

# Experimental Ecology of Plant-Microbiome Interactions

## Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät  
der Eberhard Karls Universität Tübingen  
zur Erlangung des Grades eines  
Doktors der Naturwissenschaften  
(Dr. rer. nat.)

vorgelegt von  
Jun Hee Jung  
aus Seoul, Republik Korea

Tübingen  
2025

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der  
Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:

18.06.2025

Dekan:

Prof. Dr. Thilo Stehle

1. Berichterstatter/-in:

Prof. Dr. Oliver Bossdorf

2. Berichterstatter/-in:

Prof. Dr. Johannes F. Scheepens

## Table of Contents

Declaration of author contributions	4
Summary	5
Zusammenfassung	7
Chapter 1	9
General introduction	
Chapter 2	21
Understanding plant microbiomes requires a genotype × environment framework	
Chapter 3	26
Leaf bacterial diversity promotes growth and pathogen resistance in <i>Arabidopsis thaliana</i>	
Chapter 4	61
Temperature fluctuations alter the <i>Arabidopsis thaliana</i> leaf microbiome	
Chapter 5	87
Discarded sequencing reads uncover natural variation in pest resistance in <i>Thlaspi arvense</i>	
Chapter 6	111
General discussion	
Acknowledgements	122

## Declaration of author contributions

This thesis, titled “Experimental ecology of plant-microbiome interactions” is based on the work and publications I produced during my PhD at the University of Tübingen, supervised by Prof. Dr. Oliver Bossdorf. Chapters 2-5 contain four independent scientific manuscripts that have already been published or will be published. My contributions to each chapter, as well as that of the other authors, are stated below:

**Chapter 2:** Published in *American Journal of Botany*, Volume 108, Issue 10 (Oct 2021) (<https://doi.org/10.1002/ajb2.1742>)

Jun Hee Jung, Frank Reis, Christina L. Richards, Oliver Bossdorf

All authors jointly developed the ideas in this essay and contributed to writing the manuscript. JHJ and FR are co-first authors.

**Chapter 3:** Unpublished manuscript

Jun Hee Jung, Johannes Kasper, Frank Reis, Nina Porath, Madalin Parepa, Julia Vorholt, Oliver Bossdorf

JK, MP, and OB conceived the experiment. JHJ, JK, FR, MP, and OB designed the experiment. JHJ, JK, FR, and NP ran the experiment with input from MP and OB. JHJ and JK analyzed the experiment with input from JV, MP, and OB. JHJ wrote the manuscript with input from all authors.

**Chapter 4:** Unpublished manuscript

Jun Hee Jung, Ana Montalvao, Natalie Betz, Madalin Parepa, Detlef Weigel, Oliver Bossdorf

JHJ and OB conceived the experiment. JHJ ran the experiment with input from OB. JHJ, AM, NB produced sequencing data with input from DW and OB. JHJ ran the analysis with input from AM, MP, and OB. JHJ and OB wrote the manuscript with input from MP.

**Chapter 5:** Published in *eLife*, Volume 13 (Oct 2024) (<https://doi.org/10.7554/eLife.95510.3>)

Dario Galanti, Jun Hee Jung, Caroline Müller, Oliver Bossdorf

JHJ, DG and OB designed the study. DG and CM collected the data. DG and JHJ analyzed the data. DG and JHJ wrote the manuscript with revisions from all coauthors.

## Summary

With accelerating global environmental changes, plants will more often face hostile environments with increased climate fluctuations, increased biotic pressures from pathogens and herbivores, and accelerated biodiversity decline. As the plant microbiome plays an essential role in plant robustness and environmental tolerance, we must study plant-microbiome interactions in the context of changing environmental conditions, to accurately address the challenges that plants will face in the future. This requires multi-factorial ecological experiments that combine the manipulation of plants, microbes, and their environments, and to test the interactions between these components.

In this thesis, I conducted a series of ecological plant-microbiome experiments. Chapter 2 explains how the genotype-environment interaction (G x E) framework known from experimental ecology can be usefully applied to microbiome research and outlines the two perspectives taken throughout this thesis. These two outlooks are determined by which dependent variable is focused on, i.e. whether plant performance is measured after the manipulation of microbiome and environmental factors, or whether the microbiome is studied after the manipulation of plant and environmental factors. In Chapter 3 I take a plant perspective approach, and in Chapters 4 and 5 a microbiome perspective.

In the third chapter, I used a bacterial isolate collection from *Arabidopsis thaliana* leaves to create leaf microbiomes of different diversity levels, and test for their effects on plant performance. In a full factorial design, I varied the number of phylogenetic groups (1 or 3), and the total number of strains in the bacterial community (3 or 9), and I tested how this influenced plant growth and *Pseudomonas viridiflava* pathogen resistance. Using the sequence information of bacterial isolates, I also calculated the phylogenetic and functional diversities of all the bacterial communities used in the experiment. I found that plants inoculated with more diverse microbiomes had higher growth and increased pathogen resistance, and that functional diversity was overall the best predictor of plant performance.

In the fourth chapter, I investigated how temperature fluctuations affected leaf microbiome diversity and composition. I exposed *Arabidopsis* plants to a series of heat stress and recovery phases, with different frequencies of fluctuations, but with the same total temperature sum for each plant at the end of the experiment. I found that bacterial diversity was significantly lowered after temperature fluctuations, especially in the group of plants exposed to rapid, one day fluctuations, as compared to those exposed to slower fluctuations in temperature. In addition, the composition of the bacterial communities differed significantly depending on the fluctuation frequency, with the group facing rapid fluctuations more significantly affected.

In the fifth chapter, I used reads from whole genome sequencing as metagenomics reads, owing to the methodological overlap. After a common garden experiment with *Thlaspi arvense*, a relative of *A. thaliana*, I treated sequencing reads that were not aligned to the genome as microbiome-associated reads. I was then able to extract reads associated with plant herbivores (aphids and aphid-associated bacteria) and pathogens (powdery mildew) present at the time of the experiment. I used this opportunity to conduct an indirect common garden experiment wherein the microbiome reads were used as a proxy for pathogen and herbivore load, and found genomic and epi-genomic mechanisms behind plant pathogen and herbivore resistance.

My thesis demonstrates the power of ecological experiments in studying plant-microbiome interactions within the context of environmental and biodiversity changes. I found that plant microbiome diversity is important to plant performance, and that temperature fluctuations can significantly decrease microbiome bacterial diversity and change community composition. These results demonstrate the importance of incorporating plant-associated microbiomes in genotype-environment interaction research. Future experiments should focus on testing the effects of realistic climate change scenarios on plants and their associated microbiomes, and how these effects may cascade between the intertwined systems of plants and bacterial communities.

## Zusammenfassung

Mit den immer schnelleren Veränderungen der globalen Umweltbedingungen werden Pflanzen immer häufiger mit herausfordernden Bedingungen konfrontiert sein. Diese können verstärkte Klimaschwankungen, erhöhter biotischer Druck durch Krankheitserreger und Pflanzenfresser sowie ein beschleunigter Rückgang der biologischen Vielfalt sein. Da das pflanzliche Mikrobiom eine wesentliche Rolle für die Widerstandsfähigkeit und Umwelttoleranz von Pflanzen spielt, muss man die Wechselwirkungen zwischen Pflanze und Mikrobiom im Kontext sich verändernder Umweltbedingungen untersuchen, um den Herausforderungen, die Pflanzen in Zukunft bewältigen müssen, sicher gegenüberzutreten zu können. Dies erfordert multifaktorielle ökologische Experimente, die die Manipulation von Pflanzen, Mikroben und ihrer Umgebung kombinieren und die Wechselwirkungen zwischen diesen Komponenten testen.

In dieser Dissertation habe ich eine Reihe von ökologischen Pflanzen-Mikrobiom-Experimenten durchgeführt. In Kapitel 2 erläuterte ich, wie das aus der experimentellen Ökologie bekannte Konzept der Genotyp-Umwelt-Interaktion (genotype-environment interaction,  $G \times E$ ) sinnvoll auf die Mikrobiom-Forschung angewandt werden kann. Dabei beschreibe ich die beiden Perspektiven, die ich in dieser Arbeit eingenommen habe. Diese beiden Sichtweisen hängen davon ab, auf welche abhängige Variable man sich konzentriert, d. h. ob die Pflanzenleistung nach der Manipulation des Mikrobioms und der Umweltfaktoren gemessen wird oder ob das Mikrobiom nach der Manipulation der Pflanze und Umweltfaktoren untersucht wird. In Kapitel 3 konzentriere ich mich auf die Pflanzenperspektive, in den Kapiteln 4 und 5 auf die Perspektive des Mikrobioms.

Im dritten Kapitel verwendete ich eine aus den Blättern der Ackerschmalwand (*Arabidopsis thaliana*) isolierte Bakteriengemeinschaft, um Blattmikrobiome mit unterschiedlichen Diversitätsgraden zu erstellen und deren Auswirkungen auf die Pflanzenleistung zu testen. In einem voll faktoriell designten Experiment variierte ich die Anzahl der phylogenetischen Gruppen (1 oder 3) und die Gesamtzahl der Stämme in der bakteriellen Gemeinschaft (3 oder 9) und testete, wie dies das Pflanzenwachstum und die Resistenz gegen das Pathogen *Pseudomonas viridiflava* beeinflusste. Anhand der Sequenzinformationen der Bakterienisolate berechnete ich auch die phylogenetische und funktionelle Diversität aller im Versuch verwendeten Bakteriengemeinschaften. Ich fand heraus, dass Pflanzen, die mit einem diverseren Mikrobiom beimpft wurden, ein höheres Wachstum und eine höhere Resistenz gegen Krankheitserreger aufwiesen und dass die funktionelle Vielfalt insgesamt die beste Prognose für die Pflanzenleistung war.

Im vierten Kapitel untersuchte ich, wie sich Temperaturschwankungen auf die Vielfalt und Zusammensetzung des Blattmikrobioms auswirken. Ich setzte *Arabidopsis* Pflanzen einer Reihe von Hitzestress- und Erholungsphasen aus, mit unterschiedlicher Häufigkeit der Schwankungen, aber mit der gleichen Gesamttemperatursumme für jede Pflanze am Ende des Experiments. Ich stellte fest, dass die bakterielle Diversität nach den Temperaturschwankungen signifikant abnahm. Dies zeigte sich besonders deutlich in den Pflanzen die schnellen, eintägigen Schwankungen ausgesetzt waren im Vergleich zu Pflanzen mit langsameren Temperaturschwankungen. Darüber hinaus unterschied sich die Zusammensetzung der Bakteriengemeinschaften je nach Häufigkeit der Schwankungen erheblich, wobei die Gruppe, die schnellen Schwankungen ausgesetzt war, stärker betroffen war.

Im fünften Kapitel nutze ich die methodische Überschneidung in einem Experiment, das vollständige Genomsequenzierung nutzt, um aus diesem Datensatz mikrobielle Sequenzen herauszufiltern und so auch die Mikrobengemeinschaften zu untersuchen. Nach einem Gartenexperiment mit *Thlaspi arvense*, einem Verwandten von *A. thaliana*, behandelte ich Sequenzierungs-Reads, die nicht dem Pflanzengenom zugeordnet werden konnten, als Mikrobiom-assoziierte Reads. So konnte ich Reads extrahieren, die mit Pflanzenfressern (Blattläusen und ihnen assoziierten Bakterien) und Krankheitserregern (Mehltau) assoziiert waren, die zum Zeitpunkt des Experiments vorhanden waren. Ich nutzte diese Gelegenheit, um ein indirektes Gartenexperiment durchzuführen, bei dem die Mikrobiom-Reads als Stellvertreter für die Pathogen- und Fressfeindbelastung verwendet wurden, und fand genomische und epigenomische Mechanismen hinter der Resistenz von Pflanzen gegen Pathogene und Fressfeinde.

Meine Dissertation zeigt die Leistungsfähigkeit ökologischer Experimente bei der Untersuchung der Wechselwirkungen zwischen Pflanzen und Mikrobiom im Kontext von Umwelt- und Biodiversitätsveränderungen. Ich habe herausgefunden, dass die Vielfalt des Pflanzenmikrobioms wichtig für die Leistungsfähigkeit der Pflanzen ist und dass Temperaturschwankungen die bakterielle Vielfalt des Mikrobioms signifikant verringern und die Zusammensetzung der Gemeinschaft verändern können. Diese Ergebnisse zeigen, wie wichtig es ist, pflanzenassoziierte Mikrobiome in die Forschung zur Interaktion zwischen Genotyp und Umwelt einzubeziehen. Künftige Experimente sollten sich darauf konzentrieren, die Auswirkungen realistischer Szenarien des Klimawandels auf Pflanzen und die mit ihnen verbundenen Mikrobiome zu testen und zu untersuchen, wie diese Auswirkungen zwischen den miteinander verflochtenen Systemen von Pflanzen und bakteriellen Gemeinschaften kaskadenartig wirken können.

# Chapter 1

## General introduction

### *Experimental ecology with the microbiome*

Throughout their 400 million year history (Kenrick and Crane, 1997) plants have always been exposed to environmental stresses, especially due to their sessile nature. Although many definitions of stress exist, it is loosely defined as an unfavorable condition that decreases plant performance (Larcher, 1987; Lichtenthaler, 1996; Kranner et al., 2010). Larcher (1987)'s definition of plant stress includes when a plant's "adaptive capacity is overtaxed", suggesting an inbuilt tolerance toward stress. This inherent stress tolerance is related to the concept of phenotypic plasticity (Bradshaw, 1965), the ability of two genetically identical plants placed in different environmental conditions to exhibit different phenotypes. Another way to formalize phenotypic plasticity is within the framework of genotype by environment interactions (hereafter G x E), which describe phenotype as product of genotype and the environment (Gutteling et al., 2007; Christensen et al., 2021). At the population level, G x E also refers to a statistical interaction (non-parallel reaction norm), but the terms phenotypic plasticity and G x E have been used interchangeably by many researchers (Pigliucci, 2005). Phenotypic plasticity is an increasingly important topic, as global environmental changes will expose plants to escalating biotic and abiotic stresses (Bonamour et al., 2019). For instance, global climate change is expected to increase both temperature and temperature variability, and both phenotypic plasticity and evolutionary changes may allow plants to adjust to this rapidly changing world (Scheepens et al., 2018; Arnold et al., 2024). Environmental stresses usually do not occur alone, and plants often face multiple environmental changes simultaneously. Combined stresses produce different plastic responses (Lorts and Lasky, 2020), and abiotic and biotic stresses should be considered in tandem.

One of the most important biotic interactions for a plant is with their microbiome. A microbiome is broadly defined as a group of microorganisms assembled into a community (Berg et al., 2020). Plants harbor a highly diverse and abundant microbial community, and the leaf alone contains  $10^6$  to  $10^7$  microbes per  $\text{cm}^2$  (Lindow and Brandl, 2003). Understanding plant-microbiome interactions is crucial, and many studies have experimentally shown significant influence of microbiomes on plant performance, pathogen resistance, mineral uptake, competition, and stress tolerance (Fitzpatrick et al., 2019; Harbort et al., 2020; Vogel et al., 2021). Much of this evidence comes from research on soil, and the microbiome from soil facing drought can benefit plant performance even when there is no history of association between the plant and that specific microbiome (Ricks and Yannarell, 2023). The association between plants and specific microbiomes is not only the passive uptake of local microorganisms, but

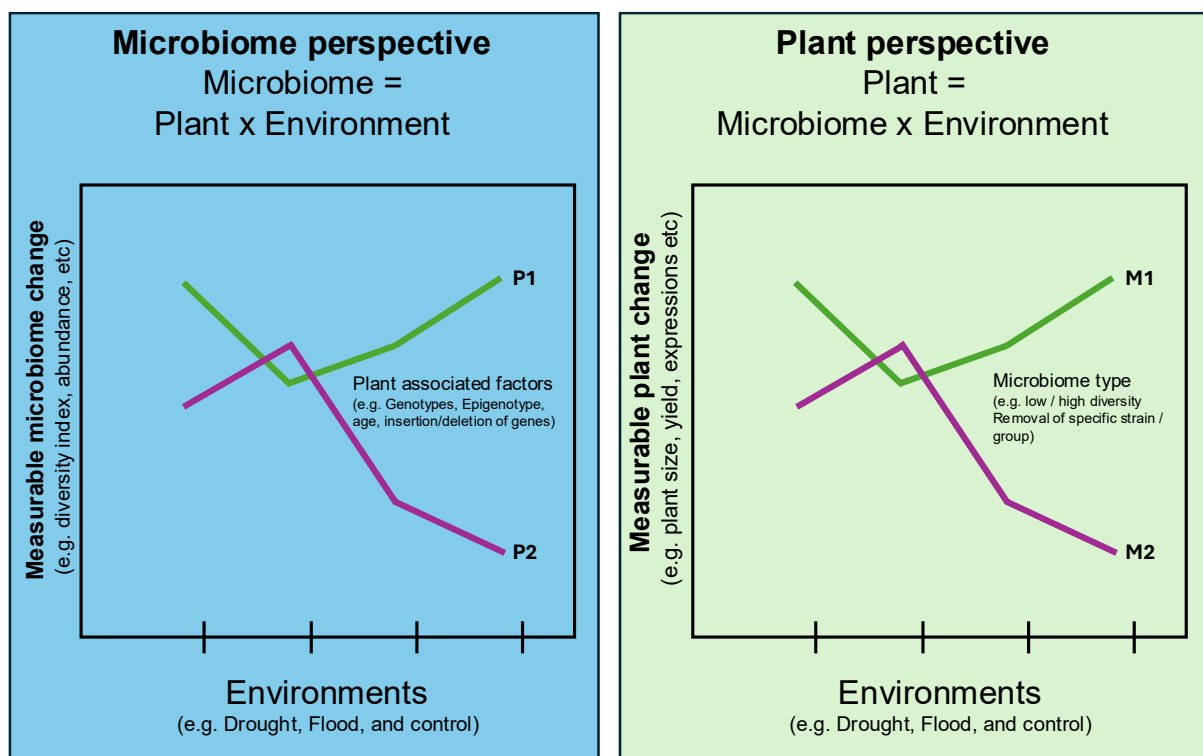
can also be an active process, and the “cry for help” hypothesis posits that plants selectively enrich their protective microbiomes (Bakker et al., 2018). Kwak et al. (2018) have found evidence of this deliberate uptake in tomatoes, as plant genotypes differed in pathogen response, with one genotype better able to recruit a beneficial microbiome than the other.

Both plants and microorganisms usually exist in communities, and it is well known that communities with greater species richness can lead to higher levels of ecosystem functioning, a relationship known as biodiversity-ecosystem functioning (BEF; Luo et al., 2022; Zheng et al., 2024). Within the context of the phytobiome, the BEF relationship means that an increase in species richness in either plants or the microbiome could lead to better ecosystem functioning in the other. Cappelli et al. (2022) have reviewed evidence for this bidirectional relationship in terms of plants and the soil microbiome, with underground microbial diversity increasing plant community function, and vice versa. There is also increasing direct and indirect evidence that leaf microbiome (phyllosphere) diversity is associated with increased plant performance (Laforest-Lapointe et al., 2017; Vogel et al., 2021; Emmenegger et al., 2023).

Despite their close and generally beneficial relationship, plants can create challenging habitats for the microbiome. The host plant exerts fine control over the bacterial community, often with a core microbiome for a particular plant species (Risely, 2020), but different plant genotypes select for specific additional microbes (Brown et al., 2020). There is further selection within plants, with plant organs also selecting for specific microbiomes; usually the leaf surface microbiome (epiphyte) differs from the inner leaf microbiome (leaf endophyte), and the leaf microbiome differs from the root microbiome (Bulgarelli et al., 2013). Once a specific microbiome is selected to live in a host, genotype and organ, it is exposed to various environmental fluctuations, with the leaf microbiome (phyllosphere) particularly exposed to strong changes in UV, temperature, and desiccation stress. Microbiomes also face biotic stresses such as plant pathogens, niche overlap between strains, and space competition (Vorholt, 2012; Koskella, 2020).

Given that the plant and microbiome mutually influence each other, and that both are constantly exposed to stress, it is essential to study plants and microbiomes as interacting systems. They both react to environmental changes, and plant responses can be accelerated or buffered by microbiome responses, and vice versa. Studying the plant microbiome system can provide a better understanding of the role of the microbiome when plants face stress, and how the microbiome could provide protection to the plant and potentially increase plant resilience in changing climates (Wipf et al., 2021; Li et al., 2023; Hoefle et al., 2024).

To answer these questions about the interactive effects between plants, the environment, and the microbiome, a G x E framework can be applied. The concept of G x E has already been used beyond genetics and phenotypic plasticity, for instance in the study of epigenetics (Bossdorf et al., 2008; Richards et al., 2010), and it can also be applied to plant microbiome research (Figure 1). When using the G X E framework in plant microbiome research, two perspectives can be adopted. First, plant microbiomes can be manipulated, and plants can be exposed to environmental stressors. The effects of varied microbiomes can then be seen in plant performance, such as growth or other metrics. We called this a ‘plant perspective’ design (Figure 1, right panel). Alternatively, a ‘microbiome perspective’ is achieved when plant factors such as genotype or age, and environmental factors like heat and water stress are manipulated, and changes to the microbiome are measured (Figure 1, left panel). An excellent microbiome perspective design was by Wagner et al. (2016), who demonstrated the interactive effects of plant genotype, age, experimental site, and year on plant microbiomes. An exemplary plant-perspective study can be found by Fitzpatrick et al. (2019), in which plants with different microbiomes were exposed to different environmental stressors.



**Figure 1.** Two different G x E perspectives in plant microbiome research in which changes in microbiome can result from interactions between plant and environment factors, or changes in plant performance result from interactions between microbiome and environment (adapted from Jung et al., 2021).

## Tools for experimental ecological microbiome research

### *Arabidopsis as a model system*

*Arabidopsis thaliana* (L.) Heynh is a diploid model plant with a small genome of around 135 MB and around 33,000 genes (The Arabidopsis Genome Initiative, 2000; Berardini et al., 2015). *A. thaliana* is a pioneer species of disturbed poor soils, spanning a vast array of ecosystems across Eurasia and Africa (Krämer, 2015). The species has long served as a useful tool in plant biology due to its short life cycle, small genome size, and ability to self-fertilize. *Arabidopsis* has been instrumental in improving our understanding of many fields of plant biology, including plant intraspecific competition (Palacio-Lopez et al., 2020), interspecific competition (Müller and Bartelheimer, 2013), and interactions with other organisms such as insects, bacteria, fungi, and viruses (e.g. Louis et al., 2012). *Arabidopsis* is also a close relative to many important crop species (Kim et al., 2018), and knowledge gained from *Arabidopsis* research can be applied to improve other, more economically-relevant plants.

A frequently studied biotic interaction within the *Arabidopsis* study system is between the plant and the bacterial pathogen *Pseudomonas syringae* DC3000. This model interaction has been used to uncover different molecular mechanisms underlying host and hemi-biotrophic pathogen interactions, pathogen effectors such as type III secretion systems, and plant hormones involved in pathogen resistance such as jasmonic acid and salicylic acid (Katagiri et al., 2002; Flors et al., 2008; Xin and He, 2013). Environmental context matters in this well-studied system, as under heat stress the virulence of *Pseudomonas* increases (Huot et al., 2017), while under drought stress virulence is lowered (Gupta et al., 2016). The microbiome is also crucial, as *Arabidopsis* grown in natural soil can withstand higher concentrations of *Pseudomonas* than those grown in sterile agar (Ishiga et al., 2011).

### *Sequencing technologies in microbiome research*

Over recent years, the study of plant microbiomes has benefitted strongly from the rapid increase in sequencing capacity and the decrease in sequencing costs, which have dropped from 30 base pairs per dollar to 3 million base pairs per dollar over the last 20 years (Wetterstrand, 2025). By using “universal primers” originally developed for comparative phylogenetics, researchers are now able to detect the presence of several kingdoms in a single PCR process by targeting conserved genes with a high degree of interspecies variation, a process known as metabarcoding. (Matchado et al., 2024). The targeted regions depend on the phylogenetic groups of interest, and a common region used in bacterial microbiome studies is the 16S SSU rRNA region (Caporaso et al., 2011). Different 16S rRNA regions are

targeted according to project requirements, for example the Human Microbiome Project used V1-3 and V3-5 regions (Peterson et al., 2009), and the Earth Microbiome Project used the V4 region (Thompson et al., 2017).

Despite their widespread use, there are limitations to conventional 16S rRNA-based primers. The main criticism of the metabarcoding approach is that it lacks resolution because of the short reads (ca. 250 bp forward and reverse reads), especially in lower taxonomic ranks (Johnson et al., 2019). Databases can also lead to biases due to incorrect or incomplete datasets, and certain groups may not be amplified consistently with the primer pairing used (Johnson et al., 2019; Loos and Nijland, 2020; Abellan-Schneyder et al., 2021). Since the concept of universal microbe primers was introduced, there has been constant improvement and updates in methodologies to try and combat these and other issues. For example, universal primers inevitably capture organelle sequences that are irrelevant to the study of the microbiome. Instead of discarding these host organelle-associated reads, researchers can now suppress them before sequencing (Lundberg et al., 2013; Mayer et al., 2021).

Metabarcoding can be supplemented with a shallow metagenomics approach: metabarcoding is a rich source of information about microbiome diversity and composition, while metagenomics can be used to extract host genotype, and the ratio of host cells to microbiome cells (Regalado et al., 2020). Used together, this method can be a cost-effective way to study plant-microbiome interactions.

Finally, data from whole genome shotgun sequencing or a restriction enzyme approach (e.g. ddRAD or GBS) can be a valuable source of information for microbiome research (Liu et al., 2017; Roman-Reyna et al., 2020). Whole genome shotgun sequencing is methodologically similar to metabarcoding, and up to 10% of the total reads are usually removed when not aligned to the reference genome (Roman-Reyna et al., 2020; Galanti et al., 2024). These discarded reads could contain microbiome information, albeit with the risk that they may also include low-quality reads of the host. Nonetheless, various recent sequencing projects have been a source of unexpected microbiome data, from laboratory contamination to plants, animals, and even to ancient DNA and herbarium samples (Sangiovanni et al., 2019; Bieker et al., 2020; Roman-Reyna et al., 2020; Yates et al., 2021; Neumann et al., 2023). These mostly untapped resources have great potential to improve our knowledge of plant microbiomes, with the benefits of lower costs and a reduced carbon footprint, especially useful for labs that cannot afford large sequencing projects (Winter et al., 2023; Paepe et al., 2024; Welburn, 2024).

### *Microbiome inoculation methods*

As well as monitoring the change in a microbiome after manipulating environmental or host factors, it is also possible to inoculate plants or soil with different communities to see the effect of microbiome diversity on the host. By applying different filter sizes, with smaller sizes excluding more microbes, filtrates of varied diversity levels can be created. Previous research has shown that inoculums created using this method can alter plant growth (Parepa et al., 2013; Wagg et al., 2019), with plants exposed to lowered soil microbial diversity suffering lower multifunctionality due to fewer taxa being present. Although the filter-reinoculate method provides a simple diversity gradient, it is difficult to select for specific phylogenetic groups, and a culture collection-based method may be better suited for this purpose.

A recent alternative to the filtration method involves the division of a complex community into individual bacterial strains, and subsequent re-construction of the strains into a new microbiome. This deliberately assembled microbiome is known as a synthetic community, hereafter abbreviated as syncom. The syncom system has garnered much interest for its ability to create a specific and tailored microbiome that can be used experimentally, and researchers have worked to create a standardized methodological approach to studying plant-microbiome interactions using syncom (Vorholt et al., 2017; Jing et al., 2024; Northen et al., 2024). Due to the labor-intensive nature of creating bacterial isolate collections and identifying them, only a few culture collections are so far available, for the model species *Arabidopsis thaliana*, *Lotus japonicus*, and *Hordeum vulgare* (Bai et al., 2015; Wippel et al., 2021; Robertson-Albertyn et al., 2022; Gómez-Pérez et al., 2023). The established cultures may reflect isolates from different parts of the microbial ecosystem such as leaf-associated (phyllosphere) and root-associated (rhizosphere) microbiomes (Bai et al., 2015).

Many of these syncom collections also have whole genome sequences and functional assays available for researchers (Bai et al., 2015; Schäfer et al., 2023; Northen et al., 2024). This large amount of information provides a unique method for studying various forms of microbiome interactions. For instance, syncom collections have been used to test the effects of strain interactions; Schäfer et al. (2023) have found that when two phylogenetically-close bacterial strains are present in a leaf, due to their large metabolic niche overlap, the interaction was negative, i.e. lowered abundance of at least one of the strains.

The syncom approach can also be used to test the role of plant genetics in dysbiosis, the perturbation of healthy plant microbiomes (Petersen and Round, 2014). As syncom experiments have a gnotobiotic approach in which microbiome composition is known (Basic and Bleich, 2019), it allows for the detection of any shifts in the community. For example, when

an *Arabidopsis* gene such as RHOHD is deleted, strains within a syncom can turn into opportunistic pathogens (Pfeilmeier et al., 2021, 2024).

### **Objectives of the thesis**

In my thesis, I test the interactions between plants and microbiomes under abiotic and biotic stressors. In Chapter 2, I outline a theoretical framework for answering ecological questions in plants and microbiomes. By approaching experiments from two perspectives, the plant and microbiome, the interaction between microorganisms and plants under different environmental conditions can be better understood.

In Chapter 3, an experiment based on a plant perspective approach is carried out. I construct 40 microbiomes by manipulating their diversity, according to both strain richness and clade diversity. The experiment is a proof-of-principle study to establish the relationship between leaf microbial diversity and plant performance, as measured by growth and *Pseudomonas* pathogen tolerance. By using information available from the At-LSPHERE leaf bacterial culture collection, I calculate predicted functional and phylogenetic diversity and then compare the statistical performance of the two indices.

In Chapter 4, I conduct a microbiome perspective experiment in which plant genotypes and the frequency of temperature fluctuations are altered. Plants begin with similar microbiome composition, as each is inoculated with a 138-member microbial community, and are exposed to varied heat stress periods. Plant size, based on image analysis, and microbiome diversity and composition, based on DNA metabarcoding, are analyzed to test the effects of temperature fluctuation throughout the experiment.

Finally, in Chapter 5 I exploit the technical overlap between whole genome sequencing (WGS) and metagenomics. By analyzing discarded WGS sequences from a common garden experiment, I am able to extract non-host reads belonging to bacteria, fungi, and eukaryotes. I then use herbivore and pathogen-associated reads for genomics and epi-genomic analysis.

## References

- Abellan-Schneyder, I., M. S. Matchado, S. Reitmeier, A. Sommer, Z. Sewald, J. Baumbach, M. List, and K. Neuhaus. 2021. Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing. *mSphere* 6: e01202.
- Arnold, P. A., S. Wang, R. F. Notarnicola, A. B. Nicotra, and L. E. B. Kruuk. 2024. Testing the evolutionary potential of an alpine plant: phenotypic plasticity in response to growth temperature outweighs parental environmental effects and other genetic causes of variation. *Journal of Experimental Botany* 75: 5971–5988.
- Bai, Y., D. B. Müller, G. Srinivas, R. Garrido-Oter, E. Potthoff, M. Rott, N. Dombrowski, et al. 2015. Functional overlap of the Arabidopsis leaf and root microbiota. *Nature* 528: 364–369.
- Bakker, P. A. H. M., C. M. J. Pieterse, R. de Jonge, and R. L. Berendsen. 2018. The Soil-Borne Legacy. *Cell* 172: 1178–1180.
- Basic, M., and A. Bleich. 2019. Gnotobiotics: Past, present and future. *Laboratory Animals* 53: 232–243.
- Berardini, T. Z., L. Reiser, D. Li, Y. Mezheritsky, R. Muller, E. Strait, and E. Huala. 2015. The arabidopsis information resource: Making and mining the “gold standard” annotated reference plant genome. *genesis* 53: 474–485.
- Berg, G., D. Rybakova, D. Fischer, T. Cernava, M.-C. C. Vergès, T. Charles, X. Chen, et al. 2020. Microbiome definition re-visited: old concepts and new challenges. *Microbiome* 8: 103.
- Bieker, V. C., F. Sánchez Barreiro, J. A. Rasmussen, M. Brunier, N. Wales, and M. D. Martin. 2020. Metagenomic analysis of historical herbarium specimens reveals a postmortem microbial community. *Molecular Ecology Resources* 20: 1206–1219.
- Bonamour, S., L.-M. Chevin, A. Charmantier, and C. Teplitsky. 2019. Phenotypic plasticity in response to climate change: the importance of cue variation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374: 20180178.
- Bossdorf, O., C. L. Richards, and M. Pigliucci. 2008. Epigenetics for ecologists. *Ecology Letters* 11: 106–115.
- Bradshaw, A. D. 1965. Evolutionary Significance of Phenotypic Plasticity in Plants. In E. W. Caspari, and J. M. Thoday [eds.], *Advances in Genetics*, 115–155. Academic Press.
- Brown, S. P., M. A. Grillo, J. C. Podowski, and K. D. Heath. 2020. Soil origin and plant genotype structure distinct microbiome compartments in the model legume *Medicago truncatula*. *Microbiome* 8: 139.
- Bulgarelli, D., K. Schlaeppli, S. Spaepen, E. V. L. van Themaat, and P. Schulze-Lefert. 2013. Structure and Functions of the Bacterial Microbiota of Plants. *Annual Review of Plant Biology* 64: 807–838.
- Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer, and R. Knight. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences* 108: 4516–4522.
- Cappelli, S. L., L. A. Domeignoz-Horta, V. Loaiza, and A.-L. Laine. 2022. Plant biodiversity promotes sustainable agriculture directly and via belowground effects. *Trends in Plant Science* 27: 674–687.
- Christensen, K. A., J. Le Luyer, M. T. T. Chan, E. B. Rondeau, B. F. Koop, L. Bernatchez, and R. H. Devlin. 2021. Assessing the effects of genotype-by-environment interaction on epigenetic, transcriptomic, and phenotypic response in a Pacific salmon. *G3: Genes|Genomes|Genetics* 11: jkab021.
- Emmenegger, B., J. Massoni, C. M. Pestalozzi, M. Bortfeld-Miller, B. A. Maier, and J. A. Vorholt. 2023. Identifying microbiota community patterns important for plant protection using synthetic communities and machine learning. *Nature Communications* 14: 7983.
- Fitzpatrick, C. R., Z. Mustafa, and J. Viliunas. 2019. Soil microbes alter plant fitness under

- competition and drought. *Journal of Evolutionary Biology* 32: 438–450.
- Flors, V., J. Ton, R. van Doorn, G. Jakab, P. García-Agustín, and B. Mauch-Mani. 2008. Interplay between JA, SA and ABA signalling during basal and induced resistance against *Pseudomonas syringae* and *Alternaria brassicicola*. *The Plant Journal: For Cell and Molecular Biology* 54: 81–92.
- Galanti, D., J. H. Jung, C. Müller, and O. Bossdorf. 2024. Discarded sequencing reads uncover natural variation in pest resistance in *Thlaspi arvense*. *eLife* 13.
- Gómez-Pérez, D., M. Schmid, V. Chaudhry, Y. Hu, A. Velic, B. Maček, J. Ruhe, et al. 2023. Proteins released into the plant apoplast by the obligate parasitic protist *Albugo* selectively repress phyllosphere-associated bacteria. *New Phytologist* 239: 2320–2334.
- Gupta, A., S. K. Dixit, and M. Senthil-Kumar. 2016. Drought Stress Predominantly Endures *Arabidopsis thaliana* to *Pseudomonas syringae* Infection. *Frontiers in Plant Science* 7.
- Gutteling, E. W., J. a. G. Riksen, J. Bakker, and J. E. Kammenga. 2007. Mapping phenotypic plasticity and genotype–environment interactions affecting life-history traits in *Caenorhabditis elegans*. *Heredity* 98: 28–37.
- Harbort, C. J., M. Hashimoto, H. Inoue, Y. Niu, R. Guan, A. D. Rombolà, S. Kopriva, et al. 2020. Root-Secreted Coumarins and the Microbiota Interact to Improve Iron Nutrition in *Arabidopsis*. *Cell Host & Microbe* 28: 825-837.e6.
- Hoefle, D., M. Sommer, B. Wassermann, M. Faticov, D. Serra, G. Berg, A. J. M. Tack, and A. Abdelfattah. 2024. Oak seedling microbiome assembly under climate warming and drought. *Environmental Microbiome* 19: 62.
- Huot, B., C. D. M. Castroverde, A. C. Velásquez, E. Hubbard, J. A. Pulman, J. Yao, K. L. Childs, et al. 2017. Dual impact of elevated temperature on plant defence and bacterial virulence in *Arabidopsis*. *Nature Communications* 8: 1808.
- Ishiga, Y., T. Ishiga, S. R. Uppalapati, and K. S. Mysore. 2011. *Arabidopsis* seedling flood-inoculation technique: a rapid and reliable assay for studying plant-bacterial interactions. *Plant Methods* 7: 32.
- Jing, J., P. Garbeva, J. M. Raaijmakers, and M. H. Medema. 2024. Strategies for tailoring functional microbial synthetic communities. *The ISME Journal* 18: wræ049.
- Johnson, J. S., D. J. Spakowicz, B.-Y. Hong, L. M. Petersen, P. Demkowicz, L. Chen, S. R. Leopold, et al. 2019. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications* 10: 5029.
- Jung, J. H., F. Reis, C. L. Richards, and O. Bossdorf. 2021. Understanding plant microbiomes requires a genotype × environment framework. *American Journal of Botany* 108: 1820–1823.
- Katagiri, F., R. Thilmony, and S. Y. He. 2002. The *Arabidopsis Thaliana*-*Pseudomonas Syringae* Interaction. *The Arabidopsis Book / American Society of Plant Biologists* 1: e0039.
- Kenrick, P., and P. R. Crane. 1997. The origin and early evolution of plants on land. *Nature* 389: 33–39.
- Kim, C.-K., Y.-J. Seol, S. Perumal, J. Lee, N. E. Waminal, M. Jayakodi, S.-C. Lee, et al. 2018. Re-exploration of U's Triangle Brassica Species Based on Chloroplast Genomes and 45S nrDNA Sequences. *Scientific Reports* 8: 7353.
- Koskella, B. 2020. The phyllosphere. *Current Biology* 30: R1143–R1146.
- Krämer, U. 2015. Planting molecular functions in an ecological context with *Arabidopsis thaliana*. *eLife* 4: e06100.
- Kranner, I., F. V. Minibayeva, R. P. Beckett, and C. E. Seal. 2010. What is stress? Concepts, definitions and applications in seed science. *New Phytologist* 188: 655–673.
- Kwak, M.-J., H. G. Kong, K. Choi, S.-K. Kwon, J. Y. Song, J. Lee, P. A. Lee, et al. 2018. Rhizosphere microbiome structure alters to enable wilt resistance in tomato. *Nature Biotechnology* 36: 1100–1109.
- Laforest-Lapointe, I., A. Paquette, C. Messier, and S. W. Kembel. 2017. Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. *Nature* 546:

- 145–147.
- Larcher, W. 1987. Streß bei Pflanzen. *Naturwissenschaften* 74: 158–167.
- Li, J., K. A. Bates, K. L. Hoang, T. E. Hector, S. C. L. Knowles, and K. C. King. 2023. Experimental temperatures shape host microbiome diversity and composition. *Global Change Biology* 29: 41–56.
- Lichtenthaler, H. K. 1996. Vegetation Stress: an Introduction to the Stress Concept in Plants. *Journal of Plant Physiology* 148: 4–14.
- Lindow, S. E., and M. T. Brandl. 2003. Microbiology of the Phyllosphere. *Applied and Environmental Microbiology* 69: 1875–1883.
- Liu, M. Y., P. Worden, L. G. Monahan, M. Z. DeMaere, C. M. Burke, S. P. Djordjevic, I. G. Charles, and A. E. Darling. 2017. Evaluation of ddRADseq for reduced representation metagenome sequencing. *PeerJ* 5: e3837.
- Loos, L. M. van der, and R. Nijland. 2020. Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. *Molecular Ecology* 30: 3270.
- Lorts, C. M., and J. R. Lasky. 2020. Competition × drought interactions change phenotypic plasticity and the direction of selection on *Arabidopsis* traits. *New Phytologist* 227: 1060–1072.
- Louis, J., V. Singh, and J. Shah. 2012. *Arabidopsis thaliana*—Aphid Interaction. *The Arabidopsis Book / American Society of Plant Biologists* 10: e0159.
- Lundberg, D. S., S. Yourstone, P. Mieczkowski, C. D. Jones, and J. L. Dangl. 2013. Practical innovations for high-throughput amplicon sequencing. *Nature Methods* 10: 999–1002.
- Luo, S., R. D. Bardgett, B. Schmid, D. Johnson, K. Png, U. Schaffner, H. Zhou, et al. 2022. Historical context modifies plant diversity–community productivity relationships in alpine grassland. *Journal of Ecology* 110: 2205–2218.
- Matchado, M. S., M. Rühlemann, S. Reitmeier, T. Kacprowski, F. Frost, D. Haller, J. Baumbach, and M. List. 2024. On the limits of 16S rRNA gene-based metagenome prediction and functional profiling. *Microbial Genomics* 10: 001203.
- Mayer, T., A. Mari, J. Almario, M. Murillo-Roos, H. Syed M. Abdullah, N. Dombrowski, S. Hacquard, et al. 2021. Obtaining deeper insights into microbiome diversity using a simple method to block host and nontargets in amplicon sequencing. *Molecular Ecology Resources* 21: 1952–1965.
- Müller, B., and M. Bartelheimer. 2013. Interspecific competition in *Arabidopsis thaliana*: root hairs are important for competitive effect, but not for competitive response. *Plant and Soil* 371: 167–177.
- Neumann, G. B., P. Korcuć, M. Reißmann, M. J. Wolf, K. May, S. König, and G. A. Brockmann. 2023. Unmapped short reads from whole-genome sequencing indicate potential infectious pathogens in German Black Pied cattle. *Veterinary Research* 54: 95.
- Northen, T. R., M. Kleiner, M. Torres, Á. T. Kovács, M. H. Nicolaisen, D. M. Krzyżanowska, S. Sharma, et al. 2024. Community standards and future opportunities for synthetic communities in plant–microbiota research. *Nature Microbiology* 9: 2774–2784.
- Paepe, M. D., L. Jeanneau, J. Mariette, O. Aumont, and A. Estevez-Torres. 2024. Purchases dominate the carbon footprint of research laboratories. *PLOS Sustainability and Transformation* 3: e0000116.
- Palacio-Lopez, K., C. M. King, J. Bloomberg, and S. M. Hovick. 2020. Natural selection on traits and trait plasticity in *Arabidopsis thaliana* varies across competitive environments. *Scientific Reports* 10: 21632.
- Parepa, M., U. Schaffner, and O. Bossdorf. 2013. Help from under ground: soil biota facilitate knotweed invasion. *Ecosphere* 4: art31.
- Petersen, C., and J. L. Round. 2014. Defining dysbiosis and its influence on host immunity and disease. *Cellular Microbiology* 16: 1024–1033.
- Peterson, J., S. Garges, M. Giovanni, P. McInnes, L. Wang, J. A. Schloss, V. Bonazzi, et al. 2009. The NIH Human Microbiome Project. *Genome Research* 19: 2317–2323.
- Pfeilmeier, S., G. C. Petti, M. Bortfeld-Miller, B. Daniel, C. M. Field, S. Sunagawa, and J. A.

