



**HAL**  
open science

# The Genetic Architecture of Adaptation to Leaf and Root Bacterial Microbiota in *Arabidopsis thaliana*

Fabrice Roux, Léa Frachon, Claudia Bartoli

► **To cite this version:**

Fabrice Roux, Léa Frachon, Claudia Bartoli. The Genetic Architecture of Adaptation to Leaf and Root Bacterial Microbiota in *Arabidopsis thaliana*. *Molecular Biology and Evolution*, 2023, 40 (5), pp.msad093. 10.1093/molbev/msad093 . hal-03806731v1

**HAL Id: hal-03806731**



**<https://hal.science/hal-03806731v1>**

Submitted on 12 Sep 2023 (v1), last revised 11 Jul 2023 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The Genetic Architecture of Adaptation to Leaf and Root Bacterial Microbiota in *Arabidopsis thaliana*

Fabrice Roux <sup>\*,1</sup> Léa Frachon <sup>1,2</sup> and Claudia Bartoli<sup>1,3</sup>

<sup>1</sup>Fabrice Roux, Laboratoire des Interactions Plantes-Microbes-Environnement, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, CNRS, Université de Toulouse, Castanet-Tolosan, France

<sup>2</sup>Department of Systematic and Evolutionary Botany, University of Zurich, Zürich, Switzerland

<sup>3</sup>Institute for Genetics, Environment and Plant Protection (IGEPP), INRAE, Institut Agro AgroCampus Ouest, Université de Rennes 1, Le Rheu, France

\*Corresponding author: E-mail: [fabrice.roux@inrae.fr](mailto:fabrice.roux@inrae.fr).

Associate editor: Emily Josephs

## Abstract

Understanding the role of the host genome in modulating microbiota variation is a need to shed light on the holobiont theory and overcome the current limits on the description of host-microbiota interactions at the genomic and molecular levels. However, the host genetic architecture structuring microbiota is only partly described in plants. In addition, most association genetic studies on microbiota are often carried out outside the native habitats where the host evolves and the identification of signatures of local adaptation on the candidate genes has been overlooked. To fill these gaps and dissect the genetic architecture driving adaptive plant-microbiota interactions, we adopted a genome-environment association (GEA) analysis on 141 whole-genome sequenced natural populations of *Arabidopsis thaliana* characterized in situ for their leaf and root bacterial communities in fall and spring, and a large range of nonmicrobial ecological factors (i.e., climate, soil, and plant communities). A much higher fraction of among-population microbiota variance was explained by the host genetics than by nonmicrobial ecological factors. Importantly, the relative importance of host genetics and nonmicrobial ecological factors in explaining the presence of particular operational taxonomic units (OTUs) differs between bacterial families and genera. In addition, the polygenic architecture of adaptation to bacterial communities was highly flexible between plant compartments and seasons. Relatedly, signatures of local adaptation were stronger on quantitative trait loci (QTLs) of the root microbiota in spring. Finally, plant immunity appears as a major source of adaptive genetic variation structuring bacterial assemblages in *A. thaliana*.

**Key words:** plant-microbiota interactions, pathogens, genome-environment association, community ecology, local adaptation.

## Introduction

To cope with the growing global demand for food supplies and the need to reduce reliance on pesticides, a more eco-efficient, sustainable, and environmentally friendly agriculture is required (Keating et al. 2010). In the global change context, both crops and wild plant species face extreme and largely unpredictable abiotic stresses (such as heat waves) as well as an increase in the number and severity of epidemics (Bebber 2015; Desaint et al. 2021). Altogether, this calls for concrete interventions to improve the potential of plants to cope with multiple abiotic and biotic stresses.

The plant microbiota is defined as a set of microorganisms of a particular host compartment (i.e., rhizosphere, roots, stem, leaves, and flowers). Often referred to as the second host genome in the context of the holobiont/hologenome theory (Rosenberg and Zilber-Rosenberg 2018), the plant microbiota mainly originates from the soil

compartment (Bulgarelli et al. 2013; Bai et al. 2022), even if a large fraction of microbes of the phyllosphere can originate from the aerial sphere (Müller et al. 2016). Plant-associated microbes are crucial for plant health because they 1) mobilize and make accessible essential nutrients (e.g., nitrogen and phosphate), 2) provide resistance to abiotic stresses (such as drought), and 3) confer direct (production of antimicrobial components) or indirect (elicitation of immune defense) pathogen protection (Berendsen et al. 2012; Bulgarelli et al. 2013; Pieterse et al. 2014; Jacoby et al. 2017; Escudero-Martinez and Bulgarelli 2019; Trivedi et al. 2020; Glick and Gamalero 2021; Bai et al. 2022). The plant microbiota is therefore a promising lever to develop innovative eco-friendly agroecosystems (Busby et al. 2017; Toju et al. 2018; Mitter et al. 2019).

While numerous studies reported the strong influence of abiotic (i.e., climate and physicochemical agronomic

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Open Access

properties) (Müller et al. 2016; Fitzpatrick et al. 2020) and biotic (i.e., presence of herbivores and neighboring plants) (Humphrey and Whiteman 2020; Kong et al. 2021; Santos-García et al. 2020; Meyer et al. 2022) factors on plant microbiota diversity and composition, there was a growing interest during the last decade to estimate the effect of plant genetics on microbiota (Bergelson, Brachi, et al. 2021). Two main approaches have been adopted to unravel the genetic and molecular plant mechanisms controlling microbiota assembly. The first and most common approach is based on the use of artificial mutations, including mutant and transgenic lines (Bergelson, Brachi, et al. 2021). By testing 218 artificial genetic lines in 48 studies conducted on crops and wild species, several pathways affecting microbial assembly were identified and include 1) external and internal physical barriers in both the leaf (e.g., wax and cuticle) and root (e.g., suberin and lignin) compartments (Salas-González et al. 2021), 2) pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) that prevents dysbiosis by keeping commensal microbes at a low absolute abundance (Chen et al. 2020), 3) hormonal pathways related to salicylic acid, jasmonic acid, ethylene, and strigolactones (Lebeis et al. 2015; Hou et al. 2021), 4) mineral nutrient homeostasis (Zhang et al. 2019), which may require a fine coordination with physical barriers (Salas-González et al. 2021) and immunity (Castrillo et al. 2017), 5) excretion of plant secondary metabolites in rhizosphere, roots, or flowers (Huang et al. 2019), and 6) symbiosis (Wang et al. 2021).

The second approach exploits natural genetic variation segregating among or within plant species (Bergelson, Brachi, et al. 2021). While the importance of host genetics in shaping natural variation in microbial communities has been a long-lasting debate (Roux and Bergelson 2016), an ever-increasing number of studies reported significant microbiota differences between closely related species or among genotypes within a given species (when grown in the same environment). The typical values of these significant genotype effects within plant species on microbiota composition range from 5% to 30% (Schlaeppli et al. 2014; Bergelson, Brachi, et al. 2021). Following the detection of significant heritability estimates, seven genome-wide association studies (GWAS) conducted in common gardens and using microbial community descriptors as plant traits, have been performed in *Arabidopsis thaliana* (Horton et al. 2014; Bergelson et al. 2019; Brachi et al. 2022), maize (Walters et al. 2018), rice (Roman-Reyna et al. 2020), sorghum (Deng et al. 2021), and switchgrass (VanWallendael et al. 2022). These GWAS revealed a highly polygenic architecture, suggesting a control of natural microbiota assembly by an extensive number of quantitative trait loci (QTLs) with a small effect. A similar result was recently obtained with traditional linkage mapping performed in barley (Escudero-Martinez et al. 2022) and tomato (Oyserman et al. 2022).

While informative, the number of genetic association studies that report signatures of local adaptation on QTLs associated with microbial communities remains scarce

(Brachi et al. 2022). In addition, because the relative effect of host genetics on the microbiota can highly depend on the plant habitat, in particular the inoculum source (e.g., agricultural soil) (Robertson-Albertyn et al. 2017; Hubbard et al. 2018; Fabiańska et al. 2020), setting up genetic association studies in a single common garden can generate partial conclusions on the host genetics controlling for microbiota (Oyserman et al. 2021). One complementary approach to tackle these issues is to conduct genome-environment association (GEA) analysis. With the goal of identifying genetic variants associated with ecological variation across tens to hundreds of natural populations, GEA analysis is a powerful genome scan method to identify genes potentially involved in adaptive processes (De Mita et al. 2013). The development of next-generation sequencing (NGS) technologies combined with the availability of public databases on abiotic variables, in particular climatic variables, resulted in a recent burst of GEA studies reporting in plants the identification of adaptive QTLs to abiotic variation, from a worldwide scale (Hancock et al. 2011; Lasky et al. 2015; Bay et al. 2017; López-Hernández and Cortés 2019) to a regional scale (Pluess et al. 2016; Frachon et al. 2018). For biotic factors, GEA analysis requires to controlling for the abiotic environment that can be correlated with biotic variation (Frachon et al. 2019). While this fundamental step of removing the effect of habitat filtering increases the rate of GEA signals only associated with biotic variations, it also allows identifying the abiotic factors that are the main predictors of the biotic variation (Frachon et al. 2019). Although GEA analysis on microbiota has been largely adopted in humans and wild animal species (Blekhnman et al. 2015; Gomez et al. 2017; Weissbrod et al. 2018; Suzuki et al. 2019; Bubier et al. 2021; Ryu and Davenport 2022), GEA analysis on any type of biotic interactions remains in its infancy in wild plant species (Frachon et al. 2023). For instance, a GEA analysis on plant community descriptors revealed a high degree of biotic specialization of *A. thaliana* to members of its plant interaction network at the genetic level (Frachon et al. 2019). In addition, the genetic architecture of local adaptation to plant community diversity and composition was not predictable from the genetic architecture of local adaptation to the abundance of individual plant companion species (Frachon et al. 2019). Similar conclusions were reached when growing a GWA mapping population of *A. thaliana* in the presence of bispecific and plurispecific interactions in a common garden (Libourel et al. 2021).

In this study, by combining microbial community ecology and population genomics, we adopted a GEA approach to establish a genomic map of local adaptation to bacterial communities of the leaf and root compartment of 141 whole-genome sequenced natural populations of *A. thaliana* located southwest of France (Bartoli et al. 2018; Frachon et al. 2018). Because plant-associated microbes rapidly change within the host life cycle (Copeland et al. 2015; Beilsmith et al. 2021), bacterial communities were characterized in fall and spring, thereby allowing testing whether the strength of local adaptation to bacterial

communities differs between seasons (Bartoli et al. 2018). We also tested whether the strength of local adaptation differs between microbiota and potential pathobiota (i.e., the ensemble of potential phytopathogens). To control for the confounding effects of the abiotic environment on microbiota and potential pathobiota, the 141 natural populations of *A. thaliana* were characterized for a set of 17 biologically meaningful climate and soil variables (Frachon et al. 2018; Frachon et al. 2019). Because companion species can strongly shape the microbial communities of a focal plant species (Geremia et al. 2016; Meyer et al. 2022), we also controlled for the confounding effects of 49 plant community descriptors (Frachon et al. 2019). Finally, we cross-validated our study by testing for a significant overlap between the lists of candidate genes identified by our GEA analysis and the lists of candidate genes identified by two independent published common garden experiments conducted on two GWA mapping populations of *A. thaliana* phenotyped for host-microbiota interactions.

## Results and Discussion

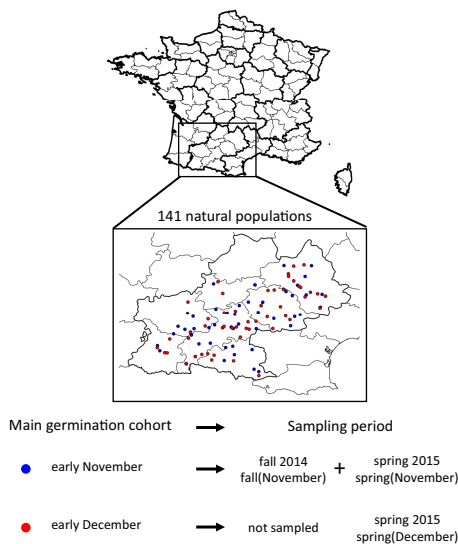
### Working with a Large Number of Nonmicrobial Ecological Factors to Disentangle Adaptive Genetic Variants Associated with Microbiota Descriptors

In this study, we focused on 141 natural populations of *A. thaliana* inhabiting contrasting ecological habitats in the southwest of France (fig. 1A) (Bartoli et al. 2018; Frachon et al. 2018). Importantly, the 141 populations

strongly differed in their main germination cohort in fall 2014 (early November vs. early December) (fig. 1A) (Bartoli et al. 2018). We, therefore, defined three seasonal groups for sampling, hereafter named 1) “fall (November)” corresponding to 73 populations sampled in November/December 2014, 2) “spring (November)” corresponding to 72 populations already sampled in fall and resampled in early spring (February/March 2015), and 3) “spring (December)” corresponding to 66 populations only sampled in early spring (February/March 2015) (fig. 1A, table 1, and supplementary table S1, Supplementary Material online).

The 141 populations have been previously whole-genome sequenced using a Pool-Seq approach (resulting in the identification of 4,781,661 single nucleotide polymorphisms [SNPs]) and characterized for a set of 6 climate variables, 14 soil physicochemical variables, and 49 descriptors of plant communities (fig. 1 and supplementary tables S1–3, Supplementary Material online, Data Sets 1–6) (Frachon et al. 2018; Frachon et al. 2019). These natural populations of *A. thaliana* have been also characterized for the leaf and root bacterial communities of on average four plants per “population × seasonal group” combination (fig. 1C). For this, we used a metabarcoding approach based on a fragment of the *gyrB* gene, which has a deeper taxonomic resolution than fragments of the 16S rRNA gene (Barret et al. 2015). This metabarcoding approach led to the identification of 278,333 operational taxonomic units (OTUs), a number that was further reduced to 6,627 OTUs after two filtering steps while still representing

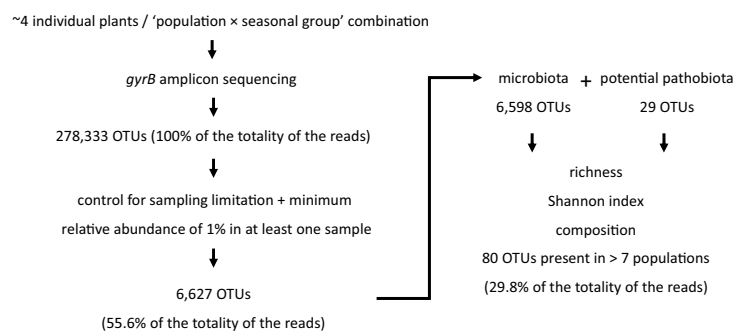
#### A Map of the natural populations and sampling period



#### B Genomic characterization (Frachon et al. 2018)

- Pool-Seq approach based on ~16-17 plants per population
- Raw allele frequencies per population for 4,781,661 SNPs

#### C Characterization of bacterial communities (Bartoli et al. 2018)



#### D Characterization of non-microbial ecological factors

- 6 uncorrelated climate variables (two temperature-related variables and four precipitation-related variables) (Frachon et al. 2018)
- 14 soil physico-chemical variables describing the main soil agronomic properties related to plant growth (Frachon et al. 2019)
- 49 descriptors of plant communities (richness, Shannon index, composition, density of *A. thaliana*, presence/absence of the most prevalent companion plant species) (Frachon et al. 2019)

**Fig. 1.** Inputs for GEA analysis of microbiota/potential pathobiota. (A) Location of the 141 *A. thaliana* populations in the Midi-Pyrénées region (southwest of France). Blue and red dots correspond to populations with a main germination cohort early November and early December, respectively. (B) Genomic characterization of the 141 *A. thaliana* populations using a Pool-Seq approach. (C) Characterization of microbiota and potential pathobiota of the 141 *A. thaliana* populations. (D) Nonmicrobial ecological characterization of the 141 *A. thaliana* populations.

