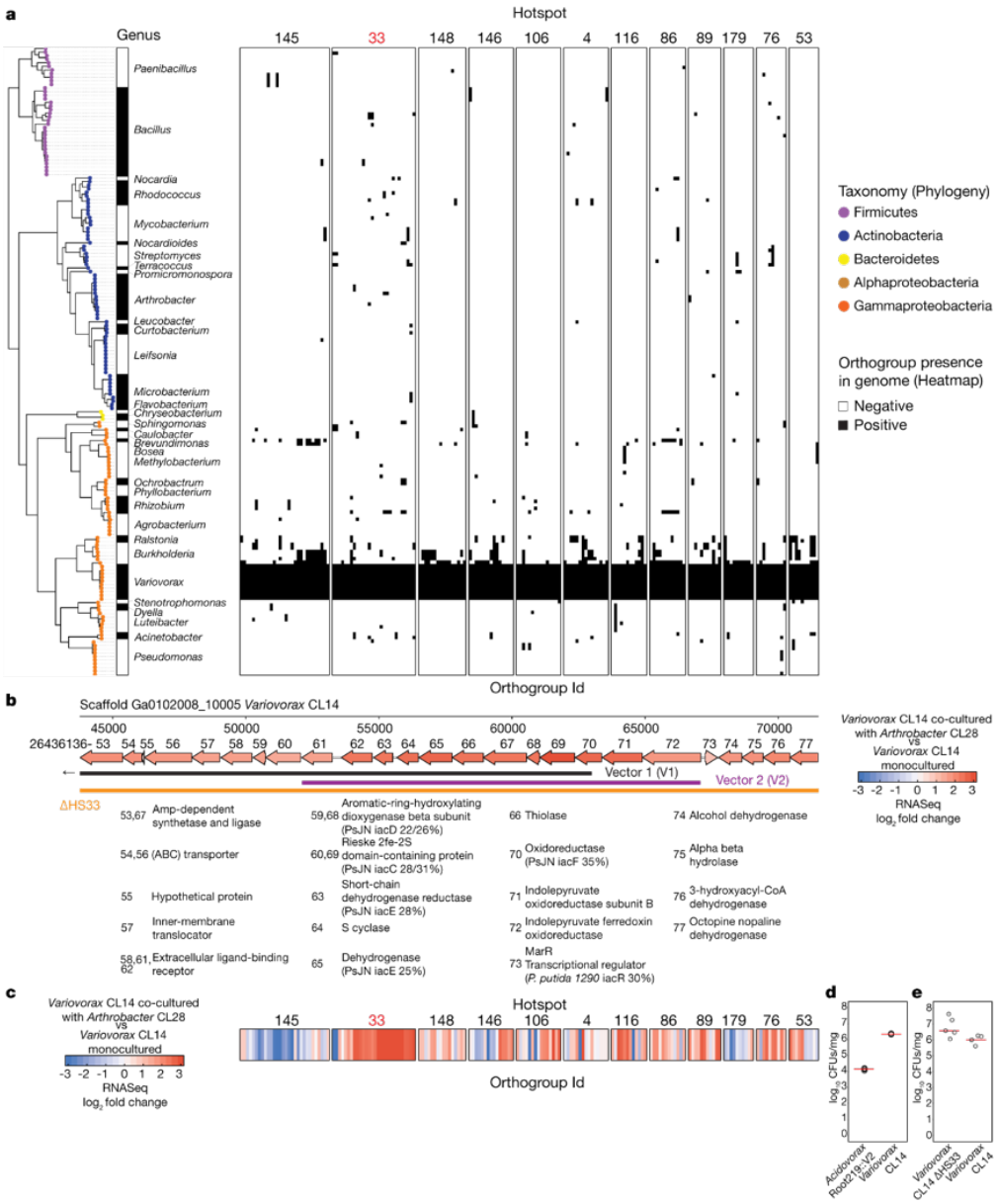


Figure 4.13 - **Detection of CL28-responsive *Variovorax*-unique operons.**

a) Presence–absence matrix denoting the distribution of 12 *Variovorax*-unique hotspots containing at least 10 genes across the 185 members of the synthetic community. Hotspots are defined using the *Variovorax* CL14 genome as a reference. Phylogeny of the 185 members of the synthetic community is shown to the left of the matrix. We determined the presence of an orthogroup based on a hidden Markov model profile scanning of each core *Variovorax* (genus) orthogroup across the 185 genomes in the synthetic community. **b)** A map of the auxin-degrading hotspot 33. Genes are annotated with the last two digits of their IMG gene identifier (26436136XX) and their functional assignments are shown below the map, including per cent identity of any to genes from a known auxin degradation locus. Genes are coloured by the log2-transformed fold change in their

transcript abundance in *Variovorax* CL14 cocultured with *Arthrobacter* CL28 versus *Variovorax* CL14 monoculture, as measured by RNA-seq (shown in c). The overlap of this region with vectors 1 and 2 and the region knocked out in *Variovorax* CL14 ΔHS33 are shown below the map. Vector 1 extends beyond this region. **c)** Results of RNA-seq on *Variovorax* CL14 transcripts. *Variovorax* CL14 was cocultured with *Arthrobacter* CL28 versus *Variovorax* CL14 monoculture. Only *Variovorax*-unique genomic hotspots are presented, aligned with the genes in a. Genes are coloured by the log2-transformed fold change in their transcript abundance in *Variovorax* CL14 cocultured with *Arthrobacter* CL28 vs *Variovorax* CL14 monoculture. Note uniform upregulation of genes in cluster 33. **d)** log-transformed CFUs of *Variovorax* CL14 or with the *Acidovorax* gain-of-function strain *Acidovorax* root219::V2 normalized to root weight. Each of these two strains was co-inoculated with *Arthrobacter* CL28 onto 7-d-old seedlings and collected after 12 days of growth. CFU counts were determined on LB plates containing 100 μg ml−1 ampicillin, for which *Arthrobacter* CL28 is susceptible and *Variovorax* CL14 and *Acidovorax* root219::V2 are naturally resistant. n = 3 biologically independent samples. Significance was determined using a two-sided Student's t-test. P = 2.57 × 10−5. Line represents median. **e)** log-transformed CFUs of *Variovorax* CL14 or of the loss-of-function strain *Variovorax* CL14 ΔHS33 normalized to root weight. Each of these two strains was inoculated individually onto 7-d-old seedlings and collected after 12 d of growth. CFU counts were determined on LB plates. n = 5 biologically independent samples. Significance was determined using a two-sided Student's t-test. P = 0.049 (mutant is slightly higher). Red lines represent median.

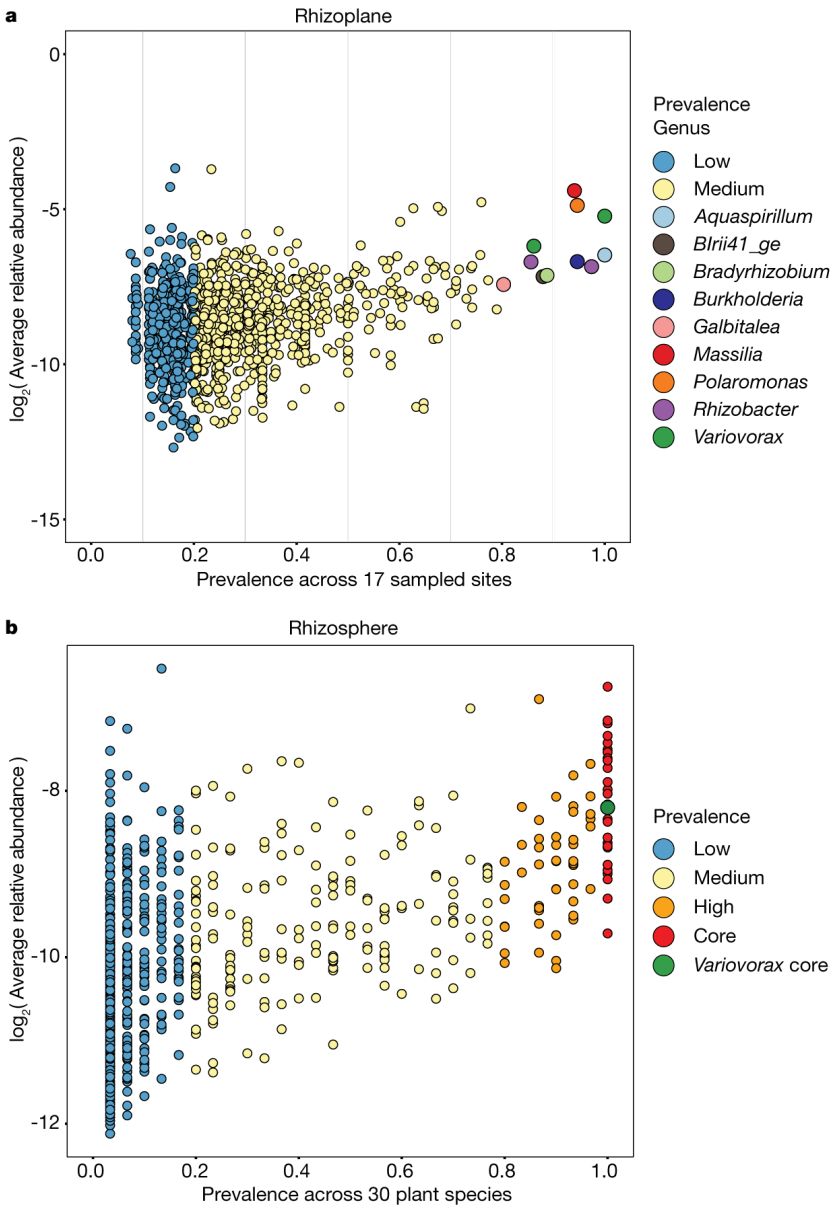


Figure 4.14 - ***Variovorax*** **are highly prevalent across naturally occurring Arabidopsis microbiomes and across 30 plant species.**

a) Correlation plot of data reanalysed from (Thiergart et al., 2020), comparing bacterial ASV prevalence to log-transformed relative abundance in *A. thaliana* rhizoplane samples taken across 3 years in 17 sites in Europe. **b)** Correlation plot of data reanalysed from (Fitzpatrick et al., 2018), comparing bacterial ASV prevalence to log-transformed relative abundance across 30 phylogenetically diverse plant species grown in a common garden experiment.

4.6 Methods

4.6.1 *Arabidopsis* with bacterial synthetic community microcosm across four stress gradients

4.6.1.1 Bacterial culture and plant inoculation

The 185-member bacterial synthetic community used here contains genome-sequenced isolates obtained from surface-sterilized Brassicaceae roots, nearly all *Arabidopsis thaliana*, planted in two soils from North Carolina (USA). A detailed description of this collection and isolation procedures can be found in (Levy et al., 2018). One week before each experiment, bacteria were inoculated from glycerol stocks into 600 µl KB medium in a 96-deep-well plate. Bacterial cultures were grown at 28 °C, shaking at 250 rpm. After 5 days of growth, cultures were inoculated into fresh medium and returned to the incubator for an additional 48 h, resulting in 2 copies of each culture (7 days old and 48 h old). We adopted this procedure to account for variable growth rates of different members of the synthetic community and to ensure that nonstationary cells from each strain were included in the inoculum. After growth, 48-h and 7-day plates were combined and optical density of cultures was measured at 600 nm (OD600) using an Infinite M200 Pro plate reader (TECAN). All cultures were then pooled while normalizing the volume of each culture to OD600 = 1. The mixed culture was washed twice with 10 mM MgCl2 to remove spent medium and cell debris, and vortexed vigorously with sterile glass beads to break up aggregates. OD600 of the mixed, washed culture was then measured and normalized to OD600 = 0.2. The synthetic community inoculum (100 µl) was spread on 12 × 12-cm vertical square agar plates with amended Johnson medium (JM)1 without sucrose before transferring seedlings.

4.6.1.2 *In vitro* plant growth conditions

All seeds were surface-sterilized with 70% bleach, 0.2% Tween-20 for 8 min, and rinsed 3 times with sterile distilled water to eliminate any seed-borne microorganisms on the seed surface. Seeds were stratified at 4 °C in the dark for 2 days. Plants were germinated on vertical square 12 × 12-cm agar plates with JM containing 0.5% sucrose, for 7 days. Then, 10 plants were transferred to each of the agar plates inoculated with the synthetic community. The composition of JM in the agar plates was amended to produce environmental variation. We added to the previously reported phosphate concentration gradient (0, 10, 30, 50, 100 and 1,000 μM Pi) (Finkel et al., 2019) three additional environmental gradients: salinity (50, 100, 150 and 200 mM NaCl), pH (5.5, 7.0 and 8.2) and incubation temperature (10, 21 and 31 °C). Each gradient was tested separately, in two independent replicas. Each condition included three synthetic community + plant samples, two no-plant controls and one no-bacteria control. Thus, the total sample size for each condition was n = 6. Previous publications (Castrillo et al., 2017; Finkel et al., 2019; Herrera Paredes et al., 2018) have shown that an $n \geq 5$ is provides sufficient power for synthetic community profiling. Plates were placed in randomized order in growth chambers and grown under a 16-h dark/8-h light regime at 21-°C day/18-°C night for 12 days. Upon collection, DNA was extracted from roots, shoots and the agar substrate. Here and hereafter, all measurements were taken from distinct samples.

4.6.1.3 DNA extraction

Roots, shoots and agar were collected separately, pooling 6–8 plants for each sample. Roots and shoots were placed in 2-ml Eppendorf tubes with 3 sterile glass beads. These samples were washed three times with sterile distilled water to remove agar particles and weakly associated microorganisms. Tubes containing the samples were stored at −80 °C until processing. Root and shoot samples were lyophilized for 48 h using a Labconco freeze-dry system and pulverized using

a tissue homogenizer (MPBio). Agar from each plate was collected in 30-ml syringes with a square of sterilized Miracloth (Millipore) at the bottom and kept at −20 °C for 1 week. Syringes were then thawed at room temperature and samples were squeezed gently through the Miracloth into 50-ml falcon tubes. Samples were centrifuged at maximum speed for 20 min and most of the supernatant was discarded. The remaining 1–2 ml of supernatant, containing the pellet, was transferred into clean 1.5-ml Eppendorf tubes. Samples were centrifuged again, supernatant was removed and pellets were stored at −80 °C until DNA extraction. DNA extractions were carried out on ground root and shoot tissue and agar pellets using 96-well-format MoBio PowerSoil Kit (MOBIO Laboratories; Qiagen) following the manufacturer's instruction. Sample position in the DNA extraction plates was randomized, and this randomized distribution was maintained throughout library preparation and sequencing.

4.6.1.4 Bacterial 16S rRNA sequencing

We amplified the V3–V4 regions of the bacterial 16S rRNA gene using the primers 338F (5′-ACTCCTACGGGAGGCAGCA-3′) and 806R (5′-GGACTACHVGGGTWTCTAAT-3′). Two barcodes and six frameshifts were added to the 5′ end of 338F and six frameshifts were added to the 806R primers, on the basis of a previously published protocol36. Each PCR reaction was performed in triplicate, and included a unique mixture of three frameshifted primer combinations for each plate. PCR conditions were as follows: 5 µl Kapa Enhancer, 5 µl Kapa Buffer A, 1.25 µl 5 µM 338F, 1.25 µl 5 µM 806R, 0.375 µl mixed plant rRNA gene-blocking peptide nucleic acids (PNAs) (1:1 mix of 100 µM plastid PNA and 100 µM mitochondrial PNA36), 0.5 µl Kapa dNTPs, 0.2 µl Kapa Robust Taq, 8 µl dH2O, 5 µl DNA; temperature cycling: 95 °C for 60 s; 24 cycles of 95 °C for 15 s; 78 °C (PNA) for 10 s; 50 °C for 30 s; 72 °C for 30 s; 4 °C until use. Following PCR clean up, using AMPure beads (Beckman Coulter), the PCR product was indexed using 96 indexed

806R primers with the Kapa HiFi Hotstart readymix with the same primers as above; temperature cycling: 95 °C for 60 s; 9 cycles of 95 °C for 15 s; 78 °C (PNA) for 10 s; 60 °C for 30 s; 72 °C for 35 s; 4 °C until use. PCR products were purified using AMPure XP magnetic beads (Beckman Coulter) and quantified with a Qubit 2.0 fluorometer (Invitrogen). Amplicons were pooled in equal amounts and then diluted to 10 pM for sequencing. Sequencing was performed on an Illumina MiSeq instrument using a 600-cycle V3 chemistry kit. DNA sequence data for this experiment are available at the NCBI Bioproject repository (accession PRJNA543313). The abundance matrix, metadata and taxonomy are available at https://github.com/isaisg/variovoraxRGI.

4.6.1.5 16S rRNA amplicon sequence data processing

Synthetic community sequencing data were processed with MT-Toolbox (Yourstone et al., 2014). Usable read output from MT-Toolbox (that is, reads with 100% correct primer and primer sequences that successfully merged with their pair) were quality-filtered using Sickle by not allowing any window with Q score under 20. The resulting sequences were globally aligned to a reference set of 16S rDNA sequences extracted from genome assemblies of members of the synthetic community. For strains that did not have an intact 16S rDNA sequence in their assembly, we sequenced the 16S rRNA gene using Sanger sequencing. The reference database also included sequences from known bacterial contaminants and Arabidopsis organellar sequences. Sequence alignment was performed with USEARCH v.7.109039 (Edgar, 2010) with the option usearch_global at a 98% identity threshold. On average, 85% of sequences matched an expected isolate. Our 185 isolates could not all be distinguished from each other based on the V3–V4 sequence and were thus classified into 97 unique sequences. A unique sequence encompasses a set of identical (clustered at 100%) V3–V4 sequences coming from a single or multiple isolates.

Sequence mapping results were used to produce an isolate abundance table. The remaining unmapped sequences were clustered into operational taxonomic units (OTUs) using UPARSE (Edgar, 2013) implemented with USEARCH v.7.1090, at 97% identity. Representative OTU sequences were taxonomically annotated with the RDP classifier (Q. Wang et al., 2007) trained on the Greengenes database (DeSantis et al., 2006) (4 February 2011). Matches to Arabidopsis organelles were discarded. The vast majority of the remaining unassigned OTUs belonged to the same families as isolates in the synthetic community. We combined the assigned unique sequence and unassigned OTU count tables into a single count table. In addition to the raw count table, we created rarefied (1,000 reads per sample) and relative abundance versions of the abundance matrix for further analyses.

