

# **Plant-associated microbial community diversity and network feature analysis**

Cumulative dissertation

for the attainment of the title of doctor  
in the Faculty of Mathematics and Natural Sciences  
at the Heinrich Heine University Düsseldorf

presented by

**Rui Guan**

from Shanxi, China

Düsseldorf, November 2022



from the institute for Computer Science  
at the Heinrich Heine University Düsseldorf

Published by permission of the  
Faculty of Mathematics and Natural Sciences at  
Heinrich Heine University Düsseldorf

Thesis reviewers:

Prof. Dr. Gunnar W. Klau

Prof. Dr. Laura Rose

Date of the oral examination: 04/05/2023



## **Statement of authorship**

I hereby certify that this dissertation is the result of my own work. No other person's work has been used without due acknowledgement. This dissertation has not been submitted in the same or similar form to other institutions. I have not previously failed a doctoral examination procedure.

Rui Guan



## Summary

Microbes are present in most of the environments on earth. By interacting with each other, macro-organisms, and the surrounding habitat, they form diverse microbial communities denoted as microbiota. In soil, microbes contribute largely to the genetic diversity of their ecosystem, and are involved in diverse processes, such as nutrient cycling and vegetation dynamics. Soil-derived microbial communities associate to plant hosts and form the plant microbiota, which promotes nutrient uptake and protects the roots from pathogens.

To understand microbiota composition and diversity, a variety of approaches have been developed, from classical strain isolation to culture-independent methods. By analysing high-throughput sequencing data derived from these experimental approaches, we can survey the microbial composition and determine the factors that affect the community assembly. In this thesis, I show how various bioinformatic tools can be applied to the analysis of the community profiling of natural and synthetic microbial communities, particularly of those associated with a plant host. Specifically, using the model plant *Arabidopsis thaliana*, we explored the mechanism of how root microbiota promotes nutrient uptake as well as the interplay between host innate immunity and microbiota regarding growth and defence. By comparing with other hosts, including the photosynthetic plant *Lotus japonicus* and model alga *Chlamydomonas reinhardtii*, we demonstrated host preference and shared features of their microbiota.

As one of the most widely cultivated crops, maize has been an important model organism for microbiota research to understand the effects of plant breeding on the establishment of root microbiota. Furthermore, the relationship between multiple kingdom root microbiota, abiotic factors such as soil management, and plant growth is still unclear. In order to explore these interactions, we characterized the root microbial communities of maize grown in two long-term experimental fields under four soil managements. The sampling spanned from the vegetative to reproductive growth stage and included four inbred lines with one phosphate transporter mutant *pht1;6*. We found that, at the phylum level, microbial communities from different root compartments converged for bacteria, but not fungi. We also identified stable root microbial taxa that persisted through the host growth and these taxa were accompanied by dynamic members that covaried with root metabolites. By comparing wild-type and mutant plants, we discovered a potential plant growth phase-specific interaction between arbuscular mycorrhizal fungal symbiosis, root lipid status, and soil phosphate availability. Together, our work sheds light on the spatio-temporal dynamics of maize root-associated microbiota and its impact on plant physiology and fitness.

To better investigate the biological meaning behind the increasing amount of plant microbiota data, we developed a novel diversity and network analysis workflow into an open-access R package named '*mina*'. We integrated a large-scale plant- and alga-associated microbiota dataset, to which we applied the developed workflow. We extracted the representative community compositions and inferred a co-occurrence network based on them. Higher-order features, namely clusters of connected microbes in the network, were introduced to diversity analysis and decreased the unexplained variance compared to traditional diversity measurements. To assist the comparative analysis of microbial networks, we established an approach that relies on the calculation of network spectral distances and Monte Carlo permutation significance tests. We differentiated networks

constructed from samples originating from various conditions and identified the features with the highest contribution to the network differentiation.

In summary, we show that by analysing microbial community profiling data, we gain insights into the assembly and function of plant microbiota. This is not limited to the natural conditions in ecological surveys but also applies to the reconstituted synthetic microbiota systems. With the novel diversity and network analysis tools that we developed, we can better describe microbiota diversity and determine distinctive features that drive the dynamics of microbial communities.

## Zusammenfassung

Mikroben sind in den meisten Lebensräumen der Erde zu finden. Indem sie miteinander, mit Makroorganismen und mit dem sie umgebenden Lebensraum interagieren, bilden sie vielfältige mikrobielle Gemeinschaften, die als Mikrobiota bezeichnet werden. Im Boden tragen die Mikroben in hohem Maße zur genetischen Vielfalt ihres Ökosystems bei und sind an verschiedenen Prozessen wie dem Nährstoffkreislauf und der Vegetationsdynamik beteiligt. Aus dem Boden stammende mikrobielle Gemeinschaften verbinden sich mit Pflanzenwirten und bilden die Pflanzenmikrobiota, die die Nährstoffaufnahme fördert und die Wurzeln vor Krankheitserregern schützt.

Um die Zusammensetzung und Vielfalt der Mikrobiota zu verstehen, wurde eine Vielzahl von Ansätzen entwickelt, von der klassischen Stammsolierung bis hin zu kulturunabhängigen Methoden. Durch die Analyse von Hochdurchsatz-Sequenzierungsdaten, die aus diesen experimentellen Ansätzen stammen, können wir die mikrobielle Zusammensetzung erfassen und die Faktoren bestimmen, die die Zusammensetzung der Gemeinschaft beeinflussen. In dieser Arbeit zeige ich, wie verschiedene bioinformatische Werkzeuge für die Analyse der Zusammensetzung natürlicher und synthetischer mikrobieller Gemeinschaften, insbesondere solcher, die mit einem pflanzlichen Wirt assoziiert sind, eingesetzt werden können. Anhand der Modellpflanze *Arabidopsis thaliana* untersuchten wir den Mechanismus, durch den die Wurzelmikrobiota die Nährstoffaufnahme fördert, sowie das Zusammenspiel zwischen der angeborenen Immunität des Wirts und der Mikrobiota in Bezug auf Wachstum und Abwehr. Durch Vergleiche mit anderen Wirten, einschließlich der photosynthetischen Pflanze *Lotus japonicus* und der Modellalge *Chlamydomonas reinhardtii*, konnten wir eine Wirtspräferenz assoziierter Bakterien und gemeinsame Merkmale ihrer Mikrobiota nachweisen.

Als eine der am weitesten verbreiteten Kulturpflanzen ist Mais ein wichtiger Modellorganismus für die Erforschung der Mikrobiota, um die Auswirkungen der Pflanzenzüchtung auf die Etablierung der Wurzelmikrobiota zu verstehen. Darüber hinaus ist der Zusammenhang zwischen der Wurzelmikrobiota mehrerer Reiche, abiotischen Faktoren wie der Bodenbewirtschaftung und dem Pflanzenwachstum noch unklar. Um diese Wechselwirkungen zu erforschen, haben wir die mikrobiellen Wurzelgemeinschaften von Maispflanzen charakterisiert, die in zwei Langzeitversuchsfeldern unter vier Bodenbewirtschaftungsformen angebaut wurden. Die Probenahme erstreckte sich von der vegetativen bis zur reproduktiven Wachstumsphase und umfasste vier Inzuchtlinien mit einer Phosphattransporter-Mutante *pht1;6*. Wir stellten fest, dass die mikrobiellen Gemeinschaften aus verschiedenen Wurzelkompartimenten auf Phylum-Ebene bei Bakterien, nicht aber bei Pilzen konvergieren. Wir identifizierten auch stabile mikrobielle Taxa in der Wurzel, die während des Wirtswachstums bestehen blieben und von dynamischen Taxa begleitet wurden, die aber mit Wurzelmetaboliten kovariierten. Durch den Vergleich von Wildtyp und Mutanten entdeckten wir eine potenzielle pflanzenwachstumsphasenspezifische Interaktion zwischen arbuskulärer Mykorrhizapilzsymbiose, Wurzellipidstatus und P-Verfügbarkeit im Boden. Unsere Arbeit beleuchtet die räumlich-zeitliche Dynamik der Maiswurzel-assoziierten Mikrobiota und ihre Auswirkungen auf die Pflanzenphysiologie und -fitness.

Um die biologische Bedeutung hinter der zunehmenden Menge an Pflanzenmikrobiota-Daten besser untersuchen zu können, haben wir einen neuartigen Diversitäts- und Netzwerkanalyse-Workflow in einem frei zugänglichen R-Paket namens 'mina' entwickelt. Wir haben einen großen pflanzen- und

algenassoziierten Mikrobiota-Datensatz integriert, auf den wir den entwickelten Arbeitsablauf angewendet haben. Wir extrahierten die repräsentativen Gemeinschaftszusammensetzungen und leiteten daraus ein *co-occurrence*-Netzwerk ab. Merkmale höherer Ordnung, nämlich Cluster miteinander verbundener Mikroben im Netzwerk, wurden in die Diversitätsanalyse eingeführt und verringerten die unerklärte Varianz im Vergleich zu herkömmlichen Diversitätsmessungen. Zur Unterstützung der vergleichenden Analyse von mikrobiellen Netzwerken haben wir einen Ansatz entwickelt, der auf der Berechnung von spektralen Netzwerkabständen und Monte-Carlo-Permutations-Signifikanztests beruht. Wir differenzierten Netzwerke, die aus Proben verschiedener Bedingungen gebildet wurden, und identifizierten die Merkmale mit dem höchsten Beitrag zur Netzwerkdifferenzierung.

Zusammenfassend lässt sich sagen, dass wir durch die Analyse von Daten zur Erstellung von Profilen mikrobieller Gemeinschaften Einblicke in den Aufbau und die Funktion der pflanzlichen Mikrobiota gewinnen können. Dies gilt nicht nur für die natürlichen Bedingungen in ökologischen Untersuchungen, sondern auch für die rekonstituierten synthetischen Mikrobiota-Systeme. Mit den neuartigen Diversitäts- und Netzwerkanalysewerkzeugen, die wir entwickelt haben, können wir die Diversität der Mikrobiota besser beschreiben und die charakteristischen Merkmale bestimmen, die die Dynamik der mikrobiellen Gemeinschaften bestimmen.

## List of figures and tables

Figure 2.1 Coumarin biosynthesis affects root microbiota composition in a naturally calcareous soil.

Figure 2.2 Coumarin biosynthesis is important for plant growth in a naturally calcareous soil.

Figure 2.3 Coumarin biosynthesis restructures the root microbiota at the ASV level.

Figure 2.4 Overnight growth of *Burkholderiaceae* bacterial strains in the presence of scopoletin or fraxetin.

Figure 2.5 Taxonomically diverse root commensals improve iron-limiting plant performance.

Figure 2.6 Iron-limiting growth rescue activity of SynCom strains in mono-association.

Figure 2.7 Microbiota-mediated plant growth rescue occurs via the reductive import of iron.

Figure 2.8 Plant biosynthesis and secretion of fraxetin is necessary for microbiota-mediated growth rescue.

Figure 2.9 Supplementation with fraxetin restores microbiota-mediated growth rescue of f6'h1 and s8h plants.

Figure 2.10 A bacterial SynCom improves plant iron nutrition and relieves the iron deficiency response.

Figure 2.11 The bacterial SynCom modulates a subset of defence genes in a coumarin-dependent manner.

Figure 3.1 *pWER::FLS2-GFP* transgenic *Arabidopsis* plants are hypersensitive to flg22.

Figure 3.2 *At-RSPHERE* root commensals exhibit strain-specific variations to suppress flg22-mediated RGI in *pWER::FLS2-GFP* plants.

Figure 3.3 Activation of immunity by flg22 affects community establishment.

Figure 3.4 SynCom colonization and flg22 treatment induce root transcriptomic changes in WT Col-0 plants.

Figure 3.5 Imbalance of specific bacteria impacts plant susceptibility to opportunistic *Pseudomonas* pathogens.

Figure 3.6 The balance between non-suppressive and suppressive strains integrates with plant innate immunity.

Figure 4.1 Diversity of *Lotus* and *Arabidopsis* root-associated bacterial communities.

Figure 4.2 Abundance and recovered *Lotus* and *Arabidopsis* root-associated bacterial OTUs.

Figure 4.3 Whole-genome phylogeny of the *Lotus* and *Arabidopsis* core culture collections.

Figure 4.4 Scheme of competition experiments.

Figure 4.5 Constrained PCoA and aggregated RA of the commensal communities in *Lj*- and *At*-associated root compartments.

Figure 4.6 Constrained PCoA and aggregated RA of the commensal communities in *Lotus* and *Arabidopsis* roots.

Figure 4.7 Constrained PCoA and aggregated RA of the commensal communities in *Lotus* and *Arabidopsis* roots and toothpick control.

Figure 4.8 SynCom-specific transcriptional outputs in *Lotus* and *Arabidopsis* roots.

Figure 4.9 Setup of the sequential inoculation experiment.

Figure 4.10 Invasion and persistence of commensal bacteria.

Figure 4.11 Host preference of individual commensal strains across gnotobiotic experiments.

Figure 4.12 Host preference is linked to invasiveness.

Figure 5.1 Schematic description of the greenhouse experiment.

Figure 5.2 Comparison of bacterial community structures associated with *At* roots and the *Cr* phycosphere in natural soil.

Figure 5.3 Culture-independent analysis of phycosphere- and root-associated communities in a natural soil.

Figure 5.4 Phylogeny of *16S* rRNA sequences of the most abundant OTUs found in *At* roots and *Cr* phycosphere community profiles.

Figure 5.5 Conservation of bacterial orders of the root and phycosphere microbiota across photosynthetic organisms.

Figure 5.6 Schematic description of the mesocosm experiments.

Figure 5.7 Mesocosm experiments recapitulate the establishment of phycosphere communities by *Cr* across soil types and growth media.

Figure 5.8 Phycosphere reconstitution using bacterial SynComs derived from the *Cr*-SPHERE core culture collection.

Figure 5.9 Schematic description of the cross-inoculation experiment.

Figure 5.10 Root and phycosphere bacteria colonize *At* and *Cr* and assemble into taxonomically equivalent communities.

Figure 5.11 Schematic description of the split co-cultivation system.

Figure 5.12 Physical proximity to *Cr* is required for the establishment of phycosphere bacterial communities.

Figure 6.1 Experimental design of field-grown maize microbiota survey.

Figure 6.2 Soil properties and microbial diversity of unplanted soil in long-term experimental fields.

Figure 6.3 Soil physicochemical properties, heterogeneity, and plant growth phase drive shifts in microbial soil biota composition.

Figure 6.4 Root compartment, plant growth phase, and soil management shape the root-associated microbiota in field-grown maize.

Figure 6.5 Stable bacterial and fungal OTUs are enriched from rhizosphere to root compartment over plant growth, irrespective to soil management.

Figure 6.6 Bacterial community profiles at the phylum level.

Figure 6.7 Plant growth phase shapes bacterial community at the phylum level.

Figure 6.8 Root metabolites and total element compositions are affected by plant growth phase.

Figure 6.9 The root associated microbiota covaries with root lipid profile over plant growth.

Figure 6.10 Genotype-dependent effect of phosphate depletion on root lipids, microbiota, and plant biomass.

Figure 7.1 Community diversity of plant-associated microbiota.

Figure 7.2 Taxonomic profiling of community structure at the phylum rank.

Figure 7.3 Occupancy and accumulative relative abundance of microbes in root-associated compartments.

Figure 7.4 Distance between the diversity of full and examined subset community members.

Figure 7.5 Density and degree distribution of the global network.

Figure 7.6 Network cluster-based community diversity analysis.

Figure 7.7 Permutation-based network comparison.

Figure 7.8 Overview of the R package '*mina*'.

Figure 7.9 Bootstrap-permutation network analysis of plant-associated microbiota.

Figure 7.10 Community diversity and taxa distribution of CAS microbiota.

Figure 7.11 Occupancy and the corresponding cumulated aRA distribution of taxa and network clusters.

Figure 7.12 Network distances between different CAS-associated conditions.

Figure 7.13 Distance decrease of each permutation group compared to the actual network distance.

Figure 7.14 Correlation between partially permutation distance change and other features.

Figure 7.15 Network distance change for all inter-condition comparisons.

Figure 7.16 Pre-processing of bacterial amplicon sequencing data.

Figure 7.17 Pre-processing of fungal ITS reads by pair-end sequencing.

Table 7.1 Dataset overview.

Table 7.2 Unexplained variance of ASV- and NCL-based diversity analyses.

Table 7.3 Sample number of CAS-derived conditions.



Summary	I
Zusammenfassung	III
List of figures and tables	V
<b>Chapter 1 Introduction</b>	
1.1 Introduction to microbiota studies	1
1.1.1 Microbes and microbial communities	1
1.1.2 Microbiota research: from isolates to communities	2
1.1.3 Deconstruction and reconstruction of microbiota	3
1.2 Plant-associated microbiota research	4
1.2.1 Plant-associated microbial communities	4
1.2.2 Plant-microbiota interactions	5
1.2.3 Synthetic plant-associated microbial communities	6
1.3 Data processing and analysis of microbial community surveys	7
1.3.1 Processing of community profiling data	7
1.3.2 Microbial community diversity analysis	8
1.3.3 Microbial community network analysis	8
1.4 Thesis overview	10
1.4.1 Outline	10
1.4.2 Publications and author contributions	10
<b>Chapter 2 Root-secreted coumarins and the microbiota interact to improve iron nutrition in Arabidopsis</b>	
2.1 Abstract	11
2.2 Introduction	11
2.3 Results	12
2.3.1 Coumarin biosynthesis is important for plant growth and root microbiota composition in iron-limiting soil	12
2.3.2 Coumarin biosynthesis restructures the root microbiota at the ASV level	14
2.3.3 Taxonomically diverse root commensals improve iron-limiting plant performance	16
2.3.4 Microbiota-mediated plant growth rescue occurs via reductive import of iron and requires fraxetin secretion	19
2.3.5 Coumarins and the microbiota interact to alleviate iron starvation	22
2.4 Discussion	25
2.5 Materials and methods	26
2.5.1 Experimental model and subject details	26
2.5.2 Plant growth conditions	27
2.5.3 Coumarin antimicrobial activity	27
2.5.4 Gnotobiotic system for iron limitation	27

2.5.5 Microbial community data analysis	28
2.5.6 RNA extraction and RNA-seq analysis	29
2.6 Data and code availability	30
2.7 Author contributions	30

## Chapter 3 Coordination of microbe-host homeostasis by crosstalk with plant innate immunity

3.1 Abstract	31
3.2 Introduction	31
3.3 Results	32
3.3.1 Taxonomically widespread ability of root commensals to interfere with defense-associated growth inhibition	32
3.3.2 Activation of immunity shapes root microbiota establishment	35
3.3.3 Root transcriptomic changes and dampening of immunity by suppressive SynComs	37
3.3.4 Suppressive and non-suppressive commensals differentially impact plant susceptibility to opportunistic pathogens	40
3.4 Discussion	42
3.5 Materials and methods	44
3.5.1 Growth conditions for plants and culture conditions for bacteria	44
3.5.2 Screening for RGI suppressive strains in monoassociation	44
3.5.3 Microbial community profiling and data processing	45
3.6 Data and code availability	46
3.7 Author contributions	46

## Chapter 4 Host preference and invasiveness of commensals in the Lotus and Arabidopsis root microbiota

4.1 Abstract	47
4.2 Introduction	47
4.3 Results	48
4.3.1 Host-species-specific bacterial culture collections	48
4.3.2 Host-preference of commensal synthetic communities	50
4.3.3 Host factors driving preferred associations in the root microbiota	53
4.3.4 SynCom-specific transcriptional responses of Lj and At roots	55
4.3.5 Invasiveness and persistence in the root microbiota.	56
4.4 Discussion	59
4.5 Materials and methods	60
4.5.1 Bacterial and plant material	60
4.5.2 Establishment of the <i>L. japonicus</i> bacterial culture collection	61

4.5.3 Culture-independent community profiling	62
4.5.4 Multi-species microbiota reconstitution experiments	62
4.5.5 SynCom invasion experiments	63
4.5.6 Collection of root exudates and millifluidics experiment	64
4.5.7 Mono-associations of SynCom members with host plants	64
4.5.8 Absolute quantification of bacteria	65
4.5.9 Processing of 16S rRNA gene amplicon data	65
4.5.10 Host preference and invasiveness indices	66
4.5.11 Bacterial genome assembly, annotation, and phylogenomic analysis	66
4.5.12 RNA-sequencing and data analysis	67
4.5.13 Statistics and reproducibility	68
4.6 Date and code availability	68
4.7 Author contributions	68

## Chapter 5 Shared features and reciprocal complementation of the *Chlamydomonas* and *Arabidopsis* microbiota

5.1 Abstract	69
5.2 Introduction	69
5.3 Results	70
5.3.1 <i>C. reinhardtii</i> assembles a distinct microbiota from the surrounding soil	70
5.3.2 The <i>C. reinhardtii</i> phycosphere and the plant root share a core microbiota	72
5.3.3 Reconstitution of phycosphere communities using reductionist approaches	74
5.3.4 Cr- and At-derived SynComs form taxonomically equivalent communities on either host	78
5.3.5 Physical proximity is required for the assembly of phycosphere communities and promotion of Cr growth	80
5.4 Discussion	82
5.5 Materials and methods	84
5.5.1 Culture-independent bacterial 16S rRNA sequencing and data analysis	84
5.5.2 Isolation and genome sequencing of <i>Chlamydomonas</i> -associated bacteria	85
5.5.3 Genome assembly, annotation and comparative analyses of the Cr-SPHERE culture collections	86
5.5.4 Preparation of SynCom inocula and analysis of amplicon sequencing data	86
5.5.5 Multi-species microbiota reconstitution experiments	87
5.5.6 Split co-cultivation system	88
5.6 Data and code availability	89
5.7 Author contributions	89

## Chapter 6 Maize field study reveals covaried microbiota and metabolic changes in roots over plant growth

6.1 Abstract	91
--------------	----

6.2 Introduction	91
6.3 Results	94
6.3.1 Dynamics of the soil and root-associated microbiota diversity	94
6.3.2 Stable root-associated microbial taxa over host growth	97
6.3.3 Microbial community assembly patterns at different phylogenetic levels	98
6.3.4 Plant growth phase is a major driver of both root metabolism and root microbiota dynamics	100
6.3.5 Genotype-dependent response of root microbiota, metabolism, and plant biomass to soil P availability	103
6.4 Discussion	105
6.4.1 Soil physicochemical properties drive soil and plant-associated microbial community shifts across different microbial kingdoms	105
6.4.2 Temporal changes in maize root-associated bacterial communities covary with root metabolite dynamics	106
6.4.3 Soil P availability induces changes in root metabolism, microbiota, and plant performance	107
6.5 Materials and methods	108
6.5.1 Experimental design	108
6.5.2 Soil and Plant sampling	109
6.5.3 Plant biomass and root metabolism measurement	109
6.5.4 Soil and root microbial community profiling	109
6.5.5 Amplicon sequencing data processing and microbial community diversity analysis	109
6.5.6 Predictive taxa for root lipid identification and network construction	110
6.6 Data and code availability	110
6.7 Author contributions	110
<b>Chapter 7 Studying plant-associated microbial communities using novel diversity and network analyses</b>	
7.1 Abstract	113
7.2 Introduction	113
7.3 Results	116
7.3.1 Meta-analysis of the plant microbiota	116
7.3.2 Network analysis of the plant microbiota	120
7.3.3 R package ‘mina’: microbial community diversity and network analysis	121
7.3.4 The R package ‘mina’ provides novel insights into plant microbiota	125
7.3.5 Assembly of alga- and root-associated microbiota in Cologne Agricultural Soil	126
7.4 Discussion	133
7.4.1 Core and representative members of plant microbiota	133
7.4.2 Community diversity analysis based on higher-order features	134
7.4.3 Network comparison at the local and global scales	134
7.5 Materials and Methods	135

7.5.1 Pre-processing of microbial community profiling data	135
7.5.2 Sequence and diversity analysis for microbial communities	137
7.5.3 Representative ASVs selection	138
7.5.4 Network inference and clustering	138
7.5.5 Community diversity and network analysis using ‘mina’	138
7.5.6 CAS-associated microbial community analysis	139
7.6 Data and code availability	139
7.7 Author contributions	139
Supplementary figures and tables	141
Abbreviations	155
References	157



# Chapter 1 Introduction

## 1.1 Introduction to microbiota studies

### 1.1.1 Microbes and microbial communities

Microbes, the living organisms that are only visible with a microscope, are present in almost all environments on earth. The number of prokaryotic microbes on planet Earth may be as many as  $4$  to  $6 \times 10^{30}$ , accounting for more than half of the carbon biomass of living organisms in the global ecosystem (Whitman *et al.*, 1998). Bacteria, the earliest used term for prokaryotic microbes, were found to be composed of multiple primary kingdoms, including typical bacteria and archaeobacteria, using phylogenetic analysis based on ribosomal RNA (rRNA) sequences (Woese & Fox, 1977). The latter was proposed to be a new domain of life named Archaea (Woese *et al.*, 1990) and together with Eukarya, which contains another microbial kingdom, Fungi, they form the widely accepted three-domain system for the tree of life.

When inhabiting the same environment, microbes from multiple kingdoms constantly interact with each other in diverse ways and assemble into complex communities, denoted as microbiota. Along with their surrounding biotic and abiotic factors, these communities constitute the microbial ecosystems (Raes & Bork, 2008). Among different habitats, soil and marine environments are considered to be the two most heavily populated with microbes and these two habitats contribute equally to the global net primary production of carbon (Field, 1998). In soil, microbes account for a large proportion of genetic diversity of the ecosystem and are involved in various processes, including nutrient cycling and decomposition, promoting plant productivity, and influencing vegetation dynamics (van der Heijden *et al.*, 2008). In the ocean, microbes of diverse kingdoms account for the majority of marine biomass with up to half of the carbon fixed by photosynthesis in the ocean being consumed by heterotrophic bacteria (Azam *et al.*, 1983). These environmental microbiota are typically composed of highly diverse free-living microbes and thus are extremely complex and dynamic.

For the microbiota that associated with specific host, for instance, human gut or plant root, the diversity of community has usually been found to be lower. However, these microbes interact not only with each other, but also with their hosts by, for example, promoting nutrient uptake (Musso *et al.*, 2011) or protecting the host from pathogens (Smith & Goodman, 1999). Though the concept was initially introduced to describe the relationship between a host and a single symbiont (Margulis, 1991), the term *holobiont* then came into broad use to refer to this symbiotic relationship between host macro-organisms and commensal microbiota. Based on this idea, the concept of a *hologenome*, which aggregates the host and associated microbial genomes, was proposed, as behaving as a discrete entity during adaptive evolution (Rosenberg *et al.*, 2007). Other hypotheses such as coevolution concept emphasized the feedbacks and adaptations between hosts and their associated microbiota (Zaneveld *et al.*, 2008), especially in the context of metabolic collaboration (Wilson & Duncan, 2015). Taken together, understanding the composition of microorganisms and their interactions within the microbiota is of fundamental importance for revealing the molecular mechanism underpinning the function and evolution of these ecosystems.

### 1.1.2 Microbiota research: from isolates to communities

To understand the principles governing microbial community assembly, approaches were developed to identify the involved microbes, which have included mainly culture-dependent methods such as strain isolation. However, for microbes present in extreme environments, such as the deep sea, at high temperatures or in anaerobic environments, isolation is time- and labor-intensive and sometimes problematic due to technical difficulties. As of today, it is thought that nearly half of the microbial taxa (at 97% 16S rRNA gene sequence similarity) have yet to be successfully cultivated (Martiny, 2019). Because of these unculturable bacteria, most of the ecological laws observed in plants and animals in macro-ecosystems, such as the relationships between body size and abundance in ecology (White *et al.*, 2007), have yet to be tested for microbial communities. Therefore, to better understand microbial ecology, effective approaches to characterize the complete community structure are needed.

Due to their conservation across living organisms and their intercalated mix of highly conserved and variable regions, rRNA genes have been successfully used for the identification and phylogenetic research of microorganisms without isolation (Woese & Fox, 1977). Generally, small subunit 16S and 18S rRNA genes are the most widely used markers for prokaryotic and eukaryotic microbes. Primers designed to bind to conserved regions have been used to amplify the variable sequences by PCR. Afterwards, these amplicon fragments, which reveal the genetic variation among species, are aligned to each other, allowing taxonomical classification. However, the presence of multiple copies and intra-organism variation in rRNA genes in bacterial genomes complicate correct assessment of relative abundance and taxonomic assignment when estimating the diversity of microbial communities. Therefore, conserved single-copy protein-coding genes have been proposed as alternatives. Comparative analysis of these protein markers shows good agreement with rRNA genes, with an increased resolution (Roux *et al.*, 2011). For fungi, the internal transcribed spacer (ITS) region has been used as a universal marker to identify diverse fungi, with higher sensitivity compared to rRNA or protein-coding genes (Schoch *et al.*, 2012). This region includes the ITS1, 5.8S rRNA gene, and ITS2 segments and is usually located in between the small subunit and large subunit of ribosomal RNA genes. Unfortunately, since the length of ITS in bacteria and archaea is highly variable and a large portion of rRNA genes are unlinked, its applicability is limited to fungal organisms (Brewer *et al.*, 2020).

The development of high-throughput DNA sequencing technologies facilitates microbial community surveys by allowing culture-independent approaches such as community profiling, meta-genomic, and meta-transcriptomic screenings. Among these technologies, amplicon sequencing of marker genes has become the most popular method for ecological surveys of microbiota, partially because of its low costs. Consequently, explosive growth in the archiving of raw next-generation sequencing data in all major DNA databases was observed (Kodama *et al.*, 2012). The analysis of high-throughput reads enables the acquisition of more comprehensive and accurate ecological information than the traditional culture method (Gupta *et al.*, 2019). For instance, a global study across diverse habitats, including both normal and extreme conditions, showed that salinity is the major driver of bacterial diversity for environmental microbial samples (Lozupone & Knight, 2007). In another study, the analysis of 16S amplicon data obtained from samples collected in 68 different locations during the Tara Oceans expedition showed that the most

important factors determining epipelagic microbial community composition in the ocean were depth and temperature (Sunagawa *et al.*, 2015). Moreover, using common laboratory and computational protocols, the Earth Microbiome Project carried out by hundreds of researchers worldwide achieved the complete characterization of Earth's microbial diversity (Thompson *et al.*, 2017).

In parallel to the advances mentioned above in the characterization of environmental microbial communities, rapid developments in the analysis of host-associated microbiota using cultivation-independent methods have also been made. The human gut microbiome has been the most intensively studied. The structures of the human microbiomes at diverse body sites (18 and 15 for women and men, respectively) were characterized (Huttenhower *et al.*, 2012), and long-term studies of fecal microbiota samples collected for up to five years showed a stable bacterial composition (Faith *et al.*, 2013). Factors that alter the human gut microbiome such as diet (David *et al.*, 2014), age, geography, cultural traditions (Yatsunenko *et al.*, 2012), and non-antibiotic medicines (Maier *et al.*, 2018) have also been extensively investigated. However, as mentioned above, 16S rRNA sequences are highly conserved across kingdoms, which would result in confounding results due to contamination from host plastid and mitochondrial sequences when studying host-associated microbiota. To prevent such contamination in these studies, approaches such as peptide nucleic acid PCR clamps were developed to enrich for microbial sequences without introducing bias (Lundberg *et al.*, 2013).

Later, the emerging long-read technology allows the sequencing of full-length genes rather than only the amplified regions of marker genes. However, the information on genetic variation provided by amplicon sequencing is limited to the marker gene and to single nucleotide resolution. To compensate for this, meta-genomic and meta-transcriptomic sequencing have also been applied in microbiota studies. With these approaches, total DNA or RNA is extracted from the surveyed samples and then deep-sequenced, allowing the capture of all genes merely present or actively transcribed within the given samples, which is then denoted as the microbiome. Using this strategy, functions of bacterial genes that are enriched in plant root-associated habitats were identified, such as iron mobilization (Bulgarelli *et al.*, 2015); parallel analysis of human microbiomes across individuals and body sites at the genomic and transcriptomic levels has revealed varying degrees of similarity, indicating divergence between functional potential and activity (Franzosa *et al.*, 2014). Besides, the above-mentioned Tara Oceans expedition was able to identify the core functionality of the ocean microbiota (Sunagawa *et al.*, 2015) and demonstrated their substantial overlap with the human gut microbiome (Li *et al.*, 2014). Taken together, in microbial studies, community profiling allows the characterization of the community structure, meta-genomics provides information about the functional capacity of the microbiota assemblage, and meta-transcriptomics data identifies the actively expressed genes. Therefore, these strategies allow us to identify the organisms and genes that possibly contribute to the establishment of microbial communities. However, further studies of underlying molecular mechanisms are essential to validate the proposed microbiota assembly process.

### 1.1.3 Deconstruction and reconstruction of microbiota

To better understand microbiota after culture-independent surveys, e.g., elucidating the functions of identified microbes or genes that primarily shape the community structure, systematic collections of

isolates are needed. Community members have been isolated using classical methods and this process was denoted as microbiota “deconstruction”. With these monoculture isolates, pairwise interactions could be assessed and used to predict higher-order interactions in complex communities (Venturelli *et al.*, 2018). For “reconstruction” of the microecosystem, culture strains with known genetics have been assembled into synthetic microbial communities (SynComs), which soon became the model system of choice, due to their high tractability as well as reduced complexity compared to most of the microbial communities formed naturally (De Roy *et al.*, 2014). For instance, community members of kefir grain and fermented milk were isolated and used to investigate pairwise interactions via metabolic modelling and laboratory validation (Blasche *et al.*, 2021). Standard protocols were also developed for isolation and cataloguing of bacteria and fungi from cheese rinds, followed by reconstruction into tractable SynComs with representative strains to quantify microbiota dynamics (Cosetta & Wolfe, 2020). High-throughput anaerobic culturing techniques for gut microbiota were established a decade ago and inspired the culturing renaissance (Vrancken *et al.*, 2019). Similar approaches were implemented in environmental systems; for example, to better understand the role of the soil microbiota in carbon and nitrogen cycling, a model community was constructed from the natural soil biota (McClure *et al.*, 2020). Besides these progresses, a comprehensive evaluation of the extent to which SynComs recapitulate the diversity and dynamics of their original natural communities across micro-ecosystems is still missing.

## 1.2 Plant-associated microbiota research

### 1.2.1 Plant-associated microbial communities

Plants live together and interact with diverse microbes in the form of complex communities, known as the plant microbiota, which colonize and interact with the plant host. These microbial communities are classified into distinct groups based on, for example, whether colonization is aboveground or belowground; and how close they are physically to the host. Among categories, phyllosphere and rhizosphere microbiota are the most studied compartments. The former represents the total aboveground surface of a plant that provides a niche for the colonization of microorganisms (Preece & Dickinson, 1971) that mostly include non-pathogenic bacteria (Vorholt, 2012). The rhizosphere was first introduced as a concept by Lorenz Hiltner in 1904 to refer to the zone around roots with the most intense bacterial activity (Curl & Truelove, 1986). Later, other eukaryotic microbes were also found in this root compartment, such as fungi and oomycetes, and the term rhizosphere was extended to denote the micro-ecosystem inhabited by roots, soil, and the soil biota and the site of their interactions with each other (Lynch & de Leij, 2012). Taken together, the rhizosphere, rhizoplane (a term suggested by F.E. Clark in 1949 representing the surface of root), and the root endosphere, constitute the root-associated microbiota, whose community members are primarily derived from the soil.

With the development of high-throughput sequencing, an increasing number of studies systematically characterizing the plant microbiota were conducted, which have aimed at understanding the structure, assembly pattern, and factors driving the assembly of the plant-associated microbiota. An early study already showed that plant roots could modulate rhizosphere conditions (such as changing the pH) in response to nutrition deficiency and regulate the abundances of bacteria living there (Marschner *et al.*, 1986). This prompted exploration of the co-

evolution and adaptation of hosts and associated microbes as well as their interactions. The model plant *Arabidopsis thaliana*, commonly known as thale cress, was introduced as a model for plant-microbe interactions, not only due to its well-studied genetics but also because it is a widespread plant native to different continents, including Europe, Asia, and Africa. These original habitats of *A. thaliana* span a huge range of geographical locations and soil physicochemical properties. By profiling more than 600 *A. thaliana* plants, the root microbiota was found to be distinct from the soil biota, and the conservation of taxa across diverse soils suggested that the assembly of root microbial communities might be driven by core ecological principles (Lundberg *et al.*, 2012). Additionally, root-inhabiting bacterial communities were also defined by the soil type from which the microbes were derived (Bulgarelli *et al.*, 2012). A comprehensive investigation of European *Arabidopsis* populations revealed strong effects of soil origin and geographic location on the composition of root bacteria and filamentous eukaryotic communities, respectively (Thiergart *et al.*, 2020).

### 1.2.2 Plant-microbiota interactions

The plant microbiome was found to benefit the host by promoting nutrient uptake and augmenting immune functions (Musso *et al.*, 2011). The most striking examples of the beneficial effects are mutualistic symbionts, including the colonization of nitrogen-fixing rhizobia and mycorrhizal fungi in the roots, organisms which have co-evolved with their plant hosts for ~400 million years (Martin *et al.*, 2017). The characterization of the rhizosphere bacteria associated with sunflower plants demonstrates wide production of siderophore, which helps transporting certain compounds such as ferric iron (Ambrosini *et al.*, 2012). By regulating the development of the diffusion barriers, which consist of specialized cell layers in root, bacterial communities stabilize the mineral nutrients in the plant (Salas-González *et al.*, 2021). When soil nutrients are limiting, for example through a reduction in the supply of phosphorus, *A. thaliana* roots engage in mutualistic interactions with microbes and consequently modulate microbial community structures (Robbins *et al.*, 2018). However, the mechanism underlying these interactions and the extent to which this process promotes beneficial traits in response to nutritional limitation remain to be determined. In contrast to the beneficial root microbiota, pathogens that colonizing the root can disrupt host-microbe homeostasis and cause diverse diseases. Therefore, plants have evolved an innate immune system to protect them against these microbes. Due to the restriction of resource, plant needs to invest into either growth or defense, formulating the phenomenon known as growth-defense trade-off (Huot *et al.*, 2014). However, it remains unclear how plants differentiate beneficial, commensal, and pathogenic microbes and mediate between growth and defense with the presence of root microbiota.

Though most of the high-throughput profiling studies focus on the bacterial members of the plant microbiota, microbes from other kingdoms, particularly fungi, play an important role in shaping plant-associated microbial communities. A suppressive effect of bacteria on a fungal root pathogen was described in the rhizosphere of sugar beet (Mendes *et al.*, 2011), indicating the existence of complex multi-kingdom interactions. These inter-kingdom relationships within the root microbiota were extensively examined at the community level in *A. thaliana* roots, and it was found that the root-associated bacterial community protects the host against pathogenic filamentous eukaryotes, including fungi and oomycetes (Durán *et al.*, 2018). A survey of the root microbiota associated with legume *Lotus japonicus* also revealed a microbial interkingdom interaction; bacterial intraradical

colonization was found to be dependent on arbuscular mycorrhizal fungi (AMF) infection (Thiergart *et al.*, 2019).

Moreover, root-associated microbes derived from surrounding soil were found to mobilize nutrients for their plant hosts, thus reducing the need for additional fertilization (Pii *et al.*, 2015). Therefore, understanding the composition and function of the plant microbiota can contribute to sustainable agriculture, through, for example, the engineering of microbes that promote crop health and fitness. Owing to their global geographic spread and economic significance, the root microbiota of maize and rice were surveyed to determine the effect of geographic location, soil type, host genetics, and growth stage in the field. Rhizosphere samples collected at weekly intervals suggest a continuously varying community structure over the life cycle of maize (Walters *et al.*, 2018), while rice samples collected in a similar way indicated a dynamic vegetative phase followed by a stable reproductive root microbiota (Edwards *et al.*, 2018). Additionally, an effect of field site on root microbiota assembly was reported for host crops. Intensive farming practices, for example, the use of mineral fertilizers, has been broadly applied in recent decades to promote plant performance. However, in the long-term, these practices can change soil physico-chemical properties, degrade soil quality, and indirectly affect the soil microbiome (Bender *et al.*, 2016). These results are helping to untangle the complex interactions between crops and soil-borne microbes.

### 1.2.3 Synthetic plant-associated microbial communities

Based on microbial community alteration patterns, we developed hypotheses and models to explain the assembly of microbiota. However, it is necessary to validate these proposed hypotheses and models in tractable systems. Additionally, to gain fundamental systemic insights, such as into molecular mechanisms and functional interactions, highly controllable and reproducible experimental conditions are essential. To address these issues, the concept of a synthetic community (SynCom) was introduced to plant microbiota studies as an approach to recapitulate the key features of natural communities in the laboratory (Vorholt *et al.*, 2017). Large-scale isolation of plant-associated microbes allows the assembly and design of SynComs that allow researchers to survey the evolution and ecological features of these microbial assemblages. More than 400 root- and leaf-derived bacteria of *A. thaliana* were isolated and the genome drafts indicated a large overlap of functional capacity between these microbes (Bai *et al.*, 2015). By inoculating germ-free *A. thaliana* with microbial SynComs of different kingdoms, including bacteria, fungi, and oomycetes, Duran *et al.* showed that the bacterial microbiota is essential for plant survival when filamentous microbial eukaryotes are present (Durán *et al.*, 2018). The effects of plant metabolites on the root microbiota were studied by examining community shifts in SynComs (Voges *et al.*, 2019). The colonization of synthetic bacterial communities at the roots of *A. thaliana* of multiple genotypes and under variable phosphate conditions shows differential and highly deterministic assembly patterns (Castrillo *et al.*, 2017; Finkel *et al.*, 2019). By conducting drop-out experiments, where specific microbes were left out from communities formed by isolated native bacteria, a priority effect governing phyllosphere microbiota assembly was observed (Carlström *et al.*, 2019). A further study has also shown that *Variovorax* could reverse the growth inhibition of *Arabidopsis* roots caused by other bacteria in both the monoculture and community context, through a auxin-degradation operon conserved in the genus (Finkel *et al.*, 2020). Taken together, these SynCom-based studies provide mechanism

insights into the plant-microbe interactions and complement the discoveries gained from natural communities.

## 1.3 Data processing and analysis of microbial community surveys

### 1.3.1 Processing of community profiling data

When next-generation sequencing was first developed, the mainstream technology was single-end sequencing with maximum read lengths of around ~600 bp (for Roche 454, GS FLX Titanium system, 2008). The quality of the reads using this approach decreased dramatically toward the end of the sequences, and this considerably restricted the sequences that could be amplified, sequenced, and compared for amplicon profiling. Later, the Illumina sequencing platform, in particular MiSeq, was developed, which yields 300-bp sequences for pair-end reads. With higher throughput and lower costs, this technology has become widely used in amplicon sequencing studies. These rapid improvements in sequencing technology benefit and foster community profiling; however, such technology also creates a demand for corresponding computational tools for data processing and analysis, especially when data generated from different platforms are involved.

For the profiling of a bacterial taxon, the amplified 16S rRNA gene region has relatively conservative length. For example, for a MiSeq sequencing run of a V5 to V7 region amplicon (amplified by 799F and 1192R primer pairs), more than 94.14% of reads have a length within the range 359–389 bp. Typically, these sequencing reads are first filtered and trimmed for quality control. Afterwards, for pair-end reads, the forward and reverse reads are then merged based on their overlapping regions and are then assigned to each sample according to barcode sequences. Subsequently, chimaeras produced during PCR are identified and removed. For filamentous microbes, including fungi and oomycetes, the ITS region is used as marker gene. Since the whole region spanning 650 to 900 bp ([Horton & Bruns, 2001](#)) is longer than the sequencing length of most next-generation platforms, only ITS1 or ITS2 regions are sequenced, and the variation in length needs to be taken into account.

To identify community members represented by the amplified and sequenced regions of marker genes, the concept of the operational taxonomic unit (OTU), originally introduced by Robert R. Sokal and Peter H. A. Sneath ([Sokal & Sneath, 1963](#)), was re-implemented in microbial ecology studies. Sequences with higher similarity than the arbitrary threshold, commonly 97%, are clustered together into one OTU that encompasses organisms from the same species or genus. This procedure decreases the random noise caused by sequencing errors. However, the non-deterministic and abundant-dependent feature of clustering algorithm also reduces the reproducibility of the OTU-based approach and hinders cross-referencing between studies. Therefore, with the increases in sequencing throughput and accuracy, unique DNA sequences obtained after quality control and error correction are now used directly to represent community members, denoted as amplicon sequence variants (ASVs). Under this scheme, raw sequencing data are processed similarly as before, with the differences that extra error correction is applied, and no clustering is performed. Since ASVs capture exact nucleotide sequence variations, they represent distinct biological entities and can easily be compared between different studies.

### 1.3.2 Microbial community diversity analysis

To compare the community structure, diversity analysis has become a ubiquitous approach in ecological surveys of microbiota. Alpha-diversity, which was initially used to assess the richness of species in macro-communities (Whittaker, 1960), has been applied to assess community differences between conditions by characterizing the intra-community diversity via measures of the richness and evenness. Additional approaches taken include the Shannon index, which was originally developed to evaluate the evenness of words distributed in strings (Shannon, 1948), the Simpson index, which assesses the diversity by measuring the possibility of obtaining the same species when randomly taking two entities from the community (Simpson, 1949), and the phylogenetic diversity, which takes into account the phylogenetic distances between community members (Faith, 1992). However, the sequencing depth, i.e., the sample size or library size, affects the diversity estimation, especially the assessment of richness (Willis, 2019). Therefore, it is essential to normalize the data before calculating and comparing the alpha-diversity between samples. Rarefaction, a process that randomly picks up a specific number of reads from all samples, is commonly used to reduce the bias caused by sequencing depth. However, some researchers advise against its use since it results in a high rate of false discoveries (McMurdie & Holmes, 2014), while others claim that the reported flaws are an artefact of the data simulation (Weiss *et al.*, 2017).

To measure differences in composition between communities, the concept of beta-diversity was introduced (Whittaker, 1960), which involves the comparison of presence/absence or relative abundance of all compositions between samples. Since the abundance tables of microbial communities are highly sparse, i.e., the presence of many zeros in the OTU/ASV table, the Euclidean distance is not often applied (Huggerth & Andersson, 2017). Instead, Bray-Curtis dissimilarity is one of the most commonly used statistics for quantification, where the more members two communities share, the less Bray-Curtis dissimilarity between the compared samples (Bray & Curtis, 1957). The index takes into account phylogenetic context, such as UniFrac, which has also been proposed as a metric for comparing the diversities between communities (Lozupone *et al.*, 2006). Afterwards, dimensionality reduction approaches such as principal component or coordinate analysis (PCA or PCoA) need to be implemented before visualization. The former is restricted to Euclidean distance, which as mentioned earlier, is not commonly used in microbial community studies, and the latter can be applied together with any dissimilarity measurement. Typically, the first few components or coordinates already explain a relatively large amount of the variation between samples; therefore, the separation or clustering of samples shown by the PCoA plot can describe the dissimilarities between microbial communities adequately. Additionally, the variance between samples explained by each known factor and the combination of factors can be statistically tested by permutational multivariate analysis of variance (PERMANOVA, Anderson, 2017).

### 1.3.3 Microbial community network analysis

Composition-based diversity analysis allows thorough characterization of community structure and differences between communities. However, throughout this process, interactions between the microbes and the system dynamics are ignored. To overcome this limitation, network analysis was introduced into profiling data analysis. As a theory based on mathematical graphs, network analysis

has been applied to diverse research fields, including social science and ecological systems, such as the food web (Dunne *et al.*, 2002; Krause *et al.*, 2003), plant-pollinator interactions (Olesen *et al.*, 2007), and plant-animal mutualisms (Bascompte *et al.*, 2003). In these systems, nodes represent the objects or biological organisms, and their relationships are indicated by edges connecting them. For microbial community networks, community members are designated as nodes and interactions between microbes, i.e., edges, most commonly, are inferred by comparing the co-occurrence or covariance of microbes across samples. The number of samples, therefore, has an impact on the robustness of the constructed network, and, in particular, on the reliability of the inferred edges.

Traditionally, Pearson or Spearman correlation coefficients are calculated between pairwise microbes. The former evaluates the linear relationship between two community members among samples and the latter focuses more on the monotonic property of compared variables. However, due to the compositional nature of the microbiota profiling data, i.e., only the relative abundances of community members are available, applying simple correlations to these abundances leads to spurious results (Aitchison, 1982). Therefore, novel methods with extra data transformation, such as SparCC (Friedman & Alm, 2012), were developed to reduce the bias introduced by compositional effects. By assuming sparse features of a microbial network and combining these with graphical model-based data transformation, SPIEC-EASI yields better performance for network inference (Kurtz *et al.*, 2015). Besides, workflows aimed at improving accuracy by integrating multiple of the previously mentioned approaches, such as CoNet and NetCoMi, were also specifically developed for microbiome data (Faust & Raes, 2016; Peschel *et al.*, 2021). However, evaluation of these strategies has shown varied sensitivity and precision when applied to microbial data sets and a golden benchmark, which is based on the ground truth of higher-order microbial interactions in a complex system, is still missing (Weiss *et al.*, 2016; Faust, 2021).

Due to the numerous members of natural microbial communities, the inferred networks are typically highly dimensional, and, therefore, direct comparison of adjacent matrices, which describe the whole network's nodes and edges, is rarely implemented. Alternatively, features such as network density, centrality, clustering coefficients, and path length, which are governed by robust principles (Albert & Barabási, 2002), have usually been computed and used to describe the topological structure of networks. Comparison of time-lagged bacterial networks of ocean planktonic samples collected from different depths revealed that layer dynamics closer to the surface drive organism changes in deep layers (Cram *et al.*, 2015). Co-occurrence network analysis indicated strong body site specialization of the human microbiome (Faust *et al.*, 2012) and non-random association as well as common life-history strategies of global soil microbiota (Barberán *et al.*, 2012). Despite these insights, the interpretation of microbial networks and validation of interactions remain challenging due to the complexity of micro-ecosystems. Experimental studies of microbe-microbe interactions mainly address pairwise relationships, although higher-order species interactions, i.e., the interactions between organisms when a third member is present, have been found to be influential for community dynamics (Bailey *et al.*, 2016).

## 1.4 Thesis overview

### 1.4.1 Outline

The author of this thesis has developed and implemented standardized frameworks for profiling data analysis of natural and synthetic microbial communities to reveal the assembly and diversity of plant microbiota. In [Chapter 2](#), a plant-beneficial interaction between *A. thaliana* and the root microbiota, dependent on the secretion of plant-derived coumarins under iron deprivation, is described. Crosstalk between root commensal bacteria and host innate immunity is extensively described in [Chapter 3](#) and how those commensals are adapted to specific host species is explored in [Chapter 4](#). As an analogous microenvironment to the rhizosphere, the phycosphere microbiota of modal alga *Chlamydomonas reinhardtii* was examined and compared with plant root microbiota, and this work is described in [Chapter 5](#). In [Chapter 6](#), by tracking the root-associated microbial community changes during the development of field-grown maize, authors found that field management, soil property, plant growth, and root metabolism were highly dynamic and covaried with each other, influencing root microbiota establishment. In the last chapter ([Chapter 7](#)), a framework of microbial community diversity and network analysis is presented and applied to plant-associated community profiling datasets, including thousands of bacterial and fungal samples integrated from the department's previous studies. The ecological networks from samples obtained under different experimental conditions were thoroughly compared to deepen our understanding of the principles governing plant microbiota assembly.

### 1.4.2 Publications and author contributions

Chapters 2 to 6 are adapted from the publications in chronological order that the author of this thesis (R.G.) contributed to. In the projects presented in Chapters 2 to 5, R.G. performed the bacterial amplicon sequencing data processing and analysis. For the project described in Chapter 6, R.G. conducted all sequencing data analysis; and contributed primarily to data interpretation, visualization, and manuscript writing. R.G. conducted the project reported in Chapter 7 and is preparing the corresponding manuscript. Details of author contributions for each project are listed at the end of each chapter.

# Chapter 2 Root-secreted coumarins and the microbiota interact to improve iron nutrition in *Arabidopsis*

## 2.1 Abstract

Plants benefit from associations with a diverse community of root-colonizing microbes. Deciphering the mechanisms underpinning these beneficial services are of interest for improving plant productivity. We report a plant-beneficial interaction between *Arabidopsis thaliana* and the root microbiota under iron deprivation that is dependent on the secretion of plant-derived coumarins. Disrupting this pathway alters the microbiota and impairs plant growth in iron-limiting soil. Furthermore, the microbiota improves iron-limiting plant performance via a mechanism dependent on plant iron import and secretion of the coumarin fraxetin. This beneficial trait is strain-specific yet functionally redundant across phylogenetic lineages of the microbiota. Transcriptomic and elemental analyses revealed that this interaction between commensals and coumarins promotes growth by relieving iron starvation. These results show that coumarins improve plant performance by eliciting microbe-assisted iron nutrition. We propose that the bacterial root microbiota, stimulated by secreted coumarins, is an integral mediator of plant adaptation to iron-limiting soils.

## 2.2 Introduction

Plant roots are colonized by a diverse community of microbes, collectively termed the root microbiota, originating from the surrounding soil biome (Bulgarelli *et al.*, 2012; Lundberg *et al.*, 2012; Bai *et al.*, 2015). The structure of these communities is shaped by soil edaphic factors and root-secreted photosynthates and secondary metabolites (Berendsen *et al.*, 2012; Bulgarelli *et al.*, 2013). The root microbiota provides indirect protection against soil-borne fungal pathogens (Durán *et al.*, 2018; Carrión *et al.*, 2019) and is thought to improve host nutrition by improving the bioavailability of nutrients (Hacquard *et al.*, 2015). However, the extent to which plants can selectively alter their microbiota and harness these beneficial traits in response to nutritional stress is unknown.

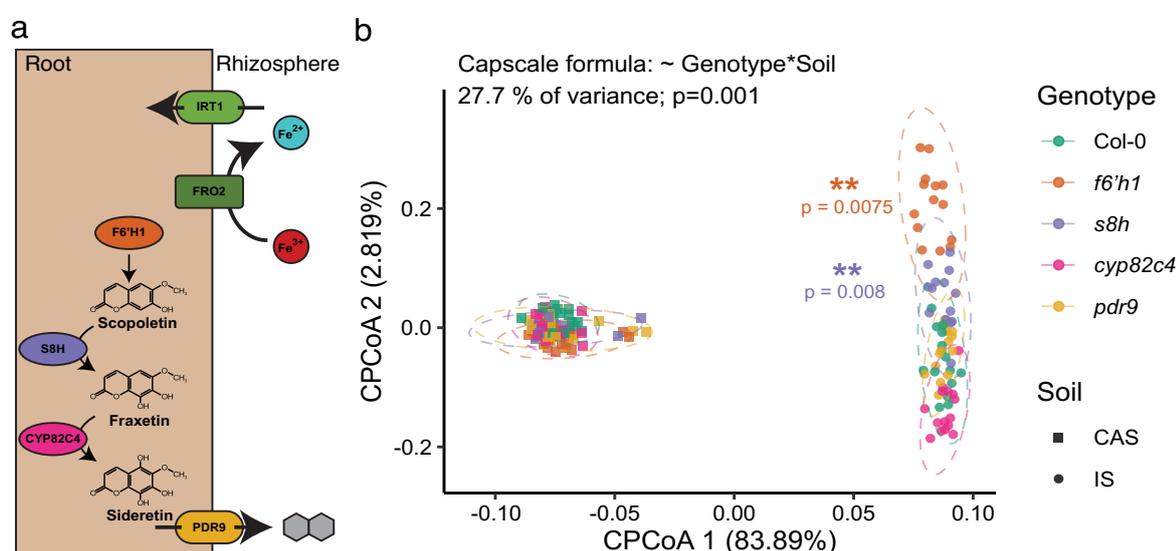
Iron is an essential mineral nutrient of plants, acting as a catalyst in many biological processes including photosynthesis and respiration. Although it is an abundant element in most soils, its bioavailability is often limiting due to its extremely low solubility at neutral and alkaline pH, as in calcareous soils containing a high proportion of calcium carbonate. Iron deficiency results in stunted plant growth and leaf chlorosis, and is responsible for decreased crop yields and nutrient content in ~30% of arable land (Morrissey & Guerinot, 2009). As such, there is great economic interest in improving plant iron nutrition. In response to iron-limiting conditions, non-graminaceous plants, such as *A. thaliana*, mount an iron starvation response which is coordinated by FER-LIKE IRON DEFICIENCY INDUCED TRANSCRIPTION FACTOR (FIT) (Jakoby *et al.*, 2004; Colangelo & Guerinot, 2004) and a suite of associated basic helix-loop-helix (bHLH) type transcription factors (Ivanov *et al.*, 2012). This response serves to improve the solubility of iron through rhizosphere acidification by H<sup>+</sup>-ATPase AHA2 (Santi & Schmidt, 2009) and reduction of iron(III) to more-soluble iron(II) by plasma membrane protein FERRIC REDUCTION OXIDASE 2 (FRO2)

(Robinson *et al.*, 1999). Iron(II) is imported into the root epidermis by IRON-REGULATED TRANSPORTER1 (IRT1) (Vert *et al.*, 2002).

The secretion of coumarins, phenolic secondary metabolites deriving from the general phenylpropanoid pathway, is also induced by iron starvation and is thought to contribute to iron nutrition through direct mobilization of recalcitrant iron pools (Rodríguez-Celma & Schmidt, 2013; Schmid *et al.*, 2014; Fourcroy *et al.*, 2014, 2016; Schmidt *et al.*, 2014; Sisó-Terraza *et al.*, 2016; Siwinska *et al.*, 2018; Tsai *et al.*, 2018; Rajniak *et al.*, 2018). Three main coumarin compounds are produced in *A. thaliana* via a linear biosynthetic pathway (Figure 2.1 a) (Tsai *et al.*, 2018; Rajniak *et al.*, 2018). FERULOYL-COA 6-HYDROXYLASE1 (F6'H1) synthesizes scopoletin (Kai *et al.*, 2008), which can be converted to fraxetin by SCOPOLETIN 8-HYDROXYLASE (S8H) (Tsai *et al.*, 2018; Rajniak *et al.*, 2018), and further converted to sideretin by CYTOCHROME P450, FAMILY 82C4 (CYP82C4) (Rajniak *et al.*, 2018). Each of these coumarins can be exported by the ABC transporter PLEIOTROPIC DRUG RESISTANCE 9 (PDR9) (Fourcroy *et al.*, 2014), though other efflux pumps may also contribute to the export of some coumarins (Ziegler *et al.*, 2017). Coumarin secretion was recently shown to influence the structure of root microbial communities in artificially limed soil (Stringlis *et al.*, 2018a) and synthetic media (Voges *et al.*, 2019). However, the impact of coumarin secretion on the root microbiota in soils with different mineral nutrient availabilities and the consequences for plant productivity remain undefined.

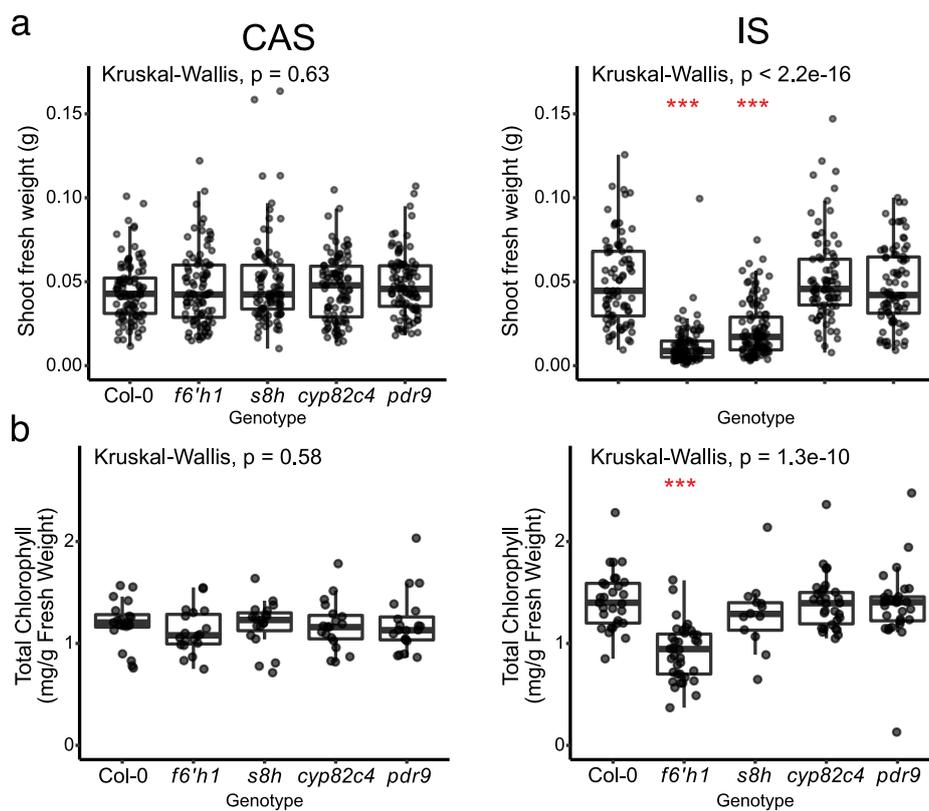
## 2.3 Results

### 2.3.1 Coumarin biosynthesis is important for plant growth and root microbiota composition in iron-limiting soil



**Figure 2.1 Coumarin biosynthesis affects root microbiota composition in a naturally calcareous soil.** (a) Diagram of pathways for coumarin biosynthesis and export, and reductive uptake of iron in Arabidopsis. (b) Constrained ordination of root bacterial community composition of coumarin pathway mutants, constrained for the interaction between soil and genotype. Ellipses delineate multivariate normal distribution at 95% confidence. Data are from one representative experiment of three (Col-0 n = 17, 15, *f6'h1* n = 18, 14, *s8h* n = 15, 14, *cyp82c4* n = 18, 15, and *pdr9* n = 17, 14 in CAS and IS, respectively). P values represent significance of separations between genotypes within each soil determined by pairwise PERMANOVA. Only *f6'h1* (orange) and *s8h* (purple) were significantly separated from Col-0 in IS.

To assess potential links between coumarin secretion, plant growth, and the root-associated microbiota, we grew *A. thaliana* Col-0 wild-type (WT) plants and mutants with defects in coumarin biosynthesis or export (Figure 2.1a) on two soils with contrasting iron availability. These mutants were previously shown to have abolished coumarin biosynthesis at their respective steps in the pathway (Schmid *et al.*, 2014; Ziegler *et al.*, 2017; Tsai *et al.*, 2018; Rajniak *et al.*, 2018). Cologne agricultural soil (CAS), obtained from a local site, is slightly acidic with pH 6.5, at which iron is sufficiently available. We also obtained soil from a vineyard in Italy, termed here Italian Soil (IS), which is alkaline and calcareous (pH 7.5, 9.7 % of active CaCO<sub>3</sub>). These conditions significantly reduce the availability of iron, despite this soil having a higher total iron content than CAS (Figure S1 in the published [journal version article](#)).



**Figure 2.2 Coumarin biosynthesis is important for plant growth in a naturally calcareous soil.** (a) SFW and (b) total chlorophyll content of coumarin pathway mutants grown in a non-calcareous (CAS) and a calcareous (IS) soil. Statistical significance was determined by Kruskal-Wallis; each mutant was compared to Col-0 by Wilcoxon Ranked Sum post-hoc. Significance is indicated by red asterisks (\*, \*\*, \*\*\*, indicate  $p < 0.05$ , 0.01, and 0.001, respectively). For shoot fresh weight measurements, Col-0  $n = 171$ , 204, *f6'h1*  $n = 168$ , 272, *s8h*  $n = 93$ , 113, *cyp82c4*  $n = 164$ , 209, and *pdr9*  $n = 172$ , 169 in CAS and IS, respectively. Chlorophyll content was measured from pooled leaf samples, (Col-0  $n = 35$ , 29, *f6'h1*  $n = 34$ , 36, *s8h*  $n = 19$ , 14, *cyp82c4*  $n = 34$ , 30, and *pdr9*  $n = 35$ , 30 in CAS and IS, respectively).

We observed a decrease in shoot fresh weight (SFW) and leaf chlorophyll content in *f6'h1* and *s8h* plants grown on IS, whereas the measured performance parameters of all genotypes were indistinguishable on CAS (Figure 2.2). A similar growth deficit was observed in *f6'h1* plants grown on other alkaline soils isolated from geographically diverse sites, and could be improved by supplementation with solubilized iron (Figure S1 in the published [journal version article](#)). These results show that coumarin biosynthesis is important for growth in naturally iron-limiting soils. The

performance of *pdr9* plants, however, was indistinguishable from WT on both CAS and IS; thus, coumarin export via the ABC transporter PDR9 was not crucial for growth in iron-limiting soils (Figure 2.1, 2.2 and Figure S1 in the published [journal version article](#)). This is in contrast to reported germ-free growth on synthetic media (Rodriguez-Celma *et al.*, 2013; Fourcroy *et al.*, 2014). These data suggest that in soil, sufficient coumarin export may occur via additional members of the diversified and promiscuous ABC transporter family (Ziegler *et al.*, 2017; Borghi *et al.*, 2019).

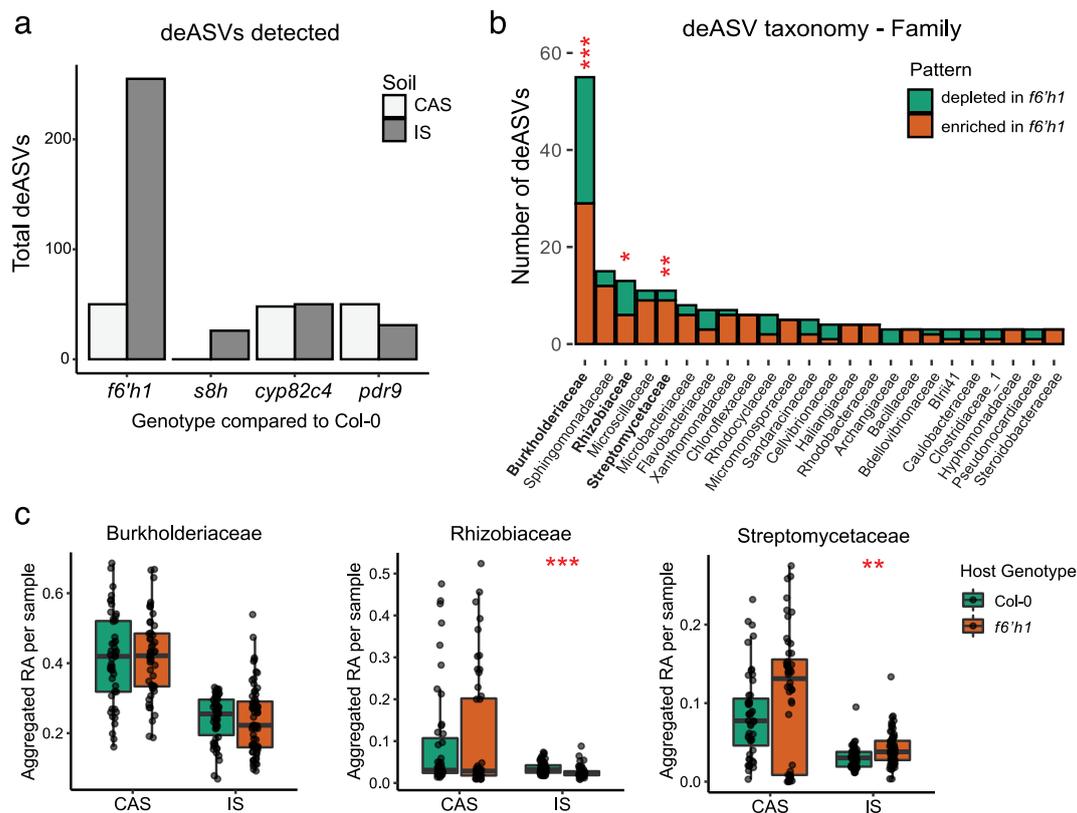
The root-associated bacterial microbiota of plants grown on CAS and IS was assessed by culture-independent *16S rRNA* gene amplicon sequencing and analysed at the amplicon sequence variant (ASV) level. Alpha-diversity was greater in IS than CAS for both unplanted soil and root samples, but did not vary by genotype (Figure S2 in the published [journal version article](#)). Unconstrained principal coordinate analysis (PCoA) of Bray-Curtis distances between samples showed that the soil type and batch were the largest drivers of divergence between samples (Figure S2 in the published [journal version article](#)). PCoA analysis of beta-diversity constrained (CPCoA) for the interaction between genotype and soil type revealed a significant separation of *f6'h1* and *s8h* plants from other genotypes when grown on IS, but not on CAS (Figure 2.1b). Analysis of bacterial community profiles from three batches of each soil type confirmed that *f6'h1* plants separate significantly from other genotypes when grown in IS, but not in CAS (Figure S2 in the published [journal version article](#)). Together, these results indicate that coumarin biosynthesis, especially of scopoletin and fraxetin, is important for plant growth and determining root microbiota composition in a naturally iron-limiting calcareous soil, but is dispensable in an iron-replete soil. Furthermore, this illustrates that the interaction between soil type and plant genotype can serve as a major determinant of root microbiota structure, explaining 27.7% of community variation (Figure 2.1b).

### 2.3.2 Coumarin biosynthesis restructures the root microbiota at the ASV level

In order to explore the nature of the observed changes in community structure, we determined which ASVs were differentially enriched (deASVs) in each mutant genotype compared to WT in each soil. For this analysis, we pooled data from three batches of each soil (except *s8h*, which was included in only one batch). The greatest number of deASVs was observed in *f6'h1* plants, with significantly more detected when grown in IS than CAS (260 deASVs in IS, 50 in CAS; Figure 2.3a). The impact of deASVs on the microbiota structure in terms of relative abundance was also greatest in *f6'h1* plants grown in IS (Figure S3 in the published [journal version article](#)).

(Figure on next page) **Figure 2.3 Coumarin biosynthesis restructures the root microbiota at the ASV level.** (a) Number of deASVs detected in indicated mutants compared to Col-0 in each soil. Data are pooled from three experiments (except *s8h*, which was included in only one), and filtered for ASVs found in at least three samples with RA > 0.05%. Differential enrichment was calculated using a negative binomial generalized log-linear model at an FDR-adjusted p value of 0.05. (b) Family-level taxonomic classification of deASVs in *f6'h1* plants growing on IS. Colours indicate if deASVs were enriched or depleted in *f6'h1* compared to Col-0. Hypergeometric enrichment test was performed to determine if each family was over- or under-represented in deASV list compared to all detected ASVs. Red asterisks indicate significance with FDR-adjusted p values. (c) Sample-wise aggregated relative abundance of the top three families most significantly over-represented in deASVs: *Burkholderiaceae*, *Rhizobiaceae*, and *Streptomycetaceae*. Each data point represents the average RA aggregated at the family level in a single sample. Significance between

genotypes in each soil was determined by Wilcoxon Ranked Sum test. \*, \*\*, and \*\*\* in (b) and (c) indicate  $p \leq 0.05$ , 0.01, and 0.001, respectively.

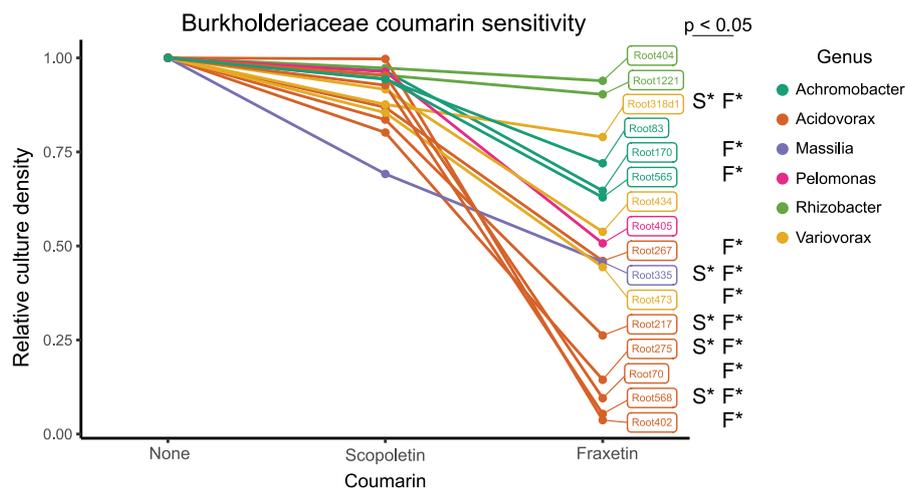


(Caption on previous page) **Figure 2.3 Coumarin biosynthesis restructures the root microbiota at the ASV level.**

Taxonomic analysis revealed that multiple bacterial families were significantly over- or under-represented within the deASV subset for *f6'h1* plants in IS compared to the full list of detected ASVs (Figure 2.3b and S3 in the published [journal version article](#)). *Burkholderiaceae* was the most prevalent family detected within the deASVs (56 deASVs; 2.65-fold enriched in deASVs subset compared to all ASVs detected). However, a correlation between coumarin production and deASV relative abundance could not be generalized at family-level resolution; most families contained deASVs which were more abundant in *f6'h1* plants compared to WT, as well as deASVs which were less abundant. Some patterns were observed at the genus level (Figure S3 in the published [journal version article](#)), but due to the overall low number of deASVs in each genus, their statistical significance remains unclear. This indicates that coumarin production restructures the root microbiota at the ASV level within multiple bacterial families. Despite being the family containing the most deASVs, the relative abundance of *Burkholderiaceae* was not significantly altered in coumarin-deficient plants (Figure 2.3c). The relative abundances of the next two most-impacted families, *Rhizobiaceae* and *Streptomyetaceae*, however, were slightly but significantly altered in *f6'h1* plants on IS. This indicates that, at the ASV level, disruption of coumarin biosynthesis has a quantitative impact within multiple root-associated commensal families, but with relatively minor effects on the microbiota structure at higher taxonomic ranks. This suggests the existence of ASV-level compensatory mechanism(s) within bacterial families which, during root microbiota establishment, maintain higher taxonomic structure in coumarin-deficient plants on iron-limiting soil. An increase in the number of deASVs was also observed for *s8h* plants grown in IS compared

to CAS (Figure 2.3a), though fewer deASVs overall were detected compared to other mutants as this genotype was only included in one experimental replicate. Consistently, the taxonomic profile of *s8h* deASVs was also enriched for *Burkholderiaceae*, and had considerable, but not complete, overlap with the deASVs detected on *f6'h1* plants in the same experiment (Figure S3 in the published [journal version article](#)). This indicates that production of both scopoletin and fraxetin impact the root microbiota.

As various coumarin compounds have been shown to exert antimicrobial activity, we examined the coumarin sensitivity of a panel of root commensal *Burkholderiaceae* strains, the most prevalent family within the deASVs on *f6'h1* and *s8h* plants. These strains were previously isolated from roots of *Arabidopsis* growing in CAS soil (Bai *et al.*, 2015). Bacterial growth was quantified in the presence of 50  $\mu$ M scopoletin or fraxetin, a concentration within the physiological range of coumarins observed within roots (Siwinska *et al.*, 2014; Siwinska *et al.*, 2018; Stringlis *et al.*, 2018). We observed prevalent growth inhibition to a range of degrees in the presence of fraxetin, and to a lesser extent in response to scopoletin (Figure 2.4). This indicates that fraxetin exerts variable antimicrobial activity on *Burkholderiaceae* strains. This strain-specific variation potentially explains part of the ASV-level shifts within *Burkholderiaceae* observed between WT and coumarin-deficient plants in iron-limiting soil.



**Figure 2.4 Overnight growth of *Burkholderiaceae* bacterial strains in the presence of scopoletin or fraxetin.** Optical density (OD) of cultures was normalized to the OD of each strain in the absence of coumarins. Significant differences ( $p \leq 0.05$  by Tukey's HSD) in growth compared to the control are indicated for scopoletin (S\*) and fraxetin (F\*) to the right of each strain. Data are averages of 2–4 experiments, each with 2–3 technical replicates, per strain.

### 2.3.3 Taxonomically diverse root commensals improve iron-limiting plant performance

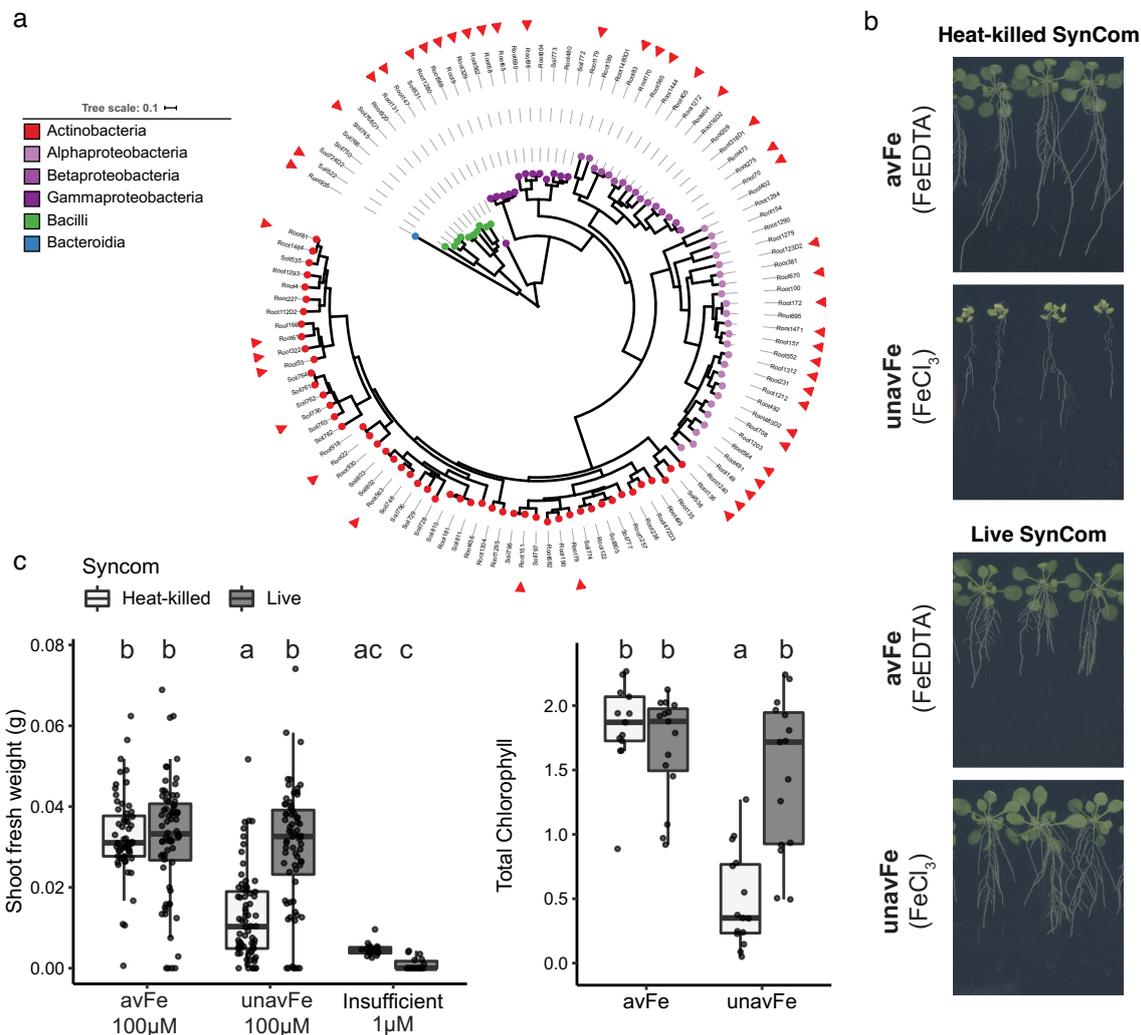
To assess the impact of root commensals on plant performance under iron-limiting conditions, we employed an agar medium-based gnotobiotic system which allows control over both iron mobility and the presence of bacterial commensals. In this system, nutrient medium (1/2 MS) is strongly buffered at pH 7.4, similar to the pH of iron-limiting IS soil. Iron is provided at 100  $\mu$ M in one of two forms: available iron (avFe) FeEDTA, a complex which retains solubility even at alkaline pH, or an unavailable form (unavFe) FeCl<sub>3</sub>, which forms Fe(OH)<sub>3</sub> and is highly insoluble at alkaline pH. Providing unavFe mimics iron-limiting conditions in calcareous and alkaline soils such as IS; iron

is present but recalcitrant due to extremely low solubility and must be actively mobilized for utilization. Using this gnotobiotic system, we reconstituted plants with a synthetic community (SynCom) of bacterial commensals to assess the impact of the microbiota on iron-limiting plant performance. To achieve this, we took advantage of a diverse culture collection of bacterial commensals isolated from *A. thaliana* roots grown in CAS (Bai *et al.*, 2015). We designed a taxonomically diverse SynCom of 115 members which reflects the root bacterial diversity observed by culture-independent methods at high taxonomic ranks (Figure 2.5a; Table S2 in the published [journal version article](#)) SFW and chlorophyll content were measured as readouts of plant performance and as a proxy measurement for iron nutritional status. A growth deficit and leaf chlorosis (Figure 2.5b, c), symptoms of iron starvation, were observed in plants grown on unavFe. Furthermore, elemental and transcriptomic analyses revealed decreased leaf iron content and induction of iron import components FRO2 and IRT1, and repression of ferritins, iron storage proteins (Figure S7 in the published [journal version article](#)). Together, these data confirm a robust induction of iron deficiency in this experimental system. We found that reconstitution of the microbiota with a bacterial SynCom greatly improved both SFW and chlorophyll content of plants grown on unavFe. This beneficial interaction was unexpected, given that iron starvation-induced coumarins exert antimicrobial activity against some commensals (Figure 2.4; Wang *et al.*, 2014; Yang *et al.*, 2016; Stringlis *et al.*, 2018a; Voges *et al.*, 2019). Iron starvation was also induced by reducing the supply of soluble iron to an insufficient amount (1  $\mu$ M FeEDTA). However, commensal-mediated improvement of plant performance was not observed in plants grown on media containing only insufficient soluble iron (Figure 2.5c). Plant growth limitation on unavailable iron and growth rescue by a live SynCom were maintained when plant roots were shielded from light (Figure S4 in the published [journal version article](#)), supporting the robustness of this experimental system. These results suggest that bacterial commensals can improve plant performance by improving access to immobile sources of iron.

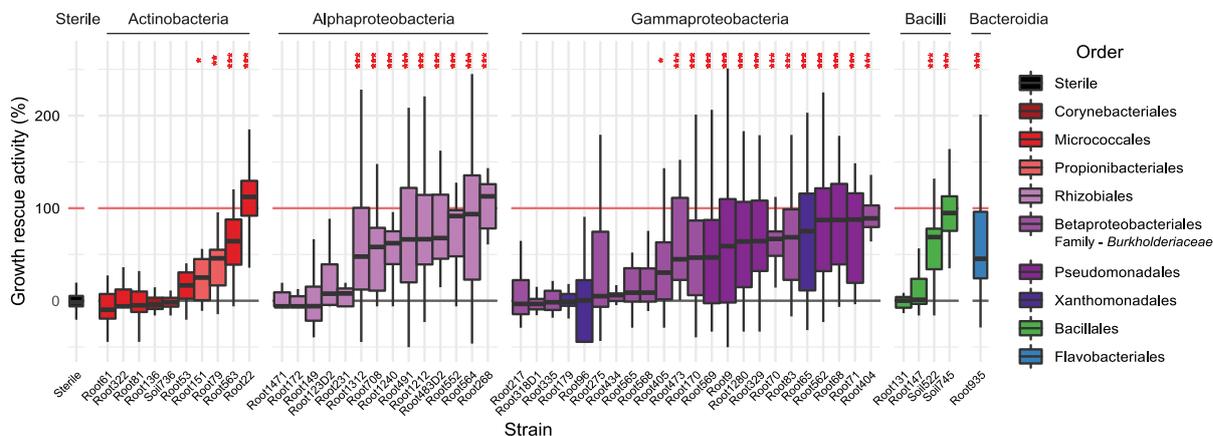
To survey the ability of various taxa to improve iron-limiting plant growth, we grew plants in mono-associations with bacterial strains on unavFe. Fifty-three taxonomically diverse SynCom strains were tested for their ability to rescue iron-limiting plant growth (Figure 2.5a, red arrows; Table S2 in the published [journal version article](#)). Within each broader taxonomic lineage, we observed growth-rescuing strains as well as strains lacking this ability (Figure 2.6), demonstrating the ubiquity of this beneficial activity as well as the strain-specific variation within all core taxonomic lineages of the root microbiota. Thus, in a community context, the capacity of bacterial commensals to rescue plant growth under iron-starved conditions is functionally redundant. Furthermore, the ability of these strains to grow in the presence of scopoletin and fraxetin was found not to correlate with plant growth rescue capacity, indicating that this plant-beneficial trait does not require selection via coumarin antimicrobial activity (Figure S4 in the published [journal version article](#)).