55.6% of the microbiota (fig. 1C and supplementary table S2, Supplementary Material online). By allowing to distinguish bacterial OTUs at the species level, the *gyrB* marker facilitated the identification of 29 OTUs belonging to the potential pathobiota (fig. 1C and supplementary table S2, Supplementary Material online) (Bartoli et al. 2018). The pathogenic behaviour of the potential pathobiota was determined, among others, by 1) the presence of disease symptoms on both host and nonhost plants in response to most of the strains isolated from the three OTUs composing ~75% of the potential pathobiota (Bartoli et al. 2018), 2) the absence of disease symptoms on *A. thaliana* in response to the most abundant OTUs of the microbiota (Ramírez-Sánchez, Gibelin-Viala, et al. 2022), and 3) the relative abundance of the potential pathobiota being significantly three times higher in *A. thaliana* plants with visible disease symptoms than in asymptomatic *A. thaliana* plants when sampled in situ (Bartoli et al. 2018).

For each “plant compartment × seasonal group” combination, we focused, for both microbiota and potential pathobiota, on descriptors of community diversity (richness and Shannon index) and community composition (approximated by the two first principal components from a simple unconstrained principal coordinate analysis [PCoA]), which largely and significantly differ among the natural populations of *A. thaliana* considered in this study (Bartoli et al. 2018). Descriptors of community diversity and composition also significantly differed between the leaf and root compartments, albeit this difference was highly dependent on the natural population of *A. thaliana* considered (Bartoli et al. 2018). In addition, we focused on the presence/absence of the most prevalent OTUs (representing a total relative abundance of ~30% of the microbiota and potential pathobiota, fig. 1C). Altogether, a total of 194 descriptors of bacterial communities were submitted to GEA analysis (table 1, Data Sets 1–6). To fine map QTLs associated with descriptors of microbiota and potential pathobiota, we combined a Bayesian hierarchical model (BHM) that decreases the rate of false genotype-ecology positive associations by making the analyses robust to complex demographic histories

(Gautier 2015), with a local score (LS) approach allowing the detection of QTLs with small effects (Bonhomme et al. 2019). The SNPs underlying the QTLs identified by the combined BHM-LS approach are hereafter named top SNPs. The efficiency of the LS approach was demonstrated in GWAS conducted in *A. thaliana*, with the fine mapping down to the gene level and the functional validation of four QTLs associated with quantitative disease resistance to the bacterial pathogen *Ralstonia solanacearum* (Aoun et al. 2020; Demirjian et al. 2022). In addition, the LS approach was successfully applied in a recent GWAS on leaf bacterial communities characterized on 200 Swedish accessions of *A. thaliana* grown in four native habitats in Sweden (Brachi et al. 2022).

### GEA Revealed a Polygenetic Architecture of Adaptation to Bacterial Communities

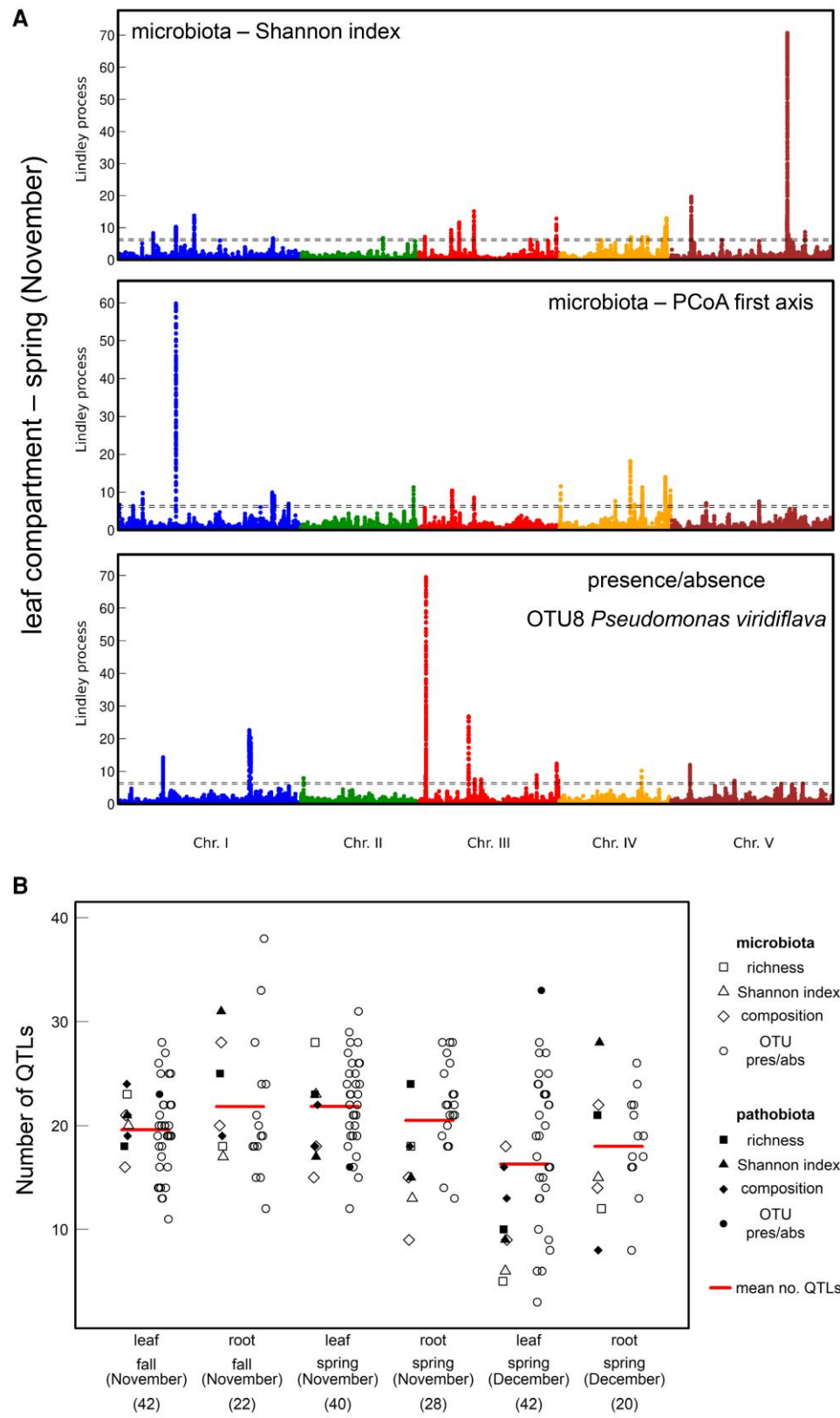
For each “plant compartment × seasonal group” combination, our GEA approach successfully detected QTLs associated with the diversity and composition of bacterial communities and the presence/absence of a particular OTU (supplementary fig. S1, Supplementary Material online). For instance, a clearly defined association peak (i.e., the presence of tens of top SNPs in a genomic region of few kbs) was detected at the end of chromosome 5 and the beginning of chromosome 1 for variation in microbiota diversity and composition in the leaf compartment of the “spring (November)” seasonal group, respectively (fig. 2A). Similarly, a clearly defined association peak was detected at the beginning of chromosome 3 for the presence/absence of *Pseudomonas viridiflava*, one of the most prevalent and abundant bacterial pathogens identified in natural populations of *A. thaliana* in several geographical regions (Karasov et al. 2014, 2018; Bartoli et al. 2018) (fig. 2A).

Overall, our study revealed a highly polygenetic architecture (i.e., number of detected QTLs > 10) for 178 out of the 194 descriptors of bacterial communities (fig. 2B and supplementary table S4, Supplementary Material online). The detection of on average ~19.6 QTLs per descriptor (median = 20, min = 3, and max = 38) (fig. 2B) is in line with the polygenetic architecture reported in GWAS

**Table 1.** Number of Bacterial Community Descriptors Investigated with GEA Analysis for Each of the Six “Plant Compartment × Seasonal Group” Combinations.

Category	Community descriptors	Fall (November)		Spring (November)		Spring (December)	
		Leaf (n = 73)	Root (n = 73)	Leaf (n = 69)	Root (n = 72)	Leaf (n = 66)	Root (n = 66)
Microbiota	Richness	1	1	1	1	1	1
	Shannon index	1	1	1	1	1	1
	Composition (PCoA axis)	2	2	2	2	2	2
	Presence/absence OTUs	33	15	31	21	33	13
Pathobiota	Richness	1	1	1	1	1	1
	Shannon index	1	1	1	1	1	1
	Composition (PCoA axis)	2	1	2	1	2	1
	Presence/absence OTUs	1	0	1	0	1	0

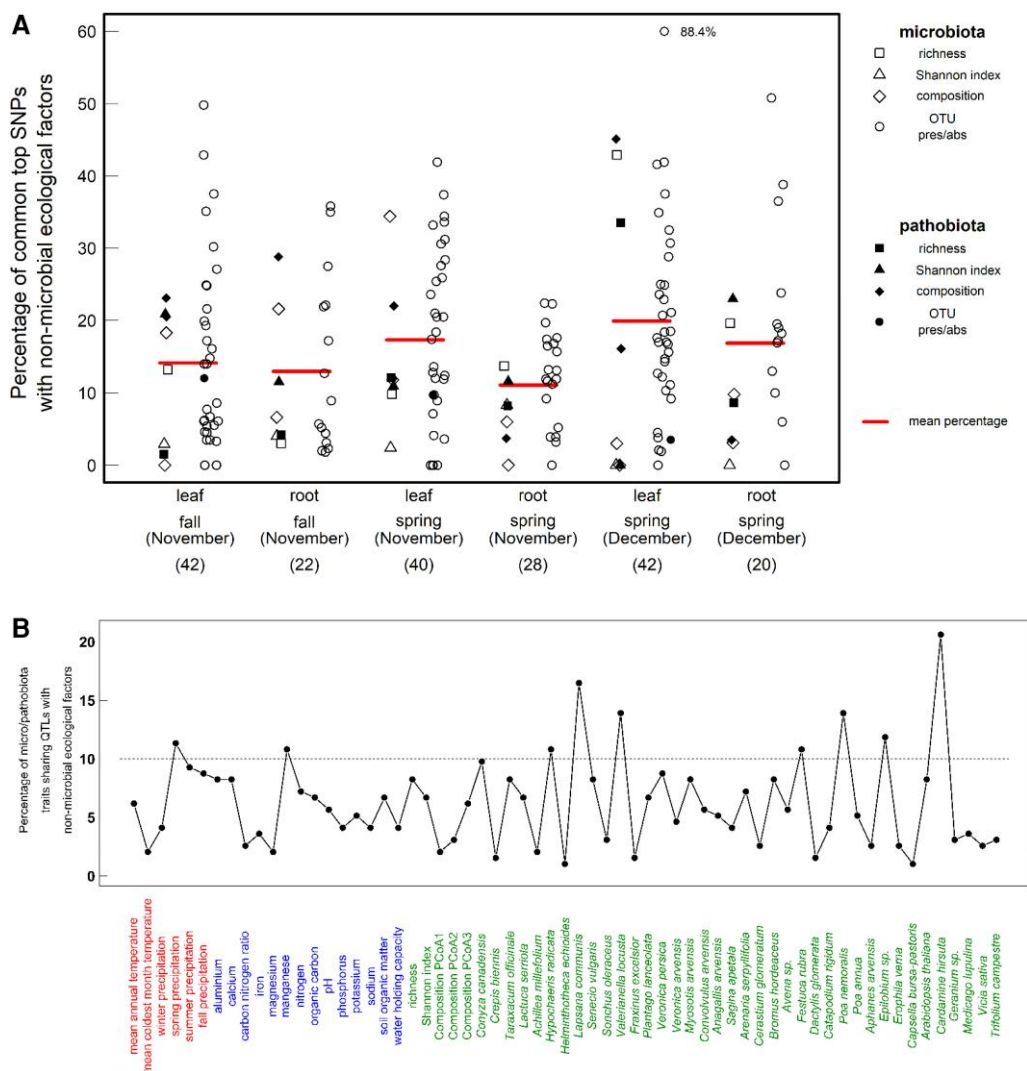
Month names in brackets correspond to the main germination cohort of the populations considered. Numbers in brackets for the leaf and root compartment correspond to the number of populations. For community composition, only PCoA axes exhibiting significant variation among populations were considered in this study.



**Fig. 2.** The genetic architecture of microbiota/potential pathobiota. (A) Manhattan plots illustrating the power of a BHM-LS approach to describe the genetic architecture of microbiota/potential pathobiota diversity, microbiota/potential pathobiota composition, and the presence/absence of a particular OTU. The x-axis indicates along the five chromosomes, the physical position of the 1,470,777 SNPs considered for the leaf compartment in the “spring (November)” seasonal group. The y-axis corresponds to the values of the Lindley process corresponding to the  $-\log_{10}(P)$  of the tuning parameter  $\zeta = 2$ . The dashed lines indicate the minimum and maximum of the five chromosome-wide significance thresholds. (B) Jitter plots illustrating the diversity in the number of QTLs among microbiota/potential pathobiota traits within each “plant compartment  $\times$  seasonal group” combination. Month names in brackets correspond to the main germination cohort of the populations considered. Numbers in brackets correspond to the number of descriptors of microbiota and potential pathobiota considered for each “plant compartment  $\times$  seasonal group” combination.

conducted on microbial communities (Horton et al. 2014; Walters et al. 2018; Bergelson et al. 2019; Roman-Reyna et al. 2020; Deng et al. 2021; Brachi et al. 2022; VanWallendael et al. 2022). Differences in the number of QTLs among the six “plant compartment  $\times$  seasons’

were weakly significant (generalized linear model [GLM],  $F = 2.30$ ,  $P = 0.0471$ ). No differences in the number of QTLs were found between microbiota and potential pathobiota descriptors (GLM,  $F = 0.50$ ,  $P = 0.8063$ ) (fig. 2B).



**Fig. 3.** Interactions between descriptors of microbiota/potential pathobiota and nonmicrobial ecological factors at the genetic level. (A) Percentage of top SNPs associated with microbiota/potential pathobiota descriptors that are common to top SNPs associated with nonmicrobial ecological variables (including climate variables, soil physicochemical properties, and descriptors of plant communities) for each “plant compartment × seasonal group” combination. Month names in brackets correspond to the main germination cohort of the populations considered. Numbers in brackets correspond to the number of descriptors of microbiota and potential pathobiota considered for each “plant compartment × seasonal group” combination. (B) Percentage of microbiota/potential pathobiota descriptors ( $n = 194$ ) sharing QTLs with each nonmicrobial ecological variable across the six “plant compartment × seasonal group” combinations. Climate variables, soil physicochemical properties, and descriptors of plant communities are represented in red, blue, and green, respectively. The dashed black line corresponds to the threshold of 19 microbiota/potential pathobiota traits, that is,  $\sim 10\%$  of the total number of microbiota/potential pathobiota traits.

