# Elucidating the genetic basis and environmental regulation of root-microbe associations in maize

**Dissertation**

zur Erlangung des Grades

Doktorin der Agrarwissenschaften (Dr. agr.)

der Landwirtschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

von

**Danning Wang**

aus

Liaoning, China

Bonn, 2024

Referent: Prof. Dr. Frank Hochholdinger

Korreferent: Prof. Dr. Mika Tarkka

Tag der mündlichen Prüfung: 19.04.2024


Angefertigt mit Genehmigung der Landwirtschaftlichen Fakultät der Universität Bonn

# Table of contents

# List of figures

# List of supplementary information for Chapter 2

(EG_BLUP) across treatments.

# List of supplementary information for Chapter 3

# Abbreviations

| | |
|---|---|
| **AMF** | Arbuscular mycorrhizal fungi |
| **ASV** | Amplicon sequence variant |
| **DAO** | Differential abundance OTU |
| **DEG** | Differentially expressed gene |
| **GO** | Gene ontology |
| **GWAS** | Genome-wide association study |
| **IAA** | Indole-3-acetic acid |
| **KEGG** | Kyoto Encyclopedia of Genes and Genomes |
| **LR** | Lateral roots |
| *lrt1* | *Lateral rootless 1* |
| **MTAs** | Marker-trait associations |
| **OTU** | Operational taxonomic unit |
| **PCA** | Principal component analysis |
| **PCoA** | Principal coordinate analysis |
| **PERMANOVA** | Permutational multivariate analysis of variance |
| **PGPR** | Plant growth-promoting rhizobacteria |
| **PR** | Primary roots |
| **QTL** | Quantitative trait locus |
| **RA** | Relative abundance |
| **RF** | Random forest |
| **RNA** | Ribonucleic acid |
| *rtcs* | *Rootless concerning crown and seminal roots* |
| *rth3* | *Roothairless 3* |
| *rth5* | *Roothairless 5* |
| *rth6* | *Roothairless 6* |
| *rum1* | *Rootless with undetectable meristems 1* |
| **SNP** | Single nucleotide polymorphism |
| **SparCC** | Sparse correlations for compositional data |
| **WGCNA** | Weighted correlation network analysis |
| **WT** | Wild type |

# Abstract/Zusammenfassung

## Abstract

Microorganisms play a critical role in promoting plant growth and performance, especially under environmental stresses. Specific root-associated microbes have been demonstrated to influence root system development, regulate plant nutrient homeostasis and protect hosts against biotic and abiotic stresses. Utilization of microbiomes has been proposed as a strategy to improve food production and support sustainable agroecosystems. However, the question how the genetic framework underlying maize root development influences its microbiome assembly across different environmental conditions and to what extent the microbiome influences host performance remains largely unknown, especially at the population level. In particular, the degree to which the host function affects the abundance and enrichment of specific microbes remains obscure. Better understanding the genetic basis and environmental regulation of host-microbe interactions may promote crop performance and resilience in the context of climate change.

In the first part of this thesis, we characterized the root and rhizosphere microbiome composition of 129 diverse *Zea* accessions (teosintes, landraces, inbred lines and hybrids) grown under control, nitrogen-, phosphorus- and water-limited conditions. Biostatistics and co-variant analyses demonstrated that the genotype had a larger impact on the rhizosphere than root microbiome under abiotic stresses. Genomic and environmental prediction models indicated that environmental factors of the native region where the maize genotypes are originally from improved the prediction accuracy of specific microbiome abundances under phytochamber conditions. The allelic variation of one of significant SNPs S4_10445603 identified by environmental genome-wide association analyses was linked to both the predicted abundance of the keystone bacteria *Massilia* and the availability of total soil nitrogen in their source environments where the maize landrace germplasm was collected. Moreover, we identified a novel gene (Zm00001d048945) encoding microtubule organization processes near the SNP S4_10445603 and validated it with independent Mu-transposon insertion mutants with lateral root defects, reflecting a causal linkage of lateral root development and enrichment of *Massilia* in two independent nitrogen-poor soil experiments. Furthermore, root inoculation experiments using specific bacterial isolates demonstrated that *Massilia* alone contributed to lateral root development, and shoot biomass promotion under low nitrogen conditions.

In the second part of this thesis, to better understand whether gene expression is associated with

enrichment of specific microbes underlying lateral root development, we characterized host transcriptome and bacterial community composition across different root compartments using a diverse panel of root type mutants (e.g. lateral root and root hair mutants) in maize. Integrated transcriptomic and microbial data analyses demonstrated that mutations affecting lateral root development had the largest effect on host gene expression and microbiome assembly, as compared to mutations affecting other root types. Further network association analyses demonstrated that the keystone bacteria *Massilia* in lateral roots are associated with root functional genes involved in flowering and overall plant biomass production. A further experiment validated that the interactions of *Massilia* with genes functioning in reproduction was driven by developmental stages. Taking advantage of microbial inoculation experiments using a maize early flowering mutant, we confirmed that *Massilia*-driven maize growth promotion indeed depends on flowering time.

In conclusion, specifically selected microbes by host genotype and environmental factors can establish beneficial interactions with their host plants. These beneficial microbes altered root architecture at early stages and later stages of development e.g. flowering time to promote plant growth and performance especially when facing nitrogen deficient stresses. These coherent findings provide strong genetic linkage and breeding potential to improve plant performance and resilience in future low-input agroecosystems.

## Zusammenfassung

Mikroorganismen spielen eine entscheidende Rolle bei der Förderung des Pflanzenwachstums und der Pflanzenleistung, insbesondere unter biotischem und abiotischem Stress. Es wurde gezeigt, dass bestimmte wurzelassoziierte Mikroorganismen die Entwicklung des Wurzelsystems beeinflussen, die Nährstoffhomöostase der Pflanze regulieren und den Wirt vor biotischem und abiotischen Stress schützen. Die Verwendung von Mikrobiomen gilt als Strategie zur Verbesserung der Lebensmittelproduktion und zur Unterstützung nachhaltiger Agrarökosysteme. Allerdings ist die Frage, wie der genetische Rahmen, der der Maiswurzelentwicklung zugrunde liegt, den Aufbau des Mikrobioms unter verschiedenen Umweltbedingungen beeinflusst und inwieweit das Mikrobiom die Leistung des Wirts beeinflusst, noch weitgehend unbekannt, insbesondere auf Populationsebene.Vor allem ist unklar, inwieweit der Wirt die Häufigkeit und Anreicherung bestimmter Mikroorganismen beeinflussen kann. Ein besseres Verständnis der genetischen Grundlagen und der Umweltregulierung der Wechselwirkungen zwischen Wirt und Mikroorganismen kann künftig die Leistung und Widerstandsfähigkeit von Nutzpflanzen im Kontext des Klimawandels fördern.

Im ersten Teil dieser Arbeit wurde die Zusammensetzung des Wurzel- und Rhizosphären-Mikrobioms von 129 verschiedenen Maislinien (Teosinte, Landrassen, Inzuchtlinien und Hybriden) charakterisiert, die unter kontrollierten, Stickstoff- und Phosphormangel sowie unter wasserlimitierten Bedingungen angezogen wurden. Biostatistik- und Kovariantenanalysen zeigten, dass der Mais Genotyp unter abiotischem Stress einen größeren Einfluss auf die Rhizosphäre hat als das Wurzelmikrobiom. Genom- und Umweltvorhersagemodelle zeigten, dass Umweltfaktoren der Region, aus der die Maisgenotypen ursprünglich stammen, die Vorhersagegenauigkeit der spezifischen Mikrobiomhäufigkeit unter kontrollierten Bedingungen verbesserten. Die durch umweltbezogene genomweite Assoziationsanalyse identifizierte allelische Variation des Top-SNP steht in Zusammenhang sowohl mit der vorhergesagten Häufigkeit des Schlüsselbakteriums *Massilia*, als auch mit der Verfügbarkeit des Gesamtstickstoffs im Boden in der Ursprungsumgebung des gesammelten Keimplasmas von Mais-Landrassen. Darüber hinaus haben wir ein neues Gen (Zm00001d048945) identifiziert, das für Prozesse der Mikrotubuli-Organisation in der Nähe des SNP S4_10445603 kodiert, und es mit unabhängigen Mu-Transposon-Insertionsmutanten mit Seitenwurzeldefekten validiert, was einen kausalen Zusammenhang zwischen Seitenwurzelentwicklung und Anreicherung von *Massilia* in zwei unabhängigen Experimenten mit stickstoffarmen Böden widerspiegelt. Darüber hinaus zeigten Inokulierungsversuche von

Wurzeln mit spezifischen Bakterienisolaten, dass allein *Massilia* zur Lateralwurzelentwicklung und zur Förderung der Sprossbiomasse unter Bedingungen mit niedrigem Stickstoffgehalt beitrug.

Um besser zu verstehen, ob die Genexpression mit der Anreicherung spezifischer Mikroben, die der Seitenwurzelentwicklung zugrunde liegen, zusammenhängt, haben wir im zweiten Teil dieser Arbeit das Wirts-Transkriptom und die Zusammensetzung der bakteriellen Gemeinschaft in verschiedenen Wurzelkompartimenten unter Verwendung einer Reihe von Wurzeltyp-Mutanten (z. B. Seitenwurzel- und Wurzelhaarmutanten) in Mais charakterisiert. Integrierte transkriptomische und mikrobielle Datenanalysen zeigten, dass Mutationen, die die Lateralwurzelentwicklung beeinflussen, im Vergleich zu Mutationen, die andere Wurzeltypen betreffen, den größten Einfluss auf die Genexpression und den Aufbau des Mikrobioms des Wirtes hatten. Weitere Netzwerkassoziationsanalysen ergaben, dass die Schlüsselbakterien *Massilia* in Lateralwurzeln mit Wurzelfunktionsgenen assoziiert sind, die an der Blütenentwicklung und der Produktion Pflanzenbiomasse beteiligt sind. Es konnte gezeigt werden, dass die Interaktionen von *Massilia* mit Genen, die eine Rolle in der Reproduktion spielen, durch das Entwicklungsstadium bestimmt werden. Durch mikrobielle Inokulationsexperimente einer frühblühenden Maismutante konnte bestätigt werden, dass die *Massilia*-gesteuerte Förderung des Maiswachstums tatsächlich von der Blütezeit abhängt.

Zusammenfassend lässt sich sagen, dass speziell nach Wirtsgenotyp und Umweltfaktoren selektierte Mikroorganismen vorteilhafte Wechselwirkungen mit ihren Wirtspflanzen aufbauen können. Diese nützlichen Mikroorganismen veränderten die Wurzelarchitektur im frühen Stadium und die Blütenentwicklung im späteren Stadium, um das Pflanzenwachstum und die Leistung zu fördern, insbesondere bei Stress durch Stickstoffmangel. Diese kohärenten Ergebnisse zeigen eine starke genetische Verknüpfung und ein Züchtungspotenzial zur Verbesserung der Pflanzenleistung und Widerstandsfähigkeit in zukünftigen Agrarökosystemen mit geringem Input.

# 1 Introduction

## 1.1 *Zea mays* − a crop model plant

### 1.1.1 Maize is a main crop feeding the world

Maize (*Zea mays* L.) is reported as the highest yielding of all crops with 1,21 billion tons of grain yield in 2021 (FAOSTAT, http://faostat.fao.org). It is widely used for human consumption as a staple food in many regions of the world due to its different varieties including waxy corn, popcorn and sweet corn. Especially in Latin America, the Caribbean and Sub-Saharan Africa, maize provides about one-third of the calorie intake for the local population (https://www.croptrust.org). Maize is also grown for feeding animals and producing industry material such as alcohol, sweeteners, and syrup products (https://www.croptrust.org). Furthermore, maize is increasingly used for the production of bio-ethanol fuel, which is an environment-friendly energy source (Baeyens et al., 2015). Nevertheless, it is necessary to improve maize performance and productivity for future low-input agricultural system under the pressure of population growth, limited resources and global climate change.

### 1.1.2 Maize is an important genetic model

Maize is a monocotyledonous plant that belongs to the grass family *Gramineae* (*Poaceae*) (Strable & Scanlon, 2009). The initial domestication of maize, began 9,000 years ago in the Balsas river valley in Mexico when farmers started to collect the seeds of the wild grass teosinte (*Zea mays* ssp. *parviglumis*; Matsuoka et al., 2002). After hybridization with another wild grass teosinte (*Zea mays* ssp. *mexicana*) in the highlands of central Mexico, the admixed maize rapidly spread across South America and North America (Yang et al., 2023), growing under diverse environmental and climatic conditions, ranging from tropical forests to dry cold areas, from sea level to highlands >3,000 meters. Maize is now grown across a wider area than any other crops (Hake & Ross-Ibarra, 2015) due to its diverse usage and flexible adaptation to local environments.

During maize domestication, diverse open-pollinated maize varieties designated "landraces" were selected either by local environments or by farmers looking for desirable seeds and grain characteristics (Cleveland & Soleri, 2007). In Mexico and Central America, and the Caribbean, most farmers still grow their own landrace populations (Bellon & Hellin, 2011). Repeated human and natural selection within a given environment could contribute to adaptation of landrace populations to local climates (Mercer et al., 2008). Modern inbred varieties are derived from landraces by self-pollinating maize plants for multiple generations (Shull, 1908). However,

inbred varieties showed depressed phenotypic traits, while cross-pollinated maize (hybrids) displayed vigorous growth and performance (Darwin, 1876). The superior performance of hybrid maize makes it the dominant form of maize grown in modern agriculture (Duvick, 2001).

Maize has been positioned as an ideal model crop for genetic and genomic research due to its easy growth, controlled pollination, extensive nucleotide diversity, a vast collections of mutant stocks and genic collinearity with related grasses (Hake & Ross-Ibarra, 2015; Nannas & Dawe, 2015). Maize is easy to grow in the paper roll system, in the phytochamber, in the greenhouses and agricultural fields. The floral development in maize is monoecious, where male flowers develop in the tassel and female flowers are in the ear, making it easy to perform self- or controlled cross-pollinations (Hake & Ross-Ibarra, 2015). Moreover, the genetic architecture of maize is very complex and diverse due to its naturally outcrossing mode of inheritance, which provides valuable material for studies of genome evolution (Rafalski & Morgante, 2004; Wallace et al., 2014). In addition, maize genetics research has been well established and substantially promoted by a broad collection of genetic mutants (Neuffer et al., 1997). Since the 1990's, high-throughput mutagenesis programs have been launched to develop collections of maize mutants (Strable & Scanlon, 2009). Previous studies have determined candidate genes associated with complex phenotypes, such as flowering time (Thornsberry et al., 2001), kernel starch biosynthesis (Wilson et al., 2004), and seed carotenoid content (Harjes et al., 2008) by using quantitative trait locus (QTL) mapping and maize mutants. Furthermore, various genetic loci have been identified associated with root architectural traits (Burton et al., 2014), resistance to pathogens (Ding et al., 2008), a subset of the phyllosphere microbiome (Wallace et al., 2018), and a specific rhizosphere microbiome (Meier et al., 2022). Better understanding the environmental regulation and genetic basis of host plant control of their phenotypes will promote the identification of important plant genes in enhancing crop resilience in the context of sustainable agriculture.

## 1.2 Maize root system

### 1.2.1 Complexity of the root system

The root system of maize is composed of multiple root types, which are formed and play important functions at different growth stages (Hochholdinger et al., 2004). Maize root system consists of embryonic and post-embryonic roots. Embryonic roots include a single primary root and a variable number of seminal roots, which initiate endogenously from the embryo (Figure 1.1 A; Feldman, 1994; Hochholdinger, 2009) and play a major role in the acquisition of water and nutrients from the soil during the early stage after gemination (Hochholdinger et al., 2004;

Lynch, 1995). Post-embryonic roots are composed of aboveground brace roots and underground crown roots, which are responsible for the majority of water and nutrient uptake during late development, meanwhile contribute to anchorage of the maize plant (Hochholdinger et al., 2004; McCully & Canny, 1988). All these different root types are able to form lateral roots, which are also post-embryonic roots and further substantially increase the absorbing surface of the overall root system (McCully & Canny, 1988).
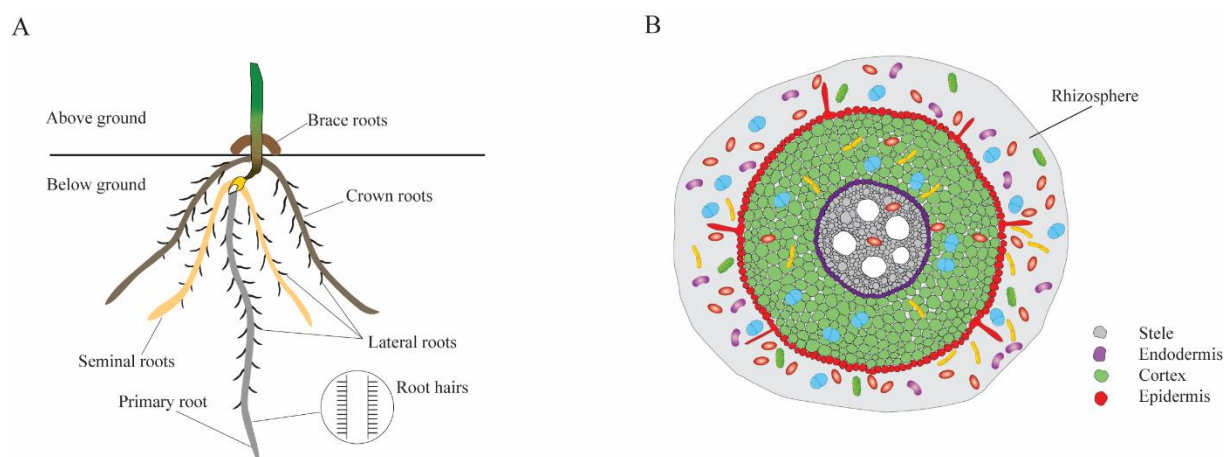


**Figure 1.1 The root system of maize. A**: Root types of maize. **B**: Anatomical structure of maize roots. For details see text. Note that brace roots are visible at the adult stage.

All mature roots show a similar anatomical structure (Hochholdinger, 2009) as illustrated in Figure 1.1 B. Radially from inside to outside, the tissues are: central cylinder (stele), endodermis, cortex, and epidermis which can form root hairs in connection with a narrow region soil surrounding the roots named as "rhizosphere" (Marschner, 2012). The central cylinder is responsible for longitudinally transporting water and nutrients to the shoot, while the major function of the cortex is a radial flow of water and nutrients and provides consistent protection against environmental stresses (Hochholdinger, 2009). From the apical part to the basal part of a single root, the root is longitudinally divided into the root cap, meristematic zone, elongation zone and maturation zone (Hochholdinger, 2009). Root hairs and lateral roots are formed in the maturation zone.

### 1.2.2 Maize root mutants and encoded genes

Several classical maize mutants have been identified in association with growth defects in different root types in the last decades. These genes and their molecular functions have been identified based on their mutant phenotypes (Hochholdinger et al., 2018). Figures of different root mutants used in this thesis are shown in Figure 1.2.
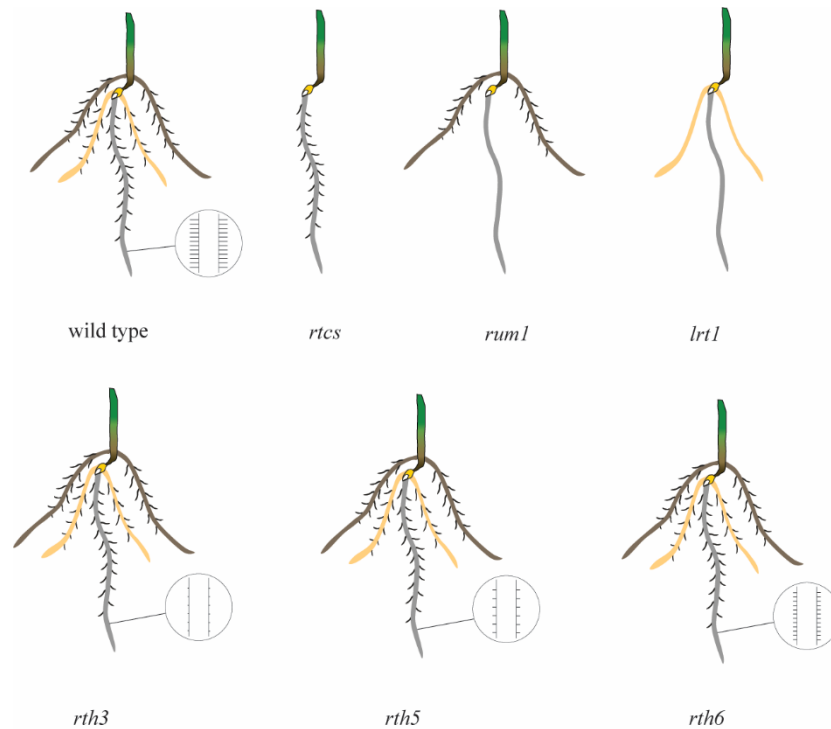
**Figure 1.2 Maize root mutants.** For details see text.

The *rtcs* (*rootless concerning crown and seminal roots*) mutant only contains a primary root and its lateral roots without any other root types (Hetz et al., 1996). Map-based cloning demonstrated that the *rtcs* gene encodes a member of the plant-specific family of LBD transcription factors involved in auxin signal transduction controlling shoot-borne and seminal root initiation (Taramino et al., 2007). The *lrt1* (*lateral rootless 1*) mutant has no lateral roots on all embryonic roots and lacks the first tier of crown roots at the early stage, and this defect can not be induced by auxin (Hochholdinger & Feix, 1998). A recent study indicates that the *lrt1* gene encodes a 209 kDa homolog of the DDB1-CUL4-ASSOCIATED FACTOR (DCAF) subunit of the CUL4-based E3 ubiquitin ligase (CRL4) complex localized in the nucleus (Baer et al., 2023). Interestingly, the lateral root defect in *lrt1* can be rescued by inoculation with the arbuscular mycorrhizal fungus, *Glomus mosseae* or growth in high phosphate environment (Paszkowski & Boller, 2002). In contrast, the *rum1* (*rootless with undetectable meristems 1*) mutant is defective in seminal roots and lateral roots on the primary root (Woll et al., 2005). Map-based cloning demonstrated that *rum1* encodes a canonical Aux/IAA protein that is a central regulator of auxin signaling (Von Behrens et al., 2011). Root hairs are tubular extensions of trichoblast cells of the epidermis and are instrumental for nutrient and water uptake. In the past, several mutants which lack root hairs or display reduced root hair length have been identified in maize (Hochholdinger et al., 2018). The proteins encoded by the genes *rth3*, *rth5*, and *rth6* are functionally linked during maize root hair formation. RTH5 produces apoplastic

4

superoxide (Nestler et al., 2014), which results in cell wall loosening via hydroxyl radicals and also controls the expression of cellulose biosynthesis genes. After cell wall loosening, the transmembrane cell wall protein RTH6 (Li et al., 2016), which corresponds to cellulose synthase-like D5 (CSLD5), is arranged in rosettes and synthesizes cellulose at the plasma membrane that is extruded to the inner side of the cell wall in the root hair tip and thus reinforces the tubular shaft. Finally, the GPI-anchored COBRA-like cell wall protein RTH3 (Hochholdinger et al., 2008) is involved in the organization of the synthesized cellulose.

The complex root system of maize is essential for its high productivity because it can efficiently extract water and nutrients from soil and transport them to the cobs (Lynch, 2013). Researchers have put a lot of effort to improve plant growth and resistance to biotic and abiotic stresses through genetic engineering (Hochholdinger, 2016; Meister et al., 2014) or microbial manipulation (Angelard et al., 2010) to modify root traits. However, the genetic basis of the root system and its association with beneficial microorganisms is largely unexplored, thus providing a great potential to improve maize productivity especially under stress conditions.

## 1.3 The root-associated microbiome is essential for plant performance

Soil serves as the major source of microorganisms in terrestrial ecosystems, thus soil microorganisms are crucial in the one health concept because they contribute to plant, animal and human health (Banerjee & van der Heijden, 2023). In plants, the root is the major interface influenced by soil microorganisms. The root-associated microbiome plays an important role in promoting plant performance and health under biotic and abiotic stresses (Cheng et al., 2019; Oldroyd & Leyser, 2020). Specific root-associated microorganisms have been shown to have causal impacts on root development (Finkel et al., 2020; Hodge et al., 2009; Yu & Hochholdinger, 2018), and nutrient uptake (Almario et al., 2017; Poole et al., 2018; Salas-González et al., 2021; Yu et al., 2021) in different plant species.

### 1.3.1 The microbiome influences root development

Although root systems are genetically determined, their architecture can be strongly changed by soil microorganisms, particularly beneficial rhizobacteria and arbuscular mycorrhizal fungi (AMF) (Yu & Hochholdinger, 2018). Specific bacteria have been shown effects on the alteration of root biomass and architecture in different plant species after inoculating with plant growth-promoting rhizobacteria (PGPR) (Carvalho et al., 2014; Zamioudis et al., 2013). There are also a number of studies reporting increased root branching and volume in response to AMF (Fusconi, 2014; Gutjahr et al., 2009; Hodge et al., 2009) and non-mycorrhizal fungi (Hiruma et

al., 2016) in plants.

The most common effect of PGPR on root architecture is the proliferation of lateral roots and root hairs but inhibition of primary root growth, which results in increment in root weight (Dahmani et al., 2020; El Zemrany et al., 2007; Jochum et al., 2019; Walker et al., 2012), shoot weight (Asari et al., 2017; Cassán et al., 2009; Etesami & Alikhani, 2016) and yield (Çakmakçi et al., 2001; Fallik & Okon, 1996). In maize, *Azospirillum brasilense* inoculated plants have increased root and shoot dry weight (Cassán et al., 2009) as well as increased yield at low levels of nitrogen in field experiments (Fallik & Okon, 1996). *Azospirillum lipoferum* inoculated maize plants had longer primary root length, higher root fresh weight and plant height (El Zemrany et al., 2007; Jacoud et al., 1998). A recent study demonstrated that both *Bacillus* and *Enterobacter* isolates significantly increased root length, root surface area and the number of root tips of maize and wheat plants under drought stress (Jochum et al., 2019). *Bacillus megaterium* was reported to cause substantial alterations of the root architecture in *Arabidopsis thaliana* (López-Bucio et al., 2007) and significantly improve the yield in maize (Efthimiadou et al., 2020).

Previous studies have shown that AMF preferentially colonized lateral roots and had negligible effects on primary roots of dicots or crown roots of monocots (Gutjahr et al., 2009; Yu & Hochholdinger, 2018). The most dramatic influence of AMF colonization on root system architecture was observed in the mutant *lrt1* of maize, resulting in the induction of very bushy lateral roots under low phosphate conditions (Paszkowski & Boller, 2002). However, the AMF effect on the root system is different across different maize or soybean genotypes (Wang et al., 2011; Zhu et al., 2005), thus suggesting that part of the response is due to genetic variation (Gutjahr et al., 2009). In addition, the non-mycorrhizal fungus *Colletotrichum tofieldia* was reported to promote root elongation under low phosphorus conditions in *Arabidopsis* (Hiruma et al., 2016).

### 1.3.2 The microbiome facilitates nutrient uptake

Nitrogen and phosphorus are the main nutrients that plants require from soil to support growth and reproduction. However, nitrogen is easily lost in the soil by leaching or denitrification (Vance, 2001), while phosphorus is usually insoluble and cannot be directly acquired by plants (Vance et al., 2003). Previous studies have shown that the plant-associated microbiome strongly influences nutrient use efficiency by mineralizing organic nutrients and transforming inorganic nutrients (Marschner, 2012; Zhang et al., 2019). There are a vast number of microorganisms with nitrogen-fixing and phosphate-solubilizing properties that include bacterial members

belonging to the taxa *Azorhizobium*, *Bradyrhizobium*, *Burkholderia*, *Cupriavidus*, *Mesorhizobium*, *Pseudomonas*, *Rhizobium* and *Sinorhizobium* (Pham et al., 2017; Poole et al., 2018) and fungal members belonging to the mycorrhizal taxon (Etesami et al., 2021), and the non-mycorrhizal taxon *Helotiale* (Almario et al., 2017) from diverse crop species.

One of the best characterized nitrogen-fixing bacteria is the rhizobia in legumes. In this symbiosis between rhizobia and legumes, plants provide carbon and energy in the form of dicarboxylic acids, and in return, the rhizobia secrete ammonium from gaseous nitrogen that the plant uses to synthesize amino acids (Poole et al., 2018). Previous studies showed that nitrogen-fixing bacteria contributed to a large portion of nitrogen supply in sugarcane (Mirza et al., 2001). Moreover, cooperation between mycorrhizal fungi and soil microbial communities doubled nitrogen acquisition in the model grass *Brachypodium distachyon* (Hestrin et al., 2019). In maize, Yu et al. (2021) reported that maize roots secreted more flavonoids to attract members of the bacterial taxon *Massilia* that induced lateral root growth as well as improved shoot biomass and nitrogen accumulation in *lrt1* mutants under low nitrogen conditions.

Phosphorus is the second macronutrient that plants acquire from soil. Microorganisms in soil can solubilize inorganic phosphate or mineralize organic phosphate to make it available for plants through the release of organic acids (Gyaneshwar et al., 2002). The best example of microorganism-driven phosphorus uptake is AMF (Behie & Bidochka, 2014; Wang et al., 2017). The phosphorus status of plant tissues grown with AMF is higher than that of tissues grown without AMF under the same condition (Fusconi et al., 2005). Non-mycorrhizal fungi can also promote plant growth and phosphorus uptake in *Arabis alpine* (Almario et al., 2017).

## 1.4 Host genetics influence on microbial community assembly

The soil type is the main driver of the composition of the bacterial community in both the rhizosphere and the endosphere of different plant species (Bulgarelli et al., 2012; Peiffer et al., 2013; Schreiter et al., 2014; Thiergart et al., 2020). Soil type and climate had a comparable influence on fungal community composition of roots in *Arabidopsis thaliana* (Thiergart et al., 2020). However, under identical soil and climate conditions, the plant genotype drives the structure and function of root microbial community, thus demonstrating that roots are able to filter their microbiomes in a defined environment (Haney et al., 2015; Lundberg et al., 2012). Furthermore, root microbial community composition can be divergent at the level of subspecies, e.g. different cultivars of sorghum (Schlemper et al., 2017) and rice (Singh et al., 2022).

An investigation of 18 plant species including maize suggests that the dissimilarity of bacterial

communities in the root endosphere strongly correlates with host phylogenetic distance (Naylor et al., 2017). Similarly, domestication also strongly shaped microbial diversity in the rhizosphere of various plant species (Pérez-Jaramillo et al., 2016). A distinct microbiome composition and functions were shown in domesticated barley (*Hordeum vulgare*) compared to its wild accession (Bulgarelli et al., 2015). Some gene families affecting host-microbe interactions showed the evidence of positive selection in domesticated barley (Bulgarelli et al., 2015). A study on domestication of wheat showed that the rhizosphere microbiome was shifted from a slow growing and fungi dominated community to a fast growing and bacteria dominated community (Yue et al., 2023).

In maize, recent studies demonstrated that the maize rhizosphere microbial community has also been substantially affected by domestication (Brisson et al., 2019; Szoboszlay et al., 2015) and modern hybrid breeding (Favela et al., 2021; Wagner et al., 2020). The maize progenitor teosinte showed significantly higher bacterial abundance and diversity in the rhizosphere compared to modern maize, sweet corn and popcorn inbred lines (Szoboszlay et al., 2015). Brisson et al. (2019) reported that the microbial community composition in hybrids was more different from teosinte and inbred lines. Similar observations were shown in the rhizosphere between inbred lines and hybrids, even some microbial features showed a heterosis effect like normal phenotypes (Wagner et al., 2020). Moreover, a recent study demonstrated that rhizosphere bacteria differed among different varieties released from 1949 to 1986, while fungi did not significantly change (Favela et al., 2021). These results indicate that root-associated microbial assembly may have co-evolved with host genetic variation, domestication and modern breeding.

Genome-wide association study (GWAS) is a powerful approach to identify genetic loci that associate with complex traits. Recently, GWAS was widely used for microbial traits in *Arabidopsis* (Bergelson et al., 2019; Horton et al., 2014), sorghum (Deng et al., 2021), and maize (Meier et al., 2022; Wallace et al., 2018). In sorghum, 49 rhizosphere OTUs belonging to diverse orders were identified in associations with host genetic loci (Deng et al., 2021). In maize, only two *Methylobacteria* OTUs on leaf were identified associating with host genetic loci (Wallace et al., 2018), while Meier et al. (2022) identified 622 genetic loci that were significantly linked to 104 microbial traits in the maize rhizosphere. Moreover, recent studies demonstrated that host genetics has a potential regulation on microbiome assembly from the rhizosphere to the root (Edwards et al., 2015; Thiergart et al., 2020).

## 1.5 Aims of this study

This thesis aims to investigate how host plant genetics and environmental factors regulate the root-microbiome association, how plants gain benefits from microbiome under stress conditions and to what extent some specific taxa can be beneficial for improving performance and resilience under abiotic stresses in maize. The following objectives are examined in this thesis:

1. Examine genotype effects on the overall microbial community of rhizosphere and root using a large maize population.
2. Identify root-associated microbial taxa that are heritable and associated with host genomic regions.
3. Investigate the plant source environmental impacts on the root-associated microbiome assembly through domestication of maize.
4. Identify specific microbial taxa that can improve plant performance under different stress conditions.
5. Explore the associations between beneficial microbial taxa and host gene functions in different root types or tissues.

# 2 Heritable microbiome variation is correlated with source environment in locally adapted maize varieties

Xiaoming He[1,2,3,#], Danning Wang[2,3,#], Yong Jiang[4,#], Meng Li[5,#], Manuel Delgado-Baquerizo[6,#], Chloee McLaughlin[5], Caroline Marcon[3], Li Guo[3], Marcel Baer[3], Yudelsy A.T. Moya[7], Nicolaus von Wirén[7], Marion Deichmann[8], Gabriel Schaaf[8], Hans-Peter Piepho[9], Zhikai Yang[10], Jinliang Yang[10], Bunlong Yim[11], Kornelia Smalla[11], Sofie Goormachtig[12,13], Franciska T. de Vries[14], Hubert Hüging[15], Mareike Baer[16], Ruairidh J. H. Sawers[5,*], Jochen C. Reif[4,*], Frank Hochholdinger[3,*], Xinping Chen[1,*], Peng Yu[2,3,*]

[1] College of Resources and Environment, and Academy of Agricultural Sciences, Southwest University (SWU), 400715 Chongqing, P. R. China

[2] Emmy Noether Group Root Functional Biology, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, 53113 Bonn, Germany

[3] Crop Functional Genomics, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, 53113 Bonn, Germany

[4] Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), 06466 Gatersleben, Germany

[5] Department of Plant Science, Pennsylvania State University, State College, PA 16802, USA

[6] Laboratorio de Biodiversidad y Funcionamiento Ecosistémico. Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS), CSIC, Av. Reina Mercedes 10, E-41012, Sevilla, Spain

[7] Department of Physiology and Cell Biology, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), 06466 Gatersleben, Germany

[8] Plant Nutrition, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, 53115 Bonn, Germany

[9] Biostatistics Unit, University of Hohenheim, 70599 Stuttgart, Germany

[10] Department of Agronomy and Horticulture, University of Nebraska-Lincoln, NE 68583 Lincoln, United States

[11] Institute for Epidemiology and Pathogen Diagnostics, Julius Kühn-Institut – Federal Research Centre for Cultivated Plants (JKI), Messeweg 11–12, D-38104 Braunschweig, Germany

[12] Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium

[13] Center for Plant Systems Biology, VIB, Ghent, Belgium

[14] Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands

[15] Crop Science Group, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, 53115 Bonn, Germany

[16] Institute of Nutrition and Food Sciences, Department of Food Microbiology and Hygiene, University of Bonn, 53115, Bonn, Germany

[#] These authors equally contributed to this work.

[*] To whom correspondence should be addressed:

rjs6686@psu.edu

reif@ipk-gatersleben.de

hochholdinger@uni-bonn.de

chenxp2017@swu.edu.cn

yupeng@uni-bonn.de (lead contact)

**Running title:** Heritable maize microbiome against abiotic stress

**Key words:** Abiotic stress, maize, microbiome, root, rhizosphere

**Author contributions**

P.Y., X.C. and F.H. designed the study; P.Y. coordinated and managed the whole project; X.H. performed the culture and harvest of the phytochamber experiments. D.W. analysed the microbiome data and performed all statistical analysis; Y.J. and J.C.R., performed the genetic analysis; C.Mc. and R.J.H.S. performed machine learning and environmental genome-wide association analysis; M.D.B. performed ecological analysis; B.Y. and K.S. contributed bacterial strains from maize; X.H. and M.B. performed bacterial inoculation experiments; X.H. and L.G. extracted all the DNA samples; M.L., Z. Y. and J. Y performed the genomic prediction analysis; P.Y. and H-P P. discussed and designed the large pot experiment; C.Ma. and F.H. contributed the Mu-transposon induced lines; M.D., G.S., Y.A.T.M. and N.v.W. conducted the soil and plant nutrient analyses. H.H. performed the preparation of soil from Dikopshof long-term experimental station; X.H., D.W., Y.J., M.L., M.D.B., R.J.H.S., J.C.R., X.C., F.H. and P.Y. wrote the paper. All authors read and approved the final version of the manuscript.

## Abstract

Beneficial interactions with microorganisms are pivotal for crop performance and resilience. However, it remains unclear how heritable the microbiome is with respect to the host plant genotype and to what extent host genetic mechanisms can modulate plant-microbiota interactions in the face of environmental stresses. Here, we surveyed 3,168 root and rhizosphere microbiome samples from 129 *Zea* accessions, sourced from diverse habitats and grown under control and different soil stress conditions. We quantified soil treatment and host genotype effects on the microbiome. Plant genotype and source environment were predictive of microbiome abundance. Genome wide association analysis identified host genetic variants linked to both rhizosphere microbiome abundance and source environments. We identified transposon insertions in a candidate gene linked to both the abundance of a keystone bacterium *Massilia* in our controlled experiments and total soil nitrogen at source environments. Isolation and controlled inoculation of *Massilia* alone can contribute to lateral root development, whole plant biomass production and adaptation to low nitrogen availability. We conclude that locally adapted maize varieties exert patterns of genetic control on their root and rhizosphere microbiomes that follow variation in their home environments, consistent with a role in tolerance to prevailing stress.

## Introduction

Microorganisms that colonize the rhizosphere surrounding plant roots, root surfaces and internal tissues play an important role in promoting plant health and fitness under biotic and abiotic stresses[1,2]. Specific features of the root microbiome have been shown to modify root architecture[3], regulate nutrient homeostasis[4], protect against stress[1] and impact ecosystem function[5]. Although the overall root microbiome is largely shaped by soil properties[6], small host-mediated changes in microbiome composition can have large effects on plant fitness[7-9]. Modification of crop microbiomes has been proposed as a strategy to promote food security while supporting sustainable agroecosystems[10,11]. However, the extent to which host genetic mechanisms can modulate the microbiome under different environmental conditions and the mechanistic basis of any such control remains poorly characterized.

The diversity of traditional crop varieties (landraces) provides a powerful resource to investigate heritable adaptive variation in crops[12-14]. Long term selection in diverse, and often challenging environments can reveal subtle signals linking plant genetic and phenotypic variation to local conditions. Maize (*Zea mays.* ssp. *mays*) is an excellent model for investigating the genetic basis and environmental signature of plant-microbe interactions due to the extensive climatic and edaphic variation across its range[15]. The domestication of maize began 9,000 years ago when Mexican farmers started to collect the seeds of the wild grass teosinte[16] (*Zea mays* ssp. *parviglumis*). During maize domestication and improvement, the root system expanded its functionality and complexity[17,18]. Recent studies have highlighted that the maize rhizosphere microbial community has been significantly impacted by domestication[19,20] and, more recently, by modern hybrid breeding[21,22]. Similarly, wheat and barley domestication have been shown to have reshaped the community structure and metabolic functions of the rhizosphere microbiome[23,24]. These studies highlight the impact of indirect selection on plant-microbe interactions during domestication and selective breeding. Nonetheless, better understanding the genetic basis and environmental regulation of host-microbiota associations under abiotic stresses may promote crop resilience in the context of more sustainable agronomic practices.

Here, we profiled 3,168 root and rhizosphere microbiome samples from 129 diverse *Zea* accessions grown under control, nitrogen-, phosphorus- and water-limited conditions using 16S rRNA gene and ITS1 gene sequencing. We assessed how the native habitats (source) of traditional varieties was predictive of root and rhizosphere microbiota assembly under our soil treatments. Understanding how plant traits modulate their microbiome to enhance tolerance to

environmental constraints, the extent to which this plant trait-microbe association is heritable under abiotic stresses, and how this association is encoded in the genetic program provides novel insights into establishment of beneficial host-microbiome associations. Such insights are a prerequisite for the generation of environment-tailored cultivars that recruit favourable microbial consortia for increasing agricultural productivity, resilience to climate change, and sustainability.

## Results

### The maize microbiome responds strongly to abiotic stresses

Our goal was to investigate how plant genotype impacts host-microbiome associations and their capacity to influence plant performance under soil stress conditions. We used 16S rRNA gene and ITS1 gene sequencing to characterize the root and rhizosphere microbiome of 129 *Zea* accessions (11 teosintes, 97 landraces, 11 maize inbred lines and 10 maize hybrids) (Supplementary Fig. 1) grown in control-, low phosphorous-, low nitrogen-, and drought-treatments in a soil sourced from a long-term field experimental station located in Dikopshof (50°48′21″N, 6°59′9″E) and representative of local maize growing areas (See Methods; Supplementary Fig. 2). We sampled root and rhizosphere compartments from the first whorl of shoot-borne crown roots (Supplementary Fig. 3), in addition to collecting bulk soil. Microbial community composition differed across samples for both bacteria and fungi, with compartment (bacteria: $R^2 = 0.756$, $P = 1.0e-4$; fungi: $R^2 = 0.402$, $P = 1.0e-4$) explaining the largest proportion of the variation followed by stress treatment (bacteria: $R^2 = 0.052$, $P = 1.0e-4$; fungi: $R^2 = 0.021$, $P = 1.0e-4$) (Supplementary Fig. 4). Among all samples, plant genotype (bacteria: $R^2 = 0.010$, $P = 7.0e-4$; fungi: $R^2 = 0.050$, $P = 1.0e-4$) was less important than either compartment or treatment (Supplementary Fig. 4). Within each compartment, treatment had a larger effect on bacteria ($R^2 = 0.32$, $R^2 = 0.26$, $P < 0.001$) than fungi ($R^2 = 0.13$, $R^2 = 0.11$, $P < 0.001$), in both rhizosphere and root (Fig. 1a and b). In contrast, genotype had a larger effect on fungi ($R^2 = 0.12$, $R^2 = 0.098$, $P < 0.001$) than bacteria ($R^2 = 0.064$, $R^2 = 0.058$, $P < 0.001$) in both rhizosphere and root. In the rhizosphere and roots, we observed significantly (Kruskal-Wallis test, Dunn's *post-hoc* test with BH adjusted, $P < 0.05$) lower bacterial α-diversity under drought stress and nitrogen deficiency compared to control conditions (Supplementary Fig. 5a). A significant effect of phosphorus on bacterial α-diversity was seen in the rhizosphere but not in the root compartment (Supplementary Fig. 5a). For fungal α-diversity, the only significant treatment difference was (Kruskal-Wallis test, Dunn's *post-hoc* test with BH adjusted $P < 0.05$) lower diversity under nitrogen deficiency than control conditions in the root (Supplementary

Fig. 5b). Interestingly, we only detected a significant ($R^2$ = 0.21, $P$ <0.01) interaction effect between treatment and genotype for rhizosphere bacteria (Fig. 1a). Such detected interaction effect between treatment and genotype indicates that there may be specific genotypes with adaptive role on bacterial microbiota in the rhizosphere.
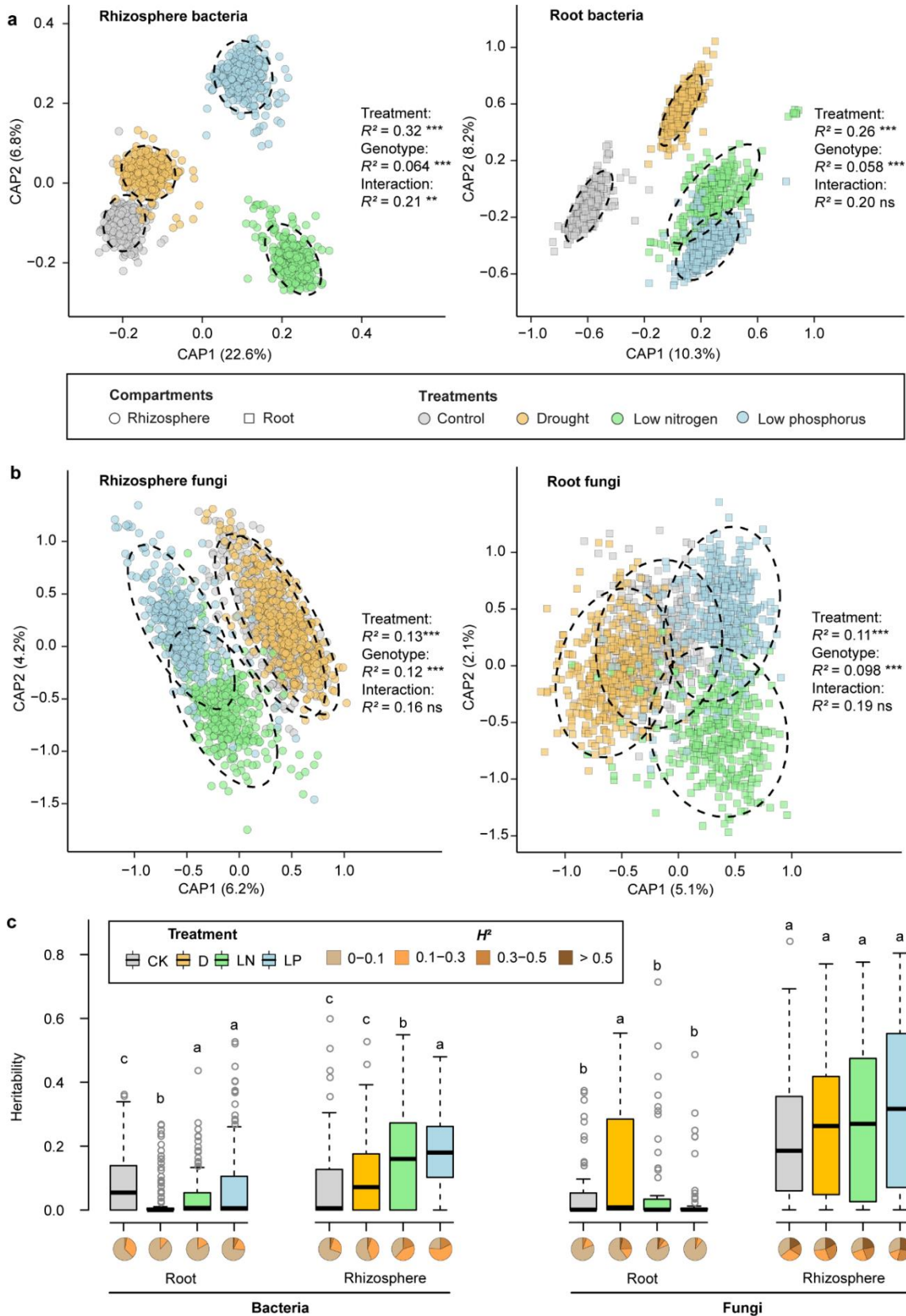
**Figure 1. Overall diversity and heritability of the microbiome under abiotic stresses. a**, Constrained analysis of principal coordinate (CAP) ordination using Bray–Curtis dissimilarity with permutational analysis of variance (PERMANOVA) was applied to visualize significant bacterial microbiota differences in the rhizosphere and root across four treatments and genotypes ($n = 129$). Compartments are shape coded. Only ASVs with reads >10 in ≥6 samples were included in the dataset. **b**, Constrained analysis of principal coordinate (CAP) ordination using Bray–Curtis dissimilarity with permutational analysis of variance (PERMANOVA) was applied to visualize significant fungal microbiota differences in the rhizosphere and root across four treatments and genotypes ($n = 129$). Datapoints for bacteria ($n = 3,138$) and fungi ($n = 3,168$) are color coded according to the four treatments. Compartments are shape coded. Only ASVs with reads >10 in ≥6 samples were included in the dataset. Ellipses represent an 90% confidence level. \*\*\*, $P < 0.001$, \*\*, $P < 0.01$, ns, not significant. **c**, Heritability estimates of individual families under four treatments for both bacteria and fungi. The broad-sense heritability ($H^2$) was calculated using highly abundant bacterial ($n = 131$) and fungal ($n = 59$) families across all samples. CK, control; D, drought; LN, low nitrogen; LP, low phosphorus. Significances are indicated among treatment groups for each compartment with Benjamini-Hochberg adjusted $P < 0.05$ (Kruskal-Wallis test, Dunn's *post-hoc* test). Boxes span from the first to the third quartiles, centre lines represent the median values and whiskers show data lying within 1.5× interquartile range of the lower and upper quartiles. Data points at the ends of whiskers represent outliers. The pie charts indicate the proportional distributions of heritability frequencies.

## Keystone genera define the major differences in the microbiome

Overall, we identified 815 bacterial amplicon sequence variants (ASVs) and 248 fungal ASVs with high quality after removing the chimeric and organelle sequences (Supplementary Datasets 1−2). Bacterial and fungal ASVs which expressed ≤0.05% relative abundance within ≤5% samples were removed from downstream analysis (Supplemental datasets 3−4). Keystone microbial taxa or ASVs are defined as the drivers of microbiome structure and function[5]. We identified putative keystone microbes among the highly abundant ASVs using co-occurrence network analysis of relative abundance data (See Methods and Supplementary Datasets 5−7). Overall, the number of associations and accumulative weights of ASVs were largely positive within the bacterial or fungal networks, but negative in the inter-kingdom network (Supplementary Fig. 6; Supplementary Dataset 5). This is consistent with previous reports that inter-kingdom interactions determine the overall assembly, stability, and fitness of the root microbiome in *Arabidopsis*[25]. We also observed that a high proportion of the negative inter-kingdom associations were conserved across the stress treatments (Supplementary Fig. 7; Supplementary Dataset 6). Among interacting ASVs, keystone taxa or ASVs were defined as those with a hub score greater than 0.1. In the rhizosphere compartment, we identified 20 keystone genera across all treatments, from which the bacterial genera *Sphingomonas*, *Massilia* and *Lysobacter* were the most represented at the ASV level (Supplementary Fig. 8a). In the root compartment, we detected 5 keystone ASVs that belong to *Massilia*, which is the only genus conserved across control and low nitrogen treatments, while 19 keystone ASVs belonging to *Streptomyces* were conserved across control and drought stress treatments (Supplementary Fig. 7 and 8b; Supplementary Dataset 7). Functional prediction indicated that these bacterial genera are involved in ureolysis (*Massilia*) and aerobic chemoheterotrophy (*Streptomyces*) (Supplementary Dataset 8). The fungal keystone taxa were mainly predicted to be decomposers

(37%) and pathogens (25%; Supplementary Dataset 9). Overall, our co-occurrence network analyses revealed strong negative correlations between bacterial and fungal ASVs in maize rhizosphere and roots, while only specific keystone bacterial members such as *Massilia* are enriched as drivers of microbiota composition in the roots in nitrogen deficient soil.

**Stress treatments result in a less diverse but more heritable microbiome**

We first investigated whether abiotic stresses will influence maize performance by estimation of shoot dry biomass and relative chlorophyll content (SPAD). As shown in Supplementary Fig. 9, drought and nutrients deficiency significantly inhibit maize growth (Supplementary Fig. 9a, One-Way ANOVA, Tukey's HSD, $P$ <0.05), while drought and nitrogen deficiency significantly reduce the relative leaf greenness (Supplementary Fig. 9b, One-Way ANOVA, Tukey's HSD, $P$ <0.05), but not under phosphorus deficiency. These shoot phenotype analyses indicate that maize plants undergone a stressed situation under our well-controlled conditions. To estimate the overall influence of the plant genotype on microbiome composition, we estimated the correlation between the plant genetic distance matrix and the microbiome distance matrix using 97 plant genotypes, for both root and rhizosphere. There was a significant correlation (Mantel's statistics) between the bacterial communities and plant genotypes in both compartments (Rhizosphere: R = 0.32, $P$ = 1.0e−4; Root: R = 0.16, $P$ = 0.0079). Fungi only displayed a significant correlation with the plant genotype in the rhizosphere (R = 0.23, $P$ = 1.0e−4; Supplementary Fig. 10). We estimated the broad-sense heritability ($H^2$) for the microbiome at different microbial taxonomic levels and for individual ASVs across the experiment, and then separately for each compartment and treatment combination (Supplementary Dataset 10; see Methods). Across treatments, the average $H^2$ was higher for the rhizosphere microbiome (Family: $H^2$ = 0.15; Genus: $H^2$ = 0.14; ASV: $H^2$ = 0.16) than the root microbiome (Family: $H^2$ = 0.052; Genus: $H^2$ = 0.049; ASV: $H^2$ = 0.052) at the level of family (Fig. 1c), genus (Supplementary Fig. 11a) and ASV (Supplementary Fig. 11b). Nutrient stress significantly (Kruskal-Wallis test, Dunn's *post-hoc* test with BH adjusted $P$ <0.05) increased the average $H^2$ (control, $H^2$ = 0.078; low nitrogen, $H^2$ = 0.16; low phosphorus, $H^2$ = 0.18) of the bacterial rhizosphere microbiome, but not of the fungal microbiome (Fig. 1c). To identify specific plant genetic loci affecting the microbiome, we performed genome-wide association analysis (GWA) for the most heritable ($H^2$ >0.1) microbes at the level of α-diversity, family, genus and individual ASV (Supplementary Dataset 11). We did not recover significant markers in association with overall measures of microbial α-diversity (Shannon index) (Supplementary Dataset 12). We did, however, identify significant (BH adjusted $P$ <0.05) associations with individual ASVs and taxa at the family and genus levels in both root and rhizosphere

compartments (Supplementary Dataset 11). We detected a total of 533 (Supplementary Fig. 12a) marker-trait associations (MTAs) with bacteria, and 283 MTAs with fungi (Supplementary Fig. 12b; MTAs with the proportion of explained phenotypic variance higher than 5%). We extracted a list of 567 genes linked to these significant markers (Supplementary Dataset 13) and assigned them to functionally classes using gene ontology (GO) analysis (Supplementary Dataset 14). We found six GO terms to be enriched: "nuclear export", "RNA transport", "mRNA export from nucleus", "purine ribonucleotide catabolic process", "regulation of organelle organization" and "purine ribonucleoside catabolic process" (Supplementary Dataset 14). Overall in our experiment, these data indicate the heritable impact of the host genotype on bacterial communities is conserved under both control and stress conditions.

**Plant source environments predict the root and rhizosphere microbiome**

To address the hypothesis that variation in the plant microbiome reflects adaptation to the native environment, we assessed the potential of climatic and edaphic descriptors of source environments to predict the microbial abundance in our standardized phytochamber experiments (Supplementary Fig. 2; Supplementary Dataset 15). To reduce the complexity of the microbiome data, we first applied a modified use of unbiased weighted correlation network analysis (WGCNA)[26] and identified network associations clustered into fourteen distinct modules of highly correlated microbial ASVs in the root (Supplementary Fig. 13a; Supplementary Dataset 16). We next assigned a relative "eigentaxa" score[27] to each sample and determined the biological relevance between microbial modules and shoot phenotypic traits i.e. shoot dry biomass, shoot nitrogen concentration and shoot nitrogen content. In particular, six of the identified modules were positively associated with shoot dry biomass and nitrogen content (Supplementary Fig. 13b). Interestingly, the taxa of the "darkred" module ($P = 3e-6$) were specifically enriched in *Massilia* (*Oxalobacteraceae*) (Supplementary Fig. 13c). We then sought evidence of covariation among microbial modules and environmental descriptors i.e. total nitrogen, phosphorus retention rate and annual precipitation in the places of origin for maize varieties. Notably, the "darkred" module that was enriched in *Massilia* was significantly negatively ($R = -0.28$, $P = 0.0039$) correlated with total soil nitrogen of the original sites of collection, exclusively in the low nitrogen treatment of the controlled experiment (Supplementary Fig. 14), thus reflecting the potentially selectiveness of *Massilia* when the soil nitrogen is limited in the natural habitats.

