Ecological genomics of adaptation of arabidopsis thaliana in a spatially heterogeneous environment

Léa Frachon

Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :
Léa Frachon

Le mercredi 19 juillet 2017

Titre :

Génomique écologique de l'adaptation d'Arabidopsis thaliana dans un environnement hétérogène

---

ED SEVAB : Interactions plantes-microorganismes

Unité de recherche :

Laboratoire des Interactions Plantes Micor-organismes

Directeur(s) de Thèse :

Fabrice Roux

Rapporteurs :

Irène Till-Bottraud (DR1 CNRS) - Joëlle Ronfort (DR2 INRA) - Juliette de Meaux (Professeur d'université)

Autre(s) membre(s) du jury :

Christophe Thébaud (Professeur d'université, Président de Jury)
Maxime Bonhomme (Maître de conférence, Examinateur)

Remerciements

Parce que je n'étais pas faite pour les études d'après ma prof de Terminale… Je tiens à remercier toutes les personnes qui ont cru en moi et qui m'ont soutenu lors de ces trois années de thèse.

Je tiens d'abord à remercier mon jury de thèse Christophe Thébaud, Irène Till-Bottraud, Joëlle Ronfort, Juliette De Meaux et Maxime Bonhomme, de prendre le temps d'évaluer mon travail. Je tiens également à remercier tous les membres de mes deux comités de thèse qui ont été plein de bons conseils ; Nina Hautekèete, Dominique Roby, Valérie Le Corre et Mathieu Gautier. Enfin je remercie les financeurs de ma thèse : la région Midi-Pyrénées et le LabEx TULIP.

Je continue ces remerciements par toi Fabrice, sans qui cette incroyable aventure aurait été complètement différente ! J'ai toujours une phrase en tête, entendu lors d'une pause au GEPV : « les étudiants en thèse avec Fabrice ont énormément de chance ». A l'époque, j'étais à mille lieux de penser qu'un jour c'est moi qui aurai ce privilège. En effet, tu m'as fait confiance dès le début alors que je n'y connaissais pas grand-chose en génomique ! Merci de m'avoir soutenue tout au long de ces trois ans dans mes projets scientifiques (ma thèse, l'organisation du colloque SPE, la création de l'AJS, la vulgarisation scientifique) mais également dans mes projets personnels (notamment avec la rénovation de la « pénichette »). Tu as toujours été très présent, notamment sur le terrain et en cette fin de thèse. Je ne pense pas qu'il y ait beaucoup de directeurs de thèse qui prennent autant de temps pour leurs étudiants. Pour tout cela je te remercie infiniment. Et « oui » tes doctorants sont privilégiés !

Affronter cette thèse a était rendu possible grâce à toute une équipe que je remercie grandement, par ordre de rencontre:

- Etienne ; merci d'avoir égaillé mes premiers mois de thèse avec ta bonne humeur et ton humour, tes « conneries » m'ont quand même bien manquées après !
- Claudia ; Ah Claudia !!! Qu'aurai été cette thèse sans toi ! Déjà scientifiquement, tu m'as beaucoup apporté et tu as toujours été là quand j'en avais besoin. Et puis viens notre amitié qui a commencé lors de nos pauses sportives assez intensives. Merci pour ta présence, tes nombreux conseils et tous ces bons moments passés ensemble.
- Cyril ; merci pour ta gentillesse et ta bonne humeur au quotidien. Tu as toujours le petit mot qui fait plaisir et valorise. Et puis parce que la vie à part le « poney-piscine », c'est aussi « tapply », merci !
- Baptiste ; merci pour ta patience pour m'expliquer la bio mol… Et merci pour ton aide sur le terrain.
- Jaishree ; Thanks for your happiness those last few months ! Have a good luck for your thesis!
- Je tiens aussi à remercier tous les stagiaires qui sont passés dans l'équipe ; Taeken, Arnaud, Mylène, Eve, Kevin et Thomas.

Parce que si j'en suis ici, c'est un peu votre faute… un immense merci à vous Nina et Yves. Mille mercis de m'avoir mis sur les voies de la recherche, sans vous je n'aurais surement jamais osé pousser les portes d'un labo ! Un remerciement tout particulier à Nina que je considère un peu comme ma mère scientifique. Tu as été là dans des moments difficiles et tu m'as toujours

«reboosté». Que ce soit avec les huiles essentielles à ma soutenance de M2, pour ma recherche de thèse et maintenant de postdoc ou dans ma vie perso. Merci !

Je tiens également à remercier toutes les personnes avec qui j'ai pu collaborer durant cette thèse ;

- Merci Mathieu Gautier de m'avoir formé à BayPass et surtout d'avoir toujours été ultra réactif à toutes mes questions.
- Merci beaucoup Dominique Roby pour ton aide sur toute la partie gènes candidats.
- I greatly thank Annie Schmitt for your hospitality in Davis laboratory during two months, your kindness, the time you took for me and the very constructive scientific exchanges. Thank you also to Miki and the Schmitt team. Thank you too Sharon Strauss, Jennifer Gremer and Daniel Runcie for your helpful scientific advices.
- Merci également à Jérôme Gouzy et Sébastien Carrère pour la partie traitement bioinfo, et merci à Ludo pour ta patience !

Je remercie les directeurs successifs du laboratoire LIPM Dominique Roby puis Claude Bruand. Je tiens également à remercier de manière plus générale tout le personnel du laboratoire, notamment ;

- Fabrice, Claudette, Jean Luc, Marine pour toute leur aide sur les différentes manips.
- Christophe et Philippe pour la fabrication du quadrat pour le terrain.
- L'ensemble du service gestion, notamment Christophe que j'ai très souvent sollicité pour mes ordres de missions en pagaille …
- Soon et Isabelle qui ont toujours été là pour mes petits problèmes informatiques !
- Dominique, Christian et Florent pour votre gentillesse.
- Ariane, Aude, Thomas, Vincent, Pierre D et Lucas merci de votre enthousiasme pour l'organisation des journées SPE ! Ce n'était pas toujours facile, mais on a géré !!
- Merci également à Alissounette, Camille, Popo, Pierre B, Gaelle, Corine, Eliane, Mireille, Marielle, Carine, Medhi, Sylvie, Céline, Richard, Marta, Patrice, Fabienne, Sandra, Alice…
- Merci à tous les membres fondateurs de l'AJS, cette aventure a été formidable. Merci Harold (pour ta motivation à toutes épreuves !), Aude, Manon, Hélène, Mathilde, Rémi, Franck, Ariane….
- Florence et Shérifa, les femmes de ménages qui, tous les jours, passent avec un sourire et toujours une petite attention.

Finir cette thèse a également été possible grâce à de nombreuses personnes extérieures à la science ;

Merci donc à tous mes amis pour vos soutiens pendant ces trois années.

- Les « plus anciens » ; Poupoule, Delphine, Yohan, Réjou, Clémentine & Seb, Laure & PY.
- Les « Lillois » : Zaza, Julie, Florence, Arnaud, Nico, Justine, Sophie, Cynthia & Pierre, Benjamin, Franciane.
- Les « dérivés du labo » : Alissounette (merci pour ta folie ! Du premier aux derniers jours de ma thèse tu auras été d'un grand soutien), Camille, Popo (toujours pétillante et pleine d'attentions), Carlos, Pipou…

Remerciements

- Tous les grimpeurs, merci de m'avoir permis de m'évader dans la bonne humeur ; Ivanna (tout simplement mille mercis d'avoir été présente pour moi, tant de chose partagées ces deux dernières années), Anne, Clarisse, Quentin (Merci pour ces ptites pauses en fin de thèse indispensable à ma survie !), Jérôme, Sylvain T, Loïc, Adrien, Marie, Maud, Sylvain F (merci pour les tableaux renversés), David R, David P, Thomas, Yohan, Alice, Maria, Lisa et tous les autres. Merci à vous tous, vous êtes supers !!!
- Les « runneurs » du midi, indispensable pour se vider la tête : Claudia, Romain, Popo, Adrien, Marina, Yolaine, Mathew…
- Les musiciens ; Monsieur Massot, Catherine, Harold, Julien (Merci de m'avoir soutenu la dernière année ; toujours les bons mots pour me redonner confiance dans des moments clés de ma thèse !)
- Les Américains ; Jared, Monica, Hayley & Masha, Creg & Molly, Corey, Logan, Marla, Dana & Kent…. Thank you for all, you're amazing !
- Enfin un immense merci à tous mes voisins portuaires qui ont été d'un formidable soutien les dernières semaines, vos sourires et vos attentions m'ont vraiment aidé à tenir le coup jusqu'au bout… Merci : Mélo, Alvaro, (Pascal et Valérie)x2, Abde, Greg, Quentin, Manu, Baptiste, Victor, Xavier, Mumu, Patrice, Michel, Anique, Christophe, Thibaut, Phiphi, Guy, Rosalia, Jean-François, Sisi, Carine, Agnès, Céline, Luc, Eric, René… Un merci ++ à Céline pour les massages shiatsu qui m'ont bien détendu ;) et enfin un grand merci à Noé qui m'a soutenu ces derniers mois, tu as été d'une gentillesse et d'une patience extrême avec moi !

Je finirai ces remerciements par ma famille. Un immense merci à « mes parents Toulousains » Thibault et Solenn ! Sans vous et vos princesses ma vie ici aurait été tout autre ! Merci pour tout ce que vous avez fais pour moi pendant ces trois années, c'est juste inestimable ! Merci à Cléanne et Elia, j'en suis complètement gaga, elles sont formidables !

Enfin, je remercie ma famille de leurs soutiens à toutes épreuves depuis de longues années ! Merci au G4 + bop' power indispensable à ma survie ! Merci à Noé(mie!) et Margot d'être toujours là pour moi, c'est tellement important de savoir que je peux compter sur mes sœurs n'importe quand ! Merci Sylvain d'être là pour toutes les trois et de m'avoir un peu poussé à aller à la fac après la prépa ! Merci papa pour avoir accepté ce défi un peu fou de retaper le Lenoma qui m'a permis de vivre dans un cadre exceptionnel. Merci à toi et Lucile pour tous ces petits moments d'évasion à la Source. Enfin, un grand merci à toi maman d'avoir toujours cru en moi, de toujours nous avoir poussé vers le haut et d'avoir voulu pour moi et mes sœurs ce qu'il y avait de meilleur.

« *C'est un beau fruit, mais il n'est pas mûr, et nous serons morts avant que le soleil de la pratique et de l'expérience ne l'ait mûri* »

Etienne de Montgolfier

Table des matières

Liste des articles scientifiques et des communications orales

Publications scientifiques

NB: Les articles marqués d'une étoile sont directement liés à ma thèse

* 1. **Frachon L.**[*], Libourel C.[*], Villoutreix R., Carrère S., Glorieux C., Huard-Chauveau C., Navascues M., Gay L., Vitalis R., Baron E., Amsellem L., Bouchez D., Vidal M., Le Corre V., Roby D., Bergelson J. & Roux F. Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time (Re-soumis à *Nature Ecology and Evolution*). [*]Authors contributed equally to this work.

* 2. Bartoli C.[*], **Frachon L.**[*], Barret M., Rigal M., Zanchetta C., Bouchez O., Carrère S. & Roux F. *In situ* relationships between microbiota and potential pathobiota in *Arabidopsis thaliana* (Invitation à resoumettre dans *eLife*). [*]Authors contributed equally to this work.