### A Non-Negligible Fraction of QTLs of Microbiota/Potential Pathobiota Was Associated with Variation in Abiotic Environment and Plant Communities

To disentangle GEA signals for a given bacterial community descriptor from GEA signals for nonmicrobial ecological factors, we applied our BHM-LS approach on 69 climate variables, soil physicochemical properties, and plant community descriptors (supplementary tables S1 & 3, Supplementary Material online, Data Set 1–6). A non-negligible fraction of top SNPs associated with descriptors of microbiota and potential pathobiota was common to variation in abiotic environment and plant communities. On average,  $\sim 15.7\%$  of the top SNPs associated with descriptors of microbiota and potential pathobiota were

also associated with nonmicrobial ecological variables (median = 13.2%, min = 0%, and max = 84%) (fig. 3A). We may, however, caution that these values certainly represent lower-bound estimates due to the undefined number of nonmicrobial ecological variables that have not been characterized on the 141 natural populations of *A. thaliana* used in this study. For instance, while a correlation study suggested drought metrics as the strongest predictors of leaf microbiota composition of *A. thaliana* at the European continental scale (Karasov et al. 2022), an experimental study demonstrated that insect herbivory can reshape the leaf microbiota assemblage of the Brassicaceae species *Cardamine cordifolia* (Humphrey and Whiteman 2020). In addition, while soil physicochemical properties

and plant community descriptors considered in our study were directly assessed from sampling within the 141 populations during the sampling period for microbiota characterization (Frachon et al. 2019), the six uncorrelated climatic variables considered in this study were averaged over the 2003–2013 period (i.e., 10 years before the sampling period for microbiota characterization) from a public database at a grid resolution of ~600 months (Frachon et al. 2018). These mean climatic values might be not representative of the local weather experienced by the plants between fall 2014 and spring 2015.

Seven out of the nine nonmicrobial ecological variables sharing top SNPs with more than 10% of descriptors of microbiota and potential pathobiota correspond to the presence/absence of companion plant species (fig. 3B), which is in line with the neighborhood effects (also known as associational effects) on microbial transmission (Worrlich et al. 2019; Meyer et al. 2022), particularly well documented for bacterial pathogens (Parker et al. 2015; Makiola et al. 2022). For instance, a clearly defined association peak located on chromosome 3 was common between the presence/absence of OTU6—recently identified as *Pseudomonas siliginis* (Ramírez-Sánchez, Gibelin-Viala, et al. 2022)—in the root compartment of the “fall (November)” seasonal group and the presence/absence of *Cardamine hirsuta*, closely related to *A. thaliana* (Hay and Tsiantis 2016) (fig. 4A). Another example is a clearly defined association peak located on chromosome 1 being common between the composition in bacterial pathogens in the leaf compartment of the “fall (November)” seasonal group and the richness in companion plant species (fig. 4B).

In agreement with the impact of precipitation and drought on the microbiota of phyllosphere (Zhu et al. 2022), the main climate variables sharing top SNPs with descriptors of microbiota and potential pathobiota correspond to precipitations (fig. 3B). For instance, a clearly defined association peak located on chromosome 1 was common between the presence/absence of OTU3—recently identified as a new species of the *Oxalobacteraceae* family (Ramírez-Sánchez, Gibelin-Viala, et al. 2022)—in the leaf compartment of the “fall (November)” seasonal group and precipitation in fall (fig. 4C). On the other hand, manganese and calcium concentrations are the main soil physicochemical properties sharing top SNPs with descriptors of microbiota and potential pathobiota (fig. 3B). For instance, microbiota composition in the leaf compartment in the “spring (November)” seasonal group and soil calcium concentration shared a clearly defined association peak located on chromosome 5 (fig. 4D). Soil calcium concentration was found to impact bacterial community structures both in soils (Neal and Glendinning 2019; Tang et al. 2019) and in plants (Li et al. 2018; Mittelstrass et al. 2021), whereas the soil manganese concentration was already suggested as a main driver of root microbial communities in *A. thaliana* at the continental scale (Thiergart et al. 2020).

For each seasonal group, the percentage of common SNPs between descriptors microbiota/potential

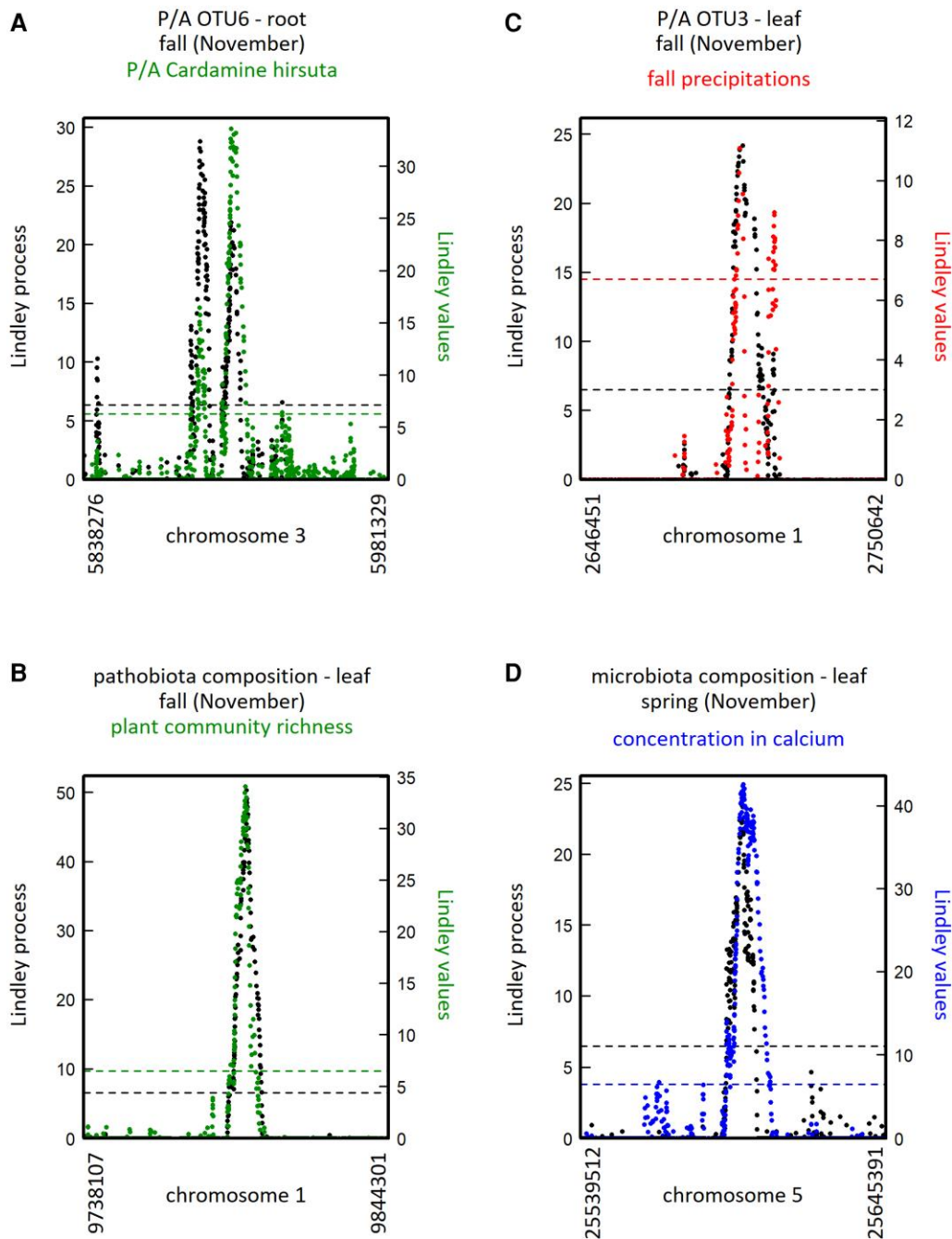
pathobiota and nonmicrobial ecological variables was on average higher in the leaf compartment than in the root compartment (fig. 3A), albeit not significant (GLM,  $F = 2.06$ ,  $P = 0.1526$ ). No differences were detected between microbiota and potential pathobiota descriptors (GLM,  $F = 0.06$ ,  $P = 0.8082$ ) (fig. 3A).

### The Relative Importance of Host Genetics and Nonmicrobial Ecology in Explaining Microbiota/Potential Pathobiota Variation Differs among Bacterial OTUs at the Family and Genus Levels

To tease apart the relative importance of host genetics and nonmicrobial ecological variables in explaining microbiota/potential pathobiota variation, we ran a multiple linear regression model on each descriptor of microbiota and potential pathobiota by considering 1) all the QTLs that were specific to the descriptor of microbiota and potential pathobiota under consideration (i.e., not common with nonmicrobial ecological variables) and 2) each nonmicrobial ecological variable sharing QTLs with the microbiota/potential pathobiota trait under consideration (see Material and Methods). We also considered in this model the first axis of a genomic principal component analysis (PCA) as a proxy of the potential effects of the demographic history of *A. thaliana* on descriptors of microbiota and potential pathobiota in southwest of France.

Across the 194 descriptors of microbiota and potential pathobiota, a much higher fraction of among-population variance was explained by host genetics (mean = 35.4%, median = 35.0%) than by ecology (mean = 5.2%, median = 3.8%) (fig. 5A and supplementary table S4, Supplementary Material online). Several hypotheses can be advanced to explain the higher total percentage of variance explained by the QTLs detected in our regional study in comparison with many former GWAS conducted on plant bacterial microbiota using worldwide mapping panels. Firstly, as previously mentioned, some QTLs detected by our GEA analysis can be associated with nonmicrobial ecological variables uncharacterized on the 141 natural populations of *A. thaliana* considered in this study. Secondly, as previously advised for GWAS (Bergelson and Roux 2010) and observed for quantitative phenotypic traits (Brachi et al. 2013; Lopez-Arboleda et al. 2021), working at a regional scale should reduce the main drawbacks of GEA analyses (e.g., genetic and allelic heterogeneity, presence of rare alleles, and allele frequency autocorrelation) often observed at large geographical scales, thereby allowing a better description of the genetic architecture at smaller geographical scales. Accordingly, in contrast with a recent study conducted at a continental scale (Karasov et al. 2022), a very small fraction of microbiota/potential pathobiota variation was on average explained by the demographic history of *A. thaliana* in our study (mean = 2.52%, median = 1.23%) (supplementary fig. S2 and table S4, Supplementary Material online). Finally, as previously demonstrated in a traditional linkage mapping study based on a mouse advanced intercross line mapping





**Fig. 4.** Zoom spanning four QTL regions common between descriptors of microbiota/potential pathobiota and nonmicrobial ecological variables. (A) Overlapping QTLs between the presence/absence of OTU6 in the root compartment for the “fall (November)” seasonal group and the presence/absence of the plant species *Cardamine hirsuta*. (B) Overlapping QTLs between the composition of potential pathobiota in the leaf compartment for the “fall (November)” seasonal group and plant community richness. (C) Overlapping QTLs between the presence/absence of OTU3 in the leaf compartment for the “fall (November)” seasonal group and fall precipitations (mean value over the 2003–2013 period). (D) Overlapping QTLs between microbiota composition of the leaf compartment for the “spring (November)” seasonal group and soil calcium concentrations. Black and colored dots correspond to Lindley values for microbiota/potential pathobiota descriptors and nonmicrobial ecological variables, respectively. Lindley values correspond to the  $-\log_{10}(p)$  of the tuning parameter  $\zeta = 2$ . The black and colored dashed lines indicate the corresponding chromosome-wide significance threshold for microbiota/potential pathobiota descriptors and nonmicrobial ecological variables, respectively. Month names in brackets correspond to the main germination cohort of the populations considered.

population phenotyped for gut microbiota composition (Benson et al. 2010), host genetics appears to mainly control microbiota at the lowest taxonomic levels such as the species level, which is in line with the identity of QTLs of

disease resistance being highly dependent on the pathogenic species/strain considered (Debieu et al. 2016; Wang et al. 2018; Demirjian et al. 2023). The deeper taxonomic resolution conferred by the *gyrB* gene used in our study

in comparison with the 16S rRNA gene used in former GWAS conducted on plants might have allowed to better describe the host genetic architecture associated with bacterial microbiota variation.

In line with the polygenic architecture detected in our study, the total fraction of variance explained by host genetics resulted from multiple QTLs, each explaining on average 1.99% ([supplementary fig. S2 and table S5, Supplementary Material online](#)). On average, the small QTL effects detected by GEA are similar to the QTL effect sizes identified in GWAS and traditional linkage mapping studies on microbiota ([Bergelson, Brachi, et al. 2021; Oyserman et al. 2022](#)). However, as previously observed in animals ([Benson et al. 2010](#)), a larger range of QTL effects was identified in our study in comparison with former association genetic studies conducted on plants. For instance, the top SNP located on position 19,335,417 bp on chromosome 2 (SNP 2\_19335417), SNP 1\_12691514, and SNP 5\_5191015 explained alone 55.4%, 35.5%, and 33.5% of the presence/absence of OTU4—recently identified as a new species of the family *Comamonadaceae* ([Ramírez-Sánchez, Gibelin-Viala, et al. 2022](#)) (root compartment of the “fall [November]” seasonal group), microbiota diversity (leaf compartment of the “spring [December]” seasonal group), and potential pathobiota composition (leaf compartment of the “fall [November]” seasonal group), respectively ([fig. 5B–D and supplementary table S5, Supplementary Material online](#)).

Amongst the nonmicrobial ecological variables, the variance in descriptors of microbiota and potential pathobiota was on average more explained by plant community descriptors (mean = 3.75%) than by climate variables (mean = 0.66%) and soil physicochemical properties (mean = 0.78%), suggesting the plant neighborhood as the main ecological driver of microbiota/potential pathobiota of *A. thaliana* populations located southwest of France ([supplementary table S4, Supplementary Material online](#)). A similar pattern was observed between the leaf and root compartments ([supplementary table S4, Supplementary Material online](#)).

As previously mentioned, microbiota/potential pathobiota variation was poorly explained by the demographic history of *A. thaliana* in our study ([supplementary fig. S2 and table S4, Supplementary Material online](#)). Note, however, that testing the effect of the demographic history of *A. thaliana* on microbiota and potential pathobiota descriptors might also be viewed as estimating the relative fraction of microbiota and potential pathobiota variance explained by host genetic variation that results from demographic processes impacting the effective population size independently of natural selection. The fraction of among-population variance explained by host genetics related to the detected QTLs might therefore not fully represent the total effect of host genetics on microbiota and potential pathobiota.

