HAL
archives-ouvertes.fr

Adaptation to plant communities across the genome of Arabidopsis thaliana

Léa Frachon, Baptiste Mayjonade, Claudia Bartoli, Nina-Coralie Hautekèete, Fabrice Roux

▶ **To cite this version:**

Léa Frachon, Baptiste Mayjonade, Claudia Bartoli, Nina-Coralie Hautekèete, Fabrice Roux. Adaptation to plant communities across the genome of Arabidopsis thaliana. Molecular Biology and Evolution, Oxford University Press (OUP), 2019, 36 (7), pp.1442-1456. 10.1093/molbev/msz078 . hal-02265398

HAL Id: hal-02265398
https://hal.archives-ouvertes.fr/hal-02265398

Submitted on 11 Oct 2019

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.

Article. MBE section discoveries.

Title: Adaptation to plant communities across the genome of *Arabidopsis thaliana*

Frachon L.[1,2,3¶], Mayjonade B.[1¶], Bartoli C.[4¶], Hautekeete N.C[5]. & Roux F.[1*]

Author affiliations

[1] LIPM, Université de Toulouse, INRA, CNRS, Castanet-Tolosan, France

[2] Dipartimento di Biologia, Università degli Studi di Napoli Federico II, Naples, Italy

[3] Department of Systematic and Evolutionary Botany, University of Zürich, Zürich, Switzerland

[4] IGEPP, INRA, AGROCAMPUS OUEST, Université Rennes, 35650 Le Rheu, France

[5] Laboratoire Evolution, Ecologie et Paléontologie, CNRS UMR 8198, Université de Lille, F-59655 Villeneuve d'Ascq Cedex, France

[¶] These authors contributed equally to this work.

[*]***Corresponding author***: Fabrice Roux

LIPM, Université de Toulouse, INRA, CNRS, Castanet-Tolosan, France

E-mail: fabrice.roux@inra.fr

Phone number: +33 5 61 28 55 57

***Short title*:** Genomic adaptation to plant communities

Keywords: plant-plant interactions, Genome – Environment Association, community ecology, local adaptation

Abstract

Despite the importance of plant-plant interactions on plant community dynamics and crop yield, our understanding of the adaptive genetics underlying these interactions is still limited and deserves to be investigated in the context of complex and diffuse interactions occurring in plant assemblages. Here, based on 145 natural populations of *Arabidopsis thaliana* located in south-west of France and characterized for plant communities, we conducted a Genome-Environment Association analysis to finely map adaptive genomic regions of *A. thaliana* associated with plant community descriptors. To control for correlated abiotic environment effects, we also characterized the populations for a set of biologically meaningful climate and soil variables. A non-negligible fraction of top SNPs was associated with both plant community descriptors and abiotic variables, highlighting the importance of considering the actual abiotic drivers of plant communities to disentangle genetic variants for biotic adaptation from genetic variants for abiotic adaptation. The adaptive loci associated with species abundance were highly dependent on the identity of the neighboring species suggesting a high degree of biotic specialization of *A. thaliana* to members of its plant interaction network. Moreover, the identification of adaptive loci associated with α-diversity and composition of plant communities supports the ability of *A. thaliana* to interact simultaneously with multiple plant neighbors, which in turn can help to understand the role of community-wide selection. Altogether, our study highlights that dissecting the genetic basis underlying plant-plant interactions at a regional scale while controlling for abiotic confounding factors, can help understanding the adaptive mechanisms modulating natural plant assemblages.

Introduction

Understanding the genetics underlying plant-plant interactions is a key element to understand the structure and functioning of natural communities (Whitham et al. 2006). In particular, identifying the genetic basis of plant-plant interactions can help to estimate the potential of plant species to face anthropogenic-related modifications of plant assemblages (Pierik et al. 2013), resulting in part from differences of geographic range shift among plant species under climate change (Gilman et al. 2010; Singer et al. 2013). In addition, the identification of genes associated with natural variation of response to the presence of other plants is of primary importance to improve plant breeding programs for the optimization of mixtures of crop species (i.e. ideomixes) (Litrico and Violle 2015).

However, there is still very limited information about the adaptive genetic loci associated with natural variation of plant-plant interactions (Subrahmaniam et al. 2018). Firstly, the number of Quantitative Trait Loci (QTL) studies focusing on plant-plant interactions is relatively limited in comparison to other types of biotic interactions (virus, bacteria, fungi, oomycetes, herbivores…) (Bartoli and Roux 2017; Subrahmaniam et al. 2018). Similarly, despite the importance of interspecific competition in mediating plant community structure, diversity and dynamics (Baron et al. 2015), only two Genome Wide Association studies (GWAS) reported the fine mapping of genomic regions underlying the response of a focal species to the presence of a competitive species (Baron et al. 2015; Frachon et al. 2017). In addition, due to the intractability of empirically testing interactions between a focal species and many neighboring plant species, most QTL mapping studies have considered only a single pair of interacting species (Subrahmaniam et al. 2018). However, throughout their life cycle, plants can interact separately or simultaneously with a large number of plant species (Wilson et al. 2012), suggesting that the genetics of plant-plant interactions needs to be investigated in the context of complex and diffuse interactions occurring in natural plant assemblages (Litrico and Violle 2015; Roux and Bergelson 2016). Secondly, to our knowledge, it is still unknown whether polymorphic genes involved in plant-plant interactions have been shaped by natural selection.

By identifying significant associations between genetic polymorphisms and environmental variables, Genome-Environment Association (GEA) analyses is a powerful genome scan method to identify genes potentially involved in adaptive processes of a given species (De Mita et al. 2013; Scalfi et al. 2014). In the last few years, the availability of public databases collecting estimates of abiotic factors (in particular climatic variables) and the

development of next-generation sequencing (NGS) technologies, led to a burst of GEA studies attempting to establish genomic map of local adaptation to abiotic variation, from a worldwide scale (Hancock et al. 2011; Lasky et al. 2012; Lasky et al. 2015; Bay et al. 2017; Ferrero-Serrano & Assmann 2019) to a regional scale (Pluess et al. 2016; Frachon et al. 2018). On the other hand, GEA studies performed on biotic factors are still scarce (not to say absent) in wild plant species. Two non-exclusive hypotheses can explain this paucity of GEA studies related to biotic interactions. Firstly, the power of GEA to identify true positives is positively correlated to the number of studied populations (Gautier 2015). However, given the substantial effort required to carefully characterize ecological communities, the number of studies reporting the biotic characterization of several dozens or hundreds natural populations of a given plant species is unsurprisingly limited (Züst et al. 2012; Brachi et al. 2013; Bartoli et al. 2018). Secondly, GEAs performed on biotic variables need to control for the abiotic environment that can be correlated with the variation of the biotic features under study. This step is fundamental to dissociate the genetic variants associated with the biotic variables from those associated with the only abiotic adaptation of the plant species. For example, it is well established that under various geographic scales, plant assemblages are shaped by a combination of abiotic factors (Hautekèete et al. 2014; Trivellone et al. 2017). Removing the effect of habitat filtering – by characterizing and controlling for several abiotic factors - would allow to identify the adaptive genetic basis that are specific to biotic interactions, i.e. that do not result from an indirect relationship with the abiotic environment (e.g. climate or soil factors).

Here, by combining plant community ecology and population genomics, we adopted a GEA approach to establish a genomic map of potential adaptive genetic loci associated with plant community descriptors (i.e. α-diversity, composition and abundance of plant species) in the model plant *Arabidopsis thaliana*. To achieve this goal, using a metabarcoding approach, we first characterized the plant communities associated with 145 natural populations of *A. thaliana* located in south-west of France (Bartoli et al. 2018; Frachon et al. 2018). To control for correlated effects of the abiotic environment, those populations were also characterized for a set of 17 biologically meaningful climate and soil variables. By using a Bayesian hierarchical model controlling for the genome wide effects of confounding demographic evolutionary forces (Gautier 2015), we then conducted GEA analyses with more than 1.5 million Single Nucleotide Polymorphisms (SNPs) to finely map genomic regions of *A. thaliana* associated with plant community descriptors. GEA was also performed on abiotic variables, thereby allowing to disentangle GEA signals for plant-plant interactions from GEA signals for abiotic factors. By

definition, GEA allows identifying genetic loci under local adaptation. However, to support that the loci identified by our GEA analysis have been shaped by natural selection, we additionally tested whether the SNPs that were the most associated with descriptors of plant communities were enriched in a set of SNPs subjected to adaptive spatial differentiation. Finally, we examined whether specific biological processes were overrepresented among SNPs involved in adaptation to plant communities and discussed the function of candidate genes.

Results

Working at a regional scale to investigate the adaptive genetics to plant communities

To establish a genomic map of local adaptation to plant communities, we focused on 145 *A. thaliana* populations located in the Midi-Pyrénées region (south-west of France) (**Fig. 1A**), with a median distance among the populations of 94.5 km (max = 264.7 km) (Frachon et al. 2018). The choice of working at regional scale was supported by previous observations on *A. thaliana* in France. Firstly, as already advised, the geographical scale used to identify genomic regions associated with environmental variables should be chosen according to the grain of ecological variation (Bergelson & Roux 2010). A previous study focusing on 49 natural populations of *A. thaliana* located in four climatically contrasted French regions (Brittany, Burgundy, Languedoc and north of France), showed that the variance of interspecific competition indices was mostly partitioned among populations within regions, with only 5.2% of the competition variance observed among the four regions (Brachi et al. 2013). Secondly, the main drawbacks of GEA analyses (i.e. allele frequency autocorrelation, genetic and allelic heterogeneity, rare alleles) are often observed at large geographical scales. As previously advised for GWAS, working at a small geographical scale should reduce these limitations (Bergelson & Roux 2010). Accordingly, in a GWAS performed on flowering time scored in greenhouse conditions, confounding by population structure was greatly reduced at a regional scale in France (Brachi et al. 2013), which is consistent with a linkage disequilibrium estimate (r^2=0.5) of 18bp in a highly polymorphic local population of *A. thaliana* located in Burgundy (Frachon et a. 2017). Thirdly, the genome sequencing of the 145 natural populations used in this study revealed that (i) all populations were polymorphic (in agreement with previous observations obtained on populations located in other French regions; Le Corre 2005, Brachi *et al.* 2013), and (ii) less than 11% of genomic variation was explained by geographic variation (Frachon et al. 2018), as expected from the pattern of isolation by distance observed across the species range of *A. thaliana* (Platt et al. 2010). Taken together, these observations suggest that

working at a regional scale would allow to work on contrasted plant assemblages, while limiting the effects of confounding by population structure on the power of GEA to identify true positives.

The plant neighborhood of *A. thaliana* is highly diverse and partially related to a combination of abiotic factors

Plant communities associated with the 145 natural populations of *A. thaliana* were characterized during spring 2015 (mid-May to mid-June) corresponding to the period of seed production of *A. thaliana* in south-west of France. Based on morphological characteristics, we characterized plant communities by first establishing a herbarium by collecting a representative individual of each putative species per population, resulting in 2,233 specimens. Because many specimens were sampled at the seedling stage or without the presence of reproductive organs commonly used for a morphologically-based taxonomical identification, we adopted a metabarcoding approach based on the chloroplast marker *matK* to determine the identity of the species (**Supplementary Fig. 1;** Barthet and Hilu 2007). A *matK* sequence was obtained for 97% of the specimens that were further assigned to one of the 244 plant Operational Taxonomic Units (OTUs) identified at a 98% identity cutoff (**Datasets 1 and 2, Supplementary Fig. 1**). In agreement with the deep taxonomic resolution of the marker *matK* (Barthet and Hilu 2007*)*, a large portion of specimens were identified at the species level (84.6%) (**Datasets 1 and 2, Supplementary Fig. 1**). Hereafter, we refer to the plant OTUs as specific plant community descriptors.