The resulting abundance tables were processed and analysed with functions from the ohchibi package (https://github.com/isaisg/ohchibi). An α-diversity metric (Shannon diversity) was calculated using the diversity function from the vegan package v.2.5-343 (Dixon, 2003). We used ANOVA to test for differences in α-diversity between groups. β-Diversity analyses (principal coordinate analysis and canonical analysis of principal coordinates (CAP)) were based on Bray–Curtis dissimilarity calculated from the relative abundance matrices. We used the capscale function from the vegan R package v.2.5-3 (Dixon, 2003) to compute the CAP. To analyze the full dataset (all fractions and all abiotic treatments), we constrained by fraction and abiotic treatment while conditioning for the replica and experiment effect. We explored the abiotic conditions effect inside each of the four abiotic gradients tested (phosphate, salinity, pH and temperature). We performed the fraction–abiotic interaction analysis within each fraction independently, constraining for the abiotic conditions while conditioning for the replica effect. In addition to CAP, we performed PERMANOVA using the adonis function from the vegan package v.2.5-3 (Dixon, 2003). We used

the package DESeq2 v.1.22.1 (Love et al., 2014) to compute the enrichment profiles for unique sequences present in the count table.

We estimated the fraction effect across all the abiotic conditions tested by creating a group variable that merged the fraction variable and the abiotic condition variable together (for example, Root_10Pi, Substrate_10Pi). We fitted the following model specification using this group variable:

$$\text{Abundance} \sim \text{rep} + \text{experiment} + \text{group}$$

From the fitted model, we extracted—for all levels within the group variables—the following comparisons: substrate versus root and substrate versus shoot. A unique sequence was considered statistically significant if it had a FDR-adjusted P value < 0.05. All scripts and dataset objects necessary to reproduce the synthetic community analyses are deposited in the following GitHub repository: https://github.com/isaisg/variovoraxRGI.

4.6.1.6 Co-occurrence analysis

The relative abundance matrix (unique sequences × samples) was standardized across the unique sequences by dividing the abundance of each unique sequence in its sample over the mean abundance of that unique sequence across all samples. Subsequently, we created a dissimilarity matrix based on the Pearson correlation coefficient between all the pairs of strains in the transformed abundance matrix, using the cor function in the stats base package in R. Finally, hierarchichal clustering (method ward.D2, function hclust) was applied over the dissimilarity matrix constructed above.

4.6.1.7 Heat map and family enrichment analysis

We visualized the results of the generalized linear model (GLM) testing the fraction effects across each specific abiotic condition tested using a heat map. The rows in the heat map were ordered according to the dendrogram order obtained from the unique sequences co-occurrence analysis. The heat map was coloured on the basis of the log2-transformed fold change output by the GLM model. We highlighted in a black shade the comparisons that were significant (q value < 0.05). Finally, for each of the four modules, we computed for each family present in that module a hypergeometric test testing if that family was overrepresented (enriched) in that particular module. Families with an FDR-adjusted P value < 0.1 are visualized in the figure.

4.6.2 Deconstructing the synthetic community to four modules of co-occurring strains

4.6.2.1 Bacterial culture and plant-inoculation

Strains belonging to each module A, B, C and D ('Co-occurrence analysis' in '*Arabidopsis* with bacterial synthetic community microcosm across four stress gradients') were grown in separate deep 96-well plates and mixed as described in 'Bacterial culture and plant-inoculation' in '*Arabidopsis* with bacterial synthetic community microcosm across four stress gradients'. The concentration of each module was adjusted to OD600 = 0.05 (1/4 of the concentration of the full synthetic community). Each module was spread on the plates either separately, or in combination with another module at a total volume of 100 µl. In addition, we included a full synthetic community control and an uninoculated control, bringing the number of synthetic community combinations to 12. We performed the experiment in two independent replicates and each replicate included five plates per synthetic community combination.

4.6.2.2 *In vitro* plant growth conditions

Seed sterilization and germination conditions were the same as in 'In vitro plant growth conditions' in '*Arabidopsis* with bacterial synthetic community microcosm across four stress gradients'. Plants were transferred to each of the synthetic-community-inoculated agar plates containing JM without sucrose. Plates were placed in randomized order in growth chambers and grown under a 16-h dark/8-h light regime at 21-°C day/18-°C night for 12 days. Upon collection, root morphology was measured.

4.6.2.3 Root and shoot image analysis

Plates were imaged 12 days after transfer, using a document scanner. Primary root length elongation was measured using ImageJ (Schindelin et al., 2012) and shoot area and total root network were measured with WinRhizo software (Regent Instruments).

4.6.2.4 Primary root elongation analyses

Primary root elongation was compared across the no bacteria, full synthetic community, single modules and pairs of modules treatments jointly using a two-sided ANOVA model controlling for the replicate effect. We inspected the normality assumptions (here and elsewhere) using qqplots and Shapiro tests. Differences between treatments were indicated using the confidence letter display derived from the Tukey's post hoc test implemented in the package emmeans (Searle et al., 1980).

4.6.3 Inoculating plants with all synthetic community isolates separately

4.6.3.1 Bacterial culture and plant inoculation

Cultures from each strain in the synthetic community were grown in KB medium and washed separately ('Bacterial culture and plant inoculation' in '*Arabidopsis* with bacterial

synthetic community microcosm across four stress gradients'), and OD600 was adjusted to 0.01 before spreading 100 µl on plates. We performed the experiment in two independent replicates and each replicate included one plate per each of the 185 strains. In vitro growth conditions were the same as described in 'In vitro plant growth conditions' in 'Deconstructing the synthetic community to four modules of co-occurring strains'. Upon collection, root morphology was measured ('Root and shoot image analysis' in 'Deconstructing the synthetic community to four modules of co-occurring strains'). Isolates generating an average main root elongation of <3 cm were classified as RGI-inducing strains.

4.6.4 Tripartite plant–microorganism–microorganism experiments

4.6.4.1 Experimental design

To identify strains that revert RGI (Figure 4.1d), we selected all 18 non-RGI-inducing strains in module A and co-inoculated them with each of four RGI-inducing strains, one from each module. The experiment also included uninoculated controls and controls consisting of each of the 22 strains inoculated alone, amounting to 95 separate bacterial combinations.

To confirm the ability of *Variovorax* and *Burkholderia* to attenuate RGI induced by diverse bacteria (Figure 4.7a), three RGI-suppressing strains were co-inoculated with a selection of 18 RGI-inducing strains. The experiment also included uninoculated controls and controls consisting of each of the 21 strains inoculated alone. Thus, the experiment consisted of 76 separate bacterial combinations. We performed each of these two experiments in two independent replicates and each replicate included one plate per each of the strain combinations.

4.6.4.2 Bacterial culture and plant-inoculation

All strains were streaked on agar plates, then transferred to 4 ml liquid KB medium for over-night growth. Cultures were then washed, and OD600 was adjusted to 0.02 before mixing and spreading 100 µl on each plate. Upon collection, root morphology was measured ('Root and shoot image analysis' in 'Deconstructing the synthetic community to four modules of co-occurring strains') and plant RNA was collected and processed from uninoculated samples, and from samples with *Variovorax* CL14, the RGI-inducing strain *Arthrobacter* CL28 and the combination of both ('Experimental design' in 'Tripartite plant–microorganism–microorganism experiments').

4.6.4.3 Primary root elongation analysis

We fitted ANOVA models for each RGI-inducing strain we tested. Each model compared the primary root elongation with the RGI-inducing strains alone against root elongation when the RGI-inducing strain was co-inoculated with other isolates. The P values for all the comparisons were corrected for multiple testing using FDR.

4.6.4.3 RNA extraction

RNA was extracted from *A. thaliana* seedlings following previously published methods (Logemann et al., 1987). Four seedlings were pooled from each plate and 3–5 samples per treatment were flash frozen and stored at −80 °C until processing. Frozen seedlings were ground using a TissueLyzer II (Qiagen), then homogenized in a buffer containing 400 µl of Z6-buffer; 8 M guanidine HCl, 20 mM MES, 20 mM EDTA at pH 7.0. Four hundred µl phenol:chloroform:isoamylalcohol, 25:24:1 was added, and samples were vortexed and centrifuged (20,000g, 10 min) for phase separation. The aqueous phase was transferred to a new 1.5-ml Eppendorf tube and 0.05 volumes of 1 N acetic acid and 0.7 volumes 96% ethanol were

added. The RNA was precipitated at −20 °C overnight. Following centrifugation (20,000g, 10 min, 4 °C), the pellet was washed with 200 µl sodium acetate (pH 5.2) and 70% ethanol. The RNA was dried and dissolved in 30 µl of ultrapure water and stored at −80 °C until use.

4.6.4.4 Plant RNA sequencing

Illumina-based mRNA-seq libraries were prepared from 1 µg RNA following previously published methods (Herrera Paredes et al., 2018). mRNA was purified from total RNA using Sera-mag oligo(dT) magnetic beads (GE Healthcare Life Sciences) and then fragmented in the presence of divalent cations (Mg2+) at 94 °C for 6 min. The resulting fragmented mRNA was used for first-strand cDNA synthesis using random hexamers and reverse transcriptase, followed by second-strand cDNA synthesis using DNA polymerase I and RNaseH. Double-stranded cDNA was end-repaired using T4 DNA polymerase, T4 polynucleotide kinase and Klenow polymerase. The DNA fragments were then adenylated using Klenow exo-polymerase to allow the ligation of Illumina Truseq HT adapters (D501–D508 and D701–D712). All enzymes were purchased from Enzymatics. Following library preparation, quality control and quantification were performed using a 2100 Bioanalyzer instrument (Agilent) and the Quant-iT PicoGreen dsDNA Reagent (Invitrogen), respectively. Libraries were sequenced using Illumina HiSeq4000 sequencers to generate 50-bp single-end reads.

4.6.4.5 RNA-seq read processing

Initial quality assessment of the Illumina RNA-seq reads was performed using FastQC v.0.11.7. Trimmomatic v.0.36 (Bolger et al., 2014) was used to identify and discard reads containing the Illumina adaptor sequence. The resulting high-quality reads were then mapped against the TAIR10 Arabidopsis reference genome using HISAT2 v.2.1.0 (Kim et al., 2015) with

default parameters. The featureCounts function from the Subread package (Liao et al., 2013) was then used to count reads that mapped to each one of the 27,206 nuclear protein-coding genes. Evaluation of the results of each step of the analysis was performed using MultiQC v.1.1 (Ewels et al., 2016). Raw sequencing data and read counts are available at the NCBI Gene Expression Omnibus accession number GSE131158.

4.6.5 *Variovorax* drop-out experiment

4.6.5.1 Bacterial culture and plant-inoculation

The entire synthetic community, excluding all 10 *Variovorax* isolates and all 5 *Burkholderia* isolates, was grown and prepared as described in 'Bacterial culture and plant inoculation' in '*Arabidopsis* with bacterial synthetic community microcosm across four stress gradients'. The *Variovorax* and *Burkholderia* isolates were grown in separate tubes, washed and added to the rest of the synthetic community to a final OD600 of 0.001 (the calculated OD600 of each individual strain in a 185-member synthetic community at a total of OD600 of 0.2), to form the following five mixtures: (i) full community: all *Variovorax* and *Burkholderia* isolates added to the synthetic community; (ii) *Burkholderia* drop-out: only *Variovorax* isolates added to the synthetic community; (iii) *Variovorax* drop-out: only *Burkholderia* isolates added to the synthetic community; (iv) *Variovorax* and *Burkholderia* drop-out: no isolates added to the synthetic community; (v) uninoculated plants: no synthetic community. The experiment consisted of six plates per synthetic community mixture, amounting to 30 plates. Upon collection, root morphology was measured and analysed ('Root and shoot image analysis' in 'Deconstructing the synthetic community to four modules of co-occurring strains', and in 'Primary root elongation analysis' in 'Tripartite plant–microorganism–microorganism experiments'); and bacterial DNA ('DNA extraction' and 'Bacterial 16S rRNA sequencing' in '*Arabidopsis* with bacterial synthetic

community microcosm across four stress gradients') and plant RNA ('RNA extraction' and 'Plant RNA sequencing' in 'Tripartite plant–microorganism–microorganism experiments') were collected and processed.

4.6.6 *Variovorax* drop-out under varying abiotic contexts

4.6.6.1 Bacterial culture and plant-inoculation

The composition of JM in the agar plates was amended to produce abiotic environmental variation. These amendments included salt stress (150 mM NaCl), low phosphate (10 μM phosphate), high pH (pH 8.2) and high temperature (plates incubated at 31 °C), as well as an unamended JM control. Additionally, we tested a different medium (1/2-strength Murashige and Skoog (MS)) and a soil-like substrate. As a soil-like substrate, we used calcined clay (Diamond Pro), prepared as follows: 100 ml of clay was placed in Magenta GA7 jars. The jars were then autoclaved twice. Forty ml of liquid JM was added to the Magenta jars, with the corresponding bacterial mixture spiked into the media at a final OD600 of 5 × 10−4. Four 1-week old seedlings were transferred to each vessel, and vessels were covered with Breath-Easy gas permeable sealing membrane (Research Products International) to maintain sterility and gas exchange.

The entire synthetic community, excluding all 10 *Variovorax* isolates, was grown and prepared as described in 'Bacterial culture and plant inoculation' in '*Arabidopsis* with bacterial synthetic community microcosm across four stress gradients'. The *Variovorax* isolates were grown in separate tubes, washed and added to the rest of the synthetic community to a final OD600 of 0.001 (the calculated OD600 of each individual strain in a 185-member synthetic community at an OD600 of 0.2), to form the following five mixtures: (i) full community: all *Variovorax* isolates

added to the synthetic community; (ii) *Variovorax* drop-out: no isolates added to the synthetic community; (iii) uninoculated plants: no synthetic community.

We inoculated all 3 synthetic community combinations in all 7 abiotic treatments, amounting to 21 experimental conditions. We performed the experiment in 2 independent replicates and each replicate included 3 plates per experimental conditions, amounting to 63 plates per replicate. Upon collection, root morphology was measured ('Root and shoot image analysis' in 'Deconstructing the synthetic community to four modules of co-occurring strains'); and Bacterial DNA ('DNA extraction, 'Bacterial 16S rRNA sequencing' and '16S rRNA amplicon sequence data processing' in 'Arabidopsis with bacterial synthetic community microcosm across four stress gradients') and plant RNA ('RNA extraction', 'Plant RNA sequencing' and 'RNA-seq read processing' in 'Tripartite plant–microorganism–microorganism experiments') were collected and processed.

4.6.6.2 Root image analysis

For agar plates, roots were imaged as described in 'Root and shoot image analysis' in 'Deconstructing the synthetic community to four modules of co-occurring strains'. For calcined clay pots, four weeks after transferring, pots were inverted, and whole root systems were gently separated from the clay by washing with water. Root systems were spread over an empty Petri dish and scanned using a document scanner.

4.6.6.3 Primary root elongation and total root network analysis

Primary root elongation was compared between synthetic-community treatments within each of the different abiotic contexts tested independently. Differences between treatments were

indicated using the confidence letter display derived from the Tukey's post hoc test implemented in the package emmeans.

4.6.6.4 Bacterial 16S rRNA data analysis

To be able to compare shifts in the community composition of samples treated with and without the *Variovorax* genus, we in silico-removed the 10 *Variovorax* isolates from the count table of samples inoculated with the full community treatment. We then merged this count table with the count table constructed from samples inoculated without the *Variovorax* genus (*Variovorax* drop-out treatment). Then, we calculated a relative abundance of each unique sequence across all the samples using the merged count matrix. Finally, we applied CAP over the merged relative abundance matrix to control for the replica effect. In addition, we used the function adonis from the vegan (Dixon, 2003) R package to compute a PERMANOVA test over the merged relative abundance matrix and we fitted a model evaluating the fraction and synthetic community (presence of *Variovorax*) effects over the assembly of the community.

4.6.7 *Variovorax* drop-out under varying biotic contexts

4.6.7.1 Bacterial culture and plant-inoculation

Strains belonging to modules A (excluding *Variovorax*), C and D were grown in separate wells in deep 96-well plates and mixed as described in 'Bacterial culture and plant inoculation' in '*Arabidopsis* with bacterial synthetic community microcosm across four stress gradients'. The concentration of each module was adjusted to OD600 = 0.05 (1/4 of the concentration of the full synthetic community). The *Variovorax* isolates were grown in separate tubes, washed and added to the rest of the synthetic community to a final OD600 of 0.001.

In a separate experiment, the 35-member synthetic community used in (Castrillo et al., 2017) was grown, excluding *Variovorax* CL14, to create a taxonomically diverse, *Variovorax*-free subset of the full 185-member community. The concentration of this synthetic community was adjusted to OD600 = 0.05. The *Variovorax* isolates were grown in separate tubes, washed and added to the rest of the synthetic community to a final OD600 of 0.001.

These two experiments included the following mixtures (Figure 4.8b): (i) module A excluding *Variovorax*; (ii) module C; (iii) module D; (iv) module A including *Variovorax*; (v) module C + all 10 *Variovorax*; (vi) module D + all 10 *Variovorax*; (vii) 35-member synthetic community excluding the one *Variovorax* found therein; (viii) 34-member synthetic community + all 10 Variovorax; (ix) uninoculated control. The experiment with modules A, C and D was performed in two independent experiments, with two plates per treatment in each. The experiment with the 34-member synthetic community was performed once, with 5 plates per treatment. Upon collection, root morphology was measured ('Root and shoot image analysis' in 'Deconstructing the synthetic community to four modules of co-occurring strains').

4.6.7.2 Primary root elongation analysis

We directly compared differences between the full synthetic community and *Variovorax* drop-out treatment using a t-test and adjusting the P values for multiple testing using FDR.

4.6.8 Phylogenetic inference of the synthetic community and *Variovorax* isolates

To build the phylogenetic tree of the synthetic community isolates, we used the previously described super matrix approach (Levy et al., 2018). We scanned 120 previously defined marker genes across the 185 isolate genomes from the synthetic community using the hmmsearch tool from the hmmer v.3.1b2 (Eddy, 2011). Then, we selected 47 markers that were present as single-

copy genes in 100% of our isolates. Next, we aligned each individual marker using MAFFT (Katoh et al., 2002) and filtered low-quality columns in the alignment using trimAl (Capella-Gutiérrez et al., 2009). Then, we concatenated all filtered alignments into a super alignment. Finally, FastTree v.2.1 (M. N. Price et al., 2010) was used to infer the phylogeny using the WAG model of evolution. For the tree of the relative of *Variovorax*, we chose 56 markers present as single copy across 124 Burkholderiales isolates and implemented the same methodology described above.