- Vorholt. 2021. The plant NADPH oxidase RBOHD is required for microbiota homeostasis in leaves. *Nature Microbiology* 6: 852–864.
- Pfeilmeier, S., A. Werz, M. Ote, M. Bortfeld-Miller, P. Kirner, A. Keppler, L. Hemmerle, et al. 2024. Leaf microbiome dysbiosis triggered by T2SS-dependent enzyme secretion from opportunistic *Xanthomonas* pathogens. *Nature Microbiology* 9: 136–149.
- Pigliucci, M. 2005. Evolution of phenotypic plasticity: where are we going now? *Trends in Ecology & Evolution* 20: 481–486.
- Regalado, J., D. S. Lundberg, O. Deusch, S. Kersten, T. Karasov, K. Poersch, G. Shirsekar, and D. Weigel. 2020. Combining whole-genome shotgun sequencing and rRNA gene amplicon analyses to improve detection of microbe–microbe interaction networks in plant leaves. *The ISME Journal* 14: 2116–2130.
- Richards, C. L., O. Bosssdorf, and K. J. F. Verhoeven. 2010. Understanding natural epigenetic variation. *The New Phytologist* 187: 562–564.
- Ricks, K. D., and A. C. Yannarell. 2023. Soil moisture incidentally selects for microbes that facilitate locally adaptive plant response. *Proceedings of the Royal Society B: Biological Sciences* 290: 20230469.
- Risely, A. 2020. Applying the core microbiome to understand host–microbe systems. *Journal of Animal Ecology* 89: 1549–1558.
- Robertson-Albertyn, S., J. C. Abbott, F. Concas, L. H. Brown, J. N. Orr, T. S. George, and D. Bulgarelli. 2022. Genome-Annotated Bacterial Collection of the Barley Rhizosphere Microbiota. *Microbiology Resource Announcements* 11: e01064-21.
- Roman-Reyna, V., D. Pinili, F. N. Borja, I. L. Quibod, S. C. Groen, N. Alexandrov, R. Mauleon, and R. Oliva. 2020. Characterization of the Leaf Microbiome from Whole-Genome Sequencing Data of the 3000 Rice Genomes Project. *Rice* 13: 72.
- Sangiovanni, M., I. Granata, A. S. Thind, and M. R. Guarracino. 2019. From trash to treasure: detecting unexpected contamination in unmapped NGS data. *BMC Bioinformatics* 20: 168.
- Schäfer, M., A. R. Pacheco, R. Künzler, M. Bortfeld-Miller, C. M. Field, E. Vayena, V. Hatzimanikatis, and J. A. Vorholt. 2023. Metabolic interaction models recapitulate leaf microbiota ecology. *Science* 381: eadf5121.
- Scheepens, J. F., Y. Deng, and O. Bosssdorf. 2018. Phenotypic plasticity in response to temperature fluctuations is genetically variable, and relates to climatic variability of origin, in *Arabidopsis thaliana*. *AoB PLANTS* 10: ply043.
- The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
- Thompson, L. R., J. G. Sanders, D. McDonald, A. Amir, J. Ladau, K. J. Locey, R. J. Prill, et al. 2017. A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature* 551: 457–463.
- Vogel, C. M., D. B. Potthoff, M. Schäfer, N. Barandun, and J. A. Vorholt. 2021. Protective role of the *Arabidopsis* leaf microbiota against a bacterial pathogen. *Nature Microbiology* 6: 1537–1548.
- Vorholt, J. A. 2012. Microbial life in the phyllosphere. *Nature Reviews Microbiology* 10: 828–840.
- Vorholt, J. A., C. Vogel, C. I. Carlström, and D. B. Müller. 2017. Establishing Causality: Opportunities of Synthetic Communities for Plant Microbiome Research. *Cell Host & Microbe* 22: 142–155.
- Wagg, C., K. Schlaeppli, S. Banerjee, E. E. Kuramae, and M. G. A. van der Heijden. 2019. Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning. *Nature Communications* 10: 4841.
- Wagner, M. R., D. S. Lundberg, T. G. del Rio, S. G. Tringe, J. L. Dangl, and T. Mitchell-Olds. 2016. Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nature Communications* 7: 12151.
- Welburn, J. 2024. How to reduce your lab’s carbon footprint. *Journal of Cell Science* 137: jcs263722.
- Wetterstrand, K. 2025. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing

- Program (GSP). Website [www.genome.gov/sequencingcostsdata](http://www.genome.gov/sequencingcostsdata).
- Winter, N., R. Marchand, C. Lehmann, L. Nehlin, R. Trapannone, D. Rokvić, and J. Dobbelaere. 2023. The paradox of the life sciences. *EMBO reports* 24: e56683.
- Wipf, H. M.-L., T.-N. Bui, and D. Coleman-Derr. 2021. Distinguishing Between the Impacts of Heat and Drought Stress on the Root Microbiome of *Sorghum bicolor*. *Phytobiomes Journal* 5: 166–176.
- Wipfel, K., K. Tao, Y. Niu, R. Zgadzaj, N. Kiel, R. Guan, E. Dahms, et al. 2021. Host preference and invasiveness of commensal bacteria in the Lotus and Arabidopsis root microbiota. *Nature Microbiology* 6: 1150–1162.
- Xin, X.-F., and S. Y. He. 2013. *Pseudomonas syringae* pv. tomato DC3000: A Model Pathogen for Probing Disease Susceptibility and Hormone Signaling in Plants. *Annual Review of Phytopathology* 51: 473–498.
- Yates, J. A. F., T. C. Lamnidis, M. Borry, A. A. Valtueña, Z. Fagernäs, S. Clayton, M. U. Garcia, et al. 2021. Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. *PeerJ* 9: e10947.
- Zheng, L., K. E. Barry, N. R. Guerrero-Ramírez, D. Craven, P. B. Reich, K. Verheyen, M. Scherer-Lorenzen, et al. 2024. Effects of plant diversity on productivity strengthen over time due to trait-dependent shifts in species overyielding. *Nature Communications* 15: 2078.

## Chapter 2

**Understanding plant microbiomes requires a genotype × environment framework**

# Understanding plant microbiomes requires a genotype $\times$ environment framework

Jun Hee Jung<sup>1</sup> | Frank Reis<sup>1</sup> | Christina L. Richards<sup>1,2</sup> | Oliver Bossdorf<sup>1</sup>

<sup>1</sup>Plant Evolutionary Ecology, University of Tübingen, Auf der Morgenstelle 5, 72076 Tübingen, Germany

<sup>2</sup>Department of Integrative Biology, University of South Florida, Tampa, FL 33620, USA

## Correspondence

Oliver Bossdorf, Plant Evolutionary Ecology, Institute of Evolution and Ecology, Auf der Morgenstelle 5, University of Tübingen, 72076 Tübingen, Germany.  
Email: [oliver.bossdorf@uni-tuebingen.de](mailto:oliver.bossdorf@uni-tuebingen.de)

## KEYWORDS

evolutionary ecology, G  $\times$  E, genotype-by-environment interactions, microbial communities, multifactorial experiments, natural variation, pathogens, stress tolerance

Plant microbiomes have become one of the hottest topics in plant biology. Driven by the increased availability of microbiome sequencing methods, analyses of plant-associated microbiomes have been skyrocketing during the last decade. They have generally taken one of two main perspectives: (1) a focus on the microbiome itself, where researchers describe microbiome diversity and attempt to understand its drivers (Figure 1A), or (2) a focus on the consequences of microbiomes, where researchers analyze effects of microbiomes on plants (Figure 1B). Below, we briefly discuss these two perspectives, and we argue that for both a genotype-by-environment (G  $\times$  E) framework will be key for achieving a deeper and more general understanding of plant microbiomes.

## TWO PERSPECTIVES IN PLANT MICROBIOME RESEARCH

Studies with a microbiome focus typically describe the diversity and composition of the root or leaf microbiomes of plants, and they often test influences of plant characteristics or environmental conditions on the microbiomes (Figure 1A). For instance, previous studies have shown that microbiome composition varies within and among plant species, with significant influences of plant genotype (Agler et al., 2016; Wagner et al., 2016; Bowen et al., 2017; Bergelson et al., 2019) and phylogeny (Fitzpatrick et al., 2018) and that not only plant tissue but also plant age and developmental stage (Chaparro et al., 2014; Wagner et al., 2016) influence plant microbiomes. In addition to plant characteristics, environmental conditions

also play a role in microbiome development. For instance, soil conditions and root exudates strongly influence root microbiomes (Fitzpatrick et al., 2018; Hu et al., 2018; Sasse et al., 2018), and leaf microbiomes vary predictably among different habitats (Agler et al., 2016; Wagner et al., 2016).

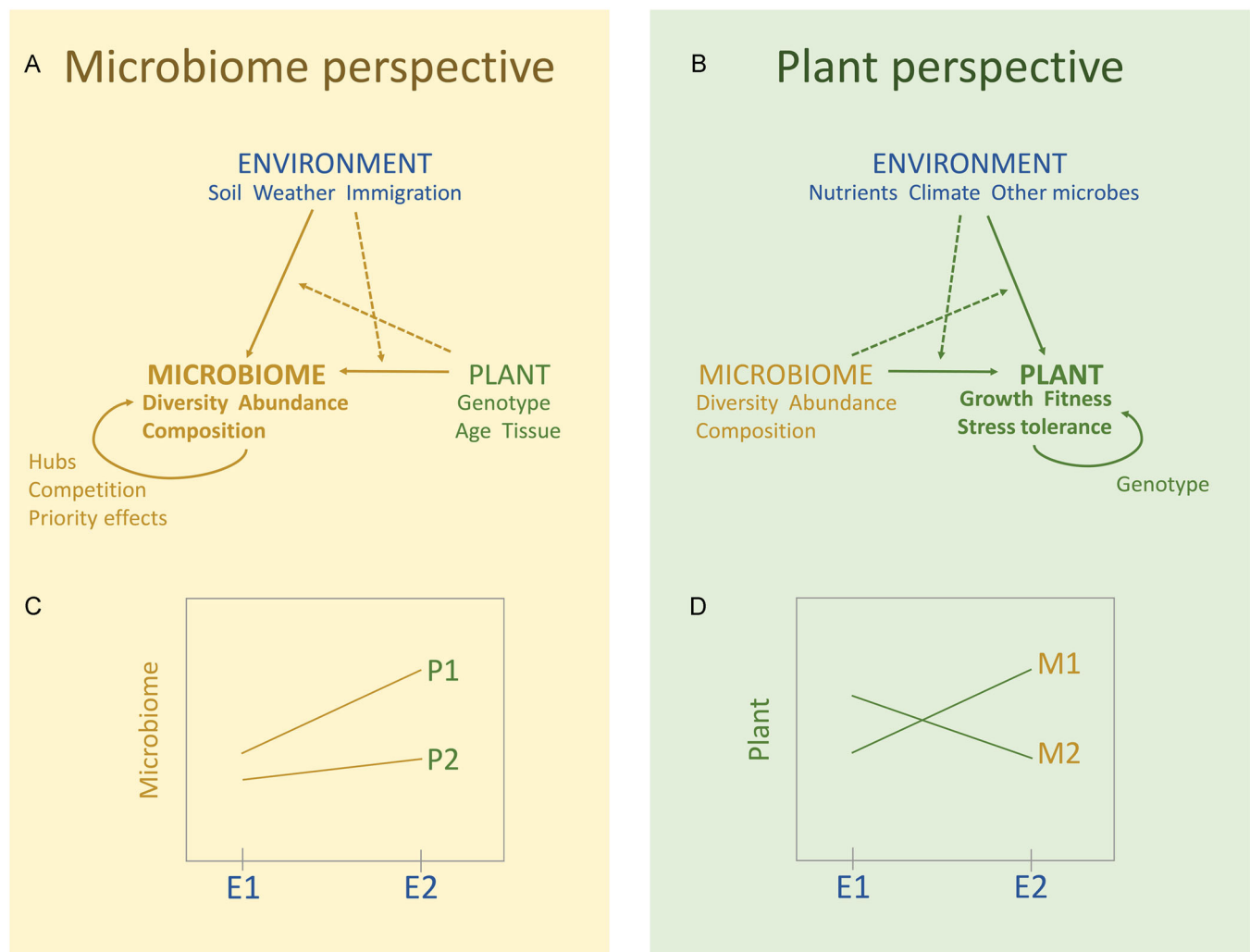
Microbiome studies with a plant focus, in contrast, are interested in how the microbiome of a plant affects its growth or environmental tolerances. They test—through experiments or association patterns—how plant performance is influenced by the presence or composition of microbiota, sometimes also taking plant or environmental characteristics into account (Figure 1B). Studies with individual microbial taxa have demonstrated that these can promote growth and stress tolerance of plants (Lugtenberg and Kamilova, 2009) and influence pathogen and herbivore resistance (Pieterse et al., 2014; Hu et al., 2018). More recent studies with complex microbiomes have confirmed these effects: inoculation with diverse microbiota altered plant growth and physiology (Fitzpatrick et al., 2019; Belimov et al., 2020), phenology (Panke-Buisse et al., 2015), and pathogen resistance (Berendsen et al., 2018; Berg and Koskella, 2018), sometimes in a genotype- or environment-dependent fashion (Berg and Koskella, 2018; Belimov et al., 2020).

In summary, recent research has demonstrated the ubiquity and importance of plant microbiomes, but this work has also shown that microbiomes are complex and influenced by a range of plant and environmental factors. Another challenge is that drivers of microbiome variation often interact. For instance, in a multi-site field experiment with different genotypes of *Boechera stricta*, Wagner et al. (2016) found that genotype and age effects on bacterial microbiomes were often

Jun Hee Jung and Frank Reis contributed equally.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *American Journal of Botany* published by Wiley Periodicals LLC on behalf of Botanical Society of America.



**FIGURE 1** Two main perspectives in plant microbiome research, and reaction norm plots illustrating possible  $G \times E$  interactions tested in each. With a microbiome perspective, researchers usually study community-level characteristics of the microbiome and test effects of plant genotype or other plant characteristics (P1/P2) and environmental conditions (E1/E2). With a plant perspective, the dependent variables are measures of plant performance, and experiments test influences of microbiomes (M1/M2), environmental conditions, and their interactions. The dashed lines in the upper graphs indicate indirect effects where some drivers of plant or microbiome variation alter plant or microbiome responses to others. Note that, while the bottom graphs display only categorical variables, the  $G \times E$  framework can be equally applied to continuous explanatory variables

site-specific. In an experiment with natural accessions of *Arabidopsis thaliana*, Fitzpatrick et al. (2019) found that the effects of a natural soil microbiome on plant fitness depended not only on the plant genotype but also on the ecological context. Similarly, Petipas et al. (2020) showed that the effects of soil microbes on the growth and fitness of *Hypericum perforatum* were habitat- and plant-genotype-specific.

In spite of the many and often interacting drivers of microbiome diversity and microbiome effects, the vast majority of previous studies focused on one or few drivers, often in reductionist laboratory settings, and there have been few solid multifactorial studies to date that allowed testing for interactions between different factors. As a result, the generality of many previous results, and their transferability to more natural environmental conditions, remains uncertain, and we are still far from truly understanding natural plant

microbiomes. A significant step forward could be to embrace an important conceptual framework from evolutionary ecology: that of genotype-by-environment interactions.

## ADOPTING A $G \times E$ FRAMEWORK

Genotype-by-environment ( $G \times E$ ) interactions are statistical interactions between the effects of genotypes and environment on phenotypes, that is, when phenotypic differences among genotypes depend on the environment in which they are tested, or when phenotypic responses to environment depend on the genotype. The  $G \times E$  concept has long been central to plant evolutionary ecology (Nicotra et al., 2010), and a large body of research has often found strong  $G \times E$  interactions in many plant species, to the extent that

genotype effects may be strong in some but absent in other environments, and phenotypic responses to environment are sometimes opposite for different genotypes. Similar results in animal research confirmed that  $G \times E$  interactions are the rule in natural populations and that caution is needed with generalizing from single-factor studies.  $G \times E$  experimental designs are essential for understanding and documenting local adaptation, and they are also relevant for adjusting agricultural and forestry practices to changing environments (Nicotra et al., 2010).

The classic experimental approach to testing  $G \times E$  interactions is a common garden experiment where multiple genotypes are replicated across different environments in a multifactorial design, so that the generality of genotype and environment effects, as well as their interactions, can be statistically tested. The results of such experiments are often visualized through reaction norm plots that show genotype-specific responses to environment or other relevant interactions (Figure 1C, D). We can easily apply these concepts and experimental approaches to a multifactorial study of plant microbiomes, both for the microbiome and the plant perspective.

Microbiome-focused studies with a  $G \times E$  character essentially treat the microbiome as an “extended phenotype” of the plant that is subject to the same complex influences as other plant phenotypes. Such studies will, for example, test the influence of plant genotype on plant microbiomes under different environmental conditions, or they will examine environmental effects on plant microbiomes across multiple plant genotypes (Figure 1C). For instance, field experiments can transplant multiple plant genotypes into different habitats and test the interactive effects of genotype and habitat on spontaneous microbiome development (Wagner et al., 2016). In laboratory experiments, different plant genotypes can be inoculated with identical microbial communities and their divergence followed under different growth conditions. An extension of the  $G \times E$  concept may also include plant factors other than genotype (e.g., plant tissue, age, or even species). More complex studies may include several environmental factors and/or additional microbial drivers of microbiome composition (Figure 1A).

Plant-focused studies with a  $G \times E$  framework generally test microbiome effects on plant performance or stress tolerance not only for one narrow type of experimental set-up but across a range of different environments and/or multiple plant genotypes (Figure 1B). For instance, beneficial or pathogenic effects of different microbial inoculates can be studied under several, controlled levels of resource availability or abiotic stress (Fitzpatrick et al., 2019). For more realistic tests, seedlings can be inoculated with different microbial communities and planted in a range of natural habitats (Petipas et al., 2020). Depending on one's perspective and strategy of data analysis, these approaches will examine how microbiome effects on plants are modulated by environmental influences or, equally important, how plant responses to the environment (phenotypic plasticity; environmental tolerances) are modulated by microbes

(Figure 1D). Finally, the environmental component in such experiments may also include additional biotic factors such as competitors (Fitzpatrick et al., 2019), herbivores, or other (background) microbiota, which will allow for testing microbe–microbe interactions or other complex biotic interactions.

## CONCLUSIONS

The study of plant microbiomes is an important frontier in current plant biology, with many open questions, particularly from an ecological-evolutionary (Koskella et al., 2017; Fitzpatrick et al., 2020) and agricultural perspective (Toju et al., 2018). Irrespective of whether the focus is on the plant or the microbiome, studies that are too simple may overlook important interactions between different factors, and they run the risk of overestimating or overgeneralizing results. A more thorough understanding of plant microbiomes will require working with a broader range of plant genotypes and nonmodel species, and to take a  $G \times E$  perspective and explicitly test the generality of plant–microbiome interactions across multiple interacting drivers. Such studies will not only be relevant for researchers with an ecological-evolutionary mindset, but they will ultimately also allow those with a more reductionist focus on functional and molecular details to better understand the robustness of their insights and refine their models of plant–microbe interactions.

## ACKNOWLEDGMENTS

We thank Pam Diggle and two anonymous reviewers for helpful comments on previous versions of this manuscript. Our work was supported by the Deutsche Forschungsgemeinschaft (DFG; projects BO 3241/10-1 and BO 3241/11-1 to O.B.) as part of the priority programme 2125 Deconstruction and Reconstruction of the Plant Microbiota “DECrypT”, and by the German Federal Ministry of Education and Research (BMBF; MOPGA project 306055 to C.L.R.). Open Access funding enabled and organized by Projekt DEAL.

## AUTHOR CONTRIBUTIONS

All authors jointly developed the ideas in this essay and contributed to writing the manuscript.

## ORCID

Christina L. Richards  <https://orcid.org/0000-0001-7848-5165>

Oliver Bossdorf  <http://orcid.org/0000-0001-7504-6511>

## REFERENCES

- Agler, M. T., J. Ruhe, S. Kroll, C. Morhenn, S.-T. Kim, D. Weigel, and E. M. Kemen. 2016. Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biology* 14: e1002352.
- Belimov, A. A., A. I. Shaposhnikov, D. S. Syrova, A. A. Kichko, P. V. Guro, O. S. Yuzikhin, T. S. Azarova, et al. 2020. The role of symbiotic microorganisms, nutrient uptake and rhizosphere bacterial

- community in response of pea (*Pisum sativum* L.) genotypes to elevated Al concentrations in soil. *Plants* 9: 1801.
- Berendsen, R. L., G. Vismans, K. Yu, Y. Song, R. de Jonge, W. P. Burgman, M. Burmölle, et al. 2018. Disease-induced assemblage of a plant–beneficial bacterial consortium. *ISME Journal* 12: 1496–1507.
- Berg, M., and B. Koskella. 2018. Nutrient- and dose-dependent microbiome-mediated protection against a plant pathogen. *Current Biology: CB* 28: 2487–2492.
- Bergelson, J., J. Mittelstrass, and M. W. Horton. 2019. Characterizing both bacteria and fungi improves understanding of the *Arabidopsis* root microbiome. *Scientific Reports* 9: 24.
- Bowen, J. L., P. J. Kearns, J. E. K. Byrnes, S. Wigginton, W. J. Allen, M. Greenwood, K. Tran, et al. 2017. Lineage overwhelms environmental conditions in determining rhizosphere bacterial community structure in a cosmopolitan invasive plant. *Nature Communications* 8: 433.
- Chaparro, J. M., D. V. Badri, and J. M. Vivanco. 2014. Rhizosphere microbiome assemblage is affected by plant development. *ISME Journal* 8: 790–803.
- Fitzpatrick, C. R., J. Copeland, P. W. Wang, D. S. Guttman, P. M. Kotanen, and M. T. J. Johnson. 2018. Assembly and ecological function of the root microbiome across angiosperm plant species. *Proceedings of the National Academy of Sciences, USA* 115: E1157–E1165.
- Fitzpatrick, C. R., Z. Mustafa, and J. Viliunas. 2019. Soil microbes alter plant fitness under competition and drought. *Journal of Evolutionary Biology* 32: 438–450.
- Fitzpatrick, C. R., I. Salas-González, J. M. Conway, O. M. Finkel, S. Gilbert, D. Russ, P. J. P. L. Teixeira, and J. L. Dangl. 2020. The plant microbiome: from ecology to reductionism and beyond. *Annual Review of Microbiology* 74: 81–100.
- Hu, L., C. A. M. Robert, S. Cadot, X. Zhang, M. Ye, B. Li, D. Manzo, et al. 2018. Root exudate metabolites drive plant–soil feedbacks on growth and defense by shaping the rhizosphere microbiota. *Nature Communications* 9: 2738.
- Koskella, B., L. J. Hall, and C. J. E. Metcalf. 2017. The microbiome beyond the horizon of ecological and evolutionary theory. *Nature Ecology & Evolution* 1: 1606–1615.
- Lugtenberg, B., and F. Kamilova. 2009. Plant-growth-promoting rhizobacteria. *Annual Review of Microbiology* 63: 541–556.
- Nicotra, A. B., O. K. Atkin, S. P. Bonser, A. M. Davidson, E. J. Finnegan, U. Mathesius, P. Poot, et al. 2010. Plant phenotypic plasticity in a changing climate. *Trends in Plant Science* 15: 684–692.
- Panke-Buisse, K., A. C. Poole, J. K. Goodrich, R. E. Ley, and J. Kao-Kniffin. 2015. Selection on soil microbiomes reveals reproducible impacts on plant function. *ISME Journal* 9: 980–989.
- Petipas, R. H., A. C. Wruck, and M. A. Geber. 2020. Microbe-mediated local adaptation to limestone barrens is context dependent. *Ecology* 101: e03092.
- Pieterse, C. M. J., C. Zamioudis, R. L. Berendsen, D. M. Weller, S. C. M. Van Wees, and P. A. H. M. Bakker. 2014. Induced systemic resistance by beneficial microbes. *Annual Review of Phytopathology* 52: 347–375.
- Sasse, J., E. Martinoia, and T. Northen. 2018. Feed your friends: Do plant exudates shape the root microbiome? *Trends in Plant Science* 23: 25–41.
- Toju, H., K. G. Peay, M. Yamamichi, K. Narisawa, K. Hiruma, K. Naito, S. Fukuda, et al. 2018. Core microbiomes for sustainable agroecosystems. *Nature Plants* 4: 247–257.
- Wagner, M. R., D. S. Lundberg, T. G. Del Rio, S. G. Tringe, J. L. Dangl, and T. Mitchell-Olds. 2016. Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nature Communications* 7: 12151.

**How to cite this article:** Jung, J. H., F. Reis, C. L. Richards, and O. Bossdorf. 2021. Understanding plant microbiomes requires a genotype × environment framework. *American Journal of Botany* 108(10): 1820–1823. <https://doi.org/10.1002/ajb2.1742>

## Chapter 3

### **Leaf bacterial diversity promotes growth and pathogen resistance in *Arabidopsis thaliana***

**Leaf bacterial diversity promotes growth and pathogen resistance in *Arabidopsis thaliana***

Jun Hee Jung<sup>1</sup>, Johannes Kasper<sup>1,2</sup>, Frank Reis<sup>1</sup>, Nina Porath<sup>1</sup>, Madalin Parepa<sup>1</sup>, Julia Vorholt<sup>3</sup>, Oliver Bossdorf<sup>1</sup>

<sup>1</sup> Plant Evolutionary Ecology, Institute of Evolution and Ecology, University of Tübingen, Germany

<sup>2</sup> Vegetation Ecology and Conservation Biology, Institute of Ecology, University of Bremen, Germany

<sup>3</sup> Institute of Microbiology, ETH Zurich, Zurich, Switzerland

## Abstract

An increase in biodiversity increases the function and stability of an ecosystem, with this relationship observed in a wide range of communities, especially in terrestrial environments. Plants are inhabited by a diverse range of microorganisms and this microbial community is referred to as the microbiome. Next-generation sequencing has enabled research of plant-associated microbiomes, and findings have shown that increases in bacterial diversity in both roots and leaves improve the performance of their ecosystem, i.e. the host plant. However, evidence from the deliberate manipulation of microbiome diversity on plant performance is rare. Here, we show that when a plant leaf is inoculated with a leaf-associated microbiome with higher diversity, the plant has increased growth and tolerance toward the pathogen *Pseudomonas viridiflava*. The increase in diversity was achieved by a random selection of strains in a full factorial combination of the total number of unique bacterial strains (3 or 9) and the number of phylogenetic clades (1 or 3). We found that an increase in the phylogenetic clade diversity as well as the increase in number of strains in a microbial community increased plant growth, and clade diversity was important for pathogen tolerance. We calculated the predicted functional and phylogenetic diversity of each microbiome using the strain genetic information. Functional diversity was better able to explain both growth and pathogen tolerance than phylogenetic diversity. Using the phylogenetic clade information, we found that interactions between clades had both positive and negative effects on plant growth. Our results demonstrate that the microbiome of various levels of diversity can be assembled and a positive relationship between diversity and performance exists. We anticipate more research linking leaf microbial diversity and plant performance, especially across different environmental factors, as well as precise interactions between individual strains or their clades.

Keywords: phyllosphere, biodiversity, ecosystem function, microbiome, pathogen resistance

## Introduction

Global biodiversity is in rapid decline (Barnosky et al., 2011). The sharp decline in biodiversity is worrying, as decades of research have shown a clear positive relationship between biodiversity and the ecosystem productivity, function, and stability (Loreau et al., 2001; Hooper et al., 2005; Cardinale et al., 2012). There have been several proposed mechanisms behind this positive relationship (referred to as biodiversity-ecosystem function, BEF), including: insurance hypothesis (species with similar function can compensate each other), facilitation effect (abiotic and biotic stress mitigation by the community changing the biotic and abiotic conditions), and complementarity by increased efficiency in resource partitioning (Ali, 2023; Eisenhauer et al., 2024). To capture these relationships, indices such as species richness, functional diversity, and phylogenetic diversity are often used to approximate biodiversity. Species richness (number of species) is a simple method that has been criticized for lacking predictive nature (Loreau et al., 2001; van der Plas, 2019). Functional diversity refers to the number of ecological functions (Tilman et al., 2014), however functional features can be numerous and difficult to measure. Phylogenetic diversity (PD) , which uses species' evolutionary distance to calculate diversity, has been proposed as a proxy for functional diversity (Faith, 1992, 2018). There have been several papers, meta-analyses, and reviews comparing the different diversity measurement methods, and to date there has been no clear agreement on the most effective index for capturing the BEF relationship (Cadotte et al., 2008, 2011; Venail et al., 2015; van der Plas, 2019).

Although the BEF relationship between plants and terrestrial ecosystems is well documented, less is known about the relationship between the plant microbial community (microbiome) and the host plant, seen as its own ecosystem. The plant microbiome has a close-knit relationship with host health and functioning (Gao et al., 2021). Current evidence suggests that the leaf microbiome is critical for plant function, positively affecting plant growth and pathogen tolerance (Firrincieli et al., 2020; Emmenegger et al., 2023; Hoefle et al., 2024). This beneficial relationship is bidirectional; an increase in plant richness (species or functional) increases leaf bacterial diversity, which in turn creates a forest community with higher productivity (Laforest-Lapointe et al., 2017). It is therefore crucial to understand which factors affect plant microbial communities. For instance, plant and microbiome interactions are tissue specific, with tissues having different microbiome compositions (Edwards et al., 2015; Firrincieli et al., 2020). Each tissue poses unique challenges to the microbiome, for example, microorganisms living in plant leaves face poor nutrient availability, increased UV stress, temperature and water fluctuations, and attack from pathogens (Chaudhry et al., 2021; Almario et al., 2022). However, despite this importance, establishing a direct link between the biodiversity of microbiomes to plant

performance is difficult. Unlike a terrestrial community wherein plant species richness can be easily manipulated, for a plant-microbial community a gnotobiotic system is required, with increased experimental difficulty.

Recent progress in the development of synthetic microorganism communities (syncom) offers a unique opportunity to study plant-microbial interactions (Vorholt et al., 2017). A synthetic community refers to a gnotobiotic system in which a sterile plant is inoculated with a known microbial community from a bacterial isolate originally from the same host (Northen et al., 2024). A typical syncom strain collection also often has rich supplementary information such as their genome sequence (information provided at <https://www.at-sphere.com/ccquery/ccview/>, an example of a syncom collection). In model species such as *Arabidopsis thaliana*, barley, and *Lotus japonicus*, bacterial isolate collections are available to assemble *de novo* communities of syncom (Bai et al., 2015; Wippel et al., 2021; Robertson-Albertyn et al., 2022). The ability to assemble new communities can be used to manipulate community diversity and see the effect on plant performance, especially in plant-pathogen interactions. For example, *Pseudomonas viridiflava* isolated from *Arabidopsis* displays a large variation in pathogenicity, ranging from pathogenic to commensal (Karasov et al., 2018). When commensal *Pseudomonas* is co-inoculated with pathogenic *Pseudomonas*, the commensal bacteria elicit the host response to suppress the pathogenic *Pseudomonas* (Shalev et al., 2022b). The syncom approach can therefore provide insight into both plant-microbiome and plant-microbiome-microbiome interactions.

Here, we used an *Arabidopsis thaliana* leaf bacteria collection (At-LSPHERE; Bai et al., 2015) and assembled 40 microbiomes by the factorial manipulation of total strain number within the community and the number of clades in the community. The assembled microbiomes were used to test the effect on the performance of host *Arabidopsis thaliana* (growth and *Pseudomonas viridiflava* pathogen resistance) based on image analysis measuring the plant size over multiple time points. Furthermore, by using available sequence resources (Bai et al., 2015; Vogel et al., 2021; Emmenegger et al., 2023), we calculated the initial phylogenetic and functional diversity of the microbiome community, and compared two types of indices in explaining our growth and disease resistance data. We asked the following questions: 1) Does an increase in leaf microbial diversity increase the host performance (growth and *Pseudomonas* resistance) and if so, which factor is more important?; 2) Which of the continuous diversity indices best explain plant performance?; 3) Is there any effect of microbiome diversity on plant performance, and what effect does the presence of multiple clades have on plant performance?

## Methods

### *Study system*

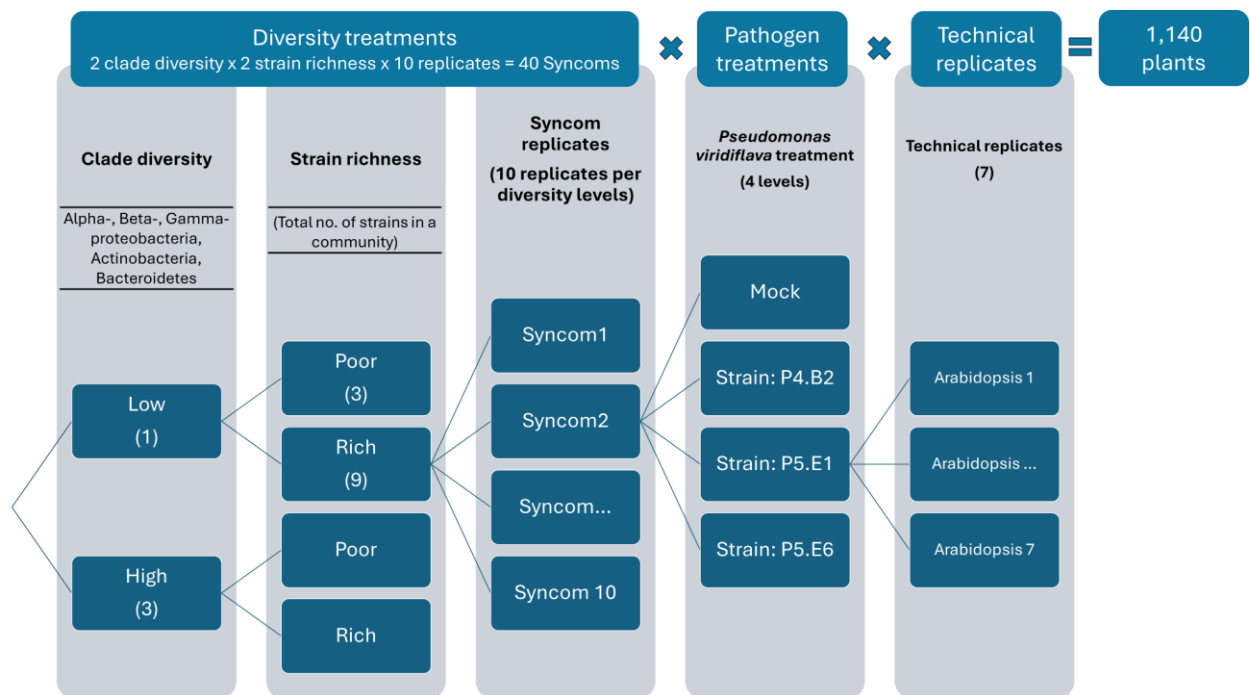
We worked with *Arabidopsis thaliana* genotype Columbia (Col-0). The leaf-associated bacterial strains used in the experiment were a 138-strain subset of a At-LSPHERE leaf-associated bacterial collection (Bai et al., 2015; subset defined from Pfeilmeier et al., 2024, <https://www.at-sphere.com/>, appendix 1). Three strains of *Pseudomonas viridiflava* (P4.B2, P5.E1, and P5.E6; NCBI RefSeq IDs: GCF\_900599365.1, GCF\_900590775.1, and GCF\_900599395.1, respectively) from a *Pseudomonas* ATU5 clade collection were also used (Karasov et al., 2018).

### *Overall experiment design*

We performed a full factorial experiment in which plants were inoculated with microbiomes and then subjected to *Pseudomonas viridiflava* pathogens (Figure 1). We randomly inoculated 11 day-old *Arabidopsis thaliana* with microbiomes from one of the four diversity levels, and after 7 days we randomly infected the 18 day-old plants with either one of three strains of *Pseudomonas* pathogen or a sterile solution. After 7 days post infection, 25 day-old plants were harvested. One week between inoculation and infection allowed the effect of microbiome on plant growth and pathogen resistance to be tested. Images of the plants were taken on the days of inoculation, infection, and harvest to calculate the plant size.

The four microbiome diversity levels were created by a two-way full factorial combination of clade diversity and strain richness (Figure 1). The clade diversity levels were either high (having 3 clades in a microbiome) or low (having only 1 clade in a microbiome), and in high clade diversity communities, equal proportions of each clade were present. A clade was defined as the highest phylogenetic group containing at least 10 strains. The phyla Proteobacteria, Actinobacteria, and Bacteroidetes consisted of 10 or more strains. To select three random clades, the phylum Proteobacteria was further divided into lower taxonomic ranks (class): Alphaproteobacteria, Betaproteobacteria, and Gammaproteobacteria. Strain richness was either rich (nine strains in a microbiome) or poor (three total strains in a microbiome), and strains were randomly selected. For each one of the four resulting microbiome diversity levels (two strain richness and two clade diversity levels), 10 microbiome replicates were created, bringing the total number of microbiomes created to 40. *Pseudomonas* treatments contained one of the three *Pseudomonas* strains P4.B2, P5.E1, P5.E6, or a sterile mock solution as a control.

In total 1,120 plants were grown, consisting of 40 microbiomes used for inoculation, four pathogen strains used for infection, and seven replicates per treatment combination (Figure 1).



**Figure 1.** Schematic for the experimental design. The five grey boxes refer to the different sources of replicates (synthetic communities (syncom) and technical) and the three-way full factorial design (clade diversity, strain richness, and pathogen treatments). The blue boxes above refer to the interaction between the two microbiome diversity levels and the pathogen treatments, and their resulting total number of plants used in this experiment. Ellipses refer to values skipped between the above and below boxes to conserve space (used in “syncom replicates” and “technical replicates”).

#### *Arabidopsis thaliana* growth condition

Two *Arabidopsis thaliana* genotype Col-0 seeds were sown onto six-well cell culture plates (Thermo scientific, USA) containing half-strength Murashige and Skoog (MS) media with MES buffer (Duchefa Biochemie, Netherlands). Plants were grown under sterile conditions with subsequent inoculation and infection carried out at a clean bench, and plates were sealed with Parafilm between every step to ensure minimal contamination. The seeds were stratified in darkness at 4 °C for three days, and plants were grown in a 21 °C day 18 °C night cycle under long day (14 h Day 8 h Night) light conditions.

#### *Microbiome assembly and inoculation*

The bacterial strains were streaked out on R2A-agar with 0.5% v/v methanol (HiMedia, India) at 22 °C for at least 4 days. From the inoculated plate, we took a loopful of bacteria using a 1- $\mu$ L sterile inoculation loop and resuspended the bacterial material in 1 mL of 10 mM MgCl<sub>2</sub> by pipetting up and down to break down the biological mass. Bacterial suspensions were mixed

together according to their randomized groups (Figure 2, Table S2). An equal volume of liquid R2A-MeOH with 25% v/v glycerol was added to bacterial suspension (12.5% v/v final glycerol concentration) and kept at -80 °C. For the inoculation of the microbial community, the bacterial inoculum was rapidly defrosted and washed twice with 10 mM MgCl<sub>2</sub>. The optical density at 600 nm (OD<sub>600</sub>) of washed cells was adjusted to 0.02 using 10 mM MgCl<sub>2</sub>. The individual plants were drip inoculated with 200 µl of inoculum.

### *Plant infection*

Three *Pseudomonas viridiflava* strains (p5.E1, p4.B2, and p5.E6) belonging to OTU/ATU 5 were initially grown overnight in 2 ml LB medium with 100 ng/mL nitrofurantoin (NF) at 28 °C (Karasov et al., 2018; Duque-Jaramillo et al., 2023). After the overnight growth phase, *Pseudomonas* strains were diluted 1:10 and allowed to grow exponentially for four hours in fresh media. The *Pseudomonas* strains were then diluted to the final OD<sub>600</sub> of 0.01 in 10 mM MgSO<sub>4</sub>. 18 day-old plants were inoculated with 100 µl of the *Pseudomonas* inoculum, and control plants were mock inoculated with sterile 10 mM MgSO<sub>4</sub>.

### *Plant and diversity measurement*

Plant growth and response to pathogen infection was measured by image analysis of photos taken using a digital camera (Nikon D750, Japan) mounted on a tripod. The plant images were taken in a dark room, with a light plate underneath the cell culture plates (Rex Leuchtplatten, Germany). The green pixel counting was done using Python 3.11 using packages *OpenCV2*, *PlantCV*, and *Matplotlib* (Bradski, 2000; Hunter, 2007; Fahlgren et al., 2015). Each plant image was first brightened and converted from an RGB image into CIELAB color space. Non-green pixels were masked and remaining pixels were counted as “green pixels”, which were used as a proxy for plant size. The codes used for the analysis are available on GitHub: <https://github.com/junhee-jung>.

In total, six diversity indices were calculated: functional dispersion (FDis), functional divergence (FDiv), Rao's Q (RaoQ), functional richness (FRic), functional evenness (FEve) or Faith's phylogenetic diversity (PD). For functional diversity, we downloaded coding sequences of the strains from the NCBI assembly database. Based on the downloaded sequences, eggNOG-mapper v2 (Huerta-Cepas et al., 2019; Cantalapiedra et al., 2021; eggno-mapper.embl.de/) was used for functional prediction. After functional prediction, we performed functional enrichment at KEGG module level (Yu et al., 2012; Wu et al., 2021). Afterwards, the R package *fundiversity* was used to calculate functional diversity of the microbial communities (Grenié and Gruson, 2023). For phylogenetic diversity, we used 16s V5-V7

sequence based NJ-tree of the strains and the package *picante* to calculate Faith's phylogenetic diversity (Kembel et al., 2010).

### *Data analysis*

All statistical analyses were carried out using R (version 4.3.2; R Core Team, 2023) with associated packages and RStudio (2024.04.0 Build 735). We analyzed the results in three parts: the effect of microbiome diversity levels on plant performance (growth and *Pseudomonas* pathogen resistance, hereafter pathogen resistance), the effect of continuous diversity index on plant performance, and the effect of clade interactions on plant performance.

To analyze the effect of microbiome diversity on plant growth, the 18 day-old plants (after microbiome inoculation only) and 25 day-old plants (after microbiome inoculation and pathogen infection) were analyzed, using 11 day-old plant size (pre-inoculation) as a covariate. Clade diversity (2 levels), strain richness (2 levels), and their interaction were used as fixed factors.

The effect of microbiome diversity on pathogen resistance was analyzed using 25 day-old plant size with a covariate of 11 day-old or 18 day-old plant size. Clade diversity (2 levels), strain richness (2 levels), pathogen treatment (4 levels; 3 strains and 1 mock infection) and their interactions were used as fixed factors. 18 day-old plant size was used as a covariate to reduce the effect of the microbiome on plant growth prior to pathogen inoculation and focus on microbiome-pathogen interactions.

To analyze the effect of continuous indices on plant performance, we first confirmed that there were differences between the numerical index of the four microbiome diversity levels with *post-hoc* Tukey's Honestly Significant Difference (HSD) test. We analyzed the effect of continuous diversity on plant growth and pathogen resistance similarly to our microbiome diversity analysis. The fixed effects clade diversity, strain richness, and their interactions were replaced with functional diversities (FDis (Functional Dispersion), RaoQ (Rao's Q), Functional Divergence (FDiv), FRic (Functional richness), FEve (Functional evenness) or Faith's Phylogenetic diversity (PD)). The package *Performance* (Lüdecke et al., 2021) was used to determine the model fit using  $R^2$ , the Akaike information criterion (AIC), and the Bayesian information criterion (BIC). The model with highest  $R^2$  and lowest of AIC, BIC, and RMSE were selected as the model of best fit. The best-fitting model was analyzed for plant growth (25 day-old plant size and 18 day-old plant size, with 11 day-old plant size as covariate) and pathogen resistance (25 day-old plant size, with 18 day-old or 11 day-old plant size as covariate).

Based on the knowledge that there was significant influence of the clade on growth and pathogen resistance, we analyzed the effect of clade interactions; we tested how presence or absence (i.e. 2 levels) of a clade interacted with the presence or absence of another clade on plant growth and pathogen resistance. The initial model for growth had the equation: *Plant size (25 day-old OR 18 day-old) ~ Covariate:Plant size (11 day-old) + (Actinobacteria + Alphaproteobacteria + Bacteroidetes + Betaproteobacteria + Gammaproteobacteria)<sup>2</sup>*. The pathogen resistance model had the equation: *Plant size (25 day-old) ~ Covariate:Plant size (11 day-old or 18 day-old) + (Actinobacteria + Alphaproteobacteria + Bacteroidetes + Betaproteobacteria + Gammaproteobacteria)<sup>2</sup> \* Pathogen treatment*. Models contained only up to two-way clade interactions as fixed effects due to the lack of statistical power. We used the *MASS* package and stepwise regression model selection test to obtain the statistically-optimized model (Venables and Ripley, 2002). For the growth model, two-way significant clade interactions were analyzed and visualized further, and for the pathogen resistance, significant three-way clade-pathogen interactions.