(Figure on next page) **Figure 2.5 Taxonomically diverse root commensals improve iron-limiting plant performance.** (a) Phylogenetic tree of 115-strain SynCom derived from At-RSPHERE culture collection (Bai *et al.*, 2015) used for microbiota reconstitution. Red arrows indicate strains used in Figure 2.6. (b) Representative images of plants grown for two weeks on media containing available (avFe) and unavailable (unavFe) forms of iron inoculated with live SynCom or heat-killed control. (c) SFW and total chlorophyll quantification of Col-0 plants after two weeks of growth on indicated iron conditions. Data are pooled from

three experiments with avFe and unavFe: n=42–54 plants per condition, and chlorophyll measured in pooled samples, n=13–15 per group. Insufficient iron data are from one experiment, n = 18 plants. Letters indicate significant pairwise differences between groups ( $p\text{-adj} \leq 0.05$  by Dunn's pairwise comparison with Bonferroni correction for SFW, and Tukey's HSD corrected for multiple comparisons for chlorophyll content).



(Caption on previous page) **Figure 2.5 Taxonomically diverse root commensals improve iron-limiting plant performance.**



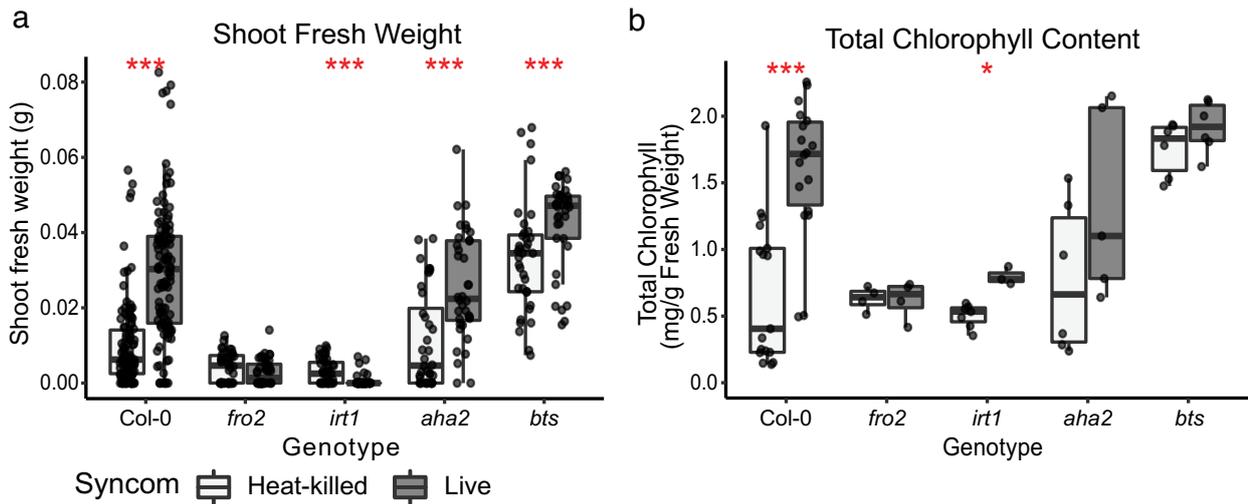
(Caption on next page) **Figure 2.6 Iron-limiting growth rescue activity of SynCom strains in mono-association.**

(Figure on previous page) **Figure 2.6 Iron-limiting growth rescue activity of SynCom strains in mono-association.** SFW was measured and plotted as percent growth rescue of bacteria-inoculated plants on unavFe compared to the growth deficit between sterile plants on avFe vs unavFe. Black and red lines indicate 0% (axenic plants on unavFe) and 100% growth rescue (axenic plants on avFe), respectively. Data are pooled from 1-4 experiments per strain and normalized to respective sterile controls (n = 18 plants per experiment). Asterisks indicate significance from sterile plants by Wilcoxon Ranked Sum test with fdr-adjustment (\*, \*\*, \*\*\*, indicate  $p\text{-adj} \leq 0.05, 0.01, \text{ and } 0.001$ , respectively).

#### 2.3.4 Microbiota-mediated plant growth rescue occurs via reductive import of iron and requires fraxetin secretion

We utilized *A. thaliana* mutant lines deficient in genes involved in iron uptake and homeostasis to determine their importance for bacteria-mediated growth rescue under iron limitation. Mutants in components of the reduction-based iron uptake system (*fro2* and *irt1*), rhizosphere acidification (*aha2*), and a negative regulator of the iron starvation response (*bts*) were grown on unavFe in the presence of a live SynCom, or a heat-killed SynCom as a negative control. Genotypes *fro2* and *irt1* displayed an exaggerated growth deficit and leaf chlorosis (Figure 2.7), consistent with their reported hypersensitivity to iron starvation (Robinson *et al.*, 1999; Vert *et al.*, 2002). In contrast to WT plants, addition of the bacterial SynCom was unable to improve the phenotype of these iron import mutants. Performance of *irt1* plants was further reduced in the presence of the SynCom. This indicates that without the iron import channel, plants may be unable to compete with bacterial commensals for access to the already limited pool of available iron. Moreover, *bts* plants, which are tolerant to iron deficiency (Selote *et al.*, 2014; Hindt *et al.*, 2017), were larger than WT plants on unavFe, not chlorotic, and still displayed slightly improved performance when inoculated with SynCom. No difference was observed between *aha2* and WT controls, however, indicating that plant-mediated rhizosphere acidification is not rate-limiting for commensal-mediated plant growth rescue in strongly buffered alkaline conditions. When grown on avFe, the SynCom did not improve plant performance in any of these genotypes (Figure S5 in the published [journal version article](#)). These results validate our gnotobiotic system for microbiota reconstitution under iron-limiting conditions, confirm that the growth limitation and chlorosis on unavFe is due to iron starvation, and suggest that improved plant performance in the presence of commensals depends on the plant's endogenous system for iron reduction and import.

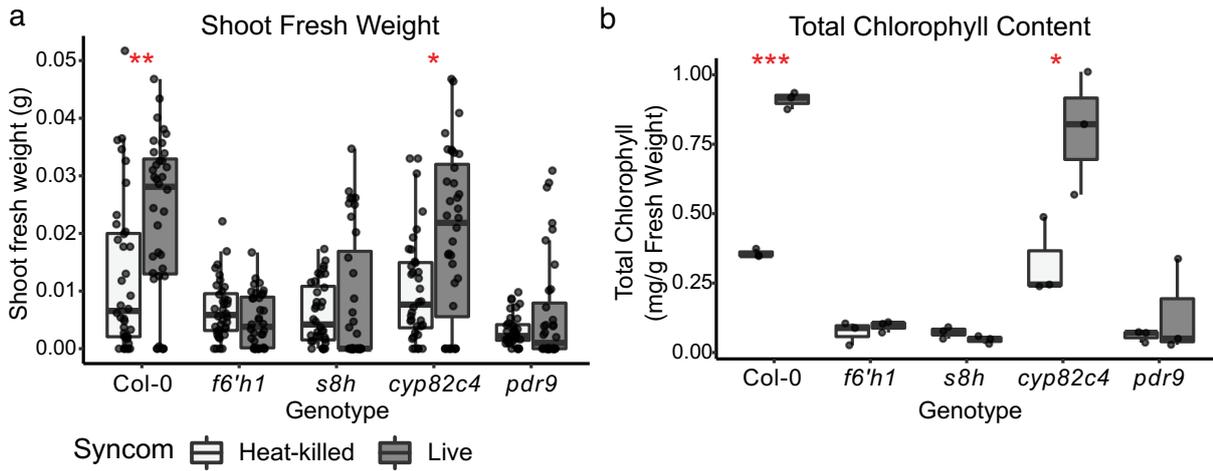
We next investigated the role of coumarins in commensal-mediated plant growth rescue under iron limitation. The addition of SynCom provided no benefit to plant growth or chlorophyll content of *f6'h1*, *s8h*, or *pdr9* plants grown on unavFe (Figure 2.8). In contrast, the SynCom improved performance of *cyp82c4* plants similar to WT. These data suggest that plant biosynthesis of scopoletin and/or fraxetin (catalysed by F6'H1 and S8H, respectively) and their secretion (through PDR9) are required for bacteria-mediated plant growth rescue under iron limitation, while sideretin (synthesized by CYP82C4) is dispensable. No growth promotion by the SynCom was observed in these genotypes grown on avFe (Figure S5 in the published [journal version article](#)).



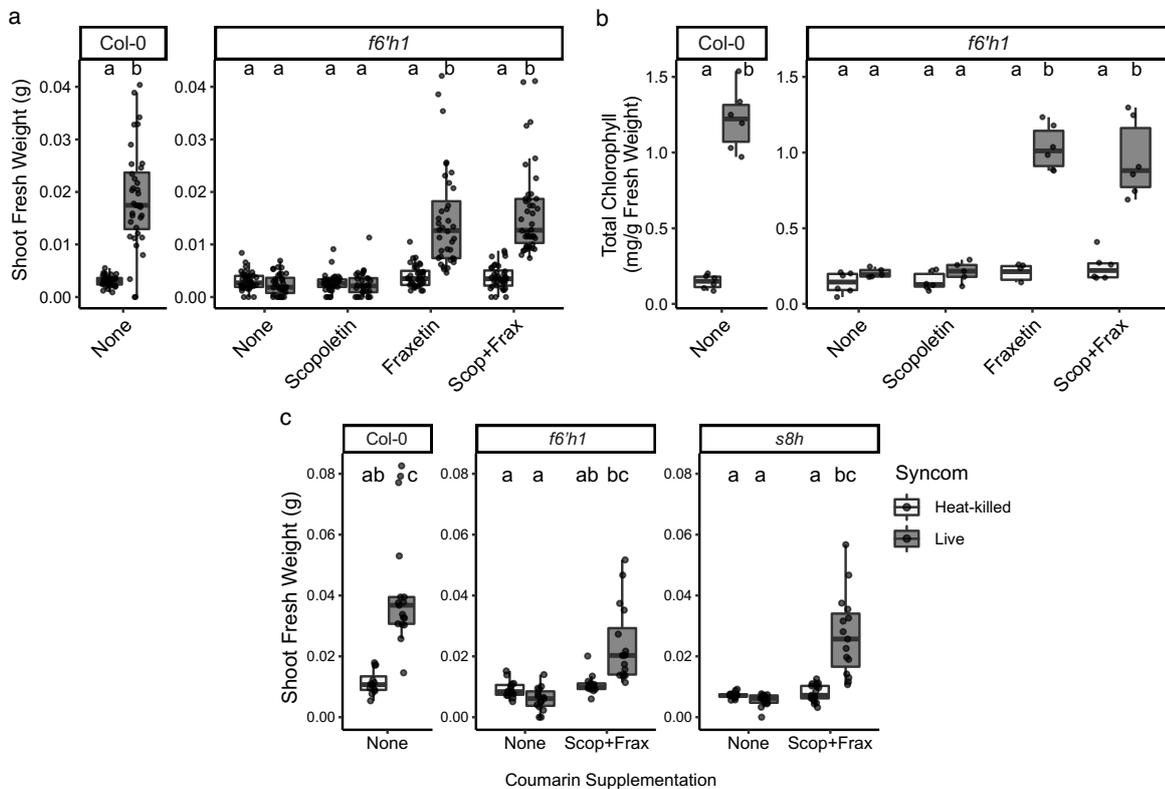
**Figure 2.7 Microbiota-mediated plant growth rescue occurs via the reductive import of iron. (a)** SFW and **(b)** leaf chlorophyll content of indicated mutants in the reductive import of iron pathway grown on unavFe media inoculated with heat-killed or live bacterial SynCom. Total chlorophyll content was measured in pooled-leaf samples from six plants. Data are from two independent experiments per genotype ( $n = 36$  plants, 6 chlorophyll samples). Each experiment included Col-0 control ( $n = 90$  plants, 18 chlorophyll samples). Asterisks indicate significance between heat-killed- and live SynCom inoculated groups by Wilcoxon Ranked Sum test for SFW and Student's T-test for chlorophyll content (\*, \*\*, \*\*\*, indicate  $p \leq 0.05$ , 0.01, and 0.001, respectively).

To further assess the roles of scopoletin and fraxetin in commensal-mediated plant growth rescue, we chemically complemented *f6'h1* plants by supplementing the growth medium with each compound (Figure 2.9a, b). Addition of scopoletin did not improve plant performance, while fraxetin fully restored the ability of the SynCom to improve both plant growth and leaf chlorophyll content in coumarin-deficient *f6'h1* plants. That scopoletin was unable to complement *f6'h1* plants suggests that external scopoletin is not sufficiently taken up by roots and converted to fraxetin after secretion. Furthermore, the ability of SynCom to rescue growth of *s8h* plants was also fully restored by supplementation with scopoletin and fraxetin (Figure 2.9c). This confirms that fraxetin is the necessary coumarin structure type for commensal-mediated growth rescue. Notably, supplementation with coumarins failed to rescue growth or chlorophyll content of germ-free *f6'h1* or *s8h* plants at 50  $\mu\text{M}$ , a concentration lower than has been used to rescue iron starvation by directly mobilizing iron. This indicates that commensal-mediated improvement in iron-limiting plant growth is induced by fraxetin concentrations lower than are required for sufficient mobilization of iron in axenic conditions. Together, these results confirm that secreted fraxetin is both necessary and sufficient to elicit growth rescue activity from bacterial commensals under iron limitation. These findings argue for an indirect activity of fraxetin in recruiting commensal-mediated mobilization of recalcitrant iron pools in addition to its direct iron-mobilizing activity.

(Figure on next page) **Figure 2.8 Plant biosynthesis and secretion of fraxetin is necessary for microbiota-mediated growth rescue. (a)** SFW and **(b)** leaf chlorophyll content of indicated coumarin biosynthesis and export mutants grown on unavFe media inoculated with heat-killed or live bacterial SynCom. SFW data are from two experiments ( $n = 36$  plants). Chlorophyll content is from one experiment ( $n=3$  pooled leaf samples). Asterisks indicate significance between heat-killed- and live SynCom-inoculated groups by Wilcoxon Ranked Sum test for SFW and Student's T-test for chlorophyll content (\*, \*\*, \*\*\*, indicate  $p \leq 0.05$ , 0.01, and 0.001, respectively).



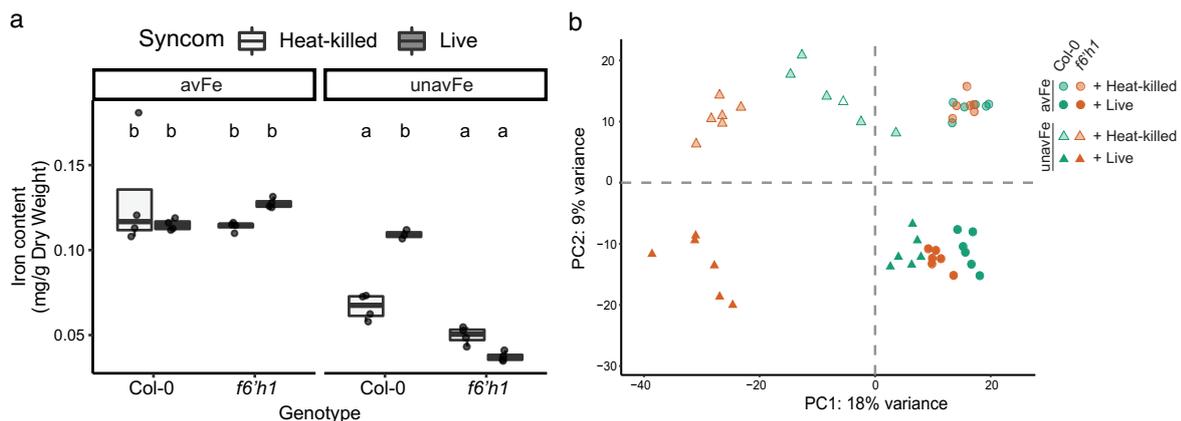
(Caption on next page) **Figure 2.8 Plant biosynthesis and secretion of fraxetin is necessary for microbiota-mediated growth rescue.**



**Figure 2.9 Supplementation with fraxetin restores microbiota-mediated growth rescue of *f6'h1* and *s8h* plants.** (a) SFW and (b) leaf chlorophyll content of Col-0 plants, and *f6'h1* plants grown on unavFe supplemented with 50  $\mu$ M scopoletin and/or fraxetin and inoculated with heat-killed or live SynCom. (c) SFW of Col-0, and *f6'h1* and *s8h* plants grown on unavFe supplemented with 50  $\mu$ M scopoletin and fraxetin and inoculated with heat-killed or live SynCom. Data in (a) and (b) are from two experiments ( $n = 30$ -42 plants, 6 pooled-leaf chlorophyll samples). Data in (c) are from a single experiment ( $n = 18$  plants). Letters indicate significant pairwise differences between groups ( $p$ -adj  $\leq 0.05$  by Dunn's pairwise comparison with Bonferroni correction for SFW, and Tukey's HSD corrected for multiple comparisons for chlorophyll content).

### 2.3.5 Coumarins and the microbiota interact to alleviate iron starvation

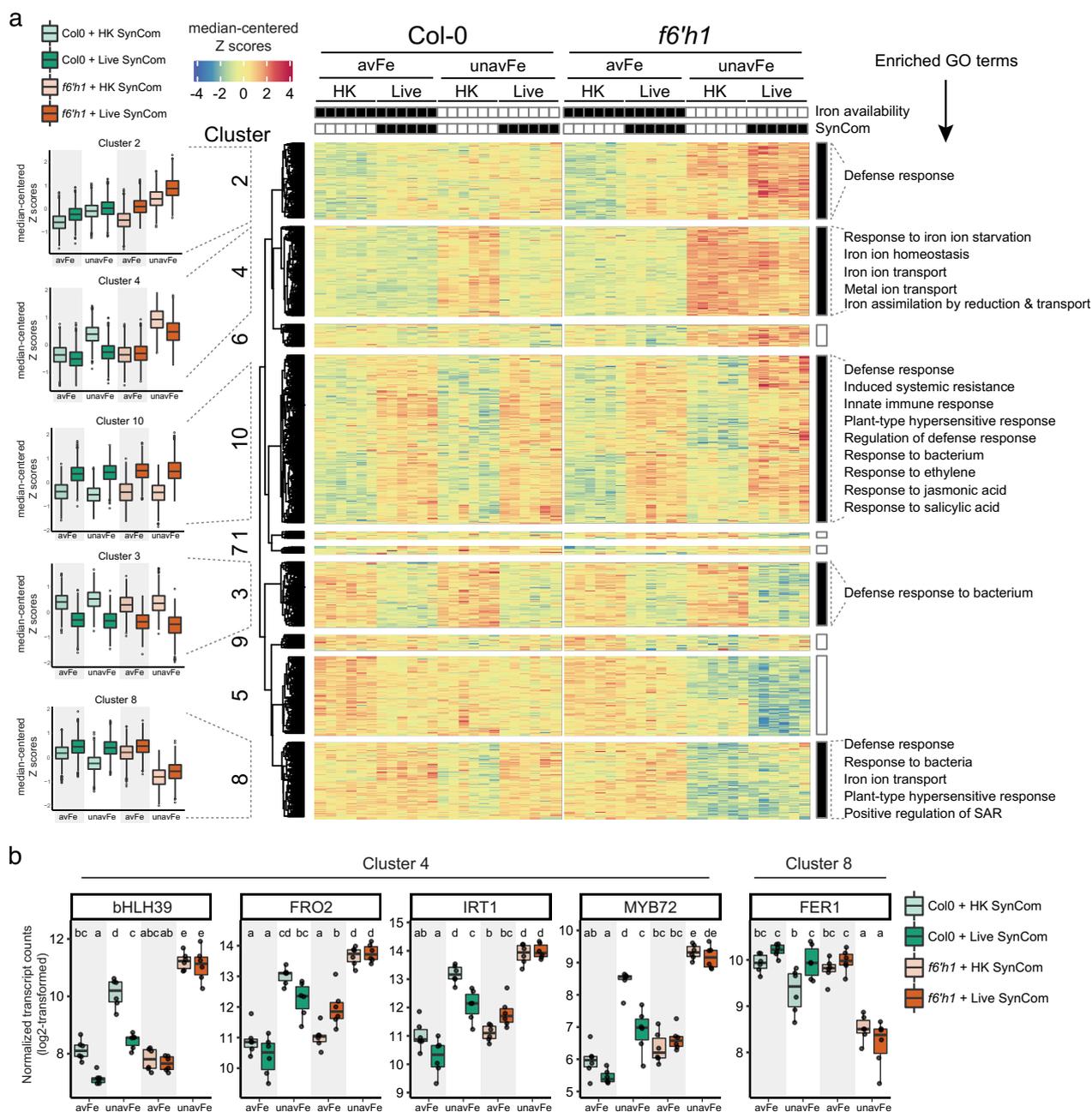
To determine if the observed iron starvation symptoms and their improvement by root commensals correlated with plant iron status, we measured leaf elemental content by inductively coupled plasma mass spectrometry (ICP-MS) (Figure 2.10a and Figure S6 in the published [journal version article](#)). Leaf iron concentration was significantly reduced on unavFe, confirming iron deficiency. Furthermore, addition of live SynCom to WT, but not *f6'h1* plants, restored plant iron content to replete avFe levels. However, addition of SynCom had no impact on plant iron content on avFe. These results confirm that the plant growth rescue activity by commensals during iron starvation is due to improved iron nutrition. We next sought to determine if this nutritional benefit was a result of microbial stimulation of the plant iron deficiency response or increased iron availability in the presence of the SynCom. To identify plant pathways responding to the presence of coumarins and bacterial commensals under different iron regimes, we performed analysis of the whole-root transcriptome of WT and *f6'h1* seedlings in our gnotobiotic system. An earlier time point (1 week) was chosen to observe potential stimulation of iron deficiency response genes by the SynCom (Verbon *et al.*, 2019) that may lead to plant growth rescue. At this time point, SynCom-mediated plant growth rescue was observed in WT, but growth was still comparable between WT and *f6'h1* plants (Figure S7 in the published [journal version article](#)).



**Figure 2.10 A bacterial SynCom improves plant iron nutrition and relieves the iron deficiency response.** (a) Shoot iron content of Col-0 and *f6'h1* plants grown on avFe and unavFe media with a live SynCom or heat-killed control measured by ICP-MS (n = 3-4 pooled plant samples per group). (b) PCA ordination of sample distances between root transcriptional profiles of Col-0 and *f6'h1* plants grown for one week on avFe or unavFe media inoculated with a live SynCom or heat-killed control. Data are from two pooled experiments (n = 6 samples pooled from 6 plant roots each). Letters in (a) indicate significant pairwise differences between groups (p-adj ≤ 0.05 by Tukey's HSD corrected for multiple comparisons).

Plotting transcriptome sample distances by PCA (Figure 2.10b) revealed that the supplied iron form was the largest determinant of dissimilarity (PC1, 18% of variance), followed by SynCom status (PC2, 9% of variance). When grown on avFe, both Col-0 and *f6'h1* plants clustered together, but separation was observed based on SynCom status. This indicates that a live SynCom impacts host transcriptional responses when iron is available, independently of plant coumarin status. The transcriptomes of plants grown on unavFe, however, were distinct from those of plants grown on avFe, and further separated based on both genotype and SynCom status. When inoculated with heat-killed SynCom, both WT and *f6'h1* on unavFe separated from the avFe cluster (upper-left quadrant). A larger genotype-driven separation was observed between Col-0 and *f6'h1* plants when

inoculated with live SynCom on unavFe. Remarkably, WT plants inoculated with live SynCom on unavFe clustered closely with SynCom-inoculated WT plants on avFe (lower-right quadrant), while *f6'h1* plants remained in the lower-left quadrant, clearly separated from the avFe cluster. This pattern indicates that the transcriptional responses to growth on unavFe are more pronounced in *f6'h1* plants than WT, consistent with their hypersensitivity to iron starvation. Indeed, more iron-responsive differentially-expressed genes (DEGs) were detected in *f6'h1* than WT plants (Figure S7 in the published [journal version article](#)). Furthermore, the number of iron-responsive DEGs indicates that the iron starvation-induced response was mitigated by the addition of SynCom in Col-0 but not *f6'h1* plants.



**Figure 2.11 The bacterial SynCom modulates a subset of defence genes in a coumarin-dependent manner. (a)** Heat map of median-centered Z scores for 2,440 DEGs identified across samples, arranged by *k*-means clustering. Significantly enriched iron homeostasis-related and defence-related GO terms of pertinent clusters are indicated on right of heat map. GO analysis was performed by comparing the indicated DEG cluster to the whole transcriptome ( $p\text{-adj} \leq 0.05$ ). **(b)** Expression of select iron deficiency response marker

genes. Data are log<sub>2</sub>-transformed, normalized counts. Letters in **(b)** indicate significant pairwise differences between groups ( $p\text{-adj}\leq 0.05$  by Tukey's HSD corrected for multiple comparisons).

We performed *k*-means clustering of all transcripts based on expression pattern and identified differentially-expressed genes (DEGs) to investigate the interaction between iron availability, SynCom, and genotype ([Figure 2.11a](#)). The identified clusters showed iron- and SynCom-responsive gene sets (plots left of heat map clusters). Cluster 4 revealed a set of genes activated in axenically-grown Col-0 and *f6'h1* plants in response to unavFe, which were more strongly induced in *f6'h1* plants. Furthermore, their expression is reduced to homeostatic levels in Col-0 plants upon addition of SynCom but remained elevated in *f6'h1* plants. This gene set corresponds to iron-responsive genes that were also responsive to SynCom in a coumarin-dependent manner. Gene ontology analysis revealed that this cluster is enriched for genes belonging to the iron starvation response, iron homeostasis, and metal ion transport ([Figure 2.11a](#), annotations right of heatmap). Genes in cluster 8 displayed the inverse pattern: downregulated on unavFe and restored in the presence of SynCom in Col-0, but not *f6'h1*. We compared our DEGs to a list of 25 previously identified core iron starvation marker genes ([Mai et al., 2016](#)) ([Figure S7](#) in the published [journal version article](#)). Of the 12 genes reported to be induced under iron starvation, 11 were found in cluster 4, while 7 out of 13 genes reported to be down-regulated under iron starvation were present in cluster 8. The cluster assignment and expression patterns ([Figure 2.11b](#) and [Figure S7](#) in the published [journal version article](#)) of selected iron homeostasis regulators and coumarin biosynthesis genes revealed that these genes are iron- and SynCom-responsive in a genotype-dependent manner. Importantly, the addition of live SynCom did not stimulate expression of iron deficiency genes. Iron starvation-upregulated genes (including bHLH39, FRO2, IRT1, and MYB72), as well as iron starvation-downregulated genes (including FER1) displayed expression patterns consistent with alleviation of the iron deficiency response by the SynCom in WT plants. This indicates that, rather than biostimulation of plant iron uptake, iron nutrition is improved by a commensal mechanism. In addition, the inability of the SynCom to alleviate the iron-starvation response in *f6'h1* roots (cluster 4 and 8 genes; [Figure 2.11](#)) supports a mechanism by which secreted coumarins are required to elicit microbiota-mediated iron nutrition, rather than two parallel mechanisms exerted by coumarins and the microbiota independently. Together, these data reveal a robust induction of iron starvation in plants grown on unavFe, which is alleviated by commensals in WT, but not in *f6'h1* plants. This implies the existence of a coumarin- and microbiota-dependent mechanism that improves plant performance via iron nutrition.

Analysis of SynCom-responsive genes also revealed a core set of DEGs common to both genotypes independently of iron availability (clusters 3 and 10, [Figure 2.11a](#); [Figure S7](#) in the published [journal version article](#)). These clusters were enriched for genes related to defence responses and response to bacteria. Interestingly, genes associated with these terms were also significantly enriched in clusters 2 and 8, the expression patterns of which are dependent on host genotype, iron status and SynCom ([Figure 2.11a](#)). The presence of immune-related genes in clusters 2 and 8 suggests that, in addition to the core SynCom-responsive genes, a subset of defence genes is regulated by the presence of commensals in a coumarin-dependent manner. Collectively, these results show that both coumarin secretion and the root microbiota have profound impacts on plant transcriptional responses to iron deprivation. Furthermore, this emergent interaction between

coumarins and the microbiota improves plant performance through bolstered iron nutrition, resolving the iron starvation response and regulating a subset of defence-related genes.

## 2.4 Discussion

Our results reveal unexpected impacts of root-secreted coumarins governing plant-bacteria interactions, including soil type-dependent alteration of root microbiota composition, elicitation of a commensal-mediated mechanism of plant iron nutrition, and regulation of a subset of defence genes. We show that *fb'h1* and *s8h* plants, which are deficient in the biosynthesis of scopoletin and fraxetin, assemble an altered root bacterial community. Individual members of the *Burkholderiaceae*, core members of the plant root microbiota that often exert plant-beneficial activities (Eberl & Vandamme, 2016; Thiergart *et al.*, 2020), are impacted by fraxetin in a strain-specific manner in both roots and growth in culture. This strain-specific sensitivity may in part underlie ASV-level changes in abundance on iron-limiting IS. The greater impact on the microbiota observed in *fb'h1* compared to *s8h* plants, in terms of deASVs detected and their RA, suggests that both scopoletin and fraxetin impact bacterial microbiota assembly. Indeed, a metagenome analysis indicated an altered microbial multi-kingdom assemblage and provided evidence for selective scopoletin anti-fungal activity against soil-borne fungal pathogens *in vitro* (Carpinella *et al.*, 2005; Kai *et al.*, 2006; Ba *et al.*, 2017; Stringlis *et al.*, 2018b). However, we cannot exclude that iron malnutrition in *fb'h1* plants has additional consequences on the microbiota.

Plant performance data, coupled with elemental content and transcriptomic analysis, confirms that benefits conferred by commensals under iron limitation occur via improved iron nutrition. In contrast to this beneficial interaction, under low phosphate concentrations *A. thaliana* must compete with a bacterial SynCom for access to the macronutrient, requiring integration of phosphate starvation and defence responses (Castrillo *et al.*, 2017). We similarly observe emergent effects of coumarins and the microbiota on a subset of defence-related genes, indicating potential trade-offs between growth and defence. Of note, the impact of commensal communities on plant performance when phosphate is present in unavailable forms, as is characteristic of many soils, has not been tested. Importantly, in our system, growth promotion by microbes is observed only when iron is present but immobile, conditions characterizing most iron-limiting soils. Thus, our results highlight the importance of studying plant nutritional phenotypes in systems closely mimicking natural conditions, including the presence of commensals and defined forms of soil minerals that are unavailable to the plant host. To the best of our knowledge, experimental evidence for clear plant nutritional benefits from commensals in a community context has not been reported before. This is a significant step forward in understanding how plant nutrition and productivity can be bolstered by harnessing endogenous soil microbes.

The presence of this beneficial activity across all core taxonomic lineages of the *A. thaliana* bacterial microbiota suggests that the underlying molecular mechanism(s) evolved independently rather than by common descent. As growth rescue depends on plant expression of the iron reductive import machinery, but does not involve microbial stimulation of its expression, the mechanism(s) must function upstream of reduction and import at the root surface. Multiple bacterial molecules can mobilize insoluble iron and are potentially utilized by plants, including siderophores and other metabolites, though the ability of plants to utilize microbially-mobilized iron in the context of an

intact microbiota has yet to be shown. The nutritional benefits provided by commensals requires plant-secreted fraxetin, but is independent of its antimicrobial activity. This suggests that the impact of coumarins on root microbiota composition and on commensal-mediated iron nutrition may be independent mechanisms. This would indicate that coumarins have multiple roles in mediating host-microbe interactions. Importantly, our results were obtained using a bacterial culture collection derived from *A. thaliana* roots grown in CAS, in which coumarin status did not affect plant performance or microbiota structure. The observation that taxonomically diverse commensals isolated from an iron-replete soil are capable of rescuing iron-limiting plant growth further suggests the involvement of microbial functions which are ubiquitous across soil types and can be elicited by fraxetin. Future work with commensal culture collections derived from soils with contrasting iron availability are required to directly link plant phenotypes in natural calcareous soils and gnotobiotic systems.

Root-secreted coumarins are inducible under iron starvation and mediate an interaction between the host and commensals that improves host iron nutrition. This genotype-environment interaction strongly suggests that the root microbiota is an integral component of plant edaphic adaptation to growth in iron-limiting soil. Quantitative variation in coumarin production has been demonstrated among *A. thaliana* accessions (Siwinska *et al.*, 2014; Tsai *et al.*, 2018), and was shown to correlate with performance under iron-limitation (Tsai *et al.*, 2018). As coumarins are both ubiquitously present and chemically diverse among flowering plants (Bourgaud *et al.*, 2014; Rajniak *et al.*, 2018; Krieger *et al.*, 2018), our findings provide an ecological framework for examining the consequences of their evolutionary diversification on microbiota-mediated mineral nutrition of plant hosts.

## 2.5 Materials and methods

### 2.5.1 Experimental model and subject details

Cologne agricultural soil (CAS) was obtained from a local site (GPS code : 50.958 N, 6.856 E) that has not been exposed to agriculture for >15 years. Italian soil (IS) was obtained from an organic vineyard in Tebano, Italy (GPS code : 44.292 N, 11.784 E) which has been maintained since 2007 without irrigation or fertilization. Soil was homogenized, sieved, and stored at 4°C until used in experiments. All *A. thaliana* genotypes used in this study were in the Columbia wild-type (Col-0, N60000) background. Mutants related to coumarin biosynthesis (*f6'h1-1*, SALK\_132418C; *s8h-2*, SM\_3\_23443; *cyp82c4-1*, SALK\_001585) and export (*pdr9-2*, SALK\_050885), regulation of the iron starvation response (*bts-1*, SALK\_016526), and iron reductive import (*aha2-4*, SALK\_082786; *fro2*, also known as *frd1-1* (Robinson *et al.*, 1999), and *irt1-1* (Vert *et al.*, 2002)) were employed in this study and are available from the Arabidopsis Biological Resource Center (ABRC). Each of these genes are expressed in roots.

All bacterial strains used in this study were previously described (Bai *et al.*, 2015) and are summarized in Table S2 in the published [journal version article](#). Species phylogenetic trees were generated with iTOL version 5.5 (Letunic & Bork, 2007) from previous whole genome taxonomic classification (Bai *et al.*, 2015). Each of these strains was originally isolated from *A. thaliana* roots grown in CAS soil. Strains were stored in 20% glycerol stocks and cultured on 50% tryptic soy agar (TSA) plates and 50% tryptic soy broth (TSB) at 25°C.

### 2.5.2 Plant growth conditions

Seeds were surface-sterilized with 70% ethanol for 15 minutes under agitation, followed by two washes with 70% ethanol, one with 100% ethanol, and three with sterile distilled water. Sterilized seeds were stratified at 4°C in the dark for 2–3 days either imbibed in water (for soil experiments) or on agar media plates (for agar-media experiments) before transfer to growth conditions.

In soil system, surface-sterilized, stratified seeds were germinated in 7x7 cm square pots filled with CAS or IS. Pots were watered from the top with non-sterile distilled water every 2 days. Plants were grown in the greenhouse under long-day conditions (16hrs light, 8 hours dark). Pots were distributed at random within trays. and shuffled periodically to minimize edge and location effects. In agar-media system, surface-sterilized seeds were sown on plates containing 1% agar (Agar, granulated, Difco) in 50% Murashige and Skoog (MS) medium with vitamins (2.2 g/L, Duchefa Biochemie) supplemented with 0.5% sucrose. After two days of stratification at 4°C, plates were positioned vertically in a climate chamber (Panasonic, MLR-352) and grown for six days (10 hours light, 21 °C; 14 hours dark, 19 °C). Uniform seedlings were then transferred to experimental condition plates prepared fresh on the day of seedling transfer.

### 2.5.3 Coumarin antimicrobial activity

The antimicrobial activity of scopoletin and fraxetin (Sigma Aldrich) against single bacterial strains was assayed in liquid culture in 50% tryptic soy broth (TSB, 15g/L; Sigma Aldrich). Scopoletin and fraxetin stocks were prepared in sterile DMSO (Sigma Aldrich) and stored at -80°C. Bacterial colonies were picked from TSA plates into liquid TSB and grown for 2–3 days at 25°C with 180 rpm agitation. Liquid cultures were subcultured by diluting 1:100 into fresh TSB and incubated for 1–2 hours. In a clear flat-bottom 96-well microtiter plate, 100 µl of subculture were added to 100 µl of fresh TSB media supplemented with scopoletin or fraxetin for a final 50 µM concentration, or equivalent DMSO negative control. The microtiter plate was sealed with a clear adhesive film to prevent evaporation. Growth was monitored kinetically in a microplate reader (Infinite M200 PRO, Tecan) with 30 seconds of shaking followed by measurement of optical density (OD) at 600 nm in four locations per well every 30–60 minutes for 18–20 hours. The OD in each experiment was expressed as the average of triplicate wells per condition. Relative growth (Figure 2.4) was calculated by dividing the average final OD measurement of each strain and indicated condition by the average OD in the coumarin-free control.

### 2.5.4 Gnotobiotic system for iron limitation

Iron limitation was achieved with a modified MS medium prepared from stock solutions. Stock solutions were prepared of ethylenediaminetetraacetic acid ferric sodium salt (Fe(III)EDTA, Sigma) in distilled water, and 100mM Fe(III)Cl<sub>3</sub> (Merck) in 10 mM HCl (to prevent precipitation), sterile-filtered and stored at 4°C protected from light. A 2M stock solution of 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid, N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES) buffer (Roth) was prepared, and the pH was adjusted with KOH until a dilution to 10 mM in 50% MS resulted in a pH of 7.4 (approximately pH 8.2 for stock solution) and stored at 4°C.

As a base medium, modified 50% MS media without iron or pH buffer (750  $\mu\text{M}$   $\text{MgSO}_4$ , 625  $\mu\text{M}$   $\text{KH}_2\text{PO}_4$ , 10.3 mM  $\text{NH}_4\text{NO}_3$ , 9.4 mM  $\text{KNO}_3$ , 1.5 mM  $\text{CaCl}_2$ , 55 nM  $\text{CoCl}_2$ , 53 nM  $\text{CuCl}_2$ , 50  $\mu\text{M}$   $\text{H}_3\text{BO}_3$ , 2.5  $\mu\text{M}$   $\text{KI}$ , 50  $\mu\text{M}$   $\text{MnCl}_2$ , 520 nM  $\text{Na}_2\text{MoO}_4$ , 15  $\mu\text{M}$   $\text{ZnCl}_2$ , and 9.4mM  $\text{KCl}$ ) was prepared from individual stock solutions. Base media with 1% agar was autoclaved and cooled to 50°C before adding iron source (final 100  $\mu\text{M}$ ) and HEPES (final 10mM, pH 7.4) with constant stirring. Media were allowed to cool to ~45°C, and 45 ml were measured into a conical tube. Live or heat-killed bacteria or SynCom (preparation see below) were added to a final  $\text{OD}_{600\text{nm}} = 0.0001$ , corresponding to approximately  $10^5$  cells/ml. For coumarin complementation experiments, scopoletin and fraxetin were added to a final concentration of 50  $\mu\text{M}$ , or equivalent DMSO-only control. Media was mixed thoroughly by inverting, poured into 12\*12cm square petri dishes, dried with an open lid for 30 minutes, then allowed to solidify. Seedlings were transferred to experimental plates (six plants per plate, three replicate plates per experiment). Plates were returned to the growth chamber and grown vertically with random shuffling and re-distribution every 2–3 days for uniformity. After two weeks, SFW was measured and chlorophyll and root samples were collected. Plant growth rescue in mono-association assays (Figure 2.6) was expressed as percent growth rescue of the differential between axenic plant growth on avFe and unavFe using the formula:

$$\% \text{ Growth Rescue} = \frac{SF\ W_{\text{inoculated on unavFe}} - SF\ W_{\text{axenic on unavFe}}}{SF\ W_{\text{axenic on avFe}} - SF\ W_{\text{axenic on unavFe}}} * 100 \%$$

### 2.5.5 Microbial community data analysis

For 16S profiling, root samples were harvested and libraries were processed as in (Thiergart *et al.*, 2020). Plant roots were harvested at the early flowering stage, average 37 days after sowing. Paired-end Illumina sequencing was performed in-house using the MiSeq sequencer and custom sequencing primers. Forward and reverse sequencing reads were demultiplexed separately according to the barcode sequence and output in individual fastq files per sample. A denoising pipeline, DADA2 (v1.12.1) (Callahan *et al.*, 2016) was used to obtain the final ASV table. Raw sequencing reads were truncated to 260 bp for the forward reads and 240 bp for the reverse reads and filtered with the following parameters: *maxN=0*, *maxEE=c(2,2)*, *truncQ=2*, *rm.phix=TRUE*. Subsequently, error rates were inferred from filtered reads until convergence or exceeding a maximum consistent number of 20. Sequence variants were then inferred from the trimmed, filtered and error-corrected sequences and ASVs were obtained by merging the forward and reverse sequences together. Finally, chimeras were identified and removed when an ASV could be mapped to the left- and right-segments from two other, more abundant ASVs. Finally, the ASV table was generated by aggregating the number of reads mapped to each variant.

Analyses and visualization were performed in the R statistical environment (Version 4.0.1). Analysis was performed on samples with a sequencing depth of at least 2,000 high-quality reads. Alpha and beta diversity were calculated on ASV count tables that were rarefied to 2,000 reads. Alpha diversity (Shannon index) was calculated with using the “diversity” function in *vegan* (R package version 2.5-6) and differences were tested with ANOVA. Bray-Curtis dissimilarity index was calculated using the “vegdist” function in *vegan* and used for unconstrained ordination by Principal Coordinate Analysis (PCoA). Constrained PCoA (CPCoA) was performed with the “capscale” function in *vegan*, using the square-root distances of Bray-Curtis dissimilarity index. For ordination constrained on the interaction between genotype and soil type (Figure 2.1b), the formula

used was “distance.matrix ~Host.Genotype\*Soil”. Statistical significance of genotype separations was determined using adonis function and pairwise PERMANOVA with 999 permutations using vegan and the RVAideMemoire package.

Analyses and visualization were all performed in the R statistical environment (R Development Core Team, 2010). Analysis was performed on samples with a sequencing depth of at least 2,000 high-quality reads. Relative abundance (RA) was calculated using non-rarefied ASV count data. To calculate deASVs between coumarin pathway mutants and Col-0 WT, data were pooled from three experiments (except *s8h*, which was included in only one experiment), and filtered for ASVs found in at least three samples at a RA > 0.05%. Statistically significant differential enrichment was determined with the *edgeR* package (Robinson *et al.*, 2010) (version 3.24.3) using pair-wise genotype comparisons in a negative binomial generalized log-linear model at an FDR-adjusted p value of 0.05. Taxonomic classification of ASVs and culture collection strains was performed using the Silva 132 database (released Dec. 2017) (Quast *et al.*, 2012). Hypergeometric enrichment test was performed using the *stats* package in R. Each family was tested for over- or under-representation in the deASVs set by comparing to the list of all detected ASVs. Red asterisks indicate significance at an FDR-adjusted p value of 0.05.

#### 2.5.6 RNA extraction and RNA-seq analysis

For transcriptomic analysis, 6-day-old *A. thaliana* seedlings were transferred to avFe or unavFe media with live or heat-killed SynCom as above and grown for eight days. Roots from six plants (one plate) were pooled for one replicate, with a total of three replicates per condition in each of two experiments (final n=6). Roots were homogenized with Lysing Matrix E and TissueLyser II (30 beats per second for 2x30 s; Qiagen) and RNA was extracted with the miRNeasy Mini Kit (Qiagen) according to the manufacturer’s instructions. RNA quality was determined using a 2100 Bioanalyzer (Agilent Technologies). Preparation of Illumina sequencing libraries was conducted by the Max Planck Genome Center. Sequences were generated using the Illumina HiSeq2500 platform. Approximately 20M paired-end reads with a length of 150 bp were obtained per sample in one experiment, and 8M per sample in the second.

Initial paired-end RNA-Seq reads were pre-processed using *fastp* (Chen *et al.*, 2018). High quality reads were aligned to *A. thaliana* reference transcriptome (TAIR 10) using *kallisto* (Bray *et al.*, 2016) with default settings. After removal of low abundant transcripts, 35,886 transcripts were imported using the *tximport* R package (Soneson *et al.*, 2016). Batch effects were detected and removed using the *SVA* R package (Leek *et al.*, 2012). Differential gene expression analysis was performed using *DESeq2* (Love *et al.*, 2014) by comparing the avFe and unavFe treatment with live or heat-killed SynCom in WT and *f6’h1* plants, respectively. DEGs were selected with the threshold  $\log_2\text{FoldChange} > \log_2(1.5)$  and an adjusted p-value < 0.05.

Scaled counts normalized to library size were generated using *DESeq2* (‘rlog’ function) and transformed as median-centered z-score (by transcripts, ‘scale’ function). Then, z-scores were used to conduct k-means clustering for all transcripts. The cluster number (k=10) was determined by sum of squared error and Akaike information criterion. Transcripts with similar expression patterns were grouped in the same cluster. Differentially expressed transcripts and cluster results were visualized using heatmaps generated with the *ComplexHeatmap* package in R (Gu *et al.*, 2016). Gene

expression in individual plots ([Figure S7](#) in the published [journal version article](#)) was plotted using scaled counts data. Gene-set enrichment analyses were performed with the *goseq* package ([Young et al., 2010](#)) with gene ontology annotations from the Gene Ontology Consortium ([Ashburner et al., 2000](#); The Gene Ontology, 2019) (September 2019).

## 2.6 Data and code availability

Raw sequencing data of 16S rRNA profiling experiments and RNA-Seq transcriptomic data have been deposited in the European Nucleotide Archive (ENA) under the accession number PRJEB3866. Datasets and scripts necessary to reproduce analyses and generate figures have been deposited to Mendeley Data with the DOI: 10.17632/tkdn6zbx7k.1. Scripts for RNASeq analysis and heatmap generation are available at [https://github.com/YulongNiu/MPIPZ\\_CJ\\_RNASeq](https://github.com/YulongNiu/MPIPZ_CJ_RNASeq).

## 2.7 Author contributions

Harbort and Hashimoto et al., 2020

P.S.-L., H.I., and M.H. conceived the project. C.J.H., M.H., H.I., and P.S.-L. designed the experiments. A.D.R. provided IS. M.H. and C.J.H. performed the root microbiota profiling experiments in natural soils. C.J.H., [R.G.](#), and R.G.-O. analyzed the microbiota profiling data. C.J.H. performed SynCom reconstitution experiments, bacterial growth assays, and RNA-seq experiments. RNA-seq data analysis was performed by Y.N., R.G.-O., and C.J.H.; R.G.-O., [R.G.](#), and Y.N. provided bioinformatic tools and support. S.K. advised and performed ICP-MS analyses. E.S.S. and M.J.E.E.V. provided scientific advice at an early stage of the project and feedback on the manuscript. P.S.-L. and R.G.-O. supervised the project. C.J.H. and P.S.-L. wrote the paper.

# Chapter 3 Coordination of microbe-host homeostasis by crosstalk with plant innate immunity

## 3.1 Abstract

Plants grown in natural soil are colonized by phylogenetically structured communities of microbes known as the microbiota. Individual microbes can activate microbe-associated molecular pattern (MAMP)-triggered immunity (MTI), which limits pathogen proliferation but curtails plant growth, a phenomenon known as the growth-defense trade-off. We report that in mono-associations, 41% (62/151) of taxonomically diverse root bacterial commensals suppress *Arabidopsis thaliana* root growth inhibition (RGI) triggered by immune-stimulating MAMPs or damage-associated molecular patterns. Amplicon sequencing of bacteria 16S rRNA genes reveals that immune activation alters the profile of synthetic communities (SynComs) comprised of RGI-non-suppressive strains, while the presence of RGI-suppressive strains attenuates this effect. Root colonization by SynComs with different complexities and RGI-suppressive activities alters the expression of 174 core host genes, with functions related to root development and nutrient transport. Furthermore, RGI-suppressive SynComs specifically downregulate a subset of immune-related genes. Pre-colonization with RGI suppressive SynComs, or mutation of one commensal-downregulated transcription factor, *MYB15*, renders plants more susceptible to opportunistic *Pseudomonas* pathogens. Our results suggest that RGI-non-suppressive and suppressive root commensals modulate host susceptibility to pathogens by either eliciting or dampening MTI responses, respectively. This interplay buffers the plant immune system against pathogen perturbation and defense-associated growth inhibition, ultimately leading to commensal-host homeostasis.

## 3.2 Introduction

Ubiquitous interactions within, and between, microbial communities and their plant hosts often shape host phenotypes and drive community diversification, leading to the conceptualization of plants and their associated microbes as discrete ecological units, or holobionts (Hassani *et al.*, 2018). Analysis of *Arabidopsis thaliana* grown in different locations has shown that plants accommodate a conserved core microbiota, microbial assemblages that represent a subset of microbes from the surrounding soil seeding inocula (Bulgarelli *et al.*, 2012; Lundberg *et al.*, 2012; Schlaeppi *et al.*, 2014; Thiergart *et al.*, 2020). While most microbiota members are commensals, a small number provide beneficial services for the host (Bulgarelli *et al.*, 2013; Jacoby *et al.*, 2017), or become pathogenic under favorable conditions. Recent studies have shed light on how specialized metabolites (Lebeis *et al.*, 2015; Hu *et al.*, 2018; Huang *et al.*, 2019; Koprivova *et al.*, 2019) and abiotic stresses (Castrillo *et al.*, 2017; Harbort *et al.*, 2020) influence host-associated microbiota. However, how microbe-host homeostasis is maintained upon perturbation remains poorly understood.

Plants have evolved a sophisticated innate immune system to protect themselves against pathogens. One arm of this system is activated by the extracellular perception of microbe/pathogen-associated molecular patterns (M/PAMPs) that are recognized by host pattern recognition receptor (PRRs). For example, the bacterial flagellin-derived epitope flg22 is detected by the cognate PRR FLS2. Both

pathogenic and beneficial bacteria can carry flg22 epitope variants (Stringlis *et al.*, 2018), resulting in MAMP-triggered immunity (MTI) (Saijo *et al.*, 2018; Wan *et al.*, 2019). MTI effectively restricts pathogen proliferation (Zipfel *et al.*, 2004), but, if unrestrained, may result in plant growth penalties, a phenomenon known as the growth-defense trade-off (Huot *et al.*, 2014). Pathogens have evolved diverse mechanisms to suppress MTI (Jones & Dangl, 2006), however, this property is not limited to harmful bacteria, as a previous report has shown that commensal Alphaproteobacteria from the *Arabidopsis* root culture collection (At-RSPHERE) (Bai *et al.*, 2015) can also override flg22-mediated root growth inhibition (RGI) (Garrido-Oter *et al.*, 2018). Similarly, the beneficial rhizobacterium *Pseudomonas simiae* suppresses more than half of the MAMP-triggered transcriptional responses in mono-association with *Arabidopsis*, possibly through acidification of the rhizosphere (Stringlis *et al.*, 2018; Yu *et al.*, 2019). However, how plants tolerate a rich diversity of commensals without compromising effective resistance to pathogens is unknown. Here, we used a bottom-up approach to show that phylogenetically diverse root commensals can modulate plant immunity, and that their combined interactions in community contexts coordinate commensal-host homeostasis under pathogen challenges (Hacquard *et al.*, 2017; Teixeira *et al.*, 2019).

### 3.3 Results

#### 3.3.1 Taxonomically widespread ability of root commensals to interfere with defense-associated growth inhibition

To facilitate screening of individual root commensals of the At-RSPHERE culture collection, we took advantage of a flg22-hypersensitive line, *pWER::FLS2-GFP* (Wyrsh *et al.*, 2015; Emonet *et al.*, 2021), in which the flg22 receptor is overexpressed but restricted to the root epidermis. This hypersensitivity leads to an enhanced signal-to-noise ratio for flg22-mediated RGI (Figure 3.1). After three weeks of co-culturing with individual bacterial isolates and flg22, 41% of the strains (62 out of 151) were found to interfere with RGI. RGI-suppressive activity was detected across all four phyla of the microbiota – Actinobacteria, Proteobacteria, Bacteroidetes, and Firmicutes – but was overrepresented among Actinobacteria and Gammaproteobacteria commensals (Figure 3.2a). Viable plate counting confirmed that the RGI non-suppressive strains still colonize roots in mono-associations (Extended Data Fig. 1b in the published [journal version article](#)). In contrast, only three strains, *Streptomyces* strains 107 and 187, and *Pseudomonas* 401, had detrimental impacts on *Arabidopsis* in mono-associations, with *Pseudomonas* 401 most severely compromising plant growth (Extended Data Fig. 1c in the published [journal version article](#)).



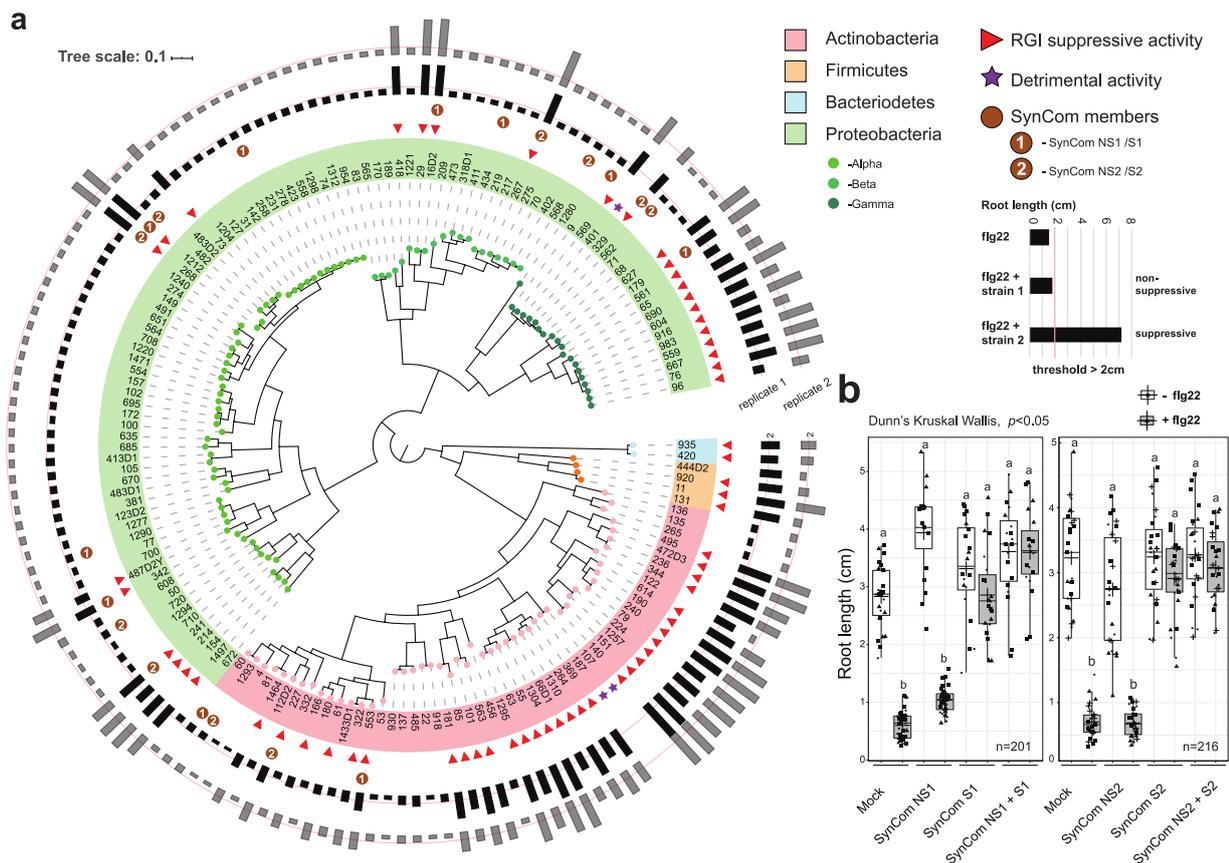
(Caption on next page) **Figure 3.1** *pWER::FLS2-GFP* transgenic *Arabidopsis* plants are hypersensitive to flg22.

(Figure on previous page) **Figure 3.1 *pWER::FLS2-GFP* transgenic Arabidopsis plants are hypersensitive to flg22.** Representative images of 3-week-old plants treated with 1  $\mu$ M flg22 grown on agar plate.

To examine whether root-derived bacteria were also able to suppress RGI elicited by an endogenous plant-derived danger-associated molecular pattern (DAMP), we treated plants with the DAMP *Atpep1*, which induces RGI and immune responses (Poncini *et al.*, 2017). Using *Atpep1*-treated Col-0 wild type (WT) plants, we found that 12 out of 13 suppressive strains, representing members from diverse taxa, retained the capacity to interfere with RGI, while none of the eight non-suppressive strains elicited this effect (Extended Data Fig. 2a in the published [journal version article](#)). Thus, phylogenetically diverse root commensals can suppress both DAMP- and MAMP-induced RGI. One isolate, *Caulobacter* strain 342, suppressed flg22- but not *Atpep1*-mediated RGI (Extended Data Fig. 2b in the published [journal version article](#)), suggesting the existence of at least two modes of RGI suppression: one interfering with both MAMP- and DAMP-induced RGI, the other possibly specific to flg22 perception.

Although germ-free *pWER::FLS2-GFP* (Wyrsh *et al.*, 2015) plants respond to flg22 treatment with enhanced RGI, compared to Col-0 on synthetic medium, no growth differences were noted between these two genotypes when grown on natural soil (Extended Data Fig. 2c in the published [journal version article](#)). Given that root growth in natural soil likely proceeds in the face of chronic exposure to MAMPs and DAMPs, as well as colonization by both suppressive and non-suppressive commensals, we speculated that the aforementioned RGI suppression phenotype may act as a dominant community trait. To test this hypothesis, we composed four independent but taxonomically similar 5-member synthetic communities (SynComs) with contrasting capacities for RGI suppression, i.e., non-suppressive SynComs (SynCom NS1 and NS2) and suppressive SynComs (SynCom S1 and S2; Supplementary Table 1 in the published [journal version article](#)). We observed RGI-suppressive activity in plants inoculated with the suppressive SynComs, but not with the non-suppressive SynComs. Furthermore, full RGI suppressive activity was retained when these commensals were combined as 10-member SynComs (Figure 3.2 b). A recent study showed that auxin-mediated RGI could be rescued by *Variovorax* commensals (Finkel *et al.*, 2020). However, our four tested SynComs neither induced RGI to a level comparable to flg22 treatment, nor did the presence of *Variovorax* 434 in SynCom NS1 rescue the flg22-mediated RGI phenotype (Figure 3.2b). Therefore, we conclude that RGI is mainly caused by flg22 treatment, and is widely suppressed by At-RSPHERE members that function dominantly in our setup.

(Figure on next page) **Figure 3.2 *At-RSPHERE* root commensals exhibit strain-specific variations to suppress flg22-mediated RGI in *pWER::FLS2-GFP* plants.** (a) Phylogenetic tree showing the distribution of strains exhibiting RGI suppressive activity. The outer rings represent root lengths of 3-week-old plants germinated on plates supplemented with 1  $\mu$ M flg22, and individual strains ( $OD_{600}=0.0005$ ), two independent replicates. The threshold for suppressive activity is indicated by the red line, i.e., root length > 2cm. (b) The impact of 4 independent 5-member SynComs (Supplementary Table 1 in the published [journal version article](#)), differential in RGI suppressive activity, on flg22-mediated RGI. 2-week-old plants were germinated on plates supplemented with 1  $\mu$ M flg22, and the indicated SynCom. Shapes represent three (SynCom NS1 +S1) and four (SynCom NS2 +S2) independent replicates in (b). n=total number of biological samples. Different letters indicate statistical significance (two-sided Dunn's Kruskal Wallis,  $p<0.05$ ). The box plots centre on the median and extend to the 25th and 75th percentiles, and the whiskers extend to the furthest point within the range of 1.5 $\times$  the interquartile range.



(Caption on previous page) **Figure 3.2 *At-RSPHERE* root commensals exhibit strain-specific variations to suppress *flg22*-mediated RGI in *pWER::FLS2-GFP* plants.**

We speculated that the co-occurrence of RGI-non-suppressive and suppressive strains might reflect a need for commensal microbes to dampen plant immunity to balance root growth and defense trade-offs. Thus, we asked if a single suppressive strain is sufficient to achieve full RGI suppression. We found that the addition of diverse individual suppressive strains to a 5-member non-suppressive SynCom resulted in only partial RGI suppression ([Extended Data Fig. 2d](#) in the published [journal version article](#)). This result suggests that the identity of suppressive commensals, and the input proportion of suppressive to non-suppressive strains, affect RGI suppression capacity quantitatively.

The ability of specific strains to differentially suppress *Atpep1*- and *flg22*-mediated RGI prompted us to investigate the mechanisms underlying this biological process. Previously, commensal *Pseudomonas* spp. in mono-associations were shown to acidify the growth medium, rendering plants insensitive to *flg22* (Yu *et al.*, 2019). To determine if acidification is responsible for RGI suppression by our SynCom, we measured the growth medium pH of plants co-inoculated with different SynComs. We observed average reductions in pH, ranging from 5.18 in mock treated plants, to 4.62 and 3.97 in the presence of a SynCom S1 and NS1, respectively. This lack of correlation between RGI suppression and growth medium acidification, suggests that this mechanism is unlikely to explain suppression in our community setup. We then investigated whether type III secretion, a well-characterized virulence mechanism among Gram-negative bacteria pathogens, was required for the suppressive activity of root commensals tested.

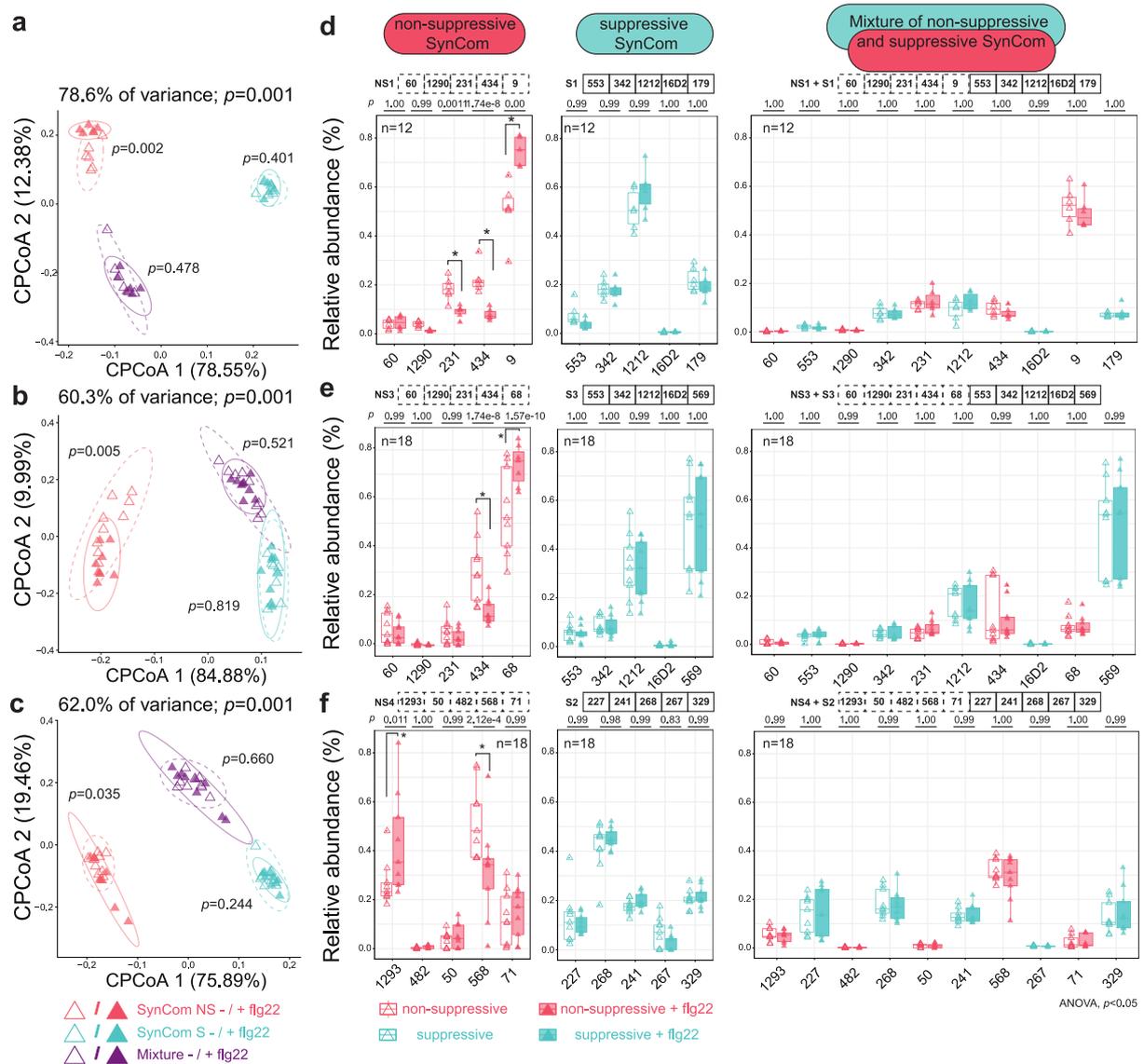
Intriguingly, *hrcC*, a gene essential for a functional type III secretion system in pathogenic *Pseudomonas*, is dispensable for RGI suppression mediated by suppressive *Pseudomonas* strain 569 (Extended Data Fig. 2e, f in the published [journal version article](#)). Next, we investigated whether root commensals can target the step upstream of *flg22* perception. Only the culture filtrate of *Janibacter* 101, an Actinobacteria member, but not that of three other suppressive strains, de-repressed both *flg22* and *Atpep1*-induced RGI (Extended Data Fig. 3a-c in the published [journal version article](#)). Heat treatment and filtration of the culture filtrate showed that the molecule(s) responsible for the suppressive activity retained in the supernatant of *Janibacter* 101 is heat-labile, and is larger than 3 kDa (Extended Data Fig. 3d in the published [journal version article](#)). Mass spectrometry analysis further revealed that the filtrates of 101 and a closely related suppressive *Janibacter*, 563, elicited a significant reduction of intact *flg22* peptide, as compared to three other tested suppressive commensals (Extended Data Fig. 3e in the published [journal version article](#)). Thus, the ability of these two *Janibacter* strains to suppress MTI resembles the activity of pathogenic bacteria (Pel *et al.*, 2014), and is associated with an extracellular mechanism that can modify/degrade *flg22* peptide. Together, these data reveal that commensals use diverse mechanisms to suppress elicitor-mediated RGI.

### 3.3.2 Activation of immunity shapes root microbiota establishment

To determine if plant immunity affects microbiota establishment, we performed reconstitution experiments with gnotobiotic plants grown on an agar matrix. We designed three taxonomically similar SynComs with contrasting RGI suppression capacities, for community profiling experiments with strain-specific resolution (total of six SynComs; SynComs used in experiment 1 and 2 differ in two Gammaproteobacteria, and SynComs used in experiment 3 are composed of entirely different strains, see [Supplementary Table 1](#) in the published [journal version article](#)). Principal Coordinate Analyses (PCoA) of Bray-Curtis dissimilarities revealed that root-associated bacterial communities were distinct from the corresponding unplanted or planted matrix samples (Extended Data Fig. 4 in the published [journal version article](#)), regardless of the SynCom composition and plant genotypes (Col-0 and *pWER::FLS2-GFP*). Constrained PCoA revealed that *flg22* treatment elicited a consistent community shift in plants inoculated with non-suppressive SynComs, while samples from those inoculated with suppressive SynComs remained together. Consistent with a dominant effect of RGI suppression, roots inoculated with 10-member mixed communities (suppressive plus non-suppressive SynComs) were not affected by *flg22* treatment (Figure 3.3a-c; Extended Data Fig. 5a-c in the published [journal version article](#)). This is consistent with another report showing that roots of *fls2* mutant and Col-0 plants have similar community profiles, consisting of a mixed 32-member SynCom (Wippel *et al.*, 2021).

To dissect the contribution of individual strains to the overall community shift, we quantified the relative abundance (RA) of individual strains. The detection of non-suppressive commensals as the most abundant strains in the mixed SynComs suggests that the ability to dominate in a community is not necessarily coupled to RGI suppression (Figure 3.3d-f; Extended Data Fig. 5d-f in the published [journal version article](#)). However, the RA of specific strains in a community was impacted by plant immunity. For example, *flg22* treatment led to altered RA of *Pseudomonas* 9/68 (up) and *Variovorax* 434 (down), while *Microbacteriaceae* 60 was unaltered (experiment 1, 2). Similarly, *flg22* treatment altered the RA of *Microbacteriaceae* 1293 (up) and *Comamonadaceae*

568 (down), while *Pseudomonas* 71 was unaffected (experiment 3; Figure 3.3d-f). A similar trend was also detected in Col-0 (Extended Data Fig. 5d-f in the published journal version article), although the effect was more pronounced in *pWER::FLS2-GFP* plants, possibly due to enhanced MTI and/or altered root architecture. In addition, we found that flg22 treatment reduced within-sample diversity of non-suppressive SynComs, (experiment 1, 2; Extended Data Fig. 5g in the published journal version article), suggesting that immune activation can affect the distribution of specific strains in community contexts.



**Figure 3.3 Activation of immunity by flg22 affects community establishment.** (a-c) Constrained coordination of the microbial profile of *pWER::FLS2-GFP* root samples showing the corresponding community shift of non-suppressive SynCom, upon flg22 treatment. Ellipses correspond to Gaussian distributions fitted to each cluster (95% confidence interval).  $p$ -values next to ellipses indicate statistical significance based on a PERMANOVA test between untreated and flg22-treated samples of each SynCom (permutation = 999,  $p < 0.05$ ). (d-f) Relative abundance of strains upon flg22 treatment. Experiment 1: (a,d); experiment 2: (b,e); experiment 3: (c,f). These experiments were conducted using different SynComs and repeated twice with consistent result. The corresponding strains used in each SynCom are indicated on top of (d-f). Values in brackets are eigenvalues explained by the Principal Component (PC). Colors indicate SynComs used and shapes indicate flg22 treatment.  $n$ =total number of biological samples collected from three independent replicates. Asterisks indicate statistical significance (two-sided ANOVA,

$p < 0.05$ ) and the  $p$  values are provided on top of each graph. The box plots centre on the median and extend to the 25th and 75th percentiles, and the whiskers extend to the furthest point within the range of 1.5× the interquartile range.

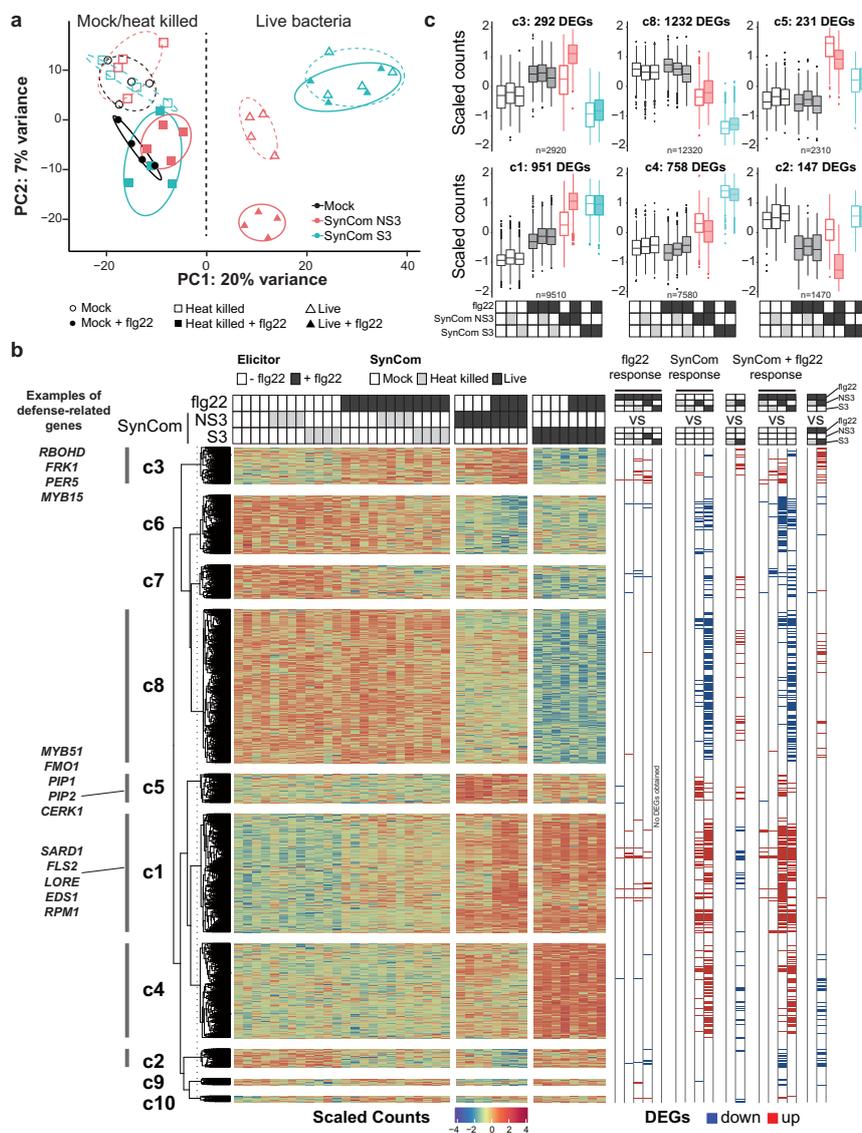
### 3.3.3 Root transcriptomic changes and dampening of immunity by suppressive SynComs

Although flg22-mediated RGI is closely associated with immune activation, its role as a *bona fide* immune output is unclear. Here, we sought to explore how inoculation with suppressive or non-suppressive SynComs affected the root transcriptome of plants treated with flg22 and grown on an agar matrix (Supplementary Table 2-7 in the published [journal version article](#)). Principal component analyses (PCA) at the transcriptome level revealed distinct expression patterns between Col-0 plants inoculated with live bacteria, compared with germ-free plants (PC1, 20% variance; Figure 3.4a). Interestingly, the transcriptional output of roots inoculated with these two taxonomically similar SynComs was clearly distinguishable after two weeks of co-cultivation, even in the absence of flg22 treatment (triangles, Figure 3.4a). In addition, we observed a separation according to the immune status of the plants, triggered by flg22 exposure, in all samples treated with heat-killed bacteria as well as with the non-suppressive SynCom (PC2, 7% variance; Figure 3.4a). By contrast, flg22 treatment of plants colonized by the suppressive SynCom did not elicit significant changes. Independent transcriptome experiments using *pWER::FLS2-GFP* plants confirmed these results (Extended Data Fig. 6 and Supplementary Table 2-4 in the published [journal version article](#)).

Next, we performed *k*-means clustering of differentially expressed genes (DEGs) involved in the flg22 response, the SynCom response, or both (Figure 3.4b and Supplementary Table 5-7 in the published [journal version article](#)). We observed three large clusters (2,221 DEGs) that were induced (c4 and c5) or suppressed (c8) by live bacteria, independent of flg22 treatment (Figure 3.4b,c). Gene ontology (GO) enrichment analyses showed that the SynCom-responsive clusters were primarily enriched in functions related to detoxification, root development, nutrient transport, and response to hypoxia (Extended Data Fig. 7 in the published [journal version article](#)). To determine whether similar GO terms could also be identified in experiments with more complex SynComs, we compared our data with two independent *Arabidopsis* root transcriptome studies that employed SynComs consisting of both suppressive and non-suppressive commensals (35 members Teixeira *et al.*, 2021; 115 members Harbort *et al.*, 2020). Despite differences in technical setups and SynCom complexities, we identified 174 common SynCom-responsive DEGs in the absence of flg22 that were related to the same biological functions mentioned above (Extended Data Fig. 8 and Supplementary Table 8-12 in the published [journal version article](#)).

Importantly, we found a flg22-inducible cluster (c3), which was significantly upregulated by the non-suppressive SynCom, but downregulated by the suppressive community (Figure 3.4b,c), in a pattern matching the RGI phenotype of the plant (Figure 3.2b) and the bacterial community shifts (Figure 3.3). As expected, a portion of defense-related genes were enriched in c3, e.g., *PER5*, *FRK1* and *RBOHD* (70 genes; Figure 3.4b). However, additional defense-related DEGs were found outside c3 and were upregulated by flg22 treatment, even in the presence of the suppressive SynCom (348 genes; Figure 3.4b). Previously characterized examples include regulators of antimicrobial camalexin, e.g., *MYB51* (Frerigmann *et al.*, 2015, 2016; c5); systemic acquired resistance, e.g., *FMO1* (c5) and *SARD1* (c1) (Zhang *et al.*, 2010; Hartmann *et al.*, 2018); and endogenous peptides amplifying MTI, e.g., *PIP1* and *PIP2* (Hou *et al.*, 2014; c5; Figure 3.4b).

Recent work showed that MAMP responsiveness in germ-free roots was gated by the expression of damage-induced PRRs (Zhou *et al.*, 2020). However, the sustained expression of *FLS2* (c1) in the presence of SynComs indicates that RGI suppression is not due to *FLS2* downregulation (Figure 3.4b). The ability of diverse root commensals to suppress *Atpep1*-mediated RGI (Extended Data Fig. 2a in the published [journal version article](#)) also highlights the interference from *FLS2*-independent pathway(s). An independent study by Teixeira *et al.* also identified a cluster of DEGs that was highly induced in axenic *Arabidopsis* by flg22-treatment, but suppressed by the presence of a 35-member SynCom consisting of suppressive and non-suppressive root commensals (Teixeira *et al.*, 2021)(Extended Data Fig. 9 in the published [journal version article](#)). Remarkably, this cluster showed the largest overlap with our cluster c3, with 58 common DEGs (at least 21 were defense-related) that were downregulated by both SynComs (Extended Data Fig. 9 in the published [journal version article](#)). Even though we have shown that two *Janibacter* strains can degrade/modify flg22 extracellularly, the downregulation of only a subset of flg22-mediated responses suggests that direct removal of the flg22 peptide is insufficient to account for the differential suppressive activities observed.



(Caption on next page) **Figure 3.4 SynCom colonization and flg22 treatment induce root transcriptomic changes in WT Col-0 plants.**

(Figure on previous page) **Figure 3.4 SynCom colonization and flg22 treatment induce root transcriptomic changes in WT Col-0 plants.** (a) PCA plot separating samples inoculated with SynComs and flg22. Ellipses correspond to *t*-distributions fitted to each cluster (70% confidence interval). (b) Heat-map (middle) and DEGs ([Supplementary Table 5-7](#) in the published [journal version article](#)) obtained by pairwise comparison (right). *k*-means clusters (*k* = 10) are marked on the left. (c) Scaled counts of transcripts in six clusters and their expression patterns upon treatments. *n*=total number of biological samples collected from four replicates. The corresponding transcriptome data of *pWER::FLS2-GFP* plants are presented in [Extended Data Fig. 6](#) (in the published [journal version article](#)), and [Supplementary Table 2-4](#) (in the published [journal version article](#)). Colors used in (a) and (c) match those used in [Figure 3.3](#), and correspond to different SynComs. The box plots centre on the median and extend to the 25th and 75th percentiles, and the whiskers extend to the furthest point within the range of 1.5× the interquartile range.

We further validated our findings by examining the expression of two flg22-inducible defense marker genes ([Millet et al., 2010](#); [Wyrsh et al., 2015](#); [Castrillo et al., 2017](#)) in roots of *Arabidopsis* by qPCR, in the presence of other suppressive SynComs. *PER5* and *FRK1* remained significantly elevated two weeks after co-inoculation with flg22 and a non-suppressive SynCom, but not with a suppressive SynCom ([Extended Data Fig. 10a](#) in the published [journal version article](#)). A non-suppressive SynCom alone also significantly induced their expressions, indicating that non-suppressive commensals stimulate specific root immune responses. As expected, a 10-member mixed SynCom, which was shown to suppress flg22-mediated RGI ([Figure 3.2b](#)), did not significantly induce the expression of *PER5* and *FRK1*. ([Extended Data Fig. 10b](#) in the published [journal version article](#)). We then examined whether the suppressive SynCom exerted an effect on the root defense response at earlier time points, upon flg22 stimulation. Intriguingly, we observed significant induction of *PER5*, *FRK1*, and *MYB15*, one hour after flg22 treatment. However, suppressive SynCom, in contrast to non-suppressive SynCom, specifically downregulated the expression of these three genes after 24 hours ([Extended Data Fig. 10c](#) in the published [journal version article](#)), suggesting that SynComs can modulate defense responses as early as 1 day after stimulation.

To determine if MTI has a direct impact on commensal proliferation independent of any microbe-microbe interactions, we focused on transcription factors (TFs) and investigated the contributions of the top three candidates identified in our dataset, *WRKY30*, *MYB15*, and *WRKY28* (cluster c3; [Extended Data Fig. 10d-e](#) in the published [journal version article](#)). Null mutants of *WRKY30* and *WRKY28* have not been reported, and our attempts to knock out these TFs *via* CRISPR failed, suggesting that these genes are essential for plant viability ([Zhao et al., 2018](#); [Zou et al., 2019](#)). We therefore focused on *MYB15*, a positive regulator of defense against the foliar pathogen *PtoDC3000* ([Chezem et al., 2017](#)). In *myb15-1* plants, elimination of this single TF led to a significantly enhanced proliferation of the detrimental strain *Pseudomonas* 401, and the commensal *Variovorax* 434 ( $p < 0.05$ , [Extended Data Fig. 10f-g](#) in the published [journal version article](#)), which also showed a reduced RA upon flg22 treatment in community contexts ([Figure 3.3d, e](#)). Together, amplicon sequencing and transcriptome data support the idea that colonization of specific root commensals is affected by host MTI, which can be attenuated by suppressive strains.

We then tested if our suppressive SynCom can also suppress defense responses triggered by a non-proteinaceous elicitor, chitin, which is commonly found as a MAMP in the cell wall of eukaryotic fungi. We performed a time-resolved experiment to follow the expression of defense marker genes

*FRK1*, *PER5* and *chitinase* (At2g43620), after one, six and 24 hours elicitor application. Unlike flg22, chitin treatment only marginally induced *FRK1* expression after one hour, while *PER5* and *chitinase* were significantly induced up to six hours (Extended Data Fig 10h in the published [journal version article](#)). This is consistent with a previous report showing that flg22 and chitin induce both overlapping and specific root responses (Stringlis *et al.*, 2018). No stimulation of marker gene expression was detected after 24 hours chitin application. Interestingly, our SynComs exerted a cooperative effect on chitin-mediated signaling. For instance, chitin-induced *PER5* expression was further stimulated by a suppressive SynCom after one hour but this stimulation was reversed such that the non-suppressive SynCom enhanced *PER5* expression after six hours. After 24 hours, both suppressive and non-suppressive SynComs induced *PER5* expression while the chitin-triggered response was no longer detectable (Extended Data Fig 10h in the published [journal version article](#)), suggesting that our SynComs interact differently with flg22- and chitin-triggered responses.

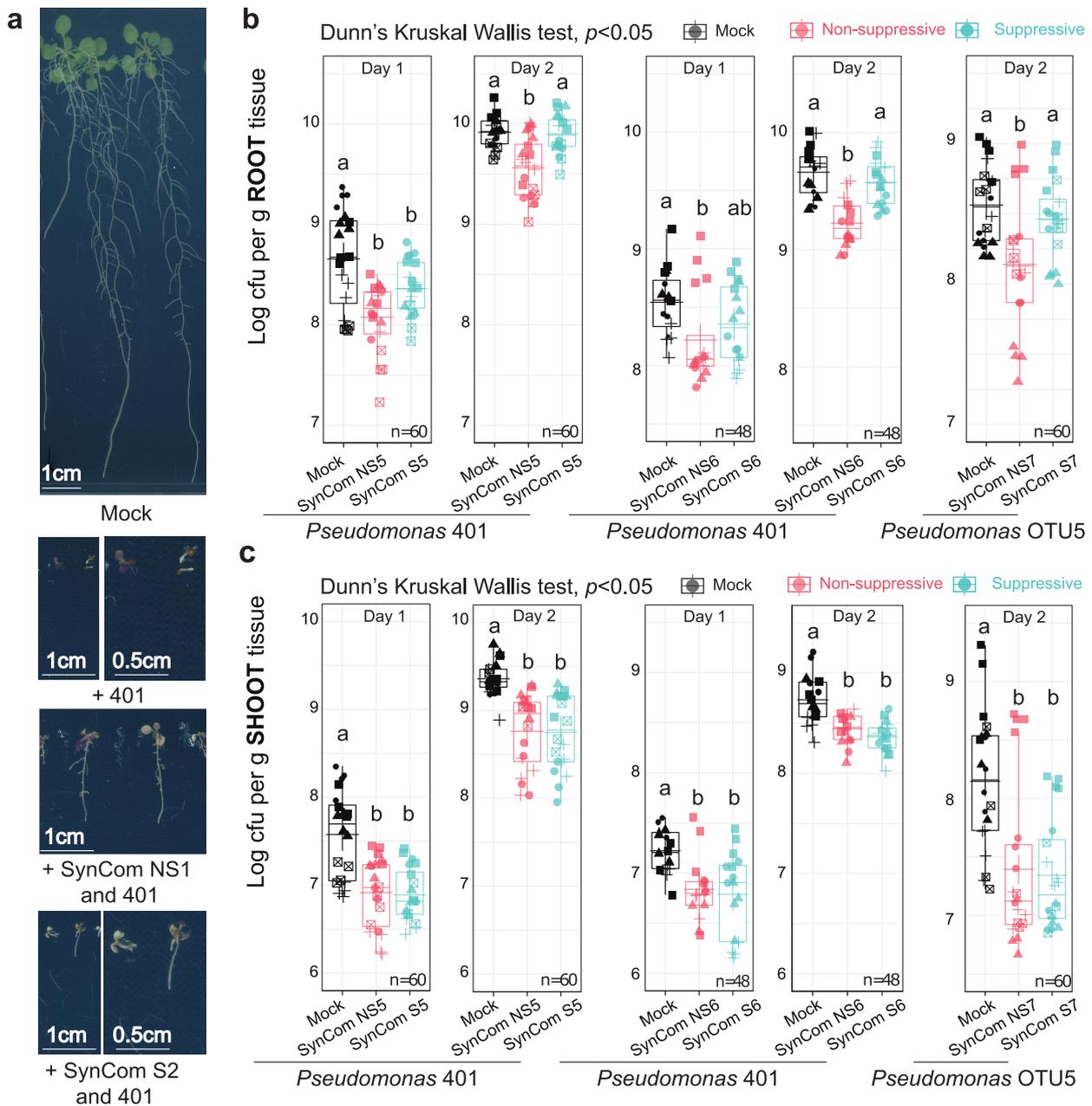
### 3.3.4 Suppressive and non-suppressive commensals differentially impact plant susceptibility to opportunistic pathogens

Since a subset of commensals dampens root immune responses, we hypothesized that colonization with a suppressive SynCom might render plants more susceptible to opportunistic pathogens. We identified three detrimental strains from *At*-RSPHERE. In particular, *Arabidopsis* plants inoculated with *Pseudomonas* 401 exhibited reduced growth and accumulated pigments in shoots reminiscent of stress-inducible anthocyanins (Extended Data Fig. 1c in the published [journal version article](#)), which indicates its pathogenic potential in a laboratory environment. Consistent with the fact that *Pseudomonas* 401 was originally isolated from healthy and asymptomatic *Arabidopsis* roots colonized by a diverse microbial community, the detrimental effect was attenuated when plants were colonized by our SynComs. Interestingly, the attenuation was stronger when plants were co-colonized with the non-suppressive SynCom, compared to the suppressive SynCom (Figure 3.5a).