4.6.9 Measuring how prevalent the RGI suppression trait is across the *Variovorax* phylogeny

4.6.9.1 Bacterial culture and plant-inoculation

Fifteen *Variovorax* strains from across the phylogeny of the genus were each co-inoculated with the RGI-inducing *Arthrobacter* CL28. All 16 strains were grown in separate tubes, then washed and OD600 was adjusted to 0.01 before mixing. Pairs of strains were mixed in 1:1 ratios and spread at a total volume of 100 µl onto agar before seedling transfer. The experiment also included uninoculated controls and controls consisting of each of the 16 strains inoculated alone. Thus, the experiment consisted of 32 separate bacterial combinations. We performed the experiment one time, which included three plates per bacterial combination. Upon collection, primary root elongation was analysed as described in ‘‘Root and shoot image analysis’ in ‘Deconstructing the synthetic community to four modules of co-occurring strains’.

4.6.10 Measuring RGI in tomato seedlings

4.6.10.1 Experimental design

This experiment included the following treatments: (i) no bacteria, (ii) *Arthrobacter* CL28, (iii) *Variovorax* CL14 and (iv) *Arthrobacter* CL28 + *Variovorax* CL14. Each treatment was

repeated in three separate agar plates with five tomato seedlings per plate. The experiment was repeated in two independent replicates.

4.6.10.2 Bacterial culture and plant inoculation

All strains were grown in separate tubes, then washed and OD600 was adjusted to 0.01 before mixing and spreading ('Bacterial culture and plant inoculation' in 'Tripartite plant–microorganism–microorganism experiments'). Four hundred µl of each bacterial treatment was spread on 20 × 20 agar plates containing JM agar with no sucrose.

4.6.10.3 *In vitro* plant growth conditions

We used tomato cultivar Heinz 1706 seeds. All seeds were soaked in sterile distilled water for 15 min, then surface-sterilized with 70% bleach, 0.2% Tween-20 for 15 min, and rinsed 5 times with sterile distilled water to eliminate any seed-borne microorganisms on the seed surface. Seeds were stratified at 4 °C in the dark for 2 days. Plants were germinated on vertical square 10 × 10 cm agar plates with JM containing 0.5% sucrose, for 7 days. Then, 5 plants were transferred to each of the synthetic-community-inoculated agar plates. Upon collection, root morphology was measured ('Root and shoot image analysis' in 'Deconstructing the synthetic community to four modules of co-occurring strains').

4.6.10.4 Primary root elongation analysis

Differences between treatments were indicated using the confidence letter display derived from the Tukey's post hoc test from an ANOVA model.

4.6.11 Determination of *Arthrobacter* CL28 colony forming units from roots

Arabidopsis seedlings were inoculated with (i) *Arthrobacter* CL28 alone, (ii) *Arthrobacter* CL28 + *Variovorax* CL14 or (iii) *Arthrobacter* CL28 + *Variovorax* B4, as described in 'Bacterial culture and plant inoculation' in 'Tripartite plant–microorganism–microorganism experiments'. Each bacterial treatment included four separate plates, with nine seedlings in each plate. Upon collection, all seedlings were placed in pre-weighed 2-ml Eppendorf tubes containing 3 glass beads, 3 seedlings per tube (producing 12 data points per treatment). Roots were weighed, and then homogenized using a bead beater (MP Biomedicals). The resulting suspension was serially diluted, then plated on LB agar plates containing 50 μg/ml of apramycin to select for *Arthrobacter* CL28 colonies and colonies were counting after incubation of 48 h at 28 °C.

4.6.12 *Arabidopsis* RNA-seq analysis

4.6.12.1 Detection of RGI-induced genes

To measure the transcriptional response of the plant to the different synthetic community combinations, we used the R package DESeq2 v.1.22.1 (Love et al., 2014). The raw count genes matrices for the drop-out and tripartite experiments were used independently to define differentially expressed genes (DEGs). For the analysis of both experiments we fitted the following model specification:

$$\text{Abundance gene} \sim \text{synthetic community}.$$

From the fitted models, we derived the following contrasts to obtain DEGs. A gene was considered differentially expressed if it had a q-value < 0.1. For the tripartite system ('Tripartite plant–microorganism–microorganism experiments'), we performed the following contrasts: *Arthrobacter* CL28 versus no bacteria (NB) and *Arthrobacter* CL28 versus *Arthrobacter* CL28

co-inoculated with *Variovorax* CL14. The logic behind these two contrasts was to identify genes that were induced in RGI plants (*Arthrobacter* CL28 versus NB) and repressed by *Variovorax* CL14. For the drop-out system ('Variovorax drop-out experiment'), we performed the following contrasts, *Variovorax* drop-out versus NB, and *Variovorax* drop-out versus full synthetic community. The logic behind these two contrasts was identical to the tripartite system: to identify genes that are associated with the RGI phenotype (*Variovorax* drop-out versus NB contrast) and repressed when *Variovorax* are present (*Variovorax* drop-out versus full synthetic community contrast).

For visualization purposes, we applied a variance stabilizing transformation (DESeq2) to the raw count gene matrix. We then standardized each gene expression (z-score) along the samples. We subset DEGs from this standardized matrix and calculated the mean z-score expression value for each synthetic community treatment.

To identify the tissue-specific expression profile of the 18 intersecting genes between the tripartite and drop-out systems, we downloaded the spatial expression profile of each gene from the Klepikova atlas (Klepikova et al., 2016) using the bio-analytic resource of plant biology platform. Then, we constructed a spatial expression matrix of the 18 genes and computed pairwise Pearson correlation between all pairs of genes. Finally, we applied hierarchical clustering to this correlation matrix.

4.6.12. Comparison with acute auxin response dataset

We applied the variance stabilizing transformation (DESeq2) to the raw count gene matrix. We then standardized each gene expression (z-score) along the samples. From this matrix, we subset 12 genes that in a previous study (uchida et al., 2018) exhibited the highest fold change

between auxin-treated and untreated samples. Finally, we calculated the mean z-score expression value of each of these 12 genes across the synthetic community treatments. We estimated the statistical significance of the trend of these 12 genes between a pair of synthetic community treatments (full synthetic community versus *Variovorax* drop-out, *Arthrobacter* CL28 versus *Arthrobacter* CL28 plus *Variovorax* CL14) using a permutation approach: we estimated a P value by randomly selecting 12 genes 10,000 times from the expression matrix and comparing the mean expression between the two synthetic community treatments (for example, full synthetic community versus *Variovorax* drop-out) with the actual mean expression value from the 12 genes reported as robust auxin markers.

4.6.13 Measuring the ability of *Variovorax* to attenuate RGI induced by small molecules IAA, 2,4-dichlorophenoxyacetic acid, ethylene (ACC), cytokinins (zeatin and 6-benzylaminopurine) and flagellin 22 peptide (flg22)

4.6.13.1 Bacterial culture and plant inoculation

We embedded each of the following compounds in JM plates: 100 nM IAA (Sigma), 100 nM 1- ACC) (Sigma), 100 nM 2,4-dichlorophenoxyacetic acid (Sigma), 100 nM flg22 (PhytoTech labs), 100 nM 6-benzylaminopurine (BAP) (Sigma) and 100 nM zeatin (Sigma). As some of these compound stocks were initially solubilized in ethanol, we included comparable amounts of ethanol in the control treatments. Plates with each compound were inoculated with one of the *Variovorax* strains CL14, MF160, B4 or YR216 or with *Burkholderia* CL11. These strains were grown in separate tubes, then washed and OD600 was adjusted to 0.01 before spreading 100 µl on plates. In addition, we included uninoculated controls for each compound. We also included unamended JM plates inoculated with the RGI-inducing *Arthrobacter* CL28 co-inoculated with each of the *Variovorax* or *Burkholderia* strains, or alone. Thus, the experiment included 42 individual treatments. The experiment was repeated twice, with three independent replicates per experiment.

Upon collection, root morphology was measured ('Root and shoot image analysis' in 'Deconstructing the synthetic community to four modules of co-occurring strains').

4.6.13.2 Primary root elongation analysis

Primary root elongation was compared between bacterial treatments within each of RGI treatments tested. Differences between treatments were estimated as described in 'Primary root elongation analysis' in 'Tripartite plant–microorganism–microorganism experiments'. We plotted the estimated means with 95% confidence interval of each bacterial treatment across the different RGI treatments.

4.6.14 *In vitro* growth of *Variovorax*

Variovorax CL14 was grown in 5-ml cultures for 40 h at 28 °C in 1× M9 minimal salts medium (Sigma M6030) supplemented with 2 mM MgSO4, 0.1 mM CaCl2, 10 μM FeSO4, and a carbon source: either 15 mM succinate alone, 0.4 mM IAA with 0.5% ethanol for IAA solubilization, or both. Optical density at 600 nm and IAA concentrations were measured at six time points. IAA concentrations were measured using the Salkowski method modified from (Gordon & Weber, 1951). One hundred μl of Salkowski reagent (10 mM FeCl3 in 35% perchloric acid) was mixed with 50 μl culture supernatant or IAA standards and colour was allowed to develop for 30 min before measuring the absorbance at 530 nm.

4.6.15 Measuring plant auxin response in vivo using a bioreporter line

4.6.15.1 Bacterial culture and plant-inoculation

Seven-day-old transgenic *Arabidopsis* seedlings expressing the DR5::GFP reporter construct (Friml et al., 2003) were transferred onto plates containing: (i) 100 nM IAA, (ii) *Arthrobacter* CL28, (iii) 100 nM IAA + *Variovorax* CL14, (iv) *Arthrobacter* CL28 + *Variovorax*

CL14, (v) uninoculated plates. For treatments (ii) and (iii), Bacterial strains were grown in separate tubes, then washed and OD600 was adjusted to 0.01. For treatment (iv), OD-adjusted cultures were mixed in 1:1 ratios and spread onto agar before seedling transfer.

4.6.15.2 Fluorescence microscopy

GFP fluorescence in the root elongation zone of 8–10 plants per treatment were visualized using a Nikon Eclipse 80i fluorescence microscope at days 1, 3, 6, 9 and 13 after inoculation. The experiment was performed in two independent replicates.

From each root imaged, 10 random 30 × 30 pixel squares were sampled and average GFP intensity was measured using imageJ (Schindelin et al., 2012). Treatments were compared within each time point using ANOVA tests with Tukey's post hoc in the R base package emmeans. For visualization purposes, we plotted the estimated means of each bacterial across the different time points

4.6.16 Measuring the dual role of auxin and ethylene perception in synthetic-community-induced RGI

4.6.16.1 Bacterial culture and plant inoculation

We transferred four 7-day-old wild-type seedling and four axr2-1seedlings to each plate in this experiment. The plates contained one of five bacterial treatments: (i) *Arthrobacter* CL28, (ii) *Arthrobacter* CL28 + *Variovorax* CL14, (iii) *Variovorax* drop-out synthetic community, (iv) full synthetic community, (v) uninoculated, prepared as described in 'Bacterial culture and plant inoculation' in '*Variovorax* drop-out under varying abiotic contexts', and in 'Bacterial culture and plant inoculation' in 'Measuring plant auxin response in vivo using a bioreporter line'. Plates were placed vertically inside sealed 86 × 68 cm Ziploc bags. In one of the bags, we placed an open water

container with 80 2.5 g sachets containing 0.014% 1-MCP (Ethylene Buster, Chrystal International BV). In the second bag we added, as a control, an open water container. Both bags were placed in the growth chamber for 12 days. After 6 days of growth, we added 32 additional sachets to the 1-MCP-treated bag to maintain 1-MCP concentrations in the air. Upon collection, root morphology was measured ('Root and shoot image analysis' in 'Deconstructing the synthetic community to four modules of co-occurring strains').

4.6.16.2 Primary root elongation analysis

Primary root elongation was standardized to the no-bacteria control of each genotype, and compared between genotype and 1-MCP treatments within the *Arthrobacter* CL28 treatment and within the *Variovorax* drop-out synthetic community treatment, independently. Differences between treatments were estimated as described in 'Primary root elongation analysis' in 'Tripartite plant–microorganism–microorganism experiments'. We plotted the estimated means with 95% confidence interval of each bacterial treatment across the four genotypes. We calculated the IQR for the full and *Arthrobacter* CL28 with *Variovorax* CL14 treatments, pooling the four genotypes and MCP treatments.

4.6.17 Preparation of binarized plant images

To present representative root images, we increased contrast and subtracted background in ImageJ (Schindelin et al., 2012), then cropped the image to select representative roots. Neighboring roots were manually erased from the cropped image.

4.6.18 Mining Variovorax genomes for auxin degradation operons and ACC-deaminase genes and comparative genomics of Variovorax genomes against the other synthetic community members

We used local alignment (BLASTp) (Altschul et al., 1990) to search for the presence of the 10 genes (*iacA*, *iacB*, *iacC*, *iacD*, *iacE*, *iacF*, *iacG*, *iacH*, *iacI* and *iacY*) from a previously characterized auxin degradation operon in a different genus (J. H. J. Leveau & Gerards, 2008) across the 10 *Variovorax* isolates in our synthetic community. A minimal set of seven of these genes (*iacA*, *iacB*, *iacC*, *iacD*, *iacE*, *iacF* and *iacI*) was shown to be necessary and sufficient for auxin degradation (Donoso et al., 2017). We identified homologues for these genes across the *Variovorax* phylogeny (Figure 4.9a) at relatively low sequence identity (27–48%). Two genes of the minimal set of seven genes did not have any homologues in most *Variovorax* genomes (*iacB* and *iacI*). In addition to the *iac* operon, we scanned the genomes for the auxin degradation operon described in (Ebenau-Jehle et al., 2012) and could not identify it in any of the *Variovorax* isolates.

We also searched for the ACC deaminase gene by looking for the KEGG orthology identifier K01505 (1-aminocyclopropane-1-carboxylate deaminase) across the IMG annotations available for all our genomes.

We used orthofinder v.2.2.1 (Emms & Kelly, 2015) to construct orthogroups (group of homologue sequences) using all the coding sequences of the 10 *Variovorax* isolates included in the full 185-member synthetic community. We aligned each orthogroup using MAFFT v.707 (Katoh et al., 2002) and proceeded to build hidden Markov model (HMM) profiles from the alignments using hmmbuild v.3.1b2 (Eddy, 2011). We then used the HMM profiles that consist of core genes (present in the 10 isolates) in the *Variovorax* genus and scanned the 175 remaining genomes in the synthetic community for these HMM profiles using hmmsearch v.31.b2 (Eddy,

2011). We then selected orthogroups that were less than 5% prevalent in the 175 remaining isolates scanned. Finally, taking the *Variovorax* CL14 genome as a reference, we built hotspots of physically adjacent selected orthogroups. We used an iterative approach that extended adjacent orthogroups if they were less than 10 kb from each other. As a final filtering step, we selected hotspots that contained more than 10 selected orthogroups.

4.6.19 *Variovorax* CL14 RNA-seq in monoculture and in coculture with *Arthrobacter* CL28

4.6.19.1 Bacterial culture

Variovorax CL14 was grown either alone or in coculture with *Arthrobacter* CL28 in 5 ml of 1/10 2× YT medium (1.6 g/l tryptone, 1 g/l yeast extract, 0.5 g/l NaCl) in triplicate. The monoculture was inoculated at OD600 of 0.02 and the coculture was inoculated with OD600 of 0.01 of each strain. Cultures were grown at 28 °C to early stationary phase (approximately 22 h) and cells were collected by centrifugation at 4,100g for 15 min and frozen at −80 °C before RNA extraction.

4.6.19.2 RNA extraction and RNA-seq

Cells were lysed for RNA extraction using TRIzol Reagent (Invitrogen) according to the manufacturer instructions. Following cell lysis and phase separation, RNA was purified using the RNeasy Mini kit (Qiagen) including the optional on column DNase Digestion with the RNase-Free DNase Set (Qiagen). Total RNA quality was confirmed on the 2100 Bioanalyzer instrument (Agilent) and quantified using a Qubit 2.0 fluorometer (Invitrogen). RNA-seq libraries were prepared using the Universal Prokaryotic RNA-Seq, Prokaryotic AnyDeplete kit (Tecan, formerly NuGEN). Libraries were pooled and sequenced on the Illumina HiSeq4000 platform to generate 50-bp single-end reads.

4.6.19.3 RNA-seq analysis

We mapped the generated raw reads to the Variovorax CL14 genome (fasta file available on https://github.com/isaisg/variovoraxRGI/blob/master/rawdata/2643221508.fna) using bowtie2 (Langmead & Salzberg, 2012) with the 'very sensitive' flag. We then counted hits to each individual coding sequence annotated for the *Variovorax* CL14 genome using the function featureCounts from the R package Rsubread (Liao et al., 2013), inputting the *Variovorax* CL14 gff file (available on https://github.com/isaisg/variovoraxRGI/blob/master/rawdata/2643221508.gff) and using the default parameters with the flag allowMultiOverlap = FALSE. Finally, we used DESeq2 (Love et al., 2014) to estimate DEGs between treatments with the corresponding fold change estimates and FDR-adjusted P values.

4.6.20 *Variovorax* CL14 genomic library construction and screening

4.6.20.1 Library construction

High-molecular weight *Variovorax* CL14 genomic DNA was isolated by phenol–chloroform extraction. This genomic DNA was partially digested with Sau3A1 (New England Biolabs), and separated on the BluePippin (Sage Science) to isolate DNA fragments >12.5 kb. Vector backbone was prepared by amplifying pBBR-1MCS262 using Phusion polymerase (New England Biolabs) with primers JMC277–JMC278, digesting the PCR product with BamHI-HF (New England Biolabs), dephosphorylating with Quick CIP (New England Biolabs), and gel extracting using the QIAquick Gel Extraction Kit (Qiagen). The prepared *Variovorax* CL14 genomic DNA fragments were ligated to the prepared pBBR1-MCS2 vector backbone using ElectroLigase (New England Biolabs) and transformed by electroporation into NEB 10-beta Electrocompetent E. coli (New England Biolabs). Clones were selected by blue–white screening

on LB plates containing 1.5% agar, 50 μg/ml kanamycin, 40 μg/ml X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside), and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 37 °C. White colonies were screened by colony PCR using Taq polymerase and JMC247–JMC270 primers to eliminate clones with small inserts. The screened library clones were picked into LB medium + 50 μg/ml kanamycin, grown at 37 °C, and stored at −80 °C in 20% glycerol. The *Variovorax* CL14 genomic library comprises approximately 3,500 clones with inserts >12.5 kb in vector pBBR1-MCS2 in NEB 10-beta *E. coli*.