We used structural equation modeling to quantify the cumulative effects of source environments, plant genetic diversity, stress treatments, domestication status and biomass on the "darkred"

module. These analyses demonstrated an impact of plant genotype and source environments on specific assemblies of microbiome. Low nitrogen treatment (3.4% standardized total effect), source mean annual temperature (2.9% standardized total effect), source precipitation (−10.9% standardized total effect) and plant genotype (10.8% standardized total effect) significantly correlated with the microbiome assemblage, notably with the abundance of the keystone genus *Massilia* (Supplementary Fig. 15). We next compared the predictive ability of a genomic model, an environmental model and a combined genomic-environmental model on ASV abundance (See Methods). Overall, prediction was better for bacterial data than for fungal data, and better for rhizosphere than root (Fig. 2a; Supplementary Fig. 16). Interestingly, microbial abundance could be predicted more accurately with descriptors of source environments or a combination of these with plant genetic markers than with genetic markers alone at the level of family, genus, and ASV (Fig. 2a; Supplementary Fig. 17−19), which is consistent with the scenario that environment plays an influential role in determining microbiome composition through its impact on host genetics. Under the conditions of our experiment, ecological modelling and prediction analyses show potential effects of source environments of locally adapted maize on the abundance of the rhizosphere bacterial community. Therefore, such local environments together with the plant genomic information will largely play positive effect on understanding the structure and function of the rhizosphere microbiome even grown under a new environment.
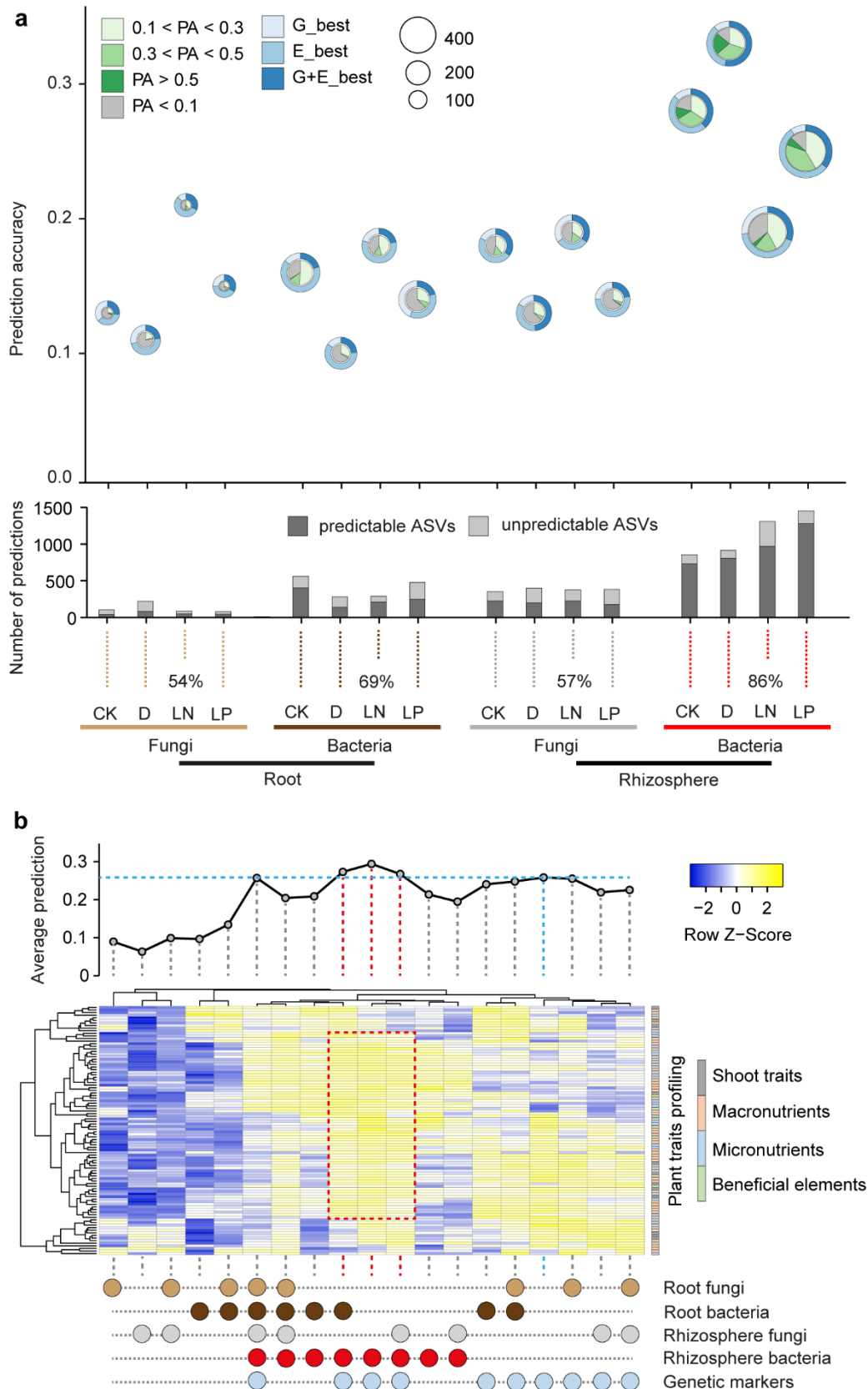
**Figure 2. Genomic, environmental and microbial prediction of host-microbe interactions and plant traits.**
**a**, Microbiome traits prediction using genetic markers and environmental characters. Inner pie charts describe the proportion of ASVs with four different magnitudes of prediction accuracies from different treatments across compartments. Outer circles define the best prediction patterns observed by applying the genetic markers (G_best) alone, environmental characters (E_best) alone or combined genetic markers and environmental characters (G+E_best). The numbers denote the average prediction accuracies for microbial ASVs from different treatments

across compartments. Only ASVs with heritability ($H^2$) >0.1 were considered in prediction analysis. PA, prediction accuracy. Bar plots indicate the proportions of predictable (PA >0.1) and unpredictable (PA <0.1) ASVs from the total predictions. CK, control; D, drought; LN, low nitrogen; LP, low phosphorus. **b**, Plant traits prediction using genetic markers and microbiome traits. Different colour dots indicate genetic markers (blue) or different microbiome traits e.g. root fungi (brown), root bacteria (dark brown), rhizosphere fungi (grey) and rhizosphere bacteria (red). Combinations of different dots indicate integration of several data matrix. A curved line describes the average prediction accuracy for plant traits using microbiome data alone, genomic data alone or combined genomic and microbiome traits data. A heatmap illustrates the standardized prediction accuracy for fitness traits across different microbiome features combined with genetic markers. Shoot traits include the biomass, leaf area and chlorophyll measured by SPAD value. Nutrient uptake properties include the concentration and content of macronutrients (nitrogen, phosphorus, potassium, calcium, magnesium and sulfur), micronutrients (iron, manganese, zinc and boron) and beneficial elements (aluminium and sodium).

**Consideration of the root-associated microbiome improves prediction of plant traits**

To assess the relationship between the microbiome and plant growth and physiology, we used a two-step strategy combining genomic prediction and Random Forest models based on source environment descriptors. First, we compared the genomic prediction ability on plant growth and nutrient accumulation traits by using plant genetic markers alone or in combination with microbiome ASVs abundance. The combination of plant genetic markers and rhizosphere bacterial community composition provided the highest average prediction ability (29%) (Fig. 2b; Supplementary Datasets 17 and 18). We confirmed this result by employing an alternative approach to fit a ridge regression mixed model, observing ~10%−15% increase of prediction accuracy when using both genetic and microbiome data (Supplementary Fig. 20). As has been previously seen in foxtail millet[28], we found that the rhizosphere microbiome combined with genetic data increased the average prediction accuracy ~7% of 11 agronomic traits compared to genetic markers alone (Supplementary Fig. 21). We then explored relationships among source environments, genetic differentiation and specific microbial taxa. As a measure for the pattern of similarity among samples, we calculated matrices of pairwise distance using the observed microbiome ASVs in different treatments, and two source environmental descriptors (*elevation* and *geographical distance*). Mantel tests were used to study the correlations between different distance matrices. On average, the correlations of inter-treatment and treatment-environment similarity patterns for bacterial communities were higher than for fungal communities (Supplementary Fig. 22). We observed that the correlation between the rhizosphere bacteria and source environments was significantly (one-tailed Student's *t*-test, *P* = 0.047) higher than that between the root bacteria and environments, although we did not detect any significant difference between such environment and rhizosphere or root fungi. To reduce dimensionality, we extracted the first five principal components (PCs) from the microbiome ASV data. We then used a Random Forest (RF) approach to predict these PCs using different environmental descriptors as explanatory variables (Supplementary Dataset 15). We observed the highest accuracy for the rhizosphere bacteria PC2 (Supplementary Fig. 23a)

using the nine most important environment predictors including *photosynthetically active radiation* and *potential evapotranspiration* (Supplementary Fig. 23b). Prediction of individual ASVs was less successful (Supplementary Fig. 24), although significant predictors were identified for specific taxa belonging to the *Oxalobacteraceae*, including *Massilia* (Supplementary Fig. 25). Based on these results, we hypothesize that a footprint of environmental adaptation on the genomes of the material tested is sufficiently constitutive to modulate plant and microbial traits when those are measured under different environmental conditions.

**A candidate gene linked to source environments, *Massilia* abundance, and root branching**

Across our samples, we detected four highly abundant (relative abundance >5%) bacterial families *Streptomycetaceae*, *Oxalobacteraceae, Pseudonocardiaceae* and *Chitinophagaceae* (Fig. 3a), and three highly abundant fungal families *Aspergillaceae*, *Trichocomaceae* and *Nectriaceae* (Supplementary Fig. 26). In particular, the bacterial taxon *Oxalobacteraceae* showed the highest heritability among all the families under nitrogen limitation in our experiment (Fig. 3b). *Oxalobacteraceae* have been previously proposed to play an important role in maize tolerance to nitrogen limitation when grown in nitrogen-deficient soils[29]. To investigate evidence for adaptive host effects on *Oxalobacteraceae* abundance, we used our existing environmental RF model to predict *Oxalobacteraceae* ASV abundance for 1,781 previously genotyped traditional maize varieties sourced from diverse environments across Mexico[15] with the 129 accessions as the training set (Fig. 4a). Among the *Oxalobacteraceae* ASVs, the best predicted (RF model $R^2 = 0.28$) was ASV37, belonging to the genus *Massilia* (*Oxalobacteraceae*), in the root under low nitrogen treatment (Supplementary Fig. 27). We ran GWA using the predicted ASV37 abundance values of the 1,781 varieties and compared the results to GWA of the 129 accessions in the training set, finding more overlap between the two than predicted by chance (Fig. 4b). Our RF predictions are derived from non-linear combinations of environmental descriptors. As such, the predicted ASV abundance GWA is essentially an analysis of genotype-environment association, with the training set defining the cline most appropriate to the trait in question. The top GWA hit for predicted ASV37 root abundance under low nitrogen (SNP S4_10445603) fell within the gene Zm00001d048945 on chromosome 4 (Fig. 4a and c; Supplementary Dataset 20). Across the 1,781 panel, the minor allele at SNP S4_10445603 was more abundant at higher predicted ASV37 abundance but lower source soil nitrogen content (Fig. 4d), suggesting the hypothesis that allelic variation in Zm00001d048945 contributes to adaptation to nitrogen-poor soil by enhancing association with *Massilia*[29] (Supplementary Fig. 28).

**Figure 3. Dominating and heritable bacterial families in the maize root and rhizosphere microbiome under abiotic stresses. a**, Maximum-likelihood phylogeny of dominant bacterial families (n >5). Circle sizes along the branches of the tree indicate the number of ASVs observed in association with microbial families. Colour coded families are clustered at the phylum level. Bar plots describe the prevalence according to the proportional sample size. The heatmaps illustrate the standardized mean relative abundance and the estimated heritability of microbial families from the rhizosphere to the root. Left and right oriented triangles represent the enrichment or depletion of microbial families from the rhizosphere to the root. Up and down oriented triangles represent the increase or decrease on heritability from the rhizosphere to the root. The significance levels were controlled at two levels (*: $P$ <0.05; **: $P$ <0.01). **b**, Phylogenetic tree of dominant bacterial ASVs ($n = 67$) of roots grown under nitrogen-poor condition. Dot size corresponds to relative abundance. Inner heatmap from outside to inside indicates heritability ($H^2$ >0.1) of individual ASVs and their annotated taxa at the genus and family levels. Red bar plots

23

describe the explained variance ($R^2$) by GWAS. The outer heatmap indicates the predictions by genomic best linear unbiased prediction (GBLUP), or based on the environmental best linear unbiased prediction (EBLUP) or prediction based on both genomics and environment (EGBLUP). Triangles indicate significant associations with the presence/absence (P/A) GWAS. Color coded tree branches of ASVs are clustered at the family level. Box plot indicates significantly higher heritability of *Oxalobacteraceae* compared to other families.
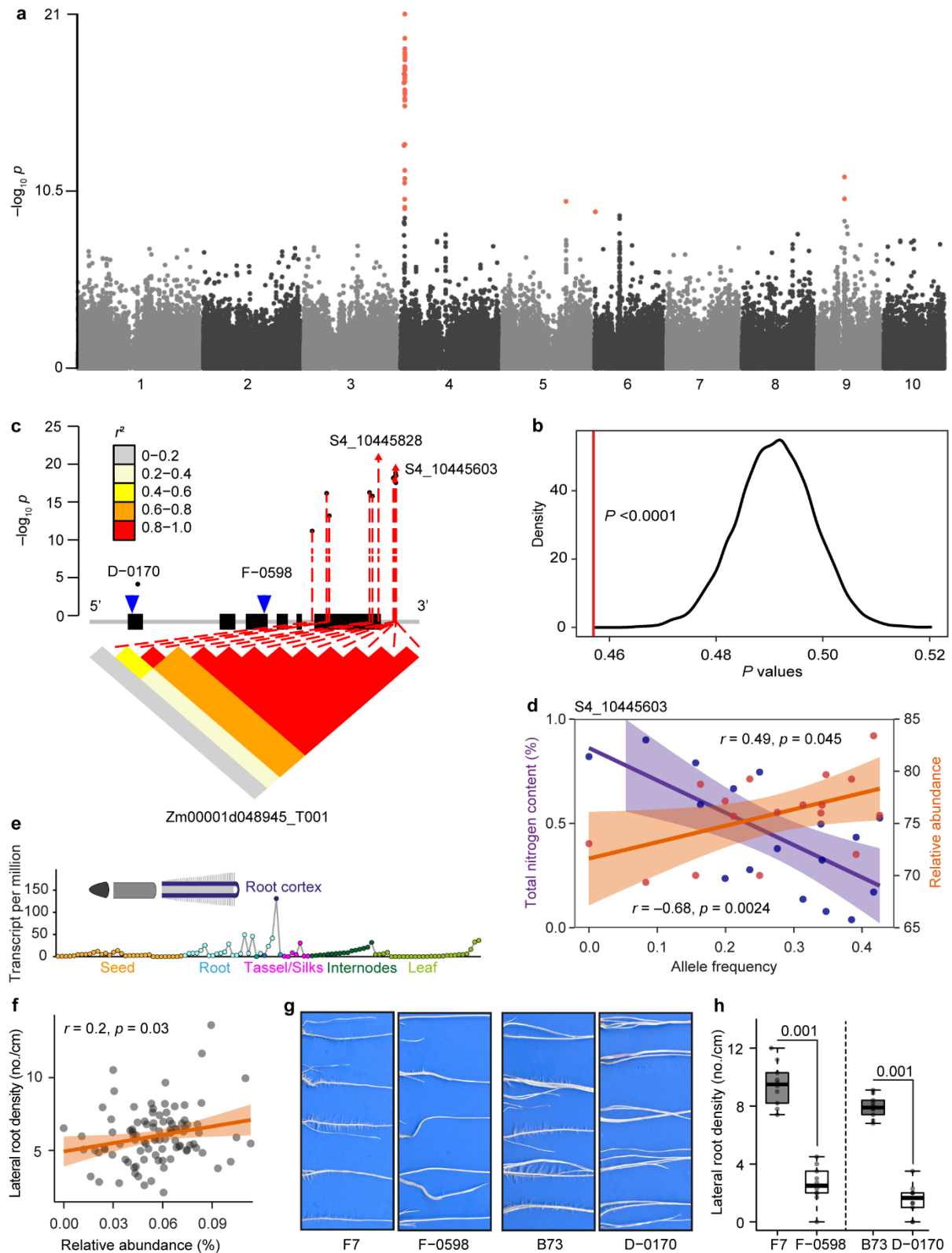


**Figure 4. Source habitats facilitate microbiome-driven root phenotypic association with nitrogen availability. a**, Manhattan plots showing environmental GWAS of specific predicted abundance of *Massilia* ASV37 by Random Forest model. **b**, Permutation test results showing significantly lower median *P* value for SNPs

from the phenotypic GWAS that are around 200kb of the top 100 SNPs of predicted GWAS (red vertical line) than the median $P$ value of random selected SNPs of phenotypic GWAS, based on 10000 permutations. **c**, Linkage disequilibrium (LD) plot for SNPs within 2.5kb of gene Zm00001d048945. Exons in the gene model are indicated by black bins. All significant SNPs are linked (red) to the LD plot ($P < 1.0 \times 10^{-7}$). Arrows indicate the positions of the peak SNPs. The colour key (grey to red) represents linkage disequilibrium values ($r^2$). Blue triangles indicate the transposon insertion positions of the two mutant alleles D−0170 and F−0598. **d**, Pearson correlation coefficient analysis of allele frequency (S4_10445603) with soil total nitrogen content (purple) and predicted relative abundance of ASV37_Root_LN (orange) across 1,781 geographical locations worldwide. **e**, Tissue-specific expression of gene Zm00001d048945 according to the eFP Browser database. **f**, Pearson correlation coefficient analysis of lateral root density with relative abundance of ASV37_Root_LN (orange) among 97 maize landraces. Scatter plots show best fit (solid line) and 95% confidence interval (colour shading) for linear regression. **g** and **h**, Root phenotypes and lateral root density of two independent Mu-transposon insertion mutant alleles (D−0170 and F−0598) in comparison to the corresponding wild types (B73 and F7) grown under the paper-roll system. Significances are indicated between wild type and mutant for different genetic backgrounds (two-tailed Student's $t$-tests). Boxes span from the first to the third quartiles, centre lines represent the median values and whiskers show data lying within 1.5× interquartile range of the lower and upper quartiles. Data points at the ends of whiskers represent outliers.

The gene Zm00001d048945 is most strongly expressed in the root cortex (Fig. 4e; https://www.maizegdb.org/gene_center/gene/Zm00001d048945) and is predicted to encode a TPX2 domain containing protein related to the WAVE-DAMPENED2 microtubule binding protein that functions in *Arabidopsis* root development[30] and lateral root initiation[31]. Using root architectural data available for the training set, we found a significant positive correlation between lateral root density and ASV37 abundance (R = 0.2, $P = 0.03$; Fig. 4f), suggesting recruitment of *Massilia* might be linked to root development. To functionally test for an effect of Zm00001d048945 on root architecture and *Massilia* abundance, we identified transposon insertional mutants in two different genetic backgrounds (Inbred: B73 and F7; Supplementary Fig. 29). Plants homozygous for transposon insertions in Zm00001d048945 showed a significant reduction in lateral root density in a paper-roll system (Fig. 4g and h). To then determine an effect on *Massilia*, we grew wild-type and mutant plants in high and low nitrogen treatments using the same soil as for our initial screen and characterized the root microbial community. In general, nitrogen limitation significantly increased the α-diversity of the root microbiome (Fig. 5a), meanwhile compartment (PERMANOVA, $R^2 = 0.37$, $P < 0.001$) plays larger effect on the microbiome composition than the treatment (PERMANOVA, $R^2 = 0.27$, $P < 0.001$) (Fig. 5b), which is consistent with the result of our initial screen. Interestingly, we only detected three bacterial taxa *Massilia*, *Muribaculaceae* and *Pseudomonas* – that differed in relative abundance between wild type and mutant plants in nitrogen-poor soil, and both mutants recruited significantly (two-tailed Wilcoxon rank-sum test, $P < 0.05$) less of the genus *Massilia* in the root than wild type plants although the overall microbial diversity did not differ ($R^2 = 0.030$, not significant; Fig. 5b and c). No difference was seen in *Massilia* abundance between wild type and mutant plants in the well fertilized soil (Supplementary Dataset 21). Taken together, we interpret these results to support a role for Zm00001d048945 in the modulation of

lateral root development and *Massilia* abundance in nitrogen deficient soil.
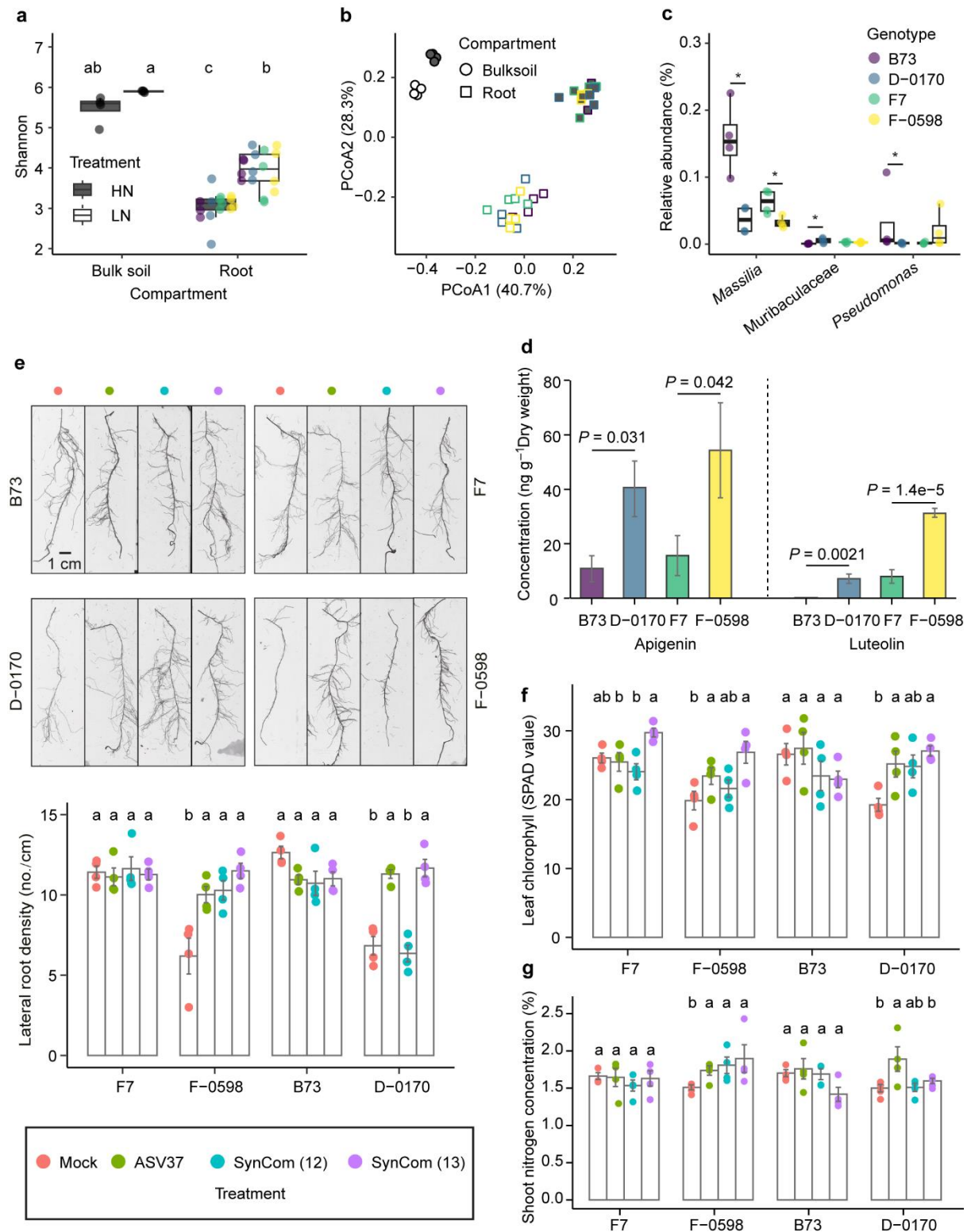


**Figure 5. *Massilia* alone can modulate lateral root development and growth performance in nitrogen-poor soil. a**, Root bacterial α-diversity (Shannon's diversity index) from mutants (D−0170 and F−0598) and wild type plants (B73 and F7) under low (LN) and high (HN) nitrogen conditions. The bulk soil collected from the unplanted pots are the control. Significances were indicated among treatments and compartments by different letters at *P* <0.05 (Kruskal-Wallis test, Dunn's *post-hoc* test). Boxes span from the first to the third quartiles, centre lines represent median values and whiskers show data lying within 1.5× interquartile range of lower and upper quartiles. Data points at the ends of whiskers represent outliers. Bulk soil, *n* = 4; Root, *n* = 16. **b**, Principal coordinate analysis (PCoA) showing the dissimilarity of bacterial β-diversity between mutants and wild type plants under low (LN)

and high (HN) nitrogen conditions. The explained variance ($R^2$) by compartment ($R^2 = 0.37$, $P$ <0.001), treatment ($R^2 = 0.27$, $P$ <0.001), genotype ($R^2 = 0.030$, not significant) and treatment:genotype interaction ($R^2 = 0.072$, $P$ <0.05) were assessed by permutational analysis of variance (PERMANOVA, $P$ <0.01). Bulk soil, $n = 4$; Root, $n = 16$. **c**, Significantly differential abundant taxa between mutants and wild type plants. Significances were indicated by asterisk at $P$ <0.05 (two-tailed Wilcoxon rank-sum test). $n = 4$. **d**, Targeted metabolite profiling indicates that the lateral root defective mutants produce more flavone than their respective wild type plants. For the extracts, $n = 6$ biologically replicates. **e**, Specific *Massilia* ASV37 is able to promote lateral root formation of lateral root defected mutants (D−0170 and F−0598) by root inoculation of different synthetic communities (SynCom). B73 and F7 are wild type plants. Representative images of 1st whorl of crown roots illustrate the more emerged lateral roots by *Massilia* strains. Different letters indicate significantly different groups (ANOVA, Tukey's HSD, $P$ <0.05). $n = 4$ biologically replicates. Scale bar = 1 cm. *Massilia* inoculations are able to alleviate the nitrogen deficient phenotype for leaf chlorophyll concentration (**f**) and shoot nitrogen concentration (**g**). Nitrogen deficient phenotype was evaluated by relative leaf chlorophyll concentration measured by the SPAD value of the last fully expanded leaf. Each individual leaf was measured 10 times. Different letters indicate significantly different groups (ANOVA, Tukey's HSD, $P$ <0.05). $n = 4$ biologically replicates. Data are mean ± s.e.m. Scale bar = 1 cm. Prior to ANOVA analysis, the observed values were checked for normal distribution and the homogeneous variance among the groups.

**Inoculations of the bacterial keystone taxon *Massilia* alone is sufficient to promote root and shoot growth**

Our previous study has shown that root-derived flavones i.e. apigenin and luteolin are important drivers for mediating the beneficial association of *Massilia* with lateral root development in maize[29]. We quantified apigenin and luteolin in the Zm00001d048945 mutants and found that the mutants accumulated significantly more apigenin and luteolin in comparison to wild type plants (Fig. 5d). Thus, together with our published work and this study, we confirmed that the potential linkage between lateral root development and *Massilia* depends on root exudation of flavones in maize. To characterize the specificity of the impact of *Massilia* on maize root and shoot growth, we performed controlled inoculation experiments with *Massilia* Isolate13 (100% sequence similarity with ASV37) alone, with a 12-member synthetic bacterial community (SynCom12) of *Massilia* isolates that did not include Isolate13, and with a 13-member *Massilia* (SynCom13) including SynCom12 and Isolate 13 (Supplementary Dataset 22). We quantified root and shoot growth in wild type and Zm00001d048945 transposon inserted mutants in nitrogen-poor soil. We found that Isolate 13 alone significantly promoted lateral root formation in Zm00001d048945 mutants (Fig. 5e). However, beneficial effect of *Massilia* is not necessary for the growth of wild type plants with well-developed lateral roots (Fig. 5e). These data together with previous finding[29] suggest that lateral root promotion might depend more on specific functions of *Massilia* even at the individual strain level. Moreover, single inoculation of *Massilia* Isolate13 significantly increased the content of chlorophyll in freshly formed leaves (Fig. 5f) and shoot nitrogen concentration (Fig. 5g) in Zm00001d048945 mutants under nitrogen deficient condition. More importantly, we confirmed that the growth promotion mediated by *Massilia* was consistent across two distinct nitrogen-poor soils, albeit fine-tuned by edaphic factors alone and/or interacting with specific inoculants (Supplementary Fig. 30).

Significantly, the microbial hub taxon *Massilia* alone can contribute to lateral root formation, biomass production and nitrogen tolerance of maize, indicating the potential value of root trait interactions with keystone microbial taxa when breeding for crop resilience. Nevertheless, further characterization of inoculated specimens may reveal the impact of strain variation on the *Massilia's* plant growth-promoting potential. Overall, our experimentally and computationally guided analyses (Supplementary Fig. 31) of host root and rhizosphere microbiome association identified maize genetic variation modulating bacterial microbiota through specific gene regulation, which facilitates synergistic interaction between root development and nitrogen deficiency (Fig. 6). Such causal association mirrored a footprint of environmental adaptation on maize genome, which confers to selective advantage and specified function of rhizosphere microbiota in the future climate constraint.
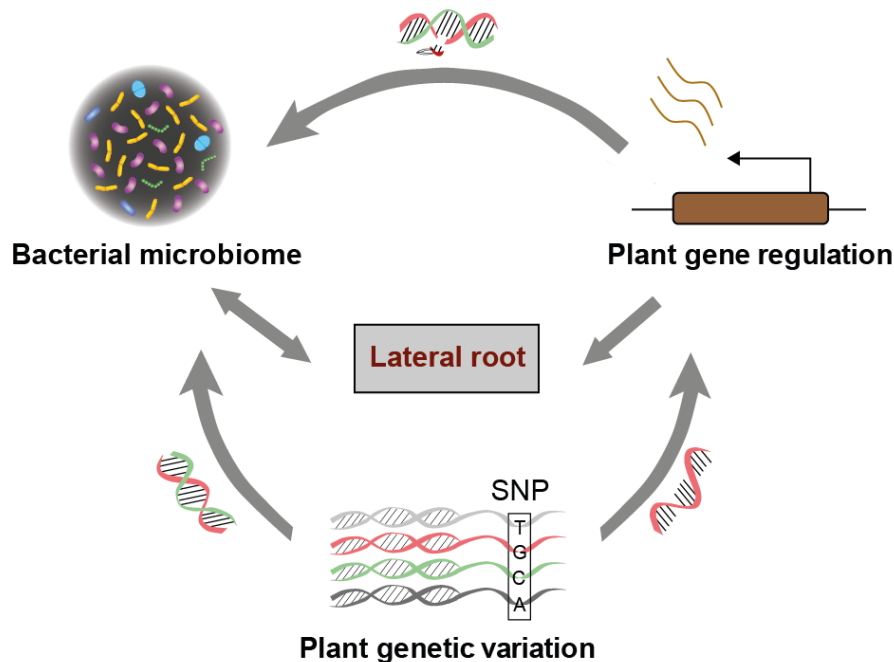


**Figure 6. Schematic illustration of the role of host plant genetic variation and gene regulation on bacterial microbiota-mediated lateral root development in maize.** Large-scale experimental and computational analyses identified genetic variation largely explain microbiome abundance and heritability under abiotic stresses, mirroring a footprint of local adaptation on the differentiation of the host genome that facilitates the modulation of plant microbiome. Genomic and environmental prediction guided genome-wide scan demonstrated specific gene regulation underlying lateral root development and recruitment of specific bacteria *Massilia*, which boost root development and plant nitrogen uptake in nitrogen-poor soil. The concept of the model is modified from[90].

## Discussion

During domestication plants have developed high productivity and environmental resilience, but may have also lost beneficial microbiome-associated traits compared with their wild relatives[9,10]. Thus, bringing back important plant traits supporting beneficial microbes from wild relatives and broader crop diversity may contribute to adaptation of crops to future climatic

challenges. In this study, we investigated the host-microbiome association and tried to understand whether and how source environments of traditional varieties relate to microbiome assembly under multiple abiotic stresses in maize. Examination of microbiomes across diverse germplasm demonstrated that plant genotype significantly impacts the microbiome, more so under abiotic stresses. Our genetic and environmental analyses support the hypothesis that plant genetic variation impacts microbiome assembly in crops[28,32-35]. Bacterial composition and structure support important functions in the rhizosphere under harsh environments[36,37] and is a heritable trait across environments[38,39]. Overall, it seems a conserved pattern of heritable taxa in cereal species such as maize[34,40], sorghum[32] and foxtail millet[28], suggesting host selection of their associated microbiota during evolution. We report here a significant improvement in plant trait prediction when combining rhizosphere microbiome with plant genetic data. Binominal regression and correlation analyses between microbial traits and source environmental variables among traditional varieties suggest that microbiome assemblage may contribute to beneficial plant trait-microbe associations underlying stress-resilience. Nonetheless, such local environmental information together with the plant genomic markers largely deepen our understanding the structure and function of the rhizosphere microbiome even in a new stress environment.

Although environmental conditions were dominant drivers of the crop microbiome, we found certain microbial taxa that were consistently influenced by genetic variability in maize, and whose abundance was correlated with plant traits. Given the plant genotypic variation and high affinity to local microbiota[39,41], we therefore hypothesize that a footprint of local adaptation on the differentiation of the host genome that facilitates the modulation of rhizosphere microbial assemblage. The endogenous genetic program that underlies root development can coordinate microbiome assembly[42] and plant mineral nutrient homeostasis[4]. Notably, we found that environment-associated alleles may promote root differentiation and microbiome-driven nitrogen deficiency tolerance in controlled conditions. These results provide strong support for a genetic basis of variation in the abundance of the bacterial taxon *Massilia* (*Oxalobacteraceae*) under nitrogen deficiency, illustrating the importance of specific bacteria for root development[3], nitrogen nutrition[43] and reciprocal interaction[29] at the strain level. Interestingly, a study in tomato identified a QTL region significantly associated with *Massilia* with larger effect size than other genera in the rhizosphere[35]. Taken together, this study advances the current understanding of the plant-trait-microbiome interactions that connect genetic variation to microbiome composition among a broad array of maize and their relatives in multiple environmental treatments, as well as identifying a specific gene with a compelling association

with both the environment and a bacterial taxon *Massilia*. Our results confirm that host genetic variation impacts keystone microbes in a consistent way across different environments. These findings help to close the knowledge gap between how plants impact the soil microbiome and how this functional interaction of the microbiome can be translated into crop resilience to nutrient limitation.

## Material and Methods

### Plant material, soil collection and growth conditions

The germplasm used in this study was selected to represent a broad diversity ranging from the maize progenitor teosinte to local open pollinating landraces and modern inbred lines and hybrids (Supplementary Dataset 23; Supplementary Fig. 1). We obtained the 11 geographically diverse teosinte accessions from the North Central Regional Plant Introduction Station (NCRPIS) and the International Maize and Wheat Improvement Center (CIMMYT). Moreover, we received the 97 landrace accessions from NCRPIS and these accessions were derived from the ten American countries which cover the major domestication areas of maize (Supplementary Fig. 1a). The modern breeding germplasm includes seven genetically diverse inbred lines[44] covering the major heterotic groups stiff-stalk and non-stiff stalk and four additional tropical inbred lines (Supplementary Fig. 1b). We have produced the ten hybrids by crossing the ten inbred lines with the reference inbred line B73 as the common mother plant (Supplementary Fig. 1c). Soil used for phytochamber pot experiments was dug from the Dikopshof long-term fertilizer field experiment established in 1904 near Cologne, Germany (50°48′21″N, 6°59′9″E) (Supplementary Fig. 2a). In this study, we collected soil subjected to three different fertilization managements including control soil fertilized with all nutrients, low nitrogen soil fertilized without nitrogen and low phosphorus soil fertilized without phosphorus as defined by[45]. The general soil type is classified as a Haplic Luvisol derived from loess above sand. Approximately the first 0-20 cm of the soil were collected and placed in a clean plastic bag. Subsequently, collected soil was dried at room temperature in clean plastic trays for about one week and sieved with a 4 mm analytical sieve (Retsch, Haan, Germany) to remove stones and vegetative debris. The sieved soil for the whole experiment was then homogenized with a MIX125 concrete mixer (Scheppach, Ichenhausen, Germany) (Supplementary Fig. 2a). The air-dried soil was ground into powder for the analysis of carbon, nitrogen, phosphorus and five metal elements (K, Fe, Mn, Cu, Zn). Soil pH was measured in deionized water (soil: solution ratio, 1:2.5 w/v) using a pH-meter 766 (Knick, Berlin, Germany). The basic physical and chemical properties of these soils are provided in Supplementary Table 1.

Local landraces used in this study were collected from local farmer fields or home gardens or natural habitats from the year 1954 to year 1994. The climate conditions and soil properties of the regeneration fields are typically representative of the natural habitats. Local landraces are open-pollinated varieties and can vary largely on seed traits. Therefore, we covered a broad geographic area but also confirmed the homogeneity of the 97 landraces concerning seed size,

seed color, and seed quality prior our phytochamber experiments (Supplementary Fig. 2b). Seeds were washed for 5 min with sterile water, followed by surface-sterilized for 2 min with 70% (v/v) ethanol, and incubated with a bleach solution (29 ml of sterile water, 15 ml of NaClO 12–13% (v/v) stock solution of 1 ml Tween 20), and rinsed 3 times with sterile deionized water[29]. Such sterilization procedure can well remove the seed-borne microbes and a certain number of seed endophytes. The sterilized seeds were pre-germinated for 3 days in a paper roll system using germination paper (Anchor Paper Co., St. Paul, MN, USA) with sterile deionized water. Then seedlings with primary roots of ca. 1–2 cm length were transferred to soil-filled pots ($7 \times 7 \times 20$ cm$^3$) in a 16/8-h light/dark, 26/18 °C cycle and were grown for 4 weeks in a walk-in phytochamber. A detailed sowing and transfer plan is provided in Supplementary Fig. 2c. No additional fertilizer was added.

**Experimental design and treatments**

The experiment was performed in a split plot design with three replications comprising four stress treatments on the main plots (trays) (Supplementary Fig. 32), e.g. fully fertilized control (CK) soil, no nitrogen fertilized low nitrogen (LN) soil, no phosphorus fertilized low phosphate (LP) and CK soil with drought (D) treatment. As controls, we used six pots without plants as 'bulk soil' samples (B), which were distributed across the main plots. Each tray contained a similar number of pots (subplots) with the different genotypes and bulk soil. The three replicates were performed at three different periods in the same phytochamber (Supplementary Fig. 32). For each stress treatment, we generated an alpha design for the genotypes and controls with three replicates and four incomplete blocks per replicate. The incomplete blocks were assigned to trays and replicates corresponded to the three replications of the experiment in time. To facilitate watering, pots subjected to the same treatment were allocated on the same tray. These trays were further randomized in the chamber. Distribution of all pots in each tray were randomized using a true random generator (excel function "RAND"), and trays were reshuffled every week in the phytochamber without paying attention to the pot labels. Since soil water availability will significantly affect the harvest of the rhizosphere and initiation of crown roots, we have performed a preliminary experiment with different water regimes (i.e. 33%, 22%, 17% water holding capacity) to ensure the establishment of suitable drought conditions and to facilitate rhizosphere harvesting and the optimal formation of the different whorls of crown roots (Supplementary Fig. 2c and 33). In brief, different volumes of sterilized water e.g. 60 ml, 40 ml, 30 ml were mixed with 500 g dry soil by spraying water and were then homogenized with a 4 mm sieve (Retsch). Each water regime was maintained by spraying water to the soil surface according to the weight loss of water during the 4-week culture. Plant height, total leaf

area, shoot and root fresh biomass from the representative genotypes B73 and Mo17 were recorded. Moreover, the multifunctional device COMBI 5000 (STEP Systems, Nuremberg, Germany) was used to measure soil variables e.g. soil moisture and electronic conductivity directly in each soil pot during each experimental run. The covariates including sample harvest time, ID of person performing DNA extraction together with the determined soil variables were collected and used for downstream data analysis (Supplementary Dataset 24).

**Characterization of native collection sites of maize landraces**

Geographical coordinates and elevation information of the collection sites for maize landraces were retrieved from the public database of the U.S. National Plant Germplasm System (https://www.grin-global.org/) and provided in Supplementary Dataset 23. Most of the landraces were received in the years 1980-1994 and were maintained by NCRPIS. To get the climate and soil variables based on the geographical coordinates for each site, we first compiled climatic and soil descriptors representative of the long-term averages of their point of origin, following methods in[46]. All used databases are publicly available and have global coverage. Data was collected from WorldClim[47], the NCEP/NCAR reanalysis project (https://psl.noaa.gov/data/reanalysis/reanalysis.shtml)[48], NASA SRB (https://asdc.larc.nasa.gov/project/SRB), Climate Research Unit (CRU)[49], SoilGrids[50] and the Global Soil Dataset (GSD)[51]. All 156 bioclimatic and soil variables were merged with the maize germplasm identity into the Supplementary Dataset 15. The related information of total soil nitrogen, available phosphorus, and annual precipitation are provided in the Supplementary Fig. 34.

**Determination of shoot phenotypic traits and ionome profile**

Aboveground phenotypic traits were determined for all 129 genotypes on the day of harvest in the phytochamber. The leaf area and chlorophyll index as measured by SPAD were determined as described accordingly[29] and are provided in Supplementary Dataset 25. The complete aboveground part of maize plants excluding the seed was harvested and heat treated at 105 °C for 30 min, dried at 70 °C to constant weight, weighed as the shoot dry biomass and then ground into powder. Approximately 6 mg of ground material was used to determine total nitrogen concentration in an elemental analyzer (Euro-EA, HEKAtech). Data were then calculated into peak areas by the software Callidus, providing quantitative results using reference material as a calibration standard. The same plant material was used to determine the concentrations of 13 additional mineral nutrients. In brief, approximately 200 mg of same ground material was weighed into polytetrafluoroethylene digestion tubes, and concentrated nitric acid (5 ml, 67–

69%; Bernd Kraft) was added to each tube. After 4 h of incubation, samples were digested under pressure using a high-performance microwave reactor (Ultraclave 4, MLS). Digested samples were transferred to Greiner centrifuge tubes and diluted with deionized (Milli-Q) water to a final volume of 8 ml. Element analysis was carried out by Inductively Coupled Plasma-Optical Emission Spectroscopy (iCAP 7400 duo; Thermo Fisher Scientific). For sample introduction a SC-4 DX autosampler with prepFAST Auto-Dilution System (ESI, Elemental Scientific) was used. A three-point external calibration curve was set from a certified multiple-standards solution (Custom Multi-Element Standard_PlasmaCAL, S-prep GmbH). The element Yttrium (ICP Standard Certipur®, Merck) was infused online and used as internal standard for matrix correction. All ionome data including concentrations and contents of all mineral nutrients are provided in the Supplementary Dataset 26.

**Root and rhizosphere samples harvest for microbiome analysis**

The root and rhizosphere samples collection were performed from 4-week-old maize plants as previously described[29]. In brief, whole root systems were carefully taken out from each pot and vigorously shaken to remove all soil not firmly attached to the roots. During this stage, most genotypes have consistently started to form the 2$^{nd}$ whorl of shoot-borne crown roots with a length of 3-10 cm. To synchronize the harvest for precise comparisons among genotypes, we collected the fully developed 1$^{st}$ whorl of shoot-borne crown roots initiated from the coleoptilar node for all maize genotypes (Supplementary Fig. 3a). These crown roots were identified similarly developmental status with mature lateral roots. Two dissected crown roots with tightly attached soil were placed into a 15 ml Falcon (Sarstedt) tube and immediately frozen in liquid nitrogen and stored at -80 °C before extraction of rhizosphere soil. The rhizosphere samples were defined and extracted into PowerBead tubes (Mo Bio Laboratories) as described previously[29]. The root samples were harvested from another crown root from the same plant that immediately washed by tap water and rinsed with three times of sterilized water followed by tissue drying and placed in PowerBead tubes (Supplementary Fig. 3b). Sample processing steps for root and rhizosphere have been performed by a designated person to avoid systematic errors. The bulk soil samples were also collected from the unplanted pots. DNA extractions were performed soon after root and rhizosphere samples were harvested, following the PowerSoil DNA isolation kit (Mo Bio Laboratories) protocol.

**Amplicon library preparation and sequencing**

Amplicon library construction was processed with a similar workflow as previously described[29] (Supplementary Fig. 3c). In brief, for bacterial 16S rRNA gene libraries, the V4 region was

amplified using the universal primers F515 (5′ GTGCCAGCMGCCGCGGTAA 3′) and R806 (5′ GGACTACHVGGGTWTCTAAT 3′) (Caporaso et al., 2011). For fungal amplicon sequencing, the ITS1 gene was amplified by the primer pair F (5′ CTTGGTCATTTAGAGGAAGTAA 3′) and R (5′ GCTGCGTTCTTCATCGATGC 3′). PCR reactions were performed with Phusion High-Fidelity PCR Master Mix (New England Biolabs) according to the manufacturer's instructions. Subsequently, only PCR products with the brightest bands at 400-450 base pairs (bp) were chosen for library preparation. Equal density ratios of the PCR products were mixed and purified with the Qiagen Gel Extraction Kit. Sequencing libraries were generated using the NEBNext Ultra DNA Library Pre Kit for Illumina, following the manufacturer's recommendations and with the addition of sequence indices. The library quality was checked on a Qubit 2.0 Fluorometer (Thermo Scientific) and Agilent Bioanalyzer 2100 system. Finally, the qualified libraries were sequenced by 250-bp paired-end reads on a MiSeq platform (Illumina).

**16S rRNA gene and ITS1 gene sequence processing**

Raw sequencing reads were processed following a similar workflow as previously described[29]. Briefly, paired-end 16S rRNA amplicon sequencing reads were assigned to samples based on their unique barcode and truncated by cutting off the barcode and primer sequence. Paired-end reads were merged using FLASH (v1.2.7)[52] and the splicing sequences were called raw tags. Sequence analyses were performed by QIIME 2 software (v2020.6)[53]. Raw sequence data were demultiplexed and quality filtered using the q2-demux plugin followed by denoising with DADA2[54] (via q2-dada2). Sequences were truncated at position 250 and each unique sequence was assigned to a different ASV. Taxonomy was assigned to ASVs using the q2-feature-classifier[55] and the classify-sklearn naïve Bayes taxonomy classifier against the SSUrRNA SILVA 99% OTUs reference sequences (v138)[56] at each taxonomic rank (kingdom, phylum, class, order, family, genus, species). Mitochondria- and chloroplast-assigned ASVs were eliminated. Out of the remaining sequences (only features with >10 reads in ≥2 samples) were kept to build an ASV table. In order to study phylogenetic relationships of different ASVs, multiple sequence alignments were conducted using mafft (via q2-alignment)[57] and the phylogenetic tree was built using fasttree2 (via q2-phylogeny)[58] in QIIME 2. Those sequences that did not align were removed. ITS1 amplicon data were processed the same as 16S amplicon data except that used the UNITE 99% ASVs reference sequences (v10.05.2021)[59] to annotate the taxonomy.

**Statistical analyses for microbial community assembly**

In consideration of experimental design, here we treated the trays as the main plots for different treatments as a random effect. There were four trays per period/replicate, and a replicate effect was considered to account for differences between the three replicates. All downstream analyses were performed in R (v4.1.0)[60]. Briefly, ASV tables were filtered with ≥10 reads in ≥2samples. For α-diversity indices, Shannon index was calculated using ASV tables rarefied to 1,000 reads. For all the following analyses ASVs which express ≤0.05% relative abundance within ≤5% samples were filtered. After filtering taxa, the samples with ≤1000 reads were also removed. Bray–Curtis distances between samples were calculated using ASV tables that were normalized using 'varianceStabilizingTransformation' function from DESeq2 (v1.34.0) package[61] in R. Constrained ordination analyses were performed using the 'capscale' function in R package vegan (v2.5-7)[62]. To test the effects of compartment, treatment and genotype on the microbial composition community, variance partitioning was performed using Bray–Curtis distance matrix between pairs of samples with a permutation-based PERMANOVA test using 'adonis' function in R package vegan[62].

**Inter-kingdom associations by co-occurrence network analysis**

The method SPIEC-EASI (SParse InversE Covariance Estimation for Ecological Association Inference) implemented in SpiecEasi (v1.1.2) R package was used to construct the inter-kingdom microbial co-occurrence network associations[63] using relative abundance of ASVs data and network was visualized by Cytoscape (v3.9.1). For this network inference, only ASVs with relative abundance >0.05% in ≥10% samples were used. The filtered bacterial and fungal ASV table were combined as the input followed by the default centered log-ratio (CLR) transformation. The neighborhood selection measured by the Meinshausen and Bühlmann (MB) method[64] was selected as the inference approach. The number of subsamples for the Stability Approach to Regularization Selection (StARS) was set to 99. The keystone taxa or ASVs were defined as those with a hub score greater than 0.1 within the co-occurrence network associations.

**Functional prediction of microbial ASVs**

FAPROTAX (v1.2.4)[65] was used to predict metabolic functions of bacteria given their taxonomy and ASV table. Python script collapse_table.py was performed to obtain the function report. FUNGuild (v1.1)[66] was used by FUNGuild.py Python script through providing taxonomy information. Functional results were extracted by matching the provided taxonomy to the FUNGuild database.

**Genotyping of 129 maize genotypes**

Genomic DNA was extracted from leaves of bulked maize seedlings subjected to different treatments and replicates for each genotype (Supplementary Fig. 3). The genetic variation across the maize genotypes was characterized using a GenoBaits Maize40K chip containing 40 K SNP markers, which was developed using a genotyping by target sequencing (GBTS) platform in maize[67]. In brief, DNA fragmentation, end-repair and adding A-tail, adapter ligation and probe hybridization were performed. After ligation of the adapters and clean up, fragment size selection was done with Beckman AMPureBeads and a PCR step to enrich the library. Quantity and quality of the libraries were determined via Qubit™ 4 Fluorometer (Invitrogen) and Agilent 2100 Bioanalyzer, respectively. In total, 129 qualified and enriched libraries were sequenced as 250-400 bp on an MGISEQ-2000 (MGI, Shenzhen, China). The quality of raw sequencing reads was assessed and filtered by fastp (version0.20.0, www.bioinformatics.babraham.ac.uk/projects/fastqc/) with the parameters (-n 10 -q 20 -u 40). The clean reads were then aligned to the maize B73 reference genome v4 using the Burrows-Wheeler Aligner (BWA) (v0.7.13, bio-bwa.sourceforge.net) with the MEM alignment algorithm. The SNPs were then called using the UnifiedGenotyper tool from Genome Analysis Toolkit (GATK, v3.5-0-g36282e4, software.broadinstitute.org/gatk) SNP caller. The genetic distance matrix was calculated based on pairwise Rogers' distance[68]. A principal component analysis (PCA) was performed based on the filtered SNPs by GCTA software[69]. A phylogenetic tree (Supplementary Fig. 35) was generated using the neighbour-joining method as implemented in Mega 10.0.4 with 1,000 bootstraps using MEGA-X[70].

**Analyses of phenotypic data**

For the three plant phenotypes (SPAD, leaf area and biomass), we first performed the outlier test using the following model for a given stress treatment:

$$y_{ijk} = \mu + \beta_{t(i)} + g_i + r_j + b_{jk} + e_{ijk}, \qquad (1)$$

where $y_{ijk}$ is the observation of the $i$-th genotype in the $k$-th block of the $j$-th complete replicate. $\mu$ is the general mean, $\beta_{t(i)}$ is the effect of the $t(i)$-th subpopulation ($t(i)$ indicates the subpopulation that the $i$-th genotype belongs to. There are four subpopulations: teosinte, landraces, inbred lines and hybrids.), $g_i$ is the effect of the $i$-th genotype, $r_j$ is the effect of the $j$-th replicate, $b_{jk}$ is the effect of the $k$-th block nested within the $j$-th replicate and $e_{ijk}$ is the residual term. All effects except the general mean were assumed to be random and follow an independent normal distribution.

After fitting the model, the residuals were standardized by the rescaled median of absolute deviation from the median (MAD) and then a Bonferroni-Holm test was applied to flag the outliers[71].

For all traits including fitness phenotypes and microbial traits, we estimated the broad-sense heritability (also referred as repeatability in this case) in each treatment. The following model was used to estimate the heritability:

$$y_{ijk} = \mu + g_i + r_j + b_{jk} + e_{ijk}, \tag{2}$$

where all notations were the same as in (1).

The heritability was calculated using the following formula:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2/R}, \tag{3}$$

where $\sigma_g^2$ and $\sigma_e^2$ are the estimated genotypic and residual variance, $R$ is the number of replications.

The best linear unbiased estimations (BLUEs) of all genotypes for each trait in each treatment were obtained by fitting Model (2) once more, assuming the general mean and genotypic effects are fixed and all other effects are random. All linear mixed models were fitted using the software ASReml-R 4.0[72].

**Statistical framework for GWAS**

Prior to GWAS, we first performed quality control for the genotypic data. In brief, the missing genotypic values were imputed using the software Beagle 5.2[73]. After imputation, we removed the markers with minor allele frequency (MAF) <0.05. As heterozygous loci were very common in our data set, we also removed markers whose maximum genotype frequency is >0.95. In total, 157,785 SNP markers were used for GWAS. For all traits, GWAS was performed separately for each treatment (i.e., using the BLUEs within the treatment as the response variable). For microbiome ASVs and alpha-diversity traits, only those with a heritability >0.1 were used for GWAS.

A standard "Q+K" linear mixed model[74] was used in GWAS. More precisely, the model is of the following form:

$$y = X\beta + ma + g + e, \tag{4}$$

where $y$ is the $n$-dimensional vector of phenotypic records (i.e. BLUEs within a certain treatment, $n$ is the number of genotypes), $\beta$ is the $k$-dimensional vector of fixed covariates

including the common intercept and the subpopulation effects. $\boldsymbol{X}$ is the corresponding $n \times k$ design matrix allocating each genotype to the subpopulation it belongs to. $a$ is the additive effect of the marker being tested, $\boldsymbol{m}$ is the $n$-dimensional vector of marker profiles for all individuals. The elements in $\boldsymbol{m}$ are coded as 0, 1 or 2, which is the number of minor alleles at the SNP. $\boldsymbol{g}$ is an $n$-dimensional random vector representing the genetic background effects. We assume that $\boldsymbol{g} \sim N(0, \boldsymbol{G}\sigma_g^2)$, where $\sigma_g^2$ is the genetic variance component, $\boldsymbol{G}$ is the VanRaden genomic relationship matrix[75]. $\boldsymbol{e}$ is the residual term and $e \sim N(0, \boldsymbol{I}\sigma_e^2)$, where $\sigma_e^2$ is the residual variance component and $\boldsymbol{I}$ is the $n \times n$ identity matrix. After solving the linear mixed model, the marker effect was tested using the Wald test statistic $W = \hat{a}^2 / \text{var}(\hat{a})$, which approximately follows a $\chi^2$-distribution with one degree of freedom.

Strictly, the model needs to be fitted once for each marker to get the precise test statistic for each marker. But to reduce the computational load, we implemented a commonly used approximate approach, namely the "population parameters previously determined" (P3D) method[76]. That is, we only fit the model once without any marker effect (the so-called "null model"), and then we fixed the estimated the variance parameters $\sigma_g^2$ and $\sigma_e^2$ throughout the testing procedure. Then, the test statistic for each marker can be efficiently calculated. GWAS was implemented using R codes developed by ourselves. The variance parameters were estimated by the Bayesian method using the package BGLR[77].

For microbial traits, the significant marker-trait association (MTA) was identified with a threshold of $P < 0.05$ after Bonferroni-Holm correction for multiple test[78]. For fitness phenotypes and alpha-diversity, we used a more liberal threshold of $P < 0.1$ after Benjamini-Hochberg correction[79]. For each trait, the proportion of phenotypic variance explained by each MTA ($R^2$) was calculated as follows: A liner regression model was fitted with all MTAs identified for the trait under consideration. Then, the sum of squares for each MTA as well as the total sum of squares was calculated by ANOVA. The $R^2$ for each MTA was estimated as the sum of squares of the MTA divided by the total sum of squares. Finally, the genes associated with these MTAs were extracted and further functionally annotated and classified according to gene ontology terms using agriGO v.2 (http://systemsbiology.cau.edu.cn/agriGOv2/).

**GWAS for the presence/absence mode**

For microbial traits, we performed in addition a GWAS based on the presence/absence mode (PA-GWAS) in each treatment. Each ASV or taxonomy is considered as present if it is present in two or more replicates. As in the GWAS for abundance, ASVs and taxa with repeatability below 0.1 were filtered out. Those with a presence rate above 95% or below 5% were

considered as non-segregated and were also excluded from the analysis. The model for PA-GWAS is a logistic linear mixed model[80]. Briefly, the model can be described as follows.

$$\text{logit}(\boldsymbol{\pi}) = \boldsymbol{X\beta} + \boldsymbol{m}a + \boldsymbol{g}, \tag{5}$$

where $\boldsymbol{X}$, $\boldsymbol{\beta}$, $\boldsymbol{m}$, $a$ and $\boldsymbol{g}$ are the same as in (6). $\boldsymbol{\pi}$ is the vector of conditional probabilities given the covariates, marker effects and the genetic background effects. More precisely, for the $i$-th individual, $\boldsymbol{\pi}_i = P(y_i = 1 | \boldsymbol{X}_i, m_i, g_i)$, where $y_i$ is the binary variable indicating the presence ($y_i = 1$) and absence ($y_i = 0$), $\boldsymbol{X}_i$ is the $i$-th row of the matrix $\boldsymbol{X}$, $m_i$ is the $i$-th entry of the vector $\boldsymbol{m}$ and $g_i$ is the $i$-th component of the random vector $\boldsymbol{g}$. The logit function is defined as $\text{logit}(x) = \ln(x/(1-x))$.

Similar to the P3D approach, a null logistic linear mixed model $\text{logit}(\boldsymbol{\pi}_0) = \boldsymbol{X\beta} + \boldsymbol{g}$ was fitted using the penalized quasi-likelihood method[81]. The estimated variance components were then fixed throughout the test procedure. A score test was applied to assess the significance of the marker effects.

The PA-GWAS was conducted using the R package GMMAT[80].

**Prediction for microbial traits using the genomic data and environmental descriptors**

To see the correlation between host genetics and microbiome assemblage, Mantel test was first performed between Rogers' genetic distance matrix and microbial composition distance matrix only for landraces. After removing the treatment effect using linear model for normalized microbial abundances, the mean value of the residual for each genotype was used to calculate the Euclidean distance. Spearman correlation method was used in mantel function in R. Permutations = 9999.

Next, we investigated the prediction abilities for all microbial traits within each treatment using both the genomic data and the environmental characters. The following three models were implemented. To eliminate the noise of subpopulation effects, we only used the 97 landraces for this part of analysis.

*Model 1 (genomic prediction).* We applied the genomic best linear unbiased prediction (GBLUP)[75] which is the most commonly used model in genomic prediction. The model can be described as follows.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{g} + \boldsymbol{e}, \tag{6}$$

where the notations are the same as in (4). Note that by the use of the VanRaden genomic relationship matrix as the covariance matrix of $\boldsymbol{g}$, it implicitly modeled the additive effects of

all markers.