Alpha-diversity largely differed among the 145 populations (**Supplementary Table 1**), with species richness ranging from 1 to 28 (mean = 12.1, median=12) and Shannon index ranging from 0.32 to 2.42 (mean = 1.40, median=1.35) (**Fig. 1B and Supplementary Fig. 2**). The plant community composition was studied by running a Principal Coordinate Analysis (PCoA) on the abundance matrix of the 44 most prevalent plant OTUs (i.e. OTUs present in more than 10 populations) (**Dataset 3**). The first three PCoA axes explained ~34% of the variation in plant composition (**Supplementary Fig. 3**). Plant composition largely differed among the 145 populations (**Fig. 1C, D, E and F, Supplementary Fig. 3**), with up to 75.9% of variance explained by the 'population' factor (**Supplementary Table 1**). The first PCoA axis (explaining 14.2% of the total variance, **Supplementary Fig. 3**) was mainly associated with annual species occurring in bare tilled, fallow or recently abandoned arable lands (EUNIS habitat I1.5, e.g. *Bromus hordeaceus*, *Cerastium glomeratum*, *Sonchus oleraceus*, *Veronica arvensis*) and with perennial species occurring in mesic grasslands (EUNIS habitat E 2, e.g.,

Crepis biennis, *Festuca rubra*, *Plantago lanceolata*) (**Dataset 4, Supplementary Fig. 4**). The second PCoA axis (11.1%, **Supplementary Fig. 3**) was significantly associated with (i) species occurring in the same habitats (I 1.5 e.g. *B. hordeaceus*, *Trifolium campestre*, E 2 e.g. *Achillea millefolium*) on the one side of the axis, and (ii) small annual and pioneer species occurring in grasslands or arable lands (E2 and I 1.5, e.g. *Poa annua*, *Anagallis arvensis*) on the other side (**Dataset 4, Supplementary Fig. 4**). The third PCoA axis (8.6%, **Supplementary Fig. 3**) was significantly associated with mainly short-lived species occurring in bare tilled lands (EUNIS I 1.51) or fallow un-inundated fields or recently abandoned arable lands with annual weed communities (EUNIS I 1.52), but with an opposition considering their soil preferences: mesic to nitrophilous species very typical of *A. thaliana* habitat preferences on the one side of the axis (e.g. *Valerianella locusta*) vs species with a preference for sandy soils on the other side (i.e. *F. rubra*, *Arenaria serpillyfolia*, *Lactuca serriola*) (**Dataset 4, Supplementary Fig. 4**). Hereafter, we refer to species richness, Shannon index and the first three PCoA axes as global plant community descriptors.

To estimate the level of dependency of the plant community descriptors towards the abiotic environment, we first characterized the 145 natural populations for 6 non-correlated climate variables and 11 non-correlated soil variables (**Fig. 2, Dataset 5**). We then run a sparse Partial Least Square Regression (sPLSR) (Lê Cao et al. 2008) on each global descriptor and the 44 most prevalent plant OTUs (**Fig. 2**). We found that up to 27% of variance of plant community descriptors can be explained by a linear combination of abiotic variables (min = 8.4%, mean = 19%, median = 19.8%) (**Fig. 2**). On average, each plant community descriptor was explained by a linear combination of 4.5 abiotic variables (min = 1, median = 5, max = 6) (**Fig. 2**). The four abiotic variables that were the most often associated with plant community descriptors (n > 20) were mean annual temperature, summer precipitations, autumn precipitations and soil pH (**Fig. 2**). Among these four variables, pH was the one mostly associated with plant community descriptors, suggesting that pH is one of the main abiotic variables driving variation in plant assemblages associated with *A. thaliana* in the studied ~31,000 km² spatial scale (**Fig. 2**).

Identification of adaptive genetic loci associated with plant community descriptors

To identify associations between genetic polymorphisms of *A. thaliana* and plant community descriptors, we adopted a GEA approach based on within-population allele frequency previously estimated in the 145 *A. thaliana* populations for 1,519,748 SNPs (i.e. one SNP every 78 bp) (Frachon et al. 2018). Variation of plant community descriptors was weakly

correlated to genomic variation among the 145 populations (**Supplementary Table 2**). According to a sensitivity analysis testing for the performance of GEA analyses based on these populations (Frachon et al. 2018), these weak correlation estimates suggest that a large fraction of the SNPs that were the most associated with variation of plant community descriptors corresponds to true positives.

To disentangle GEA signals for a given plant community descriptor from GEA signals for abiotic variables associated with this plant community descriptor (**Fig. 2)**, GEA analysis was performed on both plant community descriptors and abiotic variables. Our results showed that the proportion of shared top SNPs varied among global (alpha-diversity and composition) and specific (abundance of the 44 most prevalent plant OTUs) plant community descriptors. When considering the 0.1% top SNPs, global descriptors were sharing a high fraction of SNPs with abiotic variables (min=6.5%, max=19.5%, mean=14.4%, median=19.5%, **Table 1**), mainly with soil variables (**Dataset 6**). On the other hand, although a lower fraction of top SNPs was shared between specific plant community descriptors and abiotic variables (min=0, max=4.5%, mean=0.8%, median=0.6%, **Table 1**), some of the shared SNPs were amongst the most associated with these plant community descriptors (**Fig. 3**). For example, the abundance of *Sonchus oleraceus* (OTU27, Asteraceae) was associated with soil pH (**Fig. 2, Fig. 3A**). The SNP that was the most associated with the abundance of *S. oleraceus* was also one of the top SNPs associated with pH (**Fig. 3B, C**). Similarly, the abundance of *Papaver rhoeas* (OTU114, Papaveraceae) was associated with summer precipitations (**Fig. 2, Fig. 3D**). An association peak for the abundance of *P. rhoeas* located at the beginning of chromosome 2 corresponded also to an association peak detected for summer precipitations (**Fig. 3E, F**).

In order to establish a genomic map of local adaptation to plant communities, the top SNPs shared with abiotic variables were therefore not taken into consideration for the rest of the results presented in this study. For species abundance, the genetic architecture was highly dependent on the plant OTU identity (**Fig. 4 and Dataset 6**). For example, an association peak was detected at the beginning of chromosome 3 for the abundance of *Convolvulus arvensis* (OTU83, Convolvulaceae), present in 16.5% of the *A. thaliana* populations, whereas two association peaks were detected at the end of chromosomes 3 and 5 for the abundance of *Poa nemoralis* (OTU149, Poaceae), present in 19.3% of the populations. We also observed that the genetic architecture largely differed between two species belonging to the same botanical family, and at a deeper taxonomy level to the same genus (**Dataset 6**). For example, an association peak was detected at the beginning of chromosome 3 for the abundance of *S.*

oleraceus (OTU27, Asteraceae), present in 34% of the *A. thaliana* populations, whereas an association peak was detected at the beginning of chromosome 1 for the abundance of *Helminthotheca echioides* (OTU16, Astereaceae), present in 12% of the *A. thaliana* populations. Interestingly, we also detected association peaks for α-diversity and composition related descriptors, suggesting the presence of generalist QTLs underlying simultaneous interactions with several neighboring species (**Fig. 4 and Dataset 6**). Moreover, when considering the 152 top SNPs, less than 21% of the top SNPs associated with a given global descriptor were also associated with specific descriptors (min = 12.5% for the third PCoA axis, max = 28.3% for the first PCoA axis), suggesting a flexible genetic architecture between specific and global descriptors.

To support that the loci identified by GEA have been shaped by natural selection, we performed a genome-wide selection scan by estimating the XtX measure of genetic differentiation among the 145 populations. For a given SNP, the XtX measures the variance of the standardized population allele frequencies, which result from a rescaling based on the covariance matrix of population allele frequencies. This step allows correcting for the genome-wide effects of confounding demographic evolutionary forces (Gautier 2015). The 0.1% upper tail of the spatial differentiation distribution displayed a significant enrichment (up to 10.05) for SNPs associated with almost two-thirds of the plant community descriptors, including species richness, Shannon index, the first and third PCoA axes and the abundance of 23 OTUs belonging to 11 botanical families (**Table 1**). Interestingly, similar enrichment values were observed with or without considering the top SNPs shared between plant community descriptors and abiotic variables (**Table 1**). Altogether, these signatures of selection across the genome support our GEA results suggesting that *A. thaliana* is locally adapted to its associated plant communities.

Identity of candidate genes underlying local adaptation to plant communities

To identify candidate genes associated with plant community descriptors, we adopted three non-exclusive approaches. Firstly, we focused on genes associated with more than three plant community descriptors, resulting in a final list of 30 candidate genes (**Datasets 7 and 8**). Secondly, we examined which biological processes (BPs) were overrepresented among the 152 SNPs the most associated with plant community descriptors. We found a significant enrichment in BPs (up to 81.1 fold for the BP 'Wnt signaling pathway') for the third PCoA axis and the abundance of 17 plant OTUs (**Supplementary Table 3**), leading to the identification of 73

unique candidate genes across 29 BPs (**Dataset 9**). Thirdly, we examined which functional classes (FCs) were overrepresented among the genes identified by the two previous approaches. Using the MapMan classification (Provart & Zhu 2003), we identified three significantly over-represented FCs, i.e. hormone metabolism (P = 0.00004), cell (P = 0.00075) and signaling (P = 0.00198). By combining the results of the three approaches, we found that several candidate genes were involved in responses to shade (**Dataset 8 and 9**), either through *signaling pathways of light perception* such as (i) *TUBULIN BETA-1 CHAIN* (*TUB1*) (Leu et al. 1995), and (ii) *CASEIN KINASE I-LIKE 3* (*CK1.3*) and *CASEIN KINASE I-LIKE 4* (*CK1.4*) (Tan et al. 2013), or through *hormone signaling pathways* including (i) auxin with the auxin-responsive genes *SMALL AUXIN UPREGULATED67* (*SAUR67*) and *SAUR68* (Roig-Villanova et al. 2007; Swain et al. 2017), (ii) ethylene with *XAP5 CIRCADIAN TIMEKEEPER* (*XCT*) involved in blue light-dependent ethylene responses (Ellison et al. 2011), and (iii) brassinosteroid with the bHLH transcription factor *HOMOLOG OF BEE2 INTERACTING WITH IBH 1* (*HBI1*) (Fan et al. 2014). Another major category of candidate genes was related to plant immunity (**Datasets 8 and 9**). In particular, we identified key components of the pathogen-associated molecular patterns PAMP-triggered immunity (PTI) pathway including (i) several receptor-like kinases (RLKs) such as *FLAGELLIN-SENSING 2* (*FLS2*) (Gómez-Gómez & Boller 2000), *BRI1-ASSOCIATED RECEPTOR KINASE* (*BAK1*) (Li et al. 2002), *BAK1-INTERACTING RECEPTOR-LIKE KINASE 1* (*BIR1*) (Gao et al. 2009) and *POLLEN-SPECIFIC RECEPTOR-LIKE KINASE 5* (*PRK5*) (Wrzaczek et al. 2015), (ii) two WRKY transcriptions factors, namely *WRKY8* (Chen et al. 2010) and *WRKY48* (Xing et al. 2008), and (iii) the heterotrimeric G-protein beta subunit *AGB1* (Lorek et al. 2013) and the calmodulin-domain protein kinase *CKP5* (Dubiella et al. 2013). Finally, other candidate genes were related to root development such as the bHLH transcription factor *LRL1* and the arabinogalactan protein *AGP14*, which are both involved in root hair development (Xu et al. 2005; Lin et al. 2011).