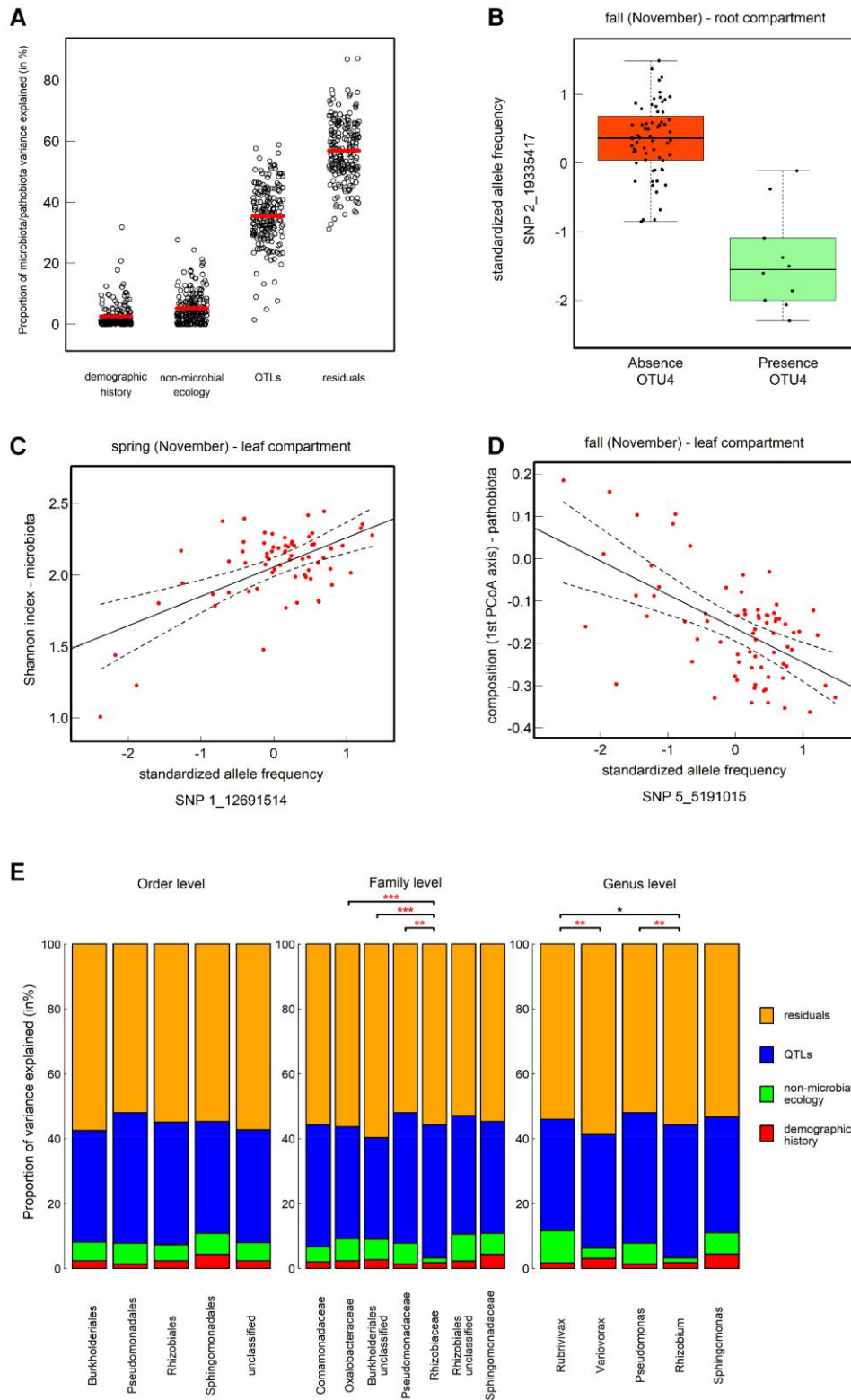
Across the 194 descriptors of microbiota and potential pathobiota, a substantial fraction of variance remained unexplained (mean = 56.9%, median = 56.4%) ([supplementary](#)

[fig. S2 and table S4, Supplementary Material online](#)). This unexplained variance may originate from 1) explanatory nonmicrobial ecological variables that were not characterized on the 141 natural populations of *A. thaliana*, as previously mentioned, 2) QTLs with a very small effect that remains undetected due to a lack of power given the number of populations used for GEA analysis ([table 1](#)), 3) the presence of epistatic QTLs not identified by our BHM-LS approach, and/or 4) stochastic processes of dispersal and drift that can drastically alter the community structure ([Hanson et al. 2012; Wagner 2021](#)).

The relative importance of host genetics, ecology, and demographic history in explaining microbiota/potential pathobiota variation was similar between leaf and root compartments, between microbiota and potential pathobiota, and between the three seasonal groups ([supplementary fig. S3 and table S6, Supplementary Material online](#)). Importantly, the relative importance of host genetics, ecology, and demographic history differs among bacterial OTUs at the family and genus level ([fig. 5E and table S6, Supplementary Material online](#)). For instance, the relative importance of host genetics (in comparison with nonmicrobial ecological variables) in explaining the variance of microbiota/potential pathobiota OTUs was significantly higher for OTUs of the *Rhizobium* genus than for OTUs of the *Pseudomonas* genus ([fig. 5E](#)). Interestingly, a phylogenetic signal at the order level was observed in GWAS conducted on the rhizospheric microbiota of sorghum and the leaf microbiota of maize, with heritable microbes that phylogenetically clustered ([Walters et al. 2018; Deng et al. 2021](#)). Similarly, heritable microbial taxa tend to be phylogenetically clustered in diverse animal species ([Ryu and Davenport 2022](#)). Altogether, these results suggest that some bacterial taxonomic groups are more under the genetic control of *A. thaliana* than others, thereby providing candidate bacterial taxa to be investigated for genomic signals of c-evolution with *A. thaliana* through, for example, free-phenotyping co-GWAS ([Bartoli and Roux 2017; Demirjian et al. 2023](#)) or controlled evolutionary experiments. Why some taxonomic groups might be more under host genetic control than other taxonomic groups remains however an intriguing and open question that deserves further investigations.

### The Genetic Architecture of Association with Bacterial Communities Is Highly Flexible between Plant Compartments and Seasons

After retrieving candidate genes located in the vicinity of the top SNPs specific to the descriptors of microbiota and potential pathobiota (i.e., not common with nonmicrobial ecological variables), we observed a highly flexible genetic architecture between the six “plant compartment × seasonal group” combinations, with 76.1% of unique candidate genes being specific to a “plant compartment × seasonal group” combination ([fig. 6A, supplementary table S7, Supplementary Material online](#)). As illustrated with OTU5 recently confirmed to belong to the *Pseudomonas moraviensis* species

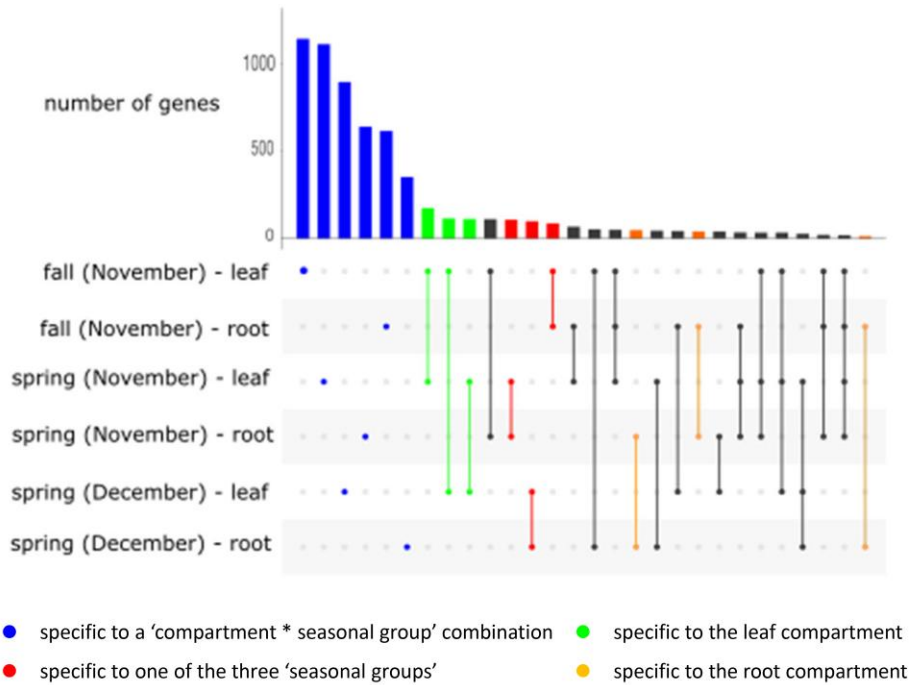


**Fig. 5.** Relative importance of host genetics and nonmicrobial ecological variables in explaining the variation of microbiota/potential pathobiota descriptors. (A) Cumulative PVE of microbiota/potential pathobiota descriptors by the demographic history of *A. thaliana*, host genetics with QTLs specific to microbiota/potential pathobiota descriptors and nonmicrobial ecological variables (including climate variables, soil physico-chemical properties, and descriptors of plant communities). Residuals correspond to the percentage of variance unexplained by the three other categories of variables. For each category, one dot corresponds to one of the 194 microbiota/potential pathobiota descriptors. (B) Box-plot illustrating the relationship between the standardized allele frequencies (corrected for the effect of population structure) of a SNP located at position 19,335,417 on chromosome 2 and the presence of OTU4 in the root compartment of the “fall (November)” seasonal group. (C) Relationship between the standardized allele frequencies (corrected for the effect of population structure) of a SNP located at position 12,691,514 on chromosome 1 and Shannon index of the microbiota in the leaf compartment of the “spring (November)” seasonal group. (D) Relationship between the standardized allele frequencies of a SNP located at position 5,191,015 on chromosome 5 and potential pathobiota composition in the leaf compartment of the “fall (November)” seasonal group. (E) Relative importance of PVE by demographic history, host genetics, nonmicrobial ecology, and residuals, among taxonomic groups at the order, family, and genus taxonomic levels. *P*-values were obtained from a chi-squared test applied on each pairwise comparison among taxonomic groups. \**P* < 0.05, \*\**P* < 0.001, and \*\*\**P* < 0.001. Red asterisks indicate significant *P*-values after correcting for multiple testing with a false discovery rate (FDR) at a nominal level of 5%. Month names in brackets correspond to the main germination cohort of the populations considered.

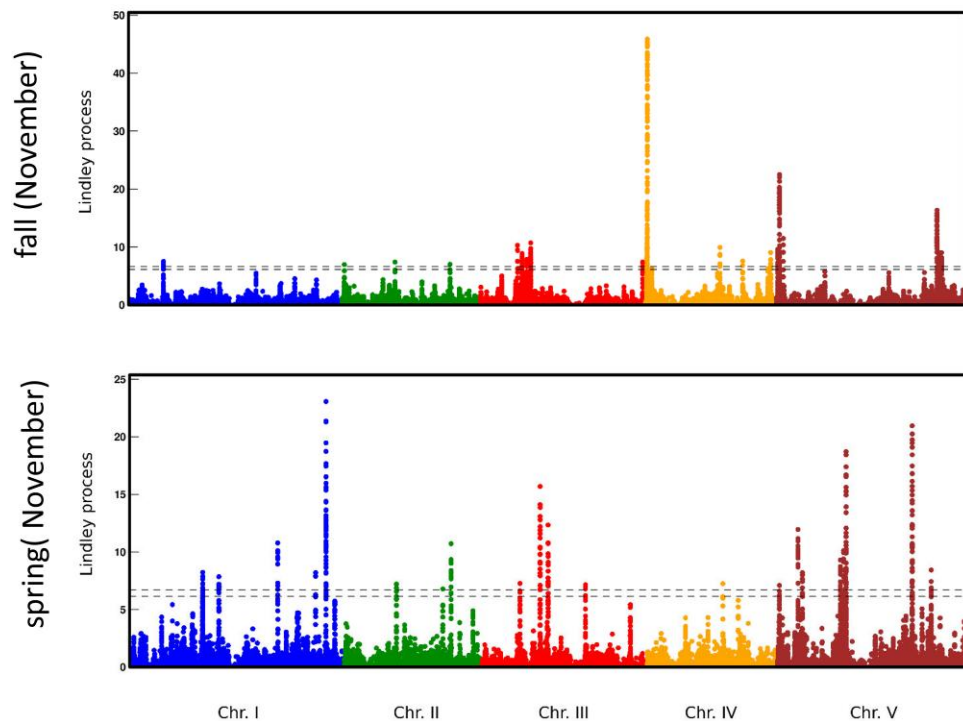
Downloaded from <https://academic.oup.com/mbe/article/40/5/msad093/7128101> by INRA Centre Val de Loire user on 12 September 2023

A

**Fig. 6.** Flexibility of the genetic architecture between the six “plant compartment × seasonal group” combinations. (A) An UpSet plot showing the intersections of the lists of unique candidate genes associated with microbiota/potential pathobiota variation, each list corresponding to each “plant compartment × seasonal group” combination. The number of candidate genes that are specific to a single “plant compartment × seasonal group” combination and common between at least two “plant compartment × seasonal group” combinations is represented by single blue dots and dots connected by a solid line, respectively. “Fall (November)—leaf”: number of unique candidate genes  $N = 1918$ ; “fall (November)—root”:  $N = 1052$ ; “spring (November)—leaf”:  $N = 1885$ ; “spring (November)—root”:  $N = 1156$ ; “spring (December)—leaf”:  $N = 1457$ ; “spring (December)—root”:  $N = 705$ . (B) Manhattan plot illustrating the flexibility of genetic architecture associated with the presence/absence of the OTU5 in the root compartment between fall and spring. The x-axis indicates along the five chromosomes the physical position of the 1,392,959 SNPs and 1,382,414 SNPs considered for the root compartment in the “fall (November)” and “spring (November)” seasonal groups, respectively. The y-axis corresponds to the values of the Lindley process corresponding to the  $-\log_{10}(p)$  of the tuning parameter  $\zeta = 2$ . The dashed lines indicate the minimum and maximum of the five chromosome-wide significance thresholds. Month names in brackets correspond to the main germination cohort of the populations considered.



B

presence/absence – OTU5 – *Pseudomonas moraviensis*

(Ramírez-Sánchez, Gibelin-Viala, et al. 2022) (fig. 6B), a highly flexible genetic architecture for a specific descriptor of microbiota and potential pathobiota was observed between fall and spring when considering the same set of natural populations (i.e., fall [November] and spring [November] seasonal groups). This is in line with the

strong genetic variation in seasonal microbial community succession observed in diverse plant species including *A. thaliana* (Copeland et al. 2015; Bartoli et al. 2018; Beilsmith et al. 2021; VanWallendael et al. 2022) and the dynamics of genetic architecture along the infection stages when *A. thaliana* accessions are challenged with bacterial

pathogens (Aoun et al. 2017, 2020; Demirjian et al. 2022). For instance, the atypical meiotic cyclin *SOLO DANCERS* identified as a candidate gene underlying a QTL with a transient effect in the response to the bacterial pathogen *Ralstonia solanacearum* was confirmed by a reverse genetic approach to only confer disease resistance during the intermediate stages of infection (Aoun et al. 2020). Interestingly, the identity of candidate genes also strongly differs between the two seasonal groups “spring (November)” and “spring (December)” in both plant compartments, thereby suggesting a relationship between germination timing and the interplay between host genetics and microbiota/potential pathobiota. While germination timing was found to influence natural selection on life-history traits in *A. thaliana* (Donohue 2002; Donohue et al. 2005), the relationship between germinating timing and microbiota/potential pathobiota assemblages has been seldom reported (Soldan et al. 2021) and deserves further investigations. We should however caution that the different seasonal groups might be genetically distinct. A lack of overlap in QTLs among the three seasonal groups could therefore result from different genetic variants being present in each seasonal group.