Recent reports suggest a positive correlation between disease progression in natural *Arabidopsis* populations and bacterial biomass (Karasov *et al.*, 2018, 2019). To determine whether *Pseudomonas* 401 virulence is related to enhanced plant colonization, we quantified its absolute abundance on *pWER::FLS2-GFP* pre-colonized with suppressive or non-suppressive SynComs. Plants already colonized by suppressive SynComs harbored significantly higher *Pseudomonas* 401 titers compared to plants pre-colonized with non-suppressive SynComs (Figure 3.5b, c). Interestingly, this SynCom-dependent difference appeared to be limited to roots since 401 growth in shoots was similarly restricted by co-colonization with either community (Figure 3.5b, c). Even though we are not excluding an impact of microbe-microbe interactions through niche competition, none of the individual strains antagonized 401 *in vitro* (Extended Data Fig. 10 in the published [journal version article](#)). This data further suggests that the underlying growth differences are unlikely to be the result of antibiosis.

To determine whether SynComs modulate plant susceptibility to a characterized opportunistic pathogen prevalent in natural *A. thaliana* populations, and exclude the possibility that differential impact of root commensals in roots and shoots is a result due to niche specialization of 401 on roots only, we inoculated plants with the opportunistic *Pseudomonas* leaf pathogen OTU5 (isolate p5.e6) (Extended Data Fig. 1c in the published [journal version article](#)). Plants colonized by suppressive

SynComs supported higher growth of *Pseudomonas* OTU5 compared to plants colonized by non-suppressive SynComs, and this SynCom-specific effect was again observed only in roots but not in shoots (Figure 3.5b, c). Together with the RNASeq and targeted PCR data, these results suggest that pre-colonization with non-suppressive SynComs activated root immunity and this correlates with reduced growth of the tested opportunistic pathogens, whereas suppressive SynComs failed to provide pathogen protection.



**Figure 3.5 Imbalance of specific bacteria impacts plant susceptibility to opportunistic *Pseudomonas* pathogens.** (a) Symptoms of 3-week-old WT plants germinated with the indicated SynCom and *Pseudomonas* 401. Bacterial titer of *Pseudomonas* 401 and OTU5 on the roots (b) and shoots (c) of *pWER::FLS2-GFP* plants, pre-colonized with the indicated SynComs for 2 weeks. No *flg22* was used in the above experiments. Shapes represent five (SynCom NS5 + S5), four (NS6 + S6) and five (NS7 + S7) biological replicates in (b).  $n$ =total number of biological samples. Different letters indicate statistical significance (two-sided Dunn's Kruskal Wallis,  $p < 0.05$ ). The box plots centre on the median and extend to the 25th and 75th percentiles, and the whiskers extend to the furthest point within the range of 1.5× the interquartile range.

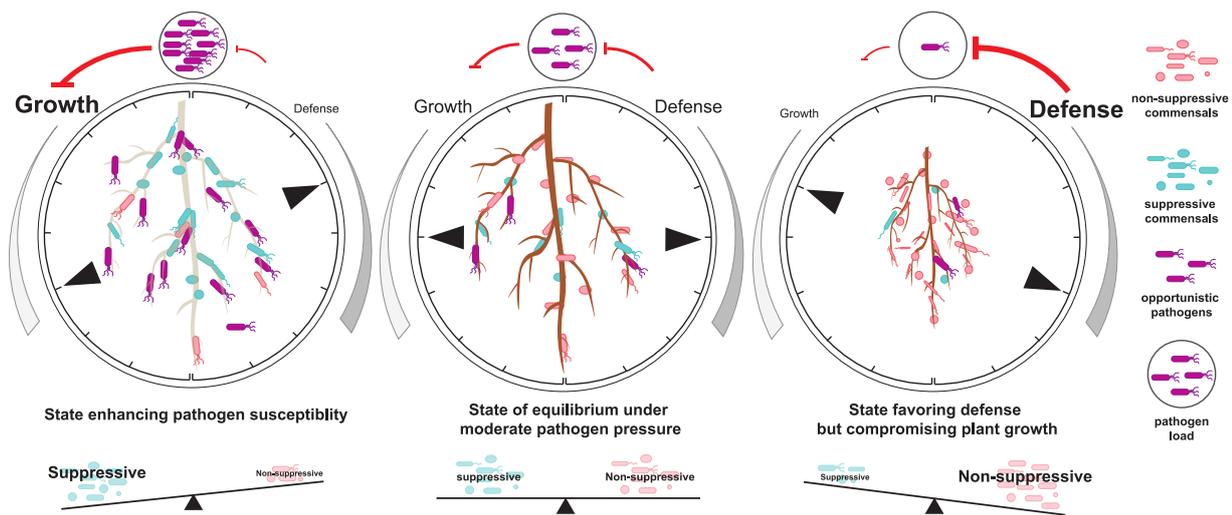
### 3.4 Discussion

In nature, a subset of soil-dwelling bacteria colonizes roots seemingly without influencing host traits, and are thus often considered as commensals. Using a bottom-up approach, we show here that phylogenetically diverse commensals, representing the core of the *Arabidopsis* root microbiota (Bai *et al.*, 2015), share the capacity to suppress host defense responses, a microbial trait that is dominant in our community setup, and is thus easily overlooked in nature. The functional redundancy of members of the root microbiota to interfere with the host's immune response is consistent with our observation that mixed communities consisting of non-suppressive and suppressive strains resist flg22-mediated community changes otherwise detected in non-suppressive SynCom-colonized plants. This finding might explain why immune-related *Arabidopsis* mutants, tested in natural soil, show relatively mild changes in root microbiota composition (Lebeis *et al.*, 2015). Here we showed that reduction of apoplastic pH and potential type-III secretion system-dependent mechanisms are insufficient to fully account for our MTI suppression phenotype. Two closely related *Janibacter* strains, but not three other tested suppressive commensals, were shown to degrade/modify flg22. Even though we cannot rule out the possible involvement of specific plant metabolites produced upon interactions with the suppressive bacteria, our findings suggest that root commensals can employ multiple mechanisms to target host immune responses, rendering this community trait difficult to overcome by the plant host.

To date, information on *Arabidopsis* root transcriptomic changes evoked by commensals are limited to mono-associations (Stringlis *et al.*, 2018; Garrido-Oter *et al.*, 2018), leaving a gap in our understanding of how plant roots respond to commensal communities that can reach a steady-state as early as 13 days after inoculation (Edwards *et al.*, 2015). We unexpectedly found that after 2 weeks, root colonization by taxonomically similar commensal SynComs, differing in their capacity to suppress RGI, elicited 2,221 DEGs (cluster c4, c5, c8) with remarkable overall similarity. These changes, associated with presumably steady-state SynComs, contrast with the subtle response to heat-killed SynComs or flg22 treatment alone, reflecting an impact of active commensal colonization on host transcriptional outputs beyond plant responses to chronic exposure to MAMPs. In addition, we observed robust enrichment of specific GO terms related to root development, nutrient transport, response to hypoxia, and detoxification across experimental setups and SynCom complexities (Harbort *et al.*, 2020; Teixeira *et al.*, 2021). Indeed, rhizobacteria alone are known to modulate root traits (López-Bucio *et al.*, 2007; Zamioudis *et al.*, 2013; Garrido-Oter *et al.*, 2018). In return, root-secreted photoassimilates feed up to 20% of root-associated bacteria (Hernández *et al.*, 2015). Since photoassimilates have been shown to serve as sources of organic carbon that limit bacterial growth (Aldén *et al.*, 2001), we thus speculate that enrichment of these GOs is associated with altered nutrient flux, and reduced oxygen due to microbial respiration in roots. Although our SynComs are taxonomically diverse with predicted varied metabolic repertoires (Bai *et al.*, 2015; Levy *et al.*, 2018), convergence to core transcriptomic outputs indicate integrated responses to a state of “community commensalism”.

The zigzag model of the plant immune system proposes that effective resistance is the result of quantitative outputs above a certain threshold following MAMP perception (Jones & Dangl, 2006). Colonization by suppressive SynComs led to the down-regulation of a subset of flg22-induced genes (Figure 3.4, cluster c3), whereas colonization by non-suppressive SynComs alone stimulated

these genes and further upregulated their expression together with *flg22*. Thus, the responsiveness of these defense-associated genes to SynCom colonization differs greatly with respect to the ability of the bacterial community to suppress RGI. However, roots in nature are co-colonized by both groups of commensals, and our experiments point to a quantitative output that is dependent on their ratio. Intriguingly, recent studies reported that 42% (Yu *et al.*, 2019), and 28% (Teixeira *et al.*, 2021) of commensals from two other *Arabidopsis* root-derived culture collections quench early and late *flg22*-induced responses in mono-associations, respectively (Yu *et al.*, 2019). Together with our study, this confirms a potential of the root microbiota to modulate plant growth-defense traits.



**Figure 3.6** The balance between non-suppressive and suppressive strains integrates with plant innate immunity. This “rheostat model” buffers the system against pathogen challenge and defense-associated trade-off.

We hypothesize that the imbalance between non-suppressive and suppressive commensals might reduce plant fitness under stress conditions. Indeed, plants pre-colonized by suppressive SynComs are as susceptible as germ-free plants to opportunistic *Pseudomonas* pathogens, whereas plants associated with non-suppressive SynComs are more resistant, but prone to MAMP-induced RGI. The observed defense-associated community shifts and potentially reduced alpha-diversity might hinder provision of microbiota-derived beneficial services (Saleem *et al.*, 2019), or exert a detrimental impact on the host under dysbiosis (Chen *et al.*, 2020). We thus propose a rheostat model (Figure 3.6), in which a balance between commensals with contrasting MTI modulating activities constitutes an integral feature of the holobiont to buffer plant resistance to pathogen perturbation and defense-associated growth reduction.

It remains unclear which factors govern the state of equilibrium and the corresponding ratio between non-suppressive and suppressive strains. However, based on the result obtained using three pairs of SynComs (Figure 4.3), the initial input composition plays a key role to define the equilibrium state. The timing of colonization (Edwards *et al.*, 2015) and abiotic factors (Hartman & Tringe, 2019) likely also exert an influence. Consequently, these complex interactions allow community coexistence, and ultimately establish microbe-host homeostasis. Accordingly, their ratio will impact the amplitude and/or might set the threshold for effective resistance in the zigzag model. Plants in nature are influenced by fluctuating stresses and are colonized by more diverse microbial communities that modulate plant physiology through multiple mechanisms, including the

modulation of phytohormone signaling (de Zélicourt *et al.*, 2018; Finkel *et al.*, 2020). As our SynComs are constructed based on their ability to suppress flg22-mediated responses, we found that they exert a synergistic effect on chitin-mediated responses. Selective modulation of chitin and flg22-mediated responses e.g. by the cytoplasmic receptor-like kinase PBL27 has been reported (Shinya *et al.*, 2014). The characterization of molecular mechanisms underlying immunosuppression by root commensals may identify immunity components that are specific to one or integrate multiple upstream signaling pathways. Another future task will be to test whether the rheostat model also applies to communities with different traits to alleviate abiotic stresses.

### 3.5 Materials and methods

*Arabidopsis thaliana* ecotype Columbia (Col-0, CS60000), *pepr1pepr2* were lab stocks. *myb15* (Chezem *et al.*, 2017) (SALK\_151976) was a gift from Nicole Clay (Yale University, USA). The transgenic line *pWER::FLS2-GFP* (Wyrsh *et al.*, 2015) (*fls2*: SAIL691\_C04 background) was provided by Niko Geldner (Université de Lausanne, Switzerland). Chitin was purchased from Sigma (C9752).

flg22 (QRLSTGSRINSAKDDAAGLQIA) and *Atpep1* (ATKVKAKQRGKEKVSSGRPGQHN) peptides were synthesized by EZbiolab.

#### 3.5.1 Growth conditions for plants and culture conditions for bacteria

*Arabidopsis* seeds were surface-sterilized in 70% ethanol twice for 5 min each followed by a brief wash with 100% ethanol. Seeds were then washed three times with sterile water. Cold-stratified seeds were sowed on agar plates (1%, Difco Agar Granulated, BD Biosciences, discontinued) or Bacto agar (BD Biosciences) prepared with half-strength Murashige and Skoog (MS) medium (Duchefa) and 0.1g/L 2-(*N*-morpholino)ethanesulfonic acid (MES, pH 5.7). Sugars were not provided as an additional carbon source unless otherwise specified. Plants were grown under short-day conditions (10 hr light, 14 hr dark) at 21°C/19°C cycle, 65% relative humidity and light intensity of 120 mE m<sup>-2</sup> sec<sup>-1</sup>. For experiments involving *myb15-1*, surfaced sterilized seeds were sowed on half-strength MS agar plates supplemented with 5g/L sucrose.

Information on individual strains used can be found at *At*-RSPHERE (<http://www.at-sphere.com/>). OTU5(p5.e6) (Karasov *et al.*, 2018) was kindly provided by Detlef Weigel (Max Planck Institute for Developmental Biology, Tübingen, Germany). Bacterial strains were prepared by taking an aliquot from the glycerol stock, followed by incubation on 50% tryptic soy broth (TSB) agar plate (Sigma-Aldrich, USA) at 25 °C from one to four days. Before the start of the experiments, strains were cultured in 50% TSB medium to saturation, and subcultured to log phase with fresh medium in a 1:5 ratio. Bacterial cultures were pelleted by centrifugation at 8k g for 5 min, followed by two washes with 10mM MgSO<sub>4</sub>.

#### 3.5.2 Screening for RGI suppressive strains in monoassociation

After washing, bacteria were diluted with 10mM MgSO<sub>4</sub> to an OD<sub>600</sub> concentration of about 0.1. A total of 150 μl bacterial suspension was added to still warm 50ml half strength MS agar medium at a final bacterial concentration of OD<sub>600</sub>= 0.0005. A final concentration of 1 μM flg22 was added accordingly. Plates were dried for 2 hrs before approximately 15 surface-sterilized *pWER::FLS2-*

*GFP* seeds were sowed on each plate. The expression of the *flg22* receptor *FLS2* in *pWER::FLS2-GFP* is limited to the root epidermis such that potential inter-organ shoot-to-root signal upon *flg22* perception is minimized. Plates were sealed with 3M tape and transferred to the phytochamber for incubation. One week after germination, plants with delayed germination were removed and the plates were trimmed to about 10 remaining plants. Pictures were taken 3 weeks after incubation and the primary root lengths were quantified by ImageJ. Shoots were separated from the roots and fresh shoot weight of individual plants was taken. For experiments using 1  $\mu$ M *Atpep1*, wild-type Col-0 plants were used instead.

A phylogenetic tree of selected strains from *At-RSPHERE* was performed previously (Bai *et al.*, 2015) and visualized by iTOL (Letunic & Bork, 2019). Strains leading to a rescue of RGI with root length longer than 2cm (average root length of germ-free *flg22*-treated *pWER::FLS2-GFP* plants=1.53cm; n=37) upon coinoculation with 1  $\mu$ M *flg22*, and exhibiting consistent suppressive activity across two biological replicates, were considered as “suppressive”. Suppressive strains were indicated with a red triangle in Figure 1a. For the inoculation of SynCom, each bacterium was inoculated to a final concentration of OD<sub>600</sub>= 0.0005, i.e., for a 5-member SynCom, the total bacteria added was OD<sub>600</sub>=0.0025. The 5-member SynCom is composed of Actinobacteria, Alpha-, Beta- and Gamma-proteobacteria. Bacteroidetes and Firmicutes were not included in these SynComs since no strains with differential ability to suppress RGI were identified in these two phyla. Composition of SynComs used in this manuscript can be found in [Supplementary Table 1](#) in the published [journal version article](#).

### 3.5.3 Microbial community profiling and data processing

For *16S* community profiling, root samples were harvested and libraries were processed according to previously a published protocol (Bulgarelli *et al.*, 2012). Forward and reverse sequencing reads were denoised and demultiplexed separately according to the barcode sequence using QIIME (Caporaso *et al.*, 2010) with the following parameters: *phred*=30; *bc\_err*=2. After quality-filtering and merging of paired-end reads, amplicon tags were then aligned to a reference set of sequences obtained from the whole-genome assemblies of every strain included in each experiment by using USEARCH (*uparse\_ref* command) (Edgar, 2010). A count feature table for each strain was generated by using only perfect matches to the reference sequence from the genome collection. This count table was employed for subsequent diversity and enrichment analyses. Alpha and beta diversities were calculated after normalizing count tables by the total number of reads per sample. The Simpson index was obtained using the *diversity* function in the vegan package. Bray-Curtis dissimilarity index was calculated using the *vegdist* function in the vegan package (Oksanen *et al.*, 2019) and used for unconstrained ordination by Principal Coordinate Analysis (PCoA). All data were used except for biological replicate c of experiment 1 due to a potential contamination issue or PCR error. Constrained PCoA (CPCoA) was performed with the vegan *capscale* function on the Bray-Curtis dissimilarity matrices, constraining by the interaction between *flg22* treatment and SynCom variables and conditioning by technical parameters. Statistical significance of separation between community profiles according to *flg22* treatment was determined using PERMANOVA with 999 permutations (*anova.cca* function in vegan). Finally, all amplicon data was visualized using the *ggplot2* R package.

### 3.6 Data and code availability

Raw transcriptome and 16S rRNA amplicon sequencing data from this project were deposited at NCBI under the accession number GSE157128. Mass spectrometry data have been deposited to Panorama Public ([https://panoramaweb.org/fig22\\_RGI.url](https://panoramaweb.org/fig22_RGI.url)) and the ProteomeExchange (PXD020452). Source data are provided with this chapter. Scripts to reproduce figures are available at GitHub ([https://github.com/YulongNiu/MPIPZ\\_microbe-host\\_homeostasis](https://github.com/YulongNiu/MPIPZ_microbe-host_homeostasis)).

### 3.7 Author contributions

#### Ma, Niu and Jia et al., 2021

K.-W.M. and P.S.-L. conceptualized the initial project. K.-W.M., R.G.-O. and P.S.-L. designed the experiments. K.-W.M. performed the initial screening of RGI-suppressive members; 16S amplicon sequencing with the help of C.C., J.O. and Y.J.; RNA-seq experiments; and inoculation experiments involving *Pseudomonas* strains 401, 434 and OTU5 with the help of J.O. and Y.J.; Y.N. performed RNA-seq data analyses and developed all related pipelines for GO term enrichment, network studies and cross-study comparisons. K.-W.M., R.G. and R.G.-O. analyzed the 16S amplicon sequencing data. K.-W.M., Y.J., P.S.-L., H.N. and S.C.S. designed the mass spectrometry experiments. Y.J. performed the experiments. S.C.S and H.N. developed the protocol to detect flg22 peptide and S.C.S. analyzed the data. A.E. and N.G. phenotyped the growth of the pWER::FLS2-GFP transgenic line in natural soil. K.-W.M., Y.N., R.G.-O. and P.S.-L. wrote the manuscript with input from all co-authors.

# Chapter 4 Host preference and invasiveness of commensals in the *Lotus* and *Arabidopsis* root microbiota

## 4.1 Abstract

Roots of different plant species are colonised by bacterial communities, that are distinct even when hosts share the same habitat. It remains unclear to what extent the host actively selects these communities and whether commensals are adapted to a specific plant species. To address this question, we assembled a sequence-indexed bacterial culture collection from roots and nodules of *Lotus japonicus* that contains representatives of most species previously identified using metagenomics. We analysed taxonomically paired synthetic communities from *L. japonicus* and *Arabidopsis thaliana* in a multi-species gnotobiotic system and detected signatures of host preference among commensal bacteria in a community context, but not in mono-associations. Sequential inoculation experiments revealed priority effects during root microbiota assembly, where established communities are resilient to invasion by late-comers, and that host preference of commensal bacteria confers a competitive advantage in their cognate host. Our findings show that host preference in commensal bacteria from diverse taxonomic groups is associated with their invasiveness into standing root-associated communities.

## 4.2 Introduction

Plant roots associate with diverse microorganisms that are recruited from the surrounding soil biome and which assemble into structured communities known as the root microbiota. These communities provide the host with beneficial functions, such as indirect pathogen protection, or mineral nutrient mobilization (Durán *et al.*, 2018; Zhang *et al.*, 2019; Carrión *et al.*, 2019). Despite conservation at higher taxonomic ranks (Bulgarelli *et al.*, 2012; Fitzpatrick *et al.*, 2018), comparison of community profiles across diverse land plants shows a clear separation according to host species (Yeoh *et al.*, 2017; Fitzpatrick *et al.*, 2018). These patterns could be explained by a process in which the root microbiota assemble according to niches defined by plant traits, that in turn diversify as a result of plant adaptation to their environment. Alternatively, variation of microbiota profiles along the host phylogeny may be at least partially caused by co-adaptation between the plant and its associated microbial communities.

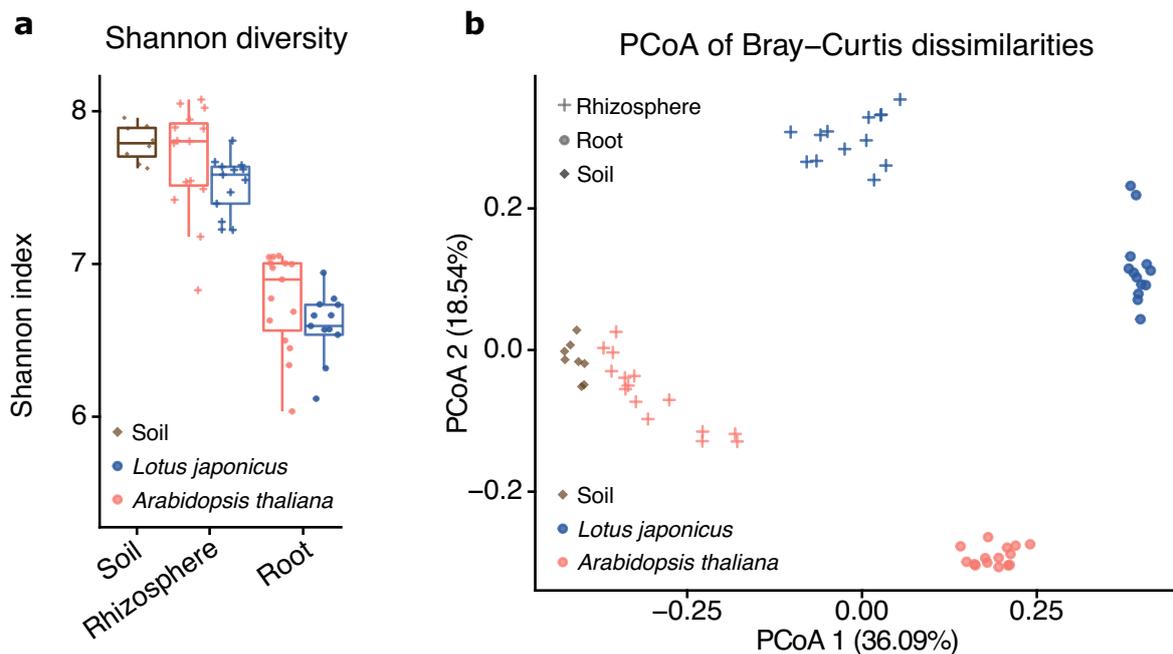
Culture-independent amplicon sequencing allows characterization of community structures and taxonomic composition but does not allow the study of phenotypes of individual community members. To overcome this fundamental limitation in microbiota studies, comprehensive culture collections of sequenced strains isolated from root and leaf tissue have been established (Bai *et al.*, 2015; Levy *et al.*, 2018; Durán *et al.*, 2018; Zhang *et al.*, 2019). Synthetic communities (SynComs) built from these collections can be used in gnotobiotic reconstitution systems of reduced complexity to explore the role of immune signaling (Lebeis *et al.*, 2015), nutritional status (Castrillo *et al.*, 2017; Zhang *et al.*, 2019), biotic and abiotic stress (Durán *et al.*, 2018) and priority effects (Carlström *et al.*, 2019) in the establishment of the root and leaf microbiota.

In order to investigate plant host preference of commensal bacteria, we assembled a collection of cultured bacterial species from the roots and nodules of the model legume *Lotus japonicus*

(hereafter *Lj*) that is comparable to the collection previously established from *Arabidopsis thaliana* (hereafter *At*) roots (Bai *et al.*, 2015) in terms of taxonomic and genomic composition, despite 125 Myr of divergence between *Lj* and *At* (Van de Peer *et al.*, 2017) whose crown groups evolved 65 and 32 Mya, respectively (Koenen *et al.*, 2021). These two collections originate from plants grown in the same soil, enabling us to design SynComs for microbiota reconstitution experiments. Using this set-up, we investigated host preference of commensal communities and the role of nitrogen-fixing nodule symbiosis, immunity and root exudation in microbiota establishment.

## 4.3 Results

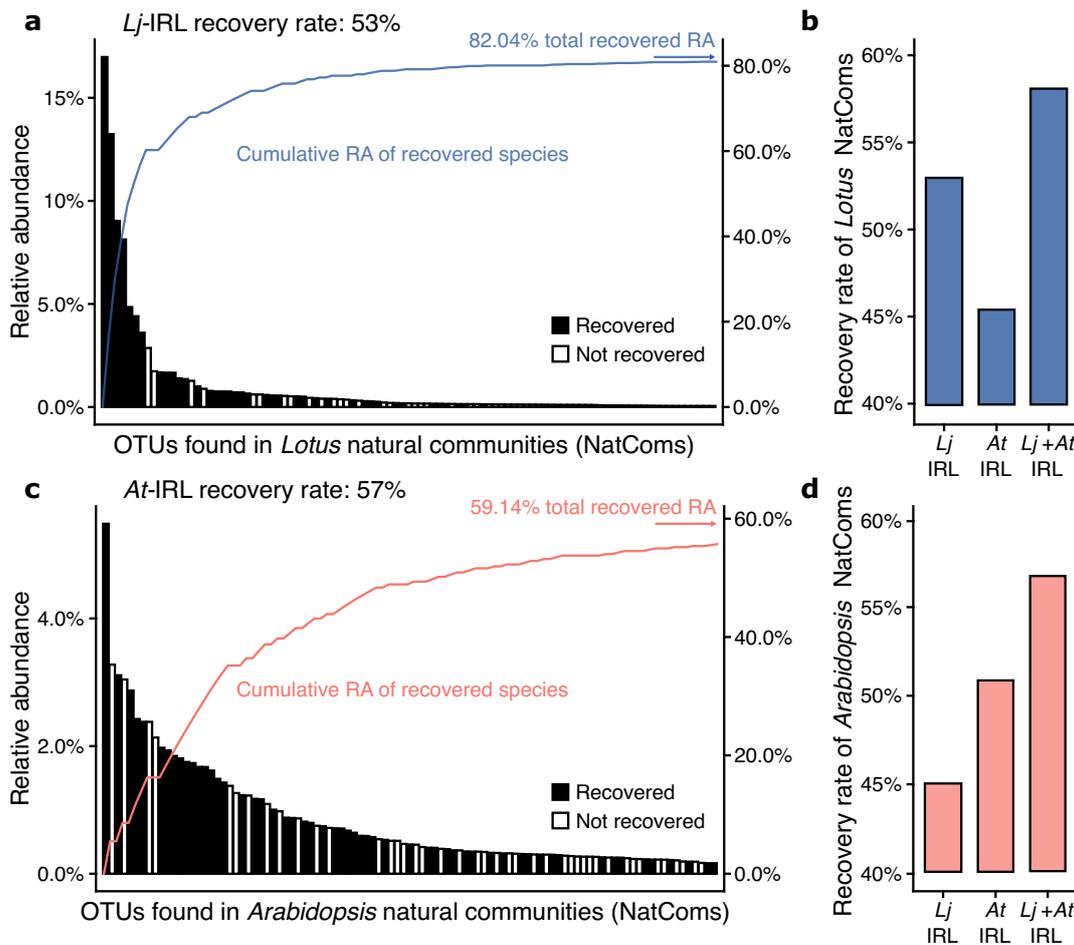
### 4.3.1 Host-species-specific bacterial culture collections



**Figure 4.1 Diversity of *Lotus* and *Arabidopsis* root-associated bacterial communities.** **a**, Alpha-diversity analysis of soil- ( $n = 8$ ), rhizosphere- ( $n = 13$  for Gifu,  $n = 15$  for Col-0), and root-associated bacterial communities ( $n = 13$  for Gifu,  $n = 15$  for Col-0) from *L. japonicus* and *A. thaliana* plants grown in natural soil (exp. A), assessed using the Shannon index. **b**, Principal coordinates analysis (PCoA) of Bray-Curtis dissimilarities of the same communities ( $n = 64$ ).

We compared the bacterial communities associated with roots of *Lj* and *At* plants grown in the same soil (exp. A; Extended Data Fig. 1a and Supplementary Table 2 in the published [journal version article](#)) (Durán *et al.*, 2018; Zhang *et al.*, 2019; Thiergart *et al.*, 2019) and confirmed that both hosts associate with communities that are clearly distinct from those of the surrounding soil (Figure 4.1). This shift is characterized by a decrease in alpha-diversity (within-sample diversity; Figure 4.1a) as well as by a separation between root, rhizosphere, and soil samples (beta-diversity; Figure 4.1b, PCoA 2). In addition, *Lj* and *At* root samples formed two distinct clusters, indicating host species-specific recruitment of commensals from identical pools of soil-dwelling bacteria (Figure 4.1b, PCoA 1), which is in line with previous studies (Zgadzaj *et al.*, 2016; Thiergart *et al.*, 2019). This separation (28% of variance;  $P = 0.001$ ) was mainly explained by the different relative abundance of Proteobacteria, Actinobacteria, Bacteroidetes (Flavobacteria and Sphingobacteria), and

Firmicutes (Bacilli) in *Lj* compared to *At* ([Extended Data Fig. 2](#) in the published [journal version article](#)).



**Figure 4.2** Abundance and recovered *Lotus* and *Arabidopsis* root-associated bacterial OTUs. **a** and **c**, Rank abundance plots of OTUs found in the *Lotus* (**c**) and *Arabidopsis* (**e**) natural root communities. Community members captured in the corresponding culture collection are depicted as black while non-recovered OTUs are shown in white. The vertical axis on the right shows the accumulated relative abundance in natural communities of all recovered OTUs. **d** and **f**, percentage of abundant OTUs (0.1% RA) associated with *Lotus* (**d**) or *Arabidopsis* (**f**) roots in nature (natural communities, NatComs) that are captured in the *Lotus* or the *Arabidopsis* indexed rhizobacterial libraries (*At*- and *Lj*-IRL).

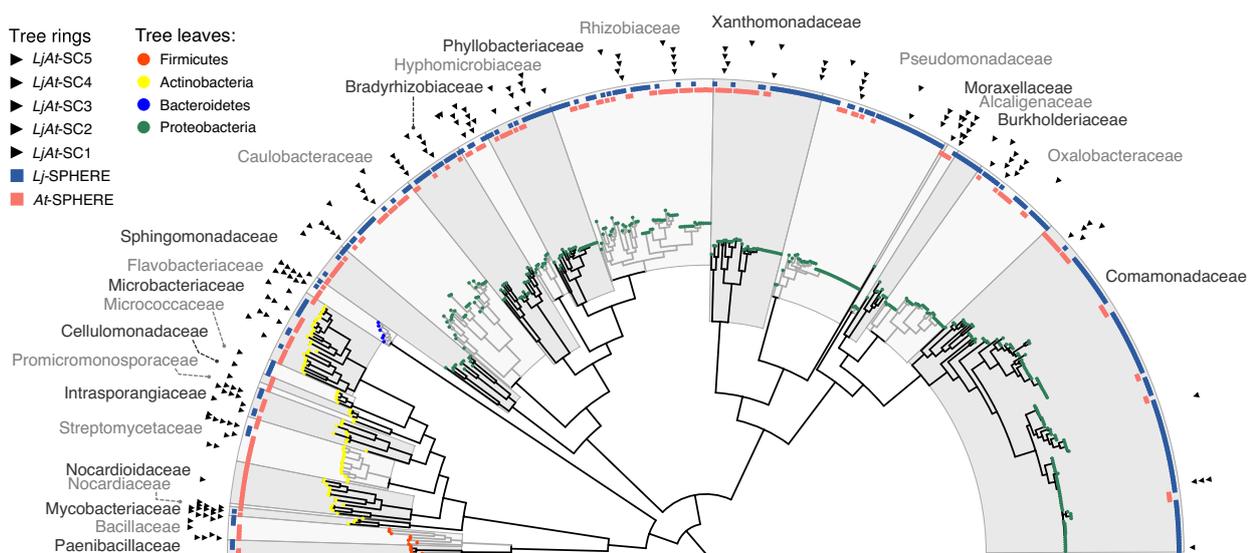
To explore the mechanisms by which different plant species associate with distinct microbial communities, we established a taxonomically and functionally diverse culture collection of the *Lj* root and nodule microbiota ([Extended Data Fig. 1b](#) in the published [journal version article](#)). A total of 3,960 colony-forming units (CFUs) were obtained and taxonomically characterized by sequencing the bacterial *16S* ribosomal RNA (rRNA; [Supplementary Data 1](#) in the published [journal version article](#)), resulting in a comprehensive sequence-indexed rhizobacterial library from *Lj* (*Lj*-IRL). In parallel, a subset of the root samples was also subjected to amplicon sequencing to obtain culture-independent community profiles for cross-referencing with the *Lj*-IRL data. In the *Lj* collection, we were able to recover up to 53% of the most abundant bacterial OTUs (Operational Taxonomic Units, defined by 97% *16S* rRNA sequence identity) found in the corresponding natural community profiles, compared with 57% for the *At* collection ([Figures 4.2a, c](#); [Supplementary Note](#) in the published [journal version article](#)). The recovered bacterial taxa in the respective collection

accounted for 82% of all sequencing reads from *Lj* root samples and 59% from *At*. Approximately 45% of the abundant OTUs found in the natural communities of one host were captured in the culture collection of the other species (Figure 4.2b, d), indicating a substantial overlap of the recovered bacterial taxa. Both culture collections include members of the Actinobacteria, Proteobacteria, Bacteroidetes, and Firmicutes, the four phyla robustly found in the root microbiota of diverse plant species (Yeoh *et al.*, 2017; Fitzpatrick *et al.*, 2018).

To establish a core *Lj* culture collection of whole-genome sequenced strains (*Lj*-SPHERE), we selected from the *Lj*-IRL a taxonomically representative subset of bacterial isolates maximizing the number of covered taxa, as previously done for *At* (Bai *et al.*, 2015). A total of 294 isolates belonging to 20 families and 124 species, including both commensal and mutualistic bacteria, were subjected to whole-genome sequencing (Supplementary Data 2 in the published [journal version article](#)). Comparative analyses of all sequenced isolates from both collections revealed an extensive taxonomic and genomic overlap between exemplars derived from *Lj* and *At* (Extended Data Fig. 3; Supplementary Note in the published [journal version article](#)). This indicates that the observed differences in natural community structures (Figure 4.1b) are likely not driven by the presence of host-specific bacterial taxonomic groups. Instead, the distinct root community profiles of the two hosts are possibly due to differences in the relative abundance of shared taxonomic groups (Extended Data Fig. 2 in the published [journal version article](#)).

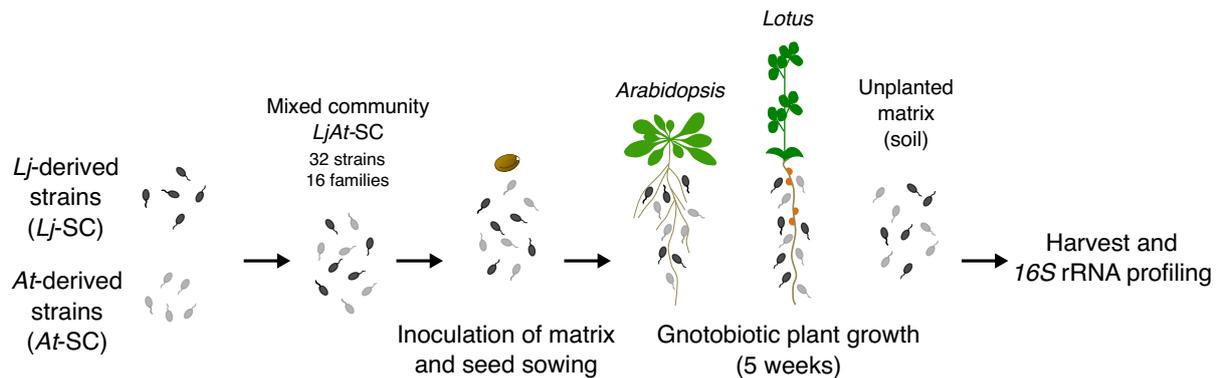
#### 4.3.2 Host-preference of commensal synthetic communities

Given the overlap between the *Lj*- and *At*-SPHERE culture collections at a high taxonomic and whole-genome level, we speculated that strain-specific phenotypic variation *in planta* could allow commensal bacteria to preferentially colonize their cognate host. In order to test this hypothesis, we designed taxonomically paired SynComs for each host, representing 16 bacterial families present in both collections (Figure 4.3).



**Figure 4.3 Whole-genome phylogeny of the *Lotus* and *Arabidopsis* core culture collections.** Maximum likelihood phylogeny, constructed from a concatenated alignment of 31 conserved, single copy genes (AMPHORA) showing the taxonomic overlap of the *Lj*-SPHERE ( $n = 294$ , blue track) and *At*-SPHERE ( $n = 194$ , red track) core culture collections. Arrows in the outer rings indicate the strains selected for four mixed communities used in reconstitution experiments.

We then combined these SynComs into a mixed community composed of 32 strains ([Supplementary Table 1](#) in the published [journal version article](#)). We allowed commensal bacteria to compete for colonization of the host from which they were derived (hereafter referred to as native strains) with strains isolated from the other plant species (non-native strains; [Supplementary Fig. 1a](#) in the published [journal version article](#)). We employed a gnotobiotic system ([Kremer \*et al.\*; Durán \*et al.\*, 2018](#)) to grow wild-type *At* (Col-0), *Lj* (Gifu), and a *Lj* mutant deficient in root nodule symbiosis (*Ljnr5*) ([Madsen \*et al.\*, 2003](#)) in the presence of the mixed community ([Figure 4.4](#)). After five weeks, we performed community profiling *via* 16S rRNA gene amplicon sequencing of the root, rhizosphere, and unplanted soil compartments.

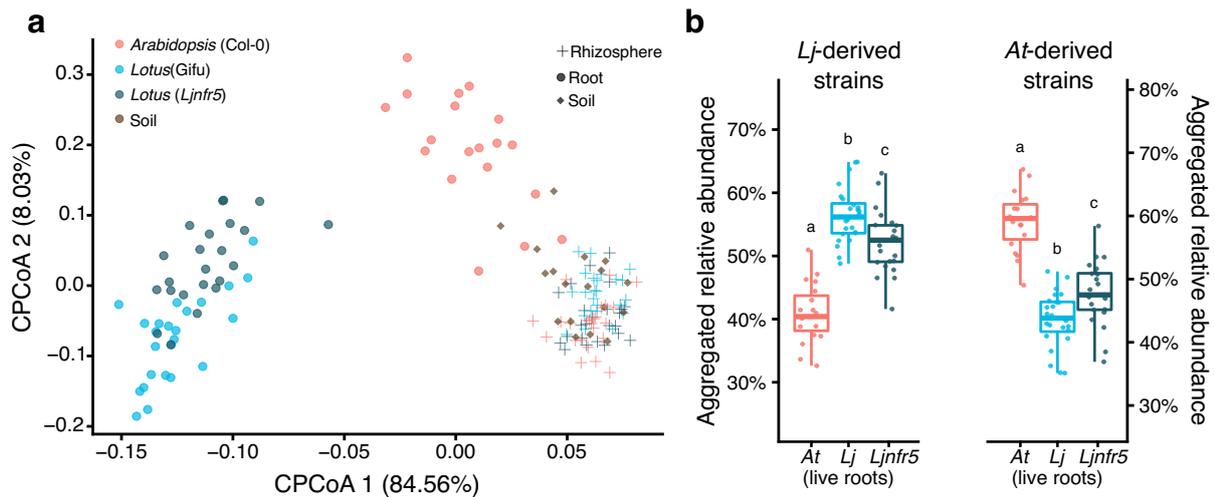


**Figure 4.4 Scheme of competition experiments.**

Analysis of community diversity revealed a significant separation of communities of root samples from those of rhizosphere and soil, which in turn clustered together (exp. B; [Figure 4.5a](#)). In addition, we observed that the two hosts are colonized by distinct root microbial communities starting from the same input, and that samples from wild-type *Lj* are differentiable from those of *Ljnr5* ([Figure 4.5a](#)). These results were confirmed by two independent, full factorial experiments using different mixed communities (exp. C and M; [Extended Data Fig. 4a, c](#) in the published [journal version article](#)). An additional experiment, where strains belonging to families found exclusively in the *Lj* or *At* culture collections (two and five families, respectively) were added to the mixed community, resulted in similar patterns of beta-diversity (exp. D; [Extended Data Fig. 4b](#) in the published [journal version article](#)). These results recapitulate the community shifts between compartment, host species, and plant genotype which were previously observed in culture-independent community profiles obtained from plants grown in natural soils ([Figure 4.1b](#)) ([Bulgarelli \*et al.\*, 2012](#); [Lundberg \*et al.\*, 2012](#); [Zgadzaj \*et al.\*, 2016](#); [Thiergart \*et al.\*, 2019](#)), thus validating our comparative reconstitution system to study host species-specific microbiota establishment.

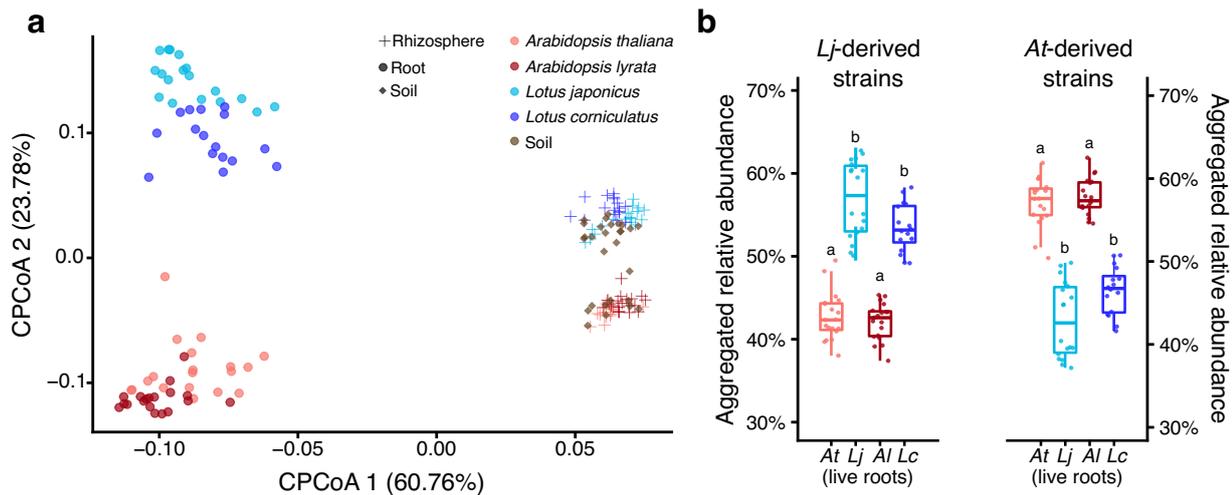
Next, we tested whether communities of commensal bacteria would preferentially colonize roots of their cognate host species (i.e., from which they were originally isolated) compared to those of the other host. We found that the aggregated relative abundance of strains from the Lj-SPHERE collection was higher in wild-type *Lj* root samples than in those of *At* ([Figure 4.5b](#); [Extended Data Fig. 4d, 4e, and 4f](#) in the published [journal version article](#)). Likewise, strains from the At-SPHERE collection were more abundant on their cognate host than on *Lj*. Commensal host preference and host species community separation was reduced but still present in the *Ljnr5* mutant ([Figure 4.5b](#)), suggesting that nodule symbiosis only partially contributes to commensal host preference. Further,

sequential *in silico* removal of individual bacterial families did not significantly alter the observed patterns of host preference at the community level ([Extended Data Fig. 5](#) in the published [journal version article](#)), indicating that host preference was not driven by a single taxonomic group. Mono-association experiments with *Lj* and *At* wild-type plants grown on agar plates revealed that most community members maintained their root colonization capacity, but did not show a significant host preference in isolation (exp. E; [Extended Data Fig. 6](#) in the published [journal version article](#)), suggesting that this commensal phenotype requires a community context. Moreover, we found that shoot biomass of both host species was not affected by these strains, confirming their commensal lifestyle in mono-associations ([Extended Data Fig. 7](#) in the published [journal version article](#)).



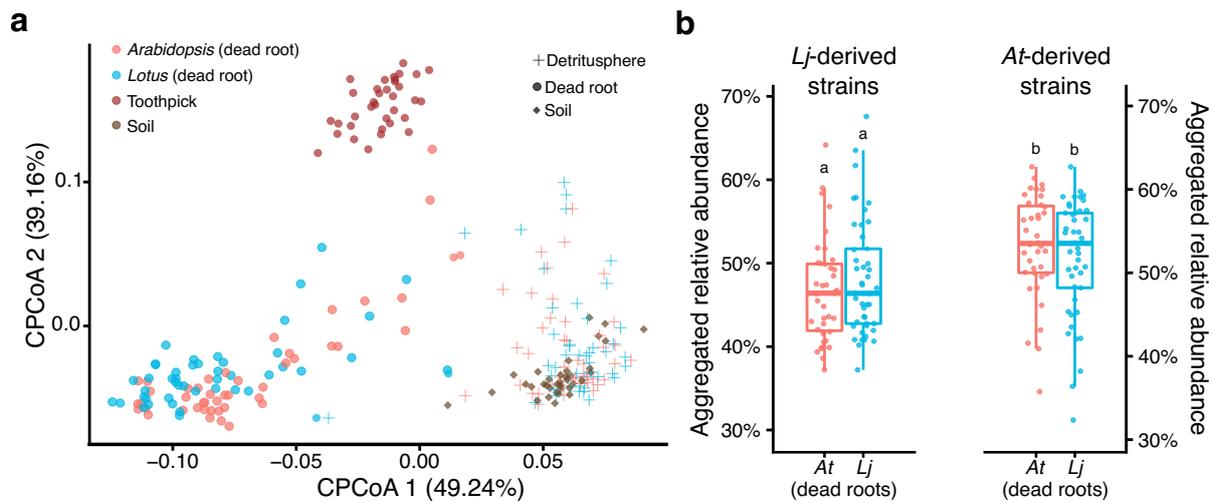
**Figure 4.5 Constrained PCoA and aggregated RA of the commensal communities in *Lj*- and *At*-associated root compartments.** (a) Constrained PCoA of Bray-Curtis dissimilarities (constrained by all biological factors and conditioned by all technical variables) of soil, rhizosphere, and root samples. *L. japonicus* wild type Gifu, *nfr5* mutant, and *A. thaliana* wild type Col-0 plants co-cultivated with the mixed community *LjAt*-SC2 (exp. B,  $n = 155$ , variance explained 53.8%,  $P = 0.001$ ). (b) Aggregated RA of the 16 *Lj*-derived and the 16 *At*-derived strains in the live roots of *Lotus* and *Arabidopsis* plants inoculated with *LjAt*-SC2 ( $n = 66$ ).

We then investigated if the phenotype of commensal host preference was conserved in a plant phylogenetic framework. We selected two additional plant species, *L. corniculatus* and *A. lyrata*, which diverged from *Lj* and *At* approximately 12.5 Mya and 13 Mya, respectively ([Beilstein et al., 2010](#); [Ojeda et al., 2014](#)), and are indigenous to the region from which the soil used to isolate these bacterial strains was collected ([Steiner & Santos, 2001](#); [Clauss & Mitchell-Olds, 2006](#)). We inoculated these four species with a mixed community of *Lj* and *At* commensals and obtained amplicon profiles of root, rhizosphere and unplanted soil samples (exp. F). We observed a significant separation between *Lotus* and *Arabidopsis* root communities ([Figure 4.6a](#);  $P=0.001$ ), and to a lesser extent between samples from the sister species within the same genus ([Extended Data Fig. 8](#) in the published [journal version article](#)), which is in line with similar results obtained from *At* relatives grown in natural sites ([Schlaeppli et al., 2014](#)). We found that the patterns of host preference observed in *Lj* and *At* were retained in their relative species ([Figure 4.6b](#)), suggesting that this community phenotype might be the result of commensal adaptation to root features conserved in a given host lineage.



**Figure 4.6 Constrained PCoA and aggregated RA of the commensal communities in *Lotus* and *Arabidopsis* roots.** Constrained PCoA of Bray-Curtis dissimilarities (constrained by all biological factors and conditioned by all technical variables) of soil, rhizosphere, and root samples. Gifu, Col-0, *A. lyrata* MN47, and *L. corniculatus* co-cultivated with *LjAt-SC3* (exp. F,  $n = 173$ , variance explained 65.1%,  $P = 0.001$ ). Aggregated RA of the 16 *Lj*-derived and the 16 *At*-derived strains in the live root of plants inoculated with *LjAt-SC3* ( $n = 72$ ).

#### 4.3.3 Host factors driving preferred associations in the root microbiota



**Figure 4.7 Constrained PCoA and aggregated RA of the commensal communities in *Lotus* and *Arabidopsis* roots and toothpick control.** (a) Dead roots of Gifu and Col-0, and toothpick cocultivated with *LjAt-SC3* (exp. J,  $n = 250$ , variance explained 43.9%,  $P = 0.001$ ). (b) Aggregated RA of the 16 *Lj*-derived and the 16 *At*-derived strains in the dead roots of *Lotus* and *Arabidopsis* plants inoculated with *LjAt-SC3* ( $n = 89$ ).

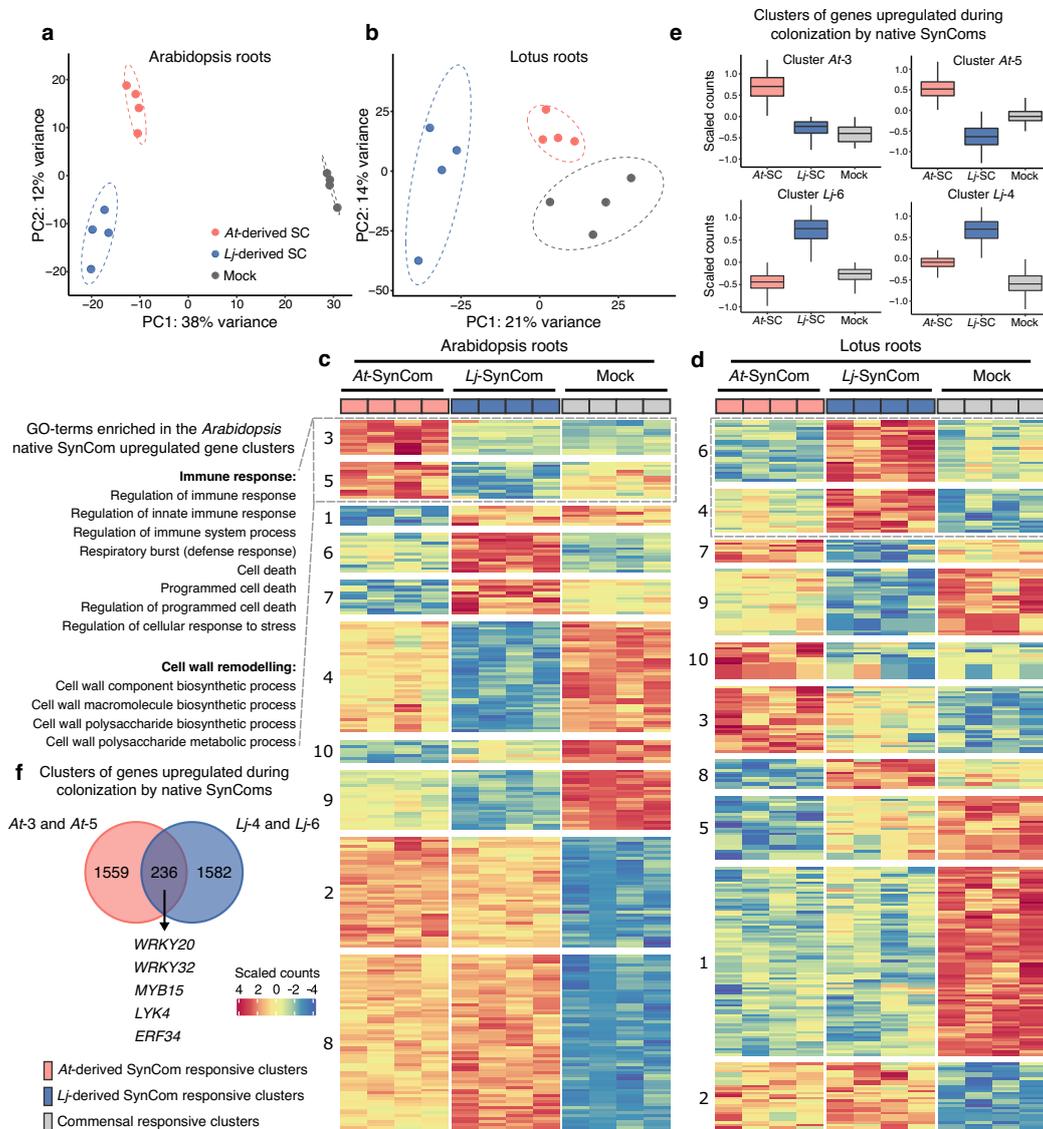
Previous studies have reported shifts in *At* leaf or root microbiota structure in mutants impaired in different host immunity pathways (Lebeis *et al.*, 2015; Chen *et al.*, 2020). We speculated that the plant immune system might also play a role in selecting commensal bacteria in a host-specific manner. We thus tested whether host mutants impaired in perception of ubiquitous microbe-associated molecular patterns (MAMPs) were also preferentially colonized by native commensal strains (exp. G). Community profiles of roots of *At* and *Lj* mutants lacking the receptor FLS2, which detects the bacterial flagellin epitope flg22 (*Ljfls2* and *Atfls2*) (Zipfel *et al.*, 2004; Mun *et al.*, 2016),

were indistinguishable from those of their respective wild types ([Extended Data Fig. 9a](#) in the published [journal version article](#)). Similar results were obtained with an *At* mutant lacking MAMP co-receptors BAK1 and BKK1 as well as CERK1 receptor kinase, known to play a role in the perception of the bacterial MAMP peptidoglycan (*Atbbc* triple mutant) ([Xin et al., 2016](#)). In addition, bacterial host preference was retained in those mutants ([Extended Data Fig. 9b](#) in the published [journal version article](#)). A separate experiment using the *dde2 ein2 pad4 sid2* (*deps*) mutant in *At*, which is simultaneously defective in all three major defense phytohormone signaling pathways (salicylic acid, jasmonate and ethylene) ([Tsuda et al., 2009](#)), showed comparable results (exp. H; [Extended Data Fig. 9c](#) and [9d](#) in the published [journal version article](#)). Together, these data suggest that the tested MAMP receptors and immune signaling pathways do not play a crucial role in preferential colonization by native commensal bacteria.

Plant root exudates contain molecular cues that can be differentially metabolized or perceived as signals by root microbiota members ([Bressan et al., 2009](#); [Zhalnina et al., 2018](#)). In particular, glucosinolates (GS), a group of nitrogen- and sulfur-containing metabolites found in root exudates throughout the family Brassicaceae, including *At*, are known to play a role in plant defense and serve as precursor of compounds that inhibit microbial growth ([Bednarek, 2012](#); [Pastorczyk & Bednarek, 2016](#); [Klein & Sattely, 2017](#)). Since legumes such as *Lj* lack genes required for GS biosynthesis, we speculated that secretion of these compounds by *At* might contribute to the observed differences in community structure. We therefore tested whether the *At cyp79b2 cyp79b3* double mutant ([Zhao, 2002](#)), which is defective in the production of microbe-inducible and tryptophan-derived metabolites, including indole GSs, was also preferentially colonized by native commensal strains (exp. H). Comparison of bacterial community profiling data suggests that indole GS had no impact on overall community structure or bacterial host preference *in planta* ([Extended Data Fig. 10](#) in the published [journal version article](#)). Notably, incubation of bacterial SynComs in root exudates from *Lj* and *At* plants in an *in vitro* millifluidics system (exp. I) resulted in small but significant community separation according to the plant genotype ([Supplementary Fig. 1a](#) in the published [journal version article](#); 5% of variance;  $P=0.002$ ). However, in this system, we observed a loss of the host preference phenotype ([Supplementary Fig. 1b](#) in the published [journal version article](#)), indicating that root exudates from axenic plants are not sufficient to recapitulate this phenomenon. This observation prompted the question of whether live root tissue was required for preferential colonization by native commensals. We profiled the bacterial communities associated with dead root material from flowering *Lj* and *At* wild-type plants and with inert lignocellulose matrices (softwood birch toothpicks) at 5, 12, and 19 days after inoculation with a mixed community (exp. J). Diversity analyses showed that dead roots and toothpicks harbored distinct microbial communities that were separated from those of soil or detritosphere (soil surrounding dead roots), independently of the timepoint ([Figure 4.7a](#)). This separation was likely driven by a significant increase in the relative abundance of Flavobacteria, a taxon associated with the capacity to decompose complex polysaccharides ([Lapébie et al., 2019](#)), and which dominates the dead root communities (53% RA on average). Unlike the large separation between living *Lj* and *At* roots (36% of the variance), we observed only a small but significant differentiation between *Lj* and *At* dead root communities (6.4% of variance;  $P=0.001$ ). Additionally, commensal host preference was undetectable in dead roots, where *Lj*- and *At*-derived strains reached similar aggregated relative abundances in root material harvested from either host ([Figure 4.7b](#)). Taken together, these results

suggest that a living root and other factors besides root exudates, such as a physical contact with the plant (i.e., host-commensal feedbacks) are required for host preference in the root microbiota.

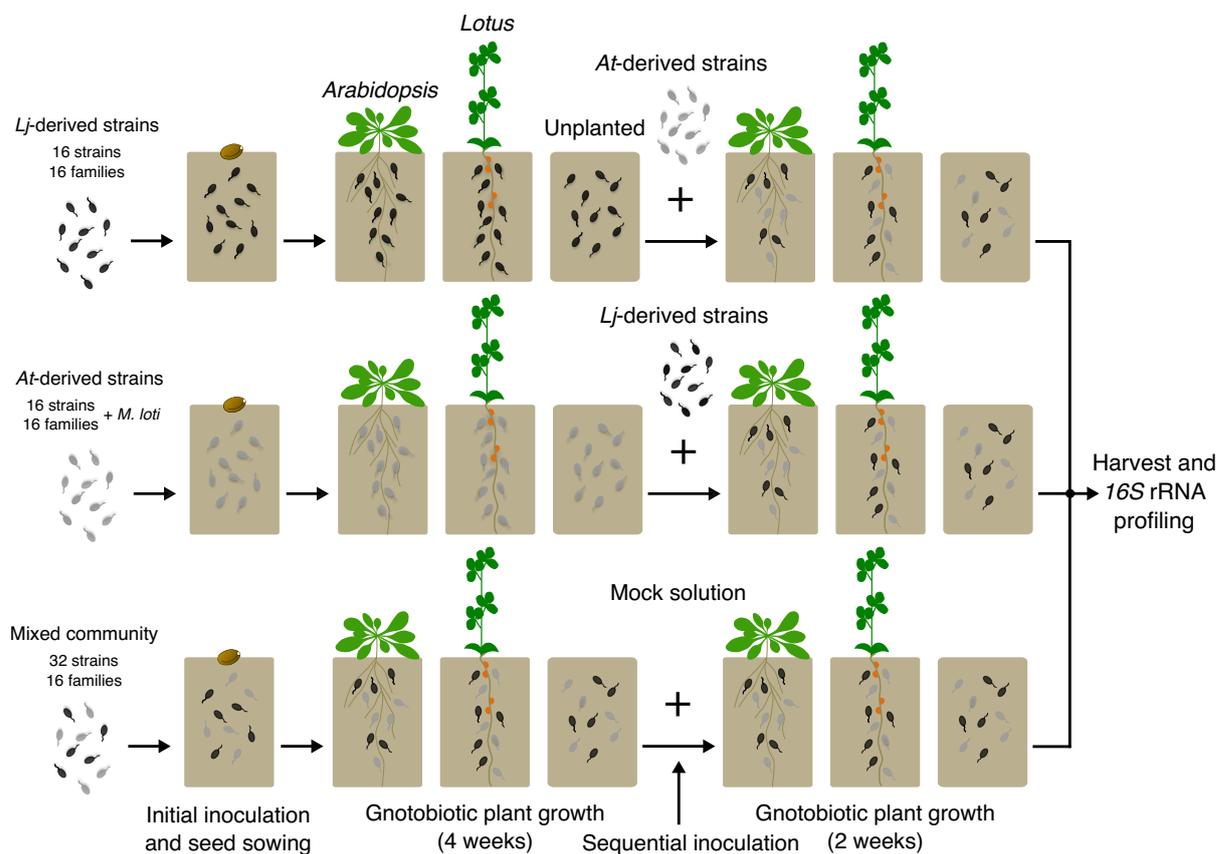
#### 4.3.4 SynCom-specific transcriptional responses of *Lj* and *At* roots



**Figure 4.8 SynCom-specific transcriptional outputs in *Lotus* and *Arabidopsis* roots.** Whole transcriptome-level Principal Component Analysis (PC) of *Arabidopsis* (a,  $n = 12$  biologically independent samples) and *Lotus* (b,  $n = 12$ ) roots after co-inoculation with host-specific SynComs (*Lj*- and *At*-SC3; exp. K). In the case of *Lotus* plants, a nodule isolate from the *Lj*-SPHERE collection was added to all treatments to prevent transcriptional outputs to be dominated by symbiosis or nitrogen-starvation responses. Heatmaps showing scaled counts of genes arranged according to  $k$ -means clustering results (only differentially expressed genes shown) for *Lotus* (c) and *Arabidopsis* (d). Distribution of expression patterns for clusters of genes upregulated after co-inoculation with native SynComs (e). Overlap in terms of homologues identified in the same clusters between the two host and a list of relevant transcription factors identified as potential key regulators of differential transcriptional responses (f).

Next, we sought to assess whether native, non-native or mixed commensal communities elicited a differential response in either host species. We grew wild-type *Lj* and *At* plants in our soil-based gnotobiotic system inoculated with *Lj*-, *At*- or mixed SynComs for five weeks (exp. K). Assessment

of plant performance revealed that treatment with commensal communities led to increased plant biomass and bacterial load compared to axenic controls, but not to differences according to SynCom treatment (Supplementary Fig. 2 in the published [journal version article](#)). Given the observation that a living root is required for commensal host preference, we conducted RNA sequencing of cross-inoculated *Lj* and *At* roots to explore host transcriptional responses that might mediate this process (exp. K). Analysis of these data showed that transcriptional outputs separated according to SynCom treatment in both hosts (Figure 4.8). Analysis of *k*-means clustering of whole transcriptomes revealed gene clusters associated with general response to bacterial colonization, as well as clusters specific to treatment with native or non-native SynComs. Among genes specifically induced by the native SynComs in both plant hosts we found several transcriptional regulators of immunity (e.g. WRKY20, WRKY32, and MYB15), well characterized MAMP receptor kinases (LYK4) and ethylene response factors (e.g. ERF34). This conserved pattern of differential response in the two plant species suggests a specific transcriptional response to native commensal communities that involves components of the host immune system. The differentially expressed transcription factors identified here constitute prime candidates for future exploration of the underlying mechanisms of differential microbiota assembly.

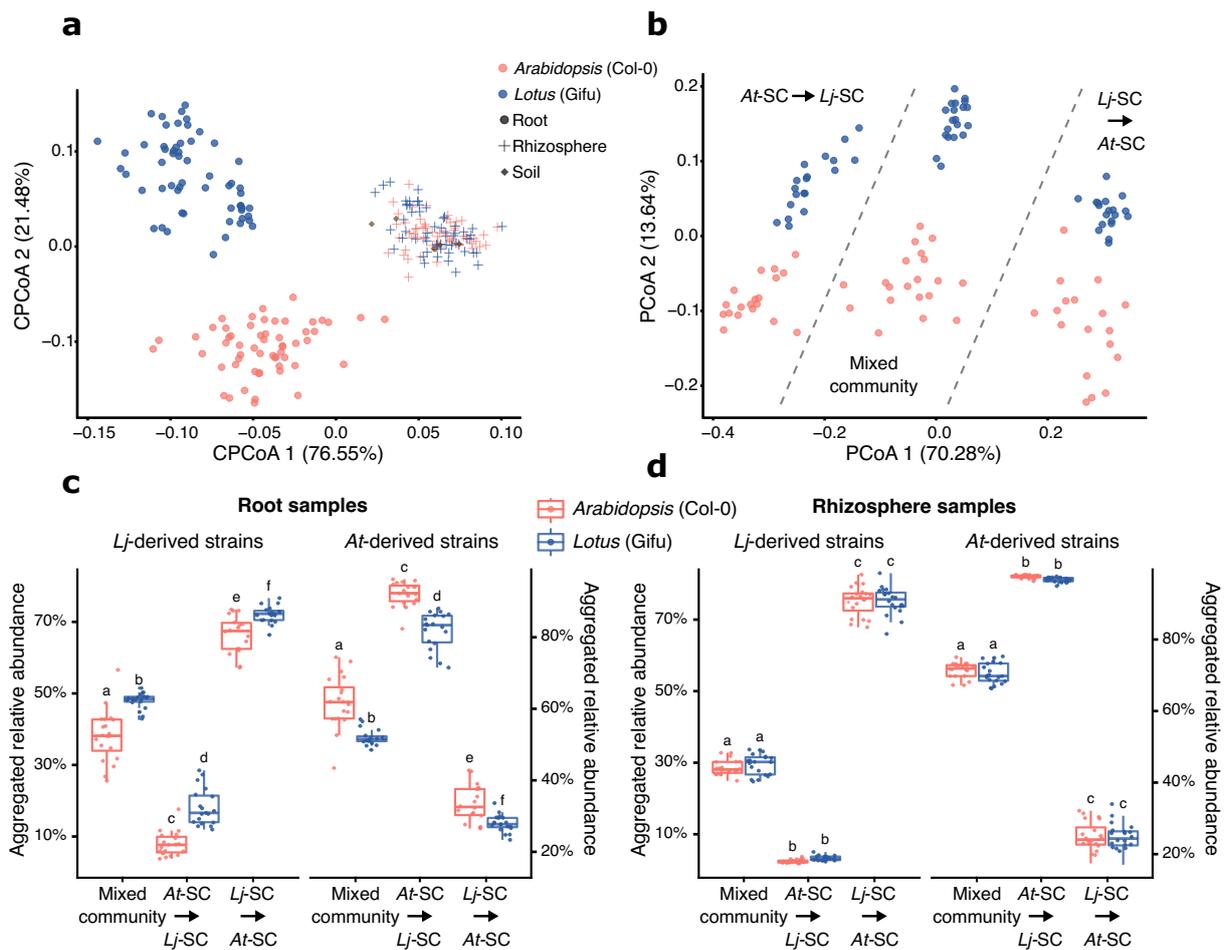


**Figure 4.9 Setup of the sequential inoculation experiment.** *L. japonicus* Gifu and *A. thaliana* Col-0 plants were co-cultivated with the mixed community *LjAt*-SC3, or individual SynComs *Lj*-SC3 and *At*-SC3, followed by inoculation with the contrasting SynCom (exp. L).

#### 4.3.5 Invasiveness and persistence in the root microbiota.

The results obtained from four independent experiments using five different mixed communities (Figure 4.4 – 4.7; Extended Data Figs. 4, 9 and 10 in the published [journal version article](#)) show

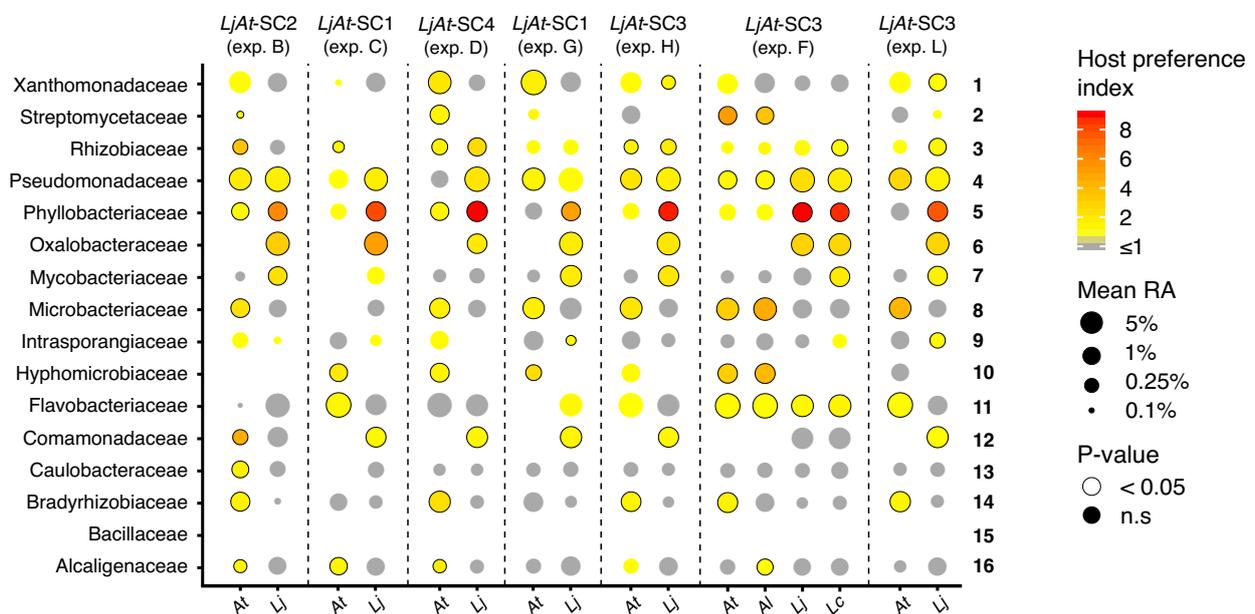
that native strains have a competitive advantage when colonizing roots of their cognate host. Ecological theory suggests that in the presence of a competitive hierarchy, the order of species arrival does not matter, as better adapted species tend to dominate irrespective of the history of the community (Fukami, 2015). To investigate the role that priority effects play in root community assembly we designed a series of sequential inoculation experiments using host-specific SynComs (exp. L; Figure 4.9). *At* and *Lj* wild-type plants were inoculated with taxonomically paired SynComs derived from *Lj* (*Lj*-SC3), *At* (*At*-SC3) or a mixed community (*LjAt*-SC3) for four weeks. Subsequently, we challenged the established root communities by adding the complementary SynCom (*At*-SC3 or *Lj*-SC3, respectively) to the soil matrix or, in the case of plants initially treated with the mixed community (*LjAt*-SC3), a mock solution (Figure 4.9). We then allowed all plants to grow for an additional two weeks before harvesting.



**Figure 4.10 Invasion and persistence of commensal bacteria.** Constrained PCoA of Bray-Curtis dissimilarities (constrained by all biological factors and conditioned by all technical variables;  $n = 267$ ; variance explained 14.7%,  $P = 0.001$ ) of soil, rhizosphere, and root samples (a), and PCoA of root samples only (b,  $n = 137$ ). Aggregated RA of the 16 *Lj*-derived and the 16 *At*-derived strains in *Lotus* and *Arabidopsis* root (c) ( $n = 120$ ) and rhizosphere (d) ( $n = 120$ ) samples in the indicated treatments.

Amplicon sequencing showed a significant separation of communities by compartment, and, within root samples, according to host species (Figure 4.10a;  $P=0.001$ ), mirroring the patterns observed in culture-independent community profiles (Figure 4.1a). Interestingly, analysis of beta-diversity of root samples at strain-level resolution revealed an effect of the treatment on community structure (Figure 4.10b), demonstrating that the order of arrival of strains affects community assembly.

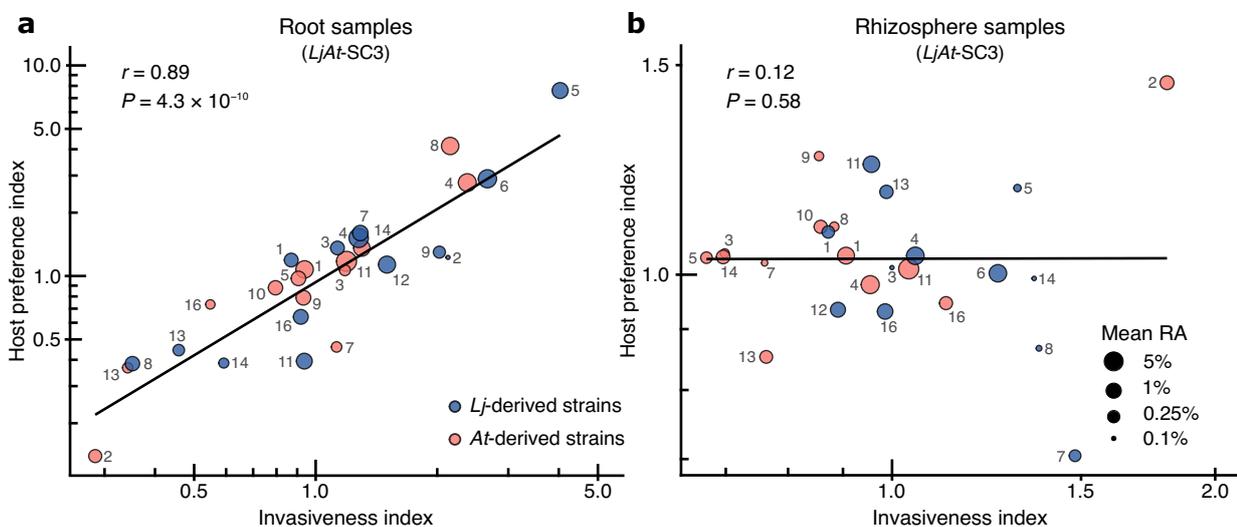
Examination of aggregated relative abundances showed that, in a competition context (i.e., initial inoculation with the mixed community *LjAt-SC3*), commensal SynComs preferentially colonized roots of their cognate host (Figure 4.10c), in line with results from the previous competition experiment shown in Figure 4.4 to Figure 4.7. However, in an invasion context, early-arriving SynComs invariably reached higher proportions in the output communities compared to the late-arriving SynComs (Figure 4.10c, d). Notably, estimation of absolute bacterial abundances showed that a secondary inoculation with an invading SynCom did not result in a significant increase in total bacterial load (Supplementary Fig. 3 in the published [journal version article](#)). Together, the results from our sequential inoculation experiments (Figure 4.10b, c) are indicative of the existence of priority effects in the root and rhizosphere microbiota, a well-known phenomenon in microbial community assembly (Fukami, 2015). These effects could be explained by niche preemption, where early-arriving community members reduce the amount of resources available (e.g. nutrients, space) for latecomers (Chase, 2003); alternatively, they could be the result of a feedback process between the host and the early-arriving commensals.



**Figure 4.11 Host preference of individual commensal strains across gnotobiotic experiments.** Each strain is represented by a dot, whose color corresponds to its host preference index and whose size to its average relative abundance ( $n = 366$ ). A significant host preference (Mann-Whitney test, FDR-corrected) is depicted by a black circle around a dot.

We hypothesized that commensal bacteria would be less affected by priority effects when colonizing their cognate host, given their competitive advantage with respect to non-native strains. In order to test this, we examined aggregated relative abundances of *Lj*- and *At*-derived SynComs in the root and rhizosphere communities. We found that host-specific SynComs were better able to invade a resident community in the roots of their cognate host compared to those of the other plant species (Figure 4.10c), thus reducing the strength of the priority effects. However, in the rhizosphere compartment of either plant species, host-specific SynComs showed neither host preference in a competition context nor differences in their ability to invade standing communities (Figure 4.10d). However, it is also possible that root communities did not reach equilibrium two weeks after invasion, and that the observed patterns could change over time.

We then tested if host preference was directly linked to invasiveness and to what extent these traits were found in individual community members. First, we quantified the strength of host preference by calculating the ratio between the relative abundance of each strain in their cognate host compared to the other plant species (host preference index, details in part 4.5 Materials and methods). Notably, although *Lj* root samples did not include nodules, but possibly contained incipient symbiotic events, the strains with the highest host preference index were the nitrogen-fixing *Lj* strains belonging to the Phyllobacteriaceae family (Figure 4.11), indicating that host preference of mutualistic rhizobia is not limited to nodule tissue. In addition, multiple other commensal strains showed significant host preference, with members of the families Pseudomonadaceae, Oxalobacteriaceae, Rhizobiaceae and Microbacteriaceae robustly displaying a high host preference index. Interestingly, members of these last two bacterial families also had a significant impact on community structure during invasion in a recent study with phyllosphere bacteria<sup>12</sup>. Next, we calculated an invasiveness index by comparing the ability of each strain to invade a standing community on their cognate host compared to the other plant species. We found a strong correlation between host preference and invasiveness of commensal bacteria which is independent of their relative abundance ( $r = 0.89$ ;  $P = 4.3 \times 10^{-10}$ ; Figure 4.12a). In contrast, this correlation was absent in the rhizosphere samples (Figure 4.12b), indicating that the link between these two bacterial traits is mediated by host attributes that do not extend to the rhizosphere. Together, our data show that host preference is prevalent in commensal bacteria from diverse taxonomic groups and that this trait is tightly linked to invasiveness and together play a role during root microbiota assembly.



**Figure 4.12 Host preference is linked to invasiveness.** Correlation between host preference and invasiveness index for each strain in root (**a**,  $n = 115$ ) and rhizosphere samples (**b**,  $n = 119$ ), respectively, obtained from the sequential inoculation experiment (exp. L). The color of each point designates the host of origin of each strain and the size denotes its mean relative abundance (log<sub>2</sub>-transformed). Each point is labeled with a numeric identifier that corresponds to the strains in Figure 4.11 (*LjAt-SC3*).

#### 4.4 Discussion

The current concept of host specificity in plant-microbe interactions was originally developed based on studies using microorganisms with either pathogenic or mutualistic lifestyles. Recently, it has been shown that soilborne, nitrogen-fixing *Ensifer meliloti* mutualists can adapt to local host

genotypes in only five plant generations, and proliferate to greater abundances in hosts with shared evolutionary histories (Batstone *et al.*, 2020). We show here that in the *Lj* and *At* root microbiota, there is a gradient of host preference among commensals belonging to diverse taxonomic lineages. Maintenance of host preference in the sympatric relative species *L. corniculatus* and *A. lyrata* raises the possibility that these commensals might have adapted to host features conserved in the respective plant genera. Alternatively, the observed host preference patterns might be the consequence of other ecological processes, such as ecological fitting, whereby organisms are able to colonize and persist in a new environment using traits that they already possess (Agosta & Klemens, 2008). Diversification of plant traits as a result of adaptation to edaphic or other environmental factors is expected to result in new host features that constitute novel root niches for microbial colonization. It is also possible that host diversification is partly driven by the adaptation of plants to commensal microbiota in soils with contrasting properties. However, the observation that in our experimental conditions, colonization by native or non-native bacterial SynComs had no impact on plant growth suggests that host preference is the result of microbial adaptation to host features instead of co-evolution. However, it is possible that a significant impact on host fitness might be observed in long-term experimentation, or in the presence of biotic or abiotic stresses, which were absent in the tested conditions, or in direct competition with other plant species. This latter hypothesis is supported by the observation that similarity between the root microbiota of different species affects competitive plant-plant interactions and has an impact on host performance through plant-soil feedback<sup>5</sup>. Future experimentation using multi-species gnotobiotic systems and varying environmental conditions will serve to test these hypotheses.

In aquatic and terrestrial ecosystems, microbial traits such as growth rate, antagonistic activity or resource use efficiency are known determinants of invasiveness (Litchman, 2010; Kinnunen *et al.*, 2016). In microbial communities associated with eukaryotic organisms, the ability to interact with the host might also be required for successful invasion. Our results indicate that native commensals have a competitive advantage when invading standing communities in the root but not in soil or rhizosphere. One possibility is that increased invasiveness by native bacteria is enabled by the existence of unfilled host species-specific root niches that can be occupied by late-comers. Alternatively, direct interaction of commensals with their host may be required to trigger the formation of host species-specific root niches, which could be linked to the specific transcriptional reprogramming in roots observed during colonization by native SynComs. This latter hypothesis is further supported by the observation that bacterial SynComs colonizing dead roots or incubated in root exudates *in vitro* showed no significant host preference. Our study provides a framework to test these hypotheses and to investigate the molecular basis of host preference in multiple taxa of the bacterial root microbiota in comparison with host adaptation mechanisms in plant pathogens and mutualists.

## 4.5 Materials and methods

### 4.5.1 Bacterial and plant material

Bacterial strains were grown in tryptic soy broth (15 g/L, TSB, Sigma-Aldrich) liquid medium or on agar plates containing 15 g/L of Bacto Agar (Difco) at 25°C. *Mesorhizobium* strains LjNodule210, LjNodule215, and LjNodule218, isolated from *L. japonicus* root nodules, were cultured in TY

medium (5 g/L tryptone, 3 g/L yeast extract) supplemented with 10 mM CaCl<sub>2</sub> or in YMB medium (5 g/L mannitol, 0.5 g/L yeast extract, 0.5 g/L K<sub>2</sub>HPO<sub>4</sub>·3H<sub>2</sub>O, 0.2 g/L MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.1 g/L NaCl). The composition of synthetic bacterial communities (SynComs) is listed in [Supplementary Table 1](#) in the published [journal version article](#). *L. japonicus* ecotype Gifu B-129 was used as wild type. Symbiosis-deficient mutant *nfr5-2* (Madsen *et al.*, 2003), and flagellin receptor-deficient mutant *fls2* (LORE1-30003492) (Mun *et al.*, 2016) were derived from the Gifu B-129 genotype. For *A. thaliana*, ecotype Columbia-0 was used as wild type. Mutant genotypes *fls2* (Zipfel *et al.*, 2004), *bbc* (Xin *et al.*, 2016), *deps* (Tsuda *et al.*, 2009), and *cyp79b2 cyp79b3* (Zhao, 2002) were available in our seed stock. *L. corniculatus* seeds, cultivated in the North-Western German lowland, were retrieved from Rieger-Hofmann GmbH, Blaufelden-Raboldshausen, Germany. *A. lyrata* MN47 seeds were a gift from Prof. Juliette de Meaux, University of Cologne.