Discussion

The identification of the genetic basis underlying adaptation to current environmental conditions is essential to estimate the potential of wild plant and crop species to persist in presence of novel and fluctuating environmental conditions. In the context of climate change, most GEA analyses have focused on the identification of adaptive genetic loci associated with climate variation (Bay et al. 2017), which has certainly been facilitated by public databases availability of climate estimates. However, the adaptive potential response of a plant species to global changes is not only dependent on the direct interaction between the plant species and its abiotic environment, but also on the interactions between the plant species and its biotic partners (Roux and Bergelson 2016). In this study, we therefore aimed at identifying adaptive genetic loci associated with plant community descriptors in 145 natural *A. thaliana* populations. We found that *A. thaliana* can inhabit diverse and contrasted plant assemblages in south-west of France. This observation is in line with previous studies reporting (i) the potential interactions of *A. thaliana* with a large number of plant species in natural communities in other French geographical regions (Brachi et al. 2013), and (ii) the extensive genetic diversity of *A. thaliana* for the response to interspecific competition observed in common garden experiments (Baron et al. 2015; Bartelheimer et al. 2015; Frachon et al. 2017).

Importance of controlling for abiotic factors

In the light of the impact of habitat type on plant community assemblages (Trivellone et al. 2017), controlling for abiotic factors appears indispensable for establishing a genomic map of local adaptation to natural plant communities. Indeed, by characterizing the 145 *A. thaliana* populations for a set of 17 biologically meaningful climate and soil variables, we found that a non-negligible fraction (~19-20%) of variation in plant community descriptors was explained by a combination of climatic and soil factors. In particular, in agreement with previous studies (Dubuis et al. 2013; Gould et al. 1999; Kelly and Goulden 2008), we identified temperature, precipitations and soil pH as the main drivers of diversity and composition of plant communities inhabited by *A. thaliana*. In accordance, as illustrated for *S. oleraceus* and *P. rhoeas*, GEA results showed that soil pH and summer precipitations were sharing some of the top SNPs with plant species abundance.

Altogether, the non-negligible portion of SNPs shared among plant community descriptors and abiotic variables highlights the importance of disentangling GEA signals for biotic adaptation from GEA signals for abiotic adaptation. However, we need to stress out that the abiotic environment cannot be reduced to the only factors described in this study and that co-association with other abiotic variables might be also considered as an indirect effect of the genetic basis identified in this study. While characterizing the whole abiotic environment is a lofty challenge, we advocate that GEA studies performed on biotic interactions need to control for as many abiotic variables as possible. In a complementary way, GWA studies based on the growth of hundreds of whole-genome sequenced accessions within natural plant assemblages transplanted in a common garden, might be a powerful approach to help understanding and validating whether genetic loci associated with either specific plant species neighbors or community diversity/composition are not indirectly linked to the associated abiotic environment. A deeply characterization of the *in situ* abiotic environment in conjunction with a validation of the associated candidate genes by adopting a controlled common garden approach should be advocated for future studies aiming at characterizing the adaptive genetics of plant-plant interactions.

The genomic architecture of adaptation to plant communities

In agreement with a GWAS performed in field conditions with *A. thaliana* (Baron et al. 2015), we found that genomic regions associated with species abundances largely differed among species belonging to different botanical families. We further found that the genetic architecture of *A. thaliana* related to species abundances largely differed among species belonging to the same botanical family, suggesting a high degree of biotic specialization at a regional scale. In addition, we finely mapped adaptive QTLs associated with α-diversity and composition of plant communities, which is in line with the fine mapping of genomic regions in *A. thaliana* associated with α-diversity and composition of leaf microbial communities in a GWAS performed in field conditions (Horton et al. 2014). Interestingly, the QTLs identified for these global descriptors largely differed from the QTLs identified for species abundance. Such a flexible genetic architecture between monospecific and plurispecific interactions has been also recently observed in a GWAS performed in greenhouse conditions on a set of 91 local whole-genome sequenced accessions of *A. thaliana* (Libourel et al. 2019). Those accessions were subjected to monospecific and plurispecific competition treatments based on all one-way,

two-way and three-way combinations of three species frequently associated with *A. thaliana* in natural plant communities in France, *i.e. Poa annua*, *Stellaria media* and *Veronica arvensis* (Baron et al. 2015). The authors found that almost 80% of top SNPs associated with the response of *A. thaliana* to the simultaneous presence of *P. annua*, *S. media* and *V. arvensis* were not detected in the three monospecific interaction treatments. The identification of QTLs that are specific to descriptors of plant assemblages highlights the ability of *A. thaliana* to interact simultaneously with multiple plant neighbors, which in turn can help to understand the role of community-wide selection.

Candidate genes underlying plant-plant interactions at a regional scale

The identity of the candidate genes identified in this study is in line with known molecular mechanisms of plant-plant interactions (Pierik et al. 2013; Subrahmaniam et al. 2018). For example, as sessile organisms, plants compete for both above- and below-ground resources and have several detection mechanisms to identify the presence of neighboring plants. Accordingly, we identified several candidate genes (i) involved in response to altered light environment, generally referred to the shade avoidance syndrome (SAS) or (ii) related to nutrient foraging, thereby linked to below-ground competitive ability. Interestingly, we also identified biological processes associated with plant immunity (**Datasets 9**), which is in line with several transcriptomic studies reporting expression changes of numerous genes associated with defense pathways against microbial pathogens and insects in different types of plant-plant interactions (Subrahmaniam et al. 2018). In addition, the identification of several RLKs involved in PTI is also in agreement with a GWAS on competitive response of *A. thaliana* grown in monospecific and plurispecific neighborhoods, in which 12 out of the 18 candidate genes supporting a significant enrichment for the functional class 'signaling' encode RLKs (Libourel et al. 2019). This enrichment in immunity-related molecular mechanisms may be explained by three non-exclusive hypotheses. Firstly, the growth-defense theory predicts that light perception by photoreceptors actives SAS and reduces defenses against bio-aggressors (Ballaré and Pierik 2017). However, this theory has been recently challenged by several studies reporting an independent regulation of the SAS-related pathways and defense pathways (Subrahmaniam et al. 2018). Secondly, the presence of neighboring species modifies the rhizosphere microbiota surrounding a focal plant (Sanon et al. 2009; Hawkins & Crawford 2018), which may in turn trigger pathways involved in perception of microbes. Thirdly, RLKs

may directly perceive damage-associated molecular patterns (DAMPs) that potentially result from cell wall modification and degradation in *A. thaliana* in presence of neighboring species (Subrahmaniam et al. 2018). For example, a 20-aa fragment of the extracellular *Arabidopsis* protein GRIM REAPER (GRI) is known to directly interact with *PRK5* (Wrzaczek et al. 2015), an atypical RLK associated with species richness and composition of plant communities in our study (**Dataset 8**).

Conclusion

While GEA analysis has been proved to be highly powerful for identifying genetic basis associated with climate adaptation in *A. thaliana* (Hancock et al. 2011, Lasky et al. 2012, Frachon et al. 2018, Ferrero-Serrano & Assmann 2019), our study reveals the potential of GEA analysis to unravel the adaptive genetics underlying plant-plant interactions in the context of realistic community complexity. However, we have to stress out that the main conclusions drawn from our study might be specific to the geographical region studied. Whether similar conclusions patterns are observed either in other geographical regions of similar size or at a larger geographical scale remains an open question and will deserve a careful characterization of a new set of hundreds of populations both at the genomic level and for plant community descriptors.

Because no study reported the cloning of a gene involved in response to interspecific competition in plants so far (to our knowledge), the candidate genes identified in this study undoubtedly represent key candidate genes for functional analysis, which in turn can help to dissect the genetic and molecular mechanisms underlying co-evolutionary processes between *A. thaliana* and its plant interacting network.

Materials and Methods

Characterization of plant communities

We focused on 168 natural populations of *A. thaliana* located in the Midi-Pyrénées region (Bartoli et al. 2018, Frachon et al. 2018). Plant communities associated with *A. thaliana* were characterized during spring 2015 (mid-May to mid-June), that is the period of seed production of *A. thaliana* in south-west of France. Due to anthropogenic perturbations such as herbicide spraying and mowing, we were not able to characterize plant communities associated with 23 *A. thaliana* natural populations.

To characterize plant communities, two 50 x 50 cm quadrats divided into 25 smaller squares (10 cm x 10 cm) were established in two representative areas of each *A. thaliana* population, with the exception of (i) the large population CLAR-A in which three quadrats were established and (ii) three very small populations (BAGNB-B, DAMI-A and MERV-B) in which only a single quadrat was established. Based on morphological aspects, we first determined the number of putative plant species present in each quadrat. We then estimated the abundance of each putative plant species per quadrat by summing the number of individuals (N) estimated in each of the 25 squares according to the following scale: "1". real count if $N \leq 5$, "2". $N = 10$ if $5 < N \leq 10$, "3". $N = 20$ if $10 < N \leq 20$, "4". $N = 50$ if $20 < N \leq 50$ and "5". $N = 70$ if $N > 50$. It should be noted that *A. thaliana* was not present in seven populations at the time of plant community characterization, despite its presence in early-spring 2015 (Bartoli et al. 2018). A herbarium was established by collecting a representative individual of each putative plant species per population, resulting in 2,233 specimens.

Many specimens were sampled at the seedling stage or without the presence of flowers or fruits (i.e. reproductive organs commonly used for a morphological-based identification of plant species). We therefore adopted a metabarcoding approach based on the chloroplast marker *matK* to determine the identity of the specimens at the species (even sub-species) level. A detailed procedure of the metabarcoding approach is given in ***Supplementary Information*** (**Supplementary Fig. 1**). We obtained a *matK* sequence (> 500bp) for 97% of the specimens (n = 2,166). The *matK* sequences were clustered into OTUs using the software USEARCH with a 98% identity cutoff (i.e. 98% similarity between two OTUs) (Edgar 2010; Edgar 2013), resulting in a total of 244 plant OTUs. To identify the species name of OTUs, a reference sequence was retrieved for each plant OTU with the command *–uparse* of the software USEARCH and blasted on NCBI against Nucleotide Collection (nr/nt) database (percentage of identity: mean = 99.1, median = 99.7; alignment length in bp: mean = 739, median = 771;

alignment length in percent: mean = 99.4, median = 100) (**Datasets 1 and 2**). In addition, the presence of the plant species in the Midi-Pyrénées region was checked by using the databases of The French Botany Network (www.tela-botanica.org) and The Global Biodiversity Information Service (www.gbif.org).