The highly flexible genetic architecture observed between plant compartments in our study (fig. 6A) is in agreement with one GWAS conducted on a set of worldwide accessions of *A. thaliana* on the leaf (Horton et al. 2014) and root (Bergelson et al. 2019) compartments, with the detection of a very small overlap in candidate genes associated with descriptors of bacterial and fungal communities between the two plant compartments (Bergelson et al. 2019). Such a compartment-dependent genetic architecture is in line with the adaptive differences in microbial community diversity and composition observed among plant niches, from rhizosphere soils to plant canopies (Müller et al. 2016; Cregger et al. 2018).

Most of the remaining candidate genes were either specific to the leaf compartment and common between the three seasonal groups or specific to a given seasonal group and common between the leaf and root compartments (fig. 6A). On the other hand, very few candidate genes were specific to the root compartment and common between the three seasonal groups (fig. 6A). These candidate genes shared between seasons and/or plant compartments might represent “core” genes affecting microbiota/potential pathobiota variation independently of plant organs, plant developmental stage and environmental conditions.

Beyond a highly flexible genetic architecture between plant compartments and between seasons, we also observed a highly flexible genetic architecture among descriptors of microbiota and potential pathobiota for each “plant compartment × seasonal group” combination (supplementary fig. S4, Supplementary Material online). For instance, for the potential pathobiota in the leaf compartment in the “spring (December)” seasonal group, the identity of candidate genes strongly differs between community diversity, community composition, and the

presence/absence of *P. viridiflava* (supplementary fig. S5, Supplementary Material online). As previously observed in GWAS and GEAS conducted on plant-plant interactions in *A. thaliana* (Baron et al. 2015; Frachon et al. 2019; Libourel et al. 2021), these results suggest a high degree of biotic specialization of *A. thaliana* to members of its bacterial interaction network, as well as the genetic ability of *A. thaliana* to interact simultaneously with multiple bacterial members.

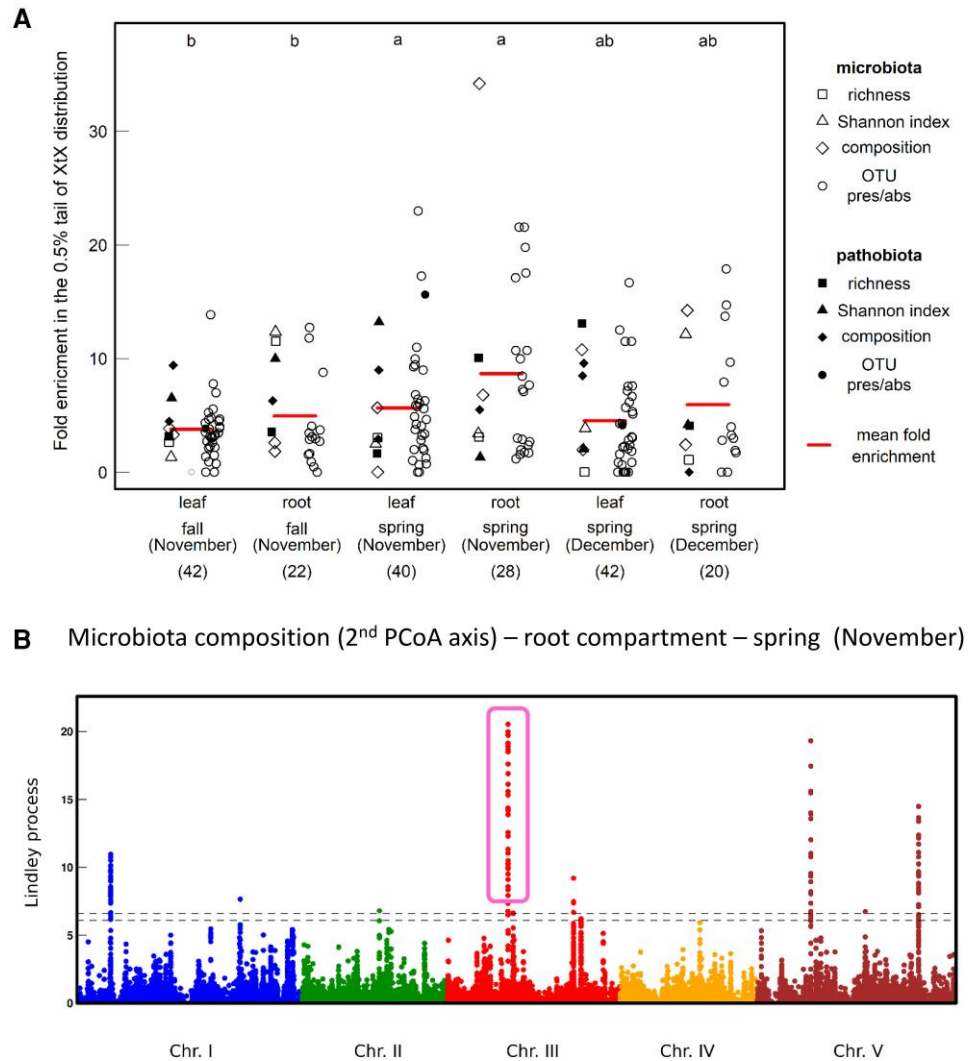
Altogether, as previously observed (Frachon et al. 2017), the highly flexible genetic architecture observed in this study should allow local populations to quickly adapt to fast environmental changes without affecting the level of genetic diversity present in local populations. Therefore, in line with the ever-changing complexity of biotic interactions observed in nature (Bergelson, Kreitman, et al. 2021), our study reinforces the need to conduct association genetic studies on diverse plant compartments and seasons to obtain a full picture of the host genetics controlling natural variation of microbiota/potential pathobiota.

### The Strength of Signatures of Local Adaptation on QTLs of Microbiota/Potential Pathobiota Depends on Plant Compartment and Season

By definition, GEA allows identifying genetic loci under local adaptation. However, in order to support that the loci identified by our GEA analysis have been shaped by natural selection, we additionally tested whether the top SNPs specific to descriptors of microbiota and potential pathobiota were enriched in a set of SNPs subjected to adaptive spatial differentiation. To do so, we first performed for each “plant compartment × seasonal group” combination a genome-wide selection scan by estimating a Bayesian measure of spatial genetic differentiation (XtX) among the natural populations of *A. thaliana*, with SNPs with the highest XtX values considered as the main candidates for signatures of local adaptation (Gautier 2015). The 0.5% upper tail of the spatial genetic differentiation distribution displayed a significant enrichment (up to 34.2) for top SNPs of almost two-thirds of the descriptors of microbiota and potential pathobiota (fig. 7A and supplementary table S8, Supplementary Material online). For instance, a strong signature of local adaptation was identified on SNPs located in the 5' region of *MYB15* (fig. 7B), a transcription factor involved in the coordination of microbe-host homeostasis in *A. thaliana* (Ma et al. 2021).

No differences in the mean fold enrichment for signatures of local adaptation were detected between microbiota and potential pathobiota descriptors (GLM,  $F = 0.44$ ,  $P = 0.7260$ ). However, the mean fold enrichment for signatures of local adaptation largely differed between the three seasonal groups (GLM,  $F = 5.06$ ,  $P = 0.0072$ ), with top SNPs presenting stronger signatures of local adaptation in spring than in fall when considering the same set of populations, that is, populations from the “fall (November)” and “spring (November)” seasonal groups (fig. 7A). Top SNPs of the “spring (December)” seasonal

**Fig. 7.** Signatures of local adaptation acting on the host genetics of microbiota/potential pathobiota variation. (A) Fold enrichment of the top SNPs in the 0.5% tail of the genome-wide spatial differentiation (XtX) distribution. Each dot corresponds to one of the 194 descriptors of microbiota and potential pathobiota. Different upper letters indicate different groups according to the “plant compartment × seasonal group” combination after a Tukey correction for multiple pairwise comparisons. Numbers in brackets correspond to the number of descriptors of microbiota and potential pathobiota considered for each “plant compartment × seasonal group” combination. (B) Manhattan plot highlighting a QTL associated with root microbiota composition, located in the middle of chromosome 3 and presenting a strong signature of local adaptation. The x-axis indicates along the five chromosomes, the physical position of the 1,382,414 SNPs considered for the root compartment of the “spring (November)” seasonal group. The y-axis corresponds to the values of the Lindley process corresponding to the  $-\log_{10}(p)$  of the tuning parameter  $\zeta = 2$ . The dashed lines indicate the minimum and maximum of the five chromosome-wide significance thresholds. Month names in brackets correspond to the main germination cohort of the populations considered.



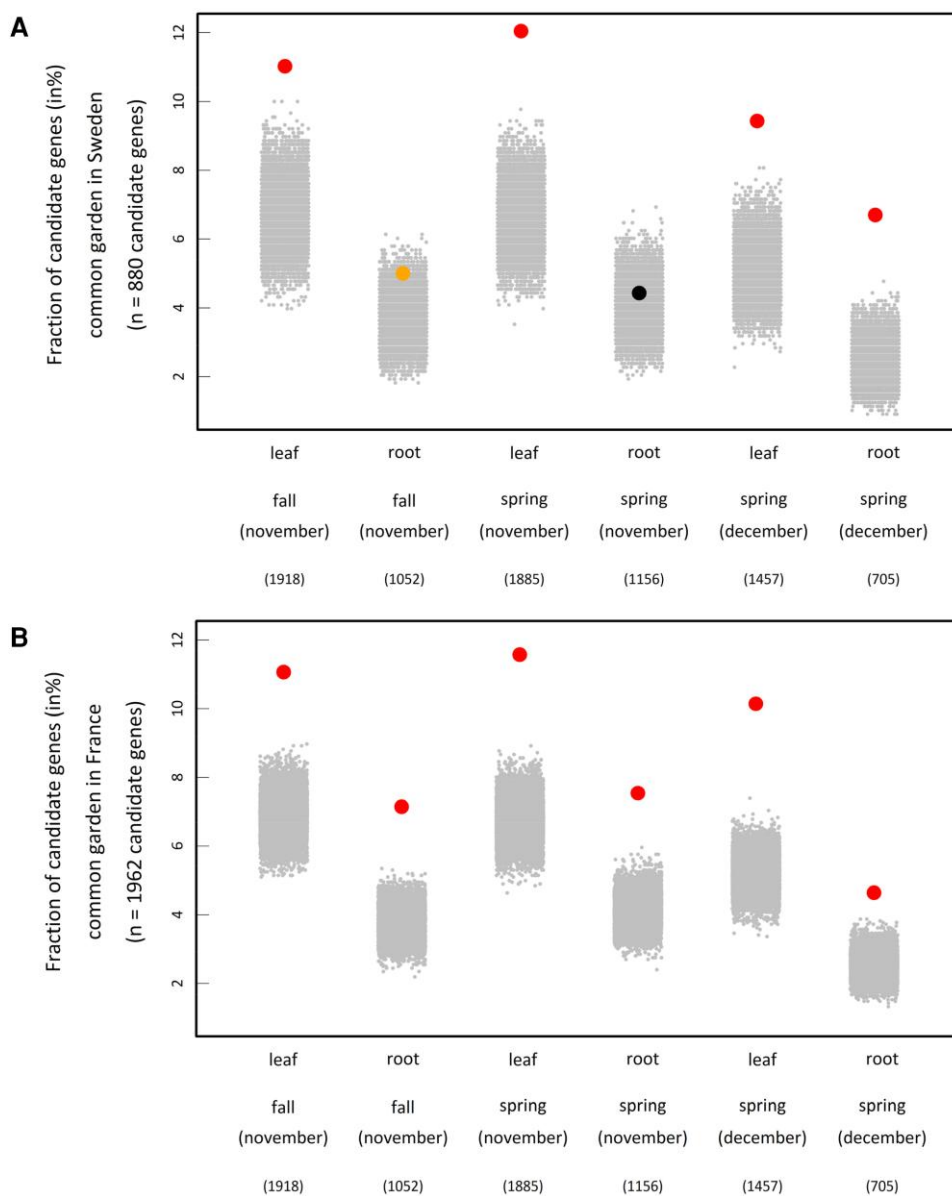
group presented signatures of local adaptation that were intermediate between the two other seasonal groups, suggesting that germinating timing affects not only the genetic architecture underlying microbiota/potential pathobiota variation but also the strength of local adaptation acting on candidate genes.

In addition, the mean fold enrichment for signatures of local adaptation was significantly higher for the root compartment than for the leaf compartment (GLM,  $F = 6.11$ ,  $P = 0.0143$ ). Combined with the observation that the percentage of common SNPs between microbiota/potential pathobiota and nonmicrobial ecological variables was on average higher in the leaf compartment than in the root compartment (fig. 3A), the difference between the leaf and root compartments in the strength of local adaptation acting on candidate genes suggests a higher

adaptive genetic regulation of the root microbiota than that of the leaf microbiota.

### Cross-validation of Our GEA Approach through Results Obtained from GWAS Conducted in Common Gardens

Our GEA analysis on microbiota suffers from the quasi-impossibility of measuring the entire set of nonmicrobial ecological variables acting on natural populations of *A. thaliana*, thereby precluding the identification of all SNPs common between descriptors of microbiota/potential pathobiota and nonmicrobial ecological variables. To circumvent this issue, we, therefore, tested whether our list of candidate genes associated with descriptors of microbiota and potential pathobiota significantly overlaps



**Fig. 8.** Cross-validation of the candidate genes identified by our GEA analysis with candidate genes identified by GWAS conducted in two common gardens. (A) Fraction of candidate genes obtained from a GWAS performed on the leaf bacterial communities of 200 Swedish *A. thaliana* accessions grown in the native habitats of four natural populations of *A. thaliana* in Sweden (Brachi et al. 2022) overlapping with GEA candidate genes. (B) Fraction of candidate genes obtained from a GWAS performed on the vegetative growth response of 162 accessions from southwest of France to 13 bacterial strains belonging to 7 of the 12 most prevalent and abundant leaf nonpathogenic bacterial species isolated from the leaf compartment of *A. thaliana* in the same geographical region (Ramírez-Sánchez, Duflos, et al. 2022) overlapping with GEA candidate genes. Large red, purple, orange, and black dots correspond to the fraction of overlapped GWA candidate genes with  $P < 0.001$ ,  $P < 0.01$ ,  $P < 0.05$ , and  $P > 0.05$ , respectively. Numbers in brackets correspond to the number of unique candidate genes  $n$  for each “plant compartment  $\times$  seasonal group” combination. Smaller gray dots represent 10,000 random sampling of  $n$  genes across the entire set of 27,206 genes present across the five chromosomes of *A. thaliana*. Month names in brackets correspond to the main germination cohort of the populations considered.

with candidate genes identified in two GWAS conducted on *A. thaliana* in common gardens under ecologically relevant conditions. The first GWAS was conducted on 200 Swedish accessions characterized for the bacterial and fungal communities in the leaf compartment in the native habitat of four populations located in Sweden (Brachi et al. 2022). The resulting list of 880 candidate genes located in the vicinity of the top SNPs associated with bacterial hubs in Sweden significantly overlapped with the lists of candidate genes identified by our GEA analysis for the leaf compartment, whatever the seasonal group (fig. 8A and supplementary table S9, Supplementary Material online). The fraction of candidate genes detected in Sweden and overlapping with the lists of candidate genes identified by our GEA analysis for the root compartment

was smaller and less significant for each seasonal group (fig. 8A).