*Model 2 (prediction purely based on the environmental characters).* In this model, the genetic effects were replaced by the effects of the environmental characters, which were modeled in a similar way to the GBLUP. More precisely, the model has the following form:

$$y = X\beta + l + e, \tag{7}$$

where $l$ is the $n$-dimensional random vector representing the E-determined values for all individuals. We assume that $l \sim N(0, \Sigma\sigma_l^2)$ where $\sigma_l^2$ is the corresponding variance component, $\Sigma$ is a covariance matrix. Assuming that $L$ is the $n \times s$ matrix of standardized environmental character records ($s$ is the number of environmental characters), we have $\Sigma = LL'/c$ where $c$ is the mean of all diagonal elements in the matrix $LL'$.

*Model 3 (prediction based on both genomics and environmental characters).* In this approach, we combined the genomic data and the Es in a multi-kernel model, which is of the following form:

$$y = X\beta + g + l + e, \tag{8}$$

where the notations were inherited from (6) and (7).

The prediction abilities of the above three models were assessed in a leave-one-out cross-validation scenario. That is, each individual was predicted once using a training set consisting of all other individuals. Thus, for each trait the prediction model was fitted $n$ times. After we obtained the predicted values of all individuals, the prediction ability was calculated as the correlation between the predicted and observed values. The standard error was estimated using the bootstrap approach[82].

All prediction models were implemented using the R package BGLR[77] and rrBLUP[83].

**Prediction for plant phenotypes using the genomic and microbiome data**

We explored the possibility of predicting the three fitness phenotypes and ionome traits in each treatment using the genomic data and microbiomes. As in the last subsection, we focused on the subpopulation of 97 landraces.

*Scenario 1 (prediction based on microbiomes only).* In this scenario, we considered 9 cases, in which the phenotypes were predicted using bacteria in the root sample (BA_RO), in the rhizosphere sample (BA_RH), fungi in the root sample (FU_RO), in the rhizosphere sample (FU_RH), bacteria in both samples (BA), fungi in both samples (FU), both types of microbiomes in the root sample (RO), in the rhizosphere sample (RH), and both types of

microbiomes in both samples (ALL). The model can be uniformly described as follows:

$$y = \mathbf{1}_n \mu + \sum_{i=1}^{k} \boldsymbol{m}_i + \boldsymbol{e}, \qquad (9)$$

where $\boldsymbol{m}_i$ is an $n$-dimensional trait values for all individuals determined by a certain type of microbiome in a specific sample, $k$ can be 1 (BA_RO, BA_RH, FU_RO, FU_RH), 2 (BA, FU, RO, RH), or 4 (ALL), other notations are the same as in (8). We assume that $\boldsymbol{m}_i \sim N(0, \boldsymbol{V}_i \sigma_{m_i}^2)$, where $\sigma_{m_i}^2$ is the corresponding variance component, $\boldsymbol{V}_i$ is a covariance matrix derived from the microbiome ASVs. Assuming that $\boldsymbol{M}_i$ is the $n \times t$ matrix of standardized records of microbiome ASVs (t is the number of different ASVs), we have $\boldsymbol{V}_i = \boldsymbol{M}_i \boldsymbol{M}_i' / c_i$ where $c_i$ is the mean of all diagonal elements in the matrix $\boldsymbol{M}_i \boldsymbol{M}_i'$.

*Scenario 2* (*prediction based on both microbiomes and genomic data*). In this scenario, the 9 cases in Scenario 1 were combined with genomic data (G_BA_RO, G_BA_RH, G_FU_RO, G_FU_RH, G_BA, G_FU, G_RO, G_RH, G_ALL). The models are of the following form:

$$y = \mathbf{1}_n \mu + \boldsymbol{g} + \sum_{i=1}^{k} \boldsymbol{m}_i + \boldsymbol{e}, \qquad (10)$$

where the notations were adopted from (8) and (11).

As in the last subsection, the prediction abilities were evaluated in a leave-one-out cross-validation scenario. Prediction models were implemented using the R package BGLR.

**Effects of source environmental factors on specific microbial assemblies**

To better characterize different microbial assemblies and taxonomic shifts among variables, we applied a modified use of unbiased weighted correlation network analysis (WGCNA) to understand the biological relevance between microbial compositions and phenotypic traits[26,27]. WGCNA is a data-driven method that clusters ASVs to different modules based on weighted correlations between ASVs. First, we used WGCNA (1.72.1) in R to identify different microbial modules of bacterial ASVs whose differential representation was correlated across treatment groups and genotypes within root or rhizosphere compartment. For robust construction of co-expressed microbial networks, we filtered and normalized ASVs table as described above. The soft thresholding power β was automatically selected and used to calculate adjacency matrix. To minimize the effects of noise and false associations, we transformed the adjacency matrix into a topological overlap matrix (TOM) with selected power and calculated the corresponding dissimilarities (dissTOM) as 1 – TOM. For hierarchical clustering of ASVs, we used dissTOM as a distance measure and set the minimum module size (number of ASVs to 5) to detect

modules. Next, we quantified the similarities of entire modules and their "eigentaxa"[27] were calculated and subsequently used to associate with phenotypic traits (shoot dry biomass, shoot nitrogen concentration and content). We chose modules that have a Spearman correlation coefficient >0.1 and $P$ value <0.05 with different traits as significant associated modules. Network visualization was performed in Cytoscape (3.8.0) only for significant modules. Width and transparency of edges are proportional to weight exported from WGCNA.

To further explore the relationship between source environment of native habitats and specific microbial modules among landraces, we performed network association analyses between these microbial modules and sourced environment factors i.e. total nitrogen, phosphorus retention and annual precipitation. We then conducted a correlation network conformed by taxa associated with the root and rhizosphere microbiomes. We calculated all pairwise Spearman correlation coefficients among these microbial eigentaxa and sourced environmental factors, and kept all positive or negative correlations (R >0.1 or R <−0.1, $P$ <0.05). Total nitrogen, phosphorus retention and mean annual precipitation were obtained from the WorldClim database (https://www.worldclim.org/) as explained above. Structural equation modelling (SEM) was conducted to provide a system-level understanding on the direct and indirect associations between environmental factors, the proportion of modules and that of selected taxa from above-explained analyses. Because some of the variables introduced were not normally distributed, we used bootstrap tests in these SEMs. We evaluated the fit of these models using the model $\chi^2$-test, the root mean squared error of approximation and the Bollen–Stine bootstrap test.

**Environmentally adaptive loci and microbiome relatedness across abiotic stresses**

To determine if the environmentally associated loci are contributing to microbiome adaptation to abiotic stresses, we used a representative set of natural varieties e.g. 97 landraces accessions covering typical geographical range. Prior to analysis, PCA was conducted based on the BLUEs for each treatment and compartment to extract major sources of variance from bacterial and fungal microbial community data. The first five PCs were obtained for downstream analyses. PCA was performed using the prcomp function in R. In addition, we selected 18 individual ASVs belonging to *Oxalobacteraceae* to be predicted by Random Forest models. To improve model accuracy, feature selection was conducted prior to model building to eliminate unimportant or redundant environmental variables by identifying those with significant associations to an outcome variable. The feature selection method Boruta was employed to identify environmental aspects that describe significant variation in the PCs and ASVs using

Boruta::boruta() (v7.0.0)[84].

The subset of boruta-identified environmental variables (Supplementary Dataset 15) for each ASV were used for Random Forest model construction. This model works under the expectation that a response variable can be described by several explanatory variables through the construction of decision trees. Thus, each Random Forest model is representative of the non-linear, unique combination of explanatory variables that describe variation in a response variable. Random Forest models were built using RandomForest::randomForest() function under default parameters, 5000 trees were built and one third of the number of explanatory variables were tried at each split[85]. Random Forest models were trained with 80% of the data and validated with the remaining 20% test set. Model success was evaluated with percent error explained, Nash-Sutcliffe efficiency (NSE), mean absolute error (MAE), and mean squared error (MSE). Using constructed Random Forest models, ASVs were predicted for 1,781 genotyped landraces in Mexico. These landraces were genotyped as a part of the Seeds of Discovery project (SeeD).

We conducted genome wide association studies (GWAS) to measure the associations between SNPs of landrace genotypes and predicted microbial traits, as well as the associations between SNPs and the environmental variables used to predict the microbial traits. SNPs were filtered for minor allele frequency >1%. We applied the method as previously described[86], using a linear model to fit the genotypic data and each microbial trait and environmental variable for Mexican landrace accessions. The first five eigenvectors of the genetic relationship matrix were included in the model to control for population structure. To control for the number of false positive tests, we re-calibrated the $p$-values using the false discovery rate (FDR) control algorithm[87] and selected significant SNPs based on the calibrated results. To test if GWA hits based on the prediction is significantly better in capturing top GWA hits of observed data than random, we conducted a permutation test and compared the median p-value of GWA hits of observed data that are around 200kb of the top 100 prediction-based GWA hits and the median p-value of random selected GWA hits based on 10000 permutations.

**Association of allele frequency with soil nitrogen and microbial taxa**

To identify whether the microbiome is associated with environment and maize phenotypes, we performed allelic variation analysis of Zm00001d048945 using an SNP dataset of CIMMYT landraces accessions obtained from a previous study[15]. We extracted the genotypic information of top SNPs of the target gene Zm00001d048945 for all tested landraces. We divided maize landraces into 20 groups based on the total soil nitrogen content (%) of their sampling sites[51].

We calculated the mean total nitrogen, the minor allele frequencies (MAF) of the target SNPs, and the mean predicted ASV abundance for each group of landraces. Pearson correlation was conducted to test the correlations between MAF and total nitrogen content, and between MAF and ASV abundance.

**Candidate gene validation by independent transposon insertion alleles**

Gene expression for Zm00001d048945 was explored in qTeller (https://qteller.maizegdb.org/), which allows to compare gene expression across different tissues from multiple data sources. Gene expression data was extracted from different organs (seed, root, tassel/silk, internodes and leaf) and specific tissues such as the root meristematic zone, elongation zone, stele and cortex. The gene encoded protein annotation was inferred from UniProt database (https://www.uniprot.org/). We next identified potential loss-of-function mutations by exploring the sequence indexed collection *BonnMu*[88]. Induced maize mutants of the *BonnMu* resource derive from Mutator-tagged $F_2$−families in various genetic backgrounds, such as B73 and F7. We identified two insertion lines, *BonnMu*−8−D−0170 (B73) and *BonnMu*−F7−2−F−0598 (F7), harboring insertions 1,264 bp upstream of the start codon ATG and in the second exon of Zm00001d048945, respectively. These two families were phenotyped in paper-roll culture[29] and the seedling plants were scanned using the scanner Expression 12000XL (Epson, Suwa, Japan). Lateral roots were counted and the density was normalized with the measure number of lateral roots per cm length of primary root. Statistical analyses were performed by pair-wise Students *t* test with *F* statistics.

**Association of relative abundance of *Massilia* with lateral root density**

To understand the relationship between *Massilia* and the formation of lateral roots, root system architecture and morphology of 97 maize landraces was scanned with an Epson Expression 12000XL scanner. Lateral root density was determined by manual calculation as the number of emerged lateral roots per length (cm) of the main root. The linear correlation was plotted between lateral root density and relative abundance data of *Massilia* ASVs using R (v4.1.0).

**Microbiome community and targeted metabolites profiling in lateral root mutants and wild types**

To explore how the gene Zm00001d048945 mutation altered the abundance of *Massilia*, we have grown these transposon mutants and wild types in the soil pots as described in previous screen experiment and performed the 16S rRNA amplicon sequencing for the crown roots. Prior to the soil pot experiment, we genotyped for the target transposon insertion for these mutants

and wild types. In detail, approximate 200 mg of leaf sample from each individual plant was harvested and immediately frozen in liquid nitrogen, and ground to fine powder for DNA extraction. DNA was extracted using the CTAB method. The primers were designed using Primer 3 and BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) based on B73 reference genome 5.0. Plants carrying the mutant allele were identified by genotyping with an outward-facing primer recognizing the terminal inverted repeat (TIR) of the Mutator transposon, TIR6 (5′-AGAGAAGCCAACGCCAWCGCCTCYATTTCGTC-3′) and the Zm00001d048945 gene-specific primer (Forward, 5'-CGT CGG ATG AAA GCC TCA AG-3'; Reverse, 5'-AAA CCT GAA GCG AGC GTG TA-3'). The seedlings that confirmed homozygous positive for the insertion were used for samples harvest and sequencing. The microbial community analyse and bioinformatic pipeline was applied according to our previous analysis. Moreover, we profiled the targeted metabolite analysis for flavones i.e. apigenin and luteolin using the same root type based on our in-house established protocol[29].

**Synthetic community, root bacterial inoculation and plant biomass, root architecture and nitrogen tolerance assay**

Bacterial strains were isolated using R2A media supplemented with 100 µg mL-1 Cyclohexamid from the rhizosphere or rhizoplane of maize roots in the soil column experiment[89]. Isolates were picked randomly from plates with colony forming units (CFUs) range between 30-100 CFUs. Among 480 isolates, only strains affiliated to *Massilia* were used in the present study. To explore effects of specific *Massilia* ASV37 mapped on the gene Zm00001d048945 on root development and nitrogen uptake, a growth promotion assay by inoculation with either a single inoculation or a synthetic community of *Massilia* isolates (Supplementary Dataset 16) was performed on two maize wild types (B73 and F7) and their mutants (D−0170 and F−0598) in nitrogen-poor soil pots. Before inoculation of these *Massilia* strains, we first mapped the sequences of in total 13 *Massilia* strains (~1000 bp) to the 16S sequence of the GWAS mapped ASV37 (250 bp) using BLASTn (v2.6.0) with default parameters and we chose the 100% alignment Isolate13 as the candidate used in the following experiment. We applied three different inoculation strategies e.g. single inoculation using Isolate13 which has 100% identity with *Massilia* ASV37 (SynCom1), 12 *Massilia* isolates excluding Isolate13 and all 13 *Massilia* isolates including of Isolate13 (SynCom2) under nitrogen-poor condition. We performed inoculation experiments using two different soil types, one is the soil from Dikopshof as described in the large cultivation experiment and another natural soil was dug from a natural field at Campus Klein-Altendorf (University of Bonn), then sieved, homogenized and mixed with 50% quartz sand (WF 33, Quarzwerke Weferlingen,

Germany) to reduce the nitrogen content of the recipient soil. Prior to establishment of soil pot experiment, the soil mixture was sterilized with detailed protocol to ensure the best effect of removal of soil microbes. In brief, the soil was moistened with sterilized water and autoclaved in a liquid cycle for 25 min at 121 °C, then leave the soil at room temperature for at least 24 hours for a rest. Then moist the soil again if it looks dry with sterilized water and autoclave a second time in a liquid cycle for 25min 121 °C followed by incubation in the oven for 24 hours. Repeat the autoclaving steps if the soil still looks dry. Meanwhile, took samples at each autoclaving step to make sure not only that the soil was sterile, but also to rule out contaminations. However, the effect of autoclaving soil-derived byproducts on microbiota is not known, we therefore leave the autoclaved soil rest for at least 1 week. The seed sterilization, isolates preparation, root inoculation and growth assay were done according as previously reported[29]. Different genotypes were grown in the phytochamber (16/8 h light/dark and 26/18 °C) for 1 month and plants were harvested, and the length and weight of crown root, lateral root density and shoot fresh weight were determined. Chlorophyll content was determined as the average of 10 measurements with a SPAD-502 chlorophyll metre (Konica Minolta) in the middle third of the newest expanded leaf in the longitudinal direction. The whole experiment was done using two different soil types. The linear correlation was plotted between different root traits and shoot fresh weight and chlorophyll content using R (v4.1.0).

**Data availability**

All raw maize genotyping data, bacterial 16S and fungal ITS1 gene data in this paper were deposited in the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under the BioProject ID PRJNA889703 and PRJNA1015142. The SSUrRNA database from SILVA database (release 138, 2020, https://www.arb-silva.de/) and UNITE database (v8.3, 2021, https://unite.ut.ee/) were used for analysing the bacterial 16S and fungal ITS1 gene sequences, respectively. We deposited customized scripts in the following GitHub repository: https://github.com/Danning16/MaizeMicrobiome2022. All statistical data are provided with this paper.

## Acknowledgement

# References

1. Cheng, Y. T., Zhang, L. & He, S. Y. Plant-Microbe Interactions Facing Environmental Challenge. *Cell Host Microbe* **26**, 183-192 (2019).

2. Oldroyd, G. E. D. & Leyser, O. A plant's diet, surviving in a variable nutrient environment. *Science* **368**, eaba0196 (2020).

3. Finkel, O. M. et al. A single bacterial genus maintains root growth in a complex microbiome. *Nature* **587**, 103-108 (2020).

4. Salas-Gonzalez, I. et al. Coordination between microbiota and root endodermis supports plant mineral nutrient homeostasis. *Science* **371**, eabd0695 (2021).

5. Banerjee, S., Schlaeppi, K. & van der Heijden, M. G. A. Keystone taxa as drivers of microbiome structure and functioning. *Nat. Rev. Microbiol.* **16**, 567-576 (2018).

6. Bulgarelli, D., Schlaeppi, K., Spaepen, S., Van Themaat, E. V. L. & Schulze-Lefert, P. Structure and functions of the bacterial microbiota of plants. *Ann. Rev. Plant Biol.* **64**, 807–838 (2013).

7. Bulgarelli, D. et al. Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature* **488**, 91-95 (2012).

8. Lundberg, D. S. et al. Defining the core *Arabidopsis* thaliana root microbiome. *Nature* **488**, 86-90 (2012).

9. Haney, C. H., Samuel, B. S., Bush, J. & Ausubel, F. M. Associations with rhizosphere bacteria can confer an adaptive advantage to plants. *Nat. Plants* **1**, 15051 (2015).

10. de Vries, F. T., Griffiths, R. I., Knight, C. G., Nicolitch, O. & Williams, A. Harnessing rhizosphere microbiomes for drought-resilient crop production. *Science* **368**, 270-274 (2020).

11. Singh, B. K., Trivedi, P., Egidi, E., Macdonald, C. A. & Delgado-Baquerizo, M. Crop microbiome and sustainable agriculture. *Nat. Rev. Microbiol.* **18**, 601-602 (2020).

12. Meyer, R. S. & Purugganan, M. D. Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* **14**, 840-852 (2013).

13. Cordovez, V., Dini-Andreote, F., Carrión, V. J. & Raaijmakers, J. M. Ecology and evolution of plant microbiomes. *Annu. Rev. Microbiol.* **73**, 69–88 (2019).

14. Raaijmakers, J. M. & Kiers, E. T. Rewilding plant microbiomes. *Science* **378**, 599-600 (2022).

15. Navarro, J. A. R. et al. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat. Genet.* **49**, 476-480 (2017).

16. Hake, S. & Ross-Ibarra, J. Genetic, evolutionary and plant breeding insights from the

domestication of maize. *eLife* **4**, e05861 (2015).

17. Yu, P., Gutjahr, C., Li, C. & Hochholdinger, F. Genetic control of lateral root formation in Ccereals. *Trends Plant Sci.* **21**, 951-961 (2016).

18. Hochholdinger, F., Yu, P. & Marcon, C. Genetic control of root system development in maize. *Trends Plant Sci.* **23**, 79-88 (2018).

19. Szoboszlay, M. et al. Comparison of root system architecture and rhizosphere microbial communities of Balsas teosinte and domesticated corn cultivars. *Soil Biol. Biochem.* **80**, 34-44 (2015).

20. Brisson, V. L., Schmidt, J. E., Northen, T. R., Vogel, J. P. & Gaudin, A. C. M. Impacts of Maize Domestication and Breeding on Rhizosphere Microbial Community Recruitment from a Nutrient Depleted Agricultural Soil. *Sci. Rep.* **9**, 15611 (2019).

21. Wagner, M. R., Roberts, J. H., Balint-Kurti, P. & Holland, J. B. Heterosis of leaf and rhizosphere microbiomes in field-grown maize. *New Phytol.* **228**, 1055-1069 (2020).

22. Favela, A., Bohn, M. O. & Kent, A. D. Maize germplasm chronosequence shows crop breeding history impacts recruitment of the rhizosphere microbiome. *ISME J.* **15**, 2454-2464 (2021).

23. Bulgarelli, D. et al. Structure and Function of the Bacterial Root Microbiota in Wild and Domesticated Barley. *Cell Host Microbe* **17**, 392-403 (2015).

24. Yue, H. et al. Plant domestication shapes rhizosphere microbiome assembly and metabolic functions. *Microbiome* **11**, 70 (2023).

25. Durán, P. et al. Microbial interkingdom interactions in roots promote *Arabidopsis* survival. *Cell* **175**, 973-983 (2018).

26. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

27. Caslin, B. et al. Alcohol shifts gut microbial networks and ameliorates a murine model of neuroinflammation in a sexspecific pattern. *Proc. Natl. Acad. Sci. USA.* **116**, 25808-25815 (2019).

28. Wang, Y. et al. GWAS, MWAS and mGWAS provide insights into precision agriculture based on genotype-dependent microbial effects in foxtail millet. *Nat. Commun.* **13**, 5913 (2022).

29. Yu, P. et al. Plant flavones enrich rhizosphere *Oxalobacteraceae* to improve maize performance under nitrogen deprivation. *Nat. Plants* **7**, 481-499 (2021).

30. Yuen, C. et al. WVD2 and WDL1 modulate helical organ growth and anisotropic cell expansion in Arabidopsis. *Plant Physiol.* **131**, 493-506 (2003).

31. Qian, Y., Wang, X., Liu, Y., Wang, X. & Mao, T. HY5 inhibitsilateral root initiation in Arabidopsis through negative regulation of the microtubule-stabilizing protein TPXL5. *Plant Cell* **35**, 1092-1109 (2023).

32. Deng, S. et al. Genome wide association study reveals plant loci controlling heritability of the rhizosphere microbiome. *ISME J.* **15**, 3181-3194 (2021).

33. Escudero-Martinez, C. et al. Identifying plant genes shaping microbiota composition in the barley rhizosphere. *Nat. Commun.* **13**, 3443 (2022).

34. Meier, M. A. et al. Association analyses of host genetics, root-colonizing microbes, and plant phenotypes under different nitrogen conditions in maize. *eLife* **11**, e75790 (2022).

35. Oyserman, B. O. et al. Disentangling the genetic basis of rhizosphere microbiome assembly in tomato. *Nat. Commun.* **13**, 3228 (2022).

36. Ramirez, K. S. et al. Detecting macroecological patterns in bacterial communities across independent studies of global soils. *Nat Microbiol.* **3**, 189-196 (2018).

37. Ling, N., Wang, T. & Kuzyakov, Y. Rhizosphere bacteriome structure and functions. *Nat. Commun.* **13**, 836 (2022).

38. Walters, W. A. et al. Large-scale replicated field study of maize rhizosphere identifies heritable microbes. *Proc. Natl. Acad. Sci. USA.* **115**, 7368-7373 (2018).

39. Edwards, J. A. et al. Genetic determinants of switchgrass- root-associated microbiota in field sites spanning its natural range. *Curr. Biol.* **33**, 1926-1938.e6 (2023).

40. Peiffer, J. A. et al. Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proc. Natl. Acad. Sci. USA.* **110**, 6548-6553 (2013).

41. Thiergart, T. et al. Root microbiota assembly and adaptive differentiation among European *Arabidopsis* populations. *Nat. Ecol. Evol.* **4**, 122-131 (2020).

42. Gonin, M. et al. Plant microbiota controls an alternative root branching regulatory mechanism in plants. *Proc. Natl. Acad. Sci. USA.* **120**, e2301054120 (2023).

43. Zhang, J. et al. *NRT1.1B* is associated with root microbiota composition and nitrogen use in field-grown rice. *Nat. Biotechnol.* **37**, 676-684 (2019).

44. Baldauf, J. A. et al. Single-parent expression is a general mechanism driving extensive complementation of non-syntenic genes in maize hybrids. *Curr. Biol.* **28**, 431-437.e4 (2018).

45. Rueda-Ayala, V. et al. Impact of nutrient supply on the expression of genetic improvements of cereals and row crops - A case study using data from a long-term fertilization experiment in Germany. *Eur. J. Agron.* **96**, 34-46 (2018).

46. Lasky, J. R. et al. Genome-environment associations in sorghum landraces predict adaptive

traits. *Sci. Adv.* **1**, e1400218 (2015).

47. Zomer, R. J., Trabucco, A., Bossio, D. A. & Verchot, L. V. Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agric. Ecosyst. Environ.* **126**, 67-80 (2008).

48. Kalnay, E. et al. The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteorol. Soc.* **77**, 437-472 (1996).

49. New, M., Lister, D., Hulme, M. & Makin, I. A high-resolution data set of surface climate over global land areas. *Clim. Res.* **21**, 1-25 (2002).

50. Hengl, T. et al. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One* **12**, e0169748 (2017).

51. Shangguan, W., Dai, Y., Duan, Q., Liu, B. & Yuan, H. A global soil data set for earth system modeling. *J. Adv. Model. Earth Syst.* **6**, 249-263 (2014).

52. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957-2963 (2011).

53. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852-857 (2019).

54. Callahan, B. J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581-583 (2016).

55. Bokulich, N. A. et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2′s q2-feature-classifier plugin. *Microbiome* **6**, 90 (2018).

56. Yilmaz, P. et al. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* **42**, D643-D648 (2014).

57. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059-3066 (2002).

58. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2-approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).

59. Abarenkov, K. et al. UNITE QIIME release for Fungi. Version 10.05.2021. UNITE Community (2021).

60. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ (2021).

61. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

62. Oksanen, J. et al. Vegan: Community Ecology Package. R package version 2.5-7.

https://CRAN.R-project.org/package=vegan (2020).

63. Kurtz, Z. D. et al. Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLoS Comput. Biol.* **11**, e1004226 (2015).

64. Meinshausen, N. & Buehlmann, P. High-dimensional graphs and variable selection with the Lasso. *Ann. Stat.* **34**, 1436-1462 (2006).

65. Louca, S., Parfrey, L. W. & Doebeli, M. Decoupling function and taxonomy in the global ocean microbiome. *Science* **353**, 1272-1277 (2016).

66. Nguyen, N. H. et al. FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* **20**, 241-248 (2016).

67. Guo, Z. et al. Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. *Mol. Breed.* **39**, 37 (2019).

68. Rogers, J.S. Measures of genetic similarity and genetic distance. In: Studies in Genetics VII, University of Texas Publication 7213, Austin, 145-153 (1972).

69. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76-82 (2011).

70. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547-1549 (2018).

71. Bernal-Vasquez, A., Utz, H. F. & Piepho, H. Outlier detection methods for generalized lattices: a case study on the transition from ANOVA to REML. *Theor. Appl. Genet.* **129**, 787-804 (2016).

72. Butler, D. G., Cullis, B. R., Gilmour A. R., Gogel, B. G. & Thompson, R. ASReml-R Reference Manual Version 4. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK (2017).

73. Browning, B. L., Zhou, Y. & Browning, S. R. A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**, 338-348 (2018).

74. Yu, J. M. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genet.* **38**, 203-208 (2006).

75. VanRaden, P. M. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* **91**, 4414-4423 (2008).

76. Zhang, Z. et al. Mixed linear model approach adapted for genome-wide association studies. *Nature Genet.* **42**, 355-360 (2010).

77. Pérez, P. & de Los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **198**, 483-495 (2014).

78. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65-70

(1979).

79. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **57**, 289-300 (1995).

80. Chen, H. et al. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *Am. J. Hum. Genet.* **98**, 653-666 (2016).

81. Breslow, N. E. & Clayton, D. G. Approximate inference in generalized linear mixed models. *J. Am. Stat. Assoc.* **88**, 9-25 (1993).

82. Efron, B. Bootstrap methods: another look at the jackknife *Ann. Stat.* **7**, 1-26 (1979).

83. Carley, C. A. S. et al. Automated tetraploid genotype calling by hierarchical clustering. *Theor. Appl. Genet.* **130**, 717-726 (2017).

84. Kursa, M. B. & Rudnicki, W. R. Feature selection with the Boruta package. *J. Stat. Softw.* **36**, 1-13 (2010).

85. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News* **2**, 18–22 (2002).

86. Gates, D. J. Single-gene resolution of locally adaptive genetic variation in Mexican maize. Preprint at https://doi.org/10.1101/706739 (2019).

87. François, O., Martins, H., Caye, K. & Schoville, S. D. Controlling false discoveries in genome scans for selection. *Mol. Ecol.* **25**, 454-469 (2016).

88. Marcon, C. et al. BonnMu: a sequence-indexed resource of transposon-induced maize mutations for functional genomics studies. *Plant Physiol.* **184**, 620-631 (2020).

89. Yim, B. et al. Soil texture is a stronger driver of the maize rhizosphere microbiome and extracellular enzyme activities than soil depth or the presence of root hairs. *Plant Soil* **478**, 229-251 (2022).

90. Luca, F., Kupfer, S. S., Knights, D., Khoruts, A. & Blekhman, R. Functional Genomics of Host-Microbiome Interactions in Humans. *Trends Genet.* **34**, 30-40 (2018).

# 3 Enrichment of the bacterial taxon *Massilia* in lateral roots is associated with flowering in maize

Danning Wang[1,2], Xiaoming He[1,2], Marcel Baer[1,2], Klea Lami[1,2,3], Baogang Yu[1,2], Alberto Tassinari[4], Silvio Salvi[4], Gabriel Schaaf[3], Frank Hochholdinger[2], Peng Yu[1,2,*]

[1] Emmy Noether Group Root Functional Biology, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn 53113, Germany

[2] Crop Functional Genomics, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn 53113, Germany

[3] Plant Nutrition, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn 53113, Germany

[4] Department of Agricultural and Food Sciences, University of Bologna, 40127 Bologna, Italy

[*] To whom correspondence should be addressed:

yupeng@uni-bonn.de

**Running title:** Spatial organization of root-bacteria association in maize

**Key words:** lateral roots, maize, *Massilia*, rhizosphere microbiome, root transcriptome

**Author contributions**

P.Y. conceived and coordinated the study; D.W. performed the integrative analysis of the transcriptome and microbiome data and performed all statistical analysis; X.H. and M.B. performed the soil inoculation experiments; X.H., K.L., B.Y. and G.S. conducted the soil and plant nutrient analyses; A.T. and S.S. contributed the maize flowering mutant; D.W., F.H. and P.Y. drafted the manuscript. All authors read and approved the final version of the manuscript.

**Manuscript is in revision in *Microbiome*.**

## Abstract

Beneficial interactions between plants and soil microorganisms are critical for crop fitness and resilience. However, it remains obscure how microorganisms are enriched across different root compartments and to what extent such recruited microbiomes determine crop performance. Here, we surveyed the root transcriptome via RNA sequencing, and the root and rhizosphere microbiome via full-length (V1−V9) 16S rRNA gene sequencing of genetically distinct monogenic root mutants of maize (*Zea mays* L.) under different nutrient-limiting conditions. Integrated transcriptomic and microbial analyses demonstrated that mutations affecting lateral root development had the largest effect on host gene expression and microbiome assembly, as compared to mutations affecting other root types. Cooccurrence and trans-kingdom network interactions analysis demonstrated that the keystone bacterial taxon *Massilia* (*Oxalobacteraceae*) in lateral roots is associated with root functional genes involved in flowering development and overall plant biomass. We further observed that the developmental stage drives the differentiation of the rhizosphere microbial assembly, especially the interactions of the keystone bacteria *Massilia* with functional genes in reproduction. Taking advantage of microbial inoculation experiments using a maize early flowering mutant, we confirmed that *Massilia*-driven maize growth promotion indeed depends on flowering time. We conclude that specific microbiota supporting lateral root formation could enhance crop performance by mediating functional gene expression underlying plant flowering time in maize.

## Introduction

The complex root system of cereals is essential for the efficient uptake of water and minerals and thus for their productivity (Lynch, 2013). Therefore, root systems offer great potential for crop improvement in unfavorable environments (Lynch, 2022). Root system architecture is shaped by intrinsic genetically encoded regulators and an enormous developmental plasticity that allows continuous adjustment of the root stock to fluctuating environmental conditions (Gruber et al., 2013; Motte et al., 2019; Yu & Hochholdinger, 2023). Mutant analyses have revealed that root-type-specific genetic regulators determine root system architecture in cereals (Coudert et al., 2010; Hochholdinger et al., 2018). A plethora of studies have highlighted lateral roots and root hairs as the major determinants of root system architecture (Rogers & Benfey, 2015). They substantially increase the root surface and are therefore instrumental for foraging nutrients and water resources in crops (Marzec et al., 2015; Yu et al., 2016) and have great potential in adaptation to unfavorable conditions such as nutrient-deficient soil.

Lateral roots are initiated post-embryonically from pericycle cells deep inside of all root types, while root hairs are tubular extensions of epidermal cells at the root surface (Hochholdinger et al., 2004; Yu et al., 2016). Molecular cloning of genes underlying maize root formation has demonstrated that key elements of auxin signal transduction, such as LOB domain and Aux/IAA proteins, are instrumental for seminal, shoot-borne, and lateral roots initiation (Hochholdinger et al., 2018). Moreover, genetic analyses have demonstrated that genes related to exocytotic vesicle docking, cell wall loosening, and cellulose synthesis and organization control root hair elongation and/or initiation (Hochholdinger et al., 2018).

Emerging lateral roots and root hairs are important sites for the release of exudates to the rhizosphere. A broad range of substrates and signaling molecules are secreted by roots to communicate with rhizosphere-inhabiting microorganisms (Aira et al., 2010; Bouffaud et al., 2012; Haichar et al., 2014; Peiffer et al., 2013). The release of easily decomposable exudates by roots leads to higher microbial density and activity in the rhizosphere compared to the bulk soil (Marschner, 2012). Thus, root hairs are also a major determinant of both rhizosphere formation, i.e., the proportion of soil modified by the roots (Delhaize et al., 2015) and function, i.e., the metabolic reactions taking place at the root-soil interface (Pausch et al., 2016). Several publications indicate that plant growth-promoting rhizobacteria are able to manipulate primary root development (Garrido-Oter et al., 2018; Poitout et al., 2017) and lateral root formation (Gutiérrez-Luna et al., 2010; López-Bucio et al., 2007; Zamioudis et al., 2013) in *Arabidopsis thaliana*. In crops, root- and rhizosphere-associated microbiota contribute to alleviate overall

plant nutrient stress (Yu et al., 2021; Zhang et al., 2019). Nevertheless, the cross-kingdom interplay between plants and microbes at the root-soil interface for structuring rhizosphere-associated microbial communities and their potential impact on plant growth and nutrient acquisition has so far received little attention in crops.

In this study, we elucidated whether and how consistent patterns between root development and associated microbiome have emerged by employing genetically distinct monogenic root mutants in a model crop maize (*Zea mays* L.). We profiled root transcriptome using RNA sequencing, and root-and rhizosphere-associated microbiome community assemblage using full-length (V1−V9) 16S rRNA gene sequencing. We assessed trans-kingdom network associations and how specific bacteria are assembled spatially through different root types by interacting with root genes. Understanding how root traits modulate their microbiome, and how plant-microbe associations influence plant development provides novel insights into establishment of beneficial host–microbiome associations in enhancing tolerance to environmental constraints.

## Results

### The root transcriptome synchronizes with bacterial community assemblage across root compartments

To understand whether and how root development affects microbiome assembly across different compartments along a single root, we examined different root and rhizosphere compartments in the primary root (Fig. 1A). The root compartments included the primary root without lateral roots, lateral roots, as well as separated cortex and stele tissues of the root differentiation zone (Fig. 1A). Moreover, we also extracted the closely attached rhizospheres from both primary and lateral roots separately, and the bulk soil from the unplanted pot as the control. These compartments were sampled from genetically diverse monogenic maize root mutants (*rum1, lrt1, rtcs, rth3, rth5,* and *rth6*) and their respective wild type B73. We conducted the study in four biological replicates and under three nutrient conditions of natural soil: control soil with sufficient nutrients, low nitrogen soil, and low phosphorus soil. We performed transcriptome analysis via RNA sequencing for root compartments and conducted bacterial microbiome analysis via 16S full-length (V1−V9) rRNA gene sequencing for both the root and rhizosphere compartments (Fig. 1A). We kept the OTUs with relative abundance (RA) >0.1% in at least two samples. Samples with <200 reads were removed from downstream analysis. After filtering, there were in total 1098 abundant OTUs from 388 samples (Table S1). For maize genes, there were 29335 genes after keeping genes expressed >10 reads in >4 samples. Among

them, only 239 OTUs (22%) were conserved for all different compartments (Fig. S1A), while most genes (99%) displayed overlapping expression (Fig. S1B). Bacterial richness (measured by α-diversity Shannon's index) varied significantly among different compartments (Benjamini-Hochberg adjusted *P* <0.05, Kruskal-Wallis test, Dunn's *post-hoc* test) and reduced from bulk soil, to the rhizosphere and roots, and further to different tissues (Fig. 1B). However, neither genotype (Fig. S2A) nor soil nutrient conditions (Fig. S2B) had consistent effects on the α-diversity within compartments, except for the *rtcs* mutant with respect to the rhizosphere extracted from lateral roots (Fig. S2A) and except for the rhizosphere obtained from lateral roots and primary roots of all possible genotypes under low phosphorus (Fig. S2B). Such divergence might be explained by the substantial differences in the bacterial community of bulk soil under different nutrient conditions (Fig. S2B). To investigate the impact of different compartments, genotypes and nutrient conditions on bacterial community composition, we performed a principal coordinate analysis (PCoA) for bacterial abundance. Regardless of the mutants and nutrient conditions, a strong shift in bacterial community composition was observed in different spatial compartments (PERMANOVA, $R^2 = 0.53$, *P* = 0.0001) along a single root (Fig. 1C). Genotype (PERMANOVA, $R^2 = 0.02$, *P* = 0.0001) and nutrient treatment (PERMANOVA, $R^2 = 0.01$, *P* = 0.0001) explained a small part but still significant variance of bacterial community composition (Fig. 1C). Similarly, a principal component analysis (PCA) illustrated the highest transcriptomic dissimilarity among different compartments (PERMANOVA, $R^2 = 0.40$, *P* = 0.0001, Fig. 1D). Notably, the genotype was responsible for a much stronger variance (PERMANOVA, $R^2 = 0.17$, *P* = 0.0001) of gene expression than of bacterial community composition. Primary root and lateral root differed substantially from the cortex and stele tissues for both the transcriptomes and bacterial assemblies (Fig. 1C, 1D). These data suggest that root transcriptomic changes specifically synchronize with spatial patterns of microbiome assembly during root development.
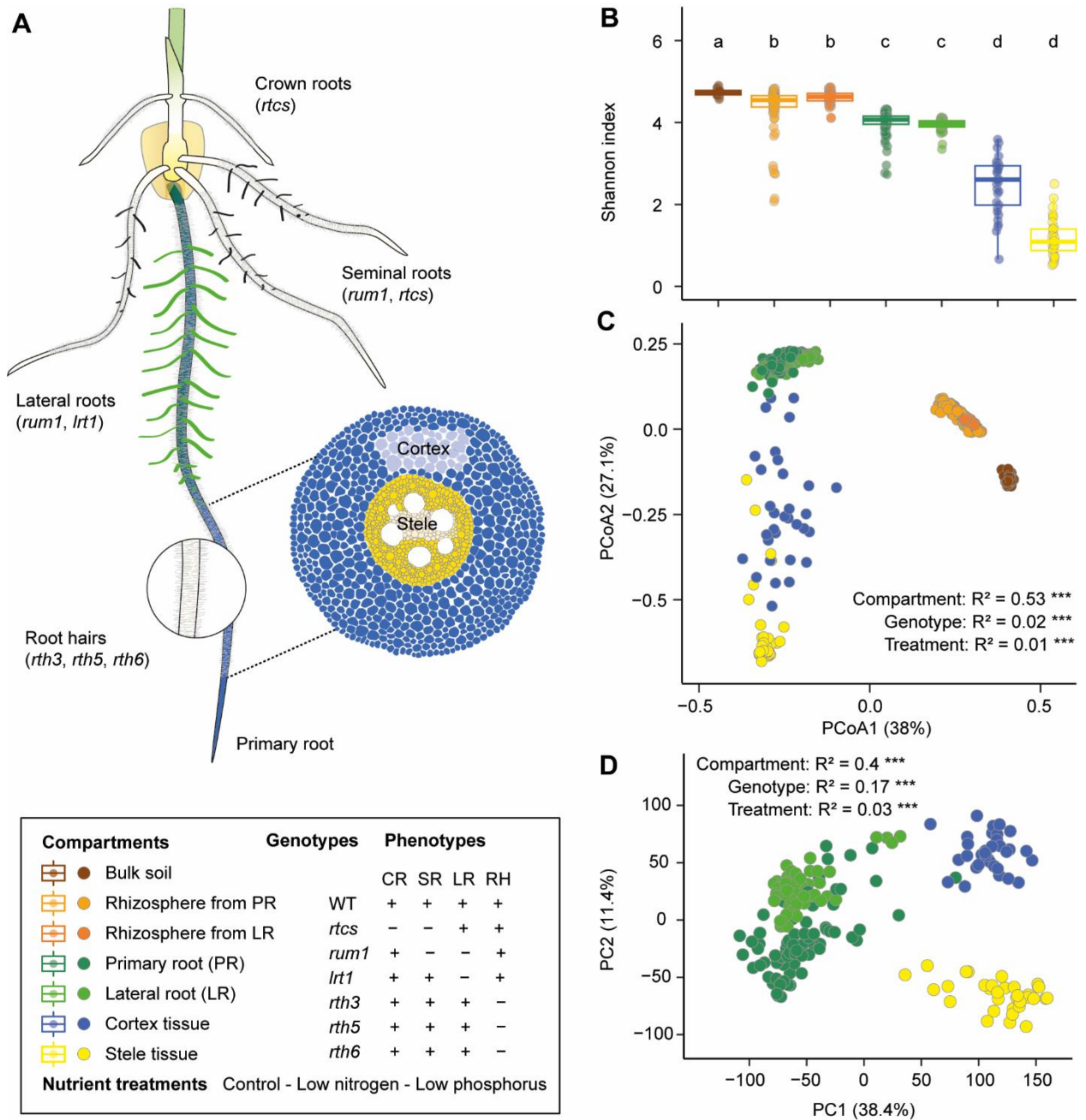
**Figure 1. Overall bacterial diversity and gene expression patterns across the rhizosphere and root compartments. A**, Schematic illustration of maize seedling root system consisting of different root types and tissue (cortex and stele) patterning along the primary root. The cloned genes with known functions affecting root phenotypes are highlighted in brackets. The meristematic and elongation zone were removed during the sampling. The differentiation zone with root hairs and lateral root primordia were physically peeled off to separate the stele and cortex tissue. The differentiation zone with emerged lateral roots was dissected and separated as the lateral roots and primary root without lateral roots. *rum1*, *rootless with undetectable meristem 1*; *rtcs*, *rootless concerning crown and seminal roots*; *lrt1*, *lateral rootless 1*; *rth*, *roothairless*. **B**, Spatial shift of bacterial α-diversity (Shannon's diversity index) across the rhizosphere and root compartments along the primary root. Significances were indicated among different compartments by different letters (Benjamini-Hochberg adjusted $P$ <0.05, Kruskal-Wallis test, Dunn's *post-hoc* test). Boxes span from the first to the third quartiles, centre lines represent median values and whiskers show data lying within 1.5× interquartile range of lower and upper quartiles. Data points at

the ends of whiskers represent outliers. **C**, Principal coordinate analysis (PCoA) showing the dissimilarity of bacterial β-diversity across the rhizosphere and root compartments. **D**, Principal component analysis (PCA) illustrating the transcriptomic shift across different root compartments along the primary root. For both bacterial and transcriptomic dissimilarity matrix data, the explained variance by compartments, genotypes and treatments were assessed by permutational analysis of variance (PERMANOVA, $P$ <0.001). The sample sizes are indicated as below: Bulk soil ($n = 42$); Rhizosphere from PR ($n = 78$); Rhizosphere from LR ($n = 57$); Primary root ($n = 84$); Lateral root ($n = 60$); Cortex tissue ($n = 33$); Stele tissue ($n = 34$).

## Lateral roots dramatically influence host gene expression and bacterial community composition

We then examined, within each compartment, the impact of the genotype and nutrient treatment on host gene expression and bacterial abundance. Using a PERMANOVA test, we observed that the genotype had consistently more impact ($R^2 = 0.27−0.52$) on gene expression for each compartment than soil nutrient condition ($R^2 = 0.08−0.12$) (Fig. S3 and S5B, Table S2). In contrast, both the genotype and soil nutrient condition had comparable impact on bacterial composition (Fig. S4 and S5A, Table S2). More specifically, the genotype explained more of the variance of bacterial composition of the primary root rhizosphere as well as of the primary root, cortex and stele ($R^2 = 0.14−0.27$) than the soil nutrient status ($R^2 = 0.08−0.16$). In contrast, for lateral roots and their rhizosphere, the soil nutrient condition ($R^2 = 0.16−0.27$) explained more variance of the bacterial composition than the genotype ($R^2 = 0.11−0.13$).

To compare the mutation effects of root hairs and lateral roots on transcriptomic changes and microbiome assemblage within each compartment, we classified all genotypes into three groups: group 1 (WT and *rtcs*) with lateral roots and root hairs, group 2 (*rum1* and *lrt1*) with root hairs but no lateral roots and group 3 (*rth3*, *rth5* and *rth6*) with lateral roots but no root hairs. As shown by PCoA and PCA, mutations that lead to lateral root defects (PERMANOVA, $R^2 = 0.2−0.25$, $P = 5e−04$) have much stronger effects on the microbiome community composition and transcriptomic changes than mutations that result in impaired root hairs (PERMANOVA, $R^2 = 0.045−0.1$, $P = 5e−04$) (Fig. 2A-C). We further performed pair-wise differential abundance/expression analyses for both bacteria and host genes between each mutant and wild type B73.
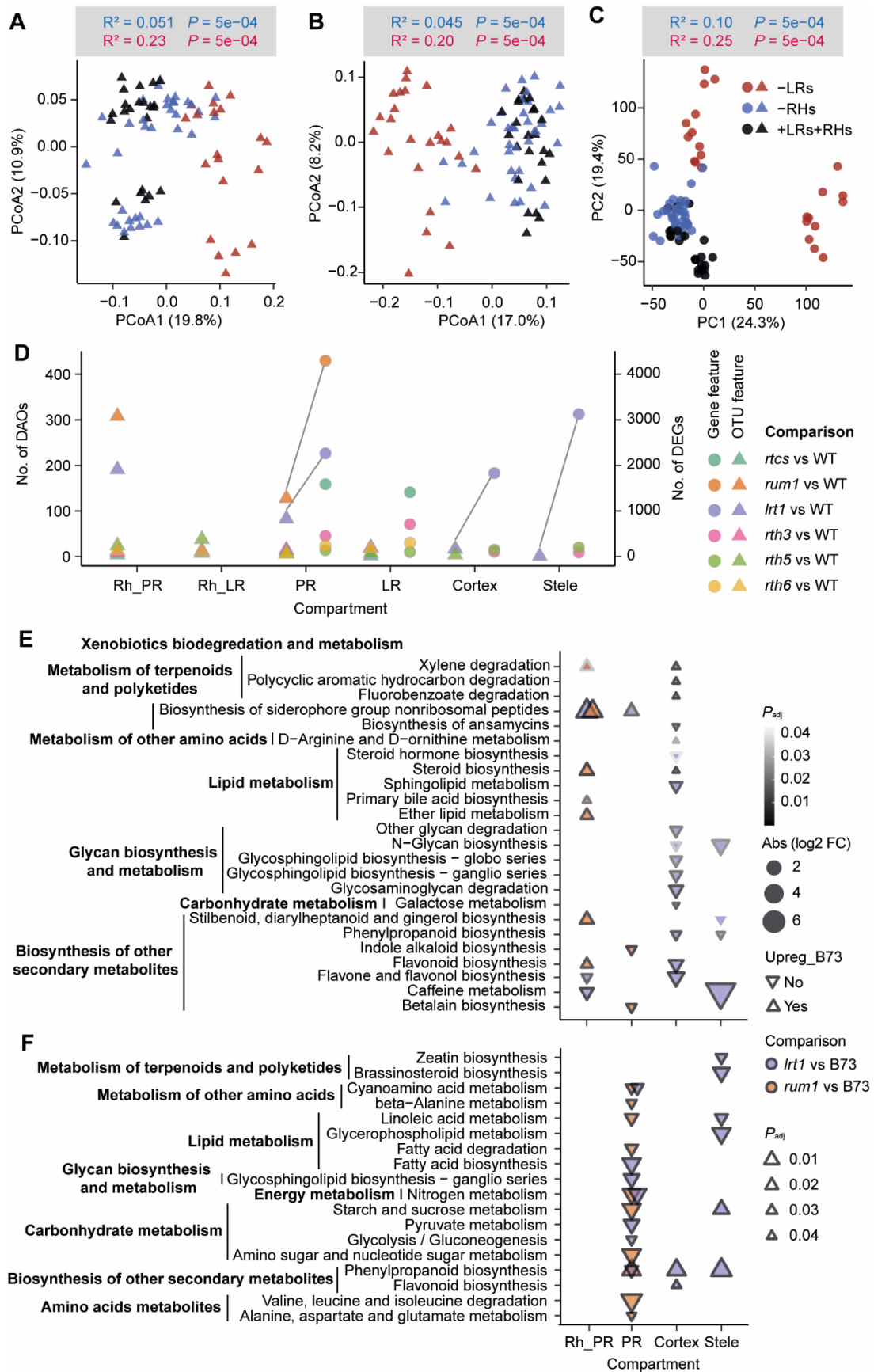
**Figure 2. Lateral roots determine microbiome assemblage and transcriptomic changes across different compartments.** Principal coordinate analysis (PCoA) showing the effects of root hairs and lateral roots defects on the dissimilarity of bacterial β-diversity across the rhizosphere (**A**) and root (**B**) compartments. **C**, Principal

component analysis (PCA) illustrating the effects of root hairs and lateral roots defects on transcriptomic shift across different root compartments along the primary root. For both, the bacterial and transcriptomic dissimilarity matrix data, the explained variance by the effects from root hair and lateral roots were assessed by permutational analysis of variance (PERMANOVA, $P <0.001$). **D**, Pair-wise comparison of differentially abundant OTUs (DAOs) and differentially expressed genes (DEGs) between each monogenic mutant and wild type (WT). The triangles and dots indicated the microbiome and transcriptome features respectively. Rh_PR, Rhizosphere from primary root; Rh_LR, Rhizosphere from lateral root; PR, Primary root; LR, Lateral root. OTU, operational taxonomic unit. **E**, Prediction of the functional potential of the bacterial microbiota using the PICRUSt tool from the lateral root mutants in comparison to the wild type plants. Gradient colors of the edges indicate the significance degree. The sizes of triangles indicate the relative abundance. Directions of triangles indicate up- and down-regulation in wild type B73. **F**, Prediction of root metabolism using the KEGG pathway analysis from the lateral root mutants in comparison to the wild type plants. The size of the triangles indicates the significance degree. The purple and orange color indicates the *lrt1* and *rum1* compared with wild type B73 respectively.

For the microbiome of primary roots and their rhizosphere, the genotypes *rum1* and *lrt1*, both are defective in lateral root formation, displayed 100−300 more differentially abundant OTUs (operational taxonomic units) in comparison to the wild type. In contrast, genotypes that are defective in root hair formation (*rth3*, *rth5* and *rth6*) showed much less changes (10−20 OTUs) (Fig. 2D). With respect to gene expression in the primary root, the lateral rootless mutants showed 2000−4300 differentially expressed genes in comparison to the wild type, while root hairless mutants displayed <500 differentially expressed genes (Fig. 2D). In particular, we identified >1800 and >3000 differentially expressed genes in the cortex and stele tissue of the *lrt1* mutant in comparison to the wild type (Fig. 2D). Functional prediction of these differentially abundant OTUs demonstrated enrichment of a substantial proportion of metabolic pathways in cortex tissue of the *lrt1* mutant (Fig. 2E, Table S3). Among the secondary metabolites, flavone and flavonol biosynthesis were the most differentially regulated metabolic pathways (Fig. 2E). Moreover, the indole alkaloid biosynthesis pathway was significantly enriched in the *rum1* mutant (Fig. 2E), which is consistent with the function of RUM1 as an Aux/IAA protein involved in auxin signaling. We also annotated the differentially expressed genes according to the KEGG pathways to understand the metabolic differences between the studied root compartments, which highlight the importance of nitrogen and cyanoamino acid metabolism pathways enriched in the primary root of both lateral rootless mutants (Fig. 2F, Table S4). Together, these analyses highlight the enrichment of maize transcripts associated with nitrogen metabolism and potential roles of microbial flavonoids driven root–microbe interactions along spatial root compartments. This also may explain the unexpectedly strong effect of the lateral root mutations on rhizospheric community composition.

**Lateral roots and their rhizosphere recruit highly complex bacterial networks**

Next, we explored the co-occurrence networks between bacterial OTUs within each compartment. To reduce the impact of rare OTUs, only OTUs with a relative abundance of >0.1% in ≥10% of samples were kept for network construction. Networks were built using SparCC algorithm for each compartment. Network correlation was calculated using the default centered log-ratio (CLR) transformed filtered bacterial table based on 100 bootstraps. Among these co-occurrence networks, the complexity of networks, total number of interactions and interacted OTUs decreased from soil, via the rhizosphere, to the root, and to the tissues (Fig. S6 and S7). We then calculated the hub score for each OTU in each network and nodes which were ranked in the top 10 were considered as keystone OTUs. We found that keystone OTUs belonging to phyla *Gemmatimodadetes*, *Planctomycetes*, and *Bacteroidetes* in the soil, keystone OTUs belonging to phylum *Proteobacteria* in the rhizosphere and primary root and keystone OTUs belonging to phyla *Proteobacteria* and *Bacteroidetes* in the lateral root and cortex tissue (Table S5). Interestingly, we detected the two hubs OTU3535 (*Massilia*) and OTU5737 (*Pseudoduganella*) belonging to *Oxalobacteraceae* in both lateral roots and cortex tissue (Fig. S6E, F).

We also examined the 10 most abundant bacterial families for each compartment. Some families belonging to *Gemmatimonadaceae*, *Acidobacteriaceae*, *Geobacteraceae* and *Planctomycetaceae* were most abundant in soil and gradually decreased from the rhizosphere to root, cortex and stele (Fig. S8; Table S6). The highly abundant families *Sphingomonadaceae* (RA = 0.07) and *Xanthomonadaceae* (RA = 0.04) were detected as the indicator taxa in the rhizosphere from the primary and lateral roots (Fig. S8; Table S6). *Chitinophagaceae* were enriched in both roots (RA = 0.20) and rhizosphere (RA = 0.18) (Fig. S8). *Sinobacteraceae* and *Polyangiaceae* were specifically enriched in both lateral roots (RA = 0.06; RA = 0.07) and primary roots (RA = 0.05; RA = 0.07) (Fig. 3A). *Streptomycetaceae* were specifically enriched in lateral roots (RA = 0.11), which is significantly higher than the abundance in primary roots (RA = 0.07) and cortex (RA = 0.06) (Fig. 3A). Interestingly, *Comamonadaceae* (RA = 0.02) were highly enriched in all root tissues but not in soil, while *Burkholderiaceae* were highly enriched in cortex (RA = 0.45) and stele (RA = 0.83) as the indicator species (Fig. S8; Table S6). In particular, *Oxalobacteraceae* were significantly enriched in primary roots (RA = 0.05), lateral roots (RA = 0.04) and cortex (RA = 0.08), and were also detected as an indicator taxon in the cortex (Fig. 3A, Table S6). These network and enrichment analyses demonstrate that different compartments recruit distinguished bacterial communities and that only specific bacteria inhabit distinct root tissues in maize.
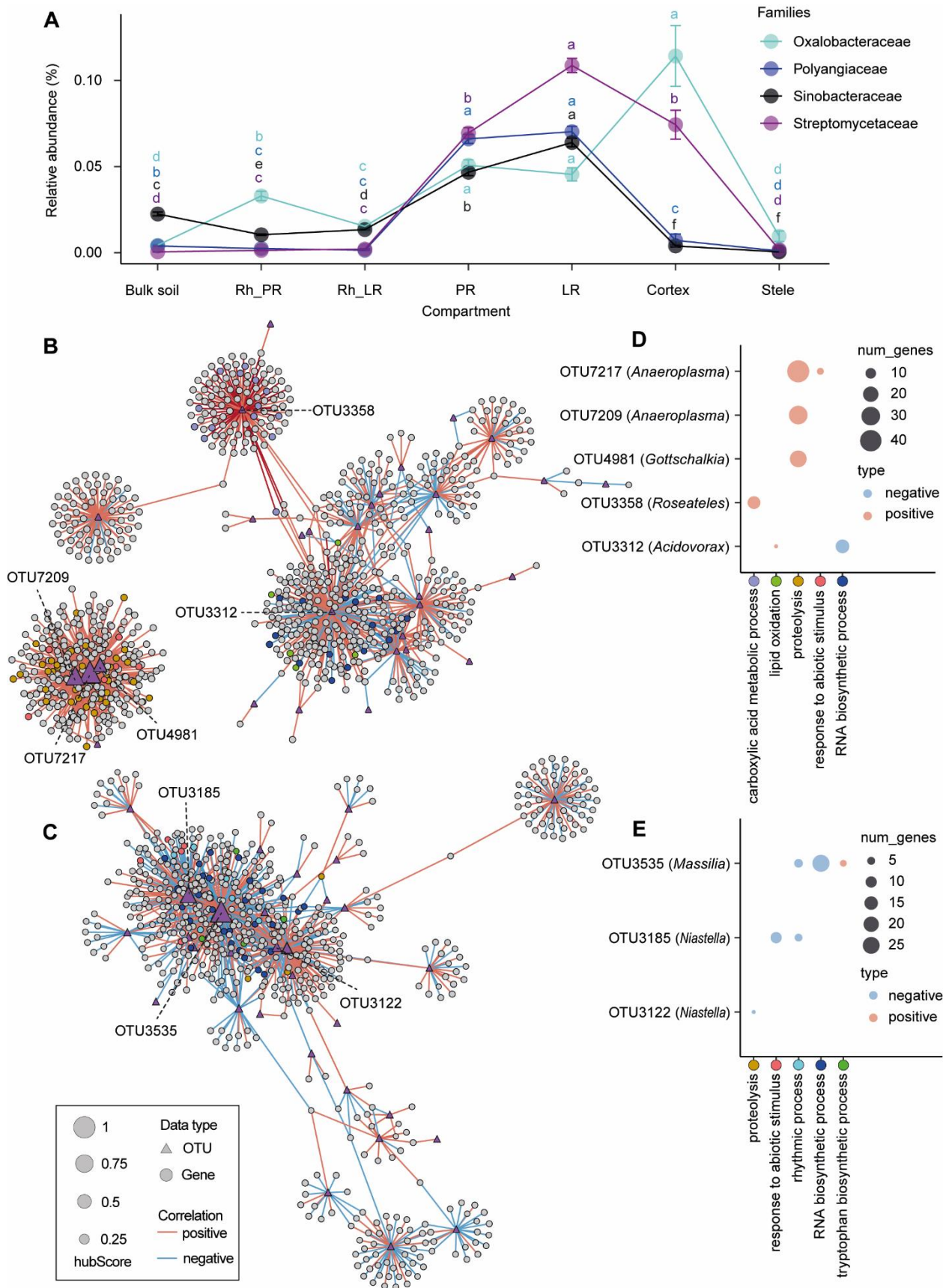
**Figure 3. Trans-kingdom interactions between root genes and bacterial OTUs in root microbiota. A**, Selective enrichment of different bacterial families from the bulk soil, to the rhizosphere, root and root tissues. Significances were indicated among different compartments by different letters for each family (Benjamini-

Hochberg adjusted *P* <0.05, Kruskal-Wallis test, Dunn's *post-hoc* test). Spearman correlations between plant genes and bacterial OTUs in the primary root (**B**) and lateral roots (**C**). The triangles and dots indicated the bacterial OTUs and gene features respectively. The size of the triangles indicates the hub score. Red and blue solid lines indicate positive and negative correlations respectively. Only the hub OTUs connected with genes with significant plant gene ontology (GO) terms are labelled accordingly. Scatter plots illustrating the hub OTUs interacting with functional genes enriched in specific plant GO terms in the primary root (**D**) and lateral roots (**E**). Different color dots indicate different plant GO terms, referring to the network relationships. The size of the dots indicates the number of genes enriched in specific plant GO terms. Red and blue dots indicate positive and negative associations respectively.