We established an abundance matrix of the 244 plant OTUs across the 145 populations (**Dataset 3**). For each quadrat, we estimated the species richness and Shannon index by using the functions 'specnumber' and 'diversity' in the package 'vegan' (Oksanen et al. 2016) in the R environment (Version 1.0.136 – © 2009-2016 Rstudio, Inc.). To estimate plant community composition in each quadrat, we performed a Principal Coordinate Analysis (PCoA) (function 'pcoa' in the *R* package 'ape') (Paradis et al. 2004) on a Bray-Curtis dissimilarity matrix (function 'vegdist' in the *R* package 'vegan') based on the abundance matrix of the 44 most prevalent OTUs (i.e. presence in more than 10 populations).

Characterization of the climate and soil variables

To control for the putative relationships between plant communities and abiotic factors, our initial set of 168 *A. thaliana* populations were characterized for both climate and soil variables. Climatic characterization was previously performed in Frachon et al. (2018). Briefly, an average value across the 2003-2013 annual data of 21 climatic variables was obtained from the ClimateEU databases (hhtp://tinyurl,com/ClimateEU). As described in Frachon et al. (2018), only six climate variables (mean annual temperature, mean coldest month temperature and precipitations in winter, spring, summer, and autumn) were considered here because (i) they did not display inter-correlation (i.e. Spearman's *rho* lower than 0.8) and (ii) they presented the most obvious link to the ecology of *A. thaliana*. For the characterization of soil properties, we followed the procedure described in Brachi *et al.* (2013) by characterizing for each population two samples of the upper soil layer for 13 chemical properties (content of total N and organic C, C/N ratio, content of organic matter, concentrations of P2O5, K, Ca, Mg, Mn, Al, Na and Fe) and one physical property (maximal water holding capacity, WHC) (**Dataset 5**). The two soil samples were collected in autumn 2014 and winter 2015 and were dried in a laboratory dry oven at 50°C for 10 hours and 90% of ventilation. As for climate data, inter-correlation between two soil variables was avoided by considering only variables that did not display a Spearman's *rho* greater than 0.8 (**Dataset 10**), thereby leading to the removal of content of organic C, content of organic matter and concentration of Mn. The characterization

of some of the 11 remaining soil variables failed for seven *A. thaliana* populations, due to the low quantity of soil available in their native habitats (i.e. stone wall, parking lot….). For these populations, we assigned the mean value of each soil variable obtained from the other populations (**Dataset 5**).

Statistical analyses

To test whether the five plant community descriptors (species richness, Shannon diversity and the first three PCoA components) differed among the 145 populations, we ran the following mixed model under the SAS environment with inference performed using ReML estimation (PROC MIXED procedure in SAS9.3, SAS Institute Inc., Cary, North Carolina, USA):

$$Y_i = \mu_{trait} + population_i + \varepsilon_i \qquad (1)$$

where '*Y*' is one of the five descriptors of plant communities, 'μ' is the overall mean; 'population' accounts for differences among populations; 'ε' is the residual term. A Wald test was used to estimate the random effect 'population'. For each plant community descriptor, Best Linear Unbiased Predictions (BLUPs) were obtained for each population.

To test whether the five plant community descriptors displayed a geographic pattern, we ran the following model (PROC GLM procedure in SAS9.3, SAS Institute Inc., Cary, North Carolina, USA):

$$Y_{ijk} = latitude_i + longitude_j + altitude_k + latitude_i*longitude_j + latitude_i*altitude_k + longitude_j*altitude_k + latitude_i*longitude_j*altitude_k + \varepsilon_{ijk} \qquad (2)$$

where '*Y*' is one of the five plant community descriptors; 'longitude', latitude' and 'altitude' have been retrieved from the GPS coordinates of the 145 populations (**Dataset 3**; Frachon et al. 2018); 'ε' is the residual term.

In order to investigate the relationships between plant communities and abiotic factors, a sparse Partial Least Square Regression (sPLSR) (Lê Cao et al. 2008; Carrascal et al. 2009) was adopted to maximize the covariance between each of the 49 following plant community descriptors (species richness, Shannon index, coordinates on three first PCoA axes and abundance of the 44 most prevalent OTUs) and linear combinations of the six climate factors

and 11 soil factors. sPLSR was run using the mixOmics package implemented in the *R* environment (Lê Cao et al. 2008; Lê Cao et al. 2010). sPLSR results were validated by plotting the Root Mean Square Error of Prediction (Maestre 2004; Lê Cao et al. 2008). For the abiotic environment, we calculated the final loadings for the ten abiotic variables with the highest initial loadings on the first component. Following the procedure described in Bartoli et al. (2018), we tested the significance of the abiotic variables included in the linear combinations by adopting a Jackknife resampling approach by leaving out 10% of the samples 1,000 times. We only considered as significant abiotic variables with a loading value above 0.2 in more than 75% of the resampled matrices.

Genomic characterization and data filtering

Based on a Pool-Seq approach, a representative picture of within-population genetic variation across the genome was previously obtained for the 145 populations (Frachon et al. 2018). Briefly, to perform *A. thaliana* DNA extraction, a portion of a rosette leaf was sampled from 2,574 plants collected in late February – early March 2015 (approximately 16 plants per populations, min = 5 plants, max = 16 plants, mean = 15.32 plants, median = 16 plants). Individual samples were placed in 96-well S-block plates containing beads, immersed in liquid nitrogen and crushed under cold conditions by using Mixer Mill MM 300 Retsch®. DNA was extracted on each individual by following the protocol described in Brachi et al. (2013) and further quantified with Quant-iTTM PicoGreen® dsDNA Assay Kit with a qPCR ABI7900 machine. DNA quantification values were used to establish 145 equimolar pools of the individuals of each population, which were then used to build Illumina libraries as described in Frachon et al. (2018). Libraries were sequenced by using Illumina HiSeq 3000 sequencer with a paired-end read length of 2 ×150 bp at the Get-PlaGe core facility (INRA, Toulouse, France). Raw genomic data for each *A. thaliana* population are available at the NCBI Sequence Read Archive (SRA) under the study accession number SRP103198. Raw reads were mapped against the reference genome Col-0 with glint tool (version 1.0.rc8.779) and mapped reads further filtered with SAMtool (Li et al. 2009). SNPCalling was performed with SAMtool and VarScan mpileup2snp (Koboldt et al. 2012) as described in Frachon et al. (2018).

Following Frachon et al. (2018), the matrix of population allele frequencies was trimmed according to five successive criteria, resulting in a final number of 1,519,748 SNPs: (i) removing SNPs with missing values in more than seven populations, (ii) 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), removing SNPs with a mean relative coverage depth

across the 145 populations above 1.5 to take into account multiple gene copies in the 145 populations that map to a unique gene copy in the reference genome Col-0, (iii) removing SNPs with a standard deviation of allele frequency across the populations below 0.004, (iv) removing SNPs with a mean relative coverage depth across the 145 populations below 0.5 to take into account indels that correspond either genomic regions present in Col-0 but absent in the 145 populations or genomic regions present in the 145 populations but absent in Col-0, and (v) removing SNPs with the alternative allele present in less than 11 populations to take into account bias in GWA/GEA analysis due to rare alleles (Bergelson & Roux 2010).

Genome-Environment Association analysis on plant community descriptors

We performed a GEA analysis between the set of 1,519,748 SNPs and 49 plant community descriptors (species richness, Shannon index, coordinates on three first PCoA axes and abundance of the 44 most prevalent OTUs) based on a Bayesian hierarchical model, which deals with Pool-Seq data and is implemented in the program BayPass (Gautier 2015). This model explicitly accounts for the scaled covariance matrix of population allele frequencies (**Ω**), which makes the analyses robust to complex demographic histories, thereby allowing decreasing drastically the rate of false positives (Gautier 2015). In this study, we run the core model to evaluate the association between allele frequencies along the genome and the 49 plant community descriptors. For each SNP, we estimated a Bayesian Factor (BF_{is} measured in deciban units) and the associated regression coefficient (Beta_is, β_i) using an Importance Sampling algorithm (Gautier 2005). The full posterior distribution of the parameters was obtained based on a Metropolis–Hastings within 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-iterations burn-in period. The 49 plant community descriptors 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 plant community descriptor. The results presented in this study correspond to the average BF_{is} and β_i values across these three repeats.

As previously performed in Frachon et al. (2018), we parallelized the genome-environment analysis by dividing the full data set into 32 sub-data sets, each containing 3.125% of the 1,519,748 SNPs (ca, 51,000 SNPs taken every 32 SNPS along the genome). Pairwise comparisons of the 32 resulting covariance matrices confirmed that all estimates were consistent with highly correlated elements. In agreement, the pairwise FMD distances among the 32 covariance matrices (Förstner and Moonen 2003) had a narrow range of variation (from

2.02 to 2.27). BF_{is} and the associated regression coefficients $\hat{\beta}_t$ between SNP allele frequencies and variation of plant community descriptors were therefore estimated for each SNP by analyzing in parallel the 32 sub-data sets described above, but with the same matrix $\hat{\Omega}_1$ (obtained on the first sub-data set) as an estimate of the scaled covariance matrix of population allele frequencies.

To identify plant community descriptors for which the core model poorly converged, we calculated for each plant community descriptor a non-parametric correlation coefficient (Spearman's *rho*) between the Bayes Factor (BF_{is}) and the posterior mean of the absolute regression coefficient β_i (Beta_is). We discarded four plant community descriptors (OTU4, OTU7, OTU67 and OTU203) with a correlation coefficient below 0.75.

Comparing abiotic and biotic GEA signals

To evaluate the putative cross-effect of the abiotic environment on the genetic basis associated with plant descriptors, we run the core model to evaluate the association between allele frequencies of the 1,519,748 SNPs and the 6 climate and 11 soil variables. A GEA analysis for the 6 climatic variables was previously performed in Frachon et al. (2018) but was based on the whole set of 168 natural populations of *A. thaliana* located in the Midi-Pyrénées region. To be consistent with the GEA analysis performed on the plant community descriptors, we decided to rerun a climate related GEA analysis based on the subset of 145 natural populations used in this study. The core model poorly converged for the concentration in aluminum and water holding capacity (i.e. Spearman correlation's *rho* between BF_{is} and β_i value < 0.75), that were therefore removed from further analyses based on GEA results.

To compare abiotic and biotic GEA signals, we estimated for each plant community descriptor the proportion of top SNPs (i.e. 0.1% of the SNPs with the highest BF_{is}, n = 1,520) that were common to the 0.1% top SNPs identified by GEA analysis performed on climate and soil variables.

Enrichment analyses and identification of candidate genes associated with plant community descriptors

For supporting local adaptation signals identified by GEA analyses and identifying candidate genes associated with plant community descriptors, we only considered top SNPs that were detected for plant community descriptors but not for abiotic variables (i.e. SNPs that were specific to plant community descriptors).

For supporting local adaptation signals identified by GEA analyses, we first performed a genome-wide selection scan by estimating the XtX measure of spatial genetic differentiation among the 145 populations. For a given SNP, the XtX measures the variance of the standardized population allele frequencies, that is, rescaled using **Ω** and across population allele frequencies (Günther & Coop 2013, Gautier 2015). Correcting for the genome-wide effects of confounding demographic evolutionary forces such as genetic drift and gene flow, therefore allows for a robust identification of SNPs that are highly differentiated among populations (Gautier 2015). Following the methodology described in Brachi et al. (2015), we then tested for each plant community descriptor whether the top SNPs (0.1% upper tail of the BF_{is} distribution trimmed for SNPs associated with abiotic variables) were significantly enriched in the 0.1% upper tail of the XtX distribution. Based on Hancock et al. (2011), statistical significance of enrichment was assessed by running 10,000 null circular permutations of the top plant community related SNPs along the genome.