The second GWAS was conducted in a common garden located southwest of France on 162 whole-genome sequenced accessions of *A. thaliana* originating from 54 out of the 141 natural populations used in our study (Ramírez-Sánchez, Duflos, et al. 2022). Those accessions were challenged in field conditions with 13 bacterial strains belonging to seven of the 12 most abundant and prevalent leaf OTUs across our natural populations in southwest of France and phenotyped for diverse vegetative growth-related traits such as the rosette surface area and relative growth rate (Ramírez-Sánchez, Duflos, et al. 2022). Whatever the “plant compartment  $\times$  seasonal group” combination considered, a significant overlap was

detected between the list of 1,962 unique candidate genes located in the vicinity of the top SNPs associated with the response of vegetative growth-related traits to the 13 bacterial strains and the list of candidate genes identified by GEA analysis in our study (fig. 8B and supplementary table S10, Supplementary Material online).

In addition, based on the list of candidate genes of each “plant compartment × seasonal group” combination, the main enriched biological pathways identified in our study (e.g., cell, development, hormone metabolism, secondary metabolism, signaling, and transport) (supplementary fig. S6 and table S11, Supplementary Material online) correspond to the main biological pathways (with the exception of symbiosis) identified by the use of artificial mutations and by previous GWAS conducted on *A. thaliana* (Bergelson, Brachi, et al. 2021) as being involved in microbiota regulation.

Altogether, the similar lists of candidate genes and biological pathways detected between diverse mapping populations and using different approaches in association genetics indicate cross-validation of our GEA results by GWAS conducted in ecologically relevant conditions. We nonetheless stress that the candidate genes identified by our GEA approach and not identified by the two GWAS mentioned above can still be associated with nonmicrobial ecological variables not characterized in this study.

## Conclusion

The GEA study conducted here on descriptors of microbiota/potential pathobiota and with an unprecedented number of nonmicrobial ecological variables is complementary to GWAS carried out previously on *A. thaliana*. In particular, we investigated 1) the level of flexibility of the genetic architecture between seasons and between in situ germination timings, 2) the relative importance of host genetics and nonmicrobial ecological variables in explaining microbiota/potential pathobiota variation, 3) the identity of the nonmicrobial ecological variables as the main predictors of descriptors of microbiota and potential pathobiota, and 4) the variation between plant compartments and seasonal groups, in the strength of local adaptation acting on candidate genes. Importantly, no differences were observed between microbiota and potential pathobiota descriptors, thereby suggesting a similar adaptive genetic control of *A. thaliana* towards its pathogenic and nonpathogenic members, including members with already described beneficial effects (Ramírez-Sánchez, Gibelin-Viala, et al. 2022).

One of the next steps is then to apply our approach to fungal communities that have not been characterized on our set of natural populations yet. Establishing a genomic map of local adaptation to fungal communities might allow testing whether the genetic bases largely differ between bacterial and fungal communities, as previously documented in two GWAS conducted on *A. thaliana* (Horton et al. 2014; Bergelson et al. 2019; Brachi et al. 2022). It might also help to identify the adaptive genetic bases of common interactions among OTUs, including

mutualism, antagonism, aggression, and altruism, within and across kingdoms (He et al. 2021). Another fundamental step would be to functionally validate the most promising candidate genes identified by our GEA approach, that is, candidate genes with the strongest allelic effects, presenting signatures of local adaptation and identified in other GWAS performed on *A. thaliana*-microbiota interactions. Such a functional validation would represent a first step for a better understanding of the biomolecular networks underlying an adaptive host response to its microbiota and its pleiotropic effects on other phenotypic traits.

## Materials and Methods

### Descriptors of Bacterial Communities in Leaves and Roots

The bacterial communities of 1,903 leaf and root samples collected in fall 2014 and spring 2015 across 163 out of 168 natural populations of *A. thaliana* located southwest of France (Frachon et al. 2018) were characterized with a *gyrB*-based metabarcoding approach (fig. 1C) (Bartoli et al. 2018). As previously described, leaf and root samples of each plant collected at the rosette stage were manipulated with sterilized tools and removed from any visible rhizosphere and dust, respectively, by rinsing roots and rosettes several times into individual tubes of sterilized distilled water (Bartoli et al. 2018). The epiphytic and endophytic bacterial components of either leaf or root samples were not separated (Bartoli et al. 2018).

In this study, we considered 141 natural populations of *A. thaliana* for which a set of 6 climate variables, 14 soil physicochemical variables, and 49 descriptors of plant communities was available (see below), thereby resulting in 73, 73, 69, 72, 66, and 66 populations for the “fall (November)—leaf,” “fall (November)—root,” “spring (November)—leaf,” “spring (November)—root,” “spring (December)—leaf,” and “spring (December)—root” combinations, respectively (table 1).

For each leaf and root sample collected in the 141 populations, we estimated the relative abundance (number of reads per OTU/number of total reads) of each of the 6,627 OTUs selected after controlling for sampling limitation and a minimal relative abundance of 1% in at least one sample (fig. 1C). For each “plant compartment × seasonal group” combination, we then averaged for each OTU the corresponding relative abundances per population. For each OTU, populations with a mean relative abundance above and below (or equal to) 0.5% were scored as 1 (presence of the OTU) and 0 (absence of the OTU), respectively. In this study, for statistical power in GEA analysis, we only kept OTUs present in more than seven populations, resulting in 34 OTUs, 15 OTUs, 32 OTUs, 21 OTUs, 34 OTUs, and 13 OTUs investigated for GEA analysis for the “fall (November)—leaf,” “fall (November)—root,” “spring (November)—leaf,” “spring (November)—root,” “spring (December)—leaf,” and “spring (December)—root” combinations, respectively (table 1).



Similarly, for each “plant compartment × seasonal group” combination, we averaged for both the microbiota and the potential pathobiota estimates of community richness and Shannon index previously obtained in [Bartoli et al. 2018](#) and estimates of community composition using the coordinates of the samples on the two first axes of a PCoA run on a Hellinger distance matrix based on the relative abundances of the 6,627 most abundant OTUs ([fig. 1C](#) and [supplementary table S2, Supplementary Material online](#)) ([Bartoli et al. 2018](#)). As previously described, no effect of plant age approximated by two traits (i.e., rosette diameter and number of leaves) was detected on microbiota/potential pathobiota diversity and composition, with the exception of a barely significant effect of plant age on the first PCoA axis of microbiota of the “spring (December)” season group ([Bartoli et al. 2018](#)).

### Abiotic Descriptors and Descriptors of Plant Communities

A set of 6 climate variables, 14 soil physicochemical properties, and 49 descriptors of plant communities was available for 141 out of the 168 natural populations of *A. thaliana* located southwest of France ([supplementary table S3, Supplementary Material online, Data Sets 1–6](#)). The climate variables correspond to two variables related to temperature and four variables related to precipitation ([Frachon et al. 2018](#)). The 14 soil physicochemical variables describe the main soil agronomic properties related to plant growth ([Brachi et al. 2013; Frachon et al. 2019](#)). The 49 descriptors of plant communities correspond to 1) estimates of richness and Shannon index, 2) estimates of community composition using the coordinates of the populations on the three first axes of a PCoA run on a Bray-Curtis dissimilarity matrix based on the relative abundance of the 44 most prevalent plant species, and 3) the presence/absence of the 44 most prevalent plant species, except for *A. thaliana* for which an estimate of density was kept ([Frachon et al. 2019](#)). In this study, for the purpose of statistical power in GEA analysis, we only kept companion plant species present in more than seven populations, resulting in 30, 29, 30, 31, 33, and 33 plant species investigated for GEA analysis for the “fall (November)—leaf,” “fall (November)—root,” “spring (November)—leaf,” “spring (November)—root,” “spring (December)—leaf,” and “spring (December)—root” combinations, respectively (Data Sets 1–6).

### Plant Genomic Characterization and Data Filtering

As previously described in [Frachon et al. 2018](#), a representative picture of within-population genetic variation across the genome was previously obtained for the 168 populations located southwest of France, using a Pool-Seq approach based on the individual DNA extraction of ~16 plants per population (min = 5 plants, max = 16 plants, mean = 15.32 plants, and median = 16 plants) that were different from the plants used for the characterization of bacterial communities ([fig. 1B](#)). After

bioinformatics analysis using the reference genome Col-0, the allele read count matrix (for both the reference and alternate alleles) was composed of 4,781,661 SNPs across the 168 populations ([Frachon et al. 2018](#)).

Following [Frachon et al. \(2018\)](#), for each “plant compartment × seasonal group” combination, the matrix of population allele frequencies was trimmed according to five successive criteria: 1) removing SNPs with missing values in more than two populations, 2) in order to take into account multiple gene copies in the populations that map to a unique gene copy in the reference genome Col-0, removing SNPs with a mean relative coverage depth across the populations above 1.5 after calculating for each population the relative coverage of each SNP as the ratio of its coverage to the median coverage (computed over all the SNPs), 3) in order to take into account indels that correspond to either genomic regions present in Col-0 but absent in the populations or genomic regions present in the populations but absent in Col-0, removing SNPs with a mean relative coverage depth across the populations below 0.5, 4) removing SNPs with a standard deviation of allele frequency across the populations below 0.004, and 5) in order to take into account bias in GEA analysis due to rare alleles ([Bergelson and Roux 2010](#)), removing SNPs with the alternative allele present in less than 10% of the populations. This SNP pruning resulted in a final number of 1,396,579 SNPs, 1,392,959 SNPs, 1,470,777 SNPs, 1,382,414 SNPs, 1,514,789 SNPs, and 1,514,789 SNPs for the “fall (November)—leaf,” “fall (November)—root,” “spring (November)—leaf,” “spring (November)—root,” “spring (December)—leaf,” and “spring (December)—root” combinations, respectively.

### Genome-Environment Association Analysis

For each “plant compartment × seasonal group” combination, a GEA analysis was performed between the set of pruned SNPs and each trait related to microbiota/potential pathobiota, climate, soil physicochemical properties, and plant communities, resulting in a total number of 530 traits, with 97 traits, 76 traits, 95 traits, 84 traits, 100 traits, and 78 traits for the “fall (November)—leaf,” “fall (November)—root,” “spring (November)—leaf,” “spring (November)—root,” “spring (December)—leaf,” and “spring (December)—root” combinations, respectively ([supplementary table S1, Supplementary Material online, Data Sets 1–6](#)).

To identify significant “SNP-trait” associations for each “plant compartment × seasonal group” combination, we first run a BHM 1) explicitly accounting for the scaled covariance matrix of population allele frequencies ( $\Omega$ ), which makes the analyses robust to complex demographic histories, 2) dealing with Pool-Seq data, and 3) implemented in the program BayPass ([Gautier 2015](#)). For each SNP, we estimated a Bayesian Factor (BF<sub>is</sub>, measured in deciban units) and the associated regression coefficient (Beta<sub>is</sub>,  $\beta_i$ ) using an importance sampling algorithm ([Gautier 2015](#)). The full posterior distribution of the parameters

was obtained based on a Metropolis–Hastings within the Gibbs Markov chain Monte Carlo (MCMC) algorithm. A MCMC chain consisted of 15 pilot runs of 500 iterations each. Then, MCMC chains were run for 25,000 iterations after a 2500-iteration burn-in period. The  $n$  traits were scaled (*scalecov* option) so that  $\mu = 0$  and  $\sigma^2 = 1$ . Because of the use of an importance sampling algorithm, we repeated the analyses three times for each trait and averaged  $BF_{is}$  and  $\beta_i$  values across these three repeats. As previously performed in Frachon et al. 2018, for each “plant compartment  $\times$  seasonal group” combination, we parallelized GEA analysis by dividing the full data set of pruned SNPs into 32 subdata sets, each containing 3.125% of the total number of pruned SNPs taken every 32 SNPs across the genome.

As a second step, to better characterize the genetic architecture associated with each trait, the GEA results were reanalyzed by applying a LS approach (Bonhomme et al. 2019), which allows detecting significant genomic segments by accumulating the statistical signals from contiguous genetic markers such as SNPs (Fariello et al. 2017). In addition, this LS approach increases the power of detecting QTLs with a small effect and narrows the size of QTL genomic regions (Fariello et al. 2017; Bonhomme et al. 2019). In a given QTL region, the association signal, through the  $P$ -values, will cumulate locally due to linkage disequilibrium between SNPs, which will then increase the LS (Bonhomme et al. 2019). Following Libourel et al. 2021, to apply the LS approach to the GEA results, we first ranked each SNP based on the Bayes factor values obtained across the genome (from the highest to the lowest values) for each trait. Then, each rank was divided by the total number of SNPs to obtain a  $P$ -value associated with each SNP. The LS approach was then implemented on these  $P$ -values to fine map genomic regions associated with traits. In this study, the tuning parameter  $\zeta$  was fixed at 2 expressed in  $-\log_{10}$  scale. Significant associations between SNPs and a given trait were identified by estimating a chromosome-wide significance threshold for each chromosome (Bonhomme et al. 2019).

### Estimating the Relative Importance of Host Genetics and Nonmicrobial Ecological Variables in Explaining the Variation of Microbiota/Potential Pathobiota Traits

To estimate the relative importance of 1) QTLs specific to microbiota/potential pathobiota traits, 2) the demographic history of *A. thaliana*, and 3) nonmicrobial ecological variables (i.e., abiotic environment and plant communities) in explaining the variation of microbiota/potential pathobiota traits, we run the following multiple linear regression model under the R environment for each of the 194 microbiota/potential pathobiota traits:

$$Y_{a,p,i\dots n,j\dots m,k} = \mu_a + \text{population structure}_p + \text{QTL}_i + \dots + \text{QTL}_n + \text{ECOL}_j + \dots + \text{ECOL}_m + \varepsilon_{a,p,i\dots n,j\dots m,k}$$

where  $Y$  is one of the 194 microbiota/potential pathobiota traits;  $\mu$  is the overall mean; “population structure” accounts for the effect of the demographic history of *A. thaliana* by using the coordinates of the 141 populations on the first Principal Component axis ( $PC_{\text{genomic}}$  axis 1) explaining 96.4% of the genomic variation observed among the 168 populations located southwest of France (Frachon et al. 2018); “QTL” accounts for the effect of host genetics by using the standardized allele frequencies (i.e., raw allele frequencies corrected for the effect of demographic history) (Gautier 2015) of the SNP with the highest  $BF_{is}$  value for each QTL specific to the microbiota/potential pathobiota trait considered; “ECOL” corresponds to values of nonmicrobial ecological variables (climate variables, soil physicochemical properties, and descriptors of plant communities) for which QTLs were common to QTLs associated with microbiota/potential pathobiota traits; and  $\epsilon$  is the residual term. For each microbiota/potential pathobiota trait, we obtained a percentage of variance explained (PVEs) by each model term (supplementary table S5, Supplementary Material online) and PVEs were then summed according to four categories, that is, demographic history, host genetics specific to microbiota/potential pathobiota, nonmicrobial ecology (including abiotic environment and plant communities), and residuals (supplementary table S4, Supplementary Material online).