#### 4.5.2 Establishment of the *L. japonicus* bacterial culture collection

The *L. japonicus* culture collection combines strains isolated during three independent isolation events. Bacterial isolation, DNA isolation, and identification using Illumina sequencing were performed as previously described (Bai *et al.*, 2015). Wild-type *L. japonicus* (ecotype Gifu B-129) plants were grown in natural soil (Cologne agriculture soil, CAS, batch 10 from Spring 2014, and batch 11 from spring 2015) in the greenhouse and harvested after four or eight weeks to cover different developmental stages. Root systems of 20 plants were subjected to DNA isolation and culture-independent community profiling *via* amplicon sequencing. From 45 plants, a 4-cm section of the roots was collected and rigorously washed 3x with phosphate-buffered saline (PBS; 130 mM NaCl [7.6 g/l], 7 mM Na<sub>2</sub>HPO<sub>4</sub> [1.246 g/l], 3 mM NaH<sub>2</sub>PO<sub>4</sub> [0.414 g/l], pH 7.0) and 3x with sterile water. Nodule and root parts were separated and homogenized independently. Homogenized roots from each individual plant were allowed to sediment for 15 min and the supernatant was diluted (1:20K, 1:40K, and 1:60K) with four different media: 3 g/L TSB, 50% TY, CY for enrichment of Myxococcales (3 g/L Casitone, 1.36 g/L CaCl<sub>2</sub>·2H<sub>2</sub>O, and 1 g/L yeast extract; pH adjusted to 7.2), and YAN, for enriching of Burkholderiales (10 g/L yeast extract, 1 g/L K<sub>2</sub>HPO<sub>4</sub>, and 0.5 g/L MgSO<sub>4</sub>·7H<sub>2</sub>O). Bacterial dilutions cultivated in 96-well microtiter plates. Homogenized nodules from each individual plant were directly diluted (1:20K, 1:40K and 1:60K) and cultivated in 96-well microtiter plates. This procedure was carried out for individual plants to obtain bacterial isolates from different plant roots. After 10–20 days at room temperature, plates that showed visible bacterial growth in around 30 wells were chosen for high-throughput sequencing. Bacterial isolates were identified with a two-step barcoded PCR protocol described previously (Bai *et al.*, 2015), with the difference that at the first step of the PCR, the v5-v7 fragments of the *16S* rRNA gene were amplified by the degenerate primers 799F (AACMGGATTAGATACCKG) and 1192R (ACGTCATCCCCACCTTCC), and indexing was done using Illumina-barcoded primers. The indexed *16S* rRNA amplicons were pooled, purified, and sequenced on the Illumina MiSeq platform. Strains isolated from nodules were tested for their ability to form functional nodules in *L. japonicus* Gifu plants grown on agar plates.

Cross-referencing of IRL sequences with culture-independent profiles was used to identify candidate strains for further characterization, purification, and whole-genome sequencing (WGS). Two main selection criteria were used: maximum taxonomic coverage, selecting candidates from as many taxa as possible; priority to strains whose *16S* sequences were highly abundant in the natural

communities. Whenever multiple candidates from the same phylogroup were identified, we aimed at obtaining multiple independent strains, if possible, coming from separate biological replicates to ensure they represented independent isolation events. After validation of selected strains, 294 (including 9 isolated from nodules) were successfully subjected to WGS.

#### 4.5.3 Culture-independent community profiling

Bacterial communities were profiled by amplicon sequencing of the variable v5-v7 regions of the bacterial *16S* rRNA gene. Library preparation for Illumina MiSeq sequencing was performed as described previously (Durán *et al.*, 2018). In all experiments, multiplexing of samples was performed by double-indexing (barcoded forward and reverse oligonucleotides for *16S* rRNA gene amplification).

*L. japonicus* Gifu and *A. thaliana* Col-0 were grown for five weeks in CAS soil (batch 15 from January 2020) in 7x7 cm pots alongside unplanted control pots under short-day conditions. Pots were watered with sterile water from the bottom as needed. Root, rhizosphere, and soil samples were harvested and processed as described previously (Thiergart *et al.*, 2020). In total, 15, 13, and 8 replicates were sampled for Col-0, Gifu, and unplanted controls, respectively. DNA was isolated from those samples using the MP Biomedicals FastDNA Spin Kit for Soil.

#### 4.5.4 Multi-species microbiota reconstitution experiments

We utilized the gnotobiotic FlowPot system (Kremer *et al.*; Durán *et al.*, 2018) to grow *A. thaliana* and *L. japonicus* plants with and without bacterial SynComs. In brief, the system allows for even inoculation of each growth pot with microbes by the flushing of pots with the help of a syringe attached to the bottom opening. Sterilized seeds are placed on the matrix (peat and vermiculite, 2:1 ratio), and pots are incubated under short-day conditions (10 hours light, 21°C; 14 hours dark, 19°C), standing in customized metal racks in sterile plastic boxes with filter lids (SacO2 microboxes, www.saco2.com). For SynCom preparation, bacterial commensals were grown separately in liquid culture for 2-5 days to reach high density, harvested, and washed in 10 mM MgSO<sub>4</sub>. Equivalent amounts of each strain were combined to yield the desired SynComs with an optical density (OD<sub>600</sub>) of 1. An aliquot of 200 μL of the SynCom as reference sample for the experiment start, and aliquots of 50 μL of the individual strains were taken and stored at -80°C for sequencing. The SynCom was added to the desired medium to reach a final OD<sub>600</sub> of 0.02. FlowPots were each flushed with 50 mL of inoculum (medium/SynCom mix). Generally, the medium used for inoculation was 0.25x B&D<sup>43</sup> supplemented with 1 mM KNO<sub>3</sub> for both plant species. In experiments D, F, K, and M (Supplementary Table 2 in the published [journal version article](#)), 0.5x MS (2.22 g/L Murashige+Skooog basal salts, Duchefa; 0.5 g/L MES anhydrous, BioChemica; adjusted to pH 5.7 with KOH) was used for *Arabidopsis*. The two plant species were grown in separate FlowPots side-by-side, with ten pots in total per plastic box. After five weeks of growth, roots were harvested and cleaned thoroughly from attached soil using sterile water and forceps. *Lotus* root segments containing nodules were omitted. Soil samples from planted and unplanted pots were collected as rhizosphere and soil samples, respectively. All root (epiphytic and endophytic compartments), rhizosphere, and soil samples were transferred to Lysing Matrix E tubes (FastDNA Spin Kit for Soil, MP Biomedicals), frozen in liquid nitrogen, and stored at -80°C for further processing. DNA was isolated from those samples using the FastDNA Spin Kit for Soil, and from

individual strains of the SynCom *via* quick alkaline lysis (Bai *et al.*, 2015), and subjected to bacterial community profiling or absolute quantification of bacteria. For RNA isolation, samples were harvested the same way and processed using the RNeasy Plant Mini kit (Qiagen, Hilden, Germany).

Mature root systems from Gifu and Col-0 plants grown in potting soil in the greenhouse were harvested from flowering plants (13-week old *Lotus*, 7-week old *Arabidopsis*), washed several times in water, padded on kitchen paper to remove moisture, and dried in big glass petri dishes at 120°C for one hour. Note that Gifu plants had a few small, most likely ineffective root nodules. Pieces of the dried, dead roots were planted into FlowPots under sterile conditions, and SynCom (*LjAt-SC3*) inoculation was performed as described above. Dead roots were recovered from the FlowPots after 5, 12, and 19 days of incubation, and washed and stored as described above for live roots.

#### 4.5.5 SynCom invasion experiments

FlowPots were sequentially inoculated with native and non-native strains. FlowPots were prepared as usual, with the addition of a round nylon filter (pore size 200  $\mu\text{m}$ ) at the bottom of the pot to avoid clogging of the bottom opening by matrix material. FlowPots were first inoculated with either the mixed SynCom (16 *Lj*- and 16 *At*-strains), the *At* SynCom (16 *At*-strains), the *Lj* SynCom (16 *Lj*-strains), or the mock solution (medium only). The medium used for inoculation was 0.25x B&D (Broughton & Dilworth, 1971) supplemented with 1 mM  $\text{KNO}_3$  for both plant species.

For sterilization, *A. thaliana* seeds were incubated for 5 min in 70% ethanol, then twice for 1 min in 100% ethanol, washed 5x with sterile water, and stored at 4°C in the dark for stratification. *L. japonicus* seeds were scarified by abrading the surface using sand paper, incubated for 20 min in diluted bleach, and washed 5x with sterile water. Sterilized seeds were placed on sterile Whatman paper wetted with sterile water in a squared petri dish and allowed to germinate under short-day conditions. Sterilized Col-0 seeds and germinated sterile Gifu seeds were placed on the soil surface. Note that a few drops of *Mesorhizobium* culture (*Lotus* root nodule symbiont, strain LjNodule218,  $\text{OD}_{600}$  0.02) were applied to Gifu seedlings in the *At* SynCom treatment to allow for normal root nodule symbiosis to occur and ensure healthy plant growth. After growth for four weeks, a second inoculation was performed, where a mock inoculum (medium) was added to the mixed SynCom-treated pots, the *Lj* SynCom was added to the *At* SynCom-treated pots, the *At* SynCom was added to the *Lj* SynCom-treated pots, and mock inoculum was added to the mock-treated pots. The pots were flushed in reverse by adding the inoculum from the top and applying vacuum from the bottom. On a sterile bench, FlowPots (cut 60-mL syringes with a male Luer Lok connector) were placed onto female Luer Lok connectors of a vacuum manifold (QIAvac 24 Plus, Qiagen), keeping the valves of the manifold closed. Vacuum was applied to the manifold with an attached vacuum pump. 20 mL of inoculum were carefully added to a pot with a 20-mL syringe and needle, avoiding damage of the plant shoots. Each pot was inoculated by opening and closing the corresponding valve. Pots were put back into the plastic containers and plants grown for another two weeks. Root, rhizosphere, and soil samples were harvested as described above.

#### 4.5.6 Collection of root exudates and millifluidics experiment

*Arabidopsis* and *Lotus* plants were grown in a customized hydroponic system (original design by Manuela Peukert, University of Cologne, unpublished). This sterile growth setup consists of glass jars filled with glass beads and a stainless-steel mesh on top. Nutrient solution (modified 0.25x B&D medium; Fe-EDTA instead of Fe-citrate) was poured into the jars until the beads were covered in liquid and the liquid touched the metal mesh. We employed the same medium for both plant species in order to allow for direct comparison of exudate composition, and to minimize differential effects on the bacterial community originating from different media types. We chose the *Lotus* B&D medium since *Arabidopsis* grew reasonably well in it. Sterilized and pregerminated seeds were placed onto the mesh, jars were put into sterile plastic boxes with filter lids (SacO2 microboxes), and plants were grown for five weeks. The medium containing root exudates was removed from the jars in the clean bench using a sterile metal needle and plastic syringe. After transfer to 50-mL Falcon tubes, exudates were frozen at -80°C, freeze-dried until a volume of 2-3 mL was left, thawed, and adjusted with sterile water to 5 mL. Exudates were kept at -80°C until further usage.

Bacterial incubation in root exudates was performed in a novel millifluidics system (MilliDrop Analyzer, MilliDrop, Paris, [www.millidrop.com](http://www.millidrop.com)). This drop-based system allows incubation of bacteria in very small volumes of root exudates or growth medium. In brief, bacteria and exudates or growth medium are combined in wells of a 96-well plate using a pipetting robot Freedom Evo 100 (Tecan, France). Droplets of approximately 100-200 nL are then sucked in from the wells of the loading plate by a tip on the robotic arm of the MilliDrop Analyzer, generating hundreds of droplets within an oil-filled tube, separated by air spacers. During incubation, the droplet “train” moves back and forth, so that during each round, each droplet passes a detector that counts the droplets. Culture droplets are collected after the experiment and subjected to community profiling.

The mixed community *LjAt-SC1* was used and was essentially prepared as described above for the *in planta* experiments, adjusted to OD<sub>600</sub> of 0.1 and used as input for preparation of the loading plate. Pure exudates (pH between 7.0 and 8.0) or a defined M9+carbon growth medium (1x M9 salts including phosphate buffer, 1 mM magnesium sulfate, 0.3 mM calcium chloride, 1x vitamin B solution, and artificial root exudates, pH 7.0) was used for incubation. Vitamin B solution contained 0.4 mg/L 4-aminobenzoic acid, 1 mg/L nicotinic acid, 0.5 mg/L calcium-D-pantothenate, 1.5 mg/L pyridoxine hydrochloride, 1 mg/L thiamine hydrochloride, 0.1 mg/L biotin, and 0.1 mg/L folic acid (modified from Pfennig, 1978). Artificial root exudates (modified from Baudoin *et al.* 2003) were composed of 0.9 mM glucose, 0.9 mM fructose, 0.2 mM sucrose, 0.8 mM succinic acid, 0.6 mM sodium lactate, 0.3 mM citric acid, 0.9 mM serine, 0.9 mM alanine, and 0.5 mM glutamic acid. Bacteria were incubated for three days, during which the pH of the cultures stayed stable. Droplets were collected in 6  $\mu$ L, and DNA isolated via quick alkaline lysis (Bai *et al.*, 2015), which consisted of addition of 10  $\mu$ L of buffer 1 (25 mM NaOH, 0.2 mM EDTA, pH 12), incubation at 95°C for 30 min, addition of 10  $\mu$ L of buffer 2 (40 mM Tris-HCl at pH 7.5), storage at -20°C.

#### 4.5.7 Mono-associations of SynCom members with host plants

*Lotus* seeds were sterilized and placed on sterile wet Whatman paper for germination. Seedlings were transferred to squared petri dishes containing 0.25x B&D medium (with Fe-EDTA instead of

Fe-citrate) supplemented with 3 mM KNO<sub>3</sub> and 1% Difco Bacteriological agar, and sterile filter paper was put on top of the sloped solidified medium prior to placing the seedlings to prevent root growth inside the agar. *Arabidopsis* seeds were sterilized and germinated on 0.5x MS medium plus 1% Difco Bacteriological agar. Seedlings were transferred to squared petri dishes containing 0.5x MS medium (neutral pH, buffered with 2 mM HEPES) plus 1% agar. The 32 strains of the mixed community *LjAt-SC3* were grown individually in liquid medium, harvested, and adjusted to an OD<sub>600</sub> of 0.02. Seedlings were inoculated by adding 500 μL of bacterial culture to the roots. Plants were grown for 14 days under long-day conditions (16/8 day-night cycles) at 21°C. Three biological replicates were prepared for each genotype-bacteria combination.

#### 4.5.8 Absolute quantification of bacteria

Genomic DNA was isolated from roots of plants grown in FlowPots (experiments K and L). DNA concentration was determined with the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific).

To quantify bacterial load on plant roots, the amount of bacterial DNA relative to the amount of plant DNA was determined *via* qPCR. For bacteria, the v5-v7 region of the *16S* rRNA gene was amplified using the AACMGGATTAGATACCCCKG (799F) and ACGTCATCCCCACCTTCC (1192R) primers. For Col-0, a fragment of At1g12360 was amplified using the TCCGGTCAATATTTTTGTTCG and TATAGCAGCGAAAGCCTCGT primers, and for Gifu, a fragment of the *NFR5* gene was amplified using the TCATATGATGGAGGAGTTGTCTGTT and ATATGAGCTTCGGAGCATGG primers. qPCR was performed as described previously (Lohmann *et al.*, 2010). The amount of *16S* rRNA was normalized to plant gene within each individual sample using the following equation:  $16S \text{ rRNA gene over plant gene} = 2^{-Ct(16S)} / 2^{-Ct(\text{plant})}$ .

For colony counts (exp. E), roots were harvested, washed, weighed, and crushed in 500 μl (Col-0) or 750 μl (Gifu) sterile water. Serial dilutions of 10<sup>-1</sup>, 10<sup>-2</sup>, 10<sup>-3</sup>, 10<sup>-4</sup>, and 10<sup>-5</sup> of the crushed roots were prepared in sterile water. 10 μl each were spotted onto 10% TSB agar square plates. Single colonies were counted after 1-3 days.

#### 4.5.9 Processing of *16S* rRNA gene amplicon data

Amplicon sequencing data from *L. japonicus* and *A. thaliana* roots of plants grown in CAS soil in the greenhouse, along with unplanted controls, were demultiplexed according to their barcode sequence using the QIIME (Caporaso *et al.*, 2010) pipeline. DADA2 (Callahan *et al.*, 2016) was used to process the raw sequencing reads of each sample. Unique amplicon variants (ASVs) were inferred from error-corrected reads, followed by chimera filtering, also using the DADA2 pipeline. ASVs were aligned to the SILVA database (Quast *et al.*, 2012) for the taxonomic assignment using the naïve Bayesian classifier implemented by DADA2. Raw reads were mapped to the inferred ASVs to generate a relative abundance table, which was subsequently employed for analyses of diversity and differential abundance using the R package *vegan* (Oksanen *et al.*, 2019).

Amplicon sequencing reads from the *Lotus* and *Arabidopsis* (Bai *et al.*, 2015) IRLs and from their corresponding culture-independent root community profiling were quality-filtered and demultiplexed according to their two-barcode (well and plate) identifiers using custom scripts and a combination of tools included in the QIIME (Caporaso *et al.*, 2010) and USEARCH (Edgar, 2010)

pipelines. Sequences were clustered into Operational Taxonomic Units (OTUs) with a 97% sequence identity similarity using the UPARSE algorithm, followed by identification of chimeras using UCHIME (Edgar *et al.*, 2011). Samples (wells) with fewer than 100 good quality reads were removed from the data set as well as OTUs not found in a well with at least ten reads. A purity threshold of 90% was chosen for identification of recoverable OTUs. We identified *Lj*-IRL samples matching OTUs found in the culture-independent root samples and selected a set of 294 representative strains maximizing taxonomic coverage for subsequent validation and WGS, forming the basis of the core *Lj*-SPHERE collection.

Sequencing data from SynCom experiments (including FlowPot and millifluidics experiments) were pre-processed similarly as natural community *16S* rRNA data. Quality-filtered, merged paired-end reads were then aligned to a reference set of sequences extracted from the whole-genome assemblies of every strain included in a given gnotobiotic experiment, using USEARCH (*uparse\_ref* command) (Edgar, 2013). Only sequences with a perfect match to the reference database were retained. We checked that the fraction of unmapped reads did not significantly differ between compartment, experiment or host species. We generated a count table that was employed for downstream analyses of diversity with the R package *vegan* (Oksanen *et al.*, 2019). We visualized amplicon data from all experimental systems using the *ggplot2* R package (Wickham, 2016).

#### 4.5.10 Host preference and invasiveness indices

In order to quantify the strength of the host preference of each bacterial strain individually, we calculated the ratio between the mean relative abundance of a given SynCom member in root samples of their cognate host and its mean relative abundance in root samples of the other plant species. The host preference indices depicted in Fig. 5a were calculated independently for each experiment. To avoid obtaining very high ratios due to small denominator values, strains with mean relative abundances below 0.1% in either of the two hosts were removed from the analysis. Similarly, an invasiveness index was calculated by obtaining the ratio between mean relative abundance of a strain when invading resident communities on roots of their cognate host, compared to the other plant species. The invasiveness index was calculated using samples from the sequential inoculation experiment (experiment L, Figure 4.8). The direct comparison between the two indices shown in Figure 4.10a and Figure 4.10b were calculated using samples from experiment L only, where invasion and competition treatments were performed in parallel. To test whether a SynCom member was significantly more abundant in the roots of their cognate host (i.e., significant host preference), we used the non-parametric Wilcoxon test controlling for false discovery rate (FDR) with  $\alpha = 0.05$ .

#### 4.5.11 Bacterial genome assembly, annotation, and phylogenomic analysis

Paired-end Illumina reads were first subjected to length-trimming and quality-filtering using Trimmomatic (Bolger *et al.*, 2014). Reads were assembled using the A5 assembly pipeline (Tritt *et al.*, 2012), which uses the IDBA algorithm (Peng *et al.*, 2012) to assemble error-corrected reads. Detailed assembly statistics and corresponding metadata can be found in Supplementary Data 2 in the published [journal version article](#). Genomes with multi-modal *k*-mer and GC content distributions or multiple instances of marker genes from diverse taxonomic groups were flagged as

not originating from clonal cultures. These samples were processed using a metagenome binning approach (Pasolli *et al.*, 2019). Briefly, contigs from each metagenome sample were clustered using METABAT2 (Kang *et al.*, 2019), followed by an assessment of completeness and contamination of each metagenome-assembled genome (MAG) using CheckM (Parks *et al.*, 2015). Only bins with completeness scores larger than 75% and contamination rates lower than 5% were retained and added to the collection (Supplementary Data 2 in the published [journal version article](#); designated MAG in the column ‘type’). Functional annotation of genes was conducted using Prokka and employing a custom database based on KEGG Orthologue (KO) groups (Kanehisa *et al.*, 2014) downloaded from the KEGG FTP server in November 2019. Hits to sequences in the database were filtered using an *E* value threshold of  $10 \times 10^{-9}$  and a minimum coverage of 80% of the length of the query sequence.

Genomes from the *Lj*- and *At*-SPHERE culture collections (Bai *et al.*, 2015) were searched for the presence of a set of 31 conserved, single-copy marker genes, known as AMPHORA (Wu & Eisen, 2008) genes. Sequences of each gene were aligned using Clustal Omega (Sievers *et al.*, 2011) with default parameters. Using a concatenated alignment of each gene, we inferred a maximum likelihood phylogeny using FastTree (Price *et al.*, 2010). We visualized this tree using the Interactive Tree of Life web tool (Letunic & Bork, 2019). Genomes from both collections (*Lj*-SPHERE and *At*-SPHERE) were clustered into phylogroups, roughly corresponding to a species designation (Olm *et al.*, 2020) using FastANI (Jain *et al.*, 2018) and a threshold of average nucleotide identity at the whole genome level of at least 97%.

#### 4.5.12 RNA-sequencing and data analysis

RNA isolated from FlowPot samples was subjected to quality control, library preparation, and sequencing (on the Illumina HiSeq3000 platform) at the Max Planck-Genome center, Cologne, Germany (<https://mpgc.mpiiz.mpg.de/home/>). Sequencing depth was 6 million reads per sample. Raw Illumina RNA-Seq reads were pre-processed using fastp (v0.19.10) (Chen *et al.*, 2018) with default settings for pair-end reads. High quality reads were pseudo-aligned to the *Lotus japonicus* Gifu *Arabidopsis thaliana* Col-0 transcriptome reference using kallisto (v0.46.1) (Bray *et al.*, 2016). After removal of low abundant transcripts that were not present in at least two replicates under each condition, count data were imported using the *tximport* package (Soneson *et al.*, 2016).

Differential expression analyses were performed using the *DESeq2* package (Love *et al.*, 2014). Firstly, raw counts were normalized with respect to the library size (*rlog* function) and transformed into  $\log_2$  scale. We tested for sample effects by surrogate variable (SV) analysis using the *sva* package (Leek *et al.*, 2012). Significant SVs were automatically detected and integrated into the model for differential analyses. Principal component analysis based on whole transcripts were then conducted and plotted to visualize the cluster and variance of biological replicates under each condition. Transcripts with fold-changes  $> 1.5$  and adjusted *P*-value for multiple comparisons (Benjamini–Hochberg method) equal to or below 0.05 were considered significant.

The  $\log_2$  scaled counts were normalized by the identified SVs using the *limma* package (Ritchie *et al.*, 2015) (‘removeBatchEffect’ function), and transformed as median-centered *z*-score (by transcripts, ‘scale’ function). Then *z*-scores was used to conduct *k*-means clustering for all transcripts. The cluster number ( $k = 10$ ) was determined by sum of squared error and Akaike

information criterion. Differential expressed transcripts and cluster results were visualized using heatmaps generated by *ComplexHeatmap* package (Gu et al., 2016).

Gene ontology (GO) enrichment for each cluster using the whole *Lotus* and *Arabidopsis* transcriptomes as backgrounds were performed with the *goseq* package (Young et al., 2010), which considers the transcripts length bias in RNA-Seq data. GO annotations were retrieved from the Gene Ontology Consortium (September 2019) (Ashburner et al., 2000; ‘The Gene Ontology Resource: 20 years and still GOing strong’, 2019). Significantly changed biological process GO terms (adjusted  $p$ -value < 0.05) were visualized in dot plots using the *clusterProfiler* package (Yu et al., 2012).

#### 4.5.13 Statistics and reproducibility

All experiments were performed with full factorial (biological and technical) replication. Competition experiments using SynComs were in addition repeated multiple times (see Extended Data Fig. 1 in the published [journal version article](#)) using independent bacterial communities. Whenever bacterial abundances or plant growth parameters were compared, we employed a two-sided, non-parametric Mann-Whitney test or, in the case of multiple comparisons, a Kruskal-Wallis test, followed by a Dunn’s *post hoc*. Whenever appropriate,  $P$ -values were adjusted for multiple testing using the Benjamini–Hochberg method ( $\alpha=0.005$ ). Statistical tests on beta-diversity analyses were performed using a PERMANOVA test with 5,000 random permutations. Whenever boxplots were employed in figures, data was represented as median values (horizontal line), Q1–1.5 × interquartile range (IQR; box) and Q3 + 1.5 × IQR (whiskers).

#### 4.6 Date and code availability

The strains of the *Lj*-SPHERE collection will be deposited at and will be available upon request from the Leibniz Institute DSMZ in Braunschweig, Germany. Raw *16S* rRNA amplicon reads will be deposited in the European Nucleotide Archive (ENA) under the accession number PRJEB37695. Similarly, sequencing reads and genome assemblies of the *Lj*-SPHERE core collection will be uploaded to the same database with the accession number PRJEB37696. The scripts used for the computational analyses described in this study are available at <http://www.github.com/garridoo/ljsphere>, to ensure replicability and reproducibility of these results.

#### 4.7 Author contributions

Wippel and Tao et al., 2021

K.W., K.T., S.R., P.S.-L. and R.G.-O. conceived the research and designed the experiments. K.W., K.T., R.Z. and D.B.J. established the *Lj*-SPHERE culture collection. K.W., K.T. and N.K. performed the gnotobiotic competition experiments. K.W. and E.L. conducted the *in planta* invasion and millifluidics SynCom experiments. R.G. and R.G.-O. analyzed culture-independent amplicon data. E.D. and R.G.-O. analyzed the *Lj*-IRL data. P.Z. and R.G.-O. processed bacterial whole-genome data from the *Lj*-SPHERE collection. Y.N. and R.G.-O. analyzed the transcriptome data. K.W., R.G.-O. and N.K. analyzed sequencing data from the SynCom experiments. K.W., K.T., S.R., P.S.-L. and R.G.-O. interpreted data and wrote the paper.

# Chapter 5 Shared features and reciprocal complementation of the *Chlamydomonas* and *Arabidopsis* microbiota

## 5.1 Abstract

Microscopic algae release organic compounds to the region immediately surrounding their cells, known as the phycosphere, constituting a niche for colonization by heterotrophic bacteria. These bacteria take up algal photoassimilates and provide beneficial functions to their host, in a process that resembles the establishment of microbial communities associated with the roots and rhizospheres of land plants. Here, we characterize the microbiota of the model alga *Chlamydomonas reinhardtii* and reveal extensive taxonomic and functional overlap with the root microbiota of land plants. Reconstitution experiments using synthetic communities derived from *C. reinhardtii* and *Arabidopsis thaliana* show that phycosphere and root bacteria assemble into taxonomically equivalent communities on either host. We show that provision of diffusible metabolites is not sufficient for phycosphere community establishment, which additionally requires physical proximity to the host. Our data suggests that the microbiota of photosynthetic organisms, including green algae and flowering plants, assembles according to core ecological principles.

## 5.2 Introduction

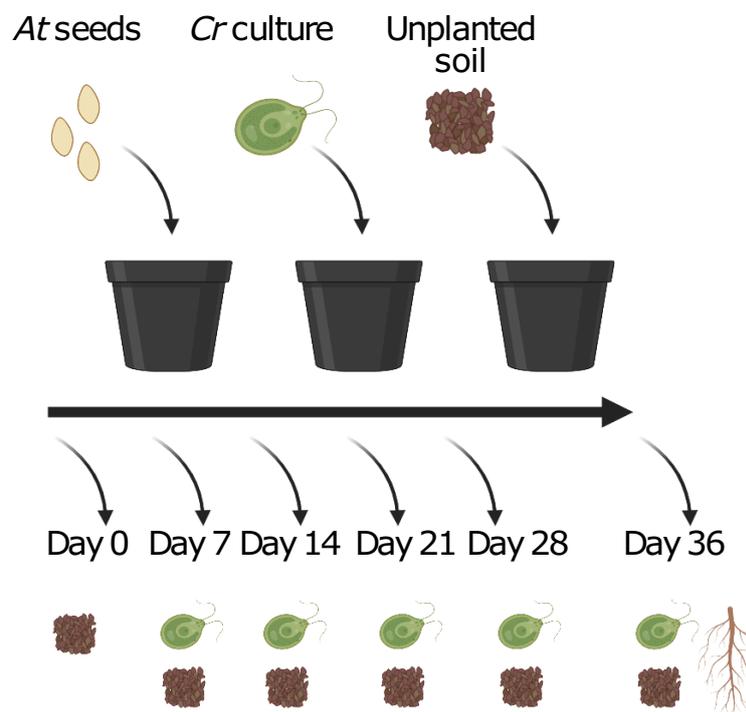
Plants associate with diverse microbes in their aerial and belowground tissues which are recruited from the surrounding environment. These microbial communities, known as the plant microbiota, provide the host with beneficial functions, such as alleviation of abiotic stresses (de Zélicourt *et al.*, 2018; Xu *et al.*, 2018; Berens *et al.*, 2019; Simmons *et al.*, 2020), nutrient mobilization (Castrillo *et al.*, 2017; Zhang *et al.*, 2019; Harbort *et al.*, 2020), or protection against pathogens (Durán *et al.*, 2018; Carrión *et al.*, 2019). Characterization of the microbiota associated with a wide range of plant species including liverworts (Alcaraz *et al.*, 2018), lycopods, ferns (Yeoh *et al.*, 2017), gymnosperms (Beckers *et al.*, 2017; Cregger *et al.*, 2018), and angiosperms (Bulgarelli *et al.*, 2012, 2015; Lundberg *et al.*, 2012; Schlaeppi *et al.*, 2014; Edwards *et al.*, 2015; Zgadzaj *et al.*, 2016; Walters *et al.*, 2018; Thiergart *et al.*, 2020) shows a strong influence of host phylogeny as well as conserved and possibly ancestral community features. Furthermore, it has been speculated that the ability to form associations with members of these communities, such as mycorrhizal fungi, was a trait required for the colonization of land by plants 450 Mya, possibly inherited from their algal ancestor (Knack *et al.*, 2015; Delaux *et al.*, 2015). Algae are also known to associate with complex bacterial communities termed phycosphere microbiota, particularly in aquatic environments (Kim *et al.*, 2014; Amin *et al.*, 2015; Seymour *et al.*, 2017; Cirri & Pohnert, 2019), where exchange of metabolites, including organic carbon (Moran *et al.*, 2016; Wienhausen *et al.*, 2017; Toyama *et al.*, 2018; Fu *et al.*, 2020), soluble micronutrients (Amin *et al.*, 2009), vitamins (Croft *et al.*, 2005; Grant *et al.*, 2014; Paerl *et al.*, 2017), and other molecular currencies (Teplitski *et al.*, 2004; Wichard *et al.*, 2015) influence algal growth and development. These parallelisms suggest that the phycosphere is analogous to the rhizosphere environment, in which secreted diffusible compounds alter soil pH, oxygen availability, concentration of antimicrobials and organic carbon, and thus support distinct microbial communities by favoring the growth of certain bacteria while restricting

proliferation of others (Bell & Mitchell, 1972; Bulgarelli *et al.*, 2013; Amin *et al.*, 2015; Krohn-Molt *et al.*, 2017; Shibl *et al.*, 2020). However, it is not yet known whether the ability to assemble a complex microbiota from the surrounding soil is also conserved in soil-borne microscopic algae, and to what extent they overlap with those of vascular plants.

In this study, we characterize the microbiota of the model green alga *C. reinhardtii* (*Cr*), and show significant taxonomic and functional similarities between the root and phycosphere microbiota. In addition, we report a comprehensive, whole-genome sequenced culture collection of *Cr*-associated bacteria that includes representatives of the major taxa found in associations with land plants. We then introduce a series of gnotobiotic systems designed to reconstruct artificial phycospheres that recapitulate natural communities using synthetic communities (SynComs) assembled from bacterial isolates. Cross-inoculation and competition experiments using the model plant *Arabidopsis thaliana* (*At*) and its associated bacterial culture collection (Bai *et al.*, 2015) indicate a degree of functional equivalence between phycosphere and root bacteria in associations with a photosynthetic host. Finally, we show that physical proximity between *Cr* and its microbiota is required for the establishment of fully functional phycosphere communities, suggesting that this process is not exclusively driven by the exchange of diffusible metabolites.

## 5.3 Results

### 5.3.1 *C. reinhardtii* assembles a distinct microbiota from the surrounding soil

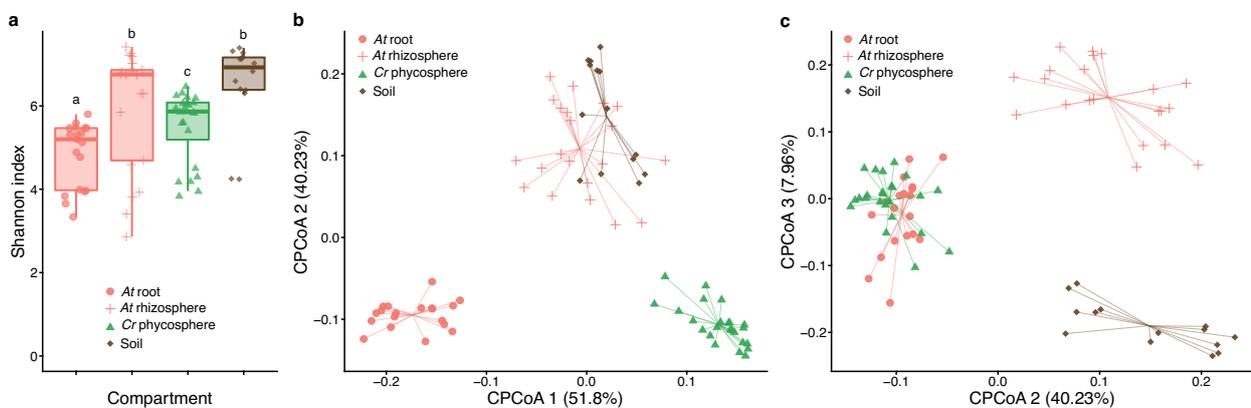


**Figure 5.1 Schematic description of the greenhouse experiment.** Pots containing CAS natural soil were either sown with *At* seeds, inoculated with *Cr* cultures or mock-treated. Samples were taken over time for bacterial community profiling.

To determine whether *Cr* shapes soil-derived bacterial communities similarly to land plants, we designed an experiment where *At* and *Cr* were grown in parallel in natural soil in the greenhouse (Figure 5.1). Briefly, pots containing Cologne Agricultural Soil (CAS) were inoculated with axenic

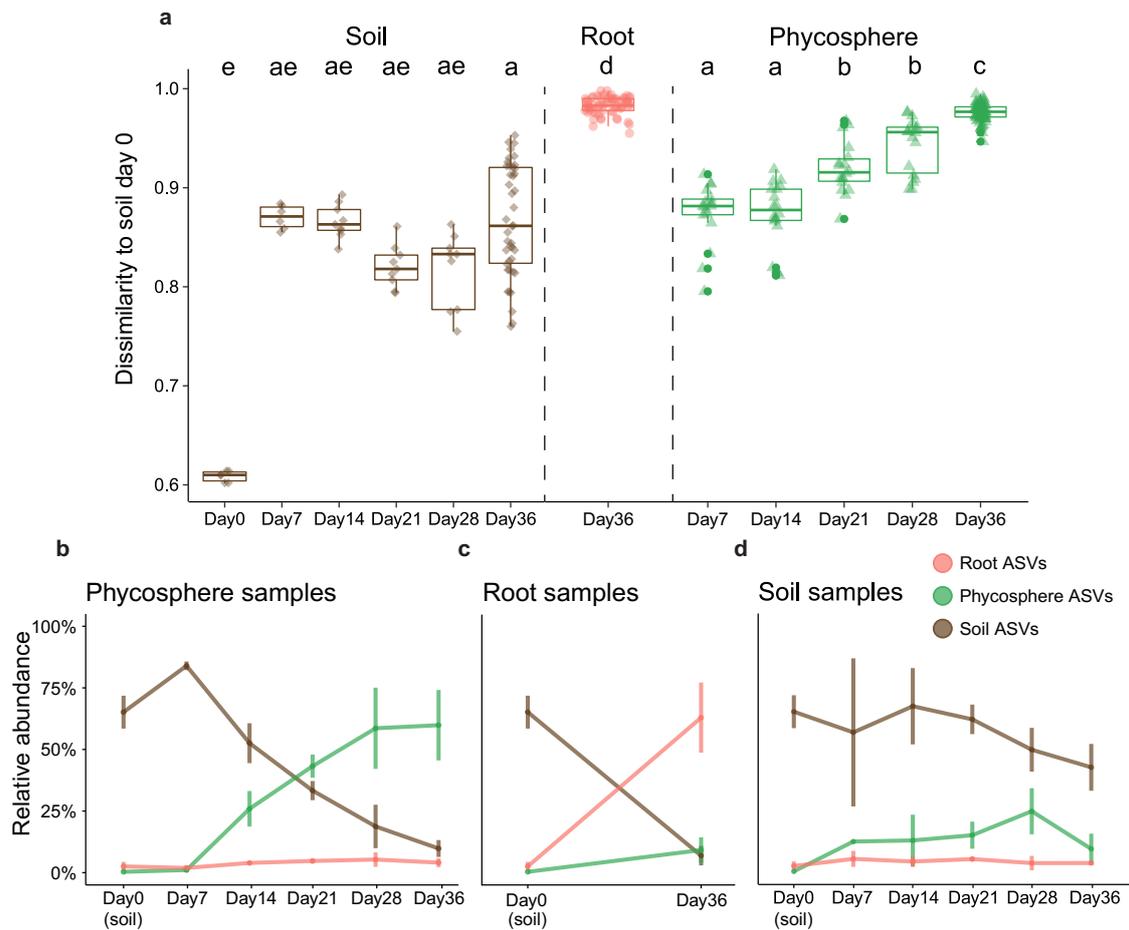
*Cr* (CC1690) cultures or sowed with surface-sterilized *At* (Col-0) seeds. We then collected samples from unplanted controls and from the surface of *Cr*-inoculated pots (phycosphere fraction) at 7-day intervals, and harvested the root and rhizosphere of *At* plants after 36 days. Bacterial communities from all compartments were characterized by *16S* rRNA amplicon sequencing.

Analysis of bacterial community profiles showed a decrease in  $\alpha$ -diversity (Shannon index) in the phycosphere and root compartments compared to the more complex soil and rhizosphere communities (Figure 5.2a). In addition, analysis of  $\beta$ -diversity revealed a significant separation by compartment, where phycosphere and root samples formed distinct clusters that were also separated from those consisting of soil and rhizosphere samples (Figure 5.2b; 22.4% of variance;  $P < 0.001$ ). Further inspection of amplicon profiles showed an overlap between root- and phycosphere-associated communities along the second and third components (Figure 5.2c), suggesting similarities between the bacterial communities that associate with *Cr* phycospheres and *At* roots.



**Figure 5.2 Comparison of bacterial community structures associated with *At* roots and the *Cr* phycosphere in natural soil.** (a) Alpha diversity estimates of soil, rhizosphere, root and phycosphere samples from *At* and *Cr* grown in CAS soil in the greenhouse. (b-c) PCoA of Bray-Curtis dissimilarities constrained by compartment (22.4% of variance explained;  $P < 0.001$ ). A separation between root, phycosphere and soil-derived samples can be observed in the first two components (b), while the root and phycosphere communities cluster together in the second and third PCoA axes (c).

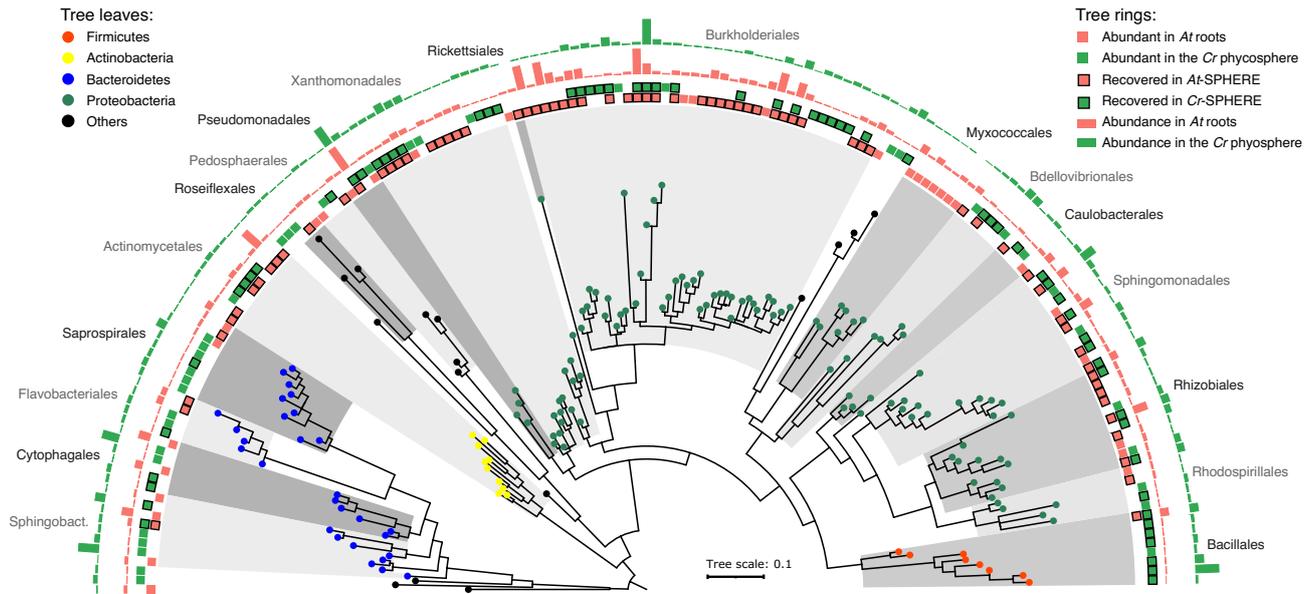
To characterize the dynamics of these microbiota assembly processes, we analyzed the time-series data from soil and phycosphere and end-point community profiles from *At* roots. This revealed a gradual recruitment of bacterial taxa from soil, leading to the formation of distinct phycosphere communities that become significantly differentiated 21 days after inoculation, which is of comparable to that observed in *At* root-associated communities at day 36 (Figure 5.3a). Subsequent enrichment analysis of amplicon sequence variants (ASVs) in each compartment, compared to unplanted soil, showed an increase in the relative abundance of *Cr*- and *At*-enriched ASVs in phycosphere and root samples, respectively. In contrast, total relative abundance of soil-enriched ASVs progressively decreased in host-associated compartments, while remaining stable in unplanted soil (Figure 5.3 b-d). Although the magnitude of the changes in bacterial community composition in the phycosphere diminishes over time, it remains unclear whether these communities reach a steady state over the duration of the experiment. Taken together, these results indicate that, similarly to *At*, *Cr* is able to recruit a subset of bacterial taxa from the surrounding soil and assemble a distinct microbiota.



**Figure 5.3 Culture-independent analysis of phycosphere- and root-associated communities in a natural soil.** (a) Bray-Curtis dissimilarities between phycosphere, root and soil communities, compared to the initial soil input (day 0). Boxplots are color-coded depending on the fraction. Significant differences are marked with different letters (Kruskal-Wallis, followed by a Dunn's *post hoc*, with Bonferroni correction). (b-d) Dynamics of relative abundances of ASVs enriched in phycosphere (b), root (c) or soil samples (d) over time, compared to initial soil input (day 0; Wilcoxon test;  $P < 0.05$ ). Curves are color-coded depending on the fraction indicated.

### 5.3.2 The *C. reinhardtii* phycosphere and the plant root share a core microbiota

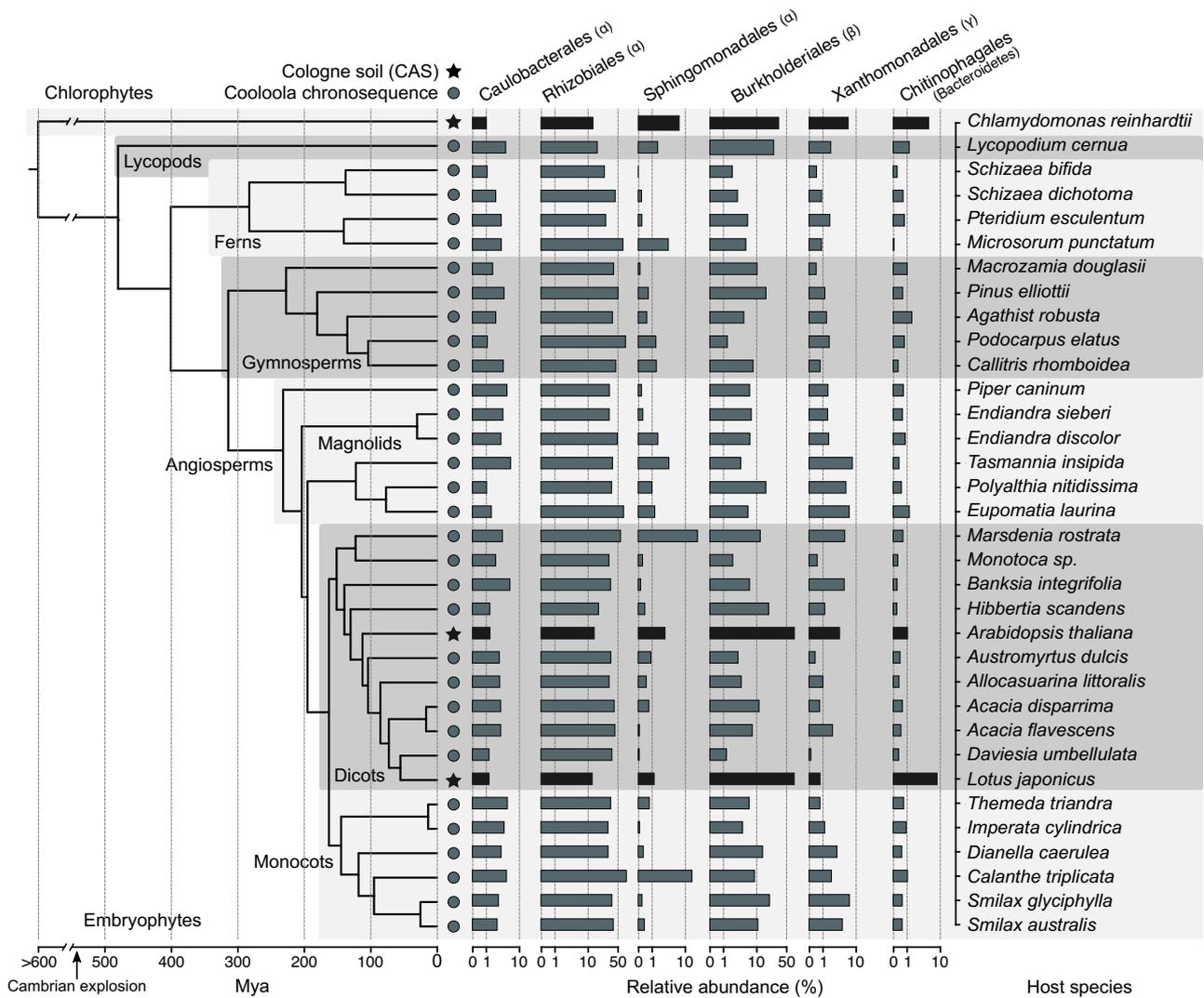
Given the observed similarities between phycosphere and root communities (Figure 5.2c), we compared the most abundant taxonomic groups found in association with the two photosynthetic hosts. We found a significant overlap between Operational Taxonomic Units (OTUs) with the highest relative abundances in either phycosphere or root samples (Figure 5.4;  $>0.1\%$  relative abundance; 32% shared;  $P < 0.001$ ), which included members of every bacterial order except Myxococcales, which were only found in large relative abundances in *At* root samples (Supplementary Data 1 in the [journal version article](#)). In line with previous descriptions of the *At* root microbiota, we observed that these host-associated communities were dominated by Proteobacteria, and also included members of the Actinobacteria, Bacteroidetes, and Firmicutes phyla. At this taxonomic level, the major difference between the two photosynthetic hosts was given by a lower contribution of Actinobacteria and a larger relative abundance of Firmicutes in the *Cr* phycosphere compared to the *At* root compartment (Figure 5.4). Given that this latter phylum is most abundant in soil, this difference may be due to the difficulty of fully separating soil particles from the phycosphere fraction during sample collection.



**Figure 5.4 Phylogeny of 16S rRNA sequences of the most abundant OTUs found in *At* roots and *Cr* phycosphere community profiles.** Leaf nodes are colored by taxonomic affiliation (phylum level). The two innermost rings (colored squares) represent abundant OTUs in each compartment. Squares highlighted with a black contour correspond to OTUs for which at least one representative bacterial strain exists in the IRL or IPL culture collections. The two outermost rings (barplots) represent log-transformed relative abundances of each OTU in *At* root or *Cr* phycosphere samples.

Next, we sought to assess whether the observed overlap in community structures between *Cr* and *At* could be extended to other land plant lineages. We performed a meta-analysis, broadening our study to include samples from phylogenetically diverse plant species found in a natural site, including lycophytes, ferns, gymnosperms, and angiosperms (Yeoh *et al.*, 2017), as well as the model legume *Lotus japonicus* (*Lj*) grown in CAS soil in the greenhouse (Thiergart *et al.*, 2019; Harbort *et al.*, 2020). First, we determined which taxonomic groups were present in each plant species ( $\geq 80\%$  occupancy and  $\geq 0.1\%$  average relative abundance) and identified a total of six bacterial orders that consistently colonize plant roots (i.e., found in every host species). These taxa include Caulobacteriales, Rhizobiales, Sphingomonadales, Burkholderiales, Xanthomonadales (Proteobacteria), and Chitinophagales (Bacteroidetes). We observed that the aggregated relative abundance of these six bacterial orders accounted for 39% of their respective communities on average (Figure 5.5). Interestingly, these taxa were also found among the most abundant in the *Cr* phycosphere (45% aggregated relative abundance), indicating that they are also able to associate with *Cr*. These results suggest the existence of a common principle for microbiota assembly across a wide phylogenetic range of photosynthetic hosts, which includes uni- and multicellular eukaryotic organisms.

(Figure on next page) **Figure 5.5 Conservation of bacterial orders of the root and phycosphere microbiota across photosynthetic organisms.** Phylogeny inferred from a multiple sequence alignment of the ribulose-bisphosphate carboxylase gene (*rbcL*) of 35 plant species and *Chlamydomonas reinhardtii*. The barplots represent the average aggregated relative abundance of the six bacterial orders found to be present in the root microbiota of each plant species (80% occupancy and  $\geq 0.1\%$  average relative abundance). Leaf nodes depicted with a star symbol denote community profiles of plants grown in CAS soil in the greenhouse (Thiergart *et al.*, 2019; Harbort *et al.*, 2020), whereas those marked with a circle were obtained from plants sampled at the Cooloola natural site chronosequence (Yeoh *et al.*, 2017).



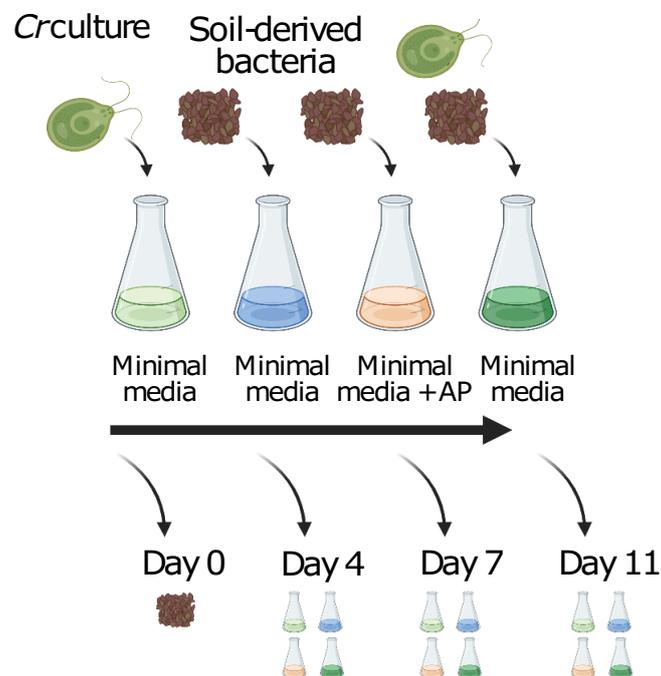
(Caption on previous page) **Figure 5.5 Conservation of bacterial orders of the root and phycosphere microbiota across photosynthetic organisms.**

### 5.3.3 Reconstitution of phycosphere communities using reductionist approaches

After the characterization of phycosphere-associated bacterial communities in natural soil, we sought to develop systems of reduced complexity that would allow controlled perturbation of environmental parameters, and targeted manipulation of microbial community composition. First, we established a mesocosm system using soil-derived microbial communities as start inocula (Figure 5.6). We co-inoculated axenic *Cr* (CC1690) cultures with microbial extracts from two soil types (CAS and Golm) in two different carbon-free media (TP and B&D), which ensures that the only source of organic carbon to sustain bacterial growth is derived from *Cr* photosynthetic activity. These phycosphere mesocosms were then incubated under continuous light for 11 days, during which we assessed *Cr* growth using cell counts, and profiled bacterial communities *via* 16S rRNA amplicon sequencing. In this system, *Cr* was able to steadily grow without a detrimental impact from co-inoculation with soil-derived bacterial extracts (Figure S3A in the [journal version article](#)).

Analysis of diversity showed that *Cr* was able to shape soil-derived bacterial communities within the first 4 days, compared to the starting inocula, and that these phycosphere communities remained stable until the end of the experiment (Figure 5.7). Interestingly, cultivation of soil-derived bacteria

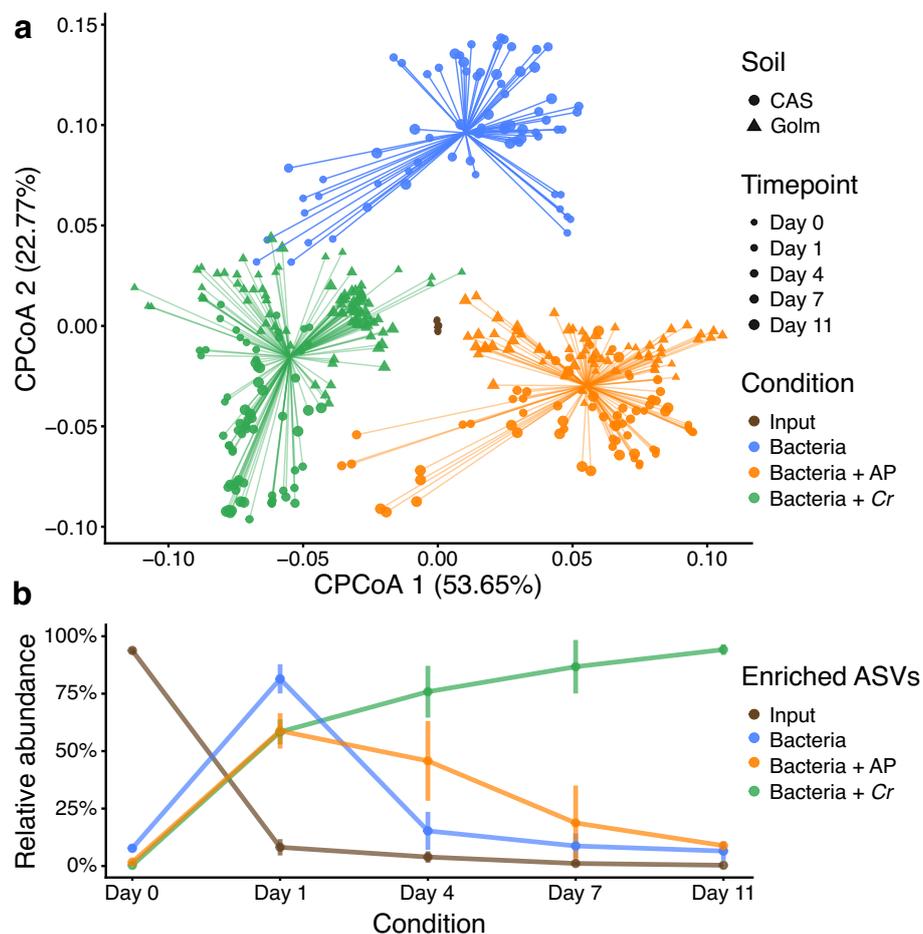
in the absence of organic carbon or supplemented with Artificial Photosynthates (AP) led to significantly differentiated bacterial communities (Figure 5.7a; 17.9% of variance;  $P < 0.001$ ). In addition, inoculation of soil-derived bacteria with heat-killed *Cr* cultures was not sufficient to recapitulate this community shift (Figure S3B in the [journal version article](#)), suggesting that the presence of live and metabolically active *Cr* is required for the establishment of synthetic phycospheres. We then tested whether larger eukaryotic microorganisms present in the soil microbial extracts, such as other unicellular algae or fungi, were also contributing to the observed changes in bacterial composition. A separate experiment, where microbial inocula were filtered through a 5  $\mu\text{m}$  pore-size membrane, showed similar bacterial community shifts compared to non-filtered extracts (Figure S3C in the [journal version article](#)). Similar to the results obtained using natural soil, the aggregated relative abundance of *Cr*-associated ASVs in the synthetic phycosphere samples increased over time, whereas ASVs enriched in the bacteria only control samples consistently decreased (Figure 5.7b). At the end of the experiment (day 11), the relative abundance of *Cr*-enriched ASVs accounted for 94% of the entire phycosphere community, in contrast to a lower contribution observed in the natural soil system (Figure 5.3b; 60% relative abundance at day 36). This pattern could be a consequence of the unintended depletion of bacteria that are not capable of metabolizing *Cr*-secreted photoassimilates in a liquid environment, and in these specific culture media. Finally, an independent mesocosm experiment using day/night light cycles showed delayed but similar patterns to those using continuous light, indicating that phycosphere community establishment may be independent of *Cr* culture synchronization (Figure S3D in the [journal version article](#)).



**Figure 5.6 Schematic description of the mesocosm experiments.** In different liquid media (minimal media or AP-containing media), *Cr* cultures were co-incubated with soil-derived bacteria. Samples were taken over time for bacterial community profiling and assessment of *Cr* growth.

Next, we aimed to control community composition in this reductionist system by establishing a *Cr*-associated bacterial culture collection following a similar approach as reported in previous studies with land plants (Lebeis *et al.*, 2015; Bai *et al.*, 2015; Garrido-Oter *et al.*, 2018; Eida *et al.*, 2018;

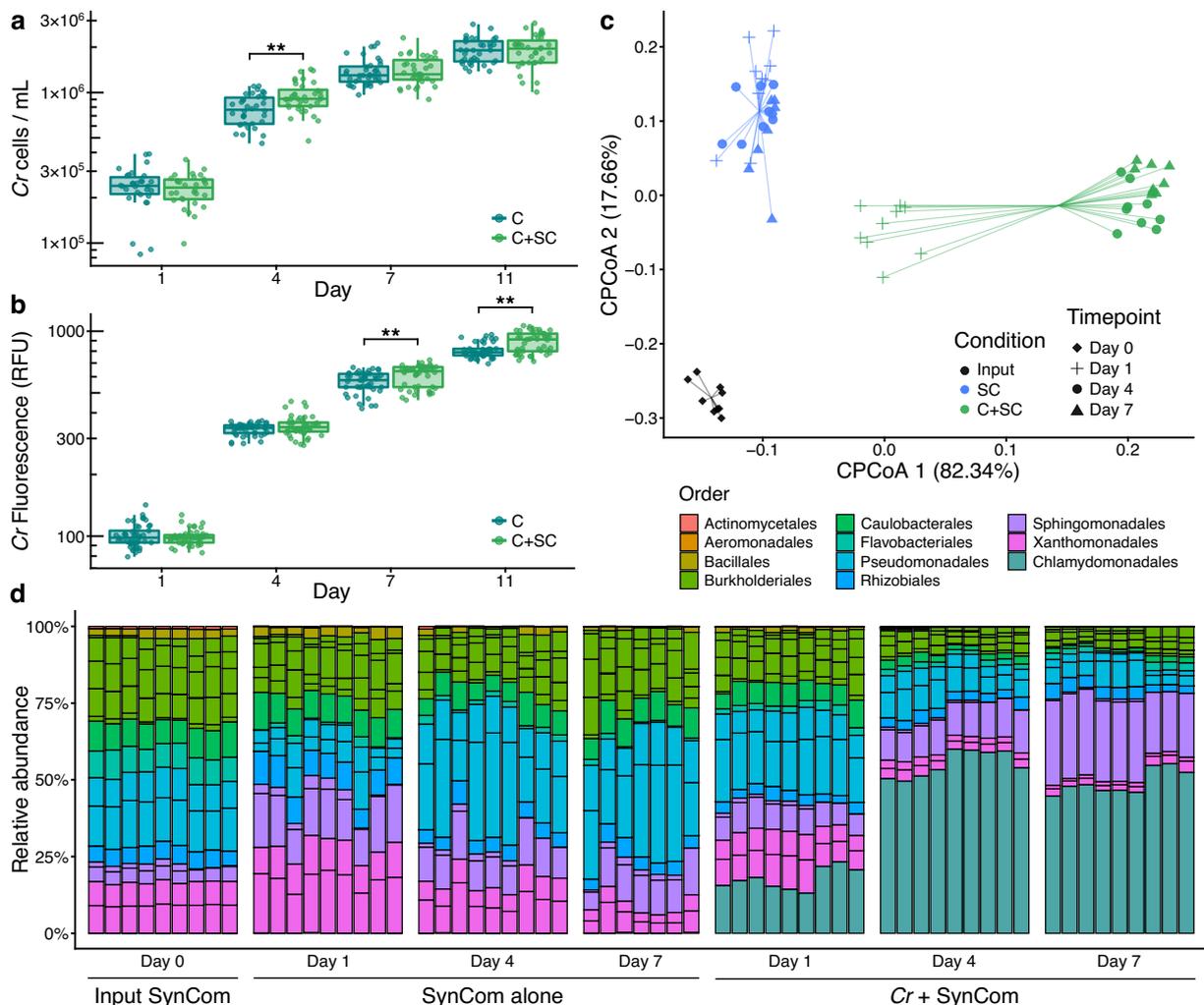
Zhang *et al.*, 2021a; Wippel *et al.*, 2021). We employed a limiting dilution approach using 7 day-old *Cr* phycospheres derived from CAS soil bacteria incubated in two minimal media (TP and B&D). The resulting sequence-indexed phycosphere bacterial library (*Cr*-IPL) contained a total of 1,645 colony forming units (CFUs), which were taxonomically characterized by 16S rRNA amplicon sequencing. Comparison of these sequencing data with the community profiling of soil phycospheres revealed that we were able to recover 62% of the most abundant bacterial OTUs found in natural communities (Figure S4A and Supplementary Data 2 in the [journal version article](#)). Recovered OTUs accounted for up to 63% of the cumulative relative abundance of the entire culture-independent community, indicating that our collection is taxonomically representative of *Cr* phycosphere microbiota. These results are comparable to the recovery rates observed in previously reported culture collections from different plant species (e.g., 57% for *A. thaliana*, Bai *et al.*, 2015; 69% for rice, Zhang *et al.*, 2019; 53% for *L. japonicus*; Wippel *et al.*, 2021).



**Figure 5.7 Mesocosm experiments recapitulate the establishment of phycosphere communities by *Cr* across soil types and growth media.** (a) PCoA analysis of Bray-Curtis dissimilarities constrained by condition (17.9% of variance;  $P < 0.001$ ) show a significant separation between start inocula (soil washes, depicted in brown), phycosphere communities (green), and soil washes incubated in minimal media (blue), or media supplemented with artificial photoassimilates (APs, depicted in orange). (b) Dynamic changes in the phycosphere community composition in terms of the aggregate relative abundances of ASVs enriched in each condition with respect to the start inocula.

To establish a core collection of phycosphere bacteria, we selected a taxonomically representative set of strains from the *Cr*-IPL covering all major taxonomic groups found in the culture-independent community profiles and subjected them to whole-genome sequencing. In total, we

sequenced the genomes of 185 bacterial isolates, classified into 42 phylogroups (97% average nucleotide identity), belonging to 5 phyla and 15 families (Supplementary Data 3 in the [journal version article](#)). Next, we performed comparative analyses of the genomes from the phycosphere core collection (*Cr*-SPHERE) with those established from soil, roots of *A. thaliana*, and roots and nodules of *L. japonicus* (*At*- and *Lj*-SPHERE) grown in the same soil (CAS). A whole-genome phylogeny of these bacterial strains showed that all major taxonomic groups that included root-derived isolates were also represented in the *Cr*-SPHERE collection, but not in the soil collection (Figure S4B in the [journal version article](#)). Importantly, the phycosphere collection also included multiple representatives of each of the six bacterial orders that were found to consistently colonize plant roots in natural environments (Figure 5.5). Next, we assessed the functional potential encoded in the genomes of the sequenced phycosphere bacteria using the KEGG orthology database as a reference (Kanehisa *et al.*, 2014). Principal coordinates analysis (PCoA) of functional distances showed that bacterial taxonomy accounted for most of the variance of the data (58.63%;  $P < 0.001$ ), compared to a much smaller impact of the host of origin of the genomes (4.22% of variance;  $P < 0.001$ ; Figure S4C in the [journal version article](#)).

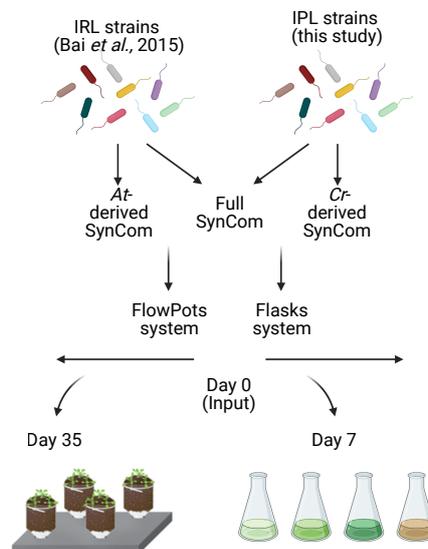


**Figure 5.8 Phycosphere reconstitution using bacterial SynComs derived from the *Cr*-SPHERE core culture collection.** *Cr* growth in the gnotobiotic system axenically ( $n = 36$ ) or in co-inoculation with the bacterial SynCom ( $n = 36$ ), measured as algal cell densities (a), and chlorophyll fluorescence (b). A Mann–Whitney test was used to assess significant differences among groups (FDR-corrected;  $P < 0.05$ ). c Strain-level beta-diversity analysis (CPCoA of Bray–Curtis dissimilarities; 40.4% of the

variance;  $P < 0.001$ ) of bacterial communities from samples obtained from a liquid-based gnotobiotic system. Samples are color-coded based on the experimental condition: input SynCom samples (black;  $n = 9$ ), synthetic phycospheres (light green;  $n = 27$ ), and SynCom only controls (blue;  $n = 25$ ). **d** Bar charts showing relative abundances of individual SynCom members across conditions and time points.

Next, we tested whether synthetic communities formed by isolates from the *Cr*-SPHERE collection could recapitulate assembly patterns of natural phycospheres under laboratory conditions. Axenic *Cr* cultures (CC1690) were inoculated with a bacterial SynCom composed of 26 strains that could be distinguished at the *16S* level and contained representative members of all major phycosphere taxonomic groups (Figure S1D and Supplementary Data 4 in the [journal version article](#)). Assessment of *Cr* growth using chlorophyll fluorescence and cell counts showed that the presence of the bacterial SynCom had no consistent beneficial or detrimental impact on *Cr* proliferation in this system (Figure 5.8a, b), similarly to what we observed in mesocosms (Figure S3A in the [journal version article](#)). Analysis of time-course amplicon profiles showed that *Cr* assembled a characteristic phycosphere community within the first 4 days of co-inoculation, which was significantly separated from both, start inocula and bacterial SynComs alone (Figure 5.8c, d). Together, these results demonstrate that we can recapitulate *Cr* assembly of distinct phycosphere communities in natural soils using culture-dependent and -independent gnotobiotic systems.

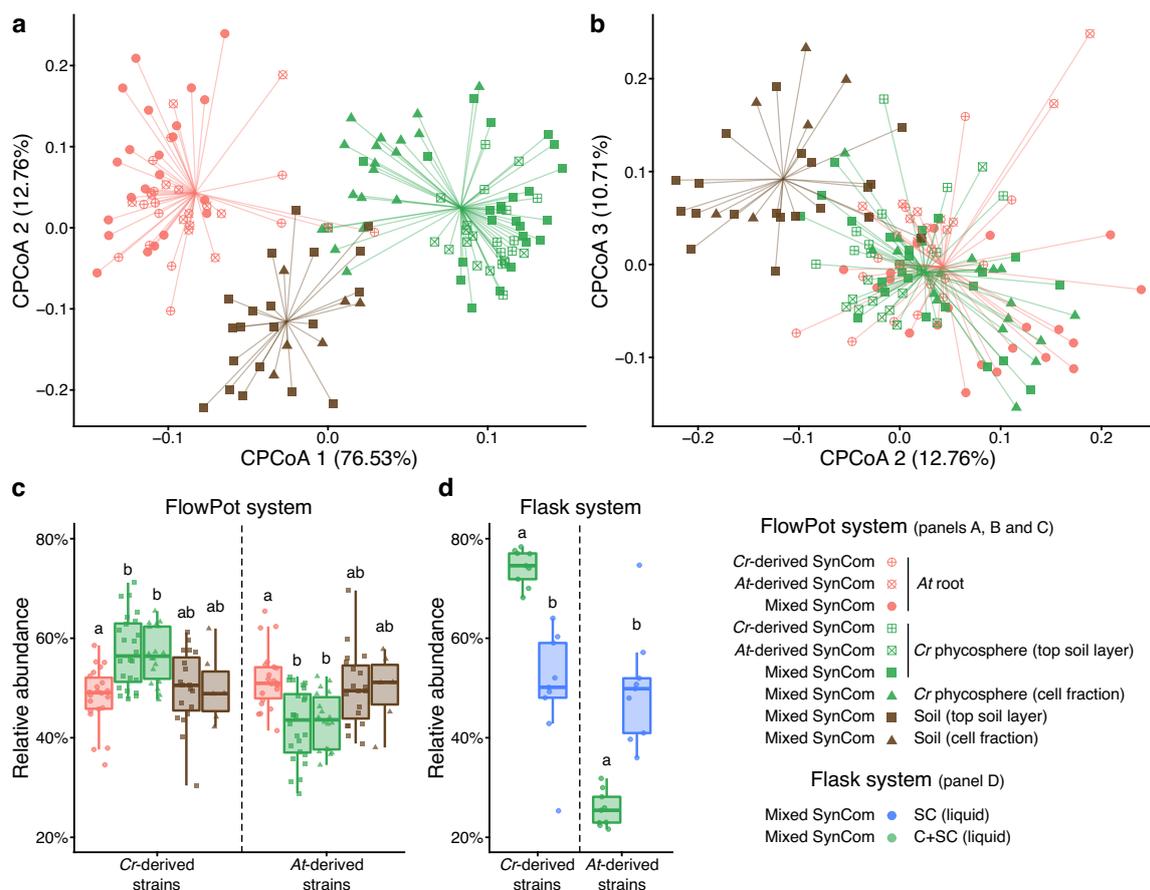
#### 5.3.4 *Cr*- and *At*-derived SynComs form taxonomically equivalent communities on either host



**Figure 5.9 Schematic description of the cross-inoculation experiment.** From the core culture *Cr*-SPHERE collection, 26 representative strains were selected, pooled together and co-incubated with *Cr* in minimal media for 7 days.

Given the similarity between phycosphere and root communities observed in natural soils (Figure 5.2), and the taxonomic and functional overlap across genomes from their corresponding core collections (Figure S4B, C in the [journal version article](#)), we hypothesized that SynComs with the same taxonomic composition would assemble into similar communities, regardless of their origin. To test this hypothesis, we used a soil-based gnotobiotic system in which we could grow *Cr* and *At* in parallel, in addition to the previously described liquid-based system. We designed taxonomically-

paired SynComs composed of strains from either the IPL (*Cr*-SPHERE) or IRL (*At*-SPHERE) bacterial culture collections. In these SynComs we included one representative strain from each bacterial family shared between the two collections ( $n=9$ ), ensuring that they could be differentiated by their *16S* rRNA sequences (Supplementary Data 4 in the [journal version article](#)). We then inoculated axenic *Cr* cultures and *At* seeds with either IPL, IRL or mixed (IPL+IRL) SynComs and allowed to colonize either host for four weeks (Figure 5.9 and Figure S1E in the [journal version article](#)). Next, we harvested the root, soil, and phycosphere fractions, measured host growth, and performed *16S* rRNA amplicon sequencing. Assessment of growth parameters (cell counts for bacteria and *Cr*, chlorophyll content for *Cr* and shoot fresh weight for *At*) showed no significant differences across SynCom treatments (Figure S5 in the [journal version article](#)).



**Figure 5.10 Root and phycosphere bacteria colonize *At* and *Cr* and assemble into taxonomically equivalent communities.** (a-b) Beta diversity analysis of soil, root, and phycosphere community profiles obtained from gnotobiotic *At* and *Cr*, inoculated with bacterial SynComs derived from *At* roots (*At*-SPHERE), *Cr* (*Cr*-SPHERE) or mixed (*At*- and *Cr*-SPHERE), grown in the FlowPot system analysis (CPCoA of Bray-Curtis dissimilarities aggregated at the family level; 16.4% of the variance;  $P<0.001$ ). Similar as in natural soils (Figure 5.2 b, c), root and phycosphere samples were significantly separated from soil and from each other in the first two axes, while overlapping in the second and third components. (c-d) Aggregated relative abundances of *At*- and *Cr*-derived strains in the mixed SynCom show ectopic colonization and signatures of host preference in a soil-derived (FlowPot, panel c), and liquid-based (flask, panel d) gnotobiotic system.

However, analysis of community profiles of the mixed SynComs showed that *Cr* and *At* assemble distinct communities that could also be clearly separated from unplanted soil (Figure 5.10a). Similar to what we observed in natural soil (Figure 5.2c), there was an overlap between phycosphere and

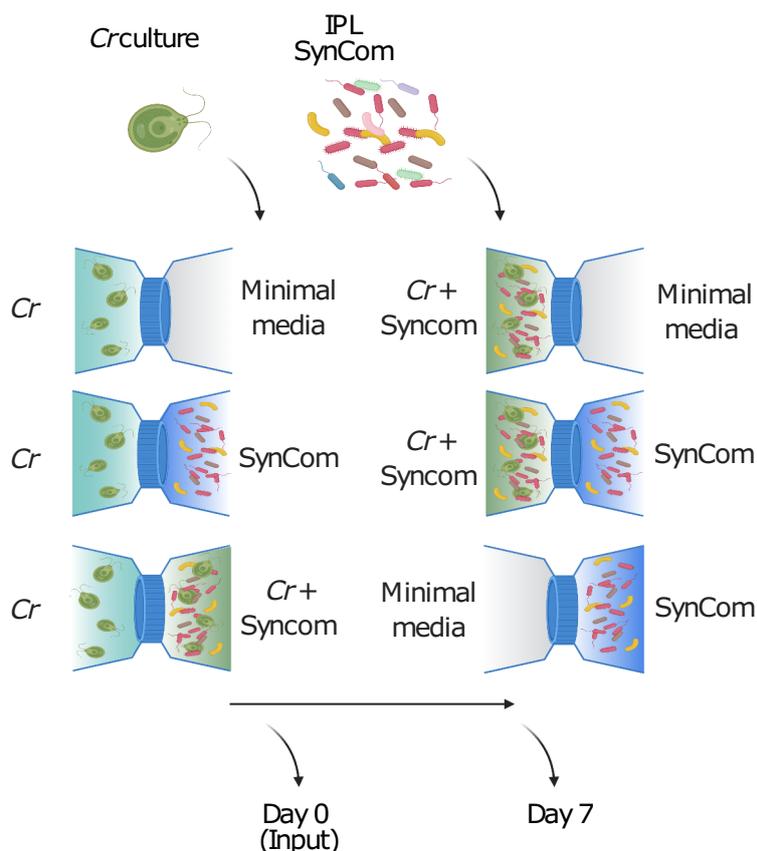
root samples, which clustered together along the second and third components (Figure 5.10b). Interestingly, analysis of community composition at the family level showed that all SynComs (*Cr*-, *At*-derived, and mixed) formed taxonomically indistinguishable root or phycosphere communities, independently of their host of origin (Figure 5.10a, b). Furthermore, analysis of aggregated relative abundances from mixed communities showed that phycosphere-derived strains could successfully colonize *At* roots (48.32% relative abundance), and root-derived strains established associations with *Cr* in both soil and liquid systems (42.94% and 25.70% relative abundance, respectively; Figure 5.10c, d). Despite this capacity for ectopic colonization, we observed significant signatures of host preference in SynComs from the two culture collections, indicated by the fact that *Cr*-derived strains reached higher aggregated relative abundances in the phycosphere compared to the root, while the opposite pattern was identified for *At*-derived bacteria (Figure 5.10c). This tendency was accentuated in the liquid system, where *Cr* bacteria outcompeted *At* strains in the presence of the algae but not when they were incubated alone (Figure 5.10d). Taken together, these results suggest the presence of conserved features in bacterial members of the *Cr* and *At* microbiota at a high taxonomic level, with signatures of host preference at the strain level.

### 5.3.5 Physical proximity is required for the assembly of phycosphere communities and promotion of *Cr* growth

Next, we sought to investigate whether the observed formation of distinct phycosphere communities is driven by the secretion of diffusible photoassimilates and to what extent physical proximity to bacteria is required to establish other forms of interactions. To test this hypothesis, we developed a gnotobiotic split co-cultivation system where synthetic phycospheres could be grown photoautotrophically (Figure 5.11). In this system, two growth chambers were connected through a 0.22  $\mu\text{m}$ -pore polyvinylidene fluoride (PVDF) membrane that allows diffusion of compounds but not passage of bacterial or algal cells. We co-cultivated axenic *Cr* cultures (C), bacterial SynComs (SC), and synthetic phycospheres (C+SC) in these split chambers containing minimal carbon-free media (TP) in multiple pair-wise combinations (Figure 5.11; Supplementary Data 4 in the [journal version article](#)).

Analysis of 16S rRNA amplicon profiles after 7 days of incubation revealed that SC and C+SC samples were distinguishable from the input bacterial SynComs (Figure 5.12a). In addition, samples clustered according to the presence of *Cr* in the same compartment, causing SC and C+SC samples to be significantly separated, independently of the community present in the neighboring chamber (Figure 5.12a, indicated by colors; 21.4% of variance;  $P < 0.001$ ). Comparison of amplicon profiles of samples taken from chambers containing C+SC further showed a significant impact of the content of the neighboring compartment in community structures (Figure 5.12b, indicated by shapes; 39.5% of variance;  $P < 0.001$ ). Interestingly, we also observed that the presence of *Cr* in the neighboring compartment was sufficient to change SC communities where the bacterial SynCom was incubated alone (Figure 5.12c; SC|C or SC|C+SC versus SC|–;  $P = 0.001$ ), possibly by secreting diffusible compounds or inducing changes in the composition of the culture medium (e.g., minerals, pH). Furthermore, SC communities where *Cr* was present in the neighboring compartment could be differentiated depending on whether *Cr* was in direct contact with bacteria or grown axenically (Figure 5.12c; SC|C versus SC|C+SC). These community shifts could be explained by competition

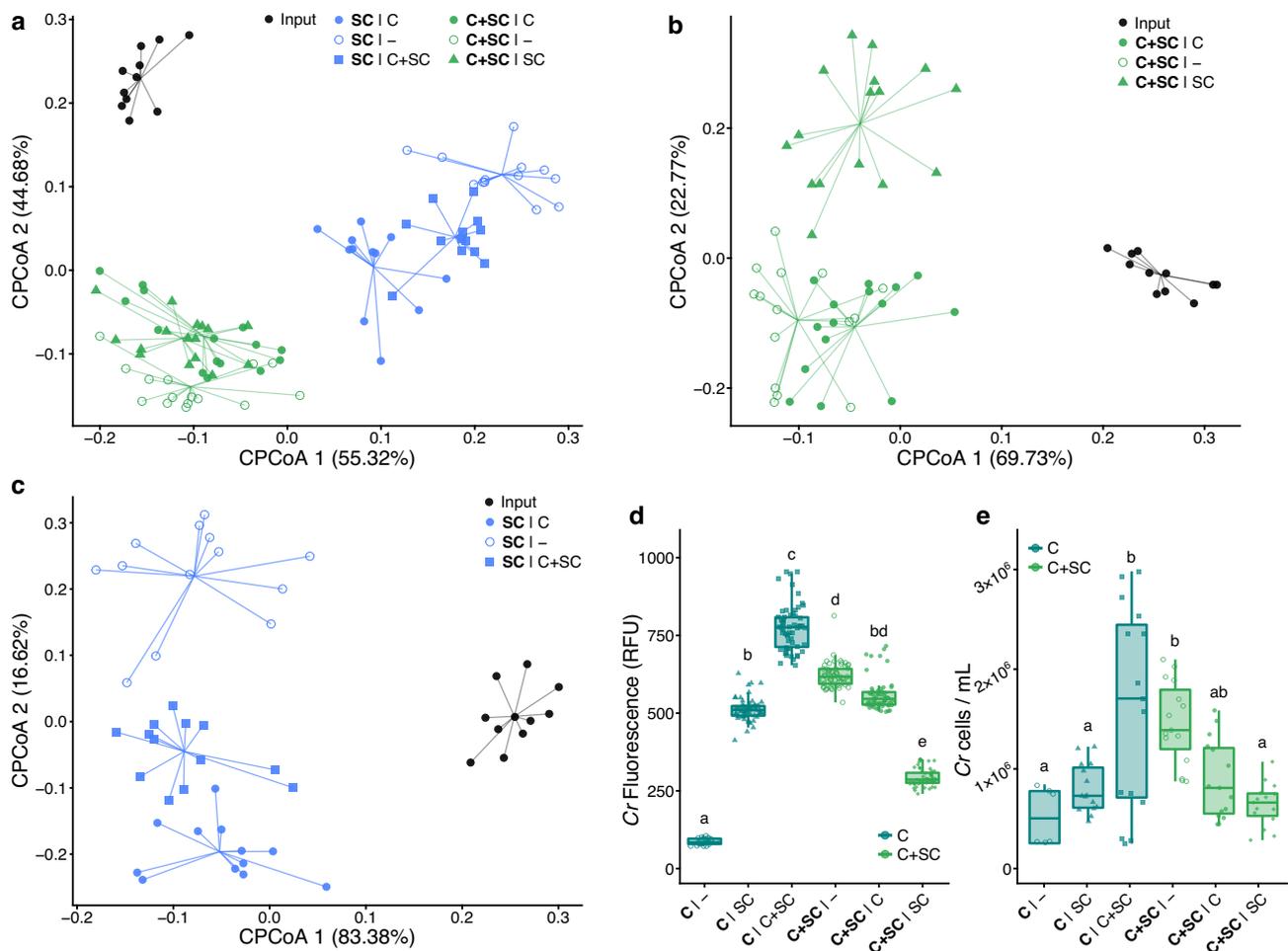
for diffusible metabolites with the neighboring compartment containing the SynCom together with the algae (C+SC), or by physiological changes in *Cr* induced by physical proximity with bacteria.



**Figure 5.11 Schematic description of the split co-cultivation system.** Co-cultivation chamber separated by a 0.22  $\mu\text{m}$  filter were used to inoculate *Cr*, the phycosphere SynCom alone or both together in different combinations. Samples were harvested after 7 days for community profiling and *Cr* growth measurements.

In parallel to bacterial community profiles, we assessed *Cr* growth by measuring chlorophyll fluorescence and algal cell counts in all vessels. We observed significant differences in the growth of axenic *Cr* cultures depending on the contents of the neighboring chamber, where the bacterial SynCom alone (CISC) had a positive impact on the microalgae compared to the control (C|–; [Figure 5.12d, e](#)). Remarkably, the presence of a synthetic phycosphere in the neighboring compartment had the strongest positive impact on axenic *Cr* cultures (C|C+SC; [Figure 5.12d, e](#)), suggesting that changes in bacterial community composition driven by physical proximity to *Cr* lead to a beneficial impact on algal growth. In addition, chlorophyll fluorescence and cell counts of synthetic phycospheres (C+SC) were higher when no other microorganisms were incubated in the neighboring chamber (C+SC|– versus C+SC|C or C+SC|SC; [Figure 5.12d, e](#)), possibly due to competition for diffusible compounds. An additional full-factorial replicate experiment using a modified version of this split co-cultivation system showed consistent results both in community structures and *Cr* growth parameters ([Figure S6 in the online article](#)), despite of a large technical variation in cell density measurements ([Figure 5.12d](#)). Together, these results indicate that physical proximity of bacteria to *Cr* is required for assembly and growth of phycosphere communities, which in turn may benefit host growth by providing metabolites and / or other compounds including carbon dioxide, which in this experimental setup is likely limiting autotrophic growth of *Cr*. Future

experimentation with synthetic phycospheres composed by SynComs designed using combinatorial approaches, coupled with metabolomic and transcriptomic profiling, will be needed to decipher the molecular and genetic mechanisms driving these interactions.