To identify the candidate genes associated with variation in plant community descriptors, we adopted three non-exclusive approaches. Firstly, using the TAIR 10 database (https://www.arabidopsis.org/), we retrieved for each plant community descriptor all the annotated genes located within a 2kb region around each of the 152 top SNPs. Secondly, we tested for each plant community descriptor whether the152 top SNPs were over-represented in each of the 736 Gene Ontology Biological Processes from the GOslim set (Consortium 2008), using a window size of 5kb on each side of each top SNP. Following Hancock et al. (2011) and Wang et al. (2018), statistical significance of the enrichment was assessed by running 10,000 null circular permutations of the 152 top SNPs along the genome. For each plant community descriptor, we only considered BPs that had at least 5 hits. For each significantly enriched BP at a $P < 0.001$, we retrieved the identity of all the genes containing top SNPs. Thirdly, a gene list combining the genes underlying significantly enriched biological processes and genes associated with more than three plant community descriptors was used after removal of duplicates, to identify significantly over-represented functional classes ($P < 0.05$ after Bonferroni correction) with the classification superviewer tool on the university of Toronto website (http://bar.utoronto.ca/ntools/cgibin/ntools_classification_superviewer.cgi) using the MAPMAN classification (Provart & Zhu, 2003).

Acknowledgments

This work was funded by the Région Midi-Pyrénées (CLIMARES project) and the LABEX TULIP (ANR-10-LABX-41, ANR-11-IDEX-0002-02).

Author contributions

L.F., C.B. and F.R. planned and designed the research. L.F. and F.R. conducted fieldwork. L.F. and B.M. performed DNA extraction and generated the sequencing data. L.F., B.M. and C.B. performed the bioinformatics analysis. L.F., B.M. and N.H. identified the species/genus name of the OTUs. L.F. and F.R. performed the statistical analyses and the genome-environment analysis. L.F. and F.R. performed the enrichment analyses and the identification of candidate genes. N.H. guided the statistical analysis. L.F., C.B. and F.R. wrote the manuscript, with contributions from B.M. and N.H.

References

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.

Bartelheimer M, Schmid C, Storf J, Hell K, Bauer S. 2015. Interspecific competition in *Arabidopsis thaliana*: a knowledge gap is starting to close. Prog Bot. 76:303-319.

Barthet MM, Hilu KW. 2007. Expression of matK: functional and evolutionary implications. Am J Bot. 94:1402–1412.

Bartoli C, Roux F. 2017. Genome-Wide Association studies in plant pathosystems: toward an ecological genomics approach. Front Plant Sci. 8:763.

Bartoli C, Frachon L, Barret M, Rigal M, Zanchetta C, Bouchez O, Carrere S, Roux F. 2018. In situ relationships between microbiota and potential pathobiota in *Arabidopsis thaliana*. ISME J. 12:2024–2038.

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.

Bergelson J, Roux F. 2010. Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. Nature Rev. Genet. 11:867-879.

Brachi B, Villoutreix R, Faure N, Hautekeete 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.

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. Nat. Acad. Sci. U.S.A. 112**:**4032–4037.

Carrascal LM, Galvá I, Gordo O. 2009. Partial least squares regression as an alternative to current regression methods used in ecology. Oikos. 118:681-690.

Chen F, Li B, Demone J, Charron JB, Shi X, Deng XW. 2014. Photoereceptor partner FHY1 has an independent role in gene modulation and plant development under far-red light. Proc Natl Acad Sci USA. 111:11888–11893.

Consortium GO. (2008). The gene ontology project in 2008. Nucleic Acids Res. 36:D440–D444. doi: 10.1093/nar/gkm883.

De Mita S, Thuillet AC, 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.

Dubiella U, Seybold H, Durian G, Komander E, Lassig R, Witte CP, Schulze WX, Romeis T. 2013. Calcium-dependent protein kinase/NADPH oxidase activation circuit is required for rapid defense signal propagation. Proc. Nat. Acad. Sci. U.S.A. 110:8744-8749.

Dubuis A, Giovanettina S, Pellissier L, Pottier J, Vittoz P, Guisan A. 2013. Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. J Veg Science. 24:593-606.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 26:2460–2461.

Edgar RC. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods. 10:996-998.

Ellison CT, Vandenbussche F, Van der Straeten D, Harmer SL. 2011. XAP5 CIRCADIAN TIMEKEEPER regulates ethylene responses in aerial tissues of Arabidopsis. Plant Physiol. 155:988-999.

Fan M, Bai MY, Kim JG, Wang T, Oh E, Chen L, Park CH, Son SH, Kim SK, Mudgett MB, Wang ZY. 2014. The bHLH transcription factor HBI1 mediates the trade-off between growth and pathogen-associated molecular pattern–triggered immunity in Arabidopsis. Plant Cell 26:828-841.

Ferrero-Serrano Á, Assmann SM. 2019. Phenotypic and genome-wide association with the local environment of *Arabidopsis*. Nat. Ecol. Evol. 3:274-285.

Förstner W, Moonen B. 2003. A metric for covariance matrices. In: Grafarend EW, Krumm FW, Schwarze VS, eds. Springer-Verlag. Berlin/Heidelberg, Germany, 299-309.

Frachon L, Libourel C, Villoutreix R, Carrère S, Glorieux C, Huard-Chauveau C, Navascués M, Gay L, Vitalis R, Baron E, Amsellem L, Bouchez O, Vidal M, Le Corre V, Roby D, Bergelson J, Roux F. 2017. Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time. Nat Ecol Evol. 1:1551-1561.

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.

Gao M, Wang X, Wang D, Xu F, Ding X, Zhang Z, Bi D, Chang YT, Chen S, Li X, Zhang Y. 2009. Regulation of cell death and innate immunity by two receptor-like kinases in Arabidopsis. Cell Host Microbe. 6:34-44.

Gautier M. 2015. Genome-wide scan for adaptive divergence and association with population-specific covariates. Genetics. 201:555–1579.

Gilman SE, Urban MC, Tewksbury J, Gilchrist GW, Holt RD. 2010. A framework for community interactions under climate change. Trends Ecol Evol. 25:325-331.

Gould WA, Walker MD. 1999. Plant communities and landscape diversity along a Canadian Arctic river. J Veg Science. 10: 537-548.

Günther T, Coop G. 2013. Robust identification of local adaptation from allele frequencies. Genetics. 195:205-220.

Gómez-Gómez L, Boller T. 2000. FLS2: An LRR Receptor–like Kinase Involved in the Perception of the Bacterial Elicitor Flagellin in *Arabidopsis*. Mol. Cell. 5:1003-1011.

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.

Hautekéete N-C, Frachon L, Luczak C, Toussaint B, Van Landuyt W, Van Rossum F, Piqout Y. 2015. Habitat type shapes long-term plant biodiversity budgets in two densely populated regions in north-western Europe. Diversity Distrib. 21:631-642.

Hawkins AP, Crawford KM. 2018. Interactions between plants and soil microbes may alter the relative importance of intraspecific and interspecific plant competition in a changing climate. AoB Plants 10:ply039.

Horton MW, Bodenhausen N, Beilsmith K, Meng D, Muegge BD, Subramanian S, Vetter MM, Vilhjalmsson BJ, Nordborg M, Gordon JI, Bergelson J. 2014. Genome-wide association study of *Arabidopsis thaliana* leaf microbial community. Nat. Comm. 5:5320.

Lasky JR, Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH. 2012. Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geographt and climate. Mol. Ecol. 21: 5512-5529.

Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, Bonnette J, Juenger TE, Hyma K, Acharya C, Mitchell SE, Buckler ES, Brenton Z, Kresovich S, Morris GP. 2015. Genome-environment associations in sorghum landraces predict adaptive traits. Sci. Adv. 1:e1400218.

Lê Cao KA, Rossouw D, Robert-Granié C, Besse P. 2008. A sparse PLS for variable selection when integrating Omics Data. Stat. Appl. Genet. Mol. Biol. 7: Article 35.

Lê Cao KA, Meugnier E, McLachlan G. 2010. Integrative mixture of experts to combine clinical factors and gene markers. *Bioinformatics* 26: 1192–1198.

Le Corre V. 2005. Variation at two flowering time genes within and among populations of *Arabidopsis thaliana*: comparison with markers and traits. Mol. Ecol. 14:4181-4192.

Leu WM, Cao XL, Wilson TJ, Snustad DP, Chua NH. 1995. Phytochrome A and phytochrome B mediate the hypocotyl-specific downregulation of TUB1 by light in arabidopsis. Plant Cell. 7:2187-2196.

Li J, Wen J, Lease KA, Doke JT, Tax FE, Walker JC. 2002. BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. Cell. 110:213-222.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079.

Libourel C, Baron E, Lenglet J, Amsellem L, Roby D, Roux F. 2019. The genomic architecture of competitive response of *Arabidopsis thaliana* is highly flexible between monospecific and plurispecific neighborhoods. bioRxiv. doi: http://dx.doi.org/10.1101/536953.

Litrico I, Violle C. 2015. Diversity in plant breeding: a new conceptual framework. Trends Plant Sci. 20:604-613.

Lorek J, Griebel T, Jones AM, Kuhn H, Panstruga R. 2013. The role of Arabidopsis heterotrimeric G-protein subunits in MLO2 function and MAMP-triggered immunity. Mol. Plant Microbe Interact. 26:991-1003.

Kelly AE, Gloulden ML. 2008. Rapid shifts in plant distribution with recent climate change. Proc. Nat. Acad. Sci. U.S.A. 105:11823-11826.

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22:568-576.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2016. vegan: Community Ecology Package. (https://CRAN.R-project.org/package=vegan).

Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics. 20:289–290.

Pierik R, Mommer L, Voesenek LACJ. 2013. Molecular mechanisms of plant competition: neighbour detection and response strategies. Funct Ecol. 27:841–853.

Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, Ågren J, Bossdorf O, Byers D, Donohue K, Dunnning M, Holub EB, Hudson A, Le Corre V, Loudet O, Roux F, Warthmann N, Weigel D, Rivero L, Scholl R, Nordborg M, Bergelson J, Borevitz JO. 2010. The scale of population structure in *Arabidopsis thaliana.* Plos Genet. 6:e1000843.

Pluess AR, Frank A, Heiri C, Lalagüe H, Vendramin GG, Oddou-Muratorio. 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. Comput. Mol. Biol. 2003:271–272.

Roig-Villanova I, Bou-Torrent J, Galstyan A, Carretero-Paulet L, Portolés S, Rodriguez-Concepcion M, Martinez-Garcia JF. 2007. Interaction of shade avoidance and auxin responses: a role for two novel atypical bHLH proteins. EMBO J. 26:4756–4767.

Roux F, Bergelson J. 2016. Chapter Four – 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.

Sanon A, Andrianjaka ZN, Prin Y, Bally R, Thioulouse J, Comte G, Duponnois R. 2009. Rhizosphere microbiota interfers with plant-plant interactions. Plant Soil 321:259-278.