4.6.20.2 Library screening for IAA degradation

To screen the *Variovorax* CL14 genomic library for IAA degradation, the *E. coli* clones were grown in LB medium containing 50 μg/ml kanamycin, 1 mM IPTG, 0.05 mg/ml IAA, and 0.25% ethanol from IAA solubilization for 3 days at 37 °C. Salkowski reagent (10 mM FeCl3 in 35% perchloric acid) was mixed with culture supernatant 2:1 and colour was allowed to develop for 30 min before measuring the absorbance at 530 nm. Two clones from the library (plate 8A well E8 and plate2A well F10, henceforth vector 1 and vector 2, respectively) were identified as degrading IAA. The *Variovorax* CL14 genes contained in vectors 1 and 2 were inferred by isolating plasmid from these clones using the ZymoPURE II Plasmid Midiprep Kit (Zymo Research) and Sanger-sequencing the insert ends using primers JMC247 and JMC270. Double digest of the purified plasmids with SacI and EcoRV confirmed the size of the inserts. Vector 1 contains a 35-kb insert and vector 2 contains a 15-kb insert (nucleotide coordinates 29,100–64,406 and 52,627–67,679, respectively, from *Variovorax* CL14 scaffold Ga0102008_10005) (Figures 4.4a and 4.13b). The genes in both inserts are in the same direction as the IPTG0inducible Lac promoter used to drive LacZ-alpha expression for blue–white screening on pBBR1-MCS2.

4.6.21 *Acidovorax* Root219::EV and *Acidovorax* root219::V2 construction and screening

Triparental mating was used to mobilize vector 2 or control EV pBBR1-MCS2 from *E. coli* to *Acidovorax* root219. Donor NEB 10-beta *E. coli* containing the vector for conjugation and helper strain *E. coli* pRK201363 were grown in LB medium containing 50 μg/ml kanamycin at 37 °C. *Acidovorax* root219 was grown in 2xYT medium containing 100 μg/ml ampicillin at 28 °C. Bacteria were washed 3 times with 2xYT medium without antibiotics, mixed in a ratio of approximately 1:1:10 donor:helper:recipient, centrifuged and resuspended in 1/10 the volume and plated as a pool on LB agar plates without antibiotics and grown at 28 °C. Eighteen to thirty h later, exconjugantes were streaked on LB agar plates containing 50 μg/ml kanamycin and 100 μg/ml ampicillin to select only *Acidovorax* root219 containing the conjugated vector. The resulting strains are designated *Acidovorax* root219::EV containing empty vector pBBR1-MCS2 and *Acidovorax* root219::V2 containing vector 2. In vitro IAA degradation was performed as in 'In vitro growth of *Variovorax'* using M9 medium with carbon sources: 15 mM succinate, 0.1 mg/ml IAA and 0.5% ethanol with the addition of 50 μg/ml kanamycin and 1 mM IPTG. Primary root elongation measurement was performed as described in 'Root and shoot image analysis' in 'Deconstructing the synthetic community to four modules of co-occurring strains', on MS medium with 1 mM IPTG and RGI induced by either 100 nM IAA or *Arthrobacter* CL28. *Acidovorax* root219::V2 root colonization was compared to *Variovorax* CL14 colonization by plating a subset of ground root samples from the root elongation experiment (see 'Determination of *Arthrobacter* CL28 colony forming units from roots' for root collection and processing protocol) on LB agar plates containing 100 μg/ml ampicillin, for which *Arthrobacter* CL28 is susceptible and *Variovorax* CL14 and *Acidovorax* Root219 are not. Number of colony-forming units (CFUs) was normalized to root weight (Figure 4.13d).

4.6.22 *Variovorax* hotspot 33 knockout construction and screening

4.6.22.1 Knockout suicide vector pJMC158 construction

The unmarked deletion mutant *Variovorax* CL14 Δ2643613653–2643613677 (*Variovorax* CL14 ΔHS33) was constructed based on a genetic system developed for *Burkholderia* spp. and its suicide vector pMo13064

The vector backbone was amplified from pMo130 using primers JMC203–JMC204 with Q5 DNA polymerase (New England Biolabs), cleaned up and treated with DpnI (New England Biolabs). One-kb regions for homologous recombination flanking *Variovorax* CL14 genes 2643613653–2643613677 were amplified using Q5 Polymerase (New England Biolabs) and primers JMC533–JMC534 and JMC535–JMC536. The vector was assembled with Gibson Assembly Mastermix (New England Biolabs) at 50 °C for 1 h, transformed into NEB 5-alpha chemically competent *E. coli* (New England Biolabs), and plated on LB agar with 50 μg/ml kanamycin. pJMC158 DNA was isolated from a clone using the ZR Plasmid Miniprep Classic Kit (Zymo Research), sequence confirmed, and transformed into biparental mating strain *E. coli* WM3064. *E. coli* strain WM3064 containing pJMC158 was maintained on LB containing 50 μg/ml kanamycin and 0.3 mM diaminopimelic acid (DAP) at 37 °C.

4.6.22.2 Conjugative transfer of pJMC158 into *Variovorax* CL14

For biparental mating, E. coli WM3064 containing pJMC158 was grown as above, and *Variovorax* CL14 was grown in 2xYT medium containing 100 μg/ml ampicillin. Each strain was washed separately 3 times with 2xYT medium, then mixed at ratios between 1:1–1:10 donor:recipient, centrifuged and resuspended in approximately 1/10 the volume and plated in a single pool on LB agar containing 0.3 mM DAP and grown at 28 °C overnight. Exconjugants were

streaked onto LB plates containing 100 μg/ml ampicillin, 50 μg/ml kanamycin lacking DAP and grown at 28 °C to select *Variovorax* CL14 strains that incorporated suicide vector pJMC158. First crossover strains were subsequently purified once by restreaking and then individual colonies grown in LB with 100 μg/ml ampicillin, 50 μg/ml kanamycin.

4.6.22.3 Resolution of pJMC158 integration and knockout strain purification and verification

To resolve the integration of pJMC158, first crossover strains were grown once in LB medium containing 100 μg/ml ampicillin and 1 mM IPTG then plated on medium containing 10 g/l tryptone, 5 g/l yeast extract, 100 g/l sucrose, 1.5% agar, 100 μg/ml ampicillin and 1 mM IPTG. Colonies were picked into the same liquid medium and grown once. The resulting strains were screened by PCR using Q5 polymerase for deletion of genes 2643613653–2643613677 using primers JMC568–JMC569.These strains were subsequently plate-purified at least 3 times on LB 100 μg/ml ampicillin plates. To ensure strain purity, PCR primers were designed to amplify from outside into the genes that were deleted (primer pairs JMC571–JMC569 and JMC568–JMC570). These PCR reactions were performed using Q5 polymerase with wild-type *Variovorax* CL14 as a control. All genomic DNA used for screening PCR was isolated using the Quick-DNA miniprep kit (Zymo Research). The resulting knockout strain was designated *Variovorax* CL14 ΔHS33.

4.6.22.4 Screening of *Variovorax* CL14ΔHS33

In vitro IAA degradation was performed as in 'In vitro growth of *Variovorax'* using M9 medium with carbon sources: 15 mM succinate, 0.1 mg/ml IAA and 0.5% ethanol. Primary root elongation measurement was performed as described in 'Root and shoot image analysis' in 'Deconstructing the synthetic community to four modules of co-occurring strains', on MS medium with 1mM IPTG and RGI induced by either 100 nM IAA or *Arthrobacter* CL28. To control for

pleiotropic colonization effects, CFU counts were obtained for both CL14 and CL14 ΔHS33 in binary-association with the plant (see 'Determination of *Arthrobacter* CL28 colony-forming units from roots' for root collection and processing protocol). CL14 ΔHS33 colonization was not impaired (Figure 4.13e).

4.6.23 Screening existing 16S rRNA census data for *Variovorax*

4.6.23.1 Natural Arabidopsis populations across Europe (Thiergart et al., 2020)

We used the DADA2 (Callahan et al., 2016) pipeline to create amplicon sequence variants (ASVs) from the raw reads published in (Thiergart et al., 2020) We then used the naive Bayes classifier implemented in mothur (Schloss et al., 2009) to taxonomically classify each ASV using the SILVA 132 database (Quast et al., 2013). To determine prevalence and relative abundance of ASVs, we followed the same approach as in (Thiergart et al., 2020). The bacterial ASV table was restricted to samples having >1,000 reads. The table was transformed to relative abundance (RA) by dividing each value in a sample by the total reads in that sample. An ASV was considered to be present in a sample if it had an RA >0.01% in that particular sample. To calculate the average relative abundance (y-axis in Figure 4.14a, b) independently of prevalence, we calculated the mean RA using only the sample for which each ASV was considered present. We used the same classification scheme to colour the ASVs across the scatter plots shown. ASVs present in >80% of the sites on average are considered as widespread ASVs.

4.6.23.2 Different plant species in the same soil (Fitzpatrick et al., 2018)

We used the ASV table provided within (Fitzpatrick et al., 2018) and followed the same pipeline described in 'Natural Arabidopsis populations across Europe'. For each sample in the dataset, an ASV was considered to be present if that ASV had an RA >0.01% in that particular

sample. To calculate the average relative abundance (y-axis in Figure 4.14 a, b) independently of prevalence, we calculated the mean RA using only the samples for which each ASV was considered present. For each one of the thirty plant species in this data set, an ASV was considered present in that plant species if it was present in >70% of samples coming from that plant species. An ASV was considered core if it was present in all 30 plant species surveyed in this experiment.

CHAPTER 5: THE FUTURE OF THE PLANT MICROBIOME FIELD[5]

Introduction

When I started my graduate training, the utilization of synthetic ecology as a framework to elucidate molecular mechanisms was a pioneering methodology applied solely by a handful of research groups. Five years afterwards, at the twilight of my doctorate, the deployment of amenable microcosms and synthetic communities has become a common tool across the whole plant microbiome field. The democratization of this research framework has brought new challenges to overcome and new frontiers of knowledge to discover. Both ecological and reductionist approaches have led to important discoveries in plant microbiome research, yet in isolation, these approaches are limited. Mechanism is impossible to discern using amplicon or metagenomic surveys, and ecological importance is often unknown with reductionist approaches. Given the labor, cost, and time associated with large-scale surveys and detailed mechanistic studies, integrating these approaches may not always be feasible in a single research group, let alone a single project. Instead, integration needs to take place across research groups to advance the understanding of the mechanistic underpinnings and consequences of variation observed in plant microbiota in natural and managed habitats. In this chapter I will highlight some of the current methodological challenges and new frontiers of knowledge faced by the plant microbiome field.

[5]A portion of the results presented in this chapter appeared as an article in:
Fitzpatrick CR*, Salas-González I*, Conway JM, Finkel OM, Gilbert S, Russ D, Teixeira PJPL, Dangl JL. The Plant Microbiome: From Ecology to Reductionism and Beyond. Annu Rev Microbiol. 2020 Sep 8;74:81-100.

5.1 Revisiting bacterial comparative genomics in the light of reverse ecology

The explosion in culturomics, propelled by the development of highly parallelizable isolation methods (Nichols et al., 2010) and the decreased cost of genome-sequencing, had led to the accumulation of microbial cultured collections isolated from diverse plant hosts and conditions. The increasing availability of both the cultured isolates and their associated genomic data, will permit to test novel hypothesis concerning microbial adaptation to distinct plant hosts and co-adaptation of both, host and microbes, to particular agronomically-relevant environments such as low nutrient soils.

Computationally, the development of novel genomic-based reverse-ecology approaches (Arevalo et al., 2019) permit one to leverage the available genomic data of closely-related isolates (same bacterial species) to delineate signatures of bacterial adaptation. By intersecting this reverse ecology approach with isolates metadata, such as host or environmental source of isolation, it is possible to identify the genomic features associated with the bacterial adaptation to specific plant hosts or environmental conditions. In addition, by reconstructing groups of homologs between bacterial species, it is feasible to estimate the generality of the traits that exhibit signatures of adaptation across the species of bacterial genus with sufficient genomic representation in the plant-microbe culture collections.

Currently, it is possible to test this evolutionary comparative genomics framework, using *Pseudomonas* species, isolated from *Arabidopsis thaliana* wild populations across Europe (Karasov et al., 2018), and other bacterial genus with relevant host-associated phenotypes such as *Variovorax* (maintaining root growth in the context of a microbial community) and *Methylobacterium* isolates (plant growth promoters) (Krug et al., 2020).

5.2 An eco-evolutionary framework to universalize SynCom-derived knowledge

One of the biggest challenges of deploying synthetic communities resides in the fact that each designed community carries unique meta-properties because of the unique genomic architectures of the members within the community. Due to these unique meta-properties, the findings obtained from SynCom-based experiments lack universality in scenarios where specific microbe-microbe interactions are missing in a given SynCom. In order to alleviate the context-dependency of SynCom-derived results, it is important to establish a methodological framework that increases the reproducibility and scalability of these findings. Here, I present an eco-evolutionary methodological framework (Figure 5.1) that maximizes scalability and reproducibility of SynCom-derived findings:

1) Design of SynComs, with highly reproducible and characterized phenotypes, that serve as control treatments for SynCom-based experiments. The utilization of control SynComs would reduce technical noise inherent to the biological systems and increase the interpretability of SynCom-based results by establishing a biologically better control that the current utilized axenic (NB) control.

2) Establishment of experimentally-independent standardized statistics to permit the comparability of SynCom-derived effects across multiple studies. The adoption of these universal metrics in conjunction with deployment of control SynComs would facilitate to contrast effects (e.g Plant mutants affecting microbiome assembly) via meta-analysis.

3) When possible, SynCom-based claims should be supported by ecological evidence obtained from wild surveys employing either marker-based or metagenomic approaches.

4) SynCom-based claims obtained for a particular host-species should be extended across other plant-species. This extended evolutionary approach, besides demonstrating generality of a

particular microbial effect, can also identify novel mechanisms of plant-microbiome interactions by exploring plant lineage-specific specializations distributed across the plant phylogeny.

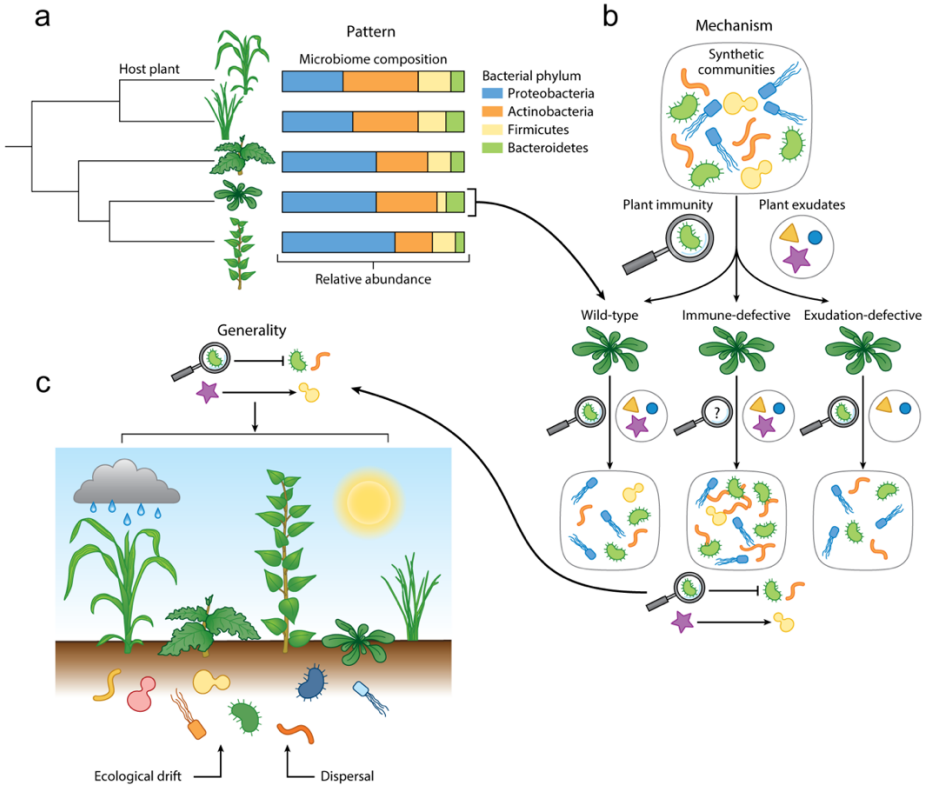


Figure 5.1 - **A three-part research plan to advance our understanding of the plant microbiome under natural and managed settings.**

a) The plan begins with describing patterns in the distribution of microbial taxa or genes using census experiments. **b)** The mechanism or mechanisms underlying these patterns are then investigated using reductionist experiments. **c)** Finally, evaluating the operative role of the identified mechanisms across diverse systems demonstrates generality. For example, numerous census experiments reveal that the composition of plant microbiota, including those found in leaves and roots, varies among host plants. Reductionist studies, which use a variety of approaches including synthetic communities paired with targeted mutants of model plant species, support the mechanism that innate plant immunity and plant exudation profiles shape leaf and root microbiota. However, these mechanisms have yet to be evaluated across a wide range of settings, which could include different host plant populations or species, variation in the abiotic and biotic environment, and different resident microbiota from which the plant microbiome is derived. Moreover, the ecological processes of dispersal and drift occurring in conjunction with selection may obscure the operation of the identified mechanisms under more complex settings

5.3 The plant microbiome project

Analogous to the Human Microbiome Project (Turnbaugh et al., 2007), the plant microbiome field requires to develop a standardizable, scalable and accessible repository to store and distribute the data from plant-microbiome ecological surveys, either amplicon-based or metagenomics-based, and the genomic information form plant-derived microbial isolates. The creation of this community resource would facilitate the deployment of the aforementioned eco-evolutionary reproducible research framework ultimately generating novel findings by integrating knowledge across multiple studies

REFERENCES

Abreu, M. E., & Munne-Bosch, S. (2009). Salicylic acid deficiency in NahG transgenic lines and sid2 mutants increases seed yield in the annual plant Arabidopsis thaliana. *Journal of Experimental Botany*, *60*(4), 1261–1271. https://doi.org/10.1093/jxb/ern363

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amend, A. S., Cobian, G. M., Laruson, A. J., Remple, K., Tucker, S. J., Poff, K. E., Antaky, C., Boraks, A., Jones, C. A., Kuehu, D., Lensing, B. R., Pejhanmehr, M., Richardson, D. T., & Riley, P. P. (2019). Phytobiomes are compositionally nested from the ground up. *PeerJ*, *2019*(3). https://doi.org/10.7717/peerj.6609

Ames, B. N. (1966). [10] Assay of inorganic phosphate, total phosphate and phosphatases. *Methods in Enzymology*, *8*(C), 115–118. https://doi.org/10.1016/0076-6879(66)08014-5

Angel, R., Soares, M. I. M., Ungar, E. D., & Gillor, O. (2010). Biogeography of soil archaea and bacteria along a steep precipitation gradient. *ISME Journal*, *4*(4), 553–563. https://doi.org/10.1038/ismej.2009.136

Arevalo, P., VanInsberghe, D., Elsherbini, J., Gore, J., & Polz, M. F. (2019). A Reverse Ecology Approach Based on a Biological Definition of Microbial Populations. *Cell*, *178*(4), 820-834.e14. https://doi.org/10.1016/j.cell.2019.06.033