* 3. **Frachon L.**[*], Bartoli C.[*], Carrère S., Bouchez O., Chaubet A., Gautier M., Roby D., Roux F. A genomic map of adaptation to local climate in *Arabidopsis thaliana* (*En révision dans New Phytologist*). [*]Authors contributed equally to this work.

* 4. **Frachon L.**[*], Mayjonade B.[*], Bartoli C.[*], Hautekeete N.C. & Roux F. Adaptation to plant communities across the genome of *Arabidopsis thaliana.* (La soumission sera effectuée une fois que le manuscrit n°3 sera accepté pour publication). *Authors contributed equally to this work.

5. Hautekèete N.C., **Frachon L.**, Luczak C., Toussaint B., Van Landuyt W., Van Rossum F. & Piquot Y. (2015) Habitat type shapes long-term plant biodiversity budgets in two densely populated regions in north-western Europe. *Diversity and Distributions* **21**: 631-642.

6. Tayeh A., Hufbauer R.A., Estoup A., Ravigne V., **Frachon L.** & Facon B. (2015) Biological invasion and biological control select for different life histories. *Nature Communications* **6**:7268.

Communications scientifiques

- *Organisation*

Membre du comité d'organisation des « 8èmes journées des doctorants SPE » à Toulouse (30 juin – 2 juillet 2016)

- *Communications orales*

1. **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. (October 2016) Tracking the genetic bases of contemporary evolution in a spatially heterogeneous environment. International conference sfécologie, Marseille (France). **Talk**

2. **Frachon L.** (August 2016) The adaptive genetics of *Arabidopsis thaliana* in heterogeneous environments. Lab meeting in the group of Johanna Schmitt, Davis University (California, USA). **Séminaire**

3. **Frachon L.** (June 2016) The adaptive genetics of *Arabidopsis thaliana* in heterogeneous environments. LIPM meeting Toulouse (France). **Séminaire**

4. **Frachon L.** (March 2015) Plant-plant interactions: identification of genetics bases and characterization of their associated signatures of selection in *A. thaliana*. LIPM meeting Toulouse (France). **Séminaire**

- *Posters*

1. **Frachon L.**, C. Bartoli, B. Mayjonade, T. Wijmer & F. Roux (December 2016) The selective agents acting on *Arabidopsis thaliana* depend on the type of habitat. British Ecological Society, Liverpool (UK).

2. **Frachon L.**, R. Villoutreix, C. Libourel, E. Baron, S. Carrère, C. Glorieux, L. Amsellem, V. Le Corre, J. Gouzy, J. Bergelson & F. Roux (May 2016) Adaptive genomics to fine-grained spatial heterogeneity in a natural population of *Arabidopsis thaliana*. Journée des doctorants de l'école doctorale SEVAB, Toulouse (France).

3. **Frachon L.**, R. Villoutreix, C. Libourel, E. Baron, S. Carrère, C. Glorieux, L. Amsellem, V. Le Corre, J. Gouzy, J. Bergelson & F. Roux (July 2015) Adaptive genomics to fine-grained spatial heterogeneity in a natural population of *Arabidopsis thaliana*. 7ème journée des doctorants du département INRA- SPE, Rennes (France).

Introduction générale

A. Adaptation aux changements globaux

Importance de l'effet des changements globaux sur la biodiversité

Année 2016 : les géologues ont validé le changement d'ère géologique dû à l'impact de l'Homme sur la planète (Waters *et al.* 2016). L'Anthropocène place ainsi l'Homme comme la source majeure des changements sur l'écosystème terrestre.

Ce n'est plus à prouver : l'Homme a un réel impact sur la planète. Dès le Pléistocène, les activités humaines auraient conduit à l'extinction de la mégafaune *via* une chasse excessive (Martin & Steadman 1999, Wolverton 2010, Johnson 2002). Les pratiques agricoles ont modifié la diversité des communautés biotiques dès le Néolithique (Lopez-Garcia *et al.* 2013). Le développement des moyens de transport ont par la suite augmenté les échanges entre les continents, entraînant une modification importante de la répartition géographique des principales espèces cultivées et de nombreuses espèces sauvages (Beinart & Middleton 2004), certaines d'entre elles étant décrites comme envahissantes dès le 19ème siècle (Richardson & Pysek 2007) comme l'Ajonc d'Europe introduit dès 1825 sur l'île de la Réunion (Udo *et al.* 2017) ou encore le lapin introduit en Australie en 1859 (Ratcliffe 1959).

Depuis quelques décennies, les activités humaines se sont multipliées et intensifiées à un rythme sans précédent (**Figure 1**), modifiant ainsi en profondeur (i) la diversité, la structure et la dynamique des communautés biotiques, et par conséquent (ii) le fonctionnement des écosystèmes, notamment *via* une altération des services écosystémiques (Chapin III *et al.* 2000, Sala *et al.* 2000, Millennium Ecocystem Assessment 2005). Avec un taux d'extinction du nombre d'espèces depuis 1900 de 100 à 1 000 fois plus important que les 1 à 10 millions d'années précédentes (Pimm *et al.* 1995, Pimm *et al.* 2014), le déclin observé de la biodiversité est sans précédent, affectant plus de trois quarts de la surface des biomes terrestres (Ellis *et al.* 2012).

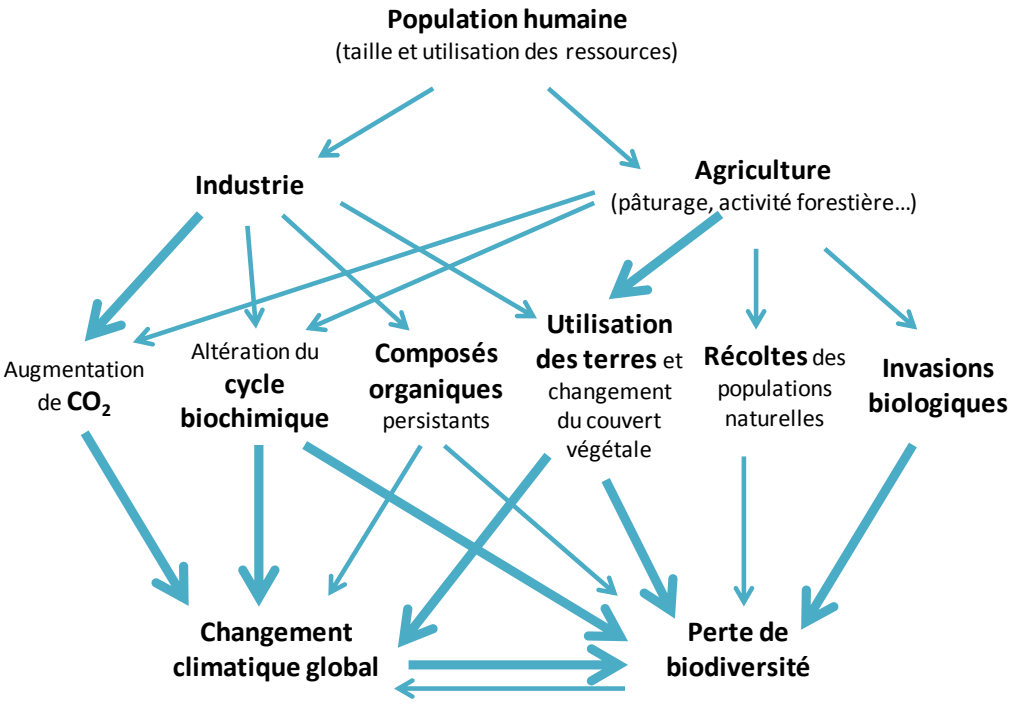


Figure 1: Composantes des changements globaux menant à un déclin de la biodiversité. L'utilisation des terres comprend la perte et la fragmentation des habitats, l'intensification des pratiques agricoles, l'étalement urbain et l'érosion des sols. Figure modifiée d'après Vitousek *et al.* (1997).

Parmi les modifications environnementales majeures liées aux activités humaines récentes, nous pouvons citer le changement climatique observé à une échelle mondiale. Le climat étant un des principaux facteurs abiotiques déterminant les aires de distribution géographique des espèces, une augmentation de température de seulement 1°C entraîne à la fois un déplacement des niches climatiques et une remontée importante des espèces vers de plus hautes latitudes (Thuiller 2007, Kelly & Goulden 2008, Felde *et al.* 2012). A une échelle régionale ou à une échelle des paysages, une forte progression du tourisme et des échanges commerciaux a entraîné une augmentation du nombre d'introductions d'espèces exotiques dans les pays (Beinart & Middleton 2004, Carruthers *et al.* 2011), perturbant potentiellement les interactions au sein des communautés biotiques (Elton 1958, Gilman *et al.* 2010). Ainsi, depuis le début du siècle, de nombreuses épidémies sur les plantes cultivées résultent de sauts géographiques d'espèces pathogènes, comme le chancre bactérien du kiwi causé par la bactérie *Pseudmonas syringae* pv. *actinidiae* qui s'est propagée à travers le monde *via* des plantules importées de kiwi (Bartoli & Roux 2017). Aux mêmes échelles géographiques, des modifications de l'environnement comme la perte des habitats et leur fragmentation (Aguilar *et al.* 2006), l'intensification des pratiques agricoles ou l'urbanisation,

vont entraîner l'apparition de nouvelles barrières géographiques auxquelles les espèces n'ont jamais été confrontées. Par exemple, de par la construction de routes et de bâtiments qui divisent et isolent les habitats naturels, les zones urbaines deviennent de véritables barrières pour les espèces animales et végétales (Van Rossum & Triest 2010). Par ailleurs, selon les régions géographiques du globe, le dépôt d'azote atmosphérique, l'augmentation de CO_2, la production et la prolifération de composés organiques persistants (e.g. fluorochrome) ou la surexploitation des stocks de ressources biologiques ont été signalés comme des changements globaux ayant des effets non négligeables sur la biodiversité (Vitousek *et al.* 1997, Sala *et al.* 2000).

L'hétérogénéité des changements globaux en cours et les différences d'échelles spatiales auxquelles ils sont observés vont se superposer aux filtres environnementaux existants (**Figure 2**), augmentant ainsi la multiplicité des pressions de sélection et leurs interactions auxquelles devront faire face les espèces.