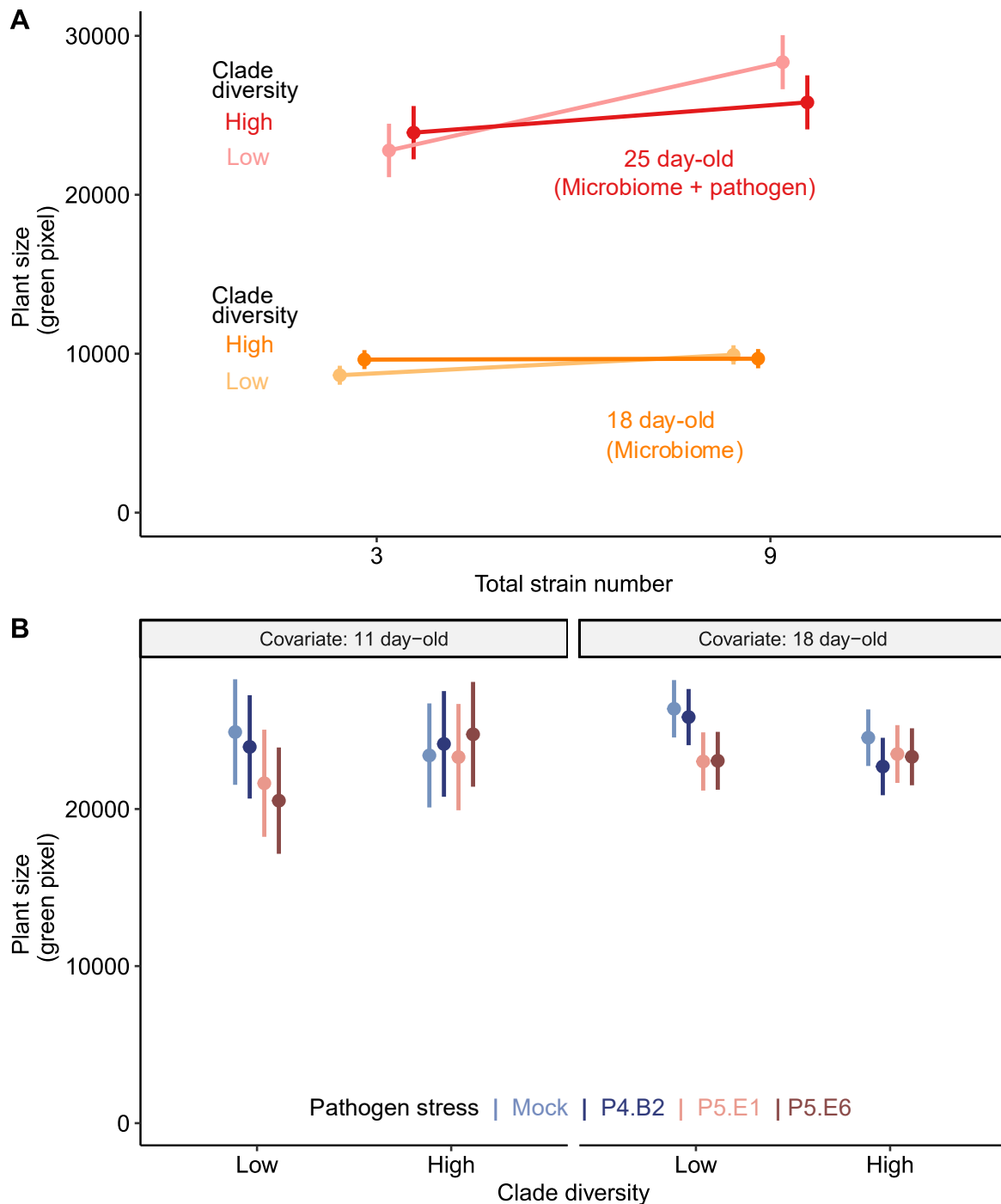
Finally, “Anova” from the package *car* was used to test the model (Fox and Weisberg, 2019). For the visualization of the plant performance analysis, predicted effect size based on GLM was visualized using *ggeffects* and *ggplot2* packages (Wickham, 2016; Lüdtke, 2018). For the growth analysis, only the predicted effect size based on two-way interaction was visualized for clade interaction. For pathogen resistance, only significant three-way interaction was visualized. The correlation between continuous indices were calculated and plotted using the “*GGally*” package (Schloerke et al., 2024).

## Results

### *The categorical increase in leaf bacterial diversity improves plant performance*

An increase in categorical microbiome diversity had a positive effect on plant growth. For plant growth analysis of 18 day-old plants (size after microbiome inoculation only) and 25 day-old plants (size after microbiome inoculation and pathogen infection), the use of the covariate (initial size, plant size at 11 days) corrected for the initial size differences between the plants (Table 1). Clade diversity alone did not have an influence on plant growth at either time point, but an increase in strain richness did have a positive influence on growth for both time points (Table 1, Figure 2A). There was a significant positive interaction between clade diversity and strain richness (Table 1, Figure 2A). The plant's response to the pathogen significantly depended on the covariates used (Table 1, Figure 2B). Overall, inoculation with *Pseudomonas* led to reduced plant size, and the effect was more pronounced when the previous effects of the microbiome diversity on growth was corrected for by using the covariate of plant size at

18 days old. There was a significant interaction between clade diversity and pathogen infection only when the covariate of 18 day-old plant size was used (Table 1).



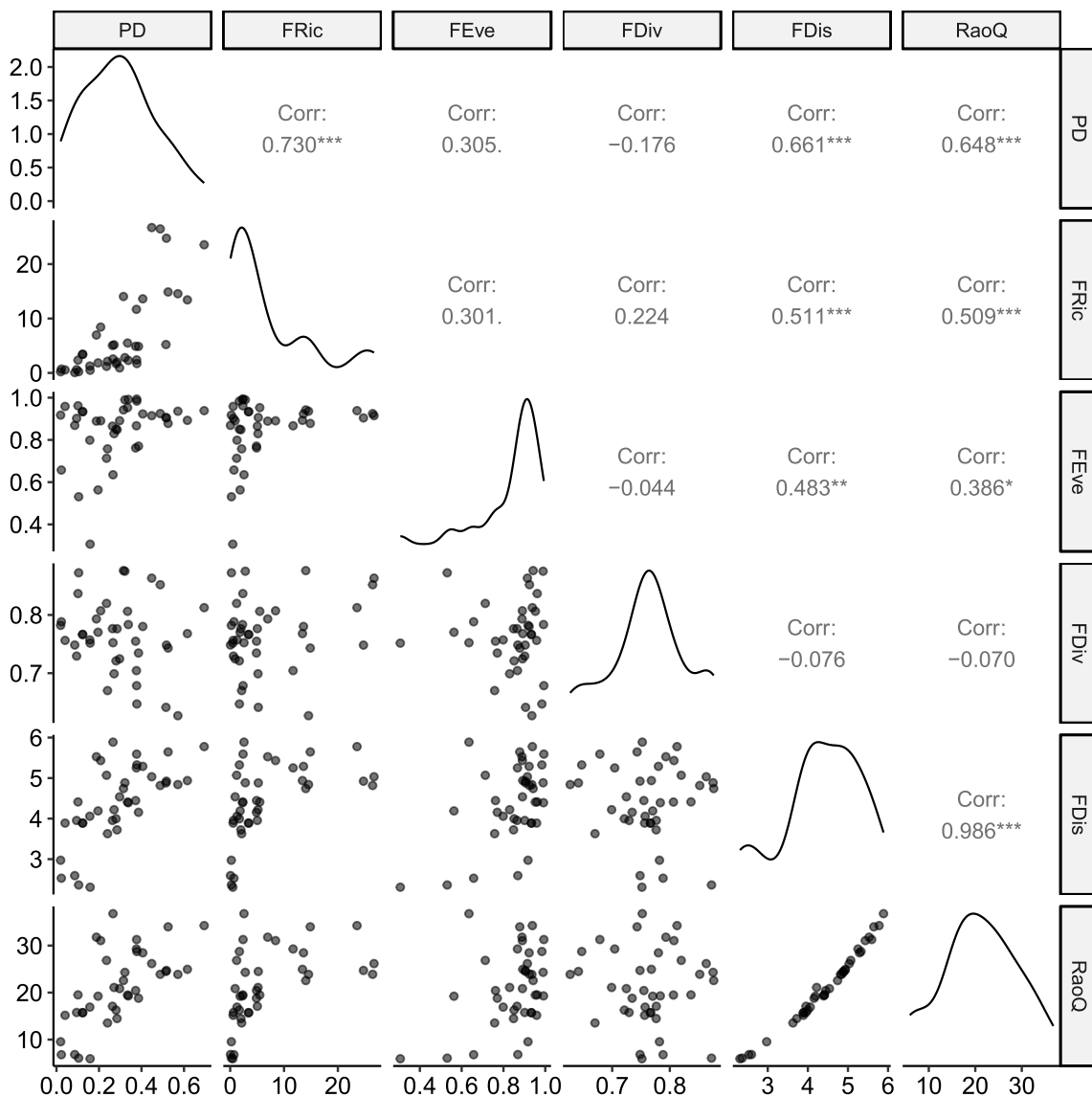
**Figure 2.** Predicted plant size from a generalized linear model depending on the growth analysis (A) and pathogen resistance analysis (B). The error bars represent 95% confidence intervals (A and B) and reaction norm is shown for strain richness and clade diversity interaction (A). In A, Low and high clade diversity are represented by lighter and darker red (plant size at 25 days; with microbiome and pathogen treatment) and orange (plant size at 18 days; with microbiome). In B, each point represents plant size at 25 days in different pathogen stress conditions including a mock treatment (light blue). In B, strain richness is not included for the pathogen resistance analysis, as no significant interaction was found between strain richness and pathogen stress or strain richness, clade diversity, and pathogen stress.

**Table 1.** Anova table of a generalized linear model testing for growth and pathogen resistance based on four categorical microbiome diversity levels. For growth analysis, plant sizes at two ages (18 and 25 days old) were used as response variables and were analyzed using clade diversity, strain richness, and their interaction as fixed effects, with plant size at 11 days old as a covariate. For pathogen resistance, 25 day-old plant size was used as a response variable, and was analyzed using clade diversity (CD), strain richness (SR), pathogen stress (PS), and their interaction as fixed effect. Plant sizes at 11 and 18 days old were used as covariates. The term pathogen stress refers to the four pathogen treatments: 3 strains and 1 mock treatment. Both *F* and *P* values (*Pr(>F)*) were rounded up to 3 decimal places. The statistically significant terms are highlighted in bold. Df = degree of freedom, *F* = *F*-value.

Growth				Pathogen resistance								
Plant age	18 day-old (microbiome)			25 day-old (microbiome + pathogen)								
Covariate used	11 day-old plant (no microbiome)			11 day-old plant			11 day-old plant			18 day-old		
	Df	<i>F</i>	<i>Pr(&gt;F)</i>	Df	<i>F</i> value	<i>Pr(&gt;F)</i>	Df	<i>F</i>	<i>Pr(&gt;F)</i>	Df	<i>F</i>	<i>Pr(&gt;F)</i>
Covariate	1	383.591	<b>&lt;0.001</b>	1	179.027	<b>&lt;0.001</b>	1	180.320	<b>&lt;0.001</b>	1	3178.110	<b>&lt;0.001</b>
Clade diversity (CD)	1	1.541	0.215	1	0.626	0.429	1	0.617	0.432	1	10.917	<b>&lt;0.001</b>
Strain richness (SR)	1	4.745	<b>0.030</b>	1	18.695	<b>&lt;0.001</b>	1	19.184	<b>&lt;0.001</b>	1	24.108	<b>&lt;0.001</b>
Pathogen stress (PS)							3	2.938	<b>0.032</b>	3	3.931	<b>0.008</b>
CD:SR	1	3.959	<b>0.047</b>	1	4.482	<b>0.034</b>	1	4.592	<b>0.032</b>	1	1.081	0.299
CD:PS							3	0.331	0.803	3	2.634	<b>0.049</b>
SR:PS							3	1.057	0.367	3	0.397	0.755
CD:SR:PS							3	0.929	0.426	3	1.045	0.372
Residuals	1089			1086			1077			1077		

*Functional dispersion is the best-performing continuous index*

There was a significant positive correlation between functional diversities and phylogenetic diversity of microbiomes (Figure 3; Functional dispersion: ANOVA<sub>Four diversity levels</sub> ( $F$  3,36) = 29.8,  $p < 0.001$ ; Phylogenetic diversity: ANOVA<sub>Four diversity levels</sub> ( $F$  3,36) = 52.5,  $p < 0.001$ ). As expected, the *post-hoc* Tukey test showed that both higher strain richness and clade diversity increased functional and phylogenetic diversity. The increase in clade diversity contributed to a bigger increase in both measures of diversity compared to the strain richness, and the microbiomes with high clade diversity and rich strain richness had the highest diversity values (statistics not shown).



**Figure 3.** Pairwise relationship between functional diversity indices and phylogenetic diversity (PD). Functional diversity indices used were FRic (Functional richness), FEve (Functional evenness), FDiv (Functional divergence), FDis (Functional dispersion), and Rao's Q (RaoQ). Pearson's correlation coefficient is shown (Corr:) in the top right panels, with significance represented by asterisks ( $P=0.1-0.5$ ;  $P=0.05-0.01^*$ ;  $P=0.01-0.001^{**}$ ).

The analysis of model fit quality and goodness of fit showed that, out of the six continuous indices tested, functional dispersion (FDis) was statistically the best at explaining plant growth and *Pseudomonas* resistance (Table S3). Following this, we further investigated the relationship between functional dispersion on plant growth and pathogen resistance. Similar to the previous analysis of the microbiome diversity levels on growth and pathogen resistance, covariate levels were highly significant (Table 2). The growth analysis showed that in plants at two ages (18 or 25 days; inoculated with microbiome only or microbiome and pathogen respectively) FDis had a significant positive influence on plant growth. Upon analysis of the effect of functional diversity on pathogen resistance, we did not observe any interaction between pathogen stress and FDis (Table 2).

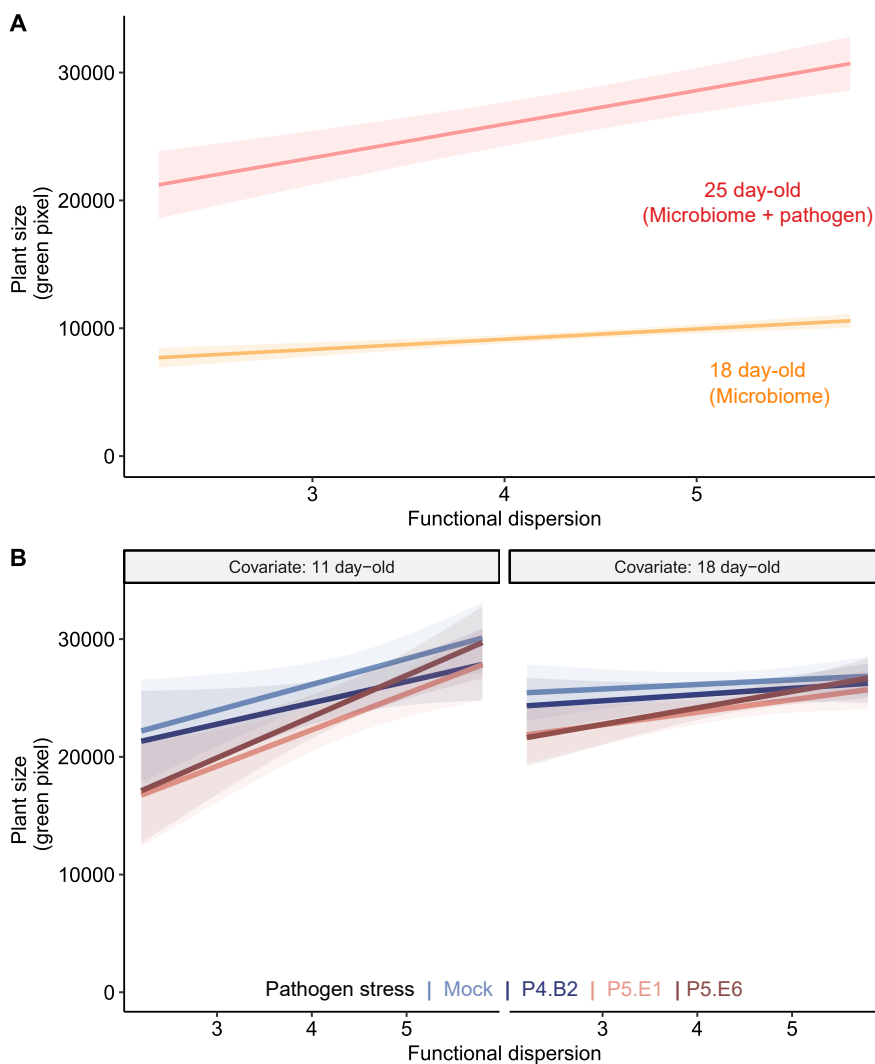
**Table 2.** Anova table of a generalized linear model testing for growth and pathogen resistance based on functional dispersion (FDis). For growth analysis, plant sizes at two ages (18 and 25 day-old) were used as a response variable and were analyzed using functional dispersion (FDis) as a fixed effect. For pathogen resistance, 25 day-old plant size was used as a response variable and was analyzed using FDis, pathogen stress (PS), and their interaction as fixed effects. The term pathogen stress refers to the three *Pseudomonas viridiflava* treatments and the mock control treatment. *P* values (*Pr(>F)*) are rounded up to 3 decimal places. The statistically significant terms are highlighted in bold. Df = degree of freedom, *F* = *F*-value.

Growth				Pathogen resistance								
Plant age	18 day-old (microbiome only)			25 day-old (microbiome + pathogen)								
Covariate	11 day-old plant (pre-microbiome inoculation)						18 day-old plant					
	Df	F	<i>Pr(&gt;F)</i>	Df	F value	<i>Pr(&gt;F)</i>	Df	F	<i>Pr(&gt;F)</i>	Df	F	<i>Pr(&gt;F)</i>
Covariate	1	385.2	<b>&lt;0.001</b>	1	176.2	<b>&lt;0.001</b>	1	179.2	<b>&lt;0.001</b>	1	3048	<b>&lt;0.001</b>
FDis	1	23.4	<b>&lt;0.001</b>	1	32.1	<b>&lt;0.001</b>	1	32.2	<b>&lt;0.001</b>	1	10.46	<b>0.001</b>
PS							3	2.843	<b>0.037</b>	3	3.697	<b>0.012</b>
FDis:PS							3	0.701	0.551	3	0.833	0.476
Residuals	1091			1091			1085			1085		

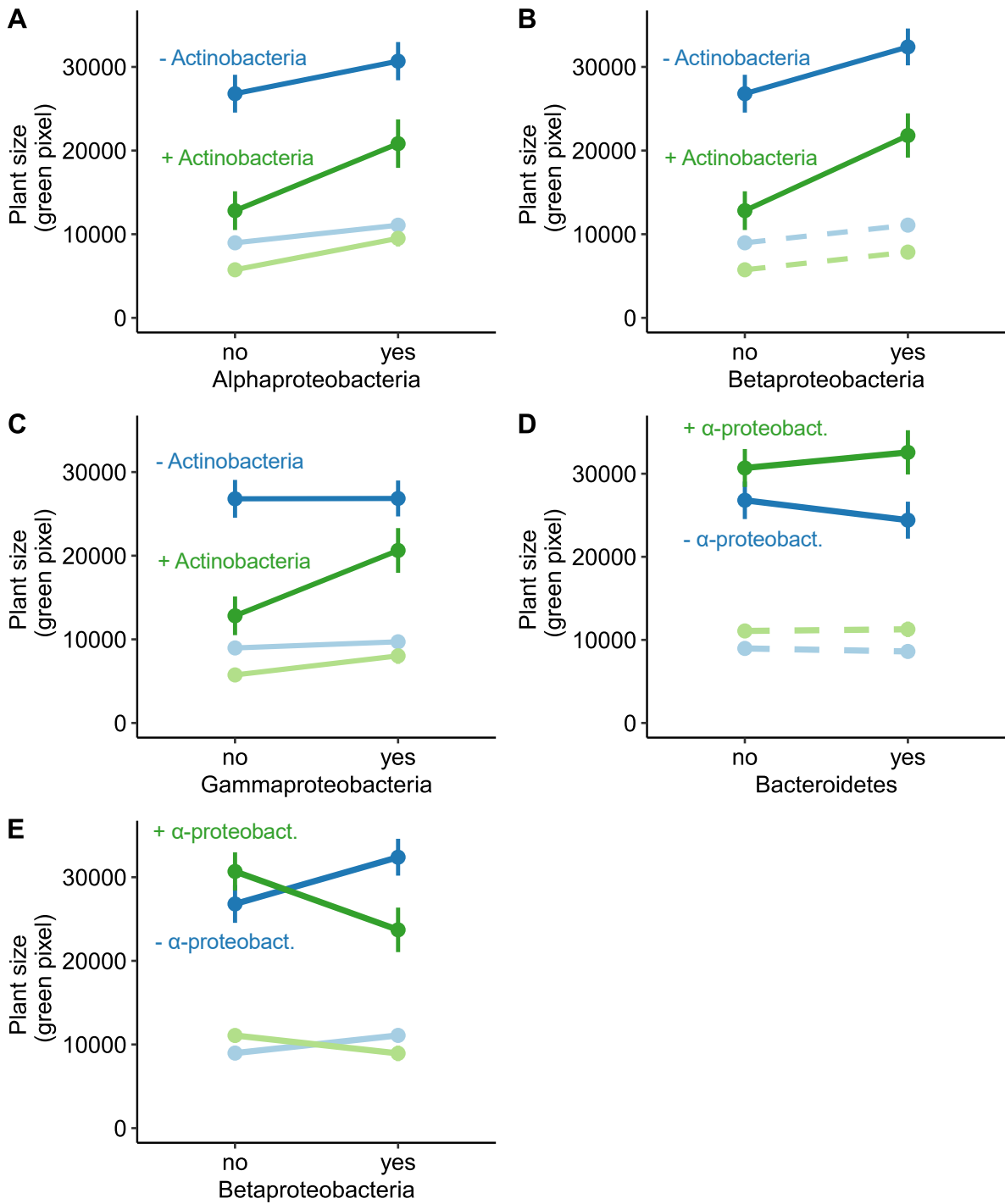
### *Clade-clade interaction can improve plant fitness*

First, stepwise regression reduced the 10 possible interaction terms to three and five interacting terms for growth analysis at 18-day-old plants (microbiome inoculation only) and 25 day-old plants (microbiome inoculation and *Pseudomonas viridiflava* infection) respectively (Table S4 and S5). The remaining interacting terms were Alphaproteobacteria:bacteroidetes and Actinobacteria:Gammaproteobacteria for both time points, and Alphaproteobacteria:Bacteroidetes and Actinobacteria:Betaproteobacteria only for 25 day-old plant growth analysis. For the clade-clade-pathogen interaction, only one term remained: Alphaproteobacteria:Bacteroidetes:pathogen stress (Table S4 and S5).

We found that plant growth was lower when only Actinobacteria was present, but when Actinobacteria was present alongside other clades, there was a positive interaction (Figure 5A-C). The presence of Alphaproteobacteria was positive on plant growth, had a positive interaction on plant growth when co-present with Actinobacteria or Bacteroidetes (Figure 5A and D), but had a negative effect on plant growth when co-present with Betaproteobacteria (Figure 5E). Finally, there was only one significant interaction term between clades and *Pseudomonas*, Alphaproteobacteria:Bacteroidetes:Pseudomonas (Table S4 and S5). We found that plants resisted *Pseudomonas* better when both clades were present together rather than individually (Table S6 and Figure S2).



**Figure 4.** Significant relationships between plant size and functional dispersion (FDIs). The analysis of microbiome inoculation (18 day-old plant; orange line) and microbiome-and-pathogen inoculation (25 day-old plant; red line) on growth (A) and the effect of *Pseudomonas viridiflava* infection on the plant resistance when two plant size covariates (11 and 18 days) are used (B). Solid trend lines represent model predictions with shaded areas representing 95% confidence intervals.



**Figure 5.** The effect of interaction between two clades on plant growth. Green and blue lines refer to the presence and absence of a clade respectively. Light green and blue indicate plant size at 18 days (inoculated with microbiome only), and dark green and blue indicate plant size at 25 days (microbiome inoculation and *Pseudomonas viridiflava* infection). The dashed lines refer to interaction terms that did not remain after model selection at 18 days but remained at 25 day selection (B and D). Predicted means and 95% confidence intervals (error bars) from generalized linear models (A and B) with their reaction norms are shown.

## Discussion

The role of plant biodiversity in ecosystem functioning is well studied, but the link between plant microbiome diversity and overall plant function is less well understood. An area especially lacking in research is how diversity in the leaf microbiome community contributes to host plant performance. Through our factorial experimental design, we inoculated *Arabidopsis thaliana* leaves with microbiomes of four increasing levels of bacterial diversity. We then measured how these levels of diversity affected host performance, determined by plant growth and pathogen resistance.

### *Leaf bacterial diversity improves plant performance*

We used a factorial design to create four total combinations of microbiome diversity, with two levels each of strain richness and clade diversity. By randomly selecting the clade and strains, we created 40 unique bacterial communities with a stepwise increase in diversity. After examining the interaction between clade diversity and species richness, we found that plants always benefited from an increase in the number of strains in a microbiome, and while having higher clade diversity was better for plant growth, it was not as important as strain richness. Previous research on bacterial strains in At-LSPHERE showed phylogenetically close strains had similar resource usage (Schäfer et al., 2022, 2023). Therefore, an increase in strain richness in our experiment could have resulted in more resilient bacterial communities that could compensate strain losses effectively, as seen in the functional redundancy and insurance hypotheses (Eisenhauer et al., 2023, 2024).

As functional redundancy and other BEF principles can be highly environmentally dependent (Fetzer et al., 2015; Ali, 2023), we added pathogen stress to confirm whether an increase in leaf microbial diversity led to plants having greater pathogen resistance. We found that, after *Pseudomonas* infection, plants inoculated with more diverse microbiomes performed as well as control plants, whereas those inoculated with less diverse microbiomes had lower sizes. Therefore we cautiously suggest that a microbiome with high diversity can confer a greater degree of pathogen resistance.

The *Pseudomonas* strains used in the experiment are found to infect *Arabidopsis* in nature (Karasov et al., 2018), and when infected onto a sterile plant they are capable of killing it (Karasov et al., 2018; Duque-Jaramillo et al., 2023). However, when these strains are present in a plant alongside other non-pathogenic bacteria, including other commensal *Pseudomonas* strains, the effect is reduced, with strains lowering biomass but not killing the plant (Ma et al., 2021; Shalev et al., 2022b, a). Furthermore, another experiment with the At-LSPHERE bacterial collection suggests that some of the At-LSPHERE strains can confer full protection against a related *Pseudomonas* pathogen (*Pseudomonas syringae* DC3000), and the

increase in leaf bacterial diversity conferred higher protection against *Pseudomonas* (Vogel et al., 2021). In this context, we hypothesize that in our experiment, the increase in clade diversity could have increased the probability of the microbiomes containing a strain conferring pathogen resistance. However, greater species richness in our experiment may not have captured a wide enough range of bacterial function, as bacterial function shows a level of phylogenetic conservatism (Schäfer et al., 2022, 2023).

#### *Functional diversity could better capture ecosystem functioning*

Despite biodiversity-ecosystem function theory being well established, the method of estimating ecosystem function has not been universally accepted (van der Plas, 2019), with both functional and phylogenetic diversity indices being used to measure biodiversity and explain ecosystem function (van der Plas, 2019). We converted the categorical diversity levels of microbiomes in our experiment into both predicted phylogenetic and functional diversities. Using the continuous indices, we confirmed that there were significant differences between the four groups, with microbiomes with higher strain richness and clade diversity corresponding to higher functional and phylogenetic diversity. We also found that there was a good correlation between functional diversity and phylogenetic diversity within microbiomes (Faith, 2018). However, we found that functional diversity was statistically better at explaining plant growth and disease resistance than phylogenetic diversity (Mazel et al., 2018; E-Vojtkó et al., 2023). This is in contrast to findings where phylogenetic diversity and functional diversity perform relatively similarly in statistical models (similar AIC and R<sup>2</sup>; Flynn et al., 2011; Hao et al., 2018; Le Bagousse-Pinguet et al., 2019).

Different resources were used to calculate each index, with phylogenetic diversity calculated using a relatively short 16S rRNA region, and functional diversity based on the predicted function of the whole genome. In such a scenario, the index that was more information-rich may have been better at capturing the overall BEF relationship, i.e. functional dispersion that captured the variance of species abundance in trait space (Daniel and Rooney, 2022).

#### *The effect of clade-clade interaction on plant performance*

Varying clade diversity levels allowed us to test clade-clade interactions. The bacterial clades used in this experiment have often been identified as core microbial taxa in both roots and leaves within various ecosystems, including in agricultural soil (Kembel et al., 2014; Acosta et al., 2020; Castellano-Hinojosa and Strauss, 2021; Ling et al., 2022). The prevalence of these taxa across different ecosystems and plant groups suggests a long-term, co-evolutionary relationship between plants and microbiomes, and their potential importance across varied environments.

We found that plant growth increased with the presence of Actinobacteria when in tandem with other clades, but when present alone, Actinobacteria strains did not have a positive effect on plant growth. However, previous research indicates that Actinobacteria strains have great potential to be biofertilizers, and can be used to increase plant fitness against pathogens (Bao et al., 2021; Boukhatem et al., 2022; Ebrahimi-Zarandi et al., 2022; Narsing Rao et al., 2022). We speculate that, when present with other phylogenetic groups, Actinobacteria could have an additive effect, while alone their benefits could be limited. Similar results have previously been found, with Actinobacteria having a potential additive effect with other Alphaproteobacteria on pathogen resistance in *Arabidopsis* (Emmenegger et al., 2023).

We also found an adverse additive effect, as plant growth was lower when Alphaproteobacteria and Betaproteobacteria occurred together in a microbiome compared to when either was present alone. This negative interaction may stem from the close phylogenetic relationship between the two groups, with similar resource utilization potentially leading to niche overlap. Increased competition could then result in a decrease in microbial productivity from antagonistic interactions, and lead to reduced plant performance (Mayfield and Levine, 2010; Romdhane et al., 2022; Schäfer et al., 2023; Schlechter et al., 2023). However, as we did not confirm the realized microbial community composition, we cannot be certain that this effect is due to niche overlap.

## **Conclusion**

In summary, we found that leaf-associated microbial communities follow a widely established BEF relationship in which an increase in microbiome diversity improves plant performance, measured through increased growth as well as improved resistance to *Pseudomonas* pathogens. We were also able to identify functional diversity as a more effective index to explain the BEF relationship as compared to phylogenetic diversity. This study helps us to understand the relationship between plants and their leaf microbial communities, and to an extent the interaction between clades in the microbiome. To the best of our knowledge, this is one of the first experiments in which leaf bacterial diversity has been deliberately increased in a full factorial design. Further studies could measure final microbiome composition to observe whether communities remain stable throughout experiments, and could focus on further increasing species richness experimentally, to more closely reflect natural systems. These strategies would improve our understanding of the functions of leaf microbiome diversity on plant performance, a crucial area of emerging research.

## Acknowledgements

Cecilia Heyworth for the grammar edits and proofreading. Deutsche Forschungsgemeinschaft (DFG) as part of the priority programme 2125 Deconstruction and Reconstruction of the Plant Microbiota, “DECrypT” (grant 401829393 to OB; PhD fellowship of JHJ). We thank Prof. Dr. Julia Vorholt for the At-LSPHERE bacterial isolates and we thank Prof. Dr. Detlef Weigel for *Pseudomonas viridiflava* strains. The authors acknowledge support by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG.

## References

- Acosta, K., J. Xu, S. Gilbert, E. Denison, T. Brinkman, S. Lebeis, and E. Lam. 2020. Duckweed hosts a taxonomically similar bacterial assemblage as the terrestrial leaf microbiome. *PLoS ONE* 15: e0228560.
- Ali, A. 2023. Biodiversity–ecosystem functioning research: Brief history, major trends and perspectives. *Biological Conservation* 285: 110210.
- Almario, J., M. Mahmoudi, S. Kroll, M. Agler, A. Placzek, A. Mari, and E. Kemen. 2022. The Leaf Microbiome of Arabidopsis Displays Reproducible Dynamics and Patterns throughout the Growing Season. *mBio* 13: e02825-21.
- Bai, Y., D. B. Müller, G. Srinivas, R. Garrido-Oter, E. Potthoff, M. Rott, N. Dombrowski, et al. 2015. Functional overlap of the Arabidopsis leaf and root microbiota. *Nature* 528: 364–369.
- Bao, Y., J. Dolfing, Z. Guo, R. Chen, M. Wu, Z. Li, X. Lin, and Y. Feng. 2021. Important ecophysiological roles of non-dominant Actinobacteria in plant residue decomposition, especially in less fertile soils. *Microbiome* 9: 84.
- Barnosky, A. D., N. Matzke, S. Tomiya, G. O. U. Wogan, B. Swartz, T. B. Quental, C. Marshall, et al. 2011. Has the Earth’s sixth mass extinction already arrived? *Nature* 471: 51–57.
- Boukhatem, Z. F., C. Merabet, and H. Tsaki. 2022. Plant Growth Promoting Actinobacteria, the Most Promising Candidates as Bioinoculants? *Frontiers in Agronomy* 4.
- Bradski, G. 2000. The OpenCV Library. *Dr. Dobb’s Journal of Software Tools*.
- Cadotte, M. W., B. J. Cardinale, and T. H. Oakley. 2008. Evolutionary history and the effect of biodiversity on plant productivity. *Proceedings of the National Academy of Sciences* 105: 17012–17017.
- Cadotte, M. W., K. Carscadden, and N. Mirotchnick. 2011. Beyond species: functional diversity and the maintenance of ecological processes and services. *Journal of Applied Ecology* 48: 1079–1087.
- Cantalapiedra, C. P., A. Hernández-Plaza, I. Letunic, P. Bork, and J. Huerta-Cepas. 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* 38: 5825–5829.
- Cardinale, B. J., J. E. Duffy, A. Gonzalez, D. U. Hooper, C. Perrings, P. Venail, A. Narwani, et al. 2012. Biodiversity loss and its impact on humanity. *Nature* 486: 59–67.
- Castellano-Hinojosa, A., and S. L. Strauss. 2021. Insights into the taxonomic and functional characterization of agricultural crop core rhizobiosomes and their potential microbial drivers. *Scientific Reports* 11: 10068.
- Chaudhry, V., P. Runge, P. Sengupta, G. Doehlemann, J. E. Parker, and E. Kemen. 2021. Shaping the leaf microbiota: plant–microbe–microbe interactions. *Journal of Experimental Botany* 72: 36–56.

- Daniel, J., and R. C. Rooney. 2022. Functional dispersion of wetland birds, invertebrates, and plants more strongly influenced by hydroperiod than each other. *Ecosphere* 13: e3971.
- Duque-Jaramillo, A., N. Ulmer, S. Alseekh, I. Bezrukov, A. R. Fernie, A. Skirycz, T. L. Karasov, and D. Weigel. 2023. The genetic and physiological basis of *Arabidopsis thaliana* tolerance to *Pseudomonas viridiflava*. *New Phytologist* 240: 1961–1975.
- Ebrahimi-Zarandi, M., R. Saberi Riseh, and M. T. Tarkka. 2022. Actinobacteria as Effective Biocontrol Agents against Plant Pathogens, an Overview on Their Role in Eliciting Plant Defense. *Microorganisms* 10: 1739.
- Edwards, J., C. Johnson, C. Santos-Medellín, E. Lurie, N. K. Podishetty, S. Bhatnagar, J. A. Eisen, and V. Sundaresan. 2015. Structure, variation, and assembly of the root-associated microbiomes of rice. *Proceedings of the National Academy of Sciences of the United States of America* 112: E911–920.
- Eisenhauer, N., J. Hines, F. T. Maestre, and M. C. Rillig. 2023. Reconsidering functional redundancy in biodiversity research. *npj Biodiversity* 2: 1–4.
- Eisenhauer, N., K. Mueller, A. Ebeling, G. Gleixner, Y. Huang, A.-M. Madaj, C. Roscher, et al. 2024. The multiple-mechanisms hypothesis of biodiversity–stability relationships. *Basic and Applied Ecology* 79: 153–166.
- Emmenegger, B., J. Massoni, C. M. Pestalozzi, M. Bortfeld-Miller, B. A. Maier, and J. A. Vorholt. 2023. Identifying microbiota community patterns important for plant protection using synthetic communities and machine learning. *Nature Communications* 14: 7983.
- E-Vojtkó, A., F. de Bello, Z. Lososová, and L. Götzenberger. 2023. Phylogenetic diversity is a weak proxy for functional diversity but they are complementary in explaining community assembly patterns in temperate vegetation. *Journal of Ecology* 111: 2218–2230.
- Fahlgren, N., M. Feldman, M. A. Gehan, M. S. Wilson, C. Shyu, D. W. Bryant, S. T. Hill, et al. 2015. A Versatile Phenotyping System and Analytics Platform Reveals Diverse Temporal Responses to Water Availability in *Setaria*. *Molecular Plant* 8: 1520–1535.
- Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation* 61: 1–10.
- Faith, D. P. 2018. Phylogenetic Diversity and Conservation Evaluation: Perspectives on Multiple Values, Indices, and Scales of Application. In R. A. Scherson, and D. P. Faith [eds.], *Phylogenetic Diversity: Applications and Challenges in Biodiversity Science*, 1–26. Springer International Publishing, Cham.
- Fetzer, I., K. Johst, R. Schäwe, T. Banitz, H. Harms, and A. Chatzinotas. 2015. The extent of functional redundancy changes as species' roles shift in different environments. *Proceedings of the National Academy of Sciences* 112: 14888–14893.
- Firincieli, A., M. Khorasani, A. C. Frank, and S. L. Doty. 2020. Influences of Climate on Phyllosphere Endophytic Bacterial Communities of Wild Poplar. *Frontiers in Plant Science* 11.
- Flynn, D. F. B., N. Mirotchnick, M. Jain, M. I. Palmer, and S. Naeem. 2011. Functional and phylogenetic diversity as predictors of biodiversity–ecosystem-function relationships. *Ecology* 92: 1573–1581.
- Fox, J., and S. Weisberg. 2019. *An R Companion to Applied Regression*. Third. Sage, Thousand Oaks CA.
- Gao, M., C. Xiong, C. Gao, C. K. M. Tsui, M.-M. Wang, X. Zhou, A.-M. Zhang, and L. Cai. 2021. Disease-induced changes in plant microbiome assembly and functional adaptation. *Microbiome* 9: 187.
- Grenié, M., and H. Gruson. 2023. fundiversity: a modular R package to compute functional diversity indices. *Ecography* 2023: e06585.
- Hao, M., C. Zhang, X. Zhao, and K. von Gadow. 2018. Functional and phylogenetic diversity determine woody productivity in a temperate forest. *Ecology and Evolution* 8: 2395–2406.
- Hoefle, D., M. Sommer, B. Wassermann, M. Faticov, D. Serra, G. Berg, A. J. M. Tack, and

- A. Abdelfattah. 2024. Oak seedling microbiome assembly under climate warming and drought. *Environmental Microbiome* 19: 62.
- Hooper, D. U., F. S. Chapin, J. J. Ewel, A. Hector, P. Inchausti, S. Lavorel, J. H. Lawton, et al. 2005. Effects of Biodiversity on Ecosystem Functioning: A Consensus of Current Knowledge. *Ecological Monographs* 75: 3–35.
- Huerta-Cepas, J., D. Szklarczyk, D. Heller, A. Hernández-Plaza, S. K. Forslund, H. Cook, D. R. Mende, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* 47: D309–D314.
- Hunter, J. D. 2007. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* 9: 90–95.
- Karasov, T. L., J. Almario, C. Friedemann, W. Ding, M. Giolai, D. Heavens, S. Kersten, et al. 2018. Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. *Cell Host & Microbe* 24: 168-179.e4.
- Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26: 1463–1464.
- Kembel, S. W., T. K. O'Connor, H. K. Arnold, S. P. Hubbell, S. J. Wright, and J. L. Green. 2014. Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. *Proceedings of the National Academy of Sciences of the United States of America* 111: 13715–13720.
- Laforest-Lapointe, I., A. Paquette, C. Messier, and S. W. Kembel. 2017. Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. *Nature* 546: 145–147.
- Le Bagousse-Pinguet, Y., S. Soliveres, N. Gross, R. Torices, M. Berdugo, and F. T. Maestre. 2019. Phylogenetic, functional, and taxonomic richness have both positive and negative effects on ecosystem multifunctionality. *Proceedings of the National Academy of Sciences* 116: 8419–8424.
- Ling, N., T. Wang, and Y. Kuzyakov. 2022. Rhizosphere bacteriome structure and functions. *Nature Communications* 13: 836.
- Loreau, M., S. Naeem, P. Inchausti, J. Bengtsson, J. P. Grime, A. Hector, D. U. Hooper, et al. 2001. Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges. *Science* 294: 804–808.
- Lüdtke, D. 2018. ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software* 3: 772.
- Ma, K.-W., Y. Niu, Y. Jia, J. Ordon, C. Copeland, A. Emonet, N. Geldner, et al. 2021. Coordination of microbe–host homeostasis by crosstalk with plant innate immunity. *Nature Plants* 7: 814–825.
- Mayfield, M. M., and J. M. Levine. 2010. Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters* 13: 1085–1093.
- Mazel, F., M. W. Pennell, M. W. Cadotte, S. Diaz, G. V. Dalla Riva, R. Grenyer, F. Leprieur, et al. 2018. Prioritizing phylogenetic diversity captures functional diversity unreliably. *Nature Communications* 9: 2888.
- Narsing Rao, M. P., K. Lohmaneeratana, C. Bunyoo, and A. Thamchaipenet. 2022. Actinobacteria–Plant Interactions in Alleviating Abiotic Stress. *Plants* 11: 2976.
- Northen, T. R., M. Kleiner, M. Torres, Á. T. Kovács, M. H. Nicolaisen, D. M. Krzyżanowska, S. Sharma, et al. 2024. Community standards and future opportunities for synthetic communities in plant–microbiota research. *Nature Microbiology* 9: 2774–2784.
- van der Plas, F. 2019. Biodiversity and ecosystem functioning in naturally assembled communities. *Biological Reviews* 94: 1220–1245.
- R Core Team. 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Robertson-Albertyn, S., J. C. Abbott, F. Concas, L. H. Brown, J. N. Orr, T. S. George, and D. Bulgarelli. 2022. Genome-Annotated Bacterial Collection of the Barley Rhizosphere Microbiota. *Microbiology Resource Announcements* 11: e01064-21.

- Romdhane, S., A. Spor, J. Aubert, D. Bru, M.-C. Breuil, S. Hallin, A. Mounier, et al. 2022. Unraveling negative biotic interactions determining soil microbial community assembly and functioning. *The ISME Journal* 16: 296–306.
- Schäfer, M., A. R. Pacheco, R. Künzler, M. Bortfeld-Miller, C. M. Field, E. Vayena, V. Hatzimanikatis, and J. A. Vorholt. 2023. Metabolic interaction models recapitulate leaf microbiota ecology. *Science* 381: eadf5121.
- Schäfer, M., C. M. Vogel, M. Bortfeld-Miller, M. Mittelviehhaus, and J. A. Vorholt. 2022. Mapping phyllosphere microbiota interactions in planta to establish genotype-phenotype relationships. *Nature microbiology* 7: 856–867.
- Schlechter, R. O., E. J. Kear, M. Bernach, D. M. Remus, and M. N. P. Remus-Emsermann. 2023. Metabolic resource overlap impacts competition among phyllosphere bacteria. *The ISME Journal* 17: 1445–1454.
- Schloerke, B., D. Cook, J. Larmarange, F. Briatte, M. Marbach, E. Thoen, A. Elberg, and J. Crowley. 2024. GGally: Extension to 'ggplot2'.
- Shalev, O., H. Ashkenazy, M. Neumann, and D. Weigel. 2022a. Commensal *Pseudomonas* protect *Arabidopsis thaliana* from a coexisting pathogen via multiple lineage-dependent mechanisms. *The ISME Journal* 16: 1235–1244.
- Shalev, O., T. L. Karasov, D. S. Lundberg, H. Ashkenazy, P. Pramoj Na Ayutthaya, and D. Weigel. 2022b. Commensal *Pseudomonas* strains facilitate protective response against pathogens in the host plant. *Nature Ecology & Evolution* 6: 383–396.
- Tilman, D., F. Isbell, and J. M. Cowles. 2014. Biodiversity and Ecosystem Functioning. *Annual Review of Ecology, Evolution, and Systematics* 45: 471–493.
- Venables, W. N., and B. D. Ripley. 2002. Modern Applied Statistics with S. Fourth. Springer, New York.
- Venail, P., K. Gross, T. H. Oakley, A. Narwani, E. Allan, P. Flombaum, F. Isbell, et al. 2015. Species richness, but not phylogenetic diversity, influences community biomass production and temporal stability in a re-examination of 16 grassland biodiversity studies. *Functional Ecology* 29: 615–626.
- Vogel, C. M., D. B. Potthoff, M. Schäfer, N. Barandun, and J. A. Vorholt. 2021. Protective role of the *Arabidopsis* leaf microbiota against a bacterial pathogen. *Nature Microbiology* 6: 1537–1548.
- Vorholt, J. A., C. Vogel, C. I. Carlström, and D. B. Müller. 2017. Establishing Causality: Opportunities of Synthetic Communities for Plant Microbiome Research. *Cell Host & Microbe* 22: 142–155.
- Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
- Wippel, K., K. Tao, Y. Niu, R. Zgadzaj, N. Kiel, R. Guan, E. Dahms, et al. 2021. Host preference and invasiveness of commensal bacteria in the *Lotus* and *Arabidopsis* root microbiota. *Nature Microbiology* 6: 1150–1162.
- Wu, T., E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, et al. 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* 2: 100141.
- Yu, G., L.-G. Wang, Y. Han, and Q.-Y. He. 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* 16: 284–287.

## Supplementary information

**Table S1.** Information of the strains used to create a syncom. Phylum, Family, and Genus information is from Pfeilmeier et al 2024. Clade used shows five experimental clades.