**Figure 5.12 Physical proximity to *Cr* is required for the establishment of phycosphere bacterial communities.** Beta-diversity analyses of Bray-Curtis dissimilarities of SynComs grown in a split gnotobiotic system show a significant separation of samples regarding physical proximity to *Cr* (21% of variance;  $P < 0.001$ , **a**), or the content of the neighboring vessel (39.4–39.5% of variance;  $P < 0.001$ , panels **b-c**). (**d-e**) *Cr* growth across conditions measured using relative chlorophyll fluorescence (RFU; panel **d**) and algal cell densities (panel **e**).

## 5.4 Discussion

Microscopic algae release photoassimilated carbon to the diffusible layer immediately surrounding their cells, which constitutes a niche for heterotrophic bacteria. Microbes from the surrounding environment compete for colonization of this niche and assemble into complex communities that play important roles in global carbon and nutrient fluxes. These ecological interactions have been well studied in aquatic environments, where each year approximately 20 Gt of organic carbon fixed by phytoplankton are taken up by heterotrophic bacteria (Moran *et al.*, 2016), which can account for up to 82% of all algal-derived organic matter (Horňák *et al.*, 2017). For multiple species of green algae, optimal growth in turn requires interactions with their associated phycosphere bacteria, which can provide beneficial services to their host, such as mobilization of non-soluble iron (Amin *et al.*,

2009), or exogenous biosynthesis of organic compounds such as vitamins (Croft *et al.*, 2005; Paerl *et al.*, 2017). Despite the known importance of these interactions in marine environments, the role of algae-bacterial associations in terrestrial ecosystems remains understudied. This gap in our understanding could be explained by the fact that aquatic phytoplankton are more readily noticed and more amenable to systematic study compared to edaphic microalgae. However, exploring the role of soil-borne unicellular photosynthetic organisms as hosts of complex microbial communities could expand our understanding of carbon and energy fluxes in terrestrial ecosystems.

The results from our culture-independent and gnotobiotic experiments using the ubiquitous algae *Cr*, which was originally isolated from soil (Sasso *et al.*, 2018), illustrate that green algae can recruit and sustain the growth of heterotrophic, soil-borne bacteria. This process resembles the establishment of the microbial communities that associate with the roots and rhizospheres of land plants, suggesting common organizational principles shared between chlorophytes and embryophytes. Our in-depth characterization of the *Cr* microbiota shows clear differences as well as striking similarities in the taxonomic affiliation of abundant root and phycosphere community members (Figure 5.3). Notably, these similarities are found despite biochemical differences between extracellular organic carbon compounds released by *At* roots and *Cr*, as well as by differences in cell wall composition, which in the case of the plant root mostly consists of complex polysaccharides such as cellulose, whereas in *Cr* it is primarily composed of (glyco)proteins (Harris *et al.*, 2009). Among the bacterial lineages shared between the root and phycosphere microbiota, we found groups that are known to establish intimate interactions with multicellular plants, ranging from symbiotic to pathogenic, such as Rhizobia, *Pseudomonas*, *Burkholderia*, or *Xanthomonas* (Suárez-Moreno *et al.*, 2012; Karasov *et al.*, 2018; Garrido-Oter *et al.*, 2018; Timilsina *et al.*, 2020). Meta-analyses of available data from multiple studies further confirm this pattern by revealing the presence of a set of six bacterial orders, found as abundant members not only in the root communities of all analyzed land plants, but also in the *Cr* phycosphere (Figure 5.5). These findings suggest that the capacity to associate with a wide range of photosynthetic organisms is a common trait of these core bacterial taxa, which might predate the emergence of more specialized forms of interaction with their host. This hypothesis was implicitly tested in our cross-inoculation gnotobiotic experiments, where bacterial strains originally isolated from the roots of *At* or the phycosphere of *Cr* competed for colonization of either host (Figure 5.9). The observation that *Cr*-derived strains could colonize *At* roots in a competition setup, whereas *At*-derived bacterial SynComs also populated *Cr* phycospheres (Figure 5.10c) supports the existence of shared bacterial traits for establishing general associations with photosynthetic hosts. Despite these patterns of ectopic colonization, we also detected significant signatures of host preference, illustrated by the observation that native bacterial SynComs outcompeted non-native strains in the presence of either host, but not in their absence (Figure 5.12c). These findings are in line with a recent comparative microbiota study where similar results were observed for bacterial commensals from two species of land plants (*A. thaliana* and *L. japonicus*; Wippel *et al.*, 2021). In addition, SynComs composed of strains exclusively derived from the *At*- or the *Cr*-SPHERE collections, assembled into taxonomically equivalent communities on either host, which were indistinguishable at the family level (Figure 5.10a, b). Together, our findings suggest that these bacterial taxa have in common the ability to assemble into robust communities and associate with a wide range of photosynthetic organisms, including unicellular algae and flowering plants.

Carbon is assumed to be the main factor limiting bacterial growth in soil (Demoling *et al.*, 2007). Thus, secretion of organic carbon compounds by photosynthetic organisms constitutes a strong cue for the assembly of soil-derived microbial communities (Bulgarelli *et al.*, 2013; Zhalnina *et al.*, 2018; Huang *et al.*, 2019). The observed similarities between the root and phycosphere microbiota at a high taxonomic level suggest that the release of photoassimilates acts as a first organizing principle driving the formation of these communities. This hypothesis is also supported by a recent study with marine bacterial mesocosms where community composition could be partially predicted by the addition of phytoplankton metabolites (Fu *et al.*, 2020). However, the results from our split system (Figures 5.11, 5.12), where bacterial SynComs formed distinct communities and had a beneficial effect on *Cr* growth depending on their physical proximity, indicate that the provision of diffusible carbon compounds is not sufficient to explain the observed patterns of microbial diversity. In addition, shed *Cr* cell wall components, which may not be diffusible through the 0.22  $\mu\text{m}$ -pore membrane, could be degraded by bacteria only in close proximity. The importance of proximity to the algal cells could also be a consequence of gradients in concentrations and variations in the diffusivity of different compounds, which in aquatic environments is predicted to cause highly chemotactic, copiotrophic bacterial populations to outcompete low-motility oligotrophic ones (Smriga *et al.*, 2016). Together with the algal growth data, the observed variations in SynCom structures suggest that, in addition to physical proximity, bi-directional exchange of metabolic currencies and / or molecular signals may be required for the assembly and sustained growth of a phycosphere microbiota capable of providing beneficial functions to their host. Future experimentation using this system will be aimed at elucidating core molecular and ecological principles that govern interactions between photosynthetic organisms and their microbiota.

## 5.5 Materials and methods

### 5.5.1 Culture-independent bacterial *16S* rRNA sequencing and data analysis

Total DNA was extracted from the aforementioned samples using the FastDNA™ SPIN Kit for Soil following instructions from the manufacturer (MP Biomedicals, Solon, USA). DNA samples were eluted in 50  $\mu\text{L}$  nuclease-free water and used for microbial community profiling. DNA from liquid samples was extracted using alkaline lysis (Bai *et al.*, 2015). Briefly, 12  $\mu\text{L}$  of the sample were diluted in 20  $\mu\text{L}$  of Buffer I (NaOH 25 mM, EDTA(Na) 0.2mM, pH 12), mixed by pipetting and incubated at 94 °C for 30 min. Next, 20  $\mu\text{L}$  of Buffer II (Tris-HCl 40 mM, pH 7.46) were added to the mixture and stored at -20 °C.

DNA samples were used in a two-step PCR amplification protocol. In the first step, V2–V4 (341F: CCTACGGGNGGCWGCAG; 806R: GGACTACHVGGGTWTCTAAT) or V4-V7 (799F: AACMGGATTAGATACCKG; 1192R: ACGTCATCCCCACCTTCC) of bacterial *16S* rRNA were amplified. Sequencing data from *Cr* or *At* roots grown in CAS soil in the greenhouse, along with unplanted controls, were demultiplexed according to their barcode sequence using the QIIME pipeline (Caporaso *et al.*, 2010). Afterwards, DADA2 (Callahan *et al.*, 2016) was used to process the raw sequencing reads of each sample. Unique amplicon sequencing variants (ASVs) were inferred from error-corrected reads, followed by chimera filtering, also using the DADA2 pipeline. Next, ASVs were aligned to the SILVA database (Quast *et al.*, 2012) for taxonomic assignment using the naïve Bayesian classifier implemented by DADA2. Raw reads were mapped to the

inferred ASVs to generate an abundance table, which was subsequently employed for analyses of diversity and differential abundance using the R package *vegan* (Oksanen *et al.*, 2019).

Amplicon sequencing reads from the *Cr* IPL and from the corresponding mesocosm culture-independent community profiling were quality-filtered and demultiplexed according to their two-barcode (well and plate) identifiers using custom scripts and a combination of tools included in the QIIME and USEARCH (Edgar, 2010) pipelines. Next, sequences were clustered into Operational Taxonomic Units (OTUs) with a 97% sequence identity similarity using the UPARSE algorithm, followed by identification of chimeras using UCHIME (Edgar *et al.*, 2011). Samples from wells with fewer than 100 good quality reads were removed from the data set as well as OTUs not found in a well with at least ten reads.

### 5.5.2 Isolation and genome sequencing of *Chlamydomonas*-associated bacteria

Soil bacteria associated with *Cr* after co-cultivation were isolated from mesocosm cultures using a dilution-to-extinction approach (Bai *et al.*, 2015; Wippel *et al.*, 2021). Briefly, cultures containing *Cr* and bacteria from CAS soil washes as described above were incubated in TP or B&D media. After 7 days of co-cultivation mesocosm samples were fractionated by sequential centrifugation and sonication (Kim *et al.*, 2014) prior to dilution. For fractionation, cultures were centrifuged at 400×g for 5 min to recover the supernatant. The pellet was washed with 1x TE buffer followed by sonication in a water bath at room temperature for 10 min and centrifugation at 1,000×g for 5 min. The supernatant from the first and second centrifugation were pooled together and diluted at either 1:10,000 or 1:50,000. Diluted supernatants were then distributed into 96-well microtiter plates containing 20% TSB media. After 3 weeks of incubation in the dark at room temperature, plates that showed visible bacterial growth were chosen for *16S* rRNA amplicon sequencing. For identification of the bacterial isolates, a two-step barcoded PCR protocol was used as previously described (Wippel *et al.*, 2021). Briefly, DNA extracted from the isolates was used to amplify the v5-v7 fragments of the *16S* rRNA gene by PCR using the primers 799F (AACMGGATTAGATACCCCKG) and 1192R (ACGTCATCCCCACCTTCC), followed by indexing of the PCR products using Illumina-barcoded primers. The indexed *16S* rRNA amplicons were subsequently pooled, purified, and sequenced on the Illumina MiSeq platform. Recovery rates were estimated by calculating the percentage of the top 100 most abundant OTUs found in natural communities (greenhouse experiment) that had at least one isolate in the culture collection (62%), and the total aggregated relative abundances of recovered OTUs (63%).

Next, cross-referencing of IPL sequences with mesocosm profiles allowed us to identify candidate strains for further characterization, purification, and whole-genome sequencing. Two main criteria were used for this selection: first, we aimed at obtaining maximum taxonomic coverage and selected candidates from as many taxa as possible; second, we gave priority to strains whose *16S* sequences were highly abundant in the natural communities. Whenever multiple candidates from the same phylogroup were identified, we aimed at obtaining multiple independent strains, if possible, coming from separate biological replicates to ensure they represented independent isolation events. We identified IPL samples matching OTUs found in the culture-independent root samples and selected a set of 185 representative strains maximizing taxonomic coverage for

subsequent validation and whole-genome sequencing, forming the basis of the *Cr*-SPHERE collection.

### 5.5.3 Genome assembly, annotation and comparative analyses of the *Cr*-SPHERE culture collections

Paired-end Illumina reads were first trimmed and quality-filtered using Trimmomatic (Bolger *et al.*, 2014). QC reads were assembled using the IDBA assembler (Peng *et al.*, 2012) within the A5 pipeline (Tritt *et al.*, 2012). Assembly statistics and metadata from the assembled genomes can be found in [Supplementary Data 3](#) in the [journal version article](#). Genome assemblies with either multi-modal *k*-mer and G+C content distributions or multiple cases of marker genes from diverse taxonomic groups were flagged as not originating from clonal cultures. Such assemblies were then processed using a metagenome binning approach (Pasolli *et al.*, 2019). Briefly, contigs from each of these samples were clustered using METABAT2 (Kang *et al.*, 2019) to obtain metagenome-assembled genomes (MAGs). Each MAG was analyzed to assess completeness and contamination using CheckM (Parks *et al.*, 2015). Only bins with completeness scores larger than 75% and contamination rates lower than 5% were retained and added to the collection ([Supplementary Data 3](#) in the [journal version article](#); designated MAG in the column ‘type’). Classification of the bacterial genomes into phylogroups was performed by calculating pair-wise average nucleotide identities using FastANI (Jain *et al.*, 2018) and clustering at a 97% similarity threshold. Functional annotation of the genomes was conducted using Prokka (Seemann, 2014) with a custom database based on KEGG Orthologue (KO) groups (Kanehisa *et al.*, 2014) downloaded from the KEGG FTP server in November 2019. Hits to sequences in the database were filtered using an *E*-value threshold of  $10 \times 10^{-9}$  and a minimum coverage of 80% of the length of the query sequence.

The genomes from the *Cr*-, *At*- and *Lj*-SPHERE culture collections (Bai *et al.*, 2015; Wippel *et al.*, 2021) were queried for the presence of 31 conserved, single-copy marker genes, known as AMPHORA genes (Wu & Eisen, 2008). Next, sequences of each gene were aligned using Clustal Omega (Sievers *et al.*, 2011) with default parameters. Using a concatenated alignment of each gene, we inferred a maximum likelihood phylogeny using FastTree (Price *et al.*, 2010). This tree was visualized using the Interactive Tree of Life web tool (Letunic & Bork, 2019). Finally, genomes from the three collections (*Cr*-SPHERE, *At*-SPHERE and *Lj*-SPHERE) were clustered into phylogroups, roughly corresponding to a species designation (Olm *et al.*, 2020) using FastANI (Jain *et al.*, 2018) and a threshold of average nucleotide identity at the whole genome level of at least 97%. Functional comparison among the genomes from the *Cr*-, *Lj*- and *At*-SPHERE collections was performed by comparing their annotations. KO groups were gathered from the genome annotations and aggregated into a single table. Lastly, functional distances between genomes based on Pearson correlations were used for principal coordinate analysis using the *cmdscale* function in R.

### 5.5.4 Preparation of SynCom inocula and analysis of amplicon sequencing data

Bacterial cultures from the strains selected for the different SynComs ([Supplementary Data 4](#) in the [journal version article](#)) were started from glycerol stocks which were used to streak agar plates containing TSA 50% media. Plates were cultured at 25 °C for five days and later used to inoculate culture tubes with 1 mL of 50% TSB media. The tubes were incubated for six days at 25 °C and 180

RPM. After 6 days, the cultures were washed three times by centrifugation at 4,000 ×g for 5 min, the supernatant discarded, and the pellet resuspended into 2 mL of TP or TP10 media. The washed cultures were further incubated with shaking at 25 °C for an additional day. Bacterial concentration in washed cultures was determined by measuring OD<sub>600</sub> and, subsequently pooled in equal ratios. Cell counts of the pooled SynCom were measured using the Multisizer 4e and adjusted to 10<sup>6</sup>, to inoculate together with 10<sup>4</sup> cells of *Cr* (prepared as described above) in 50 mL of TP10 in 200-mL flasks. These flasks were inoculated in triplicate and three biological replicates were prepared for both bacteria and *Cr* start inocula. As controls, *Cr*-only cultures and SynCom-only cultures were incubated in parallel, and samples taken at 0, 1, 4, 7 for community profiling, and at 0, 4, 7, 14 days for *Cr* cell counts.

Sequencing data from SynCom experiments was pre-processed similarly as natural community 16S rRNA data. Quality-filtered, merged paired-end reads were then aligned to a reference set of sequences extracted from the whole-genome assemblies of every strain included in a given SynCom, using Rbec (Zhang *et al.*, 2021b). We then checked that the fraction of unmapped reads did not significantly differ between compartment, experiment or host species. Next, we generated a count table that was employed for downstream analyses of diversity with the R package vegan. Finally, we visualized amplicon data from all experimental systems using the *ggplot2* R package (Wickham, 2016).

#### 5.5.5 Multi-species microbiota reconstitution experiments

The gnotobiotic FlowPot (Kremer *et al.*, 2021) system was used to grow *Cr* or *A. thaliana* plants with and without bacterial SynComs. This system allows for even inoculation of each FlowPot with microbes by flushing of the pots with the help of a syringe attached to the bottom opening. After FlowPot assemblage, sterilization and microbial inoculation sterilized seeds were placed on the matrix (peat and vermiculite, 2:1 ratio), and pots were incubated under short-day conditions (10 hours light, 21°C; 14 hours dark, 19°C), standing in customized plastic racks in sterile ‘TP1600+TPD1200’ plastic boxes with filter lids (SacO2, Deinze, Belgium). For SynCom preparation, bacterial strains from either *Cr*- or *At*-SPHERE were grown separately in liquid culture for 2-5 days in 50% TSB media and then centrifuged at 4,000 ×g for 10 min and re-suspended in 10 mM MgCl<sub>2</sub> to remove residual media and bacteria-derived metabolites. Equivalent ratios of each strain, determined by optical density (OD<sub>600</sub>) were combined to yield the desired SynComs (Table S1 in the [journal version article](#)). An aliquot of the SynComs as reference samples for the experiment microbial inputs were stored at –80°C for further processing. SynCom bacterial cells (10<sup>7</sup>) were added to either 50 mL of TP10 or ½ MS (Duchefa Biochemie, Haarlem, Netherlands), which were then inoculated into the FlowPots using a 60 mL syringe. For *Cr*-inoculated pots, 10<sup>5</sup> of washed *Cr* cells were added to the 50mL of media with or without microbes to be inoculated into the FlowPots.

*Chlamydomonas* or *Arabidopsis* FlowPots were grown side-by-side in gnotobiotic boxes, with six pots in total per box. This experiment was repeated in three independent biological replicates. After five weeks of growth, roots were harvested and cleaned thoroughly from attached soil using sterile water and forceps. Surface of *Chlamydomonas* pots were used as phycosphere samples (cells were harvested from visibly green surface areas, top soil samples). In addition, to remove any possible

background effect from carry-over soil particles, the surface-harvested samples were washed in sterile TE supplemented with 0.1% of Triton X-100 by manually shaking in 2-mL Eppendorf tubes. Then, the tubes rested for a few minutes and the supernatant was used as “cell fraction” samples. Finally, soil from unplanted pots were collected as soil samples and treated similarly as *Chlamydomonas*-inoculated pots for microbial community comparison. All phycosphere, root (comprising both the epiphytic, and endophytic compartments), and soil (soil from unplanted pots) samples were transferred to Lysing Matrix E tubes (MP Biomedicals, Solon, USA), frozen in liquid nitrogen, and stored at  $-80^{\circ}\text{C}$  for further processing. DNA was isolated from those samples using the MP Biomedicals FastDNA™ Spin Kit for Soil, and from the input SynCom by alkaline lysis, and subjected to bacterial community profiling.

To ensure sufficient surface for phycosphere harvesting, we set up an additional experiment based on sterile peat without FlowPots. Experiments with the mixed SynCom of *Cr*- and *At*-SPHERE strains were conducted using sterile ‘TP750+TPD750’ plastic boxes (SacO2, Deinze, Belgium). Sterile soil and vermiculite were mixed in a 2:1 ratio and added to each box. Next, the boxes were inoculated by adding 95 mL of TP10 or  $\frac{1}{2}$  MS, for the *Chlamydomonas* or *Arabidopsis* boxes respectively, containing  $2 \times 10^7$  bacterial cells.

#### 5.5.6 Split co-cultivation system

Co-cultivation devices were built by adapting 150 mL Stericup-GV filtration devices (Merck Millipore, Darmstadt, Germany) harboring a  $0.22 \mu\text{m}$  filter membrane (Alvarez *et al.*, 2018). Each co-cultivation device was assembled inside a clean hood 150 mL and 100 mL of TP10 were added into the big and small chamber of the filtration device, respectively. Chambers were inoculated at different cell concentrations depending on the content of the chamber (Figure S1F in the [journal version article](#)). The concentrations used were  $10^5$  and  $10^7$  cells/mL for *Chlamydomonas* and SynCom respectively. For the C+SC condition, the inoculum concentration was the same as for individual content chambers. After inoculation the devices were transferred to a shaking platform and incubated under the same conditions used for *Cr* liquid cultures described above. Four samples per chamber were harvested for DNA extraction, fluorescence, and cell growth at the start of the incubation and 7 days after inoculation. These experiments were repeated in three independent biological replicates, containing one technical replicate each.

Additionally, a full-factorial replicate of the experiment was carried out using a custom-made co-cultivation device (Cat. #0250 045 25, WLB Laborbedarf, Möckmühl, Germany). Briefly, two 250 mL borosilicate glass bottles (Figure S1F in the [journal version article](#)) were modified by adding on the sidewall of each bottle a glass neck with a NW25 flange. The flange holds a disposable  $0.22 \mu\text{m}$ -pore PVDF Durapore filtration membrane (Merck Millipore, Darmstadt, Germany) and is kept in place by an adjustable metal clamp. In this device, each bottle holds 150 mL of TP10 and the initial cell concentrations were the same as the ones used in the previously described co-cultivation device. Similar to the Stericup system, four samples per chamber were harvested for DNA extraction. Chlorophyll fluorescence and cell growth measurements were collected at the start of the incubation and 7 days after inoculation. These experiments were repeated in three independent biological replicates, containing one technical replicate each.

## 5.6 Data and code availability

Raw sequencing will be deposited into the European Nucleotide Archive (ENA) under the accession number PRJEB43117. The scripts used for the computational analyses described in this study are available at <http://www.github.com/garridoo/crsphere>, to ensure replicability and reproducibility of these results.

## 5.7 Author contributions

### Duran and Flores-Urbe et al., 2022

P.D., J.F.-U., K.W., B.M., M.M., and R.G.-O. designed the experiments. P.D. and J.F.-U. conducted the greenhouse and soil native algae community experiments. P.D. and K.W. performed the mesocosm experiments. P.D., J.F.-U., and K.W. established the IPL bacterial library and characterized the *Cr*-SPHERE core culture collection. B.M. and M.M. isolated the environmental algal strains. P.D. conducted the experiment with the environmental algal strains. P.D. and J.F.-U. performed the synthetic community experiments. P.Z., J.F.-U., and R.G.-O. analyzed whole-genome sequencing data. P.D., J.F.-U., P.Z., R.G., and R.G.-O. analyzed the amplicon data. P.D., J.F.-U., K.W., B.M., M.M., and R.G.-O interpreted the results. P.D., J.F.-U., and R.G.-O wrote the paper.



# Chapter 6 Maize field study reveals covaried microbiota and metabolic changes in roots over plant growth

## 6.1 Abstract

Plant roots are colonized by microbial multi-kingdom communities from the surrounding soil, termed root microbiota. Despite their importance for plant growth, the relationship between soil management, the root microbiota, and plant performance remains unknown. We characterized here the maize root-associated bacterial, fungal and oomycetal communities during the vegetative and reproductive growth stages of four maize inbred lines and the phosphate transporter mutant *pht1;6*. These plants were grown in two long-term experimental fields under four contrasting soil managements, including phosphate-deficient and -sufficient conditions. We showed that the maize root-associated microbiota is influenced by soil management and changes during host growth stages. We identified stable bacterial and fungal root-associated taxa that persist through the host life-cycle. These taxa were accompanied by dynamic members that covary with changes in root metabolites. Unexpectedly, we observed an inverse stable-to-dynamic ratio between root-associated bacterial and fungal communities. We also found a host footprint on the soil biota, characterized by a convergence between soil, rhizosphere and root bacterial communities during reproductive maize growth. Our study reveals the spatio-temporal dynamic of maize root-associated microbiota and suggests that the fungal assemblage is less responsive to changes in root metabolites than the bacterial community.

## 6.2 Introduction

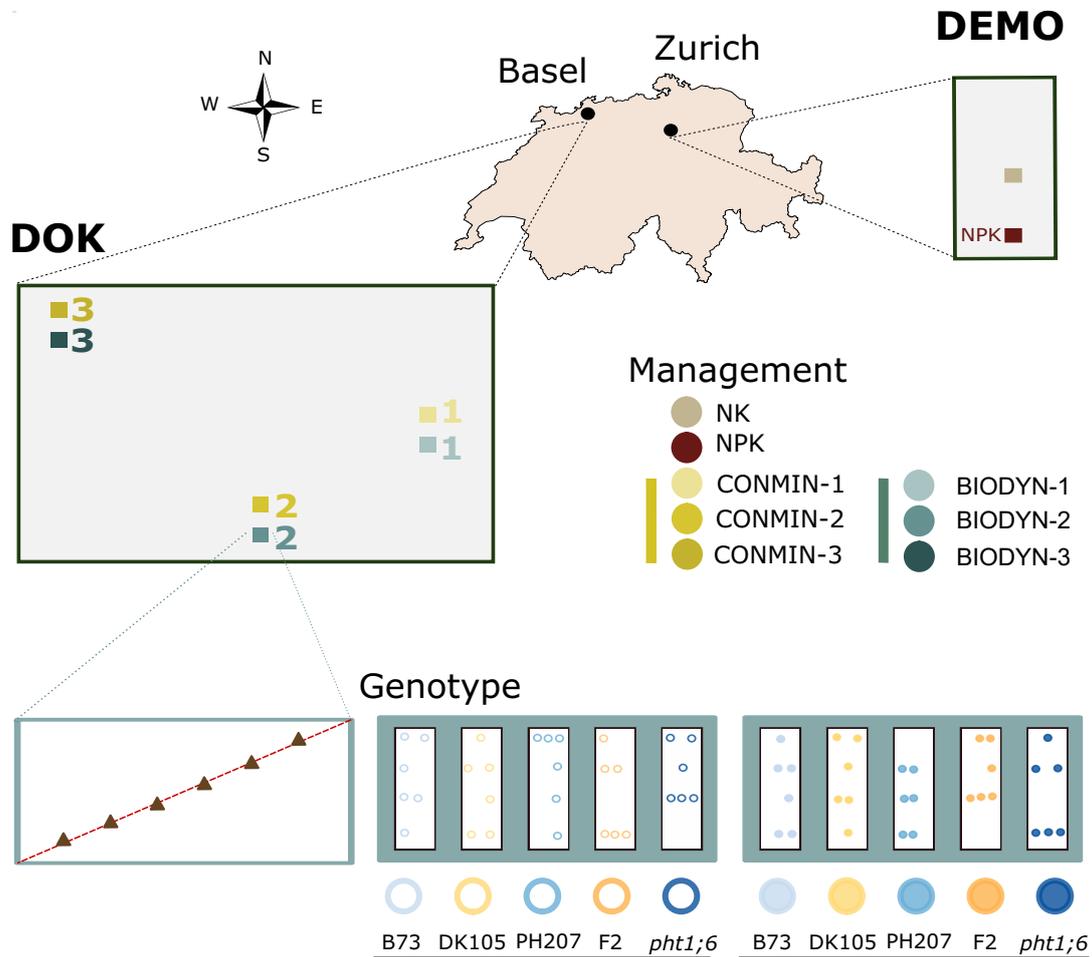
In nature, plant roots are colonized by diverse soil-dwelling microbes, which are collectively known as the root microbiota, and this microbial multi-kingdom community promotes plant growth and health (Berendsen *et al.*, 2012; Bulgarelli *et al.*, 2013; Hacquard *et al.*, 2015; Hassani *et al.*, 2018). Numerous microbiota members assist with nutrient mobilization of macro and micro-nutrients from the soil for host nutrition. For instance, in orthophosphate (P)-limiting soils, P-solubilizing rhizosphere bacteria can elevate the amount of bioavailable P, and long-distance transport of soluble P is mediated by hyphae of symbiotic arbuscular mycorrhizal fungi (AMF) or certain fungal root endophytes to the host (Bucher, 2007; Pii *et al.*, 2015; Hiruma *et al.*, 2016). In iron (Fe)-limiting calcareous soil, the bacterial root microbiota serves a critical role in mobilizing insoluble ferric iron for plant Fe nutrition (Harbort *et al.*, 2020). Lastly, interactions between microbes from different kingdoms are important for plant survival, as shown by Duran *et al.* (Durán *et al.*, 2018) who described the protective function of the root-associated bacterial community against a taxonomically broad range of soil-dwelling and harmful filamentous eukaryotes in *Arabidopsis thaliana*. Although bacteria and fungi are considered the main microbial kingdoms of the root microbiota, roots are also colonized by the kingdom Stramenopiles (formerly Oomycota), with common oomycete phytopathogenic members (Kamoun *et al.*, 2015) and some strains belonging to *Pythium spp.* that are known to promote plant growth by preventing biotic stress (Benhamou *et al.*, 2012).

Plant breeding allows the selection of the best traits that favor optimal plant fitness under high fertilization conditions, but the innate capacity of inbred lines to establish beneficial plant-microbe associations has not received as much attention. The choice of long-term soil management can indirectly affect plant growth by modifying the diversity and connectivity of root-associated microbial communities (Banerjee *et al.*, 2019; Chowdhury *et al.*, 2019; Schmidt *et al.*, 2019). Organic fertilization appears to contribute to the maintenance of more abundant and diverse soil microbial communities (Fließbach *et al.*, 2007; Francioli *et al.*, 2016). However, little is known about the relationship of soil-dwelling microbial communities with soil nutrient status, such as how limiting soil nutrients influence the capacity of plant roots to be colonized by microbes that boost plant growth (Castrillo *et al.*, 2017; Harbort *et al.*, 2020).

Though several studies have linked temporal changes in the root-associated microbiota to plant development in both field and controlled environment experiments, not much is known about how maize-associated microbiota are altered between plant growth stages (Shi *et al.*, 2015; Edwards *et al.*, 2015, 2018; Walters *et al.*, 2018). As one of the most widely cultivated crops in the world, maize has been used as a model to characterize plant-microbe interactions in agricultural contexts, specifically to assess the effects of plant genotype and age, biogeography, and soil management on microbial community assembly (Peiffer *et al.*, 2013; Walters *et al.*, 2018; Schmidt *et al.*, 2019; Wagner *et al.*, 2020). Additionally, the architecture of the maize root system is modified over time (Tai *et al.*, 2016). Fungal communities were shown to vary for axial and lateral root types and aerial roots of a particular maize landrace secrete a carbohydrate-rich mucilage enriched in diazotrophic bacteria (Yu *et al.*, 2018; Van Deynze *et al.*, 2018). In the crop root system, the different types of roots function in dissimilar manners regarding nutrient and water foraging and uptake (Rogers & Benfey, 2015; Tai *et al.*, 2016), which can affect microbial root colonization, as shown for AMF (Gutjahr *et al.*, 2009). However, the spatio-temporal variability of the root microbiota of maize has not been extensively explored, and few studies have considered the diverse multi-kingdom microbial communities associated with maize or the roots of other species, in their entirety (Durán *et al.*, 2018; Brisson *et al.*, 2019; Schmidt *et al.*, 2019; Wagner *et al.*, 2020).

Rhizodeposits, including soluble root exudates, represent a major source of organic carbon for soil-dwelling bacteria surrounding roots (Jones *et al.*, 2009). Root exudates were shown to change consistently during the early vegetative and senescence developmental stages of the annual grass *Avena barbata* (Zhalnina *et al.*, 2018). Specifically, chemical succession in *A. barbata* exudates interacts with microbial metabolite substrate preferences in heterotrophic bacteria isolated from soil, in which *Avena* dominates (Zhalnina *et al.*, 2018). The age-correlated *A. barbata* exudation and microbial substrate uptake explain part of the bacterial community assembly and dynamics in the rhizosphere of this annual plant. However, in perennial *Arabidopsis thaliana* the comparison of vegetative and reproductive stage of non-flowering wild type and a naturally occurring and perpetually flowering mutant did not show any impact of flowering time on root bacterial community profiles, but a clear effect of soil residence time (Dombrowski *et al.*, 2017). This shows that the genetically determined program of flowering time (transition to reproductive growth) can be uncoupled from dynamic changes in the root microbiota. In annual *A. thaliana*, root samples from young and fruiting plants clustered together, indicating that vegetative and reproductive growth phases do not have a major effect on overall bacterial community composition (Lundberg *et al.*, 2012). By

contrast, perennial *Boechera stricta* was found to alter its root-associated bacterial community over the growing season (Wagner *et al.*, 2016). Moreover, similar results were observed in field grown rice, presenting a shift in bacterial and archaeal root microbiota, over plant growth from juvenile to adult plant stages (Edwards *et al.*, 2018). Collectively, these findings suggest plant species-specific variation in root microbiota dynamics that is not necessarily linked to the genetically programmed developmental growth stages of the host.



**Figure 6.1 Experimental design of field-grown maize microbiota survey.** Five different maize genotypes including four inbred lines (B73, DK105, PH207, F2) and one phosphate transporter mutant line (*pht1;6*), compromised in phosphate transport from fungi to the plant, were planted in two long-term experimental fields, DEMO (fertilization DEMONstration experiment, Reckenholz) and DOK (Dynamic, organic and conventional managements, Therwil), in Switzerland. Two soil managements per field were tested. NK and NPK soil fertilization were practiced in DEMO (one plot per management) and CONMIN and BIODYN were used in DOK (three plots per management). Six soil samples per plot were collected before sowing. Later, at the vegetative and reproductive stages of plant growth, six plants per genotype were harvested. For each plant, rhizosphere and root samples were collected. Additional bulk soil was sampled from each planted plot ( $n = 3$ ) at both stages.

The composition of root metabolites, which include various compound classes such as soluble carbohydrates, amino acids, fatty acids, organic acids and specialized metabolites (Bais *et al.*, 2006), change during the life-cycle of flowering plants (Walker *et al.*, 2003). In maize, several studies have shown that benzoxazinoids (BXs), specialized metabolites that display insecticidal, antimicrobial, and allelopathic activities and are predominantly secreted by roots at an early growth

stage, influence the root-associated microbiota by inhibiting colonization by specific microbial taxa and plant pathogens (Hu *et al.*, 2018; Kudjordjie *et al.*, 2019). The BX breakdown product 6-methoxy-benzoxazolin-2-one (MBOA), which accumulates in the soil, acts indirectly by altering root-associated microbiota and is necessary and sufficient to promote maize tolerance to herbivore attack in the next plant generation (Hu *et al.*, 2018). In addition, BXs regulate global maize root metabolism and influence the root microbiota via BX-dependent metabolites, especially flavonoids (Cotton *et al.*, 2019).

Here, we examined how soil-dwelling and root-associated microbial communities from four maize inbred lines (B73, PH207, DK105, F2) respond to contrasting organic (BIODYN) and mineral (NK, NPK, CONMIN) soil managements, in long-term fertilization fields at two different locations (Figure 6.1). In addition, considering the major role of soil orthophosphate availability on crop growth and the positive interaction between maize and AMF, we included the P transporter mutant *pht1;6* in our study. We surveyed 1,104 samples (of soil, rhizosphere and root compartments, at three different time points) by amplicon sequencing of bacterial, fungal and oomycete marker genes to reveal the dynamics of soil and root microbiota. Comparison of maize samples collected at vegetative and reproductive growth stages shows that the root-associated microbiota is influenced by soil management and is dynamic over the host's life-cycle. Moreover, we found a convergence of bacterial soil, rhizosphere and root communities at the phylum level over the growing season. By performing parallel profiling of root lipids, amino acids, soluble carbohydrates, and the root ionome, we show that root metabolites covary with root-associated microbial communities. Comparison of wild type and *pht1;6* mutant plants revealed a potential plant growth stage-specific link between AMF symbiosis, root lipid status and soil P availability. We discuss the potential interplay between the root microbiota, root metabolites, and soil management over the life cycle of field-grown maize, and highlight how the dynamics of plant-microbe associations could affect plant physiology and fitness depending on soil nutrient availability.

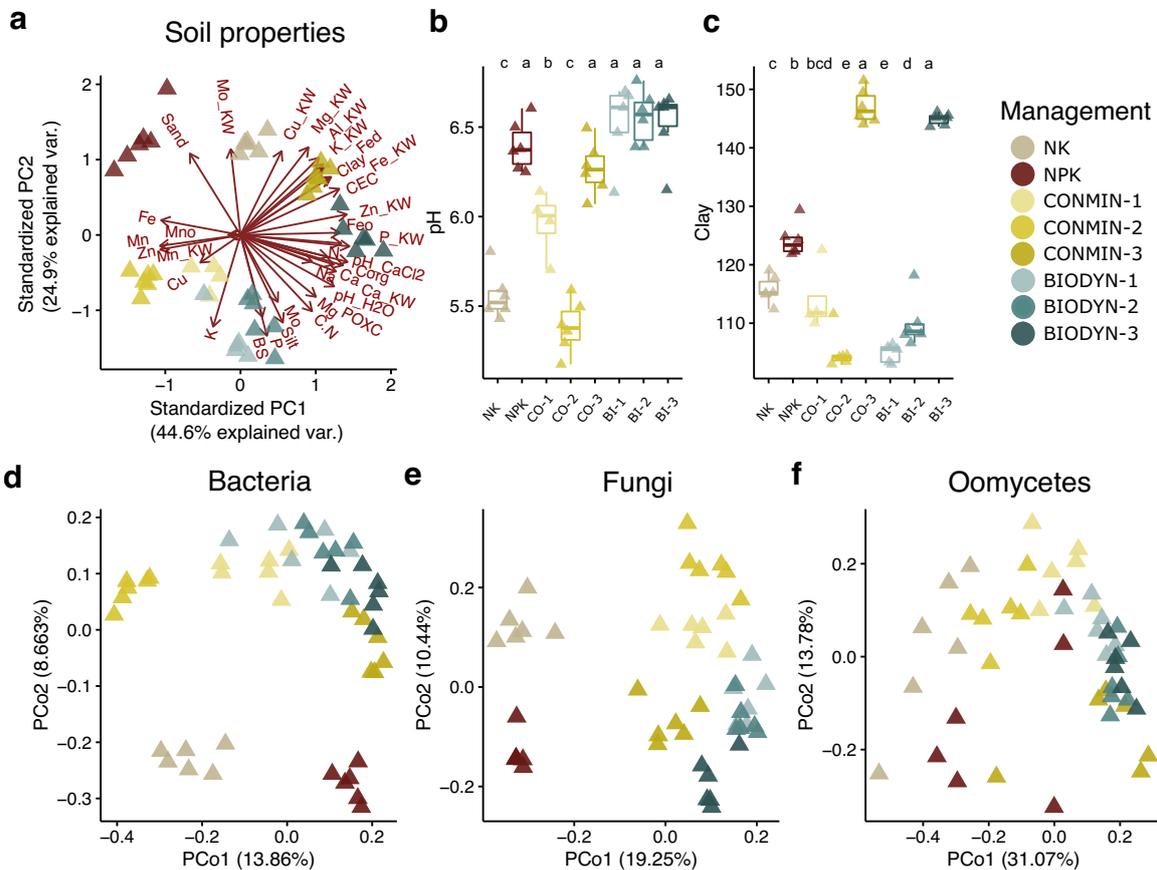
## 6.3 Results

### 6.3.1 Dynamics of the soil and root-associated microbiota diversity

We assessed the community composition of the three main microbial kingdoms – bacteria (B), fungi (F) and oomycetes (O) – in the respective soil samples: NK and NPK in DEMO field, CONMIN (Conventional solely mineral fertilized) and BIODYN (Biodynamic mixed) in DOK field (Figure 6.1). These two fields are geographically separated by approx. 100 km. First, we characterized the chemical properties of unplanted soil before sowing, and its soil-dwelling microbial communities (Figure 6.2).

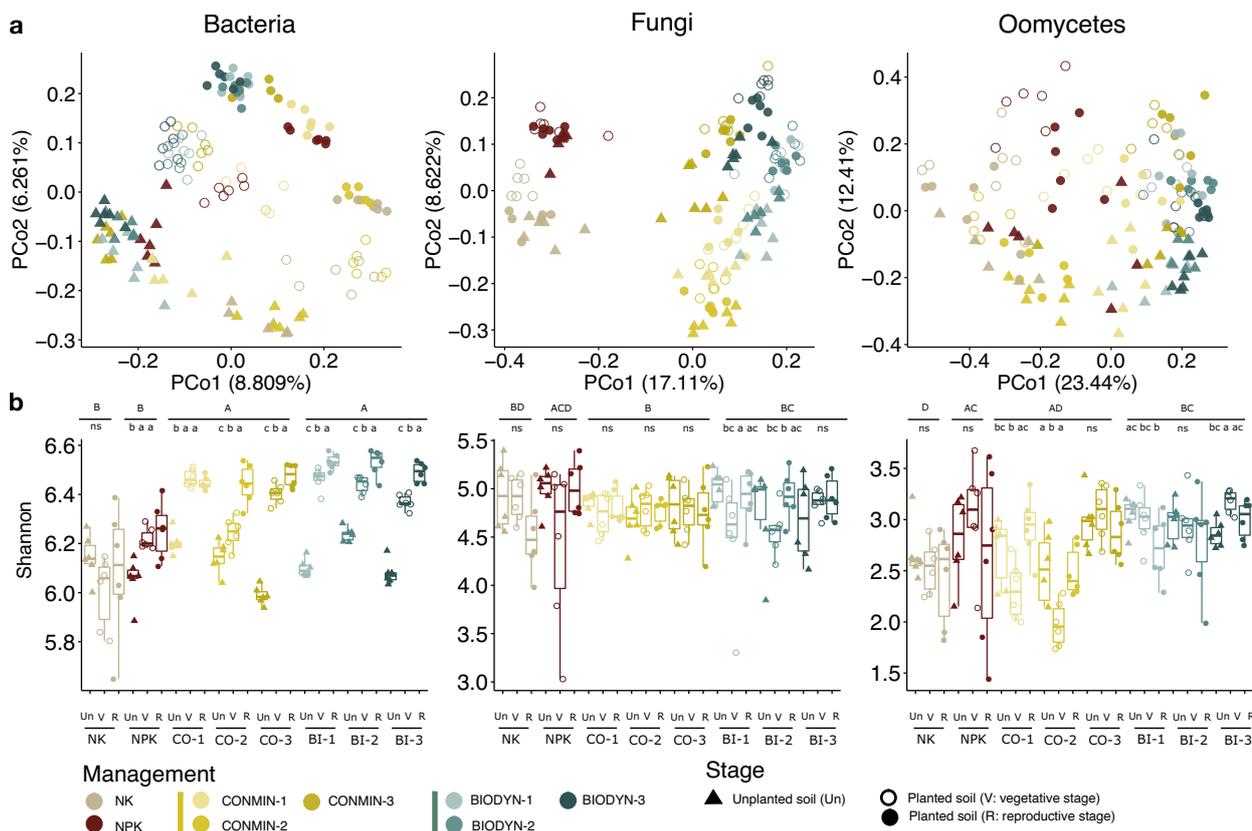
We observed a differentiation of soil physicochemical properties between soil samples from DEMO and DOK fields (Figure 6.2a), of which the former soil type is a Gleyic Cambisol and the latter is a Haplic Luvisol on alluvial loess according to FAO (WRB, 2015). Principal Coordinate Analysis (PCoA) of Bray-Curtis dissimilarities between samples based on ASVs revealed that samples clustered together by field and plot (Figure 6.2d, e, f). The bacterial biota of soil was more affected by variability in soil physicochemical features between plots (PCo1, 13.86%) than variations between fields (PCo2, 8.66%). Filamentous eukaryotic community structure, meanwhile, was more strongly determined by field location (PCo1, 19.25%). Except for these well-known factors effect

on the microbial kingdoms, an impact of soil properties on unplanted bulk soil biota was observed, where bacterial communities of plots with the lowest pH values (5.97 ± 0.13 in NK and 5.91 ± 0.16 in CONMIN-2, [Figure 6.2b](#)) were clustered together ([Figure 6.2d](#)), and microbial communities from the plots with the highest clay content ( $P < 0.05$ , CONMIN-3 and BIODYN-3, [Figure 6.2c](#)) exhibited lower dissimilarities in all three kingdoms ([Figure 6.2d, e, f](#)).



**Figure 6.2 Soil properties and microbial diversity of unplanted soil in long-term experimental fields.** (a) PCA of Euclidean distance of soil properties (34 properties in 48 samples were analyzed). Variability of pH (b) and clay content (c) of the soil between plots in each field ( $n = 48$ , Wilcoxon test was used for statistical analysis with FDR correction,  $P < 0.05$ ). Beta-diversity (Bray-Curtis dissimilarity) of bacterial (d), fungal (e), and oomycetal (f) communities in unplanted soil collected before sowing ( $n = 48$ ).

In planted soil, besides the field, management and plot effects, microbial communities were also clustered by the corresponding host growth stage when sampling ([Figure 6.3a](#)). Although plant growth phase influenced the diversity of soil microbes from all three kingdoms with relatively high explained variance ratios (B: 10.10%, F: 7.28% and O: 10.71%;  $P < 0.001$ , PERMANOVA), this effect was more clearly observed in bacterial communities ([Figure 6.3a](#)). Furthermore, the separation between fields was observed for fungi only. For oomycetes, the community shift was largely driven by soil management. Moreover, soil bacterial communities exhibited varying alpha-diversities, associated to soil chemistry and plant growth stage ([Figure 6.3b](#)). The Shannon index was significantly higher in the DOK field (CONMIN and BIODYN) compared to the DEMO field (NK and NPK) for bacteria ( $P < 0.001$ ). In addition, regardless of soil management, we found a significant increase in bacterial diversity in the soil over time (with the exception of the NK condition), while alpha-diversity for fungi and oomycetes remained relatively stable.

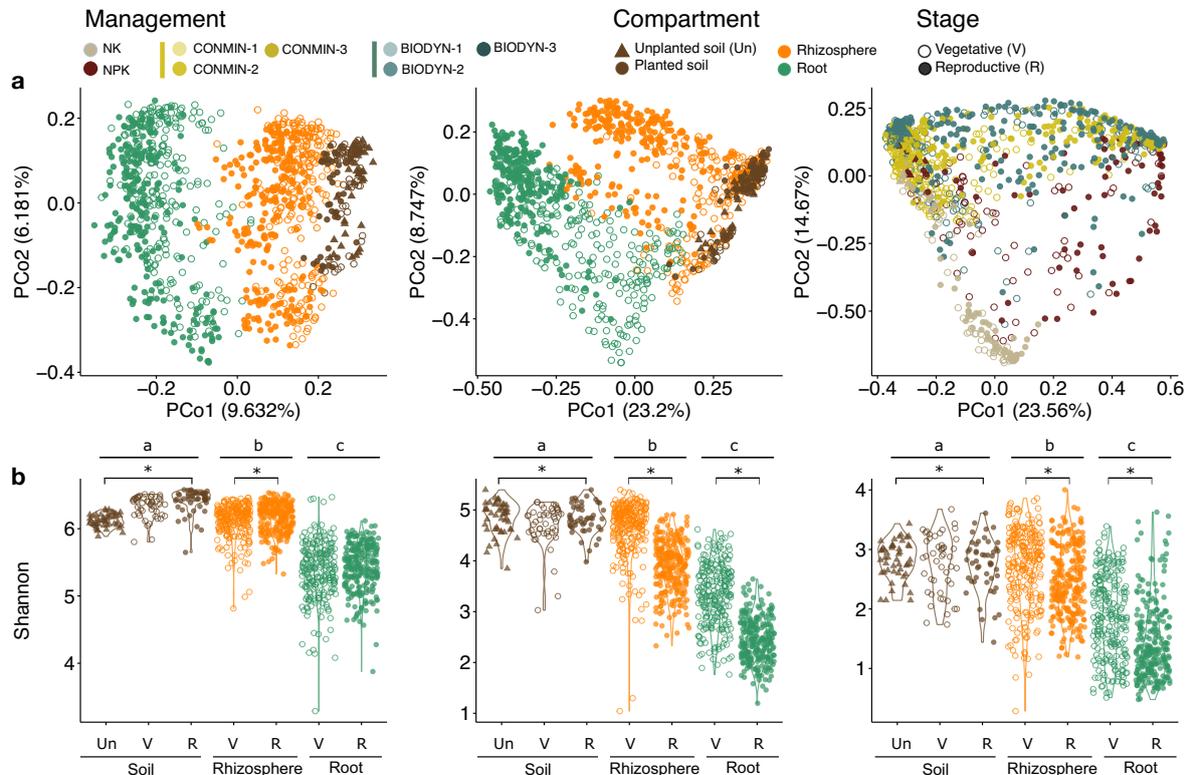


**Figure 6.3 Soil physicochemical properties, heterogeneity, and plant growth phase drive shifts in microbial soil biota composition.** (a) PCoA indicates that soil bacterial community shifts are driven by plant growth phase and intra-field soil heterogeneity, whereas fungal and oomycetal communities are affected by biogeography and soil management. (b) Soil microbial alpha-diversity (Shannon index) is impacted by plant growth phase and soil management (three plots per management in DOK field,  $n = 6$  per plot; CO: CONMIN, BI: BIODYN, Un: Unplanted soil, V: vegetative stage, R: reproductive stage). The Wilcoxon test was used for statistical analysis with FDR correction. Capital letters indicate significant differences between managements; lowercase letters indicate significant differences between time points.

We extended the characterization of all three microbial kingdom communities to all tested compartments, including soil, rhizosphere and root, at vegetative and reproductive growth phases (Figure 6.4). Bacterial and fungal communities were clustered by compartment (Figure 6.4a), and distinguished by host growth stage and management practices along the fourth and third axis, respectively (Figure S6.1a, b). The oomycetal communities showed a larger dispersion and partially clustered by soil management (Figure 6.4a). The main drivers observed to be responsible for variations in the soil and root-associated microbiota were confirmed by PERMANOVA (Figure S6.1c). For all microbial kingdoms, we observed a decrease in diversity (Shannon indices) from soil to rhizosphere and root (Figure 6.4b). However, the bacterial alpha-diversity increased in both soil and rhizosphere from the vegetative to the reproductive growth phase, while remaining stable in the root compartment over both growth stages. In contrast, for fungi and oomycetes, a decrease in diversity was observed in both rhizosphere and root over the growing season.

In both root-associated compartments (rhizosphere and root), filamentous eukaryotic communities (fungi and oomycetes) were separated by soil management, whereas bacterial samples additionally clustered by the local heterogeneity of soil chemistry at the plot scale (Figure S6.2a, b). Moreover,

soil management was also found to influence the alpha-diversity of the rhizosphere and root microbiota (Figure S6.2c, d), and the effect of plant growth stage shown earlier in Figure 6.4b was independently observed under most types of management. Additionally, a small but significant effect of plant genotype on microbiota was observed in roots (explained variance by genotype, B: 2.37%, F: 3.04%, O: 5.22%;  $P < 0.001$ ).

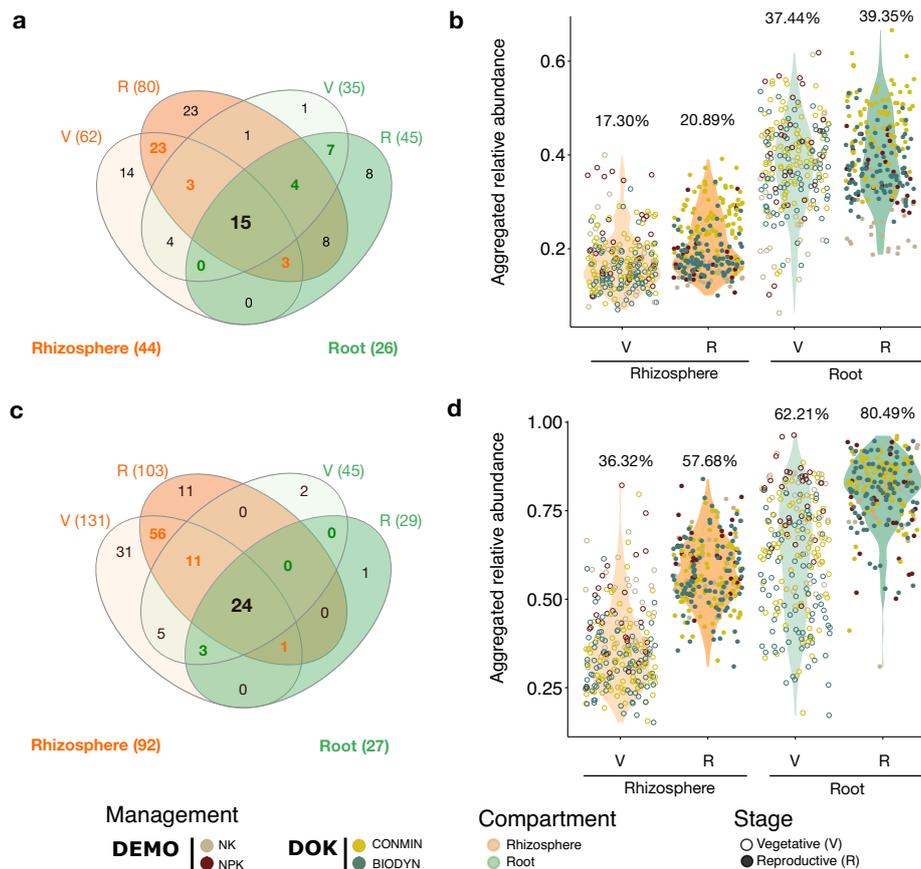


**Figure 6.4 Root compartment, plant growth phase, and soil management shape the root-associated microbiota in field-grown maize.** (a) PCoA of all harvested samples, including all three compartments (soil, rhizosphere and root), four soil managements (NK; NPK; CONMIN plot 1, 2, and 3; and BIODYN plot 1, 2, and 3) and three sampling times (before sowing, vegetative stage and reproductive stage). For bacteria, fungi and oomycetes,  $n = 1,079, 1,103,$  and  $1,103$  respectively. (b) Alpha-diversity (Shannon index) of all samples. Wilcoxon test ( $P < 0.05$ ) was used for statistical analysis with FDR correction. Capital letters indicate significant differences between compartments; asterisks indicate significant differences between different plant growth phases within each compartment.

### 6.3.2 Stable root-associated microbial taxa over host growth

Despite the aforementioned dynamics of microbial communities over the growing season, we were also able to investigate the stability of root-associated microbial members (Figure 6.5) by examining the widespread (found in  $> 80\%$  samples in the corresponding condition) taxa that are persistent, i.e. detected at both the vegetative and reproductive plant growth stages. We identified 26 stable bacterial OTUs in the root compartment (Figure 6.5a), consisting of Proteobacteria and Actinobacteria (16 and 10 OTUs, respectively) and representing more than half of the entire root community (50.94% aggregated Relative Abundance; aRA). Furthermore, 15 of the 26 stable OTUs were shared between root and rhizosphere compartments. These 15 OTUs accounted in the rhizosphere for 17.30% and 20.89% aRA in vegetative and reproductive growth phases (Figure 6.5d), respectively, and their aRA increased in the root compartment (to 37.44% and 39.35%, respectively) independent of field location or soil management (Figure 6.5b). This finding indicates

their progressive enrichment when moving from rhizosphere to root compartments. Most of these OTUs were assigned to Proteobacteria (12 / 15 OTUs), and the remaining three to Actinobacteria. A similar pattern was found for the fungal community. We identified 24 stable fungal OTUs between rhizosphere and root compartments (Figure 6.5c), irrespective of management practices and field location and representing on average 80.49% aRA in reproductive roots (Figure 6.5d). Most of these members were affiliated to four classes of Ascomycota (23/24 OTUs), namely Dothideomycetes, Eurotiomycetes, Leotiomycetes and Sordariomycetes. Interestingly, we found an inverse stable-to-dynamic ratio of bacterial and fungal root-associated assemblages (0.65 and 4.13 in roots at reproductive stage, respectively; calculation described in Figure 6.5 legend).

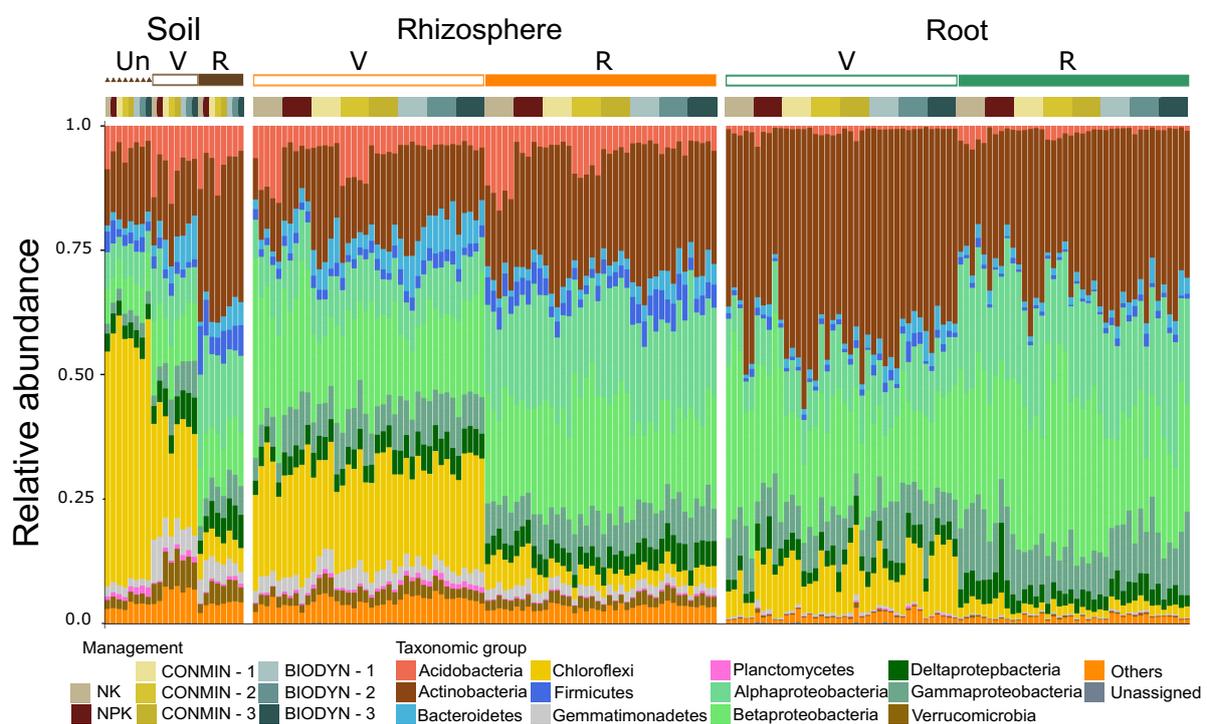


**Figure 6.5 Stable bacterial and fungal OTUs are enriched from rhizosphere to root compartment over plant growth, irrespective to soil management.** The Venn diagram shows the number of bacterial (a) and fungal (c) OTUs found in more than 80% of samples in each compartment at both plant growth stages (including NK, NPK, CONMIN and BIODYN soil managements and B73, DK105, PH207 and F2 plant genotypes). Relative abundance of stable OTUs were aggregated and demonstrated for bacteria (b) and fungi (d). Stable-to-dynamic ratio was calculated as the ratio of aRA between stable and dynamic community members, shown in (b) and (d). For example, the stable aRA of root-associated bacteria at the reproductive stage is 39.35% (b), hence the dynamic aRA is  $(100 - 39.35) = 60.65\%$ , thus the stable-to-dynamic ratio for bacteria in the root at the reproductive stage is  $60.65 / 39.35 = 0.65$ .

### 6.3.3 Microbial community assembly patterns at different phylogenetic levels

To better understand the community composition and to identify the principles governing community differentiation patterns, we compared community profiles and diversity at different taxonomic levels (Figures 6.6, 6.7). Examination of the relative abundance (RA) values of each individual bacterial phylum demonstrated an unexpected convergence over time of microbial

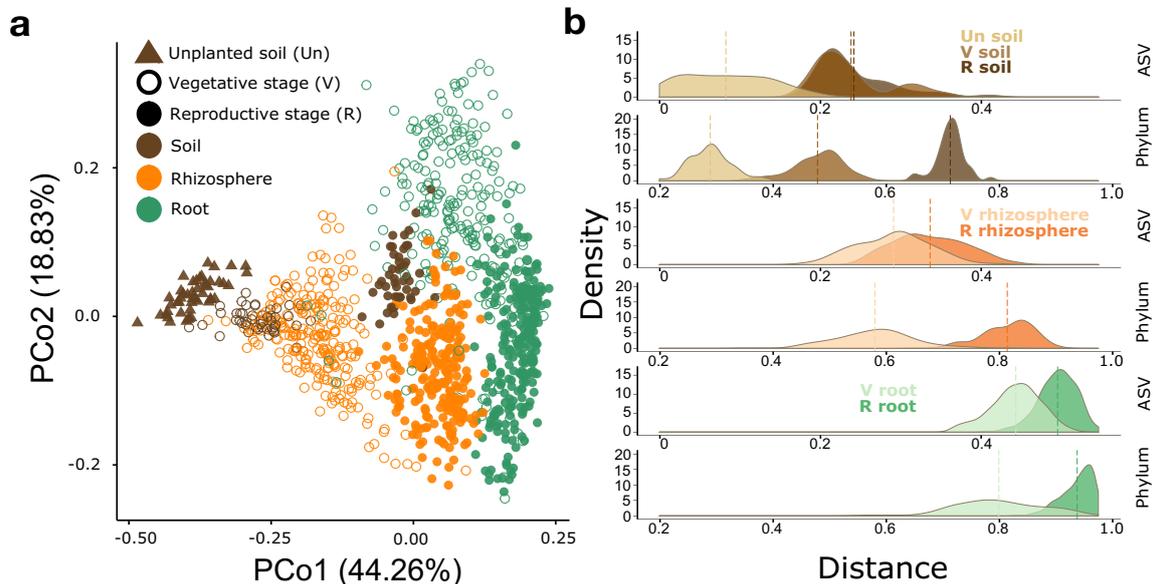
communities in the root microbiota, rhizosphere and planted soil (Figure 6.6). In soil samples, RAs of Chloroflexi and Planctomycetes was decreased over time, accounting for the main difference between the microbial biota of unplanted and planted soil. A significant enrichment of four taxonomic groups, namely Actinobacteria, Alpha-, Beta- and Gamma-proteobacteria, was identified in root samples compared to other compartments, independently of plant growth ( $P < 0.001$ ). Moreover, a progressive decrease in the RA of Acidobacteria, Bacteroidetes, Chloroflexi, Deltaproteobacteria, Firmicutes, Gemmatimonadetes, Planctomycetes and Verrucomicrobia was observed in a gradient from soil to rhizosphere to root samples, from the vegetative to the reproductive stage. In soil and rhizosphere, we also observed that Acidobacteria was significantly ( $P < 0.001$ ) enriched in the plots with the lowest pH values (NK and CONMIN-2) compared to the other plots under the same soil management, contributing to the higher similarities between those bacterial communities observed before (Figure 6.3a).



**Figure 6.6 Bacterial community profiles at the phylum level.** The relative abundance (RA) of the 12 most abundant bacterial taxonomic groups. Taxonomic group “Others” gathers bacterial phyla with less than 0.1% RA.

In order to characterize the effect of plant growth stage on bacterial community structure at a higher taxonomic level, beta-diversity analyses were performed at the phylum level (Figure 6.7). As previously shown for bacteria at the ASV level, the dissimilarities between samples were largely due to compartment, and communities of samples from soil, rhizosphere, and root were separated from each other. At the phylum level, however, samples from different growth stages within each compartment were separated. In particular, the later soil and rhizosphere communities were harvested, the more similar these communities became to those from roots. To quantify the observed effect of the plant on community structure across compartments, we calculated the distance between each sample, and the initial (unplanted soil) and final (reproductive phase of the root samples) bacterial communities. The distances were then compared between compartments and sampling times at the ASV and phylum levels (Figure 6.7b). Within each compartment, an obvious

separation between samples from different stages was found only at the phylum level, consistent with what we observed from the PCoA (Figures 6.4a and 6.7a). For instance, distance distributions of soil samples from the vegetative and reproductive stages were almost overlapping at the ASV level (0.0037 average distance difference; n.s.) but were significantly separated (0.23;  $P < 0.001$ ) at the phylum level.



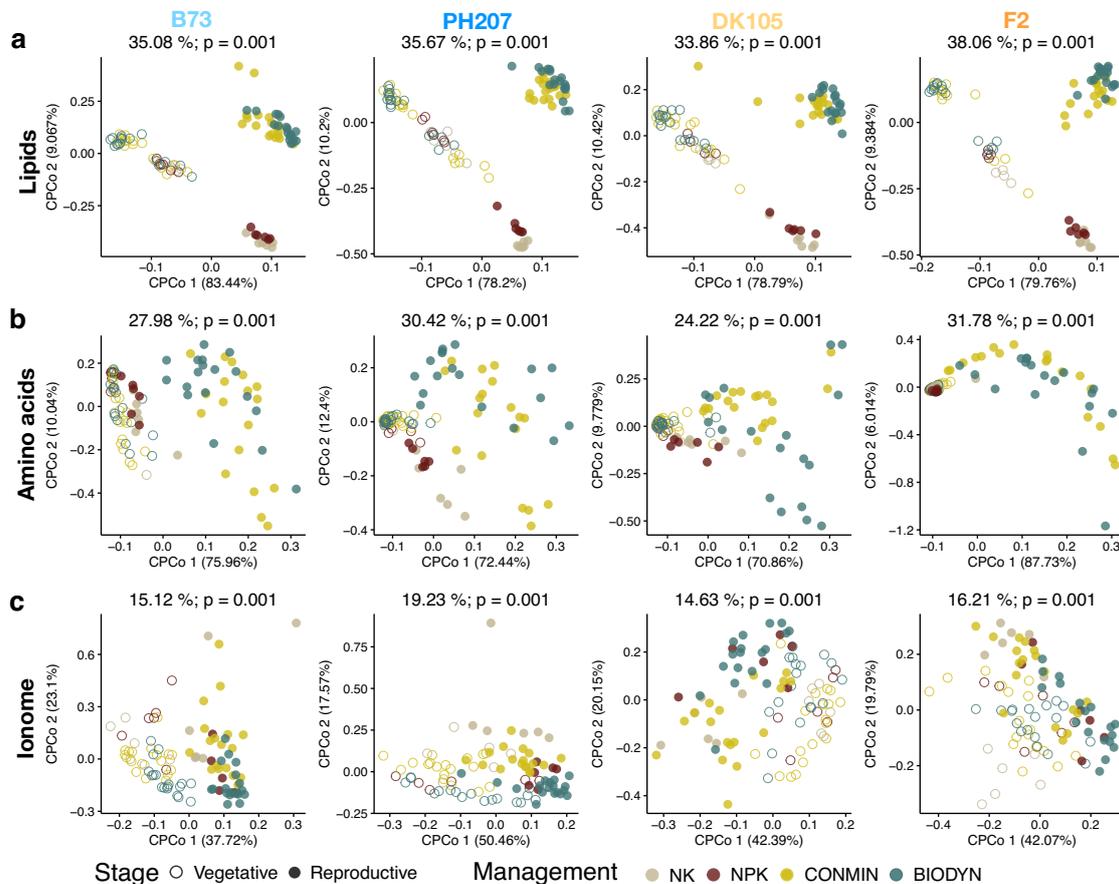
**Figure 6.7 Plant growth phase shapes bacterial community at the phylum level.** (a) PCoA of Bray-Curtis dissimilarity based on the RA of each phylum between communities ( $n = 1,104$ ). (b) Comparison of bacterial communities at the ASV and phylum levels. “Distance” in the x-axis indicating the Euclidean distance between the Bray-Curtis dissimilarity of each sample to the initial (unplanted soil) and final (reproductive root) condition community. The vertical line indicates the average distance of corresponding condition. Un: unplanted soil, V: vegetative soil and R: reproductive soil.

The structure of fungal communities was shaped by the compartment and the plant growth stage, which is highlighted by the significant ( $P < 0.001$ ) enrichment of Glomeromycota in the root (Figure S6.3a). The RA of this phylum decreased over the course of plant growth in rhizosphere and root ( $P < 0.001$ ). We also performed beta-diversity analysis at the phylum level for fungal communities (Figure S6.3b), which revealed an influence of compartment and host growth stage that is similar to the ASV level (Figure 6.4a). This pattern was further confirmed by the distance distributions of samples from different compartments and growth stages (Figure S6.3c). For rhizosphere samples, the average distances between vegetative and reproductive samples at ASV and phylum levels were similar (0.16 and 0.15, respectively). For root samples, a stronger separation at the ASV level than phylum level was observed (0.15 and 0.048, respectively;  $P < 0.001$ ). Taken together, ASV and phylum level distance distributions differ significantly for bacteria but not fungi.

### 6.3.4 Plant growth phase is a major driver of both root metabolism and root microbiota dynamics

To assess the temporal dynamics of plant metabolism under different soil management regimes, we characterized the root metabolome and ionome of wild-type plants at vegetative and reproductive growth stages, respectively (Figure 6.8). Of all examined metabolite classes, the profile of root

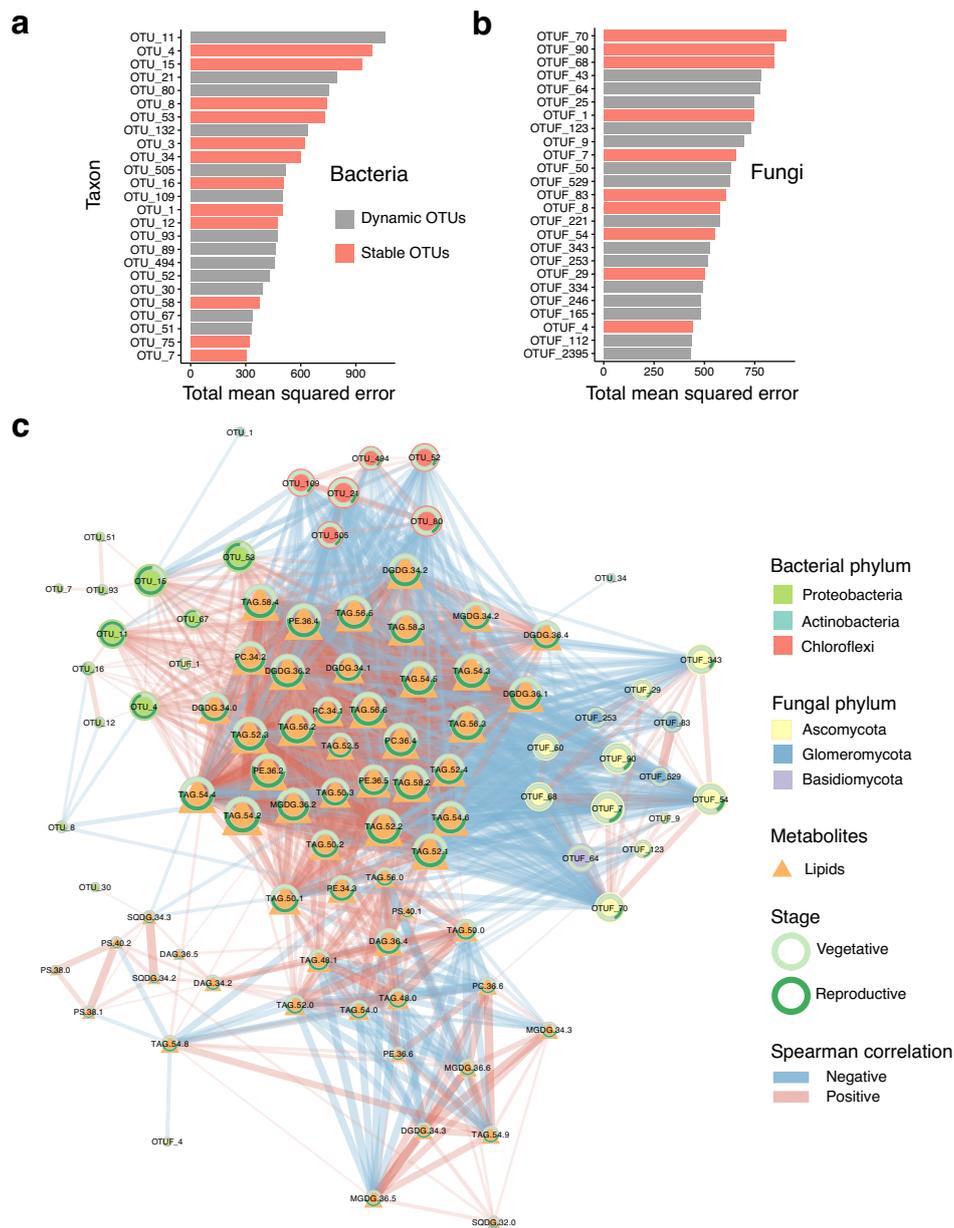
lipids was the most strongly affected by the plant growth phase (along the 1<sup>st</sup> axis), followed by field location and soil management, explaining 34 - 38% of the variance between samples of the four tested maize inbred lines (Figure 6.8a). When analyzing the individual lipids, we observed a general increase at the reproductive compared to the vegetative stage for most of lipid classes, irrespective of soil management and genetic background (Figure S6.4). This response was more evident when comparing the soil managements CONMIN and BYODYN at both developmental stages. In this case, levels of most of the annotated monogalactosyldiacylglycerols (MGDGs), digalactosyldiacylglycerols (DGDGs) and triacylglycerols (TAGs) were substantially increased at the reproductive stage. An effect of soil management was also evident, especially at the vegetative stage. For instance, levels of most of the MGDGs, DGDGs, and TAGs were reduced under CONMIN and BIODYN compared to NK and NPK. Some exceptions were MGDGs 34.3, 36.5, and 36.6 and the phosphatidylcholine PC 36.6 that showed an opposing response. At the reproductive stage, phosphatidylserine (PS), sulfoquinovosyldiacylglycerol (SQDG), and diacylglycerol (DAG) lipids also decreased under CONMIN and BIODYN compared to NK and NPK. We found similar but lower overall profile changes for amino acids (24 - 32% of variance for different inbred lines; Figure 6.8b) and the ionome (15 - 19% of variance for different inbred lines; Figure 6.8c), with plant growth phase again being the main explanatory factor. We also determined the sugar composition of roots, but the corresponding metabolite profiles were limited to the DEMO field (Figure S6.5). This confirmed that host growth phase is the most important explanatory variable for root metabolite dynamics of all metabolite classes tested.



(Caption on next page) **Figure 6.8 Root metabolites and total element compositions are affected by plant growth phase.**

(Figure on previous page) **Figure 6.8 Root metabolites and total element compositions are affected by plant growth phase.** Constrained PCoA (CPCoA) based on Euclidean distances between samples regarding the root lipid (a), amino acids (b), and ionic (c) profiles of the four inbred lines. PCoA was constrained by soil management and plant growth phase, for four inbred lines (59 lipid compounds were analyzed, n = 361; 15 amino acid compounds were analyzed, n = 376, 20 total elements were analyzed, n = 384). Each single point represents one analyzed root sample.

We then tested for covaried microbial taxa and root lipids, and identified a group of predictive taxa including both stable (widespread and persistent from vegetative to reproductive stage) and dynamic OTUs (Figure 6.9).



**Figure 6.9 The root associated microbiota covaries with root lipid profile over plant growth.** The most discriminatory (with highest mean squared error calculated by Random Forest, n = 25) bacterial (a) and fungal OTUs (b) for lipid in root and their phylum affiliation were shown. Salmon corresponds to the persistent OTUs in root shown in Fig. 2 and grey indicates non-persistent OTUs. (c) Network of the most predictive bacterial and fungal OTUs and lipids compounds. Strong (abs > 0.5) and significant (P < 0.05) Spearman correlations were kept as edges, n = 25 OTUs for both microbial kingdoms. Circle of each node shows the ratio of average RA of corresponding OTU or lipid between vegetative (light green) and reproductive stage (dark green) root.

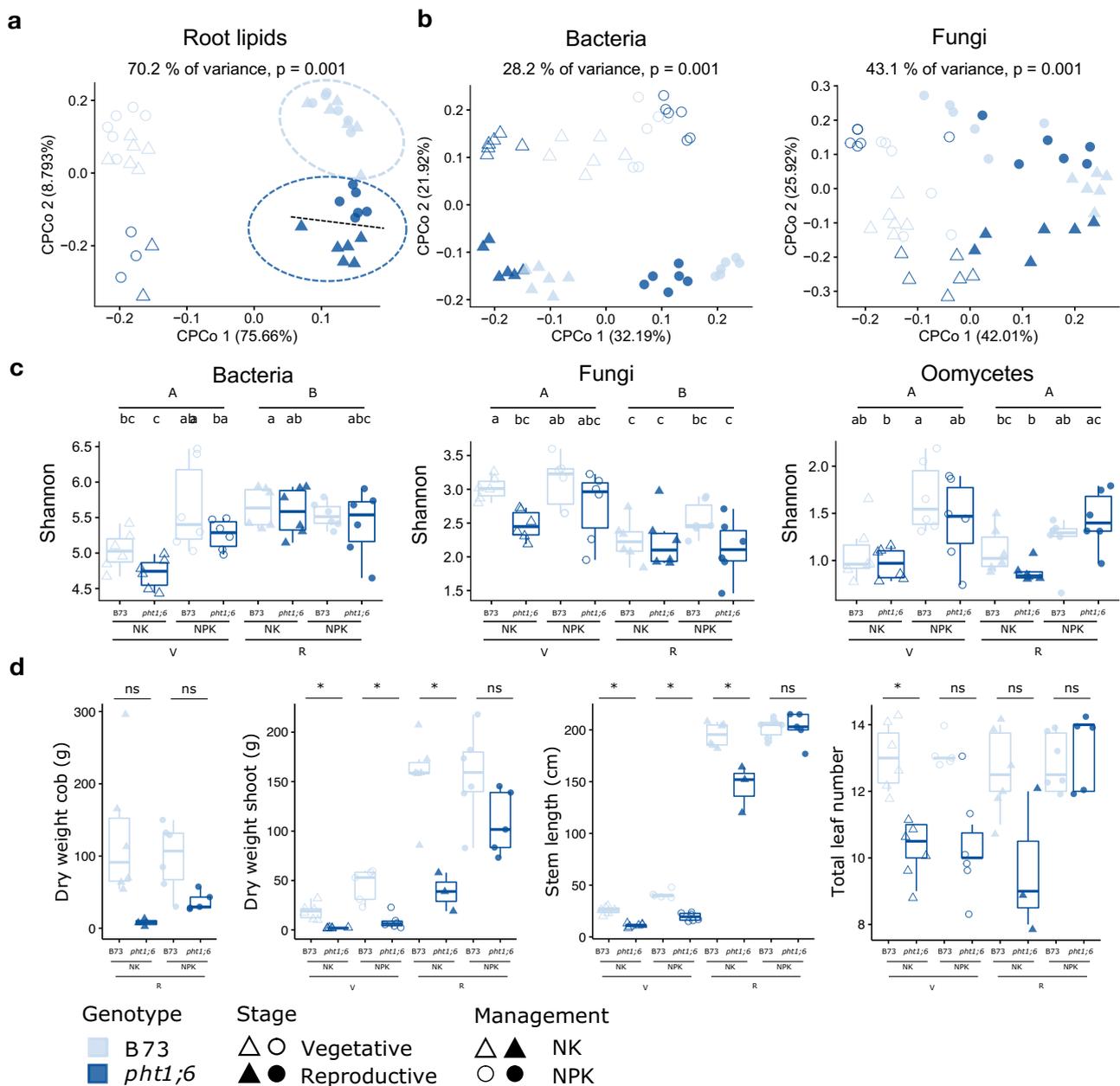
Approximately half of the most significantly covaried taxa (OTUs with top 25 max mean squared error value) were stable bacterial (12 OTUs, 10 from Proteobacteria and two Actinobacteria, out of 26 stable root OTUs in [Figure 6.5a](#)) and fungal (10 OTUs, nine Ascomycota and one Glomeromycota, out of 27 stable root OTUs in [Figure 6.5c](#)) taxa in the root ([Figure 6.9a, b](#)). We identified the microbial taxa that were significantly ( $P < 0.05$ ) and strongly ( $|r| > 0.5$ ) correlated with lipids ([Figure 6.9c](#)). Predictive Proteobacteria taxa were mainly found to be positively correlated with lipid profiling and Chloroflexi, a phylum that was decreased during plant growth in the root ([Figure 6.6](#)), was negatively correlated to most of the lipids. For fungal predictive OTUs, except for one Ascomycota, most taxa were negatively correlated to lipids ([Figure 6.9c](#)). Taken together, our data indicate that both the root metabolome and the root microbiota respond most strongly to plant growth phase, and that there is a correlation between stable and dynamic root microbial taxa and root lipids during plant growth.

### 6.3.5 Genotype-dependent response of root microbiota, metabolism, and plant biomass to soil P availability

To assess the response of the host plant, root metabolism and root microbiota to nutrient availability, we introduced a P transporter mutant *pht1;6* of B73, whose transport of bioavailable P from AMF to the plant host is impaired, thereby compromising the establishment of maize AMF symbiosis ([Willmann et al., 2013](#)). We firstly compared the degree of mycorrhizal root colonization and the RA of Glomeromycota in root samples of wild-type and mutant, under different managements and at different growth stages ([Figure S6.6](#)). As expected, in the *pht1;6* mutant we found a significant decrease in mycorrhizal colonization (based on microscopy) and RA of Glomeromycota (based on amplicon sequencing) at the vegetative stage ([Figure S6.6](#)). Surprisingly, at the reproductive stage of wild-type plants, the mycorrhizal colonization ratio, which was calculated only in fine roots by microscopy, increased ([Figure S6.6a](#)), while the average RA of Glomeromycota in the entire root system decreased ([Figure S6.6b](#)).

Afterwards, to specify the P effect, we examined root lipid, root microbiota and plant phenotypic data of samples derived from wild-type B73 and the mutant *pht1;6* plants grown in NK and NPK soil managements only ([Figure 6.10](#)). Root lipid compositions were mainly influenced by growth stage (1<sup>st</sup> axis) and plant genotype (2<sup>nd</sup> axis) ([Figure 6.10a](#)). In addition, at the reproductive stage, P soil amendment (NPK versus NK managements) differentiated lipid profiles in *pht1;6* roots ([Figure 6.10a](#)), indicating a potential plant growth phase-specific link between AMF symbiosis, root lipid status and soil P availability. The soil P availability-dependent change in root lipid profiles was paralleled by alterations in root-associated bacterial and fungal communities ([Figure 6.10b](#)). The main drivers of the bacterial community shift were P amendment (1<sup>st</sup> axis), followed by growth phase and plant genotype (13.39%, 9.47% and 3.34% of variance explained, respectively), whereas for fungi, samples clustered primarily according to plant growth stage (25.54%). We also detected a trend of lower bacterial and fungal alpha diversity in *pht1;6* roots compared to wild type, and in NK compared to NPK grown plants at the vegetative stage ([Figure 6.10c](#)), indicating that AMF and soil P deficiency have a broad effect on bacterial and fungal root-associated assemblages. This effect was undetectable at the reproductive growth stage, suggesting compensatory changes in the maize root microbiota over time. Consistent with this interpretation, we found that marked differences in

tested plant performance parameters in *phl1;6* compared to wild-type plants, seen at the vegetative growth phase, were mostly compensated at the reproductive growth stage (Figure 6.10d).



**Figure 6.10 Genotype-dependent effect of phosphate depletion on root lipids, microbiota, and plant biomass.** (a) The CPCoA based on Euclidean distance of the root lipid profile of the wild type (WT) and P transporter-defective line *phl1;6* at both growth stage (71 lipid compounds were analyzed,  $n = 41$ , data could be found in the folder “00.data/meta\_data” on GitHub). (b) The CPCoA based on Bray-Curtis dissimilarity of the root bacterial and fungal communities ( $n = 48$  for bacterial and fungal communities) and constrained by plant genotype, plant growth stage, and soil management (depletion/amendment of P on soil). (c) Effect of genotype, plant growth stage, and P fertilization on microbial community alpha-diversity. ( $n = 6$  per plot; V: vegetative stage, R: reproductive stage). (d) Effect of genotype, plant growth stage, and P fertilization on four plant biomass parameters ( $n = 43$ ). Wilcoxon test was used for statistical analysis followed by FDR correction ( $P < 0.05$ ). Capital letters indicate significant differences between growth stages; lowercase letters indicate significant differences between genotypes in specific soil management, within each growth stage.

## 6. 4 Discussion

### 6.4.1 Soil physicochemical properties drive soil and plant-associated microbial community shifts across different microbial kingdoms

Our results showed that in a long-term agricultural system, soil biota composition was strongly impacted by abiotic factors determined by geographical location (DEMO: Fertilization DEMONstration experiment and DOK: Dynamic, Organic and Conventional fields; with approx. 100 km distance between them), soil management (including mineral NK, NPK, CONventional MINeral, and organic BIODYNamic) and plot location (one plot in DEMO and three plots in DOK for each management). These factors were responsible for variability in soil physicochemical properties and further differentially shaped the communities of bacteria, fungi and oomycetes in rhizosphere and root compartments.

We found a slightly higher alpha-diversity for the bacterial communities in DOK (Figure 6.3b), possibly linked to differences in the size distribution of soil aggregation between the DEMO and DOK fields. The particle size influences the oxygen concentration and moisture availability of the soil micro environment (Fierer, 2017), and therefore will affect all microbes living there. Moreover, differences in climatic conditions, such as rainfall and air temperature, likely contribute to the effect of geographical location. Consistent with recently characterized natural populations of *A. thaliana* and cultivated and native *Agave* species (Coleman-Derr *et al.*, 2016; Thiergart *et al.*, 2020), we observed that the community of filamentous eukaryotes is stronger influenced by geographical location compared to the bacterial community (Figures 6.2, 6.3, 6.4). However, because the two fields were not managed identically, this confounds the effects of geographical and management factors.