Scalfi M, Mosca E, Di Pierro EA, Troggio M, Vendramin GG, Sperisen C, La Porta N, Neale DB. 2014. Micro- and macro-geographic scale effect on the molecular imprint of selection and adaptation in Norway Spruce. PLoS ONE 9:e115499.

Singer A, Travis JMJ, Johst K. 2013. Interspecific interactions affect species and community responses to climate shifts. Oikos. 122:358–366.

Subrahmaniam HJ, Libourel C, Journet EP, Morel JB, Muños, Niebel A, Raffaele S, Roux F. 2018. The genetics underlying natural variation of plant-plant interactions, a beloved but forgotten member of the family of biotic interactions. Plant J. 93:747-770.

Swain S, Jiang HW, Hsieh HL. 2017. FAR-RED INSENSITIVE 219/JAR1 contributes to shade avoidance responses of Arabidopsis seedlings by modulating key shade signaling components. Front Plant Sci. 8:1901.

Tan ST, Dai C, Liu HT, Xue HW. 2013. Arabidopsis Casein Kinase 1 proteins CK1.3 and CK1.4 phosphorylate Cryptochrome 2 to regulate blue light signaling. Plant Cell. 25:2618-2632.

Trivellone V, Bougeard S, Giavi S, Krebs P, Balseiro D, Dray S, Moretti M. 2017. Factors shaping community assemblages and species co-occurrence of different trophic levels. 7:4745-4754.

Wang M, Roux F, Bartoli C, Huard-Chauveau C, Meyer C, Lee H, Roby D, McPeek MS, Bergelson J. 2018. Two-way mixed-effects methods for joint association analysis using both host and pathogen genomes. Proc. Nat. Acad. Sci. U.S.A. 115:E5440-E5449.

Whitham TG, Bailey JK, Schweitzer JA, Shuster SM, Bangert RK, LeRoy CJ, Lonsdorf EV, Allan GJ, DiFazio SP, Potts BM, Fischer DG, Gehring CA, Lindroth RL, Marks JC, Hart SC, Wimp GM, Wooley SC. 2006. A framework for community and ecosystem genetics: from genes to ecosystems. Nat Rev Genet. 7:510-523.

Wilson JB, Peet RK, Dengler J, Pärtel M. 2012. Plant species richness: the world records. J Veg Sci. 24:796-802.

Wrzaczek M, Vainonen JP, Stael S, Tsiatsiani L, Help-Rinta-Rahko H, Gauthier A, Kaufholdt D, Bollhöner B, Lamminmäki A, Staes A, Gevaert K, Tuominen H, Van Breusegem F, Helariutta Y, Kangasjärvi J. 2015. GRIM REAPER peptide binds to receptor kinase PRK5 to trigger cell death in Arabidopsis. EMBO J. 34:55-66.

Xing DH, Lai ZB, Zheng ZY, Vinod KM, Fan BF, Chen ZX. 2008. Stress- and pathogen-induced Arabidopsis WRKY48 is a transcriptional activator that represses plant basal defense. Mol. Plant. 1:459-470.

Xu CR, Liu C, Wang YL, Li LC, Chen WQ, Xu ZH, Bai SN. 2005. Histone acetylation affects expression of cellular patterning genes in the Arabidopsis root epidermis. Proc Natl Acad Sci USA. 102:14469–14474.

Züst T, Heichinger C, Grossniklaus U, Harrington R, Kliebenstein DJ, Turnbull LA. 2012. Natural enemies drive geographic variation in plant defenses. Science. 338:116-119.

Figure legends.

Figure 1. Species richness and composition of plant communities associated with *A. thaliana*. (*A*) The Midi-Pyrénées region (south-west of France) is represented in red on a European map. (*B*) Geographic variation of species richness. (*C*) Geographic variation of plant community composition represented by the first PCoA axis. (*D*), (*E*) and (*F*) Photographs illustrating the degree of differentiation in plant community assemblages among the 145 natural populations of *A. thaliana*. *Images in panels (D), (E) and (F): courtesy of L. Frachon*.

Figure 2. Relationships between plant community descriptors and climate and soil factors based on a sparse partial lest-square regression (sPLSR). PVE (%): percentage of variance in plant community descriptors explained by abiotic factors. Only abiotic variables with a loading value above 0.2 in more than 75% of the 1000 Jackknife resampled matrices were considered as significant. The color gradient indicates the strength of the loading values of the abiotic variables (yellow: 0.20 < loadings < 0.50, orange: 0.50 < loadings < 0.65, red: 0.65 < loadings <1). MAT = mean annual temperature (C°), MCMT = mean coldest month temperature (C°), PPT_wt = winter precipitations (mm), PPT_sp = spring precipitations (mm), PPT_sm = summer precipitations (mm), PPT_at = autumn precipitations (mm), N = nitrogen, ratio_CN = carbon-nitrogen ratio, WHC = water holding capacity.

Figure 3. Illustration of GEA signals shared between plant community descriptors and abiotic factors. The upper and lower panels illustrate GEA signals obtained from the abundance of *Sonchus oleraceaus* (OTU 27) and *Papaver rhoeas* (OTU149), respectively. (*A*) and (*D*) Relationships between the abundance of *S. oleracus* and soil pH (*A*) and between the abundance of *Papaver rhoeas* and summer precipitations (*D*). Manhattan plots of the GEA results for the abundance of *S. oleraceus* (*B*) and *P. rhoeas* (*E*). The *x*-axis indicates the position of the 1,519,748 SNPs along the five chromosomes, represented by different shades of blue. The *y*-axis indicates the Bayes factor (BF_{is} expressed in deciban units) estimated by the core model implemented in the program BayPass. Red dots indicate top SNPs shared between OTU abundance and climate variables, whereas green dots indicate top SNPs shared between OTU abundance and soil variables. Arrows indicate two association peaks for which a close-up was performed in (*C*) and (*F*). (*C*) and (*F*) Zoom spanning association peaks containing top SNPs shared between OTU abundance and abiotic factors. In (*C*), the green point is associated with

both the abundance of *S. oleraceus* and pH, whereas in (F) all red points are associated with both the abundance of *P. rhoeas* and summer precipitations.

Figure 4. Manhattan plots of the genome-environment association results for the abundance of four plant OTUs and plant composition (third PCoA axis) after removing top SNPs shared with abiotic variables. The *x*-axis indicates the SNP position along the five chromosomes, represented by different shades of blue. The *y*-axis indicates the Bayes factor (BF_{is} expressed in deciban units) estimated by the core model implemented in the program BayPass. *Images of Sonchus oleraceus, Convolvulus arvensis and Poa nemoralis: courtesy of F. Roux. Images of Helminthotheca echioides and for PCoA3: courtesy of L. Frachon*.

Table 1. Enrichment of each plant community descriptor related top SNPs (0.1% upper tail of the BF_{is} distribution) in the 0.1% tail of the genome-wide spatial differentiation (XtX) distribution. (*A*) Number of top SNPs (out of 1,520) shared between plant community descriptors and abiotic factors. (*B*) Enrichment when considering all the 0.1% top SNPs (n = 1,520). (*C*) Enrichment when considering the 0.1% top SNPs trimmed for the top SNPs shared with abiotic variables. 'Fe' stands for fold value of enrichment. Bold values indicate significant enrichment after a false discovery rate (FDR) correction at the nominal level of 5%. *P*: ^{ns} non-significant, * $0.05 > P > 0.01$, ** $0.01 > P > 0.001$, *** $P < 0.001$

	Plant community descriptors	Species	Family	A. Number of top SNPs shared with abiotic factors	B. All top SNPs		C. All top SNPs trimmed for SNPs shared with abiotic factors	
					Fe	*P*	Fe	*P*
Specific descriptors	OTU1	Conyza canadensis	Asteraceae	3	1.32	ns	1.32	*ns*
	OTU3	Crepis biennis	Asteraceae	14	1.97	ns	2.01	*ns*
	OTU8	Lactuca serriola	Asteraceae	1	**3.29**	**	**3.29**	**
	OTU10	Achillea millefolium	Asteraceae	3	**3.95**	**	**3.96**	**
	OTU15	Hypochaeris radicata	Asteraceae	26	**5.26**	***	**4.77**	***
	OTU16	Helminthotheca echioides	Asteraceae	4	**4.60**	***	**4.63**	***
	OTU18	Lapsana communis	Asteraceae	8	**2.63**	*	**2.66**	*
	OTU20	Senecio vulgaris	Asteraceae	12	1.97	ns	1.34	*ns*
	OTU27	Sonchus oleraceus	Asteraceae	9	1.32	ns	1.33	*ns*
	OTU46	Valerianella locusta	Caprifoliaceae	5	**4.60**	***	**4.63**	***
	OTU49	Fraxinus excelsior	Oleaceae	31	0.00	ns	0.00	*ns*
	OTU65	Plantago lanceolata	Plantaginaceae	4	**3.29**	**	**3.30**	**
	OTU71	Veronica arvensis	Plantaginaceae	13	**5.26**	***	**5.35**	***
	OTU72	Galium mollugo	Rubiaceae	13	0.66	ns	0.67	*ns*
	OTU78	Myosotis arvensis	Boraginaceae	14	1.97	ns	2.01	*ns*
	OTU83	Convolvulus arvensis	Convolvulaceae	14	**9.87**	***	**10.05**	***
	OTU87	Anagallis arvensis	Primulaceae	2	**3.29**	**	**3.30**	**
	OTU88	Polygonum aviculare	Polygonaceae	11	0.66	ns	0.67	*ns*
	OTU100	Sagina apetala	Caryophyllaceae	1	**3.29**	**	**3.29**	**
	OTU109	Arenaria serpyllifolia	Caryophyllaceae	5	0.00	ns	0.00	*ns*
	OTU113	Cerastium glomeratum	Caryophyllaceae	67	**5.26**	***	**5.04**	***
	OTU114	Papaver rhoeas	Papaveraceae	28	1.97	ns	1.36	*ns*
	OTU132	Bromus hordeaceus	Poaceae	8	**5.26**	***	**5.31**	***
	OTU136	Avena sp.	Poaceae	2	**7.89**	***	**7.91**	***
	OTU143	Festuca rubra	Poaceae	10	**6.58**	***	**6.66**	***
	OTU145	Holcus lanatus	Poaceae	22	1.97	ns	2.03	*ns*
	OTU146	Dactylis glomerata	Poaceae	9	**2.63**	*	2.00	*ns*
	OTU147	Catapodium rigidum	Poaceae	9	1.97	ns	2.00	*ns*
	OTU149	Poa nemoralis	Poaceae	4	**3.95**	**	**3.97**	**
	OTU154	Poa annua	Poaceae	14	**5.92**	***	**6.03**	***
	OTU159	Aphanes arvensis	Rosaceae	1	0.00	ns	0.00	*ns*
	OTU179	Epilobium sp.	Onagraceae	16	**6.58**	***	**6.72**	***
	OTU192	Erophila verna	Brassicaceae	0	**3.29**	**	3.29	**
	OTU196	Capsella bursa-pastoris	Brassicaceae	69	1.32	ns	1.44	*ns*
	OTU198	Arabidopsis thaliana	Brassicaceae	2	**9.87**	***	**9.89**	***
	OTU202	Cardamine hirsuta	Brassicaceae	0	**2.63**	*	2.63	*
	OTU204	Geranium sp.	Geraniaceae	4	**2.63**	*	**2.65**	*
	OTU216	Medicago lupulina	Fabaceae	47	1.32	ns	1.40	*ns*
	OTU223	Vicia sativa	Fabaceae	1	1.32	ns	1.32	*ns*
	OTU234	Trifolium campestre	Fabaceae	7	**2.63**	*	**2.66**	*
Global descriptors	Species richness	-	-	242	**3.95**	**	**3.72**	**
	Shannon diversit	-	-	286	**4.60**	***	**4.99**	***
	PCoA 1	-	-	167	**5.26**	***	**4.98**	***
	PCoA 2	-	-	297	1.32	ns	1.02	*ns*
	PCOA 3	-	-	99	**7.89**	**	**7.53**	***

Figures

Figure 1.