Atamna-Ismaeel, N., Finkel, O., Glaser, F., von Mering, C., Vorholt, J. A., Koblížek, M., Belkin, S., & Béjà, O. (2012). Bacterial anoxygenic photosynthesis on plant leaf surfaces. *Environmental Microbiology Reports*, *4*(2), 209–216. https://doi.org/10.1111/j.1758-2229.2011.00323.x

Atamna-Ismaeel, N., Finkel, O. M., Glaser, F., Sharon, I., Schneider, R., Post, A. F., Spudich, J. L., von Mering, C., Vorholt, J. A., Iluz, D., Béjà, O., & Belkin, S. (2012). Microbial rhodopsins on leaf surfaces of terrestrial plants. *Environmental Microbiology*, *14*(1), 140–146. https://doi.org/10.1111/j.1462-2920.2011.02554.x

Bahram, M., Hildebrand, F., Forslund, S. K., Anderson, J. L., Soudzilovskaia, N. A., Bodegom, P. M., Bengtsson-Palme, J., Anslan, S., Coelho, L. P., Harend, H., Huerta-Cepas, J., Medema, M. H., Maltz, M. R., Mundra, S., Olsson, P. A., Pent, M., Põlme, S., Sunagawa, S., Ryberg, M., … Bork, P. (2018). Structure and function of the global topsoil microbiome. *Nature*, *560*(7717), 233–237. https://doi.org/10.1038/s41586-018-0386-6

Bai, Y., Müller, D. B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., Dombrowski, N., Münch, P. C., Spaepen, S., Remus-Emsermann, M., Hüttel, B., McHardy, A. C., Vorholt, J. A., & Schulze-Lefert, P. (2015). Functional overlap of the Arabidopsis leaf and root microbiota. *Nature*, 1–19. https://doi.org/10.1038/nature16192

Bardoel, B. W., van der Ent, S., Pel, M. J. C., Tommassen, J., Pieterse, C. M. J., van Kessel, K. P. M., & van Strijp, J. A. G. (2011). Pseudomonas Evades Immune Recognition of Flagellin in Both Mammals and Plants. *PLoS Pathogens*, *7*(8), e1002206. https://doi.org/10.1371/journal.ppat.1002206

Basler, M., Ho, B. T., & Mekalanos, J. J. (2013). Tit-for-tat: Type VI secretion system counterattack during bacterial cell-cell interactions. *Cell*, *152*(4), 884–894. https://doi.org/10.1016/j.cell.2013.01.042

Baumann, P. (2005). Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. In *Annual Review of Microbiology* (Vol. 59, pp. 155–189). Annu Rev Microbiol. https://doi.org/10.1146/annurev.micro.59.030804.121041

Baxter, I. R., Vitek, O., Lahner, B., Muthukumar, B., Borghi, M., Morrissey, J., Guerinot, M. Lou, & Salt, D. E. (2008). The leaf ionome as a multivariable system to detect a plant's physiological status. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(33), 12081–12086. https://doi.org/10.1073/pnas.0804175105

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis*, *53*(8), 474–485. https://doi.org/10.1002/dvg.22877

Berendsen, R. L., Vismans, G., Yu, K., Song, Y., de Jonge, R., Burgman, W. P., Burmølle, M., Herschend, J., H M Bakker, P. A., & J Pieterse, C. M. (2018). Disease-induced assemblage of a plant-beneficial bacterial consortium. *The ISME Journal*, *12*, 1496–1507. https://doi.org/10.1038/s41396-018-0093-1

Berens, M. L., Wolinska, K. W., Spaepen, S., Ziegler, J., Nobori, T., Nair, A., Krüler, V., Winkelmüller, T. M., Wang, Y., Mine, A., Becker, D., Garrido-Oter, R., Schulze-Lefert, P., & Tsuda, K. (2019). Balancing trade-offs between biotic and abiotic stress responses through leaf age-dependent variation in stress hormone cross-talk. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(6), 2364–2373.

https://doi.org/10.1073/pnas.1817233116

Beszteri, B., Temperton, B., Frickenhaus, S., & Giovannoni, S. J. (2010). Average genome size: a potential source of bias in comparative metagenomics. *The ISME Journal*, *4*, 1075–1077. https://doi.org/10.1038/ismej.2010.29

Bodenhausen, N., Horton, M. W., & Bergelson, J. (2013). Bacterial Communities Associated with the Leaves and the Roots of Arabidopsis thaliana. *PLoS ONE*, *8*(2), e56329. https://doi.org/10.1371/journal.pone.0056329

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bowen, J. L., Kearns, P. J., Byrnes, J. E. K., Wigginton, S., Allen, W. J., Greenwood, M., Tran, K., Yu, J., Cronin, J. T., & Meyerson, L. A. (2017). Lineage overwhelms environmental conditions in determining rhizosphere bacterial community structure in a cosmopolitan invasive plant. *Nature Communications*, *8*(1), 1–8. https://doi.org/10.1038/s41467-017-00626-0

Bowsher, A. W., Kearns, P. J., Popovic, D., Lowry, D. B., & Shade, A. (2020). Locally Adapted *Mimulus* Ecotypes Differentially Impact Rhizosphere Bacterial and Archaeal Communities in an Environment-Dependent Manner. *Phytobiomes Journal*, *4*(1), 53–63. https://doi.org/10.1094/PBIOMES-05-19-0026-R

Brumos, J., Robles, L. M., Yun, J., Vu, T. C., Jackson, S., Alonso, J. M., & Stepanova, A. N. (2018). Local Auxin Biosynthesis Is a Key Regulator of Plant Development. *Developmental Cell*, *47*(3), 306-318.e5. https://doi.org/10.1016/j.devcel.2018.09.022

Brynildsrud, O., Bohlin, J., Scheffer, L., & Eldholm, V. (2016). Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biology*, *17*(1), 238. https://doi.org/10.1186/s13059-016-1108-8

Bulgarelli, D., Garrido-Oter, R., Münch, P. C., Weiman, A., Dröge, J., Pan, Y., McHardy, A. C., & Schulze-Lefert, P. (2015). Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host and Microbe*, *17*(3), 392–403. https://doi.org/10.1016/j.chom.2015.01.011

Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F., Rauf, P., Huettel, B., Reinhardt, R., Schmelzer, E., Peplies, J., Gloeckner, F. O., Amann,

R., Eickhorst, T., & Schulze-Lefert, P. (2012). Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. In *Nature* (Vol. 488, Issue 7409, pp. 91–95). Nature Publishing Group. https://doi.org/10.1038/nature11336

Bulgarelli, D., Schlaeppi, K., Spaepen, S., van Themaat, E. V. L., & Schulze-Lefert, P. (2013). Structure and Functions of the Bacterial Microbiota of Plants. *Annual Review of Plant Biology*, *64*(1), 807–838. https://doi.org/10.1146/annurev-arplant-050312-120106

Burstein, D., Zusman, T., Degtyar, E., Viner, R., Segal, G., & Pupko, T. (2009). Genome-Scale Identification of Legionella pneumophila Effectors Using a Machine Learning Approach. *PLoS Pathogens*, *5*(7), e1000508. https://doi.org/10.1371/journal.ppat.1000508

Bustos, R., Castrillo, G., Linhares, F., Puga, M. I., Rubio, V., Pérez-Pérez, J., Solano, R., Leyva, A., & Paz-Ares, J. (2010). A central regulatory system largely controls transcriptional activation and repression responses to phosphate starvation in arabidopsis. *PLoS Genetics*, *6*(9). https://doi.org/10.1371/journal.pgen.1001102

Büttner, D., & He, S. Y. (2009). Type III protein secretion in plant pathogenic bacteria. *Plant Physiology*, *150*(4), 1656–1664. https://doi.org/10.1104/pp.109.139089

Caddell, D. F., Deng, S., & Coleman-Derr, D. (2019). Role of the plant root microbiome in abiotic stress tolerance. In *Seed Endophytes: Biology and Biotechnology* (pp. 273–311). Springer International Publishing. https://doi.org/10.1007/978-3-030-10504-4_14

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Carlström, C. I., Field, C. M., Bortfeld-Miller, M., Müller, B., Sunagawa, S., & Vorholt, J. A. (2019). Synthetic microbiota reveal priority effects and keystone strains in the Arabidopsis phyllosphere. *Nature Ecology and Evolution*, *3*(10), 1445–1454. https://doi.org/10.1038/s41559-019-0994-z

Carrión, V. J., Perez-Jaramillo, J., Cordovez, V., Tracanna, V., De Hollander, M., Ruiz-Buck, D., Mendes, L. W., van Ijcken, W. F. J., Gomez-Exposito, R., Elsayed, S. S., Mohanraju, P., Arifah, A., van der Oost, J., Paulson, J. N., Mendes, R., van Wezel, G. P., Medema, M. H.,

& Raaijmakers, J. M. (2019). Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. *Science*, *366*(6465), 606–612. https://doi.org/10.1126/science.aaw9285

Carvalhais, L. C., Dennis, P. G., Fan, B., Fedoseyenko, D., Kierul, K., Becker, A., von Wiren, N., & Borriss, R. (2013). Linking Plant Nutritional Status to Plant-Microbe Interactions. *PLoS ONE*, *8*(7), e68555. https://doi.org/10.1371/journal.pone.0068555

Cary, A. J., Liu Wennuan, & Howell, S. H. (1995). Cytokinin action is coupled to ethylene in its effects on the inhibition of root and hypocotyl elongation in Arabidopsis thaliana seedlings. *Plant Physiology*, *107*(4), 1075–1082. https://doi.org/10.1104/pp.107.4.1075

Castrillo, G., Teixeira, P. J. P. L., Paredes, S. H., Law, T. F., De Lorenzo, L., Feltcher, M. E., Finkel, O. M., Breakfield, N. W., Mieczkowski, P., Jones, C. D., Paz-Ares, J., & Dangl, J. L. (2017). Root microbiota drive direct integration of phosphate stress and immunity. *Nature*, *543*(7646), 513–518. https://doi.org/10.1038/nature21417

Cesari, S., Bernoux, M., Moncuquet, P., Kroj, T., & Dodds, P. N. (2014). A novel conserved mechanism for plant NLR protein pairs: The "integrated decoy" hypothesis. *Frontiers in Plant Science*, *5*(NOV). https://doi.org/10.3389/fpls.2014.00606

Cha, J. Y., Han, S., Hong, H. J., Cho, H., Kim, D., Kwon, Y., Kwon, S. K., Crusemann, M., Bok Lee, Y., Kim, J. F., Giaever, G., Nislow, C., Moore, B. S., Thomashow, L. S., Weller, D. M., & Kwak, Y. S. (2016). Microbial and biochemical basis of a Fusarium wilt-suppressive soil. *ISME Journal*, *10*(1), 119–129. https://doi.org/10.1038/ismej.2015.95

Chen, L., Dodd, I. C., Theobald, J. C., Belimov, A. A., & Davies, W. J. (2013). The rhizobacterium Variovorax paradoxus 5C-2, containing ACC deaminase, promotes growth and development of Arabidopsis thaliana via an ethylene-dependent pathway. *Journal of Experimental Botany*, *64*(6), 1565–1573. https://doi.org/10.1093/jxb/ert031

Cheng, Y. T., Zhang, L., & He, S. Y. (2019). Plant-Microbe Interactions Facing Environmental Challenge. In *Cell Host and Microbe* (Vol. 26, Issue 2, pp. 183–192). Cell Press. https://doi.org/10.1016/j.chom.2019.07.009

Choudoir, M. J., Barberán, A., Menninger, H. L., Dunn, R. R., & Fierer, N. (2018). Variation in range size and dispersal capabilities of microbial taxa. *Ecology*, *99*(2), 322–334. https://doi.org/10.1002/ecy.2094

Chowdhury, S. P., Hartmann, A., Gao, X., & Borriss, R. (2015). Biocontrol mechanism by root-

associated Bacillus amyloliquefaciens FZB42 – a review. *Frontiers in Microbiology*, *6*(JUL), 780. https://doi.org/10.3389/fmicb.2015.00780

Clauss, M. J., & Aarssen, L. W. (1994). Phenotypic Plasticity of Size--Fecundity Relationships in Arabidopsis Thaliana. *The Journal of Ecology*, *82*(3), 447. https://doi.org/10.2307/2261254

Cole, B. J., Feltcher, M. E., Waters, R. J., Wetmore, K. M., Mucyn, T. S., Ryan, E. M., Wang, G., Ul-Hasan, S., McDonald, M., Yoshikuni, Y., Malmstrom, R. R., Deutschbauer, A. M., Dangl, J. L., & Visel, A. (2017). Genome-wide identification of bacterial plant colonization genes. *PLOS Biology*, *15*(9), e2002860. https://doi.org/10.1371/journal.pbio.2002860

Coutinho, B. G., Licastro, D., Mendonça-Previato, L., Cámara, M., & Venturi, V. (2015). Plant-Influenced Gene Expression in the Rice Endophyte Burkholderia kururiensis M130. *Molecular Plant-Microbe Interactions MPMI*, *28*(1), 10–21. https://doi.org/10.1094/MPMI-07-14-0225-R

Cregger, M. A., Veach, A. M., Yang, Z. K., Crouch, M. J., Vilgalys, R., Tuskan, G. A., & Schadt, C. W. (2018). The Populus holobiont: Dissecting the effects of plant niches and genotype on the microbiome. *Microbiome*, *6*(1), 31. https://doi.org/10.1186/s40168-018-0413-8

Dangl, J. L., & Jones, J. D. G. (2001). Plant pathogens and integrated defence responses to infection. In *Nature* (Vol. 411, Issue 6839, pp. 826–833). Nature. https://doi.org/10.1038/35081161

De Weert, S., Vermeiren, H., Mulders, I. H. M., Kuiper, I., Hendrickx, N., Bloemberg, G. V, Vanderleyden, J., De Mot, R., & Lugtenberg, B. J. J. (2002). *Flagella-Driven Chemotaxis Towards Exudate Components Is an Important Trait for Tomato Root Colonization by Pseudomonas fluorescens* (Vol. 15, Issue 11).

De Weger, L. A., Van Der Vlugt, C. I. M., Wijfjes, A. H. M., Bakker, P. A. H. M., Schippers, B., & Lugtenberg1, B. (1987). Flagella of a Plant-Growth-Stimulating Pseudomonas fluorescens Strain Are Required for Colonization of Potato Roots Downloaded from. In *JOURNAL OF BACTERIOLOGY* (Vol. 169, Issue 6). http://jb.asm.org/

Dean, P. (2011). Functional domains and motifs of bacterial type III effector proteins and their roles in infection. In *FEMS Microbiology Reviews* (Vol. 35, Issue 6, pp. 1100–1125). FEMS Microbiol Rev. https://doi.org/10.1111/j.1574-6976.2011.00271.x

Deng, P., Wang, X., Baird, S. M., Showmaker, K. C., Smith, L., Peterson, D. G., & Lu, S. (2016). Comparative genome-wide analysis reveals that Burkholderia contaminans MS14 possesses multiple antimicrobial biosynthesis genes but not major genetic loci required for pathogenesis. *MicrobiologyOpen*, *5*(3), 353–369. https://doi.org/10.1002/mbo3.333

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., & Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, *72*(7), 5069–5072. https://doi.org/10.1128/AEM.03006-05

Desbrosses, G. J., & Stougaard, J. (2011). Root nodulation: A paradigm for how plant-microbe symbiosis influences host developmental pathways. In *Cell Host and Microbe* (Vol. 10, Issue 4, pp. 348–358). Cell Host Microbe. https://doi.org/10.1016/j.chom.2011.09.005

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. In *Journal of Vegetation Science* (Vol. 14, Issue 6, pp. 927–930). Opulus Press AB. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

Donald, J., Roy, M., Suescun, U., Iribar, A., Manzi, S., Péllissier, L., Gaucher, P., & Chave, J. (2020). A test of community assembly rules using foliar endophytes from a tropical forest canopy. *Journal of Ecology*, *108*(4), 1605–1616. https://doi.org/10.1111/1365-2745.13344

Donoso, R., Leiva-Novoa, P., Zúñiga, A., Timmermann, T., Recabarren-Gajardo, G., & González, B. (2017). Biochemical and genetic bases of indole-3-acetic acid (auxin phytohormone) degradation by the plantgrowth- promoting rhizobacterium Paraburkholderia phytofirmans PsJN. *Applied and Environmental Microbiology*, *83*(1). https://doi.org/10.1128/AEM.01991-16

Doty, S. L., Oakley, B., Xin, G., Kang, J. W., Singleton, G., Khan, Z., Vajzovic, A., & Staley, J. T. (2009). Diazotrophic endophytes of native black cottonwood and willow. *Symbiosis*, *47*(1), 23–33. https://doi.org/10.1007/BF03179967

Durán, P., Thiergart, T., Garrido-Oter, R., Agler, M., Kemen, E., Schulze-Lefert, P., & Hacquard, S. (2018). Microbial Interkingdom Interactions in Roots Promote Arabidopsis Survival. *Cell*, *175*(4), 973-983.e14. https://doi.org/10.1016/j.cell.2018.10.020

Ebenau-Jehle, C., Thomas, M., Scharf, G., Kockelkorn, D., Knapp, B., Schühle, K., Heider, J., & Fuchs, G. (2012). Anaerobic metabolism of indoleacetate. *Journal of Bacteriology*, *194*(11), 2894–2903. https://doi.org/10.1128/JB.00250-12

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Computational Biology*, *7*(10), 1002195. https://doi.org/10.1371/journal.pcbi.1002195

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Edgar, R. C. (2013). UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, *10*(10), 996–998. https://doi.org/10.1038/nmeth.2604

Edwards, J., Johnson, C., Santos-Medellín, C., Lurie, E., Podishetty, N. K., Bhatnagar, S., Eisen, J. A., Sundaresan, V., & Jeffery, L. D. (2015). Structure, variation, and assembly of the root-associated microbiomes of rice. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(8), E911–E920. https://doi.org/10.1073/pnas.1414592112

Eichinger, V., Nussbaumer, T., Platzer, A., Jehl, M. A., Arnold, R., & Rattei, T. (2016). EffectiveDB - Updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. *Nucleic Acids Research*, *44*(D1), D669–D674. https://doi.org/10.1093/nar/gkv1269

Eissfeller, V., Beyer, F., Valtanen, K., Hertel, D., Maraun, M., Polle, A., & Scheu, S. (2013). Incorporation of plant carbon and microbial nitrogen into the rhizosphere food web of beech and ash. *Soil Biology and Biochemistry*, *62*, 76–81. https://doi.org/10.1016/j.soilbio.2013.03.002

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*(1), 157. https://doi.org/10.1186/s13059-015-0721-2