Strain_ID	Phylum	Class	Order	Family	Clade_used
Leaf3	Actinobacteria	Actinomycetia	Micrococcales	Sanguibacteraceae	Actinobacteria
Leaf69	Actinobacteria	Actinomycetia	Actinomycetales	Micrococcaceae	Actinobacteria
Leaf137	Actinobacteria	Actinomycetia	Actinomycetales	Micrococcaceae	Actinobacteria
Leaf141	Actinobacteria	Actinomycetia	Actinomycetales	Micrococcaceae	Actinobacteria
Leaf145	Actinobacteria	Actinomycetia	Actinomycetales	Micrococcaceae	Actinobacteria
Leaf151	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf161	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf164	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf183	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf185	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf203	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf222	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf225	Actinobacteria	Actinomycetia	Mycobacteriales	Nocardiaceae	Actinobacteria
Leaf233	Actinobacteria	Actinomycetia	Mycobacteriales	Nocardiaceae	Actinobacteria
Leaf234	Actinobacteria	Actinomycetia	Actinomycetales	Micrococcaceae	Actinobacteria
Leaf245	Actinobacteria	Actinomycetia	Propionibacteriales	Nocardioidaceae	Actinobacteria
Leaf254	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf261	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf263	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf264	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf285	Actinobacteria	Actinomycetia	Propionibacteriales	Nocardioidaceae	Actinobacteria
Leaf288	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf289	Actinobacteria	Actinomycetia	Propionibacteriales	Nocardioidaceae	Actinobacteria
Leaf299	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf304	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf314	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf334	Actinobacteria	Actinomycetia	Micrococcales	Cellulomonadaceae	Actinobacteria

Leaf335	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf336	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf337	Actinobacteria	Actinomycetia	Actinomycetales	Micrococcaceae	Actinobacteria
Leaf350	Actinobacteria	Actinomycetia	Propionibacteriales	Nocardiodaceae	Actinobacteria
Leaf351	Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Actinobacteria
Leaf354	Actinobacteria	Actinomycetia	Mycobacteriales	Nocardiaceae	Actinobacteria
Leaf369	Actinobacteria	Actinomycetia	Geodermatophilales	Geodermatophilaceae	Actinobacteria
Leaf380	Actinobacteria	Actinomycetia	Geodermatophilales	Geodermatophilaceae	Actinobacteria
Leaf446	Actinobacteria	Actinomycetia	Propionibacteriales	Nocardiodaceae	Actinobacteria
Leaf189	Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	Bacteroidetes
Leaf82	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Bacteroidetes
Leaf180	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Bacteroidetes
Leaf201	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Bacteroidetes
Leaf313	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Bacteroidetes
Leaf359	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Bacteroidetes
Leaf394	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Bacteroidetes
Leaf405	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Bacteroidetes
Leaf41	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Bacteroidetes
Leaf132	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Bacteroidetes
Leaf176	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Bacteroidetes
Leaf216	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Bacteroidetes
Leaf250	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Bacteroidetes
Leaf326	Deinococcus- Thermus	Deinococci	Deinococcales	Deinococcaceae	
Leaf13	Firmicutes	Bacilli	Bacillales	Bacillaceae	
Leaf49	Firmicutes	Bacilli	Bacillales	Bacillaceae	
Leaf72	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	
Leaf75	Firmicutes	Bacilli	Bacillales	Bacilli	
Leaf182	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	
Leaf187	Firmicutes	Bacilli	Bacillales	Bacillales	

Leaf406	Firmicutes	Bacilli	Bacillales	Bacillaceae	
Leaf2	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf4	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf5	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf10	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf17	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf21	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf26	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf28	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf30	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf32	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf33	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf34	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf42	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf62	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf64	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Hyphomicrobiaceae	Alphaproteobacteria
Leaf67	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf68	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Rhizobiaceae	Alphaproteobacteria
Leaf85	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Methylobacteriaceae	Alphaproteobacteria
Leaf86	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Methylobacteriaceae	Alphaproteobacteria
Leaf87	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Methylobacteriaceae	Alphaproteobacteria
Leaf88	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Methylobacteriaceae	Alphaproteobacteria
Leaf90	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Methylobacteriaceae	Alphaproteobacteria
Leaf91	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Methylobacteriaceae	Alphaproteobacteria
Leaf99	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Methylobacteriaceae	Alphaproteobacteria
Leaf106	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Methylobacteriaceae	Alphaproteobacteria
Leaf118	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Methylobacteriaceae	Alphaproteobacteria
Leaf122	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Methylobacteriaceae	Alphaproteobacteria
Leaf168	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Alphaproteobacteria
Leaf208	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf226	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria

Leaf231	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf242	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf257	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf262	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Rhizobiaceae	Alphaproteobacteria
Leaf280	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Alphaproteobacteria
Leaf306	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Rhizobiaceae	Alphaproteobacteria
Leaf311	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Rhizobiaceae	Alphaproteobacteria
Leaf324	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Aurantimonadaceae	Alphaproteobacteria
Leaf339	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf343	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf344	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Bradyrhizobiaceae	Alphaproteobacteria
Leaf357	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf361	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Methylobacteriaceae	Alphaproteobacteria
Leaf363	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Alphaproteobacteria
Leaf371	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Rhizobiaceae	Alphaproteobacteria
Leaf391	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Rhizobiaceae	Alphaproteobacteria
Leaf396	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Bradyrhizobiaceae	Alphaproteobacteria
Leaf412	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Alphaproteobacteria
Leaf420	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Hyphomicrobiaceae	Alphaproteobacteria
Leaf427	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Aurantimonadaceae	Alphaproteobacteria
Leaf443	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Aurantimonadaceae	Alphaproteobacteria
Leaf454	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Aurantimonadaceae	Alphaproteobacteria
Leaf456	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Methylobacteriaceae	Alphaproteobacteria
Leaf466	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Methylobacteriaceae	Alphaproteobacteria
Leaf61	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Betaproteobacteria
Leaf73	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Betaproteobacteria
Leaf76	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Betaproteobacteria
Leaf78	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Betaproteobacteria
Leaf126	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Betaproteobacteria
Leaf139	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Betaproteobacteria
Leaf160	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Betaproteobacteria

Leaf177	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Betaproteobacteria
Leaf220	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Betaproteobacteria
Leaf265	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Betaproteobacteria
Leaf267	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Betaproteobacteria
Leaf274	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Betaproteobacteria
Leaf400	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Betaproteobacteria
Leaf416	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	Betaproteobacteria
Leaf15	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Gammaproteobacteria
Leaf50	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	Gammaproteobacteria
Leaf51	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	Gammaproteobacteria
Leaf53	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	Gammaproteobacteria
Leaf58	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Gammaproteobacteria
Leaf59	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Gammaproteobacteria
Leaf70	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Gammaproteobacteria
Leaf83	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Gammaproteobacteria
Leaf127	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Gammaproteobacteria
Leaf129	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Gammaproteobacteria
Leaf130	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Gammaproteobacteria
Leaf131	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Gammaproteobacteria
Leaf434	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Gammaproteobacteria

---

**Table S2.** The composition of 40 syncoms. Numbers used in S1 to S9 refers to strains used according to Strain\_ID (Supplementary table 1). Actino- = Actinobacteria;  $\alpha$ - = Alphaproteobacteria; Bact- = Bacteroidetes;  $\beta$ - = Betaproteobacteria;  $\gamma$ - = Gammaproteobacteria

Syncom	CDxSR	Group1	Group2	Group3	S1	S2	S3	S4	S5	S6	S7	S8	S9
1	1x3	Actino-			446	145	137						
2	1x3	$\alpha$ -			343	85	42						
3	1x3	$\beta$ -			265	177	220						
4	1x3	$\gamma$ -			127	15	131						
5	1x3	Actino-			69	137	151						
6	1x3	$\alpha$ -			99	231	420						
7	1x3	$\beta$ -			220	265	73						
8	1x3	$\gamma$ -			58	127	129						
9	1x3	Bact-			405	313	201						
10	1x3	Bact-			176	359	41						
11	1x9	Actino-			288	351	304	336	161	446	263	222	234
12	1x9	$\alpha$ -			33	68	85	339	208	5	361	87	420
13	1x9	$\beta$ -			78	267	220	160	177	416	400	139	76
14	1x9	$\gamma$ -			59	127	15	131	434	129	70	58	83
15	1x9	Actino-			336	354	145	304	289	446	314	264	161
16	1x9	$\alpha$ -			262	68	90	371	99	21	91	5	122
17	1x9	$\beta$ -			267	78	265	73	160	177	416	76	220
18	1x9	$\gamma$ -			131	70	15	59	127	83	129	58	434
19	1x9	Bact-			180	313	189	176	201	216	405	132	359
20	1x9	Bact-			313	359	176	405	82	41	250	216	189
21	3x1	Actino-	Bact-	$\alpha$ -	233	313	62						

22	3x1	Actino-	Bact-	$\beta$ -	233	201	267							
23	3x1	Actino-	Bact-	$\gamma$ -	369	82	83							
24	3x1	Actino-	$\alpha$ -	$\beta$ -	334	42	177							
25	3x1	Actino-	$\alpha$ -	$\gamma$ -	288	91	131							
26	3x1	Actino-	$\beta$ -	$\gamma$ -	314	126	127							
27	3x1	Bact-	$\alpha$ -	$\beta$ -	405	64	139							
28	3x1	Bact-	$\beta$ -	$\gamma$ -	180	220	434							
29	3x1	Bact-	$\gamma$ -	$\alpha$ -	201	129	88							
30	3x1	$\alpha$ -	$\beta$ -	$\gamma$ -	106	76	131							
31	3x3	Actino-	Bact-	$\alpha$ -	254	233	225	201	216	359	361	396	17	
32	3x3	Actino-	Bact-	$\beta$ -	145	289	203	405	313	250	220	400	160	
33	3x3	Actino-	Bact-	$\gamma$ -	351	151	314	189	201	394	434	127	58	
34	3x3	Actino-	$\alpha$ -	$\beta$ -	151	183	263	339	5	122	220	139	267	
35	3x3	Actino-	$\alpha$ -	$\gamma$ -	233	304	351	62	344	106	83	434	15	
36	3x3	Actino-	$\beta$ -	$\gamma$ -	264	183	261	177	160	274	58	70	15	
37	3x3	Bact-	$\alpha$ -	$\beta$ -	201	189	313	466	443	17	267	265	416	
38	3x3	Bact-	$\beta$ -	$\gamma$ -	405	180	250	274	416	220	434	131	83	
39	3x3	Bact-	$\gamma$ -	$\alpha$ -	405	201	250	83	58	70	21	427	4	
40	3x3	$\alpha$ -	$\beta$ -	$\gamma$ -	4	26	242	177	76	160	131	15	59	

**Table S3.** Akaike's information criterion (AIC), Bayesian information criterion (BIC), and R<sup>2</sup> values of the six growth models to explain plant performance (growth and pathogen tolerance) with plant size of 18 or 25 day-old plant using the continuous diversity indices as fixed effect. The statistically most efficient indices based on the lowest AIC and BIC values and highest R<sup>2</sup> value is highlighted in bold. Df = degree of freedom, *F* = *F-value*

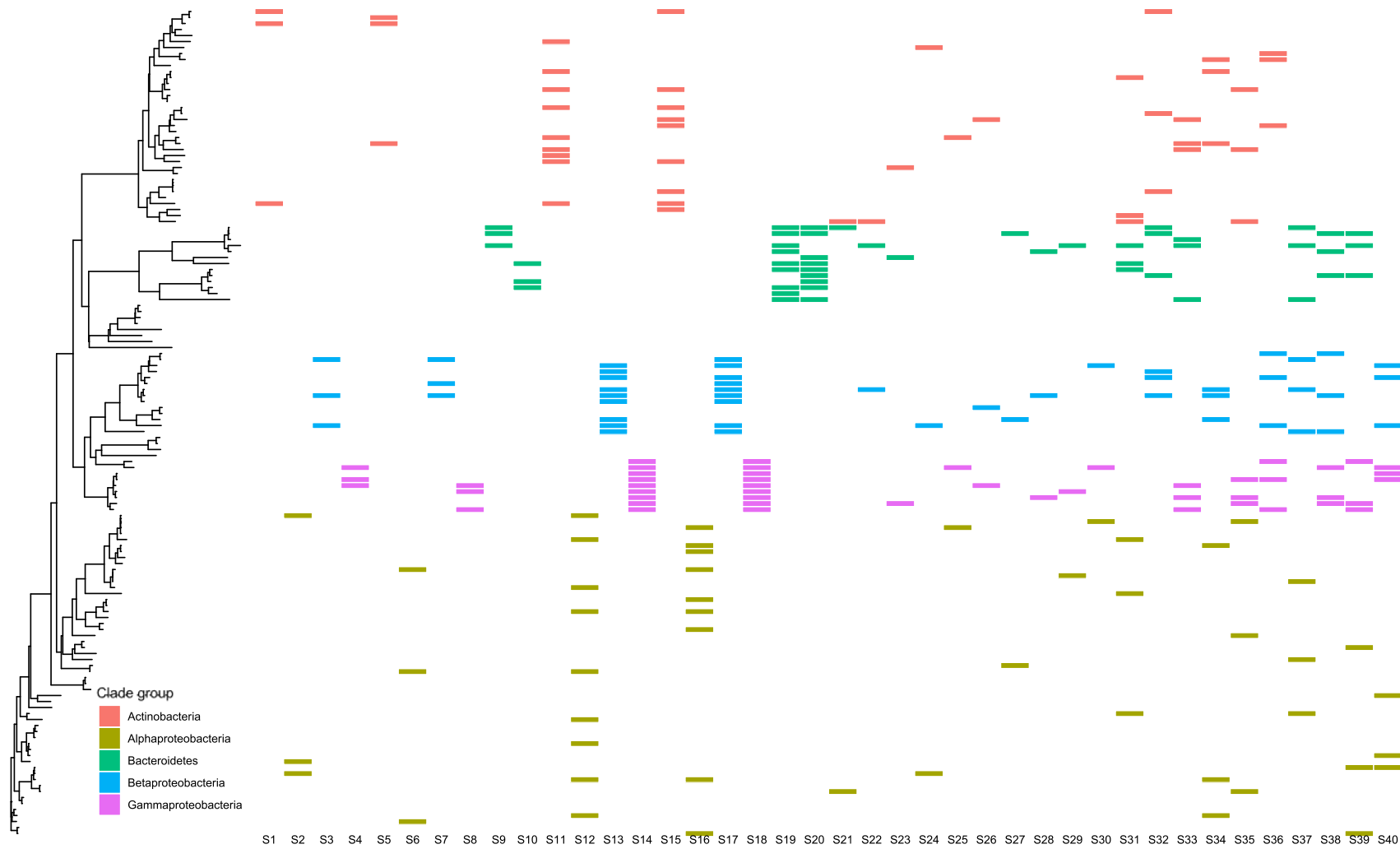
Covariate	11 day-old plant (pre-microbiome inoculation)			11 day-old plant			11 day-old plant			18 day-old plant		
	Name	AIC	BIC	R <sup>2</sup>	AIC	BIC	R <sup>2</sup>	AIC	BIC	R <sup>2</sup>	AIC	BIC
PD	21837	21857	0.234	24098	24133	0.128	24103	24153	0.129	22762	22812	0.744
FRic	21835	21855	0.236	24094	24129	0.131	24097	24147	0.134	22760	22810	0.744
FEve	21838	21858	0.234	24096	24131	0.129	24100	24150	0.131	22757	22807	0.745
FDiv	21838	21858	0.234	24098	24133	0.127	24103	24153	0.129	22764	22814	0.743
<b>FDis</b>	<b>21816</b>	<b>21836</b>	<b>0.249</b>	<b>24067</b>	<b>24102</b>	<b>0.152</b>	<b>24070</b>	<b>24120</b>	<b>0.154</b>	<b>22752</b>	<b>22802</b>	<b>0.746</b>
RaoQ	21818	21838	0.248	24070	24105	0.15	24074	24124	0.151	22754	22804	0.746

**Table S4.** Remaining terms and their interaction with other clades and pathogen strain (PS). The dash (-) represents the truncation of “bacteria” from the taxonomic groups.

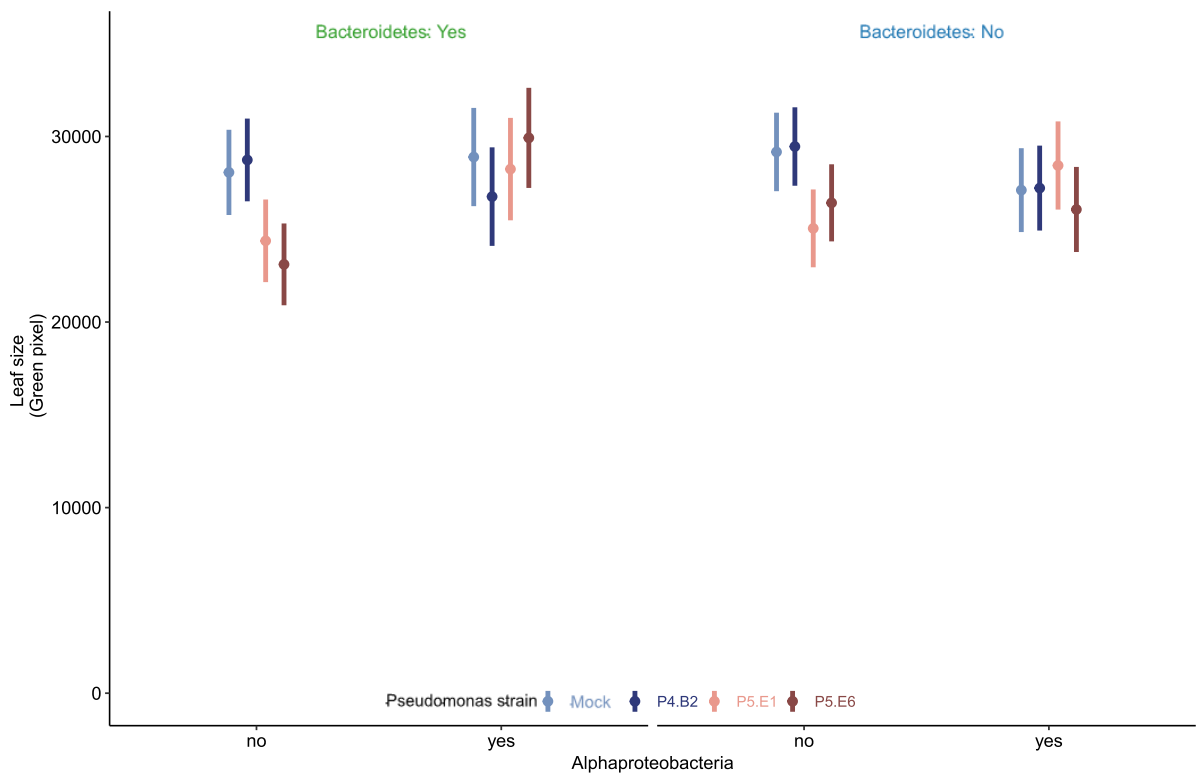
Test	Growth		Pathogen tolerance	
<b>Plant age</b>	18 day-old (microbiome only)	25 day-old (microbiome + pathogen)	25 day-old (microbiome + pathogen)	25 day-old (microbiome + pathogen)
<b>Covariate used</b>	11 day-old plant (pre- microbiome inoculation)	11 day-old plant	11 day-old plant	18 day-old
Test	Growth	Growth	Pathogen tolerance	Pathogen tolerance
<b>Remaining terms</b>	Actinobacteria (Actino -) + Alphaproteobacteria (α-) + Betaproteobacteria (β-) + Gammaproteobacteria (γ-)	Actinobacteria (Actino -)+ Alphaproteobacteria (α-) + Bacteroidetes + Betaproteobacteria (β-) + Gammaproteobacteria (γ-)	Actinobacteria (Actino -) + Alphaproteobacteria (α-) + Bacteroidetes Betaproteobacteria (β-) + Gammaproteobacteria (γ-) + PS	Actinobacteria (Actino -) + Alphaproteobacteria (α-) + + Bacteroidetes + Betaproteobacteria (β-) + Gammaproteobacteria (γ-) + PS
<b>Remaining interacting clades</b>	Actino- : α- + Actino- : γ- + α- : β-	Actino- : α-+ Actino- :β-+ Actino- :γ- + α- : Bacteroidetes + α-:β-	Actino-:α-+ Actino-:β-+ Actino-:γ- + α-:Bacteroidetes + α-:β-	Actino-:β- + Actino-:γ- + α-:Bacteroidetes + α-:β-  α-:PS + Bacteroidetes:PS + β-:PS + γ-:PS + α-:Bacteroidetes:PS + α-:β-:PS
<b>Clade interaction with pathogen stress (PS)</b>			α-:PS + β-:PS + α-:β-:PS	

**Table S5.** ANOVA table of generalized linear model of growth analysis explained by clades and their interaction using the remaining term after the model selection. Statistically significant term are highlighted in bold. Df = degree of freedom, *F* = *F*-value

Plant age	Growth						Pathogen tolerance		
	18 day-old (microbiome only)			25 day-old (microbiome + pathogen)			18 day-old plant		
	Df	F value	Pr(>F)	Df	F value	Pr(>F)	Df	F value	Pr(>F)
Covariate	11 day-old plant (pre-microbiome inoculation)						18 day-old plant		
Covariate	1	429.078	<b>&lt;0.001</b>	1	207.373	<b>&lt;0.001</b>	1	2849.492	<b>&lt;0.001</b>
Actinobacteria	1	45.029	<b>&lt;0.001</b>	1	97.723	<b>&lt;0.001</b>	1	64.854	<b>&lt;0.001</b>
Alphaproteobacteria	1	19.334	<b>&lt;0.001</b>	1	7.272	<b>0.007</b>	1	2.333	0.127
Bacteroidetes				1	0.899	0.343	1	0.861	0.354
Betaproteobacteria	1	2.588	0.108	1	6.586	<b>0.010</b>	1	5.473	<b>0.019</b>
Gammaproteobacteria	1	20.393	<b>&lt;0.001</b>	1	13.584	<b>&lt;0.001</b>	1	0.037	0.848
Pseudomonas_Strain (PS)							3	4.153	<b>0.006</b>
Actinobacteria:Alphaproteobacteria	1	7.212	<b>0.007</b>	1	6.032	<b>0.014</b>			
Actinobacteria:Betaproteobacteria				1	3.880	<b>0.049</b>	1	4.83	<b>0.028</b>
Actinobacteria:Gammaproteobacteria	1	5.228	<b>0.022</b>	1	19.071	<b>&lt;0.001</b>	1	17.091	<b>&lt;0.001</b>
Alphaproteobacteria:Bacteroidetes				1	5.798	<b>0.016</b>	1	7.28	<b>0.007</b>
Alphaproteobacteria:Betaproteobacteria	1	47.994	<b>&lt;0.001</b>	1	52.953	<b>&lt;0.001</b>	1	12.502	<b>&lt;0.001</b>
Alphaproteobacteria:PS							3	4.01	<b>0.007</b>
Bacteroidetes:PS							3	0.102	0.959
Betaproteobacteria:PS							3	0.63	0.596
Gammaproteobacteria:PS							3	2.043	0.106
Alphaproteobacteria:Bacteroidetes:PS							3	2.882	<b>0.035</b>
Alphaproteobacteria:Betaproteobacteria:PS							3	1.952	0.120
Residuals	1085			1082			1062		



**Figure S1.** The available pool of 138 strains used to assemble microbiomes in the experiment. A tile refers to a strain used in a microbiome assembly and their corresponding color indicates their clades. Microbiome S1-S10: Low clade diversity - poor strain richness; S11-S20: Low Clade diversity- rich strain richness; S21-30: High clade diversity - poor strain richness; S31-40: High clade diversity-rich strain richness



**Figure S2.** The effect of interaction between two clades on the pathogen tolerance. Generalized linear model included 18-day-old (inoculated with microbiome) plant size as covariate. Predicted means and 95% confidence intervals (error bars) from generalized linear models.

## Chapter 4

### **Temperature fluctuations alter the *Arabidopsis thaliana* leaf microbiome**

## **Temperature fluctuations alter the *Arabidopsis thaliana* leaf microbiome**

Jun Hee Jung<sup>1</sup>, Ana Montalvao<sup>2</sup>, Natalie Betz<sup>2</sup>, Madalin Parepa<sup>1</sup>, Detlef Weigel<sup>2</sup>, Oliver Bossdorf<sup>1</sup>

<sup>1</sup> Plant Evolutionary Ecology, Institute of Evolution and Ecology, University of Tübingen, Germany

<sup>2</sup> Department of Molecular Biology, Max Planck Institute for Biology Tübingen, 72076 Tübingen, Germany

## **Abstract**

Plants face constantly changing environmental conditions, and the temporal patterns of these fluctuations may affect not only plant growth, but also important interactions with closely associated microbes. Understanding the impacts of environmental fluctuations on plant microbiomes is particularly timely, as global climate change is altering temporal patterns of temperature and water availability at an ever-increasing rate, yet there is little research testing these effects. Here, we experimentally tested the effect of temperature fluctuations on the leaf bacteria of *Arabidopsis thaliana*. We exposed plants with initially identical bacterial communities to rapid, moderate, or slow fluctuations in temperature, and compared the fate of their leaf bacteria to control plants at constant temperature conditions through repeated 16S amplicon sequencing. All plants received the same total sum of temperature over the experiment. We showed that temperature fluctuations significantly impacted the diversity of leaf bacteria, with plants facing rapid temperature changes showing the strongest decrease in diversity. Temperature fluctuations also altered the species composition of leaf microbiomes, with differences increasing over time. Our results demonstrate that changes in the temporal variability of temperature alone, without a change in mean temperature, impacts the plant microbiome in multiple ways. As climate change makes temperature fluctuations increasingly likely, and given the important functions of many plant-associated microbes, future research should address the mechanisms underlying these effects, and test their generality across different plant species and environmental factors.

**Keywords:** climate change, environmental variability, heat stress, leaf bacteria, microbial diversity, microbiome

## Introduction

In their natural environments, most plants experience temporal variation in different environmental conditions, and at different temporal scales (Kaiser et al., 2018; Matsubara, 2018; Su-Zhou et al., 2024). This is especially relevant for plant leaves, which are constantly exposed to a plethora of stresses including regular fluctuations in light, temperature, and water (Chaves et al., 2002; Kaiser et al., 2018; Matsubara, 2018). Because of their sessile nature, plants have evolved an array of developmental and physiological mechanisms for adjusting to dynamic environments (Bradshaw, 1965; Arnold et al., 2022; Arenas et al., 2025). Despite this phenotypic plasticity, plants are increasingly challenged by global climate change, which not only affects mean environmental conditions but also alters the frequency and amplitude of environmental fluctuations, particularly water availability and temperature (Schär et al., 2004; Konapala et al., 2020; Olonscheck et al., 2021; Heeter et al., 2023). It is therefore important to understand the potential consequences of this intensifying environmental variability on both plants and their related microbiomes.

Environmental fluctuations can have a profound impact on plant physiological processes, from a specific physiological trait to overall plant performance. Under natural environmental fluctuations, physiological traits such as flower size can change compared to when plants are subjected to a constant temperature (Wiszniewski et al., 2022). Studies have also demonstrated how changes in environmental variability, independent from changes in average, can impact plant performance. For example, plants exposed to rapid changes in temperature from day to night showed lower biomass and lower chlorophyll amounts than those facing a more gradual change (Chiang et al., 2020). Similarly, in a water fluctuation experiment both invasive and non-invasive plants preferred stable conditions, with invasive plants preferring infrequent watering in large quantities, and native plants preferring a high frequency of watering in smaller quantities (Qin et al., 2024). In *Arabidopsis thaliana*, the timing of temperature fluctuations and genotype are the biggest influencers on plant performance across two generations (Scheepens et al., 2018a; Deng et al., 2021). In general, research has shown that factors relating to the timing and frequency of stress fluctuations can have a significant impact on plant performance, including when stress commences in the life stage of the plant, how long stress-recovery periods last, and the level of severity, or amplitude, of the stress.

While plants live in variable environments, they are also a habitat for many organisms which must tolerate these fluctuations alongside their hosts (Vorholt, 2012; Chaudhry et al., 2021; Almario et al., 2022). It is crucial to study the complex interactions between environment, plants, and their microbiomes to fully understand the impact of environmental changes. Examples can be seen in the literature, as when heat stress was found to influence beta

diversity and in particular the abundance of Actinobacteria from the root-associated microbiome of sorghum (Wipf et al., 2021).

A plant and its associated microbes have a delicate balance between proliferation and control, as many bacterial endophytes can shift from beneficial to harmful. For example, Pfeilmeier et al. (2021) showed in a knock-out *Arabidopsis* line how symbiotic microbes can become opportunistic pathogens, and further research has shown that naturally-occurring *Pseudomonas* lineages isolated from healthy *Arabidopsis* can be pathogenic in both sterile and non-sterile *Arabidopsis* (Karasov et al., 2018; Shalev et al., 2022b). When environmental stress is combined with pathogenic stress, it can also create unique scenarios for plants to respond to. Jose et al. (2024) found that a salicylic acid-producing *Arabidopsis thaliana* mutant had higher *Pseudomonas* pathogen tolerance and higher heat tolerance compared to the wild progenitor, but only when microbiome diversity was high. When the microbial diversity was lowered, heat stress caused dysbiosis (Jose et al., 2024), suggesting some level of interplay between genotype, hormone level, and environment in controlling microbial communities. Although our collective knowledge about the consequences of environmental stress on the plant microbiome is expanding, there is much yet to learn, especially about the time-scales in which the microbiome responds and the effect of temperature fluctuations on the microbiome over time.

There are two main considerations when designing environmental fluctuation experiments with a plant-microbiome system: ensuring the total sum of environmental factors received by the plants in different treatments at the end of the experiment is equal (see Parepa et al., 2013a; Scheepens et al., 2018a, b; Qin et al., 2024 for the experimental design; N.B. some of the experiments may contain treatment where different mean treatment is applied), and ensuring all plants receive the same initial microbiome. In the synthetic community (syncom) approach, a microbial culture collection can be used to assemble a microbiome, which can then be inoculated onto a plant (Vorholt et al., 2017; Northen et al., 2024). Using syncom ensures the initial microbial composition is as similar as possible, and conducting a serial microbiome harvest can enable observation of microbiome change over the course of the experiment.

Here, we designed an experiment to test the effect of varying frequency of temperature fluctuations on the composition and abundance of leaf-associated microbiomes in *Arabidopsis thaliana*. We supplemented the 16S rRNA metabarcoding with image analysis, allowing plant growth to be analyzed alongside changes in microbiome. We expected plant growth and bacterial diversity and composition to be negatively affected by heat stress and temperature fluctuation. The analysis of the plant and microbiome traits at the end of the experiment

determined: 1) whether fluctuation was stressful to the plant, 2) the effect of environmental fluctuation on bacterial diversity and composition, and 3) any change of microbiome composition over time.

## Methods

### *Plants and bacteria*

We worked with four genotypes of *Arabidopsis thaliana* - Col-0, Cvi-0, Sha and Oy-0, a subset of the Versailles “Core Collection 8” of natural accessions that has been created to maximize natural genetic diversity (McKhann et al., 2004), and that should thus provide a good idea of the breadth of responses in *A. thaliana*. We purchased the plant seeds from the Versailles *Arabidopsis* Stock Center. The bacterial community inoculum we used in our experiment consisted of 138 bacterial strains (from 30 families in 5 phyla) from the At-LSPHERE culture collection of *A. thaliana* leaf-associated bacteria (Table S1, <https://www.at-sphere.com/>, Bai et al., 2015). We used a subset of 138 strains from the full collection of 223 strains that maximized phylogenetic diversity while providing distinct enough 16S rRNA sequences (Pfeilmeier et al., 2021).

### *Experimental design*

To test for effects of temperature fluctuations on plant microbiomes, we exposed *Arabidopsis thaliana* plants with identical initial leaf bacterial communities (Table S1; see below for details) to four experimental treatments (Fig. 1a), and then followed the development of plants and bacteria through repeated photographing and image analyses, and repeated 16S amplicon sequencing. The four treatments were (1) a control treatment where the inoculated plants were grown at 24°C during the day and 21°C during the night, (2) a rapid fluctuations treatment where the conditions shifted daily between 33°C/30°C (day/night) and 15°C/12°C (day/night), (3) a moderate fluctuations treatment where the same temperature changes occurred at three-day intervals, and (4) a slow change treatment where a nine-day period at 33°C/30°C was followed by a nine-day period at 15°C/12°C (Fig. 1a). A particular quality of our experiment was that during the 18 days of temperature treatments, all four treatments accumulated the same temperature sums and thus had the same average temperatures. Our experiment thus varied only the temporal patterns of temperature conditions, not their mean. Because a day-night temperature cycle was wrongly set to 33°C/12°C on the first day of the rapid fluctuations treatment, we had to use these settings a second time on the last day of the temperature treatments, to maintain equal temperature sums in all treatments. Of each of four *Arabidopsis* genotypes, we grew a total of 50 replicate plants in each treatment, for a total of 200 plants in each treatment, and overall 800 plants in the experiment.

### *Plant cultivation and growth monitoring*

Before the start of the experiment, we surface-sterilized all seeds by storing them overnight at  $-80^{\circ}\text{C}$  followed by vapor-phase sterilizing with chlorine gas (Duque-Jaramillo et al., 2023). We sowed the seeds directly into 60 mm pots filled with a 2:1 mix of potting soil (Pikiererde, Einheitserde Werkverband e.V., Germany) and sand that had been autoclaved twice before, with at least 24 h between the two autoclavings. After three days of stratification at  $4^{\circ}\text{C}$ , the pots were placed into four plant growth chambers (Percival E-36L, Percival Scientific, USA), initially all with a day/night cycle of 12 h day at  $24^{\circ}\text{C}$  and 12 h night at  $21^{\circ}\text{C}$ . The light intensity during the day was set to 80% ( $440 \mu\text{mol m}^{-2} \text{s}^{-1}$ ). We randomly assigned 50 plants of each genotype to each chamber and then distributed the plants evenly across the two shelves in each chamber. We kept the plants under these conditions until they had germinated and developed their first true leaves. At day 14, we inoculated all plants with the bacterial community (see below), and at day 18 we started the temperature treatments by randomly selecting one climate chamber for each treatment. To track the plant growth and its response to the treatments, we took daily images of all plants, starting at day 13 (pre-inoculation) and estimated plant sizes as the numbers of plant-associated pixels counted with a Python 3.9 script using the libraries Pillow and OpenCV2 (script available on <https://github.com/junhee-jung>).

### *Bacterial inoculation*

To create the bacterial inoculum, we streaked out each of the 138 bacteria strains individually on R2A-agar plates (Himedia, India) with 0.05% MeOH. After four days incubation at room temperature, we suspended a 1  $\mu\text{l}$  sterilized loop of biological material over 500  $\mu\text{l}$  of R2A - MeOH liquid media for each strain, and then combined equal volumes of individual strain suspensions to create the 138-strain community inoculum. We added glycerin at a final concentration of 12.5% v/v, aliquoted the suspension into 1 ml tubes, and cooled it at  $4^{\circ}\text{C}$  for 24 h, then at  $-20^{\circ}\text{C}$  for 24 h, and finally stored all tubes at  $-80^{\circ}\text{C}$ . At day 14 of our experiment, we unfroze the inoculum, washed it twice using 10mM  $\text{MgCl}_2$  and diluted it to  $\text{OD}_{600}$  0.02 using  $\text{MgCl}_2$ . We then drop-inoculated all plants with 200  $\mu\text{l}$  of inoculum suspension.

### *Microbiome sequencing*

To understand the dynamics of the bacterial communities, and quantify their responses to the experimental treatments, we harvested subsets of the plants for 16S amplicon sequencing at nine time points: before inoculation on day 13 (to have an idea of the levels of spontaneous colonization), three days after inoculation on day 17 (to confirm the inoculation), and then at seven additional time points (days 18, 19, 20, 23, 26, 35 and 36 post-stratification, or days 1, 2, 3, 6, 9, 17, 18 of the treatment, see Fig. 1) during the treatment phase (Fig. 1A). At each

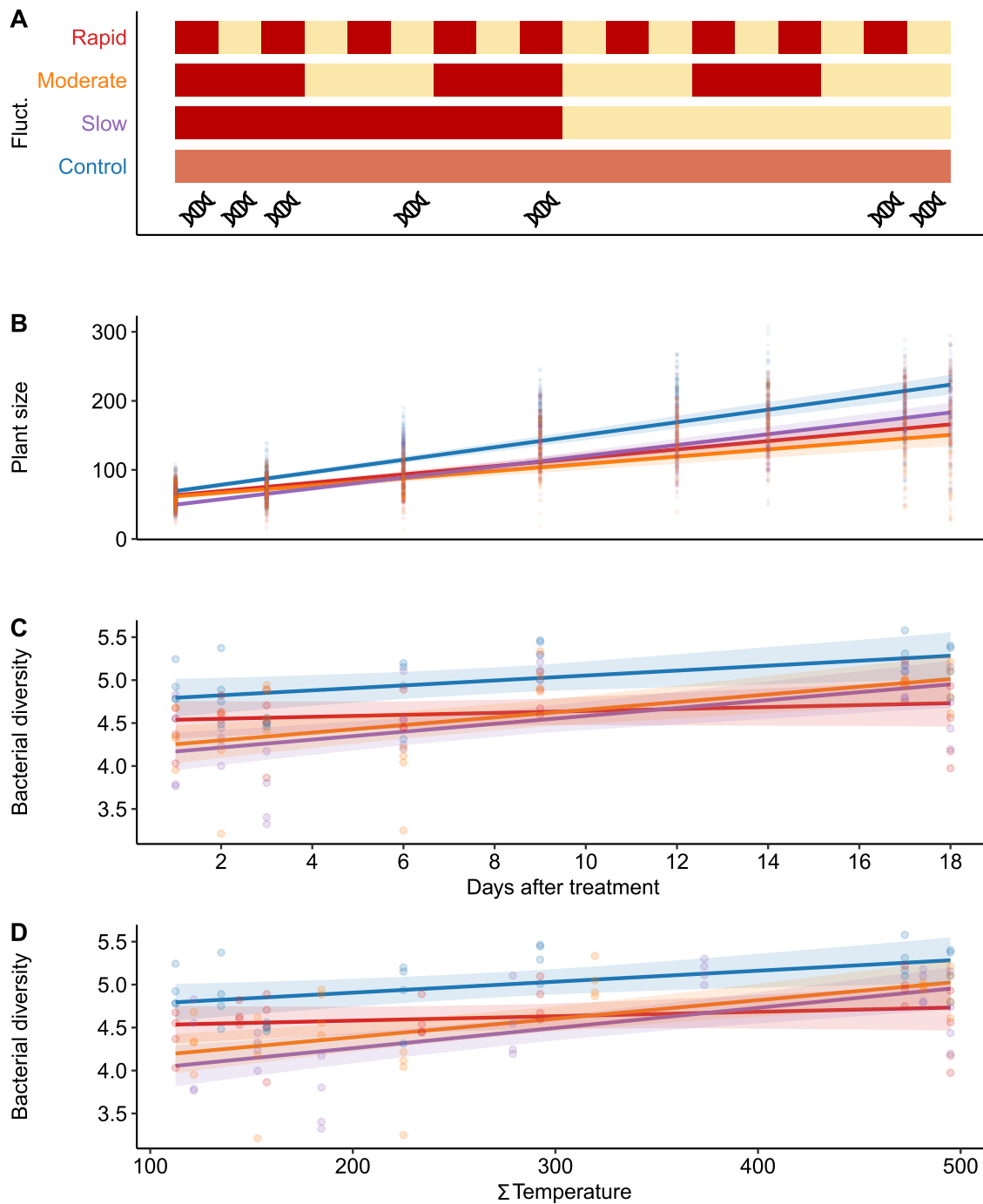
time point we harvested four replicates of each genotype per treatment, i.e. a total of 16 plants per chamber and 64 plants in total. We carefully cut the aboveground biomass of each plant and rinsed it first with sterile ddH<sub>2</sub>O for 30 s, then with 70% EtOH for 30 s, and again with sterile ddH<sub>2</sub>O for another 30 s. We then snap froze all samples in liquid nitrogen and stored them at -80°C until the DNA extraction. For the DNA extraction, we always pooled the four available plants of the same genotype per time point and treatment, since at the early harvesting time points the individual plant masses were too low for reliable DNA extraction. This reduced the total number of samples for amplicon sequencing to 144 pooled samples. We extracted all DNA using the FastDNA SPIN Kit for Soil (MP Biomedicals, Irvine, USA). As controls, we included DNA from the inoculum and a blank water sample.

For the library preparation, we used a hamPCR protocol (Lundberg et al., 2021) modified to use Nextera adapters. The first reaction contained 2.5 µL 10x Taq buffer (10mM), 1 µL dNTPs (10 mM), 0.5 µL 10 µM forward and reverse primers, 0.375 µL of mitochondria and chloroplast 16s rRNA gene blocking Peptide Nucleic Acid (PNAs; 50 µM; Lundberg et al., 2013), 5 µL of sample DNA or water (negative control), 0.2 µL Taq polymerase, and 14.55 µL PCR grade water in a total of 25 µL reactant. The primer used were combination of Nextera XT adaptors and 16S 515F-799R region primers (primer sequences for 515F: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAgtycagcmgcccgcggttaa and 799R: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGcmgggtatctaatacckgtt, Lundberg et al., 2021). The settings for the first reaction were: 2 minutes at 94°C for initial denaturation, followed by 29 cycles of 94°C for 30 seconds for denaturation, 78°C for 5 seconds for PNAs annealing, 55°C for 30 seconds for primer annealing, 72°C for 45 seconds for extension, and final 2 minute final extension at 72°C. We cleaned up the resulting PCR product through Solid Phase Reversible Immobilization (SPRI) bead clean-up (Rohland and Reich, 2012). The second PCR reaction had 40 µL volume consisting of 7 µL 5x Q5 buffer, 0.7 µL 10 µM dNTPs, 0.75 µL mitochondria and chloroplast PNAs mix, 4.78 µL PCR grade water, 0.35 µL Q5 polymerase, 0.875 µL P5 adaptor, 8.75 µL P7 adaptor, and 16.8 µL of PCR product from the first reaction. The setting for the second reaction were: 2 minutes at 94°C for initial denaturation, with 9 cycles of 94°C for 20 seconds for denaturation, 30 seconds at 65°C for primer annealing, 2 minutes at 72°C for extension, and a single 2 minute final extension at 72°C. We washed the final PCR products using SPRI bead clean-up, measured the DNA concentrations of the washed PCR products using PicoGreen (Invitrogen, Carlsbad, CA, USA), and pooled the samples at equimolar total DNA ratio. We further cleaned up the pooled samples using Bluepippin (Sage Science, Beverly, MA, USA), and sequenced them on an Illumina NextSeq 2000 system at the Max Planck Institute for Biology Tübingen Genome

Center. After the sequencing, we demultiplexed the raw data and removed the barcodes with the Illumina bcl2fastq conversion software.

### *Data analysis*

After importing the demultiplexed amplicon sequences to QIIME2 2023.05, we performed quality control (denoising) using DADA2 within the QIIME2 environment (Callahan et al., 2016; Bolyen et al., 2019), by trimming, informed by the sequence quality plots, three base pairs from forward reads and 60 from reverse reads, and removing chimeric reads. We then taxonomically classified the quality-controlled reads, based on the Silva 138 SSURef NR99 515F/806R region sequences, using the classify-sklearn function in QIIME 2 (Quast et al., 2013; Yilmaz et al., 2014; Li et al., 2021), and imported the resulting data into R version 4.3.2 with the RStudio environment (2024.04.0 Build 735), using the *file2meco* package (Liu et al., 2022). We discarded one sample with less than 5,000 reads from the subsequent analyses. Based on the numbers of OTUs in each sample and their relative abundances, we calculated the Shannon diversity (= alpha diversity) of the bacterial communities in all samples and then fitted generalized linear models with Gaussian distribution that tested for the effects of either time or temperature sums, the experimental treatment, and their interaction, on bacterial diversity. The difference in alpha diversity between the experiment treatments were calculated using *emmeans* package with Bonferroni correction (Lenth, 2025). To understand how the fluctuation treatments affected the composition (= beta diversity) of the bacterial communities, we conducted a PERMANOVA (permutational multivariate analysis of variance; Anderson, 2001) based on the Bray-Curtis distance matrix of the samples, that tested the effects of treatment, time, and their interaction on community dissimilarity, with 10,000 bootstrap permutations. We also visualized the dissimilarity of all samples, either grouped by treatment or harvest time point, through principal coordinate analysis (PCoA). All diversity analyses graphs were done with the *microeco* package in R (Liu et al., 2021), which wraps, among others, the *vegan* and *ggplot2* packages (Wickham, 2016; Oksanen et al., 2022). For the statistical analysis of the plant growth data, we used generalized least square models with a coAR1 structure in the *nlme* package (Pinheiro and Bates, 2000), with plant size (estimated by numbers of pixels) during days 18-36 (duration of temperature treatments) as the dependent variable, and plant genotype, treatment, harvesting time point, and their interactions, as fixed effects. Prior to the analysis, we square-root transformed the plant size data, because it resulted in a better-behaved distribution of residuals. Further comparison between different experimental treatments, time, and their interaction were made using *emmeans* package with Bonferroni corrections. Visualization of all of the figures were done using *ggplot2* package. The R codes for all analysis can be found at <https://github.com/junhee-jung>.