## Keystone bacteria interact with host functional genes in the root

To construct potential causal associations between bacterial OTUs and host expressed genes, we performed Spearman correlation analyses (rho >0.7 or < −0.7 with FDR adjusted *P* <0.05) for OTUs and genes expressed in ≥10 samples for rhizosphere, root, cortex and stele samples. Among those, the lateral root rhizosphere had the most complex network with 1196 nodes and 3820 edges (Fig. S9B and S10), while the primary root rhizosphere displayed the least complex network with only 544 nodes and 774 edges (Fig. S9A and S10). We identified the hub OTUs by selecting nodes that have a high hub score and where the associated genes have enriched functions. There were 25 hub OTUs from the orders *Burkholderiales*, *Chitinophagales* and *Gemmatimonadales* in the lateral root rhizosphere, but only one OTU from *Anaerolineales* in the primary root rhizosphere (Table S7). In the primary root, the hub OTUs (OTU7217, OTU7209, OTU4981, OTU3312 and OTU3358) belong to *Burkholderiales*, *Anaeroplasmatales* and *Clostridiales*, while in lateral roots OTU3535, OTU3311 and OTU3185 belong to *Burkholderiales* and *Chitinophagales*. To understand the functional genes associated with hub OTUs, we examined the gene ontology (GO) terms of those interacting genes which have positive or negative significant correlations with each hub OTU, respectively. The genes positively correlated to hub OTU6622 (*Anaerolineaceae*) in the primary root rhizosphere have functions enriched in "cellular response to auxin stimulus" (Table S7). The genes positively correlated to hubs OTU7217, OTU7209, OTU4981 in the primary root have functions enriched in "proteolysis" and "response to abiotic stimulus" (Fig. 3B, D). The genes positively correlated to hub OTU3312 (*Acidovorax*) in the primary root have enriched functions in "lipid oxidation", while genes negatively correlated to it have a function in "RNA biosynthetic processes". Finally, genes positively correlated to OTU3358 (*Roseateles_depolymerans*) are enriched in "carboxylic acid metabolic processes" (Fig. 3B, D). In lateral roots, genes positively correlated to OTU3535 (*Massilia*) are enriched in "tryptophan biosynthetic process", while negatively correlated genes have functions enriched in "rhythmic

process" and "RNA biosynthetic processes" (Fig. 3C, E). In particular, we did not detect any significant (FDR adjusted $P$ <0.05) associations between genes and OTUs in the cortex or stele tissue. These data indicate that specific keystone bacteria may play an important role in interacting with functionally different host genes within specified compartments.

**Keystone bacteria *Massilia* in lateral roots associated with plant phenology and flowering development**

To identify whether keystone bacteria in association with host genes influence the maize phenotypes, we performed weighted correlation network analyses (WGCNA) between host genes and phenotypic traits as well as between bacteria and phenotypic traits across different nutrient treatments and genotypes followed by Spearman correlation analyses between identified phenotypic traits related genes and bacteria (see Methods). Overall, low nutrient treatments significantly (Benjamini-Hochberg adjusted $P$ <0.05, Kruskal-Wallis test, Dunn's *post-hoc* test) reduced the shoot dry biomass (Fig. S11A), while low nitrogen and low phosphorus treatment significantly (Benjamini-Hochberg adjusted $P$ <0.05, Kruskal-Wallis test, Dunn's *post-hoc* test) reduced the nitrogen (Fig. S11B) and phosphorus (Fig. S11C) concentration, respectively. We performed this integrative analyses using the data in the rhizosphere and root since we did not detect any significant correlations between host genes and bacteria in the cortex or stele tissue. Genes in the primary root were clustered into 21 modules and five (MEgreen, MEyellow, MEskyblue3, MEdarkmagenta and MEsienna3) of them were significantly ($P$ <0.05) correlated with dry biomass, four (MEgreen, MEskyblue3, MEdarkmagenta and MEsienna3) of them were significantly correlated with nitrogen concentration, three (MEturquoise, MEskyblue3 and MEred) of them were significantly correlated with phosphorus concentration (Fig. S12A). In lateral roots we identified 20 modules and four (MElightgreen, MEtan, MEblack and MEgreenyellow) of them were significantly correlated with dry biomass and only module MEdarkmagenta was significantly correlated with nitrogen concentration and phosphorus concentration (Fig. S12B). For bacteria, there were 11 modules in the rhizosphere of the primary root, and five of them were significantly correlated with dry biomass, four with nitrogen concentration, three with phosphorus concentration (Fig. S13A). In the rhizosphere of lateral roots, there were also 11 modules, and four of them were significantly correlated with dry biomass, two with nitrogen concentration, and three with phosphorus concentration (Fig. S13B). In the primary root, there were only five modules, and two of them were significantly correlated with dry biomass and nitrogen concentration, while only one with phosphorus concentration (Fig. S13C). In lateral roots, there were only seven modules, and only the module MEturquoise was significantly correlated with dry biomass,

MEgreen with nitrogen concentration and MEyellow with phosphorus concentration (Fig. S13D). In total, we identified the most significant correlations between host genes and bacteria related to shoot dry biomass in the rhizosphere of lateral roots with 243 nodes and 1202 edges (rho >0.7 or < −0.7 and FDR <0.05) (Fig. 4B, Table S8), while the primary root rhizosphere had the least complex network with only 19 nodes and 34 edges (Fig. 4A, Table S8). Moreover, there were more significant (rho >0.7 or < −0.7 and FDR <0.05) correlations between host genes and bacterial OTUs in the lateral roots than in the primary root (Fig. 4D and C, Table S8). Interestingly, the relative abundance of hub OTU3535 (*Massilia*) in the lateral root was significantly positively correlated with shoot dry biomass ($R^2_{adj}$ = 0.49, $P$ = 0.0022, Fig. S14) and had significant associations with 103 genes expressed in lateral roots (Fig. 4D). We further examined the GO terms of these dry biomass-associated genes which are enriched in the GO term "circadian rhythm" (Fig. 4D). In particular, we observed the gene *night light-inducible and clock-regulated 1* (*lnk1*), which functions in response to abiotic stimulus and the gene *timing of cab expression 1* (*toc1*) which functions in flower development. For the primary root and its rhizosphere, functions of enriched GO terms were "iron transport" and "cell wall organization", respectively (Fig. 4A, C). Similarly, we performed network integration analyses for host genes, microbial OTUs, shoot nitrogen and phosphorus concentration. We found only significant correlations between host gene expression and bacterial abundance related to nutrient uptake in the rhizosphere (Fig. S15; Table S8). In the rhizosphere of primary roots, genes with GO functions in "cell wall organization" and "glucose metabolic process" had significant correlations with some keystone bacteria related to nitrogen concentration (Fig. S15A), while in the rhizosphere of lateral roots significantly enriched GO terms were e.g. "phosphate starvation", "phosphate transport" and "response to cold" for host genes, which were correlated with some keystone bacteria in association with nitrogen concentration (Fig. S15B). We also observed phosphorus concentration related genes with GO functions in "phosphate ion homeostasis" and "phosphate starvation" significantly correlated with keystone bacteria in the rhizosphere of lateral roots (Fig. S15C). These integrative results may indicate that *Massilia* in lateral roots dramatically influence maize biomass through interactions with circadian rhythm and flowering development of host genes.
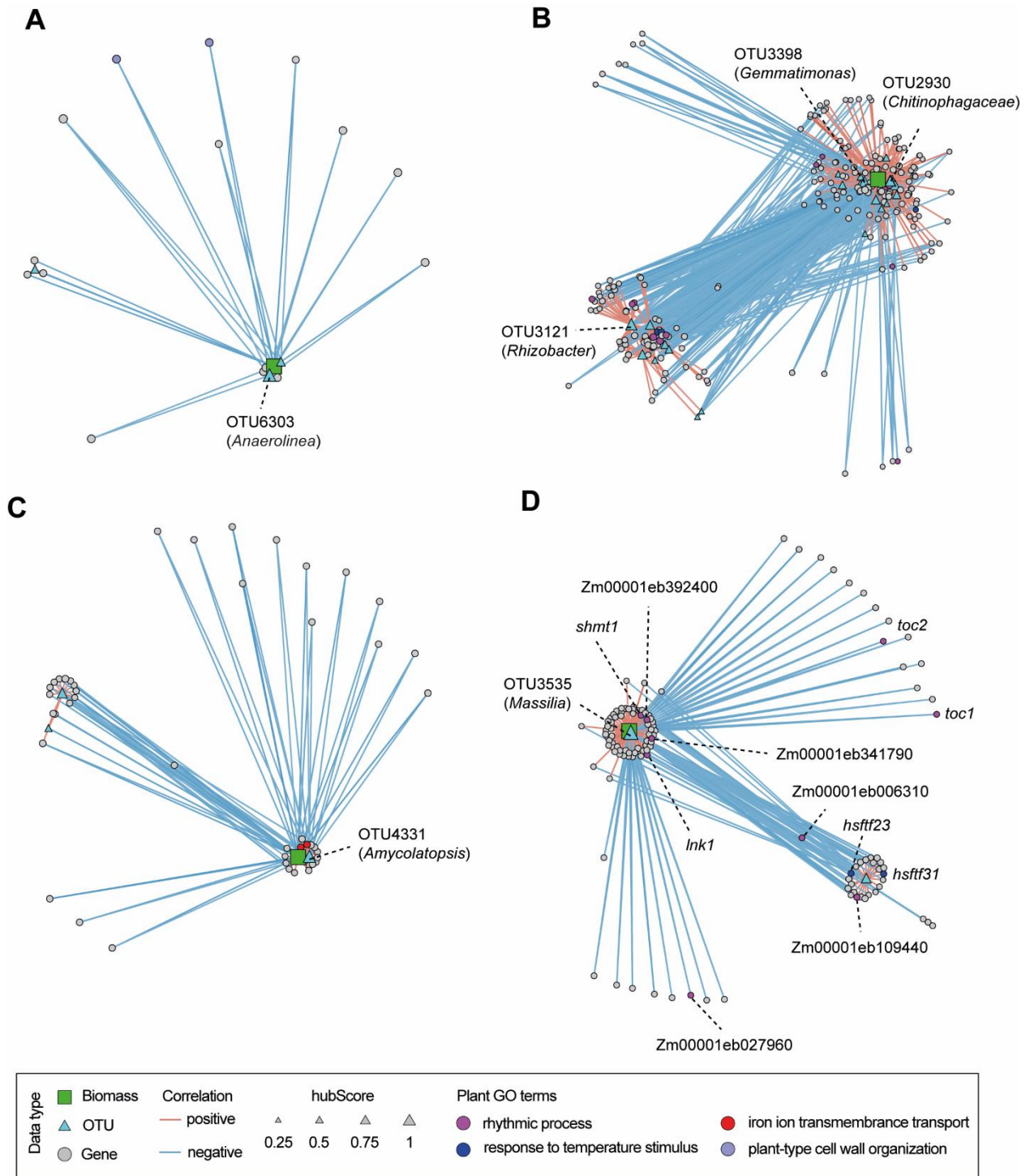
**Figure 4. Bacterial hubs prioritize the causal association with plant rhythmic process and biomass accumulation.** Spearman correlation relationships between gene expression, OTU abundance and shoot biomass in the rhizosphere from primary root (**A**), rhizosphere from lateral roots (**B**), primary root (**C**) and lateral roots (**D**). The triangles, dots and cubes indicated the microbiome, transcriptome and biomass features respectively. Red and blue solid lines indicate positive and negative correlations respectively. The size of the triangles indicates the hub score. Different color dots indicate specific plant gene ontology (GO) terms. The genus name of the hub OTUs and specific genes with GO annotation were labelled. *toc*, *timing of cab expression*; *shmt1*, *Serine hydroxymethyltransferase 1*; *lnk*, *night light-inducible and clock-regulated*; *hsftf*, *heat stress transcription factor*.

**Genetic validation of *Massilia* function in early flowering by mutant analysis in maize**

The circadian rhythm in plants refers to physiologically relevant activity cycles of various biological processes regulated by an innate circadian clock, including growth, leaf development and flowering transition. To explore whether *Massilia* is involved in maize flowering time, we performed a soil experiment using the wild type B73 and the *rtcs* mutant which contains both lateral roots and root hairs and performed 16S amplicon (V3−V4) sequencing for the whole primary root with lateral roots at two different growth stages, i.e. at seedling stage (three weeks) and at flowering stage (ten weeks). Bacterial α-diversity analysis showed microbiome richness significantly increased from seedling to flowering stage ($P$ <0.01, Wilcoxon rank-sum test) (Fig. 5A). In PCoA analysis, the developmental stage explained the large variance ($R^2 = 0.30$, $P$ <0.01, PERMANOVA) of bacterial community composition (Fig. 5B). To investigate whether *Massilia* abundance changes during maize development, we compared the relative abundance of each dominant genus in the root in control soil. In total, we detected three differentially abundant bacterial genera between stages, from which *Massilia* was the most abundant genus. Notably, its relative abundance significantly ($P$ <0.01, Wilcoxon rank-sum test) decreased from seedling to flowering stage (Fig. 5C). We next examined the root transcriptome and observed that the developmental stage effect can significantly separate the pattern of gene expression as shown by the PCA plot (Fig. S16A). We determined differential gene expression between seedling and flowering stages and discovered that 2837 genes were differentially expressed between developmental stages (Fig. S16B). To investigate whether some bacterial ASVs have significant associations with plant genes across developmental stages, Spearman correlation analysis was performed between highly abundant ASVs and differentially expressed genes. Interestingly, four hub ASVs (ASV83, ASV97, ASV102 and ASV134) that belong to the bacterial genus *Massilia* were identified significantly in association with genes annotated in functional categories such as primary and secondary metabolic processes, defense responses and flowering related pathways (Fig. 5D). We next further narrowed down the GO terms by removing the redundant ones and only highlighted the driver GO terms (Table S9). Specifically, we identified two driver terms in "recognition of pollen" and "detoxification" among the GO terms derived from ASV83 and ASV97 (Fig. 5D; Table S9). These integrated microbiome and transcriptome results suggest that hub *Massilia* might be associated with plant genes involved in maize flowering.
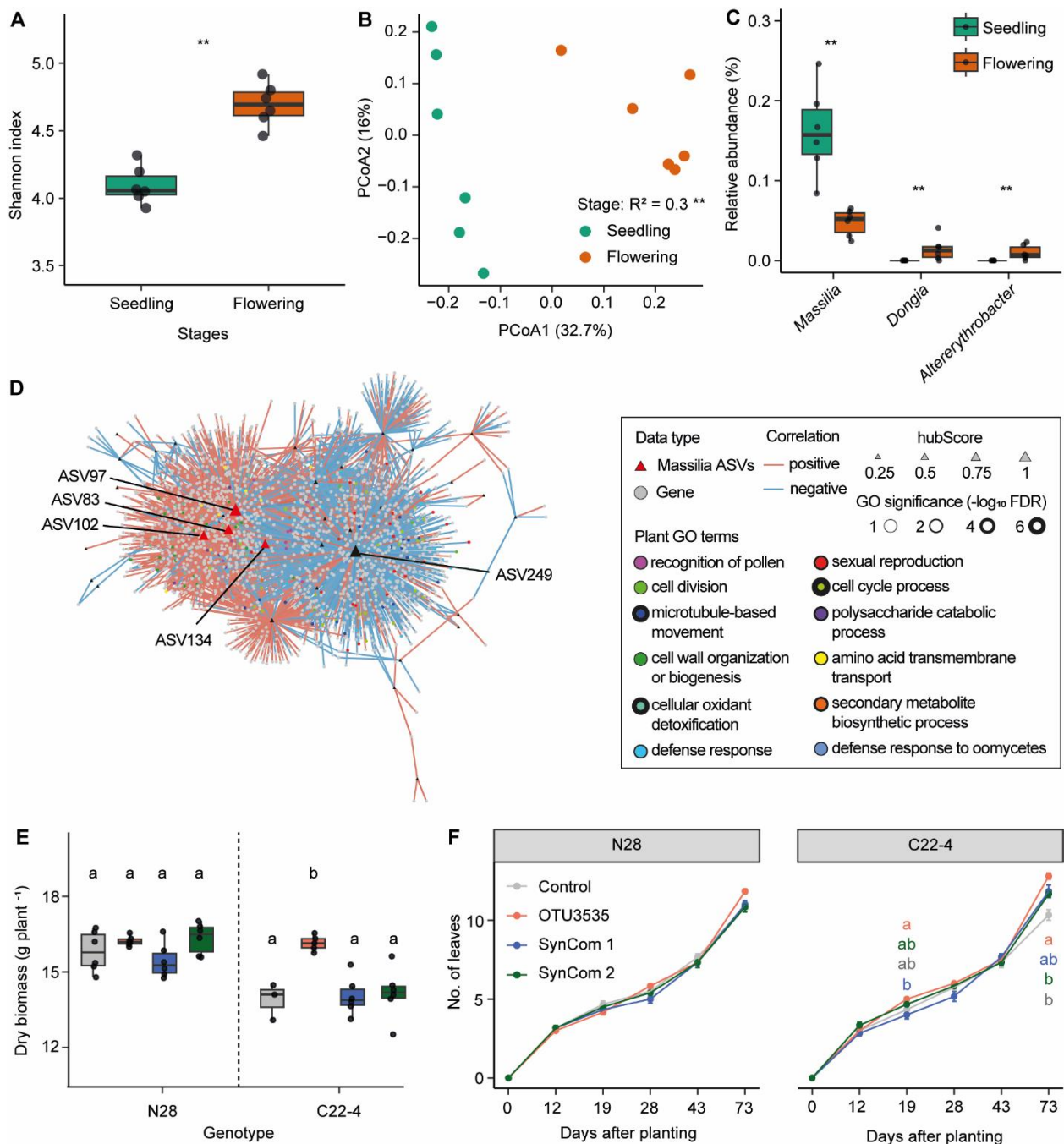
**Figure 5. The bacterial hub *Massilia* associates with maize flowering and biomass production. A**, Root bacterial α-diversity (Shannon's diversity index) at seedling and flowering stage. Significances are indicated between different stages by asterisk at $P < 0.01$ (two-tailed Wilcoxon rank-sum test). Boxes span from the first to the third quartiles, centre lines represent median values and whiskers show data lying within 1.5× interquartile range of lower and upper quartiles. Data points at the ends of whiskers represent outliers. $n = 6$. **B**, Principal coordinate analysis (PCoA) showing the dissimilarity of bacterial β-diversity between seedling and flowering stage. The explained variance ($R^2$) by different stages were assessed by permutational analysis of variance (PERMANOVA, $P < 0.01$). $n = 6$. **C**, Significantly differential abundant genera between seeding and flowering stage. Significances were indicated among different stages by asterisk at $P < 0.01$ (two-tailed Wilcoxon rank-sum test). **D**, Spearman correlation between high abundant ASVs and root differential genes between seedling and flowering stages. The triangles and dots indicated the bacterial OTUs and gene features respectively. Red and blue

solid lines indicate positive and negative correlations respectively. The size of the triangles indicates the hub score. Different color dots indicate specific plant gene ontology (GO) terms. Effects of single inoculation of *Massilia* OTU3535 and Synthetic Community (SynCom) on the biomass production (**E**) and number of leaves (**F**) of maize wild type N28 and its isogenic early flowering mutant C22-4. SynCom 1 is composed of 12 different bacterial isolates without OTU3535, while SynCom 2 is composed of 13 different bacterial isolates including of OTU3535.

To investigate whether the function of *Massilia* is associated with flowering time as a possible consequence of altered rhythmicity, we performed *Massilia* inoculation experiments for wild type N28 and the corresponding early flowering mutant C22-4 in the greenhouse. We first performed alignment analysis using previously identified OTU3535 with these four hub ASVs and confirmed with >97% sequence identity based on the BLAST alignment (Table S10). We then performed inoculation with the single OTU3535, the synthetic community with 12-member isolates derived from *Oxalobacteraceae* (SynCom 1), and a 13-member isolates including of OTU3535 (SynCom 2). Interestingly, only the inoculation of single *Massilia* OTU3535 significantly promoted maize dry biomass (Fig. 5E) and the number of leaves (Fig. 5F) in the early flowering mutant, whereas we did not observe any effects on these traits in the wild type plants. Taken together, the function of *Massilia* in promoting maize biomass and leaf number might depend on flowering time in maize.

## Discussion

Cereal root systems comprise diverse root types and specified organs such as lateral roots and root hairs supporting the high demand for water and nutrients of these plants in agricultural production (Hochholdinger et al., 2018; Yu et al., 2016). The root system determines the main chemical messengers with causal interactions with microorganisms in the root-soil interface e.g. rhizosphere (Yu et al., 2016). The structure and function of the microbiome associated with rhizosphere and root influence the health and nutrition of the host plants (Mendes et al., 2013). Thus, plant selectiveness and recruitment of soil microbes into the host through separated rhizocompartments (rhizosphere, rhizoplane and endosphere) is the critical step for initiation of such causal interactions (Attia et al., 2022; Edwards et al., 2015; Edwards et al., 2018). Nevertheless, little is known on how specific root types or root tissues selectively recruit microorganisms in space and time, and what role such enriched microorganisms play in crop performance.

In this study, we systemically dissected different root organs and rhizosphere compartments (Fig. 1) along a single root at longitudinal and transversal resolution, and subjected these samples to host-microbiome associations by investigating the root transcriptome and associated microbiome using diverse maize root mutants. Examination of microbiomes across diverse mutants demonstrated that genotypes defective in lateral root formation displayed the largest changes in overall root gene expression and bacterial community composition in both the rhizosphere and the endosphere (Fig. 2). This is in agreement with earlier observations that lateral roots largely reshaped specific gene expression and microbial colonization in crops (Gutjahr, et al., 2015; Yu & Hochholdinger, 2018).

Integrative trans-kingdom network association analyses demonstrated that the keystone bacterial taxon *Massilia* (*Oxalobacteraceae*) was highly associated with plant genes involved in plant circadian rhythm, tryptophan and RNA biosynthetic processes in lateral roots (Fig. 3). Notably, in maize, the keystone bacterial taxon *Massilia* is significantly associated with the trait biomass and confers to a substantial enrichment of genes related to the circadian rhythm e.g. flower development of plants (Fig. 4). Here, we demonstrated that lateral root specific recruitment of *Massilia* might contribute to the association with maize growth and the flowering time phenotype (Fig. 5), thus being the potential linkage between plant rhythmicity and variation of rhizosphere microbiome (Hubbard et al., 2018; Lu et al., 2018; Newman et al., 2022). The abundance of the specific bacterial taxon *Massilia* driven by root-derived flavonoids can modulate lateral root development and overall maize performance under nitrogen

deprivation (Yu et al., 2021) and such causal interaction between root and rhizosphere is genetically regulated (He et al., 2023). More importantly, the abundance of *Massilia* is highly heritable in comparison to other bacterial taxa in the root under nitrogen limited conditions (He et al., 2023), thus suggesting a great potential to breed stress-resilient crop root systems that more readily accommodate beneficial keystone bacteria. Overall, the identification of important root traits supporting beneficial microbes might contribute to the adaptation of crops to future climatic challenges.

## Materials and Methods

### Maize mutants, soil preparation and experimental design

Genetically distinct root mutants and wild type B73 were used in this study. Among them, the *rtcs* mutant displays crown and seminal root defects; *rum1* displays seminal and lateral root defects; *lrt1* and *rth3-rth6* display lateral root and root hair defects respectively (Hochholdinger et al., 2018). C22-4 is a nearly isogenic line of wild type N28. C22-4 carries Gaspe Flint introgressions on Chromosome 8 and shows an early flowering phenotype (Salvi et al., 2007). We used the soil that was dug from a long-term fertilizer field experiment in Dikopshof (50°48′21″N, 6°59′9″E). We have collected three different soils e. g. the control soil with fully fertilized nutrients (CK), low nitrogen soil fertilized without nitrogen (LN) and low phosphorus soil fertilized without phosphorus (LP) as described by He et al. (2023). The freshly dug soil was air-dried and sieved through a 4 mm sieve ready for use. The soil pot (7cm × 7cm × 19cm) experiment was carried out in a complete randomized design and comprised seven genotypes and three nutrient treatments. We prepared additional empty soil pots without plants grown as the "bulk soil" samples. All pots in each culture tray were completely randomized using a true random generator (excel function "RAND") and trays were reshuffled every week in the phytochamber without paying attention to the pot labels. Specifically, we maintained the soil water content by weighing the soil pots according to the loss of water every two days. For the whole 4-week culture, only sterilized water was applied to avoid the potential contamination.

### Sample collection and tissue separation

The root and rhizosphere samples were harvested from 4-week-old maize plants for all genotypes grown under different nutrient treatments. In detail, the whole root systems were carefully taken out from each pot and vigorously shaken to remove all soil not firmly attached to the roots. To synchronize the harvest for precise comparison among different maize mutants, we specifically focused on the primary root which is present in all mutants. We separately dissected the lateral roots from the primary root as previously described (Yu & Hochholdinger, 2018). Moreover, we carefully removed all big particles from these primary or lateral roots to avoid the contamination. Only the root organs with tightly attached soil were placed into a 15 ml Falcon (Sarstedt) tube and immediately frozen in liquid nitrogen and stored at -80 °C before extraction of rhizosphere soil. The rhizosphere samples were defined and extracted into PowerBead tubes (MP Biomedical) as described previously (Yu et al., 2021). The root samples were harvested from a separate plant and the attached soil were washed immediately by tap water and rinsed with three times of sterilized water followed by tissue drying and placed in

PowerBead tubes. Moreover, the stele and cortex tissue from the differentiation zone of the primary root were peeled off separately by hand as previously described (Yu et al., 2015). Please note that we were not be able to separate these two tissues in the zone with emerged lateral roots. We collected the bulk soil samples from the unplanted pots as the control.

**RNA sequencing and bacterial 16S rRNA gene sequencing**

The frozen rhizosphere samples were extracted from the primary root and lateral roots respectively as described previously (Yu et al., 2021). DNA extractions were performed soon after root and rhizosphere samples were harvested, following the FastDNA™ SPIN Kit for Soil (MP Biomedical) protocol. Total RNA was isolated from the primary root, lateral root, cortex and stele tissues samples using the RNeasy Plus Universal Mini Kit (Qiagen). Both DNA and RNA were qualified and quantified via Agilent RNA or DNA Chips (Agilent Technologies). The complementary DNA libraries for RNA-seq were constructed with the MGIEasy RNA library construction kit. Cluster preparation and PE150 read sequencing were performed on a DNBSEQ-G400 system. Amplicon library construction was processed with a similar workflow as previously described (Yu et al., 2021). In brief, the full length (V1−V9 region) of the 16S rRNA genes was sequenced on a Pacbio Sequel II (PacBio Biosciences Inc.) using the forward primer (27F) with anchor sequence 5′-TTTCTGTTGGTGCTGATATTGCAGRGTTYGATYMTGGCTCAG-3′ and reverse primer (1492R) with anchor sequence 5′-ACTTGCCTGTCGCTCTATCTTCCGRGYTACCTTGTTACGACTT-3′ (Klindworth et al., 2013). For validation experiments, the amplicon libraries were sequenced at the V3-V4 regions of the 16S rRNA gene were amplified using a 16S (V3–V4) Metagenomic Library Construction Kit for NGS (Takara Bio Inc., Kusatsu, Japan). The following primers were used: 341F with overhang adapter 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3′ and 806R with overhang adapter 5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACHVGGGTWTCTAAT-3′. The second PCR was performed using the Nextera® XT Index Kit (Illumina, San Diego, CA, USA) for sample multiplexing with index adapters. The libraries were sequenced on the MiSeq™ platform using the MiSeq™ Reagent Kit v3 (2 × 250 bp; Illumina).

**Transcriptome data analysis**

Processing and trimming of raw RNA-seq reads was performed as described previously (Yu et al., 2021). In brief, the low-quality sequences and low-complexity polyA tails were eliminated.

Subsequently, we built the reference genome index and mapped the sequences to the maize reference genome sequence v.5 (http://ftp.ensemblgenomes.org/pub/plants/current/fasta/zea_mays/dna/Zea_mays.Zm-B73-REFERENCE-NAM-5.0.dna.toplevel.fa.gz) by HISAT2 (v2.1.0) software (Kim et al., 2019). All commands and default parameters were used in HISAT2. All downstream analyses were performed in R (v4.2.2) (R Core Team, 2022). Then all bam files generated by HISAT2 were input to 'featureCounts' function (Liao et al., 2014) in R package Rsubread (v2.12.3) using maize reference annotation v.5 (http://ftp.ensemblgenomes.org/pub/plants/current/gtf/zea_mays/Zea_mays.Zm-B73-REFERENCE-NAM-5.0.53.chr.gtf.gz) to generate the gene expression table. Chimeric reads and reads mapped to more than one position in the genome were removed. Only genes represented by a minimum of ten mapped reads in ≥4 samples were declared as expressed and considered for downstream analyses. Before statistical analyses, data were normalized by library size using the DESeq2 (v1.38.3) package (Love et al., 2014) in R. A principal component analysis was performed using the 'prcomp' function in base R (v4.2.2). To test the marginal effects of compartment, treatment and genotype on the transcriptome, a permutation-based PERMANOVA test was performed with the Euclidean distance matrix between pairs of transcriptomic samples using 'adonis2' function in R package vegan (v2.6.4) (Oksanen et al., 2020). All plots were produced using R package ggplot2 (v3.4.2) (Wickham, 2016).

**Bacteria full length 16S rRNA gene sequence processing and data analysis**

16S rRNA gene (V1−V9) raw sequencing reads were processed as following steps. Sequencing reads were assigned to samples based on their unique barcode and truncated by cutting off the barcode and primer sequence which were called raw tags. Raw sequence data were quality filtered and deduplicated using Usearch fastx_uniques command. OTUs (Operational taxonomic units) were generated by UPARSE (Edgar, 2013) algorithm in the Usearch software (v11.0.667) with parameters for full-length sequences. Sequences were clustered based on 97% identity and assigned to a different OTU using Usearch cluster_otus. Taxonomy was assigned to OTUs using the BLCA software (v2.2) (Gao et al., 2017) against the NCBI 99% OTUs reference sequences (20170709) at each taxonomic rank (kingdom, phylum, class, order, family, genus, species). During the clustering process, the chimeric sequences were removed using UCHIME to filter the final OTU sequences using the RDP "gold" sequences (Edgar et al., 2011). Mitochondria, chloroplast and phylum-unassigned OTUs were eliminated.

All downstream statistical analyses were performed in R (v4.2.2) (R Core Team 2022). Briefly,

OTU tables were filtered with ≥0.1% relative abundance in ≥2 samples. For α-diversity indices, Shannon index was calculated using OTU tables rarefied to 200 reads. The OTUs which expressed ≤0.1% relative abundance within ≤5% samples and the samples with ≤200 reads were removed. Bray–Curtis distances between samples were calculated using OTU tables that were normalized using 'varianceStabilizingTransformation' function from DESeq2 (v1.38.3) package (Love et al., 2014) in R. If not specified, the following data analysis is based on the normalized OTU table. Principal coordinate analyses were performed using the 'ordinate' function in R package phyloseq (v1.42.0) (McMurdie & Holmes, 2013). To test the marginal effects of compartment, treatment and genotype on the microbial composition community, a permutation-based PERMANOVA test was performed using Bray–Curtis distance matrix between pairs of bacterial samples using 'adonis2' function in R package vegan (v2.6.4) (Oksanen et al., 2020). All plots were produced using R package ggplot2 (v3.4.2) (Wickham, 2016).

**Differential gene expression and functional characterization**

We determined differential gene expression between each mutant and wild type (B73) using the 'DESeq' function in DESeq2 R package (v1.38.3). Subsequently, we determined the number of differentially expressed genes for each comparison by controlling the FDR adjusted *P* values of pairwise *t*-tests <0.05 and a fold change of >2. We then functionally classified differential gene expression patterns according to GO terms using g:Profiler (Raudvere et al., 2019). The GO annotation system is based on four structured vocabularies that describe gene products in terms of their associated biological processes, cellular components, molecular functions and KEGG pathways in a species-independent manner. Subsequently, we performed a gene set enrichment analysis to discover significantly overrepresented functional categories.

**Functional prediction of bacterial OTUs**

PICRUSt2 (v2.5.2) (Douglas et al., 2020) was used to predict functional pathways present in bacterial community from the 16S rRNA marker gene. KEGG orthology (KO) abundance table was predicted through the PICRUSt2 pipeline with input of OTU-table and OTUs sequences. To keep only abundant KOs, KOs which express >0.01% relative abundance in >5% samples were kept. After filtering, individual KOs were summarized at KEGG pathway level 3 using the categorize_by_function_l3 function and KEGG mapping file provided by PICRUSt group. KEGG pathway level 2 and level 1 were manually curated from the KEGG website (https://www.genome.jp/kegg). Then we identified differential KEGG pathways between each mutant and wild type (B73) using the 'DESeq' function in DESeq2 R package (v1.38.3).

Subsequently, we determined the number of differentially abundance pathways for each comparison by controlling the FDR adjusted $P$ values of pairwise $t$-tests <0.05 and a $\log_2$ fold change of >0.5.

**Construction of bacterial cooccurrence networks**

To explore the potential associations among different bacterial OTUs, we constructed bacterial cooccurrence networks within each compartment using the high abundant OTUs >0.1% in at least 10% samples. These filtered OTUs were used as the input for SparCC algorithm (Friedman & Alm, 2012). This analysis was performed with default parameters and 100 bootstraps were used to infer $P$ values. The correlations >0.4 or <−0.4 ($P$ <0.05) were kept for network construction. Networks were visualized using spring embedded layout with weight in Cytoscape (v3.8.0) (Shannon et al., 2003).

**Trans-kingdom interactions between root genes and bacterial OTUs**

We next built the causal interactions between root expressed genes and root/rhizosphere abundant OTUs. To reduce false positive correlations, only genes expressed >5 reads in at least 10 samples were used as gene input table and only OTUs expressed in at least 10 samples were kept for OTU input table. After normalization using 'varianceStabilizingTransformation' function from DESeq2 package for both gene table and OTU table, spearman correlation was calculated by 'corr.test' function from psych package (v2.3.3) (Revelle, 2023) in R for each compartment. This function provides the correlation coefficient and the corresponding FDR adjusted $P$ values. Spearman correlations with rho value >0.7 or <−0.7 and adjusted $P$ values <0.05 were kept as significant correlations. The above significant correlations were input to Cytoscape (v3.8.0) (Shannon et al., 2003) for network visualization. Node hub scores were calculated using the 'hub_score' function from igraph package (v1.4.2) (Csárdi et al., 2023) in R.

**Integration of host transcriptome, bacterial community and phenotypic traits**

Network-based analysis is the most biologically interpretable tool available to analyze association among variables, such as relationships between microbial compositions, gene expression and phenotypic traits (Wang et al., 2019). Weighted correlation network analysis (WGCNA) is a data-driven method that clusters genes to different modules based on weighted correlations between gene transcripts (Langfelder & Horvath, 2008). To identify keystone bacteria which significantly associate with phenotypic traits and also with host transcriptome in each compartment, we performed WGCNA for both bacteria OTUs and host root expressed

genes in R: (1) cluster gene/OTU co-expression modules among different genotypes across three nutrient conditions; (2) associate module eigengene/eigenOTU with phenotypic traits; (3) select genes/OTUs with high membership value from modules which significantly associate with phenotypic traits and high correlation coefficient with phenotypic traits; (4) correlate the selected genes and OTUs to determine key OTUs and genes.

For robust construction of co-expression networks, we filtered and normalized genes/OTUs table as described above. The soft thresholding power β was automatically selected and used to calculate adjacency matrix. To minimize the effects of noise and false associations, we transformed the adjacency matrix into a topological overlap matrix (TOM) with selected power and calculated the corresponding dissimilarities (dissTOM) as 1 − TOM. For hierarchical clustering of genes/OTUs we used dissTOM as a distance measure and set the minimum module size (number of genes) to 30 (number of OTUs to 3) to detect modules. To quantify the co-expression similarities of entire modules, their eigengenes/eigenOTUs were calculated and subsequently used to associate with phenotypic traits. We chose modules with $P$ values <0.05 as significantly associated modules. Then we calculated Spearman correlation between normalized genes/OTUs expression and phenotypic traits as well as gene/OTU membership value using 'signedKME' function from WGCNA package in above selected significant modules. The key genes/OTUs were determined by selecting the overlapping genes/OTUs between genes/OTUs which have Spearman correlation coefficients >0.7 or < −0.7 and $P$ values <0.05 and genes/OTUs which have membership value >0.7 or <−0.7. Then we did Spearman correlation between the key genes and OTUs to find significant associations by selecting *rho* >0.7 or <−0.7 and FDR adjusted $P$ values <0.05. Network visualization was performed in Cytoscape as described above. For each keystone OTU, we functionally classified the interacted root genes that enriched into different GO terms using g:Profiler (Raudvere et al., 2019).

**Validation experiment 1: comparison of different growth stages**

To verify if *Massilia* is associated with maize developmental stages, we grew wild type B73 in control soil and harvested their root samples at seedling stage and flowering stage. RNA and DNA extractions were performed as described above. We then performed the RNAseq and 16S rRNA (V3−V4) gene sequencing for the root samples. Raw reads were processed following a similar workflow as previously described (Yu et al., 2021). Briefly, paired-end 16S rRNA amplicon sequencing reads were assigned to samples based on their unique barcode and truncated by cutting off the barcode and primer sequence. Sequence analyses were performed by QIIME 2 software (v2020.6)(Bolyen et al., 2019). Raw sequence data were demultiplexed

and quality filtered using the q2-demux plugin followed by denoising with DADA2 (Callahan et al., 2016) (via q2-dada2). Sequences were truncated at position 250 and each unique sequence was assigned to a different ASV. Taxonomy was assigned to ASVs using the q2-feature-classifier (Bokulich et al., 2018) and the classify-sklearn naïve Bayes taxonomy classifier against the SSUrRNA SILVA 99% OTUs reference sequences (v138) (Quast et al., 2013) at each taxonomic rank (kingdom, phylum, class, order, family, genus, species). Mitochondria- and chloroplast-assigned ASVs were eliminated. Out of the remaining sequences (only ASVs with >5 reads in ≥2 samples) were kept to build an ASV table.

All downstream analyses were performed in R (v4.2.2) (R Core Team, 2022). For α-diversity indices, the Shannon index was calculated using ASV table rarefied to 1000 reads. For all following analyses abundant ASVs were used, so ASVs which express ≤0.05% relative abundance within ≤5% samples were removed. After filtering taxa, the samples with ≤1000 reads were also removed. Bray-Curtis distances between samples were calculated using the ASV table that was normalized using the 'varianceStabilizingTransformation' function from DESeq2 (v1.38.3) package (Love et al., 2014) in R. If not specified, the following data analysis is based on the normalized ASV table. Principal coordinate analyses were performed using the 'ordinate' function in R package phyloseq (v1.42.0) (McMurdie & Holmes, 2013). To test the marginal effects of treatment and genotype on the microbial composition community, a permutation-based PERMANOVA test was performed using Bray–Curtis distance matrix between pairs of bacterial samples using 'adonis2' function in R package vegan (v2.6.4) (Oksanen et al., 2020). To identify genera differentially expressed between stages, ASVs were grouped by genus and unidentified genera were removed. Only genera that express >0.1% in ≥3 samples were kept. Then relative abundance of each genus was compared between stages using Wilcoxon rank sum test (Haynes, 2013). All plots were produced using R package ggplot2 (v3.4.2) (Wickham, 2016). To the end, we performed similar Spearman correlation analyses using the differentially expressed genes and differentially abundant OTUs for seedling and flowering stage. Network visualization, definition of keystone OTU and functional classification of GO terms were performed as described above.

**Validation experiment 2: *Massilia* inoculation in early flowering maize mutants**

To explore the potential effects of *Massilia* on plant performance, we carried out a growth promotion assay for wild type N28 and its early flowering time mutant C22-4 by inoculation with a single *Massilia* isolate, or synthetic communities with different members of isolates e.g. SynCom 1 with 12 members, SynCom 2 with 13 members. Before the inoculation experiment,

we first aligned the sequences of different *Massilia* strains to the 16S (V1−V9) sequence of the hub OTU3535 using BLASTn (v2.6.0) (Camacho et al., 2009) with default parameters. We used the same soil as in the previous soil pot experiment and sterilized it according to an established protocol (Yu et al., 2021). Seed sterilization, isolates preparation, root inoculation and growth assay were performed as previously reported (He et al., 2023). The wild type plants and mutants were grown in the greenhouse (16/8 h light/dark and 26/18 °C) for 1 month. Then plants were harvested, and the shoot dry biomass and number of fully developed leaves was determined.

**Data availability**

All raw plant RNA-seq data, bacterial 16S sequencing data reported in this paper were deposited in the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under accession no. PRJNA1018308. RNA-seq reads were mapped to the maize reference genome sequence v.5 (http://ftp.ensemblgenomes.org/pub/plants/current/fasta/zea_mays/dna/Zea_mays.Zm-B73-REFERENCE-NAM-5.0.dna.toplevel.fa.gz) and were annotated based on the reference gene models v.5 (http://ftp.ensemblgenomes.org/pub/plants/current/gtf/zea_mays/Zea_mays.Zm-B73-REFERENCE-NAM-5.0.53.chr.gtf.gz). The GO terms were annotated using g:Profiler (https://biit.cs.ut.ee/gprofiler/gost). We deposited all raw OTU table, OTU taxonomy, gene table and customized scripts including all downstream analysis and the association of gene, bacteria and phenotypic traits in the following GitHub repository:https://github.com/Danning16/Integrated-analysis-of-bacteria-and-transcriptome-along-the-primary-root-in-maize.

# References

Aira, M., Gómez-Brandón, M., Lazcano, C., Bååth, E., & Domínguez, J. (2010). Plant genotype strongly modifies the structure and growth of maize rhizosphere microbial communities. *Soil Biology and Biochemistry*, *42*(12), 2276–2281.

Attia, S., Russel, J., Mortensen, M. S., Madsen, J. S., & Sørensen, S. J. (2022). Unexpected diversity among small-scale sample replicates of defined plant root compartments. *ISME Journal*, *16*(4), 997–1003.

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., Huttley, G. A., & Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*, *6*(1), 1–17.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857.

Bouffaud, M. L., Kyselková, M., Gouesnard, B., Grundmann, G., Muller, D., & Moënne-Loccoz, Y. (2012). Is diversification history of maize influencing selection of soil bacteria by roots? *Molecular Ecology*, *21*(1), 195–206.

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

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 1–9.

Coudert, Y., Périn, C., Courtois, B., Khong, N. G., & Gantet, P. (2010). Genetic control of root development in rice, the model cereal. *Trends in Plant Science*, *15*(4), 219–226.

Csárdi, G., Nepusz, T., Traag, V., Kirill, S., Zanini, H., Noom, F., & Müller, D. (2023). *igraph: Network Analysis and Visualization in R*.

Delhaize, E., Rathjen, T. M., & Cavanagh, C. R. (2015). The genetics of rhizosheath size in a multiparent mapping population of wheat. *Journal of Experimental Botany*, *66*(15), 4527–4536.

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., & Taylor,

Christopher M Huttenhower, Curtis Langille, M. G. (2020). PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology*, *38*(6), 685--688.

Edgar, R. C. (2013). UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, *10*(10), 996–998.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, *27*(16), 2194–2200.

Edwards, J., Santos-Medellín, C. M., Liechty, Z. S., Nguyen, B., Lurie, E., Eason, S., Phillips, G., & Sundaresan, V. (2018). Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life cycle in field-grown rice. *PLoS Biology*, *16*(2), 1–28.

Edwards, J., Johnson, C., Santos-Medellín, C., Lurie, E., Podishetty, N. K., Bhatnagar, S., Eisen, J. A., Sundaresan, V., & Jeffery, L. D. (2015). Structure, variation, and assembly of the root-associated microbiomes of rice. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(8), e911–e920.

Friedman, J., & Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLoS Computational Biology*, *8*(9), 1–11.

Gao, X., Lin, H., Revanna, K., & Dong, Q. (2017). A Bayesian taxonomic classification method for 16S rRNA gene sequences with improved species-level accuracy. *BMC Bioinformatics*, *18*(1), 1–10.

Garrido-Oter, R., Nakano, R. T., Dombrowski, N., Ma, K. W., McHardy, A. C., & Schulze-Lefert, P. (2018). Modular traits of the rhizobiales root microbiota and their evolutionary relationship with symbiotic rhizobia. *Cell Host and Microbe*, *24*(1), 155-167.

Gruber, B. D., Giehl, R. F. H., Friedel, S., & von Wirén, N. (2013). Plasticity of the *Arabidopsis* root system under nutrient deficiencies. *Plant Physiology*, *163*(1), 161–179.

Gutiérrez-Luna, F. M., López-Bucio, J., Altamirano-Hernández, J., Valencia-Cantero, E., De La Cruz, H. R., & Macías-Rodríguez, L. (2010). Plant growth-promoting rhizobacteria modulate root-system architecture in *Arabidopsis thaliana* through volatile organic compound emission. *Symbiosis*, *51*(1), 75–83.

Gutjahr, C., Sawers, R. J. H., Marti, G., Andrés-Hernández, L., Yang, S. Y., Casieri, L., Angliker, H., Oakeley, E. J., Wolfender, J. L., Abreu-Goodger, C., & Paszkowski, U. (2015). Transcriptome diversity among rice root types during asymbiosis and interaction

with arbuscular mycorrhizal fungi. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(21), 6754–6759.

Haichar, F. el Z., Santaella, C., Heulin, T., & Achouak, W. (2014). Root exudates mediated interactions belowground. *Soil Biology and Biochemistry*, *77*, 69–80.

Haynes, W. (2013). Wilcoxon rank sum test. *Encyclopedia of Systems Biology*, *3*(1), 2354--2355.

He, X., Wang, D., Jiang, Y., Li, M., Delgado-Baquerizo, M., McLaughlin, C., Marcon, C., Guo, L., Baer, M., Moya, Y. A. T., von Wirén, N., Deichmann, M., Schaaf, G., Piepho, H.-P., Yang, Z., Yang, J., Yim, B., Smalla, K., Goormachtig, S., … Yu, P. (2023). Heritable microbiome variation is correlated with source environment in locally adapted maize varieties. *BioRxiv*. https://doi.org/10.1101/2023.01.10.523403

Hochholdinger, F., & Feix, G. (1998). Early post-embryonic root formation is specifically affected in the maize mutant *lrt1*. *Plant Journal*, *16*(2), 247–255.

Hochholdinger, F., Wen, T. J., Zimmermann, R., Chimot-Marolle, P., Da Costa E Silva, O., Bruce, W., Lamkey, K. R., Wienand, U., & Schnable, P. S. (2008). The maize (*Zea mays* L.) *roothairless3* gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. *Plant Journal*, *54*(5), 888–898.

Hochholdinger, F., Woll, K., Sauer, M., & Dembinsky, D. (2004). Genetic dissection of root formation in maize (*Zea mays*) reveals root-type specific developmental programmes. *Annals of Botany*, *93*(4), 359–368.

Hochholdinger, F., Yu, P., & Marcon, C. (2018). Genetic control of root system development in maize. *Trends in Plant Science*, *23*(1), 79–88.

Hubbard, C. J., Brock, M. T., Van Diepen, L. T., Maignien, L., Ewers, B. E., & Weinig, C. (2018). The plant circadian clock influences rhizosphere community structure and function. *ISME Journal*, *12*(2), 400–410.

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, *41*(1), 1–11.

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*.

Liao, Y., Smyth, G. K., & Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930.

López-Bucio, J., Millán-Godínez, M., Méndez-Bravo, A., Morquecho-Contreras, A., Ramírez-Chávez, E., Molina-Torres, J., Pérez-Torres, A., Higuchi, M., Kakimoto, T., & Herrera-Estrella, L. (2007). Cytokinin receptors are involved in alkamide regulation of root and shoot development in *Arabidopsis*. *Plant Physiology*, *145*(4), 1703–1713.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21.

Lu, T., Ke, M., Lavoie, M., Jin, Y., Fan, X., Zhang, Z., Fu, Z., Sun, L., Gillings, M., Peñuelas, J., Qian, H., & Zhu, Y. G. (2018). Rhizosphere microorganisms can influence the timing of plant flowering. *Microbiome*, *6*(1), 1–12.

Lynch, J. (2013). Steep, cheap and deep: An ideotype to optimize water and N acquisition by maize root systems. *Annals of Botany*, *112*(2), 347–357.

Lynch, J. (2022). Harnessing root architecture to address global challenges. *Plant Journal*, *109*(2), 415–431.

Marschner, P. (2012). Rhizosphere biology. In *Marschner's Mineral Nutrition of Higher Plants (Third Edition)*. Academic Press, 369–388.

Marzec, M., Melzer, M., & Szarejko, I. (2015). Root hair development in the grasses: What we already know and what we still need to know. *Plant Physiology*, *168*(2), 407–414.

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, *8*(4).

Mendes, R., Garbeva, P., & Raaijmakers, J. M. (2013). The rhizosphere microbiome: Significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiology Reviews*, *37*(5), 634–663.

Motte, H., Vanneste, S., & Beeckman, T. (2019). Molecular and environmental regulation of root development. *Annual Review of Plant Biology*, *70*, 465–488.

Newman, A., Picot, E., Davies, S., Hilton, S., Carré, I. A., & Bending, G. D. (2022). Circadian rhythms in the plant host influence rhythmicity of rhizosphere microbiota. *BMC Biology*, *20*(1), 1–15.

Oksanen, A. J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P. R., Hara, R. B. O., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Szoecs, E. (2020). *vegan: Community Ecology Package*. https://cran.r-project.org/package=vegan

Pausch, J., Loeppmann, S., Kühnel, A., Forbush, K., Kuzyakov, Y., & Cheng, W. (2016). Rhizosphere priming of barley with and without root hairs. *Soil Biology and Biochemistry*, *100*, 74–82.

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

Poitout, A., Martinière, A., Kucharczyk, B., Queruel, N., Silva-Andia, J., Mashkoor, S., Gamet, L., Varoquaux, F., Paris, N., Sentenac, H., Touraine, B., & Desbrosses, G. (2017). Local signalling pathways regulate the *Arabidopsis* root developmental response to *Mesorhizobium loti* inoculation. *Journal of Experimental Botany*, *68*(5), 1199–1211.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*, 590–596.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*, 191–198.

Revelle, W. (2023). *psych: Procedures for Psychological, Psychometric, and Personality Research*. https://cran.r-project.org/package=psych

Rogers, E. D., & Benfey, P. N. (2015). Regulation of plant root system architecture: Implications for crop advancement. *Current Opinion in Biotechnology*, *32*, 93–98.

Salvi, S., Sponza, G., Morgante, M., Tomes, D., Niu, X., Fengler, K. A., Meeley, R., Ananiev, E. V., Svitashev, S., Bruggemann, E., Li, B., Hainey, C. F., Radovic, S., Zaina, G., Rafalski, J. A., Tingey, S. V., Miao, G. H., Phillips, R. L., & Tuberosa, R. (2007). Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(27), 11376–11381.

Shannon, P., Markiel, A., Ozier, O., Baliga, N., Wang, J., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504.

Wang, Q., Wang, K., Wu, W., Giannoulatou, E., Ho, J. W. K., & Li, L. (2019). Host and microbiome multi-omics integration: applications and methodologies. *Biophysical Reviews*, *11*(1), 55–65.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Yu, P., Baldauf, J. A., Lithio, A., Marcon, C., Nettleton, D., Li, C., & Hochholdinger, F. (2016). Root type-specific reprogramming of maize pericycle transcriptomes by local high nitrate results in disparate lateral root branching patterns. *Plant Physiology*, *170*(3), 1783–1798.

Yu, P., Eggert, K., von Wirén, N., Li, C., & Hochholdinger, F. (2015). Cell type-specific gene expression analyses by RNA sequencing reveal local high nitrate-triggered lateral root initiation in shoot-borne roots of maize by modulating auxin-related cell cycle regulation. *Plant Physiology*, *169*(1), 690–704.

Yu, P., He, X., Baer, M., Beirinckx, S., Tian, T., Moya, Y. A. T., Zhang, X., Deichmann, M., Frey, F. P., Bresgen, V., Li, C., Razavi, B. S., Schaaf, G., von Wirén, N., Su, Z., Bucher, M., Tsuda, K., Goormachtig, S., Chen, X., & Hochholdinger, F. (2021). Plant flavones enrich rhizosphere *Oxalobacteraceae* to improve maize performance under nitrogen deprivation. *Nature Plants*, *7*(4), 481–499.

Yu, P., & Hochholdinger, F. (2018). The role of host genetic signatures on root–microbe interactions in the rhizosphere and endosphere. *Frontiers in Plant Science*, *9*, 1–5.

Yu, P., & Hochholdinger, F. (2023). Genetic and environmental regulation of root growth and development. In *Marschner's Mineral Nutrition of Plants*. LTD.

Zamioudis, C., Mastranesti, P., Dhonukshe, P., Blilou, I., & Pieterse, C. M. J. (2013). Unraveling root developmental programs initiated by beneficial *Pseudomonas* spp. bacteria. *Plant Physiology*, 162, 304–318.

Zhang, J., Liu, Y. X., Zhang, N., Hu, B., Jin, T., Xu, H., Qin, Y., Yan, P., Zhang, X., Guo, X., Hui, J., Cao, S., Wang, X., Wang, C., Wang, H., Qu, B., Fan, G., Yuan, L., Garrido-Oter, R., … Bai, Y. (2019). *NRT1.1B* is associated with root microbiota composition and nitrogen use in field-grown rice. *Nature Biotechnology*, *37*(6), 676–684.

# 4 Discussion

The microorganisms living in soil are the main source of the root and rhizosphere microbiome in plants. The relationship between microorganisms and plants can be mutually beneficial. Microorganisms facilitate plants to acquire mineral nutrients (Salas-González et al., 2021), and protect against biotic and abiotic stresses (Cheng et al., 2019; Oldroyd & Leyser, 2020), while plants provide nutrient-rich resources such as carbohydrates, amino acids, and vitamins as energy source for microorganisms (Lanfranco et al., 2018; Poole et al., 2018). Soil type and environmental factors are the main drivers of the composition and structure of microbial communities in the rhizosphere. Previous studies have shown that the genotype had a strong impact on its microbial community composition with small sample size experiments in maize (Brisson et al., 2019; Favela et al., 2021; Wagner et al., 2020; Wallace et al., 2018; Walters et al., 2018). However, how the maize genotype and its source environmental factors influence its microbiome assembly and to what extent the microbiome improves the host performance under abiotic stresses is largely unknown, especially at the population level. In particular, what host functional genes affect the abundance and enrichment of specific microbes remains obscure.

In this thesis, we investigated the genotype and plant source environmental effects on the microbial diversity and specific microbial members in the rhizosphere and root by employing a large number of genotypes in maize with different domestication status under different abiotic stresses. We also explored how much the root and rhizosphere microbiome variation is attributed to host genetics, and which specific microbes are associated with host genomic variation underlying root development. Through diverse root type mutants in maize and validation experiments, we demonstrated that the keystone bacterial taxon *Massilia* may play an important role in associating with lateral root development as well as promoting plant performance and tolerance to nitrogen deficiency.