Espinoza, C., Liang, Y., & Stacey, G. (2017). Chitin receptor <scp>CERK</scp> 1 links salt stress and chitin-triggered innate immunity in Arabidopsis. *The Plant Journal*, *89*(5), 984–995. https://doi.org/10.1111/tpj.13437

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Fabiańska, I., Gerlach, N., Almario, J., & Bucher, M. (2019). Plant-mediated effects of soil phosphorus on the root-associated fungal microbiota in Arabidopsis thaliana. *New Phytologist*, *221*(4), 2123–2137. https://doi.org/10.1111/nph.15538

Faure, D., Vereecke, D., & Leveau, J. H. J. (2009). Molecular communication in the rhizosphere. In *Plant and Soil* (Vol. 321, Issues 1–2, pp. 279–303). Springer. https://doi.org/10.1007/s11104-008-9839-2

Fibach-Paldi, S., Burdman, S., & Okon, Y. (2012). Key physiological properties contributing to rhizosphere adaptation and plant growth promotion abilities of Azospirillum brasilense. *FEMS Microbiology Letters*, *326*(2), 99–108. https://doi.org/10.1111/j.1574-6968.2011.02407.x

Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(3), 626–631. https://doi.org/10.1073/pnas.0507535103

Fierer, N., Lauber, C. L., Ramirez, K. S., Zaneveld, J., Bradford, M. A., & Knight, R. (2012). Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME Journal*, *6*(5), 1007–1017. https://doi.org/10.1038/ismej.2011.159

Finkel, O. M., Burch, A. Y., Elad, T., Huse, S. M., Lindow, S. E., Post, A. F., & Belkin, S. (2012). Distance-decay relationships partially determine diversity patterns of phyllosphere bacteria on Tamrix trees across the sonoran desert. *Applied and Environmental Microbiology*, *78*(17), 6187–6193. https://doi.org/10.1128/AEM.00888-12

Finkel, O. M., Delmont, T. O., Post, A. F., & Belkin, S. (2016). Metagenomic signatures of bacterial adaptation to life in the phyllosphere of a salt-secreting desert tree. *Applied and Environmental Microbiology*, *82*(9), 2854–2861. https://doi.org/10.1128/AEM.00483-16

Finkel, O. M., Salas-González, I., Castrillo, G., Spaepen, S., Law, T. F., Teixeira, P. J. P. L., Jones, C. D., & Dangl, J. L. (2019). The effects of soil phosphorus content on plant microbiota are driven by the plant phosphate starvation response. *PLoS Biology*, *17*(11). https://doi.org/10.1371/journal.pbio.3000534

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., & Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279–D285. https://doi.org/10.1093/nar/gkv1344

Fitzpatrick, C. R., Copeland, J., Wang, P. W., Guttman, D. S., Kotanen, P. M., & Johnson, M. T. J. (2018). Assembly and ecological function of the root microbiome across angiosperm plant species. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(6), E1157–E1165. https://doi.org/10.1073/pnas.1717617115

Friml, J., Vieten, A., Sauer, M., Weijers, D., Schwarz, H., Hamann, T., Offringa, R., & Jürgens, G. (2003). Efflux-dependent auxin gradients establish the apical-basal axis of Arabidopsis. *Nature*, *426*(6963), 147–153. https://doi.org/10.1038/nature02085

Gadjeva, M., Takahashi, K., & Thiel, S. (2004). Mannan-binding lectin - A soluble pattern recognition molecule. *Molecular Immunology*, *41*(2–3), 113–121. https://doi.org/10.1016/j.molimm.2004.03.015

Gallart, M., Adair, K. L., Love, J., Meason, D. F., Clinton, P. W., Xue, J., & Turnbull, M. H. (2018). Genotypic variation in Pinus radiata responses to nitrogen source are related to changes in the root microbiome. *FEMS Microbiology Ecology*, *94*(6). https://doi.org/10.1093/femsec/fiy071

Gao, R., Krysciak, D., Petersen, K., Utpatel, C., Knapp, A., Schmeisser, C., Daniel, R., Voget, S., Jaeger, K.-E., & Streit Biocenter Klein Flottbek, W. R. (2015). *Genome-Wide RNA Sequencing Analysis of Quorum Sensing-Controlled Regulons in the Plant-Associated Burkholderia glumae PG1 Strain*. https://doi.org/10.1128/AEM.01043-15

Gao, Z., Karlsson, I., Geisen, S., Kowalchuk, G., & Jousset, A. (2019). Protists: Puppet Masters of the Rhizosphere Microbiome. In *Trends in Plant Science* (Vol. 24, Issue 2, pp. 165–176). Elsevier Ltd. https://doi.org/10.1016/j.tplants.2018.10.011

GARDES, M., & BRUNS, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. *Molecular Ecology*, *2*(2), 113–118. https://doi.org/10.1111/j.1365-294X.1993.tb00005.x

Gilbert, B., & Levine, J. M. (2017). Ecological drift and the distribution of species diversity. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1855). https://doi.org/10.1098/rspb.2017.0507

Gilbert, S., Xu, J., Acosta, K., Poulev, A., Lebeis, S., & Lam, E. (2018). Bacterial Production of Indole Related Compounds Reveals Their Role in Association Between Duckweeds and Endophytes. *Frontiers in Chemistry*, *6*(JUL), 265. https://doi.org/10.3389/fchem.2018.00265

Gómez-Gómez, L., Felix, G., & Boller, T. (1999). A single locus determines sensitivity to bacterial flagellin in Arabidopsis thaliana. *The Plant Journal*, *18*(3), 277–284. https://doi.org/10.1046/j.1365-313X.1999.00451.x

Gómez, P., Paterson, S., De Meester, L., Liu, X., Lenzi, L., Sharma, M. D., McElroy, K., &

Buckling, A. (2016). Local adaptation of a bacterium is as important as its presence in structuring a natural microbial community. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms12453

González, E., Solano, R., Rubio, V., Leyva, A., & Paz-Ares, J. (2005). PHOSPHATE TRANSPORTER TRAFFIC FACILITATOR1 is a plant-specific SEC12-related protein that enables the endoplasmic reticulum exit of a high-affinity phosphate transporter in Arabidopsis. *Plant Cell*, *17*(12), 3500–3512. https://doi.org/10.1105/tpc.105.036640

Gordon, S. A., & Weber, R. P. (1951). COLORIMETRIC ESTIMATION OF INDOLEACETIC ACID. *Plant Physiology*, *26*(1), 192–195. https://doi.org/10.1104/pp.26.1.192

Gottel, N. R., Castro, H. F., Kerley, M., Yang, Z., Pelletier, D. A., Podar, M., Karpinets, T., Uberbacher, E., Tuskan, G. A., Vilgalys, R., Doktycz, M. J., & Schadt, C. W. (2011). Distinct Microbial Communities within the Endosphere and Rhizosphere of Populus deltoides Roots across Contrasting Soil Types †. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, *77*(17), 5934–5944. https://doi.org/10.1128/AEM.05255-11

Gould, S. J., & Vrba, E. S. (1982). Exaptation—a Missing Term in the Science of Form. *Paleobiology*, *1*(N/A), 4–15. https://doi.org/10.1017/S0094837300004310

Grainger, T. N., Letten, A. D., Gilbert, B., & Fukami, T. (2019). Applying modern coexistence theory to priority effects. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(13), 6205–6210. https://doi.org/10.1073/pnas.1803122116

Gransee, A., & Merbach, W. (2000). Phosphorus dynamics in a long-term P fertilization trial on Luvic Phaeozem at Halle. *Journal of Plant Nutrition and Soil Science*, *163*(4), 353–357. https://doi.org/10.1002/1522-2624(200008)163:4<353::AID-JPLN353>3.0.CO;2-B

Griffiths, R. I., Thomson, B. C., James, P., Bell, T., Bailey, M., & Whiteley, A. S. (2011). The bacterial biogeography of British soils. *Environmental Microbiology*, *13*(6), 1642–1654. https://doi.org/10.1111/j.1462-2920.2011.02480.x

Grinberg, M., Orevi, T., Steinberg, S., & Kashtan, N. (2019). Bacterial survival in microscopic surface wetness. *ELife*, *8*. https://doi.org/10.7554/eLife.48508

Hacquard, S., Garrido-Oter, R., González, A., Spaepen, S., Ackermann, G., Lebeis, S., McHardy, A. C., Dangl, J. L., Knight, R., Ley, R., & Schulze-Lefert, P. (2015). Microbiota and host nutrition across plant and animal kingdoms. In *Cell Host and Microbe* (Vol. 17, Issue 5, pp. 603–616). Cell Press. https://doi.org/10.1016/j.chom.2015.04.009

Hacquard, S., Kracher, B., Hiruma, K., Münch, P. C., Garrido-Oter, R., Thon, M. R., Weimann, A., Damm, U., Dallery, J. F., Hainaut, M., Henrissat, B., Lespinet, O., Sacristán, S., Ver Loren Van Themaat, E., Kemen, E., McHardy, A. C., Schulze-Lefert, P., & O'Connell, R. J. (2016). Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms11362

Hacquard, S., Spaepen, S., Garrido-Oter, R., & Schulze-Lefert, P. (2017). Interplay between Innate Immunity and the Plant Microbiota. *Annual Review of Phytopathology*, *55*, 565–589. https://doi.org/10.1146/annurev-phyto-080516-035623

Hadjithomas, M., Chen, I. M. A., Chu, K., Ratner, A., Palaniappan, K., Szeto, E., Huang, J., Reddy, T. B. K., Cimermančič, P., Fischbach, M. A., Ivanova, N. N., Markowitz, V. M., Kyrpides, N. C., & Pati, A. (2015). IMG-ABC: A knowledge base to fuel discovery of biosynthetic gene clusters and novel secondary metabolites. *MBio*, *6*(4). https://doi.org/10.1128/mBio.00932-15

Haft, D. H., Selengut, J. D., & White, O. (2003). The TIGRFAMs database of protein families. In *Nucleic Acids Research* (Vol. 31, Issue 1, pp. 371–373). https://doi.org/10.1093/nar/gkg128

Hall, A. E., Findell, J. L., Schaller, G. E., Sisler, E. C., & Bleecker, A. B. (2000). Ethylene perception by the ERS1 protein in Arabidopsis. *Plant Physiology*, *123*(4), 1449–1457. https://doi.org/10.1104/pp.123.4.1449

Haney, C. H., Samuel, B. S., Bush, J., & Ausubel, F. M. (2015). Associations with rhizosphere bacteria can confer an adaptive advantage to plants. *Nature Plants*, *1*(6). https://doi.org/10.1038/nplants.2015.51

Hardoim, P. R., van Overbeek, L. S., Berg, G., Pirttilä, A. M., Compant, S., Campisano, A., Döring, M., & Sessitsch, A. (2015). The Hidden World within Plants: Ecological and Evolutionary Considerations for Defining Functioning of Microbial Endophytes. *Microbiology and Molecular Biology Reviews*, *79*(3), 293–320. https://doi.org/10.1128/mmbr.00050-14

Hartmann, A., Rothballer, M., & Schmid, M. (2008). Lorenz Hiltner, a pioneer in rhizosphere microbial ecology and soil bacteriology research. *Plant and Soil*, *312*(1–2), 7–14. https://doi.org/10.1007/s11104-007-9514-z

Hernández, M., Dumont, M. G., Yuan, Q., & Conrad, R. (2015). Different bacterial populations associated with the roots and rhizosphere of rice incorporate plant-derived carbon. *Applied and Environmental Microbiology*, *81*(6), 2244–2253. https://doi.org/10.1128/AEM.03209-

Herrera Paredes, S., Gao, T., Law, T. F., Finkel, O. M., Mucyn, T., Teixeira, P. J. P. L., Salas González, I., Feltcher, M. E., Powers, M. J., Shank, E. A., Jones, C. D., Jojic, V., Dangl, J. L., & Castrillo, G. (2018). Design of synthetic bacterial communities for predictable plant phenotypes. *PLoS Biology*, *16*(2). https://doi.org/10.1371/journal.pbio.2003962

Hershey, D. M., Lu, X., Zi, J., & Peters, R. J. (2014). Functional conservation of the capacity for ent-kaurene biosynthesis and an associated operon in certain rhizobia. *Journal of Bacteriology*, *196*(1), 100–106. https://doi.org/10.1128/JB.01031-13

Hiei, Y., Ohta, S., Komari, T., & Kumashiro, T. (1994). Efficient transformation of rice (Oryza sativa L.) mediated by Agrobacterium and sequence analysis of the boundaries of the T-DNA. *The Plant Journal*, *6*(2), 271–282. https://doi.org/10.1046/j.1365-313X.1994.6020271.x

Hiruma, K. (2019). Roles of plant-derived secondary metabolites during interactions with pathogenic and beneficial microbes under conditions of environmental stress. In *Microorganisms* (Vol. 7, Issue 9). MDPI AG. https://doi.org/10.3390/microorganisms7090362

Hiruma, K., Gerlach, N., Sacristán, S., Nakano, R. T., Hacquard, S., Kracher, B., Neumann, U., Ramírez, D., Bucher, M., O'Connell, R. J., & Schulze-Lefert, P. (2016). Root Endophyte Colletotrichum tofieldiae Confers Plant Fitness Benefits that Are Phosphate Status Dependent. *Cell*, *165*(2), 464–474. https://doi.org/10.1016/j.cell.2016.02.028

Ho, B. T., Dong, T. G., & Mekalanos, J. J. (2014). A view to a kill: The bacterial type VI secretion system. In *Cell Host and Microbe* (Vol. 15, Issue 1, pp. 9–21). Cell Press. https://doi.org/10.1016/j.chom.2013.11.008

Hogenhout, S. A., Van Der Hoorn, R. A. L., Terauchi, R., & Kamoun, S. (2009). Emerging concepts in effector biology of plant-associated organisms. In *Molecular Plant-Microbe Interactions* (Vol. 22, Issue 2, pp. 115–122). The American Phytopathological Society . https://doi.org/10.1094/MPMI-22-2-0115

Hu, J., Wei, Z., Friman, V. P., Gu, S. H., Wang, X. F., Eisenhauer, N., Yang, T. J., Ma, J., Shen, Q. R., Xu, Y. C., & Jousset, A. (2016). Probiotic diversity enhances rhizosphere microbiome function and plant disease suppression. *MBio*, *7*(6). https://doi.org/10.1128/mBio.01790-16

Huang, A. C., Jiang, T., Liu, Y. X., Bai, Y. C., Reed, J., Qu, B., Goossens, A., Nützmann, H. W., Bai, Y., & Osbourn, A. (2019). A specialized metabolic network selectively modulates Arabidopsis root microbiota. *Science*, *364*(6440). https://doi.org/10.1126/science.aau6389

Hueck, C. J. (1998). Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants. In *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS* (Vol. 62, Issue 2). http://mmbr.asm.org/

Hultman, J., Waldrop, M. P., Mackelprang, R., David, M. M., McFarland, J., Blazewicz, S. J., Harden, J., Turetsky, M. R., McGuire, A. D., Shah, M. B., VerBerkmoes, N. C., Lee, L. H., Mavrommatis, K., & Jansson, J. K. (2015). Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature*, *521*(7551), 208–212. https://doi.org/10.1038/nature14238

Humphrey, P. T., & Whiteman, N. K. (2020). Insect herbivory reshapes a native leaf microbiome. *Nature Ecology and Evolution*, *4*(2), 221–229. https://doi.org/10.1038/s41559-019-1085-x

Huntemann, M., Ivanova, N. N., Mavromatis, K., James Tripp, H., Paez-Espino, D., Palaniappan, K., Szeto, E., Pillay, M., Chen, I. M. A., Pati, A., Nielsen, T., Markowitz, V. M., & Kyrpides, N. C. (2015). The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Standards in Genomic Sciences*, *10*(1). https://doi.org/10.1186/s40793-015-0077-y

Hurtado-McCormick, V., Kahlke, T., Petrou, K., Jeffries, T., Ralph, P. J., & Seymour, J. R. (2019). Regional and microenvironmental scale characterization of the zostera muelleri seagrass microbiome. *Frontiers in Microbiology*, *10*(MAY), 1011. https://doi.org/10.3389/fmicb.2019.01011

Ives, A. R., & Garland, T. (2010). Phylogenetic logistic regression for binary dependent variables. *Systematic Biology*, *59*(1), 9–26. https://doi.org/10.1093/sysbio/syp074

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457–D462. https://doi.org/10.1093/nar/gkv1070

Karasov, T. L., Almario, J., Friedemann, C., Ding, W., Giolai, M., Heavens, D., Kersten, S., Lundberg, D. S., Neumann, M., Regalado, J., Neher, R. A., Kemen, E., & Weigel, D. (2018). Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. *Cell Host and Microbe*, *24*(1), 168-179.e4. https://doi.org/10.1016/j.chom.2018.06.011

Kassen, R., Buckling, A., Bell, G., & Ralney, P. B. (2000). Diversity peaks at intermediate productivity in a laboratory microcosm. *Nature*, *406*(6795), 508–512. https://doi.org/10.1038/35020060

Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317

Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D., & Penin, A. A. (2016). A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *The Plant Journal*, *88*(6), 1058–1070. https://doi.org/10.1111/tpj.13312

Knief, C., Delmotte, N., Chaffron, S., Stark, M., Innerebner, G., Wassmann, R., Von Mering, C., & Vorholt, J. A. (2012). Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *ISME Journal*, *6*(7), 1378–1390. https://doi.org/10.1038/ismej.2011.192

Koskella, B., & Taylor, T. B. (2018). Multifaceted Impacts of Bacteriophages in the Plant Microbiome. *Annual Review of Phytopathology*, *56*(1), 361–380. https://doi.org/10.1146/annurev-phyto-080417-045858

Koyama, A., Maherali, H., & Antunes, P. M. (2019). Plant geographic origin and phylogeny as potential drivers of community structure in root-inhabiting fungi. *Journal of Ecology*, *107*(4), 1720–1736. https://doi.org/10.1111/1365-2745.13143

Kremer, J. M., Paasch, B. C., Rhodes, D., Thireault, C., Froehlich, J. E., Schulze-Lefert, P., Tiedje, J. M., & He, S. Y. (2018). FlowPot axenic plant growth system for microbiota research. In *bioRxiv* (p. 254953). bioRxiv. https://doi.org/10.1101/254953

Krug, L., Morauf, C., Donat, C., Müller, H., Cernava, T., & Berg, G. (2020). Plant Growth-Promoting Methylobacteria Selectively Increase the Biomass of Biotechnologically Relevant Microalgae. *Frontiers in Microbiology*, *11*, 427. https://doi.org/10.3389/fmicb.2020.00427