Through modification of the edaphic parameters within a given field, soil management also modifies the soil biota (Edwards *et al.*, 2015; Hartmann *et al.*, 2015; Schmidt *et al.*, 2019). These effects are mainly apparent with respect to community structure and are not as strong with regard to alpha-diversity (Hartmann & Widmer, 2006). Consistent with this, in the DEMO field, mineral NK and NPK fertilizations possessed distinguishable soil biota communities (Figures 6.2, 6.3, 6.4), probably depending on soil P level, but resulted in non-significant differences of bacterial and fungal alpha-diversity. This indicates that the alpha-diversity of the microbial soil biota is largely resilient to long-term P nutrient supplementation, as previously suggested by another P fertilization field experiment (Robbins *et al.*, 2018). In the DOK field, comparison of CONMIN and BIODYN soil properties featured higher pH and elevated contents of C<sub>org</sub>, POXC, total N, C/N, and available P in the organic treatment. Previous studies in the same field showed a positive effect of farmyard manure (FYM), used in BIODYN management, on microbial abundance and activity (Widmer *et al.*, 2006; Birkhofer *et al.*, 2008), and community composition (Hartmann *et al.*, 2015). In line with the concept of *r*-/*K* selection theory in ecology applied to soil microbial communities, in which selection of bacterial communities is driven by nutrient availability and their ecological strategies, we found oligotrophs such as Acidobacteria (*K*-strategist) enriched in CONMIN (Fierer *et al.*, 2007). The abundant nutrients provided by FYM in BIODYN plots promotes high-growth-rate microbes defined as copiotrophs, such as some Bacteroidetes (*r*-strategist), which may be able to degrade complex organic compounds (Lapébie *et al.*, 2019).

At a local scale within each management, multiple soil properties were highly variable between different plots (Figure 6.2). Bacteria were more impacted than filamentous eukaryotes by heterogeneity in soil chemistry at the plot scale (Figure 6.3). The sensitivity of bacteria to pH, which has also been reported previously (Fierer & Jackson, 2006; Rousk *et al.*, 2010), and to clay, highlights the importance of edaphic parameters for bacterial community structure. A shift of around one pH unit led to an enrichment of Acidobacteria in the NK and CONMIN-2 plots, regardless of their contrasting managements and field locations (Figure 6.6).

Horizontal gene transfers from fungi to oomycetes converge within the radiation of oomycetes capable of colonizing plant tissues and are associated with the transition to their predominant phytopathogenic lifestyle (Richards *et al.*, 2011). We observed a strong effect of soil management on the root-associated microbiota (Figures 6.3, 6.4), notably for oomycetes, suggesting that adapted agricultural practices can be instrumental in reducing damage caused by these ubiquitous phytopathogens. We found that the interaction between host genotype and management also explained a large degree of the variance in communities. This interaction suggests that plant microbiota composition is not only impacted by soil management, but also determined by the degree to which specific plant genotypes are adapted to specific environments.

#### 6.4.2 Temporal changes in maize root-associated bacterial communities covary with root metabolite dynamics

We have shown that the effect of soil management and plant genotype interaction on shaping root microbiota was relatively stable between vegetative and reproductive growth stages for the three microbial kingdoms. Time-course experiments in rice have shown a rapid acquisition and stable taxonomic structure of the bacterial root microbiota within 14 days after transplantation from sterile media to soil (Edwards *et al.*, 2015). The root-associated bacterial communities of these two-week-old seedlings were most similar to the root microbiota of six-week-old rice plants, which was interpreted as evidence that the root microbiota in vegetatively growing rice might approach a steady-state (Edwards *et al.*, 2015). A subsequent study comparing the bacterial and archaeal root microbiota of field grown rice during three growing seasons, including four cultivars, revealed changes in microbial composition from the vegetative to the reproductive stage, and identified predictive microbiota reflecting plant age (Edwards *et al.*, 2018). A plant age-dependent variation in rhizosphere bacterial communities of field-grown maize over a 20-week period with weekly sampling, spanning vegetative and reproductive growth phases, showed gradual rather than two-stage community shifts (Walters *et al.*, 2018). In light of these findings, the distinctive microbial profiles detected in our study at vegetative and reproductive growth stages for rhizosphere and root compartments are likely two snapshots of a gradual maize root-associated microbiota dynamic over time. In field-grown maize with weekly samplings, a core of seven bacterial OTUs shared in all rhizosphere samples was identified (Walters *et al.*, 2018). In this study, we found 15 stable bacterial OTUs, shared between rhizosphere and root compartments as well as between vegetative and reproductive growth phases, representing approximately one third of the root microbiota (Figure 6.5). A comparison of the seven common rhizosphere bacterial OTUs found in US fields with the 15 OTUs identified here in European field-grown maize shows an overlap at the rank of family (5/7 US and 7/15 European OTUs with shared family assignment, respectively), including Bradyrhizobiaceae, Comamonadaceae, Pseudomonadaceae and Sinobacteraceae. This overlap

indicates the existence of stable community members that define the maize bacterial root microbiota across two continents, despite different host genotypes, soil managements, climates and soil types.

The gradual shift in microbiota composition over maize growth (Walters *et al.*, 2018), and the uncoupling of genetically determined flowering time in *A. alpina* from soil residence time-dependent changes in its root microbiota (Dombrowski *et al.*, 2017), makes it unlikely that genetically determined vegetative and reproductive growth stages control the temporal dynamics of the bacterial root microbiota. Instead, dynamics of root metabolites over the growing season (Figure 6.8) could drive bacterial succession in a ‘shell’ surrounding a stable microbiota core, enabled by bacterial immigration from the soil biome. According to our study, this dynamic shell comprises up to two thirds of the bacterial root microbiota. This model is supported by a study in monocotyledonous *A. barbata*, in which it is shown that chemical succession in root exudation over the growing season explains part of the bacterial community assembly and dynamics in the rhizosphere (Zhalnina *et al.*, 2018). However, we found that the majority of the fungal root community is stable over the tested maize growth phases, indicating that the root-associated fungi are less responsive to changes in root metabolites.

At a high taxonomic level, we found a shift in the bacterial community towards a structure resembling that of later-stage roots in all of the compartments, including planted soil (Figures 6.6, 6.7). The large expansion of the maize root system over time in the field might extend the spatial chemical gradient from the plant to the soil, inducing this shift. This plant footprint was observed only at the phylum level, indicating that the root shapes bacterial communities based on their conserved metabolic (functional) potential, independently of intra-species (ASV-level) diversity (Bai *et al.*, 2015). We propose that this plant footprint on the bacterial communities is partially driven by the dynamics of root metabolites over time that spread via exudation beyond the rhizosphere into the soil. BXs, and its stable degradation product MBOA, are candidate maize root-secreted chemicals that could contribute to this mechanism (Hu *et al.*, 2018; Cotton *et al.*, 2019; Kudjordjie *et al.*, 2019). By contrast, for fungal communities, we found similar assembly patterns at the ASV and phylum levels over the growing season across all tested compartments, corroborating that fungi are more resistant to changes in root metabolism. It remains to be tested whether this difference reflects fundamentally different preferences of bacteria and fungi for root-derived nutrients and/or sensitivity to phytochemicals.

#### 6.4.3 Soil P availability induces changes in root metabolism, microbiota, and plant performance

By comparing wild-type and P transporter *pht1;6* mutant plants we showed a potential plant growth phase-specific link between AMF symbiosis, root lipid status and soil P availability (Figure 6.10). In AMF-deficient *pht1;6* plants during reproductive maize growth, root lipid status varied the most in response to P availability. Establishment of AMF symbiosis in roots is promoted under P-limiting conditions (NK), and during symbiosis the fungi can provide the dominant route for plant P supply (Smith *et al.*, 2003; Willmann *et al.*, 2013). In return, host-derived lipids are a major source of organic carbon delivered to fatty acid auxotrophic AMF for fungal growth (Bravo *et al.*, 2017; Jiang *et al.*, 2017; Luginbuehl *et al.*, 2017; Keymer *et al.*, 2017).

We found that root lipid status between wild type and AMF-deficient *pht1;6* plants differed more strongly in NK compared to NPK conditions, which is likely linked to a perturbed cross-kingdom lipid transfer from host to AMF in the mutant roots. We also observed that the performance of *pht1;6* mutants compared to wild type was reduced as previously shown (Willman 2013, Fabianska 2020), and more severely at the vegetative growth phase. These performance differences were partially compensated at the reproductive growth phase under P-sufficient condition (NPK). Lack of full growth compensation points to either additional AMF-independent functions of the Pht1;6 transporter for maize growth and/or additional beneficial AMF symbiosis activities independently of P supply to the host. Additionally, we found that contrasting P availability induced changes in the bacterial root microbiota that exceeded host growth phase-dependent microbiota variation in both wild type and *pht1;6* plants. This indicates that the bacterial root microbiota actively responds to host P status as shown in *A. thaliana* by a direct integration of P stress and plant immune responses (Castrillo *et al.*, 2017; Finkel *et al.*, 2019), thereby likely contributing to overall host performance.

This work provides insights into the spatio-temporal dynamic of maize root-associated microbiota by revealing an inverse stable-to-dynamic ratio between root-associated bacterial and fungal communities over the growing season. Future development of a gnotobiotic maize growth system and defined (synthetic) microbial communities should allow direct tests of whether changes in root metabolites drive succession in the dynamic shell of the bacterial root community, whereas stable bacterial and fungal root microbiota members have adapted to host metabolite alterations.

## 6.5 Materials and methods

### 6.5.1 Experimental design

The experiment was performed on two long-term fields located in Switzerland: DEMO (Fertilization DEMONstration experiment Agroscope, 47°25'31" N, 8°30'59" E; MAT 9.4 °C, MAP 1031 mm, Zürich-Reckenholz; established in 1987) and DOK (Dynamic, Organic and Conventional managements, (47°30'09" N, 7°32'21" E; MAT 10.5 °C, MAP 842 mm, Therwil; established in 1978; Hartmann *et al.*, 2015), where different soil managements (encompasses a combination of soil fertilization and agricultural practice) were applied (Figure 6.1). The soil type at DEMO is a Gleyic Cambisol and at DOK is a Haplic Luvisol field according to FAO (WRB, 2015). In the DEMO field, NPK (Nitrogen, Phosphate and Potassium) and NK (Nitrogen and Potassium) management were compared. In DOK, the Biodynamic mixed (BIODYN) and the Conventional solely mineral fertilized (CONMIN) managements, with three replicate plots dispersed within the field (at location 1, 2, 3), were used.

Within each plot, five maize genotypes were planted: inbred lines B73 and PH207 from genetic pool dent and DK105 and F2 from flint, in addition to mutant line *pht1;6* derived from B73 with a mutation in the mycorrhiza-specific P transporter gene Pht1;6 (Willmann *et al.*, 2013). The examined genetic pools correspond to a classification that considers the structure of the grain and differentiation in traits such as flowering time and cold tolerance (Unterseer *et al.*, 2016). For each genotype, six rows (with 75 cm distance) of five plants each (15 cm distance) were distributed per plot (n = 30). Plants were grown and harvested at two time points, at seven weeks (July 2017) and 15 weeks (September 2017) after sowing, corresponding approximately to the vegetative (V6) and reproductive (R2) stages of plant development, respectively (Abendroth *et al.*, 2011).

### 6.5.2 Soil and Plant sampling

For soil properties measurement, six disturbed soil samples (2 kg) from each plot were collected before sowing (n = 48). Topsoil samples were collected from 5 to 20 cm below surface along a linear transect to assess the effect of spatial variability within each plot, and 38 soil parameters were measured. For soil microbiota analysis, a subsample of 50 g from each soil sample was flash-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until analysis.

For each genotype, six healthy plants were harvested at each time point. To prevent any border or cross-genotype effect, plants in the middle of the plot (harvesting area) with at least one plant from the border or a different genotype were harvested, preferentially. After removing most of the attached soil, the representative sample of the root system (including different types of roots) was collected and flash-frozen in liquid nitrogen. Additionally, for each planted plot at each time point, 10 g of bulk soil (n = 3) were collected (-5 to -20 cm from surface) between two rows of plants for microbiota analyses.

### 6.5.3 Plant biomass and root metabolism measurement

For each harvested plant, the stem length, leaf number, and developmental status were determined. After measuring the fresh weight, the shoot was cut into pieces, dried at  $65^{\circ}\text{C}$  at Reckenholz over several days and ground, and then the dry weight was measured. Cobs of each plant were weighed similarly. For metabolome (lipid, sugar and free amino acids) and total elemental composition, frozen washed root tissues were homogenized into fine powder and 50 to 100 mg were used. Fine roots of maize plants were harvested at vegetative and reproductive stages in 70% EtOH and mycorrhizal colonization was assessed using ink staining.

### 6.5.4 Soil and root microbial community profiling

The collected root samples were fractionated into rhizosphere and endosphere (here after referred to as root) fractions in the laboratory with a protocol adapted from [Bulgarelli et al., \(2012\)](#). Total genomic DNA was extracted from bulk soil, rhizosphere and root samples from at least 250 mg of material using the FastDNA SPIN Kit for Soil (MP Biomedicals, Solon, USA). The concentration of DNA samples was measured by fluorescence (Quant-IT<sup>TM</sup>Picogreen, Invitrogen, Oregon, USA). DNA samples diluted to  $3.5\text{ ng}/\mu\text{l}$  were amplified in triplicate in a two-step PCR using specific primer sets for profiling of bacterial (V5–V7 region of 16S rRNA; 799F/1192R), fungal (ITS2; fITS7/ITS4) and oomycetal (ITS1; ITS1-O/5.8s-O-rev) communities, as described in [Robbins et al., \(2018\)](#). Illumina sequencing was performed at the Cologne Center for Genomics (CCG) using the MiSeq platform and custom sequencing primers.

### 6.5.5 Amplicon sequencing data processing and microbial community diversity analysis

The sequenced amplicon profiling data were processed with workflow based on DADA2 (v1.12.1, [https://github.com/Guan06/DADA2\\_pipeline](https://github.com/Guan06/DADA2_pipeline)) ([Callahan et al., 2016](#)). Forward and reverse reads were demultiplexed. For bacterial samples, raw sequencing reads were subsequently truncated to 260 bp (forward) or 240 bp (reverse) and filtered with “maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE”. For fungal and oomycetal communities, we mapped the primers to the sequencing reads to trim the non-amplified region and then filtered with “maxN=0, maxEE=c(2, 2)”. After

learning the error rates, ASVs were generated by merging the corrected forward and reverse reads, and chimeras were removed.

Taxonomy of ASVs were assigned by the naïve Bayesian classifier (Wang *et al.*, 2007), using SILVA (v132), UNITE (release 02.02.2019) and an in-house database described in Durán *et al.*, (2018) for bacteria, fungi and oomycetes, respectively. ASV tables were rarefied to depth 1,000, 10,000 and 2,000 for bacteria, fungi and oomycetes, respectively. The Shannon index of alpha-diversity was then calculated as the average value of 999 independent rarefactions. The average was then computed as alpha-diversity of communities. Bray-Curtis dissimilarity (BC) between samples was calculated based on rarefied ASV tables for beta-diversity analysis at the ASV level. Permutational multivariate analysis of variance (PERMANOVA) was performed with the *adonis()* function in R package *vegan* (Oksanen *et al.*, 2019). Bacterial OTUs were clustered with identity of 97% from ASVs.

For the diversity analysis at the phylum level, relative abundance of ASVs belonging to the same phylum were summed up to obtain the relative abundance of the corresponding phylum, and based on this, the BC matrix was calculated. Mean BC between each sample, and samples from unplanted soil (the most soil-like condition,  $x_n$ ) and reproductive-stage root (the most root-like condition,  $y_n$ ) were calculated. Samples with minimum  $x_n$  were defined as the starting point ( $x_0, y_0$ ) of the community and the Euclidean distance between this sample and all other samples was calculated. Additionally, the relative abundance of each phylum was compared between conditions by analysis of variance (ANOVA) and Tukey's post-hoc test (De Mendiburu & Yassen, 2020).

#### 6.5.6 Predictive taxa for root lipid identification and network construction

To identify the predictive taxa for root lipid dynamics, we used function 'randomForest()' from R package *randomForest* (Liaw & Wiener, 2002) with parameter 'ntree = 1000' using OTU table (for bacteria and fungi respectively) and lipid profiling dataset in the model. Afterwards, mean squared error (MSE) was calculated with function 'importance()' using 'scale = TRUE'. Taxa with highest MSE ( $n = 25$ ) was then extracted as predictive OTUs. We calculated Spearman correlation between predictive OTUs and lipids and filtered the correlation matrix by keeping only significant ( $P < 0.05$ ) and strong (with absolute value  $> 0.5$ ) coefficients. Subsequently, the network was visualized by Cytoscape (v 3.8.2, Shannon, 2003).

### 6.6 Data and code availability

The raw sequencing data described in the manuscript were uploaded to the European Nucleotide Archive under accession number PRJEB44300. The modified DADA2 pipeline for data processing ([https://github.com/Guan06/DADA2\\_pipeline](https://github.com/Guan06/DADA2_pipeline)) and scripts as well as clean data for visualization ([https://github.com/Guan06/Bourceret and Guan et al 2022](https://github.com/Guan06/Bourceret_and_Guan_et_al_2022)) are both available on GitHub.

### 6.7 Author contributions

Bourceret and Guan et al., 2022

S.S., T.M., A.F., U.S., M.B. and P.S.-L. conceived the project. A.B., S.S., K.D., N.G. designed the experiment with support of the RECONSTRUCT project consortium. J.M. managed the field experiment. A.B., K.D., A.O., J.H. collected soil and root material, and measured plant performance

parameters. A.B. performed microbial community profiling, K.D. carried out soil sampling and physico-chemical measurements before sowing. A.O. and D.B.M. characterized root lipid profiles, J.H. characterized root sugar and free amino acid profiles. N.G. determined root ionome profiles and conducted microscopic mycorrhizal colonization analyses. R.G. and A.B. analyzed the data with support of R.G.-O. R.G. and A.B. created the Figures. A.B., R.G., and P.S.-L. wrote the manuscript with support from S.S. and R.G.-O.



# Chapter 7 Studying plant-associated microbial communities using novel diversity and network analyses

## 7.1 Abstract

With the rapid development of sequencing technologies, an increasing number of plant microbiome datasets have been generated. At present, microbiota diversity analyses are conducted by comparing changes in community composition across samples. However, these approaches ignore interactions between microbiota members and limit the study of community dynamics. To better understand the assembly of plant microbiota, we integrated extensive amplicon datasets and developed a framework for microbial community diversity and network analysis. Examining this combined dataset showed the absence of well-defined core microbiota at the level of amplicon sequence variant. Therefore, we proposed to extract the most abundant and prevalent members as representative microbiota and demonstrated that they capture the most community diversity and network dynamics. Based on the selected members, we inferred a large-scale co-occurrence network, from which microbes with co-varying abundances were clustered into groups for diversity measurement. We show that the unexplained variance was decreased compared to traditional composition-based methods. Furthermore, we introduced a bootstrap- and permutation-based statistical approach to compare microbial networks from diverse conditions at the global and local scales; through the latter, we extract the distinctive features contribute to microbial community dynamics. We provide these computational tools as an open-access R package, named ‘*mina*’.

## 7.2 Introduction

Microbes live in most ecosystems and form complex communities by interacting with each other and their surrounding environment (Raes & Bork, 2008). The assembly and stability of these microbial communities are affected by both biotic and abiotic environmental factors, such as the presence of a host and the availability of nutrients. The composition of these microbial communities could be determined by amplifying and sequencing marker genes, for instance rRNA genes, conserved single-copy protein genes, and internal transcribed spacer regions (Woese & Fox, 1977; Roux *et al.*, 2011; Schoch *et al.*, 2012). The development of sequencing technology makes the high-throughput profiling of microbial communities possible (Woese & Fox, 1977; Roux *et al.*, 2011; Schoch *et al.*, 2012). Using this approach, the composition of a community can be represented by operational taxonomic units (OTUs; Sokal & Sneath, 1963) or amplicon sequence variants (ASVs; Callahan *et al.*, 2017), extracted from sequences. OTUs are obtained by clustering the sequencing reads with arbitrary thresholds (e.g., the most commonly used 97% sequence identity), and ASVs are the error-corrected reads that can be distinguished by single nucleotide difference, therefore provide a finer resolution of component estimation. Typically, analysis of these data includes estimating within and between sample diversities (alpha- and beta-diversity, respectively) based on community profiles. Conventionally, alpha-diversity is evaluated by comparing richness and distribution evenness of potential community members, i.e., OTUs or ASVs. Beta-diversity is quantified using distances or dissimilarities calculated by comparing the abundance of each OTU or ASV between samples (Whittaker, 1960). Novel approaches which taking into account the

phylogenetics or microbial interactions are also developed for beta-diversity analysis (Lozupone *et al.*, 2006; Schmidt *et al.*, 2017).

To distinguish the communities from different conditions and thereafter identify the factors that affect the microbiota assembly, concepts such as “core members” were introduced (Turnbaugh *et al.*, 2009; Lundberg *et al.*, 2012). By focusing on specific members for hypothesis validation, the subsequent analysis is simplified due to the highly diverse and sparse properties of the microbiota from most natural conditions (Shade & Handelsman, 2012). For instance, in a plant root-associated microbiota study, OTUs that are enriched in *A. thaliana* roots and detected in different soil types were identified as “core microbiota” to facilitate the design of plant-associated SynComs (Lundberg *et al.*, 2012). Moreover, a “core microbiome” of the human gut at the gene rather than organismal level was detected among 154 individuals (Turnbaugh *et al.*, 2009), indicating the assembly of gut microbiota might be a function driven process. Meanwhile, the factors determining the compositions of core microbiota has been evaluated for marine sponge microbiotas, highlighting the effect of habitat (Astudillo-García *et al.*, 2017). Examining the abundance-occupancy distribution of micro-organisms offers the approach to prioritize core member selection for plant-associated and other microbiomes studies (Shade & Stopnisek, 2019). Though significant differences were detected in diversity analyses, the major conclusions were not affected by changes in the determining criteria (Astudillo-García *et al.*, 2017). However, a consensus definition of the core microbiota is still missing, so is a thorough assessment of to which extent the diversities are skewed when only considering those core compositions. Moreover, due to the fact that most of the core members are defined by taxonomic units, their contribution to the functional capacity of the whole community across different ecosystems are predominantly unknown (Lemanceau *et al.*, 2017).

In addition to the compositional approaches that characterize community structure and measure the differences between samples, the dynamics of the system such as interactions between microbes can also be examined by inferring co-occurrence networks (Barberán *et al.*, 2012; Faust *et al.*, 2015; Mandakovic *et al.*, 2018; Banerjee *et al.*, 2019; Mamet *et al.*, 2019; Huang *et al.*, 2019b; Lima *et al.*, 2020; Zamkovaya *et al.*, 2021; Yuan *et al.*, 2021). In these microbial community networks, nodes represent community members and relationships between microbes are indicated by undirected edges, which are inferred by comparing the covariance of microbes across samples. However, the compositional nature of the microbial community profiles, i.e., only the relative abundances are available, leads to spurious results when applying simple correlations (Aitchison, 1982). Therefore, novel methods with extra data transformation, such as SparCC (Friedman & Alm, 2012) and SPIEC-EASI (Kurtz *et al.*, 2015), or integrated workflows, including CoNet (Faust & Raes, 2016) and NetCoMi (Peschel *et al.*, 2021), were developed to reduce the bias introduced by compositional effect. However, due to the lack of a ground truth, these newly developed approaches cannot be proved to perform better than classical correlation methods (such as Pearson and Spearman) for microbiota studies, which usually involve the complex natural communities (Weiss *et al.*, 2016; Hirano & Takemoto, 2019).

Another challenge in microbial network inference is performing data analyses when the number of available samples is low. Due to the labour and sequencing cost, the number of conditions and replicates for each condition is usually limited. Typically, thousands to tens of thousands of community members were represented by OTUs or ASVs for plant-associated microbiota,

especially in rhizosphere and soil, while only tens to hundreds of samples were available (Barberán *et al.*, 2012). To compensate for this, community profiles are usually filtered by prevalence (Röttgers & Faust, 2018; Zamkovaya *et al.*, 2021; Yuan *et al.*, 2021) before constructing the network to reduce the sparsity of the data. Concepts such as the above-mentioned core microbiota are also applied for this purpose. Although disregarding the unknown or unassigned community members is also an option, it comes with information leakage according to a recent study that showed the unknown or unassigned microbes captured by sequencing have an important effect on the network topology (Zamkovaya *et al.*, 2021).

To compare networks from different environments, typical methods focus on the comparison of specific topology features inferred from the adjacency matrices, including clustering complexity (Xiong *et al.*, 2021), clustering coefficient (Mandakovic *et al.*, 2018; Yuan *et al.*, 2021), density (Faust *et al.*, 2015), centrality (Mamet *et al.*, 2019; Lima *et al.*, 2020; Zamkovaya *et al.*, 2021), and connectivity (Banerjee *et al.*, 2019; Huang *et al.*, 2019b). To calculate these network features, edges are firstly filtered according to criteria based on such as  $P$ -value (Ma *et al.*, 2020; Lima *et al.*, 2020), correlation coefficient (Mamet *et al.*, 2019), or the top-ranking (Faust *et al.*, 2012). To overcome this cut-off determination, approaches such as random matrix theory was applied to generate the thresholds for edge selection automatically (Yuan *et al.*, 2021). Moreover, the statistical tests for these comparisons are highly dependent on the non-parametric permutation procedure, therefore are computationally expensive and time-consuming (Peschel *et al.*, 2021). Alternatively, approaches that directly quantify global network structural differences have been developed (e.g., spectral distances), but are not commonly used for the comparison of microbial networks. A likely reason for this is the lack of methods to assess whether these distances are statistically significant.

After the network comparison, factors that contribute to the distances between networks of different conditions are typically the most biological relevant feature to be examined. For example, key genes were recovered by applying differential network analysis algorithms on the gene co-expression data (Lichtblau *et al.*, 2016), while similar approaches are still missing in microbial network studies, most likely due to the lack of corresponding computational tools. As one of the most comprehensive workflows for microbial network analysis, NetComi offers the quantification of differences in connections of single taxa or groups of taxa between conditions (Peschel *et al.*, 2021). However, it is still unclear how to decide on the taxa that to be compared and whether taxa are the best units for the assessment of contribution to the network dynamics.

In this chapter, we describe a computational and statistical framework implemented in an R package called ‘mina’ (microbial community diversity and network analysis), which is developed for microbial community data processing with a focus on network comparison. By applying this approach to the plant root microbiota dataset that we integrated from published studies (Zgadzaj *et al.*, 2016; Durán *et al.*, 2018; Thiergart *et al.*, 2019, 2020; Harbort *et al.*, 2020; Wippel *et al.*, 2021), we show that ‘mina’ diversity analysis was able to decrease the unexplained variance ratio. We introduce a test based on Spectral distances and Monte Carlo permutation that can be used to statistically assess differences between ecological networks. Using this method, we are able to compare networks constructed from samples collected under different conditions that associated with the same soil type (Cologne Agricultural Soil, CAS), obtain novel patterns of community

network dynamics between environments, and identify the distinctive features that contributed significantly to the network deviation.

## 7.3 Results

### 7.3.1 Meta-analysis of the plant microbiota

#### 7.3.1.1 Dataset overview

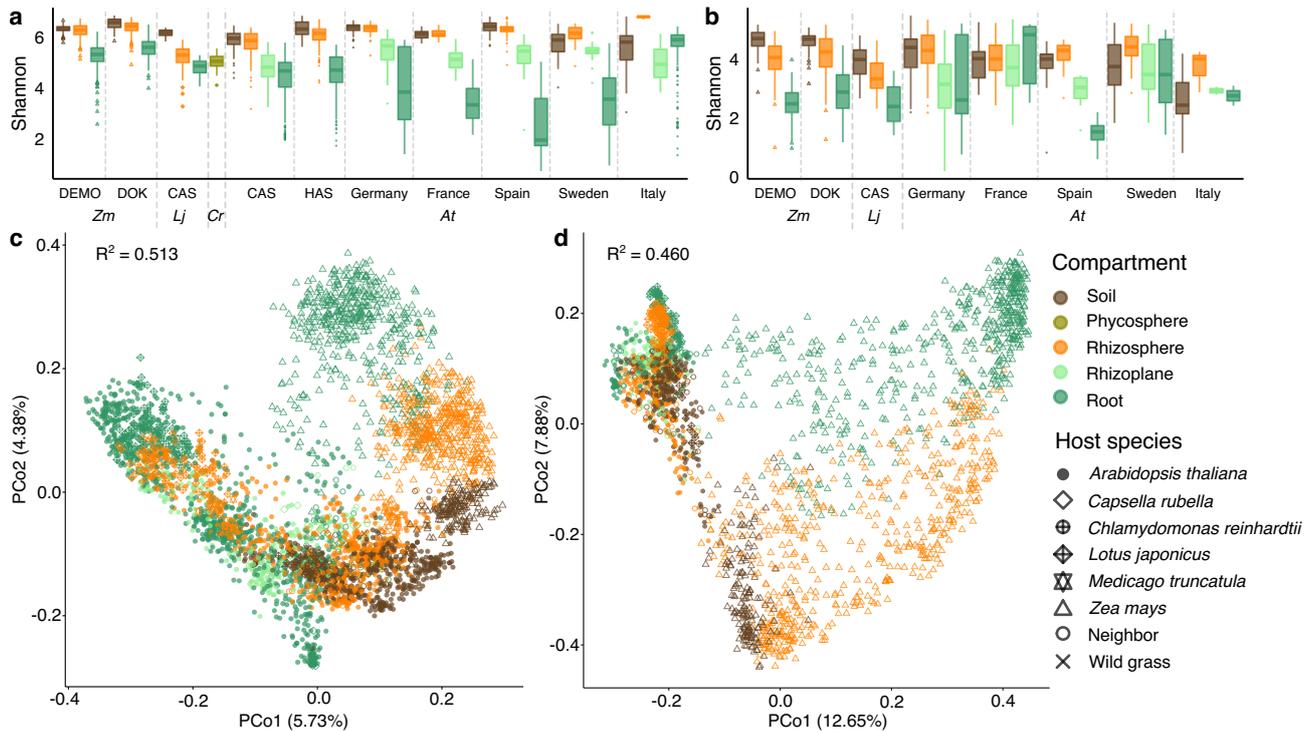
By colonizing and interacting with plant, microbes derived from soil assemble into multi-kingdom complex communities, denoted as root microbiota. Based on the physical distance to the plant host, these communities are categorized into different compartment groups, including endosphere/root, rhizoplane, rhizosphere, and soil. To understand the principles governing the assembly of plant-associated microbiota, we assembled a large-scale dataset for both bacteria and fungi. Published datasets of former studies (Zgad Zaj *et al.*, 2016; Durán *et al.*, 2018; Thiergart *et al.*, 2019, 2020; Harbort *et al.*, 2020; Wippel *et al.*, 2021) using the same primer set (bacteria: 16S rRNA gene V5 to V7; fungi: ITS2 region) were included. We integrated these samples, comprising 3,809 bacterial 16S rRNA and 2,232 fungal ITS2 amplicon profiles, as well as their corresponding meta-data, into a large-scale dataset spanning diverse soil types, host species and microhabitats (Table 7.1).

**Table 7.1 Dataset overview.** Conditions with less than 30 samples are shown as “Others”, including different compartments from host species *Capsella Rubella* and *Medicago truncatula*. Neighbour indicates the neighbouring grass beside the wild *A. thaliana* described in Thiergart *et al.*, 2020.

Host	Compartment	Bacteria	Fungi
<i>Zea mays l</i>	Rhizosphere	498	540
	Root	497	541
<i>Lotus japonicus</i>	Rhizosphere	113	88
	Root	113	78
<i>Chlamydomonas reinhardtii</i>	Phycosphere	48	NA
	Rhizosphere	486	186
<i>Arabidopsis thaliana</i>	Rhizoplane	209	107
	Root	909	101
	Rhizosphere	42	44
Neighbor	Rhizoplane	36	29
	Root	32	42
NA	Soil	760	476
Others	NA	66	NA
Total		3809	2232

#### 7.3.1.2 Data processing and diversity analysis

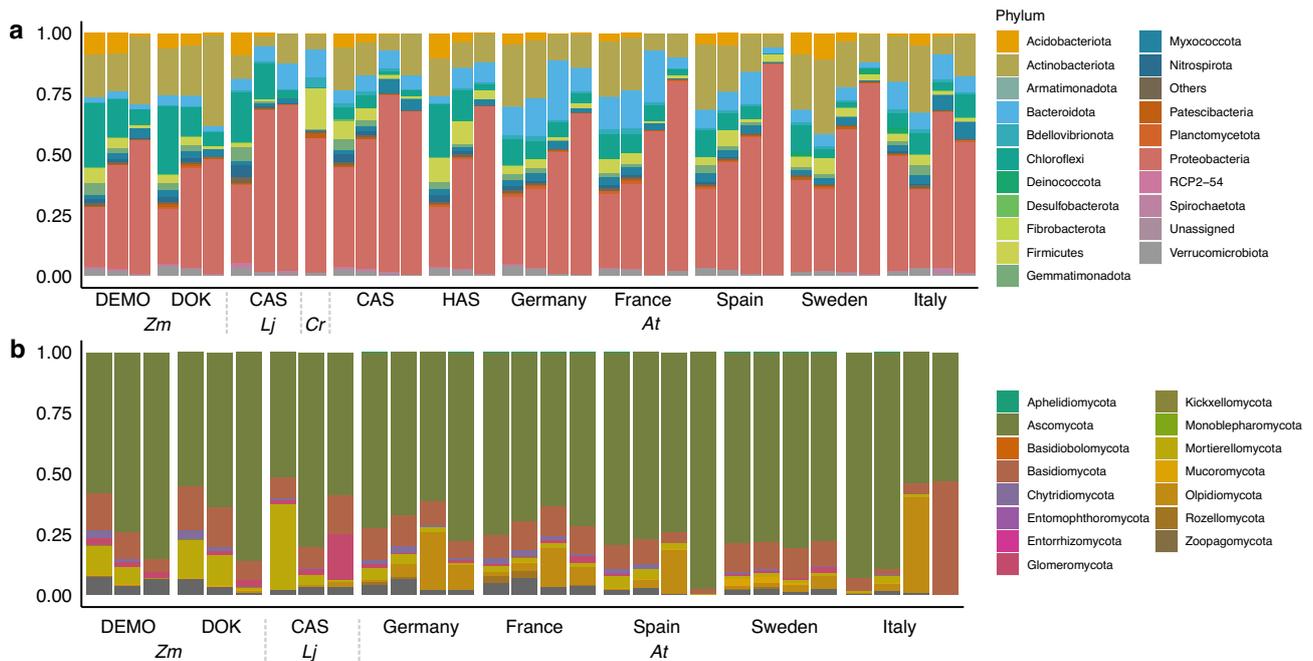
Raw data analysis was performed by applying standardized quality filtering step, followed by error-correction using DADA2 (Callahan *et al.*, 2016), leading to the identification of 42,060 bacterial and 9,337 fungal ASVs (details described in the method Section 7.5.1). Analyses of alpha-diversity (within-sample diversity) showed a decrease in the complexity of both bacterial and fungal microbial assemblages from the highly diverse soil communities to the rhizosphere, rhizoplane and root compartments (Figure 7.1a, b). This result is in line with previous studies, which show a decrease in alpha-diversity as the strength of the association with the host increase, in a pattern which is thought to reflect a process of recruitment by the plant of selected community members (Bai *et al.*, 2015; Zgad Zaj *et al.*, 2016; Durán *et al.*, 2018; Thiergart *et al.*, 2019, 2020).



**Figure 7.1 Community diversity of plant-associated microbiota.** Average Shannon indices of 999 times rarefaction and calculation were shown for bacterial (a) and fungal (b) samples under each condition. PCoA of Bray-Curtis dissimilarities between bacterial (c) and fungal (d) communities are shown here.  $R^2$  indicates the variance between samples which cannot be explained by compartment, soil type, host species, host genotype, and experimental condition (natural site or green house).

We examined the community structure at the higher taxonomic rank and consistently with previous studies (Bai *et al.*, 2015; Zgadzaj *et al.*, 2016; Durán *et al.*, 2018; Thiergart *et al.*, 2019, 2020), the bacterial phyla Proteobacteria, Bacteroidota, Actinobacteriota and Chloroflexi were the most abundant (Figure 7.2a). As the most abundant phylum, aggregated RAs of Proteobacteria increased from soil to rhizosphere, rhizoplane, and root, ending with up to an average of 47.37% - 87.17% for different hosts. On the contrary, Chloroflexi showed an opposite trend from soil to root, with RAs decreasing from 4.86% - 27.77% in the soil to 0.39% - 9.46% in the root. Notably, higher RAs of Firmicutes, accounting for 16.67% on average, were found in the phycosphere of *Chlamydomonas*, contrasting to a relative abundance of 1.64% - 9.53% in soil and 0.78% - 9.28% in rhizosphere conditions. For fungal communities, the most abundant phyla were Ascomycota, Basidiomycota, Mortierellomycota and Olpidiomyocota (Figure 7.2b). With absolute dominant RA, the Ascomycota increased its portion from soil to root, up to 97.83% in the root samples of *Arabidopsis* collected in Spain. For the maize-associated samples, a decrease of Basidiomycota from soil to root were observed (9.49% and 10.43% for DEMO and DOK soil types respectively). Mortierellomycota was detected particularly abundant in CAS soil, with an aggregated RA of 34.72%. Olpidiomyocota was mainly enriched in the rhizoplane of *Arabidopsis*, consisting of 15.72% to 39.39% in different soil types, with an outlier of 2.50% in Sweden samples.

(Figure on next page) **Figure 7.2 Taxonomic profiling of community structure at the phylum rank.** The most abundant (with aRA > 0.1%) bacterial (a) and fungal (b) phyla are shown and samples from different host species, soil types, and compartments are compared here.

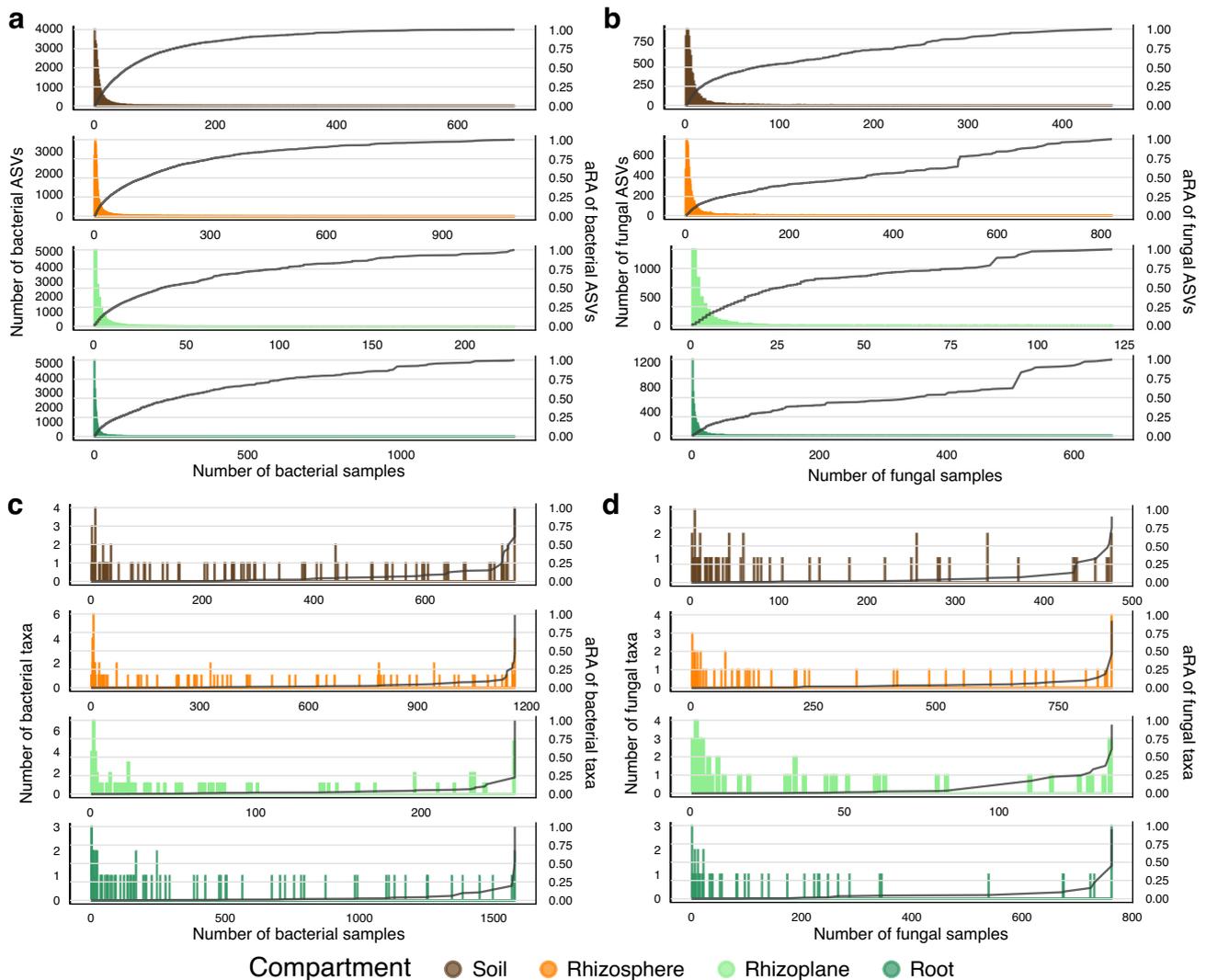


(Caption on previous page) **Figure 7.2 Taxonomic profiling of community structure at the phylum rank.**

### 7.3.1.3 A core microbiota was observed only at the higher taxonomic ranks

To explore changes in community composition across microhabitats, we analyzed the beta-diversity (between-sample diversity) of bacterial and fungal amplicon profiles. We observed a strong effect of compartment and host species in differentiation microbial communities, which were the most important drivers of diversity in the dataset (Figure 7.1c, d). Despite these patterns, a large percentage of the variance of the data (51.3% for bacteria and 46.0% for fungi) was not explained by any biological (including compartment, soil type, experimental condition, host species and genotype) or technical factor (namely sequencing run). We hypothesized that this unaccounted variance could be caused by stochastic variation of rare taxa (i.e., found only in a subset of samples). Therefore, we investigated the prevalence of bacterial and fungal ASVs in our dataset and found that the ASV prevalence across samples followed an exponential distribution within each environment, with most amplicon tags observed only in a small subset of samples (Figure 7.3a, b). We observed an almost complete absence of core ASVs (i.e., found in the majority of samples) in our dataset.

Later, we examined the distribution of taxa at different taxonomic ranks and core taxa were distinguished at the higher levels. For example, clear cores were observed for all compartments at the class rank, indicated by both increased number and contribution to the accumulated RA (aRA) of taxa detected in most of the samples within each condition (Figure 7.3c-d). Moreover, we observed similar results when assessing the prevalence of microbes in different compartments from each host species, for both kingdoms. Specifically, samples from the rhizosphere compartment show stronger enrichment of widespread taxa in lower taxonomic levels compared to others within each host species (Figure S7.1). Notably, for bacterial communities from the root of *Lotus japonicus*, the core was observed already at the genus rank (Figure S7.1).



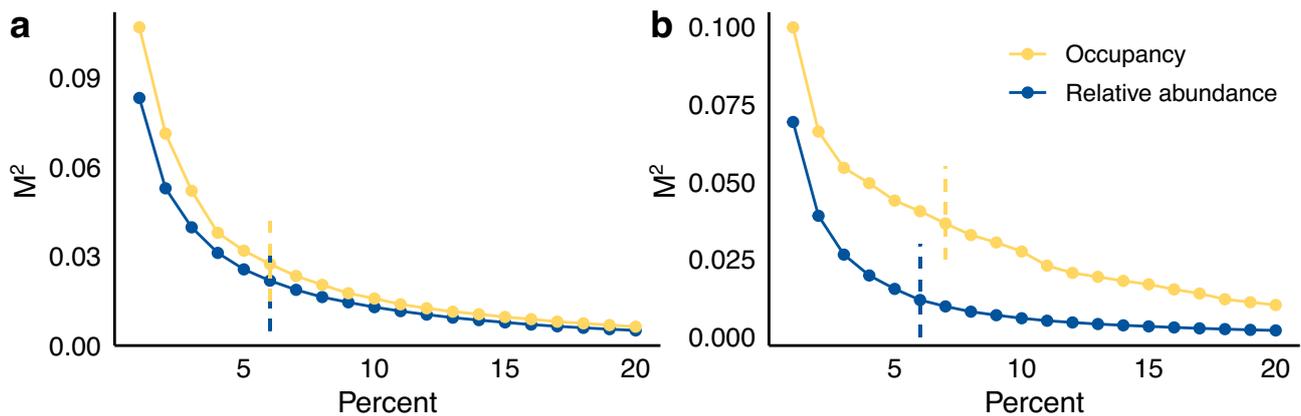
**Figure 7.3 Occupancy and accumulative relative abundance of microbes in root-associated compartments.** The distribution of bacterial ASVs (a), fungal ASVs (b), bacterial Classes (c) and fungal Classes (d) are shown here.

#### 7.3.1.4 Representative members capture most of the community diversity

Due to the observation of a general lack of a core bacterial and fungal core microbiota at the ASV level in our dataset, we sought an alternative selection criterion to identify representative ASVs and remove the effect of highly variable and rare taxa. Recently, a strategy to select representative microbiota members based on abundance-occupancy distributions was proposed (Shade & Stopnisek, 2019). Following this approach, we ranked ASVs according to their relative abundance (RA) and occupancy, and used Procrustes Analysis to quantify their contributions to the patterns of beta-diversity observed for the entire dataset (detailed methods in section 7.5.3).

We identified 2,047 bacterial (ranked top 6% RA and occupancy) and 370 fungal (ranked top 6% RA and 7% occupancy) representative ASVs (repASVs, Figure 7.4), whose aggregated relative abundances reached 45.28% and 54.12% in soil, respectively, and accounted for the majority of the community in the rhizosphere (58.19% and 71.20%), rhizoplane (61.64% and 49.18%), and root (72.33% and 77.88%) compartments (Figure S7.2a, b). A similar trend was observed for each host species, where repASVs dominated their respective communities, with small variation between different soil types (Figure S7.2c, d). Alpha- and beta-diversity analyses based only on repASVs

recapitulated the patterns observed for the entire dataset (Figures 7.1, S7.3). Bray-Curtis dissimilarities based on repASVs showed high consistency with the results based on all ASVs ( $M^2=0.031$  for bacteria and  $M^2=0.038$  for fungi when compared by Procrustes Analysis), including a clear separation of samples according to compartment and host species (Figures 7.1c, d; S7.3a, b). Importantly, the percentage of unexplained variance using only repASVs decreased by 8.2% for bacteria (from 51.3% to 43.1%) and by 7.0% for fungi (from 46.0% to 39.0%). Together, the results from these analyses indicate that, despite the absence of a core microbiota at the ASV level, we can identify microbial taxa that are representative of the whole community. This approach has the advantage of increasing the variance of the data explained by biological factors, and reducing that associated with technical factors.

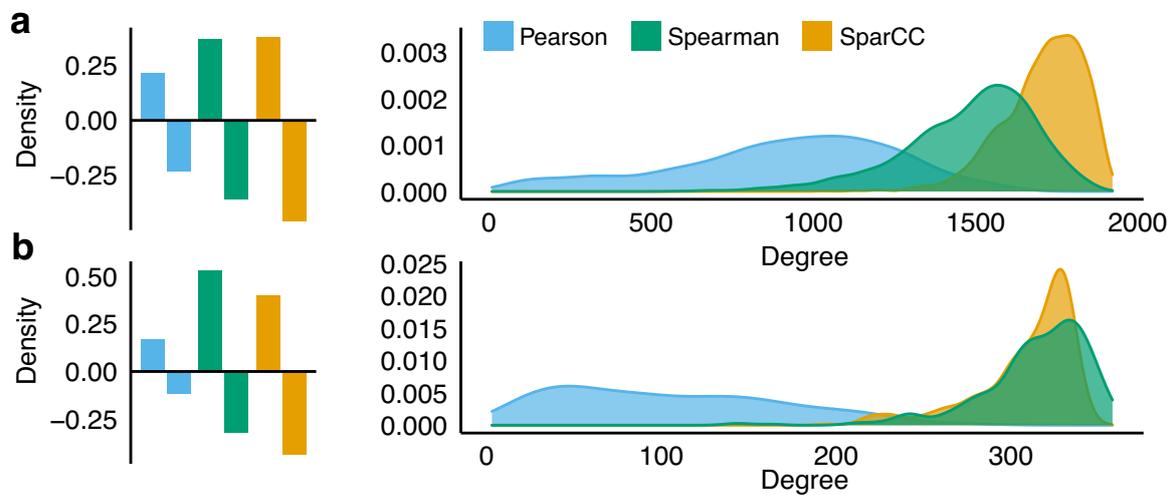


**Figure 7.4 Distance between the diversity of full and examined subset community members.** Subset members of bacterial (A) and fungal (B) communities were chosen and assessed for both RA and occupancy with thresholds from 1% to 20%. Vertical dash lines indicate the determined threshold for representative ASVs.

### 7.3.2 Network analysis of the plant microbiota

To reduce the noise introduced by rare taxa, we implemented the network analysis on the integrated dataset for bacteria and fungi respectively with determined repASVs, which in turn decreased the computing time and resource. Correlation coefficients between pairwise representative community members were computed according to their co-variance among samples, denoted as the edges in the global networks that connecting the nodes, i.e., microbial taxa that was represented by ASVs (detailed methods in section 7.5.4). To compare the connections of nodes between networks inferred from different correlation coefficients, we calculated the density for both positive and negative edges (Figure 7.5a, b) and compared the distribution of node degree, which represents the number of edges that connecting to each node (Figure 7.5c, d). A noteworthy bias introduced by the network construction method was observed for both kingdoms (Figure 7.5), where the use of Pearson correlation coefficients lead to more sparsely connected networks, while networks inferred using SparCC had the highest portion of negative connections. To verify if this observed bias is caused by the heterogenous and complexity of samples integrated for network construction, we inferred the networks separately for each compartment samples. We then examined the same features for the constructed networks individually and observed similar variances between applied correlation coefficients (Table S7.1). Even though we noted the density differences between compartments, due to the bias and random pattern demonstrated by compared methods, extreme cautions need to be

paid when trying to conclude any meaningful insights. To conclude, despite the complexity of samples used for network inference, the bias introduced by different correlation coefficients prevents the drawing of convincing conclusions when comparing networks only based on mentioned features such as the network density and degree distribution.



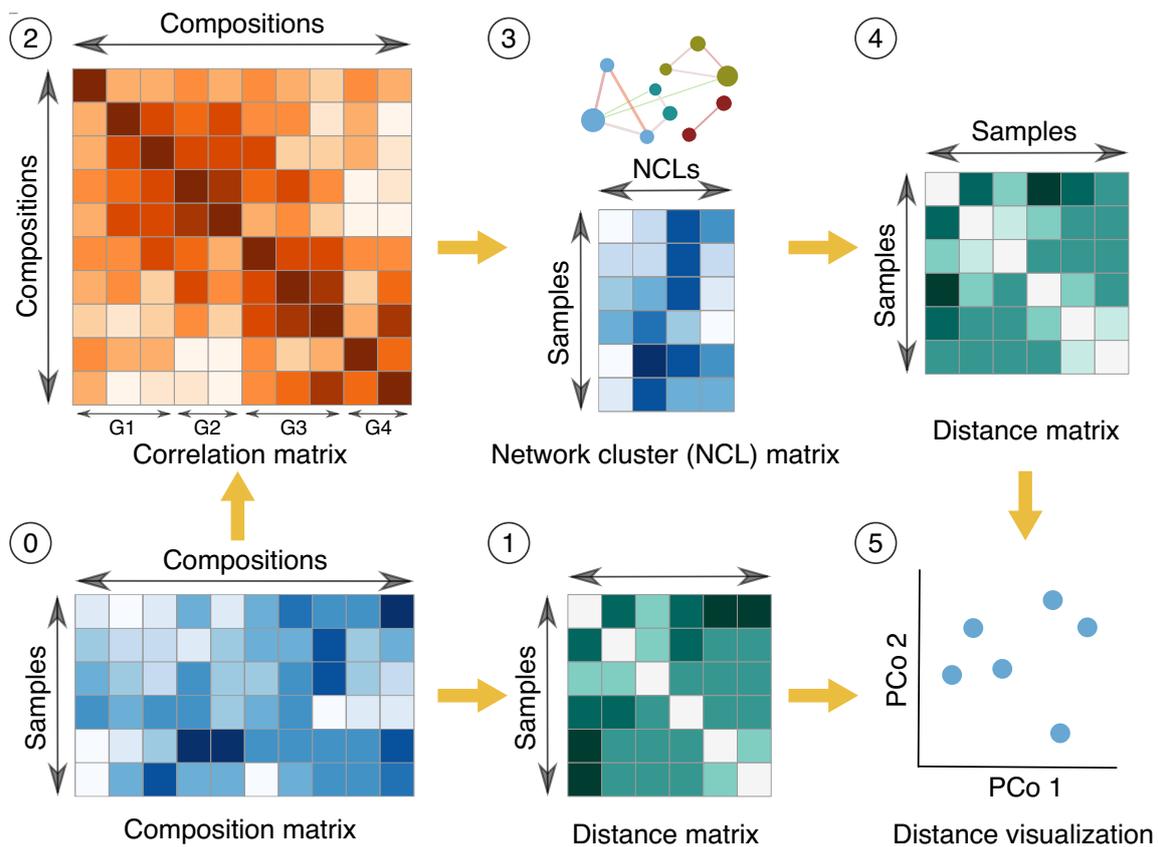
**Figure 7.5 Density and degree distribution of the global network.** Bacterial (a) and fungal (b) networks were constructed based on the co-vary of repASVs in all samples using different correlation coefficients.

### 7.3.3 R package ‘mina’: microbial community diversity and network analysis

#### 7.3.3.1 Network cluster-based community diversity analysis

The diversity analyses based on community compositions measured the differences between samples at the individual taxa level. Additionally, microbe-microbe interaction dynamics were characterised by co-occurrence networks. To provide a more comprehensive analysis of our dataset, we combined these two approaches and developed a network-based diversity index (Figure 7.6). First, we generated microbial interaction networks based on pairwise ASV correlation coefficients (Step 2 in Figure 7.6). Next, we grouped co-occurring ASVs into network clusters (NCLs) using two alternative algorithms: Markov cluster (MCL, Enright, 2002) and affiliation propagation (AP, Frey & Dueck, 2007). Afterwards, the number and RA of network nodes (i.e. ASVs), belonging to the same network cluster were aggregated as the unweighted and weighted RA of the corresponding cluster, respectively (Step 3 in Figure 7.6). Later, pairwise distances or dissimilarities between samples were calculated based on the RAs of NCLs and used for community diversity analyses, similar as before (Steps 4, 5 in Figure 7.6). The implementation of the corresponding workflow was standardized as the diversity analysis part of the R package ‘mina’ (details in Section 7.3.3.3).

(Figure on next page) **Figure 7.6 Network cluster-based community diversity analysis.** Conventionally, distance matrix (1) is calculated from composition matrix (0), in which the relative abundance of each composition in each sample was characterized, and afterwards visualized by PCoA (5). We introduce here steps (2 to 4) to first compute the correlation between pairwise compositions (2) and then cluster the closely co-varied members (3), based on which the aggregated number or relative abundance are added up. Based on the obtained NCL matrix, the distance between samples were calculated (4) and visualized (5).



(Caption on previous page) **Figure 7.6 Network cluster-based community diversity analysis.**

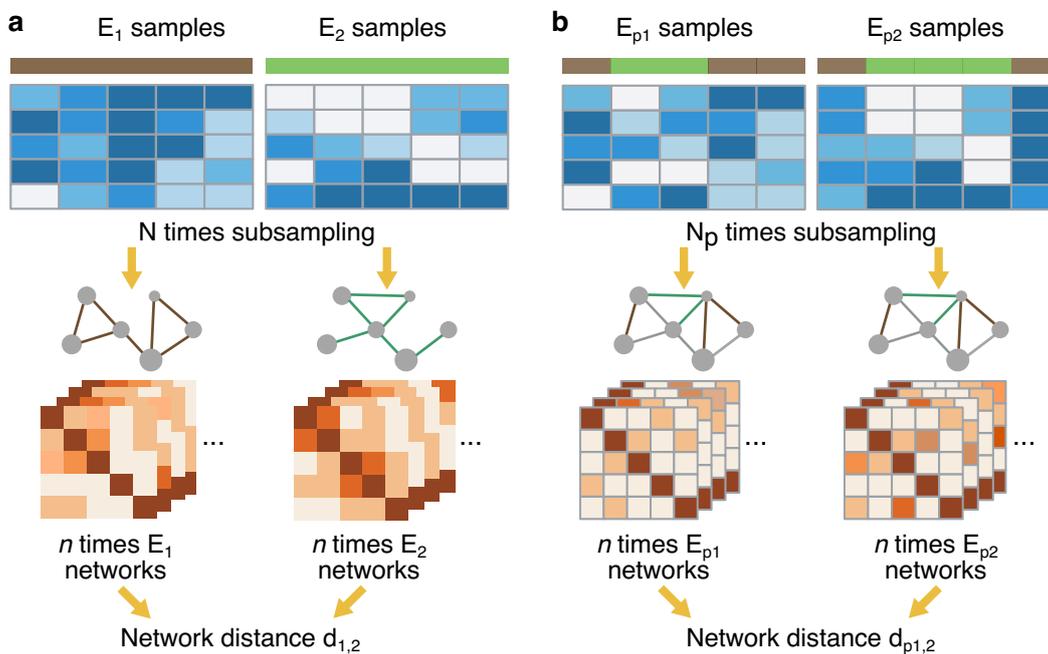
### 7.3.3.2 Network comparison and discriminative feature selection

As previously shown in section 7.3.2, network analysis has been broadly applied to the microbial profiling data to compare the higher-level dynamics between different conditions. However, the prevalent tools are mainly focused on the comparison between derived network properties (Faust *et al.*, 2015; Mandakovic *et al.*, 2018; Huang *et al.*, 2019a; Banerjee *et al.*, 2019; Mamet *et al.*, 2019; Lima *et al.*, 2020; Xiong *et al.*, 2021; Zamkovaya *et al.*, 2021), which either overwhelmingly biased by the network construction method or represent only a few perspectives of the network structure. To allow the application of alternative approach that quantify structural differences, we have developed a Monte-Carlo permutation test (Nichols & Holmes, 2002) that can be used to statistically assess the calculated network differences between microbial networks at the global and local scales (Methods and Figure 7.7).

To perform this analysis, samples were initially assigned to groups according to their real experimental set-up (referred as observed dataset later, Figure 7.7a). Afterwards, same samples were randomly labelled and formed the permuted dataset (Figure 7.7b). To avoid biases given by an uneven number of samples from different conditions when inferring networks, count matrices with the same dimensions were subsampled in parallel for the observed and permuted datasets respectively. Based on which, co-occurrence networks were constructed, and their distances were then computed. We applied here two measurements for network comparison at the global level, Spectral and Jaccard distance. For each pairwise comparison, eigenvectors of networks were calculated from their Laplacian matrix ( $\mathbf{L}$ ), which was obtained by subtracting adjacency matrix ( $\mathbf{A}$ , the correlation network matrix) by the degree matrix ( $\mathbf{D}$ ). Based on this, Euclidean distance between the first  $k$  (adjust according to the dimension of correlation matrix) eigenvalues of two Laplacian

matrices was computed as the Spectral distance between networks. Alternatively, the Jaccard distance was calculated by dividing the sum of matrix contrast by the sum of the maximum of absolute value between two networks represented by correlation matrices.

For the evaluation of the significance value, we compared the observed distances between networks from different environments or conditions to the estimated null distribution, which were obtained by calculating pair-wise distances between permuted networks (Figure 7.7b). The  $P$ -value of the test is obtained by calculating the proportion of distances between permuted networks ( $d_p$ ) greater than the observed values ( $d$ ) with formula  $P = (N_{d > d_p} + 1) / (N + 1)$  where  $N_{d > d_p}$  is the times when  $d$  is larger than  $d_p$ ; and  $N$  is the time of pairwise comparisons between the real and permuted distances. To obtain a meaningful significance evaluation, a large  $N$  is needed. To reduce the demanded computing time and resource for network inference and distance calculation, we increased the number of subsampling for both original and permutation datasets simultaneously.



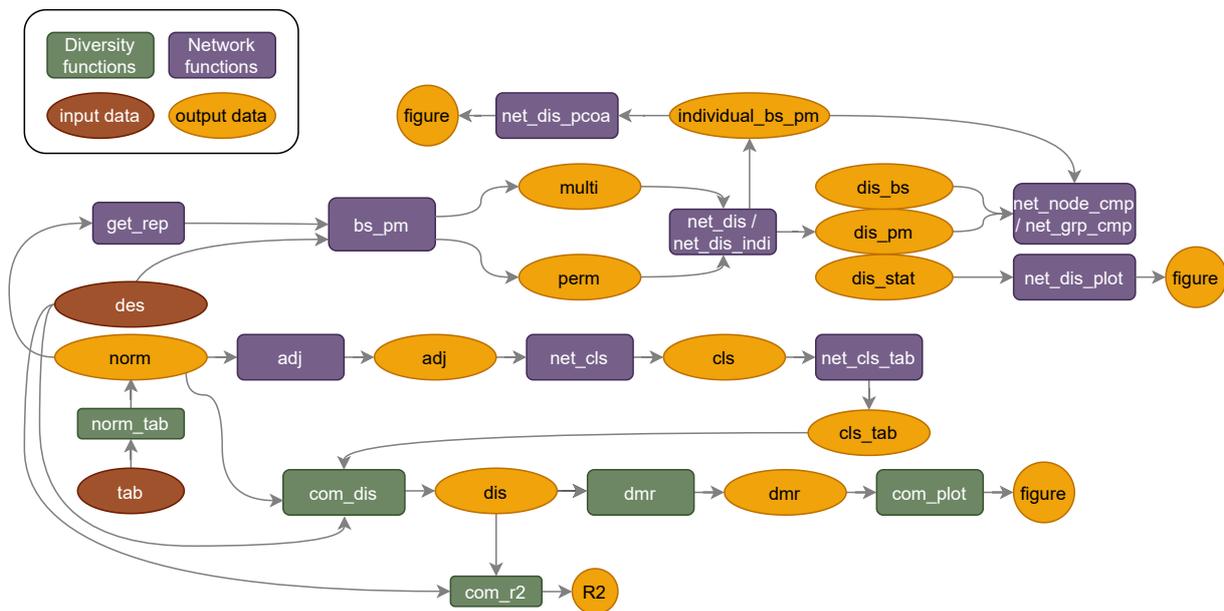
**Figure 7.7 Permutation-based network comparison.** Samples were labelled to different groups in the observed ( $E_1$  and  $E_2$ ; **a**) and permuted ( $E_{p1}$  and  $E_{p2}$ ; **b**) dataset according to their experimental setup. Sample labels in the permutation dataset were randomly assigned. For each condition,  $N$  times subsampling were applied. Afterwards, correlation networks were inferred for both observed ( $E_1$  network and  $E_2$  network) and permuted ( $E_{p1}$  network and  $E_{p2}$  network) datasets and their distance ( $d_{1,2}$  and  $d_{p1,2}$ ) were calculated respectively. The significance of network distance was then evaluated *via* comparing to the distance between randomized networks.

Furthermore, to validate the influence of a specific community member or group of members on network distance, we also generated the corresponding node-specific permutation dataset, where the metadata of samples were only randomized for specific community compositions. By comparing the distances between observed networks and partially permuted networks, the impact of these nodes could be quantified and statistically evaluated.

### 7.3.3.3 Integrating the microbial analysis framework into an R package

To ensure the repeatability and reproducibility, we integrated the above-mentioned data processing and novel diversity and network analysis methods into an R package ‘*mina*’ (for microbial

community diversity and network analysis; <https://bioconductor.org/packages/release/bioc/html/mina.html>). As indicated by the package name, the ‘*mina*’ workflow could be divided into two main parts: community diversity analysis (green functions; Figure 7.8), and network analysis (purple functions; Figure 7.8). By processing the microbial community profiling results, usually represented by a table that stores the counted number of OTU or ASV within each sample, ‘*mina*’ calculates and compares the network cluster-based diversities of microbiota. Moreover, ‘*mina*’ constructs and statistically evaluates the networks inferred for groups determined by the metadata of samples. We could apply this workflow to other non-microbiota related datasets straightforwardly, as long as a quantitative table describes the relative abundance of features in samples and a descriptive table defines the metadata of each sample available, such as the high-throughput transcriptomic data in a classical case-control set-up experiment.



**Figure 7.8 Overview of the R package ‘*mina*’.** Ellipse and rectangles indicate fields and functions of the object ‘*mina*’ respectively. Color indicates different types of functions and data attribute.

The whole package was built around a data structure object, also called ‘*mina*’, which contains all the relevant features as its slots and can be accessed for all steps in the workflow. The data object ‘*mina*’ expects the present of count data such as the commonly used OTU or ASV table to indicate the abundance of each community member in each sample (‘*tab*’; Figure 7.8). In addition, a descriptive metadata table, including the group information of samples, is also required for the most downstream analysis as input (e.g., comparison between treatments; ‘*des*’ in Figure 7.8). With defined count data ‘*tab*’ and metadata ‘*des*’, community diversity and network analyses could be performed following the workflow and the detailed vignette of the package (<https://bioconductor.org/packages/release/bioc/vignettes/mina/inst/doc/mina.html>). Alternatively, the user can also perform each step on self-defined feature matrices (e.g., ASV/OTU tables; detailed could be found in the vignette) independently without using the ‘*mina*’ object. Notably, the ‘*mina*’ package is not limited to the application in plant microbiota but also applicable to all types of microbial community studies.

### 7.3.4 The R package ‘*mina*’ provides novel insights into plant microbiota

We applied ‘*mina*’ to the integrated plant microbiota dataset and assessed the novel network cluster-based diversity analysis. Moreover, we computed network distances between different conditions (soil type, compartment, and host species) and statistically assessed their significances. To identify the most influential taxon for network differentiation, we further evaluated the contribution of each taxon individually.

#### 7.3.4.1 ‘*mina*’ diversity analysis decreased the unexplained variance of plant microbiota

Assessment of bacterial and fungal community diversity using ‘*mina*’ (detailed methods in section 7.5.5.1) showed a considerable decrease of unexplained variance compared to the conventional composition-based method (14-34% decrease for bacteria and 8-31% for fungi; Table 7.2). Particularly, networks inferred using SparCC and clustered using Affinity Propagation outperformed all others, with a decrease of 20% unexplained variance for bacteria and 15% for fungi (Table 7.2). At the same time, the separation of samples according to biological factors (e.g., compartment, host species) was more evident using network cluster-based diversity indices (Figure S7.4), suggesting an increased signal-to-noise ratio.

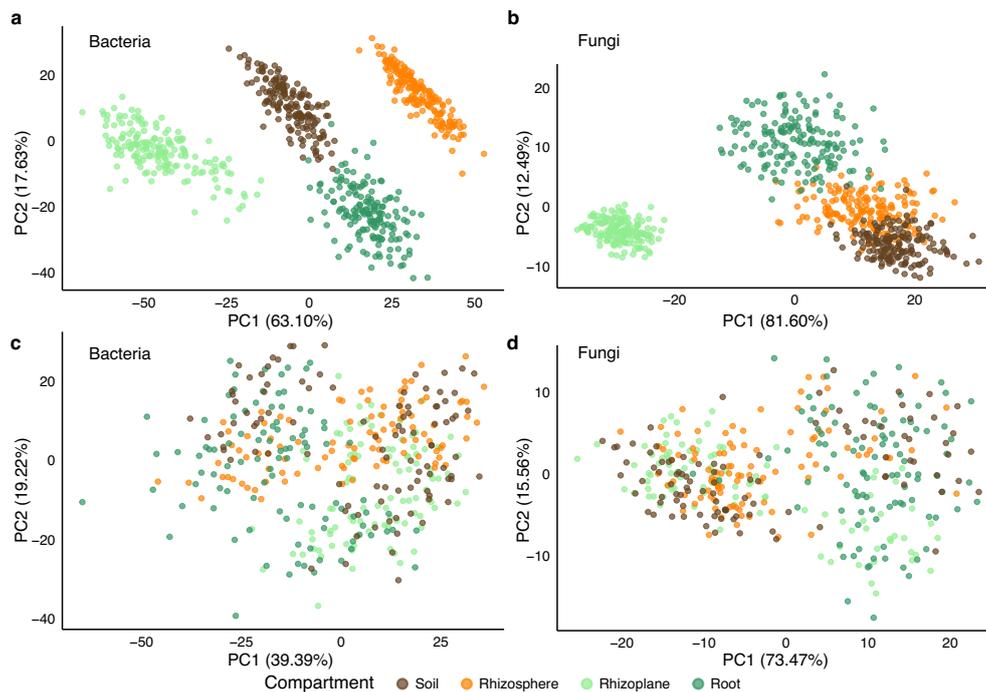
**Table 7.2 Unexplained variance of ASV- and NCL-based diversity analyses.** P = 0 was used for AP clustering. Column No. indicates the number of ASVs or network clusters applied for diversity analysis; column % is the unexplained variance ratio.

Network	Clustering	Bacteria		Fungi	
		No.	%	No.	%
/	ASVs	42,060	51.3	9337	46.0
	RepASVs	2047	43.1	370	39.0
Pearson	MCL	45	15.7	36	26.2
	AP	191	29.4	58	31.5
Spearman	MCL	4	10.6	4	7.6
	AP	81	23.1	21	24.1
SparCC	MCL	3	9.3	2	10.7
	AP	78	23.1	29	23.6

#### 7.3.4.2 Network dynamics of plant microbiota between different compartments

We further compared the microbial networks that we inferred for different compartment groups for bacteria and fungi, respectively, using ‘*mina*’. Spectral distances between networks were visualized with PCA (Figure 7.9). As expected, networks of different groups were separated from each other for both kingdoms. For bacteria, networks of root and soil are closer to each other, indicating similar interactions between microbes in these two compartments. For fungi, the rhizosphere and soil networks overlap, showing a similar connection topology. Particularly, rhizoplane samples derived distinctive networks from others in both kingdoms, which could be due to the single host species of this compartment. In the corresponding permutation dataset, where the samples were randomly labelled, the dividing patterns between conditions were disappeared. Although the subsampling process was randomized and therefore the computation of the network distance process was non-deterministic, highly consistent results were obtained between different runs when repeating the analyses. We also performed the network comparison step with higher number subsampling process (99 times for both observed and permuted datasets, Figure S7.5), and the

results showed a similar separation as observed before (Figure 7.9), indicating that the small number subsampling is already capable of capturing the network dynamics.



**Figure 7.9 Bootstrap-permutation network analysis of plant-associated microbiota.** Observed (a bacteria, b fungi) and permuted (c bacteria, d fungi) networks are represented by points. Subsampling time  $n = 33$  for both observed and permuted datasets for both kingdoms.

Furthermore, we examined the mean pairwise distance between compartments and their significance. Surprisingly, the distance between some of the comparisons was not significant (Figure S7.6), such as between root and rhizosphere networks of bacteria (average  $d = 47$ ,  $P = 0.059$ ) and fungi (average  $d = 19$ ,  $P = 0.14$ ), indicating the stringency of the  $F$ -test. Besides, as observed in Figure 7.9, networks of fungal rhizosphere and soil are close to each other, and their corresponding distance was found to be non-significant ( $d = 14$ ,  $P = 0.20$ ).

### 7.3.5 Assembly of alga- and root-associated microbiota in Cologne Agricultural Soil

To focus on the impact of compartment and host species on microbial network structure variation, we applied these methods to a subset of samples derived from studies conducted in the greenhouse using the same soil type (Cologne Agricultural Soil; CAS), where environmental factors were kept stable. We compared 854 plant root- (including *A. thaliana* and *L. japonicus*) and soil-borne alga- (*C. reinhardtii*) associated communities (Table. 7.3).

#### 7.3.5.1 CAS-associated microbial community diversities

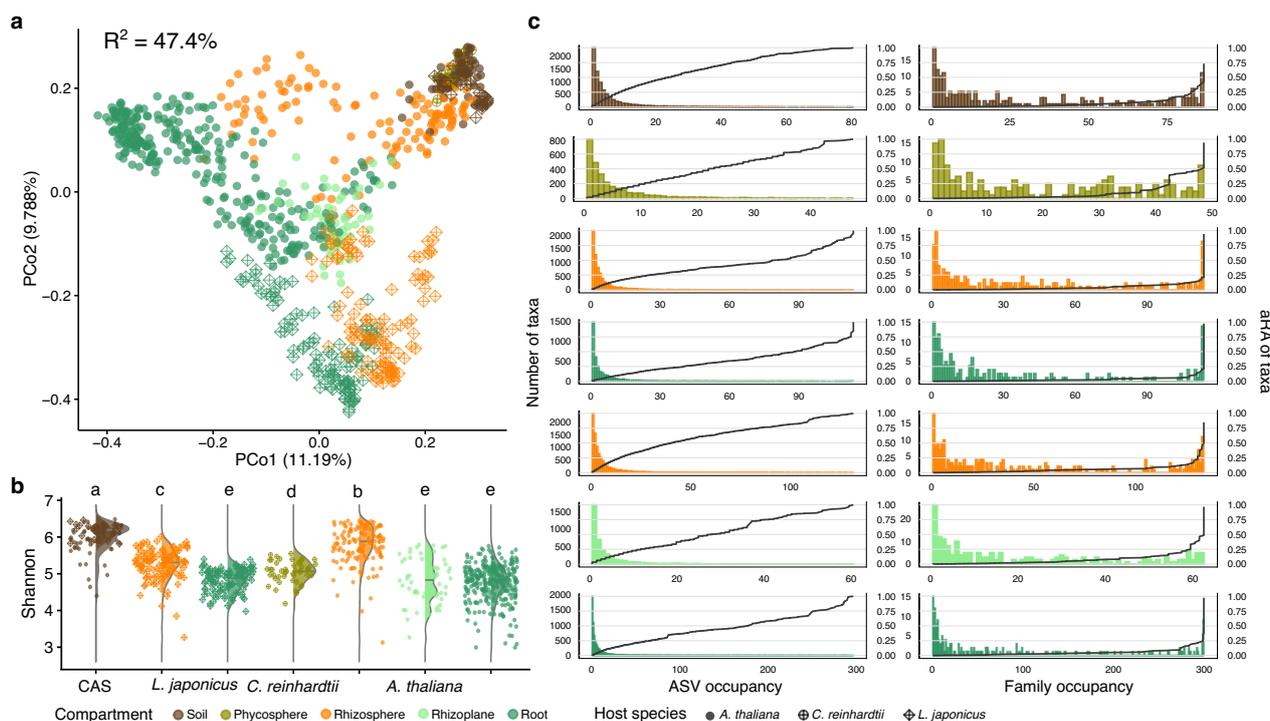
We first investigated the diversities of CAS-derived root- and alga-associated microbial communities (Figure 7.10) and observed a clear separation of samples from different compartments, as well as a clustering of samples derived from the same host species (Figure 7.10a), indicating the diversification of microbiota structures associated with different photosynthetic organisms. Phycosphere samples associated with *C. reinhardtii* clustered closer to soil compared to rhizosphere compartments, indicating a lower host effect for the alga than for the other two plant hosts. Furthermore, we assessed the unexplained variance of diversities between samples and at the ASV

level, the  $R^2$  value was up to 47.4%. Moreover, Shannon indices were compared between samples and the same decreasing from soil to rhizosphere to root for both *L. japonicus* (referred as *Lj* later) and *A. thaliana* (referred as *At* henceforth) was observed (Figure 7.10b) as reported earlier (Section 7.3.1.1). Notably, a slightly higher ( $P < 0.001$ ) alpha diversity of phycosphere samples were identified compared to root compartments of *Lj* and *At*, consistent with the observation in Chapter 5 (Figure 5.2).

**Table 7.3** Sample number of CAS-derived conditions.

Hose Species	Compartment	Sample
	Soil	86
<i>Chlamydomonas reinhardtii</i>	Phycosphere	48
<i>Lotus japonicus</i>	Rhizosphere	113
	Root	113
<i>Arabidopsis thaliana</i>	Rhizosphere	133
	Rhizoplane	62
	Root	299

We examined the occupancy distribution of community members under each condition and confirmed a similar pattern at all taxonomic ranks as reported earlier (Section 7.3.1.2). Particularly, we compared the occupancy and the accumulated RA distribution of taxa at the ASV and family levels (Figure 7.10c). At the ASV level, most of the compositions were identified in very few samples, and they account for a relatively small proportion of aRA for the whole community, usually less than 25% (Figure 7.10c). Consistent with what we described earlier, at the higher taxonomic level, a group of core taxa emerged for each condition. Notably, these taxa with high occupancy account for a large proportion of aRA, indicated by the steep increase of the aRA curve in the end (Figure 7.10c).



(Caption on next page) **Figure 7.10** Community diversity and taxa distribution of CAS microbiota.

(Figure on previous page) **Figure 7.10 Community diversity and taxa distribution of CAS microbiota.** (a) PCoA of beta-diversity between samples.  $R^2$  is the ratio of community diversity variance that cannot be explained by compartment, soil batch, host species, and host genotype; (b) Alpha-diversity of samples from each condition calculated based on all ASVs; (c) occupancy and accumulated RA at different taxonomic ranks, i.e., ASVs and families, in each condition. From upper to lower, conditions are soil; phycosphere; rhizosphere and root of *L. japonicus*; rhizosphere, rhizoplane, and root of *A. thaliana*. Shannon diversities between conditions were tested with the Wilcoxon test with FDR correction.

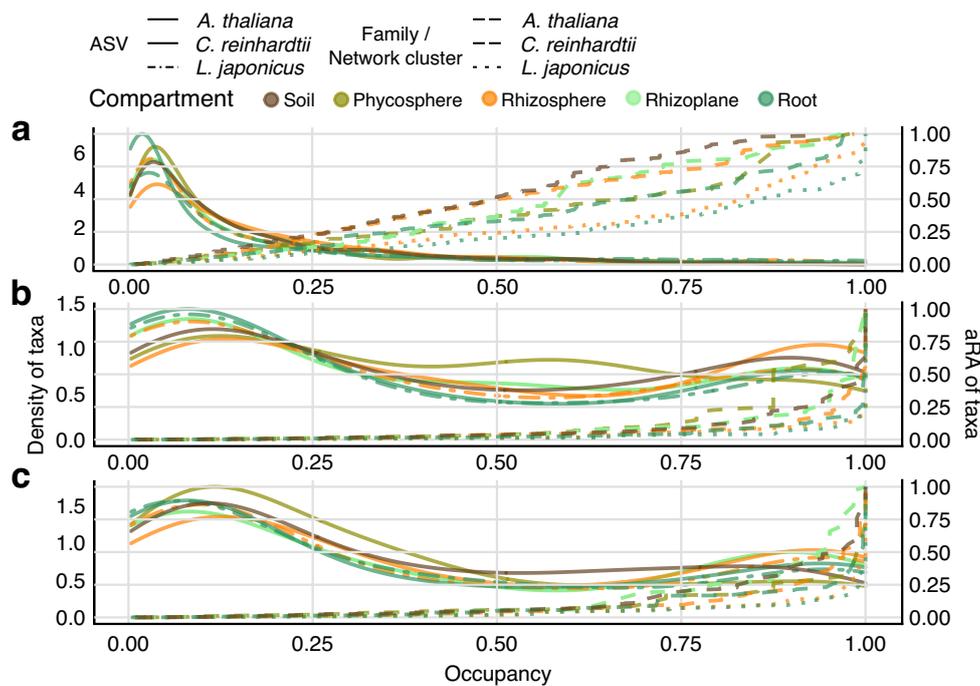
#### 7.3.5.2 A reduced set of repASVs recapitulates whole-community microbiota diversity patterns

Previously, we have shown that repASVs of the integrated dataset that identified using occupancy and RA criteria were able to recapitulate most of the community diversities (Section 7.3.1.3). To assess whether these repASVs, derived from the whole dataset, maintain its representative for the CAS dataset, we checked the aRA, the Shannon diversity, and community dissimilarities captured by these predefined repASVs (Figure S7.7). In most samples, more than half of the entire community was represented by the repASVs, and the proportion increased to on average 80.94% and 81.40% in *Lj* and *At* roots, respectively (Figure S7.7a). Similar patterns were observed when calculating the alpha-diversities using only repASVs (Figure S7.7b) comparing to the whole communities (Figure 7.10b). Notably, the diversity differences between soil and other root-associated compartments are decreased, suggesting the non-repASVs with relatively low RA and occupancy contributed to the high alpha-diversity of soil microbial communities. Moreover, among all the conditions, the phycosphere possessed the lowest aRA and alpha diversity, as well as the largest decrease of alpha diversity when leaving out rare ASVs. This indicates a different community shift compared to other plant-associated compartments and/or the bias caused by the number of samples from the specific condition included in the dataset when identifying repASVs (Table 7.1). We also compared community beta-diversity basing on repASVs (Figure S7.7c) and obtained highly consistent patterns as before (Figure 7.10a). Samples from different compartments were separated, and different host species drove the community shift distinctively. Meanwhile, the  $R^2$  decreased slightly (from 47.4% to 40.7%) when comparing sample dissimilarities based only on the repASVs, indicating the reduced noise when excluding rare taxa for diversity analysis.

#### 7.3.5.3 CAS derived microbial community diversity analysis using ‘mina’

We surveyed the occupancy and accumulated RA of representative members, i.e. repASVs, only (Figure 7.11). At the ASV level, a substantial portion of aRA was contributed by prevalent repASVs within each condition, indicated by the increasing slope of the cumulative aRA curve when occupancy increased. For example, in rhizosphere and root samples of *Lj*, more than half of the aRA were contributed by repASVs present in more than 75% of the samples (Figure 7.11a). At the family level, as expected, we observe the enrichment of prevalent taxa, where the number of taxa with high occupancy was increased, and these taxa accounted for a large portion of aRA, shown by the cumulative aRA curve (Figure 7.11b). a substantial portion of aRA was also contributed by more prevalent repASVs within each condition at the ASV level. For example, in rhizosphere and root samples of *Lj*, more than half of the aRA were contributed by repASVs present in more than 75% of the samples (Figure 7.11a).

(Figure on next page) **Figure 7.11 Occupancy and the corresponding cumulated aRA distribution of taxa and network clusters.** Scaled density distribution of CAS, *Cr*-, *At*- and *Lj*-associated microbiota at the ASV (a), family (b), and network clusters (c) levels of repASVs.



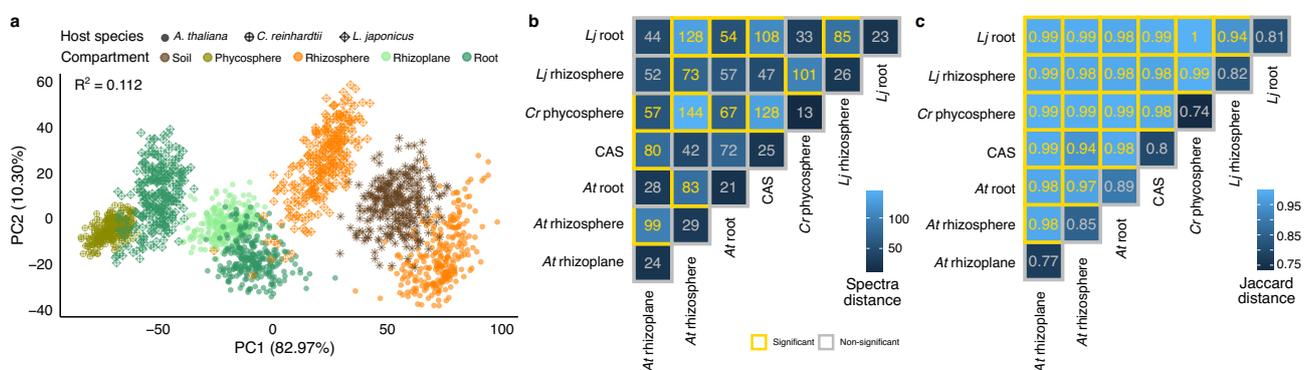
(Caption on next page) **Figure 7.11 Occupancy and the corresponding cumulated aRA distribution of taxa and network clusters.**

Moreover, a CAS-associated network was constructed using repASVs and communities were then compared based on the derived network clusters by ‘*mina*’. Similar as we observed in [Section 7.3.3.2](#), the unexplained variance ratio between different conditions was decreased for NCL-based diversity analysis. NCLs clustered by AP from the Spearman correlation network outperformed others by reducing the most  $R^2$  (from 0.474 to 0.240, [Table S7.2](#)) while maintaining the dissimilarities patterns ([Figure S7.8](#)). Moreover, we surveyed the occupancy and accumulated RA of those network clusters ([Figure 7.11c](#)) and compared them with the distribution of community compositions at both the ASV ([Figure 7.11a](#)) and family ranks ([Figure 7.11b](#)). A comparable distribution pattern of occupancy and aRA among conditions was observed for network clusters, where a precipitate increase of aRA was observed, contributed by the clusters of microbes with higher occupancy ([Figure 7.11c](#)).