## 4.1 Significant impact of the plant genotype on microbiome composition and structure

To discover host genotype effects on the root and rhizosphere microbiome in maize, we employed two different strategies including permutational multi-variate analysis of variance (PERMANOVA) and correlation analyses between host genetic distance and microbiome composition. In chapter 2, principal coordinate analyses followed by PERMANOVA indicated that the genotype had a small but significant influence, and larger effect on fungi ($R^2 = 0.12$, $R^2 = 0.098$, $P < 0.001$) than bacteria ($R^2 = 0.064$, $R^2 = 0.058$, $P < 0.001$) in both rhizosphere and

root across different soil treatments. Several studies in maize have also shown that the genotype has a small impact on the bacterial or fungal community composition in the rhizosphere (Lund et al., 2022; Peiffer et al., 2013; Wagner et al., 2020). Specifically, Peiffer et al. (2013) demonstrated a ~5% genotype effect on the rhizosphere bacterial community using 27 inbred lines grown in 5 fields, while they observed a ~18% genotype effect within each field. Lund et al. (2022) detected a ~30% genotype effect on the rhizosphere bacterial community using 6 landraces under a single controlled condition. Wagner et al. (2020) reported that the genotype has similar effects on bacteria ($R^2 = 0.066$, $P = 0.001$) and fungi ($R^2 = 0.063$, $P = 0.001$) of the maize rhizosphere using 7 inbred lines and 11 hybrids in two fields. The highly variable genotypes, sample size, sampling time and different environments made it very difficult to compare the genotype effect on microbial community, which might be hidden by different soil and climate conditions. Alternatively, we detected only a significant soil treatment and genotype interaction effect on the rhizosphere bacteria, demonstrating that some specific genotypes have a stronger influence on rhizosphere bacteria under specific stress conditions than under control conditions.

At the population level, we detected a stronger correlation between the host genetic distance and the rhizosphere than that with the root in both bacteria and fungi respectively, using the Mantel statistical test after controlling soil treatments. However, there was no significant correlation between the maize kinship matrix and bacterial composition in the rhizosphere (Peiffer et al., 2013). One possible explanation is that the rhizosphere microbiome is indirectly influenced by host genotype which may be hindered by environmental effects. Nevertheless, recent studies in maize showed that the host genetic distance was correlated with the rhizosphere microbiome dissimilarity (Deng et al., 2021; Lund et al., 2022; Wagner et al., 2020), which is consistent with our results.

Moreover, our results in chapter 3 demonstrated that the genotype explained more of the variance of bacterial composition of the primary root rhizosphere, the primary root, cortex and stele ($R^2 = 0.14−0.27$) than that of the lateral root rhizosphere and the lateral roots ($R^2 = 0.11−0.13$). All these data indicate that the genotype effect is small but significant, which may depend on different environments, compartments and specific root tissues.

## 4.2 Stresses increased broad-sense heritability in the rhizosphere bacteria

Broad-sense heritability is defined as the proportion of phenotypic variability that is attributable to genetic factors (Klug et al., 2016). Higher heritability estimates suggest that genetic

variability has a larger influence on the variability of microbial abundance. Our results in chapter 2 showed, in general, that heritability in the rhizosphere is higher than in the root at family, genus and amplicon sequence variant (ASV) level, which is consistent with our above PERMANOVA and correlation test results showing stronger effects between host genetics and the rhizosphere than the root. In our study, the maximum heritability of bacteria at ASV level was 0.63 in the rhizosphere under control soil and 0.76 in the root under low phosphorus soil. In previous studies in maize the maximum heritability was 0.25 across five fields and 0.54 in one field (Walters et al., 2018), whereas the maximum heritability was 0.76 under high nitrogen and 0.84 under low nitrogen in one field (Meier et al., 2022). The maximum heritability of fungi at ASV level in our experiment was 0.88 in the rhizosphere under control soil and 0.83 in the root under drought soil. The reason for a relatively higher heritability compared to previous studies can be due to the controlled climate conditions and homogenous soil property in our phytochamber. Thus, reduced environmental variance may result in higher heritability (Peiffer et al., 2013).

Furthermore, both of our results and Meier et al. (2022) showed that heritability varies considerably within family, genus and ASV levels. At the family level, the heritability of *Oxalobacteraceae* is significantly higher than that of other families in roots under low nitrogen soil. This suggests that specific taxa may have very strong associations with host genetics (Peiffer et al., 2013). In addition, nutrient stresses increased heritability especially in the rhizosphere bacteria in chapter 2. However, this is not the same scenario in fungi. Similarly, Meier et al. (2022) found that bacteria were comparatively more heritable under low nitrogen than high nitrogen conditions. These two population studies coherently demonstrate a much stronger selectiveness of host genetics on the rhizosphere bacteria than other compartments under stress conditions. A potential explanation is that the function of the rhizosphere confers to a major defense layer to stresses and may be substantially adjusted by plants through root exudation when they face stresses (Yu et al., 2021).

## 4.3 Amplicon sequence variants from *Massilia* showed high associations with host genomic loci

Based on the causality of microbiome heritability ($H^2 > 0.1$), we performed GWAS to elucidate the associations between host genomic regions and specific microbial abundance at the family, genus, and ASV level under each soil condition. We detected a total of 42 significant associations between 32 bacterial traits and 35 QTLs in the rhizosphere, 491 associations between 227 bacterial traits and 193 QTLs in the root, 170 associations between 121 fungal

traits and 107 QTLs in the rhizosphere, and 113 associations between 51 fungal traits and 81 QTLs in the root. While Wallace et al. (2018) only identified two *Methylobacteria* OTUs associating with host genetic loci in maize leaf, Meier et al. (2022) identified 622 genetic loci that were significantly linked to 104 bacterial ASVs in maize rhizosphere. These results highlight that host genetics confers to a much stronger selectiveness on microbial assembly in the rhizosphere/root than in the leaf. One potential explanation could be a direct pathway for plant-microbe interactions via root exudation (Doornbos et al., 2012), while the leaf microbiota is mainly mediated by the indirect pathway of stomata openness (Sohrabi et al., 2023). Furthermore, a study in tomato identified a QTL region significantly associated with *Massilia* in the rhizosphere with a larger effect size than other genera (Oyserman et al., 2022). Their findings support our data that several ASVs that belong to *Massilia* had higher heritability and larger explained variance by host genomic loci than other ASVs in the root under low nitrogen condition. This may suggest that some microbes belonging to *Massilia* have co-evolved with maize during domestication and their abundances are genetically controlled by host genomic regions.

## 4.4 The keystone bacterial taxon *Massilia* drives the structure and function of the microbial community in maize roots

In chapter 2, we identified the bacterial genus *Massilia* as one of the most represented keystone taxa within bacteria-fungi trans-kingdom networks across different soil treatments in both rhizosphere and root compartments. In addition, in chapter 3, we also identified *Massilia* as the keystone taxon within bacterial networks in the lateral root and cortex compartments. The bacterial taxon *Oxalobacteraceae* to which *Massilia* belongs was significantly enriched in primary roots (RA = 0.05), lateral roots (RA = 0.04) and cortex (RA = 0.08), and was detected as an indicator taxon in the cortex. This may indicate that *Massilia* plays a key role in driving the structure of the microbial community in the whole maize root system. Furthermore, in lateral roots, host genes significantly correlated with *Massilia* were enriched in functions "tryptophan biosynthetic process", "rhythmic process" and "RNA biosynthetic processes". We further detected significant correlations between some ASVs belonging to *Massilia* and genes in associations with maize developmental stages in chapter 3. These data further indicate that *Massilia* may influence host plant metabolism, growth and performance during its whole life cycle. In plants, various members belonging to the *Rhizobiales* and *Burkholderiales* orders (to which *Massilia* belongs to) were identified as keystone taxa in different studies (Banerjee et al., 2018). Two recent studies in maize identified several potential keystone taxa which had

significant associations with yield and nutrient level (Lang et al., 2021; Wang et al., 2022), however, there is no overlap with our keystone taxa since they used different soil types from diverse origins in experiments and different strategies to define keystone taxa. Nevertheless, *Massilia* has been identified as a keystone taxon in crops (Lewin et al., 2021), played key role in protecting against pathogens in rice (Li et al., 2021), and improved salt tolerance in maize (Krishnamoorthy et al., 2016). Therefore, functional validation experiments are necessary to verify the key role of the potential keystone taxa.

## 4.5 Plant source environmental factors improved the genomic prediction accuracy of microbial abundance

To address the hypothesis that variation in the plant microbiome reflects local adaptation to the native environment, we assessed the potential effect of climatic and edaphic descriptors of the source environment on the microbiome in our phytochamber experiments. Firstly, we clustered the microbial ASVs into different co-abundant modules and investigated the correlations between microbial modules and source environmental factors using weighted correlation network analysis (WGCNA, Langfelder & Horvath, 2008). Interestingly, we found that a specific *Massilia* enriched module was significantly negatively (R = −0.28, *P* = 0.0039) correlated with total soil nitrogen content of the original sites of collection, exclusively in the low nitrogen soil condition, thus mirroring a potential selectiveness of *Massilia* when soil nitrogen is limited in the natural habitats. Future studies need to address the abundance of *Massilia* taxon and its relationship with the availability of soil nitrogen when plants are grown in their native environments. Nevertheless, our finding highlights great potential correlations between microbial diversity and environmental legacy effects using the natural population.

To further understand the importance of environmental factors on the explanation of microbial abundance, we applied the genomic prediction strategy and compared the prediction ability of a genomic model, an environment model as well as a combined genomic and environment model on the variation of microbial abundance. Our results in chapter 2 demonstrated that microbial abundance could be predicted more accurately with source environmental factors or a combination of these with plant genetic markers than with genetic markers alone. Although there are no comparable studies available currently, Edwards et al. (2023) demonstrated that microbes tend to colonize their host plants when they are grown in their native environments (Edwards et al., 2023). This supports our result that the native environment of maize improved the prediction accuracy of microbial abundance, which may suggest that the environment plays an influential role in determining microbiome composition through its impact on host genetics.

## 4.6 Environmental genome-wide association study facilitates the identification of genetic loci associated with the keystone taxon *Massilia*

To deepen the understanding of potential host adaptive genome in association with specific microbes, we specifically focused on the bacterial family *Oxalobacteraceae* due to its high abundance and prevalence, its highest heritability in low nitrogen soil, and its important role in maize tolerance to nitrogen deficiency (Yu et al., 2021). We applied our existing environmental Random Forest (RF) model with the 129 accessions as a training set to predict the abundance of ASVs in *Oxalobacteraceae* for 1,781 previously genotyped traditional maize varieties sourced from diverse environments across Mexico ( Romero Navarro et al., 2017). Among all predicted ASVs, the best predicted (RF model $R^2 = 0.28$) was ASV37, belonging to the genus *Massilia*, in the root under low nitrogen treatment. Then we identified significant associations between the predicted abundance of ASV37 and host genetic loci using the 1,781 varieties via GWAS, with a top association at SNP S4_10445603 near the gene Zm00001d048945. Moreover, across the 1,781 panel, the minor allele frequency at SNP S4_10445603 was positively correlated with the predicted abundance of ASV37 but negatively correlated with source soil nitrogen content. This may indicate that the abundance of *Massilia* is genetically controlled by the host through its native environmental impact. Moreover, this gene Zm00001d048945 is predicted to encode TPX2 domain containing protein related to the WAVE-DAMPENED2 microtubule binding protein which was recently demonstrated to involve in lateral root initiation in *Arabidopsis* (Qian et al., 2023). Consistently, maize mutants for transposon insertions in the gene Zm00001d048945 in our experiments showed a significant reduction in lateral root density. This data supports the hypothesis that allelic variation in the gene Zm00001d048945 contributes to adaptation to nitrogen-poor soil by enhancing associations with *Massilia*.

## 4.7 A lateral root mutation dramatically influences host gene expression and bacterial community composition

To investigate the effect of specific root phenotypes on microbiome assembly, we used lateral root mutants, root hair mutants and wild types to compare the mutation effects of lateral roots and root hairs on transcriptome and microbiome assemblage. As shown by PCoA and PCA in chapter 3, mutations that lead to lateral root defects ($R^2 = 0.2-0.25$, $P = 5e-04$) had much stronger effects on the bacterial community composition and transcriptomic changes than mutations that result in impaired root hairs ($R^2 = 0.045-0.1$, $P = 5e-04$). This observation was

in line with the differential expression/abundance analyses between each mutant and wild type. There were about ten times more differentially expressed genes (DEGs) / differentially abundant OTUs (DAOs) in lateral root mutants than in other mutants compared to wild types. Our co-occurrence network analysis also supports the notion that lateral root related bacterial communities had the most complex co-abundance network. This is consistent with the most complex connections between prokaryotes and protists on lateral roots in maize (Rüger et al., 2021) as well as the finding that AMF preferentially colonize large lateral roots in rice (Gutjahr et al., 2015). One possible explanation is that lateral roots are the most active area of exudation which attracts more microorganisms from soil, and that lateral root primordia provide space for microbiome invasion (Baudoin et al., 2002; Park et al., 2004). Both functional prediction analysis and KEGG pathways of DEGs in chapter 3 demonstrated that a substantial proportion of metabolic pathways were enriched in lateral root mutants. Furthermore, the correlation analyses in chapter 3 suggested that there were more associations between host gene expression and bacterial abundance in lateral roots than in primary roots. This supports the notion that lateral roots may drive the function of microbial community to facilitate the host to uptake nutrients from soil.

## 4.8 The bacterial taxon *Massilia* alone can contribute to lateral root formation, biomass production and nitrogen tolerance in maize

We next tried to explore the direct evidence of gene-encoded root trait variation and specific microbiome assembly. Using Zm00001d048945 transposon-insertion mutants from two different genetic backgrounds, we compared the bacterial relative abundance between wild type and mutant plants, and detected that *Massilia* was the most abundant genus that significantly decreased in both mutants in comparison to the wild type plants in nitrogen-poor soil. These results together with the above suggest that the gene Zm00001d048945 modules lateral root development meanwhile contributes to variation of *Massilia* abundance in nitrogen deficient soil. Our previous study has shown that root-derived flavones i.e. apigenin and luteolin are important drivers for mediating the beneficial association of *Massilia* with lateral root development in maize (Yu et al., 2021). Therefore, we quantified apigenin and luteolin in the Zm00001d048945 mutants and found that the mutants accumulated significantly more apigenin and luteolin in comparison to wild type plants. Thus, we confirmed that the potential linkage between lateral root development and *Massilia* depends on root exudation of flavones in maize. A recent study demonstrated the importance of the bacteria genus *Variovorax* for root development in *Arabidopsis* by manipulating plant hormone level (Finkel et al., 2020). Another

study demonstrated that *Pseudomonas* promoted lateral root formation by stimulating both lateral root initiation and lateral root outgrowth through auxin signaling pathway in *Arabidopsis* (Zamioudis et al., 2013). Their results together with our findings demonstrate that specific bacterial taxa may affect root development and architecture by interacting with specific metabolic pathways of host plants. Furthermore, our two separate inoculation experiments in chapter 2 provided a strong support for the promotion of lateral root formation and improved growth performance via inoculation of single *Massilia* in mutants in nitrogen-poor soil, thus highlighting a great potential value of root trait interactions with keystone microbial taxa underlying plant fitness under abiotic stresses.

Moreover, in chapter 3, we demonstrated that *Massilia* had significant associations with 103 genes expressed in lateral roots, meanwhile was strongly associated with shoot biomass through an integrated network analysis. The enriched GO term of those genes was "circadian rhythm" and the gene *timing of cab expression 1* (*toc1*) functioning in flower development was identified. Previous studies in *Arabidopsis* demonstrated that rhizosphere microbiota modulated flowering time by IAA production and can also promote plant growth by improving nitrogen availability in soil (Lu et al., 2018; Panke-Buisse et al., 2015). This supports our results that tryptophan metabolic processes or IAA biosynthesis may drive the causal interactions with specific beneficial microbes to improve plant growth in nitrogen limited soil. In addition, Lu et al. (2018) showed that addition of IAA resulted in delayed flowering time and downregulated gene expression of flowering related genes including *toc1* gene. Similarly, our further experiment verified that specific ASVs that belong to *Massilia* had significant correlations with genes functioning in flowering related pathways such as pollen recognition. Through *Massilia* inoculation experiments, we detected a single inoculation of *Massilia* significantly promoted maize dry biomass and the number of leaves in the early flowering mutants, whereas we did not observe any effects on these traits in the wild type plants. This may suggest that the function of *Massilia* in promoting biomass and leaf number might depend on flowering time in maize.

Taken together, we found that environmentally-adapted alleles during domestication of maize may promote root development and associated microbiomes to improve the plant performance. Our results confirm that host genetics impacts keystone microbe assembly and root development under soil nitrogen stress conditions to achieve optimal plant growth. These findings help to close the knowledge gap between how plants select and enrich the soil microbiome, and how this functional interaction between host plants and their microbiomes can be translated into crop resilience to nutrient limitation. These results also provide new insights

to the relationship between plant genetics, root development, associated microbiomes and plant fitness.

## 4.9 Future perspectives

Overall, the aim of this doctoral thesis was to investigate the beneficial interactions between host genetics and its microbial community that further promote plant growth and performance especially under stress conditions. The major finding of this thesis was that there are specific bacterial taxa which promote root development and also shoot biomass under low nitrogen conditions, but that this promotion is only specific to some maize genotypes. The mechanism behind this still remains unclear. To better understand the causal relationships between host genetics, microbial community and host performance, future experiments are needed by using specific plant mutants, microbe mutants and integrative analyses of multi-omics data including host gene expression, host metabolomes, and microbial pan-genomics.

The most challenging part to identify the causal relationship between host plants and their associated microbiome is that the microbiome does not only interact with host plants and environments, but also with other microbial members. To verify the specific influence of some taxa, a collection of microbial strains needs to be isolated from maize root systems and form a synthetic community to perform inoculation experiments. Until now our lab only has 17 strains from the genus *Massilia*. Therefore, using these taxa to construct a community and verify the function of some taxa is limited. Moreover, the mechanism underlying the association between flowering related genes and *Massilia* is still unclear. To elucidate the causal relationship, the whole genomic information and biochemical properties of *Massilia* need to be investigated at the molecular level.

Our finding that linkage between host genetic loci and variation of *Massilia* abundance underlying root development is encouraging. However, our experiments were performed in our phytochamber with homogenous soil and climatic conditions using maize genotypes mostly from landraces. Therefore, translational work using more maize genotypes from inbred lines and hybrids is necessary to comprehensively investigate this relationship in different fields.

# 5 References

Almario, J., Jeena, G., Wunder, J., Langen, G., Zuccaro, A., Coupland, G., & Bucher, M. (2017). Root-associated fungal microbiota of nonmycorrhizal *Arabis alpina* and its contribution to plant phosphorus nutrition. *Proceedings of the National Academy of Sciences*, **114**, e9403–e9412.

Angelard, C., Colard, A., Niculita-Hirzel, H., Croll, D., & Sanders, I. R. (2010). Segregation in a mycorrhizal fungus alters rice growth and symbiosis-specific gene transcription. *Current Biology*, **20**, 1216–1221.

Asari, S., Tarkowská, D., Rolčík, J., Novák, O., Palmero, D. V., Bejai, S., & Meijer, J. (2017). Analysis of plant growth-promoting properties of *Bacillus amyloliquefaciens* UCMB5113 using *Arabidopsis thaliana* as host plant. *Planta*, **245**, 15–30.

Baer, M., Taramino, G., Multani, D., Sakai, H., Jiao, S., Fengler, K., & Hochholdinger, F. (2023). Maize *lateral rootless 1* encodes a homolog of the DCAF protein subunit of the CUL4-based E3 ubiquitin ligase complex. *New Phytologist*, **237**, 1204–1214.

Baeyens, J., Kang, Q., Appels, L., Dewil, R., Lv, Y., & Tan, T. (2015). Challenges and opportunities in improving the production of bio-ethanol. *Progress in Energy and Combustion Science*, **47**, 60–88.

Banerjee, S., Schlaeppi, K., & van der Heijden, M. G. A. (2018). Keystone taxa as drivers of microbiome structure and functioning. *Nature Reviews Microbiology*, **16**, 567–576.

Banerjee, S., & van der Heijden, M. G. A. (2023). Soil microbiomes and one health. *Nature Reviews Microbiology*, **21**, 6–20.

Baudoin, E., Benizri, E., & Guckert, A. (2002). Impact of growth stage on the bacterial community structure along maize roots, as determined by metabolic and genetic fingerprinting. *Applied Soil Ecology*, **19**, 135–145.

Behie, S. W., & Bidochka, M. J. (2014). Nutrient transfer in plant-fungal symbioses. *Trends in Plant Science*, **19**, 734–740.

Bellon, M. R., & Hellin, J. (2011). Planting hybrids, keeping landraces: Agricultural modernization and tradition among small-scale maize farmers in Chiapas, Mexico. *World Development*, **39**, 1434–1443.

Bergelson, J., Mittelstrass, J., & Horton, M. W. (2019). Characterizing both bacteria and fungi

improves understanding of the *Arabidopsis* root microbiome. *Scientific Reports*, **9**, 24.

Brisson, V. L., Schmidt, J. E., Northen, T. R., Vogel, J. P., & Gaudin, A. C. M. (2019). Impacts of maize domestication and breeding on rhizosphere microbial community recruitment from a nutrient depleted agricultural soil. *Scientific Reports*, **9**, 15611.

Bulgarelli, D., Garrido-oter, R., Münch, P. C., Weiman, A., Dröge, J., Pan, Y., Mchardy, A. C., & Schulze-Lefert, P. (2015). Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host and Microbe*, **17**, 392–403.

Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F., Rauf, P., Huettel, B., Reinhardt, R., Schmelzer, E., Peplies, J., Gloeckner, F. O., Amann, R., Eickhorst, T., & Schulze-Lefert, P. (2012). Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature*, **488**, 91–95.

Burton, A. L., Johnson, J. M., Foerster, J. M., Hirsch, C. N., Meredith, C. R. B., Kaeppler, S. M., Brown, K. M., & Lynch, J. P. (2014). QTL mapping and phenotypic variation for root architectural traits in maize ( *Zea mays* L .). *Theoretical and Applied Genetics*, **127**, 2293–2311.

Çakmakçi, R., Kantar, F., & Sahin, F. (2001). Effect of $N_2$-fixing bacterial inoculations on yield of sugar beet and barley. *Journal of Plant Nutrition and Soil Science*, **164**, 527–531.

Carvalho, T. L. G., Balsemão-Pires, E., Saraiva, R. M., Ferreira, P. C. G., & Hemerly, A. S. (2014). Nitrogen signalling in plant interactions with associative and endophytic diazotrophic bacteria. *Journal of Experimental Botany*, **65**, 5631–5642.

Cassán, F., Perrig, D., Sgroy, V., Masciarelli, O., Penna, C., & Luna, V. (2009). *Azospirillum brasilense* Az39 and *Bradyrhizobium japonicum* E109, inoculated singly or in combination, promote seed germination and early seedling growth in corn (*Zea mays* L.) and soybean (*Glycine max* L.). *European Journal of Soil Biology*, **45**, 28–35.

Cheng, Y. T., Zhang, L., & He, S. Y. (2019). Plant-microbe interactions facing environmental challenge. *Cell Host and Microbe*, **26**, 183–192.

Cleveland, D. A., & Soleri, D. (2007). Extending Darwin's analogy: Bridging differences in concepts of selection between farmers, biologists, and plant breeders. *Economic Botany*, **61**, 121–136.

Dahmani, M. A., Desrut, A., Moumen, B., Verdon, J., Mermouri, L., Kacem, M., Coutos-Thévenot, P., Kaid-Harche, M., Bergès, T., & Vriet, C. (2020). Unearthing the plant growth-promoting traits of *Bacillus megaterium* RmBm31, an endophytic bacterium isolated from root nodules of *Retama monosperma*. *Frontiers in Plant Science*, **11**, 124.

Darwin, C. (1876). *The effects of cross and self fertilisation in the vegetable kingdom*. Ed. by **J Murray**. 1st ed. London: John Murray.

Deng, S., Caddell, D. F., Xu, G., Dahlen, L., Washington, L., Yang, J., & Coleman-Derr, D. (2021). Genome wide association study reveals plant loci controlling heritability of the rhizosphere microbiome. *ISME Journal*, **15**, 3181–3194.

Ding, J., Wang, X., Chander, S., Yan, J., & Li, J. (2008). QTL mapping of resistance to Fusarium ear rot using a RIL population in maize. *Molecular Breeding*, **22**, 395–403.

Doornbos, R. F., Van Loon, L. C., & Bakker, P. A. H. M. (2012). Impact of root exudates and plant defense signaling on bacterial communities in the rhizosphere. *Agronomy for Sustainable Development*, **32**, 227–243.

Duvick, D. N. (2001). Biotechnology in the 1930s: The development of hybrid maize. *Nature Reviews Genetics*, **2**, 69–74.

Edwards, J., Saran, U. B., Bonnette, J., MacQueen, A., Yin, J., Nguyen, T. uyen, Schmutz, J., Grimwood, J., Pennacchio, L. A., Daum, C., Glavina del Rio, T., Fritschi, F. B., Lowry, D. B., & Juenger, T. E. (2023). Genetic determinants of switchgrass-root-associated microbiota in field sites spanning its natural range. *Current Biology*, **33**, 1926-1938.

Edwards, J., Johnson, C., Santos-Medellín, C., Lurie, E., Podishetty, N. K., Bhatnagar, S., Eisen, J. A., & Sundaresan, V. (2015). Structure, variation, and assembly of the root-associated microbiomes of rice. *Proceedings of the National Academy of Sciences*, **112**, e911–e920.

Efthimiadou, A., Katsenios, N., Chanioti, S., Giannoglou, M., Djordjevic, N., & Katsaros, G. (2020). Effect of foliar and soil application of plant growth promoting bacteria on growth, physiology, yield and seed quality of maize under Mediterranean conditions. *Scientific Reports*, **10**, 21060.

El Zemrany, H., Czarnes, S., Hallett, P. D., Alamercery, S., Bally, R., & Jocteur Monrozier, L. (2007). Early changes in root characteristics of maize (*Zea mays*) following seed inoculation with the PGPR *Azospirillum lipoferum* CRT1. *Plant and Soil*, **291**, 109–118.

Etesami, H., & Alikhani, H. A. (2016). Co-inoculation with endophytic and rhizosphere bacteria allows reduced application rates of N-fertilizer for rice plant. *Rhizosphere*, **2**, 5–12.

Etesami, H., Jeong, B. R., & Glick, B. R. (2021). Contribution of arbuscular mycorrhizal fungi, phosphate–solubilizing bacteria, and silicon to P uptake by plant. *Frontiers in Plant Science*, **12**, 699618.

Fallik, E., & Okon, Y. (1996). The response of maize (*Zea mays*) to *Azospirillum* inoculation in various types of soils in the field. *World Journal of Microbiology and Biotechnology*, **12**, 511–515.

Favela, A., Bohn, M. O., Kent, A. D., & Kent, A. D. (2021). Maize germplasm chronosequence shows crop breeding history impacts recruitment of the rhizosphere microbiome. *The ISME Journal*, **15**, 2454–2464.

Feldman, L. (1994). The maize root. In *The Maize Handbook*. Springer New York, 29–37.

Finkel, O. M., Salas-González, I., Castrillo, G., Conway, J. M., Law, T. F., Teixeira, P. J. P. L., Wilson, E. D., Fitzpatrick, C. R., Jones, C. D., & Dangl, J. L. (2020). A single bacterial genus maintains root growth in a complex microbiome. *Nature*, **587**, 103–108.

Fusconi, A. (2014). Regulation of root morphogenesis in arbuscular mycorrhizae: What role do fungal exudates, phosphate, sugars and hormones play in lateral root formation? *Annals of Botany*, **113**, 19–33.

Fusconi, A., Lingua, G., Trotta, A., & Berta, G. (2005). Effects of arbuscular mycorrhizal colonization and phosphorus application on nuclear ploidy in *Allium porrum* plants. *Mycorrhiza*, **15**, 313–321.

Gutjahr, C., Casieri, L., & Paszkowski, U. (2009). *Glomus intraradices* induces changes in root system architecture of rice independently of common symbiosis signaling. *New Phytologist*, **182**, 829–837.

Gutjahr, C., Sawers, R. J. H., Marti, G., Andrés-Hernández, L., Yang, S. Y., Casieri, L., Angliker, H., Oakeley, E. J., Wolfender, J. L., Abreu-Goodger, C., & Paszkowski, U. (2015). Transcriptome diversity among rice root types during asymbiosis and interaction with arbuscular mycorrhizal fungi. *Proceedings of the National Academy of Sciences*, **112**, 6754–6759.

Gyaneshwar, P., Kumar, G. N., Parekh, L. J., & Poole, P. S. (2002). Role of soil microorganisms in improving P nutrition of plants. *Plant and Soil*, **245**, 83–93.

Hake, S., & Ross-Ibarra, J. (2015). Genetic , evolutionary and plant breeding insights from the domestication of maize. *ELife*, **4**, e05861.

Haney, C. H., Samuel, B. S., Bush, J., & Ausubel, F. M. (2015). Associations with rhizosphere bacteria can confer an adaptive advantage to plants. *Nature Plants*, **1**, 1–9.

Harjes, C. E., Rocheford, T. R., Bai, L., Brutnell, T. P., Kandianis, C. B., Sowinski, S. G., Stapleton, A. E., Vallabhaneni, R., Williams, M., Wurtzel, E. T., Yan, J., & Buckler, E. S. (2008). Natural genetic variation in *lycopene epsilon cyclase* tapped for maize biofortification. *Science*, **319**, 330–333.

Hestrin, R., Hammer, E. C., Mueller, C. W., & Lehmann, J. (2019). Synergies between mycorrhizal fungi and soil microbial communities increase plant nitrogen acquisition. *Communications Biology*, **2**, 233.

Hetz, W., Hochholdinger, F., Schwall, M., & Feix, G. (1996). Isolation and characterization of *rtcs*, a maize mutant deficient in the formation of nodal roots. *Plant Journal*, **10**, 845–857.

Hiruma, K., Gerlach, N., Sacristán, S., Nakano, R. T., Hacquard, S., Kracher, B., Neumann, U., Ramírez, D., Bucher, M., O'Connell, R. J., & Schulze-Lefert, P. (2016). Root endophyte *Colletotrichum tofieldiae* confers plant fitness benefits that are phosphate status dependent. *Cell*, **165**, 464–474.

Hochholdinger, F. (2009). The maize root system: morphology, anatomy, and genetics. In *Handbook of Maize: Its Biology*. Springer, New York, 145–160.

Hochholdinger, F. (2016). Untapping root system architecture for crop improvement. *Journal of Experimental Botany*, **67**, 4431–4433.

Hochholdinger, F., & Feix, G. (1998). Early post-embryonic root formation is specifically affected in the maize mutant *lrt1*. *The Plant Journal*, **16**, 247–255.

Hochholdinger, F., Wen, T. J., Zimmermann, R., Chimot-Marolle, P., Da Costa E Silva, O., Bruce, W., Lamkey, K. R., Wienand, U., & Schnable, P. S. (2008). The maize (*Zea mays* L.) *roothairless3* gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. *The Plant Journal*, **54**, 888–898.

Hochholdinger, F., Woll, K., Sauer, M., & Dembinsky, D. (2004). Genetic dissection of root

formation in maize (*Zea mays*) reveals root-type specific developmental programmes. *Annals of Botany*, **93**, 359–368.

Hochholdinger, F., Yu, P., & Marcon, C. (2018). Genetic control of root system development in maize. *Trends in Plant Science*, **23**, 79–88.

Hodge, A., Berta, G., Doussan, C., Merchan, F., & Crespi, M. (2009). Plant root growth, architecture and function. *Plant and Soil*, **321**, 153–187.

Horton, M. W., Bodenhausen, N., Beilsmith, K., Meng, D., Muegge, B. D., Subramanian, S., Vetter, M. M., Vilhjálmsson, B. J., Nordborg, M., Gordon, J. I., & Bergelson, J. (2014). Genome-wide association study of *Arabidopsis thaliana* leaf microbial community. *Nature Communications*, **5**, 5320.

Jacoud, C., Faure, D., Wadoux, P., & Bally, R. (1998). Development of a strain-specific probe to follow inoculated *Azospirillum lipoferum* CRT1 under field conditions and enhancement of maize root development by inoculation. *FEMS Microbiology Ecology*, **27**, 43–51.

Jochum, M. D., McWilliams, K. L., Borrego, E. J., Kolomiets, M. V., Niu, G., Pierson, E. A., & Jo, Y. K. (2019). Bioprospecting plant growth-promoting rhizobacteria that mitigate drought stress in grasses. *Frontiers in Microbiology*, **10**, 2106.

Klug, W. S., Cummings, M. R., Spencer, C. A., & Palladino, M. A. (2016). Quantitative genetics and multifactorial traits. In *Concepts of genetics*. Pearson education, 638–653.

Krishnamoorthy, R., Kim, K., Subramanian, P., Senthilkumar, M., Anandham, R., & Sa, T. (2016). Arbuscular mycorrhizal fungi and associated bacteria isolated from salt-affected soil enhances the tolerance of maize to salinity in coastal reclamation soil. *Agriculture, Ecosystems & Environment*, **231**, 233-239.

Lanfranco, L., Fiorilli, V., & Gutjahr, C. (2018). Partner communication and role of nutrients in the arbuscular mycorrhizal symbiosis. *New Phytologist*, **220**, 1031–1046.

Lang, M., Zou, W., Chen, X., Zou, C., Zhang, W., Deng, Y., Zhu, F., Yu, P., & Chen, X. (2021). Soil microbial composition and *phoD* gene abundance are sensitive to phosphorus level in a long-term wheat-maize crop system. *Frontiers in Microbiology*, **11**, 605955.

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 1–13.

Lewin, S., Francioli, D., Ulrich, A., & Kolb, S. (2021). Crop host signatures reflected by co-association patterns of keystone Bacteria in the rhizosphere microbiota. *Environmental Microbiome*, **16**, 1–18.

Li, C., Cao, P., Du, C., Zhang, X., Bing, H., Li, L., ... & Wang, X. (2021). *Massilia rhizosphaerae* sp. nov., a rice-associated rhizobacterium with antibacterial activity against *Ralstonia solanacearum*. International journal of systematic and evolutionary microbiology, **71**, 005009.

Li, L., Hey, S., Liu, S., Liu, Q., McNinch, C., Hu, H. C., Wen, T. J., Marcon, C., Paschold, A., Bruce, W., Schnable, P. S., & Hochholdinger, F. (2016). Characterization of maize *roothairless6* which encodes a D-type cellulose synthase and controls the switch from bulge formation to tip growth. *Scientific Reports*, **6**, 34395.

López-Bucio, J., Millán-Godínez, M., Méndez-Bravo, A., Morquecho-Contreras, A., Ramírez-Chávez, E., Molina-Torres, J., Pérez-Torres, A., Higuchi, M., Kakimoto, T., & Herrera-Estrella, L. (2007). Cytokinin receptors are involved in alkamide regulation of root and shoot development in *Arabidopsis*. *Plant Physiology*, **145**, 1703–1713.

Lu, T., Ke, M., Lavoie, M., Jin, Y., Fan, X., Zhang, Z., Fu, Z., Sun, L., Gillings, M., Peñuelas, J., Qian, H., & Zhu, Y. G. (2018). Rhizosphere microorganisms can influence the timing of plant flowering. *Microbiome*, **6**, 1–12.

Lund, M., Agerbo, J., Jazmín, R., Sawers, R., Gilbert, M. T. P., & Barnes, C. J. (2022). Rhizosphere bacterial communities differ among traditional maize landraces. *Environmental DNA*, *4*(6), 1241–1249.

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

Lynch, J. P. (1995). Root architecture and plant productivity. *Plant Physiology*, **109**, 7–13.

Lynch, J. P. (2013). Steep, cheap and deep: An ideotype to optimize water and N acquisition by maize root systems. *Annals of Botany*, **112**, 347–357.

Marschner, P. (2012). Rhizosphere biology. In *Marschner's Mineral Nutrition of Higher Plants (Third Edition)*. Academic Press, 369–388.

Matsuoka, Y., Vigouroux, Y., Goodman, M. M., G, J. S., Buckler, E., & Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences*, **99**, 6080–6084.

McCully, M. E., & Canny, M. J. (1988). Pathways and processes of water and nutrient movement in roots. *Plant and Soil*, **111**, 159–170.

Meier, M. A., Xu, G., Lopez-Guerrero, M. G., Li, G., Smith, C., Sigmon, B., Herr, J. R., Alfano, J. R., Ge, Y., Schnable, J. C., & Yang, J. (2022). Association analyses of host genetics, root-colonizing microbes, and plant phenotypes under different nitrogen conditions in maize. *eLife*, **11**, e75790.

Meister, R., Rajani, M. S., Ruzicka, D., & Schachtman, D. P. (2014). Challenges of modifying root traits in crops for agriculture. *Trends in Plant Science*, **19**, 779–788.

Mercer, K., Martínez-Vásquez, Á., & Perales, H. R. (2008). Asymmetrical local adaptation of maize landraces along an altitudinal gradient. *Evolutionary Applications*, **1**, 489–500.

Mirza, M., Ahmad, W., Latif, F., Haurat, J., Bally, R., Normand, P., & Malik, K. (2001). Isolation, partial characterization, and the effect of plant growth-promoting bacteria (PGPB) on micro-propagated sugarcane *in vitro*. *Plant and Soil*, **237**, 47–54.

Nannas, N. J., & Dawe, R. Kelly. (2015). Genetic and genomic toolbox of *Zea mays*. *Genetics*, **199**, 655–669.

Naylor, D., Degraaf, S., Purdom, E., & Coleman-Derr, D. (2017). Drought and host selection influence bacterial community dynamics in the grass root microbiome. *ISME Journal*, **11**, 2691–2704.

Nestler, J., Liu, S., Wen, T. J., Paschold, A., Marcon, C., Tang, H. M., Li, D., Li, L., Meeley, R. B., Sakai, H., Bruce, W., Schnable, P. S., & Hochholdinger, F. (2014). *Roothairless5*, which functions in maize (*Zea mays* L.) root hair initiation and elongation encodes a monocot-specific NADPH oxidase. *Plant Journal*, **79**, 729–740.

Neuffer, M. G., Coe, E. ., & Wessler, S. . (1997). *Mutants of Maize*. New York: Cold Spring Harbor Laboratory Press.

Oldroyd, G. E. D., & Leyser, O. (2020). A plant's diet, surviving in a variable nutrient environment. *Science*, **368**, eaba0196.

Oyserman, B. O., Flores, S. S., Grif, T., Pan, X., Wijk, E. Van Der, Pronk, L., Lokhorst, W.,

Nur, A., Paulson, J. N., Movassagh, M., Stopnisek, N., Kupczok, A., Cordovez, V., Carrión, V. J., Ligterink, W., Snoek, B. L., Medema, M. H., & Raaijmakers, J. M. (2022). Disentangling the genetic basis of rhizosphere microbiome assembly in tomato. *Nature Communications*, **13**, 3228.

Panke-buisse, K., Poole, A. C., Goodrich, J. K., Ley, R. E., & Kao-kniffin, J. (2015). Selection on soil microbiomes reveals reproducible impacts on plant function. *ISME Journal*, **9**, 980–989.

Park, W. J., Hochholdinger, F., & Gierl, A. (2004). Release of the benzoxazinoids defense molecules during lateral- and crown root emergence in *Zea mays*. *Journal of Plant Physiology*, **161**, 981–985.

Paszkowski, U., & Boller, T. (2002). The growth defect of *lrt1*, a maize mutant lacking lateral roots, can be complemented by symbiotic fungi or high phosphate nutrition. *Planta*, **214**, 584–590.

Peiffer, J. A., Spor, A., Koren, O., Jin, Z., Tringe, S. G., Dangl, J. L., Buckler, E. S., & Ley, R. E. (2013). Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proceedings of the National Academy of Sciences*, **110**, 6548–6553.

Pérez-Jaramillo, J. E., Mendes, R., & Raaijmakers, J. M. (2016). Impact of plant domestication on rhizosphere microbiome assembly and functions. *Plant Molecular Biology*, **90**, 635–644.

Pham, V. T. K., Rediers, H., Ghequire, M. G. K., Nguyen, H. H., De Mot, R., Vanderleyden, J., & Spaepen, S. (2017). The plant growth-promoting effect of the nitrogen-fixing endophyte *Pseudomonas stutzeri* A15. *Archives of Microbiology*, *199*, 513–517.

Poole, P., Ramachandran, V., & Terpolilli, J. (2018). Rhizobia: From saprophytes to endosymbionts. *Nature Reviews Microbiology*, **16**, 291–303.

Qian, Y., Wang, X., Liu, Y., Wang, X., & Mao, T. (2023). HY5 inhibits lateral root initiation in *Arabidopsis* through negative regulation of the microtubule-stabilizing protein TPXL5. *Plant Cell*, **35**, 1092–1109.

Rafalski, A., & Morgante, M. (2004). Corn and humans: Recombination and linkage disequilibrium in two genomes of similar size. *Trends in Genetics*, **20**, 103–111.

Romero Navarro, J. A., Willcox, M., Burgueño, J., Romay, C., Swarts, K., Trachsel, S.,

Preciado, E., Terron, A., Delgado, H. V., Vidal, V., Ortega, A., Banda, A. E., Montiel, N. O. G., Ortiz-Monasterio, I., Vicente, F. S., Espinoza, A. G., Atlin, G., Wenzl, P., Hearne, S., & Buckler, E. S. (2017). A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nature Genetics*, **49**, 476–480.

Rüger, L., Feng, K., Dumack, K., Freudenthal, J., Chen, Y., Sun, R., Wilson, M., Yu, P., Sun, B., Deng, Y., Hochholdinger, F., Vetterlein, D., & Bonkowski, M. (2021). Assembly patterns of the rhizosphere microbiome along the longitudinal root axis of maize (*Zea mays* L.). *Frontiers in Microbiology*, **12**, 614501.

Salas-González, I., Reyt, G., Flis, P., Custódio, V., Gopaulchan, D., Bakhoum, N., Dew, T. P., Suresh, K., Franke, R. B., Dangl, J. L., Salt, D. E., & Castrillo, G. (2021). Coordination between microbiota and root endodermis supports plant mineral nutrient homeostasis. *Science*, **371**, eabd0695.

Schlemper, T. R., Leite, M. F. A., Lucheta, A. R., Shimels, M., Bouwmeester, H. J., van Veen, J. A., & Kuramae, E. E. (2017). Rhizobacterial community structure differences among sorghum cultivars in different growth stages and soils. *FEMS Microbiology Ecology*, **93**, 1–11.

Schreiter, S., Ding, G.-C., Heuer, H., Neumann, G., Sandmann, M., Grosch, R., Siegfried, K., & Smalla, K. (2014). Effect of the soil type on the microbiome in the rhizosphere of field-grown lettuce. *Frontiers in Microbiology*, **5**, 144.

Shull, G. H. (1908). The Composition of a field of maize. *Journal of Heredity*, **1**, 296–301.

Singh, A., Kumar, M., Chakdar, H., Pandiyan, K., Kumar, S. C., Zeyad, M. T., Singh, B. N., Ravikiran, K. T., Mahto, A., Srivastava, A. K., & Saxena, A. K. (2022). Influence of host genotype in establishing root associated microbiome of *indica* rice cultivars for plant growth promotion. *Frontiers in Microbiology*, **13**, 1033158.

Sohrabi, R., Paasch, B. C., Liber, J. A., & He, S. Y. (2023). Phyllosphere microbiome. *Annual Review of Plant Biology*, **74**, 539–568.

Strable, J., & Scanlon, M. J. (2009). Maize (*Zea mays*): A model organism for basic and applied research in plant biology. *Cold Spring Harbor Protocols*, **2**, 33–41.

Szoboszlay, M., Lambers, J., Chappell, J., Kupper, J. V, Moe, L. A., & Mcnear, D. H. (2015). Comparison of root system architecture and rhizosphere microbial communities of Balsas teosinte and domesticated corn cultivars. *Soil Biology & Biochemistry Journal*, **80**, 34–44.

Taramino, G., Sauer, M., Stauffer, J. L., Multani, D., Niu, X., Sakai, H., & Hochholdinger, F. (2007). The maize (*Zea mays* L.) *RTCS* gene encodes a LOB domain protein that is a key regulator of embryonic seminal and post-embryonic shoot-borne root initiation. *Plant Journal*, **50**, 649–659.

Thiergart, T., Durán, P., Ellis, T., Garrido-Oter, R., Kemen, E., Roux, F., Alonso-Blanco, C., Ågren, J., Schulze-Lefert, P., & Hacquard, S. (2020). Root microbiota assembly and adaptive differentiation among European *Arabidopsis* populations. *Nature Ecology and Evolution*, **4**, 122–131.

Thornsberry, J. M., Goodman, M. M., Doebley, J., Kresovich, S., Nielsen, D., & Buckler, E. S. (2001). *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics*, **28**, 286–289.

Vance, C. P. (2001). Symbiotic nitrogen fixation and phosphorus acquisition. Plant nutrition in a world of declining renewable resources. *Plant Physiology*, **127**, 390–397.

Vance, C. P., Uhde-Stone, C., & Allan, D. L. (2003). Phosphorus acquisition and use: Critical adaptations by plants for securing a nonrenewable resource. *New Phytologist*, **157**, 423–447.

Von Behrens, I., Komatsu, M., Zhang, Y., Berendzen, K. W., Niu, X., Sakai, H., Taramino, G., & Hochholdinger, F. (2011). *Rootless with undetectable meristem 1* encodes a monocot-specific AUX/IAA protein that controls embryonic seminal and post-embryonic lateral root initiation in maize. *Plant Journal*, **66**, 341–353.

Wagner, M. R., Roberts, J. H., Balint-kurti, P., & Holland, J. B. (2020). Heterosis of leaf and rhizosphere microbiomes in field-grown maize. *New Phytologist*, **228**, 1055–1069.

Walker, V., Couillerot, O., von Felten, A., Bellvert, F., Jansa, J., Maurhofer, M., Bally, R., Moënne-Loccoz, Y., & Comte, G. (2012). Variation of secondary metabolite levels in maize seedling roots induced by inoculation with *Azospirillum*, *Pseudomonas* and *Glomus* consortium under field conditions. *Plant and Soil*, **356**, 151–163.

Wallace, J. G., Larsson, S. J., & Buckler, E. S. (2014). Entering the second century of maize quantitative genetics. *Heredity*, **112**, 30–38.

Wallace, Jason G., Kremling, K. A., Kovar, L. L., & Buckler, E. S. (2018). Quantitative genetics of the maize leaf microbiome. *Phytobiomes*, **2**, 208–224.

Walters, W. A., Jin, Z., Youngblut, N., Wallace, J. G., Sutter, J., & Zhang, W. (2018). Large-scale replicated field study of maize rhizosphere identifies heritable microbes. *Proceedings of the National Academy of Sciences*, **115**, 7368–7373.

Wang, J. L., Liu, K. Lou, Zhao, X. Q., Gao, G. F., Wu, Y. H., & Shen, R. F. (2022). Microbial keystone taxa drive crop productivity through shifting aboveground-belowground mineral element flows. *Science of the Total Environment*, **811**, 152342.

Wang, W., Shi, J., Xie, Q., Jiang, Y., Yu, N., & Wang, E. (2017). Nutrient exchange and regulation in arbuscular mycorrhizal symbiosis. *Molecular Plant*, **10**, 1147–1158.

Wang, X., Pan, Q., Chen, F., Yan, X., & Liao, H. (2011). Effects of co-inoculation with arbuscular mycorrhizal fungi and rhizobia on soybean growth as related to root architecture and availability of N and P. *Mycorrhiza*, **21**, 173–181.

Wilson, L. M., Whitt, S. R., Ibáñez, A. M., Rocheford, T. R., Goodman, M. M., & Buckler, E. S. (2004). Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell*, **16**, 2719–2733.

Woll, K., Borsuk, L. A., Stransky, H., Nettleton, D., Schnable, P. S., & Hochholdinger, F. (2005). Isolation, characterization, and pericycle-specific transcriptome analyses of the novel maize lateral and seminal root initiation mutant *rum1*. *Plant Physiology*, **139**, 1255–1267.

Yang, N., Wang, Y., Liu, X., Jin, M., Vallebueno-Estrada, M., Calfee, E., Chen, L., Dilkes, B. P., Gui, S., Fan, X., Harper, T. K., Kennett, D. J., Li, W., Lu, Y., Luo, J., Mambakkam, S., Menon, M., Snodgrass, S., Veller, C., … Ross-Ibarra, J. (2023). Two teosintes made modern maize. *Science*, **382**, eadg8940.

Yu, P., He, X., Baer, M., Beirinckx, S., Tian, T., Moya, Y. A. T., Zhang, X., Deichmann, M., Frey, F. P., Bresgen, V., Li, C., Razavi, B. S., Schaaf, G., von Wirén, N., Su, Z., Bucher, M., Tsuda, K., Goormachtig, S., Chen, X., & Hochholdinger, F. (2021). Plant flavones enrich rhizosphere *Oxalobacteraceae* to improve maize performance under nitrogen deprivation. *Nature Plants*, **7**, 481–499.

Yu, P., & Hochholdinger, F. (2018). The role of host genetic signatures on root–microbe interactions in the rhizosphere and endosphere. *Frontiers in Plant Science*, **9**, 1–5.

Yue, H., Yue, W., Jiao, S., Kim, H., Lee, Y.-H., Wei, G., Song, W., & Shu, D. (2023). Plant domestication shapes rhizosphere microbiome assembly and metabolic functions.

*Microbiome*, **11**, 1–19.

Zamioudis, C., Mastranesti, P., Dhonukshe, P., Blilou, I., & Pieterse, C. M. J. (2013). Unraveling root developmental programs initiated by beneficial *Pseudomonas* spp. bacteria. *Plant Physiology*, **162**, 304–318.

Zhang, J., Liu, Y. X., Zhang, N., Hu, B., Jin, T., Xu, H., Qin, Y., Yan, P., Zhang, X., Guo, X., Hui, J., Cao, S., Wang, X., Wang, C., Wang, H., Qu, B., Fan, G., Yuan, L., Garrido-Oter, R., … Bai, Y. (2019). *NRT1.1B* is associated with root microbiota composition and nitrogen use in field-grown rice. *Nature Biotechnology*, **37**, 676–684.

Zhu, J., Kaeppler, S. M., & Lynch, J. P. (2005). Topsoil foraging and phosphorus acquisition efficiency in maize (*Zea mays*). *Functional Plant Biology*, **32**, 749–762.

# 6 Appendix

## 6.1 Supporting figures for Chapter 2
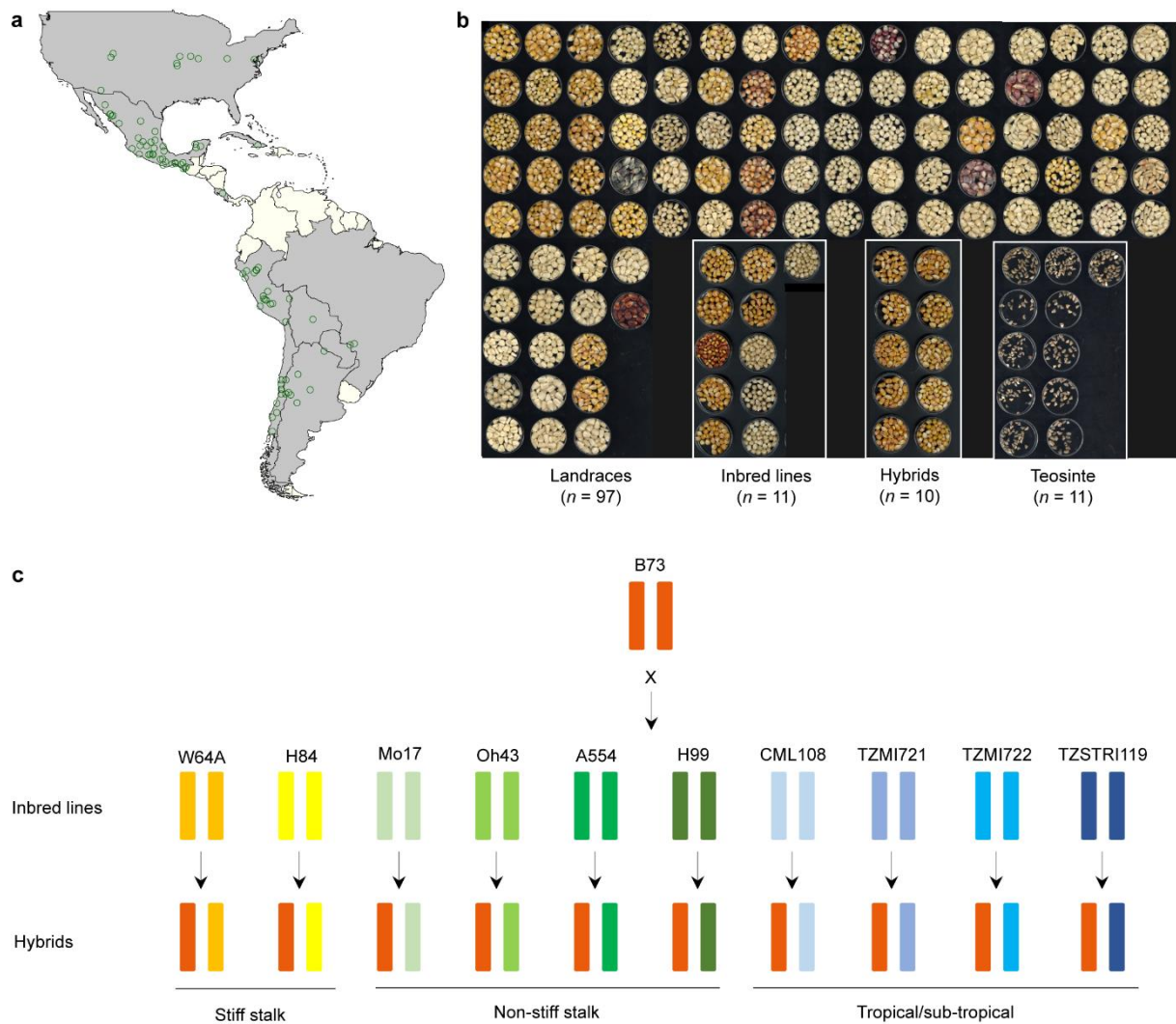


**Supplementary Figure 1. Germplasm used in this study. a**, Geographical coordinates of the 97 collected maize landraces are highlighted by green circles. **b**, Seed phenotypes of 97 landraces, 11 teosinte accessions, 11 inbred lines and 10 hybrids. **c**, Design of crosses of 10 inbred lines with the recurrent mother plant B73 to generate hybrids. The parental lines can be divided into three sub-populations (stiff stalk, non-stiff stalk, tropical/sub-tropical).
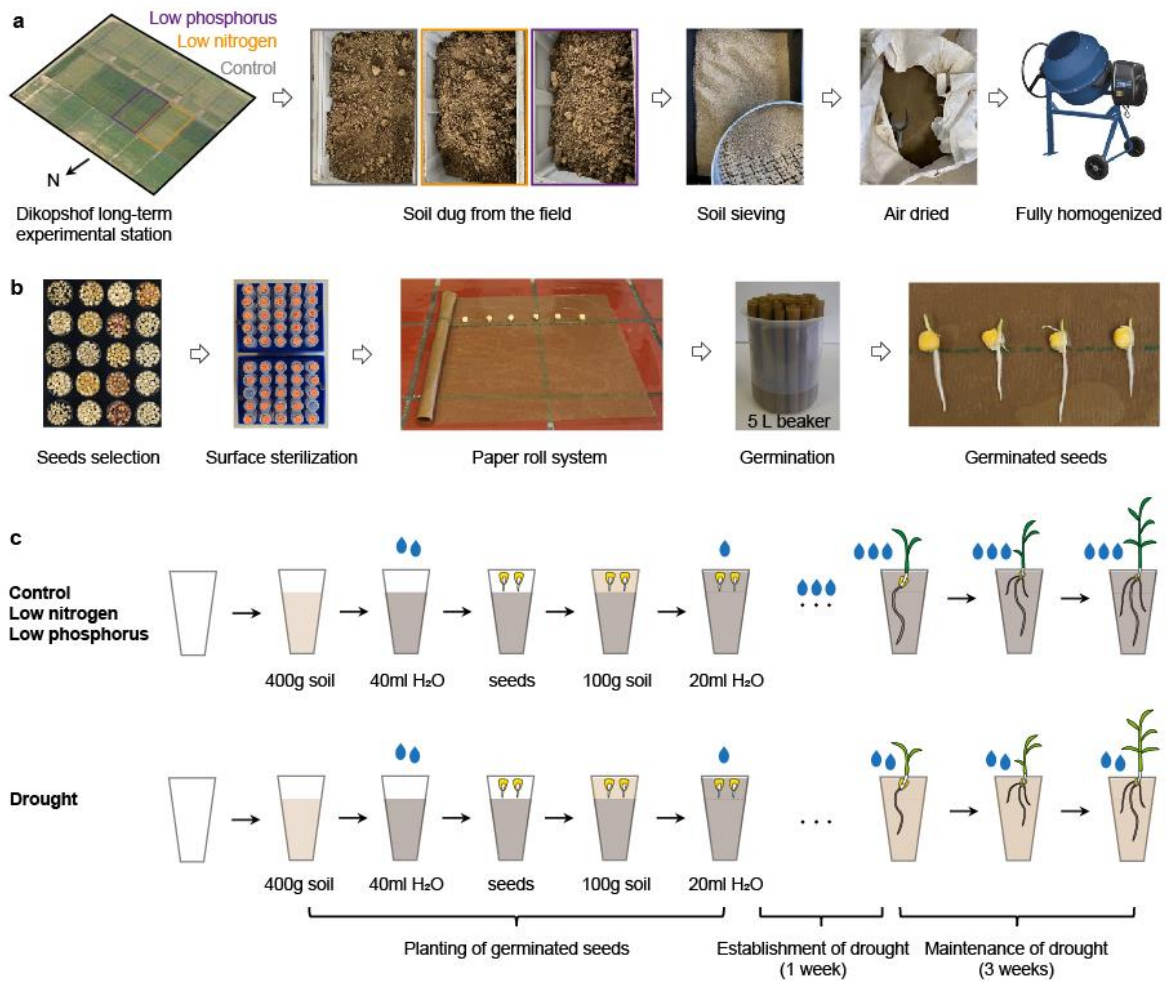
**Supplementary Figure 2. Soil pot experiments. a**, Natural soils with different nutrient levels were dug from Dikopshof long-term experimental station (see Methods). The soils were then sieved with a 4 mm sieve, air-dried and homogenized. **b**, Seed sterilization and germination. Only seeds with similar phenotypes were sterilized and germinated in a paper-roll system until the primary root had a length of 1-2 cm. **c**, Transfer to the soil-filled pots. The germinated seeds were transferred to the soil-filled pots with sterilized water. Drought was established by suspending watering for one week and maintenance of low soil water moisture at mild stress level described in Supplementary Figure 33. The plants were grown in the soil-filled pots for 4 weeks. And the soil for the drought treatment was used from the control soil.