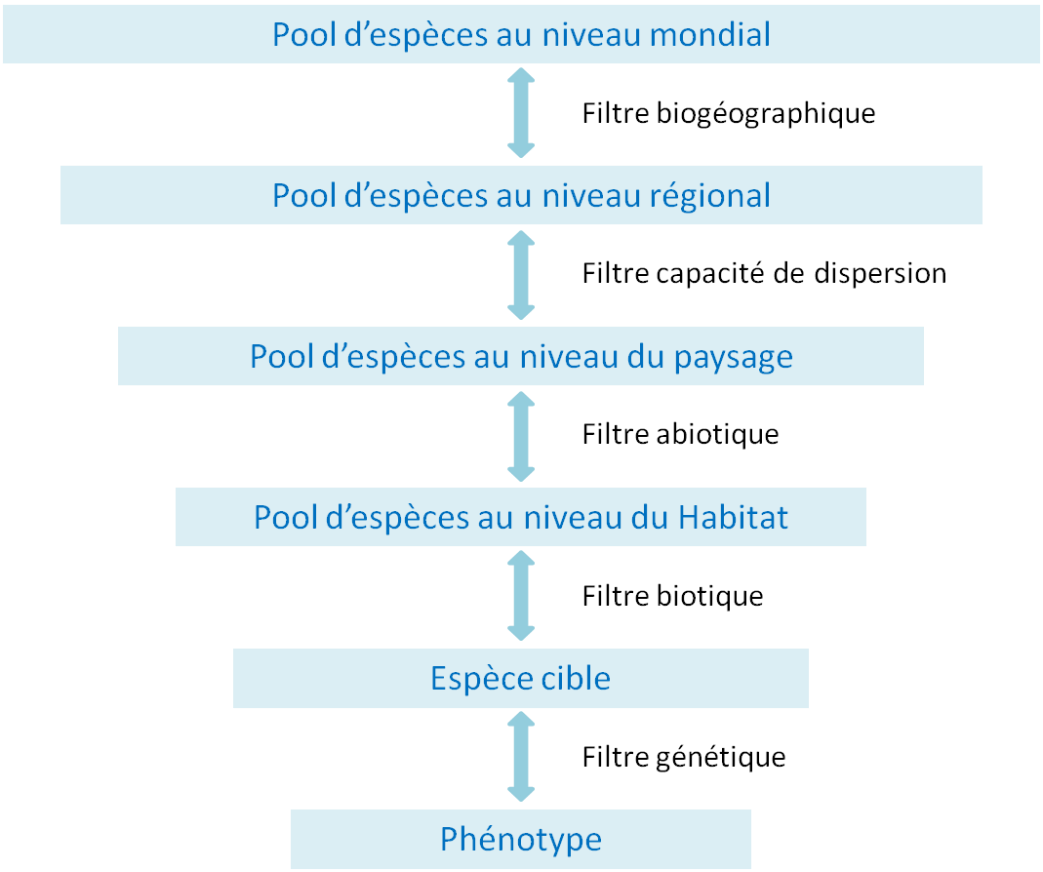


Figure 2: Différents filtres environnementaux agissant à différentes échelles spatiales. Ce sont la combinaison de différents agents sélectifs qui créeront le maillage de ces filtres environnementaux. Figure modifiée d'après Gugerli *et al.* (2013).

En présence de ce nouveau réseau de pressions de sélection, trois types de réponse non-exclusifs peuvent être adoptés par les espèces (Hansen *et al.* 2012). Sur le court terme, les individus peuvent s'acclimater aux changements de conditions environnementales *via* la plasticité phénotypique, en exprimant des phénotypes particuliers en réponse aux conditions environnementales locales. Sur le long terme, les organismes peuvent migrer vers des sites plus favorables, potentiellement sur de longues distances. Le troisième type de réponse correspond à la sélection génétique amenant à l'adaptation locale. Ci-dessous, je décris plus précisément ces trois types de réponse.

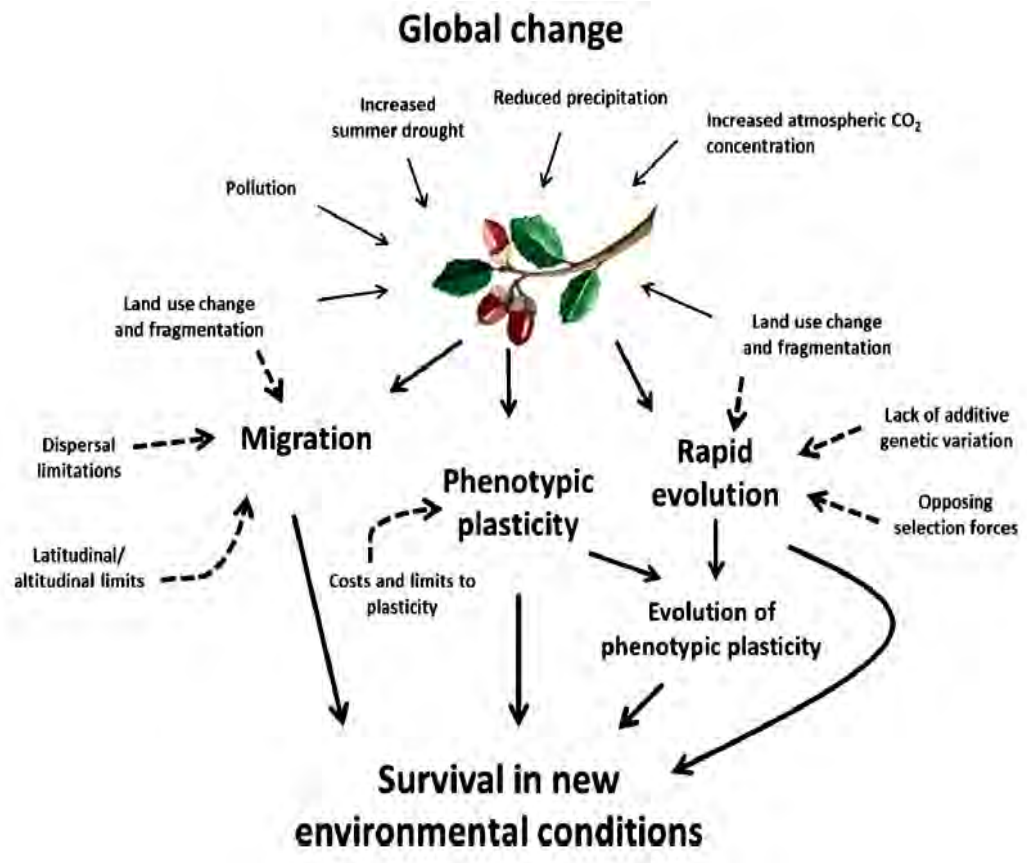


Figure 3: Réponses potentielles d'une espèce face à différents agents sélectifs, lui permettant de survivre à ce nouvel environnement. D'après Matesanz & Valladares (2014).

Se rapprocher d'un nouvel optimum phénotypique via *la plasticité phénotypique*

La plasticité phénotypique correspond à la capacité d'un génotype à produire plusieurs phénotypes en fonction de l'environnement biotique ou abiotique auquel il est

exposé (Sultan 2000, Agrawal 2001). Ainsi, la plasticité phénotypique permettrait à court terme une réponse rapide d'une espèce en modifiant son phénotype sans modification génétique (Matesanz & Valladares 2014). L'hétérogénéité environnementale favorise la plasticité phénotypique (Moran 1992, Sultan & Spencer 2002). Ceci est d'autant plus vrai dans le cas d'une hétérogénéité temporelle où tous les individus font face à une modification de l'environnement. A l'inverse, dans le cas d'une hétérogénéité spatiale, des refuges peuvent toujours persister permettant ainsi aux génotypes fixés adaptés localement de se maintenir.

Les modèles théoriques prédisent que la plasticité phénotypique adaptative peut aider les populations naturelles à se rapprocher d'un nouvel optimum phénotypique (Lande 2009, Chevin *et al.* 2010). Face aux changements globaux en cours, il est donc attendu que la plasticité phénotypique soit une réponse adaptative répandue entre les espèces. Cependant, malgré ses bénéfices théoriques, la plasticité phénotypique adaptative n'est pas aussi fréquente qu'on pourrait l'espérer (Charmantier *et al.* 2008). Cette contradiction entre théorie et observations peut résulter de coûts et de limites qui entravent l'évolution de la plasticité phénotypique adaptative. Parmi les nombreux coûts et limites répertoriés (DeWitt *et al.* 1998), nous pouvons citer dans le cadre du changement climatique le manque de fiabilité des signaux environnementaux, amenant les individus à des réponses plastiques non-adaptatives ou mal-adaptatives (van Kleunen & Fischer 2005, Chevin *et al.* 2010, Price *et al.* 2013, Murren *et al.* 2015). De tels signaux peuvent correspondre à des évènements climatiques extrêmes en dehors de la gamme historique des conditions climatiques rencontrées par les populations. Par ailleurs, la plasticité phénotypique est théoriquement favorisée dans des environnements où les variations environnementales sont régulières et prédictibles (Moran 1992, Sultan & Spencer 2002). Or, le changement climatique est non seulement associé à une augmentation moyenne des températures mais aussi à une augmentation du niveau de fluctuation des conditions climatiques entre les années.

Migrer pour fuir un milieu devenu défavorable

Face aux changements globaux, un deuxième type de réponse correspond à la migration des espèces depuis leurs milieux d'origine devenus défavorables vers de nouveaux milieux correspondant à leur niche écologique (Hansen *et al.* 2012). Ce type de réponse est

notamment attendu dans le cadre du changement climatique où un déplacement des enveloppes climatiques vers de plus hautes latitudes est d'ores et déjà observé.

Durant ces dernières décennies, il a été observé une migration des espèces avec un taux moyen de déplacement de 17.6 km/décennie et une remontée moyenne des espèces en altitude de 11 m/décennie (méta-analyse réalisée par Chen *et al.* 2011). Cependant, la dynamique de migration est très variable entre les espèces (Chen *et al.* 2011), entraînant par conséquent une modification de la diversité et de la composition des communautés et des interactions interspécifiques inhérentes (Gilman *et al.* 2010, Singer *et al.* 2013). De manière intéressante, il peut même être observé une augmentation provisoire de la biodiversité à l'échelle du paysage, avec l'arrivée rapide de nouvelles espèces qui est concomitante à une extinction plus lente d'autres espèces déjà présentes dans les communautés (Jackson & Sax 2010, Hautekèete *et al.* 2014). Ces étapes transitoires suggèrent l'importance d'étudier la dynamique de la biodiversité à différentes échelles géographiques : déclin de la biodiversité à une échelle mondiale *vs* augmentation provisoire de la biodiversité à l'échelle du paysage.

Pour suivre le déplacement géographique des enveloppes climatiques, les espèces végétales dispersant leur pollen et leurs graines sur de longues distances (i.e. espèces anémochores, espèces zoochores…) seront favorisées par rapport aux espèces végétales ayant des distances de dispersion limitées (i.e. espèces barochores). Cependant, même pour les espèces végétales avec de longues distances de dispersion, la probabilité qu'une propagule arrive dans un milieu favorable peut être fortement réduite à cause de la présence de barrières naturelles (court d'eau, montagne, forêt…) mais surtout à cause de l'augmentation de la présence de barrières anthropiques (route, urbanisation, champs…).

Adaptation locale via *la sélection génétique*

Le troisième type de réponse pour répondre aux changements globaux concerne l'adaptation locale *via* la sélection génétique (Hansen *et al.* 2012). Ce type de réponse repose sur la disponibilité dans les populations de variants génétiques qui seront sélectionnés lors de la marche adaptative vers un nouvel optimum phénotypique. Plusieurs sources peuvent être à l'origine de ces variants génétiques (Bay *et al.* 2017) :

- Immigration d'allèles à partir de populations proches (i.e. 'genetic rescue') (Hoffman & Sgrò 2011).

- Apparition de nouvelles mutations dans les populations naturelles (i.e. *de novo* mutations).
- Variation génétique préexistante dans les populations naturelles (i.e. 'standing genetic variation').

Toutefois, ces trois sources de variants génétiques n'apparaissent pas équivalentes en termes de vitesse de réponse à des changements globaux rapides. En effet, les contraintes imposées par l'attente de nouvelles mutations peut entraîner une limite adaptative rapidement atteinte par les populations (Stapley *et al.* 2010, Hancock *et al.* 2011). Les flux de gènes peuvent effectivement permettre l'arrivée dans une population d'allèles pré-adaptés à des modifications environnementales. Cependant, les flux de gènes entre populations proches restent limités sur une courte période de temps ; et ce phénomène est d'autant plus accentué par l'augmentation des barrières anthropiques comme décrit précédemment. Par ailleurs, une contrepartie associée aux flux de gènes est l'immigration simultanée d'allèles maladaptés augmentant potentiellement le fardeau génétique des populations.