Krüger, U. S., Bak, F., Aamand, J., Nybroe, O., Badawi, N., Smets, B. F., & Dechesne, A. (2018). Novel method reveals a narrow phylogenetic distribution of bacterial dispersers in

environmental communities exposed to lowhydration conditions. *Applied and Environmental Microbiology*, *84*(7). https://doi.org/10.1128/AEM.02857-17

Laforest-Lapointe, I., Messier, C., & Kembel, S. W. (2016). Tree phyllosphere bacterial communities: Exploring the magnitude of intra- and inter-individual variation among host species. *PeerJ*, *2016*(8), e2367. https://doi.org/10.7717/PEERJ.2367

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lauber, C. L., Hamady, M., Knight, R., & Fierer, N. (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology*, *75*(15), 5111–5120. https://doi.org/10.1128/AEM.00335-09

Leadbetter, J. R., & Greenberg, E. P. (2000). Metabolism of acyl-homoserine lactone quorum-sensing signals by Variovorax paradoxus. *Journal of Bacteriology*, *182*(24), 6921–6926. https://doi.org/10.1128/JB.182.24.6921-6926.2000

Leff, J. W., Jones, S. E., Prober, S. M., Barberán, A., Borer, E. T., Firn, J. L., Harpole, W. S., Hobbie, S. E., Hofmockel, K. S., Knops, J. M. H., McCulley, R. L., La Pierre, K., Risch, A. C., Seabloom, E. W., Schütz, M., Steenbock, C., Stevens, C. J., & Fierer, N. (2015). Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(35), 10967–10972. https://doi.org/10.1073/pnas.1508382112

Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, *44*(W1), W242–W245. https://doi.org/10.1093/nar/gkw290

Leveau, J. H. (2019). A brief from the leaf: latest research to inform our understanding of the phyllosphere microbiome. In *Current Opinion in Microbiology* (Vol. 49, pp. 41–49). Elsevier Ltd. https://doi.org/10.1016/j.mib.2019.10.002

Leveau, J. H. J., & Gerards, S. (2008). Discovery of a bacterial gene cluster for catabolism of the plant hormone indole 3-acetic acid. *FEMS Microbiology Ecology*, *65*(2), 238–250. https://doi.org/10.1111/j.1574-6941.2008.00436.x

Leveau, J. H. J., & Lindow, S. E. (2005). Utilization of the plant hormone indole-3-acetic acid for growth by Pseudomonas putida strain 1290. *Applied and Environmental Microbiology*,

71(5), 2365–2371. https://doi.org/10.1128/AEM.71.5.2365-2371.2005

Levy, A., Salas Gonzalez, I., Mittelviefhaus, M., Clingenpeel, S., Herrera Paredes, S., Miao, J., Wang, K., Devescovi, G., Stillman, K., Monteiro, F., Rangel Alvarez, B., Lundberg, D. S., Lu, T. Y., Lebeis, S., Jin, Z., McDonald, M., Klein, A. P., Feltcher, M. E., Rio, T. G., … Dangl, J. L. (2018). Genomic features of bacterial adaptation to plants. *Nature Genetics*, *50*(1), 138–150. https://doi.org/10.1038/s41588-017-0012-9

Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S., Schlegel, M. L., Tucker, T. A., Schrenzel, M. D., Knight, R., & Gordon, J. I. (2008). Evolution of mammals and their gut microbes. *Science*, *320*(5883), 1647–1651. https://doi.org/10.1126/science.1155725

Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*, *41*(10). https://doi.org/10.1093/nar/gkt214

Liu, F., Hewezi, T., Lebeis, S. L., Pantalone, V., Grewal, P. S., & Staton, M. E. (2019). Soil indigenous microbiome and plant genotypes cooperatively modify soybean rhizosphere microbiome assembly. *BMC Microbiology*, *19*(1), 201. https://doi.org/10.1186/s12866-019-1572-x

Logemann, J., Schell, J., & Willmitzer, L. (1987). Improved method for the isolation of RNA from plant tissues. *Analytical Biochemistry*, *163*(1), 16–20. https://doi.org/10.1016/0003-2697(87)90086-8

Long, S. R. (1989). Rhizobium-legume nodulation: Life together in the underground. In *Cell* (Vol. 56, Issue 2, pp. 203–214). Elsevier. https://doi.org/10.1016/0092-8674(89)90893-3

Louca, S., Hawley, A. K., Katsev, S., Torres-Beltran, M., Bhatia, M. P., Kheirandish, S., Michiels, C. C., Capelle, D., Lavik, G., Doebeli, M., Crowe, S. A., & Hallam, S. J. (2016). Integrating biogeochemistry with multiomic sequence information in a model oxygen minimum zone. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(40), E5925–E5933. https://doi.org/10.1073/pnas.1602897113

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12). https://doi.org/10.1186/s13059-014-0550-8

Ludwig-Müller, J. (2015). Bacteria and fungi controlling plant growth by manipulating auxin:

Balance between development and defense. In *Journal of Plant Physiology* (Vol. 172, pp. 4?12-12). Urban und Fischer Verlag GmbH und Co. KG. https://doi.org/10.1016/j.jplph.2014.01.002

Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., Tremblay, J., Engelbrektson, A., Kunin, V., del Rio, T. G., Edgar, R. C., Eickhorst, T., Ley, R. E., Hugenholtz, P., Tringe, S. G., & Dangl, J. L. (2012). Defining the core Arabidopsis thaliana root microbiome. *Nature*, *488*(7409), 86–90. https://doi.org/10.1038/nature11237

Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., Tremblay, J., Engelbrektson, A., Kunin, V., Rio, T. G. Del, Edgar, R. C., Eickhorst, T., Ley, R. E., Hugenholtz, P., Tringe, S. G., & Dangl, J. L. (2012). Defining the core Arabidopsis thaliana root microbiome. *Nature*, *488*(7409), 86–90. https://doi.org/10.1038/nature11237

Ma, Q. H., Tian, B., & Li, Y. L. (2010). Overexpression of a wheat jasmonate-regulated lectin increases pathogen resistance. *Biochimie*, *92*(2), 187–193. https://doi.org/10.1016/j.biochi.2009.11.008

MacIntyre, D. L., Miyata, S. T., Kitaoka, M., & Pukatzki, S. (2010). The Vibrio cholerae type VI secretion system displays antimicrobial properties. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(45), 19520–19524. https://doi.org/10.1073/pnas.1012931107

Maignien, L., DeForce, E. A., Chafee, M. E., Murat Eren, A., & Simmons, S. L. (2014). Ecological succession and stochastic variation in the assembly of Arabidopsis thaliana phyllosphere communities. *MBio*, *5*(1). https://doi.org/10.1128/mBio.00682-13

Mantel, N. (1967). The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research*, *27*(2 Part 1).

Massoni, J., Bortfeld-Miller, M., Jardillier, L., Salazar, G., Sunagawa, S., & Vorholt, J. A. (2020). Consistent host and organ occupancy of phyllosphere bacteria in a community of wild herbaceous plant species. *ISME Journal*, *14*(1), 245–258. https://doi.org/10.1038/s41396-019-0531-8

Meaden, S., Metcalf, C. J. E., & Koskella, B. (2016). The effects of host age and spatial location on bacterial community composition in the English oak tree (Quercus robur). *Environmental Microbiology Reports*, *8*(5), 649–658. https://doi.org/10.1111/1758-2229.12418

Mendes, R., Kruijt, M., De Bruijn, I., Dekkers, E., Van Der Voort, M., Schneider, J. H. M., Piceno, Y. M., DeSantis, T. Z., Andersen, G. L., Bakker, P. A. H. M., & Raaijmakers, J. M. (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science*, *332*(6033), 1097–1100. https://doi.org/10.1126/science.1203980

Merbach, W., Garz, J., Schliephake, W., Stumpe, H., & Schmidt, L. (2000). The long-term fertilization experiments in Halle (Saale), Germany — Introduction and survey. *Journal of Plant Nutrition and Soil Science*, *163*(6), 629–638. https://doi.org/10.1002/1522-2624(200012)163:6<629::AID-JPLN629>3.0.CO;2-P

Miller, E. T., Svanbäck, R., & Bohannan, B. J. M. (2018). Microbiomes as Metacommunities: Understanding Host-Associated Microbes through Metacommunity Ecology. In *Trends in Ecology and Evolution* (Vol. 33, Issue 12, pp. 926–935). Elsevier Ltd. https://doi.org/10.1016/j.tree.2018.09.002

Morcillo, R. J., Singh, S. K., He, D., An, G., Vílchez, J. I., Tang, K., Yuan, F., Sun, Y., Shao, C., Zhang, S., Yang, Y., Liu, X., Dang, Y., Wang, W., Gao, J., Huang, W., Lei, M., Song, C., Zhu, J., … Zhang, H. (2020). Rhizobacterium-derived diacetyl modulates plant immunity in a phosphate-dependent manner. *The EMBO Journal*, *39*(2). https://doi.org/10.15252/embj.2019102602

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. https://doi.org/10.1111/mec.14542

Mullan, A., Quinn, J. P., & McGrath, J. W. (2002). Enhanced phosphate uptake and polyphosphate accumulation in Burkholderia cepacia grown under low-pH conditions. *Microbial Ecology*, *44*(1), 69–77. https://doi.org/10.1007/s00248-002-3004-x

Mylona, P., Pawlowski, K., & Bisseling, T. (1995). Symbiotic nitrogen fixation. In *Plant Cell* (Vol. 7, Issue 7, pp. 869–885). American Society of Plant Physiologists. https://doi.org/10.1105/tpc.7.7.869

Nagpal, P., Walker, L. M., Young, J. C., Sonawala, A., Timpte, C., Estelle, M., & Reed, J. W. (2000). AXR2 encodes a member of the Aux/IAA protein family. *Plant Physiology*, *123*(2), 563–573. https://doi.org/10.1104/pp.123.2.563

Naylor, D., Degraaf, S., Purdom, E., & Coleman-Derr, D. (2017). Drought and host selection influence bacterial community dynamics in the grass root microbiome. *ISME Journal*, *11*(12), 2691–2704. https://doi.org/10.1038/ismej.2017.118

Nett, R. S., Montanares, M., Marcassa, A., Lu, X., Nagel, R., Charles, T. C., Hedden, P., Rojas, M. C., & Peters, R. J. (2017). Elucidation of gibberellin biosynthesis in bacteria reveals convergent evolution. *Nature Chemical Biology*, *13*(1), 69–74. https://doi.org/10.1038/nchembio.2232

Newcombe, G., Harding, A., Ridout, M., & Busby, P. E. (2018). A Hypothetical Bottleneck in the Plant Microbiome. *Frontiers in Microbiology*, *9*, 1645. https://doi.org/10.3389/fmicb.2018.01645

Nichols, D., Cahoon, N., Trakhtenberg, E. M., Pham, L., Mehta, A., Belanger, A., Kanigan, T., Lewis, K., & Epstein, S. S. (2010). Use of ichip for high-throughput in situ cultivation of "uncultivable microbial species▽. *Applied and Environmental Microbiology*, *76*(8), 2445–2450. https://doi.org/10.1128/AEM.01754-09

Nilsson, R. H., Larsson, K. H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F. O., Tedersoo, L., Saar, I., Kõljalg, U., & Abarenkov, K. (2019). The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research*, *47*(D1), D259–D264. https://doi.org/10.1093/nar/gky1022

Ofek-Lalzar, M., Sela, N., Goldman-Voronov, M., Green, S. J., Hadar, Y., & Dror Minz, &. (2014). *ARTICLE Niche and host-associated functional signatures of the root surface microbiome*. https://doi.org/10.1038/ncomms5950

Osborn, M. J., Rosen, S. M., Rothfield, L., Zeleznick, L. D., & Horecker, B. L. (1964). Lipopolysaccharide of the gram-negative cell wall. In *Science* (Vol. 145, Issue 3634, pp. 783–789). American Association for the Advancement of Science. https://doi.org/10.1126/science.145.3634.783

Pant, B. D., Pant, P., Erban, A., Huhman, D., Kopka, J., & Scheible, W. R. (2015). Identification of primary and secondary metabolites with phosphorus status-dependent abundance in Arabidopsis, and of the transcription factor PHR1 as a major regulator of metabolic changes during phosphorus limitation. *Plant Cell and Environment*, *38*(1), 172–187. https://doi.org/10.1111/pce.12378

Pantigoso, H. A., Manter, D. K., & Vivanco, J. M. (2018). Phosphorus addition shifts the microbial community in the rhizosphere of blueberry (Vaccinium corymbosum L.). *Rhizosphere*, *7*, 1–7. https://doi.org/10.1016/j.rhisph.2018.06.008

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in

R language. *Bioinformatics*, *20*(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Pascale, A., Proietti, S., Pantelides, I. S., & Stringlis, I. A. (2020). Modulation of the Root Microbiome by Plant Molecules: The Basis for Targeted Disease Suppression and Plant Growth Promotion. In *Frontiers in Plant Science* (Vol. 10). Frontiers Media S.A. https://doi.org/10.3389/fpls.2019.01741

Peiffer, J. A., Spor, A., Koren, O., Jin, Z., Tringe, S. G., Dangl, J. L., Buckler, E. S., & Ley, R. E. (2013). Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6548–6553. https://doi.org/10.1073/pnas.1302837110

Pelletier, F., Garant, D., & Hendry, A. P. (2009). Eco-evolutionary dynamics. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 364, Issue 1523, pp. 1483–1489). Royal Society. https://doi.org/10.1098/rstb.2009.0027

Peters, N. K., Frost, J. W., & Long, S. R. (1986). A plant flavone, luteolin, induces expression of Rhizobium meliloti nodulation genes. *Science*, *233*(4767), 977–980. https://doi.org/10.1126/science.3738520

Pfeilmeier, S., Caly, D. L., & Malone, J. G. (2016). Bacterial pathogenesis of plants: future challenges from a microbial perspective. *Molecular Plant Pathology*, *17*(8), 1298–1313. https://doi.org/10.1111/mpp.12427

Pickles, B. J., Wilhelm, R., Asay, A. K., Hahn, A. S., Simard, S. W., & Mohn, W. W. (2017). Transfer of 13C between paired Douglas-fir seedlings reveals plant kinship effects and uptake of exudates by ectomycorrhizas. *New Phytologist*, *214*(1), 400–411. https://doi.org/10.1111/nph.14325

Poggio, S., Abreu-Goodger, C., Fabela, S., Osorio, A., Dreyfus, G., Vinuesa, P., & Camarena, L. (2007). A complete set of flagellar genes acquired by horizontal transfer coexists with the endogenous flagellar system in Rhodobacter sphaeroides. *Journal of Bacteriology*, *189*(8), 3208–3216. https://doi.org/10.1128/JB.01681-06

Porter, S. S., Chang, P. L., Conow, C. A., Dunham, J. P., & Friesen, M. L. (2017). Association mapping reveals novel serpentine adaptation gene clusters in a population of symbiotic

Mesorhizobium. *ISME Journal*, *11*(1), 248–262. https://doi.org/10.1038/ismej.2016.88

Price, C. T., Al-Khodor, S., Al-Quadan, T., Santic, M., Habyarimana, F., Kalia, A., & Kwaik, Y. A. (2009). Molecular Mimicry by an F-Box Effector of Legionella pneumophila Hijacks a Conserved Polyubiquitination Machinery within Macrophages and Protozoa. *PLoS Pathogens*, *5*(12), e1000704. https://doi.org/10.1371/journal.ppat.1000704

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE*, *5*(3), e9490. https://doi.org/10.1371/journal.pone.0009490

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. https://doi.org/10.1093/nar/gks1219

Raghothama, K. G. (1999). Phosphate acquisition. *Annual Review of Plant Biology*, *50*, 665–693. https://doi.org/10.1146/annurev.arplant.50.1.665

Ramirez, K. S., Craine, J. M., & Fierer, N. (2012). Consistent effects of nitrogen amendments on soil microbial communities and processes across biomes. *Global Change Biology*, *18*(6), 1918–1927. https://doi.org/10.1111/j.1365-2486.2012.02639.x

Ravcheev, D. A., Khoroshkin, M. S., Laikova, O. N., Tsoy, O. V., Sernova, N. V., Petrova, S. A., Rakhmaninova, A. B., Novichkov, P. S., Gelfand, M. S., & Rodionov, D. A. (2014). Comparative genomics and evolution of regulons of the LacI-family transcription factors. *Frontiers in Microbiology*, *5*(JUN), 1–1. https://doi.org/10.3389/fmicb.2014.00294

Rebolleda-Gómez, M., Forrester, N. J., Russell, A. L., Wei, N., Fetters, A. M., Stephens, J. D., & Ashman, T. (2019). Gazing into the anthosphere: considering how microbes influence floral evolution. *New Phytologist*, *224*(3), 1012–1020. https://doi.org/10.1111/nph.16137

Rebolleda Gómez, M., & Ashman, T. (2019). Floral organs act as environmental filters and interact with pollinators to structure the yellow monkeyflower (*Mimulus guttatus*) floral microbiome. *Molecular Ecology*, *28*(23), 5155–5171. https://doi.org/10.1111/mec.15280

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A., Dodsworth, J. A., Hedlund, B. P., Tsiamis, G., Sievert, S. M., Liu, W.-T., Eisen, J. A., Hallam, S. J., Kyrpides, N. C., Stepanauskas, R., … Woyke, T. (2013). *ARTICLE Insights into the phylogeny and coding potential of microbial dark matter*. https://doi.org/10.1038/nature12352

Robbins, C., Thiergart, T., Hacquard, S., Garrido-Oter, R., Gans, W., Peiter, E., Schulze-Lefert, P., & Spaepen, S. (2018). Root-Associated Bacterial and Fungal Community Profiles of *Arabidopsis thaliana* Are Robust Across Contrasting Soil P Levels. *Phytobiomes Journal*, *2*(1), 24–34. https://doi.org/10.1094/PBIOMES-09-17-0042-R

Robles, L., Stepanova, A., & Alonso, J. (2013). Molecular Mechanisms of Ethylene-Auxin Interaction. *Molecular Plant*, *6*, 1734–1737. https://doi.org/10.1093/mp/sst113

Rochefort, A., Briand, M., Marais, C., Wagner, M.-H., Laperche, A., Vallée, P., Barret, M., & Sarniguet, A. (2019). Influence of Environment and Host Plant Genotype on the Structure and Diversity of the *Brassica napus* Seed Microbiota. *Phytobiomes Journal*, *3*(4), 326–336. https://doi.org/10.1094/PBIOMES-06-19-0031-R

Rothmeier, E., Pfaffinger, G., Hoffmann, C., Harrison, C. F., Grabmayr, H., Repnik, U., Hannemann, M., Wölke, S., Bausch, A., Griffiths, G., Müller-Taubenberger, A., Itzen, A., & Hilbi, H. (2013). Activation of Ran GTPase by a Legionella Effector Promotes Microtubule Polymerization, Pathogen Vacuole Motility and Infection. *PLoS Pathogens*, *9*(9), e1003598. https://doi.org/10.1371/journal.ppat.1003598