**Figure 1.** Experimental design and temporal trends of plant growth and bacterial diversity. (A) Illustration of the four experimental treatments, with dark red color for periods at 33°C/30°C (day/night) and yellow color for periods at 15°C/12°C. The DNA symbols at the bottom indicate the seven time points for microbiome sequencing, and the font colors on the y-axis serve as legend for the panels below. (B) Temporal trends in plant size (green pixels). (C) and (D) show changes in bacterial Shannon diversity over time or cumulative temperature sums. The trend lines in panels (B) - (D) are effect sizes and 95% confidence intervals from the statistical models.

## Results

We found that plant growth was significantly impacted by the fluctuating temperature conditions in our experiment (Table 1), with plants growing less well in all fluctuation

treatments (Fig. 1B;  $P < 0.001$  for all three contrasts between control and individual fluctuations treatments).

There was also a significant time by treatment interaction, indicating that temporal trends different between the treatments. Indeed, different slopes of the growth trends resulted in increasing growth deficits over time in the fluctuation treatments. The amplicon sequencing resulted in an average of over 60,000 DNA sequencing reads per pooled sample, of which on average 42% remained after filtering, denoising, merging and removal of chimerics, and >99.99% of these were assigned to bacteria (see Table S2 for more details). In all treatments, the Shannon diversity of bacterial communities increased over time (Fig. 1C; significant time effect in Table 2A), but under all fluctuating temperature conditions the bacterial diversity remained lower than in the control treatment ( $P < 0.01$  for all three contrasts between the control and individual fluctuation treatments), with the greatest final diversity difference between the control and the rapid fluctuations treatment (Fig. 1C). When we used the accumulated temperature sums instead of time points as explanatory variable (Table 2B), this model had greater explanatory power than the one with time points (AIC = 115.68 compared to AIC = 122.33), and also here the rapid fluctuations treatment differed from the other two fluctuations treatments: While the diversity differences between rapid fluctuations and control increased at higher temperature sums, the bacterial diversity of the slow and moderate fluctuations treatments rather “caught up” with the control over time.

The temperature fluctuations not only affected the diversity but also the composition of bacterial communities. Although the PCoA showed substantial community overlap between experimental treatments (Fig. 3A), there was still significant treatment-related clustering that explained 8.3% of the variance in species composition ( $F_{3,104} = 4.1$ ,  $P < 0.001$ ), with the greatest dissimilarity between bacterial communities in the rapid vs. slow fluctuations ( $P < 0.001$  for contrast in PERMANOVA) and slow versus control treatment ( $P = 0.003$  for contrast control vs. slow fluctuations, but only  $P = 0.050$  and  $P = 0.066$  for control vs. moderate and rapid fluctuations, respectively). As for species diversity, the species composition was also strongly influenced by harvesting time point ( $F_{1,104} = 26.6$ ,  $P < 0.001$ , 17.9% variance explained), with a gradual, succession-like change of species composition over time (Fig. 3B). The PERMANOVA also detected a significant time by treatment interaction ( $F_{3,104} = 1.89$ ,  $P = 0.005$ , 3.8% variance explained).

**Table 1.** Pseudo-ANOVA table of generalized least squares (GLS) model fit testing the effects of time point (days since inoculation), plant genotype, temperature treatment, and their interactions, on plant size. d.f. = degrees of freedom.

	d.f.	F-ratio	P-value
Time	1	2742	<0.001
Plant genotype	3	64	<0.001
Temperature treatment	3	55	<0.001
Time x Genotype	3	22	<0.001
Time x Treatment	3	90	<0.001
Genotype x Treatment	9	2	0.018
Time x Genotype x Treatment	9	2	0.029
Residual	3396		

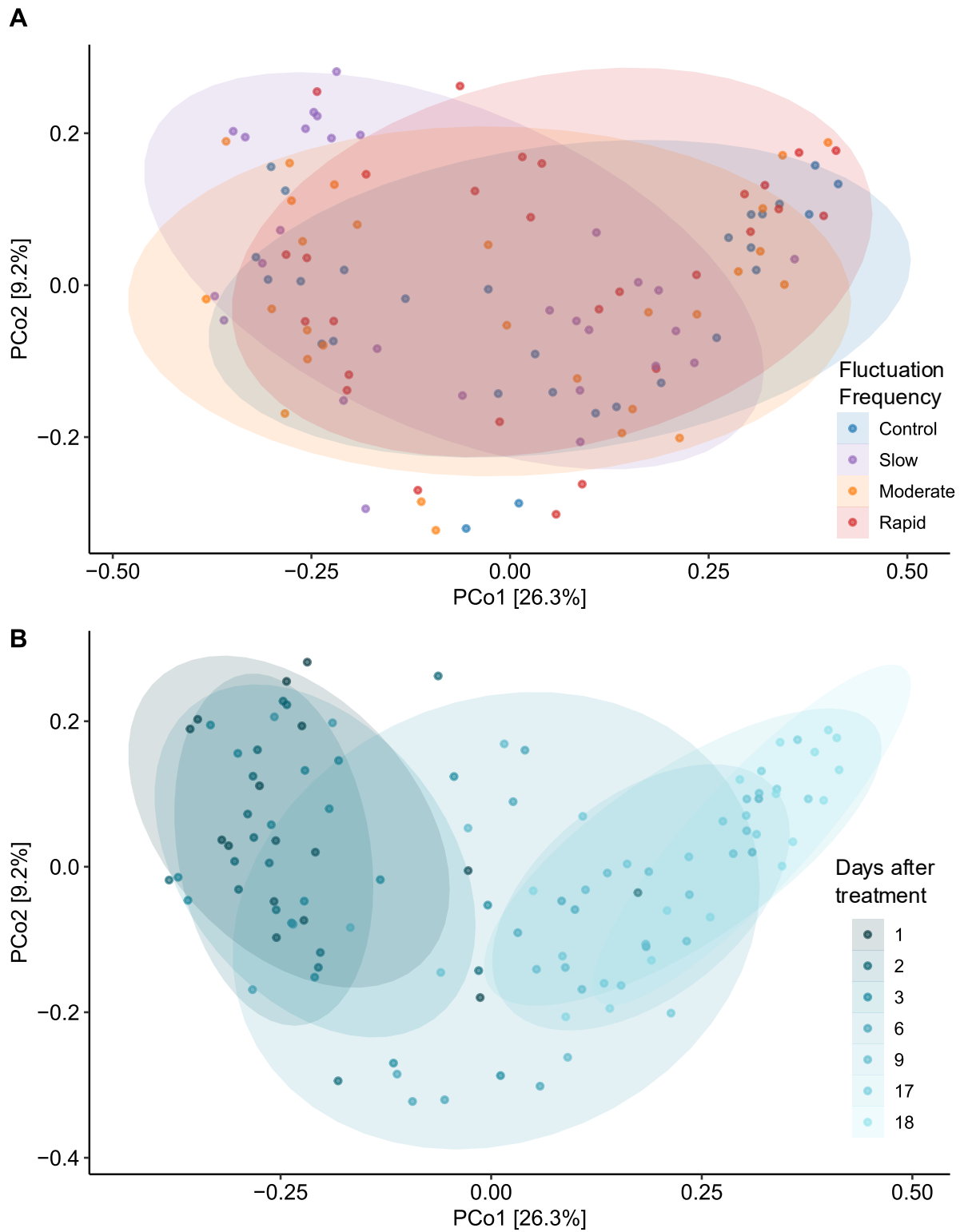
**Table 2.** ANOVA tables of generalized linear models testing the effects of either harvesting time (A) or accumulated temperature sums (B), as well as the temperature treatment, and their interactions on the estimated Shannon diversity of leaf bacterial communities in *A. thaliana*.

**A**

	d.f.	Deviance	F-ratio	P-value
Time	1	5.06	31.6	<b>&lt;0.001</b>
Treatment	3	4.25	8.84	<b>&lt;0.001</b>
Time x Treatment	3	0.933	1.94	0.127
Residual	104	16.4		

**B**

	d.f.	Deviance	F-ratio	P-value
Temperature sum	1	5.12	33.9	<b>&lt;0.001</b>
Treatment	3	4.82	10.7	<b>&lt;0.001</b>
Temperature sum x Treatment	3	1.25	2.76	<b>0.0458</b>
Residual	104	15.6		



**Figure 2.** Principal coordinate analysis (PCoA) based on Bray-Curtis dissimilarities of bacterial communities associated with the four temperature treatments (A) and with different harvesting time points (B).

## Discussion

Our study tested the effect of temperature fluctuation on plant growth, leaf microbiome bacterial diversity, and microbiome composition. We found that periods of heat stress affected plant growth, which was significantly lowered by fluctuations. Heat stress also initially caused a reduction in bacterial diversity in plants compared to those under constant temperature, but did not affect microbiome composition. At the end of the experiment, despite each plant receiving an identical temperature sum, plants exposed to fluctuating temperatures had both significantly lower bacterial diversity and altered microbiome composition, with the strongest effect in plants exposed to the rapid cycle of heat stress and recovery periods.

### *Plant growth*

During the experiment, we monitored plant growth non-destructively through pictures, which allowed us to measure treatment effects over time. Image analysis of plants provides a great proxy for plant fresh weight, and has been previously shown to accurately capture the effect of stress on plants (Shalev et al., 2022a; Duque-Jaramillo et al., 2023). Our analysis showed there was a significant decrease in growth in plants exposed to temperature fluctuations as compared to plants under a stable temperature regime, confirming that our treatment was stressful for the plant.

We observed significant variation in growth across plant genotypes, and we posit that these differences may reflect climates and their variability in the plants' original environments. Sha, Oy-0, and Col-0 performed better than Cvi, and Sha originated from environments with high temperature variability, thereby possibly giving Sha an advantage in fluctuating environments. The recovery temperature was similar to the mean annual temperature of the original habitats of Oy-0 and Col-0, therefore also potentially benefitting these genotypes in this experiment (Adams et al., 2016). For Cvi the recovery temperature was below the normal temperature in their original habitat, which could be why plants did not fully recover from heat stress periods (Scheepens et al., 2018b). Similar results have been found in other *Arabidopsis* fluctuation experiments where climate of origin was associated with varied plant response (Scheepens et al., 2018a, b; Deng et al., 2021).

Another possible explanation for genotypic differences in plant growth could be sugar production, as the levels of sugar molecules produced by plants after heat stress are genotype dependent (Zinta et al., 2018; Seydel et al., 2022). This could also have a downstream impact on the microbiome, as the plant leaf microbiome relies on sugar from the plants (Mercier and Lindow, 2000), and a change in the sugar source and quantity could alter microbiome composition and diversity.

### *Temporal changes in the microbiome*

We observed changes in bacterial composition and an increase in microbial diversity in leaf microbiomes across different time points in all temperature treatments. Plant age has a significant influence on microbiome composition, possibly due to bacterial turnover over time (Hannula et al., 2019), which could be the driver of the compositional change in our experiment.

As the experiment was not conducted in sterile conditions, it was inevitable that the leaves would be spontaneously colonized. In nature, sources of plant microbiomes include soil, seeds, water, and contact with other organisms (Vorholt, 2012; Rastogi et al., 2013). The leaf microbiome can also partially reflect the air microbiome, and microbial communities are formed through a combination of stochastic events with a priority effect (Maignien et al., 2014). In our experiment, the leaf microbiome would have been shaped mostly by the initial inoculum with the addition of spontaneous inoculation by environmental microorganisms (Maignien et al., 2014; Carlström et al., 2019).

In our study, the age of a leaf was positively correlated with leaf microbiome diversity, independent of leaf size (Wagner et al., 2016; Geyer et al., 2024; Lan et al., 2024; Li et al., 2024). It has been previously shown that the diversity of the core leaf microbiome, a group of microbial taxa shared across different environments (Risely, 2020), increases with age, potentially highlighting that older leaves can accommodate microbiomes with higher diversity (Li et al., 2024). Although bacteria are generally considered r-strategists in comparison to fungi (Lan et al., 2024), within a bacterial community slower-growing segments can be seen as K-strategists. In our experiment, it is possible that as the plants aged the portion of slow-growing bacteria increased, which could then have led to the observed increase in diversity and composition changes in the microbiome (Shi et al., 2024). As our analysis further showed that the plant microbiomes under stress from fluctuations were significantly less diverse compared to the control, this implies the microbiome community was not entirely shaped by spontaneous colonization and plant age, but also by our experimental temperature fluctuations.

### *Bacterial diversity*

From plant samples that had not experienced the recovery phase, we could evaluate the effect of heat stress on microbiome alpha diversity. Leaf microbiomes from plants subjected to heat stress had lower diversity, similar to results found by Li et al. (2023), in which heat stress negatively affected bacterial diversity in a meta-reanalysis of microbiome data. Leaf microbiome studies on *Galium album*, rice, and poplar have also shown this adverse effect (Ren et al., 2015; Aydogan et al., 2018; Firrincieli et al., 2020). Despite this, in the first heat phase, leaf microbial diversity in the slow and moderate fluctuation treatments increased after

the initial decrease. Interestingly, microbiome diversity in these groups also did not decrease again when a new heat phase started. This suggests that once the microbiome was acclimatized to heat stress it became resilient toward future temperature fluctuations (Li et al., 2023). In contrast, plants in the rapid fluctuation cycle were not able to significantly increase alpha-diversity over time, reflecting the similarly reduced performance that was found in their growth results. We speculate that the microbiome used in this experiment required at least 1 to 3 days of continuous stress to acclimatize to the fluctuation.

The decrease in diversity after the initial heat stress period suggests an interesting question: could a decrease in bacterial diversity have further knock-on effects on plant functionality? Increased microbiome diversity is associated with increased plant function in both soil and leaf-associated microbiota (Laforest-Lapointe et al., 2017; Delgado-Baquerizo et al., 2020), and leaf bacterial diversity is an important factor in plant protection against pathogens (Vogel et al., 2021; Emmenegger et al., 2023). Further evidence indicates that a decrease in soil bacterial diversity can hamper the proliferation of an invasive species, and therefore the productivity of certain plant species must be dependent upon soil microbiome diversity (Parepa et al., 2013b). Decreases in soil biodiversity have also been associated with loss of function and lower productivity within other plant communities (Hu et al., 2021; Wagg et al., 2021). With this strong evidence for microbiome diversity affecting plant performance, we hypothesize that the decrease in diversity in our experiment could have made the plants more susceptible to other stressors.

#### *Bacterial community composition*

In our experiment, the community composition of plant microbiomes showed significant change across different treatment groups, so communities showed compositional change after both time and temperature fluctuation. These results are similar to Fu et al. (2025)'s long-term experiment on plants facing increased temperatures, which has shown both functional and compositional change occurs in leaf microbiomes. One possible explanation is the "cry for help" hypothesis, in which plants under stress manipulate microbiome composition in order to recruit more beneficial microorganisms (Wang and Song, 2022). This has been found in some tomato genotypes, which selectively recruit beneficial bacteria when facing pathogen stress (Kwak et al., 2018). Alternatively, composition changes after heat stress periods could be explained by the replacement, loss, or gain of bacterial species to the leaf microbiome, and these factors have been found to contribute to compositional change in other plant systems exposed to heat stress (Hoefle et al., 2024). Finally, heat stress could exert selection pressure benefiting specific strains in a microbiome. As leaf ecosystems are oligotrophic environments, microbiomes are reliant upon metabolites produced by the plant. Heat stress alters metabolite

composition in plants (Rizhsky et al., 2004; Mishra et al., 2023) which could select for K-strategist bacterial species that prefer nutrient-poor environments (Shi et al., 2024).

### *Conclusions*

To the best of our knowledge, this is the first time the effect of temperature fluctuation has been tested on plants with deliberate microbial inoculations. We found that after the temperature fluctuation experiment, plants were not able to recover the level of bacterial diversity lost after heat stress, nor could they return to their original compositions. This effect was especially pronounced in the rapid-temperature fluctuation treatment, where we observed very little microbiome recovery in diversity. Our study shows the importance of serial time point harvesting as it better captures the change in the microbiome over time, especially regarding the decrease in microbiome diversity after initial heat stress. To improve our approach, future studies could investigate microbiome-plant systems under temperature fluctuations in which the molecular response of the plant or microbiome is captured. Additionally, due to pooling samples for DNA extraction, we could not identify the genotype effect on microbiome diversity and composition, a limitation which future studies could avoid. By incorporating microbiome-plant systems in environmental fluctuation studies, we can improve our understanding of plants and their microbiomes under increasing global climate change.

### **Acknowledgements**

Cecilia Heyworth for the grammar edits and proofreading. Frank Reis for help with inoculation procedure. Deutsche Forschungsgemeinschaft (DFG) as part of the priority programme 2125 Deconstruction and Reconstruction of the Plant Microbiota, “DECRyPT” (grant 401829393 to OB; PhD fellowship of JHJ). We thank Prof. Dr. Julia Vorholt for the At-LSPHERE bacterial isolates. The authors acknowledge support by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG.

### **References**

- Adams, W. W., J. J. Stewart, C. M. Cohu, O. Muller, and B. Demmig-Adams. 2016. Habitat Temperature and Precipitation of *Arabidopsis thaliana* Ecotypes Determine the Response of Foliar Vasculature, Photosynthesis, and Transpiration to Growth Temperature. *Frontiers in Plant Science* 7: 1026.
- Almario, J., M. Mahmoudi, S. Kroll, M. Agler, A. Placzek, A. Mari, and E. Kemen. 2022. The Leaf Microbiome of *Arabidopsis* Displays Reproducible Dynamics and Patterns

- throughout the Growing Season. *mBio* 13: e02825-21.
- Arenas, S., Y. Djabali, R. Rincent, P. Cubry, M.-L. Martin, M. Blein-Nicolas, L. Laplaze, et al. 2025. Modeling plant phenotypic plasticity and its underlying genetic architecture: a comparative study. *Journal of Experimental Botany*: e013.
- Arnold, P. A., S. Wang, A. A. Catling, L. E. B. Kruuk, and A. B. Nicotra. 2022. Patterns of phenotypic plasticity along a thermal gradient differ by trait type in an alpine plant. *Functional Ecology* 36: 2412–2428.
- Aydogan, E. L., G. Moser, C. Müller, P. Kämpfer, and S. P. Glaeser. 2018. Long-Term Warming Shifts the Composition of Bacterial Communities in the Phyllosphere of *Galium album* in a Permanent Grassland Field-Experiment. *Frontiers in Microbiology* 9.
- Bai, Y., D. B. Müller, G. Srinivas, R. Garrido-Oter, E. Potthoff, M. Rott, N. Dombrowski, et al. 2015. Functional overlap of the Arabidopsis leaf and root microbiota. *Nature* 528: 364–369.
- Bolyen, E., J. R. Rideout, M. R. Dillon, N. A. Bokulich, C. C. Abnet, G. A. Al-Ghalith, H. Alexander, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* 37: 852–857.
- Bradshaw, A. D. 1965. Evolutionary Significance of Phenotypic Plasticity in Plants. In E. W. Caspari, and J. M. Thoday [eds.], *Advances in Genetics*, 115–155. Academic Press.
- Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* 13: 581–583.
- Carlström, C. I., C. M. Field, M. Bortfeld-Miller, B. Müller, S. Sunagawa, and J. A. Vorholt. 2019. Synthetic microbiota reveal priority effects and keystone strains in the Arabidopsis phyllosphere. *Nature Ecology & Evolution* 3: 1445–1454.
- Chaudhry, V., P. Runge, P. Sengupta, G. Doehlemann, J. E. Parker, and E. Kemen. 2021. Shaping the leaf microbiota: plant–microbe–microbe interactions. *Journal of Experimental Botany* 72: 36–56.
- Chaves, M. M., J. S. Pereira, J. Maroco, M. L. Rodrigues, C. P. P. Ricardo, M. L. Osório, I. Carvalho, et al. 2002. How plants cope with water stress in the field. Photosynthesis and growth. *Annals of Botany* 89 Spec No: 907–916.
- Chiang, C., D. Bänkestad, and G. Hoch. 2020. Reaching Natural Growth: The Significance of Light and Temperature Fluctuations in Plant Performance in Indoor Growth Facilities. *Plants* 9: 1312.
- Delgado-Baquerizo, M., P. B. Reich, C. Trivedi, D. J. Eldridge, S. Abades, F. D. Alfaro, F. Bastida, et al. 2020. Multiple elements of soil biodiversity drive ecosystem functions across biomes. *Nature Ecology & Evolution* 4: 210–220.
- Deng, Y., O. Bossdorf, and J. F. Scheepens. 2021. Transgenerational effects of temperature fluctuations in Arabidopsis thaliana. *AoB PLANTS* 13: plab064.
- Duque-Jaramillo, A., N. Ulmer, S. Alseekh, I. Bezrukov, A. R. Fernie, A. Skirycz, T. L. Karasov, and D. Weigel. 2023. The genetic and physiological basis of Arabidopsis thaliana tolerance to Pseudomonas viridiflava. *New Phytologist* 240: 1961–1975.
- Emmenegger, B., J. Massoni, C. M. Pestalozzi, M. Bortfeld-Miller, B. A. Maier, and J. A. Vorholt. 2023. Identifying microbiota community patterns important for plant protection using synthetic communities and machine learning. *Nature Communications* 14: 7983.
- Firriacieli, A., M. Khorasani, A. C. Frank, and S. L. Doty. 2020. Influences of Climate on Phyllosphere Endophytic Bacterial Communities of Wild Poplar. *Frontiers in Plant Science* 11.
- Fu, G., G. Zhang, and H. Zhou. 2025. Effects of long-term experimental warming on phyllosphere epiphytic bacterial and fungal communities of four alpine plants. *Journal of Integrative Agriculture* 24: 799–814.
- Geyer, J. K., R. L. Grunberg, J. Wang, and C. E. Mitchell. 2024. Leaf age structures phyllosphere microbial communities in the field and greenhouse. *Frontiers in Microbiology* 15.

- Hannula, S. E., A. M. Kielak, K. Steinauer, M. Huberty, R. Jongen, J. R. De Long, R. Heinen, and T. M. Bezemer. 2019. Time after Time: Temporal Variation in the Effects of Grass and Forb Species on Soil Bacterial and Fungal Communities. *mBio* 10: 10.1128/mbio.02635-19.
- Heeter, K. J., G. L. Harley, J. T. Abatzoglou, K. J. Anchukaitis, E. R. Cook, B. L. Coulthard, L. A. Dye, and I. K. Homfeld. 2023. Unprecedented 21st century heat across the Pacific Northwest of North America. *npj Climate and Atmospheric Science* 6: 1–9.
- Hoefle, D., M. Sommer, B. Wassermann, M. Faticov, D. Serra, G. Berg, A. J. M. Tack, and A. Abdelfattah. 2024. Oak seedling microbiome assembly under climate warming and drought. *Environmental Microbiome* 19: 62.
- Hu, W., J. Ran, L. Dong, Q. Du, M. Ji, S. Yao, Y. Sun, et al. 2021. Aridity-driven shift in biodiversity–soil multifunctionality relationships. *Nature Communications* 12: 5350.
- Li, M. S. R., D. R. O'Rourke, B. D. Kaehler, M. Ziemski, M. R. Dillon, J. T. Foster, and N. A. Bokulich. 2021. RESCRIPt: Reproducible sequence taxonomy reference database management. *PLOS Computational Biology* 17: e1009581.
- Jose, J., E. Teutloff, S. Naseem, E. Barth, R. Halitschke, M. Marz, and M. T. Agler. 2024. Immunity and bacterial recruitment in plant leaves are parallel processes whose link shapes sensitivity to temperature stress. 2024.06.10.598336.
- Kaiser, E., A. Morales, and J. Harbinson. 2018. Fluctuating Light Takes Crop Photosynthesis on a Rollercoaster Ride. *Plant Physiology* 176: 977–989.
- Karasov, T. L., J. Almario, C. Friedemann, W. Ding, M. Giolai, D. Heavens, S. Kersten, et al. 2018. Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. *Cell Host & Microbe* 24: 168-179.e4.
- Konapala, G., A. K. Mishra, Y. Wada, and M. E. Mann. 2020. Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation. *Nature Communications* 11: 3044.
- Kwak, M.-J., H. G. Kong, K. Choi, S.-K. Kwon, J. Y. Song, J. Lee, P. A. Lee, et al. 2018. Rhizosphere microbiome structure alters to enable wilt resistance in tomato. *Nature Biotechnology* 36: 1100–1109.
- Laforest-Lapointe, I., A. Paquette, C. Messier, and S. W. Kembel. 2017. Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. *Nature* 546: 145–147.
- Lan, G., Y. Wei, X. Zhang, Z. Wu, K. Ji, H. Xu, B. Chen, and F. He. 2024. Assembly and maintenance of phyllosphere microbial diversity during rubber tree leaf senescence. *Communications Biology* 7: 1–12.
- Lenth, R. V. 2025. emmeans: Estimated Marginal Means, aka Least-Squares Means.
- Li, J., K. A. Bates, K. L. Hoang, T. E. Hector, S. C. L. Knowles, and K. C. King. 2023. Experimental temperatures shape host microbiome diversity and composition. *Global Change Biology* 29: 41–56.
- Li, J., M.-K. Jin, L. Huang, Z.-F. Liu, T. Wang, R.-Y. Chang, M. Op de Beeck, et al. 2024. Assembly and succession of the phyllosphere microbiome and nutrient-cycling genes during plant community development in a glacier foreland. *Environment International* 187: 108688.
- Liu, C., Y. Cui, X. Li, and M. Yao. 2021. microeco: an R package for data mining in microbial community ecology. *FEMS Microbiology Ecology* 97: fiae255.
- Liu, C., X. Li, F. R. P. Mansoldo, J. An, Y. Kou, X. Zhang, J. Wang, et al. 2022. Microbial habitat specificity largely affects microbial co-occurrence patterns and functional profiles in wetland soils. *Geoderma* 418: 115866.
- Lundberg, D. S., P. Pramoj Na Ayutthaya, A. Strauß, G. Shirsekar, W.-S. Lo, T. Lahaye, and D. Weigel. 2021. Host-associated microbe PCR (hamPCR) enables convenient measurement of both microbial load and community composition. M. C. Schuman, R. Bart, M. Qi, and B. E. Wolfe [eds.], *eLife* 10: e66186.
- Lundberg, D. S., S. Yourstone, P. Mieczkowski, C. D. Jones, and J. L. Dangl. 2013. Practical innovations for high-throughput amplicon sequencing. *Nature Methods* 10: 999–1002.

- Maignien, L., E. A. DeForce, M. E. Chafee, A. M. Eren, and S. L. Simmons. 2014. Ecological Succession and Stochastic Variation in the Assembly of *Arabidopsis thaliana* Phyllosphere Communities. *mBio* 5: 10.1128/mbio.00682-13.
- Matsubara, S. 2018. Growing plants in fluctuating environments: why bother? *Journal of Experimental Botany* 69: 4651–4654.
- McKhann, H. I., C. Camilleri, A. Bérard, T. Bataillon, J. L. David, X. Reboud, V. Le Corre, et al. 2004. Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. *The Plant Journal* 38: 193–202.
- Mercier, J., and S. E. Lindow. 2000. Role of Leaf Surface Sugars in Colonization of Plants by Bacterial Epiphytes. *Applied and Environmental Microbiology* 66: 369–374.
- Mishra, S., K. Spaccarotella, J. Gido, I. Samanta, and G. Chowdhary. 2023. Effects of Heat Stress on Plant-Nutrient Relations: An Update on Nutrient Uptake, Transport, and Assimilation. *International Journal of Molecular Sciences* 24: 15670.
- Northen, T. R., M. Kleiner, M. Torres, Á. T. Kovács, M. H. Nicolaisen, D. M. Krzyżanowska, S. Sharma, et al. 2024. Community standards and future opportunities for synthetic communities in plant–microbiota research. *Nature Microbiology* 9: 2774–2784.
- Oksanen, J., G. L. Simpson, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O’Hara, et al. 2022. vegan: Community Ecology Package.
- Olonscheck, D., A. P. Schurer, L. Lücke, and G. C. Hegerl. 2021. Large-scale emergence of regional changes in year-to-year temperature variability by the end of the 21st century. *Nature Communications* 12: 7237.
- Parepa, M., M. Fischer, and O. Bossdorf. 2013a. Environmental variability promotes plant invasion. *Nature Communications* 4: 1604.
- Parepa, M., U. Schaffner, and O. Bossdorf. 2013b. Help from under ground: soil biota facilitate knotweed invasion. *Ecosphere* 4: art31.
- Pfeilmeier, S., G. C. Petti, M. Bortfeld-Miller, B. Daniel, C. M. Field, S. Sunagawa, and J. A. Vorholt. 2021. The plant NADPH oxidase RBOHD is required for microbiota homeostasis in leaves. *Nature Microbiology* 6: 852–864.
- Pinheiro, J. C., and D. M. Bates eds. . 2000. Theory and Computational Methods for Linear Mixed-Effects Models. *Mixed-Effects Models in S and S-PLUS*, 57–96. Springer, New York, NY.
- Qin, W., Y. Sun, H. Müller-Schärer, and W. Huang. 2024. Responses of non-native and native plant species to fluctuations of water availability in a greenhouse experiment. *Ecology and Evolution* 14: e11692.
- Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41: D590–D596.
- Rastogi, G., G. L. Coaker, and J. H. J. Leveau. 2013. New insights into the structure and function of phyllosphere microbiota through high-throughput molecular approaches. *FEMS Microbiology Letters* 348: 1–10.
- Ren, G., C. Zhu, M. S. Alam, T. Tokida, H. Sakai, H. Nakamura, Y. Usui, et al. 2015. Response of soil, leaf endosphere and phyllosphere bacterial communities to elevated CO<sub>2</sub> and soil temperature in a rice paddy. *Plant and Soil* 392: 27–44.
- Risely, A. 2020. Applying the core microbiome to understand host–microbe systems. *Journal of Animal Ecology* 89: 1549–1558.
- Rizhsky, L., H. Liang, J. Shuman, V. Shulaev, S. Davletova, and R. Mittler. 2004. When Defense Pathways Collide. The Response of *Arabidopsis* to a Combination of Drought and Heat Stress. *Plant Physiology* 134: 1683–1696.
- Rohland, N., and D. Reich. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research* 22: 939–946.
- Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller. 2004. The role of increasing temperature variability in European summer heatwaves. *Nature* 427: 332–336.
- Scheepens, J. F., Y. Deng, and O. Bossdorf. 2018a. Phenotypic plasticity in response to temperature fluctuations is genetically variable, and relates to climatic variability of

- origin, in *Arabidopsis thaliana*. *AoB PLANTS* 10: ply043.
- Scheepens, J. F., R. Rauschkolb, R. Ziegler, V. Schroth, and O. Bossdorf. 2018b. Genotypic diversity and environmental variability affect the invasibility of experimental plant populations. *Oikos* 127: 570–578.
- Seydel, C., J. Biener, V. Brodsky, S. Eberlein, and T. Nägele. 2022. Predicting plant growth response under fluctuating temperature by carbon balance modelling. *Communications Biology* 5: 1–13.
- Shalev, O., H. Ashkenazy, M. Neumann, and D. Weigel. 2022a. Commensal *Pseudomonas* protect *Arabidopsis thaliana* from a coexisting pathogen via multiple lineage-dependent mechanisms. *The ISME Journal* 16: 1235–1244.
- Shalev, O., T. L. Karasov, D. S. Lundberg, H. Ashkenazy, P. Pramoj Na Ayutthaya, and D. Weigel. 2022b. Commensal *Pseudomonas* strains facilitate protective response against pathogens in the host plant. *Nature Ecology & Evolution* 6: 383–396.
- Shi, X., Y. Zhao, M. Xu, L. Ma, J. M. Adams, and Y. Shi. 2024. Insights into plant–microbe interactions in the rhizosphere to promote sustainable agriculture in the new crops era. *New Crops* 1: 100004.
- Su-Zhou, C., M. Durand, P. J. Aphalo, J. Martinez-Abaigar, A. Shapiguzov, H. Ishihara, X. Liu, and T. M. Robson. 2024. Weaker photosynthetic acclimation to fluctuating than to corresponding steady UVB radiation treatments in grapevines. *Physiologia Plantarum* 176: e14383.
- Vogel, C. M., D. B. Potthoff, M. Schäfer, N. Barandun, and J. A. Vorholt. 2021. Protective role of the *Arabidopsis* leaf microbiota against a bacterial pathogen. *Nature Microbiology* 6: 1537–1548.
- Vorholt, J. A. 2012. Microbial life in the phyllosphere. *Nature Reviews Microbiology* 10: 828–840.
- Vorholt, J. A., C. Vogel, C. I. Carlström, and D. B. Müller. 2017. Establishing Causality: Opportunities of Synthetic Communities for Plant Microbiome Research. *Cell Host & Microbe* 22: 142–155.
- Wagg, C., Y. Hautier, S. Pellkofer, S. Banerjee, B. Schmid, and M. G. van der Heijden. 2021. Diversity and asynchrony in soil microbial communities stabilizes ecosystem functioning. D. Weigel, and A. Mori [eds.], *eLife* 10: e62813.
- Wagner, M. R., D. S. Lundberg, T. G. del Rio, S. G. Tringe, J. L. Dangl, and T. Mitchell-Olds. 2016. Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nature Communications* 7: 12151.
- Wang, Z., and Y. Song. 2022. Toward understanding the genetic bases underlying plant-mediated “cry for help” to the microbiota. *iMeta* 1: e8.
- Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wipf, H. M.-L., T.-N. Bui, and D. Coleman-Derr. 2021. Distinguishing Between the Impacts of Heat and Drought Stress on the Root Microbiome of *Sorghum bicolor*. *Phytobiomes Journal* 5: 166–176.
- Wiszniewski, A., E. Uberegui, M. Messer, G. Sultanova, M. Borghi, G. T. Duarte, R. Vicente, et al. 2022. Temperature-mediated flower size plasticity in *Arabidopsis*. *iScience* 25: 105411.
- Yilmaz, P., L. W. Parfrey, P. Yarza, J. Gerken, E. Priesse, C. Quast, T. Schweer, et al. 2014. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Research* 42: D643–D648.
- Zinta, G., H. AbdElgawad, D. Peshev, J. T. Weedon, W. Van den Ende, I. Nijs, I. A. Janssens, et al. 2018. Dynamics of metabolic responses to periods of combined heat and drought in *Arabidopsis thaliana* under ambient and elevated atmospheric CO<sub>2</sub>. *Journal of Experimental Botany* 69: 2159–2170.

## Supplementary Information

**Table S1.** Information of the strains used to inoculate the plants. Phylum, Family, and Genus information from Pfeilmeier et al 2024.

Strain_Leaf_ID	Phylum	Family	Genus
Leaf2	Proteobacteria	Sphingomonadaceae	Novosphingobium
Leaf3	Actinobacteria	Sanguibacteraceae	Sanguibacter
Leaf4	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf5	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf10	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf13	Firmicutes	Bacillaceae	Bacillus
Leaf15	Proteobacteria	Pseudomonadaceae	Pseudomonas
Leaf17	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf21	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf26	Proteobacteria	Sphingomonadaceae	Sphingobium
Leaf28	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf30	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf32	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf33	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf34	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf41	Bacteroidetes	Sphingobacteriaceae	Pedobacter
Leaf42	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf49	Firmicutes	Bacillaceae	Bacillus
Leaf50	Proteobacteria	Enterobacteriaceae	Serratia
Leaf51	Proteobacteria	Enterobacteriaceae	Serratia
Leaf53	Proteobacteria	Enterobacteriaceae	Erwinia
Leaf58	Proteobacteria	Pseudomonadaceae	Pseudomonas
Leaf59	Proteobacteria	Pseudomonadaceae	Pseudomonas
Leaf61	Proteobacteria	Oxalobacteraceae	Duganella
Leaf62	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf64	Proteobacteria	Hyphomicrobiaceae	Devosia
Leaf67	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf68	Proteobacteria	Rhizobiaceae	Rhizobium
Leaf69	Actinobacteria	Micrococcaceae	Arthrobacter

Leaf70	Proteobacteria	Xanthomonadaceae	Stenotrophomonas
Leaf72	Firmicutes	Paenibacillaceae	Paenibacillus
Leaf73	Proteobacteria	Comamonadaceae	Acidovorax
Leaf75	Firmicutes	Bacilli	Bacillus
Leaf76	Proteobacteria	Comamonadaceae	Acidovorax
Leaf78	Proteobacteria	Comamonadaceae	Acidovorax
Leaf82	Bacteroidetes	Flavobacteriaceae	Flavobacterium
Leaf83	Proteobacteria	Pseudomonadaceae	Pseudomonas
Leaf85	Proteobacteria	Methylobacteriaceae	Methylobacterium
Leaf86	Proteobacteria	Methylobacteriaceae	Methylobacterium
Leaf87	Proteobacteria	Methylobacteriaceae	Methylobacterium
Leaf88	Proteobacteria	Methylobacteriaceae	Methylobacterium
Leaf90	Proteobacteria	Methylobacteriaceae	Methylobacterium
Leaf91	Proteobacteria	Methylobacteriaceae	Methylobacterium
Leaf99	Proteobacteria	Methylobacteriaceae	Methylobacterium
Leaf106	Proteobacteria	Methylobacteriaceae	Methylobacterium
Leaf118	Proteobacteria	Methylobacteriaceae	Methylobacterium
Leaf122	Proteobacteria	Methylobacteriaceae	Methylobacterium
Leaf126	Proteobacteria	Oxalobacteraceae	Duganella
Leaf127	Proteobacteria	Pseudomonadaceae	Pseudomonas
Leaf129	Proteobacteria	Pseudomonadaceae	Pseudomonas
Leaf130	Proteobacteria	Moraxellaceae	Acinetobacter
Leaf131	Proteobacteria	Xanthomonadaceae	Xanthomonas
Leaf132	Bacteroidetes	Sphingobacteriaceae	Pedobacter
Leaf137	Actinobacteria	Micrococcaceae	Arthrobacter
Leaf139	Proteobacteria	Oxalobacteraceae	Massilia
Leaf141	Actinobacteria	Micrococcaceae	Arthrobacter
Leaf145	Actinobacteria	Micrococcaceae	Arthrobacter
Leaf151	Actinobacteria	Microbacteriaceae	Microbacterium
Leaf160	Proteobacteria	Comamonadaceae	Acidovorax
Leaf161	Actinobacteria	Microbacteriaceae	Microbacterium
Leaf164	Actinobacteria	Microbacteriaceae	Rathayibacter

Leaf168	Proteobacteria	Caulobacteraceae	Brevundimonas
Leaf176	Bacteroidetes	Sphingobacteriaceae	Pedobacter
Leaf177	Proteobacteria	Burkholderiaceae	Burkholderia
Leaf180	Bacteroidetes	Flavobacteriaceae	Chryseobacterium
Leaf182	Firmicutes	Paenibacillaceae	Brevibacillus
Leaf183	Actinobacteria	Microbacteriaceae	Curtobacterium
Leaf185	Actinobacteria	Microbacteriaceae	Rathayibacter
Leaf187	Firmicutes	Bacillales	Exiguobacterium
Leaf189	Bacteroidetes	Cytophagaceae	Dyadobacter
Leaf201	Bacteroidetes	Flavobacteriaceae	Chryseobacterium
Leaf203	Actinobacteria	Microbacteriaceae	Microbacterium
Leaf208	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf216	Bacteroidetes	Sphingobacteriaceae	Pedobacter
Leaf220	Proteobacteria	Comamonadaceae	Variovorax
Leaf222	Actinobacteria	Microbacteriaceae	Agromyces
Leaf225	Actinobacteria	Nocardiaceae	Rhodococcus
Leaf226	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf231	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf233	Actinobacteria	Nocardiaceae	Rhodococcus
Leaf234	Actinobacteria	Micrococcaceae	Arthrobacter
Leaf242	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf250	Bacteroidetes	Sphingobacteriaceae	Pedobacter
Leaf254	Actinobacteria	Microbacteriaceae	Frigoribacterium
Leaf257	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf261	Actinobacteria	Microbacteriaceae	Curtobacterium
Leaf262	Proteobacteria	Rhizobiaceae	Rhizobium
Leaf263	Actinobacteria	Microbacteriaceae	Clavibacter
Leaf264	Actinobacteria	Microbacteriaceae	Leifsonia
Leaf265	Proteobacteria	Comamonadaceae	Pseudorhodoferax
Leaf267	Proteobacteria	Comamonadaceae	Variovorax
Leaf272	Actinobacteria	Nocardoidaceae	Aeromicrobium
Leaf274	Proteobacteria	Comamonadaceae	Pseudorhodoferax

Leaf280	Proteobacteria	Caulobacteraceae	Brevundimonas
Leaf285	Actinobacteria	Nocardioidaceae	Nocardioides
Leaf288	Actinobacteria	Microbacteriaceae	Microbacterium
Leaf289	Actinobacteria	Nocardioidaceae	Aeromicrobium
Leaf299	Actinobacteria	Microbacteriaceae	Rathayibacter
Leaf304	Actinobacteria	Microbacteriaceae	Fron dih abitans
Leaf306	Proteobacteria	Rhizobiaceae	Rhizobium
Leaf311	Proteobacteria	Rhizobiaceae	Rhizobium
Leaf313	Bacteroidetes	Flavobacteriaceae	Chryseobacterium
Leaf314	Actinobacteria	Microbacteriaceae	Plantibacter
Leaf324	Proteobacteria	Aurantimonadaceae	Aureimonas
Leaf326	Deinococcus-Thermus	Deinococcaceae	Deinococcus
Leaf334	Actinobacteria	Cellulomonadaceae	Cellulomonas
Leaf335	Actinobacteria	Microbacteriaceae	Agreia
Leaf336	Actinobacteria	Microbacteriaceae	Leifsonia
Leaf337	Actinobacteria	Micrococcaceae	Arthrobacter
Leaf339	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf343	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf344	Proteobacteria	Bradyrhizobiaceae	Bosea
Leaf350	Actinobacteria	Nocardioidaceae	Aeromicrobium
Leaf351	Actinobacteria	Microbacteriaceae	Microbacterium
Leaf354	Actinobacteria	Nocardiaceae	Williamsia
Leaf357	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf359	Bacteroidetes	Flavobacteriaceae	Flavobacterium
Leaf361	Proteobacteria	Methylobacteriaceae	Methylobacterium
Leaf363	Proteobacteria	Caulobacteraceae	Brevundimonas
Leaf369	Actinobacteria	Geodermatophilaceae	Geodermatophilus
Leaf371	Proteobacteria	Rhizobiaceae	Rhizobium
Leaf380	Actinobacteria	Geodermatophilaceae	Blastococcus
Leaf391	Proteobacteria	Rhizobiaceae	Rhizobium
Leaf394	Bacteroidetes	Flavobacteriaceae	Chryseobacterium
Leaf396	Proteobacteria	Bradyrhizobiaceae	Bradyrhizobium

Leaf400	Proteobacteria	Comamonadaceae	Acidovorax
Leaf405	Bacteroidetes	Flavobacteriaceae	Chryseobacterium
Leaf406	Firmicutes	Bacillaceae	Bacillus
Leaf412	Proteobacteria	Sphingomonadaceae	Sphingomonas
Leaf416	Proteobacteria	Methylophilaceae	Methylophilus
Leaf420	Proteobacteria	Hyphomicrobiaceae	Devosia
Leaf427	Proteobacteria	Aurantimonadaceae	Aurantimonas
Leaf434	Proteobacteria	Pseudomonadaceae	Pseudomonas
Leaf443	Proteobacteria	Aurantimonadaceae	Aurantimonas
Leaf446	Actinobacteria	Nocardioidaceae	Marmoricola
Leaf454	Proteobacteria	Aurantimonadaceae	Aurantimonas
Leaf456	Proteobacteria	Methylobacteriaceae	Methylobacterium
Leaf466	Proteobacteria	Methylobacteriaceae	Methylobacterium

**Table S2.** Summary statistics for the 16S amplicon sequencing. A) Average numbers of reads remaining a different steps of data-trimming for pooled plant samples or blanks. B) Frequencies of taxonomic assignments among the final set of reads.

<b>A</b>	Raw input	Filtered	Denosed	Merged	Non-chimeric
Samples	61484	48965	47784	41851	25699
Blanks	1526	1097	918	594	457

<b>B</b>	Bacteria	Plantae	Unassigned	Mitochondria	Archaea	Eukaryota	Chloroplast
Samples	25263	<1	<1	<1	<1	<1	<1
Blanks	454	0	0	<4	0	0	0

## Chapter 5

### **Discarded sequencing reads uncover natural variation in pest resistance in *Thlaspi arvense***

# Discarded sequencing reads uncover natural variation in pest resistance in *Thlaspi arvense*

Dario Galanti<sup>1,2\*</sup>, Jun Hee Jung<sup>1</sup>, Caroline Müller<sup>3</sup>, Oliver Bossdorf<sup>1\*</sup>

<sup>1</sup>Plant Evolutionary Ecology, Institute of Evolution and Ecology, University of Tübingen, Tübingen, Germany; <sup>2</sup>Royal Botanic Gardens, Kew, Richmond upon Thames, United Kingdom; <sup>3</sup>Chemical Ecology, Bielefeld University, Bielefeld, Germany

## eLife Assessment

This **important** study presents a significant methodological advance by leveraging previously discarded, unmapped DNA sequence reads to estimate pest infestation loads across plant accessions, and map variation in these apparent pest loads to defense genes. The bioinformatics approach is **compelling**, and the results should bear broad implications for phenotype-genotype prediction, especially regarding the use of unmapped reads for GWAS.