Il apparaît donc indispensable que les populations possèdent une variation génétique préexistante suffisante pour répondre rapidement aux changements globaux. Une population génétiquement diversifiée permet d'augmenter la probabilité que certains individus aient une combinaison génétique favorable pour répondre à de nouveaux stress environnementaux, et ainsi permettre d'atteindre un nouvel optimum phénotypique plus rapidement (Chevin *et al.* 2010, **Figure 4**).

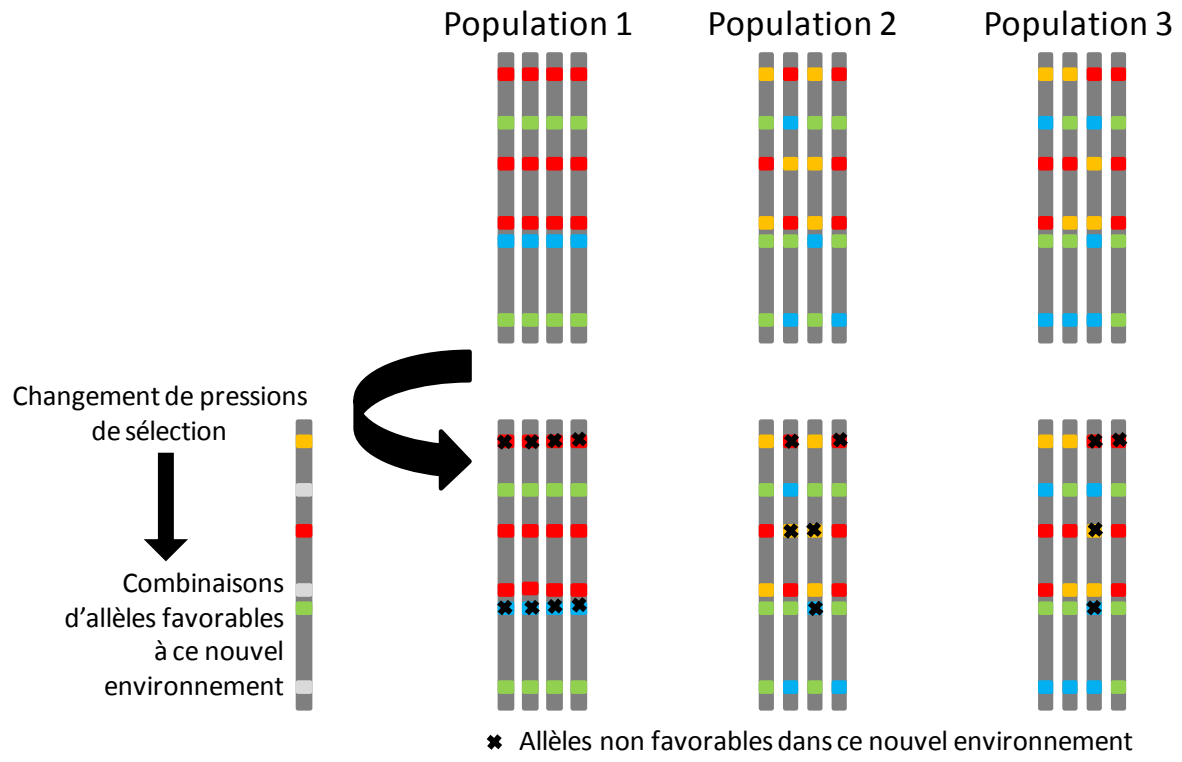


Figure 4: Illustration de l'importance de la diversité génétique préexistante dans les populations naturelles pour répondre rapidement aux changements globaux. Trois populations de quatre individus homozygotes sont représentées ici. La population 1 est génétiquement monomorphe alors que les populations 2 et 3 présentent une diversité génétique. Lors d'un changement environnemental symbolisé par la flèche noire, une combinaison d'allèles devient alors optimale. Cette combinaison « idéale » pour le nouvel environnement est représentée sur le côté gauche du schéma. Les allèles adaptatifs vis-à-vis du nouvel environnement sont colorés, alors que les allèles neutres sont en gris. La combinaison « idéale » n'étant pas présente dans la population 1, elle a de fortes chances de s'éteindre. En revanche, la combinaison « idéale » est déjà présente dans les deux autres populations, leur permettant de s'adapter rapidement à ce nouvel environnement et ainsi de se maintenir.

B. Identification des bases génétiques de l'adaptation

Comme nous l'avons vu précédemment, une réponse des espèces à des changements globaux rapides passera en partie par la sélection génétique, notamment à partir de la variation génétique préexistante. Afin de déterminer le potentiel adaptatif des populations naturelles face aux changements environnementaux d'origine abiotique et/ou biotique globaux (Bergelson & Roux 2010, Bay *et al.* 2017), un des enjeux majeurs en écologie évolutive est donc d'étudier l'architecture génétique de l'adaptation et implique de s'intéresser aux questions suivantes (liste non exhaustive) :

- Quel est le nombre de gènes sous-jacents à l'adaptation locale ?

- Quelle est la distribution des effets alléliques ?
- Quelle est l'identité des gènes adaptatifs et des fonctions biologiques associés?
- La pléiotropie et l'épistasie contribuent-elles à la marche adaptative vers un nouvel optimum phénotypique ?

En tirant bénéfice du développement récent de technologies de séquençage haut débit (Next Generation Sequencing technologies, NGS) permettant d'obtenir un nombre sans précédent de marqueurs génétiques (notamment de type Single Nucleotide Polymorphism, SNP), quatre approches complémentaires peuvent être utilisées pour étudier l'architecture génétique de l'adaptation, et plus particulièrement pour cartographier finement les gènes sous-jacents à l'adaptation (**Table 1**).

Approach	Data collected/resources required	What analysis reveals
Genetic differentiation outlier tests	Genome-wide SNPs from multiple populations	Allele frequencies for a SNP or SNPs that are differentiated across populations above what is expected from neutrality
Genetic-environment association	Genome-wide SNPs from multiple populations and environmental data for each population	Alleles at a SNP or SNPs that are associated with environmental variables over space
QTL mapping in a reciprocal transplant field experiment	Hybrids (F_2s, BCs, RILs, etc.) between locally adapted populations grown and phenotyped for fitness traits in reciprocal transplant common garden experiment	Use of hybrids allows identifying QTLs involved in local adaptation and the effect size of those QTLs on fitness; can resolve whether trade-offs at individual loci underlie local adaptation
GWAS	Genome-wide SNPs from hundreds of individuals grown in one or multiple common gardens; phenotypes and/or fitness for each individual	Identifies SNPs that are associated with traits associated with fitness measured under field conditions
Population-specific selective sweeps	Genome-wide SNPs from at least two populations and a recombination map	DNA sequences with longer-than-expected regions of extended haplotype homozygosity, which is consistent with a recent selective sweep in one of the populations

Table 1: Différentes approches permettant d'identifier les bases génétiques de l'adaptation locale (d'après Hoban *et al.* 2016). **SNP**: single nucleotide polymorphism, **QTL**: quantitative trait loci, **BC**: backcross, **RIL**: recombinant inbred lines, **GWAS**: genome-wide association studies.

Les méthodes 'genetic differentiation outlier tests' et 'population-specific selective sweeps' ne s'appuient que sur des données génomiques et cherchent à identifier des régions génomiques dont les patrons de diversité et de sélection s'écartent des attendus neutres.

Bien que largement utilisées, ces deux méthodes fournissent une liste de gènes candidats qu'il est souvent difficile de relier à des agents sélectifs potentiels ou à des traits phénotypiques adaptatifs. Pour palier ce manque, des méthodes statistiques permettant d'identifier le long du génome des polymorphismes génétiques associés soit à des variables écologiques ('Genome-Environment Association', GEA), soit à des traits phénotypiques ('QTL mapping traditionnel' et 'GWAS') ont été développées. Ci-après, je me focalise plus particulièrement sur ces méthodes d'analyses d'association génome-environnement et de QTL mapping sur des traits supposés adaptatifs.

1. Méthodes d'analyses génome-environnement

Les analyses de type association génome-environnement (Genome-Environment Association, GEA) reposent sur l'effet de gradients écologiques sélectifs spatiaux sur la variation génomique d'une espèce donnée (Lasky *et al.* 2012). En effet, un environnement hétérogène au niveau abiotique ou biotique conduit à différents optima phénotypiques locaux, qui se traduiront par une différenciation spatiale des variants génétiques sous-jacents aux phénotypes impliqués dans la réponse aux agents sélectifs. Ce type d'analyses permet donc non seulement d'identifier des gènes potentiellement impliqués dans l'adaptation mais aussi de décrire les facteurs écologiques responsables de leur divergence génétique entre les populations (Pluess *et al.* 2016, **Figure 5**).

Réalisées dans un premier temps par des approches individu-centré où un seul individu par population était caractérisé au niveau génomique, les analyses de type GEA ont été réalisées par la suite en adoptant des approches populationnelles afin de tirer bénéfice des informations apportées par la variation génétique intra-population.

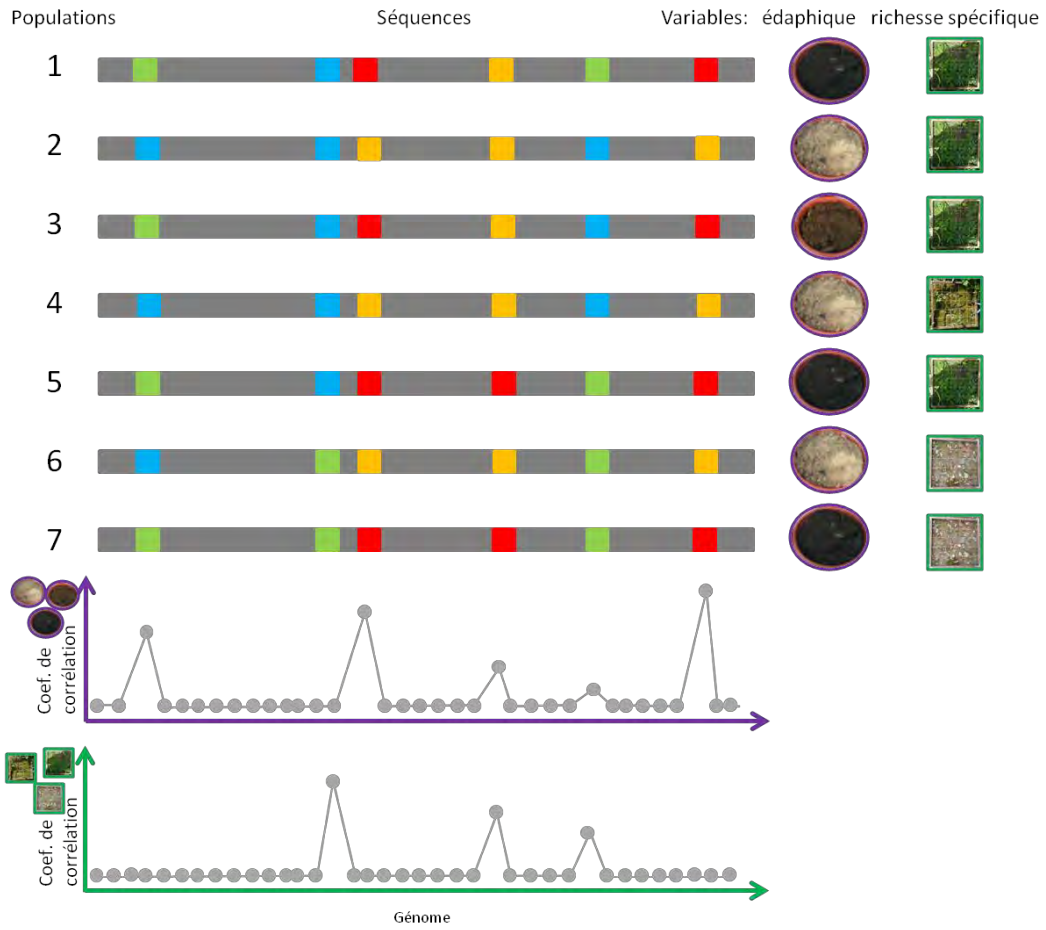


Figure 5: Principe de l'analyse d'Association Génome-Environnement (GEA). Les génomes sont représentés pour sept populations. A droite, la caractérisation d'une variable édaphique et d'un descripteur des communautés végétales est indiquée pour chacune des populations. En bas du schéma, les coefficients de corrélation entre la variation génétique et la variation écologique (violet : variable édaphique, vert : richesse spécifique) sont tracés en fonction de la position des marqueurs génétiques polymorphes le long du génome.