**a**     **Sample harvest**

1st whorl shoot-borne crown root

Seminal root
Primary root

Pooled → Leaf

Extraction → Rhizosphere

Washed → Root

**b**     **Genomic DNA extraction**

Leaf     Rhizosphere     Root

Grinding
↓
Isolation
↓
Washing
↓
Elution

Genomic DNA

**c**     **Next-generation sequencing**

Primer mix    Library and sequencing    SNP

**Genotyping by target sequencing**

V4

**Bacterial 16S rRNA gene**

18S    5.8S    28S

ITS1 ITS2

**Fungal ITS gene sequencing**

**Supplementary Figure 3. Sample harvest and sequencing strategy. a**, Harvesting of shoot, rhizosphere and root samples. Fresh leaf tip samples were dissected and pooled from several maize plants grown under different treatment conditions. Rhizosphere sample was extracted from the 1st shoot-borne crown root with tightly attached soil. The same whorl of shoot-borne roots was separately dissected and immediately washed, dried with clean tissue, followed by flash freezing in liquid nitrogen. **b**, Extraction of genomic DNA from leaf, rhizosphere and root samples. **c**, Next-generation sequencing. The genomic DNA from pooled leaf samples was genotyped by genotyping by target sequencing (see Methods). The genomic DNA from rhizosphere and root samples was sequenced by 16 rRNA (V4 region) gene and ITS1 gene sequencing for bacterial and fungal community analyses.
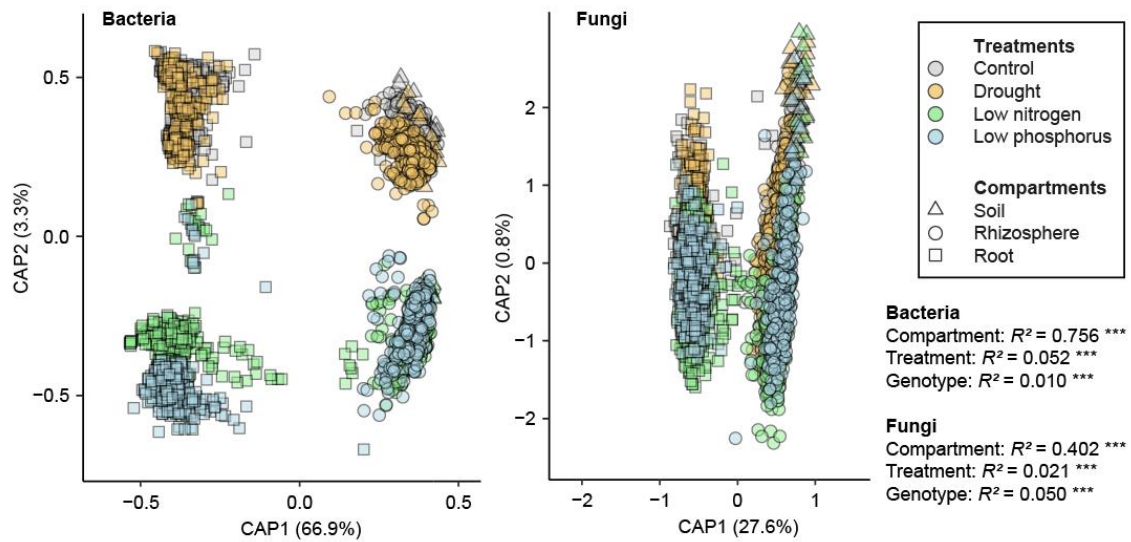
**Supplementary Figure 4. Overall diversity of the microbiome under abiotic stresses across different compartments.** Constrained analysis of principal coordinate (CAP) ordination using Bray–Curtis dissimilarity with permutational analysis of variance (PERMANOVA) was applied to visualize significant microbiome differences across three compartments, four treatments and 129 genotypes. Compartments are shape coded. Only ASVs with reads >10 in ≥6 samples were included in the dataset.
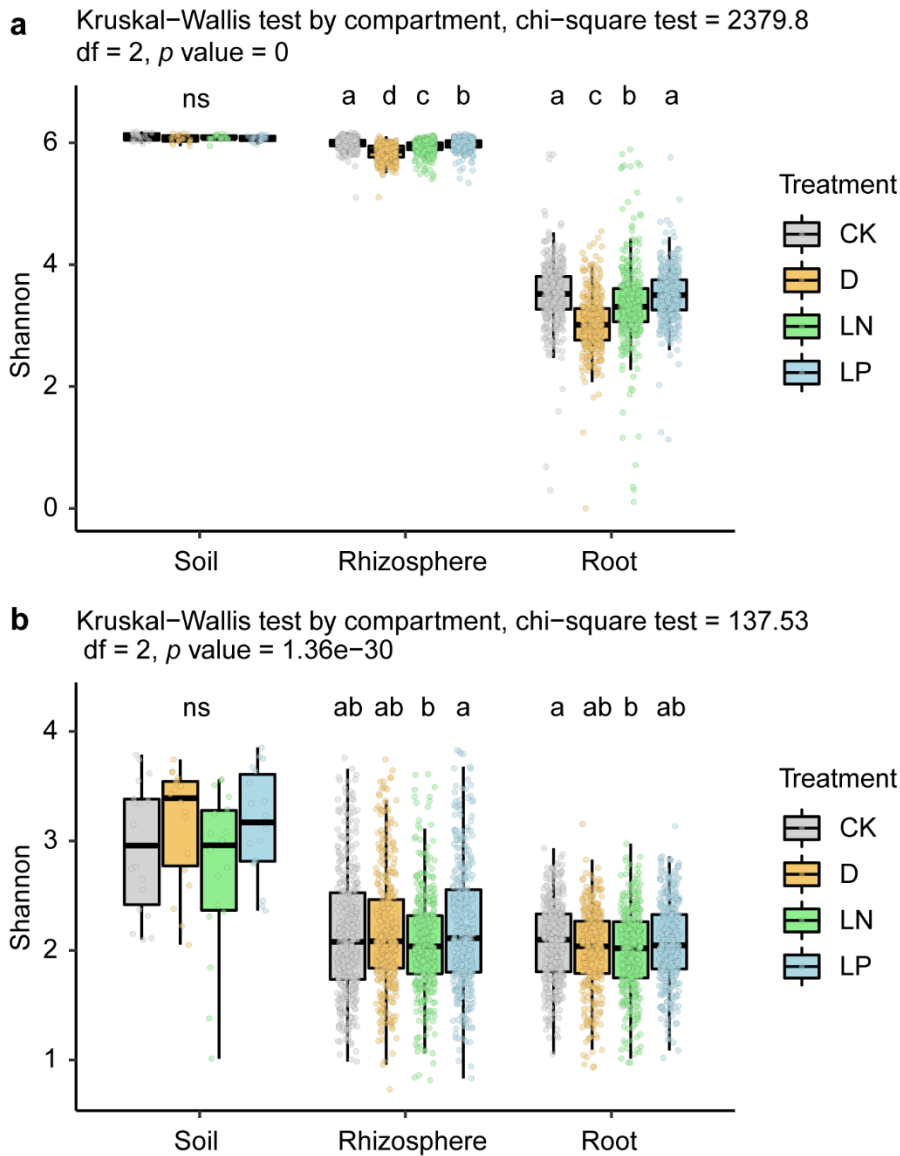
**Supplementary Figure 5. Microbial diversity across treatments in different compartments.** Changes in bacterial (**a**) and fungal (**b**) α-diversity estimated by Shannon's diversity index. Compartment significances were calculated using a two-tailed Kruskal-Wallis test with post-hoc Wilcoxon (Benjamini-Hochberg adjusted). Different letters indicate significance between different treatments (*p* <0.05). CK, control; D, drought; LN, low nitrogen; LP, low phosphorus. ns, not significant. Box plots include the median (the horizontal line inside the boxes), 95% confidence intervals (the black vertical lines) and the distribution of frequency (width of the differently colored zones). For **a** and **b**, *n* = 18 for soil under each treatment, *n* = 387 for rhizosphere and root under each treatment. All samples are biologically independent.
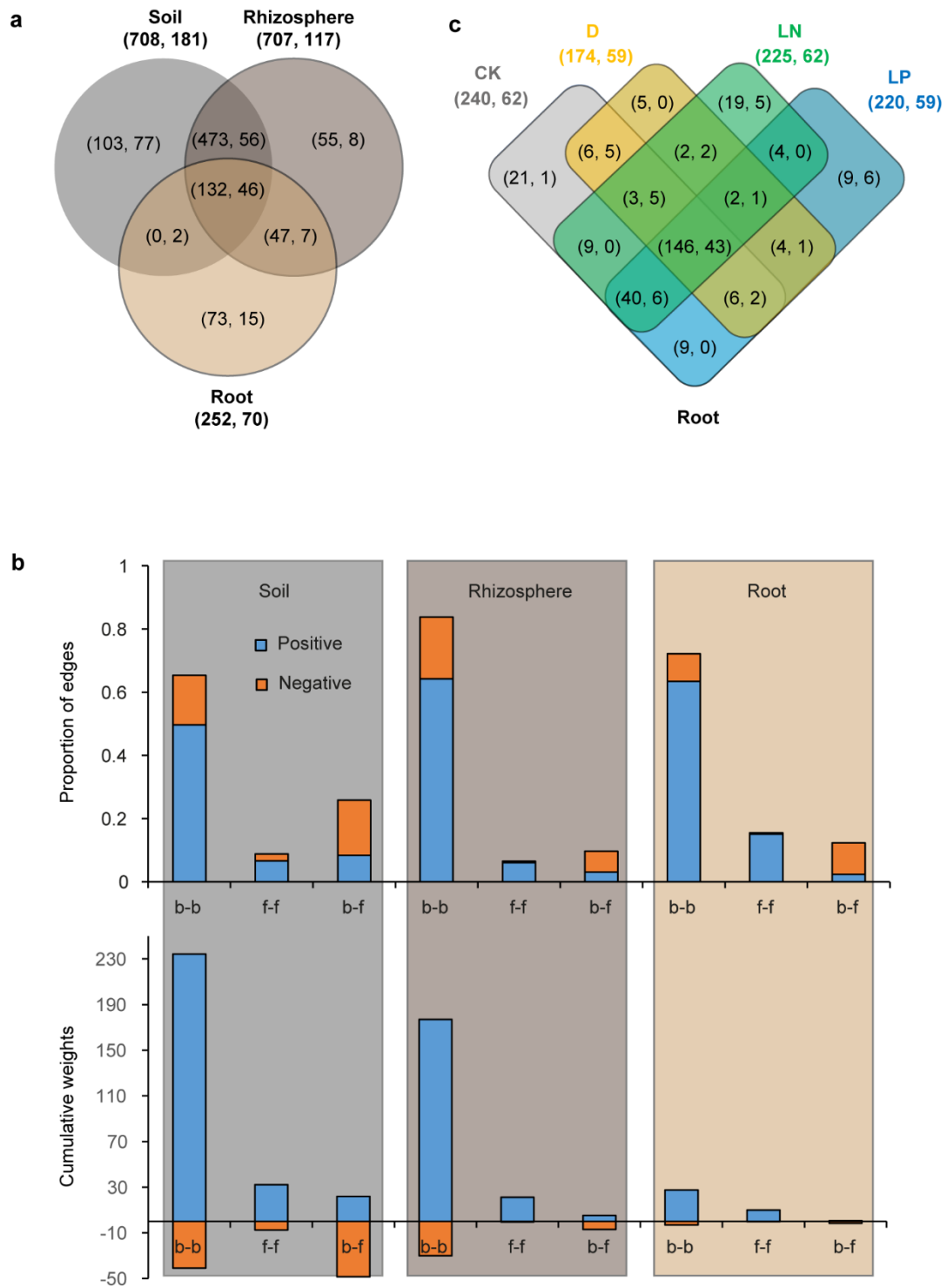
**Supplementary Figure 6. Intra- and inter-kingdom network analysis. a**, 3-way Venn diagram indicates the overlapping patterns of bacterial and fungal interacting ASVs across the compartments. The first and second number in the brackets indicate the bacterial and fungal ASVs respectively. **b**, Bar plot indicates the network associations measured by proportion of edges and cumulative weights within and between different microbial kingdoms. b-b, bacterial-bacterial interactions; f-f, fungal-fungal interactions; b-f, bacterial-fungal interactions. **c**, 4-way Venn diagram indicating the overlapping patterns of bacterial and fungal interacting ASVs across treatments in the root. The first and second number in the bracket indicate the bacterial and fungal ASVs respectively.
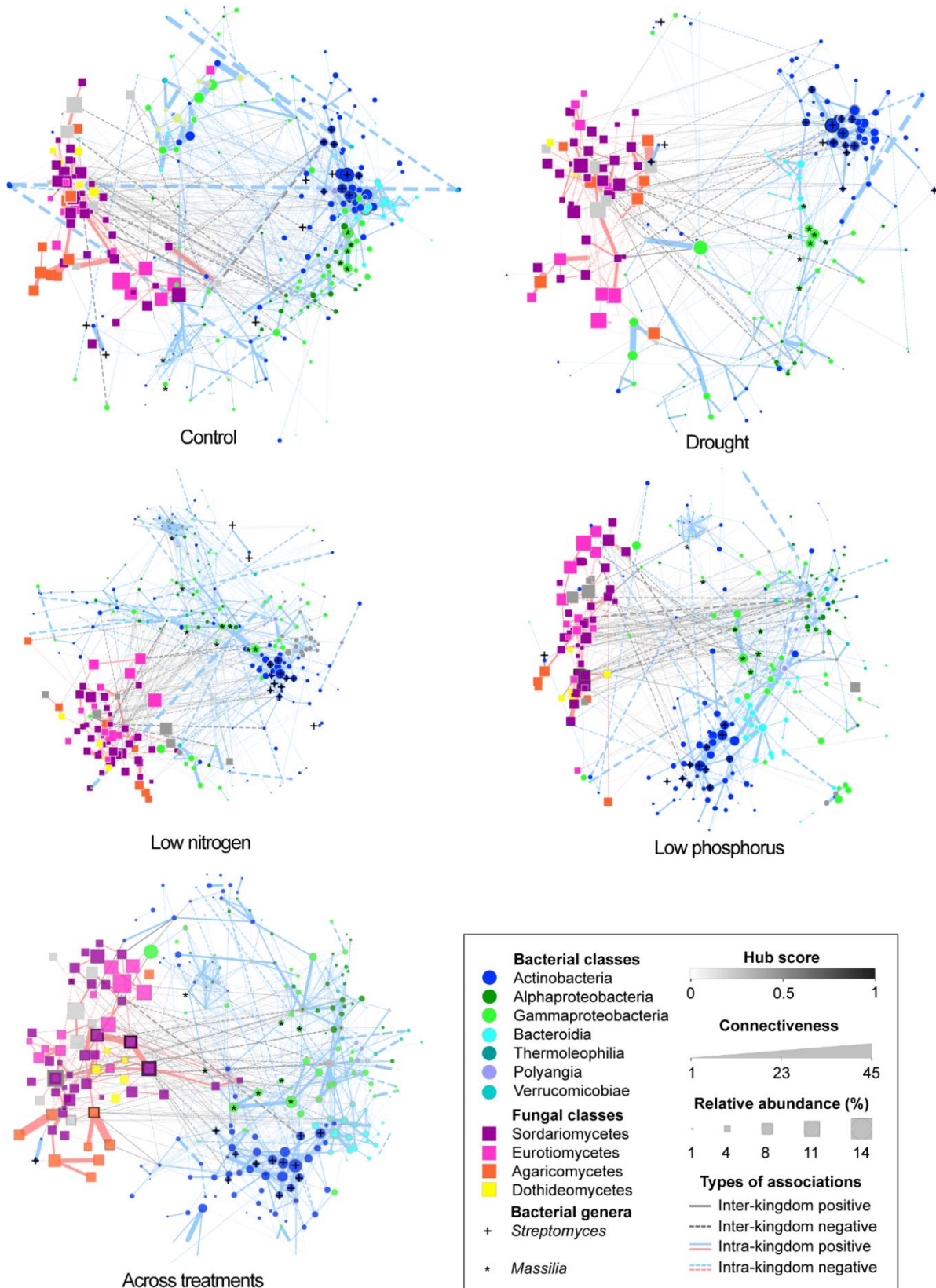
**Supplementary Figure 7. Intra- and inter-kingdom root network associations under different stress treatments.** A sparse inverse covariance estimation for ecological association inference (SPIEC-EASI) based network association analysis of root-associated microbial ASVs (RA >0.05%; Prevalence >10%) across different treatments. Each node corresponds to an ASV and was color coded according to the family taxa. The size of nodes reflects the relative abundance in the root. Edges between nodes correspond to either positive (solid lines) or negative (dotted lines) associations and color codes of edges indicate inter- or intra-kingdom associations between ASVs. Connectiveness between nodes corresponds to the degree of associations. Keystone ASVs in microbial networks were reflected by hub scores corresponding to the thickness of the nodes.
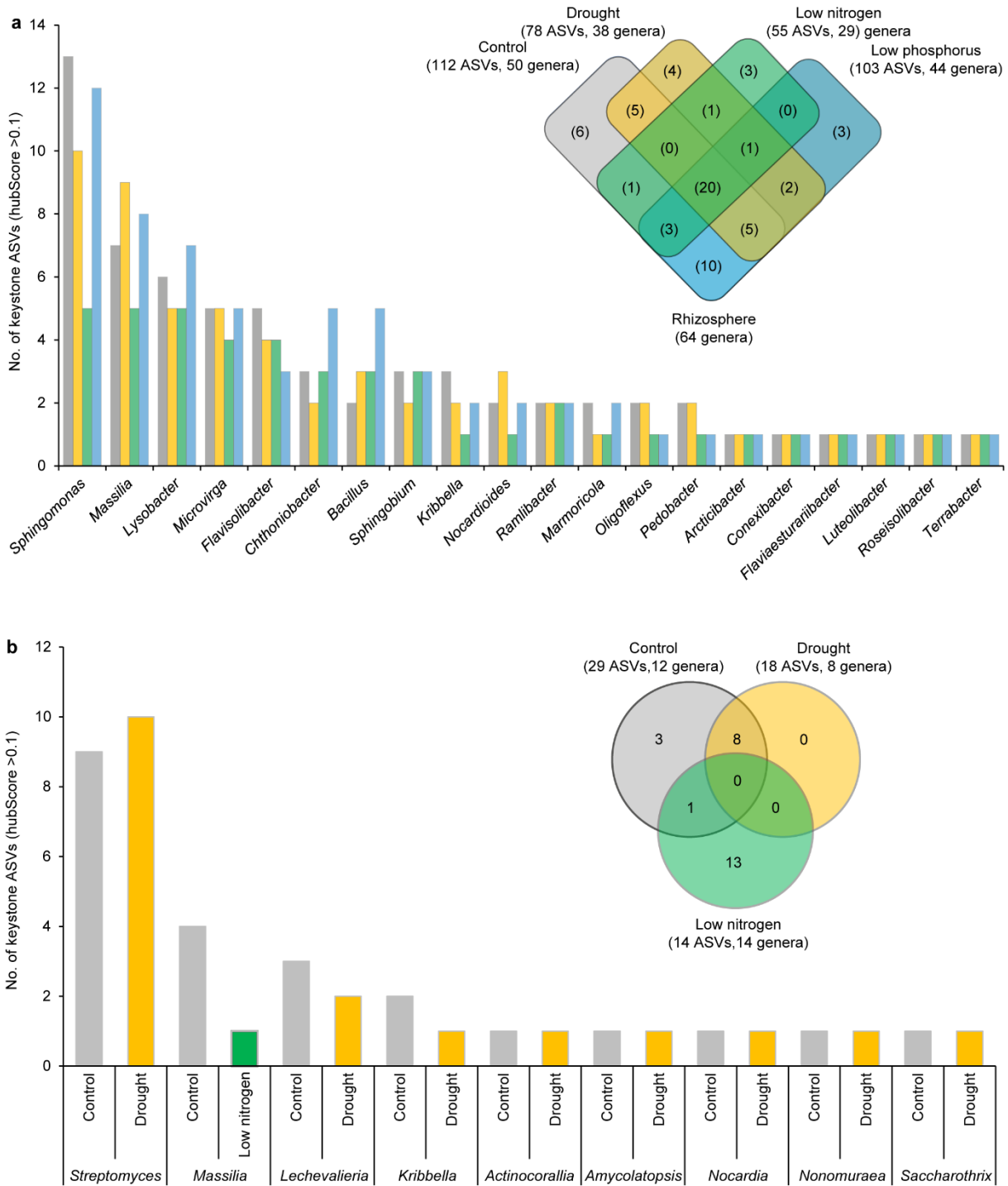
**Supplementary Figure 8. Identification of bacterial keystone ASVs in the rhizosphere (a) and root (b).** Among those ASVs interacting within the network, keystone ASVs were defined based on a hub score >0.1.
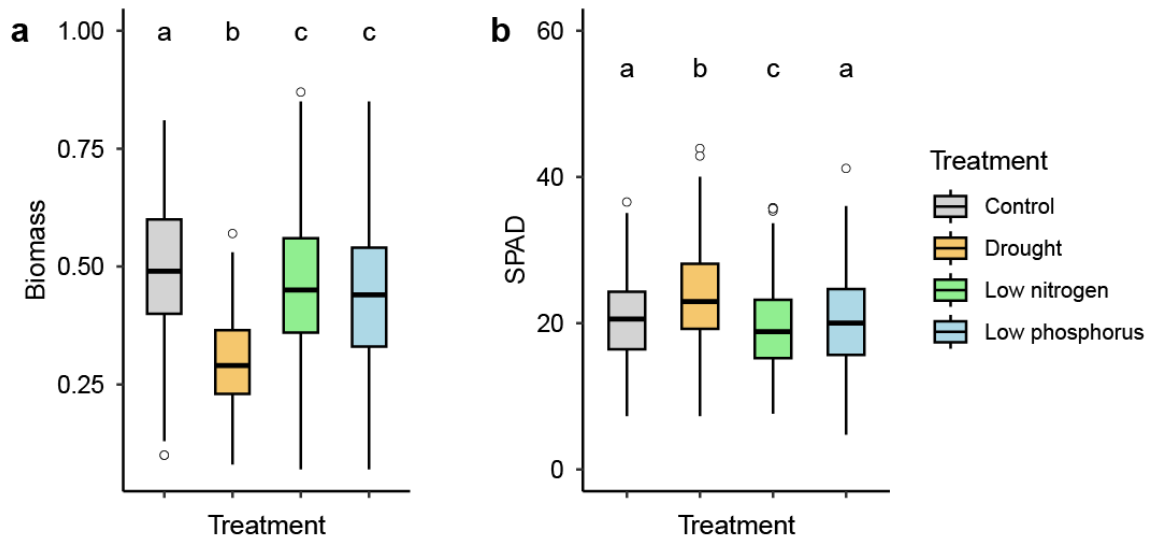
**Supplementary Figure 9. Shoot phenotype analyses measured by shoot dry biomass (a) and leaf chlorophyll content (SPAD) (b) under different stress conditions.** Different letters indicate significant differences controlled by One-Way ANOVA (Tukey's HSD, $P <0.05$).
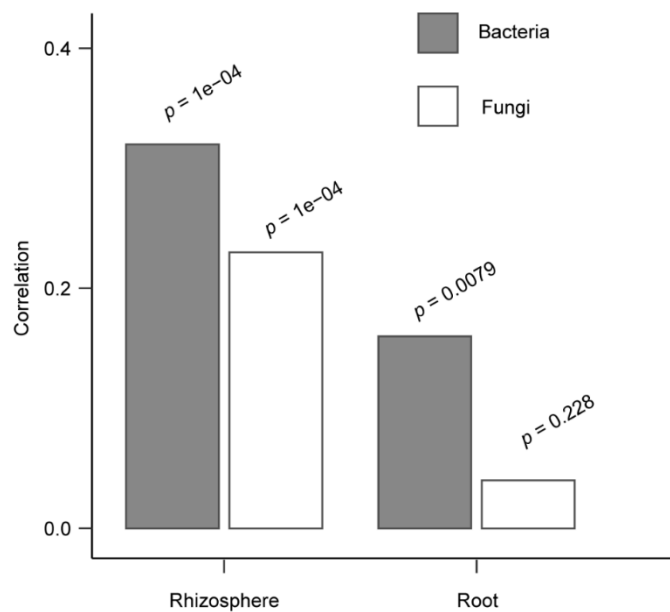