To test whether the relative PVE by these four categories differs among OTUs at the order, family, and genus taxonomic levels, we first averaged the PVE among OTUs belonging to a specific order, family, or genus (only orders, families, or genera with at least four OTUs were considered). At each taxonomic level, we then applied a chi-squared test on each pairwise comparison among taxonomic groups. A Bonferroni correction was applied to control for multiple testing. A similar approach was applied to test whether the relative PVE by these four categories differs between leaves and roots, between microbiota and potential pathobiota, and among the six seasonal groups.

### Identification of Candidate Genes Associated with Descriptors of Microbiota and Potential Pathobiota and Associated Enriched Biological Pathways

Based on a custom script (Libourel et al. 2021), we retrieved all candidate genes underlying QTLs by selecting all genes inside the QTL regions as well as the first gene upstream and the first gene downstream of these QTL regions (supplementary table S7, Supplementary Material online). The TAIR 10 database (<https://www.arabidopsis.org/>) was used as our reference. The number of candidate genes that were either specific to a single “plant compartment  $\times$  seasonal group” combination (a single microbiota/potential pathobiota descriptor) or common between several “plant compartment  $\times$  seasonal group” combinations (several microbiota/potential pathobiota descriptors) was illustrated by UpSet plots using the UpSetR package in R (Conway et al. 2017).

To identify biological pathways significantly over-represented ( $P < 0.01$ ) in each of the six “plant compartment × seasonal group” combinations, each of the six lists of unique candidate genes was submitted to the classification supervisor tool on the University of Toronto website ([http://bar.utoronto.ca/ntools/cgi-bin/ntools\\_classification\\_supervisor.cgi](http://bar.utoronto.ca/ntools/cgi-bin/ntools_classification_supervisor.cgi)) using the MAPMAN classification (Provart and Zhu 2003) (supplementary table S11, Supplementary Material online).

### Enrichment in Signatures of Local Adaptation

For supporting signals of local adaptation identified by GEA analysis, we first performed for each “plant compartment × seasonal group” combination, a genome-wide selection scan by estimating the XtX measure of spatial genetic differentiation among the populations. For a given SNP, XtX is a measure of the variance of the standardized population allele frequencies, which results from a rescaling based on the covariance matrix of population allele frequencies (Gautier 2015). Such rescaling allows for a robust identification of highly differentiated SNPs by correcting for the genome-wide effects of confounding demographic evolutionary forces, such as genetic drift and gene flow (Gautier 2015). For each of the 194 microbiota/potential pathobiota traits, we tested whether the top SNPs present signatures of local adaptation by following the methodology described in Brachi et al. 2015. More precisely, we tested whether the top SNPs were over-represented in the extreme upper tail of the XtX distribution using the following formula:

$$FE_{XtX} = \frac{n_a/n}{N_a/N}$$

with  $n$  being the number of SNPs in the upper tail of the XtX distribution. In our case, we used a threshold of 0.5%.  $n_a$  is the number of top SNPs that were also in the upper tail of the XtX distribution.  $N$  is the total number of SNPs tested genome wide and  $N_a$  is the total number of top SNPs. The statistical significance of enrichment was assessed by running 10,000 null circular permutations based on the methodology described in Hancock et al. 2011.

### Overlap with Candidate Genes Obtained from GWAS

Based on a GWAS performed on the leaf bacterial communities of 200 Swedish *A. thaliana* accessions grown in the native habitats of four natural populations of *A. thaliana* in Sweden (Brachi et al. 2022), we retrieved a list of 880 unique candidate genes underlying 209 QTLs, by selecting all genes inside the QTL regions as well as the first gene upstream and the first gene downstream of these QTL regions. To test whether the list of unique candidate genes obtained for each “plant compartment × seasonal group” combination significantly overlaps with the list of 880 candidate genes, we first estimated the percentage of the 800 candidate genes that were in common with the list of  $n$

candidate genes identified in this study. To estimate the level of significance, we then created a null distribution by randomly sampling 10,000 times,  $n$  genes across the entire set of 27,206 genes present across the five chromosomes of *A. thaliana*.

A similar approach was adopted with a list of 1,962 candidate genes identified in a GWAS performed on 162 whole-genome sequenced accessions of *A. thaliana* originating from 54 out of the 141 natural populations used in our study (Ramírez-Sánchez, Duflos, et al. 2022). The 162 accessions were phenotyped in a common garden for their response to 13 bacterial strains belonging to seven of the 12 most abundant and prevalent leaf OTUs across our natural populations in southwest of France (Ramírez-Sánchez, Duflos, et al. 2022).

### Supplementary material

Supplementary data are available at *Molecular Biology and Evolution* online.

### Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No. 951444-PATHOCOM to F.R. and grant agreement No. 1011039541-HoloE2Plant to C.B.).

### Author Contributions

F.R., L.F., and C.B. planned and designed the research. F.R. performed the statistical analyses and the genome-environment analysis. F.R. wrote the manuscript, with contributions from L.F. and C.B.

### References

- Aoun N, Desaint H, Boyrie L, Bonhomme M, Deslandes L, Berthomé R, Roux F. 2020. A complex network of additive and epistatic quantitative trait loci underlies natural variation of *Arabidopsis thaliana* quantitative disease resistance to *Ralstonia solanacearum* under heat stress. *Mol Plant Pathol*. **21**:1405–1420.
- Aoun N, Tauleigne L, Lonjon F, Deslandes L, Vailleau F, Roux F, Berthomé R. 2017. Quantitative disease resistance under elevated temperature: genetic basis of new resistance mechanisms to *Ralstonia solanacearum*. *Front Plant Sci*. **8**:1387.
- Bai B, Liu W, Qiu X, Jie Z, Jingying Z, Bai Y. 2022. The root microbiome: community assembly and its contributions to plant fitness. *J Integr Plant Biol*. **64**:230–243.
- Baron E, Richirt J, Villoutreix R, Amsellem L, Roux F. 2015. The genetics of intra- and interspecific competitive response and effect in a local population of an annual plant species. *Funct Ecol*. **29**: 1361–1370.
- Barret M, Briand M, Bonneau S, Prévieux A, Valière S, Bouchez O, Hunault G, Simoneau P, Jacquesa M-A. 2015. Emergence shapes the structure of the seed microbiota. *Appl Environ Microbiol*. **81**: 1257–1266.
- Bartoli C, Frachon L, Barret M, Rigal M, Huard-Chauveau C, Mayjonade B, Zanchetta C, Bouchez O, Roby D, Carrère S,

- et al. 2018. *In situ* relationships between microbiota and potential pathobiota in *Arabidopsis thaliana*. *ISME J.* **12**:2024–2038.
- Bartoli C, Roux F. 2017. Genome-wide association studies in plant pathosystems: toward an ecological genomics approach. *Front Plant Sci.* **8**:763.
- Bay RA, Rose N, Barrett R, Bernatchez L, Ghalambor CK, Lasky JR, Brem RB, Palumbi SR, Ralph P. 2017. Predicting responses to contemporary environmental change using evolutionary response architectures. *Am Nat.* **189**:463–473.
- Bebber DP. 2015. Range-expanding pests and pathogens in a warming world. *Annu Rev Phytopathol.* **53**:335–356.
- Beilsmith K, Perisin M, Bergelson J. 2021. Natural bacterial assemblages in *Arabidopsis thaliana* tissues become more distinguishable and diverse during host development. *MBio.* **12**:e02723–20.
- Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J, Zhang M, Oh PL, Nehrenberg D, Hua K, et al. 2010. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A.* **107**:18933–18938.
- Berendsen RL, Pieterse CMJ, Bakker PAHM. 2012. The rhizosphere microbiome and plant health. *Trends Plant Sci.* **17**:478–486.
- Bergelson J, Brachi B, Roux F, Vaillau F. 2021. Assessing the potential to harness the microbiome through plant genetics. *Curr Opin Biotechnol.* **70**:167–173.
- Bergelson J, Kreitman M, Petrov DA, Sanchez A, Tikhonov M. 2021. Functional biology in its natural context: a search for emergent simplicity. *Elife.* **10**:e67646.
- Bergelson J, Mittelstrass J, Horton MW. 2019. Characterizing both bacteria and fungi improves understanding of the *Arabidopsis* root microbiome. *Sci Rep.* **9**:24.
- Bergelson J, Roux F. 2010. Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nat Rev Genet.* **11**:867–879.
- Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, Bell JT, Spector TD, Keinan A, Ley RE, Gevers D, et al. 2015. Host genetic variation impacts microbiome composition across human body sites. *Genome Biol.* **16**:191.
- Bonhomme M, Fariello MI, Navier H, Hajri A, Badis Y, Miteul H, Samac DA, Dumas B, Baranger A, Jacquet C, et al. 2019. A local score approach improves GWAS resolution and detects minor QTL: application to *Medicago truncatula* quantitative disease resistance to multiple *Aphanomyces euteiches* isolates. *Heredity (Edinb).* **123**:517–531.
- Brachi B, Filiault D, Whitehurst H, Darne P, Le Gars P, Le Mentec M, Morton TC, Kerdaffrec E, Rabanal F, Anastasio A, et al. 2022. Plant genetic effects on microbial hubs impact host fitness in repeated field trials. *Proc Natl Acad Sci U S A.* **119**:e2201285119.
- Brachi B, Meyer CG, Villoutreix R, Platt A, Morton TC, Roux F, Bergelson J. 2015. Coselected genes determine adaptive variation in herbivore resistance throughout the native range of *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A.* **112**:4032–4037.
- Brachi B, Villoutreix R, Faure N, Hautekèete N, Piquot Y, Pauwels M, Roby D, Cuguen J, Bergelson J, Roux F. 2013. Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in *Arabidopsis thaliana*. *Mol Ecol.* **22**:4222–4240.
- Bubier JA, Chesler EJ, Weinstock GM. 2021. Host genetic control of gut microbiome composition. *Mamm Genome.* **32**:263–281.
- Bulgarelli D, Schlaeppi K, Spaepen S, Ver Loren van Themaat E, Schulze-Lefert P. 2013. Structure and functions of the bacterial microbiota of plants. *Annu Rev Plant Biol.* **64**:807–838.
- Busby PE, Soman C, Wagner MR, Friesen ML, Kremer J, Bennett A, Morsy M, Eisen JA, Leach JE, Dangl JL. 2017. Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLoS Biol.* **15**:e2001793.
- Castrillo G, Teixeira PJPL, Paredes SH, Law TF, de Lorenzo L, Feltcher ME, Finkel OM, Breakfield NW, Mieczkowski P, Jones CD, et al. 2017. Root microbiota drive direct integration of phosphate stress and immunity. *Nature.* **543**:513–518.
- Chen T, Nomura K, Wang X, Sohrabi R, Xu J, Yao L, Paasch BC, Ma L, Kremer J, Cheng Y, et al. 2020. A plant genetic network for preventing dysbiosis in the phyllosphere. *Nature.* **580**:653–657.
- Conway JR, Lex A, Gehlenborg N. 2017. Upsetr: an R package for the visualization of intersecting sets and their properties. *Bioinformatics.* **33**:2938–2940.
- Copeland JK, Yuan L, Layeghifard M, Wang PW, Guttman DS. 2015. Seasonal community succession of the phyllosphere microbiome. *Mol Plant Microbe Interact.* **28**:274–285.
- Cregger MA, Veach AM, Yang ZK, Crouch MJ, Vilgalys R, Tuskan GA, Schadt CW. 2018. The *Populus* holobiont: dissecting the effects of plant niches and genotype on the microbiome. *Microbiome.* **6**:31.
- Debieu M, Huard-Chauveau C, Genissel A, Roux F, Roby D. 2016. Quantitative disease resistance to the bacterial pathogen *Xanthomonas campestris* involves an is immune receptor pair and a gene of unknown function. *Mol Plant Pathol.* **17**:510–520.
- Demirjian C, Razavi N, Desaint H, Lonjon F, Genin S, Roux F, Berthomé R, Vaillau F. 2022. Study of natural diversity in response to a key pathogenicity regulator of *Ralstonia solanacearum* reveals new susceptibility genes in *Arabidopsis thaliana*. *Mol Plant Pathol.* **23**:321–338.
- Demirjian C, Vaillau F, Berthomé R, Roux F. 2023. Genome-wide association studies in plant pathosystems: success or failure? *Trends Plant Sci.* **28**:471–485.
- De Mita S, Thuillet A-C, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y. 2013. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol Ecol.* **22**:1383–1399.
- Deng S, Caddell DF, 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 J.* **15**:3181–3194.
- Desaint H, Aoun N, Deslandes L, Vaillau F, Roux F, Berthomé R. 2021. Fight hard or die trying: when plants face pathogens under heat stress. *New Phytol.* **229**:712–734.
- Donohue K. 2002. Germination timing influences natural selection on life-history characters in *Arabidopsis thaliana*. *Ecology.* **83**:1006.
- Donohue K, Dorn L, Griffith C, Kim E, Aguilera A, Polisetty CR, Schmitt J. 2005. Niche construction through germination cueing: life-history responses to timing of germination in *Arabidopsis thaliana*. *Evolution.* **59**:771–785.
- Escudero-Martinez C, Bulgarelli D. 2019. Tracing the evolutionary routes of plant-microbiota interactions. *Curr Opin Microbiol.* **49**:34–40.
- Escudero-Martinez C, Coulter M, Alegria Terrazas R, Foito A, Kapadia R, Pietrangelo L, Maver M, Sharma R, Aprile A, Morris J, et al. 2022. Identifying plant genes shaping microbiota composition in the barley rhizosphere. *Nat Commun.* **13**:3443.
- Fabiańska I, Pesch L, Koebeke E, Gerlach N, Bucher M. 2020. Neighboring plants divergently modulate effects of loss-of-function in maize mycorrhizal phosphate uptake on host physiology and root fungal microbiota. *PLoS One.* **15**:e0232633.
- Fariello MI, Boitard S, Mercier S, Robelin D, Faraut T, Arnould C, Recoquillay J, Bouchez O, Salin G, Dehais P, et al. 2017. Accounting for linkage disequilibrium in genome scans for selection without individual genotypes: the local score approach. *Mol Ecol.* **26**:3700–3714.
- Fitzpatrick CR, Salas-González I, Conway JM, Finkel OM, Gilbert S, Russ D, Teixeira PJPL, Dangl JL. 2020. The plant microbiome: from ecology to reductionism and beyond. *Annu Rev Microbiol.* **74**:81–100.
- Frachon L, Arrigo L, Rusman Q, Poveda L, Qi W, Scopece G, Schiestl FP. 2023. Putative signal of generalist plant species adaptation to local pollinator communities and abiotic factors. *Mol Biol Evol.* **40**(3):msad036.
- Frachon L, Bartoli C, Carrère S, Bouchez O, Chaubet A, Gautier M, Roby D, Roux F. 2018. A genomic map of climate adaptation