Approche individu-centré

Une des premières études de type GEA le long du génome chez une espèce sauvage a été réalisée par une approche individu-centré (Hancock *et al.* 2011). Pour réaliser cette étude, les auteurs se sont basés sur 948 accessions européennes de l'arabette des dames (*Arabidopsis thaliana*) génotypées pour 215k SNPs et caractérisées pour 13 variables climatiques (extrêmes et saisonnalité des températures et des précipitations). Pour identifier les régions génomiques associées aux variations climatiques, les auteurs ont utilisé un test partiel de Mantel basé sur le calcul du coefficient de corrélation de Spearman entre un SNP donné et une variable climatique, tout en contrôlant pour l'effet de l'histoire démographique d'*A. thaliana* en utilisant une matrice d'apparentement entre les 948 accessions calculée à partir des 215k SNPs. Ces analyses ont permis d'établir une carte

génomique de l'adaptation au climat chez *A. thaliana*. En effet, les SNPs les plus fortement corrélés au climat étaient significativement enrichis en variants génétiques correspondant à des changements d'acide aminé, suggérant que l'approche individu-centré de type GEA a bien permis d'identifier des loci adaptatifs.

A partir de 202 accessions de la luzerne tronquée (*Medicago truncatula*) génotypées pour environ 2 millions de SNPs, une approche individu-centré a aussi été adoptée pour identifier les loci candidats sous-jacents à l'adaptation à trois gradients climatiques, à savoir la température annuelle moyenne, les précipitations du mois le plus pluvieux et l'isothermalité (Yoder *et al.* 2014). Pour identifier les SNPs les plus associés à ces trois variables climatiques, les auteurs ont utilisé un modèle linéaire mixte incluant une matrice d'apparentement pour contrôler les effets confondants associés à l'histoire démographique de *Medicago truncatula*. En se basant sur l'étude des signatures de sélection dans les régions génomiques encadrant les SNPs les plus fortement corrélés au climat, les auteurs ont conclu que les loci sous-jacents à l'adaptation au climat étaient soumis à un balayage sélectif basé sur de la variation génétique préexistante (i.e. 'soft selective sweeps').

A partir de 1943 cultivars africains et indiens génotypés pour 404 627 SNPs et caractérisés pour des variables climatiques et édaphiques, une approche individu-centré a été récemment adoptée chez une plante d'intérêt agronomique, i.e. le sorgo commun (*Sorghum bicolor*) (Lasky *et al.* 2015). Un modèle mixte linéaire incluant une matrice d'apparentement a aussi été utilisée pour identifier les SNPs les plus associés au climat et au sol. Comme précédemment observé chez *A. thaliana*, les SNPs les plus fortement corrélés aux variables écologiques étaient significativement enrichis en variants génétiques correspondant à des changements d'acide aminé. Ces résultats ont ainsi permis aux auteurs de proposer des gènes candidats de réponse à la sécheresse et de tolérance à la toxicité à l'aluminium qui pourront être intégrés dans des programmes d'amélioration variétale.

Approche populationnelle

Bien que très puissantes, les approches individuelles négligent la variation génétique au sein des populations. Cependant, comme cela a été énoncé précédemment, il est

important de considérer la variation génétique intra-populationnelle pour obtenir une meilleure estimation du potentiel adaptatif des populations naturelles. Ainsi, une étude effectuée à partir de simulations a démontré qu'une approche populationnelle permettrait d'augmenter la puissance statistique des associations entre variation des fréquences alléliques des populations et variation des facteurs écologiques (Lotterhos & Whitlock 2015).

Ces approches populationnelles ont été peu utilisées jusqu'à présent car le séquençage de plusieurs individus par population peut se révéler très coûteux. Pour palier ce problème, une alternative a été proposée par l'utilisation de l'approche Pool-Seq. Cette approche consiste pour une population donnée, à extraire l'ADN de chaque individu, puis de créer un bulk de manière équimolaire et de séquencer ce bulk (Schlöterrer *et al.* 2014). Plusieurs études ont démontré que les fréquences alléliques estimées à partir d'une approche Pool-Seq étaient fortement corrélées aux fréquences alléliques obtenues à partir d'un séquençage individuel (Schlöterrer *et al.* 2014, Fracasetti *et al.* 2015).

Les analyses de type GEA basées sur des données de fréquences alléliques peuvent être réalisées suivant différentes méthodes (De Villemereuil *et al.* 2014, Lotterhos et Whitlock 2015), dont en voici un aperçu (liste non exhaustive):

- Méthodes basées sur la différentiation génétique entre populations naturelles vivant dans des habitats écologiques contrastés.
  (i) **BayeScan** (Foll & Gaggiotti 2008). Basée sur l'estimation d'un indice de différenciation génétique entre populations (F_{ST}), cette méthode Bayésienne permet d'identifier des marqueurs génétiques présentant des valeurs extrêmes de F_{ST} entre des populations écologiquement différentes.
  (ii) **BayeScEnv** (De Villemereuil & Gaggiotti 2015). Cette méthode (elle-aussi Bayésienne) est une extension de BayeScan incorporant une information environnementale sous la forme d'une différenciation environnementale continue entre populations. Cette méthode permet ainsi d'associer une variable environnementale spécifique à une forte différenciation génétique.

- Méthodes permettant d'estimer l'intensité d'une relation entre les fréquences alléliques et un gradient écologique.

  (iii) **Modèle mixte incluant des facteurs latents (Latent factor Mixed Model, LFMM)** (Frichot *et al.* 2013)**.** Dans un premier temps, cette méthode génère des variables latentes modélisant l'histoire démographique de l'espèce (i.e. structure génétique des populations) par une approche comparable à une Analyse en Composantes Principales. Afin de limiter le taux de faux positifs (fausses association génome-environnement), ces variables latentes sont par la suite incorporées comme co-variables dans les modèles de régression entre les fréquences alléliques et les variables écologiques.

  (iv) **BayEnv** (Coop *et al.* 2010). Pour tenir compte de l'histoire démographique de l'espèce, cette méthode Bayésienne estime dans un premier temps une matrice de covariance des fréquences alléliques entre populations à partir d'un sous-jeu de marqueurs génétiques, permettant ainsi d'estimer un modèle nul auquel les corrélations entre fréquences alléliques à un marqueur génétique donné et une variable écologique sont comparées. La significativité de cette comparaison est approximée par l'estimation d'un facteur bayésien (Bayes Factor, BF). Récemment, une nouvelle version de BayEnv (BayEnv2) a été mise à disposition afin d'intégrer les données obtenues à partir d'une approche Pool-Seq (Günther et Coop, 2013).

  (v) **BayPass** (Gautier 2015)**.** Elaborée à partir de la méthode BayEnv2, cette méthode Bayésienne permet une meilleure estimation de la matrice de covariance populationnelle *via* une modification du paramétrage des *priors* concernant la fréquence moyenne des allèles de référence. Par ailleurs, BayPass propose des stratégies alternatives de modélisation de la relation entre fréquences alléliques et variables écologiques. Par exemple, au-delà du modèle STD ('Standard covariate model') qui correspond à une extension du modèle développé dans BayEnv, BayPass propose le modèle AUX ('Auxiliary variable covariate model') qui introduit dans le modèle STD une variable auxiliaire binaire qui est attachée au coefficient de régression 'fréquences alléliques – variation écologique' de chaque marqueur génétique testé, permettant ainsi de classifier chaque marqueur génétique testé comme

associé ou non ('binary decision') à une variable écologique donnée. Comme pour BayEnv2, la méthode BayPass est aussi adaptée aux données obtenues à partir d'une approche Pool-Seq.

Les études de GEA basées sur une approche populationnelle restent peu nombreuses chez les espèces sauvages mais deviennent de plus en plus populaires. Par exemple, à partir de cinq populations d'arabette de Haller (*Arabidopsis halleri*) caractérisées pour cinq variables topo-climatiques et pour lesquelles les fréquences alléliques le long du génome ont été obtenues par une approche Pool-Seq, 175 gènes ont été identifiés (à partir de test partiels de Mantel) comme significativement associés à une ou plusieurs des cinq variables topo-climatiques (Fischer *et al.* 2013).

Une étude comparative menée sur trois espèces de chênes (*Quercus petraea*, *Quercus pubescens*, *Quercus robur*) en Suisse a été réalisée à partir de 71 populations (~20 individus par population) caractérisées (i) par une approche Pool-Seq d'amplicons de gènes candidats (Rellstab *et al.* 2016) et (ii) pour 31 variables abiotiques (topographie, climat et sol). En utilisant la méthode LFMM, les auteurs ont identifié sept gènes communs entre les 3 espèces, comme significativement associés aux mêmes facteurs abiotiques (précipitations et teneur en argile dans les sols).

A partir de 10 populations de l'épinoche à trois épines (*Gasterosteus aculeatus*) localisées dans la Mer Baltique et caractérisées pour plus de 30 000 SNPs obtenus par une approche Pool-Seq, de nombreuses régions génomiques ont été identifiées *via* la méthode BayEnv comme associées à des gradients de température et de salinité (Guo *et al.* 2015).

Nécessité de tenir compte des pressions de sélection non seulement abiotiques, mais aussi biotiques

L'approche de type GEA s'avère très puissante pour identifier les bases génétiques associées à des variables écologiques, permettant ainsi d'obtenir une meilleure compréhension des processus sous-jacents à l'adaptation locale de certaines espèces à leur environnement. Cependant, à notre connaissance, toutes les études GEA réalisées jusqu'à présent se sont intéressées uniquement à des variables abiotiques. Notamment, du fait de la

disponibilité d'un nombre important de bases de données climatiques, la majorité des études de type GEA ont porté sur l'identification de régions génomiques associées au climat (Bay *et al.* 2017). Dans une moindre mesure, des études de type GEA ont porté sur des variables édaphiques, soit par identification de régions génomiques spatialement différenciées entre des populations échantillonnées sur deux habitats édaphiques très contrastés (*Arabidopsis lyrata* sur des sols serpentiniques et des sols non-serpentiniques, Turner *et al.* 2010), soit par identification de régions génomiques associées à un gradient édaphique (Lasky *et al.* 2015, Pluess *et al.* 2016, Rellstab *et al.* 2016).