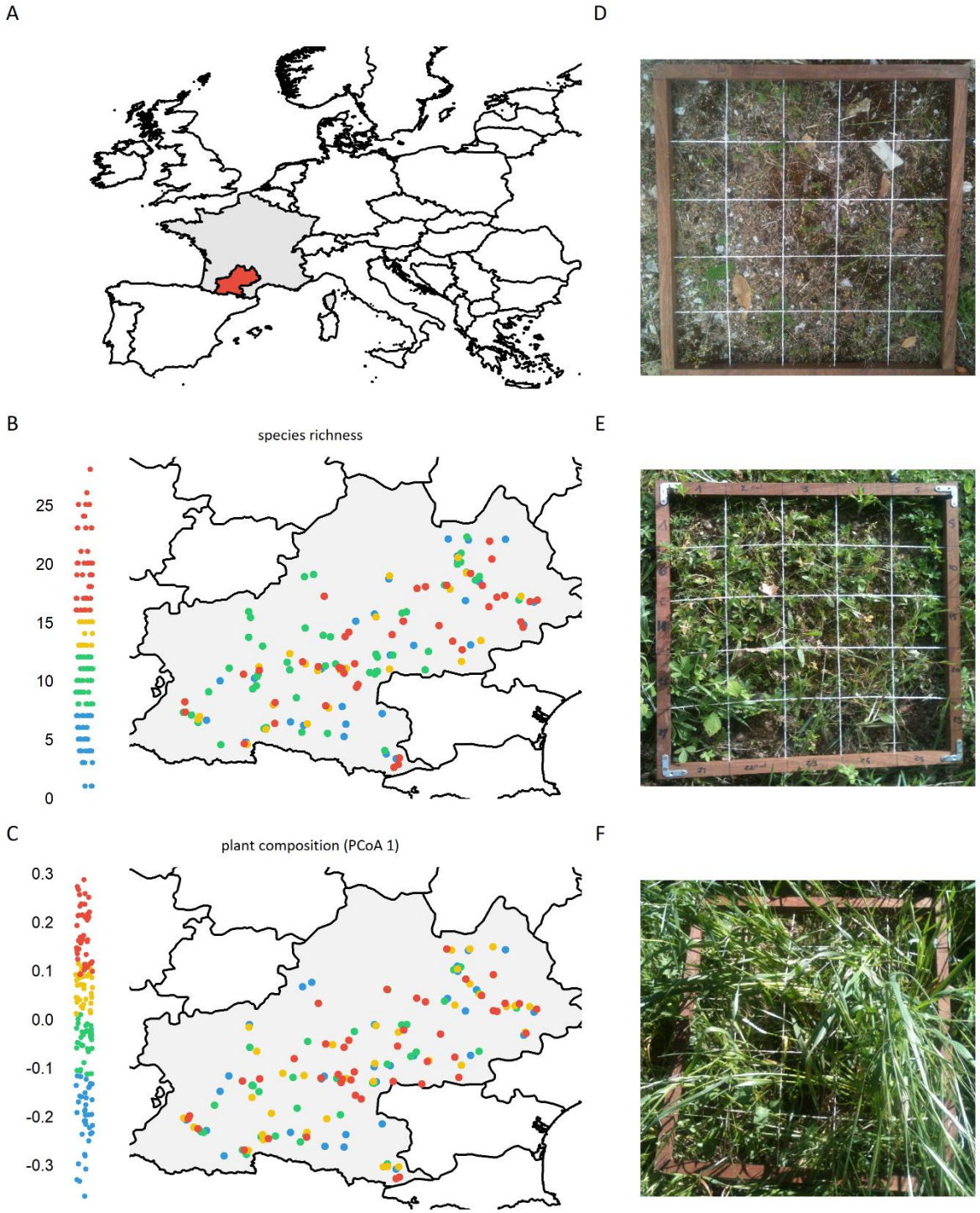
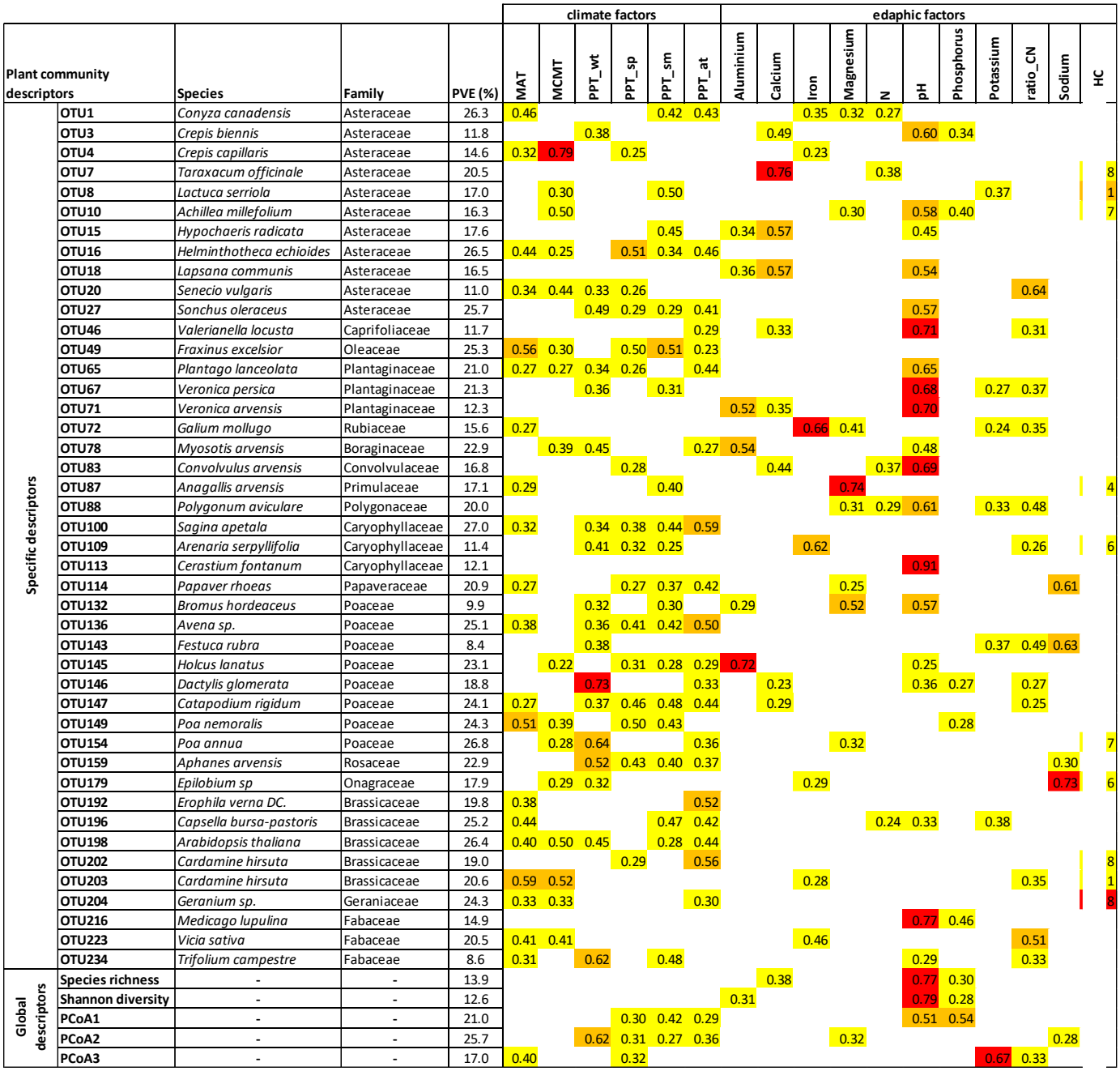


Figure 2.



					climate factors						edaphic factors										
Plant community descriptors		Species	Family	PVE (%)	MAT	MCMT	PPT_wt	PPT_sp	PPT_sm	PPT_at	Aluminium	Calcium	Iron	Magnesium	N	pH	Phosphorus	Potassium	ratio_CN	Sodium	HC
Specific descriptors	OTU1	*Conyza canadensis*	Asteraceae	26.3	0.46				0.42	0.43			0.35	0.32	0.27						
	OTU3	*Crepis biennis*	Asteraceae	11.8			0.38					0.49				0.60	0.34				
	OTU4	*Crepis capillaris*	Asteraceae	14.6	0.32	0.79		0.25					0.23								
	OTU7	*Taraxacum officinale*	Asteraceae	20.5								0.76			0.38						8
	OTU8	*Lactuca serriola*	Asteraceae	17.0		0.30			0.50									0.37			1
	OTU10	*Achillea millefolium*	Asteraceae	16.3		0.50								0.30		0.58	0.40				7
	OTU15	*Hypochaeris radicata*	Asteraceae	17.6					0.45		0.34	0.57				0.45					
	OTU16	*Helminthotheca echioides*	Asteraceae	26.5	0.44	0.25		0.51	0.34	0.46											
	OTU18	*Lapsana communis*	Asteraceae	16.5							0.36	0.57				0.54					
	OTU20	*Senecio vulgaris*	Asteraceae	11.0	0.34	0.44	0.33	0.26											0.64		
	OTU27	*Sonchus oleraceus*	Asteraceae	25.7			0.49	0.29	0.29	0.41						0.57					
	OTU46	*Valerianella locusta*	Caprifoliaceae	11.7						0.29		0.33				0.71			0.31		
	OTU49	*Fraxinus excelsior*	Oleaceae	25.3	0.56	0.30		0.50	0.51	0.23											
	OTU65	*Plantago lanceolata*	Plantaginaceae	21.0	0.27	0.27	0.34	0.26		0.44						0.65					
	OTU67	*Veronica persica*	Plantaginaceae	21.3			0.36		0.31							0.68		0.27	0.37		
	OTU71	*Veronica arvensis*	Plantaginaceae	12.3							0.52	0.35				0.70					
	OTU72	*Galium mollugo*	Rubiaceae	15.6	0.27								0.66	0.41				0.24	0.35		
	OTU78	*Myosotis arvensis*	Boraginaceae	22.9		0.39	0.45			0.27	0.54					0.48					
	OTU83	*Convolvulus arvensis*	Convolvulaceae	16.8				0.28				0.44			0.37	0.69					
	OTU87	*Anagallis arvensis*	Primulaceae	17.1	0.29				0.40					0.74							4
	OTU88	*Polygonum aviculare*	Polygonaceae	20.0										0.31	0.29	0.61		0.33	0.48		
	OTU100	*Sagina apetala*	Caryophyllaceae	27.0	0.32		0.34	0.38	0.44	0.59											
	OTU109	*Arenaria serpyllifolia*	Caryophyllaceae	11.4			0.41	0.32	0.25				0.62						0.26		6
	OTU113	*Cerastium fontanum*	Caryophyllaceae	12.1												0.91					
	OTU114	*Papaver rhoeas*	Papaveraceae	20.9	0.27			0.27	0.37	0.42				0.25						0.61	
	OTU132	*Bromus hordeaceus*	Poaceae	9.9			0.32		0.30		0.29			0.52		0.57					
	OTU136	*Avena sp.*	Poaceae	25.1	0.38		0.36	0.41	0.42	0.50											
	OTU143	*Festuca rubra*	Poaceae	8.4			0.38											0.37	0.49	0.63	
	OTU145	*Holcus lanatus*	Poaceae	23.1		0.22		0.31	0.28	0.29	0.72					0.25					
	OTU146	*Dactylis glomerata*	Poaceae	18.8			0.73			0.33		0.23				0.36	0.27		0.27		
	OTU147	*Catapodium rigidum*	Poaceae	24.1	0.27		0.37	0.46	0.48	0.44		0.29							0.25		
	OTU149	*Poa nemoralis*	Poaceae	24.3	0.51	0.39		0.50	0.43								0.28				
	OTU154	*Poa annua*	Poaceae	26.8		0.28	0.64			0.36				0.32							7
	OTU159	*Aphanes arvensis*	Rosaceae	22.9			0.52	0.43	0.40	0.37										0.30	
	OTU179	*Epilobium sp*	Onagraceae	17.9		0.29	0.32						0.29							0.73	6
	OTU192	*Erophila verna DC.*	Brassicaceae	19.8	0.38					0.52											
	OTU196	*Capsella bursa-pastoris*	Brassicaceae	25.2	0.44				0.47	0.42					0.24	0.33		0.38			
	OTU198	*Arabidopsis thaliana*	Brassicaceae	26.4	0.40	0.50	0.45		0.28	0.44											
	OTU202	*Cardamine hirsuta*	Brassicaceae	19.0				0.29		0.56											8
	OTU203	*Cardamine hirsuta*	Brassicaceae	20.6	0.59	0.52							0.28						0.35		1
	OTU204	*Geranium sp.*	Geraniaceae	24.3	0.33	0.33				0.30											8
	OTU216	*Medicago lupulina*	Fabaceae	14.9												0.77	0.46				
	OTU223	*Vicia sativa*	Fabaceae	20.5	0.41	0.41							0.46						0.51		
	OTU234	*Trifolium campestre*	Fabaceae	8.6	0.31		0.62		0.48							0.29			0.33		
Global descriptors	Species richness	-	-	13.9								0.38				0.77	0.30				
	Shannon diversity	-	-	12.6							0.31					0.79	0.28				
	PCoA1	-	-	21.0				0.30	0.42	0.29						0.51	0.54				
	PCoA2	-	-	25.7			0.62	0.31	0.27	0.36				0.32						0.28	
	PCoA3	-	-	17.0	0.40			0.32										0.67	0.33		