Ruvkun, G. B., Sundaresan, V., & Ausubel, F. M. (1982). Directed transposon Tn5 mutagenesis and complementation analysis of rhizobium meliloti symbiotic nitrogen fixation genes. *Cell*, *29*(2), 551–559. https://doi.org/10.1016/0092-8674(82)90171-4

Sahly, H., Ofek, I., Podschun, R., Brade, H., He, Y., Ullmann, U., & Crouch, E. (2002). Surfactant Protein D Binds Selectively to Klebsiella pneumoniae Lipopolysaccharides Containing Mannose-Rich O-Antigens . *The Journal of Immunology*, *169*(6), 3267–3274. https://doi.org/10.4049/jimmunol.169.6.3267

Santhanam, R., Luu, V. T., Weinhold, A., Goldberg, J., Oh, Y., & Baldwin, I. T. (2015). Native root-associated bacteria rescue a plant from a sudden-wilt disease that emerged during continuous cropping. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(36), E5013–E5120. https://doi.org/10.1073/pnas.1505765112

Santos-Medellín, C., Edwards, J., Liechty, Z., Nguyen, B., & Sundaresan, V. (2017). Drought stress results in a compartment-specific restructuring of the rice root-associated

microbiomes. *MBio*, *8*(4). https://doi.org/10.1128/mBio.00764-17

Sarris, P. F., Cevik, V., Dagdas, G., Jones, J. D. G., & Krasileva, K. V. (2016). Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biology*, *14*(1), 8. https://doi.org/10.1186/s12915-016-0228-7

Sarris, P. F., Duxbury, Z., Huh, S. U., Ma, Y., Segonzac, C., Sklenar, J., Derbyshire, P., Cevik, V., Rallapalli, G., Saucet, S. B., Wirthmueller, L., Menke, F. L. H., Sohn, K. H., & Jones, J. D. G. (2015). A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell*, *161*(5), 1089–1100. https://doi.org/10.1016/j.cell.2015.04.024

Scharf, B. E., Hynes, M. F., & Alexandre, G. M. (2016). Chemotaxis signaling systems in model beneficial plant–bacteria associations. *Plant Molecular Biology*, *90*(6), 549–559. https://doi.org/10.1007/s11103-016-0432-4

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. In *Nature Methods* (Vol. 9, Issue 7, pp. 676–682). Nat Methods. https://doi.org/10.1038/nmeth.2019

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541. https://doi.org/10.1128/AEM.01541-09

Searle, S. R., Speed, F. M., & Milliken, G. A. (1980). Population marginal means in the linear model: An alternative to least squares means. *American Statistician*, *34*(4), 216–221. https://doi.org/10.1080/00031305.1980.10483031

Sen, A., Daubin, V., Abrouk, D., Gifford, I., Berry, A. M., & Normand, P. (2014). Phylogeny of the class Actinobacteria revisited in the light of complete genomes. the orders 'Frankiales' and Micrococcales should be split into coherent entities: Proposal of Frankiales ord. nov., Geodermatophilales ord. nov., Acidothermales ord. nov. and Nakamurellales ord. nov. *International Journal of Systematic and Evolutionary Microbiology*, *64*(Pt_11), 3821–3832. https://doi.org/10.1099/ijs.0.063966-0

Shade, A., Jacques, M. A., & Barret, M. (2017). Ecological patterns of seed microbiome diversity, transmission, and assembly. In *Current Opinion in Microbiology* (Vol. 37, pp.

15–22). Elsevier Ltd. https://doi.org/10.1016/j.mib.2017.03.010

Shen, Z., Ruan, Y., Xue, C., Zhong, S., Li, R., & Shen, Q. (2015). Soils naturally suppressive to banana Fusarium wilt disease harbor unique bacterial communities. *Plant and Soil*, *393*(1–2), 21–33. https://doi.org/10.1007/s11104-015-2474-9

Shevchik, V. E., Robert-Baudouy, J., & Hugouvieux-Cotte-Pattat, N. (1997). Pectate lyase PelI of Erwinia chrysanthemi 3937 belongs to a new family. *Journal of Bacteriology*, *179*(23), 7321–7330. https://doi.org/10.1128/jb.179.23.7321-7330.1997

Snelders, N. C., Rovenich, H., Petti, G. C., Rocafort, M., Vorholt, J. A., Mesters, J. R., Seidl, M. F., Nijland, R., & Thomma, B. P. H. J. (2020). A plant pathogen utilizes effector proteins for microbiome manipulation. In *bioRxiv* (p. 2020.01.30.926725). bioRxiv. https://doi.org/10.1101/2020.01.30.926725

Sprent, J. I. (2008). 60Ma of legume nodulation. What's new? What's changing? *Journal of Experimental Botany*, *59*(5), 1081–1084. https://doi.org/10.1093/jxb/erm286

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stebbins, C. E., & Galaán, J. E. (2001). Structural mimicry in bacterial virulence. In *Nature* (Vol. 412, Issue 6848, pp. 701–705). Nature Publishing Group. https://doi.org/10.1038/35089000

Stringlis, I. A., Yu, K., Feussner, K., De Jonge, R., Van Bentum, S., Van Verk, M. C., Berendsen, R. L., Bakker, P. A. H. M., Feussner, I., & Pieterse, C. M. J. (2018). MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(22), E5213–E5222. https://doi.org/10.1073/pnas.1722335115

Sun, S.-L., Yang, W.-L., Fang, W.-W., Zhao, Y.-X., Guo, L., & Dai, Y.-J. (2018). The Plant Growth-Promoting Rhizobacterium Variovorax boronicumulans CGMCC 4969 Regulates the Level of Indole-3-Acetic Acid Synthesized from Indole-3-Acetonitrile . *Applied and Environmental Microbiology*, *84*(16), 298–316. https://doi.org/10.1128/aem.00298-18

Takase, T., Nakazawa, M., Ishikawa, A., Kawashima, M., Ichikawa, T., Takahashi, N., Shimada, H., Manabe, K., & Matsui, M. (2004). *ydk1-D* , an auxin-responsive *GH3* mutant that is involved in hypocotyl and root elongation. *The Plant Journal*, *37*(4), 471–483. https://doi.org/10.1046/j.1365-313X.2003.01973.x

Tans-Kersten, J., Huang, H., & Allen, C. (2001). Ralstonia solanacearum needs motility for invasive virulence on tomato. *Journal of Bacteriology*, *183*(12), 3597–3605. https://doi.org/10.1128/JB.183.12.3597-3605.2001

Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. In *Nucleic Acids Research* (Vol. 28, Issue 1). http://www.ncbi.nlm.nih.gov/COG

Thiergart, T., Durán, P., Ellis, T., Vannier, N., Garrido-Oter, R., Kemen, E., Roux, F., Alonso-Blanco, C., Ågren, J., Schulze-Lefert, P., & Hacquard, S. (2020). Root microbiota assembly and adaptive differentiation among European Arabidopsis populations. *Nature Ecology and Evolution*, *4*(1), 122–131. https://doi.org/10.1038/s41559-019-1063-3

Tian, Y., Zhao, Y., Wu, X., Liu, F., Hu, B., & Walcott, R. R. (2015). The type VI protein secretion system contributes to biofilm formation and seed-to-seedling transmission of Acidovorax citrulli on melon. *Molecular Plant Pathology*, *16*(1), 38–47. https://doi.org/10.1111/mpp.12159

Tkacz, A., Bestion, E., Bo, Z., Hortala, M., & Poole, P. S. (2020). Influence of plant fraction, soil, and plant species on microbiota: A multikingdom comparison. *MBio*, *11*(1). https://doi.org/10.1128/mBio.02785-19

Traore, S. M., Mcdowell, J. M., Beers, E. P., & Welbaum, G. E. (2014). *Characterization of Type Three Effector Genes of A. citrulli, the Causal Agent of Bacterial Fruit Blotch of Cucurbits*.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The Human Microbiome Project. In *Nature* (Vol. 449, Issue 7164, pp. 804–810). Nature Publishing Group. https://doi.org/10.1038/nature06244

uchida, naoyuki, takahashi, K., Iwasaki, rie, Yamada, ryotaro, Yoshimura, M., endo, takaho, Kimura, seisuke, Zhang, H., nomoto, M., tada, Y., Kinoshita, toshinori, Itami, K., Hagihara, shinya, & torii, K. (2018). Chemical hijacking of auxin signaling with an engineered auxin-tIr1 pair. *Nature CHeMICaL BIOLOGY |*, *14*. https://doi.org/10.1038/nCHeMBIO.2555

Uroz, S., Courty, P. E., & Oger, P. (2019). Plant Symbionts Are Engineers of the Plant-Associated Microbiome. In *Trends in Plant Science* (Vol. 24, Issue 10, pp. 905–916). Elsevier Ltd. https://doi.org/10.1016/j.tplants.2019.06.008

Vannette, R. L., & Fukami, T. (2017). Dispersal enhances beta diversity in nectar microbes. *Ecology Letters*, *20*(7), 901–910. https://doi.org/10.1111/ele.12787

Vargas, L., Brígida, A. B. S., Mota Filho, J. P., De Carvalho, T. G., Rojas, C. A., Vaneechoutte, D., Van Bel, M., Farrinelli, L., Ferreira, P. C. G., Vandepoele, K., & Hemerly, A. S. (2014). Drought tolerance conferred to sugarcane by association with gluconacetobacter diazotrophicus: A transcriptomic view of hormone pathways. *PLoS ONE*, *9*(12). https://doi.org/10.1371/journal.pone.0114744

Varghese, N. J., Mukherjee, S., Ivanova, N., Konstantinidis, K. T., Mavrommatis, K., Kyrpides, N. C., & Pati, A. (2015). Microbial species delineation using whole genome sequences. *Nucleic Acids Research*, *43*(14), 6761–6771. https://doi.org/10.1093/nar/gkv657

Veach, A. M., Morris, R., Yip, D. Z., Yang, Z. K., Engle, N. L., Cregger, M. A., Tschaplinski, T. J., & Schadt, C. W. (2019). Rhizosphere microbiomes diverge among Populus trichocarpa plant-host genotypes and chemotypes, but it depends on soil origin. *Microbiome*, *7*(1), 76. https://doi.org/10.1186/s40168-019-0668-8

Vessey, J. K. (2003). Plant growth promoting rhizobacteria as biofertilizers. In *Plant and Soil* (Vol. 255, Issue 2, pp. 571–586). Springer. https://doi.org/10.1023/A:1026037216893

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–5267. https://doi.org/10.1128/AEM.00062-07

Wang, X., Wei, Z., Yang, K., Wang, J., Jousset, A., Xu, Y., Shen, Q., & Friman, V. P. (2019). Phage combination therapies for bacterial wilt disease in tomato. *Nature Biotechnology*, *37*(12), 1513–1520. https://doi.org/10.1038/s41587-019-0328-3

Weidenbach, D., Esch, L., Möller, C., Hensel, G., Kumlehn, J., Höfle, C., Hückelhoven, R., & Schaffrath, U. (2016). Polarized Defense Against Fungal Pathogens Is Mediated by the Jacalin-Related Lectin Domain of Modular Poaceae-Specific Proteins. *Molecular Plant*, *9*(4), 514–527. https://doi.org/10.1016/j.molp.2015.12.009

Weller-Stuart, T., Toth, I., De Maayer, P., & Coutinho, T. (2017). Swimming and twitching motility are essential for attachment and virulence of *Pantoea ananatis* in onion seedlings. *Molecular Plant Pathology*, *18*(5), 734–745. https://doi.org/10.1111/mpp.12432

Weston, D. J., Pelletier, D. A., Morrell-Falvey, J. L., Tschaplinski, T. J., Jawdy, S. S., Lu, T.-Y., Allen, S. M., Melton, S. J., Martin, M. Z., Schadt, C. W., Karve, A. A., Chen, J.-G., Yang,

X., Doktycz, M. J., & Tuskan, G. A. (2012). Pseudomonas fluorescens Induces Strain-Dependent and Strain-Independent Host Plant Responses in Defense Networks, Primary Metabolism, Photosynthesis, and Fitness. */ 765 MPMI*, *25*(6), 765–778. https://doi.org/10.1094/MPMI

Wolf, A. B., Rudnick, M. B., de Boer, W., & Kowalchuk, G. A. (2015). Early colonizers of unoccupied habitats represent a minority of the soil bacterial community. *FEMS Microbiology Ecology*, *91*(5), 24. https://doi.org/10.1093/femsec/fiv024

Wu, M, Chatterji, S., & Eisen, J. A. (2012). Accounting For Alignment Uncertainty in Phylogenomics. *PLoS ONE*, *7*(1), 30288. https://doi.org/10.1371/journal.pone.0030288

Wu, Martin, & Eisen, J. A. (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome Biology*, *9*(10), R151. https://doi.org/10.1186/gb-2008-9-10-r151

Xiang, Y., Song, M., Wei, Z., Tong, J., Zhang, L., Xiao, L., Ma, Z., & Wang, Y. (2011). A jacalin-related lectin-like gene in wheat is a component of the plant defence system. *Journal of Experimental Botany*, *62*(15), 5471–5483. https://doi.org/10.1093/jxb/err226

Xu, L., Naylor, D., Dong, Z., Simmons, T., Pierroz, G., Hixson, K. K., Kim, Y. M., Zink, E. M., Engbrecht, K. M., Wang, Y., Gao, C., DeGraaf, S., Madera, M. A., Sievert, J. A., Hollingsworth, J., Birdseye, D., Scheller, H. V., Hutmacher, R., Dahlberg, J., … Coleman-Derr, D. (2018). Drought delays development of the sorghum root microbiome and enriches for monoderm bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(18), E4284–E4293. https://doi.org/10.1073/pnas.1717308115

Xu, R. Q., Blanvillain, S., Feng, J. X., Jiang, B. Le, Li, X. Z., Wei, H. Y., Kroj, T., Lauber, E., Roby, D., Chen, B., He, Y. Q., Lu, G. T., Tang, D. J., Vasse, J., Arlat, M., & Tang, J. L. (2008). AvrACXcc8004, a type III effector with a leucine-rich repeat domain from Xanthomonas campestris pathovar campestris confers avirulence in vascular tissues of Arabidopsis thaliana ecotype Col-0. *Journal of Bacteriology*, *190*(1), 343–355. https://doi.org/10.1128/JB.00978-07

Yamaji, Y., Maejima, K., Komatsu, K., Shiraishi, T., Okano, Y., Himeno, M., Sugawara, K., Neriya, Y., Minato, N., Miura, C., Hashimoto, M., & Namba, S. (2012). Lectin-mediated resistance impairs plant virus infection at the cellular level. *Plant Cell*, *24*(2), 778–793. https://doi.org/10.1105/tpc.111.093658

Yamauchi, Y., Hasegawa, A., Taninaka, A., Mizutani, M., & Sugimoto, Y. (2011). NADPH-dependent Reductases Involved in the Detoxification of Reactive Carbonyls in Plants * □ S. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, *286*(9), 6999–7009.

https://doi.org/10.1074/jbc.M110.202226

Yeoh, Y. K., Paungfoo-Lonhienne, C., Dennis, P. G., Robinson, N., Ragan, M. A., Schmidt, S., & Hugenholtz, P. (2016). The core root microbiome of sugarcanes cultivated under varying nitrogen fertilizer application. *Environmental Microbiology*, *18*(5), 1338–1351. https://doi.org/10.1111/1462-2920.12925

Yourstone, S. M., Lundberg, D. S., Dangl, J. L., & Jones, C. D. (2014). MT-Toolbox: Improved amplicon sequencing using molecule tags. *BMC Bioinformatics*, *15*(1). https://doi.org/10.1186/1471-2105-15-284

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS A Journal of Integrative Biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118

Yu, P., Wang, C., Baldauf, J. A., Tai, H., Gutjahr, C., Hochholdinger, F., & Li, C. (2018). Root type and soil phosphate determine the taxonomic landscape of colonizing fungi and the transcriptome of field-grown maize roots. *New Phytologist*, *217*(3), 1240–1253. https://doi.org/10.1111/nph.14893

Yuan, Z., Druzhinina, I. S., Labbé, J., Redman, R., Qin, Y., Rodriguez, R., Zhang, C., Tuskan, G. A., & Lin, F. (2016). Specialized microbiome of a halophyte and its role in helping non-host plants to withstand salinity. *Scientific Reports*, *6*. https://doi.org/10.1038/srep32467

Zhalnina, K., Dias, R., de Quadros, P. D., Davis-Richardson, A., Camargo, F. A. O., Clark, I. M., McGrath, S. P., Hirsch, P. R., & Triplett, E. W. (2014). Soil pH Determines Microbial Diversity and Composition in the Park Grass Experiment. *Microbial Ecology*, *69*(2), 395–406. https://doi.org/10.1007/s00248-014-0530-2

Zhang, Z., Wu, Y., Gao, M., Zhang, J., Kong, Q., Liu, Y., Ba, H., Zhou, J., & Zhang, Y. (2012). Disruption of PAMP-induced MAP kinase cascade by a pseudomonas syringae effector activates plant immunity mediated by the NB-LRR protein SUMM2. *Cell Host and Microbe*, *11*(3), 253–263. https://doi.org/10.1016/j.chom.2012.01.015

Zhao, X. F., Hao, Y. Q., Zhang, D. Y., & Zhang, Q. G. (2019). Local biotic interactions drive species-specific divergence in soil bacterial communities. *ISME Journal*, *13*(11), 2846–2855. https://doi.org/10.1038/s41396-019-0477-x

Zhou, J., & Ning, D. (2017). Stochastic Community Assembly: Does It Matter in Microbial Ecology? *Microbiology and Molecular Biology Reviews*, *81*(4).

https://doi.org/10.1128/mmbr.00002-17

Ziegler, J., Schmidt, S., Chutia, R., Müller, J., Böttcher, C., Strehmel, N., Scheel, D., & Abel, S. (2016). Non-targeted profiling of semi-polar metabolites in Arabidopsis root exudates uncovers a role for coumarin secretion and lignification during the local response to phosphate limitation. *Journal of Experimental Botany*, *67*(5), 1421–1432. https://doi.org/10.1093/jxb/erv539

Zúñiga, A., Poupin, M. J., Donoso, R., Ledger, T., Guiliani, N., Gutiérrez, R. A., & González, B. (2013). Quorum sensing and indole-3-acetic acid degradation play a role in colonization and plant growth promotion of arabidopsis thaliana by burkholderia phytofirmans PsJN. *Molecular Plant-Microbe Interactions*, *26*(5), 546–553. https://doi.org/10.1094/MPMI-10-12-0241-R