**Supplementary Figure 10. Mantel's statistic tests for correlation between host genotypic distance and microbiome distance.** The exact $p$ values are indicated.
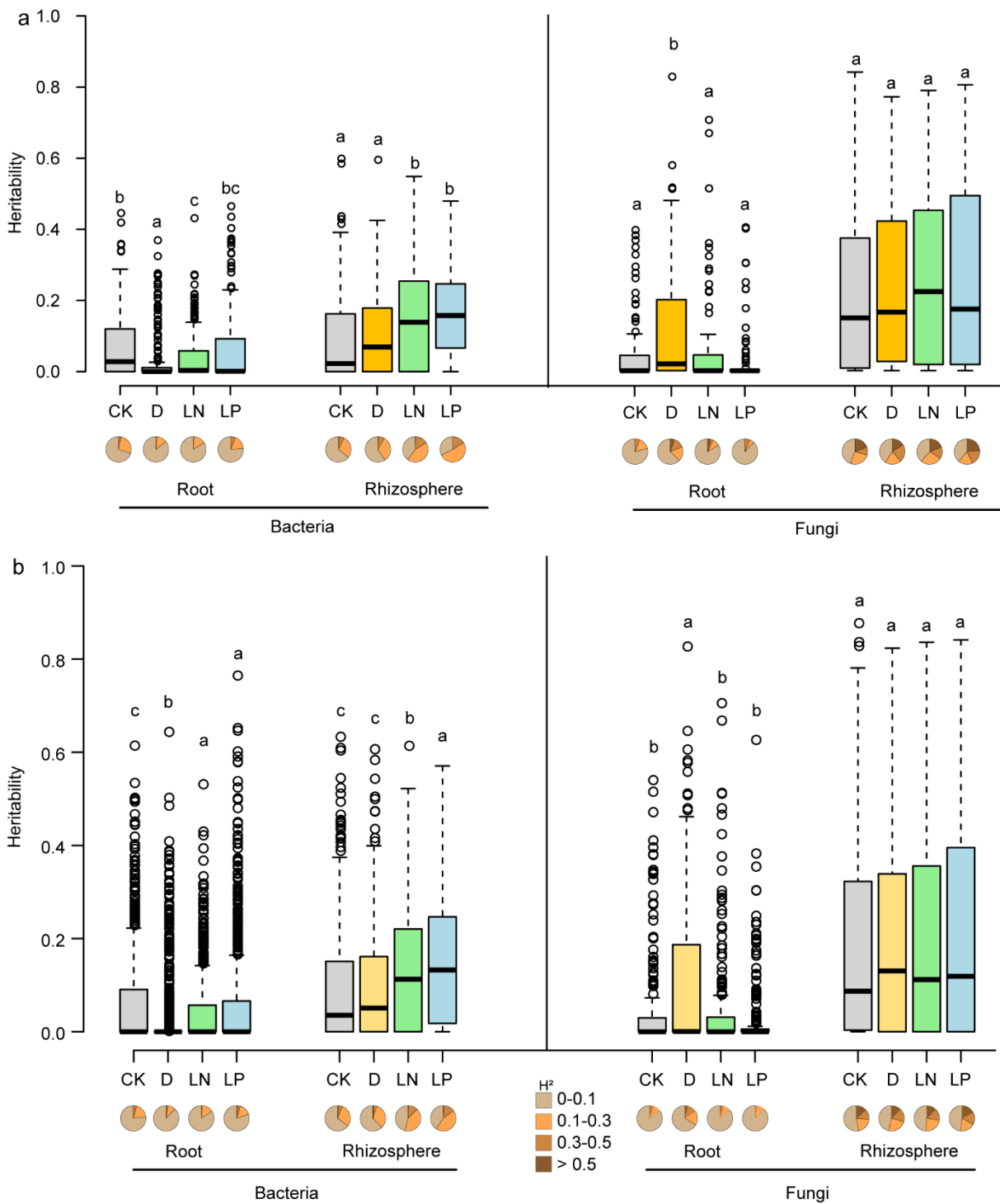
**Supplementary Figure 11. Heritability estimation of microbial traits.** Heritability of all microbial genera (**a**) and ASVs (**b**) under four treatments for both bacteria and fungi (Supplemental Dataset 9). The broad-sense heritability ($H^2$) was calculated using highly abundant bacterial ($n = 815$) and fungal ($n = 248$) ASVs across all samples. Significant differences are indicated for each compartment among treatment groups at $p < 0.05$ after Benjamini-Hochberg adjustment (Kruskal-Wallis test, Dunn's *post-hoc* test). Boxes span from the first to the third quartiles, centre lines represent the median values and whiskers show data lying within the $1.5\times$ interquartile range of the lower and upper quartiles. Data points at the ends of whiskers represent outliers. The pie chats indicate the proportional distribution of heritability frequencies. CK, control; D, drought; LN, low nitrogen; LP, low phosphorus.
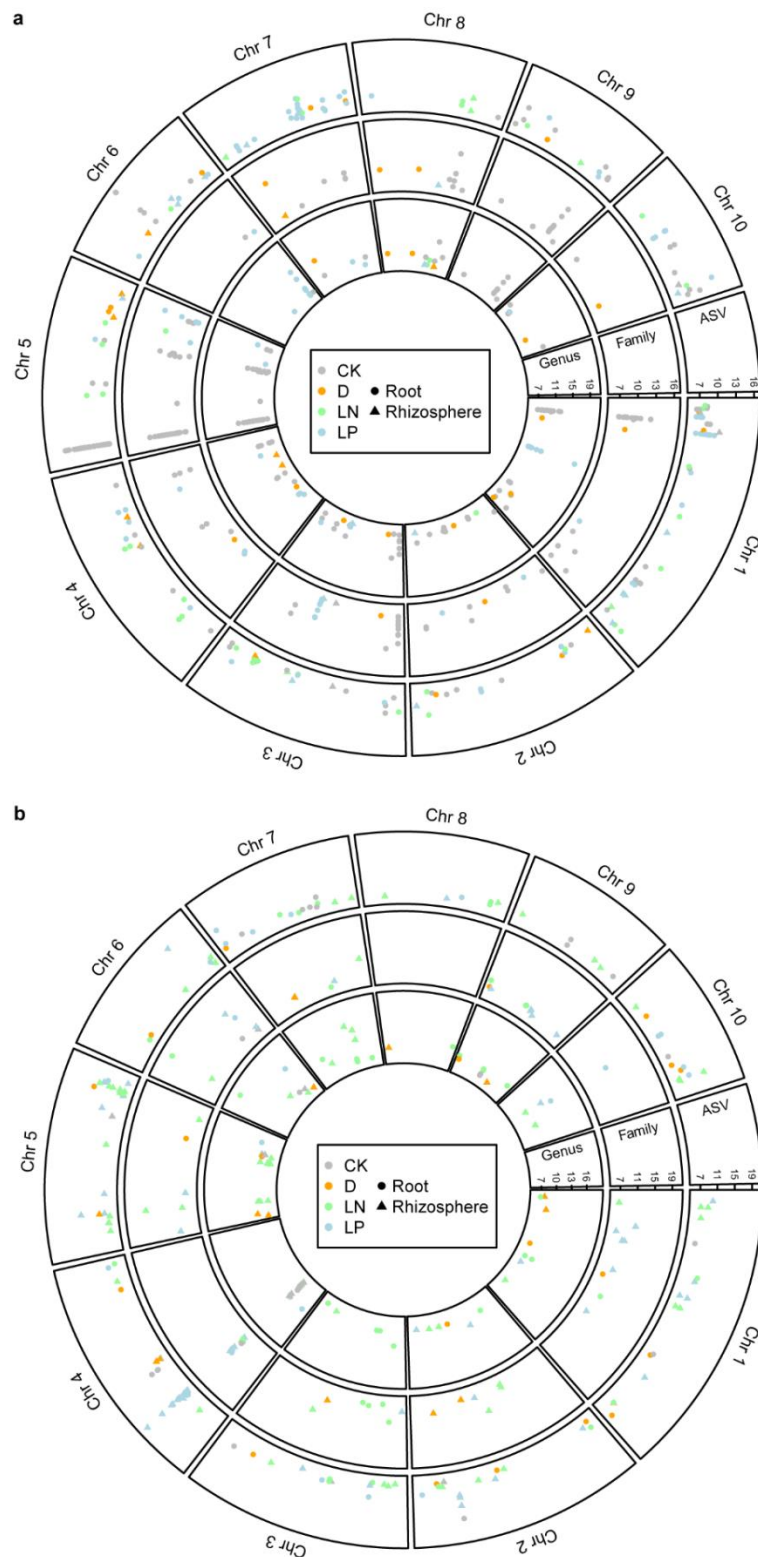
**Supplementary Figure 12. Circular plot summarizing the results of GWAS for bacterial (a) and fungal (b) traits.** Results for bacterial and fungal ASVs, families and genera are shown in three different tracks, respectively. In each track, the $-\log_{10}(P)$ values of all significant marker-trait associations (MTAs) for all traits were plotted together. Only those MTAs explaining more than 5% of the phenotypic variance were included. MTAs for traits under different treatment and compartment are indicated by different colors and symbols. CK, control; D, drought; LN, low nitrogen; LP, low phosphorus.
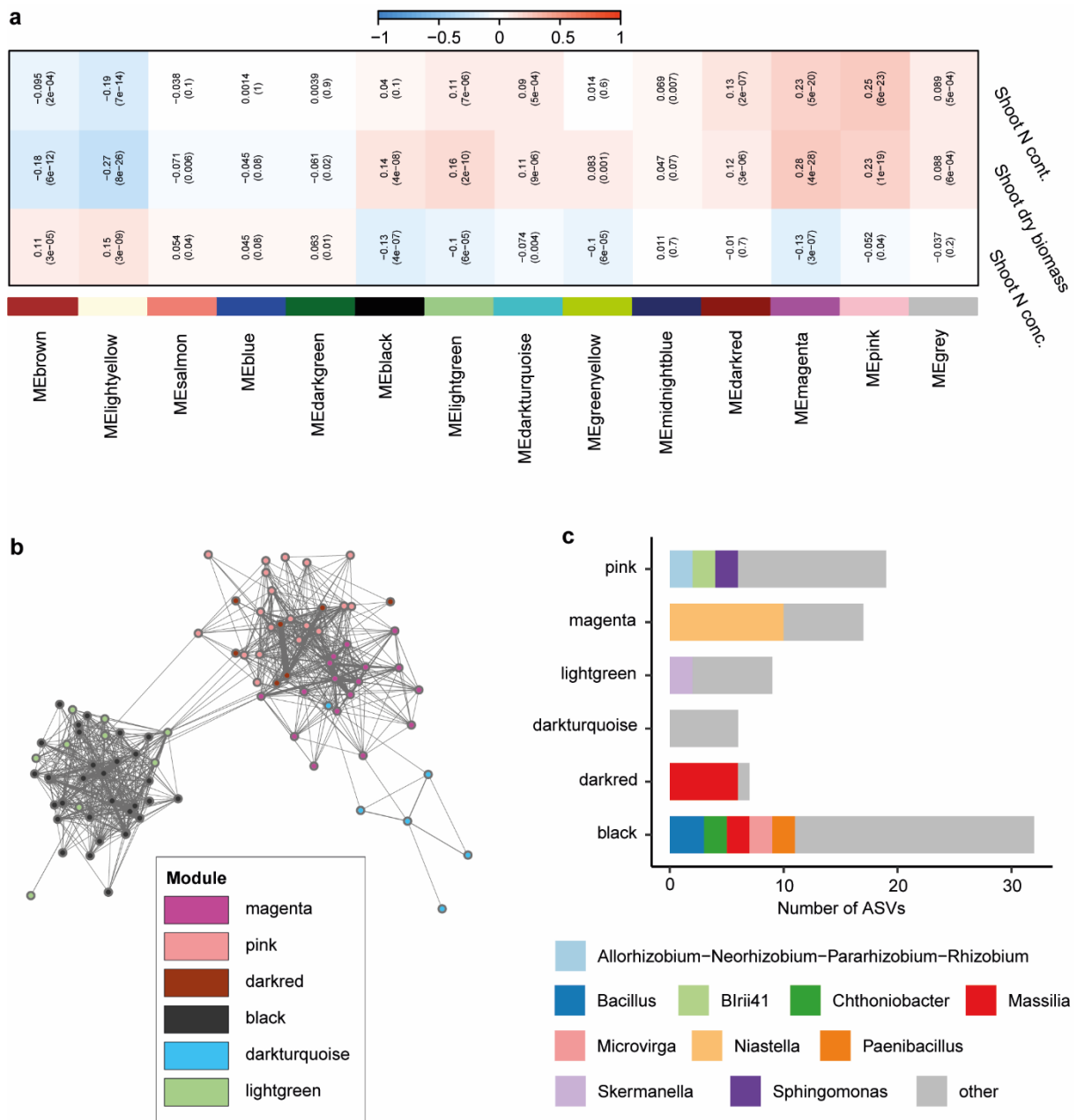
**Supplementary Figure 13. Correlation network including independent bacterial assemblies conformed by taxa associated with the root microbiomes. a,** Identification of different root bacterial assemblies represented by co-expressed modules by weighted co-expression network analysis (WGCNA). Here the module eigen (ME) ASV was used to correlate with different shoot phenotypes e.g. shoot dry biomass, shoot nitrogen concentration and content. The numbers in the brackets indicate the correlation significance. **b**, Modules correlation network considering only microbial modules positively associated with shoot phenotypes. **c**, Different module assemblies represent taxa within the ecological network including bacterial ASVs highly co-occurring with each other. To plot the composition of ASVs in each significant module at genus level, only ASVs with frequency >=2 were colored. Other ASVs were all labeled as other.
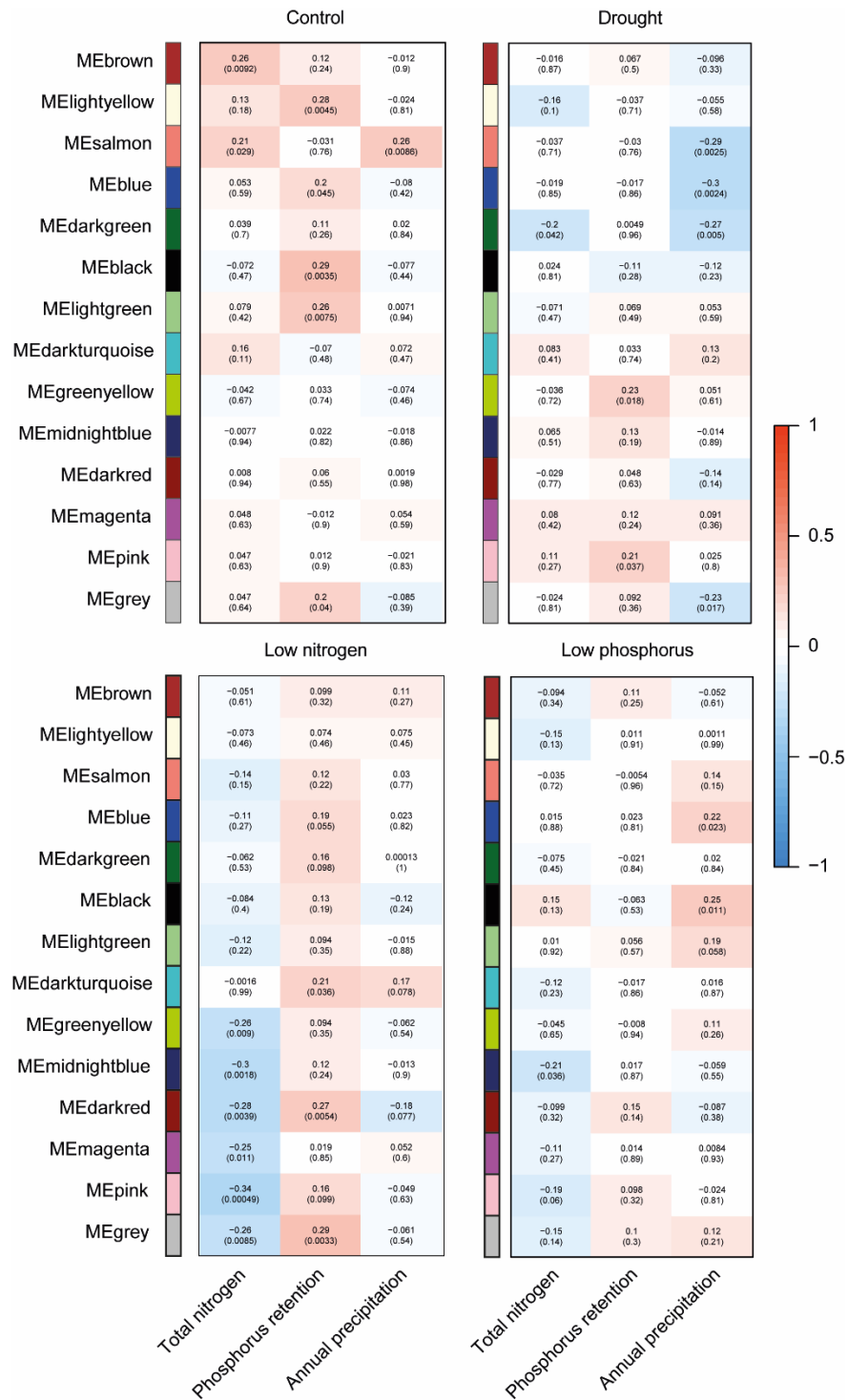
**Supplementary Figure 14. Spearman correlations between source environmental factors and different independent root bacterial assemblies in Supplementary Figure 13.** Total nitrogen, retention potential of soil phosphorus and mean annual precipitation natural availability of soil nutrients and precipitation conditions in the places of origin for maize varieties. These source environmental factors are included for correlation as indirect climatic effects as our microbiomes were determined in phytochamber experiments under different water and nutrient treatments.
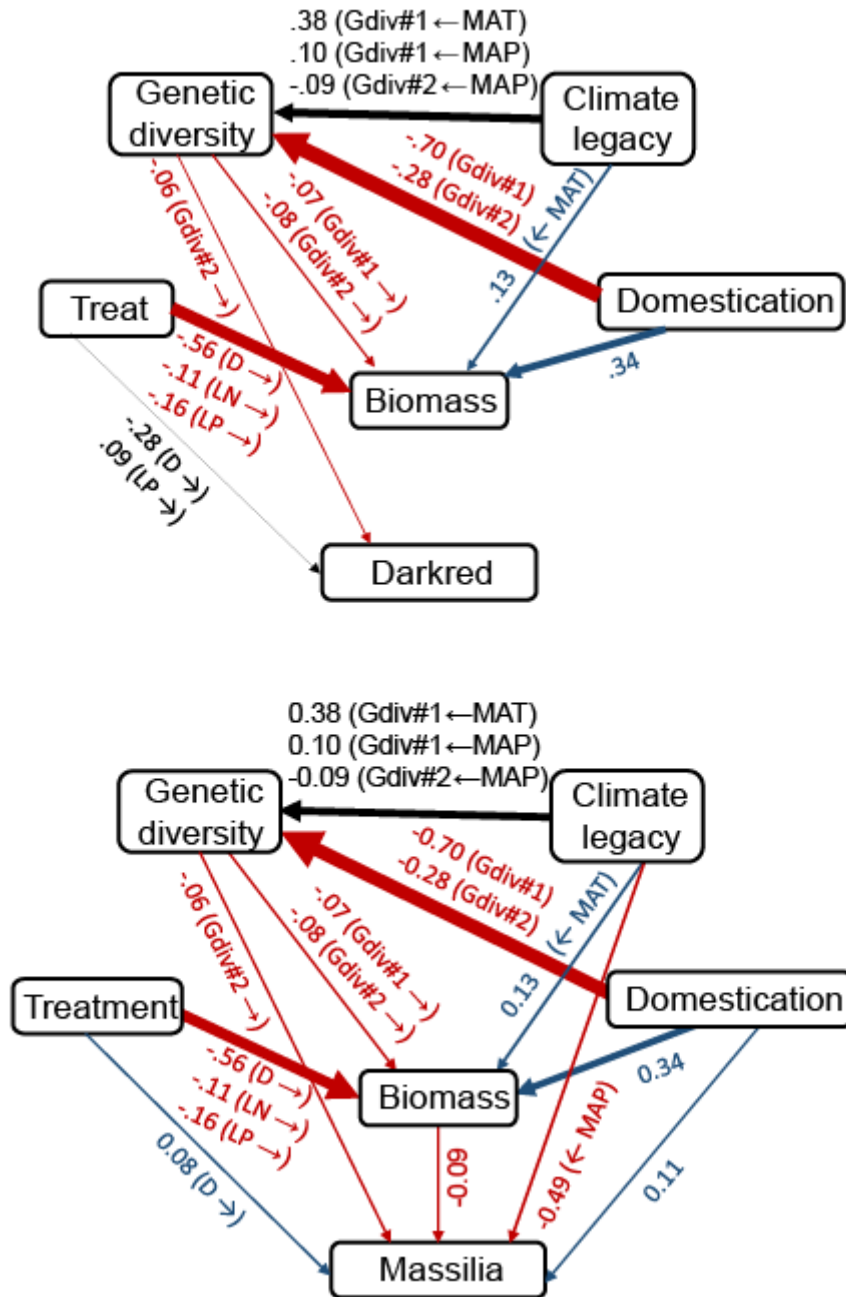
**Supplementary Figure 15. Structural equation modeling exploring the direct and indirect effects of climatic legacies, genotype diversity, treatments, domestication and biomass on the microbial "darkred" module and keystone bacteria *Massilia*.** Plant genetic diversity 1 (Gdiv#1) and 2 (Gdiv#2) are determined as the axes of a NMDS (Non-metric Multidimensional Scaling) analysis including maize genetic variability. Mean annual precipitation (MAP) and mean annual temperature (MAT) represent field temperature and precipitation conditions in the places of origin for maize varieties. They are included in our model as climatic legacies as our microbiomes were determined in phytochamber experiments under controlled conditions. CK, control; D, drought; LN, low nitrogen; LP, low phosphorus. Values indicate positive or negative Spearman correlation coefficients. Numbers adjacent to arrows are indicative of the effect size of the relationship. Only significant relationships (*P* <0.05) are included. Climate legacy, genetic diversity and treatments are included in our models as independent observable variables; however, we group them in the same box in the model for graphical simplicity. All predictors within each are allowed to co-vary. This does not apply to the model in which only one predictor for a given group is included. In this case, the name of the predictor stands alone (e.g. domestication). All models showed a good model fit. The standardized total effect (STE) of low nitrogen, source mean annual temperature, source precipitation and plant genotype (genetic diversity 2) are calculated.
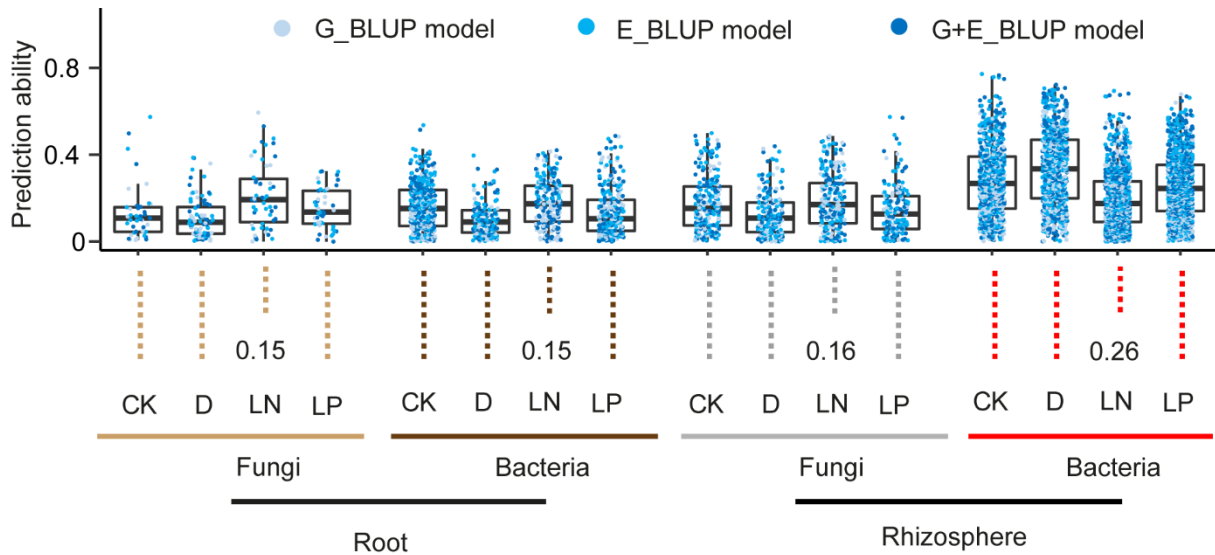
**Supplementary Figure 16. Genomic and environmental prediction of microbial ASVs.** ASV traits prediction using genetic markers and environmental characters. The numbers denote the average prediction accuracies for microbial ASVs from different treatments across compartments. Only ASVs with heritability ($H^2$) >0.1 were considered in prediction analyses. Boxes span from the first to the third quartiles, centre lines represent the median values and whiskers show data lying within 1.5× interquartile range of the lower and upper quartiles. Data points outside of whiskers represent outliers. CK, control; D, drought; LN, low nitrogen; LP, low phosphorus. G_BLUP model refers to genetic model; E_BLUP model refers to environmental model; G+E_BLUP model refers to combined genetic and environmental model.
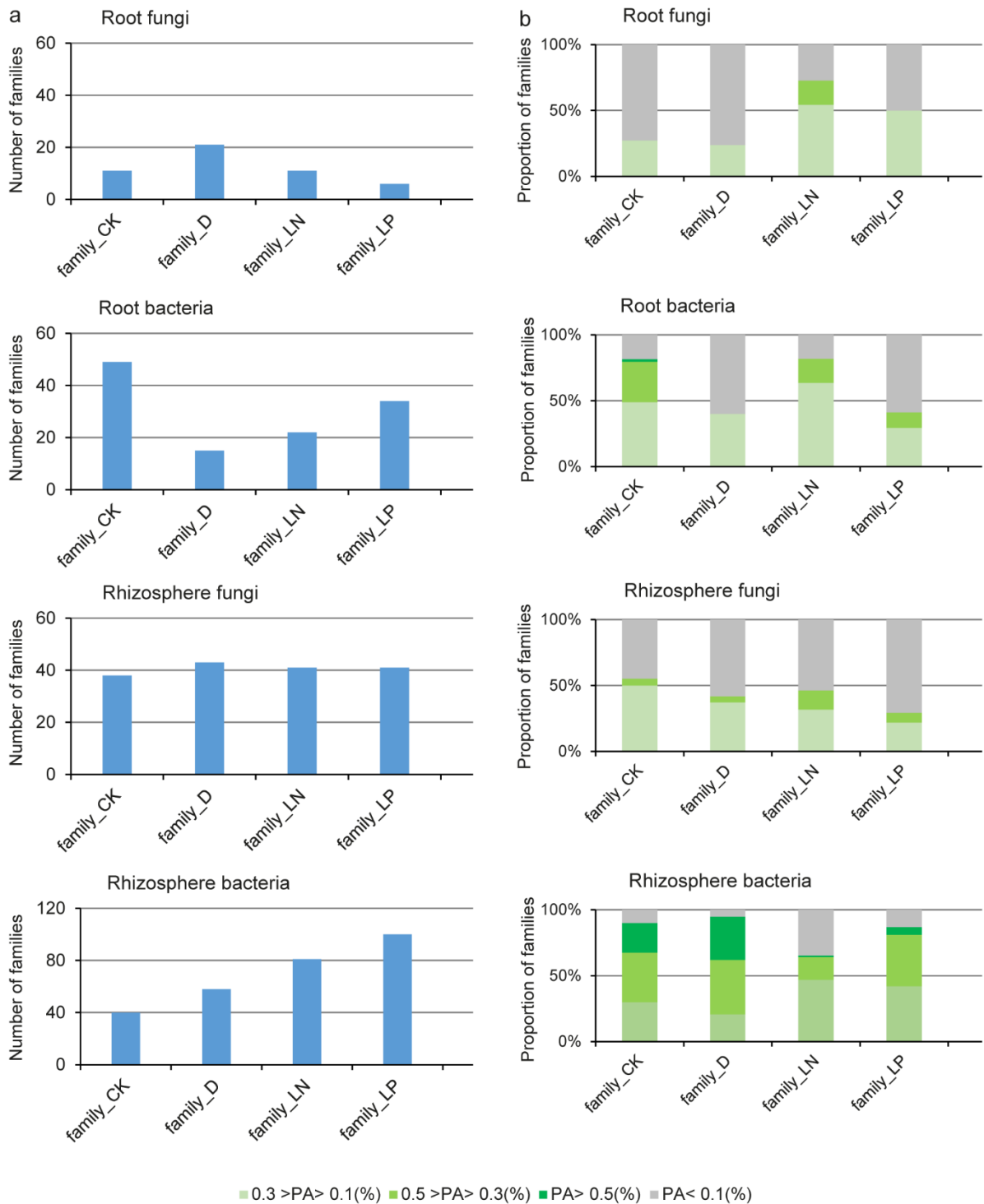
**Supplementary Figure 17. Prediction accuracy of microbial families. a**, Number of highly heritable ($H^2$ >0.1) microbial families. **b**, Degree of prediction accuracies of microbial families from different treatments across compartments. Only families with heritability ($H^2$) >0.1 were considered in prediction analyses. PA, prediction accuracy; CK, control; D, drought; LN, low nitrogen; LP, low phosphorus.
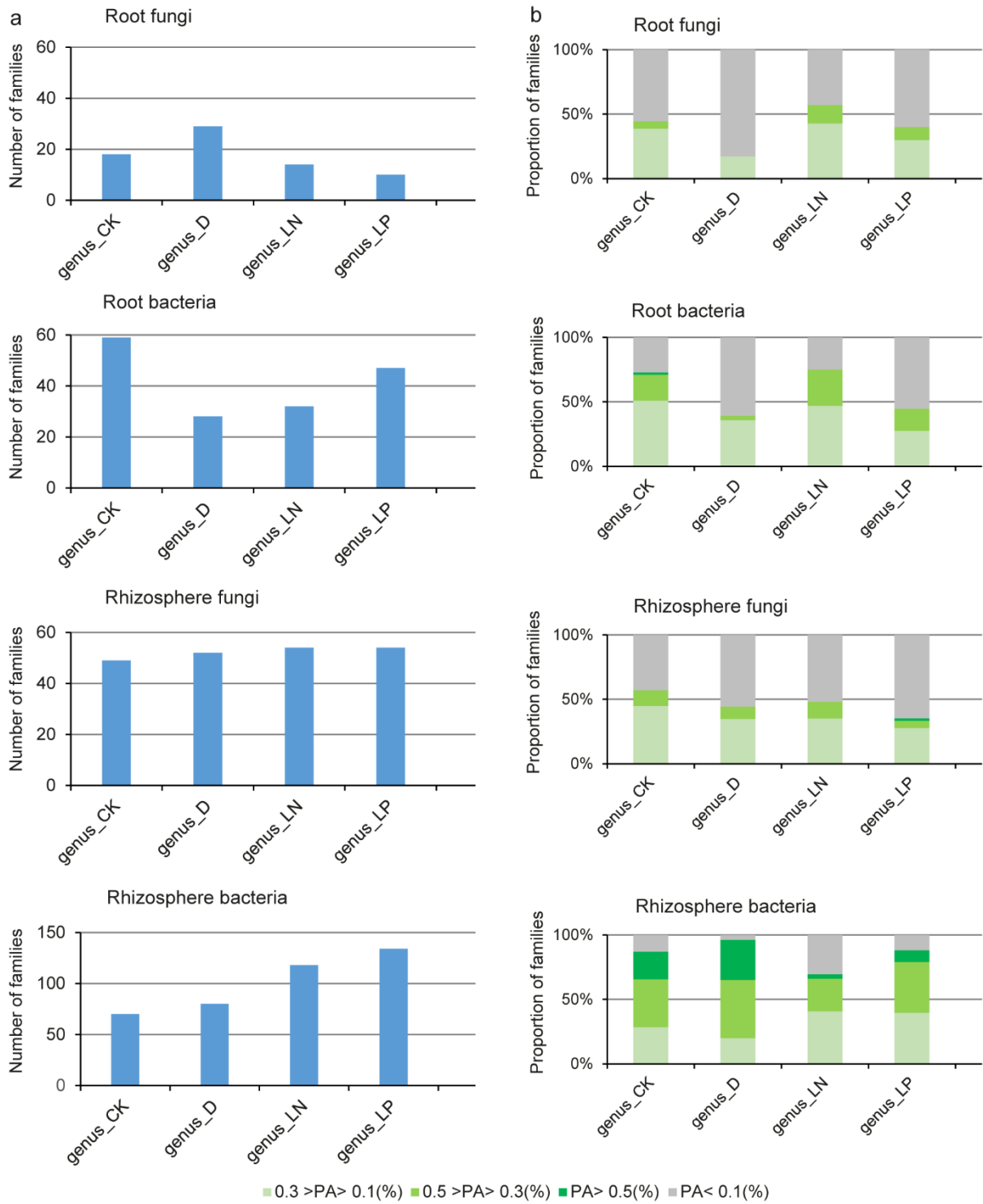
**Supplementary Figure 18. Prediction accuracy of microbial genera. a**, Number of highly heritable ($H^2$ >0.1) microbial genera. **b**, Degree of prediction accuracies of microbial genera from different treatments across compartments. Only genera with heritability ($H^2$) >0.1 were considered in prediction analyses. PA, prediction accuracy; CK, control; D, drought; LN, low nitrogen; LP, low phosphorus.
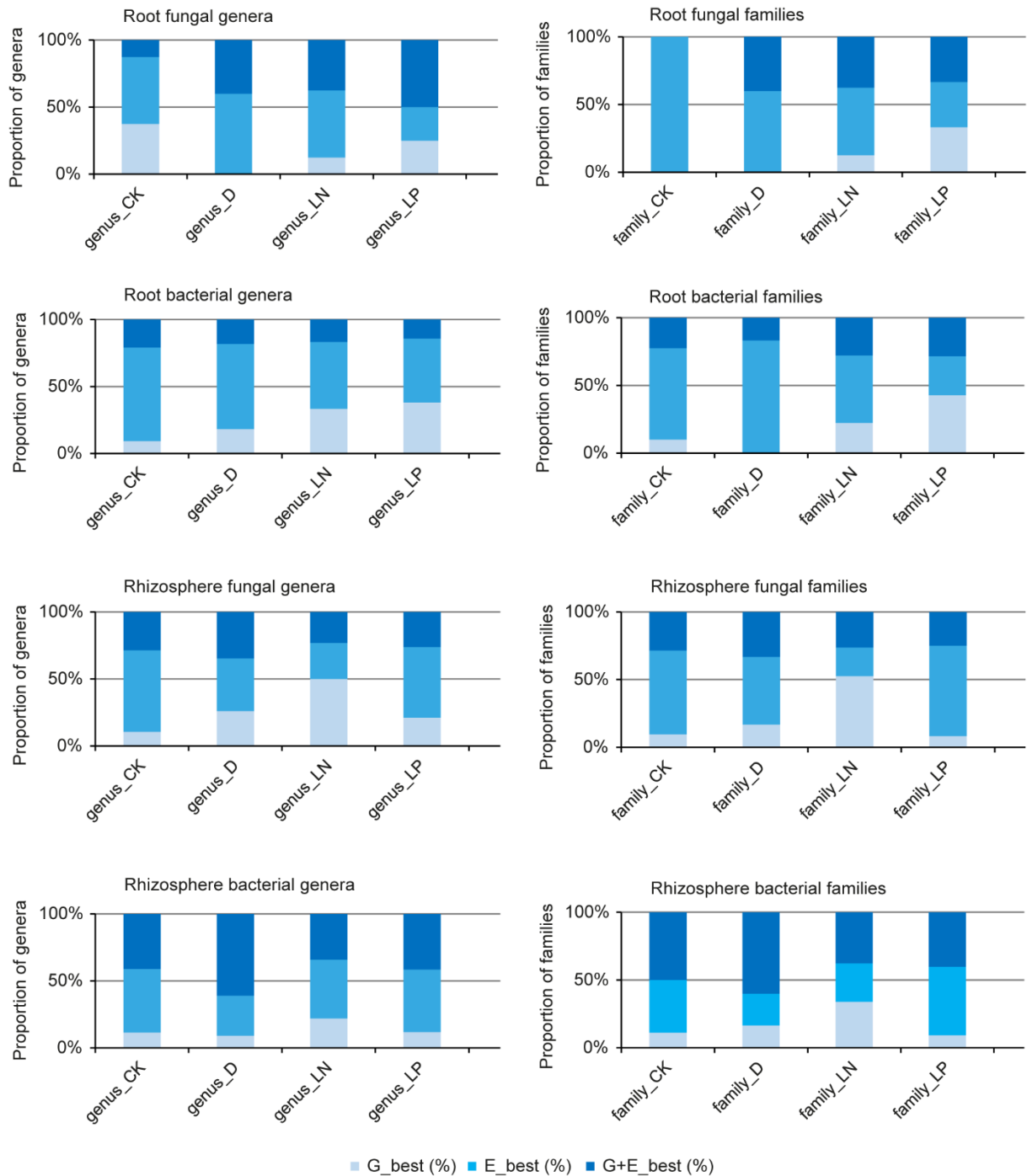
**Supplementary Figure 19. Best prediction patterns of microbial taxa.** G_best indicates best prediction by genetic makers alone. E_best indicates best prediction by environmental characters alone. G+E_best indicates best prediction by combined genetic markers and environmental characters. Only microbial families and genera with heritability ($H^2$) >0.1 were considered in prediction analyses. CK, control; D, drought; LN, low nitrogen; LP, low phosphorus.
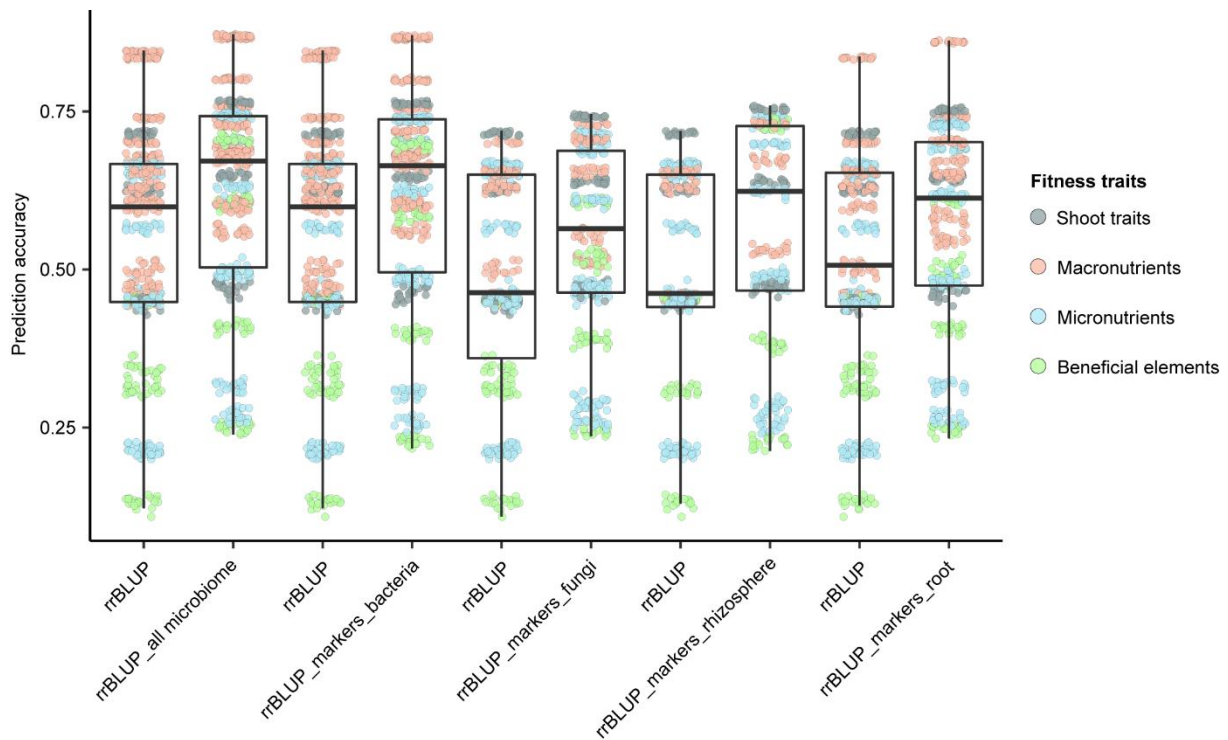
**Supplementary Figure 20. Fitness traits prediction results using host genetics data alone (rrBLUP) and both genetic and microbiome data.** In the figure, different fitness traits were aggregated into one boxplot. In addition to the SNP marker data, the following microbiome data were incorporated into the prediction models: 1) all the microbiomes, including bacteria and fungi from both rhizosphere and root compartments, 2) bacteria from both rhizosphere and root compartments, 3) fungi from both compartments, 4) bacterial and fungi from the rhizosphere, and 5) bacteria and fungi from the root.
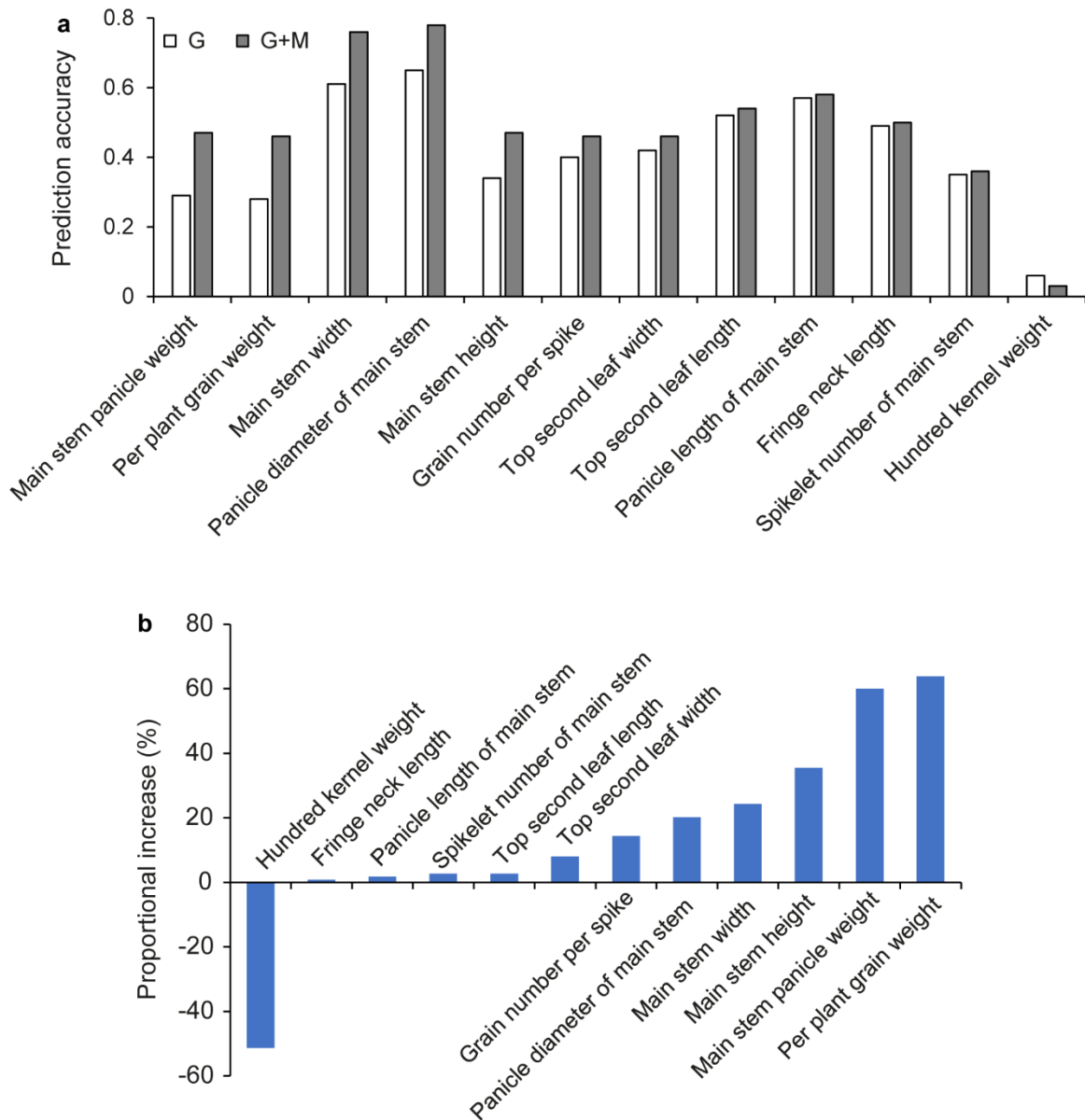
**Supplementary Figure 21. Agronomic trait prediction using genetic markers and microbiome traits in foxtail millet. a**, Prediction accuracy for different agronomic traits using genetic (G) markers alone or combined genetic and microbiome traits (G+M). **b**, Proportional increase of prediction accuracy for different agronomic traits using G+M mode over G mode. 12 agronomic traits were extracted from the publication by Wang et al. 2022 and predicted using genetic markers or combined markers and microbiome traits.
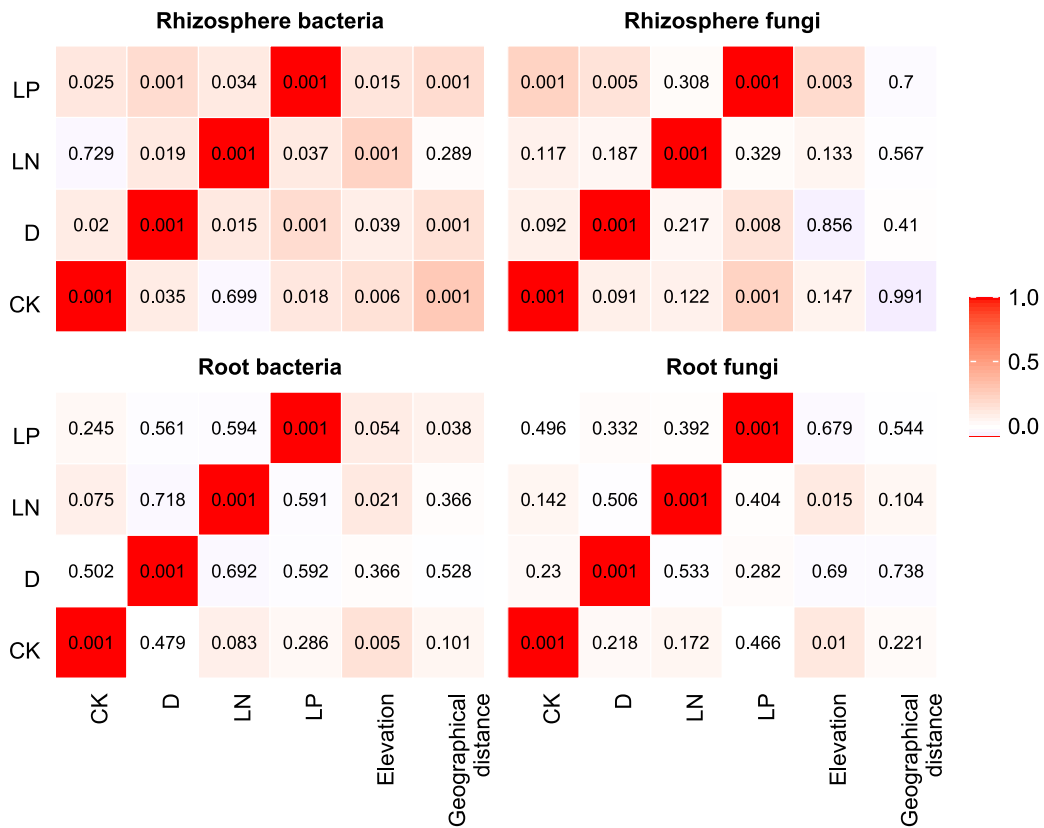
**Supplementary Figure 22. Correlation between microbial communities and local environments.** Mantel's statistical tests were performed between microbiome features and environmental characters (elevation and geographical distance). Significances were indicated by exact *p* values based on 999 permutations.
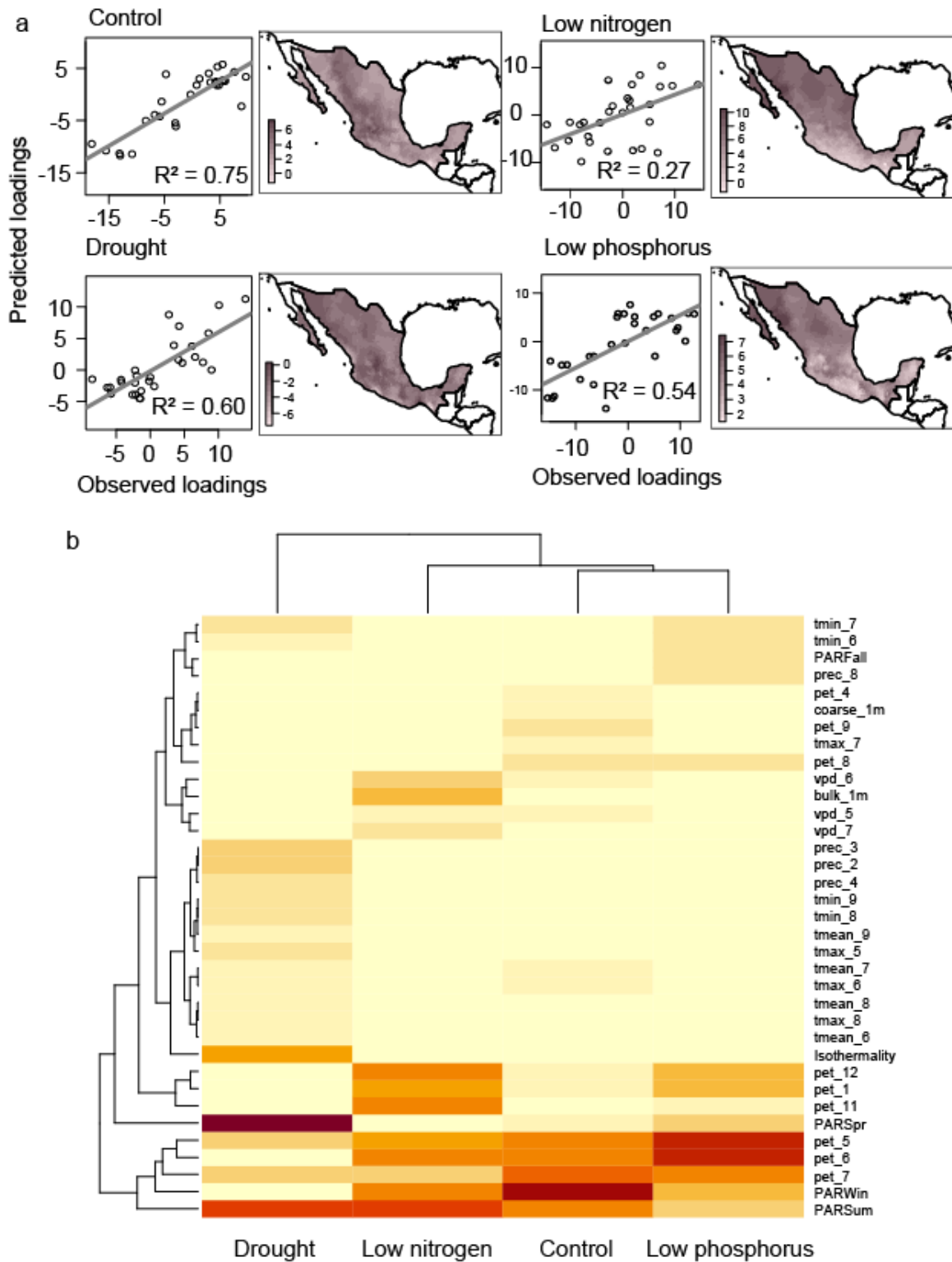
**Supplementary Figure 23. Prediction of rhizosphere bacteria PC2. a,** Random forest approach can well predict the microbiome PC2 in the rhizosphere under different conditions. **b**, Heatmap indicates the importance of environmental variables in PC2 predictions across treatments.
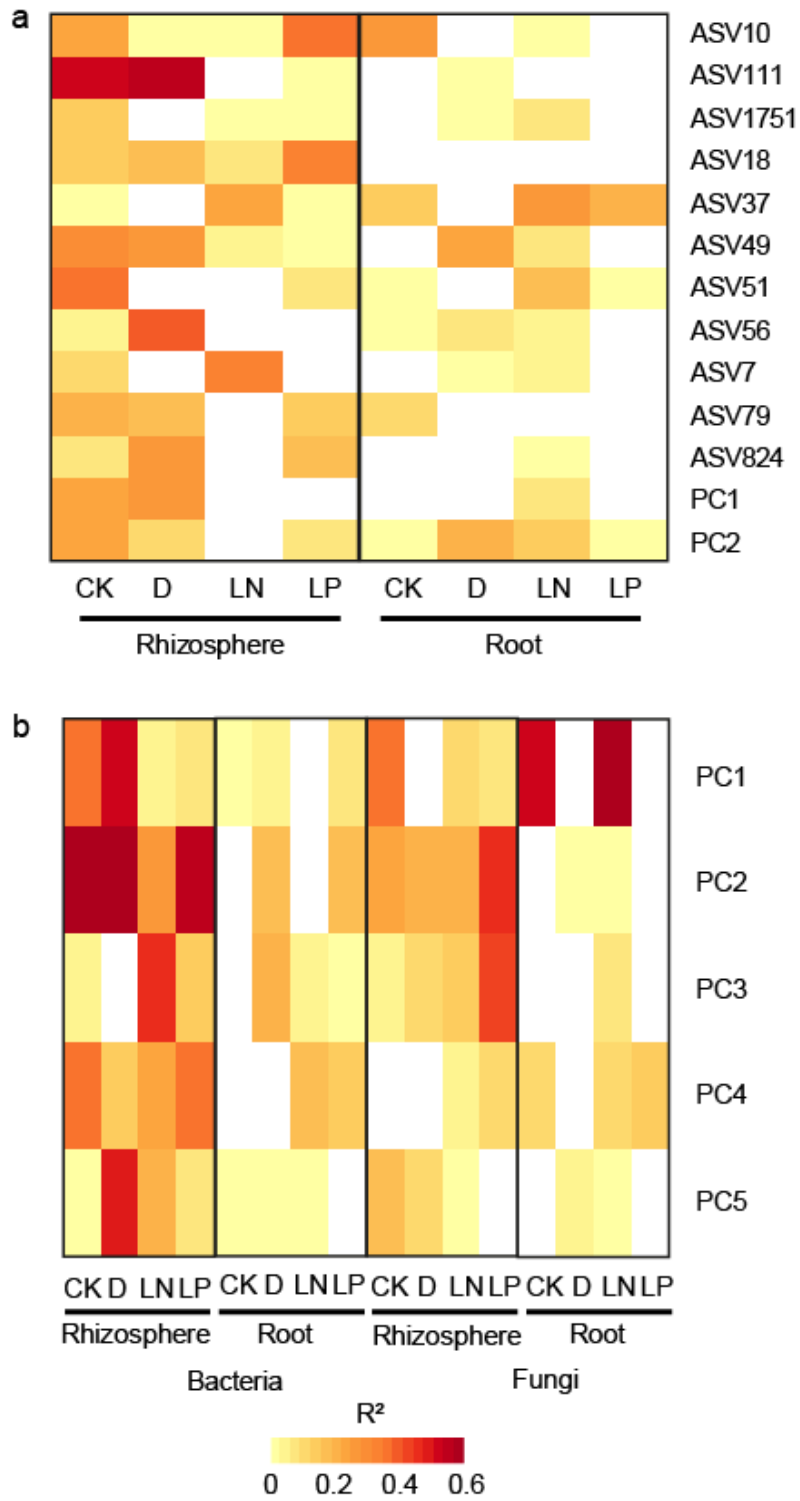
**Supplementary Figure 24. Microbiome prediction. a**, Prediction of microbial abundance by using individual *Massilia* ASVs. **b**, Prediction of microbial abundance by using the whole microbiome (top 5 PCs).
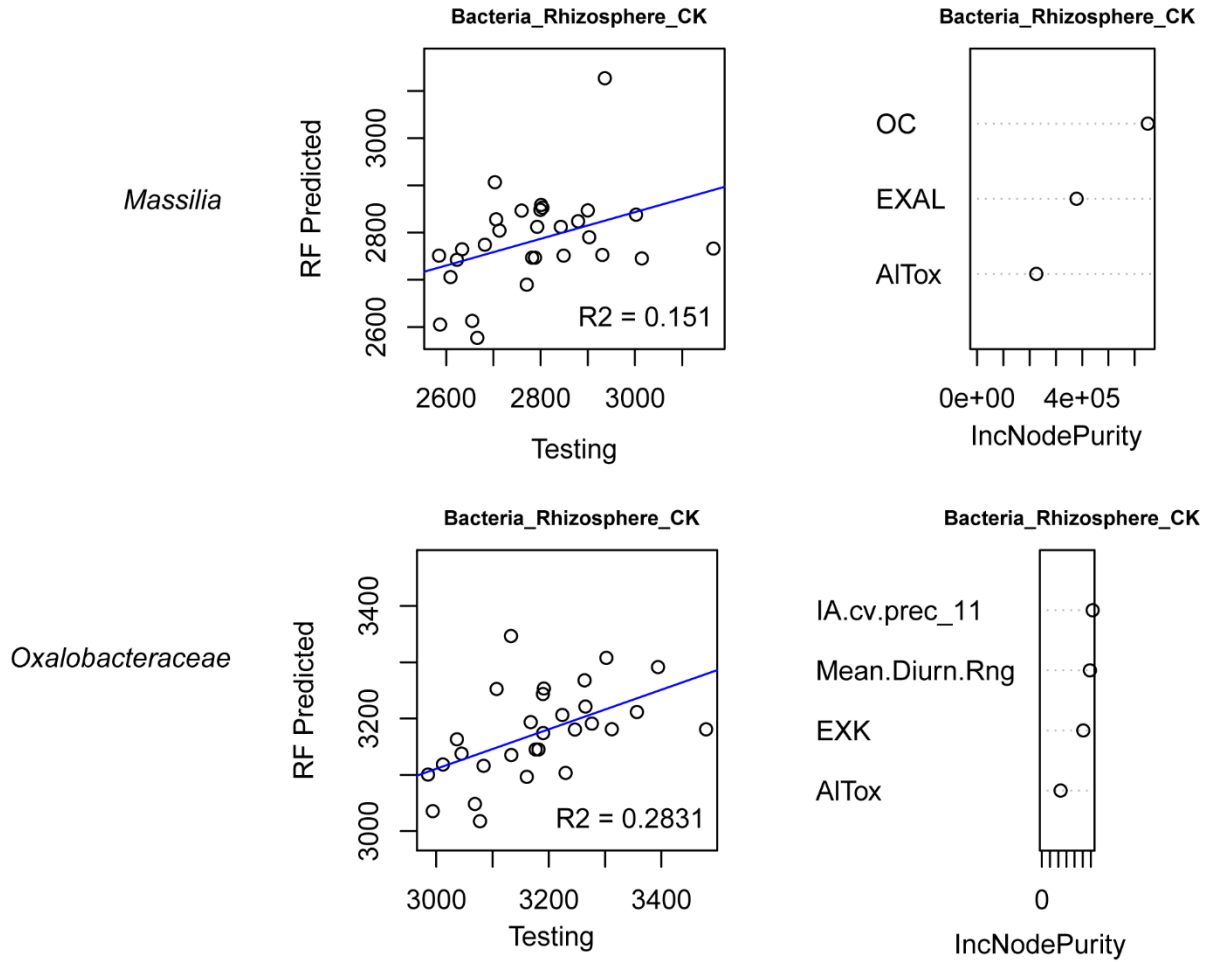
**Supplementary Figure 25. Random forest modelling correlation of testing taxa.** The environmental characters can be referred to from Supplementary Dataset 12.
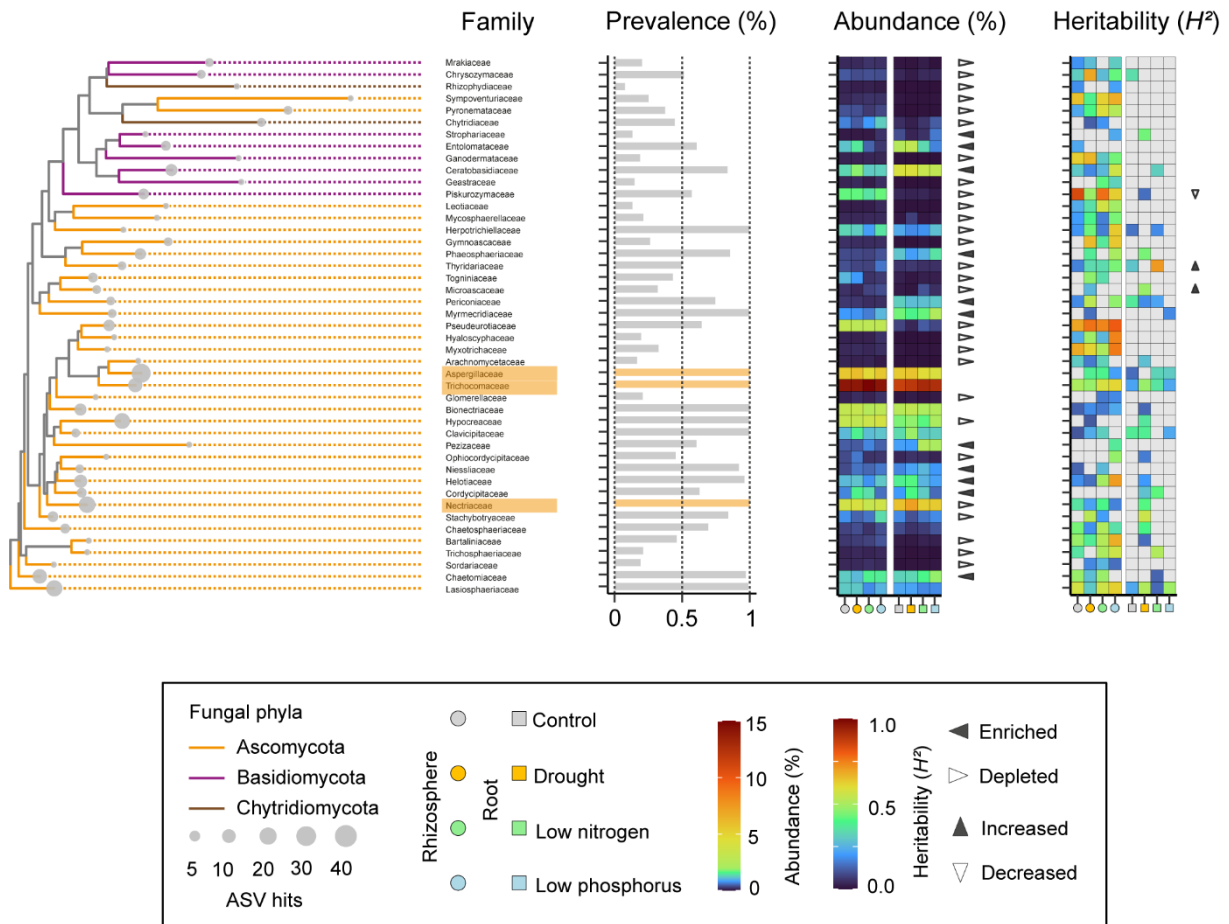
**Supplementary Figure 26. Maximum-likelihood phylogeny of dominant fungal families (n > 5).** Circle sizes along the branches of the tree indicate the number of ASVs observed in association with microbial families. Colour coded families are clustered at the phylum level. Bar plots describe the prevalence according to the proportional sample size. The heatmaps illustrate the standardized mean relative abundance and the estimated heritability of microbial families from the root to the rhizosphere. Triangles represent the enrichment or depletion of microbial families, and increased or decreased heritability from the root to the rhizosphere. The significance levels were controlled at two levels (*: $p < 0.05$; **: $p < 0.01$).
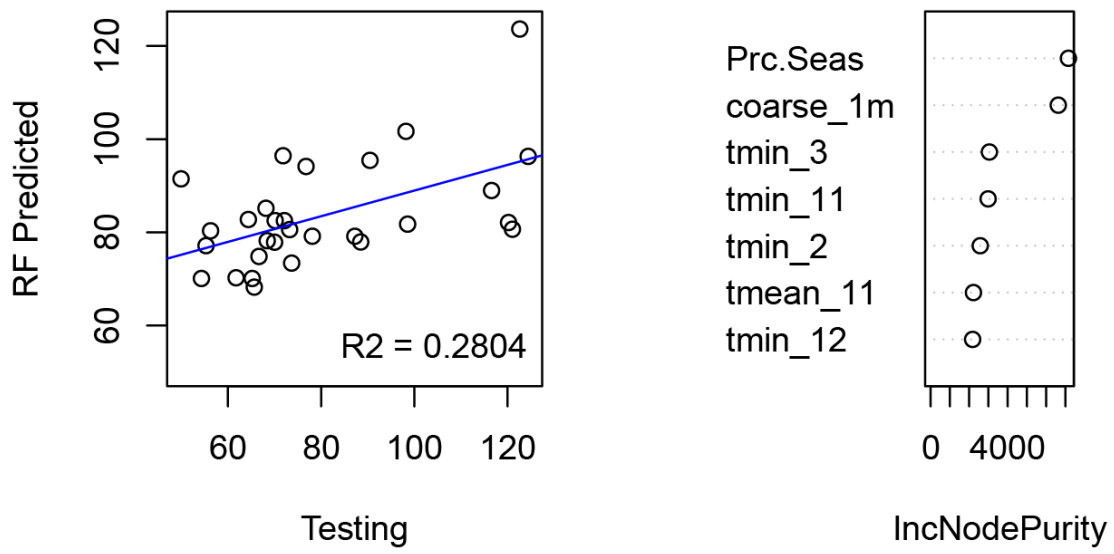
**Supplementary Figure 27. Random forest modelling correlation of ASV37 in root under nitrogen-poor conditions.** The environmental characters can be referred to from Supplementary Dataset 12.
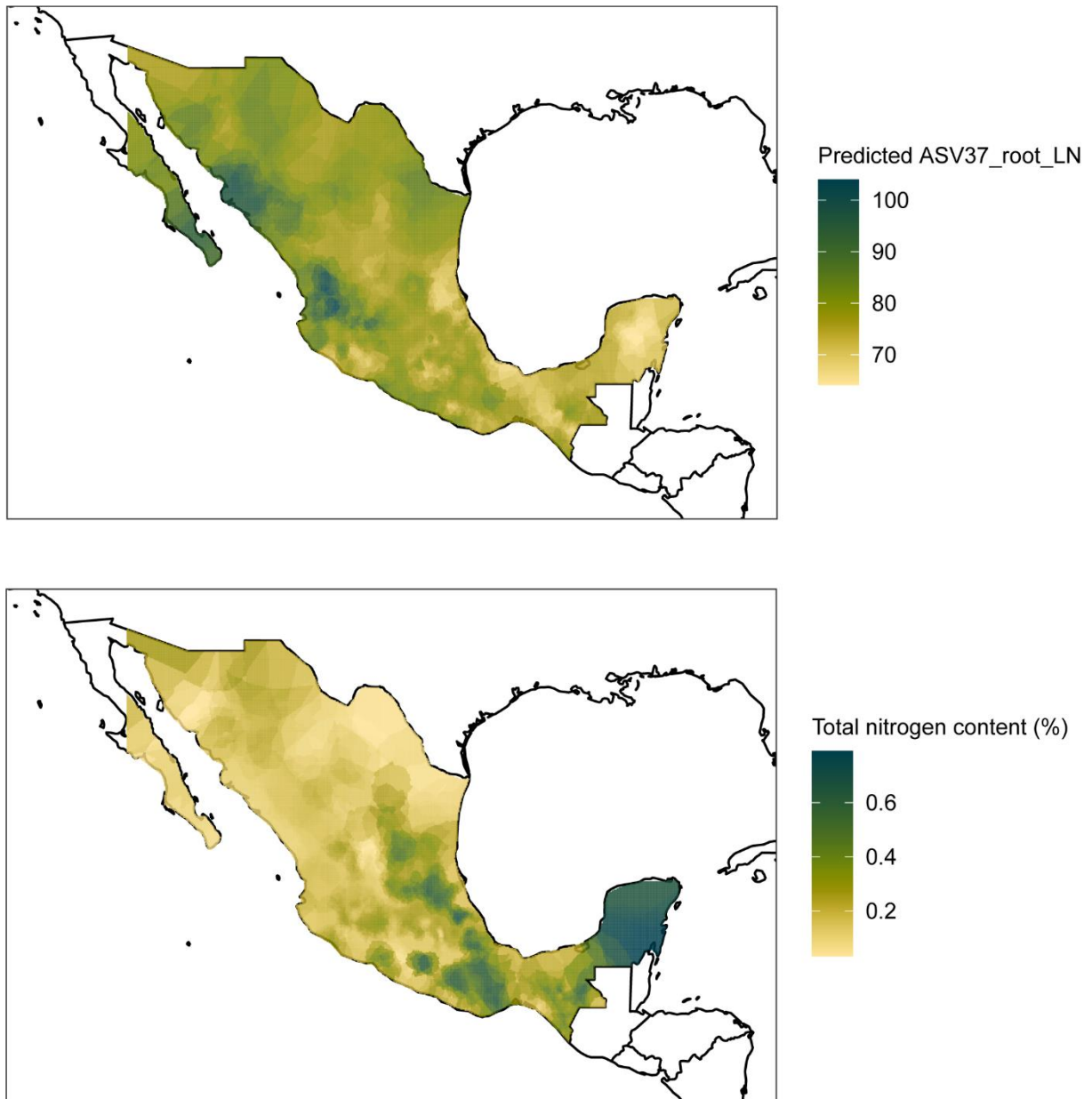
**Supplementary Figure 28. Maps showing the spatial variation of targeted ASV and total soil nitrogen contents based on 1,781 landraces.** Top panel: Predicted relative abundance of ASV37 in roots under low nitrogen; Bottom panel: Distribution of total nitrogen content in soils. Green shades correspond to higher values, whereas yellow shades correspond to lower values.
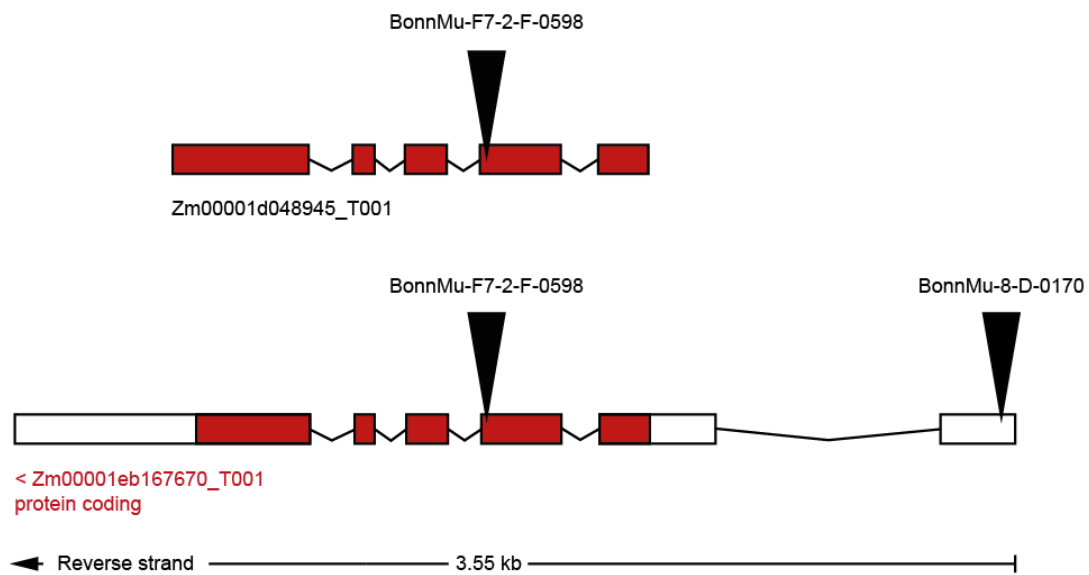
**Supplementary Figure 29. Transposon-tagged mutations of gene Zm00001d048945 (version 4).** Induced maize mutants of the BonnMu resource derive from Mutator-tagged $F_2$-families in the genetic backgrounds B73 and F7. We identified two insertion lines, BonnMu-8-D-0170 (B73) and BonnMu-F7-2-F-0598 (F7), harboring insertions 1,264 bp upstream of the start codon ATG and in the second exon of Zm00001eb167670 (version 5), respectively.
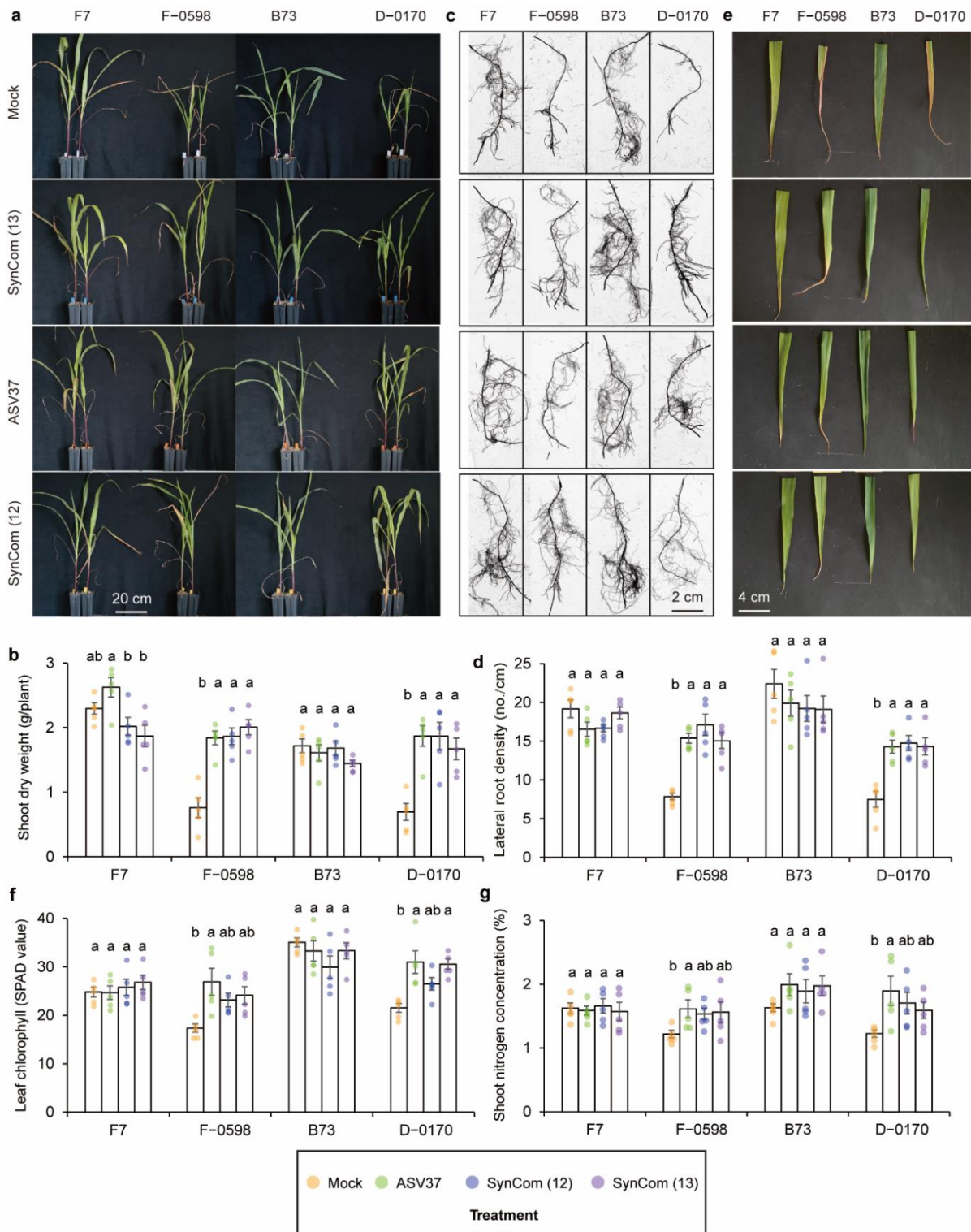
**Supplementary Figure 30. Inoculation of *Massilia* can modulate lateral root development and growth performance of lateral root defective mutants in another nitrogen-poor soil**. *Massilia* inoculations are able to enhance shoot dry weight (**a, b**) and lateral root development (**c, d**). Representative images of the whole shoot and 1st whorl of crown roots illustrate the shoot performance and more emerged lateral roots by *Massilia* strains. Nitrogen deficient phenotype (**e, f**) and shoot nitrogen concentration (**g**) was evaluated by relative leaf chlorophyll concentration measured by the SPAD value of the last fully expanded leaf. Each individual leaf was measured 10 times as the average value. D−0170 and F−0598 are the lateral root defective mutants in comparison to their respective wild type plants B73 and F7. Different letters indicate significantly different groups (ANOVA, Tukey's HSD, $P < 0.05$). $n = 5$ biologically independent samples and the error bars indicate the standard errors. Scale bars are indicated in the images. Prior to ANOVA analysis, the observed values were checked for normal distribution and the homogeneous variance among the groups.
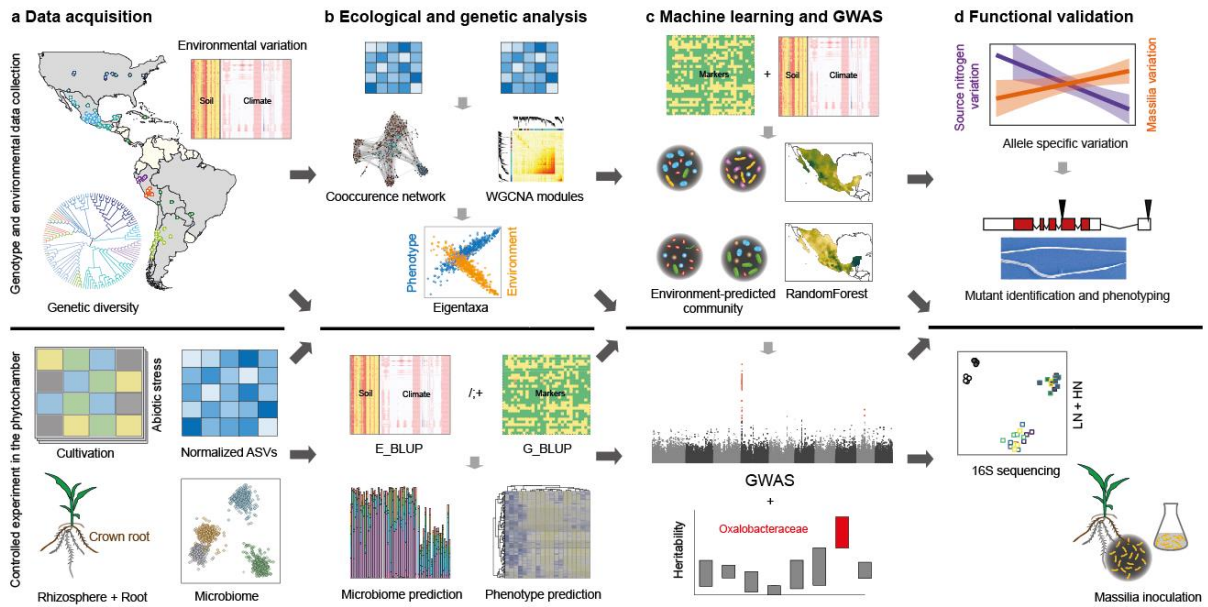
**Supplemental Fig 31. Graphic illustration of experimentally and computationally guided analyses of host root and rhizosphere microbiome association in maize. a,** Different maize locally adapted varieties (teosinte and landraces) and the environmental (soil and climate) data of their source habitats were collected from the databases. Meanwhile, these local varieties together with another modern varieties were grown under the controlled conditions with different abiotic stress treatments (control, drought, low nitrogen and low phosphorus) in the phytochamber pot experiment. The microbiome was sequenced in the compartment rhizosphere and root for both bacterial and fungal community. **b,** The normalized ASVs table was used for microbial cooccurrence network and WGCNA analyses to identify the keystone and eigentaxa, which were used to correlate with plant traits and sourced environments e.g. total nitrogen, phosphorus retention rate and annual precipitation, which corresponds to our experimental treatment. The heritability and genomic predictions using different models e.g. G_BLUP, E_BLUP, G+E_BLUP were performed to understand how the genetic and environmental variation affect the relative abundance of microbial ASVs, followed by plant trait prediction using genetic and microbiome information. **c,** RandomForest machine learning approach was used to integrate the local environmental data and genomic markers to predict the microbial abundance, which was further used for GWAS analysis based on the highly heritable microbial taxa e.g. *Oxalobacteraceae*. **d,** We finally identified the genetic variation and gene-encoded mutant defected on lateral root formation, which are negatively associated with sourced total nitrogen availability. To the end, we verified the gene regulated lateral root formation and its reciprocal interaction with specific microbial taxon *Massilia* (*Oxalobacteraceae*) by 16S rRNA sequencing and root inoculation experiments.
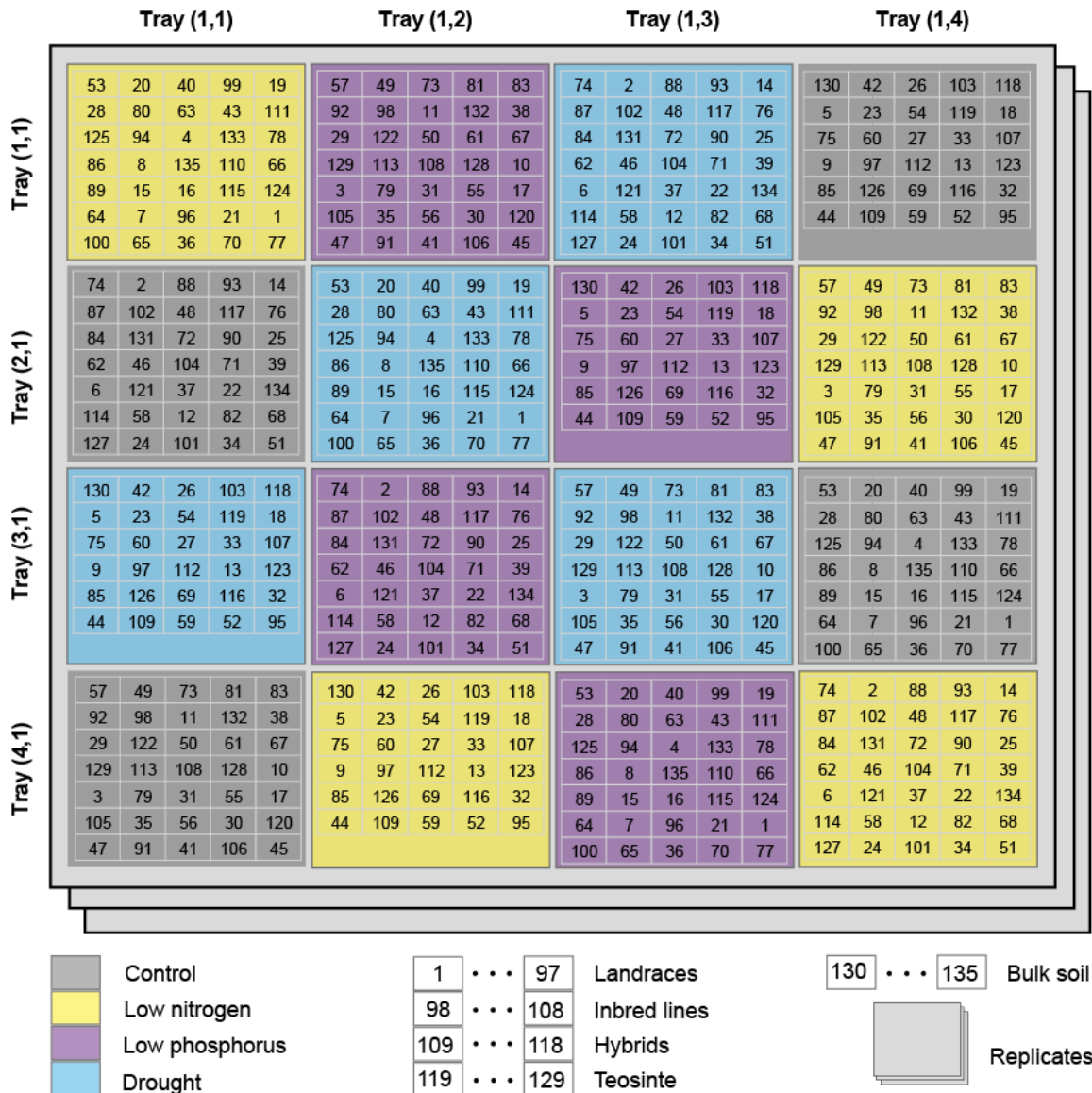
**Supplementary Figure 32. Experimental design.** The experiment was laid out in a split plot design comprising four treatments: (1) fully fertilized control soil, (2) low nitrogen (LN) soil with no nitrogen fertilizer applied, (3) low phosphate (LP) soil with no phosphate fertilizer applied and (4) control soil with drought (D) treatment. Different numbers were allocated to the different genotypes: landraces (1-97), inbred lines (98-108), hybrids (109-118) and teosinte (119-129). Six pots (130-135) without plants were set up as 'bulk soil' samples. Each tray comprises 35 pots with different germplasm groups and bulk soil pots. The pot distribution in each tray was randomized. The whole set-up was repeated three times as three biological replicates.
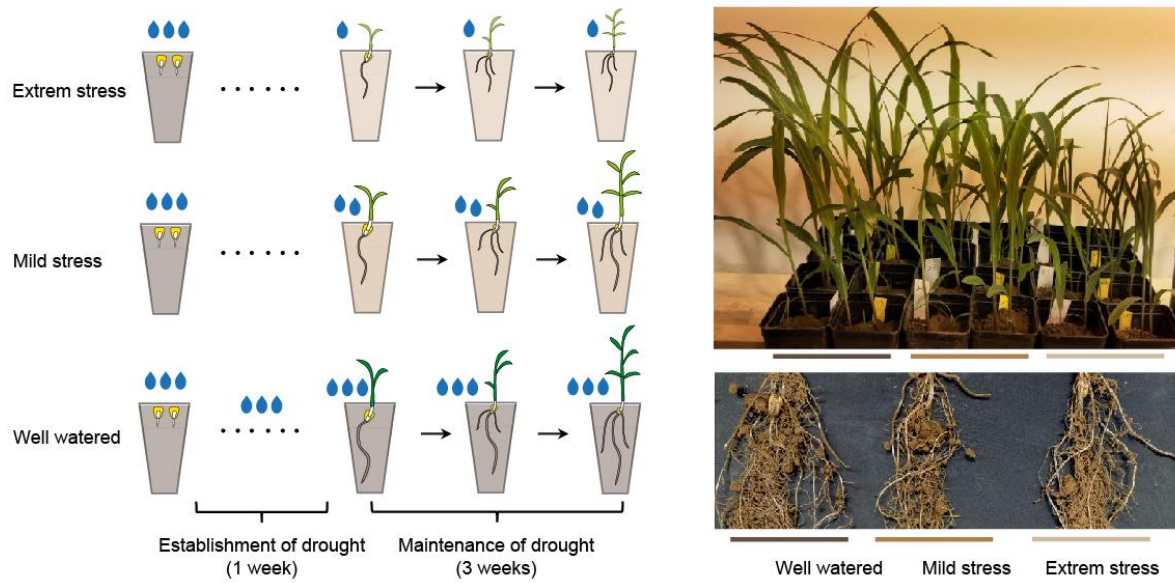
**Supplementary Figure 33. Preliminary drought experiment.** The left panel indicates the three drought experiments with different water levels. The right panel indicates the shoot and root phenotypes grown under three different drought levels.
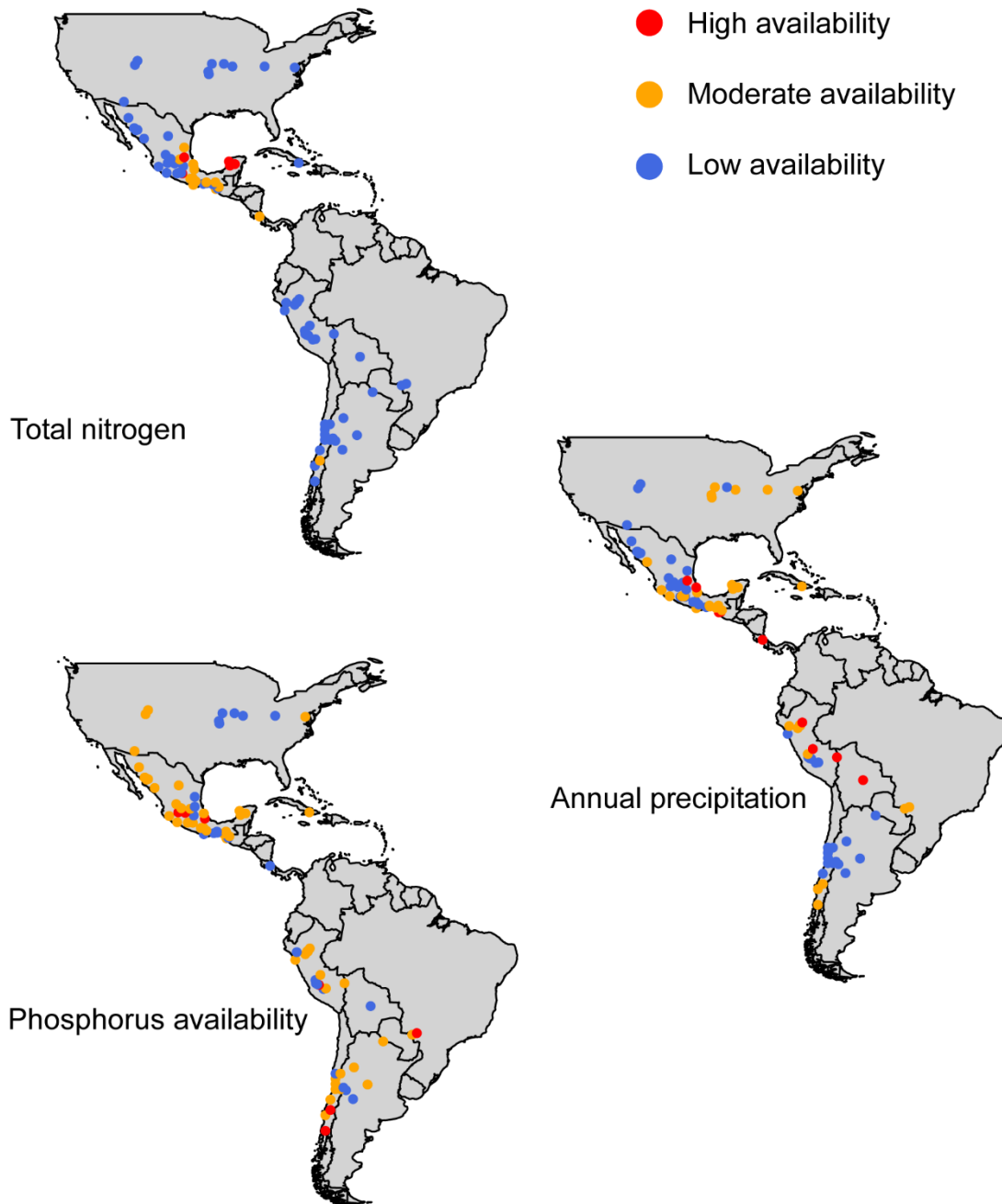
**Supplementary Figure 34. Availability of nutrients and precipitation for landraces collected in their natural habitats.** The original data is provided in Supplementary Dataset 12.
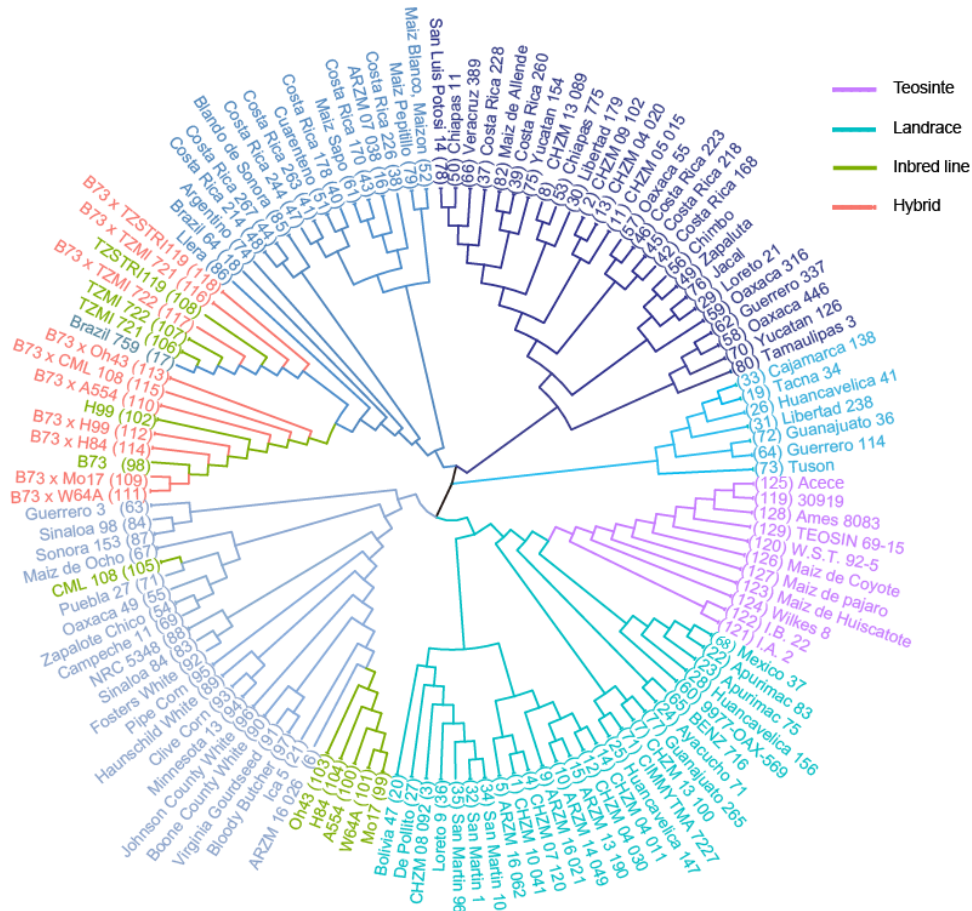
**Supplementary Figure 35. Phylogenetic relationship of the 129 maize genotypes used in this study.** A neighbour-joining tree was produced using SNP markers. Different colours correspond to different germplasm groups. Different shades of blue represent different sources of local landraces.