Or, au cours de son cycle de vie, un individu ne répond pas seulement à des conditions abiotiques. En effet, un individu interagit simultanément et séquentiellement, de manière directe ou indirecte, avec une large gamme de partenaires biotiques, dont les relations peuvent varier du mutualisme à la pathogénicité en passant par la compétition avec d'autres espèces ou avec ses congénères (Roux & Bergelson 2016). Par ailleurs, il est prédit que les interactions interspécifiques peuvent fortement affecter la réponse des espèces et des communautés biotiques aux changements climatiques (Gilman *et al.* 2010, Singer *et al.* 2013). Ainsi, de nombreux agents sélectifs potentiels y compris les facteurs biotiques doivent être considérés afin d'obtenir une vision plus complète du paysage génomique de l'adaptation locale.

2. Méthodes de QTL mapping sur des traits supposés adaptatifs

Comprendre l'architecture génétique de l'adaptation d'une espèce peut également passer par l'étude des relations existantes entre variation génétique et variation naturelle de traits phénotypiques supposés adaptatifs. Ci-dessous, je présente les deux grandes catégories de méthodes qui peuvent être utilisées pour cartographier les QTLs (Quantitative Trait Loci) associés à la variation naturelle de traits phénotypiques (Bazakos *et al.* 2017).

La première catégorie de méthodes correspond aux analyses de liaison ou bien encore appelée QTL mapping traditionnel (Bergelson & Roux 2010). Ces méthodes ont vu le jour dès la fin des années 80 et sont basées sur une carte génétique correspondant à une représentation de la position de marqueurs génétiques les uns par rapport aux autres, avec les distances entre marqueurs exprimées en termes de fréquence de recombinaison. Le QTL

mapping traditionnel est réalisé à partir de populations ségrégées, i.e. des populations artificielles découlant de croisements entre deux ou plusieurs génotypes parentaux. On comprend alors aisément que ces méthodes de cartographie QTL traditionnelle aient été plus largement adoptées chez les espèces végétales que chez les espèces animales. En effet, chez les plantes, les croisements entre génotypes d'une même espèce génèrent en moyenne un plus grand nombre de descendants que les croisements au niveau intraspécifique chez les animaux. Par ailleurs, les croisements interspécifiques peuvent être plus facilement réalisés chez les espèces végétales que chez les espèces animales, permettant de s'intéresser à l'architecture génétique d'une adaptation propre à une espèce donnée.

Plusieurs types de populations de QTL mapping traditionnel peuvent être utilisés pour cartographier les marqueurs génétiques associés à la variation naturelle phénotypique (**Figure 6**). Les premières populations de QTL mapping traditionnel qui ont été développées correspondent à des populations F2, issues généralement de l'autofécondation d'un individu F1 hybride obtenu par croisement entre deux génotypes parentaux. L'avantage de ce type de populations est qu'elles permettent une estimation de la dominance des QTLs identifiés, ce qui peut être une information précieuse pour comprendre l'architecture d'un trait adaptatif. Cependant, un inconvénient majeur des populations F2 est que chaque individu F2 phénotypé doit aussi être génotypé (à l'exception des espèces végétales avec un mode de multiplication végétative).

Pour résoudre ce problème, les populations de lignées recombinantes consanguines (Recombinant Inbred Line **RIL, Figure 6 et Figure 7**) ont été développées et correspondent à des lignées issues de descendants F2 qui ont subi plusieurs générations d'autofécondation jusqu'à obtenir des lignées quasi-homozygotes représentant une mosaïque unique des deux génomes parentaux (Bazakos *et al.* 2017). Ces populations sont devenues très populaires car les lignées RILs peuvent être phénotypées presque indéfiniment mais génotypées qu'une seule fois puisqu'il s'agit de lignées quasi-homozygotes (Savolainen *et al.* 2013, Bazakos et al. 2017). Malgré ces avantages, la diversité génétique présente au sein d'une famille RIL est limitée à la diversité génétique des deux génotypes parentaux.

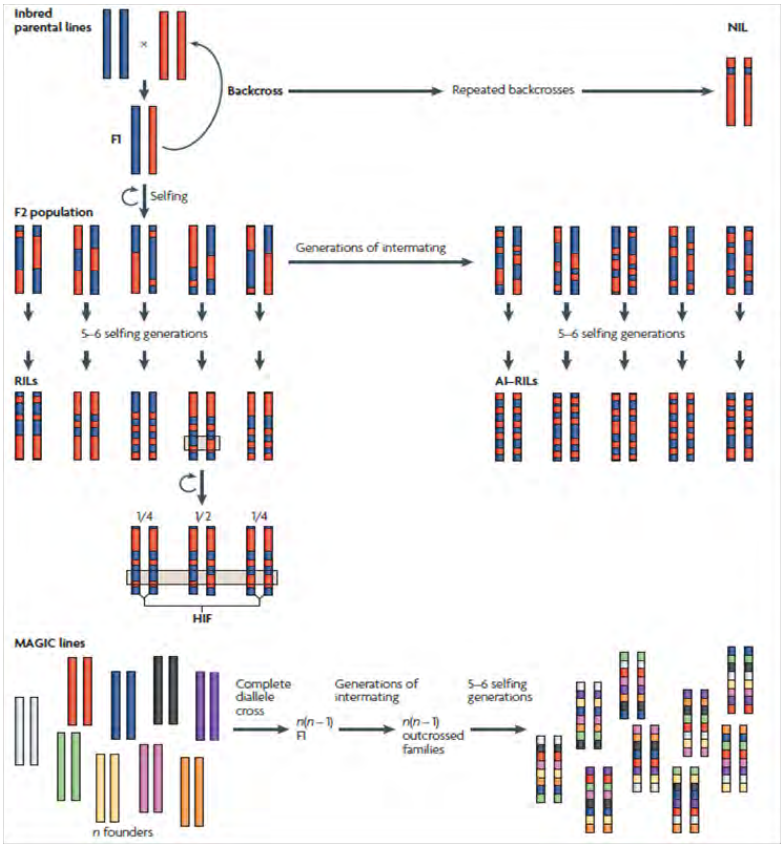


Figure 6: Illustration des différentes populations de QTL mapping traditionnel pouvant être utilisées pour cartographier les marqueurs génétiques associés à la variation naturelle phénotypique. **RILs**: Recombinant inbred line, **AI-RILS**: Advanced intercross-recombinant inbred lines, **HIF**: Heterogeneous inbred family, **MAGIC**: multiparent advanced generation inter-cross lines, **NIL**: near-isogenic line. D'après Bergelson & Roux 2010.

Ainsi, une troisième type de population a vu le jour il y a quelques années et correspond aux lignées MAGIC (multiparent advanced generation intercross, **Figure 6**) (Cavanagh *et al.* 2008) qui ont été développées à la fois chez des espèces sauvages comme *A. thaliana* (Kover *et al.* 2009) ou des espèces cultivées comme le blé ou le riz (Huang *et al.* 2012, Bandillo *et al.* 2013). Contrairement aux familles RILs, des croisements multiples sont effectués entre plusieurs génotypes parentaux et ceci sur plusieurs générations. Les descendants à partir de ce schéma de croisement sont ensuite autofécondés sur plusieurs générations, permettant de créer des lignées quasi-homozygotes comme pour les familles RILs. Ces populations MAGIC ont donc à la fois l'avantage d'être très diversifiées génétiquement mais aussi de contenir des lignées quasi-homozygotes. Par ailleurs, les générations de croisement supplémentaires permettent d'augmenter le nombre d'événements de recombinaison et donc d'obtenir une meilleure précision que les familles RILs quant à la localisation des QTLs le long du génome.

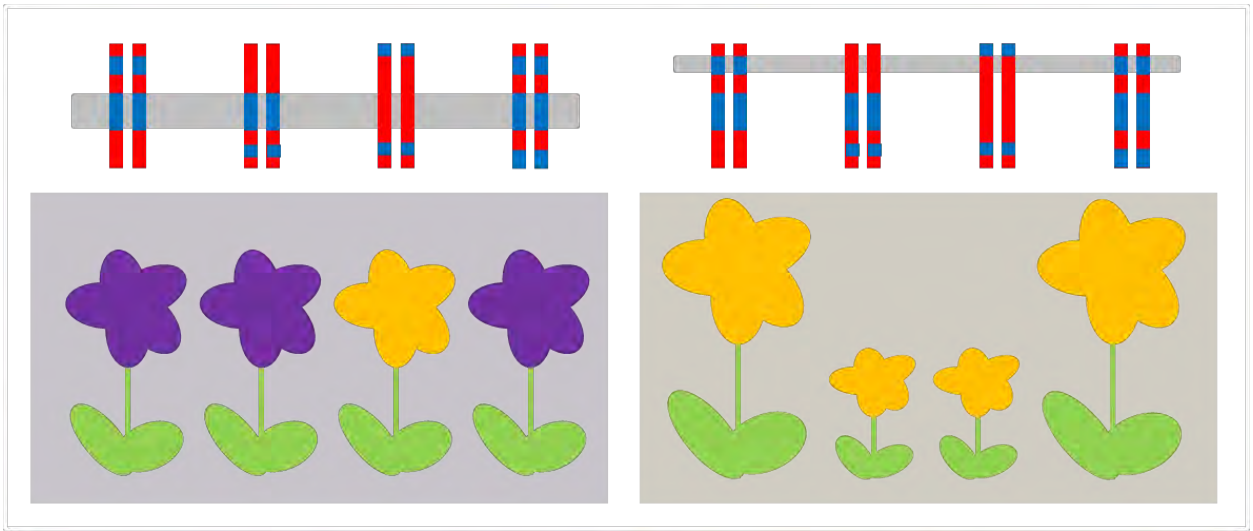


Figure 7: Illustration de la méthode de QTL mapping avec des familles RILs. Le QTL mapping permet d'associer une variation phénotypique au polymorphisme de certains marqueurs génétiques.

Malgré tout, les populations de QTL mapping traditionnel restent très peu résolutives et difficiles à mettre en place chez la plupart des espèces animales. Ainsi, la méthode de Genome-Wide Association (GWA) mapping est apparue comme une alternative puissante pour cartographier finement les régions génomiques associées à la variation phénotypique naturelle (**Figure 8**) (Mitchell-Olds & Schmitt 2006).

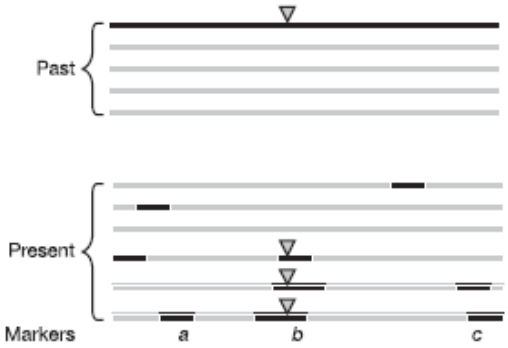


Figure 8: Illustration de la méthode de GWA mapping bénéficiant des événements de recombinaison passés. La figure du haut représente une population d'origine avec différentes versions d'un chromosome. Le chromosome noir symbolise une version ancestrale où une mutation s'est produite (représentée par un triangle) et responsable d'un nouveau phénotype. Après des milliers de générations, des événements de recombinaisons se sont accumulés. Les chromosomes de la population actuelle représentée en bas sont ainsi caractérisés par une mosaïque de régions génomiques dérivées de la population d'origine. Dans la nouvelle population, les trois individus du bas auront un phénotype différent des trois autres individus, dû à la mutation arrivée dans la population d'origine (triangle). Le marqueur moléculaire 'b' étant proche de ce polymorphisme, il sera ainsi corrélé avec le phénotype. En revanche, les marqueurs 'a' et 'c' sont distants de ce polymorphisme causal et ne seront donc pas corrélés au phénotype. D'après Mitchell-Olds & Schmitt (2006).