#### 7.3.5.4 Network dynamics between CAS-associated conditions

We compared networks of different hosts and compartments using ‘*mina*’ network analysis workflow ([Figure 7.12](#)). Assessment of network Spectra distances ([Figure 7.12a](#)) revealed a separation along the first principal component (82.97% of the variance) between microbial networks derived from soil-based microhabitats (unplanted soil and rhizosphere) and host epi- and endophytic compartments (rhizoplane, root endosphere and algal phycosphere). The observed clustering of network groups was collapsed in the permutation dataset, indicating the differentiations between compared conditions were significant ([Figure S7.9](#)). Despite the similarities of community structure between phycosphere and rhizosphere ([Figure 7.10](#)), when comparing the networks, the phycosphere were more similar to the root samples, especially to the *Lj* root ([Figure 7.12a, b](#)). This suggests that close associations with a photosynthetic host lead to significant changes in microbial co-occurrence patterns, which could be driven by higher concentrations of diverse organic carbon compounds or by direct interactions with the host.

Within each host species, network dynamics between compartments varied differently. For *At*, the rhizosphere and root compartments drove changes to different directions in the first two principal components space of the Spectra distance. Surprisingly, the distance between rhizosphere and root ( $d = 83$ ;  $P = 0.010$ ) is larger than the distance between soil and root networks ( $d = 72$ ;  $P = 0.13$ ). While for *Lj*, root compartment shifted the networks from soil ( $d = 108$ ;  $P = 0.0016$ ) in the same direction as the rhizosphere networks ( $d = 47$ ;  $P = 0.20$ ). To inspect this unexpected non-significant distance given the compared networks clustered separately in the PCA (Figure 7.12a), we examined the distribution of the distance between root and soil in the observed dataset and permuted datasets, respectively, for both plant hosts (Figure S7.10). Distances between *At* root and CAS shown a clear separation with a slight overlap in the distribution (Figure S7.10a), caused the non-significant P value due to the rigid statistical test. For *Lj*, the observed and permuted distances between root and CAS were merely intersected, therefore leading to the significance of their network divergence (Figure S7.10b).



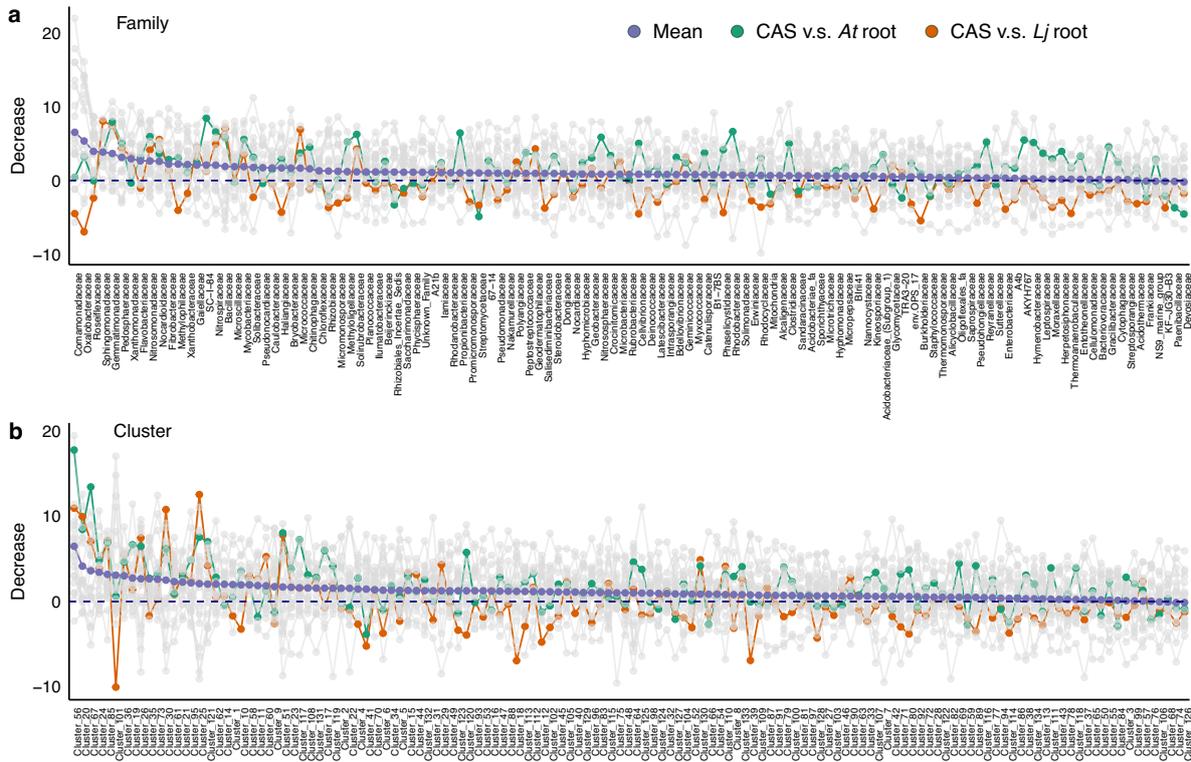
**Figure 7.12 Network distances between different CAS-associated conditions.** (a) PCA of network Spectra distance shows a clear separation between different compartments and host species. (b) Average of Spectra distance indicating a similar result as observed in (a). (c) Average of Jaccard distance between networks from different compartments and host species.  $R^2$  indicates the variance of Spectra distances between networks, which cannot be explained by compartment and host species.

We also compared the Jaccard distances between networks of different CAS-associated conditions (Figures 7.12c). Even though all the network distances were significant, due to the highly similar distances between different inter-condition comparisons, this approach failed to characterize the network differences effectively. For example, when comparing root and rhizosphere to rhizoplane of *At*, both compartments have an average distance of 0.98, while they were very different demonstrated by Spectra distance ( $d = 28$  and 99 for root and rhizosphere, respectively; Figure 7.12b). Despite the different sensitivities between two network distance measurements, similar patterns were found between them. For example, when comparing between networks of soil and rhizosphere, *Lj* shown a slightly stronger selection effect compared to *At*, with an average higher distance differences of 5 and 0.5 for Spectral and Jaccard distances, respectively (Figure 7.12b, c).

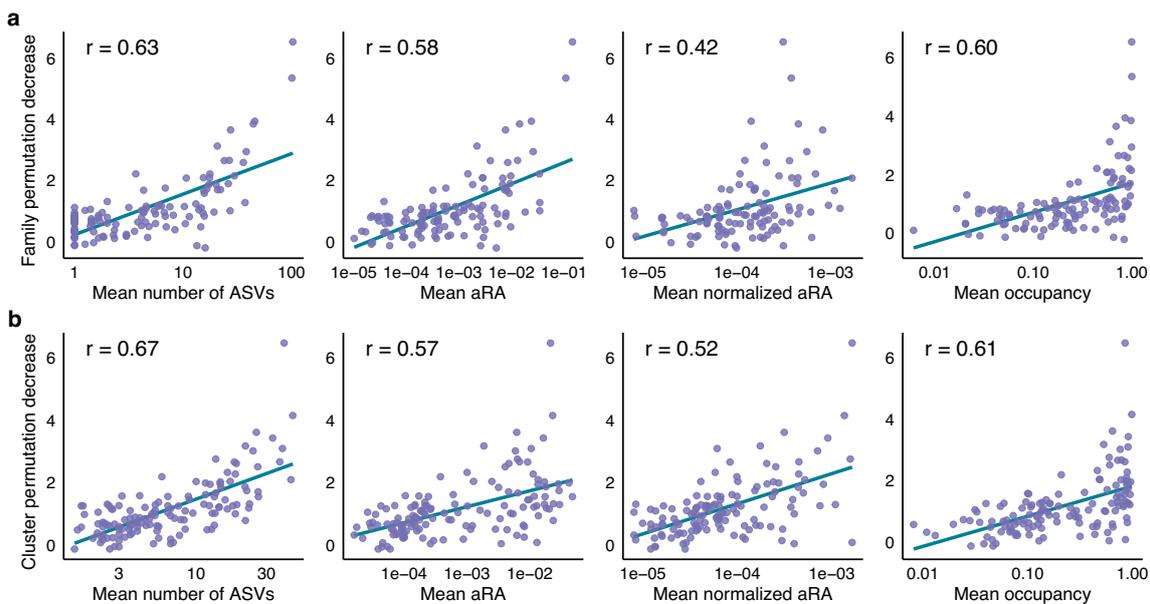
### 7.3.5.5 Distinctive features contribute to CAS-associated network dynamics

To assess the contribution of each group of nodes, i.e., family or network cluster, to the observed differences between networks, we applied the partial permutation method as described in Section 7.3.3.2. Distance changes between networks constructed from the observed and partially permuted dataset were quantified (Figure 7.13). As demonstrated by the decrease of each permutation, most

families and network clusters contributed inconsequentially, with an average of around 0. Moreover, the effect of each group on the distance between different conditions was various, such as the highlighted comparisons between soil and root from both host species (Figure 7.13).



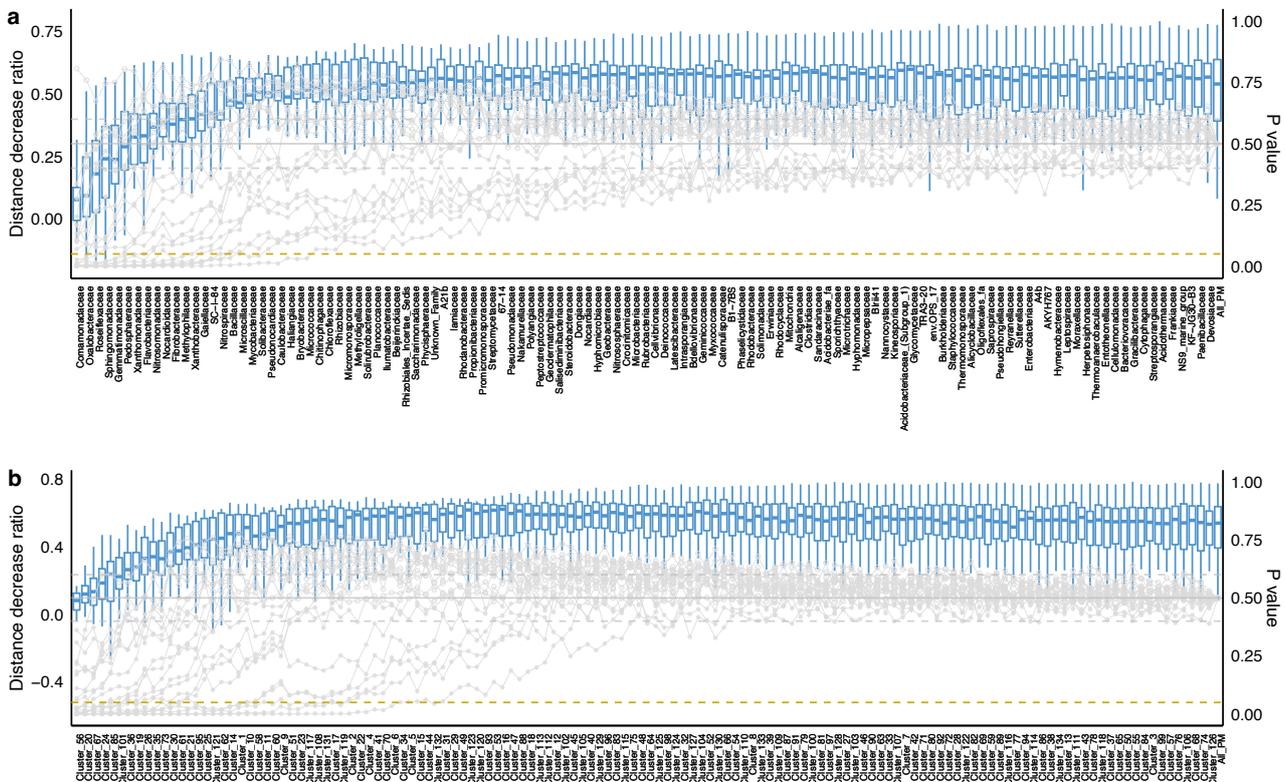
**Figure 7.13** Distance decrease of each permutation group compared to the actual network distance. Distance changes caused by each permuted family (a) and network cluster (b) were sorted by the average decrease of distance between different comparisons (shown in purple). The comparison between CAS and *At* root was shown in green, CAS and *Lj* root was in orange and other comparisons in grey.



**Figure 7.14** Correlation between partially permutation distance change and other features. Y-axis shows the average distance change for each family (a) or network cluster (b) permutation, compared to the distance between bootstrap networks inferred from the original dataset. For each panel, from left to right, the x-axis shows the mean value of features among compared conditions, including the number of present ASVs

in the group, the aRA of the group, the aRA normalized by the number of present ASVs, and the occupancy of the groups. The teal line is the linear regression of the data points. Pearson correlations are computed, and all coefficients here are significant ( $P < 0.001$ ).

To investigate the factors that might influence the change of Spectral distance when permutating each group, we calculated their correlation with features including the group size (the number of ASVs contained in the group), aRA, normalized aRA, and occupancy (Figure 7.14). For the mean distance change among all comparisons, all tested features were significantly correlated when permutating the group ( $P < 0.001$ ). Stronger coefficients were found for most features when ASVs were grouped by network cluster than taxonomy, except for the aRA (0.58 for families and 0.57 for network clusters, respectively). The number of ASVs in the group had the strongest correlation with changes in network distances, followed by node occupancy.



**Figure 7.15 Network distance change for all inter-condition comparisons.** Nodes grouped by family (a) and network cluster (b) are shown here. Distance change of cumulative group permutation datasets when nodes are grouped by network clusters. Boxplot shows the distance decrease ratio compared to the average distance calculated by bootstrapping the original dataset. The line plot shows the P-value calculated from the permutation test. The shape of points indicates the significance of the Wilcoxon test with FDR correction (full points are significant comparisons and empty points are non-significant ones). Dash line in gold shows the cutoff of significance ( $P = 0.05$ ). The solid grey line indicates 0.5 for the right side of the y-axis, and the dash lines below and above are 0.45 and 0.55, respectively.

To investigate how many permutations of families or network clusters is needed to diminish the separation between networks, we permuted the groups of nodes accumulatively, following the descending order of mean distance change when they were permuted individually (Figure 7.15). For all tested comparisons, and both grouping approaches, we observed an increasing distance change ratio in the beginning of accumulating permutation, reaching saturation after a relatively

small number of groups. The  $P$  values calculated by the Monte Carlo permutation test, i.e., counting the times when permutation distances are larger, fluctuated around 0.5 after cumulative permutation of groups, indicating the process reached the randomization (Figure 7.15). We examined the distance decrease and changes in significance of all 21 pairwise comparisons between seven CAS-associated conditions. To quantify the contribution of groups to the network separation, we identified the changing point, at which the majority (12 out of 21) comparisons became not significant anymore testing by the Wilcoxon (FDR correction) for both taxonomic and network cluster grouping. Chitinophagaceae was identified as the distinctive point for family-based partial permutation, ranking 26 among all groups (Figure 7.15a). For the network clusters, also the 26<sup>th</sup> group (Cluster\_51 here) was identified as the changing point (Figure 7.15b). Although the same number of groups was found before the changing point, nodes included in the network clusters ( $N = 736$ ) were much less than families ( $N = 985$ ). The saturation of distance decrease ratio for both groups at the determined changing point indicates the trivial effect of following groups, therefore, groups ranked before the changing point were extracted as distinctive groups. We confirmed the majority contribution of distinctive groups to the network separation by plotting the PCA of network distances when cumulatively permutating these groups, and networks of different groups were dispersed randomly (Figure S7.11a, b). Moreover, when comparing the networks basing only on these distinctive groups, similar differentiation patterns were observed, suggesting similar distinctive power of the subset groups as the full dataset (Figure S7.11c, d).

## 7.4 Discussion

In this chapter, we integrated plant-associated microbial community samples from diverse environments and showed the representative bacteria and fungi recapitulate the microbiota diversity. We introduced higher-order features to community diversity analysis to improve the comparison and developed the network evaluation to assist the understanding of system dynamics.

### 7.4.1 Core and representative members of plant microbiota

By characterizing the community structures, specifically the abundance of each composition, of plant microbiota from conditions including different compartments, host species, and soil types, we have shown here that a well-defined core microbiota was absent at the ASV level for both bacteria and fungi. This consistent long-normal distribution of species abundances between micro- and macro-ecosystems indicates that macroecological laws could be used to quantitatively describe the microbial ecosystem (Grilli, 2020). Nonetheless, common taxa were observed at the higher taxonomic ranks, similar to the observation of core root microbiome across plant phyla along a natural soil chronosequence (Yeoh *et al.*, 2017). Taken together the phylogenetically conserved bacterial metabolism (Isobe *et al.*, 2019), and functional overlap of plant microbiome across compartments and species (Bai *et al.*, 2015; Durán *et al.*, 2021; Wippel *et al.*, 2021), we hypothesize that the assembly of plant microbiota was driven by the functional capacity across various soil type and host species.

To compensate for the lack of core ASVs, we applied Procrustes Analysis to rationalize the extraction of influential members. The selected compositions confirmed the previously described abundance-occupancy distribution (Shade & Stopnisek, 2019) in turn. We further demonstrate that

they recapitulate the majority of aggregated relative abundance (aRA), community diversity and network dynamics of the whole community system, therefore, we proposed the term representative ASVs (repASVs). Moreover, we showed that these repASVs derived from the integrated large datasets maintained their representative when applied for community diversity analysis of a specific subset of samples, such as the CAS-associated conditions we assessed in this chapter. Although the bias towards the conditions that are well represented in our datasets needs to be evaluated additionally when future datasets are integrated, this repASVs set has proven to be able to be used as a base representative composition set for plant microbiota research, especially for those studies with limited sample numbers, assisting the cross-reference between different ecological surveys.

#### 7.4.2 Community diversity analysis based on higher-order features

There is a long-lasting debate between using OTU or ASV as the unit for microbial community diversity analysis. By clustering highly similar sequences (with >97% identity usually) to one unit, OTU clustering decreases the random noise introduced by sequencing and data processing. However, the fact that the clustering is a nondeterministic process makes the comparison between studies impossible. For ASVs, since each unit corresponds to one specific sequence, it could be connected to a certain biological entity and this makes it possible to compare between datasets. However, the common presence of polymorphism in some taxa (Moreno *et al.*, 2002; Větrovský & Baldrian, 2013) introduces more artefactual community members and inflation in the alpha-diversity in the analysis of community structure when using ASVs.

To take advantage of both approaches, we introduced the previously described microbial guild concept, which considering the co-varied microbes as a unit, into plant microbiota. The application of this approach in human gut microbiome research has already proven to be capable of reducing the dimension and sparsity, and assisting the identification of candidate community members that related to host health (Wu *et al.*, 2021). Under this scheme, community members were represented by ASVs, and sequences were clustered according to their covariation or cooccurrence instead of their similarity. This approach provides similar yet broader advantages against random noise compared to OTU clustering. In addition, since the different copy 16S rRNA sequences from the same strain would always co-occur, this approach also reduces the bias introduced by the presence of polymorphic copies of the amplicon sequences. Subsequently, the inferred network clusters could be interpreted as microbial guilds and thus provide novel ecological insights. Implementing the developed network cluster-based diversity analysis on the integrated plant microbiota dataset showed a decreased in the unexplained variance, indicating the higher signal-to-noise ratio of this method compared to OTU- and ASV-based approaches.

#### 7.4.3 Network comparison at the local and global scales

One of the challenges of using co-occurrence networks for the comparative study of microbial communities resides in identifying statistically significant differences between networks from different environments. Here, we established a comparison workflow based on the Spectral distance along with a Monte Carlo permutation significance test. We computed the distances between complete adjacency matrices that described the networks and then compared those distance matrices using the tools that were broadly implemented in diversity analysis to demonstrate the comparison. PCA and PERMANOVA results (Figure 7.10, 7.12) indicated we were able to differentiate the

networks and evaluate the variances that could be explained by known factors, i.e., meta-data of microbiota samples, with our workflow.

After identifying the differently connected networks, we implemented a similar concept to compare those networks at the local scale to discover the distinctive features contributing to the network distances. To reduce the number of examined feature, community compositions were grouped according to their taxonomy (at the family level) and ecological interaction (correlation network clusters). We show that network cluster, which represent microbial guild, is a better unit for comparison between microbial communities than approach based on microbial taxa (Figure 7.15). Therefore, for the previous network topology comparison studies that based on OTU, a taxonomic unit (Faust *et al.*, 2015; Mandakovic *et al.*, 2018; Banerjee *et al.*, 2019; Mamet *et al.*, 2019; Xiong *et al.*, 2021), we would expect an increased sensitivity when the network cluster based analysis are applied. We also identified the distinctive groups that contribute to network differentiations between CAS-associated conditions. Analysis of correlations between group features, including network distance change, aRA, and occupancy (Figure 7.14) indicate that the more abundant and prevalent groups also have a higher influence on the network differentiation. This could further guide the future experimental design, particularly the selecting of community compositions to assemble SynComs for the validation of data-driven hypothesis.

To summarize, we described an R package ‘*mina*’ for microbial community diversity and network analysis. By applying a network cluster-based diversity analysis on the plant microbiota dataset we integrated, we show that this new method decreases the noise while maintain the variances between samples from different environments. Moreover, using a novel statistical framework for network comparison, we show that we can identify significant differences between ecological networks and identify relevant community features. We observed different clustering patterns of plant microbiota dynamics when paralleling microbial diversity and network analysis results. By applying the same variance analysis and visualization approach developed in ‘*mina*’ workflow, we were able to demonstrate those differences more intuitively, comparing to conventional microbial network comparison methods.

## 7.5 Materials and Methods

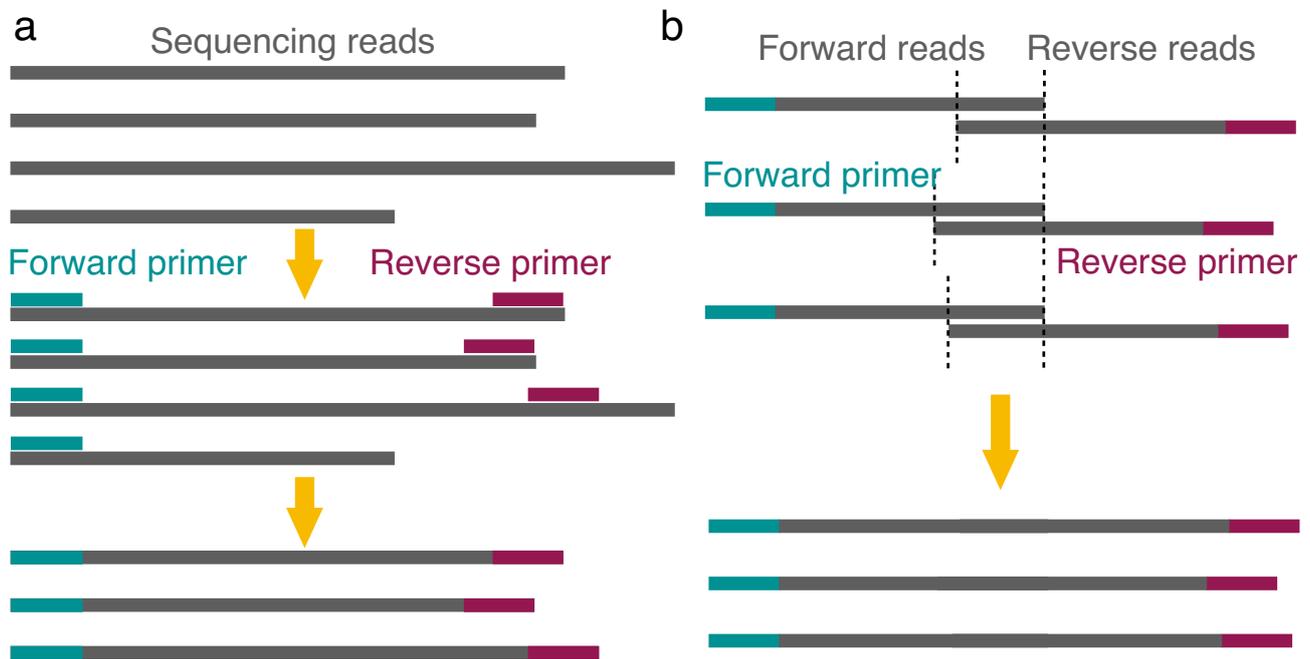
### 7.5.1 Pre-processing of microbial community profiling data

As mentioned in Chapter 1, for microbial community surveys, the available technology developed from pyrosequencing represented by Roche 454, which was widely applied in the early 2010s, to Illumina sequencers, namely MiSeq and HiSeq. The integrated dataset used in this Chapter contains samples from various studies (Zgadzaj *et al.*, 2016; Durán *et al.*, 2018; Thiergart *et al.*, 2019, 2020; Harbort *et al.*, 2020; Wippel *et al.*, 2021) amplified by the same primer set. To limit the bias between different studies and benefit cross-referencing, a standardized data processing pipeline was developed and implemented ([https://github.com/Guan06/DADA2\\_pipeline](https://github.com/Guan06/DADA2_pipeline)) to process the data the same way so that the biological variances was maintained.

#### 7.5.1.1 Bacterial 16S rRNA gene sequencing data processing

The bacterial community data processing pipeline is mainly based on DADA2 (v1.12.1) (Callahan *et al.*, 2016). For Roche 454 pyrosequencing, single-end reads are generated with lengths ranging up

to 700 bp and thus may contain sequences outside the examined region. Therefore, after demultiplexing, the forward and reverse primers were mapped to the sequencing reads and trimmed the part outside the region (Figure 7.16a). Afterwards, reads are filtered by length, longer than 200bp and shorter than 540bp, and quality, with parameter “maxN=0, maxEE = 2, truncQ=2”.

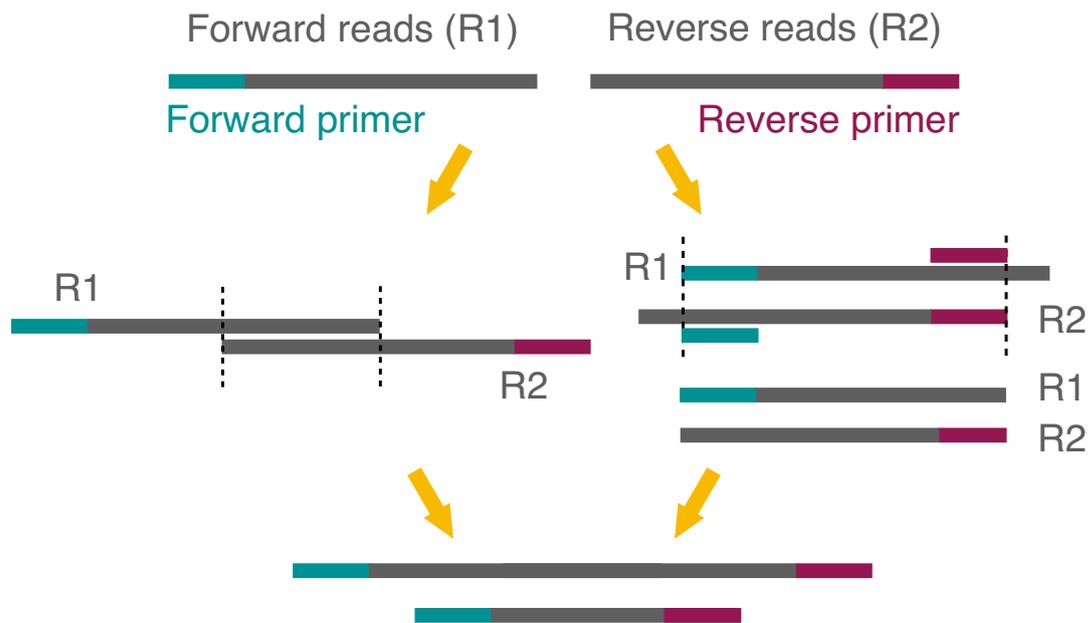


**Figure 7.16 Pre-processing of bacterial amplicon sequencing data.** Reads sequenced by Roche 454 (a) and Illumina MiSeq (b) are shown here. Regions in between dash lines in (b) are the overlap between forward and reverse reads.

Later on, the Illumina sequencing platform was widely used because of the higher throughput and lower cost. Pair-end sequencing was conducted to compensate for the shorter sequencing length compared to Roche 454 (Figure 7.16b). Raw sequencing reads were truncated to 260 bp for the forward reads, and 240 bp for the reverse reads and filtered with the following parameters: “maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE”. Subsequently, error rates were inferred from filtered reads until convergence or exceeding a maximum consistent number of 20 and corrected. Sequences of the amplified region were then obtained by merging the forward and reverse sequences together according to their overlap.

#### 7.5.1.2 Eukaryotic microbial profiling data processing

Meanwhile, for the eukaryotic microbes, we amplified the ITS region for the profiling sequence. Unlike the 16S region, the length of ITS is highly variable between different taxa. By mapping the forward and reverse primers to the sequencing reads, the length of the ITS region could be estimated (Figure 7.17). When the ITS region is shorter than the sequencing length, reverse primer sequence could be identified in forwards reads and vice versa. For the ITS regions longer than the sequencing length, an overlap sequence could be identified between forward and reverse reads and used for merging the two single-end reads. For the cases that the amplified region is shorter than sequencing length, ITS regions were kept after trimming the forward and reverse reads according to the aligned coordinates of primers.



**Figure 7.17 Pre-processing of fungal ITS reads by pair-end sequencing.** MiSeq sequencer was applied for studies mentioned in the thesis. Dash lines indicate the overlap between forward and reverse reads.

### 7.5.2 Sequence and diversity analysis for microbial communities

Natural community profiling is the isolation-independent strategy where the total DNA of plant microbiota were extracted, and specific amplicons for each microbial kingdom was sequenced. Typically, natural root-associated microbial communities are highly diverse and thus need deeper sequencing depth to characterize the composition. After pre-processing, sequences were clustered into OTUs or used as ASVs directly. In this thesis, all analyses were implemented at the ASV level, and some have OTUs as additional results to assist the cross-reference. By mapping the reads to these composition reference sequences, the aligned reads number of each community member in each sample was obtained and formatted as a matrix, named OTU or ASV table. For bacteria and fungi, sequences were mapped to *silva* (v138, [Quast et al., 2012](#)) and *unite* (release 04.02.2020, [Nilsson et al., 2019](#)) databases for taxonomic assignments, respectively.

Afterwards, alpha- and beta-diversity was performed for both bacterial and fungal communities. To remove the sequencing depth bias, normalization was introduced before the comparison of community diversities. Samples were rarefied to the same sequencing depth to remove the bias introduced by the uneven number of reads. Bacterial and fungal samples were rarefied to depth 1500 and 2000, respectively, using the *norm\_tab()* functions in '*mina*'. Afterwards, the Shannon index of each sample is calculated using the *diversity()* function from '*vegan*' ([Oksanen et al., 2019](#)) to characterize the number of community members and the evenness of their relative abundance distribution. To reduce the random error of rarefaction, this process was repeated 999 times, and the average was used as alpha-diversities of corresponding samples. Bray-Curtis dissimilarity was calculated to evaluate the dissimilarities, i.e. beta-diversity, between samples, with the *parDist()* function from '*parallelDist*' package using 80 threads considering the large ASV and sample numbers. Unexplained variance ratio was calculated as described in ([Anderson, 2001](#)) using the *get\_r2()* function in '*mina*' package.

### 7.5.3 Representative ASVs selection

Community members with the highest relative abundance and occupancy were extracted as the subset ASV table and Procrustes Analysis was implemented to compare the community shifts when only considering those ASVs. Function *procrustes()* from package *'vegan'* (Oksanen *et al.*, 2019) was used to calculate the  $M^2$  value, the sum of squared distances between paired points in the Bray-Curtis dissimilarity matrices of the compared ASV tables. ASVs ranked top  $i$  and  $j$  percentage occupancy and RA were selected as subset ASVs; and parameters  $i, j$  were evaluated from 1 to 20 separately. To account for the trade-off between community distortion and complexity, we chose parameters that caused the most considerable decrease of  $M^2$  when increasing the number of ASVs added to the subset.

### 7.5.4 Network inference and clustering

The correlation coefficient between repASVs was computed according to their co-variance between samples. Pearson and Spearman were calculated with *rcorr()* function from the *'Hmisc'* package using the renormalized ASV table after rarefaction, and SparCC was inferred by *fastspar* (Friedman & Alm, 2012; Watts *et al.*, 2019) with the raw ASV table without any rarefaction or normalization. For SparCC, 1000 times permutations were implemented to estimate the  $P$  values of the edges. Non-significant edges ( $P < 0.05$ ) were filtered to obtain the sparse adjacency matrix for later feature computation. Network density was calculated by dividing the number of edges with the biggest possible number of connections given the number of nodes. Positive and negative connections were computed separately, and therefore, both densities would distribute in the range of 0 to 0.5. The degree was calculated by counting the number of other nodes to which each ASV was connected.

Markov Clustering (MCL, Enright, 2002) and Affinity Propagation (AP, Frey & Dueck, 2007) were applied for clustering the nodes within each kingdom network using function *net\_cls()* in *'mina'*, which was implemented based on the *mcl()* function from *'MCL'* package and *apcluster()* function from *'apcluster'* package. The former method was limited to only positive edges and was applied to networks with parameter “-I 2.5” for the inflation. For AP, both positive and negative edges were considered during clustering and “p = 0” was applied for the input preference.

### 7.5.5 Community diversity and network analysis using *'mina'*

#### 7.5.5.1 Community diversity analysis using *'mina'*

We determined network clusters by inferring global bacterial and fungal networks from all samples (see also in section 7.3.2). Afterwards, Bray-Curtis dissimilarities were calculated based on the aRA of network clusters, which were aggregated RAs of repASVs assigned to the same cluster afterwards. The unexplained variance between the communities was then calculated with *get\_r2()* function and compared with the ASV-based diversity analysis.

#### 7.5.5.2 Community network analysis using *'mina'*

Samples were assigned to compartments randomly in the permutation datasets. Networks of each compartment in both original and permuted datasets were inferred with parameters “g\_size = 80, s\_size = 50, rm = FALSE, sig = TRUE, bs = 33, pm = 33” using *bs\_pm()* function from *'mina'*. To validate the stability of results, network inference with more subsampling replicates were computed,

with parameters “g\_size = 80, s\_size = 50, rm = FALSE, sig = TRUE, bs = 99, pm = 99”. Distances for the pairwise comparison of compartments was calculated with *net\_dis()* function, and both Jaccard and Spectra distances were shown in the Results session. PCA of Spectral distance was plotted for visualization, and *P* values were calculated by the Monte Carlo permutation test.

## 7.5.6 CAS-associated microbial community analysis

### 7.5.6.1 CAS dataset diversity and network comparison

For the CAS-derived root microbiota, the unexplained variance of community diversities based on all ASVs, repASVs, and network clusters were compared using *get\_r2()* function from ‘*mina*’. The global network of CAS dataset was inferred and clustered as before (Section 7.5.4).

The original and permuted networks of CAS-associated conditions were constructed the same as earlier stated (Section 7.5.5.2) for the whole dataset. Due to the limitation of sample numbers, parameter “g\_size = 40, s\_size = 20, rm = FALSE, sig = TRUE, bs = 33, pm = 33, individual = TRUE, out\_dir = bs\_pm\_dir” were used. The network matrices of all conditions inferred from the subsampling samples of both original and permuted datasets were stored in the defined “out\_dir”. Network distance was then calculated and statistically tested with the function *net\_dis()* from ‘*mina*’. As mentioned, the subsampling is completely random; therefore, the network distance calculation process is indeterministic. However, the results and conclusions from different runs are highly consistent.

### 7.5.6.2 Distinctive features selection of CAS networks

Node contrast and distance between compared conditions were calculated with *net\_node\_cmp()* function from ‘*mina*’. The connectivity of families was calculated by adding up the connectivity of the ASVs assigned to each family from the original and permuted network matrices, respectively. Group distance, including family and network cluster, were computed using *net\_grp\_cmp()* function of ‘*mina*’ and the fold change of distance between permutation and origin networks were calculated. The partial permutation was performed by randomly assigning the groups to the sample for selected group members. Afterwards, the ASV table was renormalized, and networks were then constructed.

## 7.6 Data and code availability

Data processing and visualization scripts used in this study are publicly available and could be found at [https://github.com/Guan06/DADA2\\_pipeline](https://github.com/Guan06/DADA2_pipeline) and [https://github.com/Guan06/Chapter\\_7](https://github.com/Guan06/Chapter_7).

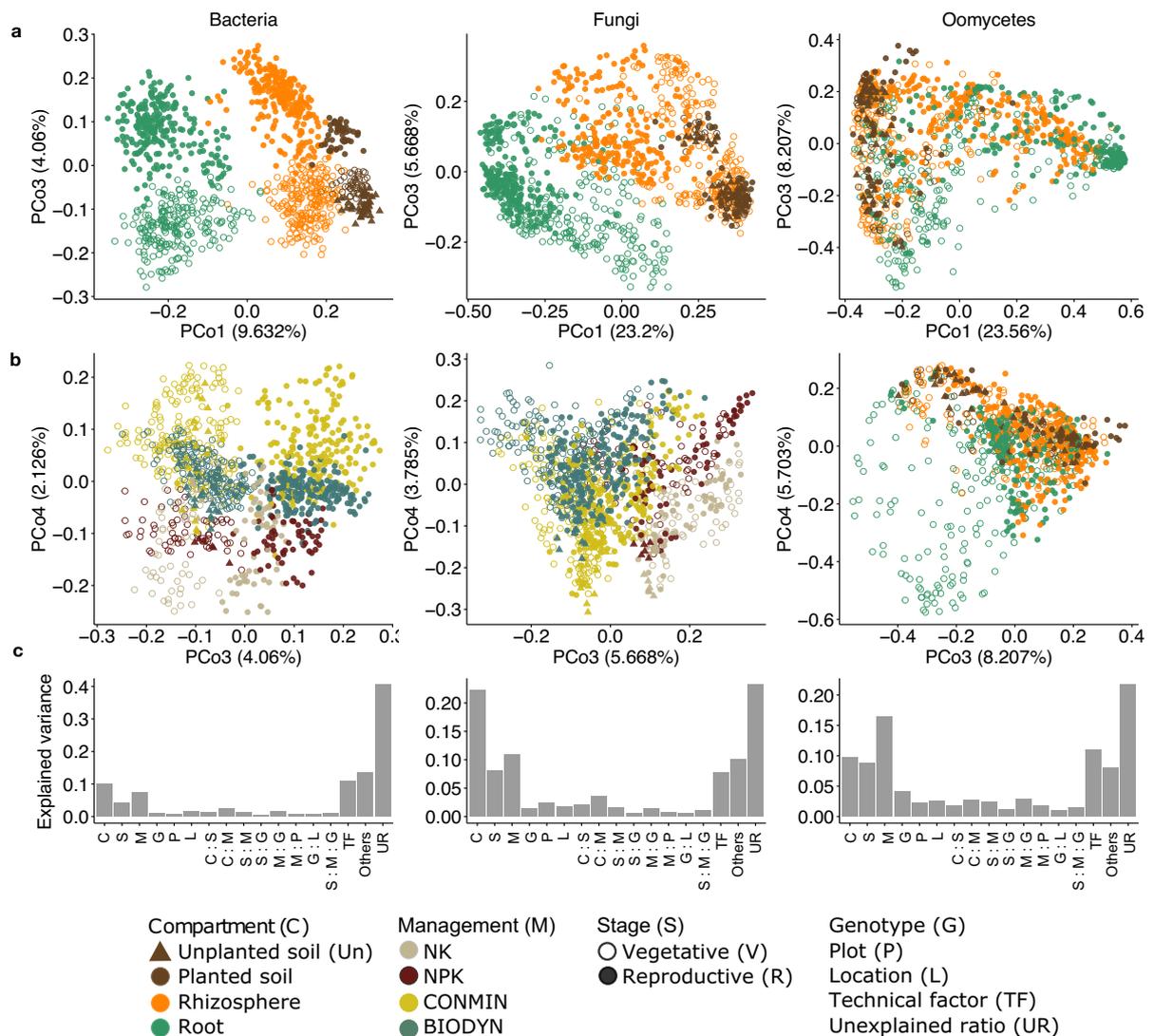
## 7.7 Author contributions

### Guan and Garrido-Oter (in preparation)

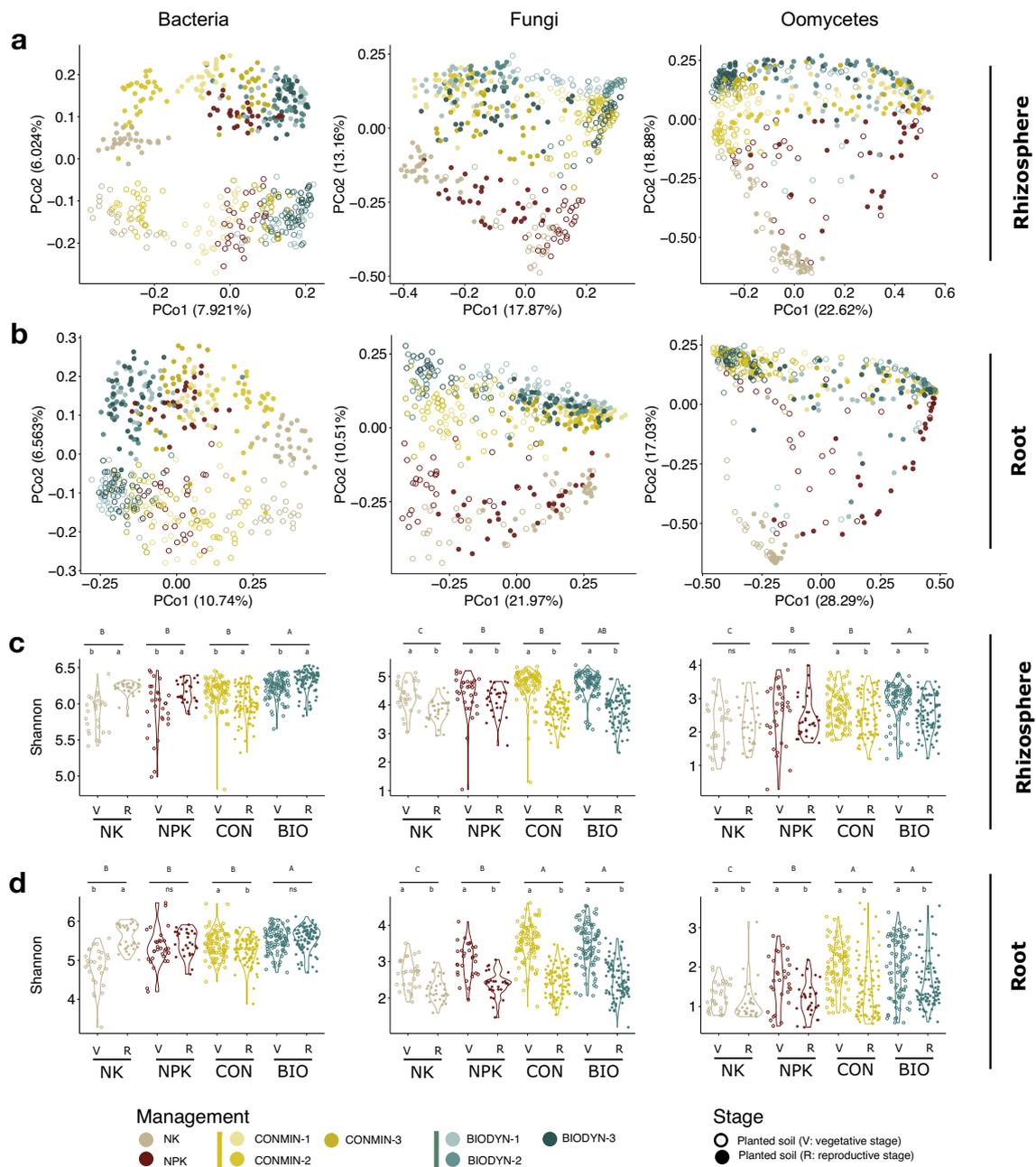
R.G.-O. conceived the project, R.G. carried out the project with the supervision of R.G.-O.; R.G. and R.G.-O. wrote the manuscript.



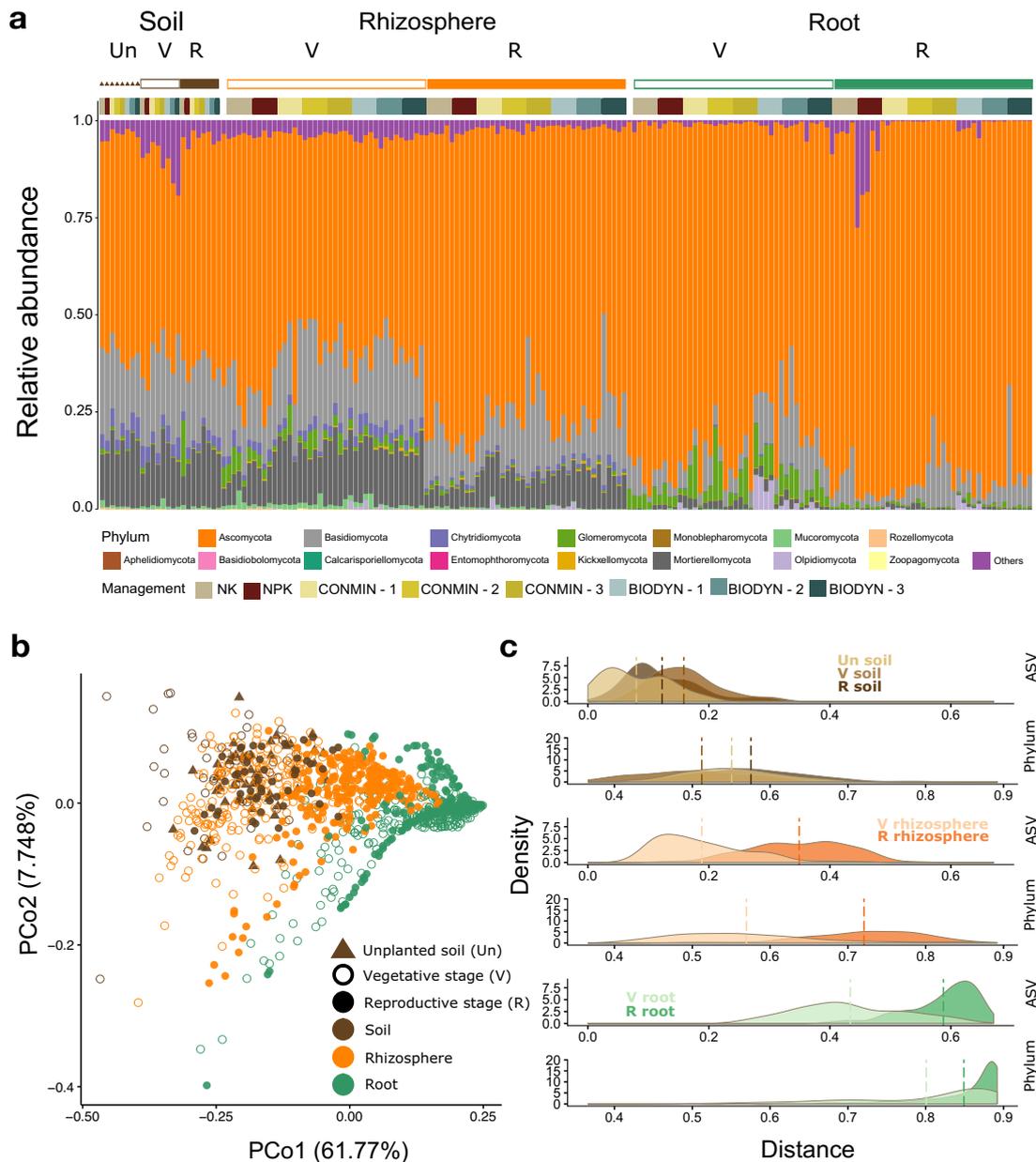
## Supplementary figures and tables



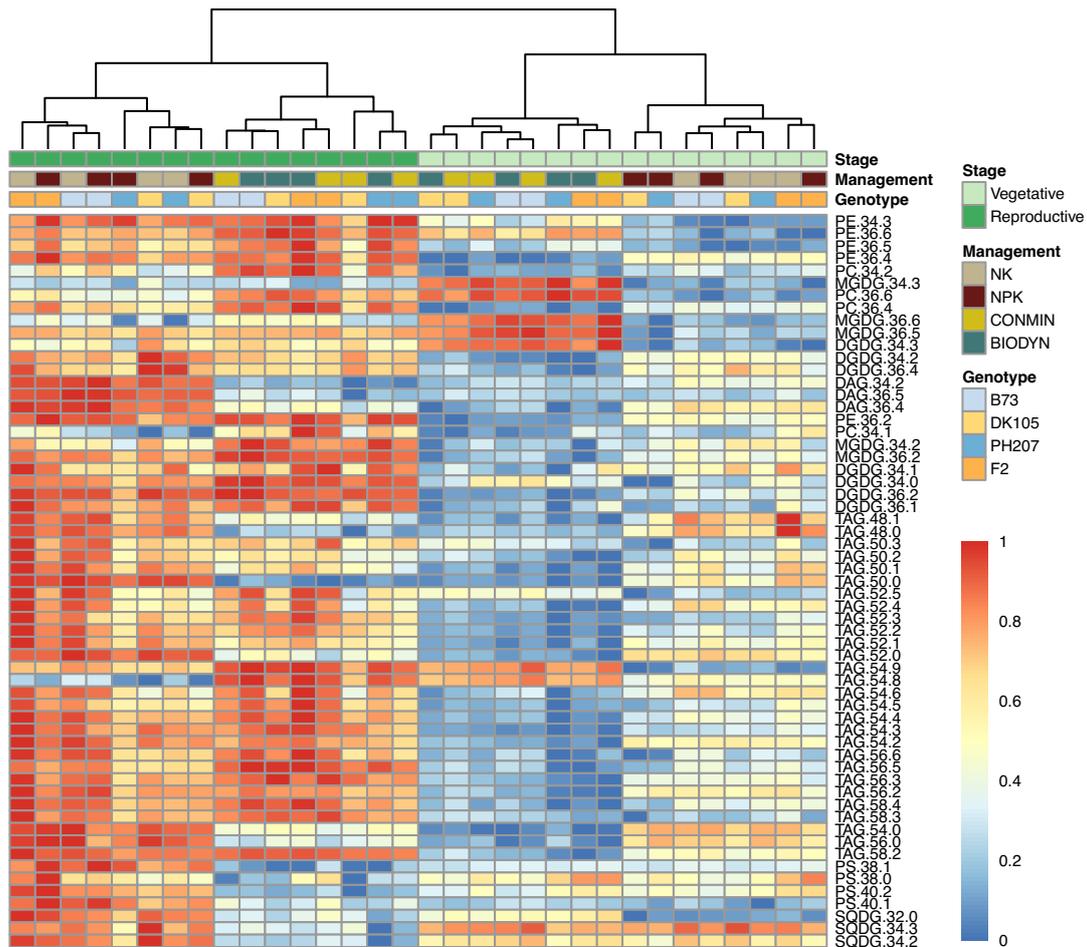
**Figure S6.1 Root compartment, soil management, and plant growth phase are the main drivers of shifts in the root-associated microbiota in field-grown maize.** (a and b) PCoA of all harvested samples for the three microbial kingdoms ( $n = 1,079, 1,103,$  and  $1,103$  samples for bacteria, fungi, and oomycetes, respectively) at PCo1 and -3 (a) and PCo3 and -4 (b). (c) The explained variance (percent) of each factor was calculated by PERMANOVA. Significant factors ( $P < 0.001$ ) explaining more than 1% of the variance are shown here, and other factors are integrated into “Others.” Technical factors (TF) include batch effects of the sequencing run and all other related parameters.



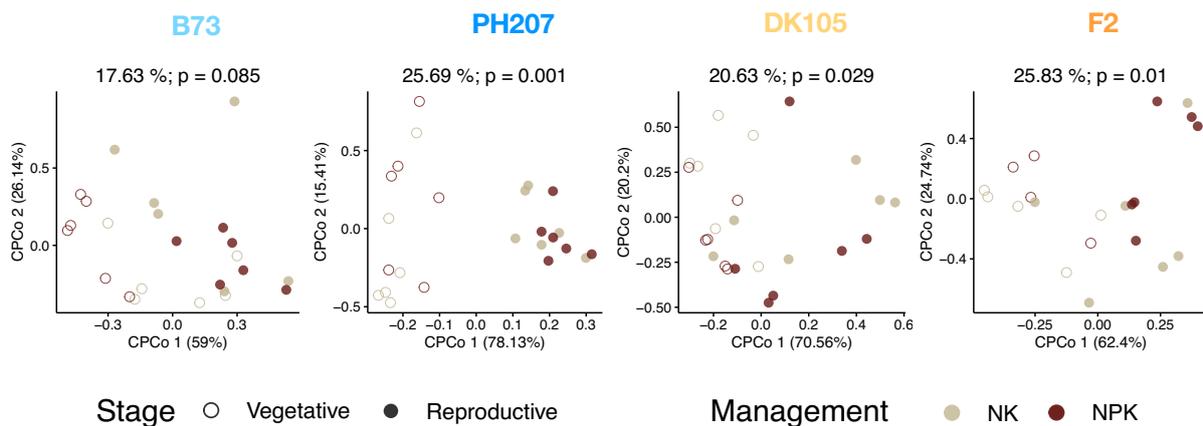
**Figure S6.2 The microbiota of rhizosphere and root is shaped by soil management and host growth.** PCoA based on Bray-Curtis dissimilarity between samples from rhizosphere (**a**) and root (**b**) of each kingdom are shown here ( $n = 480$  for both compartment). Alpha-diversity (Shannon index) of rhizosphere (**c**) and root (**d**) for all microbial kingdoms are compared. The Wilcoxon test was used for statistical analysis with FDR correction ( $P < 0.05$ ). V: vegetative stage, R: reproductive stage, NK: N and K fertilization, NPK: N, P and K fertilization, CON: CONMIN (conventional mineral fertilization) and BIO: BIODYN (biodynamic).



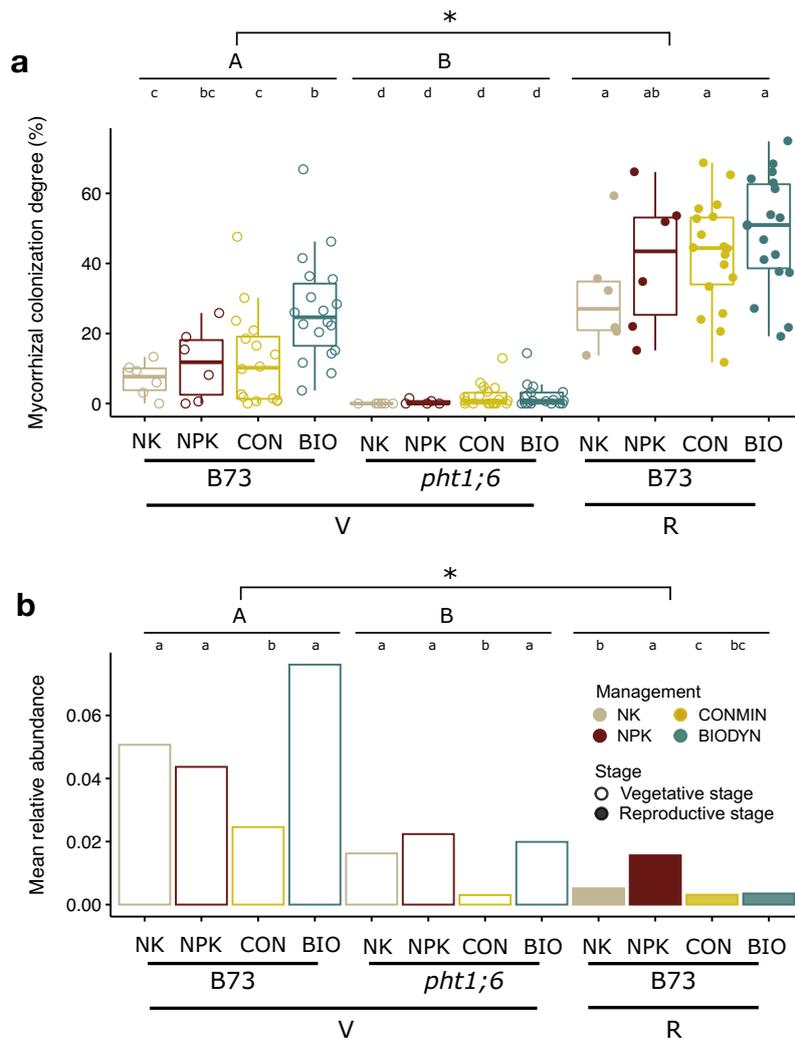
**Figure S6.3 Compartment and plant growth phase affect the fungal community similarly at the ASV and phylum levels.** (a) RA values of the 15 most abundant fungal phyla. The taxonomic group “Others” gathers phyla with <0.1% RAs. (b) PCoA of Bray-Curtis dissimilarities based on the RA of each phylum between communities. (c) Comparison of fungal communities at the ASV and phylum levels. “Distance” on the  $x$  axis indicates the Euclidean distance between the Bray-Curtis dissimilarity of each sample and those of the initial (unplanted soil) and final (reproductive root) condition communities. The vertical line indicates the average distance of the corresponding condition. Un, unplanted soil before sowing; V, vegetative soil; R, reproductive soil.



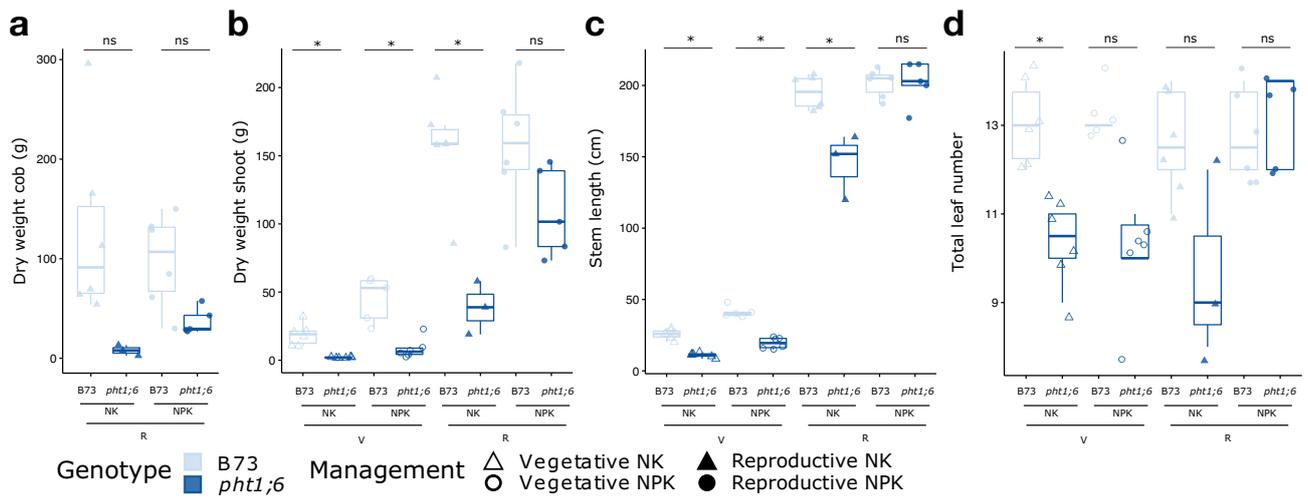
**Figure S6.4 Plant growth stage shapes root lipid profile.** Heatmap of the 59 lipid compounds in root samples from two growth stages, four soil managements and four inbred lines. Average of log-transformed and range normalized lipid profiles from each condition were shown.



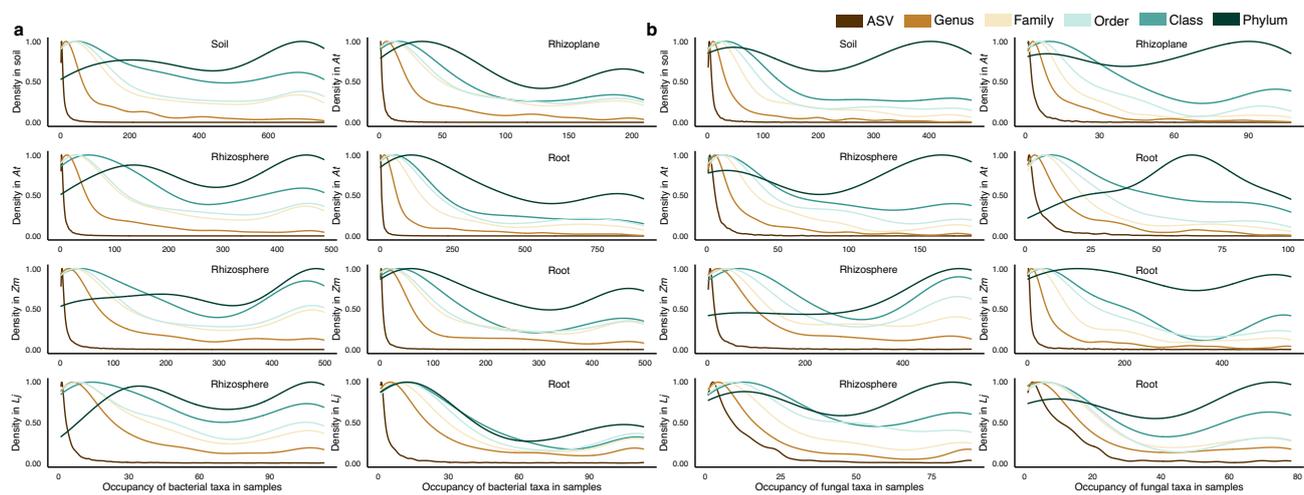
**Figure S6.5 Root sugar composition is affected by plant growth stage.** CPCA (~ Management \* Stage) based on Euclidean distance of the root sugar composition (sucrose, fructose and glucose) for the four inbred maize lines, grown in DEMO field (NK and NPK, n = 93).



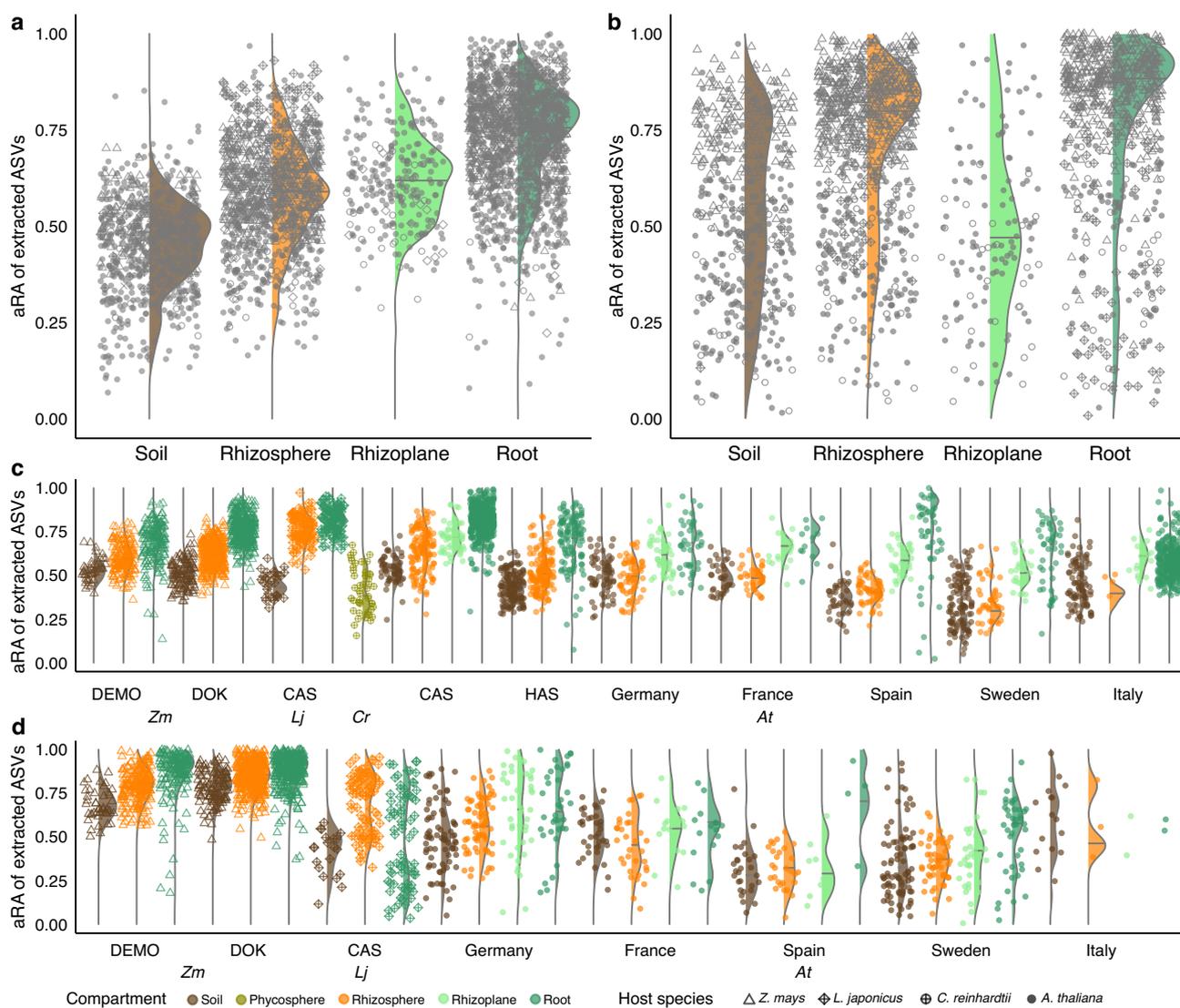
**Figure S6.6 Relative abundance of Glomeromycota and mycorrhizal colonization in roots is influenced by soil management, P transporter Pht1;6 and plant growth stage.** (a) The degree of colonization of mycorrhizal fungi over the course of plant growth in the B73 wild-type line and *pht1;6* mutant in the four soil managements. Colonization was estimated by microscopic counting on fine roots. (b) The RA of Glomeromycota in the B73 wild-type line and P transporter mutant line over the course of plant growth. Wilcoxon test was used for statistical analysis with FDR correction ( $P < 0.05$ ;  $n = 6$  for NK and NPK,  $n = 18$  for CONMIN and BIODYN). Lowercase letters indicate significant differences between managements at the two growth stages within each genotype. Capital letters indicate significant differences between two genotypes and asterisks indicate significant differences between two growth stages including all managements.



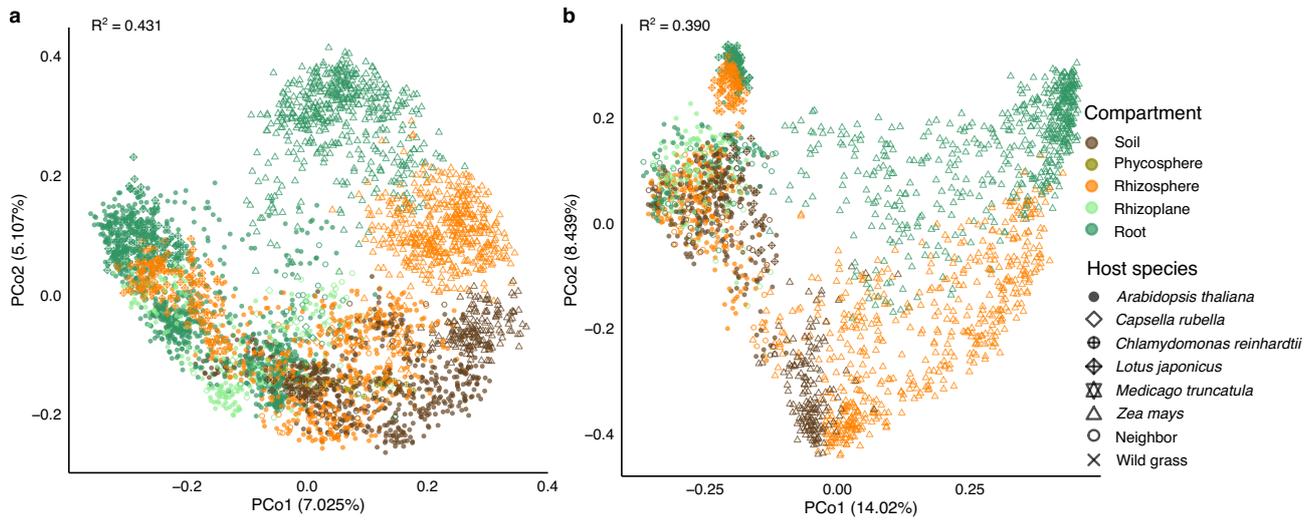
**Figure S6.7 Effects of genotype, plant growth stage, and P fertilization on plant biomass parameters.** Biomass including dry weight cob (a), dry weight shoot (b), stem length (c), and total leaf number (d) were measured ( $n = 43$ ). A Wilcoxon test followed by FDR correction was used for statistical analysis ( $P < 0.05$ ). Capital letters indicate significant differences between growth stages; lowercase letters indicate significant differences between genotypes in specific soil management, within each growth stage. ns, not significant.



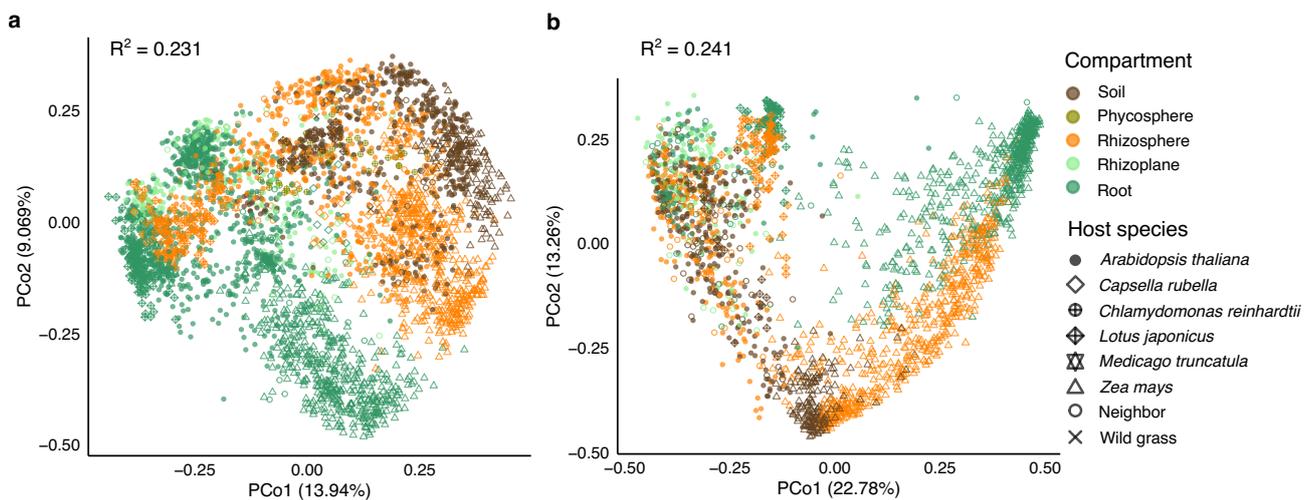
**Figure S7.1 Density plots for microbial occupancies in different compartments and host species.** Bacterial (a) and fungal (b) microbes at each taxonomic levels, from ASV to phylum, are shown here. At: *Arabidopsis thaliana*; Zm: *Zea mays*; Lj: *Lotus japonicus*.



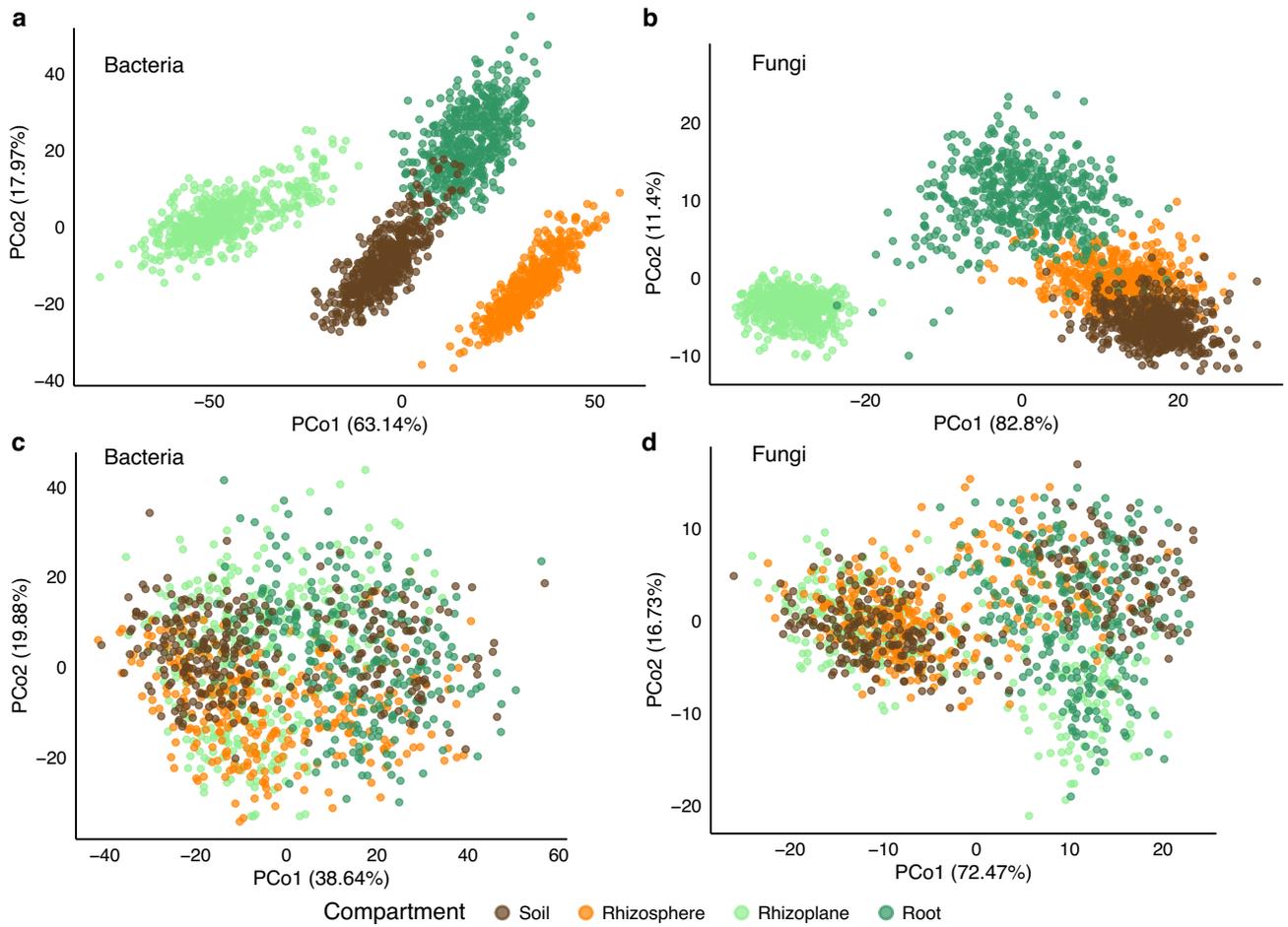
**Figure S7.2 Aggregated relative abundance of the subset ASVs selected by Procrustes Analysis.** The aRA of bacterial (a) and fungal (b) ASVs in each compartments were shown, with multiple host species indicated by the shape of the points. Samples from different compartment, host species and soil types were demonstrated separately for bacterial (c) and fungal (d) communities.



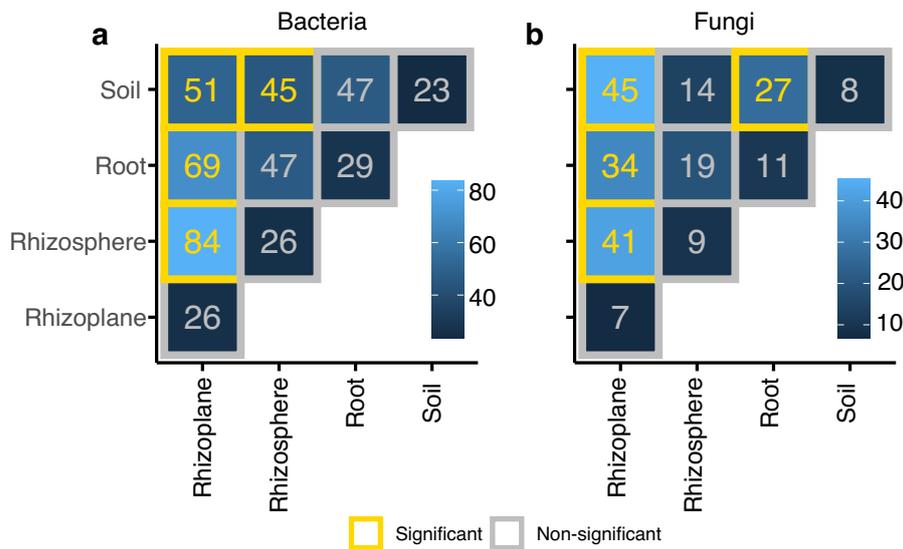
**Figure S7.3 Community diversity of plant-associated microbiota based on repASVs.** PCoA of Bray-Curtis dissimilarities between bacterial (a) and fungal (b) communities are shown here.  $R^2$  indicates the variance between samples which cannot be explained by compartment, soil type, host species, host genotype and experiment condition.



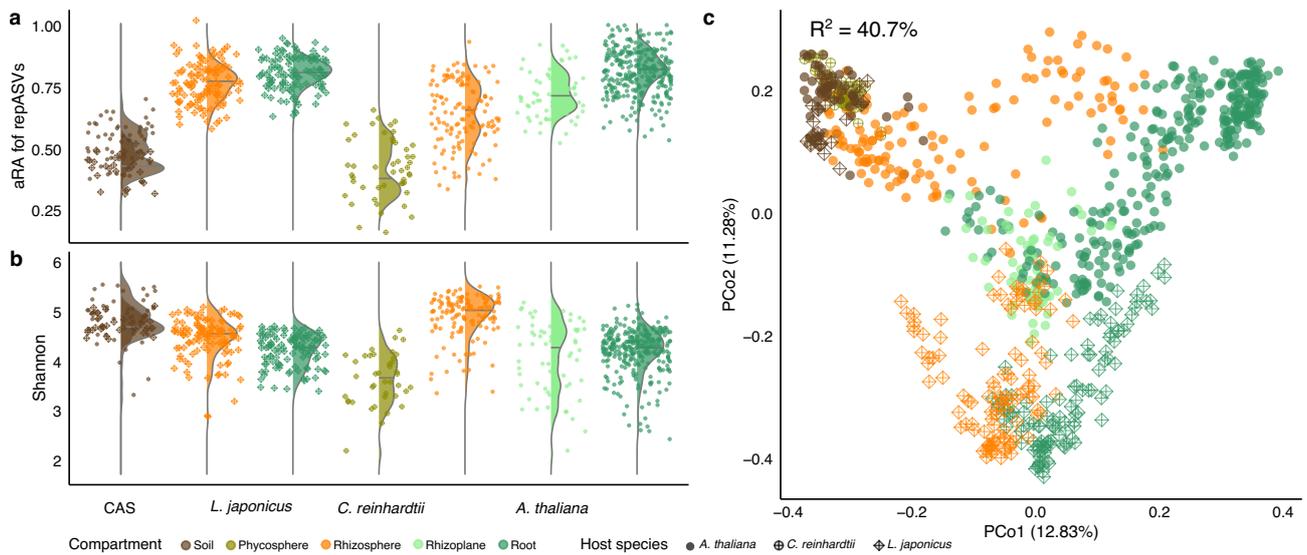
**Figure S7.4 Community diversity of plant-associated microbiota based on repASVs-based network clusters.** PCoA of Bray-Curtis dissimilarities between bacterial (a) and fungal (b) communities are shown here. Network clusters inferred from Spearman correlation matrices (with significant connections only,  $P < 0.05$ ) using AP were used for dissimilarity calculation.  $R^2$  indicates the variance between samples which cannot be explained by compartment, soil type, host species, host genotype and experiment condition.



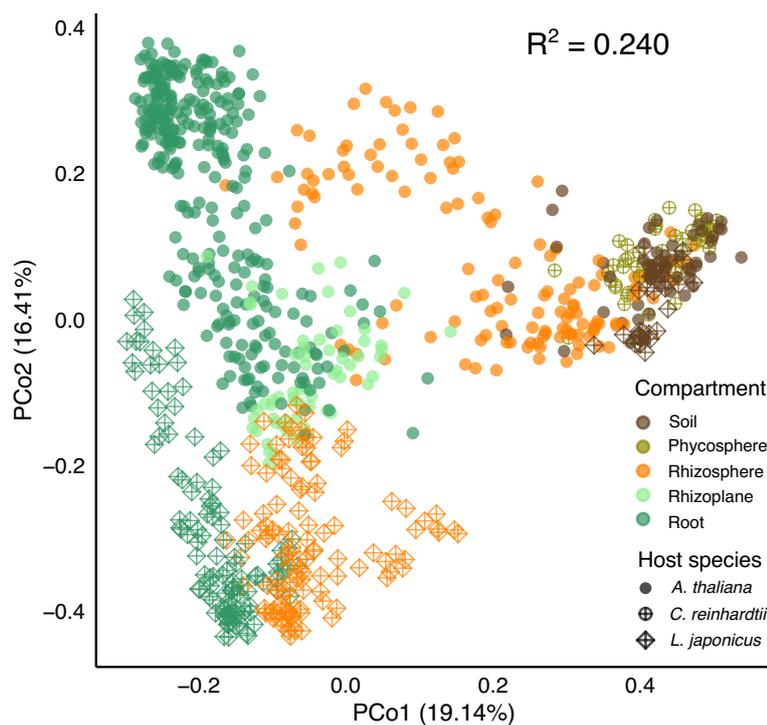
**Figure S7.5 Bootstrap-permutation network analysis of plant-associated microbiota.** Original (a bacteria, b fungi) and permuted (c bacteria, d fungi) networks are represented by points. Subsampling time  $n = 99$  for both original and permuted datasets for both kingdoms.



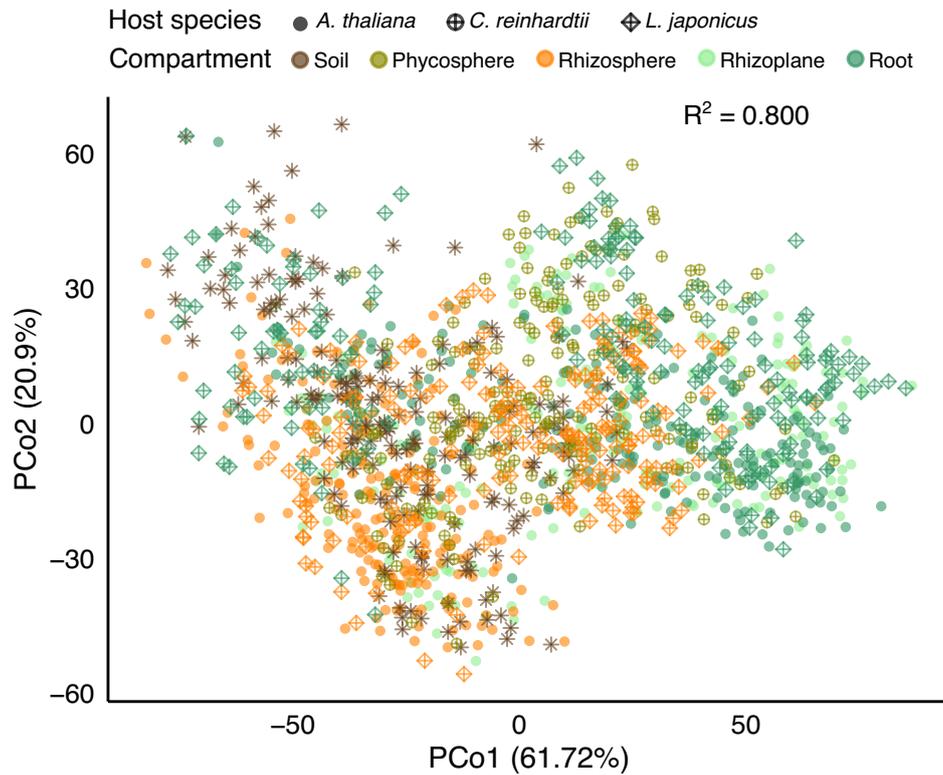
**Figure S7.6 Average Spectra distance between networks from different compartments.** Bacterial (a) and fungal (b) networks constructed from the samples of different compartments were compared here. Significances of the distances were tested with the stringent F-test.