### \*For correspondence:

dario.galanti@uni-tuebingen.de (DG);  
oliver.bossdorf@uni-tuebingen.de (OB)

**Competing interest:** The authors declare that no competing interests exist.

**Funding:** See page 17

**Sent for Review**  
07 February 2024

**Preprint posted**  
04 March 2024

**Reviewed preprint posted**  
03 April 2024

**Reviewed preprint revised**  
29 October 2024

**Version of Record published**  
19 December 2024

**Reviewing Editor:** Sergio Rasmann, University of Neuchâtel, Switzerland

© Copyright Galanti et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

**Abstract** Understanding the genomic basis of natural variation in plant pest resistance is an important goal in plant science, but it usually requires large and labor-intensive phenotyping experiments. Here, we explored the possibility that non-target reads from plant DNA sequencing can serve as phenotyping proxies for addressing such questions. We used data from a whole-genome and -epigenome sequencing study of 207 natural lines of field pennycress (*Thlaspi arvense*) that were grown in a common environment and spontaneously colonized by aphids, mildew, and other microbes. We found that the numbers of non-target reads assigned to the pest species differed between populations, had significant SNP-based heritability, and were associated with climate of origin and baseline glucosinolate contents. Specifically, pennycress lines from cold and thermally fluctuating habitats, presumably less favorable to aphids, showed higher aphid DNA load, i.e., decreased aphid resistance. Genome-wide association analyses identified genetic variants at known defense genes but also novel genomic regions associated with variation in aphid and mildew DNA load. Moreover, we found several differentially methylated regions associated with pathogen loads, in particular differential methylation at transposons and hypomethylation in the promoter of a gene involved in stomatal closure, likely induced by pathogens. Our study provides first insights into the defense mechanisms of *Thlaspi arvense*, a rising crop and model species, and demonstrates that non-target whole-genome sequencing reads, usually discarded, can be leveraged to estimate intensities of plant biotic interactions. With rapidly increasing numbers of large sequencing datasets worldwide, this approach should have broad application in fundamental and applied research.

## Introduction

Plant pests, such as pathogens and herbivores, can cause major yield losses in crops and often require the massive use of pesticides to control their damage. Natural plant populations, on the other hand, are constantly exposed to such biotic stressors and their higher genetic diversity often allows these populations to become locally adapted. Since many pest species are sensitive to climatic conditions, their pressure on plant communities is spatially heterogeneous, maintaining geographically structured

---

**eLife digest** The genetic code of organisms is made of DNA, a molecule consisting of long sequences of four different base pairs. To gain insights into the organisms' genetic information, it is necessary to establish which base pairs are in its DNA and in what order. This is known as 'sequencing', and it allows scientists to 'read out' the genetic information of an organism.

Technically, sequencing often involves shearing the organisms' DNA into smaller pieces, so that the enzymes that do the sequencing can fully 'read' each molecule of DNA. However, when DNA is isolated from an organism, for example a plant, not only the DNA from the plant will be obtained. A small portion of DNA from other organisms, including viruses, bacteria, fungi and even insects that visited the plant will also be isolated and sequenced. These 'non-target' DNA fragments are usually discarded because they do not match the reference genome of the sequenced plant.

However, the genetic information of these other organisms can provide additional insights into the plant. This is particularly true when scientists sequence a large collection of individual plants from the same species. In this case, the DNA of other organisms isolated along with each plant's own DNA can tell researchers about differences between the plants, such as whether they are able to resist a particular disease or establish symbiosis with a specific fungus.

Galanti et al. wanted to find out more about the genetic background and characteristics of a European plant called the field pennycress, *Thlaspi arvense*. To do this, they used the fact that plants from different regions would acquire different pests depending on their genetic background, and the fact that the DNA from different creatures living with the plant would be gathered when the plant DNA was collected.

First, Galanti et al. collected pennycress seeds from across Europe and grew them in the same environment, and then they let these plants be colonized by pests. Next, the researchers tested whether the DNA of pests living on the plants reflected differences in resistance to these pests, and whether that could explain why some plants were more or less resistant based on their geographic origin and genetic background.

Galanti et al. found that, in general, plants collected in warmer and thermally stable climates, where pests usually thrive, had fewer pests in the controlled environment, suggesting that these plants had developed resistance to the pests. With this information, the researchers were also able to unravel the genetic bases of resistance, finding genetic variants in the plants with pests that were close to defense genes.

These results highlight the potential of acquiring important insights from non-target DNA fragments, especially to study plant-pathogen interactions. This could be useful in plant breeding programmes.

---

genetic variation in plant defenses (Züst et al., 2012; Kerwin et al., 2015). For these reasons, natural plant populations are highly suitable to study defense mechanisms and evolution of defenses, and also a very useful source of beneficial and resistance alleles for specific pathogens and environmental conditions. This genetic variation in defense-related genes can for example be screened through genome-wide association (GWA) (Chan et al., 2010; Corwin et al., 2016; Thoen et al., 2017; Hanson et al., 2018) or approaches based on known candidate genes (Kerwin et al., 2015).

Many pest species are also highly sensitive to temporal variation in weather conditions. This temporal heterogeneity in pathogen pressure can induce plastic responses in plants, involving gene expression and epigenetic changes (Jaouannet et al., 2015; Geng et al., 2019; Annacondia et al., 2021), which may also be studied through stress experiments (Jaouannet et al., 2015; Geng et al., 2019; Annacondia et al., 2021). Some plastic epigenetic responses can have a transient stability and be transmitted to the next generations through inheritance of epigenetic marks (Kinoshita and Seki, 2014; Espinas et al., 2016; Lämke and Bäurle, 2017; He and Li, 2018). In particular, DNA methylation has been shown to respond to biotic and abiotic stresses through gene expression regulation and transposable elements (TEs) reactivation, and can be inherited across generations (Annacondia et al., 2021; Lämke and Bäurle, 2017; Roquis et al., 2021). In plants, DNA methylation can occur in the three sequence contexts CG, CHG, and CHH (H being A, T, or C), which differ in their molecular machineries depositing, maintaining, and removing methylation and consequently also in their

transgenerational stability (Law and Jacobsen, 2010; Zhang et al., 2018). While CG methylation is usually more stable across generations, CHH methylation is less stable and more responsive to stress and the sensitivity of CHG methylation lies somewhere in between (Law and Jacobsen, 2010; Zhang et al., 2018; Liu and He, 2020).

Whether inherited or induced, some strategies of plants for defense against pathogens and herbivores include: (i) physical barriers such as reinforced cell walls, leaf protective layers, or closing stomata, (ii) production of specialized (secondary) metabolites that reduce palatability or are toxic to pests, (iii) oxidative bursts, (iv) the activation of signaling cascades to induce systemic responses, and (v) RNA interference mechanisms to silence pathogen genes (Wojtaszek, 1997; War et al., 2012; Kant et al., 2015; Melotto et al., 2017; Muhammad et al., 2019). In Brassicaceae, a particularly important and diverse class of defense metabolites are glucosinolates, which often show local adaptation driven by variation in pests and can also be induced by herbivore and pathogen attacks (Züst et al., 2012; Kerwin et al., 2015; Kutyniok and Müller, 2012).

Studying natural variation in plant resistance, along with associated genetic and epigenetic variation, can identify genes involved in defense and their regulators, including vital genes whose function cannot be determined through knockout experiments. Such knowledge, and especially the discovery of natural resistance alleles, are crucial sources for the breeding of more pest-resistant crop varieties. Nevertheless, because of the diversity of resistance mechanisms and their often multigenic nature, plant defense mechanisms remain difficult to study. In particular, antixenosis (the prevention of pathogen settlement) and antibiosis (the repression of pathogen growth and reproduction) require extensive and time-consuming phenotyping, based for example on choice (Nalam et al., 2018) or settling (Annacondia et al., 2021) assays, and such assays are extremely challenging to perform on large collections. On the other hand, there are increasing numbers of large sequencing datasets, which may also be used to quantify contaminants or microbiome composition (Sangiovanni et al., 2019; Roman-Reyna et al., 2020; Gathercole et al., 2021) and thus as proxies for resistance phenotyping. In this study we investigated such usage of exogenous reads, i.e., reads not mapping to the target reference genome, as a source of information for quantifying herbivore and pathogen abundance in large collections.

We worked with field pennycress (*Thlaspi arvense*), an annual plant in the Brassicaceae family that is increasingly studied as a model species (Geng et al., 2021; Nunn et al., 2022; Hu et al., 2022; Troyee et al., 2022; Galanti et al., 2022) and new biofuel and winter cover crop (Dorn et al., 2015; Frels et al., 2019; Chopra et al., 2019; Zhao et al., 2021). In a previous study, we investigated natural epigenetic variation in a collection of 207 *Thlaspi* lines from across Europe (Galanti et al., 2022). Prior to their whole-genome (WGS) and -epigenome sequencing these lines had been grown in a common environment, an open glasshouse where the plants were spontaneously colonized by aphids and powdery mildew, as well as by other microbes. At the time of sequencing, pathogen contamination was still very limited but appeared highly variable, and preliminary analyses showed that it resulted in sizeable amounts of non-target reads assigned to the pest species, i.e., contamination of the DNA samples. Inspired by other recent studies on non-target reads (Sangiovanni et al., 2019; Roman-Reyna et al., 2020; Gathercole et al., 2021), we asked if there was systematic variation in the numbers of aphid and pathogen reads among different *T. arvense* lines, and whether these data, together with our whole-genome plant sequencing data, could provide insights into the genomic basis of plant resistance variation.

The goals of our study were thus twofold: (i) to contribute to a mechanistic understanding of pest resistance in *T. arvense*, and (ii) to explore whether non-target reads from plant sequencing can be used as proxies for studying plant biotic interactions. Considering that we are moving toward an increasingly sequencing-prone world, with more and larger datasets being generated for many species (Kajiya-Kanegae et al., 2021; Colgan et al., 2022; Habyarimana et al., 2022; Mekbib et al., 2022; Metheringham et al., 2022; Nocchi et al., 2022; Friis et al., 2024), the use of non-target reads has very broad potential.

## Results

### Reads classification and species identification

Starting from our previously published sequencing data (Galanti et al., 2022), the first step of our analysis was to separate the WGS reads of each sample into the ~99.5% mapping to the *T. arvense* reference genome (Nunn et al., 2022) and the ~0.5% that did not, hereafter called ‘exogenous reads’ (Figure 1A). Initially, we used all mapped reads for calling variants in *Thlaspi*, but after some difficulties with genome-wide association (GWA) analyses (see below) we suspected that some plant reads were false and mapped to the *T. arvense* genome only because of the high cross-taxa similarity of some genomic regions. We therefore remapped all reads to the genomes of the aphid *Acyrtosiphon pisum*, its endosymbiont *Buchnera aphidicola* and the powdery mildew *Blumeria graminis*, and found that, on average, 7.4% of the reads mapped to both *T. arvense* and at least one of the pests. We removed these ambiguous reads from our analyses and used only the *T. arvense* target reads, 92.1% on average, for variant calling (Figure 1A, Supplementary file 1).

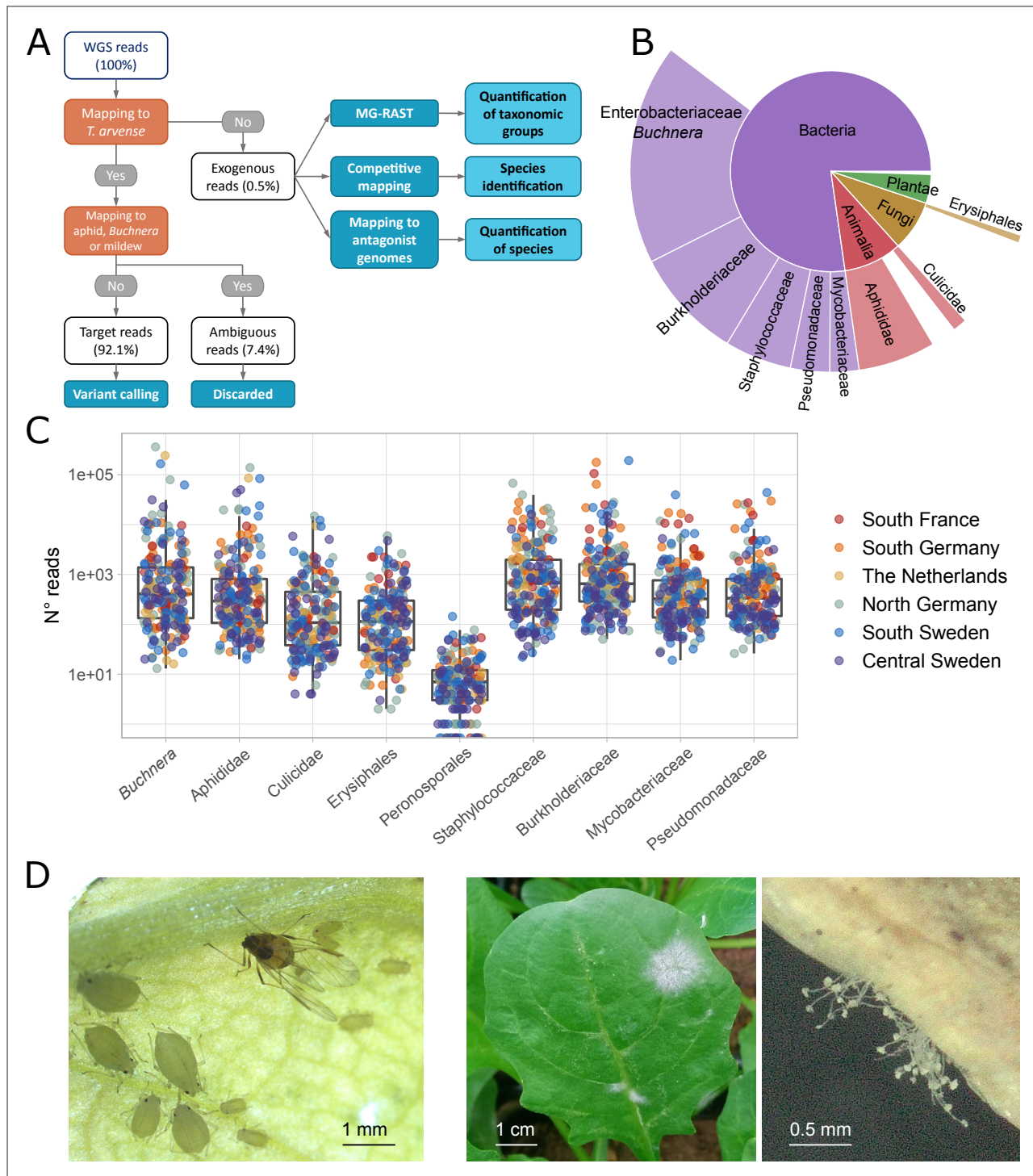
We next attempted a taxonomic classification of the exogenous reads, in multiple steps. First, we used MG-RAST (Meyer et al., 2008; Keegan et al., 2016) to assign reads to taxonomic groups based on public sequencing databases. Out of the 78% of the exogenous reads that passed the MG-RAST quality control (Supplementary file 1) the majority belonged to bacteria and smaller fractions to fungi, plants, and animals (Figure 1B and Supplementary file 2). For subsequent group-level analyses, we then focused on nine taxonomic groups that occurred consistently within our samples (Figure 1C), and that were particularly abundant or relevant: Erysiphales (fungi), Peronosporales (oomycetes), Aphididae, and Culicidae (both insects), and five bacterial families.

Visual inspection (Figure 1D) and other sources of information narrowed down the observed aphid and mildew species to a few candidates. For aphids we considered *A. pisum* (indicated by MG-RAST), *Myzus persicae* (visual match, and a generalist attacking Brassicaceae; CABI, 2021) and *Brevicoryne brassicae* (attacks Brassicaceae including *Thlaspi*; Gabrys and Pawluk, 1999). For powdery mildew we considered *B. graminis* (indicated by MG-RAST), and *Erysiphe cruciferarum* (attacks Brassicaceae; Warwick et al., 2002). To decide among these species, we then used a competitive mapping approach (Feuerborn et al., 2020), where the exogenous reads were aligned to a pseudo-reference composed of the same DNA sequences from the different candidate species (see Materials and methods for details, Supplementary file 3 and 4). The majority (77%) of the aphid reads mapped to *M. persicae*, 18% to *B. brassicae*, and 5% to *A. pisum*, while 98% of the mildew reads mapped to

**Table 1.** Population differences and SNP-based heritability for different types of exogenous read counts.

Population differences were tested with a linear model, SNP-based heritabilities (and their confidence intervals) estimated with the R package *heritability* (Kruijer and White, 2019).

Taxonomic group	Data type	Population differences (R <sup>2</sup> and p-value)	SNP-based heritability
<i>Myzus persicae</i>	Mapping to reference genome	0.245 (p=0.029)	0.190 (0.055–0.488)
<i>Buchnera aphidicola</i>	Mapping to reference genome	0.256 (p=0.016)	0.169 (0.042–0.490)
<i>Buchnera</i>	MG-RAST - genus	0.223 (p=0.090)	0.115 (0.016–0.505)
Aphididae	MG-RAST - family	0.226 (p=0.082)	0.189 (0.052–0.496)
Culicidae	MG-RAST - family	0.166 (p=0.519)	0.183 (0.055–0.465)
Erysiphales	MG-RAST - order	0.326 (p<0.001)	0.468 (0.238–0.712)
Peronosporales	MG-RAST - family	0.253 (p=0.020)	0.266 (0.096–0.553)
Staphylococcaceae	MG-RAST - family	0.390 (p<0.001)	0.301 (0.124–0.567)
Burkholderiaceae	MG-RAST - family	0.275 (p=0.005)	0.256 (0.092–0.538)
Mycobacteriaceae	MG-RAST - family	0.362 (p<0.001)	0.294 (0.120–0.560)
Pseudomonadaceae	MG-RAST - family	0.273 (p=0.006)	0.192 (0.052–0.505)



**Figure 1.** Classification of sequencing reads in *T. arvense* whole-genome sequencing (WGS) data. **(A)** Workflow of the analyses, including reads classification (orange nodes) into target, ambiguous, and exogenous reads, and downstream analysis (dark blue nodes) (see Materials and methods). **(B)** Fractions of exogenous reads assigned to different taxonomic groups by MG-RAST (Meyer et al., 2008; Keegan et al., 2016). **(C)** Read counts assigned to nine selected groups in our 207 *T. arvense* samples from different European regions. **(D)** Aphids and mildew occurring on *T. arvense* leaves during our experiment.

The online version of this article includes the following figure supplement(s) for figure 1:

**Figure supplement 1.** Pest loads in samples with or without pests in the field.

*E. cruciferarum*, with only 2% to the other mildew species (**Supplementary file 5**). Based on these results, we concluded that the plants in our experiment had been attacked by *M. persicae* and *E. cruciferarum*.

Finally, to compare the power of a large database approach (MG-RAST) vs. using specific reference genomes, we also remapped all exogenous reads to the *M. persicae* and *B. aphidicola* genome assemblies (Singh et al., 2021) ([https://www.ncbi.nlm.nih.gov/datasets/genome/GCF\\_001939165.1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_001939165.1)) and used the counts from these two mappings as additional phenotypes, besides the nine taxonomic groups selected through MG-RAST (**Table 1, Supplementary file 6**).

## Exogenous read counts are a heritable *Thlaspi* phenotype

As we had observed that aphid and mildew infections in the glasshouse were not random, but prevalent on plants from some origins than others, i.e., possibly reflecting heritable variation in plant resistance, we next tested for population differences and SNP-based heritability in pest and microbiome read counts. Prior to these analyses, to avoid biases caused by different sequencing depths, we corrected the read counts for the total numbers of deduplicated reads in each library and used the residuals as unbiased estimates of aphid, mildew, and microbe loads (see Materials and methods).

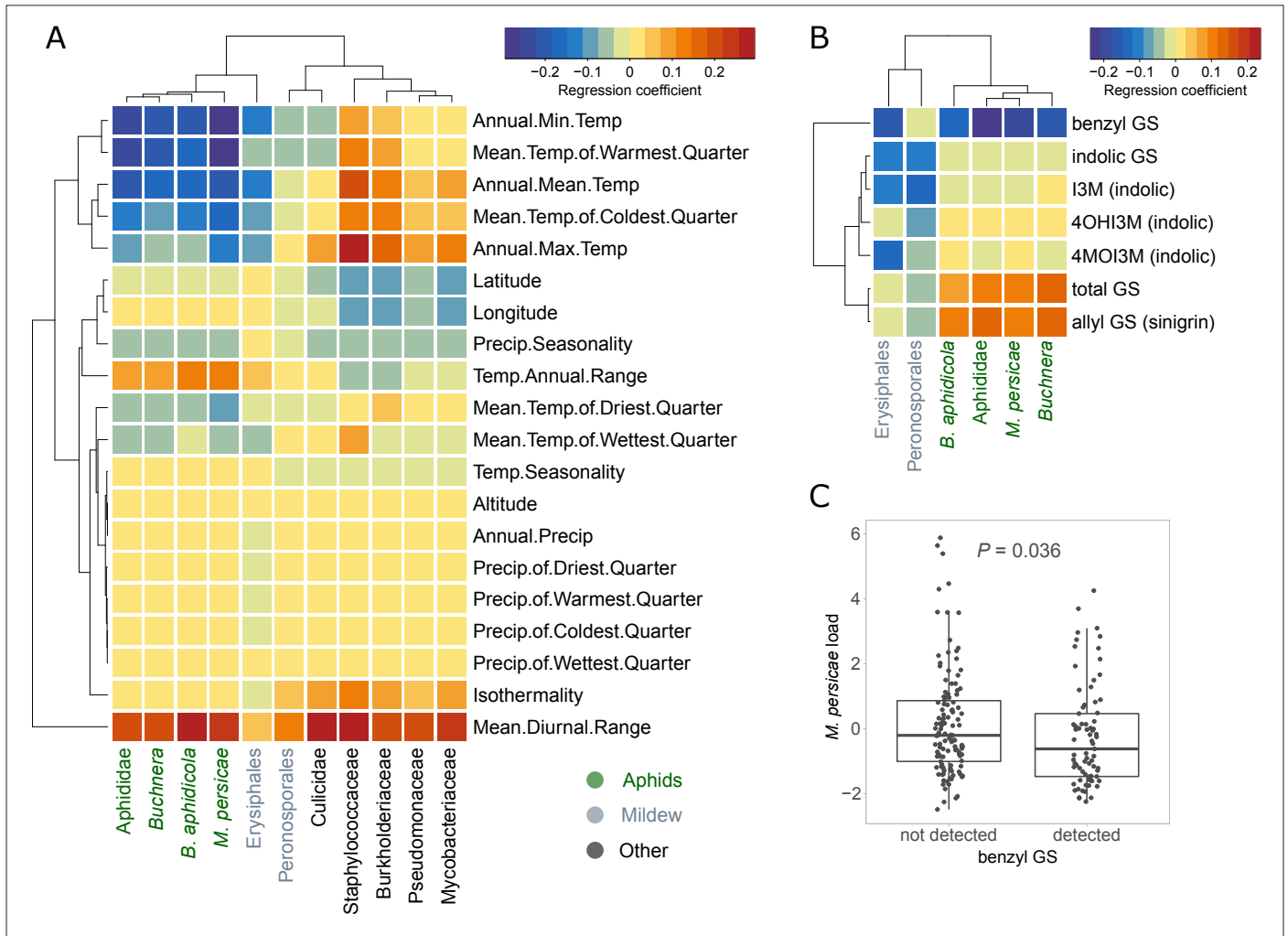
For most of the nine taxonomic groups, there were significant population effects, with 20–40% of the variance in read counts explained, as well as significant SNP-based heritability, typically in the range of 0.18–0.30 (**Table 1**). The highest heritability of 0.47 was for read counts of Erysiphales, indicating particularly strong variation for resistance to mildew. Both SNP-based heritability and population differences tended to be stronger for aphid and *Buchnera* data based on read mapping to the reference genomes than for those based on MG-RAST, demonstrating that the former method is more informative and thus preferable if high-quality genome assemblies are available.

An alternative explanation for different aphid and mildew loads in the greenhouse could be that variation in enemy densities in the field was transmitted to the greenhouse, through maternal carry-over effects, or even as seed contamination. However, we had recorded aphid and mildew occurrence during seed sampling in the field and found no significant differences in the glasshouse between the offspring of plants that had been attacked in the field vs. those that had not (**Figure 1—figure supplement 1**).

## Aphid and mildew loads correlate with climate of origin and glucosinolates content of plants

Having established that our method most likely captured variation in plant resistance, we were interested in the ecological drivers of this variation. As climate is known to be a major influence on many biotic interactions as well as plant defenses (Züst et al., 2012; Gao et al., 2019), we correlated the observed read counts with the climates of origin of the plants. We found negative correlations between aphid read counts and several temperature variables, in particular annual minimum temperature (**Figure 2A**). Aphid read counts were also positively correlated with temperature variability, i.e., the diurnal and annual ranges of temperature (**Figure 2A**). In other words, plants from warmer and more stable climates had consistently lower levels of aphid infestation in our glasshouse, possibly because these plants had evolved greater resistance under such benign climatic conditions where aphids thrive. We found similar, although weaker patterns, for the number of Erysiphales reads. The other analyzed taxonomic groups showed different and often weaker patterns of correlation with climate, except that the read counts of several bacterial groups were positively correlated with annual maximum temperature and in particular diurnal temperature range.

Since glucosinolates are major defense metabolites of Brassicaceae, and their variation could thus be an explanation for variance in plant resistance, we also tested for correlations between the baseline amounts of these metabolites and the frequencies of aphid and mildew reads. Glucosinolate levels were measured on the same *T. arvense* lines in a separate experiment not affected by pests (**Supplementary file 7**). We found positive correlations of aphid read counts with allyl glucosinolate (sinigrin), an aliphatic glucosinolate which is by far the most abundant in the leaves of *T. arvense*, and a stronger negative correlation with benzyl glucosinolates (glucotropaeolin) (**Figure 2B**). Although the baseline levels of benzyl glucosinolates were very low and probably sometimes below the detection level, plant lines where benzyl glucosinolate was detected had significantly lower aphid loads in the glasshouse



**Figure 2.** Relationships between climates of origin or glucosinolate levels of plants and the exogenous reads loads. **(A)** Correlations with bioclimatic variables. **(B)** Correlations with baseline glucosinolate (GS) levels measured in the same pennycress lines in another experiment. All correlations in **(A)** and **(B)** were done after correction for population structure. Aphid-related read counts are in green, mildew-related in gray, others in black. **(C)** Boxplot of the aphid reads residuals in samples where benzyl GS was detected vs. not. The p-value is based on the Welch's t-test for unequal variances. I3M: indol-3-yl methylGS; 4MOI3M: 4-methoxyindol-3-yl methylGS; 4OHI3M: 1-methoxyindol-3-yl methylGS.

(**Figure 2C**). We also detected three indole glucosinolates, but these did not show any significant correlations with aphid loads.

### GWA identifies peaks near defense genes

To further investigate the genetic basis of variation in aphid, mildew, and microbe loads, we next employed GWA and tested for associations between exogenous read counts and biallelic genetic variants (SNPs and short INDELs). We corrected for population structure using an isolation by state (IBS) matrix and only tested variants with minor allele frequency (MAF)>0.04 (see Materials and methods). Initially, we called genetic variants using all reads that mapped to the *T. arvense* genome and found massive peaks in some highly conserved regions of the genome, which had very high mapping coverage. We suspected that this might be because some non-*Thlaspi* reads were very similar to these highly conserved regions and, by mapping there, generated false variants only in samples containing many non-*Thlaspi* reads. We therefore identified and removed ambiguous reads prior to variant calling, which eliminated the observed massive GWA peaks, indicating that they had indeed reflected false associations (**Figure 3—figure supplement 1**).

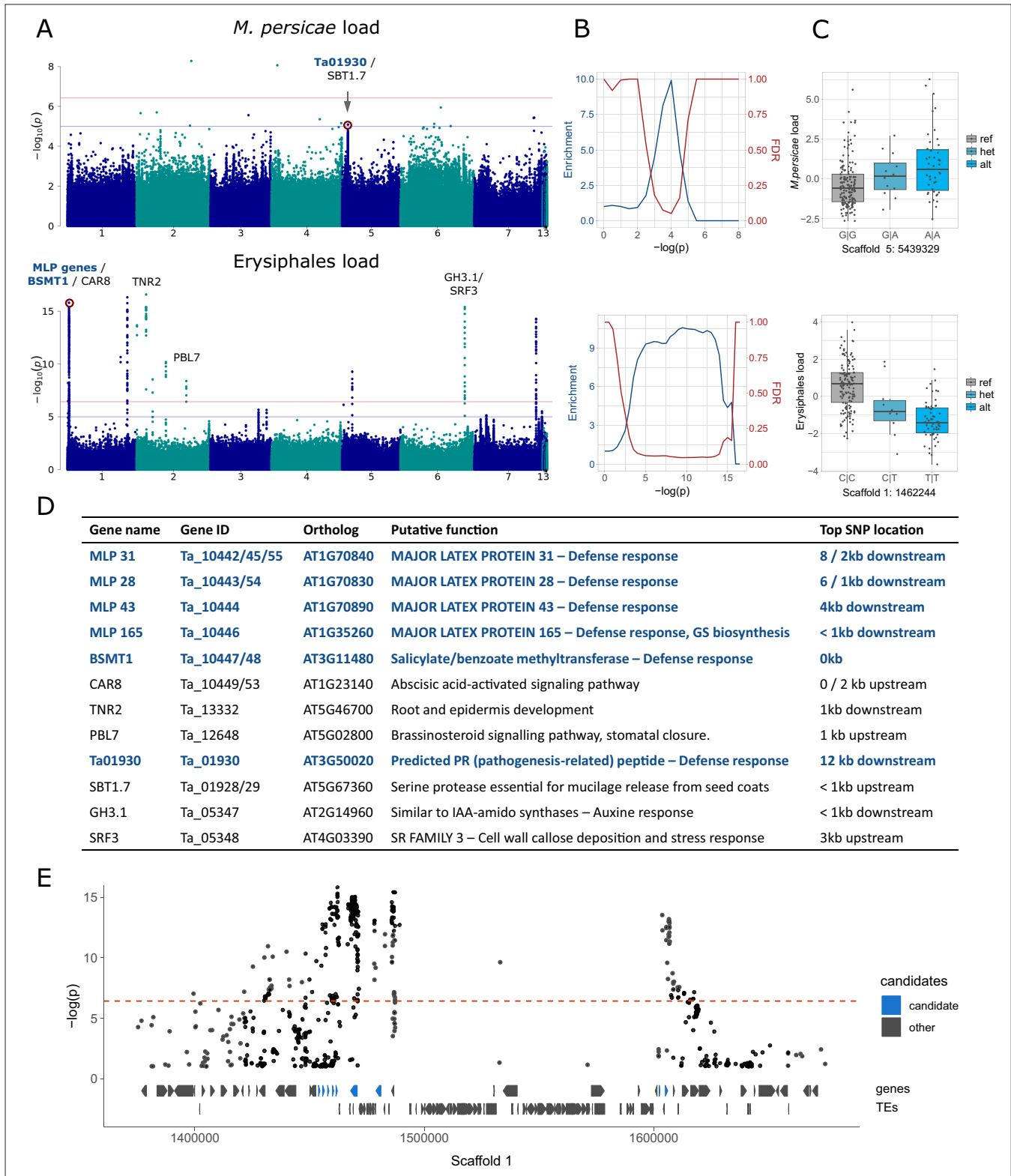
After excluding the ambiguous reads, we still found significant GWA peaks for Erysiphales but not for other types of exogenous reads (excluding isolated, unreliable variants) (**Figure 3A, Figure 3—figure supplements 2 and 3**). Nevertheless, when clear peaks were visible, regardless of their significance, they were usually located close to genes involved in plant defense response. An enrichment analysis (**Atwell et al., 2010**) confirmed that stronger variants were indeed enriched close to these defense genes (**Supplementary file 8**) for some exogenous read counts (**Figure 3B, Figure 3—figure supplements 2 and 3**). For *M. persicae* load there was a peak in the proximity of *Tarvense\_01930*, encoding a predicted pathogenesis-related peptide. The top variant in this peak had a slight but clear allelic effect on *M. persicae* load (**Figure 3C**). For Erysiphales load we detected a more persistent enrichment, with a highly significant peak in Scaffold 1, located in a region with several defense genes, including *MAJOR LATEX PROTEINS (MLP)* and two genes similar to *Arabidopsis thaliana SALICYLATE METHYLTRANSFERASE 1 (BSMT1)* (**Figure 3D and E**). This region is wide due to ancient TE colonization, but the top variants are clearly neighboring candidate genes involved in defense (**Figure 3E**). Other significant peaks for Erysiphales load were close to other genes that possibly contribute to resistance such as *PBL7*, involved in signaling and stomatal closure or *SRF3*, reinforcing cell walls by callose deposition.

## Aphid and mildew loads correlate with differential methylation at genes and transposons

Variation in phenotypes, such as our indirect estimates of pest resistance, may not only be associated with DNA sequence but also with epigenetic changes like DNA methylation. This phenotype-associated epigenetic variation can include both heritable and plastic components. The whole-genome bisulfite sequencing (WGBS) data from our previous study (**Galanti et al., 2022**) allowed us to also explore these questions and to test for associations between DNA methylation variation and pest attack. For simplicity, we limited this analysis to *M. persicae* and Erysiphales loads.

Our analysis had two steps: First we called differentially methylated regions (DMRs) between the 20 samples with the most and 20 samples with the least *M. persicae* or Erysiphales loads, and then we conducted epigenome-wide association (EWA) analyses on individual positions located within these DMRs, using the complete dataset (188 lines - see Materials and methods). This approach allowed us to target genomic regions of interest, while strongly reducing the multiple-testing problem of millions of cytosines in the whole genome and correcting for population structure. Using a relaxed false discovery rate (FDR) of 20%, we identified 162 DMRs for *M. persicae* load and 548 DMRs for Erysiphales load (**Figure 4—figure supplement 1, Supplementary file 9 and 10**). The majority of these were in the CG context, especially for *M. persicae*-related DMRs (**Figure 4A, Figure 4—figure supplement 1**). As observed previously (**Galanti et al., 2022**), DMRs in CHH were generally shorter than in the other sequence contexts (**Figure 4—figure supplement 1**). Since the genome of *T. arvense* is rich in TEs and intergenic regions, the majority of DMRs were located in those features (**Figure 4—figure supplement 1**). However, the DMR density was higher in proximity of genes and particularly in coding sequences (**Figure 4A**), and even DMRs assigned as intergenic (**Figure 4A**) were often located close to genes or promoters. In accordance with previous studies (**Geng et al., 2019; Annacondia et al., 2021**), most DMRs were hypomethylated in the infested samples (higher pathogen load), indicating that genes needed for defense might be activated through demethylation.

For a more detailed investigation, we turned to EWA, leveraging the power of the entire *Thlaspi* collection. We tested for associations between *M. persicae* or Erysiphales loads and the methylation at individual cytosines located within the DMRs. As in GWA, we corrected for population structure using an IBS matrix. For both types of pest loads, we found associations in the proximity of genes and especially within TEs, but no genomic feature was particularly enriched for low p-value associations (**Figure 4—figure supplement 2**). *M. persicae* load was associated with methylation at several genomic locations, especially TEs (**Figure 4B**), but these associations had strongly inflated p-values (**Figure 4—figure supplement 2**). For Erysiphales load the p-value distribution was more well-behaved, and we found a clear association with hypomethylation of Copia family 202 TEs upstream of *MAPKK KINASE 20 (MAPKKK20)*, a gene involved in abscisic acid (ABA) stress response and stomatal closure (**Figure 4B, C, and D**). A coverage analysis confirmed that none of the *T. arvense* lines carries insertions or deletions of the TEs upstream of *MAPKKK20*.



**Figure 3.** Genome-wide association analyses for aphid and mildew loads. We show only the results for *M. persicae* and MG-RAST Erysiphales read counts; for full results see **Figure 3—figure supplements 2 and 3**. **(A)** Manhattan plots, annotated with genes potentially affecting aphid/mildew colonization. The genome-wide significance (horizontal red line) was calculated based on unlinked variants (Sobota et al., 2015), the blue line corresponds to  $-\log(p)=5$ . **(B)** Corresponding to the Manhattan plots on the left, enrichment of a priori candidates and expected false discovery rates (as **Figure 3 continued on next page**

Figure 3 continued

in [Atwell et al., 2010](#)) for increasing significance thresholds. (C) Allelic effects of the red-marked variants in the corresponding Manhattan plots, with genotypes on the x-axes and the read count residuals on the y-axes. (D) The candidate genes marked in panel A, their putative functions and distances to the top variant of the neighboring peak. Candidates in dark blue are the a priori candidates included in the enrichment analyses and involved in defense response (GO:0006952). GS: glucosinolates. (E) Zoom-in for the Manhattan plot of Erysiphales load, around the first peak in Scaffold 1, with gene and transposable element (TE) models below, and a priori candidates in blue.

The online version of this article includes the following figure supplement(s) for figure 3:

**Figure supplement 1.** Example of a genome-wide association (GWA) peak caused by ambiguous reads.

**Figure supplement 2.** Genome-wide association (GWA) results for all exogenous reads.

**Figure supplement 3.** Genome-wide association (GWA) results for all exogenous reads.

## Discussion

Plant pests are a major threat to food safety, causing large yield losses, and new crops such as the potential biofuel plant *T. arvense* must be able to resist pathogen and herbivore attacks. A powerful source for obtaining resistant varieties is natural variation in plant defenses, but phenotyping large collections can be very time-consuming and error-prone. Here, we describe how an unplanned pest infestation in a glasshouse experiment, together with available WGS data, can be used to estimate aphid, mildew, and microbial loads, and thus variation in plant resistance. The approach is straightforward, makes use of WGS data without microbiome-specific DNA extraction, and can in principle be applied to many other situations such as field experiments. It is not error-free, but we highlight some potential pitfalls, show how to reduce noise, and illustrate its potential to detect associations with climatic, genetic, and epigenetic variation.

An important first step in our analyses was the identification and classification of pest-related reads in the plant WGS data. We began by classifying all reads as target (only mapping to *T. arvense*), ambiguous (mapping to *T. arvense* and at the same time to at least one of the pest genomes), or exogenous (not mapping to *T. arvense*) (**Figure 1A**). We demonstrated the importance of removing ambiguous reads prior to variant calling, as this prevented calling false positive variants caused by exogenous DNA that also mapped to highly conserved or repetitive sequences in the *T. arvense* genome. We then classified the exogenous reads using MG-RAST ([Meyer et al., 2008](#); [Keegan et al., 2016](#)) or by confident mapping to specific pest genomes, and selected the 11 most relevant and/or abundant taxonomic groups to focus our analyses on. To obtain unbiased pest/microbe loads we also corrected the read counts for the total number of deduplicated reads of each sample. A competitive mapping approach allowed us to identify the aphid and mildew species that had occurred in our experiment as the generalist aphid *M. persicae* and the powdery mildew *E. cruciferarum* (**Figure 2**).

Since we suspected a non-random colonization of pests and microbes in our *T. arvense* collection, we tested for population differences as well as SNP-based heritability. We found significant population differences for most pest and microbe loads, and often heritabilities above 15%, which although low, is still indicative of genetic determination ([Thoen et al., 2017](#)). Moreover, Erysiphales load had a particularly high heritability of 47% (**Table 1**). We therefore next asked what could explain the observed variation in pest loads in our experiment. As pathogen abundances in the field are often determined by climatic conditions, we expected plants originating from climates less favorable to aphids to perform worse in our glasshouse, i.e., to have higher pathogen loads. As expected, aphid counts were negatively correlated with temperature of origin (particularly minimum temperature), and positively with temperature variability (mean diurnal range and temperature annual range) (**Figure 3A**), suggesting that plants from colder and more thermally fluctuating climates, which are less favorable to aphids, were less well defended and performed worse in our glasshouse. We found similar but weaker patterns for Erysiphales load.

As we expected the observed climate-associated variation in pest loads to be at least partially explained by variation in chemical defenses, and since in *A. thaliana* glucosinolates, the main defensive compounds, are known to be geographically structured in response to aphid distributions ([Züst et al., 2012](#)), we also tested for association of aphid and mildew loads with glucosinolates in our collection. In accordance with literature on *A. thaliana* ([Kim and Jander, 2007](#)), we observed a positive correlation of aphid loads with total glucosinolates as well as with the most abundant glucosinolate sinigrin (aliphatic glucosinolate), but a negative correlation with benzyl glucosinolate (**Figure 3B**).



**Figure 4.** Differential methylation associated with aphid and mildew loads. **(A)** Differentially methylated region (DMR) densities in different genomic features when comparing the 20 samples with the most vs. the least *M. persicae* (top) or *Erysiphales* (bottom) load. CDS: coding sequences. **(B)** Manhattan plots from epigenome-wide association (EWA) analyses based on individual cytosines within DMRs, with sequence contexts in different colors and annotation of genes close to low p-value cytosines. The genome-wide Bonferroni significance thresholds (dashed red lines) were calculated *Figure 4 continued on next page*

Figure 4 continued

based on the number of DMRs. (C) Candidate genes and transposable elements (TEs) marked in panel B, their putative functions, genomic locations of associated DMRs, and whether infested samples were hyper- or hypomethylated. (D) Zoomed-in Manhattan plot for Erysiphales load around the peak in Scaffold 4, with gene and TE models given below. The CG methylation in the 20 most and least infested samples was calculated over 50 bp bins (see **Figure 4—figure supplement 2** for methylation in other contexts).

The online version of this article includes the following figure supplement(s) for figure 4:

**Figure supplement 1.** Lengths and genomic locations of differentially methylated regions (DMRs).

**Figure supplement 2.** Epigenome-wide association (EWA) enrichment in different genomic features and p-value distributions.

These findings suggest that glucosinolate composition, rather than total amount, is important for aphid defense, and that while benzyl glucosinolate might have a deterrent effect, sinigrin might on the contrary attract *M. persicae* or act as a stimulant, which would be in accordance with previous observations (Klingauf et al., 1972).

To detect genetic variants associated with pest and microbe loads, we then conducted a GWA study (GWAS). For aphid (*M. persicae*) load, we detected only one non-significant peak in Scaffold 5, close to a pathogenesis-related coding gene (**Figure 3A and D**). For Erysiphales load, however, there were several significant associations neighboring genes directly involved in defense, mostly members of the MLP family, clustered in a large peak on the first arm of Scaffold 1 (**Figure 3E**). MLP165, the closest gene to the most significant variant in the peak, is indirectly involved in glucosinolate biosynthesis in *A. thaliana* (The Arabidopsis Information Resource (TAIR), 2000), which might explain why baseline glucosinolate levels were associated with Erysiphales load (**Figure 3B**). Further GWA peaks for Erysiphales pointed toward other genes indirectly involved in the defense response through phytohormone signaling (e.g. CAR8, PBL7, GH3.1) or preventing pathogen access through cell wall reinforcement or stomatal closure (SRF3, PBL7) (**Figure 3D**). Further experiments would be necessary to confirm the functionality of these genes.

An important general insight from our GWA analyses was the frequent ambiguity of reads that mapped to both pest and host plant genomes. Such ambiguous reads generated false variants only present in samples with pest DNA, which resulted in highly significant false associations, and it was therefore important to remove these reads before variant calling. Another potential reason for sequence similarity between host and pathogens could be defense mechanisms such as RNA interference. If *T. arvense* produces small or micro-RNAs to silence pathogen genes, this would originate from genomic regions of high similarity between the host and the pathogen, and thus reflect a true association. However, a BLAST (Camacho et al., 2009) of the region in which the suspicious associations occurred did not reveal any similarity to aphid or mildew genes, but instead to the highly conserved ribosomal RNA coding regions. While genetic variants are generally inherited from parents and thus reflect evolutionary processes, DNA methylation variants can be heritable but can also reflect plastic responses to environmental stresses like herbivores or pathogens. Our data do not allow to confidently distinguish between these two sources of DNA methylation variation, and thus should be interpreted with caution, especially with regard to the directionality of associations. A beneficial DNA methylation variant is expected to be associated with lower pathogen load when already present before pathogen arrival, but with higher pathogen load when plastically induced by pathogens during the experiment. For both *M. persicae* and Erysiphales, the majority of DMRs were hypomethylated in affected samples, which is in accordance with the loss of methylation observed in *A. thaliana* and *T. arvense* upon aphid feeding, and in diploid wheat upon powdery mildew infection (Geng et al., 2019; Annacondia et al., 2021; Troyee et al., 2022), but we also detected hypermethylation at several loci. *M. persicae* load was associated with differential methylation at only few genes but several TEs, which is in accordance with the aphid or stress-induced TE reactivation observed in *A. thaliana* (Annacondia et al., 2021; Roquis et al., 2021). Erysiphales load was associated with hypomethylated Copia TEs upstream of MAPKKK20, a gene involved in ABA-mediated signaling and stomatal closure. Since stomatal closure is a known defense mechanism to block pathogen access (Melotto et al., 2017), it is tempting to conclude that hypomethylation of the MAPKKK20 promoter might induce its overexpression and consequent stomatal closure, thereby preventing mildew access to the leaf blade. Overall, we found associations between pathogen load and TE methylation that could potentially act both in *cis* (e.g. Copia TE methylation in MAPKKK20 promoter) and in *trans*, e.g., through transposon reactivation (e.g. LINE, Helitron, and Ty3/Gypsy

TEs isolated from genes). Although we cannot confidently distinguish inherited vs. induced DNA methylation variants, to our knowledge this is the first coupled GWA-EWA analysis conducted on a large natural plant collection.