En effet, en tirant bénéfice des évènements de recombinaison qui se sont accumulés sur plusieurs centaines de milliers d'années (voir millions d'années), le GWA mapping utilise le déséquilibre de liaison (DL) naturelle présent dans une collection de génotypes naturels. Les estimations du DL sont très variables entre les espèces et dépendent principalement du régime de reproduction : d'environ 1bp chez la mouche du vinaigre (*Drosophila melanogaster*) (MacKay *et al.* 2012) jusqu'à 500kb chez la variété *temperate japonica* du riz (*Oryza sativa*) qui a un régime de reproduction fortement autogame (Mather *et al.* 2007), par exemple. La contrepartie d'un DL court est la nécessité d'avoir un nombre de marqueurs génétiques suffisants afin de balayer l'ensemble du génome lors des analyses de GWA mapping. Cependant, avec le développement des NGS, cet inconvénient deviendra de plus en plus rare dans les années à venir.

Deux autres inconvénients majeurs de la méthode GWA mapping (qui ne pourront être résolus par l'utilisation de NGS) restent les faux positifs et l'hétérogénéité génétique et/ou allélique. Comme pour les analyses de GEA, les faux positifs correspondent à de fausses associations génotype-phénotype qui résultent de l'effet de l'histoire démographique de l'espèce (**Figure 9**). Comme pour les analyses de GEA, plusieurs méthodes statistiques peuvent être utilisées pour corriger ces faux positifs. L'une des plus populaires consiste à intégrer dans un modèle mixte une matrice d'apparentement entre les génotypes utilisés dans l'étude phénotypique (Kang *et al.* 2010). Bien que très performantes, ces méthodes statistiques entraînent aussi l'apparition de faux négatifs (**Figure 9**), c'est-à-dire des marqueurs génétiques réellement associés à la variation phénotypique naturelle (i.e. marqueurs causaux) mais qui sont perdus après correction pour l'effet de l'histoire démographique de l'espèce (Brachi *et al.* 2010).

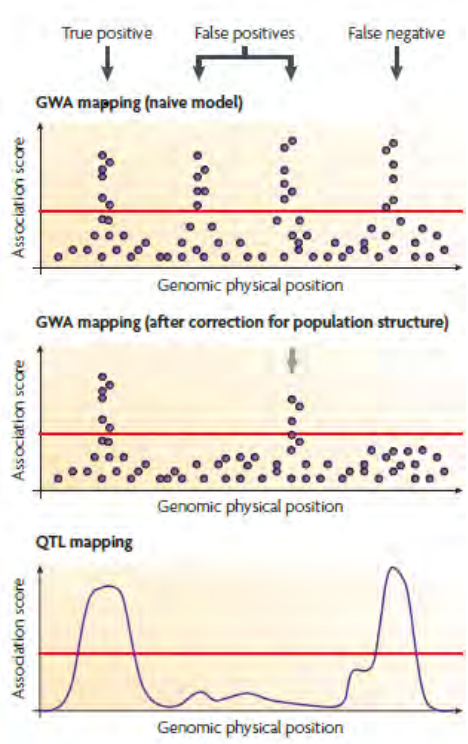


Figure 9: Illustration des faux positifs et faux négatifs dans les études de GWA mapping. D'après Bergelson & Roux (2010).

L'hétérogénéité génétique provient du fait qu'une même valeur phénotypique peut résulter de différentes combinaisons de QTLs (**Figure 10**). Par ailleurs, différents allèles à un même gène peuvent avoir le même effet et/ou des effets contrastés pour un phénotype donné, entraînant des phénomènes d'hétérogénéité allélique (**Figure 10**). Ces observations sont particulièrement bien documentées pour la date de floraison dont les bases génétiques ont été très étudiées chez différentes espèces végétales comme *A. thaliana* (Atwell *et al.* 2010) ou bien encore le maïs (*Zea mays*) (Buckler *et al.* 2009). Pour limiter les effets de l'hétérogénéité génétique et de l'hétérogénéité allélique, une solution proposée est de travailler à une échelle géographique régionale où la diversité génétique reste importante tout en étant restreinte par rapport à la diversité génétique observée sur l'ensemble de l'aire de répartition d'une espèce (Bergelson & Roux 2010). Travailler à une échelle géographique régionale présente aussi l'avantage de limiter les effets de l'histoire démographique de l'espèce considérée sur la détection des régions génomiques associées à la variation phénotypique naturelle (i.e. diminution des taux de faux positifs et de faux négatifs).

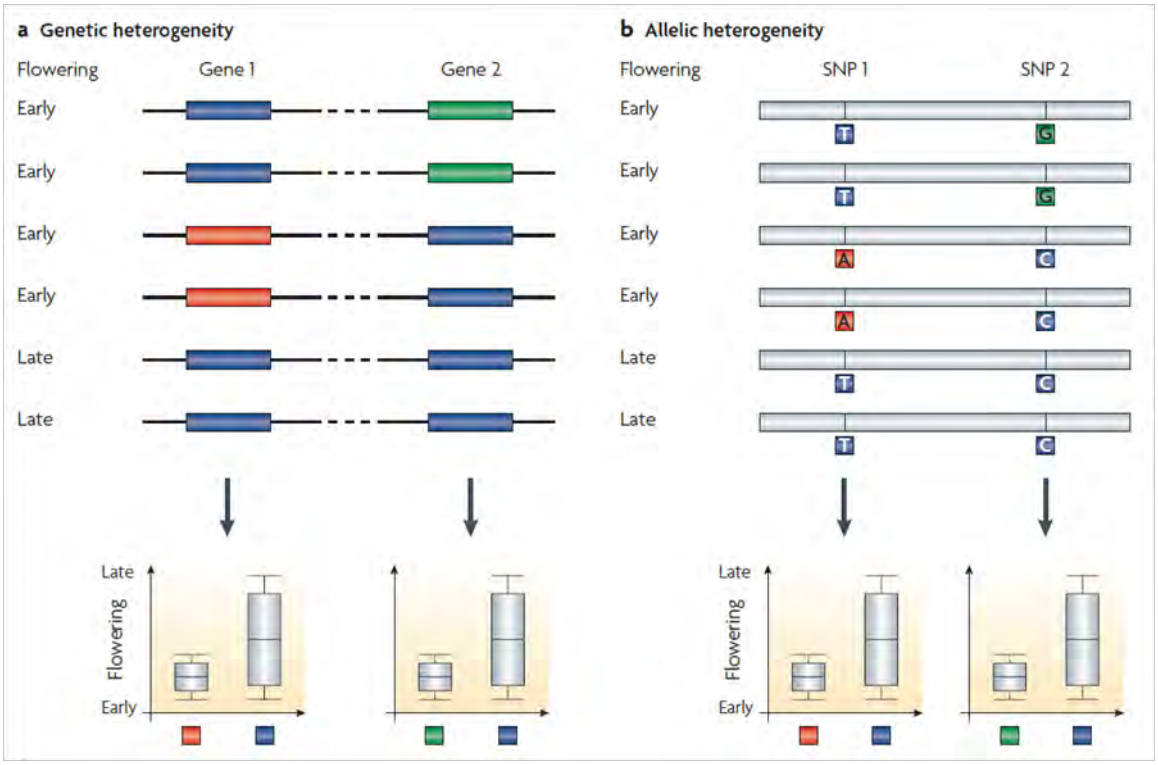


Figure 10: Illustration de l'effet de l'hétérogénéité génétique et de l'hétérogénéité allélique sur la détection de QTLs dans les études de GWA mapping. Cas de la date de floraison. D'après Bergelson & Roux (2010).

Bien que puissante et très prometteuse, (i) l'utilisation de l'approche GWA mapping reste encore l'apanage de quelques espèces sauvages modèles et de plus en plus des espèces cultivées (Bartoli & Roux 2017), et (ii) l'identité des pressions de sélection agissant sur les traits phénotypiques supposés adaptatifs est souvent suggérée mais rarement testée.

Replacer les études de QTL mapping dans un contexte écologiquement réaliste

La majorité des études de QTL mapping ont été réalisées soit dans des environnements contrôlés de laboratoire pour les espèces sauvages, soit dans des conditions agricoles (élevage ou champs cultivées) pour les espèces domestiquées. Cependant, la variation génétique d'une espèce est exposée à la sélection naturelle dans des habitats écologiquement contrastés. L'étude de l'architecture génétique de l'adaptation dans un contexte écologiquement réaliste fait ainsi appel aux approches développées en génomique écologique (**Figure 11**) (Ungerer *et al.* 2008).

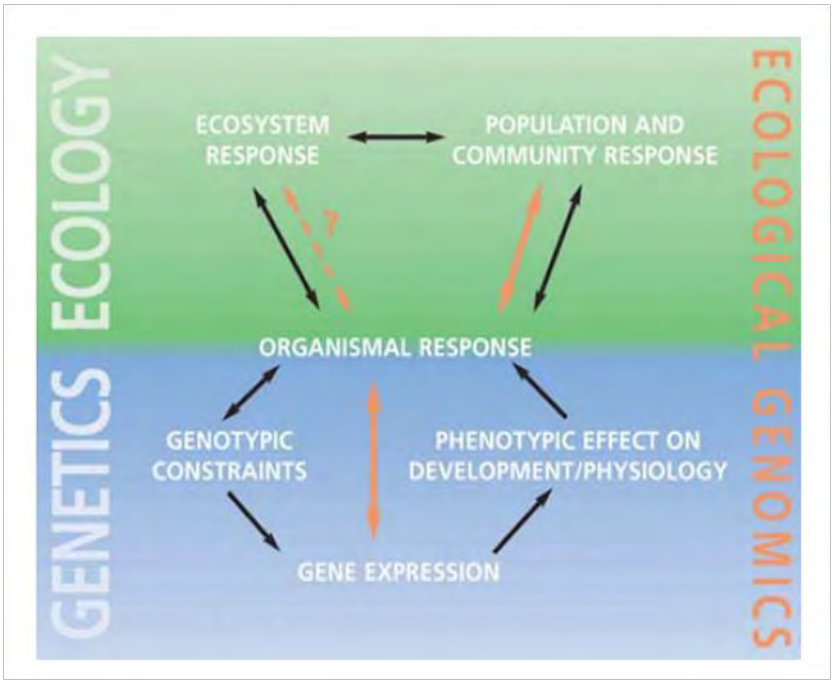


Figure 11: Figure illustrant le concept de génomique écologique. D'après Ungerer *et al.* 2008.

Il est donc important de considérer le contexte écologique dans lequel sont mesurés les phénotypes. En effet, au cours de leur cycle de vie, les individus perçoivent des signaux environnementaux complexes et variés. Ainsi, quelques études de QTL mapping réalisées chez *A. thaliana* ont pu mettre en évidence que l'architecture génétique de la date de floraison était très différente entre des conditions de phénotypage de serre et des conditions de phénotypage sur un terrain expérimental (Weinig *et al.* 2002, Brachi *et al.* 2010). Plus précisément, une étude de GWA mapping réalisée à partir de 184 accessions mondiales d'*A. thaliana* a montré qu'une majorité des gènes associés à la date de floraison mesurée sur un terrain expérimental dans le Nord de la France étaient impliqués dans la régulation de l'horloge circadienne, voie de régulation peu citée dans les études de QTL mapping de la date de floraison mesurée dans des conditions contrôlées de laboratoire (Brachi *et al.* 2010).