Figure 3.

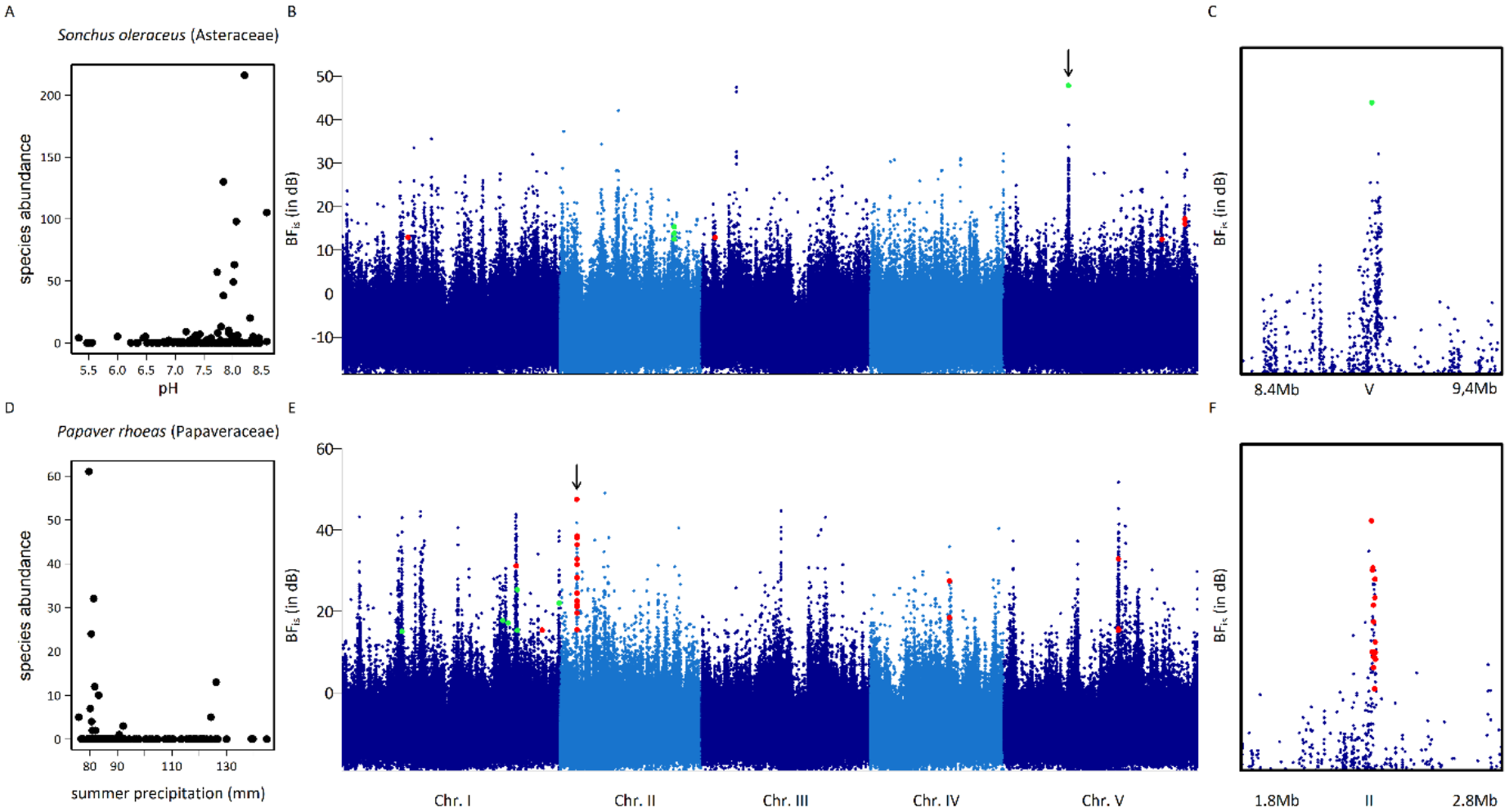
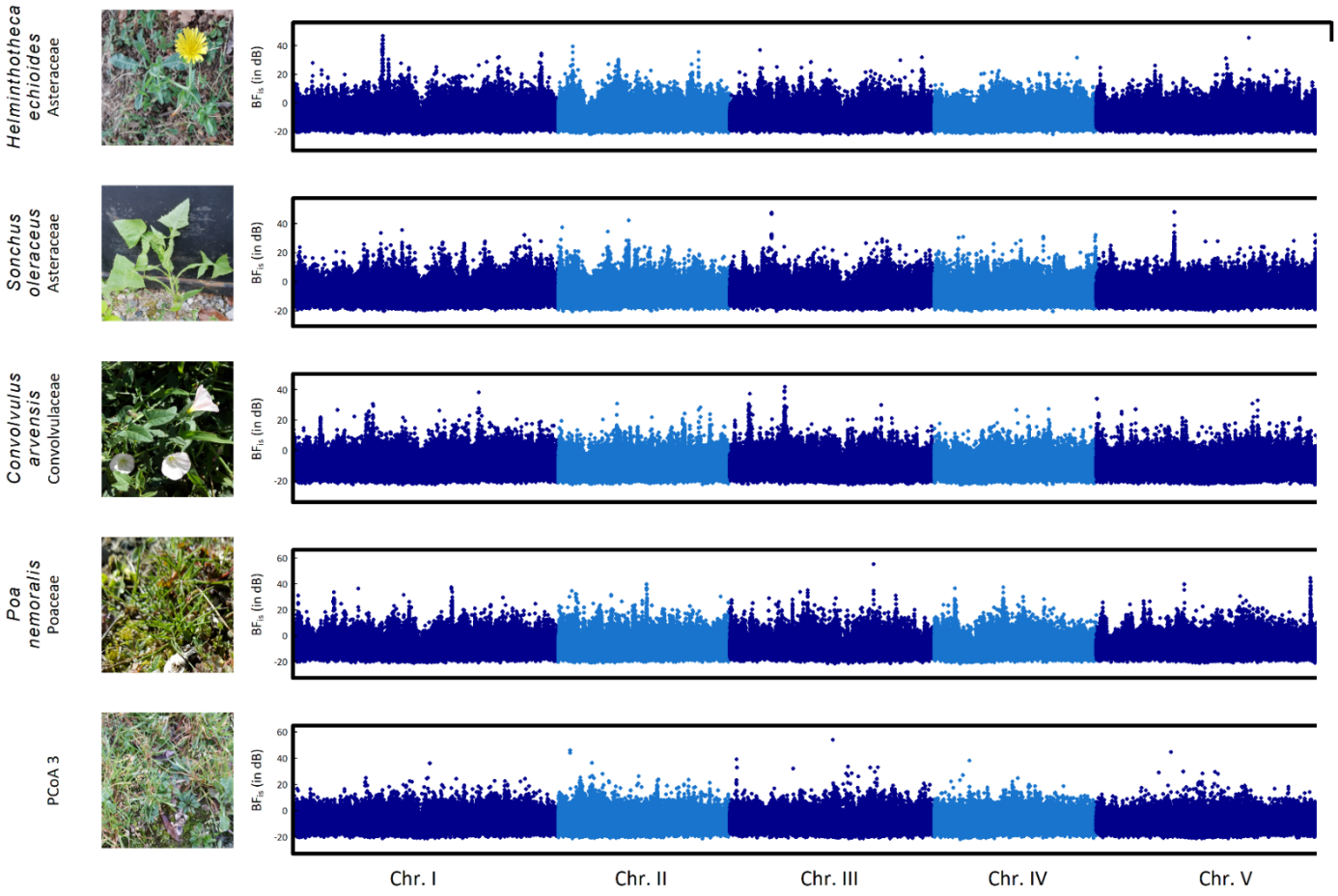

A
Sonchus oleraceus (Asteraceae)
species abundance
pH
B
BFis (in dB)
C
BFix (in dB)
8.4Mb
V
9,4Mb
D
Papaver rhoeas (Papaveraceae)
species abundance
summer precipitation (mm)
E
BFis (in dB)
Chr. I
Chr. II
Chr. III
Chr. IV
Chr. V
F
BFix (in dB)
1.8Mb
II
2.8Mb

Figure 4.


Helminthotheca echioides
Asteraceae
Sonchus oleraceus
Asteraceae
Convolvulus arvensis
Convolvulaceae
Poa nemoralis
Poaceae
PCoA 3
BF10 (in dB)
Chr. I
Chr. II
Chr. III
Chr. IV
Chr. V