In summary, our study offers first insights into the defense mechanisms of *T. arvense*, including candidate genes and alleles which may be of interest for breeding efforts in this novel biofuel and cover crop. It also provides a proof of principle that exogenous reads from large sequencing efforts, usually discarded if not mapping to the target genome, can be leveraged to extract additional information about important biotic interactions of the target species, including its antagonists and microbiome components. We combined this approach with data from a common environment experiment to show that pest and microbiome load were geographically structured, as expected from locally adapted traits, and associated with both genetic and DNA methylation variants. In principle, our approach can be applied to many other designs. For example, field-collected samples could be used to quantify geographic pathogen distributions. With the decreasing cost of sequencing and increasing large-scale and single-species sequencing projects (e.g. *Kajiya-Kanegae et al., 2021; Colgan et al., 2022; Habyarimana et al., 2022; Mekbib et al., 2022; Metheringham et al., 2022; Nocchi et al., 2022; Friis et al., 2024*), the number of datasets suitable for such analyses is set to rapidly increase in the near future.

## Materials and methods

### Plant growth and sequencing

The WGS data used in this study were already published in *Galanti et al., 2022*. Please refer to this publication for details on data generation and methods. Briefly, we collected 207 *T. arvense* accessions from 36 European populations in July 2018, and we grew their offspring in a completely randomized design, in a glasshouse at the University of Tübingen (48°32'21.3"N, 9°02'04.2"E) between August and October 2019. The glasshouse was located in biodiverse surroundings, and insects and pests could enter when the windows opened for temperature regulation. A few weeks after germination, we noticed aphid and mildew infestations. After 46 days we sampled the third or fourth true leaf of each plant and snap-froze it in liquid nitrogen. We extracted DNA using the DNeasy Plant Mini Kit (QIAGEN, Hilden, DE), sonicated (Covaris) 300 ng of genomic DNA and used the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) to prepare the libraries. Half way through the protocol we split the DNA into 1/3 for genomic libraries and 2/3 for bisulfite libraries. For the bisulfite conversion we used the EZ-96 DNA Methylation-Gold MagPrep (ZYMO) kit. We sequenced paired-end for 150 cycles using Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA) for genomic libraries and HiSeq X Ten (Illumina, San Diego, CA, USA) for bisulfite libraries.

### Reads mapping and classification

Upon demultiplexing the raw reads, we used cutadapt (*Martin, 2011*) for quality (minimum quality of 20) and adaptor trimming, excluding reads shorter than 35 bp. We used FastQC and MultiQC (*Andrews, 2010; Ewels et al., 2016*) to estimate the duplication rate, and calculated the total deduplicated reads, which we later used for correcting the number of exogenous reads. We then classified the reads based on their mapping behavior. First we aligned reads to the *T. arvense* reference genome (*Nunn et al., 2022*) with BWA-MEM v0.7.17 (*Li and Durbin, 2009*), excluding multimapping reads (-c 1) and marking duplicates with MarkDuplicatesSpark (*Van der Auwera et al., 2013; Poplin et al., 2018*). We then mapped all samples again (*Li and Durbin, 2009*) to the three putative exogenous genomes of pea aphid (*Acrophyson pisum*), the aphid symbiont *B. aphidicola* and powdery mildew (*B. graminis*), using available resources ([https://www.ncbi.nlm.nih.gov/assembly/GCF\\_005508785.2](https://www.ncbi.nlm.nih.gov/assembly/GCF_005508785.2), [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF\\_001939165.1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_001939165.1)) (*Frantzeskakis et al., 2018*). After this, we used a custom script to collect all read IDs within a sample mapping to any of the three exogenous genomes, and removed any of these reads from the *T. arvense* alignment bam files. We thus removed all ambiguous reads before proceeding with variant calling. To compare coverage of specific regions with and without ambiguous reads, we used samtools bedcov (*Danecek et al., 2021*). The numbers of reads classified by their mapping behavior are reported in **Supplementary file 1**.

## Variant calling

For variant calling we used GATK4 v4.1.8.1 (Van der Auwera et al., 2013; Poplin et al., 2018), following the best practices for germline short variant discovery (<https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels->) with few adjustments for large datasets (Galanti et al., 2022). Briefly, starting from the bam files generated after the removal of ambiguous reads, we (i) ran HaplotypeCaller, (ii) combined the resulting GVCF files with GenomicsDBImport and GenotypeGVCFs, and (iii) filtered out low-quality variants with VariantFiltration (see Galanti et al., 2022 for more details). Finally, we used vcftools v0.1.16 (Danecek et al., 2011) to retain biallelic variants with MAF>0.01 and a maximum of 10% missing genotype calls. We imputed these missing calls with BEAGLE 5.1 (Browning et al., 2018) to obtain a complete multisample vcf file.

## Identification and classification of exogenous reads

To identify exogenous reads, we extracted all unmapped reads from the bam files created aligning WGS reads to the *T. arvense* genome (Nunn et al., 2022). We selected reads with both mates unmapped (SAM flag 12) and excluded supplementary alignments (SAM flag 256 after running MarkDuplicatesSpark) with samtools (Danecek et al., 2021). We then recovered these reads from the trimmed fastq files with seqtk subset (<https://github.com/lh3/seqtk>; Li, 2024) to obtain fastq files of unmapped reads only. We used these as input for MG-RAST (Meyer et al., 2008; Keegan et al., 2016), a web-based tool for phylogenetic analysis of metagenomes.

We ran MG-RAST mostly with default parameters, without assembled reads, excluding dereplicated sequences, and dynamically trimming reads with a minimum Phred score of 15 in more than 5 consecutive bases. We set the 'sequence screening' to *A. thaliana*, the closest relative of *T. arvense* available. We used two different approaches to extract read counts. First, we classified all reads up to family level using the web-based Analysis tool from MG-RAST. We used RefSeq as query annotation database and filtered reads classified with low confidence using default settings: e-value 5, 60 %-identity, length 15, and min.abundance of 1 (Supplementary file 2). Out of the hundreds of taxonomic groups identified by MG-RAST, we selected only a small subset for follow-up analyses, based on their biological relevance, our visual observations and/or abundance: Aphididae, Culicidae, Peronosporales, Staphylococcaceae, Burkholderiaceae, Mycobacteriaceae, and Pseudomonadaceae (Table 1). Additionally, we used a custom Python script to download individual 'taxonomy' or 'sequence\_break-down' results from MG-RAST API (Paczian et al., 2019) and extracted the counts of the genus *Buchnera*, including bacterial symbionts of many aphid species, and of the order Erysiphales, to quantify the observed mildew infection (Table 1). All the code for extracting counts for all families or specific taxonomic groups are available on GitHub (<https://github.com/junhee-jung/MG-RAST-read-counter>, copy archived at Jung and Galanti, 2024).

In addition to the nine read groups selected from MG-RAST results, we also performed a highly confident mapping of exogenous reads to the *M. persicae* and *B. aphidicola* genome assemblies (Singh et al., 2021) ([https://www.ncbi.nlm.nih.gov/datasets/genome/GCF\\_001939165.1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_001939165.1)), to test whether mapping to a high-quality assembly of the exact pathogen has a higher sensitivity than MG-RAST. We mapped with BWA-MEM v0.7.17 (Li and Durbin, 2009), using a seed length of 25 bp (Robinson et al., 2017) and removing reads with MAPQ<20 and duplicates with MarkDuplicatesSpark (Van der Auwera et al., 2013; Poplin et al., 2018). We then counted all reads in the bam files.

Finally, we log transformed all read counts to approximate normality, and corrected for the total number of deduplicated reads by extracting residuals from the following linear model,  $\log(\text{read\_count} + 1) \sim \log(\text{deduplicated\_reads})$ , which allowed us to quantify non-*Thlaspi* loads, correcting for the sequencing depth of each sample.

## Exogenous reads heritability and species identification

To exclude the possibility that aphid and mildew infestation patterns were carried over from the field, through seed contamination or maternal effects, we used aphid and mildew presence/absence data collected in the field. We found no difference in aphid or mildew loads between samples with and without aphids or mildew on the original parental plant in the field (Figure 1—figure supplement 1). Nevertheless, to exclude a possible bias, we excluded one outlier sample with particularly high aphid load and aphids observed in the field (Figure 1—figure supplement 1) from the analyses.

To test for variation between populations we used a general linear model with population as a predictor. To measure SNP-based heritability, i.e., the proportion of variance explained by kinship, we used the `marker_h2()` function from the R package *heritability* (Kruijer and White, 2019), which uses a genetic distance matrix as predictor to compute REML-estimates of the genetic and residual variance. We used the same IBS matrix as for GWAS and for the correlations with climatic variables.

Even though MG-RAST classifies reads based on all taxonomic ranks, the accuracy of species identification of course strongly depends on the sequences available in the query databases. MG-RAST assigned our aphid reads to *A. pisum*, but this did not fit with our visual observations and with the poor performance of this species on Brassicaceae (Prince et al., 2014). We therefore selected three plausible aphid species and test which of these had mostly likely attacked our experiment. In addition to *A. pisum*, we tested two other aphid species commonly attacking Brassicaceae: *B. brassicae* and *M. persicae*. While not all three species have reference genomes available, all mitochondrial genomes are available on NCBI (NCBI, 1988) under accession numbers [MN232006](#), [NC\\_011594](#), and [NC\\_056270](#). We downloaded these sequences, aligned them to each other (Sievers et al., 2011), removed INDELS to retain only SNPs, and combined them into a single pseudo-reference fasta file (Supplementary file 3). We then mapped the exogenous reads from 40 randomly selected samples to this pseudo-reference, allowing for unique mappings only and counted the reads mapping to either of the three aphid species. We used the same approach for mildew except that we included only two possible species: *B. graminis*, as suggested by MG-RAST, and *E. cruciferarum* which is known to attack Brassicaceae but was not in the MG-RAST query database and seemed plausible from visual inspection (Figure 2B). For the mildew pseudo-reference (Supplementary file 4) we used the internal transcribed spacer, which is publicly available for both species on NCBI (NCBI, 1988) under accession numbers [MT644878](#) and [AF031283](#).

## Quantification of glucosinolates

Using seed material collected from the sequenced plants, we conducted a follow-up experiment to estimate the glucosinolate contents of all 207 lines in the absence of pathogens. Briefly, we sowed the seeds in Petri dishes, stratified them at 4°C in the dark for 2 weeks and transplanted the germinated seedlings to individual 9×9 cm<sup>2</sup> pots. We grew the plants in a growth chamber with a 14/10 hr light/dark cycle at 21/17°C and a relative humidity of ~45%. Two weeks after germination the plants were vernalized at 4°C for 2 more weeks in order to minimize phenological and developmental differences between winter and summer annuals. Ten days after vernalization, we collected the third or fourth true leaf and snap-froze it in liquid nitrogen. After freeze-drying, we weighed all samples and extracted the material threefold in 80% methanol, adding *p*-hydroxybenzyl glucosinolate (Phytoplan, Heidelberg, Germany) as internal standard. After centrifugation, we applied the supernatants onto ion-exchange columns with diethylaminoethyl Sephadex A25 (Sigma-Aldrich, St. Louis, MO, USA) in 0.5 M acetic acid buffer, pH 5. We added purified sulfatase, converting glucosinolates to desulfo glucosinolates. After 1 day, we eluted desulfo glucosinolates in water and analyzed them on a HPLC coupled to a DAD detector (HPLC-1200 Series, Agilent Technologies, Inc, Santa Clara, CA, USA) equipped with a Supelcosil LC 18 column (3 μm, 150×3 mm, Supelco, Bellefonte, PA, USA). We analyzed the samples with a gradient from water to methanol starting at 5% methanol for 6 min and then increased from 5% to 95% within 13 min with a hold at 95% for 2 min, followed by a column equilibration cycle. We identified different glucosinolates based on their retention times and UV spectra in comparison to respective standards and an in-house database. We integrated peaks at 229 nm and calculated respective glucosinolate concentrations in relation to the internal standard and sample dry mass, using response factors as reported by Agerbirk et al., 2015.

## Drivers of exogenous reads variation

To test for associations between glucosinolate variation, as well as climate of origin, and the observed pest loads, we extracted average bioclimatic variables for the 25 years predating our experiment for our 36 study populations from the Copernicus website (ECMWF, 2020), as described in Galanti et al., 2022. We then used the R package *lme4qtl* (Ziyatdinov et al., 2018) to run mixed models that included either bioclimatic variables or glucosinolate contents as explanatory variables, and the exogenous read counts as dependent variables, while correcting for population structure with the same IBS matrix as in GWA and EWA analyses (see below).

## GWA analysis

We conducted GWA with mixed models that corrected for population structure with a genetic IBS matrix as a random factor, as implemented in GEMMA (Zhou and Stephens, 2012). To obtain the IBS matrix we used PLINK v1.90b6.12 (Purcell et al., 2007). Starting from the imputed multisample vcf file obtained from variant calling, we pruned variants with LD>0.8 in 50 variants windows, sliding by five. To produce the genetic variants used for GWAS, we also started from the imputed multisample vcf file from variant calling and filtered out variants with MAF<0.04. As phenotypes we used the number of exogenous reads corrected for the total number of deduplicated reads, as described above.

To validate our results and test for overlap with existing gene functional annotations, we performed an enrichment analysis of variants neighboring a priori candidate genes as described in Atwell et al., 2010. Briefly, we attributed a priori candidate status to all variants located within 20 kb from orthologs (Emms and Kelly, 2019) of *A. thaliana* genes annotated with the GO term 'defense response' (GO:0006952), including nine genes similar to AtBSMT1 (Supplementary file 8). We then calculated enrichment for incremental  $-\log(p)$  thresholds as the ratio between observed frequency (significant a priori candidate/significant variants) and background frequency (total a priori candidate/total variants), and an upper bound for the FDR (Galanti et al., 2022; Atwell et al., 2010). We further assessed the significance of the enrichment through a previously established genome rotation scheme (Nordborg et al., 2005). Briefly, we calculated a null distribution of enrichments by randomly rotating the p-values and a priori candidate status of the genetic variants within each chromosome for 1 M permutations. We then assessed significance by comparing the observed enrichment at the Bonferroni threshold to the null distribution. The code for these analyses is available on <https://github.com/Dario-Galanti/multiphenogwas/tree/main/gemmaGWAS> (copy archived at Galanti, 2024d).

## Methylation and DMR calling

For the methylome analyses we used the EpiDiverse toolkit (Nunn et al., 2021), specifically designed for large WGBS datasets. We used the WGBS pipeline (<https://github.com/EpiDiverse/wgbs>; Nunn, 2022) for read mapping and methylation calling, retained only uniquely mapping reads longer than 30 bp, and obtained individual-sample bedGraph files for each sequence context. We then called DMRs using the DMR pipeline (Nunn et al., 2021), with a minimum coverage of 4x. We compared the 20 samples with the most and the least *M. persicae* and Eriysiphales loads, resulting in two sets of DMRs for each sequence context. Since this was only the first step of our methylation analysis, meant to identify potential regions of interest, we retained all DMRs with an FDR<20%. To understand the genomic preferences of DMRs, we intersected them with genomic features and calculated their densities in each by dividing their number by the total Mb covered by each genomic feature in the genome.

## EWA analysis

Following the DMR calling, we investigated methylation-phenotype relationships in more detail, using EWA. We ran EWA similarly to GWA, enabling the '-notsnp' option available in GEMMA (Zhou and Stephens, 2012), and correcting for population structure with the same IBS matrix. To exclude possible biases, we excluded all samples with a bisulfite non-conversion rate >1 (Galanti et al., 2022), which left 188 samples for analysis. To generate the methylation input files we first used custom scripts ([https://github.com/Dario-Galanti/WGBS\\_downstream/tree/main/WGBS\\_simpleworkflow](https://github.com/Dario-Galanti/WGBS_downstream/tree/main/WGBS_simpleworkflow), copy archived at Galanti, 2024c; Galanti et al., 2022) to unite individual-sample bedGraph files into unionbed files and retain positions with coverage>3 in at least 95% of the samples and a methylation difference of at least 5% in at least two samples. We then intersected the unionbed files with the DMRs of the corresponding sequence context using bedtools (Quinlan and Hall, 2010) and converted unionbed to BIMBAM format as input for GEMMA.

We ran EWA for individual positions within the DMRs and calculated Bonferroni thresholds based on the number of DMRs, assuming that cytosines within the same DMR are mostly autocorrelated. To observe in which genomic features associations with lower p-values were located, we performed enrichment analyses similar to the ones performed for defense a priori candidate genes in GWA (Atwell et al., 2010), but based on whole genomic features. Starting from all cytosines used for EWA, we calculated the background frequency as the fraction of all cytosines located in each genomic feature and then calculated the observed frequency in the same way for  $-\log(p)$  0.5 increments, with enrichment as the ratio of observed and expected frequencies. All code used for EWA and the

enrichment analysis in genomic features is available on <https://github.com/Dario-Galanti/EWAS/tree/main/gemmaEWAS> (copy archived at [Galanti, 2024e](#)).

## Acknowledgements

We thank the EpiDiverse network for its amazing support and discussions, in particular Adrián Contreras-Garrido, Bárbara Díez Rodríguez, Iris Sammarco, Adam Nunn, Daniela Ramos, and Anupoma Troyee for their close collaboration. We also thank Frank Reis for his feedback and expert tips for the project, Cecilia Heyworth for proofreading the manuscript and Karin Djendouci at Bielefeld University for support in glucosinolate analysis. We thank Peter Stadler at the University of Leipzig and David Langenberger from ecSeq, which helped with computing and hosted the EpiDiverse servers. The BinAC cluster is managed by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, and supported by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG.

## Additional information

### Funding

Funder	Grant reference number	Author
Horizon 2020 Framework Programme	764965	Oliver Bossdorf
Deutsche Forschungsgemeinschaft	401829393	Oliver Bossdorf

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Dario Galanti, Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Methodology, Writing – original draft, Writing – review and editing; Jun Hee Jung, Conceptualization, Data curation, Methodology, Writing – review and editing; Caroline Müller, Investigation, Writing – review and editing; Oliver Bossdorf, Conceptualization, Supervision, Funding acquisition, Methodology, Writing – original draft, Writing – review and editing

### Author ORCIDs

Dario Galanti [ORCID](https://orcid.org/0000-0002-6567-1505) <https://orcid.org/0000-0002-6567-1505>

Jun Hee Jung [ORCID](https://orcid.org/0000-0002-6523-4794) <https://orcid.org/0000-0002-6523-4794>

Caroline Müller [ORCID](https://orcid.org/0000-0002-8447-534X) <https://orcid.org/0000-0002-8447-534X>

Oliver Bossdorf [ORCID](https://orcid.org/0000-0001-7504-6511) <https://orcid.org/0000-0001-7504-6511>

### Peer review material

Reviewer #1 (Public review): <https://doi.org/10.7554/eLife.95510.3.sa1>

Reviewer #2 (Public review): <https://doi.org/10.7554/eLife.95510.3.sa2>

Author response <https://doi.org/10.7554/eLife.95510.3.sa3>

## Additional files

### Supplementary files

- Supplementary file 1. Classification of reads. Total reads, duplication rates, deduplicated raw reads, target, ambiguous, and exogenous reads and MG-RAST reads passing QC.
- Supplementary file 2. MG-RAST classification of exogenous reads. Results of MG-RAST classification of exogenous reads at the family level, using RefSeq as query database.
- Supplementary file 3. Competitive mapping pseudo-reference for aphid identification. Fasta file combining mitochondrial sequences of *B. Brassicaceae*, *A. pisum*, and *M. persicae*, with structural variants removed.

- Supplementary file 4. Competitive mapping pseudo-reference for mildew identification. Fasta file combining internal transcribed spacer (ITS) sequences of *B. graminis* and *E. cruciferarum*, with structural variants removed.
- Supplementary file 5. Results of competitive mapping for pest identification. Number of reads mapping uniquely to the pseudo-reference genomes of different aphid (mitochondrial DNA for either *B. Brassicaceae*, *A. pisum*, or *M. persicae*) or mildew species (internal transcribed spacer [ITS] sequence for either *B. graminis* or *E. cruciferarum*).
- Supplementary file 6. Exogenous reads used for downstream analyses. Classes of exogenous reads used in the analyses, including the nine groups from MG-RAST and two from mapping to the *M. persicae* and *B. aphidicola* reference genomes.
- Supplementary file 7. Quantification of glucosinolates. Glucosinolate concentrations in leaves of *T. arvense*, obtained from offspring of the sequenced plants, not infested by any herbivores nor pathogens.
- Supplementary file 8. A priori 'defense response' candidate genes used for the genome-wide association (GWA) enrichment analysis. List of *T. arvense* candidate genes used for the GWA enrichment analysis: orthologs of *A. thaliana* genes annotated with the GO term 'defense response'.
- Supplementary file 9. Differentially methylated regions (DMRs) based on *M. persicae* load. List of DMRs called between the 20 samples with the highest and the lowest *M. persicae* load.
- Supplementary file 10. Differentially methylated regions (DMRs) based on Erysiphales load. List of DMRs called between the 20 samples with the highest and the lowest Erysiphales load.
- MDAR checklist

#### Data availability

The seed material from the sequenced lines is available at the Nottingham Arabidopsis Stock Centre (NASC) under stock numbers N950001 to 950204. Genomic and bisulfite sequencing raw data are available on the ENA Sequence Read Archive (<https://www.ebi.ac.uk/ena>) under study accession number PRJEB50950. The reference genome and annotations were previously published by **Nunn et al., 2022**. GWA and EWA results in a format compatible with the Integrative Genomics Viewer (<https://www.igv.org/>) are available on Zenodo (<https://zenodo.org/records/10011535>). All code used in this study is available and documented on GitHub. The scripts for variant calling, filtering and imputation are on [https://github.com/Dario-Galanti/BinAC\\_varcalling](https://github.com/Dario-Galanti/BinAC_varcalling) (copy archived at **Galanti, 2024a**), and the scripts for the classification of sequencing reads and MG-RAST analysis are in [https://github.com/Dario-Galanti/Exoreads\\_treasure](https://github.com/Dario-Galanti/Exoreads_treasure) and <https://github.com/junhee-jung/MG-RAST-read-counter> respectively (copies archived at **Galanti, 2024b** and **Jung and Galanti, 2024**). The pipelines for methylation and DMR calling from WGBS data can be found on the EpiDiverse GitHub (<https://github.com/EpiDiverse>). The workflow for downstream analysis of methylation data is on [https://github.com/Dario-Galanti/WGBS\\_downstream/tree/main/WGBS\\_simpleworkflow](https://github.com/Dario-Galanti/WGBS_downstream/tree/main/WGBS_simpleworkflow) (copy archived at **Galanti, 2024c**). Finally, the scripts for running GWA and EWA analysis are on <https://github.com/Dario-Galanti/multiphenogwas/tree/main/gemmaGWAS> and <https://github.com/Dario-Galanti/EWAS/tree/main/gemmaEWAS> respectively (copies archived at **Galanti, 2024d** and **Galanti, 2024e**).

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Galanti D, Jung JH, Müller C, Bossdorf O	2023	Discarded sequencing reads uncover natural variation in pest resistance in <i>Thlaspi arvense</i>	<a href="https://doi.org/10.5281/zenodo.10011535">https://doi.org/10.5281/zenodo.10011535</a>	Zenodo, 10.5281/zenodo.10011535

The following previously published dataset was used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Galanti D, Ramos-Cruz D, Nunn A, Rodríguez-Arévalo I, Scheepens JF, Becker C, Bossdorf O	2022	Genetic and environmental drivers of large-scale epigenetic variation in <i>Thlaspi arvense</i>	<a href="https://www.ebi.ac.uk/ena/browser/view/PRJEB50950">https://www.ebi.ac.uk/ena/browser/view/PRJEB50950</a>	European Nucleotide Archive, PRJEB50950

## References

- Agerbirk N**, Olsen CE, Heimes C, Christensen S, Bak S, Hauser TP. 2015. Multiple hydroxyphenethyl glucosinolate isomers and their tandem mass spectrometric distinction in a geographically structured polymorphism in the crucifer *Barbarea vulgaris*. *Phytochemistry* **115**:130–142. DOI: <https://doi.org/10.1016/j.phytochem.2014.09.003>, PMID: 25277803
- Andrews S**. 2010. A quality control tool for high throughput sequence data | bioinformatics. Babraham Bioinformatics. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Annacondia ML**, Markovic D, Reig-Valiente JL, Scaltsoyiannes V, Pieterse CMJ, Ninkovic V, Slotkin RK, Martinez G. 2021. Aphid feeding induces the relaxation of epigenetic control and the associated regulation of the defense response in *Arabidopsis*. *The New Phytologist* **230**:1185–1200. DOI: <https://doi.org/10.1111/nph.17226>, PMID: 33475147
- Atwell S**, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Mulyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, et al. 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**:627–631. DOI: <https://doi.org/10.1038/nature08800>, PMID: 20336072
- Browning BL**, Zhou Y, Browning SR. 2018. A one-penny imputed genome from next-generation reference panels. *American Journal of Human Genetics* **103**:338–348. DOI: <https://doi.org/10.1016/j.ajhg.2018.07.015>, PMID: 30100085
- CABI**. 2021. *Myzus persicae* (green peach aphid). CABI Compendium. <https://www.cabidigitallibrary.org/doi/10.1079/cabicompendium.35642>
- Camacho C**, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**:421. DOI: <https://doi.org/10.1186/1471-2105-10-421>, PMID: 20003500
- Chan EKF**, Rowe HC, Kliebenstein DJ. 2010. Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. *Genetics* **185**:991–1007. DOI: <https://doi.org/10.1534/genetics.109.108522>, PMID: 19737743
- Chopra R**, Johnson EB, Emenecker R, Cahoon EB, Lyons J, Kliebenstein DJ, Daniels E, Dorn KM, Esfahanian M, Folstad N, Frels K, McGinn M, Ott M, Gallaher C, Altendorf K, Berroyer A, Ismail B, Anderson JA, Wyse DL, Umaslov T, et al. 2019. Progress toward the identification and stacking of crucial domestication traits in pennycress. *Plant Biology* **1**:609990. DOI: <https://doi.org/10.1101/609990>
- Colgan TJ**, Arce AN, Gill RJ, Ramos Rodrigues A, Kanteh A, Duncan EJ, Li L, Chittka L, Wurm Y. 2022. Genomic signatures of recent adaptation in a wild bumblebee. *Molecular Biology and Evolution* **39**:msab366. DOI: <https://doi.org/10.1093/molbev/msab366>, PMID: 35134226
- Corwin JA**, Copeland D, Feusier J, Subedy A, Eshbaugh R, Palmer C, Maloof J, Kliebenstein DJ. 2016. The quantitative basis of the *Arabidopsis* innate immune system to endemic pathogens depends on pathogen genetics. *PLOS Genetics* **12**:e1005789. DOI: <https://doi.org/10.1371/journal.pgen.1005789>, PMID: 26866607
- Danecek P**, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158. DOI: <https://doi.org/10.1093/bioinformatics/btr330>, PMID: 21653522
- Danecek P**, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* **10**:giab008. DOI: <https://doi.org/10.1093/gigascience/giab008>
- Dorn KM**, Fankhauser JD, Wyse DL, Marks MD. 2015. A draft genome of field pennycress (*Thlaspi arvense*) provides tools for the domestication of a new winter biofuel crop. *DNA Research* **22**:121–131. DOI: <https://doi.org/10.1093/dnares/dsu045>, PMID: 25632110
- ECMWF**. 2020. Copernicus Climate Change Service. E-OBS daily gridded meteorological data for Europe from 1950 to present derived from in-situ observations. <https://cds.climate.copernicus.eu/datasets/insitu-gridded-observations-europe?tab=overview> [Accessed February 9, 2022].
- Emms DM**, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**:238. DOI: <https://doi.org/10.1186/s13059-019-1832-y>, PMID: 31727128
- Espinás NA**, Saze H, Saijo Y. 2016. Epigenetic control of defense signaling and priming in plants. *Frontiers in Plant Science* **7**:1201. DOI: <https://doi.org/10.3389/fpls.2016.01201>, PMID: 27563304
- Ewels P**, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**:3047–3048. DOI: <https://doi.org/10.1093/bioinformatics/btw354>, PMID: 27312411
- Feuerborn TR**, Palkopoulou E, van der Valk T, von Seth J, Munters AR, Pečnerová P, Dehasque M, Ureña I, Ersmark E, Lagerholm VK, Krzewińska M, Rodríguez-Varela R, Götherström A, Dalén L, Díez-Del-Molino D. 2020. Competitive mapping allows for the identification and exclusion of human DNA contamination in ancient faunal genomic datasets. *BMC Genomics* **21**:844. DOI: <https://doi.org/10.1186/s12864-020-07229-y>, PMID: 33256612
- Frantzeskakis L**, Kracher B, Kusch S, Yoshikawa-Maekawa M, Bauer S, Pedersen C, Spanu PD, Maekawa T, Schulze-Lefert P, Panstruga R. 2018. Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. *BMC Genomics* **19**:381. DOI: <https://doi.org/10.1186/s12864-018-4750-6>, PMID: 29788921

- Frels K**, Chopra R, Dorn KM, Wyse DL, Marks MD, Anderson JA. 2019. Genetic diversity of field pennycress (*thlaspi arvense*) reveals untapped variability and paths toward selection for domestication. *Agronomy* **9**:302. DOI: <https://doi.org/10.3390/agronomy9060302>
- Friis G**, Smith EG, Lovelock CE, Ortega A, Marshall A, Duarte CM, Burt JA. 2024. Rapid diversification of grey mangroves (*Avicennia marina*) driven by geographic isolation and extreme environmental conditions in the Arabian Peninsula. *Molecular Ecology* **33**:e17260. DOI: <https://doi.org/10.1111/mec.17260>, PMID: 38197286
- Gabryś B**, Pawluk M. 1999. Acceptability of different species of Brassicaceae as hosts for the cabbage aphid. *Entomologia Experimentalis et Applicata* **91**:105–109. DOI: <https://doi.org/10.1046/j.1570-7458.1999.00471.x>
- Galanti D**, Ramos-Cruz D, Nunn A, Rodríguez-Arévalo I, Scheepens JF, Becker C, Bossdorf O. 2022. Genetic and environmental drivers of large-scale epigenetic variation in *Thlaspi arvense*. *PLOS Genetics* **18**:e1010452. DOI: <https://doi.org/10.1371/journal.pgen.1010452>, PMID: 36223399
- Galanti D**. 2024a. BinAC\_varcalling. swh:1:rev:b8473aa9b593e0553b10b4d7c065e378137a71eb. Software Heritage. [https://archive.softwareheritage.org/swh:1:dir:51cc61d8c08b8bde893dc155eb808e6d10bdb25b;origin=https://github.com/Dario-Galanti/BinAC\\_varcalling;visit=swh:1:snp:f53c52ca1905346386dd9f40ba4d16512873ad40;anchor=swh:1:rev:b8473aa9b593e0553b10b4d7c065e378137a71eb](https://archive.softwareheritage.org/swh:1:dir:51cc61d8c08b8bde893dc155eb808e6d10bdb25b;origin=https://github.com/Dario-Galanti/BinAC_varcalling;visit=swh:1:snp:f53c52ca1905346386dd9f40ba4d16512873ad40;anchor=swh:1:rev:b8473aa9b593e0553b10b4d7c065e378137a71eb)
- Galanti D**. 2024b. Exoreads\_treasure. swh:1:rev:e04d12f23171370e682380887166a0edd59b7f6f. Software Heritage. [https://archive.softwareheritage.org/swh:1:dir:c03b87531c6a1d2e0fe3717530e0a5e3c654d014;origin=https://github.com/Dario-Galanti/Exoreads\\_treasure;visit=swh:1:snp:1d33b86106dbf12eda22ec869a2779d8db28ecf6;anchor=swh:1:rev:e04d12f23171370e682380887166a0edd59b7f6f](https://archive.softwareheritage.org/swh:1:dir:c03b87531c6a1d2e0fe3717530e0a5e3c654d014;origin=https://github.com/Dario-Galanti/Exoreads_treasure;visit=swh:1:snp:1d33b86106dbf12eda22ec869a2779d8db28ecf6;anchor=swh:1:rev:e04d12f23171370e682380887166a0edd59b7f6f)
- Galanti D**. 2024c. WGBS\_downstream. swh:1:rev:393786b73a5caaa36ef3b76baeab90e832f3ba74. Software Heritage. [https://archive.softwareheritage.org/swh:1:dir:b2162c5579d0686d1621956dfb952a1447eb85c4;origin=https://github.com/Dario-Galanti/WGBS\\_downstream;visit=swh:1:snp:98754456d46cdcef0d883efca87a0961290303d5;anchor=swh:1:rev:393786b73a5caaa36ef3b76baeab90e832f3ba74](https://archive.softwareheritage.org/swh:1:dir:b2162c5579d0686d1621956dfb952a1447eb85c4;origin=https://github.com/Dario-Galanti/WGBS_downstream;visit=swh:1:snp:98754456d46cdcef0d883efca87a0961290303d5;anchor=swh:1:rev:393786b73a5caaa36ef3b76baeab90e832f3ba74)
- Galanti D**. 2024d. Multipheno\_GWAS. swh:1:rev:5b21ef1f45f51a16190fdb92b7fa43b84f69e3f5. Software Heritage. [https://archive.softwareheritage.org/swh:1:dir:b09c40d2a3e8e3e7d59eaf8131b16a5119725f8b;origin=https://github.com/Dario-Galanti/multipheno\\_GWAS;visit=swh:1:snp:abd9baab8ddd0bb37454712035ad8e078a9775d6;anchor=swh:1:rev:5b21ef1f45f51a16190fdb92b7fa43b84f69e3f5](https://archive.softwareheritage.org/swh:1:dir:b09c40d2a3e8e3e7d59eaf8131b16a5119725f8b;origin=https://github.com/Dario-Galanti/multipheno_GWAS;visit=swh:1:snp:abd9baab8ddd0bb37454712035ad8e078a9775d6;anchor=swh:1:rev:5b21ef1f45f51a16190fdb92b7fa43b84f69e3f5)
- Galanti D**. 2024e. EWAS. swh:1:rev:20928fc2e24ad6f7c2b8d2e117ce7e6ac8a672da. Software Heritage. <https://archive.softwareheritage.org/swh:1:dir:83ef651f6d57cd977526757c6bd9a25d0c020132;origin=https://github.com/Dario-Galanti/EWAS;visit=swh:1:snp:75b242395c8d563e43a48775ea0ade0b4f0efa59;anchor=swh:1:rev:20928fc2e24ad6f7c2b8d2e117ce7e6ac8a672da>
- Gao J**, Fang C, Zhao B. 2019. The latitudinal herbivory hypothesis revisited: To be part is to be whole. *Ecology and Evolution* **9**:3681–3688. DOI: <https://doi.org/10.1002/ece3.2759>, PMID: 31015958
- Gathercole LAP**, Nocchi G, Brown N, Coker TLR, Plumb WJ, Stocks JJ, Nichols RA, Denman S, Buggs RJA. 2021. Evidence for the widespread occurrence of bacteria implicated in acute oak decline from incidental genetic sampling. *Forests* **12**:1683. DOI: <https://doi.org/10.3390/f12121683>
- Geng S**, Kong X, Song G, Jia M, Guan J, Wang F, Qin Z, Wu L, Lan X, Li A, Mao L. 2019. DNA methylation dynamics during the interaction of wheat progenitor *Aegilops tauschii* with the obligate biotrophic fungus *Blumeria graminis* f. sp. *tritici*. *The New Phytologist* **221**:1023–1035. DOI: <https://doi.org/10.1111/nph.15432>, PMID: 30256420
- Geng Y**, Guan Y, Qiong L, Lu S, An M, Crabbe MJC, Qi J, Zhao F, Qiao Q, Zhang T. 2021. Genomic analysis of field pennycress (*Thlaspi arvense*) provides insights into mechanisms of adaptation to high elevation. *BMC Biology* **19**:143. DOI: <https://doi.org/10.1186/s12915-021-01079-0>, PMID: 34294107
- Habyarimana E**, Gorthy S, Baloch FS, Ercisli S, Chung G. 2022. Whole-genome resequencing of *Sorghum bicolor* and *S. bicolor* × *S. halepense* lines provides new insights for improving plant agroecological characteristics. *Scientific Reports* **12**:5556. DOI: <https://doi.org/10.1038/s41598-022-09433-0>, PMID: 35365708
- Hanson AA**, Lorenz AJ, Hesler LS, Bhusal SJ, Bansal R, Michel AP, Jiang G-L, Koch RL. 2018. Genome-wide association mapping of host-plant resistance to soybean aphid. *The Plant Genome* **11**:180011. DOI: <https://doi.org/10.3835/plantgenome2018.02.0011>, PMID: 30512046
- He Y**, Li Z. 2018. Epigenetic environmental memories in plants: establishment, maintenance, and reprogramming. *Trends in Genetics* **34**:856–866. DOI: <https://doi.org/10.1016/j.tig.2018.07.006>, PMID: 30144941
- Hu Y**, Wu X, Jin G, Peng J, Leng R, Li L, Gui D, Fan C, Zhang C. 2022. Rapid genome evolution and adaptation of *thlaspi arvense* mediated by recurrent rna-based and tandem gene duplications. *Frontiers in Plant Science* **12**:772655. DOI: <https://doi.org/10.3389/fpls.2021.772655>
- Jaouannet M**, Morris JA, Hedley PE, Bos JIB. 2015. Characterization of arabidopsis transcriptional responses to different aphid species reveals genes that contribute to host susceptibility and non-host resistance. *PLOS Pathogens* **11**:e1004918. DOI: <https://doi.org/10.1371/journal.ppat.1004918>, PMID: 25993686
- Jung JH**, Galanti D. 2024. MG RAST read counter. swh:1:rev:0cec5a5e8d15704480d4a797f6fac3bd8382f524. Software Heritage. <https://archive.softwareheritage.org/swh:1:dir:b4c2f2516718c2899ee43446aa5c4b6bdc4a9922;origin=https://github.com/junhee-jung/MG-RAST-read-counter;visit=swh:1:snp:7ca5bfc37135d76d3684e254ef193cd74a007d20;anchor=swh:1:rev:0cec5a5e8d15704480d4a797f6fac3bd8382f524>
- Kajiya-Kanegae H**, Nagasaki H, Kaga A, Hirano K, Ogiso-Tanaka E, Matsuoka M, Ishimori M, Ishimoto M, Hashiguchi M, Tanaka H, Akashi R, Isobe S, Iwata H. 2021. Whole-genome sequence diversity and association analysis of 198 soybean accessions in mini-core collections. *DNA Research* **28**:dsaa032. DOI: <https://doi.org/10.1093/dnares/dsaa032>, PMID: 33492369

- Kant MR**, Jonckheere W, Knecht B, Lemos F, Liu J, Schimmel BCJ, Villarroel CA, Ataíde LMS, Dermauw W, Glas JJ, Egas M, Janssen A, Van Leeuwen T, Schuurink RC, Sabelis MW, Alba JM. 2015. Mechanisms and ecological consequences of plant defence induction and suppression in herbivore communities. *Annals of Botany* **115**:1015–1051. DOI: <https://doi.org/10.1093/aob/mcv054>, PMID: 26019168
- Keegan KP**, Glass EM, Meyer F. 2016. MG-RAST, a metagenomics service for analysis of microbial community structure and function. *Methods in Molecular Biology* **1399**:207–233. DOI: [https://doi.org/10.1007/978-1-4939-3369-3\\_13](https://doi.org/10.1007/978-1-4939-3369-3_13), PMID: 26791506
- Kerwin R**, Feusier J, Corwin J, Rubin M, Lin C, Muok A, Larson B, Li B, Joseph B, Francisco M, Copeland D, Weinig C, Kliebenstein DJ. 2015. Natural genetic variation in *Arabidopsis thaliana* defense metabolism genes modulates field fitness. *eLife* **4**:e05604. DOI: <https://doi.org/10.7554/eLife.05604>, PMID: 25867014
- Kim JH**, Jander G. 2007. *Myzus persicae* (green peach aphid) feeding on *Arabidopsis* induces the formation of a deterrent indole glucosinolate. *The Plant Journal* **49**:1008–1019. DOI: <https://doi.org/10.1111/j.1365-3113.2006.03019.x>, PMID: 17257166
- Kinoshita T**, Seki M. 2014. Epigenetic memory for stress response and adaptation in plants. *Plant & Cell Physiology* **55**:1859–1863. DOI: <https://doi.org/10.1093/pcp/pcu125>, PMID: 25298421
- Klingauf F**, Şengonca Ç, Bennewitz H. 1972. Einfluß von sinigrin auf die nahrungsaufnahme polyphager und oligophager blattlausarten (aphididae) (effect of sinigrin on sucrose uptake by some polyphagous and oligophagous aphids (aphididae)). *Oecologia* **9**:53–57. DOI: <https://doi.org/10.1007/BF00345243>
- Kruijer W**, White I. 2019. heritability: Marker-based estimation of heritability using individual plant or plot data. CRAN. <https://cran.r-project.org/web/packages/heritability/index.html>
- Kutyniok M**, Müller C. 2012. Crosstalk between above- and belowground herbivores is mediated by minute metabolic responses of the host *Arabidopsis thaliana*. *Journal of Experimental Botany* **63**:6199–6210. DOI: <https://doi.org/10.1093/jxb/ers274>, PMID: 23045608
- Lämke J**, Bäurle I. 2017. Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. *Genome Biology* **18**:124. DOI: <https://doi.org/10.1186/s13059-017-1263-6>, PMID: 28655328
- Law JA**, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews. Genetics* **11**:204–220. DOI: <https://doi.org/10.1038/nrg2719>, PMID: 20142834
- Li H**, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. DOI: <https://doi.org/10.1093/bioinformatics/btp324>, PMID: 19451168
- Li H**. 2024. Seqtk. GitHub. <https://github.com/lh3/seqtk>
- Liu J**, He Z. 2020. Small dna methylation, big player in plant abiotic stress responses and memory. *Frontiers in Plant Science* **11**:595603. DOI: <https://doi.org/10.3389/fpls.2020.595603>, PMID: 33362826
- Martin M**. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* **17**:10. DOI: <https://doi.org/10.14806/ej.17.1.200>
- Mekbib Y**, Tesfaye K, Dong X, Saina JK, Hu GW, Wang QF. 2022. Whole-genome resequencing of *Coffea arabica* L. (Rubiaceae) genotypes identify SNP and unravels distinct groups showing a strong geographical pattern. *BMC Plant Biology* **22**:69. DOI: <https://doi.org/10.1186/s12870-022-03449-4>, PMID: 35164709
- Melotto M**, Zhang L, Oblessuc PR, He SY. 2017. Stomatal defense a decade later. *Plant Physiology* **174**:561–571. DOI: <https://doi.org/10.1104/pp.16.01853>, PMID: 28341769
- Metheringham CL**, Plumb WJ, Stocks JJ, Kelly LJ, Gorris MN, Moat J, Buggs RJA, Nichols RA. 2022. Rapid polygenic adaptation in a wild population of ash trees under a novel fungal epidemic. [bioRxiv]. DOI: <https://doi.org/10.1101/2022.08.01.502033>
- Meyer F**, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA. 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**:386. DOI: <https://doi.org/10.1186/1471-2105-9-386>, PMID: 18803844
- Muhammad T**, Zhang F, Zhang Y, Liang Y. 2019. RNA interference: a natural immune system of plants to counteract biotic stressors. *Cells* **8**:38. DOI: <https://doi.org/10.3390/cells8010038>, PMID: 30634662
- Nalam V**, Louis J, Patel M, Shah J. 2018. *Arabidopsis*-green peach aphid interaction: rearing the insect, no-choice and fecundity assays, and electrical penetration graph technique to study insect feeding behavior. *Bio-Protocol* **8**:e2950. DOI: <https://doi.org/10.21769/BioProtoc.2950>, PMID: 34395762
- NCBI**. 1988. *National Center for Biotechnology Information (NCBI)*. Bethesda (MD): National Library of Medicine (US).
- Nocchi G**, Brown N, Coker TLR, Plumb WJ, Stocks JJ, Denman S, Buggs RJA. 2022. Genomic structure and diversity of oak populations in British parklands. *PLANTS, PEOPLE, PLANET* **4**:167–181. DOI: <https://doi.org/10.1002/ppp3.10229>
- Nordborg M**, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall JD, Wang J, Zhao K, et al. 2005. The pattern of polymorphism in *Arabidopsis thaliana*. *PLOS Biology* **3**:e196. DOI: <https://doi.org/10.1371/journal.pbio.0030196>, PMID: 15907155
- Nunn A**, Can SN, Otto C, Fasold M, Díez Rodríguez B, Fernández-Pozo N, Rensing SA, Stadler PF, Langenberger D. 2021. EpiDiverse Toolkit: a pipeline suite for the analysis of bisulfite sequencing data in ecological plant epigenetics. *NAR Genomics and Bioinformatics* **3**:lqab106. DOI: <https://doi.org/10.1093/nargab/lqab106>
- Nunn A**. 2022. EpiDiverse-WGBS pipeline. 17d09d1. GitHub. <https://github.com/EpiDiverse/wgbs>