**Supplementary Table 1 Determination of soil nitrogen, phosphorus, pH and total organic carbon.**

The values presented are mean ± standard error of all samples for three nutrient treatments collected from Dikopshof in 2019. Means followed by the same letter are not significantly different at $p < 0.05$ according to LSD (ANOVA, turkey HSD, $n = 5$). Note that we used the soil from long-term experimental station, and thus the pH value in the low nitrogen soil is a bit lower than control and low phosphorus soil.

| Soil properties | Control | Low nitrogen | Low phosphorus |
|---|---|---|---|
| Total N (%) | 0.094 ± 0.0027a | 0.076 ± 0.00088b | 0.087 ± 0.00053a |
| Conc. P (mg/kg) | 69.23 ± 1.61a | 67.55 ± 1.06a | 24.50 ± 1.39b |
| pH | 6.74 ± 0.08b | 7.23 ± 0.07a | 6.97 ± 0.07ab |
| Total organic carbon (%) | 0.60 ± 0.062a | 0.56 ± 0.043a | 0.69 ± 0.010a |

# 6.2 Supporting figures for Chapter 3
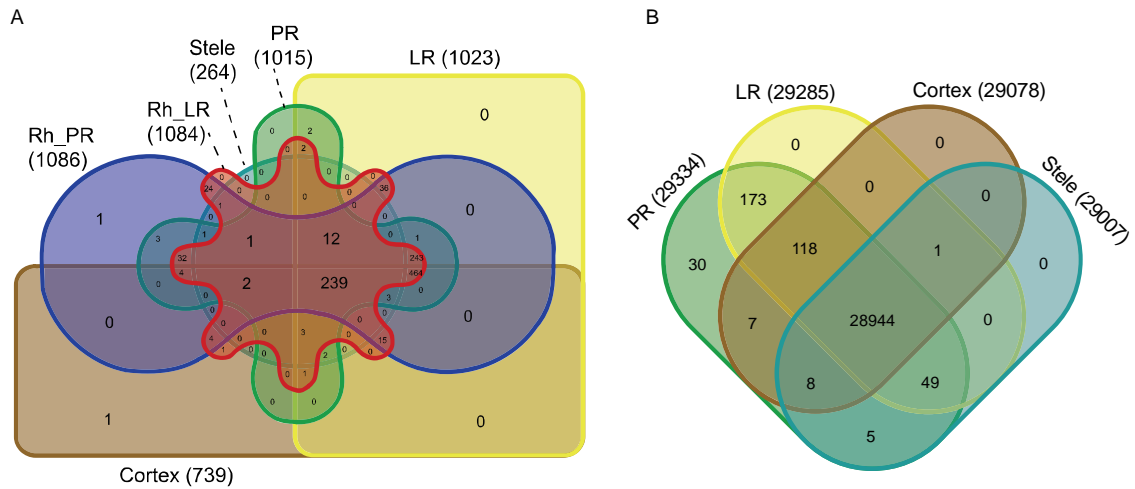


**Fig. S1. Overlapped OTUs (A) and genes (B) among all compartments.** Rh_PR, Rhizosphere from primary roots; Rh_LR, Rhizosphere from lateral roots; PR, Primary roots; LR, Lateral roots. Expressed OTUs are defined as relative abundance >0.1% in at least 5% samples. Expressed genes are defined as reads >5 in at least 4 samples.
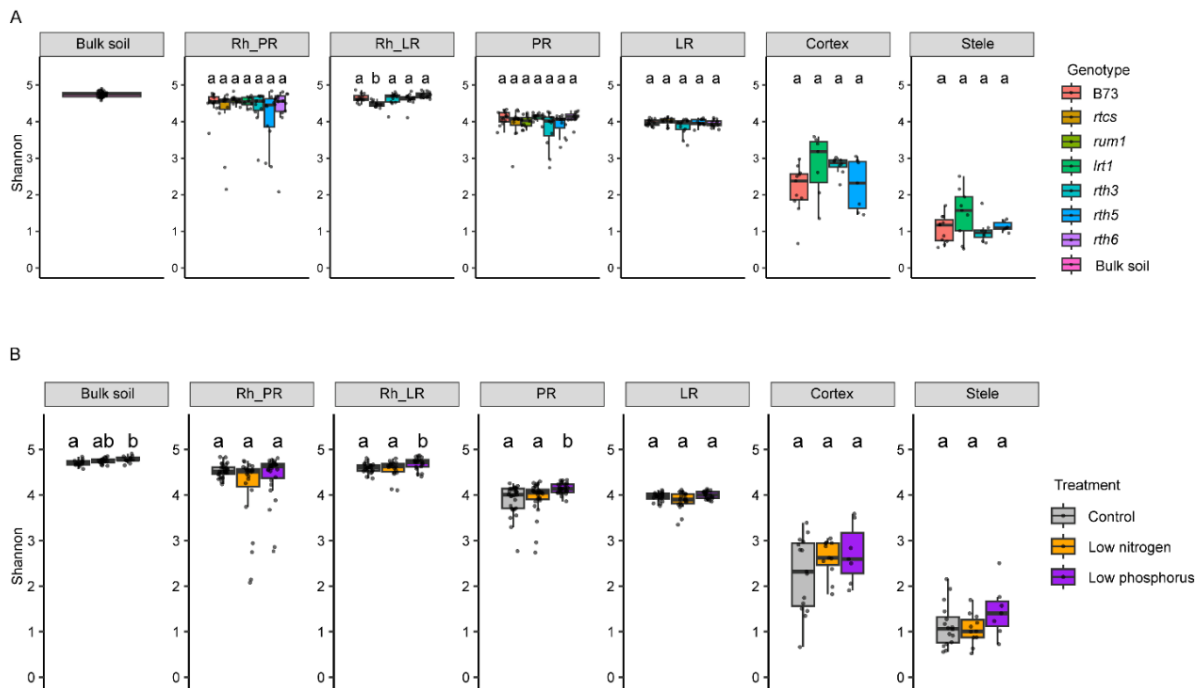


**Fig. S2. Bacterial α-diversity among genotypes (A) and treatments (B) across rhizosphere and root compartments.** α-diversity was estimated by Shannon's diversity index. Compartment significances were calculated using a Kruskal-Wallis test with *post-hoc* Dunn's test (Benjamini-Hochberg adjusted *P* <0.05). Different letters indicate significance among different genotypes or treatments (Benjamini-Hochberg adjusted *P* <0.05). Rh_PR, Rhizosphere from primary root; Rh_LR, Rhizos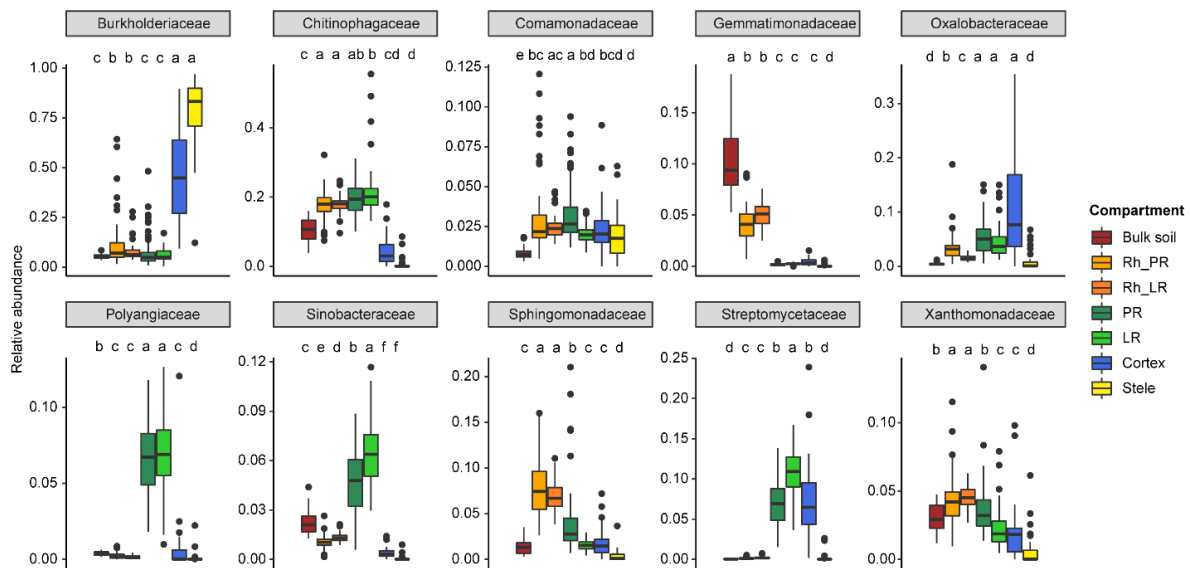phere from lateral root; PR, Primary root; LR, Lateral root. *rum1*, *rootless with undetectable meristem 1*; *rtcs*, *rootless concerning crown and seminal roots*; *lrt1*, *lateral rootless 1*; *rth*, *roothairless*. B73 is the wild type plant. Boxes span from the first to the third quartiles, centre lines represent median values and whiskers show data lying within 1.5× interquartile range of lower and upper quartiles. Data points at the ends of whiskers represent outliers.
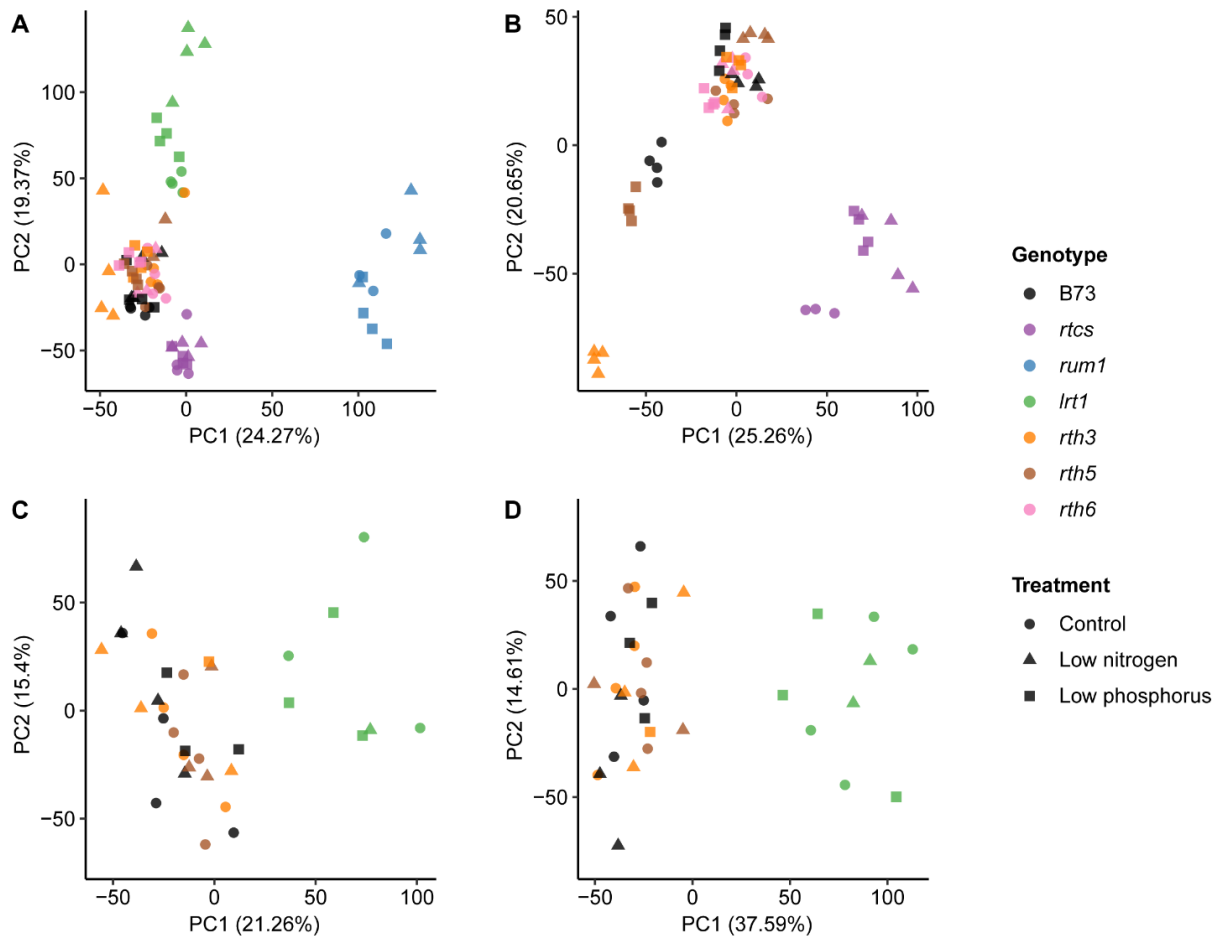
**Fig. S3. Principal component analysis (PCA) illustrating the transcriptomic dissimilarity between genotypes and treatments for each compartment. A**, Primary root; **B**, Lateral root; **C**, Cortex tissue; **D**, Stele. *rum1*, *rootless with undetectable meristem 1*; *rtcs*, *rootless concerning crown and seminal roots*; *lrt1*, *lateral rootless 1*; *rth*, *roothairless*.
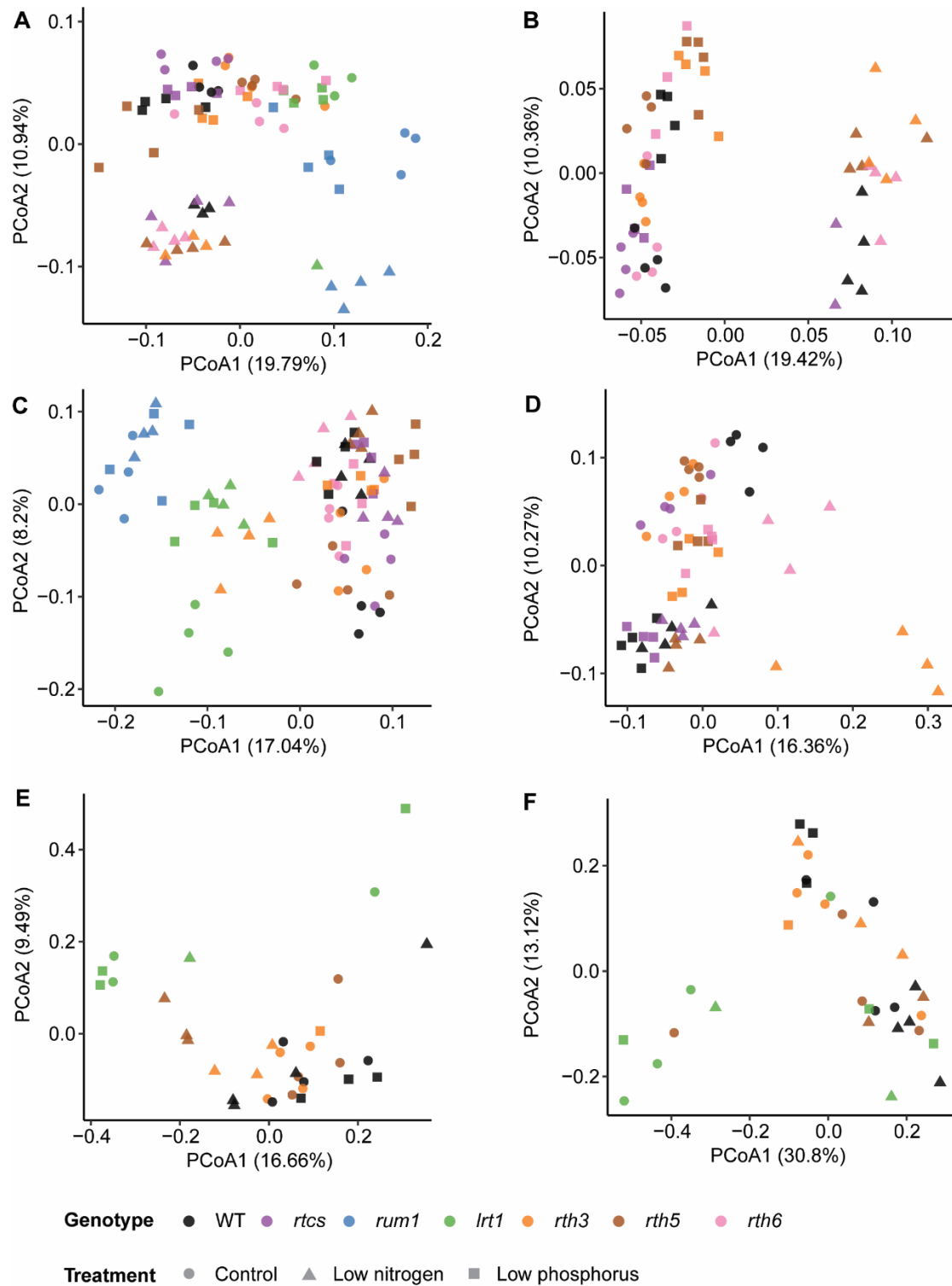
**Fig. S4. Principal coordinate analysis (PCoA) showing the dissimilarity of bacterial β-diversity for each compartment. A**, Rhizosphere from primary root; **B**, Rhizosphere from lateral root; **C**, Primary root; **D**, Lateral root; **E**, Cortex tissue; **F**, Stele tissue. *rum1*, *rootless with undetectable meristem 1*; *rtcs*, *rootless concerning crown and seminal roots*; *lrt1*, *lateral rootless 1*; *rth*, *roothairless*.
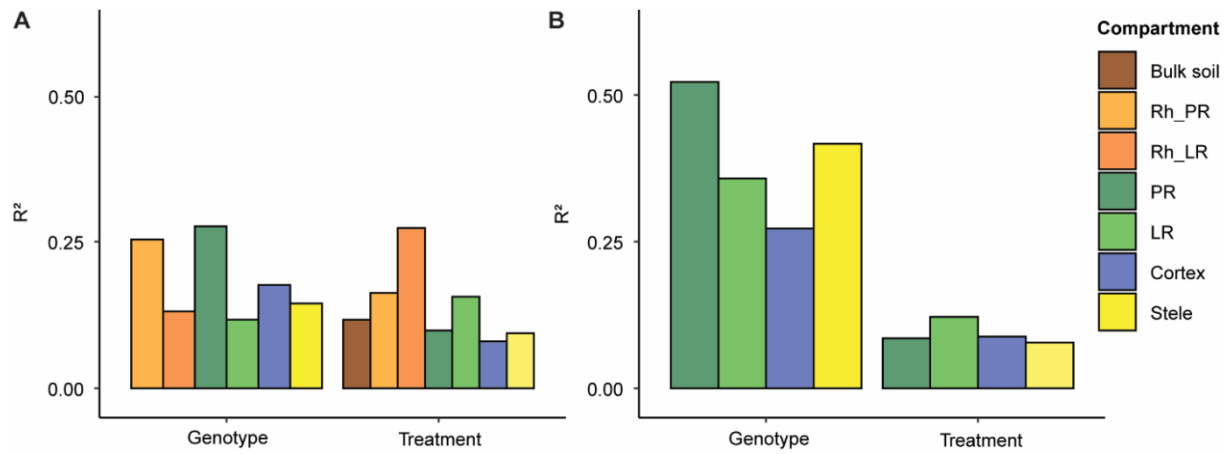
**Fig. S5. PERMANOVA results for PCoA of bacterial community composition and PCA of gene expression.** Rh_PR, Rhizosphere from primary root; Rh_LR, Rhizosphere from lateral root; PR, Primary root; LR, Lateral root.
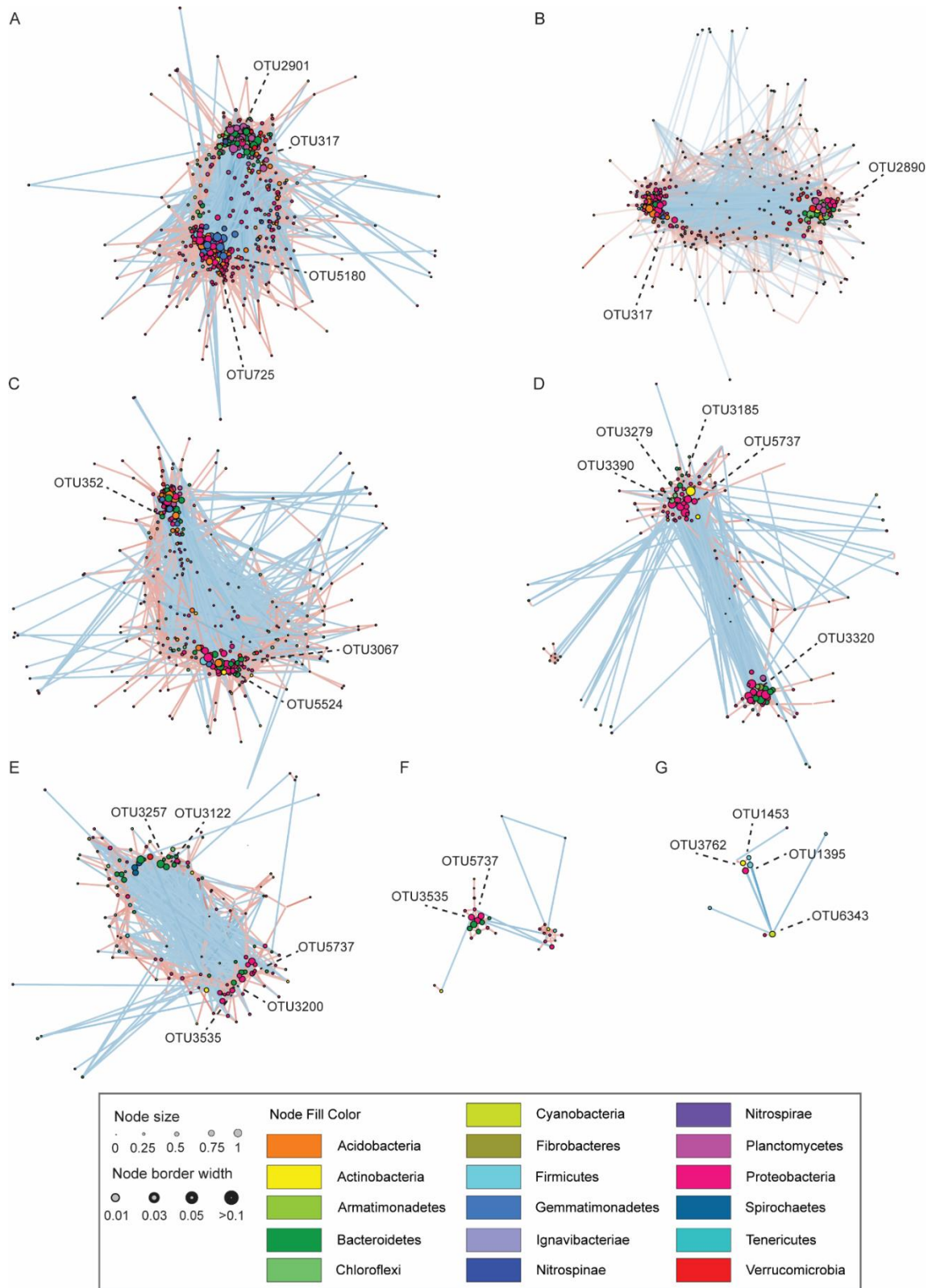
**Fig. S6. OTU-OTU co-occurrence network in soil (A), rhizosphere of primary root (B), rhizosphere of lateral root (C), primary root (D), lateral roots (E), cortex (F) and stele (G).** Nodes color represents phylum, node size is proportional to hub score and node border width is proportional to mean relative abundance. Key OTUs are labeled by OTU id. Red and blue solid lines indicate positive and negative correlations respectively.
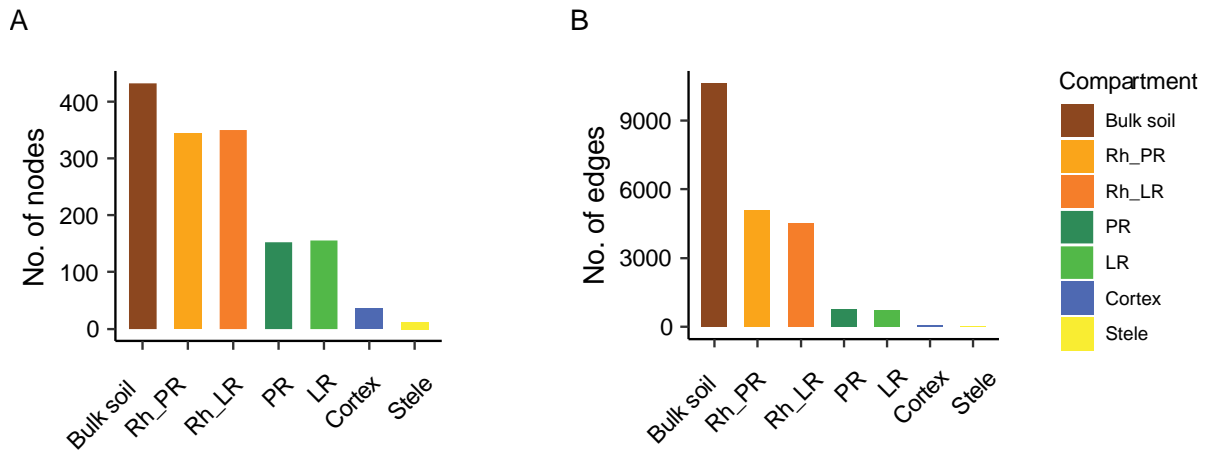
**Fig. S7**. **Number of nodes and edges of the OTU-OTU SparCC network within each compartment.** Rh_PR, Rhizosphere from primary root; Rh_LR, Rhizosphere from lateral root; PR, Primary root; LR, Lateral root.



**Fig. S8. Relative abundance of the top ten enriched families across different compartments.** Rh_PR, Rhizosphere from primary root; Rh_LR, Rhizosphere from lateral root; PR, Primary root; LR, Lateral root. Significances were indicated among different compartments by different letters for each family (Benjamini-Hochberg adjusted $P < 0.05$, Kruskal-Wallis test, Dunn's *post-hoc* test).
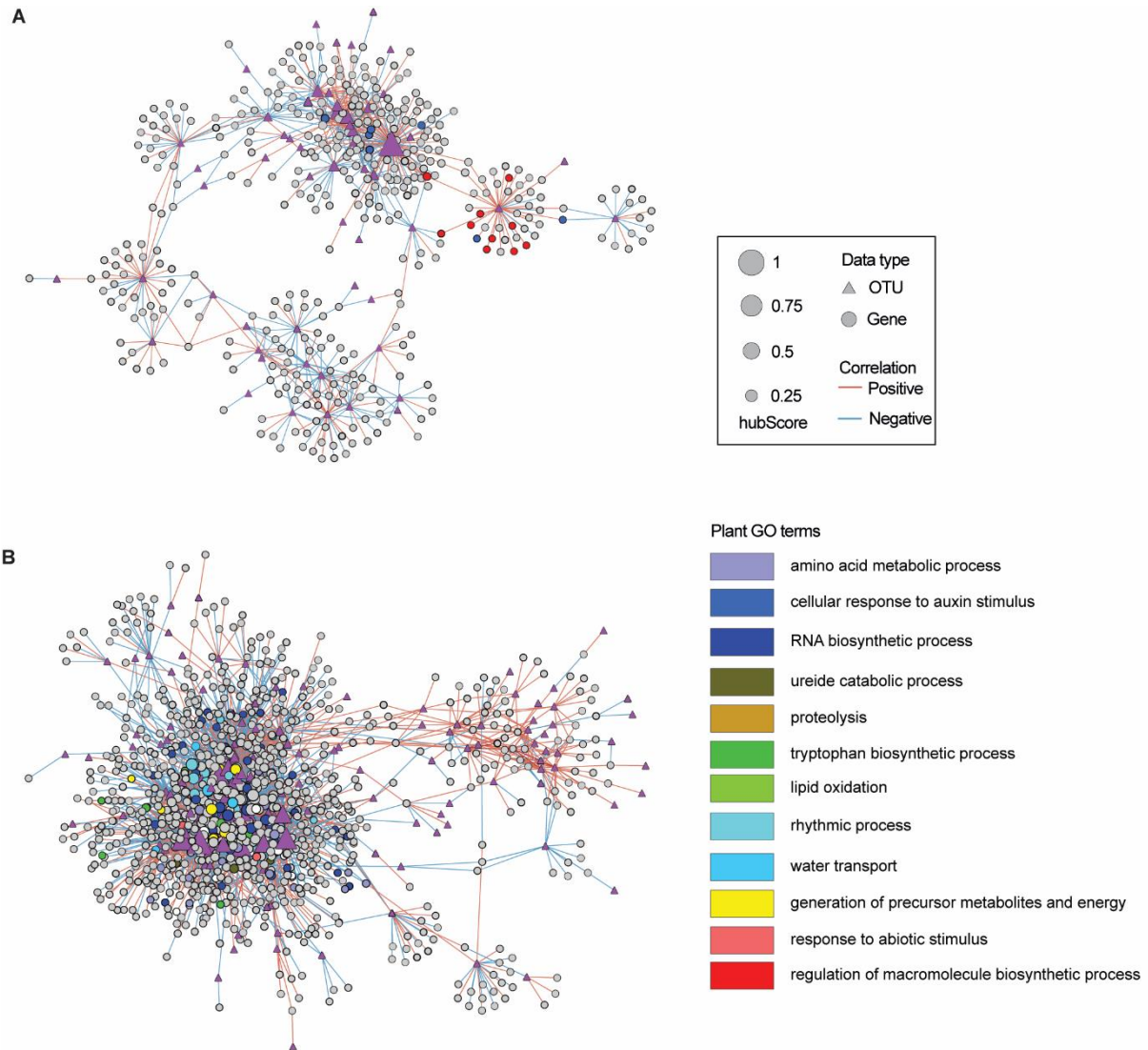
**Fig. S9. Network associations between plant genes and microbial OTUs in the rhizosphere from primary root (A) and lateral root (B).** The triangles and dots indicated the bacterial OTUs and gene features respectively. The size of the circles indicates the hub score. Red and blue solid lines indicate positive and negative correlations respectively. Only the genes with significant plant gene ontology (GO) terms connected with hub OTUs are colored accordingly.
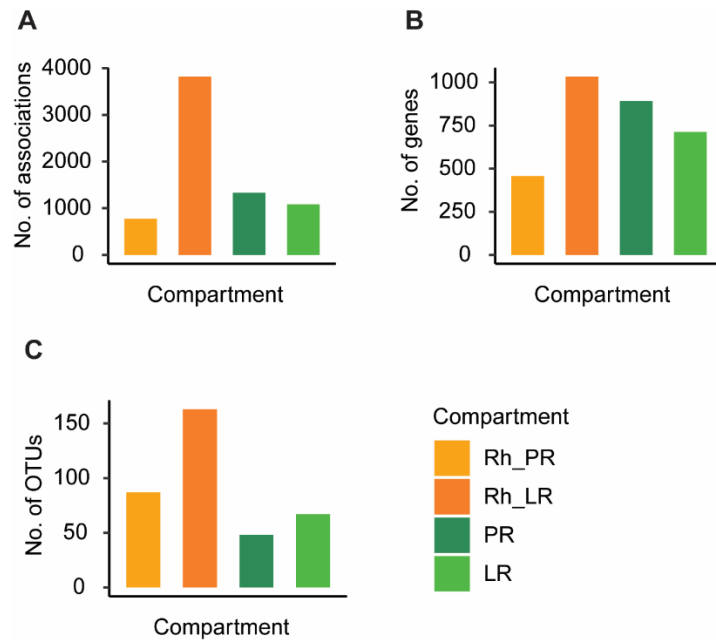
**Fig. S10. Number of edges, genes, and OTUs for each OTU-gene network.** Rh_PR, Rhizosphere from primary root; Rh_LR, Rhizosphere from lateral root; PR, Primary root; LR, Lateral root.
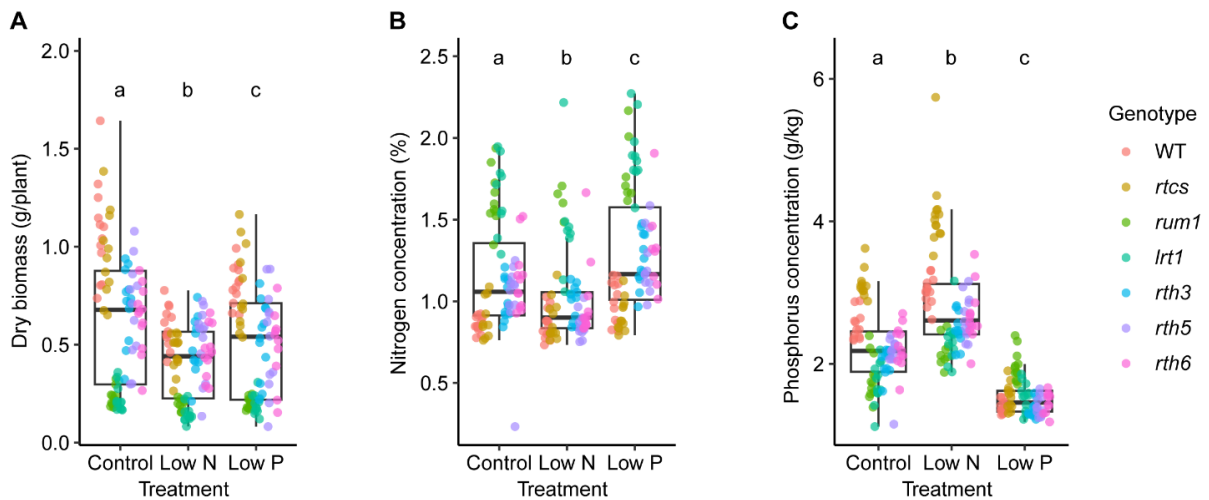


**Fig. S11. Maize phenotypic traits under different treatments. A**, Shoot dry biomass; **B**, Nitrogen concentration; **C**, Phosphorus concentration. N, nitrogen; P, phosphorus. *rum1*, *rootless with undetectable meristem 1*; *rtcs*, *rootless concerning crown and seminal roots*; *lrt1*, *lateral rootless 1*; *rth*, *roothairless*.
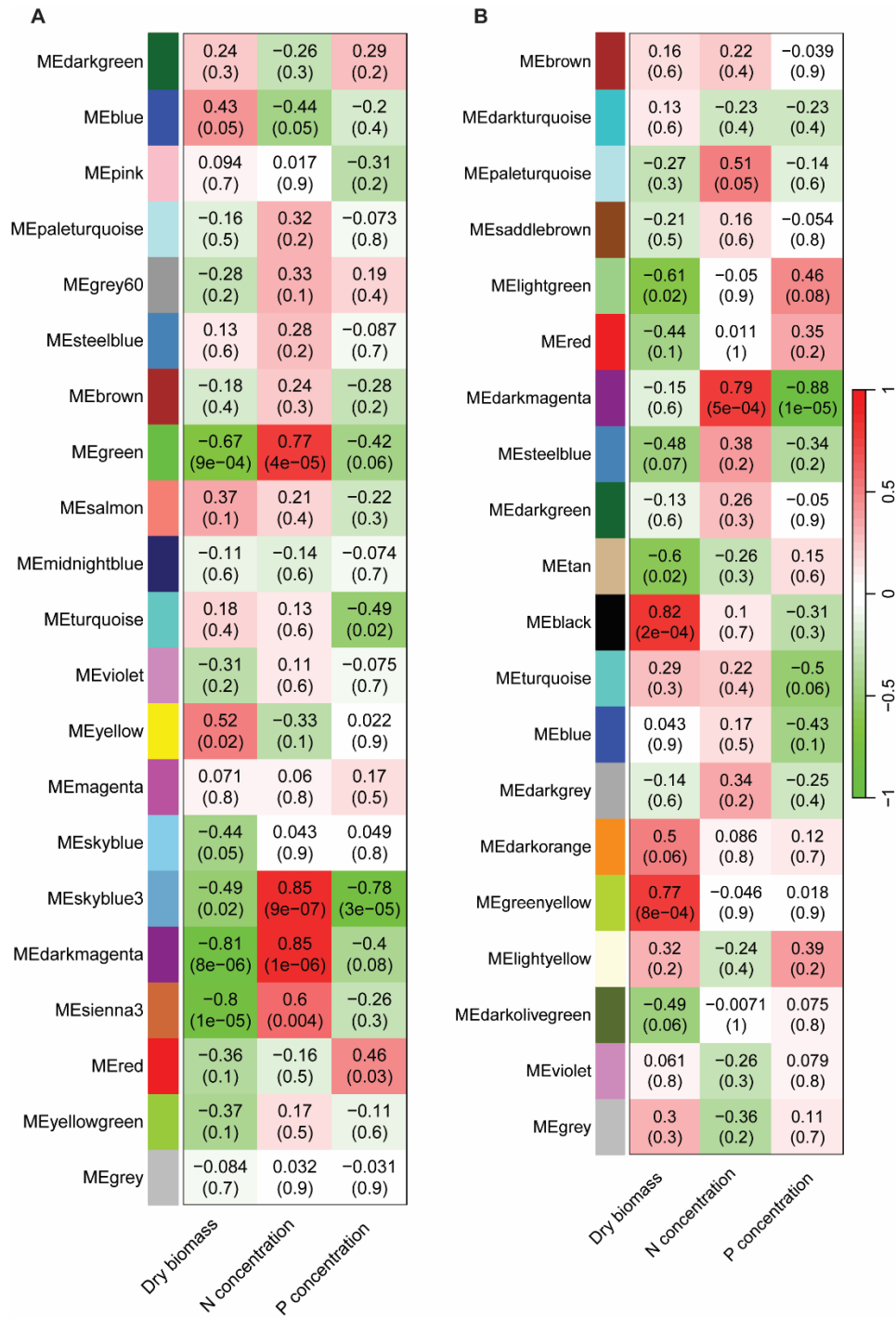
**Fig. S12. Gene WGCNA modules and their correlations with plant traits. A**, Primary root; **B**, Lateral root. The color bar represents the correlation coefficient. Correlation coefficient and its P-value in bracket are also displayed on the heatmap.
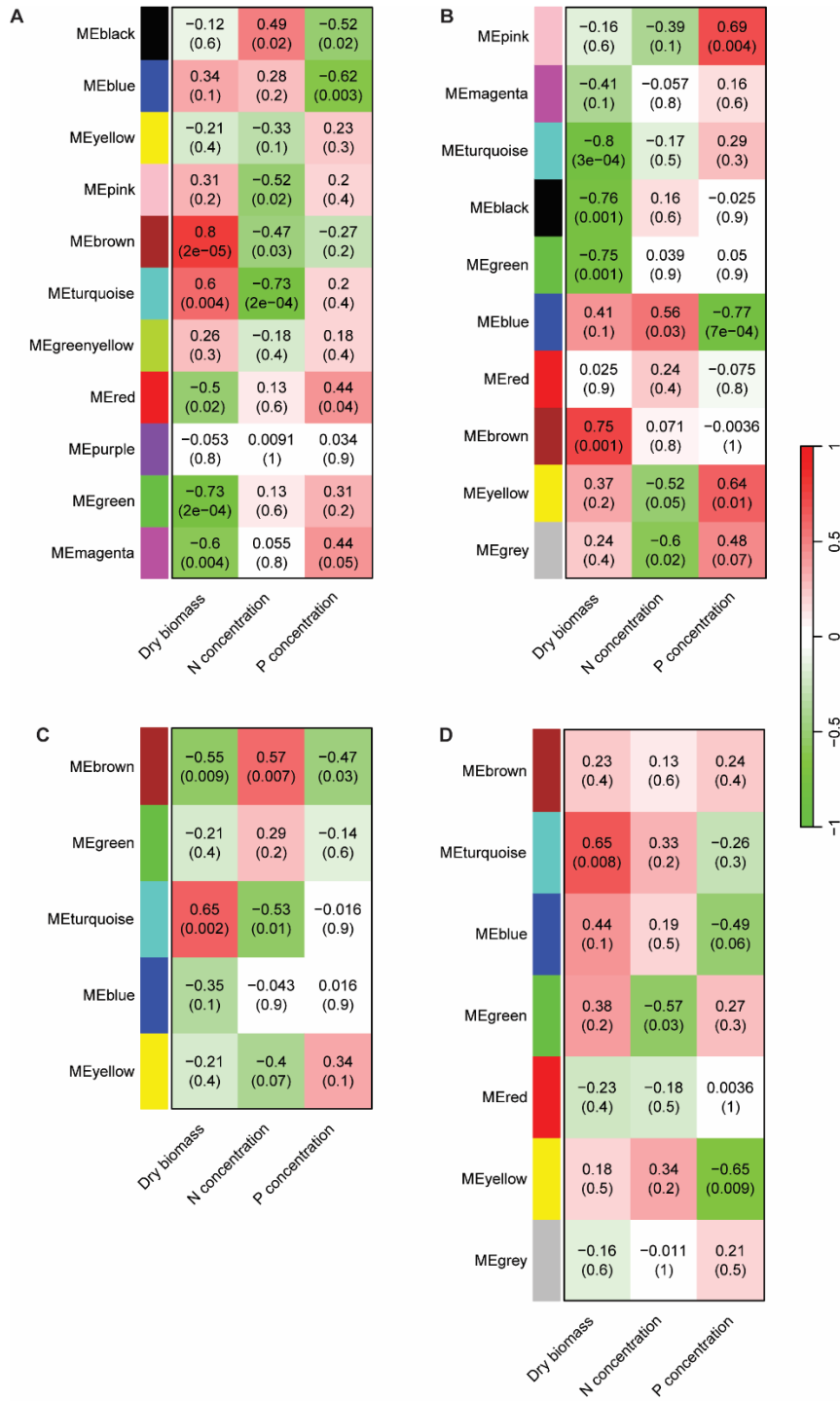
**Fig. S13. Bacterial WGCNA modules and their correlations with plant traits. A**, Rhizosphere from primary roots; **B**, Rhizosphere from lateral roots; **C**, Primary roots; **D**, Lateral roots. The color bar represents the correlation coefficient. Correlation coefficient and its P-value in bracket are also displayed on the heatmap.
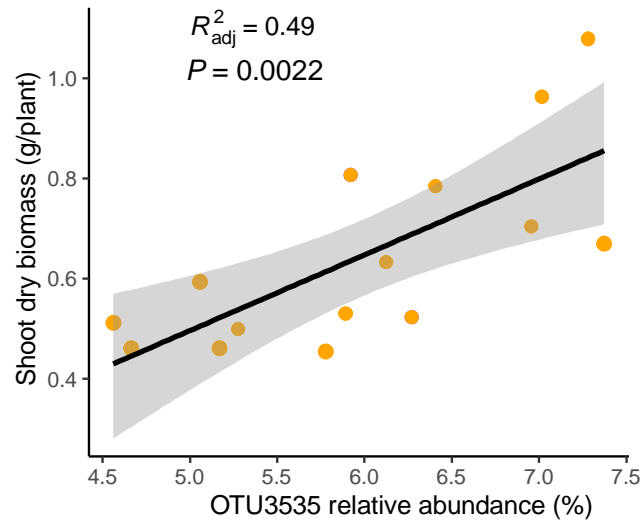
**Fig. S14. Linear correlation between shoot dry biomass and OTU3535 relative abundance (%).** Linear model was fitted using lm() function.
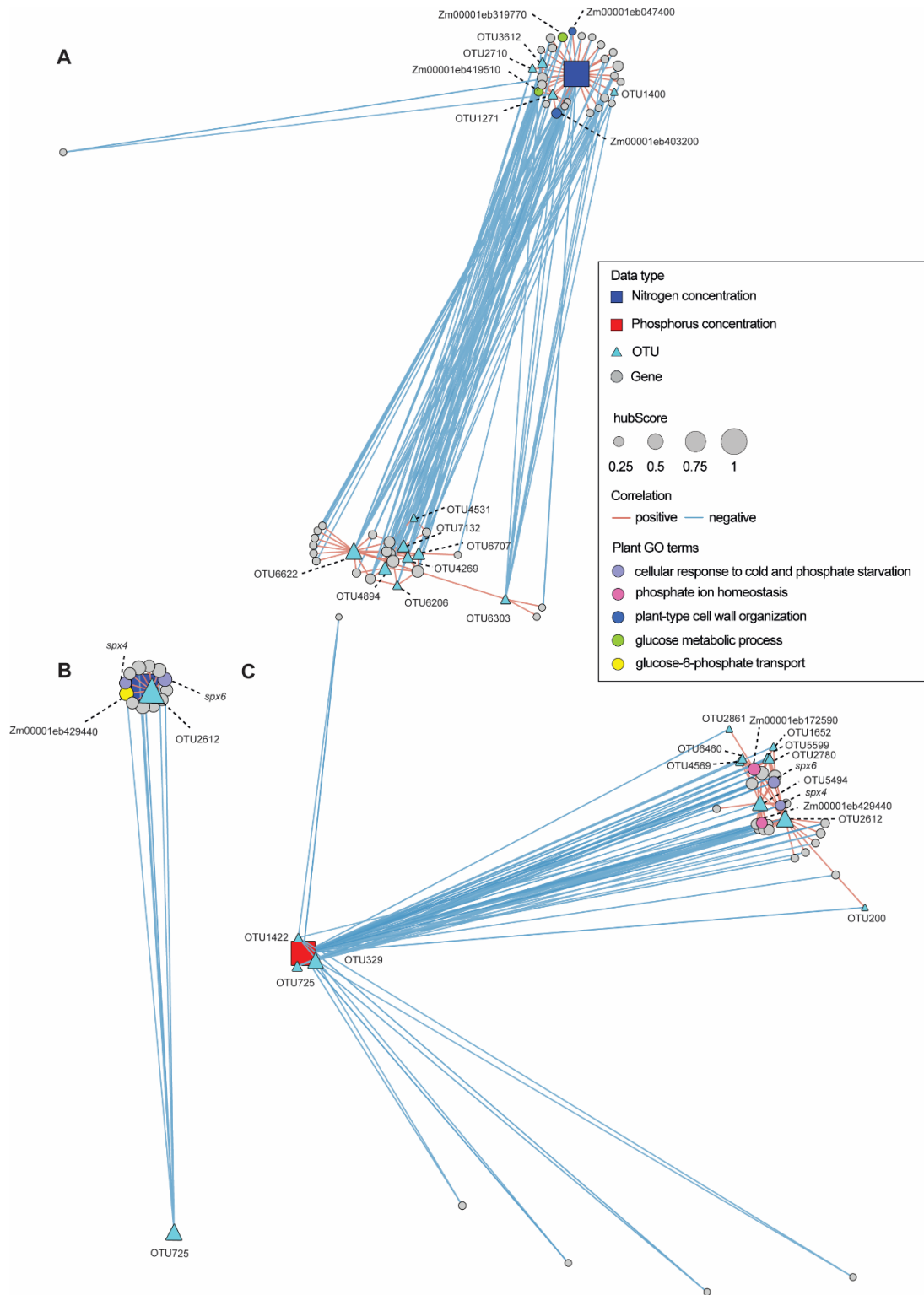
**Fig. S15. Trans-kingdom interaction network between bacterial OTUs and root genes in association with plant nutrients concentration. A**, Root gene expression, bacterial OTUs in the rhizosphere from primary root and nitrogen concentration; **B**, Root gene expression, bacterial OTUs in the rhizosphere from lateral root and nitrogen concentration; **C**, Root gene expression, bacterial OTUs in the rhizosphere from lateral root and phosphorus concentration. The triangles, dots and squares indicated the bacterial OTUs, gene features and plant phenotypic traits respectively. The size of the nodes indicates the hub score. Red and blue solid lines indicate positive and negative correlations respectively. Only the genes with significant plant gene ontology (GO) terms connected with hub OTUs are colored accordingly.

A



B
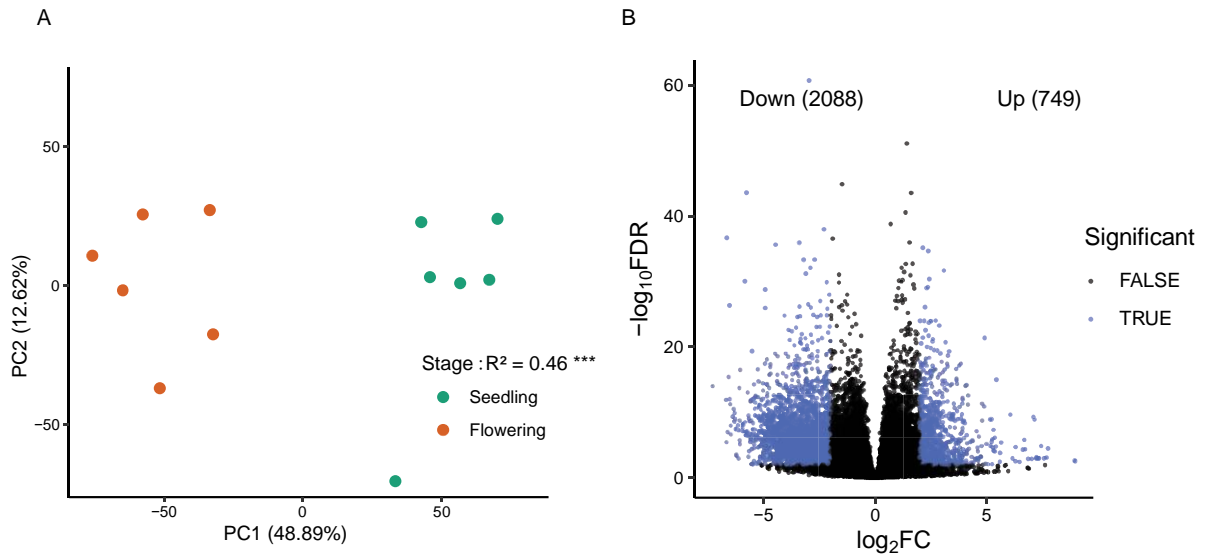


**Fig. S16. PCA plot and differentially expressed genes between flowering stage and seedling stage.** PERMANOVA test was performed to calculate the variance explained by stage in gene expression (permutations = 1999). Differentially expressed genes were determined by setting absolute value of $\log_2$Foldchange $>2$ and FDR adjusted $P <0.01$, colored in blue points.

# 7 Publications

## 7.1 Publications related to this thesis

- **Heritable microbiome variation is correlated with source environment in locally adapted maize varieties.**

  Xiaoming He[#], **Danning Wang**[#], Yong Jiang[#], Meng Li[#], Manuel Delgado-Baquerizo[#], Chloee McLaughlin, Caroline Marcon, Li Guo, Marcel Baer, Yudelsy A.T. Moya, Nicolaus von Wirén, Marion Deichmann, Gabriel Schaaf, Hans-Peter Piepho, Zhikai Yang, Jinliang Yang, Bunlong Yim, Kornelia Smalla, Sofie Goormachtig, Franciska T. de Vries, Hubert Hüging, Mareike Baer, Ruairidh J. H. Sawers, Jochen C. Reif, Frank Hochholdinger, Xinping Chen, Peng Yu. (#: co-first authors).

  *Nature Plants,* 2024, 10, 598–617. DOI: https://doi.org/10.1038/s41477-024-01654-7. The published version is reproduced in Chapter 2 of the thesis.

  **Own contribution**: I processed and mapped the 16S rRNA gene reads and conducted the downstream analyses of the microbiome data, performed the statistical analyses and interpreted the related results. I wrote the part of manuscript for which I performed the analyses, and designed and plotted the figures.

- **Enrichment of the bacterial taxon *Massilia* in lateral roots is associated with flowering in maize.**

  **Danning Wang**, Xiaoming He, Marcel Baer, Klea Lami, Baogang Yu, Alberto Tassinari, Silvio Salvi, Gabriel Schaaf, Frank Hochholdinger, and Peng Yu. *Microbiome* (2023). (In revision).

  The submitted version is reproduced in Chapter 3 of the thesis.

  **Own contribution**: I processed and mapped the 16S rRNA gene reads, RNA-seq reads and conducted all downstream analyses of the microbiome data, transcriptome data, network integration analysis, statistical analyses and interpreted the results. I wrote the whole manuscript, and designed and plotted the figures.

## 7.2 Publications unrelated to this thesis

- **Root system adaptation to water availability during maize domestication and global expansion.**

  Peng Yu[#], Chunhui Li[#], Meng Li[#], Xiaoming He[#], **Danning Wang**, Hongjie Li, Caroline Marcon, Yu Li, Sergio Perez-Limón, Xinping Chen, Manuel Delgado-Baquerizo, Robert Koller, Ralf Metzner, Dagmar van Dusschoten, Ljudmilla Borisjuk,

Iaroslav Plutenko, Audrey Mahon, Marcio F.R. Resende Jr., Silvio Salvi, Asegidew Akale, Mohanned Abdalla, Mutez Ali Ahmed, Felix Maximilian Bauer, Andrea Schnepf, Guillaume Lobet, Adrien Heymans, Kiran Suresh, Lukas Schreiber, Chloee M. McLaughlin, Chunjian Li, Manfred Mayer, Chris-Carolin Schön, Vivian Bernau, Nicolaus von Wirén, Ruairidh J. H. Sawers, Tianyu Wang, Frank Hochholdinger. *Nature Genetics* (2024). (Accepted).

- **Epidermis-specific transcriptomic responses reveal that cold mediates root hair plasticity of maize via *dreb2.1*.**
  Yaping Zhou, Annika Meyer, **<u>Danning Wang</u>**, Alina Klaus, Tyll Stöcker, Caroline Marcon, Heiko Schoof, Georg Haberer, Chris-Carolin Schön, Peng Yu, and Frank Hochholdinger.
  *Plant Physiology* (2023). (Submitted).

## 7.3 Presentations at conferences

- **Eco-evolutionary signature of root–microbiome association in maize.**
  International Conference of the German Society for Plant Sciences. Aug 28 – Sep 1, 2022. Bonn, Germany. (*Poster presentation*).

- **Maize domestication contributes to microbiome-driven root branching and nitrogen stress resilience.**
  10th International Symposium on Root Development. May 15 – 18, 2023.
  Ghent, Belgium. (*Oral presentation*).

# 8 Acknowledgement