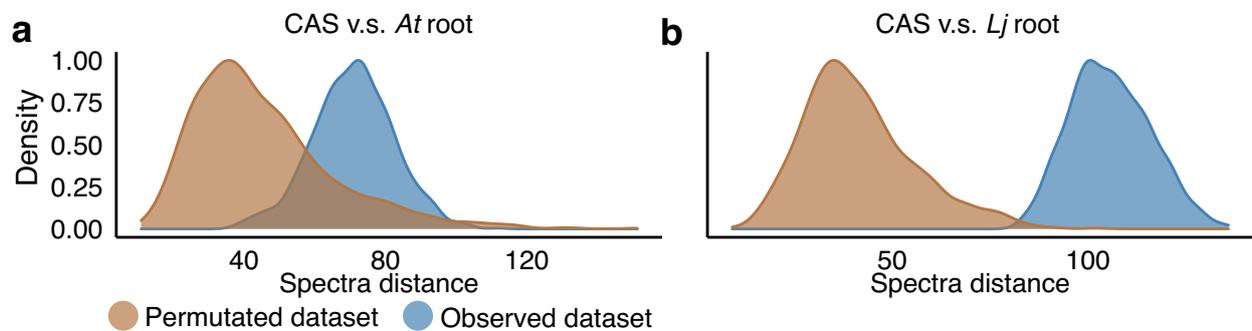
**Figure S7.7 RepASVs recapitulate the aRA and diversity of the whole community for CAS-associated samples.** (a) The aRA of repASVs comparing to the all ASVs. (b) Alpha-diversity of samples from each condition calculated based on repASVs. (c) PCoA of beta-diversity between samples, indicating compartment effect and host preference.  $R^2$  is the ratio of variance cannot be explained by compartment, soil batch, host species and genotype.



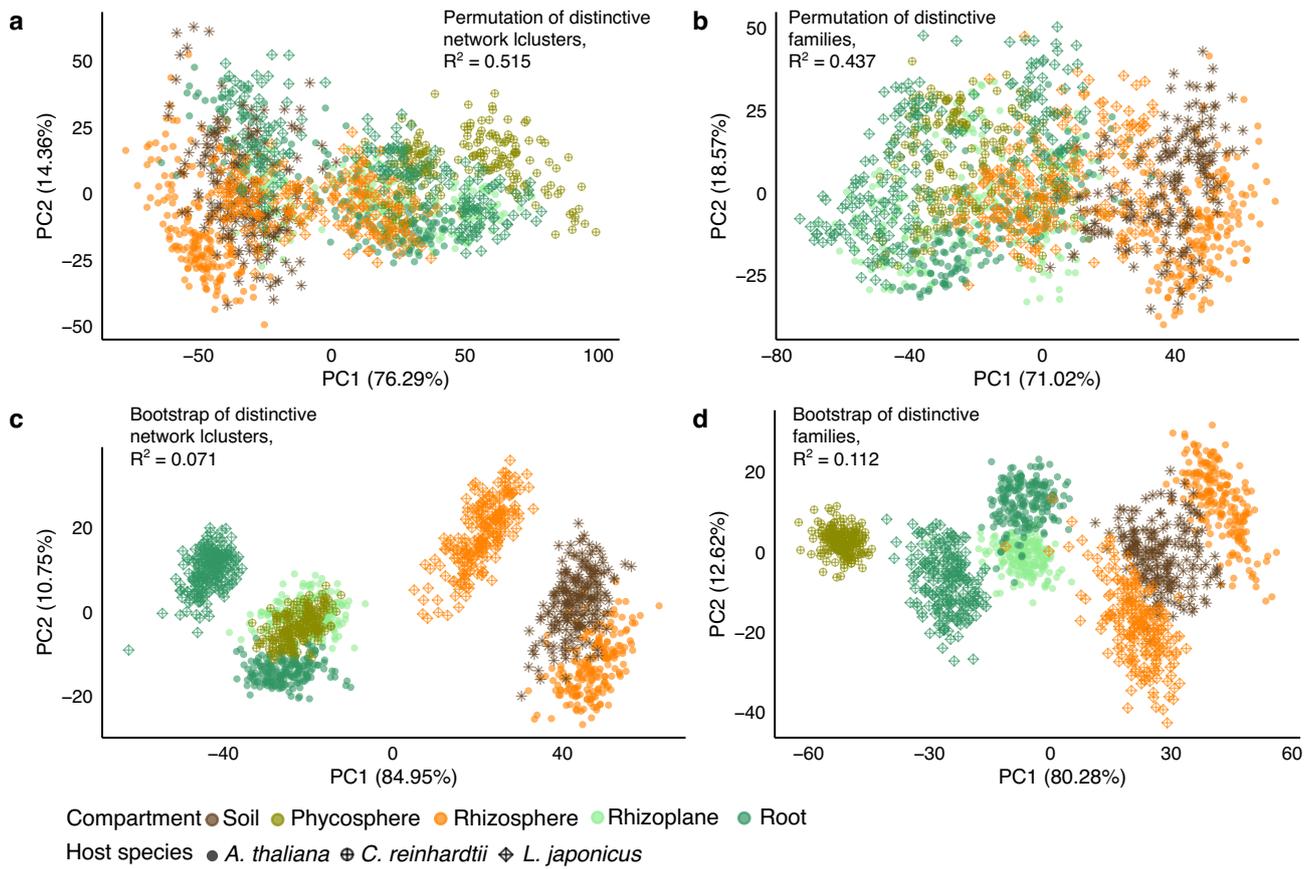
**Figure S7.8 Beta-diversity of CAS-associated communities based on network clusters.**  $R^2$  is the ratio of variance cannot be explained by compartment, soil batch, host species and genotype. Clusters were inferred by AP from the Spearman correlation networks and Bray-Curtis dissimilarities were calculated based on the aRA of network clusters.



**Figure S7.9 Network comparison between different CAS-associated conditions in the permitted dataset.** PCoA of Spectra distance show no separation when networks were constructed from the permuted dataset.  $R^2$  indicates the variance of Spectra distances between networks which cannot be explained by compartment and host species.



**Figure S7.10 Distribution of distances computed from the observed and permuted networks.** Scale density of Spectra distance between CAS and *At* root (a), CAS and *Lj* root (b) are shown here.



**Figure S7.11 Network comparison between different CAS-associated conditions of distinctive groups.** Groups before the distinctive point ( $n = 26$  for both), i.e. family Chitinophagaceae (**a**) and network Cluster\_51 (**b**), were accumulatively permuted. For (**c**) and (**d**), only distinctive groups ( $n = 26$  for both) were used for network inference and comparison. PCA of their Spectra distance was shown here.  $R^2$  indicates the variance of Spectra distances between networks that cannot be explained by compartment and host species.

**Table S7.1 Density of bacterial network inferred from all and each compartment samples.**

	Pearson		Spearman		SparCC	
	+	- (%)	+	- (%)	+	- (%)
All	21.32	22.93	36.73	35.64	39.99	43.67
Soil	12.43	6.12	19.94	13.58	31.37	34.35
Rhizosphere	17.22	15.12	26.57	25.87	34.81	38.16
Rhizoplane	7.27	0.39	12.94	2.94	9.03	12.73
Root	16.74	7.30	28.45	18.15	37.11	39.17

**Table S7.2 Unexplained variance ratio when applying network cluster based diversity analysis.**  
MCL: Marcov clustering; AP: Affinity propagation.

Network	Clustering	Number of taxa	Unexplained ratio (%)
Pearson	ASVs	14,844	47.4
	repASVs	1,774	40.7
	MCL	58	11.0
	AP	199	30.3
Spearman	MCL	26	7.5
	AP	134	24.0
SparCC	MCL	3	25.4
	AP	74	24.8

## Abbreviations

ASV	amplicon sequence variant
AMF	arbuscular mycorrhizal fungi
AP	affiliation propagation
APs	artificial photoassimilates
avFe	available iron
BC	Bray-Curtis dissimilarity
bHLH	basic helix-loop-helix
BIODYN	BIODYNamic
BXs	benzoxazinoids
CAS	Cologne agricultural soil
CFUs	colony-forming units
CONMIN	CONventional MINeral
CPCoA	constrained principal coordinate analysis
CYP82C4	CYTOCHROME P450, FAMILY 82C4
DAG	diacylglycerol
DAMP	danger-associated molecular pattern
deASVs	differentially enriched ASVs
DEGs	differentially-expressed genes
DEMO	fertilization DEMOnstration experiment
DGDGs	digalactosyldiacylglycerols
DOK	Dynamic, organic and conventional managements
ENA	European Nucleotide Archive
F6'H1	Feruloyl-CoA 6'-Hydroxylase1
FIT	Fer-like iron deficiency induced transcription factor
FRO2	Ferric reduction oxidase 2
FYM	farmyard manure
ICP-MS	inductively coupled plasma mass spectrometry
IRL	sequence-indexed rhizobacterial library
IRT1	Iron-regulated transporter1
IS	Italian soil
KO	KEGG Orthologue
M/PAMPs	microbe/pathogen-associated molecular patterns
MAG	metagenome-assembled genome
MAMP	microbe-associated molecular pattern
MBOA	6-methoxy-benzoxazolin-2-one
MCL	Markov cluster
MGDGs	monogalactosyldiacylglycerols
mina	microbial community diversity and network analysis

MTI	MAMP-triggered immunity
NK	Nitrogen and Potassium
NPK	Nitrogen, Phosphate and Potassium
OD	optical density
OTU	Operational Taxonomic Unit
PC	Principal Component
PCA	Principal component analyses
PCoA	principal coordinate analysis
PDR9	PLEIOTROPIC DRUG RESISTANCE 9
PRR	pattern recognition receptor
PS	phosphatidylserine
PVDF	polyvinylidene fluoride
RA	relative abundance
<i>rbcL</i>	ribulose-bisphosphate carboxylase gene
repASVs	representative ASVs
RGI	root growth inhibition
rRNA	ribosomal RNA
S8H	SCOPOLETIN 8-HYDROXYLASE
SFW	shoot fresh weight
SQDG	sulfoquinovosyldiacylglycerol
SV	surrogate variable
SynCom	Synthetic community
TAGs	triacylglycerols
TFs	transcription factors
TSA	tryptic soy agar
TSB	tryptic soy broth
unavFe	unavailable iron
WT	wild-type

## References

- The Gene Ontology Resource: 20 years and still GOing strong.** 2019. *Nucleic Acids Research* **47**: D330–D338.
- Abendroth LJ, Elmore RW, Boyer MJ, Markay SK.** 2011. *Corn growth and development.*
- Aitchison J.** 1982. The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society: Series B (Methodological)* **44**: 139–160.
- Agosta SJ, Klemens JA.** 2008. Ecological fitting by phenotypically flexible genotypes: implications for species associations, community assembly and evolution. *Ecology Letters* **11**: 1123–1134.
- Albert R, Barabási A-L.** 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics* **74**: 47–97.
- Alcaraz LD, Peimbert M, Barajas HR, Dorantes-Acosta AE, Bowman JL, Arteaga-Vázquez MA.** 2018. Marchantia liverworts as a proxy to plants' basal microbiomes. *Scientific Reports* **8**: 12712.
- Aldén L, Demoling F, Bååth E.** 2001. Rapid Method of Determining Factors Limiting Bacterial Growth in Soil. *Applied and Environmental Microbiology* **67**: 1830–1838.
- Alvarez L, Aliashkevich A, de Pedro MA, Cava F.** 2018. Bacterial secretion of D-arginine controls environmental microbial biodiversity. *The ISME Journal* **12**: 438–450.
- Ambrosini A, Beneduzi A, Stefanski T, Pinheiro FG, Vargas LK, Passaglia LMP.** 2012. Screening of plant growth promoting Rhizobacteria isolated from sunflower (*Helianthus annuus* L.). *Plant and Soil* **356**: 245–264.
- Amin SA, Green DH, Hart MC, Kupper FC, Sunda WG, Carrano CJ.** 2009. Photolysis of iron-siderophore chelates promotes bacterial-algal mutualism. *Proceedings of the National Academy of Sciences* **106**: 17071–17076.
- Amin SA, Hmelo LR, van Tol HM, Durham BP, Carlson LT, Heal KR, Morales RL, Berthiaume CT, Parker MS, Djunaedi B, et al.** 2015. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**: 98–101.
- Anderson MJ.** 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26**: 32–46.
- Almario J, Jeena G, Wunder J, Langen G, Zuccaro A, Coupland G, Bucher M.** 2017. Root-associated fungal microbiota of nonmycorrhizal *Arabis alpina* and its contribution to plant phosphorus nutrition. *Proceedings of the National Academy of Sciences* **114**: E9403–E9412.
- Anderson MJ.** 2017. Permutational Multivariate Analysis of Variance (PERMANOVA) . *Wiley StatsRef: Statistics Reference Online*: 1–15.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al.** 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**: 25–29.
- Astudillo-García C, Bell JJ, Webster NS, Glasl B, Jompa J, Montoya JM, Taylor MW.** 2017. Evaluating the core microbiota in complex communities: A systematic investigation. *Environmental Microbiology* **19**: 1450–1462.
- Azam F, Fenchel T, Field J, Gray J, Meyer-Reil L, Thingstad F.** 1983. The Ecological Role of Water-Column Microbes in the Sea. *Marine Ecology Progress Series* **10**: 257–263.

- Bai Y, Müller DB, Srinivas G, Garrido-Oter R, Potthoff E, Rott M, Dombrowski N, Münch PC, Spaepen S, Remus-Emsermann M, et al. 2015.** Functional overlap of the Arabidopsis leaf and root microbiota. *Nature* **528**: 364–369.
- Bairey E, Kelsic ED, Kishony R. 2016.** High-order species interactions shape ecosystem diversity. *Nature Communications* **7**: 12285.
- Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM. 2006.** The role of root exudates in rhizosphere interactions with plants and other organisms. *Annual Review of Plant Biology* **57**: 233–266.
- Banerjee S, Walder F, Büchi L, Meyer M, Held AY, Gattinger A, Keller T, Charles R, van der Heijden MGA. 2019a.** Agricultural intensification reduces microbial network complexity and the abundance of keystone taxa in roots. *ISME Journal* **13**: 1722–1736.
- Banerjee S, Walder F, Büchi L, Meyer M, Held AY, Gattinger A, Keller T, Charles R, van der Heijden MGA. 2019b.** Agricultural intensification reduces microbial network complexity and the abundance of keystone taxa in roots. *The ISME Journal* **13**: 1722–1736.
- Ba R, Alfa T, Gbaguidi F, Novidzro KM, Dotse K, Koudouvo K, Houngue U, Donou Hounsode MT, Koumaglo KH, Ameyapoh Y, et al. 2017.** Maize Fungal Growth Control with Scopoletin of Cassava Roots Produced in Benin. *International Journal of Microbiology* **2017**: 1–11.
- Barberán A, Bates ST, Casamayor EO, Fierer N. 2012.** Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME Journal* **6**: 343–351.
- Bascompte J, Jordano P, Melian CJ, Olesen JM. 2003.** The nested assembly of plant-animal mutualistic networks. *Proceedings of the National Academy of Sciences* **100**: 9383–9387.
- Batstone RT, O'Brien AM, Harrison TL, Frederickson ME. 2020.** Experimental evolution makes microbes more cooperative with their local host genotype. *Science* **370**: 476–478.
- Baudoin E, Benizri E, Guckert A. 2003.** Impact of artificial root exudates on the bacterial community structure in bulk soil and maize rhizosphere. *Soil Biology and Biochemistry* **35**: 1183–1192.
- Bender SF, Wagg C, van der Heijden MGA. 2016.** An Underground Revolution: Biodiversity and Soil Ecological Engineering for Agricultural Sustainability. *Trends in Ecology and Evolution* **31**: 440–452.
- Beckers B, Op De Beeck M, Weyens N, Boerjan W, Vangronsveld J. 2017.** Structural variability and niche differentiation in the rhizosphere and endosphere bacterial microbiome of field-grown poplar trees. *Microbiome* **5**: 25.
- Bednarek P. 2012.** Chemical warfare or modulators of defence responses – the function of secondary metabolites in plant immunity. *Current Opinion in Plant Biology* **15**: 407–414.
- Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S. 2010.** Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. *Proceedings of the National Academy of Sciences* **107**: 18724–18728.
- Bell W, Mitchell R. 1972.** Chemotactic and growth responses of marine bacteria to algal extracellular products. *The Biological Bulletin* **143**: 265–277.
- Benhamou N, le Floch G, Vallance J, Gerbore J, Grizard D, Rey P. 2012.** Pythium oligandrum: an example of opportunistic success. *Microbiology* **158**: 2679–2694.
- Berendsen RL, Pieterse CMJ, Bakker PAHM. 2012.** The rhizosphere microbiome and plant health. *Trends in Plant Science* **17**: 478–486.

- Berens ML, Wolinska KW, Spaepen S, Ziegler J, Nobori T, Nair A, Krüler V, Winkelmüller TM, Wang Y, Mine A, et al. 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* **116**: 2364–2373.
- Birkhofer K, Bezemer TM, Bloem J, Bonkowski M, Christensen S, Dubois D, Ekelund F, Fließbach A, Gunst L, Hedlund K, et al. 2008.** Long-term organic farming fosters below and aboveground biota: Implications for soil quality, biological control and productivity. *Soil Biology and Biochemistry* **40**: 2297–2308.
- Blasche S, Kim Y, Mars RAT, Machado D, Maansson M, Kafkia E, Milanese A, Zeller G, Teusink B, Nielsen J, et al. 2021.** Metabolic cooperation and spatiotemporal niche partitioning in a kefir microbial community. *Nature Microbiology* **6**: 196–208.
- Bolger AM, Lohse M, Usadel B. 2014.** Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.
- Borghi L, Kang J, de Brito Francisco R. 2019.** Filling the Gap: Functional Clustering of ABC Proteins for the Investigation of Hormonal Transport in planta. *Frontiers in Plant Science* **10**.
- Bourgaud F, Olry A, Hehn A. 2014.** Recent Advances in Molecular Genetics of Furanocoumarin Synthesis in Higher Plants. In: *Recent Advances in Redox Active Plant and Microbial Products*. Dordrecht: Springer Netherlands, 363–375.
- Bravo A, Brands M, Wewer V, Dörmann P, Harrison MJ. 2017.** Arbuscular mycorrhiza-specific enzymes FatM and RAM2 fine-tune lipid biosynthesis to promote development of arbuscular mycorrhiza. *New Phytologist* **214**: 1631–1645.
- Bray JR, Curtis JT. 1957.** An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs* **27**: 325–349.
- Bray NL, Pimentel H, Melsted P, Pachter L. 2016.** Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**: 525–527.
- Bressan M, Roncato M-A, Bellvert F, Comte G, Haichar F el Z, Achouak W, Berge O. 2009.** Exogenous glucosinolate produced by *Arabidopsis thaliana* has an impact on microbes in the rhizosphere and plant roots. *The ISME Journal* **3**: 1243–1257.
- Brewer TE, Albertsen M, Edwards A, Kirkegaard RH, Rocha EPC, Fierer N. 2020.** Unlinked rRNA genes are widespread among bacteria and archaea. *ISME Journal* **14**: 597–608.
- Brisson VL, Schmidt JE, Northen TR, Vogel JP, Gaudin ACM. 2019.** Impacts of Maize Domestication and Breeding on Rhizosphere Microbial Community Recruitment from a Nutrient Depleted Agricultural Soil. *Scientific Reports* **9**: 1–14.
- Broughton WJ, Dilworth MJ. 1971.** Control of leghaemoglobin synthesis in snake beans. *Biochemical Journal* **125**: 1075–1080.
- Bucher M. 2007.** Functional biology of plant phosphate uptake at root and mycorrhiza interfaces. *New Phytologist* **173**: 11–26.
- Bulgarelli D, Garrido-Oter R, Münch PC, Weiman A, Dröge J, Pan Y, McHardy AC, Schulze-Lefert P. 2015.** Structure and Function of the Bacterial Root Microbiota in Wild and Domesticated Barley. *Cell Host & Microbe* **17**: 392–403.

- Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, Rauf P, Huettel B, Reinhardt R, Schmelzer E, et al. 2012.** Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* **488**: 91–5.
- Bulgarelli D, Schlaeppi K, Spaepen S, van Themaat EVL, Schulze-Lefert P. 2013.** Structure and Functions of the Bacterial Microbiota of Plants. *Annual Review of Plant Biology* **64**: 807–838.
- Callahan BJ, McMurdie PJ, Holmes SP. 2017.** Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal* **11**: 2639–2643.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016.** DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**: 581–583.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, et al. 2010.** QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335–336.
- Carlström CI, Field CM, Bortfeld-Miller M, Müller B, Sunagawa S, Vorholt JA. 2019.** Synthetic microbiota reveal priority effects and keystone strains in the Arabidopsis phyllosphere. *Nature Ecology & Evolution* **3**: 1445–1454.
- Carpinella MC, Ferrayoli CG, Palacios SM. 2005.** Antifungal Synergistic Effect of Scopoletin, a Hydroxycoumarin Isolated from *Melia azedarach* L. Fruits. *Journal of Agricultural and Food Chemistry* **53**: 2922–2927.
- Carrión VJ, Perez-Jaramillo J, Cordovez V, Tracanna V, de Hollander M, Ruiz-Buck D, Mendes LW, van Ijcken WFJ, Gomez-Exposito R, Elsayed SS, et al. 2019.** Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. *Science* **366**: 606–612.
- Castrillo G, Teixeira PJPL, Paredes SH, Law TF, De Lorenzo L, Felcher ME, Finkel OM, Breakfield NW, Mieczkowski P, Jones CD, et al. 2017.** Root microbiota drive direct integration of phosphate stress and immunity. *Nature* **543**: 513–518.
- Chaparro JM, Badri D V., Bakker MG, Sugiyama A, Manter DK, Vivanco JM. 2013.** Root Exudation of Phytochemicals in Arabidopsis Follows Specific Patterns That Are Developmentally Programmed and Correlate with Soil Microbial Functions (K Wu, Ed.). *PLoS ONE* **8**: e55731.
- Chase JM. 2003.** Community assembly: when should history matter? *Oecologia* **136**: 489–498.
- Chen S, Zhou Y, Chen Y, Gu J. 2018.** fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890.
- Chen T, Nomura K, Wang X, Sohrabi R, Xu J, Yao L, Paasch BC, Ma L, Kremer J, Cheng Y, et al. 2020.** A plant genetic network for preventing dysbiosis in the phyllosphere. *Nature* **580**: 653–657.
- Chezem WR, Memon A, Li F-S, Weng J-K, Clay NK. 2017.** SG2-Type R2R3-MYB Transcription Factor MYB15 Controls Defense-Induced Lignification and Basal Immunity in Arabidopsis. *The Plant Cell* **29**: 1907–1926.
- Chowdhury SP, Babin D, Sandmann M, Jacquioid S, Sommermann L, Sørensen SJ, Fließbach A, Mäder P, Geistlinger J, Smalla K, et al. 2019.** Effect of long-term organic and mineral fertilization strategies on rhizosphere microbiota assemblage and performance of lettuce. **21**: 2426–2439.
- Cirri E, Pohnert G. 2019.** Algae–bacteria interactions that balance the planktonic microbiome. *New Phytologist* **223**: 100–106.
- Clauss MJ, Mitchell-Olds T. 2006.** Population genetic structure of *Arabidopsis lyrata* in Europe. *Molecular Ecology* **15**: 2753–2766.

- Colangelo EP, Guerinot M Lou. 2004.** The Essential Basic Helix-Loop-Helix Protein FIT1 Is Required for the Iron Deficiency Response. *The Plant Cell* **16**: 3400–3412.
- Coleman-Derr D, Desgarenes D, Fonseca-Garcia C, Gross S, Clingenpeel S, Woyke T, North G, Visel A, Partida-Martinez LP, Tringe SG. 2016.** Plant compartment and biogeography affect microbiome composition in cultivated and native Agave species. *New Phytologist* **209**: 798–811.
- Cosetta CM, Wolfe BE. 2020.** Deconstructing and Reconstructing Cheese Rind Microbiomes for Experiments in Microbial Ecology and Evolution. *Current Protocols in Microbiology* **56**.
- Cotton TEA, Pétriacq P, Cameron DD, Meselmani M Al, Schwarzenbacher R, Rolfe SA, Ton J. 2019.** Metabolic regulation of the maize rhizobiome by benzoxazinoids. *The ISME Journal* **13**: 1647–1658.
- Cram JA, Xia LC, Needham DM, Sachdeva R, Sun F, Fuhrman JA. 2015.** Cross-depth analysis of marine bacterial networks suggests downward propagation of temporal changes. *ISME Journal* **9**: 2573–2586.
- Cregger MA, Veach AM, Yang ZK, Crouch MJ, Vilgalys R, Tuskan GA, Schadt CW. 2018.** The Populus holobiont: dissecting the effects of plant niches and genotype on the microbiome. *Microbiome* **6**: 31.
- Crits-Christoph A, Diamond S, Lavy A, Matheus Carnevali PB, Banfield JF. 2020.** Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial Species Boundaries (T Woyke, Ed.). *mSystems* **5**.
- Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG. 2005.** Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* **438**: 90–93.
- Curl EA, Truelove B. 1986.** *The Rhizosphere*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling A V., Devlin AS, Varma Y, Fischbach MA, et al. 2014.** Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**: 559–563.
- de Zélicourt A, Synek L, Saad MM, Alzubaidy H, Jalal R, Xie Y, Andrés-Barrao C, Rolli E, Guerard F, Mariappan KG, et al. 2018.** Ethylene induced plant stress tolerance by Enterobacter sp. SA187 is mediated by 2-keto-4-methylthiobutyric acid production (CMJ Pieterse, Ed.). *PLOS Genetics* **14**: e1007273.
- Delaux P-M, Radhakrishnan G V., Jayaraman D, Cheema J, Malbreil M, Volkening JD, Sekimoto H, Nishiyama T, Melkonian M, Pokorny L, et al. 2015.** Algal ancestor of land plants was preadapted for symbiosis. *Proceedings of the National Academy of Sciences* **112**: 13390–13395.
- Demoling F, Figueroa D, Bååth E. 2007.** Comparison of factors limiting bacterial growth in different soils. *Soil Biology and Biochemistry* **39**: 2485–2495.
- Van Deynze A, Zamora P, Delaux P-M, Heitmann C, Jayaraman D, Rajasekar S, Graham D, Maeda J, Gibson D, Schwartz KD, et al. 2018.** Nitrogen fixation in a landrace of maize is supported by a mucilage-associated diazotrophic microbiota (E Kemen, Ed.). *PLOS Biology* **16**: e2006352.
- Dombrowski N, Schlaeppi K, Agler MT, Hacquard S, Kemen E, Garrido-oter R, Wunder J, Coupland G, Schulze-lefert P. 2017.** Root microbiota dynamics of perennial Arabis alpina are dependent on soil residence time but independent of flowering time. : 43–55.
- Dunne JA, Williams RJ, Martinez ND. 2002.** Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences* **99**: 12917–12922.
- Durán P, Flores-Urbe J, Wippel K, Zhang P, Guan R, Melkonian B, Melkonian M, Garrido-Oter R. 2022.** Shared features and reciprocal complementation of the Chlamydomonas and Arabidopsis microbiota. *Nature Communications* **13**: 406.

- 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**: 973-983.e14.
- Eberl L, Vandamme P. 2016.** Members of the genus Burkholderia: good and bad guys. *F1000Research* **5**: 1007.
- Edgar RC. 2013.** UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* **10**: 996–998.**Edgar RC. 2010.** Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011.** UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194–2200.
- Edwards J, Johnson C, Santos-Medellín C, Lurie E, Podishetty NK, Bhatnagar S, Eisen JA, Sundaresan V. 2015.** Structure, variation, and assembly of the root-associated microbiomes of rice. *Proceedings of the National Academy of Sciences* **112**: E911–E920.
- Edwards JA, Santos-medelli CM, Liechty ZS, Nguyen B, Lurie E, Eason S, Phillips G, Sundaresan V. 2018.** Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life cycle in field-grown rice. : 1–28.
- Eida AA, Ziegler M, Lafi FF, Michell CT, Voolstra CR, Hirt H, Saad MM. 2018.** Desert plant bacteria reveal host influence and beneficial plant growth properties (G Berta, Ed.). *PLOS ONE* **13**: e0208223.
- Emonet A, Zhou F, Vacheron J, Heiman CM, Déneraud Tendon V, Ma K-W, Schulze-Lefert P, Keel C, Geldner N. 2021.** Spatially Restricted Immune Responses Are Required for Maintaining Root Meristematic Activity upon Detection of Bacteria. *Current Biology* **31**: 1012-1028.e7.
- Enright AJ. 2002.** An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* **30**: 1575–1584.
- Faith DP. 1992.** Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61**: 1–10.
- Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, et al. 2013.** The Long-Term Stability of the Human Gut Microbiota. *Science* **341**: 1237439.
- Faust K. 2021.** Open challenges for microbial network construction and analysis. *The ISME Journal* **15**: 3111–3118.
- Faust K, Lima-Mendez G, Lerat J-S, Sathirapongsasuti JF, Knight R, Huttenhower C, Lenaerts T, Raes J. 2015.** Cross-biome comparison of microbial association networks. *Frontiers in Microbiology* **6**.
- Faust K, Raes J. 2016.** CoNet app: inference of biological association networks using Cytoscape. *F1000Research* **5**: 1519.
- Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C. 2012.** Microbial Co-occurrence Relationships in the Human Microbiome (CA Ouzounis, Ed.). *PLoS Computational Biology* **8**: e1002606.
- Field CB. 1998.** Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* **281**: 237–240.
- Fierer N. 2017.** Embracing the unknown: Disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology* **15**: 579–590.
- Fierer N, Bradford MA, Jackson RB. 2007.** Toward an ecological classification of soil bacteria. *Ecology* **88**: 1354–1364.

- Fierer N, Jackson RB. 2006.** The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences* **103**: 626–631.
- Finkel OM, Salas-González I, Castrillo G, Conway JM, Law TF, Teixeira PJPL, Wilson ED, Fitzpatrick CR, Jones CD, Dangl JL. 2020.** A single bacterial genus maintains root growth in a complex microbiome. *Nature* **587**: 103–108.
- Finkel OM, Salas-González I, Castrillo G, Spaepen S, Law TF, Teixeira PJPL, Jones CD, Dangl JL. 2019.** The effects of soil phosphorus content on plant microbiota are driven by the plant phosphate starvation response.
- Fitzpatrick CR, Copeland J, Wang PW, Guttman DS, Kotanen PM, Johnson MTJ. 2018.** Assembly and ecological function of the root microbiome across angiosperm plant species. *Proceedings of the National Academy of Sciences* **115**: E1157–E1165.
- Fließbach A, Mäder P, Oberholzer H-R, Gunst L. 2007.** Soil organic matter and biological soil quality indicators after 21 years of organic and conventional farming. *Agriculture, Ecosystems & Environment* **118**: 273–284.
- Fourcroy P, Sisó-Terraza P, Sudre D, Savirón M, Reyt G, Gaymard F, Abadía A, Abadía J, Álvarez-Fernández A, Briat J. 2014.** Involvement of the ABCG 37 transporter in secretion of scopoletin and derivatives by *Arabidopsis* roots in response to iron deficiency. *New Phytologist* **201**: 155–167.
- Fourcroy P, Tissot N, Gaymard F, Briat J-F, Dubos C. 2016.** Facilitated Fe Nutrition by Phenolic Compounds Excreted by the Arabidopsis ABCG37/PDR9 Transporter Requires the IRT1/FRO2 High-Affinity Root Fe 2+ Transport System. *Molecular Plant* **9**: 485–488.
- Francioli D, Schulz E, Lentendu G, Wubet T, Buscot F, Reitz T. 2016.** Mineral vs. Organic Amendments: Microbial Community Structure, Activity and Abundance of Agriculturally Relevant Microbes Are Driven by Long-Term Fertilization Strategies. *Frontiers in Microbiology* **7**: 1–16.
- Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM, Giannoukos G, Boylan MR, Ciulla D, Gevers D, et al. 2014.** Relating the metatranscriptome and metagenome of the human gut. *Proceedings of the National Academy of Sciences* **111**.
- Frerigmann H, Piślewska-Bednarek M, Sánchez-Vallet A, Molina A, Glawischnig E, Gigolashvili T, Bednarek P. 2016.** Regulation of Pathogen-Triggered Tryptophan Metabolism in *Arabidopsis thaliana* by MYB Transcription Factors and Indole Glucosinolate Conversion Products. *Molecular Plant* **9**: 682–695.
- Frey BJ, Dueck D. 2007.** Clustering by passing messages between data points. *Science* **315**: 972–976.
- Friedman J, Alm EJ. 2012.** Inferring Correlation Networks from Genomic Survey Data. *PLoS Computational Biology* **8**: 1–11.
- Fu H, Uchimiya M, Gore J, Moran MA. 2020.** Ecological drivers of bacterial community assembly in synthetic phycospheres. *Proceedings of the National Academy of Sciences* **117**: 3656–3662.
- Fukami T. 2015.** Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects. *Annual Review of Ecology, Evolution, and Systematics* **46**: 1–23.
- Garrido-Oter R, Nakano RT, Dombrowski N, Ma KW, McHardy AC, Schulze-Lefert P. 2018.** Modular Traits of the Rhizobiales Root Microbiota and Their Evolutionary Relationship with Symbiotic Rhizobia. *Cell Host and Microbe* **24**: 155-167.e5.
- Grant MA, Kazamia E, Cicuta P, Smith AG. 2014.** Direct exchange of vitamin B12 is demonstrated by modelling the growth dynamics of algal–bacterial cocultures. *The ISME Journal* **8**: 1418–1427.

- Grilli J. 2020.** Macroecological laws describe variation and diversity in microbial communities. *Nature Communications* **11**: 4743.
- Gupta S, Mortensen MS, Schjørring S, Trivedi U, Vestergaard G, Stokholm J, Bisgaard H, Krogfelt KA, Sørensen SJ. 2019.** Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing. *Communications Biology* **2**: 291.
- Gutjahr C, Casieri L, Paszkowski U. 2009.** Glomus intraradices induces changes in root system architecture of rice independently of common symbiosis signaling. *New Phytologist* **182**: 829–837.
- Gu Z, Eils R, Schlesner M. 2016.** Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**: 2847–2849.
- Hacquard S, Garrido-Oter R, González A, Spaepen S, Ackermann G, Lebeis S, McHardy AC, Dangl JL, Knight R, Ley R, et al. 2015.** Microbiota and host nutrition across plant and animal kingdoms. *Cell Host and Microbe* **17**: 603–616.
- 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.
- Harbort CJ, Hashimoto M, Inoue H, Niu Y, Guan R, Rombolà AD, Kopriva S, Voges MJEEE, Sattely ES, Garrido-Oter R, et al. 2020.** Root-Secreted Coumarins and the Microbiota Interact to Improve Iron Nutrition in Arabidopsis. *Cell Host and Microbe* **28**: 825-837.e6.
- Hartmann M, Frey B, Mayer J, Mäder P, Widmer F. 2015.** Distinct soil microbial diversity under long-term organic and conventional farming. *The ISME Journal* **9**: 1177–1194.
- Hartmann M, Widmer F. 2006.** Community structure analyses are more sensitive to differences in soil bacterial communities than anonymous diversity indices. *Applied and Environmental Microbiology* **72**: 7804–7812.
- Hassani MA, Durán P, Hacquard S. 2018.** Microbial interactions within the plant holobiont. *Microbiome* **6**: 58.
- van der Heijden MGA, Bardgett RD, van Straalen NM. 2008.** The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters* **11**: 296–310.
- Hernández M, Dumont MG, Yuan Q, Conrad R. 2015.** Different Bacterial Populations Associated with the Roots and Rhizosphere of Rice Incorporate Plant-Derived Carbon (CR Lovell, Ed.). *Applied and Environmental Microbiology* **81**: 2244–2253.
- Hindt MN, Akmakjian GZ, Pivarski KL, Punshon T, Baxter I, Salt DE, Guerinot M Lou. 2017.** BRUTUS and its paralogs, BTS LIKE1 and BTS LIKE2, encode important negative regulators of the iron deficiency response in Arabidopsis thaliana. *Metallomics* **9**: 876–890.
- Hiscox JD, Israelstam GF. 1979.** A method for the extraction of chlorophyll from leaf tissue without maceration. *Canadian Journal of Botany* **57**: 1332–1334.
- Hirano H, Takemoto K. 2019.** Difficulty in inferring microbial community structure based on co-occurrence network approaches. *BMC Bioinformatics* **20**: 329.
- Hiruma K, Gerlach N, Sacristán S, Nakano RT, Hacquard S, Kracher B, Neumann U, Ramírez D, Bucher M, O'Connell RJ, et al. 2016.** Root Endophyte Colletotrichum tofieldiae Confers Plant Fitness Benefits that Are Phosphate Status Dependent. *Cell* **165**: 464–474.
- Hornák K, Kasalický V, Šimek K, Grossart H-P. 2017.** Strain-specific consumption and transformation of alga-derived dissolved organic matter by members of the Limnohabitans -C and Polynucleobacter -B clusters of Betaproteobacteria. *Environmental Microbiology* **19**: 4519–4535.

- Horton TR, Bruns TD. 2001.** The molecular revolution in ectomycorrhizal ecology: peeking into the black-box. *Molecular Ecology* **10**: 1855–1871.
- Hou S, Wang X, Chen D, Yang X, Wang M, Turrà D, Di Pietro A, Zhang W. 2014.** The Secreted Peptide PIP1 Amplifies Immunity through Receptor-Like Kinase 7 (B Tyler, Ed.). *PLoS Pathogens* **10**: e1004331.
- Hu L, Robert CAM, Cadot S, Zhang X, Ye M, Li B, Manzo D, Chervet N, Steinger T, van der Heijden MGA, et al. 2018.** Root exudate metabolites drive plant–soil feedbacks on growth and defense by shaping the rhizosphere microbiota. *Nature Communications* **9**: 2738.
- Huang AC, Jiang T, Liu Y-X, Bai Y-C, Reed J, Qu B, Goossens A, Nützmann H-W, Bai Y, Osbourn A. 2019a.** A specialized metabolic network selectively modulates Arabidopsis root microbiota. *Science* **364**: eaau6389.
- Huang R, McGrath SP, Hirsch PR, Clark IM, Storkey J, Wu L, Zhou J, Liang Y. 2019b.** Plant–microbe networks in soil are weakened by century-long use of inorganic fertilizers. *Microbial Biotechnology* **12**: 1464–1475.
- Hugerth LW, Andersson AF. 2017.** Analysing Microbial Community Composition through Amplicon Sequencing: From Sampling to Hypothesis Testing. *Frontiers in Microbiology* **8**: 1–22.
- Huot B, Yao J, Montgomery BL, He SY. 2014.** Growth–Defense Tradeoffs in Plants: A Balancing Act to Optimize Fitness. *Molecular Plant* **7**: 1267–1287.
- Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl AM, Fitzgerald MG, Fulton RS, et al. 2012.** Structure, function and diversity of the healthy human microbiome. *Nature* **486**: 207–214.
- Isobe K, Allison SD, Khalili B, Martiny AC, Martiny JBH. 2019.** Phylogenetic conservation of bacterial responses to soil nitrogen addition across continents. *Nature Communications* **10**: 2499.
- Ivanov R, Brumbarova T, Bauer P. 2012.** Fitting into the Harsh Reality: Regulation of Iron-deficiency Responses in Dicotyledonous Plants. *Molecular Plant* **5**: 27–42.
- Jakoby M, Wang H-Y, Reidt W, Weisshaar B, Bauer P. 2004.** FRU (BHLH029) is required for induction of iron mobilization genes in Arabidopsis thaliana. *FEBS Letters* **577**: 528–534.
- Jacoby R, Peukert M, Succurro A, Koprivova A, Kopriva S. 2017.** The Role of Soil Microorganisms in Plant Mineral Nutrition—Current Knowledge and Future Directions. *Frontiers in Plant Science* **8**: 1–19.
- Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018.** High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* **9**: 5114.
- Jones JDG, Dangl JL. 2006.** The plant immune system. *Nature* **444**: 323–329.
- Jiang Y, Wang W, Xie Q, Liu N, Liu L, Wang D, Zhang X, Yang C, Chen X, Tang D, et al. 2017.** Plants transfer lipids to sustain colonization by mutualistic mycorrhizal and parasitic fungi. *Science* **356**: 1172–1175.
- Jones DL, Nguyen C, Finlay RD. 2009.** Carbon flow in the rhizosphere: carbon trading at the soil–root interface. *Plant and Soil* **321**: 5–33.
- Kai K, Mizutani M, Kawamura N, Yamamoto R, Tamai M, Yamaguchi H, Sakata K, Shimizu B. 2008.** Scopoletin is biosynthesized via ortho -hydroxylation of feruloyl CoA by a 2-oxoglutarate-dependent dioxygenase in Arabidopsis thaliana. *The Plant Journal* **55**: 989–999.
- Kai K, Shimizu B, Mizutani M, Watanabe K, Sakata K. 2006.** Accumulation of coumarins in Arabidopsis thaliana. *Phytochemistry* **67**: 379–386.

- Kamoun S, Furzer O, Jones JDG, Judelson HS, Ali GS, Dalio RJD, Roy SG, Schena L, Zambounis A, Panabières F, et al. 2015.** The Top 10 oomycete pathogens in molecular plant pathology. *Molecular Plant Pathology* **16**: 413–434.
- Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014.** Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research* **42**: D199–D205.
- Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019.** MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**: e7359.
- Karasov TL, Almario J, Friedemann C, Ding W, Giolai M, Heavens D, Kersten S, Lundberg DS, Neumann M, Regalado J, et al. 2018.** Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. *Cell Host & Microbe* **24**: 168-179.e4.
- Karasov TL, Neumann M, Duque-Jaramillo A, Kersten S, Bezrukov L, Schroepel B, Efthymia S, Lundberg DS, Regalado J, Shirsekar G, et al. 2019.** The relationship between microbial population size and disease in the Arabidopsis thaliana phyllosphere. *bioRxiv*.
- Keymer A, Pimprikar P, Wewer V, Huber C, Brands M, Bucerius SL, Delaux P-M, Klingl V, Röpenack-Lahaye E von, Wang TL, et al. 2017.** Lipid transfer from plants to arbuscular mycorrhiza fungi. *eLife* **6**: 1–33.
- Kim B-H, Ramanan R, Cho D-H, Oh H-M, Kim H-S. 2014.** Role of Rhizobium, a plant growth promoting bacterium, in enhancing algal biomass through mutualistic interaction. *Biomass and Bioenergy* **69**: 95–105.
- Kinnunen M, Dechesne A, Proctor C, Hammes F, Johnson D, Quintela-Baluja M, Graham D, Daffonchio D, Fodelianakis S, Hahn N, et al. 2016.** A conceptual framework for invasion in microbial communities. *The ISME Journal* **10**: 2773–2779.
- Klein AP, Sattely ES. 2017.** Biosynthesis of cabbage phytoalexins from indole glucosinolate. *Proceedings of the National Academy of Sciences* **114**: 1910–1915.
- Knack JJ, Wilcox LW, Delaux P-M, Ané J-M, Piotrowski MJ, Cook ME, Graham JM, Graham LE. 2015.** Microbiomes of Streptophyte Algae and Bryophytes Suggest That a Functional Suite of Microbiota Fostered Plant Colonization of Land. *International Journal of Plant Sciences* **176**: 405–420.
- Kodama Y, Shumway M, Leinonen R. 2012.** The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Research* **40**: D54–D56.
- Koenen EJM, Ojeda DI, Bakker FT, Wieringa JJ, Kidner C, Hardy OJ, Pennington RT, Herendeen PS, Bruneau A, Hughes CE. 2021.** The Origin of the Legumes is a Complex Paleopolyploid Phylogenomic Tangle Closely Associated with the Cretaceous–Paleogene (K–Pg) Mass Extinction Event (S Renner, Ed.). *Systematic Biology* **70**: 508–526.
- Koenen EJM, Ojeda DI, Steeves R, Migliore J, Bakker FT, Wieringa JJ, Kidner C, Hardy O, Pennington RT, Herendeen PS, et al. 2019.** The Origin and Early Evolution of the Legumes are a Complex Paleopolyploid Phylogenomic Tangle closely associated with the Cretaceous-Paleogene (K-Pg) Boundary. *bioRxiv*.
- Koprivova A, Schuck S, Jacoby RP, Klinkhammer I, Welter B, Leson L, Martyn A, Nauen J, Grabenhorst N, Mandelkow JF, et al. 2019.** Root-specific camalexin biosynthesis controls the plant growth-promoting effects of multiple bacterial strains. *Proceedings of the National Academy of Sciences* **116**: 15735–15744.
- Krause AE, Frank KA, Mason DM, Ulanowicz RE, Taylor WW. 2003.** Compartments revealed in food-web structure. *Nature* **426**: 282–285.

- Kremer JM, Paasch BC, Rhodes D, Thireault C, Froehlich JE, Schulze-Lefert P, Tiedje JM, Sheng Yang He.** FlowPot axenic plant growth system for microbiota research. **Kremer JM, Sohrabi R, Paasch BC, Rhodes D, Thireault C, Schulze-Lefert P, Tiedje JM, He SY.** 2021. Peat-based gnotobiotic plant growth systems for Arabidopsis microbiome research. *Nature Protocols* **16**: 2450–2470.
- Krieger C, Roselli S, Kellner-Thielmann S, Galati G, Schneider B, Grosjean J, Olry A, Ritchie D, Matern U, Bourgaud F, et al.** 2018. The CYP71AZ P450 Subfamily: A Driving Factor for the Diversification of Coumarin Biosynthesis in Apiaceous Plants. *Frontiers in Plant Science* **9**.
- Krohn-Molt I, Alawi M, Förstner KU, Wiegandt A, Burkhardt L, Indenbirken D, Thieß M, Grundhoff A, Kehr J, Tholey A, et al.** 2017. Insights into Microalga and Bacteria Interactions of Selected Phycosphere Biofilms Using Metagenomic, Transcriptomic, and Proteomic Approaches. *Frontiers in Microbiology* **8**.
- Kropat J, Hong-Hermesdorf A, Casero D, Ent P, Castruita M, Pellegrini M, Merchant SS, Malasarn D.** 2011. A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*. *The Plant Journal* **66**: 770–780.
- Kudjordjie EN, Sapkota R, Steffensen SK, Fomsgaard IS, Nicolaisen M.** 2019. Maize synthesized benzoxazinoids affect the host associated microbiome. *Microbiome* **7**: 59.
- Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA.** 2015. Sparse and Compositionally Robust Inference of Microbial Ecological Networks (C von Mering, Ed.). *PLOS Computational Biology* **11**: e1004226.
- Kvitko BH, Collmer A.** 2011. Construction of *Pseudomonas syringae* pv. tomato DC3000 Mutant and Polymutant Strains. In: 109–128.
- Lapébie P, Lombard V, Drula E, Terrapon N, Henrissat B.** 2019. Bacteroidetes use thousands of enzyme combinations to break down glycans. *Nature Communications* **10**: 2043.
- Lebeis SL, Paredes SH, Lundberg DS, Breakfield N, Gehring J, McDonald M, Malfatti S, Glavina del Rio T, Jones CD, Tringe SG, et al.** 2015. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* **349**: 860–864.
- Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD.** 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**: 882–883.
- Lemanceau P, Blouin M, Muller D, Moënne-Loccoz Y.** 2017. Let the Core Microbiota Be Functional. *Trends in Plant Science* **22**: 583–595.
- Letunic I, Bork P.** 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**: 127–128.
- Levy A, Salas Gonzalez I, Mittelviehhaus M, Clingenpeel S, Herrera Paredes S, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F, et al.** 2018. Genomic features of bacterial adaptation to plants. *Nature Genetics* **50**: 138–150.
- Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, Arumugam M, Kultima JR, Prifti E, Nielsen T, et al.** 2014. An integrated catalog of reference genes in the human gut microbiome. *Nature Biotechnology* **32**: 834–841.
- Lichtblau Y, Zimmermann K, Haldemann B, Lenze D, Hummel M, Leser U.** 2016. Comparative assessment of differential network analysis methods. *Briefings in Bioinformatics*: bbw061.
- Lima LFO, Weissman M, Reed M, Papudeshi B, Alker AT, Morris MM, Edwards RA, de Putron SJ, Vaidya NK, Dinsdale EA.** 2020. Modeling of the Coral Microbiome: the Influence of Temperature and Microbial Network (M Medina and MJ McFall-Ngai, Eds.). *mBio* **11**.

- Lindsay WL, Norvell WA. 1978.** Development of a DTPA Soil Test for Zinc, Iron, Manganese, and Copper. *Soil Science Society of America Journal* **42**: 421–428.
- Litchman E. 2010.** Invisible invaders: non-pathogenic invasive microbes in aquatic and terrestrial ecosystems. *Ecology Letters* **13**: 1560–1572.
- Lohmann GV, Shimoda Y, Nielsen MW, Jørgensen FG, Grossmann C, Sandal N, Sørensen K, Thirup S, Madsen LH, Tabata S, et al. 2010.** Evolution and Regulation of the Lotus japonicus LysM Receptor Gene Family. *Molecular Plant-Microbe Interactions* **23**: 510–521.
- López-Bucio J, Campos-Cuevas JC, Hernández-Calderón E, Velásquez-Becerra C, Farías-Rodríguez R, Macías-Rodríguez LI, Valencia-Cantero E. 2007.** Bacillus megaterium Rhizobacteria Promote Growth and Alter Root-System Architecture Through an Auxin- and Ethylene-Independent Signaling Mechanism in Arabidopsis thaliana. *Molecular Plant-Microbe Interactions* **20**: 207–217.
- Lozupone C, Hamady M, Knight R. 2006.** UniFrac - An online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7**: 1–14.
- Love MI, Huber W, Anders S. 2014.** Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550.
- Lozupone CA, Knight R. 2007.** Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 11436–11440.
- Luginbuehl LH, Menard GN, Kurup S, Van Erp H, Radhakrishnan G V., Breakspear A, Oldroyd GED, Eastmond PJ. 2017.** Fatty acids in arbuscular mycorrhizal fungi are synthesized by the host plant. *Science* **356**: 1175–1178.
- Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, Tremblay J, Engelbrektson A, Kunin V, Rio TG Del, et al. 2012.** Defining the core Arabidopsis thaliana root microbiome. *Nature* **488**: 86–90.
- Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangl JL. 2013.** Practical innovations for high-throughput amplicon sequencing. *Nature Methods* **10**: 999–1002.
- Lynch JM, de Leij F. 2012.** Rhizosphere. In: eLS. Chichester, UK: John Wiley & Sons, Ltd.
- Ma B, Wang Y, Ye S, Liu S, Stirling E, Gilbert JA, Faust K, Knight R, Jansson JK, Cardona C, et al. 2020.** Earth microbial co-occurrence network reveals interconnection pattern across microbiomes. *Microbiome* **8**: 1–12.
- MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ. 2010.** Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**: 966–968.
- Madsen EB, Madsen LH, Radutoiu S, Olbryt M, Rakwalska M, Szczyglowski K, Sato S, Kaneko T, Tabata S, Sandal N, et al. 2003.** A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* **425**: 637–640.
- Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, Brochado AR, Fernandez KC, Dose H, Mori H, et al. 2018.** Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* **555**: 623–628.
- Mai H-J, Pateyron S, Bauer P. 2016.** Iron homeostasis in Arabidopsis thaliana: transcriptomic analyses reveal novel FIT-regulated genes, iron deficiency marker genes and functional gene networks. *BMC Plant Biology* **16**: 211.

- Mamet SD, Redlick E, Brabant M, Lamb EG, Helgason BL, Stanley K, Siciliano SD. 2019.** Structural equation modeling of a winnowed soil microbiome identifies how invasive plants re-structure microbial networks. *The ISME Journal* **13**: 1988–1996.
- Mandakovic D, Rojas C, Maldonado J, Latorre M, Travisany D, Delage E, Bihouée A, Jean G, Díaz FP, Fernández-Gómez B, et al. 2018.** Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience. *Scientific Reports* **8**: 5875.
- Margulis L. 1991.** *Symbiosis as a source of evolutionary innovation: Speciation and morphogenesis*. Cambridge, MA: MIT Press.
- Marschner H, Römheld V, Horst WJ, Martin P. 1986.** Root-induced changes in the rhizosphere: Importance for the mineral nutrition of plants. *Zeitschrift für Pflanzenernährung und Bodenkunde* **149**: 441–456.
- Martin FM, Uroz S, Barker DG. 2017.** Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science* **356**: eaad4501.
- Martiny AC. 2019.** High proportions of bacteria are culturable across major biomes. *The ISME Journal* **13**: 2125–2128.
- McClure R, Naylor D, Farris Y, Davison M, Fansler SJ, Hofmockel KS, Jansson JK. 2020.** Development and Analysis of a Stable, Reduced Complexity Model Soil Microbiome. *Frontiers in Microbiology* **11**.
- McMurdie PJ, Holmes S. 2014.** Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible (AC McHardy, Ed.). *PLoS Computational Biology* **10**: e1003531.
- Mendes R, Kruijt M, De Bruijn I, Dekkers E, Van Der Voort M, Schneider JHM, Piceno YM, DeSantis TZ, Andersen GL, Bakker PAHM, et al. 2011.** Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* **332**: 1097–1100.
- Millet YA, Danna CH, Clay NK, Songnuan W, Simon MD, Werck-Reichhart D, Ausubel FM. 2010.** Innate Immune Responses Activated in Arabidopsis Roots by Microbe-Associated Molecular Patterns. *The Plant Cell* **22**: 973–990.
- Moran MA, Kujawinski EB, Stubbins A, Fatland R, Aluwihare LI, Buchan A, Crump BC, Dorrestein PC, Dyhrman ST, Hess NJ, et al. 2016.** Deciphering ocean carbon in a changing world. *Proceedings of the National Academy of Sciences* **113**: 3143–3151.
- Moreno C, Romero J, Espejo RT. 2002.** Polymorphism in repeated 16S rRNA genes is a common property of type strains and environmental isolates of the genus *Vibrio* The GenBank accession numbers for the sequences reported in this paper are AF388386 (Vp23), AF388387 (Vp16), AF388388 (F44), AF3883. *Microbiology* **148**: 1233–1239.
- Morrissey J, Guerinot M Lou. 2009.** Iron Uptake and Transport in Plants: The Good, the Bad, and the Ionome. *Chemical Reviews* **109**: 4553–4567.
- Mun T, Bachmann A, Gupta V, Stougaard J, Andersen SU. 2016.** Lotus Base: An integrated information portal for the model legume *Lotus japonicus*. *Scientific Reports* **6**: 39447.
- Musso G, Gambino R, Cassader M. 2011.** Interactions Between Gut Microbiota and Host Metabolism Predisposing to Obesity and Diabetes. *Annual Review of Medicine* **62**: 361–380.
- Nichols TE, Holmes AP. 2002.** Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping* **15**: 1–25.

- Nilsson RH, Larsson K-H, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glöckner FO, Tedersoo L, et al. 2019.** The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research* **47**: D259–D264.
- Ojeda DI, Santos-Guerra A, Oliva-Tejera F, Jaen-Molina R, Caujape-Castells J, Marrero-Rodriguez A, Cronk Q. 2014.** DNA barcodes successfully identified Macaronesian Lotus (Leguminosae) species within early diverged lineages of Cape Verde and mainland Africa. *AoB PLANTS* **6**: plu050–plu050.
- Oksanen J, Guillaume Blanchet F, Friendly M, Kindt R, Legendre P, McGlenn D, R. Minchin P, O’Hara RB, L. Simpson G, Solymos P, et al. 2019.** vegan: Community Ecology Package.
- Olesen JM, Bascompte J, Dupont YL, Jordano P. 2007.** The modularity of pollination networks. *Proceedings of the National Academy of Sciences* **104**: 19891–19896.
- Paerl RW, Bouget F-Y, Lozano J-C, Vergé V, Schatt P, Allen EE, Palenik B, Azam F. 2017.** Use of plankton-derived vitamin B1 precursors, especially thiazole-related precursor, by key marine picoeukaryotic phytoplankton. *The ISME Journal* **11**: 753–765.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015.** CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **25**: 1043–1055.
- Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P, et al. 2019.** Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* **176**: 649-662.e20.
- Pastorczyk M, Bednarek P. 2016.** *Advances in Botanical Research*. Elsevier.
- Pel MJC, van Dijken AJH, Bardoel BW, Seidl MF, van der Ent S, van Strijp JAG, Pieterse CMJ. 2014.** Pseudomonas syringae Evades Host Immunity by Degrading Flagellin Monomers with Alkaline Protease AprA. *Molecular Plant-Microbe Interactions* **27**: 603–610.
- Peiffer JA, Spor A, Koren O, Jin Z, Tringe SG, Dangl JL, Buckler ES, Ley RE. 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**: 6548–6553.
- Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012.** IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420–1428.
- Peschel S, Müller CL, von Mutius E, Boulesteix AL, Depner M. 2021.** NetCoMi: network construction and comparison for microbiome data in R. *Briefings in bioinformatics* **22**: 1–18.
- Pfennig N. 1978.** Rhodocyclus purpureus gen. nov. and sp. nov., a Ring-Shaped, Vitamin B12-Requiring Member of the Family Rhodospirillaceae. *International Journal of Systematic Bacteriology* **28**: 283–288.
- Pii Y, Mimmo T, Tomasi N, Terzano R, Cesco S, Crecchio C. 2015.** Microbial interactions in the rhizosphere: beneficial influences of plant growth-promoting rhizobacteria on nutrient acquisition process. A review. *Biology and Fertility of Soils* **51**: 403–415.
- Poncini L, Wyrsh I, Déneraud Tendon V, Vorley T, Boller T, Geldner N, Métraux J-P, Lehmann S. 2017.** In roots of Arabidopsis thaliana, the damage-associated molecular pattern AtPep1 is a stronger elicitor of immune signalling than flg22 or the chitin heptamer (OA Zobotina, Ed.). *PLOS ONE* **12**: e0185808.
- Porra RJ, Thompson WA, Kriedemann PE. 1989.** Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **975**: 384–394.

- Preece TF, Dickinson CH. 1971.** Ecology of leaf surface microorganisms.
- Price MN, Dehal PS, Arkin AP. 2010.** FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments (AFY Poon, Ed.). *PLoS ONE* **5**: e9490.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012.** The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**: D590–D596.
- Raes J, Bork P. 2008.** Molecular eco-systems biology: Towards an understanding of community function. *Nature Reviews Microbiology* **6**: 693–699.
- Rajniak J, Giehl RFH, Chang E, Murgia I, von Wirén N, Sattely ES. 2018.** Biosynthesis of redox-active metabolites in response to iron deficiency in plants. *Nature Chemical Biology* **14**: 442–450.
- Rappsilber J, Ishihama Y, Mann M. 2003.** Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Analytical Chemistry* **75**: 663–670.
- Richards TA, Soanes DM, Jones MDM, Vasieva O, Leonard G, Paszkiewicz K, Foster PG, Hall N, Talbot NJ. 2011.** Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 15258–15263.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015.** limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**: e47–e47.
- 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**: 24–34.
- Robinson MD, McCarthy DJ, Smyth GK. 2010.** edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
- Robinson NJ, Procter CM, Connolly EL, Guerinot M Lou. 1999.** A ferric-chelate reductase for iron uptake from soils. *Nature* **397**: 694–697.
- Rodríguez-Celma J, Lin W-D, Fu G-M, Abadia J, Lopez-Millan A-F, Schmidt W. 2013.** Mutually Exclusive Alterations in Secondary Metabolism Are Critical for the Uptake of Insoluble Iron Compounds by Arabidopsis and Medicago truncatula. *PLANT PHYSIOLOGY* **162**: 1473–1485.
- Rodríguez-Celma J, Schmidt W. 2013.** Reduction-based iron uptake revisited. *Plant Signaling & Behavior* **8**: e26116.
- Santi S, Schmidt W. 2009.** Dissecting iron deficiency-induced proton extrusion in Arabidopsis roots. *New Phytologist* **183**: 1072–1084.
- Rogers ED, Benfey PN. 2015.** Regulation of plant root system architecture: Implications for crop advancement. *Current Opinion in Biotechnology* **32**: 93–98.
- Rosenberg E, Koren O, Reshef L, Efrony R, Zilber-Rosenberg I. 2007.** The role of microorganisms in coral health, disease and evolution. *Nature Reviews Microbiology* **5**: 355–362.
- Röttjers L, Faust K. 2018.** From hairballs to hypotheses—biological insights from microbial networks. *FEMS Microbiology Reviews* **42**: 761–780.
- Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, Knight R, Fierer N. 2010.** Soil bacterial and fungal communities across a pH gradient in an arable soil. *The ISME Journal* **4**: 1340–1351.

- Roux S, Enault F, Bronner G, Debroas D. 2011.** Comparison of 16S rRNA and protein-coding genes as molecular markers for assessing microbial diversity (Bacteria and Archaea) in ecosystems. *FEMS Microbiology Ecology* **78**: 617–628.
- De Roy K, Marzorati M, Van den Abbeele P, Van de Wiele T, Boon N. 2014.** Synthetic microbial ecosystems: an exciting tool to understand and apply microbial communities. *Environmental Microbiology* **16**: 1472–1481.
- Saijo Y, Loo EP, Yasuda S. 2018.** Pattern recognition receptors and signaling in plant-microbe interactions. *The Plant Journal* **93**: 592–613.
- Salas-González I, Rey G, Flis P, Custódio V, Gopaulchan D, Bakhom N, Dew TP, Suresh K, Franke RB, Dangel JL, et al. 2021.** Coordination between microbiota and root endodermis supports plant mineral nutrient homeostasis. *Science* **371**.
- Saleem M, Hu J, Jousset A. 2019.** More Than the Sum of Its Parts: Microbiome Biodiversity as a Driver of Plant Growth and Soil Health. *Annual Review of Ecology, Evolution, and Systematics* **50**: 145–168.
- Sasso S, Stibor H, Mittag M, Grossman AR. 2018.** From molecular manipulation of domesticated *Chlamydomonas reinhardtii* to survival in nature. *eLife* **7**.
- Schlaeppli K, Dombrowski N, Oter RG, Ver Loren van Themaat E, Schulze-Lefert P. 2014.** Quantitative divergence of the bacterial root microbiota in *Arabidopsis thaliana* relatives. *Proceedings of the National Academy of Sciences* **111**: 585–592.
- Schmid NB, Giehl RFH, Doll S, Mock H-P, Strehmel N, Scheel D, Kong X, Hider RC, von Wiren N. 2014.** Feruloyl-CoA 6'-Hydroxylase 1-Dependent Coumarins Mediate Iron Acquisition from Alkaline Substrates in *Arabidopsis*. *PLANT PHYSIOLOGY* **164**: 160–172.
- Schmidt H, Günther C, Weber M, Spörlein C, Loscher S, Böttcher C, Schobert R, Clemens S. 2014.** Metabolome Analysis of *Arabidopsis thaliana* Roots Identifies a Key Metabolic Pathway for Iron Acquisition (G Muday, Ed.). *PLoS ONE* **9**: e102444.
- Schmidt JE, Kent AD, Brisson VL, Gaudin ACM. 2019.** Agricultural management and plant selection interactively affect rhizosphere microbial community structure and nitrogen cycling. : 1–18.
- Schmidt TSB, Matias Rodrigues JF, Von Mering C. 2017.** A family of interaction-adjusted indices of community similarity. *ISME Journal* **11**: 791–807.
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W. 2012.** Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for <em>Fungi</em>. *Proceedings of the National Academy of Sciences* **109**: 6241 LP – 6246.
- Seemann T. 2014.** Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–2069.
- Selote D, Samira R, Matthiadis A, Gillikin JW, Long TA. 2014.** Iron-Binding E3 Ligase Mediates Iron Response in Plants by Targeting Basic Helix-Loop-Helix Transcription Factors. *Plant Physiology* **167**: 273–286.
- Seymour JR, Amin SA, Raina J-B, Stocker R. 2017.** Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. *Nature Microbiology* **2**: 17065.
- Shade A, Handelsman J. 2012.** Beyond the Venn diagram: the hunt for a core microbiome. *Environmental Microbiology* **14**: 4–12.
- Shade A, Stopnisek N. 2019.** Abundance-occupancy distributions to prioritize plant core microbiome membership. *Current Opinion in Microbiology* **49**: 50–58.

- Shannon CE. 1948.** A Mathematical Theory of Communication. *Bell System Technical Journal* **27**: 379–423.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003.** Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13**: 2498–2504.
- Shi S, Nuccio E, Herman DJ, Rijkers R, Estera K, Li J, Da Rocha UN, He Z, Pett-Ridge J, Brodie EL, et al. 2015.** Successional trajectories of rhizosphere bacterial communities over consecutive seasons. *mBio* **6**: 13–20.
- Shibl AA, Isaac A, Ochsenkühn MA, Cárdenas A, Fei C, Behringer G, Arnoux M, Drou N, Santos MP, Gunsalus KC, et al. 2020.** Diatom modulation of select bacteria through use of two unique secondary metabolites. *Proceedings of the National Academy of Sciences* **117**: 27445–27455.
- Shinya T, Yamaguchi K, Desaki Y, Yamada K, Narisawa T, Kobayashi Y, Maeda K, Suzuki M, Tanimoto T, Takeda J, et al. 2014.** Selective regulation of the chitin-induced defense response by the Arabidopsis receptor-like cytoplasmic kinase PBL27. *The Plant Journal* **79**: 56–66.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al. 2011.** Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**: 539.
- Silva-Navas J, Moreno-Risueno MA, Manzano C, Pallero-Baena M, Navarro-Neila S, Téllez-Robledo B, Garcia-Mina JM, Baigorri R, Gallego FJ, del Pozo JC. 2015.** D-Root: a system for cultivating plants with the roots in darkness or under different light conditions. *The Plant Journal* **84**: 244–255.
- Simmons T, Styer AB, Pierroz G, Gonçalves AP, Pasricha R, Hazra AB, Bubner P, Coleman-Derr D. 2020.** Drought Drives Spatial Variation in the Millet Root Microbiome. *Frontiers in Plant Science* **11**.
- Simpson EH. 1949.** Measurement of Diversity. *Nature* **163**: 688–688.
- Sisó-Terraza P, Luis-Villarroya A, Fourcroy P, Briat J-F, Abadía A, Gaymard F, Abadía J, Álvarez-Fernández A. 2016.** Accumulation and Secretion of Coumarinolignans and other Coumarins in Arabidopsis thaliana Roots in Response to Iron Deficiency at High pH. *Frontiers in Plant Science* **7**.
- Siwinska J, Kadzinski L, Banasiuk R, Gwizdek-Wisniewska A, Olry A, Banecki B, Lojkowska E, Ihnatowicz A. 2014.** Identification of QTLs affecting scopolin and scopoletin biosynthesis in Arabidopsis thaliana. *BMC Plant Biology* **14**: 280.
- Siwinska J, Siatkowska K, Olry A, Grosjean J, Hehn A, Bourgaud F, Meharg AA, Carey M, Lojkowska E, Ihnatowicz A. 2018.** Scopoletin 8-hydroxylase: a novel enzyme involved in coumarin biosynthesis and iron-deficiency responses in Arabidopsis. *Journal of Experimental Botany* **69**: 1735–1748.
- Smith KP, Goodman RM. 1999.** Host variation for interactions with beneficial plant-associated microbes. *Annual Review of Phytopathology* **37**: 473–491.
- Smith SE, Smith FA, Jakobsen I. 2003.** Mycorrhizal Fungi Can Dominate Phosphate Supply to Plants Irrespective of Growth Responses. *Plant Physiology* **133**: 16–20.
- Smriga S, Fernandez VI, Mitchell JG, Stocker R. 2016.** Chemotaxis toward phytoplankton drives organic matter partitioning among marine bacteria. *Proceedings of the National Academy of Sciences* **113**: 1576–1581.
- Soneson C, Love MI, Robinson MD. 2016.** Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4**: 1521.
- Sokal RR, Sneath PHA. 1963.** *Principles of Numerical Taxonomy*. W. H. Freeman.

- Steiner JJ, Santos GG. 2001.** Adaptive Ecology of *Lotus corniculatus* L. Genotypes: I. Plant Morphology and RAPD Marker Characterizations. *Crop Science* **41**: 552–563.
- Stringlis IA, Yu K, Feussner K, de Jonge R, Van Bentum S, Van Verk MC, Berendsen RL, Bakker PAHM, Feussner I, Pieterse CMJ. 2018.** MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health. *Proceedings of the National Academy of Sciences* **115**: E5213–E5222.
- Suárez-Moreno ZR, Caballero-Mellado J, Coutinho BG, Mendonça-Previato L, James EK, Venturi V. 2012.** Common Features of Environmental and Potentially Beneficial Plant-Associated Burkholderia. *Microbial Ecology* **63**: 249–266.
- Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, et al. 2015.** Structure and function of the global ocean microbiome. *Science* **348**: 1261359–1261359.
- Tai H, Lu X, Opitz N, Marcon C, Paschold A, Lithio A, Nettleton D, Hochholdinger F. 2016.** Transcriptomic and anatomical complexity of primary, seminal, and crown roots highlight root type-specific functional diversity in maize (*Zea mays* L.). *Journal of Experimental Botany* **67**: 1123–1135.
- Teixeira PJPL, Colaianni NR, Law TF, Conway JM, Gilbert S, Li H, Salas-González I, Panda D, Del Risco NM, Finkel OM, et al. 2021.** Specific modulation of the root immune system by a community of commensal bacteria. *Proceedings of the National Academy of Sciences* **118**: e2100678118.
- Teixeira PJP, Colaianni NR, Fitzpatrick CR, Dangl JL. 2019.** Beyond pathogens: microbiota interactions with the plant immune system. *Current Opinion in Microbiology* **49**: 7–17.
- Teplitski M, Chen H, Rajamani S, Gao M, Merighi M, Sayre RT, Robinson JB, Rolfe BG, Bauer WD. 2004.** *Chlamydomonas reinhardtii* Secretes Compounds That Mimic Bacterial Signals and Interfere with Quorum Sensing Regulation in Bacteria. *Plant Physiology* **134**: 137–146.
- Thiergart T, Durán P, Ellis T, Vannier N, Garrido-Oter R, Kemen E, Roux F, Alonso-Blanco C, Ågren J, Schulze-Lefert P, et al. 2020.** Root microbiota assembly and adaptive differentiation among European *Arabidopsis* populations. *Nature Ecology and Evolution* **4**: 122–131.
- Thiergart T, Zgadzaj R, Bozsóki Z, Garrido-Oter R, Radutoiu S, Schulze-Lefert P. 2019.** *Lotus japonicus* Symbiosis Genes Impact Microbial Interactions between Symbionts and Multikingdom Commensal Communities (FM Ausubel, Ed.). *mBio* **10**: 15.
- Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, et al. 2017.** A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature* **551**: 457–463.
- Timilsina S, Potnis N, Newberry EA, Liyanapathirana P, Iruegas-Bocardo F, White FF, Goss EM, Jones JB. 2020.** *Xanthomonas* diversity, virulence and plant–pathogen interactions. *Nature Reviews Microbiology* **18**: 415–427.
- Toyama T, Kasuya M, Hanaoka T, Kobayashi N, Tanaka Y, Inoue D, Sei K, Morikawa M, Mori K. 2018.** Growth promotion of three microalgae, *Chlamydomonas reinhardtii*, *Chlorella vulgaris* and *Euglena gracilis*, by in situ indigenous bacteria in wastewater effluent. *Biotechnology for Biofuels* **11**: 176.
- Tritt A, Eisen JA, Facciotti MT, Darling AE. 2012.** An Integrated Pipeline for de Novo Assembly of Microbial Genomes (D Zhu, Ed.). *PLoS ONE* **7**: e42304.
- Tsai H-H, Rodríguez-Celma J, Lan P, Wu Y-C, Vélez-Bermúdez IC, Schmidt W. 2018.** Scopoletin 8-Hydroxylase-Mediated Fraxetin Production Is Crucial for Iron Mobilization. *Plant Physiology* **177**: 194–207.

- Tsuda K, Sato M, Stoddard T, Glazebrook J, Katagiri F. 2009.** Network Properties of Robust Immunity in Plants (GP Copenhaver, Ed.). *PLoS Genetics* **5**: e1000772.
- Turnbaugh PJ, Hamady M, Yatsunencko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, et al. 2009.** A core gut microbiome in obese and lean twins. *Nature* **457**: 480–484.
- Unterseer S, Pophaly SD, Peis R, Westermeier P, Mayer M, Seidel MA, Haberger G, Mayer KFX, Ordas B, Pausch H, et al. 2016.** A comprehensive study of the genomic differentiation between temperate Dent and Flint maize. *Genome Biology* **17**: 137.
- Van de Peer Y, Mizrachi E, Marchal K. 2017.** The evolutionary significance of polyploidy. *Nature Reviews Genetics* **18**: 411–424.
- Venturelli OS, Carr A V, Fisher G, Hsu RH, Lau R, Bowen BP, Hromada S, Northen T, Arkin AP. 2018.** Deciphering microbial interactions in synthetic human gut microbiome communities. *Molecular Systems Biology* **14**.
- Verbon EH, Trapet PL, Kruijs S, Temple-Boyer-Dury C, Rouwenhorst TG, Pieterse CMJ. 2019.** Rhizobacteria-Mediated Activation of the Fe Deficiency Response in Arabidopsis Roots: Impact on Fe Status and Signaling. *Frontiers in Plant Science* **10**.
- Vert G, Grotz N, Dédaldéchamp F, Gaymard F, Guerinot M Lou, Briat J-F, Curie C. 2002.** IRT1, an Arabidopsis Transporter Essential for Iron Uptake from the Soil and for Plant Growth. *The Plant Cell* **14**: 1223–1233.
- Větrovský T, Baldrian P. 2013.** The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses (J Neufeld, Ed.). *PLoS ONE* **8**: e57923.
- Voges MJEEE, Bai Y, Schulze-Lefert P, Sattely ES. 2019.** Plant-derived coumarins shape the composition of an Arabidopsis synthetic root microbiome. *Proceedings of the National Academy of Sciences* **116**: 12558–12565.
- Vorholt JA. 2012.** Microbial life in the phyllosphere. *Nature Reviews Microbiology* **10**: 828–840.
- Vorholt JA, Vogel C, Carlström CI, Müller DB. 2017.** Establishing Causality: Opportunities of Synthetic Communities for Plant Microbiome Research. *Cell Host & Microbe* **22**: 142–155.
- Vrancken G, Gregory AC, Huys GRB, Faust K, Raes J. 2019.** Synthetic ecology of the human gut microbiota. *Nature Reviews Microbiology* **17**: 754–763.
- Wagner MR, Lundberg DS, Del Rio TG, Tringe SG, Dangl JL, Mitchell-Olds T. 2016.** Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nature Communications* **7**.
- Wagner MR, Roberts JH, Balint-Kurti P, Holland JB. 2020.** Heterosis of leaf and rhizosphere microbiomes in field-grown maize. *New Phytologist* **228**: 1055–1069.
- Walker TS, Bais HP, Grotewold E, Vivanco JM. 2003.** Root Exudation and Rhizosphere Biology. *Plant Physiology* **132**: 44–51.
- Walters WA, Jin Z, Youngblut N, Wallace JG, Sutter J, Zhang W, González-Peña A, Peiffer J, Koren O, Shi Q, et al. 2018.** Large-scale replicated field study of maize rhizosphere identifies heritable microbes. *Proceedings of the National Academy of Sciences of the United States of America* **115**: 7368–7373.
- Wan W-L, Fröhlich K, Pruitt RN, Nürnberger T, Zhang L. 2019.** Plant cell surface immune receptor complex signaling. *Current Opinion in Plant Biology* **50**: 18–28.
- Wang H, Zou D, Xie K, Xie M. 2014.** Antibacterial mechanism of fraxetin against *Staphylococcus aureus*. *Molecular Medicine Reports* **10**: 2341–2345.

- Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007.** Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* **73**: 5261–5267.
- Watts SC, Ritchie SC, Inouye M, Holt KE. 2019.** FastSpar: rapid and scalable correlation estimation for compositional data (O Stegle, Ed.). *Bioinformatics* **35**: 1064–1066.
- Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, Xia LC, Xu ZZ, Ursell L, Alm EJ, et al. 2016.** Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *The ISME Journal* **10**: 1669–1681.
- Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, Lozupone C, Zaneveld JR, Vázquez-Baeza Y, Birmingham A, et al. 2017.** Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**: 27.
- White EP, Ernest SKM, Kerkhoff AJ, Enquist BJ. 2007.** Relationships between body size and abundance in ecology. *Trends in Ecology & Evolution* **22**: 323–330.
- Whitman WB, Coleman DC, Wiebe WJ. 1998.** Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences* **95**: 6578–6583.
- Whittaker RH. 1960.** Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs* **30**: 279–338.
- Wichard T, Charrier B, Mineur F, Bothwell JH, Clerck O De, Coates JC. 2015.** The green seaweed *Ulva*: a model system to study morphogenesis. *Frontiers in Plant Science* **6**.
- Wickham H. 2016.** ggplot2: Elegant Graphics for Data Analysis.
- Widmer F, Rasche F, Hartmann M, Fliessbach A. 2006.** Community structures and substrate utilization of bacteria in soils from organic and conventional farming systems of the DOK long-term field experiment. *Applied Soil Ecology* **33**: 294–307.
- Wienhausen G, Noriega-Ortega BE, Niggemann J, Dittmar T, Simon M. 2017.** The Exometabolome of Two Model Strains of the *Roseobacter* Group: A Marketplace of Microbial Metabolites. *Frontiers in Microbiology* **8**.
- Willis AD. 2019.** Rarefaction, Alpha Diversity, and Statistics. *Frontiers in Microbiology* **10**.
- Willmann M, Gerlach N, Buer B, Polatajko A, Nagy R, Koebke E, Jansa J, Flisch R, Bucher M. 2013.** Mycorrhizal phosphate uptake pathway in maize: vital for growth and cob development on nutrient poor agricultural and greenhouse soils. *Frontiers in Plant Science* **4**: 533.
- Wilson ACC, Duncan RP. 2015.** Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. *Proceedings of the National Academy of Sciences* **112**: 10255–10261.
- Wippel K, Tao K, Niu Y, Zgadzaj R, Kiel N, Guan R, Dahms E, Zhang P, Jensen DB, Logemann E, et al. 2021.** Host preference and invasiveness of commensal bacteria in the Lotus and Arabidopsis root microbiota. *Nature Microbiology*.
- Woese CR, Fox GE. 1977.** Phylogenetic structure of the prokaryotic domain: The primary kingdoms (archaeobacteria/eubacteria/urkaryote/16S ribosomal RNA/molecular phylogeny). *Proceedings of the National Academy of Sciences of the United States of America* **74**: 5088–5090.
- Woese CR, Kandler O, Wheelis ML. 1990.** Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America* **87**: 4576–4579.
- Wrb IWG. 2015.** World reference base for soil resources 2014. *Update*: 192.

- Wu G, Zhao N, Zhang C, Lam YY, Zhao L. 2021.** Guild-based analysis for understanding gut microbiome in human health and diseases. *Genome Medicine* **13**: 22.
- Wu M, Eisen JA. 2008.** A simple, fast, and accurate method of phylogenomic inference. *Genome Biology* **9**: R151.
- Wyrsh I, Domínguez-Ferrerías A, Geldner N, Boller T. 2015.** Tissue-specific FLAGELLIN-SENSING 2 (FLS2) expression in roots restores immune responses in Arabidopsis fls2 mutants. *New Phytologist* **206**: 774–784.
- Xin X-F, Nomura K, Aung K, Velásquez AC, Yao J, Boutrot F, Chang JH, Zipfel C, He SY. 2016.** Bacteria establish an aqueous living space in plants crucial for virulence. *Nature* **539**: 524–529.
- Xiong C, Zhu YG, Wang JT, Singh B, Han LL, Shen JP, Li PP, Wang GB, Wu CF, Ge AH, et al. 2021.** Host selection shapes crop microbiome assembly and network complexity. *New Phytologist* **229**: 1091–1104.
- Xu L, Naylor D, Dong Z, Simmons T, Pierroz G, Hixson KK, Kim Y-M, Zink EM, Engbrecht KM, Wang Y, et al. 2018.** Drought delays development of the sorghum root microbiome and enriches for monoderm bacteria. *Proceedings of the National Academy of Sciences* **115**: E4284–E4293.
- Yang L, Ding W, Xu Y, Wu D, Li S, Chen J, Guo B. 2016.** New Insights into the Antibacterial Activity of Hydroxycoumarins against *Ralstonia solanacearum*. *Molecules* **21**: 468.
- Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, et al. 2019.** Ensembl 2020. *Nucleic Acids Research*.
- Yatsunenkov T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, et al. 2012.** Human gut microbiome viewed across age and geography. *Nature* **486**: 222–227.
- Yeoh YK, Dennis PG, Paungfoo-Lonhienne C, Weber L, Brackin R, Ragan MA, Schmidt S, Hugenholtz P. 2017.** Evolutionary conservation of a core root microbiome across plant phyla along a tropical soil chronosequence. *Nature Communications* **8**: 215.
- Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010.** Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11**: R14.
- 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**: 284–287.
- Yu K, Liu Y, Tichelaar R, Savant N, Legendijk E, van Kuijk SJL, Stringlis IA, van Dijken AJH, Pieterse CMJ, Bakker PAHM, et al. 2019.** Rhizosphere-Associated Pseudomonas Suppress Local Root Immune Responses by Gluconic Acid-Mediated Lowering of Environmental pH. *Current Biology* **29**: 3913-3920.e4.
- Yu P, Wang C, Baldauf JA, 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**: 1240–1253.
- Yuan MM, Guo X, Wu L, Zhang Y, Xiao N, Ning D, Shi Z, Zhou X, Wu L, Yang Y, et al. 2021.** Climate warming enhances microbial network complexity and stability. *Nature Climate Change* **11**: 343–348.
- Zamioudis C, Mastranesti P, Dhonukshe P, Blilou I, Pieterse CMJ. 2013.** Unraveling Root Developmental Programs Initiated by Beneficial Pseudomonas spp. Bacteria. *Plant Physiology* **162**: 304–318.
- Zamkovaya T, Foster JS, de Crécy-Lagard V, Conesa A. 2021.** A network approach to elucidate and prioritize microbial dark matter in microbial communities. *The ISME Journal* **15**: 228–244.

- Zaneveld J, Turnbaugh PJ, Lozupone C, Ley RE, Hamady M, Gordon JI, Knight R. 2008.** Host-bacterial coevolution and the search for new drug targets. *Current Opinion in Chemical Biology* **12**: 109–114.
- Zgad Zaj R, Garrido-Oter R, Jensen DB, Koprivova A, Schulze-Lefert P, Radutoiu S. 2016.** Root nodule symbiosis in *Lotus japonicus* drives the establishment of distinctive rhizosphere, root, and nodule bacterial communities. *Proceedings of the National Academy of Sciences* **113**: E7996–E8005.
- Zhalnina K, Louie KB, Hao Z, Mansoori N, da Rocha UN, Shi S, Cho H, Karaoz U, Loqué D, Bowen BP, et al. 2018.** Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nature Microbiology* **3**: 470–480.
- Zhang J, Liu Y-X, Guo X, Qin Y, Garrido-Oter R, Schulze-Lefert P, Bai Y. 2021.** High-throughput cultivation and identification of bacteria from the plant root microbiota. *Nature Protocols* **16**: 988–1012.
- Zhang J, Liu Y-X, Zhang N, Hu B, Jin T, Xu H, Qin Y, Yan P, Zhang X, Guo X, et al. 2019.** NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice. *Nature Biotechnology* **37**: 676–684.
- Zhang P, Spaepen S, Bai Y, Hacquard S, Garrido-Oter R. 2021.** Rbec: a tool for analysis of amplicon sequencing data from synthetic microbial communities. *ISME Communications* **1**: 73.
- Zhang Y, Xu S, Ding P, Wang D, Cheng YT, He J, Gao M, Xu F, Li Y, Zhu Z, et al. 2010.** Control of salicylic acid synthesis and systemic acquired resistance by two members of a plant-specific family of transcription factors. *Proceedings of the National Academy of Sciences* **107**: 18220–18225.
- Zhao L, Cai H, Su Z, Wang L, Huang X, Zhang M, Chen P, Dai X, Zhao H, Palanivelu R, et al. 2018.** KLU suppresses megasporocyte cell fate through SWR1-mediated activation of WRKY28 expression in *Arabidopsis*. *Proceedings of the National Academy of Sciences* **115**: E526–E535.
- Zhao Y. 2002.** Trp-dependent auxin biosynthesis in *Arabidopsis*: involvement of cytochrome P450s CYP79B2 and CYP79B3. *Genes & Development* **16**: 3100–3112.
- Zhou F, Emonet A, Dénervaud Tendon V, Marhavy P, Wu D, Lahaye T, Geldner N. 2020.** Co-occurrence of Damage and Microbial Patterns Controls Localized Immune Responses in Roots. *Cell* **180**: 440–453.e18.
- Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. 2019.** Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* **10**: 1523.
- Ziegler J, Schmidt S, Strehmel N, Scheel D, Abel S. 2017.** *Arabidopsis* Transporter ABCG37/PDR9 contributes primarily highly oxygenated Coumarins to Root Exudation. *Scientific Reports* **7**: 3704.
- Zipfel C, Robatzek S, Navarro L, Oakeley EJ, Jones JDG, Felix G, Boller T. 2004.** Bacterial disease resistance in *Arabidopsis* through flagellin perception. *Nature* **428**: 764–767.
- Zou L, Yang F, Ma Y, Wu Q, Yi K, Zhang D. 2019.** Transcription factor WRKY30 mediates resistance to Cucumber mosaic virus in *Arabidopsis*. *Biochemical and Biophysical Research Communications* **517**: 118–124.