- Nunn A**, Rodríguez-Arévalo I, Tandukar Z, Frels K, Contreras-Garrido A, Carbonell-Bejerano P, Zhang P, Ramos Cruz D, Jandrasits K, Lanz C, Brusa A, Mirouze M, Dorn K, Galbraith DW, Jarvis BA, Sedbrook JC, Wyse DL, Otto C, Langenberger D, Stadler PF, et al. 2022. Chromosome-level *Thlaspi arvense* genome provides new tools for translational research and for a newly domesticated cash cover crop of the cooler climates. *Plant Biotechnology Journal* **20**:944–963. DOI: <https://doi.org/10.1111/pbi.13775>, PMID: 34990041
- Paczian T**, Trimble WL, Gerlach W, Harrison T, Wilke A, Meyer F. 2019. The MG-RAST API explorer: an on-ramp for RESTful query composition. *BMC Bioinformatics* **20**:561. DOI: <https://doi.org/10.1186/s12859-019-2993-0>, PMID: 31703549
- Poplin R**, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, Shakir K, Thibault J, Chandran S, Whelan C, Lek M, Gabriel S, Daly MJ, Neale B, MacArthur DG, Banks E. 2018. Scaling Accurate Genetic Variant Discovery to Tens of Thousands of Samples. [bioRxiv]. DOI: <https://doi.org/10.1101/201178v3>
- Prince D**, Mugford S, Vincent T, Hogenhout S. 2014. Pea aphid survival assays on *Arabidopsis thaliana*. *BIO-PROTOCOL* **4**:19. DOI: <https://doi.org/10.21769/BioProtoc.1251>
- Purcell S**, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**:559–575. DOI: <https://doi.org/10.1086/519795>, PMID: 17701901
- Quinlan AR**, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842. DOI: <https://doi.org/10.1093/bioinformatics/btq033>, PMID: 20110278
- Robinson KM**, Hawkins AS, Santana-Cruz I, Adkins RS, Shetty AC, Nagaraj S, Sadzewicz L, Tallon LJ, Rasko DA, Fraser CM, Mahurkar A, Silva JC, Dunning Hotopp JC. 2017. Aligner optimization increases accuracy and decreases compute times in multi-species sequence data. *Microbial Genomics* **3**:e000122. DOI: <https://doi.org/10.1099/mgen.0.000122>
- Roman-Reyna V**, Pinili D, Borja FN, Quibod IL, Groen SC, Alexandrov N, Mauleon R, Oliva R. 2020. Characterization of the leaf microbiome from whole-genome sequencing data of the 3000 rice genomes project. *Rice* **13**:72. DOI: <https://doi.org/10.1186/s12284-020-00432-1>, PMID: 33034758
- Roquis D**, Robertson M, Yu L, Thieme M, Julkowska M, Bucher E. 2021. Genomic impact of stress-induced transposable element mobility in *Arabidopsis*. *Nucleic Acids Research* **49**:10431–10447. DOI: <https://doi.org/10.1093/nar/gkab828>, PMID: 34551439
- Sangiovanni M**, Granata I, Thind AS, Guarracino MR. 2019. From trash to treasure: detecting unexpected contamination in unmapped NGS data. *BMC Bioinformatics* **20**:168. DOI: <https://doi.org/10.1186/s12859-019-2684-x>, PMID: 30999839
- Sievers F**, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**:539. DOI: <https://doi.org/10.1038/msb.2011.75>, PMID: 21988835
- Singh KS**, Cordeiro EMG, Troczka BJ, Pym A, Mackisack J, Mathers TC, Duarte A, Legeai F, Robin S, Bielza P, Burrack HJ, Charaabi K, Denholm I, Figueroa CC, Ffrench-Constant RH, Jander G, Margaritopoulos JT, Mazzoni E, Nauen R, Ramirez CC, et al. 2021. Global patterns in genomic diversity underpinning the evolution of insecticide resistance in the aphid crop pest *Myzus persicae*. *Communications Biology* **4**:847. DOI: <https://doi.org/10.1038/s42003-021-02373-x>, PMID: 34234279
- Sobota RS**, Shriner D, Kodaman N, Goodloe R, Zheng W, Gao Y-T, Edwards TL, Amos CI, Williams SM. 2015. Addressing population-specific multiple testing burdens in genetic association studies. *Annals of Human Genetics* **79**:136–147. DOI: <https://doi.org/10.1111/ahg.12095>, PMID: 25644736
- The Arabidopsis Information Resource (TAIR)**. 2000. protein\_coding. TAIR. <https://www.arabidopsis.org/locus?key=137911>
- Thoen MPM**, Davila Olivas NH, Kloth KJ, Coolen S, Huang P-P, Aarts MGM, Bac-Molenaar JA, Bakker J, Bouwmeester HJ, Broekgaarden C, Bucher J, Busscher-Lange J, Cheng X, Fradin EF, Jongsma MA, Julkowska MM, Keurentjes JJB, Ligterink W, Pieterse CMJ, Ruyter-Spira C, et al. 2017. Genetic architecture of plant stress resistance: multi-trait genome-wide association mapping. *The New Phytologist* **213**:1346–1362. DOI: <https://doi.org/10.1111/nph.14220>, PMID: 27699793
- Troyee AN**, Medrano M, Müller C, Alonso C. 2022. Variation in DNA methylation and response to short-term herbivory in *Thlaspi arvense*. *Flora* **293**:152106. DOI: <https://doi.org/10.1016/j.flora.2022.152106>
- Van der Auwera GA**, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics* **43**:11. DOI: <https://doi.org/10.1002/0471250953.bi1110s43>, PMID: 25431634
- War AR**, Paulraj MG, Ahmad T, Buhroo AA, Hussain B, Ignacimuthu S, Sharma HC. 2012. Mechanisms of plant defense against insect herbivores. *Plant Signaling & Behavior* **7**:1306–1320. DOI: <https://doi.org/10.4161/psb.21663>, PMID: 22895106
- Warwick SI**, Francis A, Susko DJ. 2002. The biology of Canadian weeds. 9. *Thlaspi arvense* L. (updated). *Canadian Journal of Plant Science* **82**:803–823. DOI: <https://doi.org/10.4141/P01-159>
- Wojtaszek P**. 1997. Oxidative burst: an early plant response to pathogen infection. *The Biochemical Journal* **322** (Pt 3):681–692. DOI: <https://doi.org/10.1042/bj3220681>, PMID: 9148737
- Zhang H**, Lang Z, Zhu JK. 2018. Dynamics and function of DNA methylation in plants. *Nature Reviews. Molecular Cell Biology* **19**:489–506. DOI: <https://doi.org/10.1038/s41580-018-0016-z>, PMID: 29784956

- Zhao R**, Yang X, Li M, Peng X, Wei M, Zhang X, Yang L, Li J. 2021. Biodiesel preparation from *Thlaspi arvense* L. seed oil utilizing a novel ionic liquid core-shell magnetic catalyst. *Industrial Crops and Products* **162**:113316. DOI: <https://doi.org/10.1016/j.indcrop.2021.113316>
- Zhou X**, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**:821–824. DOI: <https://doi.org/10.1038/ng.2310>, PMID: 22706312
- Ziyatdinov A**, Vázquez-Santiago M, Brunel H, Martínez-Perez A, Aschard H, Soria JM. 2018. lme4qt: linear mixed models with flexible covariance structure for genetic studies of related individuals. *BMC Bioinformatics* **19**:68. DOI: <https://doi.org/10.1186/s12859-018-2057-x>, PMID: 29486711
- Züst T**, Heichinger C, Grossniklaus U, Harrington R, Kliebenstein DJ, Turnbull LA. 2012. Natural enemies drive geographic variation in plant defenses. *Science* **338**:116–119. DOI: <https://doi.org/10.1126/science.1226397>, PMID: 23042895

## Chapter 6

### **General discussion**

In this thesis, I hoped to bridge the gap between microbiome and experimental ecological research by using the model plant *Arabidopsis thaliana*, the leaf bacterial isolate collection At-LSPHERE (Bai et al., 2015), and the naturally-occurring *Arabidopsis* pathogen *Pseudomonas viridiflava* (Karasov et al., 2018).

#### **Plant-microbiome relationship – it is all about perspective!**

In the second chapter of my thesis, I outlined the importance of studying plant-microbiome interactions, and how adopting either the plant or microbiome perspective can aid the investigation of complex ecological questions. I have applied this framework throughout my research by conducting experiments from both perspectives.

The plant perspective involves manipulating a plant's microbiome and observing the effects on plant performance, with the potential to also include environmental variation as a secondary factor. I carried out such a plant-perspective experiment with *Arabidopsis thaliana*, in which I manipulated microbiome composition and measured how different bacterial diversity levels affected plant growth and pathogen tolerance (Chapter 3). Microbiome perspective experiments manipulate plant variables, and then observe the effect on the plant microbiome, for example by changes in bacterial diversity and composition. In Chapter 4, I performed a microbiome-perspective experiment, in which plants of various genotypes were exposed to temperature fluctuations, and the resulting changes in microbiome were measured. Chapter 5 also partially reflected the microbiome perspective, as we extracted microbiome data from the results of an experiment that was not initially intended for microbiome research. We extracted microbiome-related reads from previous generated results from a whole genome sequencing project involving 207 *Thlaspi arvense* accessions. The non-host reads, which contained host microbiome, pathogens, herbivores, and herbivore microbiomes, were used to study host defense mechanisms.

#### **Microbiomes can be manipulated and used to test ecological theory**

The biodiversity ecosystem function (BEF) theory, in which an increase in species richness is linked to an increase in ecosystem function and stability, has been widely established across varied “macro”-ecosystems such as forests, grasslands, freshwater, and oceans (van der Plas, 2019). Recent studies have highlighted the importance of the plant microbiome in the functioning of these ecosystems, for example, Laforest-Lapointe et al. (2017) found that an increase in leaf bacterial diversity improved forest performance. Plants exist within the broader

context of an ecosystem, but they themselves are also host to complex communities of microbiomes. While the plant microbiome is often referred to as a singular unit, current research shows that plants are inhabited by different microbiomes according to tissue and organ (Turner et al., 2013; Dastogeer et al., 2020). Therefore, in Chapter 3, I treated a plant as its own “ecosystem” in terms of the BEF theory and conducted an experiment in which the effect of increased microbiome diversity on plant performance was measured in terms of growth and resistance to *Pseudomonas* pathogens.

The experiment showed that increased microbiome diversity, especially in strain richness (akin to species richness in other biodiversity experiments), increases plant growth and pathogen tolerance. These results were congruent to other findings about plant microbiomes in nature and in laboratory settings, that have shown how bacterial diversity is a key element to plant performance (Laforest-Lapointe et al., 2017; Vogel et al., 2021; Lepinay et al., 2024). Microbiome diversity manipulation research has shown the positive effect of the increase in soil microorganism diversity on community performance (Wagg et al., 2019, 2021), and leaf microbial diversity increases plant defense (Vogel et al., 2021) as well as plant performances in a forest (Laforest-Lapointe et al., 2017). In addition, in a grassland study plant productivity was linked to both higher total microbial diversity as well as the leaf microbiome diversity (Lepinay et al., 2024). Overall, my finding adds to the growing literature of BEF studies and highlights the importance of bacterial biodiversity on the plant.

Previous research has highlighted the importance of understanding microbe-microbe interaction as one of the key factors in shaping the microbiome (Almario et al., 2022; Schäfer et al., 2022). As the microbiome is important for plant function, and exists in a diverse-yet-conserved group (Schlaeppli et al., 2014; Yeoh et al., 2017), I expected the interactions to be mostly positive. However, I found that when both Alphaproteobacteria and Betaproteobacteria were present in the plant, I observed negative plant growth. This could be due to niche overlap, as the two strains could have both suffered from competition, leading to a decrease in community diversity and potentially subsequently affecting plant growth. In At-LSPHERE community, the community used in this experiment, when there was significant niche overlap between two closely-related strains, their interactions led to a decrease in bacterial abundance in one strain (Schäfer et al., 2023). This highlights the complexities of plant-microbe and plant-microbe-microbe relationships, and much more research is needed to truly understand these interactions.

This study could have been improved by measuring the final bacterial community diversity and composition. All of the diversity indices were based on the inoculum diversity, which assumes that all strains successfully colonized the plants at the same ratio as the inoculum, but current

knowledge on syncom is mixed regarding the effect of starting inoculum on the final microbiome composition. Bai et al. (2015) when using At-SPHERE found the final bacterial ratio is independent of the inoculum ratio, but Coker et al. (2022) showed that altering the ratio of initial strain mixes can have an impact on the resulting composition. Sequencing the final microbial communities, even with the low maximum diversity of 9 strains, and calculating the resulting diversity and composition directly would have given more specific results of how bacterial diversity affected the plant performance.

Nonetheless, this experiment successfully demonstrated that an increase in bacterial diversity resulted in an increase in the ecosystem (plant) function. It would be interesting to test, based on this setup, whether a positive feedback loop exists between bacterial diversity and plant fitness - would a plant with higher fitness in turn be able to support a more diverse microbial community?

### **Plant-associated microbiome change under fluctuating temperatures**

The relationship between plants and microbes predates 400 million years (Krings et al., 2007) and this association resulted in relatively stable microbiome communities, such as “common core” microbiomes (Risely, 2020), which can be brought together by hub or keystone microbes (Aglar et al., 2016; Risely, 2020; Rawstern et al., 2025). Despite this long-standing relationship, the effect of rapid environmental fluctuation on the microbiome is little known. Today, as global climate change is expected to make the environment less predictable, it is important to understand the role of the microbiome in host stress mitigation as these communities play a crucial role in plant function, as outlined in Chapter 3. In Chapter 4, I exposed plants and their microbiomes to several frequencies of temperature fluctuations, but importantly, the temperature sum received by each plant was equal regardless of the treatment.

One of the most interesting findings from Chapter 4 was that bacterial diversity was initially lowered by heat stress. The effect of warming on the host appears to lower alpha diversity across various taxa in a microbiome (Li et al., 2023), but some studies indicate that slight warming could increase bacterial diversity in the phyllosphere (Hoeffle et al., 2024; Fu et al., 2025). This discrepancy could arise from whether an increase in temperature is a stress, lowering plant fitness (Kranner et al., 2010), or not, especially when a slightly elevated temperature can actually increase plant performance (Springate and Kover, 2014). In my experiment, the plant growth analysis confirmed that the temperature fluctuation was indeed stressful for the plant, as growth was lowered, and microbiomes preferred a lower frequency of temperature fluctuations. These results agree with previous studies in which an increase in temperature significantly lowered microbiome diversity (Ren et al., 2015; Aydogan et al., 2018; Firrincieli et al., 2020).

I found that after this initial decline in diversity, bacterial communities were able to recover, but not to the same extent as communities in control plants. In addition, the microbiome under rapid fluctuation cycle was least able to recover from this loss in diversity. Based on these findings, I hypothesize that climate change could drive the loss of microbiome diversity in two ways: initial bacterial diversity loss from the heat stress and continued lower diversity if the fluctuation is too rapid, as bacterial diversity cannot recover over time. From Chapter 3, plant microbiome diversity is linked to plant function, especially in pathogen tolerance. This loss in diversity could leave plants unprepared for the next wave of stresses, unless the core microbiome fortuitously contains stress-resistant strains (Risely, 2020; Vogel et al., 2021; Emmenegger et al., 2023). These negative effects may not be universal, however, as bacterial diversity and the core taxa abundance tend to increase in plants over time, and perhaps older plant communities will be more resilient toward the coming environmental changes (Geyer et al., 2024; Li et al., 2024).

There were some aspects of the experiment that could have been improved. While I studied the difference in composition and diversity of the microbiome, I did not test how these changes affected the functions that the bacteria provide. If the experiment was carried out in a gnotobiotic manner, using a similar technique to Chapter 3, functional predictions could have been carried out (Huerta-Cepas et al., 2019). However, upon the analysis of pre-microbiome-inoculation data, I found that plants already contained highly diverse bacterial communities, and it was not possible to differentiate between the syncom strains and spontaneous inoculation, i.e. contamination. An attempt was made to predict the function solely based on barcode sequence (Douglas et al., 2020), but I could not find any significant associations between the bacterial function and the environmental treatment.

### **Reads, reuse, recycle: a microbiome perspective can be found from non-microbiome experiments**

As the cost of sequencing is rapidly decreasing (Sboner et al., 2011), more ecological knowledge is supplemented by genomic information. However, there are still several difficulties in sequencing, including developing technical know-how and the cost, as sequencing is still expensive when many samples are involved (Sboner et al., 2011; Kim et al., 2017; Schwarze et al., 2020). One way to combat these challenges is to use genomic information from previous experiments, as genomics studies could provide vast quantities of untouched microbiome data, especially when whole genome sequencing was involved. Despite the prevalence of whole genome sequencing (WGS) projects, a limited number of publications have involved successfully extracting microbiome reads (Sangiovanni et al., 2019; Roman-Reyna et al., 2020; Gathercole et al., 2021).

In Chapter 5, I used previously published sequences from non-microbiome projects to extract microbial reads (Galanti et al., 2022). I used a free online tool MG-RAST (Keegan et al., 2016) for the metagenomics as there were no associated costs and it did not require a great depth of microbiome-processing knowledge. I hoped this would remove the entry barrier to replicating the method, however, as the program is not actively maintained I would recommend using another microbiome pipeline such as Nephele: Microbiome Analysis (<https://nephele.niaid.nih.gov/>, Weber et al., 2018). This pipeline also supports assembly-free microbiome analysis, and for a discarded read analysis where the vast majority of reads are expected to be host associated, the assembly-free approach could be performed more reliably than the assembly-based metagenomics approach (Pérez-Cobas et al., 2020).

The microbiome reads extracted from this analysis included aphids, the aphid symbiont *Buchnera*, and powdery mildew (*Erysiphales*). This reflected the greenhouse conditions during the time of the experiment, as both aphids and powdery mildew were observed on *Thlaspi arvense* plants. Evidence from other aphid studies has indicated that aphids are not “stealthy” feeders and that they deposit *Buchnera* on host plants, (Byrd et al., 2023; Wolfgang et al., 2023) which could explain the symbiont presence. The origin of the aphid DNA probably stems from saliva deposition during the feeding process (Tjallingii, 2006; Thorpe et al., 2016). Correlations between the aphid and powdery mildew loads and the *Thlaspi* climate of origin suggested that plants from colder origins were less defended, as could be expected given that both pests prefer warmer climates (Hullé et al., 2010; Marçais and Desprez-Loustau, 2014), and that plants from those regions most likely evolved defenses against them. The discarded microbiome reads, when carefully sieved from the potentially contaminating reads, therefore provided insight into aphid feeding preference and effects of plant origin. More generally, this strategy can be used to learn more about other plant-herbivore or plant-pathogen systems without extra costs or consumables.

There are also some drawbacks to this approach of treating all discarded reads as metagenomic data. The DNA extraction method and the choice of commercial extraction kit is known to affect the microbiome results, with different extraction methods generating different taxonomic biases (Giangacomo et al., 2021; Nearing et al., 2021; Rauer et al., 2025). Nonetheless, DNA extraction kits designed for plant genomic DNA are effective at extracting the microbiome from *Arabidopsis thaliana* (Giangacomo et al., 2021). *Thlaspi* being a close relative to *Arabidopsis* (Nunn et al., 2022) makes it more probable that the DNA extraction method used could successfully extract the microbiome DNA. In addition, the experimental result may contain more noise due to the lack of rinsing or washing the samples. It is a common practice to have a series of washing or rinsing steps for the surface sterilization using a combination of sterile water, ethanol and/or low concentrations of bleach (Agler et al., 2016;

Lundberg et al., 2022). However, microbiome washing stages, especially when using bleach, appear to affect microbiome results (Yu et al., 2022), and useful information could have been lost during the washing process. Not washing the leaves could have provided for the general phyllosphere to be captured more completely in the microbiome analysis.

### **Reflections on methodologies**

Methods of studying microbiomes evolve quickly and constantly. I would like to highlight some techniques and experimental systems that were not present in my final experimental designs, but which nonetheless have great potential for future microbiome research. For example, in growing *Arabidopsis* Microboxes (SacO2, Belgium) can be used, a system that enables gas exchange while maintaining sterile conditions, thereby allowing a gnotobiotic experiment (Kremer et al., 2021). The main drawback for my experimental design in Chapter 3 was that taking individual pictures required working under the laminar flow cabinet which significantly added to time and workload during the imaging process.

I also found some difficulty in deciding on which substrate to use. While I used agar with MS medium in Chapter 3, it was not suitable for Chapter 4, as the high temperatures over time would desiccate the substrate. Calcined clay (Pfeilmeier et al., 2021) and soil in Microboxes were also not used, as they do not rehydrate evenly without oversaturation. Jiffy-7 (Jiffy Products, Norway, e.g. Kremer et al., 2021) peat and coconut-based pellet were also attempted, however, in later stages plants would topple over within the Microbox.

The final compromise used in Chapter 4 was to forgo a gnotobiotic system and instead minimize external microbiome sources by using double-autoclaved soil, sterilizing seeds, using standard sterile bench techniques, and harvesting plants under a laminar flow bench.

### **Conclusions and Outlook**

My thesis contributes to our knowledge of the interaction between the plants and microbiomes across different environments in the context of experimental ecology, with a focus on leaves and their associated microbiome. By incorporating both plant and microbiome perspectives throughout my experiments, I could better understand plant-microbiome relationships across different environments and stresses. Throughout this thesis I have used the latest methodological approaches to study the plant-microbiome relationship from an ecological perspective, such as creating synthetic communities and extracting microbiome reads from a previously published genomics study.

While the interaction between plant and microbiome is highly dynamic and complex, in Chapters 3 and 4, I found a link between bacterial diversity and plant performance. In Chapter 3, plants with higher bacterial diversity had higher growth and increased pathogen resistance,

and in Chapter 4, plants had lowered microbial diversity as well as reduced growth after temperature fluctuation stress. These findings from leaf bacterial communities support the relevance of biodiversity and ecosystem function theory in the plant-microbiome relationship. An ecosystem thrives with higher biodiversity, and previous research suggests that plant species richness contributes to ecosystem functioning. I found that bacterial diversity significantly contributes to the performance of a plant, which can itself be considered an ecosystem, as each is host to numerous microbial communities. Therefore, to research a broader ecosystem, we should also focus on the smallest components, i.e. the microbiomes, which will help us to study their host plants, and that will lead to a better understanding of the wider community that plants live in.

In Chapter 4 I found that plants under heat stress had lower bacterial diversity, and that rapid fluctuations in temperature can make a plant less productive. These results suggest that environmental change could lead to a decrease in plant performance due to a loss of functions previously provided by the microbial community. Under the increasing pressure that global climate change will bring to many ecosystems, including heat and pathogen stress, this loss of performance could have devastating consequences for plants. More studies are required to determine exactly the consequences that plants will likely face due to this leaf and other organ microbiome biodiversity loss. However, we still have a very limited understanding on the mechanisms behind this relationship, and further research with other sequencing techniques such as metatranscriptomics could provide further insight.

I also extracted microbiome data from discarded whole genome sequencing reads, providing a good example of possible further opportunities for microbiome research (Chapter 5). The field of ecological genomics is rapidly growing, and whole genome sequencing (WGS) offers a widest range of applications. The vast quantity of WGS data that is currently available could be, with a careful selection of the phenotype, a treasure trove of plant-microbiome data. Moreover, researchers should consider when designing a new genomics experiment, whether microbiome information can be gained alongside genomic information. This may require careful adjustments to plant harvesting and DNA extraction methods, but could significantly reduce the cost and carbon footprint associated with microbiome research.

As we learn more regarding the role of microbiome diversity and composition on the plant and *vice versa*, in future experiments, we should study microbiome function more directly. One of the methods available is WGS and, together with transcriptomic approaches, it could elucidate the precise function and mechanism behind interactions between plants and microbiomes under stress. Perhaps, just like with whole genome data, discarded reads from transcriptomic studies could also be used to extract microbiome-associated transcriptome information.

Overall, it is crucial to include microbiomes in plant experimental ecology research in order to better and more accurately understand the relationship between plants and their environments. The two perspectives can provide a simple outline for such experiments, by measuring the changes in plant performance after microbiomes are manipulated, or by analyzing changes in microbiomes when plant variables are altered. Whilst generating bacterial strain collections is difficult, once established they can be used to generate a wide range of microbiomes, and help answer questions from both perspectives. In addition, environmental factors can be incorporated to gain a richer understanding of the relationship between all three parties: plants, microbiomes, and the environment. Using these strategies, my thesis helps to answer growing demands to incorporate microbiome research into other spheres of biology, including agriculture, conservation, and improving plant resilience, areas of study that will be crucial in tackling climate change in the years to come (Trivedi et al., 2020; Peixoto et al., 2022; Shayanthan et al., 2022; Walling et al., 2025).

## References

- Agler, M. T., J. Ruhe, S. Kroll, C. Morhenn, S.-T. Kim, D. Weigel, and E. M. Kemen. 2016. Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. *PLoS Biology* 14: e1002352.
- Almario, J., M. Mahmoudi, S. Kroll, M. Agler, A. Placzek, A. Mari, and E. Kemen. 2022. The Leaf Microbiome of Arabidopsis Displays Reproducible Dynamics and Patterns throughout the Growing Season. *mBio* 13: e02825-21.
- Aydogan, E. L., G. Moser, C. Müller, P. Kämpfer, and S. P. Glaeser. 2018. Long-Term Warming Shifts the Composition of Bacterial Communities in the Phyllosphere of Galium album in a Permanent Grassland Field-Experiment. *Frontiers in Microbiology* 9.
- Bai, Y., D. B. Müller, G. Srinivas, R. Garrido-Oter, E. Potthoff, M. Rott, N. Dombrowski, et al. 2015. Functional overlap of the Arabidopsis leaf and root microbiota. *Nature* 528: 364–369.
- Byrd, D., M. Tran, J. R. Kenney, E. E. Wilson-Rankin, and K. E. Mauck. 2023. The aphid Myzus persicae (Hemiptera: Aphididae) acquires chloroplast DNA during feeding on host plants. *Environmental Entomology* 52: 900–906.
- Coker, J., K. Zhalnina, C. Marotz, D. Thiruppathy, M. Tjuanta, G. D'Elia, R. Hailu, et al. 2022. A Reproducible and Tunable Synthetic Soil Microbial Community Provides New Insights into Microbial Ecology. *mSystems* 7: e00951.
- Dastogeer, K. M. G., F. H. Tumpa, A. Sultana, M. A. Akter, and A. Chakraborty. 2020. Plant microbiome—an account of the factors that shape community composition and diversity. *Current Plant Biology* 23: 100161.
- Douglas, G. M., V. J. Maffei, J. R. Zaneveld, S. N. Yurgel, J. R. Brown, C. M. Taylor, C. Huttenhower, and M. G. I. Langille. 2020. PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology* 38: 685–688.
- Emmenegger, B., J. Massoni, C. M. Pestalozzi, M. Bortfeld-Miller, B. A. Maier, and J. A.

- Vorholt. 2023. Identifying microbiota community patterns important for plant protection using synthetic communities and machine learning. *Nature Communications* 14: 7983.
- Firriacieli, A., M. Khorasani, A. C. Frank, and S. L. Doty. 2020. Influences of Climate on Phyllosphere Endophytic Bacterial Communities of Wild Poplar. *Frontiers in Plant Science* 11.
- Fu, G., G. Zhang, and H. Zhou. 2025. Effects of long-term experimental warming on phyllosphere epiphytic bacterial and fungal communities of four alpine plants. *Journal of Integrative Agriculture* 24: 799–814.
- Galanti, D., D. Ramos-Cruz, A. Nunn, I. Rodríguez-Arévalo, J. F. Scheepens, C. Becker, and O. Bosssdorf. 2022. Genetic and environmental drivers of large-scale epigenetic variation in *Thlaspi arvense*. *PLOS Genetics* 18: e1010452.
- Gathercole, L. A. P., G. Nocchi, N. Brown, T. L. R. Coker, W. J. Plumb, J. J. Stocks, R. A. Nichols, et al. 2021. Evidence for the Widespread Occurrence of Bacteria Implicated in Acute Oak Decline from Incidental Genetic Sampling. *Forests* 12: 1683.
- Geyer, J. K., R. L. Grunberg, J. Wang, and C. E. Mitchell. 2024. Leaf age structures phyllosphere microbial communities in the field and greenhouse. *Frontiers in Microbiology* 15.
- Giangacomo, C., M. Mohseni, L. Kovar, and J. G. Wallace. 2021. Comparing DNA Extraction and 16S rRNA Gene Amplification Methods for Plant-Associated Bacterial Communities. *Phytobiomes Journal* 5: 190–201.
- Hoefle, D., M. Sommer, B. Wassermann, M. Faticov, D. Serra, G. Berg, A. J. M. Tack, and A. Abdelfattah. 2024. Oak seedling microbiome assembly under climate warming and drought. *Environmental Microbiome* 19: 62.
- Huerta-Cepas, J., D. Szklarczyk, D. Heller, A. Hernández-Plaza, S. K. Forslund, H. Cook, D. R. Mende, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* 47: D309–D314.
- Hullé, M., A. Cœur d'Acier, S. Bankhead-Dronnet, and R. Harrington. 2010. Aphids in the face of global changes. *Comptes Rendus Biologies* 333: 497–503.
- Karasov, T. L., J. Almario, C. Friedemann, W. Ding, M. Giolai, D. Heavens, S. Kersten, et al. 2018. *Arabidopsis thaliana* and *Pseudomonas* Pathogens Exhibit Stable Associations over Evolutionary Timescales. *Cell Host & Microbe* 24: 168-179.e4.
- Keegan, K. P., E. M. Glass, and F. Meyer. 2016. MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function. In F. Martin, and S. Uroz [eds.], *Microbial Environmental Genomics (MEG)*, 207–233. Springer, New York, NY.
- Kim, M., Y. Kim, L. Qian, and J. S. Song. 2017. TeachEnG: a Teaching Engine for Genomics. *Bioinformatics (Oxford, England)* 33: 3296–3298.
- Kranner, I., F. V. Minibayeva, R. P. Beckett, and C. E. Seal. 2010. What is stress? Concepts, definitions and applications in seed science. *New Phytologist* 188: 655–673.
- Kremer, J. M., R. Sohrabi, B. C. Paasch, D. Rhodes, C. Thireault, P. Schulze-Lefert, J. M. Tiedje, and S. Y. He. 2021. Peat-based gnotobiotic plant growth systems for *Arabidopsis* microbiome research. *Nature protocols* 16: 2450–2470.
- Krings, M., T. N. Taylor, H. Hass, H. Kerp, N. Dotzler, and E. J. Hermsen. 2007. Fungal endophytes in a 400-million-yr-old land plant: infection pathways, spatial distribution, and host responses. *New Phytologist* 174: 648–657.
- Laforest-Lapointe, I., A. Paquette, C. Messier, and S. W. Kembel. 2017. Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. *Nature* 546: 145–147.
- Lepinay, C., T. Větrovský, M. Chytrý, P. Dřevojan, K. Fajmon, T. Cajthaml, P. Kohout, and P. Baldrian. 2024. Effect of plant communities on bacterial and fungal communities in a Central European grassland. *Environmental Microbiome* 19: 42.
- Li, J., K. A. Bates, K. L. Hoang, T. E. Hector, S. C. L. Knowles, and K. C. King. 2023. Experimental temperatures shape host microbiome diversity and composition. *Global Change Biology* 29: 41–56.

- Li, J., M.-K. Jin, L. Huang, Z.-F. Liu, T. Wang, R.-Y. Chang, M. Op de Beeck, et al. 2024. Assembly and succession of the phyllosphere microbiome and nutrient-cycling genes during plant community development in a glacier foreland. *Environment International* 187: 108688.
- Lundberg, D. S., R. de Pedro Jové, P. Pramoj Na Ayutthaya, T. L. Karasov, O. Shalev, K. Poersch, W. Ding, et al. 2022. Contrasting patterns of microbial dominance in the *Arabidopsis thaliana* phyllosphere. *Proceedings of the National Academy of Sciences* 119: e2211881119.
- Marçais, B., and M.-L. Desprez-Loustau. 2014. European oak powdery mildew: impact on trees, effects of environmental factors, and potential effects of climate change. *Annals of Forest Science* 71: 633–642.
- Nearing, J. T., A. M. Comeau, and M. G. I. Langille. 2021. Identifying biases and their potential solutions in human microbiome studies. *Microbiome* 9: 113.
- Nunn, A., I. Rodríguez-Arévalo, Z. Tandukar, K. Frels, A. Contreras-Garrido, P. Carbonell-Bejerano, P. Zhang, et al. 2022. Chromosome-level *Thlaspi arvense* genome provides new tools for translational research and for a newly domesticated cash cover crop of the cooler climates. *Plant Biotechnology Journal* 20: 944–963.
- Peixoto, R. S., C. R. Voolstra, M. Sweet, C. M. Duarte, S. Carvalho, H. Villela, J. E. Lunshof, et al. 2022. Harnessing the microbiome to prevent global biodiversity loss. *Nature Microbiology* 7: 1726–1735.
- Pérez-Cobas, A. E., L. Gomez-Valero, and C. Buchrieser. 2020. Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. *Microbial Genomics* 6: mgen000409.
- Pfeilmeier, S., G. C. Petti, M. Bortfeld-Miller, B. Daniel, C. M. Field, S. Sunagawa, and J. A. Vorholt. 2021. The plant NADPH oxidase RBOHD is required for microbiota homeostasis in leaves. *Nature Microbiology* 6: 852–864.
- van der Plas, F. 2019. Biodiversity and ecosystem functioning in naturally assembled communities. *Biological Reviews* 94: 1220–1245.
- Rauer, L., A. De Tomassi, C. L. Müller, C. Hülpmusch, C. Traidl-Hoffmann, M. Reiger, and A. U. Neumann. 2025. De-biasing microbiome sequencing data: bacterial morphology-based correction of extraction bias and correlates of chimera formation. *Microbiome* 13: 38.
- Rawstern, A. H., D. J. Hernandez, and M. E. Afkhami. 2025. Central Taxa Are Keystone Microbes During Early Succession. *Ecology Letters* 28: e70031.
- Ren, G., C. Zhu, M. S. Alam, T. Tokida, H. Sakai, H. Nakamura, Y. Usui, et al. 2015. Response of soil, leaf endosphere and phyllosphere bacterial communities to elevated CO<sub>2</sub> and soil temperature in a rice paddy. *Plant and Soil* 392: 27–44.
- Risely, A. 2020. Applying the core microbiome to understand host–microbe systems. *Journal of Animal Ecology* 89: 1549–1558.
- Roman-Reyna, V., D. Pinili, F. N. Borja, I. L. Quibod, S. C. Groen, N. Alexandrov, R. Mauleon, and R. Oliva. 2020. Characterization of the Leaf Microbiome from Whole-Genome Sequencing Data of the 3000 Rice Genomes Project. *Rice* 13: 72.
- Sangiovanni, M., I. Granata, A. S. Thind, and M. R. Guarracino. 2019. From trash to treasure: detecting unexpected contamination in unmapped NGS data. *BMC Bioinformatics* 20: 168.
- Sboner, A., X. J. Mu, D. Greenbaum, R. K. Auerbach, and M. B. Gerstein. 2011. The real cost of sequencing: higher than you think! *Genome Biology* 12: 125.
- Schäfer, M., A. R. Pacheco, R. Künzler, M. Bortfeld-Miller, C. M. Field, E. Vayena, V. Hatzimanikatis, and J. A. Vorholt. 2023. Metabolic interaction models recapitulate leaf microbiota ecology. *Science* 381: eadf5121.
- Schäfer, M., C. M. Vogel, M. Bortfeld-Miller, M. Mittelviehhaus, and J. A. Vorholt. 2022. Mapping phyllosphere microbiota interactions in planta to establish genotype-phenotype relationships. *Nature microbiology* 7: 856–867.
- Schlaeppli, K., N. Dombrowski, R. G. Oter, E. Ver Loren van Themaat, and P. Schulze-Lefert. 2014. Quantitative divergence of the bacterial root microbiota in *Arabidopsis*

- thaliana relatives. *Proceedings of the National Academy of Sciences* 111: 585–592.
- Schwarze, K., J. Buchanan, J. M. Fermont, H. Dreau, M. W. Tilley, J. M. Taylor, P. Antoniou, et al. 2020. The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. *Genetics in Medicine* 22: 85–94.
- Shayanthan, A., P. A. C. Ordoñez, and I. J. Oresnik. 2022. The Role of Synthetic Microbial Communities (SynCom) in Sustainable Agriculture. *Frontiers in Agronomy* 4.
- Springate, D. A., and P. X. Kover. 2014. Plant responses to elevated temperatures: a field study on phenological sensitivity and fitness responses to simulated climate warming. *Global Change Biology* 20: 456–465.
- Thorpe, P., P. J. A. Cock, and J. Bos. 2016. Comparative transcriptomics and proteomics of three different aphid species identifies core and diverse effector sets. *BMC Genomics* 17: 172.
- Tjallingii, W. F. 2006. Salivary secretions by aphids interacting with proteins of phloem wound responses. *Journal of Experimental Botany* 57: 739–745.
- Trivedi, P., J. E. Leach, S. G. Tringe, T. Sa, and B. K. Singh. 2020. Plant–microbiome interactions: from community assembly to plant health. *Nature Reviews Microbiology* 18: 607–621.
- Turner, T. R., E. K. James, and P. S. Poole. 2013. The plant microbiome. *Genome Biology* 14: 209.
- Vogel, C. M., D. B. Potthoff, M. Schäfer, N. Barandun, and J. A. Vorholt. 2021. Protective role of the Arabidopsis leaf microbiota against a bacterial pathogen. *Nature Microbiology* 6: 1537–1548.
- Wagg, C., Y. Hautier, S. Pellkofer, S. Banerjee, B. Schmid, and M. G. van der Heijden. 2021. Diversity and asynchrony in soil microbial communities stabilizes ecosystem functioning. D. Weigel, and A. Mori [eds.],. *eLife* 10: e62813.
- Wagg, C., K. Schlaeppli, S. Banerjee, E. E. Kuramae, and M. G. A. van der Heijden. 2019. Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning. *Nature Communications* 10: 4841.
- Walling, L. K., M. H. Gamache, R. A. González-Pech, V. J. Harwood, A. Ibrahim-Hashim, J. H. Jung, D. B. Lewis, et al. 2025. Incorporating microbiome analyses can enhance conservation of threatened species and ecosystem functions. *Science of The Total Environment* 970: 178826.
- Weber, N., D. Liou, J. Dommer, P. MacMenamin, M. Quiñones, I. Misner, A. J. Oler, et al. 2018. Nephele: a cloud platform for simplified, standardized and reproducible microbiome data analysis. *Bioinformatics* 34: 1411–1413.
- Wolfgang, A., A. J. M. Tack, G. Berg, and A. Abdelfattah. 2023. Reciprocal influence of soil, phyllosphere, and aphid microbiomes. *Environmental Microbiome* 18: 63.
- Yeoh, Y. K., P. G. Dennis, C. Paungfoo-Lonhienne, L. Weber, R. Brackin, M. A. Ragan, S. Schmidt, and P. Hugenholtz. 2017. Evolutionary conservation of a core root microbiome across plant phyla along a tropical soil chronosequence. *Nature Communications* 8: 215.
- Yu, Y., Z. Chen, H. Xie, X. Feng, Y. Wang, and P. Xu. 2022. Overhauling the Effect of Surface Sterilization on Analysis of Endophytes in Tea Plants. *Frontiers in Plant Science* 13.

## Acknowledgements

Perhaps before I start, my love for research came partially from my amazing Bachelor thesis supervisors: two then-PhD students Aoife and Helen. They showed me how fun an experiment and investigation can be, and demonstrated the patience that science and dealing with a Bachelor's student requires. Even if my ICP-MS result was a bit silly, my love for *Arabidopsis* and genomics started from there - I hope I was fractionally as good of a PhD student as you both were.

I have spent almost six years studying microbiome - and one of my earliest memories is Cee asking me whether the word microbiome is singular or plural, i.e. do you say *A* microbiome *THE* microbiome/ microbiomes? As someone whose mother tongue does not have the concept of articles, this sowed a seed of doubt and I can confidently say that I still have no idea, but my personal view is that it should not have an article, just like "water" - can you count a community that is constantly shifting?

The office I used more than half of my twenties is a truly fascinating place - sometimes it can be seemingly the worst, with the temperature shooting up to 30+ degrees and loud ventilator whooshing noises disturbing your work. However, most days I was grateful that 1 minute away is a fabulous view of the arboretum, where the seasons are reflected in the colour of the trees' leaves, and loud frogs offer their spring choir (and sometimes you can hear roosters - I never found out where they were). To be able to work here with so much beauty nearby not many can boast, it was a blessing.

With that, thank you Oliver – thank you for giving me the opportunity to work in such an amazing space. We had our very stark disagreements, for example you enjoyed clear broth soup from Mensa while I enjoyed creamy (and tastier) soups - and you liked extra strong espresso while I much preferred extra-long coffee. Despite this, I learned so much from you, thank you for welcoming whatever new ideas I came up with, and thank you for all the patience you showed throughout my PhD journey.

Thank you also to Madalin for sharing the joys of beautiful statistics. We also both enjoyed riding bicycles and share a passion for DIYing - especially the water purifier. I will cherish all the stats you have ever taught me, and I will ignore your love for Ballistol on everything. Also thank you for offering me a chance to help in the Plant Evolutionary Ecology course - it was a unique and fun experience.

I met many wonderful scientists who I had the privilege of learning from during my PhD. Thanks to Ale in MPI who taught me *Pseudomonas* work, Ana and Natalie who helped with library prep and answered an infinite number of my questions, Sebastian and Prof. Dr. Julia Vorholt who taught me SynCom work, Dario E and Andreas G for making my time in DECRyPT meeting fun. And to all the plant evolutionary ecology people for teaching me that science is best done together. Whether it was just discussing science, going to Mensa together, or working together to solve a problem, it was a blast. Big thank you to the technicians Christianne, Eva, Julia, Sabine, and Martina who made my experiments go ever so slightly smoother. And somewhat apologies for always seemingly keeping things rather messy. Thanks also to all the FöJs that always brought new dynamics to the group. I

had the privilege of supervising the Bachelor theses of Nina P and Johannes K, and I think from them I gained more knowledge than I ever gave. I am grateful that I got to work with two very deserving scientists.

And thank you to Frank and Dario, for making the PhD journey a lot more fun and keeping me sane and healthy. Dario for accepting that I cannot spell *Thlaspi* very well and still tolerating me enough to produce a paper together. Frank, well for spending the past 6 years opposite to each other 10-5.30 sometimes more, and travelling to conferences together with the inevitable train delays. For all the other silly projects with dubious scientific purposes, like building a snowman. A PhD feels like a marathon (not that I have run one) and to have a friend more so than just a colleague running next to you, struggling together, and finishing the race together regardless of the position, it gives me great satisfaction and joy. And thank you for translating my summary into German - turns out I should have been working on my formal-scientific German as well.

All the new and old friends: Duncan, Phil(ip) and Michael for your friendship since university and all the gaming sessions throughout COVID and beyond. And to my new friends in Germany - without them my life here will be like winter, tolerable but a bit bleak - especially Bessie (best dog, a bundle of joy) and my neighbours Hanna and Basti for providing many board game moments (mostly fun, some confusing).

To all my in-laws who I spent many of my Christmases with, and who accepted their fate of having a third brother-in-law with a PhD for some reason. Special thank you to Catherine (Gran-inlaw?) for knitting me a wonderful *Arabidopsis* jumper which was based on an image of *Thlaspi*.

To my mum, dad, brother, sister-in-law and 2 adorable nephews, I have nothing to say except sorry for not visiting enough, and for being a neglectful uncle. Thank you for always being there for me whenever I missed food from home and for any (im)material assistance. I am also grateful to my grandparents, who were always very proud and supportive of every academic decision I made in life.

Lastly, to Cee, the best naturalist I know. No writing or acknowledgement can suffice in my expression of gratitude. Without whom this thesis would not have been possible and my adventure to Germany not even have started at all. For being happier, sadder, and angrier than me whenever I told you about my day. With you "I am the happiest creature in the world. Perhaps other people have said so before, but not one with such justice<sup>1</sup>" and forever my "secret standard of perfection<sup>2</sup>" in a person.

<sup>1</sup> Jane Austin, *Pride and Prejudice*, 1813

<sup>2</sup> Jane Austin, *Sense and sensibilities*, 1811