- in *Arabidopsis thaliana* at a micro-geographic scale. *Front Plant Sci.* **9**:967.
- Frachon L, Libourel C, Villoutreix R, Carrère S, Glorieux C, Huard-Chauveau C, Navascués M, Gay L, Vitalis R, Baron E, et al. 2017. Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time. *Nat Ecol Evol.* **1**:1551–1561.
- Frachon L, Mayjonade B, Bartoli C, Hautekèete N-C, Roux F. 2019. Adaptation to plant communities across the genome of *Arabidopsis thaliana*. *Mol Biol Evol.* **36**:1442–1456.
- Gautier M. 2015. Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics.* **201**:1555–1579.
- Geremia RA, Puşcaş M, Zinger L, Bonneville J-M, Choler P. 2016. Contrasting microbial biogeographical patterns between anthropogenic subalpine grasslands and natural alpine grasslands. *New Phytol.* **209**:1196–1207.
- Glick BR, Gamalero E. 2021. Recent developments in the study of plant microbiomes. *Microorganisms.* **9**:1533.
- Gomez A, Espinoza JL, Harkins DM, Leong P, Saffery R, Bockmann M, Torralba M, Kuelbs C, Kodukula R, Inman J, et al. 2017. Host genetic control of the oral microbiome in health and disease. *Cell Host Microbe.* **22**:269–278.e3.
- Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science.* **334**:83–86.
- Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. 2012. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol.* **10**:497–506.
- Hay A, Tsiantis M. 2016. *Cardamine hirsuta*: a comparative view. *Curr Opin Genet Dev.* **39**:1–7.
- He X, Zhang Q, Li B, Jin Y, Jiang L, Wu R. 2021. Network mapping of root-microbe interactions in *Arabidopsis thaliana*. *NPJ Biofilms Microbiomes.* **7**:72.
- Horton MW, Bodenhausen N, Beilsmith K, Meng D, Muegge BD, Subramanian S, Vetter MM, Vilhjálmsson BJ, Nordborg M, Gordon JJ, et al. 2014. Genome-wide association study of *Arabidopsis thaliana* leaf microbial community. *Nat Commun.* **5**:5320.
- Hou S, Thiergart T, Vannier N, Mesny F, Ziegler J, Pickel B, Hacquard S. 2021. A microbiota-root-shoot circuit favours *Arabidopsis* growth over defence under suboptimal light. *Nat Plants.* **7**:1078–1092.
- Huang AC, Jiang T, Liu Y-X, Bai Y-C, Reed J, Qu B, Goossens A, Nützmann H-W, Bai Y, Osbourn A. 2019. A specialized metabolic network selectively modulates *Arabidopsis* root microbiota. *Science.* **364**:6440.
- Hubbard CJ, Brock MT, van Diepen LT, Maignien L, Ewers BE, Weig C. 2018. The plant circadian clock influences rhizosphere community structure and function. *ISME J.* **12**:400–410.
- Humphrey PT, Whiteman NK. 2020. Insect herbivory reshapes a native leaf microbiome. *Nat Ecol Evol.* **4**:221–229.
- Jacoby R, Peukert M, Succurro A, Koprivova A, Kopriva S. 2017. The role of soil microorganisms in plant mineral nutrition—current knowledge and future directions. *Front Plant Sci.* **8**:1617.
- Karasov TL, Almario J, Friedemann C, Ding W, Giolai M, Heavens D, Kersten S, Lundberg DS, Neumann M, Regalado J, et al. 2018. *Arabidopsis thaliana* and *Pseudomonas* pathogens exhibit stable associations over evolutionary timescales. *Cell Host Microbe.* **24**:168–179.e4.
- Karasov TL, Kniskern JM, Gao L, DeYoung BJ, Ding J, Dubiella U, Lastra RO, Nallu S, Roux F, Innes RW, et al. 2014. The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature.* **512**:436–440.
- Karasov TL, Neumann M, Shirsekar G, Monroe G, Weigel D, Schwab R, Team PATHODOPSIS. 2022. Drought selection on *Arabidopsis* populations and their microbiomes. *bioRxiv.* 10.1101/2022.04.08.487684.
- Keating BA, Carberry PS, Bindraban PS, Asseng S, Meinke H, Dixon J. 2010. Eco-efficient agriculture: concepts, challenges, and opportunities. *Crop Sci.* **50**:S-109–S-119.
- Kong HG, Song GC, Sim H-J, Ryu C-M. 2021. Achieving similar root microbiota composition in neighbouring plants through air-borne signalling. *ISME J.* **15**:397–408.
- Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, Bonnette J, Juenger TE, Hyma K, Acharya C, Mitchell SE, et al. 2015. Genome-environment associations in sorghum landraces predict adaptive traits. *Sci Adv.* **1**:e1400218.
- Lebeis SL, Paredes SH, Lundberg DS, Breakfield N, Gehring J, McDonald M, Malfatti S, Glavina del Rio T, Jones CD, Tringe SG, et al. 2015. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science.* **349**:860–864.
- Li F, Zhang X, Gong J, Liu L, Yi Y. 2018. Specialized core bacteria associate with plants adapted to adverse environment with high calcium contents. *PLoS One.* **13**:e0194080.
- Libourel C, Baron E, Lenglet J, Amsellem L, Roby D, Roux F. 2021. The genomic architecture of competitive response of *Arabidopsis thaliana* is highly flexible among plurispecific neighborhoods. *Front Plant Sci.* **12**:741122.
- Lopez-Arboleda WA, Reinert S, Nordborg M, Korte A. 2021. Global genetic heterogeneity in adaptive traits. *Mol Biol Evol.* **38**:4822–4831.
- López-Hernández F, Cortés AJ. 2019. Last-generation genome-environment associations reveal the genetic basis of heat tolerance in common bean (*Phaseolus vulgaris* L.). *Front Genet.* **10**:954.
- Ma K-W, Niu Y, Jia Y, Ordon J, Copeland C, Emonet A, Geldner N, Guan R, Stolze SC, Nakagami H, et al. 2021. Coordination of microbe-host homeostasis by crosstalk with plant innate immunity. *Nat Plants.* **7**:814–825.
- Makiola A, Holdaway RJ, Wood JR, Orwin KH, Glare TR, Dickie IA. 2022. Environmental and plant community drivers of plant pathogen composition and richness. *New Phytol.* **233**:496–504.
- Meyer KM, Porch R, Muscettola IE, Vasconcelos ALS, Sherman JK, Metcalf CJE, Lindow SE, Koskella B. 2022. Plant neighborhood shapes diversity and reduces interspecific variation of the phyllosphere microbiome. *ISME J.* **16**:1376–1387.
- Mittelstrass J, Sperone FG, Horton MW. 2021. Using transects to disentangle the environmental drivers of plant-microbiome assembly. *Plant Cell Environ.* **44**:3745–3755.
- Mitter B, Brader G, Pfaffenbichler N, Sessitsch A. 2019. Next generation microbiome applications for crop production — limitations and the need of knowledge-based solutions. *Curr Opin Microbiol.* **49**:59–65.
- Müller DB, Vogel C, Bai Y, Vorholt JA. 2016. The plant microbiota: systems-level insights and perspectives. *Annu Rev Genet.* **50**:211–234.
- Neal AL, Glendinning MJ. 2019. Calcium exerts a strong influence upon phosphohydrolase gene abundance and phylogenetic diversity in soil. *Soil Biol Biochem.* **139**:107613.
- Oyserman BO, Cordovez V, Flores SS, Leite MFA, Nijveen H, Medema MH, Raaijmakers JM. 2021. Extracting the GEMs: genotype, environment, and microbiome interactions shaping host phenotypes. *Front Microbiol.* **11**:574053.
- Oyserman BO, Flores SS, Griffioen T, Pan X, van der Wijk E, Pronk L, Lohorst W, Nurfikari A, Paulson JN, Movassagh M, et al. 2022. Disentangling the genetic basis of rhizosphere microbiome assembly in tomato. *Nat Commun.* **13**:3228.
- Parker IM, Saunders M, Bontrager M, Weitz AP, Hendricks R, Magarey R, Suiter K, Gilbert GS. 2015. Phylogenetic structure and host abundance drive disease pressure in communities. *Nature.* **520**:542–544.
- Pieterse CMJ, Zamioudis C, Berendsen RL, Weller DM, Van Wees SCM, Bakker PAHM. 2014. Induced systemic resistance by beneficial microbes. *Annu Rev Phytopathol.* **52**:347–375.
- Pluess AR, Frank A, Heiri C, Lalagüe H, Vendramin GG, Oddou-Muratario S. 2016. Genome-environment association study suggests local adaptation to climate at the regional scale in *Fagus sylvatica*. *New Phytol.* **210**:589–601.

- Provart N, Zhu T. 2003. A browser-based functional classification superviewer for *Arabidopsis* genomics. *Curr Protoc Mol Biol*. 271–272.
- Ramírez-Sánchez D, Duflos R, Gibelin-Viala C, Zamar R, Vaillau F, Roux F. 2022. The genetic architecture of *Arabidopsis thaliana* in response to native non-pathogenic leaf bacterial species revealed by GWA mapping in field conditions. *Biorxiv*. 10.1101/2022.09.19.508615.
- Ramírez-Sánchez D, Gibelin-Viala C, Mayjonade B, Duflos R, Belmonte E, Pailler V, Bartoli C, Carrere S, Vaillau F, Roux F. 2022. Investigating genetic diversity within the most abundant and prevalent non-pathogenic leaf-associated bacteria interacting with *Arabidopsis thaliana* in natural habitats. *Front Microbiol*. **13**:984832.
- Robertson-Albertyn S, Alegria Terrazas R, Balbirnie K, Blank M, Janiak A, Szarejko I, Chmielewska B, Karcz J, Morris J, Hedley PE, et al. 2017. Root hair mutations displace the barley rhizosphere microbiota. *Front Plant Sci*. **8**:1094.
- Roman-Reyna V, Pinili D, Borja FN, Quibod IL, Groen SC, Alexandrov N, Mauleon R, Oliva R. 2020. Characterization of the leaf microbiome from whole-genome sequencing data of the 3000 rice genomes project. *Rice (N Y)*. **13**:72.
- Rosenberg E, Zilber-Rosenberg I. 2018. The hologenome concept of evolution after 10 years. *Microbiome*. **6**:78.
- Roux F, Bergelson J. 2016. The genetics underlying natural variation in the biotic interactions of *Arabidopsis thaliana*: the challenges of linking evolutionary genetics and community ecology. *Curr Top Dev Biol*. **119**:111–156.
- Ryu EP, Davenport ER. 2022. Host genetic determinants of the microbiome across animals: from *Caenorhabditis elegans* to cattle. *Annu Rev Anim Biosci*. **10**:203–226.
- Salas-González I, Reyt G, Flis P, Custódio V, Gopaulchan D, Bakhoun N, Dew TP, Suresh K, Franke RB, Dangl JL, et al. 2021. Coordination between microbiota and root endodermis supports plant mineral nutrient homeostasis. *Science*. **371**: eabd0695.
- Santos-García D, Mestre-Rincon N, Zchori-Fein E, Morin S. 2020. Inside out: microbiota dynamics during host-plant adaptation of whiteflies. *ISME J*. **14**:847–856.
- Schlaeppli K, Dombrowski N, Oter RG, Ver Loren van Themaat E, Schulze-Lefert P. 2014. Quantitative divergence of the bacterial root microbiota in *Arabidopsis thaliana* relatives. *Proc Natl Acad Sci U S A*. **111**:585–592.
- Soldan R, Fusi M, Cardinale M, Daffonchio D, Preston GM. 2021. The effect of plant domestication on host control of the microbiota. *Commun Biol*. **4**:936.
- Suzuki TA, Phifer-Rixey M, Mack KL, Sheehan MJ, Lin D, Bi K, Nachman MW. 2019. Host genetic determinants of the gut microbiota of wild mice. *Mol Ecol*. **28**:3197–3207.
- Tang J, Tang X, Qin Y, He Q, Yi Y, Ji Z. 2019. Karst rocky desertification progress: soil calcium as a possible driving force. *Sci Total Environ*. **649**:1250–1259.
- Thiergart T, Durán P, Ellis T, Vannier N, Garrido-Oter R, Kemen E, Roux F, Alonso-Blanco C, Ågren J, Schulze-Lefert P, et al. 2020. Root microbiota assembly and adaptive differentiation among European *Arabidopsis* populations. *Nat Ecol Evol*. **4**:122–131.
- Toju H, Peay KG, Yamamichi M, Narisawa K, Hiruma K, Naito K, Fukuda S, Ushio M, Nakaoka S, Onoda Y, et al. 2018. Core microbiomes for sustainable agroecosystems. *Nat Plants*. **4**:247–257.
- Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK. 2020. Plant-microbiome interactions: from community assembly to plant health. *Nat Rev Microbiol*. **18**:607–621.
- VanWallendaal A, Benucci GMN, da Costa PB, Fraser L, Sreedasyam A, Fritschi F, Juenger TE, Lovell JT, Bonito G, Lowry DB. 2022. Host genotype controls ecological change in the leaf fungal microbiome. *PLoS Biol*. **20**:e3001681.
- Wagner MR. 2021. Prioritizing host phenotype to understand microbiome heritability in plants. *New Phytol*. **232**:502–509.
- Walters WA, Jin Z, Youngblut N, Wallace JG, Sutter J, Zhang W, González-Peña A, Peiffer J, Koren O, Shi Q, et al. 2018. Large-scale replicated field study of maize rhizosphere identifies heritable microbes. *Proc Natl Acad Sci U S A*. **115**:7368–7373.
- Wang X, Feng H, Wang Y, Wang M, Xie X, Chang H, Wang L, Qu J, Sun K, He W, et al. 2021. Mycorrhizal symbiosis modulates the rhizosphere microbiota to promote rhizobia-legume symbiosis. *Mol Plant*. **14**:503–516.
- Wang M, Roux F, Bartoli C, Huard-Chauveau C, Meyer C, Lee H, Roby D, McPeck MS, Bergelson J. 2018. Two-way mixed-effects methods for joint association analysis using both host and pathogen genomes. *Proc Natl Acad Sci U S A*. **115**:E5440–E5449.
- Weissbrod O, Rothschild D, Barkan E, Segal E. 2018. Host genetics and microbiome associations through the lens of genome wide association studies. *Curr Opin Microbiol*. **44**:9–19.
- Worrhich A, Musat N, Harms H. 2019. Associational effects in the microbial neighborhood. *ISME J*. **13**:2143–2149.
- Zhang J, Liu Y-X, Zhang N, Hu B, Jin T, Xu H, Qin Y, Yan P, Zhang X, Guo X, et al. 2019. NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice. *Nat Biotechnol*. **37**:676–684.
- Zhu Y-G, Xiong C, Wei Z, Chen Q-L, Ma B, Zhou S-Y-D, Tan J, Zhang L-M, Cui H-L, Duan G-L. 2022. Impacts of global change on the phyllosphere microbiome. *New Phytol*. **234**:1977–1986.