Cependant, les conditions écologiques rencontrées par les plantes sur un terrain expérimental peuvent être encore très éloignées des conditions rencontrées par les populations dans leurs habitats naturels. Ainsi, une étude de QTL mapping traditionnel basée sur une famille RIL (i) issue d'un croisement entre deux populations d'*A. thaliana* localement adaptées (une population localisée en Suède et une population localisée en Italie) et (ii) phénotypée dans les habitats d'origine des populations, a permis de mettre en évidence que

deux tiers des QTLs associés au nombre de fruits étaient spécifiques aux habitats d'origine (Ågren *et al.* 2013).

Replacer les études de QTL mapping dans un contexte écologiquement réaliste permettrait donc à termes de retracer les trajectoires évolutives des traits adaptatifs dans les populations naturelles.

C. Importance de l'échelle géographique considérée

A quelle échelle géographique étudier l'adaptation locale et l'architecture génétique sous-jacente ? Cette question est loin d'être anodine. En effet, bien que l'adaptation locale ait pu être observée à des échelles spatiales très variées (de l'ordre de quelques mètres à l'échelle d'un continent) (Richardson *et al.* 2014), les études de type GEA ou de GWA mapping ont la plupart du temps été réalisées à partir d'une collection de génotypes échantillonnées à une large échelle spatiale (i.e. monde, continent). Ce constat est particulièrement bien illustré par les études de type GEA portant sur l'adaptation au climat (Hancock *et al.* 2011, Lasky et al. 2012, Yoder et al. 2014, Lasky *et al.* 2015) ou les études de GWA mapping peu importe le trait phénotypique considéré (Atwell *et al.* 2010, Huang *et al.* 2012, Bartoli & Roux 2017).

Le choix de travailler à une large échelle géographique peut être expliqué par plusieurs raisons :

- Il est communément admis que travailler sur une collection de génotypes échantillonnés sur l'ensemble de l'aire de distribution d'une espèce permettra d'avoir accès à toute la variation naturelle phénotypique présente au sein de cette espèce. Bien que cela puisse effectivement être le cas pour certains traits, d'autres traits phénotypiques peuvent présenter presque autant de diversité au sein d'une population locale que sur l'ensemble de l'aire de distribution. Chez *A. thaliana*, cela a été observé pour la résistance qualitative et quantitative à des bactéries phytopathogènes (Huard-Chauveau *et al.* 2013, Karasov *et al.* 2014) ainsi que pour des traits phénologiques comme la date de floraison (Brachi *et al.* 2013).
- Dans le cadre des changements globaux, un des objectifs est d'estimer le potentiel adaptatif des espèces afin de prédire l'évolution de la biodiversité au

cours des prochaines décennies. Ainsi, identifier les bases génétiques de l'adaptation à l'échelle de l'aire de distribution d'une espèce devrait permettre de prédire le devenir des populations locales. Or, comme nous l'avons évoqué précédemment, les bases génétiques de l'adaptation à une pression environnementale peuvent être très différentes d'une région géographique à l'autre (i.e. hétérogénéité génétique) et même entre des populations locales proches géographiquement (Hoekstra *et al.* 2006, Brachi *et al.* 2013).

- Une raison beaucoup moins scientifique concerne le coût de la caractérisation génomique des génotypes utilisés dans les études de type GEA ou de GWA mapping, nécessitant la mise en place de consortiums internationaux où chaque laboratoire veut caractériser les génotypes collectés dans son propre pays.

Toutefois, certaines études ont indiqué la complémentarité de travailler à différentes échelles spatiales. Ainsi, l'importance de travailler à différentes échelles géographiques a été mise en évidence dans une étude de type GEA réalisée sur l'arabette des Alpes (*Arabis alpina*) au sein des Alpes (Alpes européennes, Alpes françaises et trois massifs montagneux dans les Alpes françaises) (Manel *et al.* 2010). En effet, le pourcentage de loci ALFP corrélés à 8 variables climatiques était dépendant de l'échelle géographique considérée (Européenne : 12% des loci AFLP, régionale : 11% locale : variable entre 3 et 17% suivant le massif montagneux). De même, une étude de GWA mapping réalisée sur *A. thaliana* à cinq échelles géographiques différentes (mondiale, européenne, française, régionale et locale) a révélé que les bases génétiques associées à la phénologie étaient très dépendantes de l'échelle géographique considérée (Brachi *et al.* 2013). Ces études suggèrent que deux types de réponse adaptative peuvent être considérées : (i) une adaptation locale site-spécifique dû à des pressions de sélection variant à une échelle spatiale fine, et (ii) une adaptation plus générale en réponse aux pressions de sélection agissant à une plus grande échelle spatiale.

Comme préconisé dans certaines études (Bergelson & Roux 2010), l'échelle géographique à laquelle étudier l'adaptation locale et l'architecture génétique sous-jacente doit donc être déterminée en fonction des échelles de variation spatiale des pressions de sélection auxquelles est confrontée l'espèce étudiée. Déterminer l'échelle géographique à laquelle travailler nécessite donc dans un premier temps d'identifier les

pressions de sélection potentielles agissant sur une espèce donnée. Alors que les bases de données climatiques permettent une caractérisation rapide des localités où ont été échantillonnés les génotypes, la caractérisation d'autres facteurs abiotiques (comme le sol) et des facteurs biotiques restent laborieuses. Dans un deuxième temps, il s'agit d'identifier à quelle(s) échelle(s) spatiale(s) varient ces pressions de sélection potentielles. Les échelles spatiales de variation écologique étant certainement très dépendantes de la pression de sélection considérée, cela entraînera certainement une superposition de grains de l'environnement, dont la complexité devra être intégrer lors de l'interprétation des résultats obtenus à partir d'analyses de type GEA effectuées sur une variable écologique particulière ou à partir d'analyses GWA mapping effectuées sur un trait supposé adaptatif.

En complémentarité de l'échelle spatiale de variation des pressions de sélection potentielles, le choix de l'échelle spatiale d'étude dépendra aussi de la distance de dispersion de l'espèce étudiée. La dispersion moyenne des espèces est de l'ordre de quelques kilomètres par décennies (Chen *et al.* 2011), loin de l'échelle géographique utilisée dans la majorité des études de type GEA et de GWA mapping qui est de l'ordre de centaines voire milliers de kilomètres. Ainsi, considérer une plus petite échelle spatiale permettrait d'être plus cohérent avec la distance de migration des espèces.

D. Objectifs de la thèse et modèle d'étude

Un des principaux challenges en écologie évolutive est de prédire le potentiel adaptatif des populations naturelles *via* l'étude de l'architecture génétique de l'adaptation et notamment l'identification des bases génétiques de l'adaptation. Une telle connaissance du potentiel adaptif pourrait permettre de mettre en place des plans de gestion des espèces, en priorisant certaines populations ou certaines aires géographiques. D'un point de vue agronomique, étudier l'architecture génétique de l'adaptation et des bases génétiques sous-jacentes pourrait faciliter les programmes de sélection visant à sélectionner des génotypes qui maintiendront la productivité face aux changements globaux actuels, avec le challenge supplémentaire d'intégrer une diminution des intrants dans les pratiques agricoles.

Ce challenge en écologie évolutive passe par trois étapes successives que j'ai abordées durant ma thèse:

- Identification des pressions de sélection potentielles et leurs échelles spatiales de variation : comme nous l'avons mentionné précédemment, le principal agent sélectif étudié est le climat du fait de la disponibilité de bases de données climatiques. Cependant, les espèces vivent dans des milieux très hétérogènes aussi bien d'un point de vue abiotique que d'un point de vue biotique. Il est donc nécessaire de caractériser autant faire que se peut les environnements abiotiques et biotiques auxquels les populations naturelles sont confrontées. Ceci permettra non seulement d'appréhender la diversité des agents sélectifs potentiels mais aussi d'étudier l'importance relative des ces agents vis-à-vis de l'espèce étudiée. Par ailleurs, l'étude de l'échelle spatiale de variation des agents sélectifs potentiels permettra d'établir le maillage des grains de l'environnement et ainsi déterminer à quelle échelle géographique travailler pour identifier les bases génétiques de l'adaptation.
- Identification des bases génétiques associées aux agents sélectifs potentiels par une approche de génomique environnementale: le récent développement des technologies NGS couplé au développement de méthodes d'analyses statistiques puissantes représente une incroyable opportunité de réaliser des scans génomiques pour identifier les régions génomiques associées à un facteur écologique. Les résultats obtenus permettraient (i) d'avoir un aperçu de l'architecture génétique de l'adaptation, (ii) de connaître l'identité des fonctions biologiques cibles de la sélection naturelle, et (ii) de tester si des gènes peuvent être impliqués dans l'adaptation à plusieurs variables écologiques.
- Etude de la dynamique adaptative dans un milieu spatialement hétérogène par une approche de génomique écologique : les deux étapes précédentes sont basées sur des approches corrélatives visant à estimer le potentiel adaptatif des populations naturelles. Pour tester ce potentiel adaptatif, il est complémentaire d'étudier la dynamique adaptative de populations naturelles, notamment sur une courte échelle de temps étant donné la rapidité à laquelle s'effectue les changements globaux.

Arabidopsis thaliana : une espèce modèle en génomique environnemental et en génomique écologique

Durant ma thèse, j'ai abordé ces trois étapes successives de l'étude de l'adaptation en utilisant le modèle biologique *A. thaliana* (**Figure 12**). Cette plante annuelle de la famille des Brassicaceae a un régime de reproduction largement autogame (taux d'allogamie moyen = 2% mais ce taux peut varier de 0 à 20% suivant les populations naturelles ; Platt *et al.* 2010). Native d'Eurasie et naturalisée notamment en Amérique du Nord, on la trouve principalement dans des milieux rudéraux (Mitchell-Olds & Schmitt 2006, Bossdorf *et al.* 2009). En Europe, son principal cycle de vie consiste à germer à l'automne, passer l'hiver sous forme de rosette, fleurir au début du printemps et produire des graines aux mois de mai-juin.

Figure 12: Photographies d'*Arabidopsis thaliana*.

A. thaliana constitue une espèce modèle idéale pour mener des études à l'interface de plusieurs disciplines (Mitchell-Olds & Schmitt 2006). Avec (i) un cycle de vie généralement très court dans des conditions contrôlées de serre (6 semaines entre la germination de la graine et la maturation du premier fruit appelée silique ; Meinke *et al.* 1998), (ii) un petit génome (120Mb, 5 chromosomes) dont le séquençage a été achevé en 2000 (accession Columbia Col-0, The Arabidopsis Genome Initiative 2000) et (iii) une large gamme de ressources génétiques publiques pour étudier la variation phénotypique artificielle (mutants T-DNA par exemple), elle est depuis trois décennies LE modèle d'étude de choix en

génétique moléculaire (Meyerowitz & Somerville 2002). Par ailleurs, l'effort commun de plusieurs laboratoires internationaux depuis plus de 15 ans ont permis de créer des ressources génétiques importantes pour étudier la variation naturelle phénotypique et ses bases génétiques. Rendues publiques et disponibles *via* des centres de ressources génétiques (NASC, ABRC, INRA Versailles), ce sont plusieurs milliers d'accessions naturelles qui sont à la disposition de la communauté scientifique travaillant sur *A. thaliana* (Platt *et al.* 2010, Horton *et al.* 2012, The 1001 Genome Consortium 2016).

Depuis plus d'une dizaine d'années, *A. thaliana* apparaît aussi comme une espèce modèle en écologie évolutive (Gaut 2012). Sur son aire de répartition mondiale, *A. thaliana* est présente dans une grande diversité d'habitats aussi bien d'un point de vue abiotique que biotique (Jakob *et al.* 2002, Mitchell-Olds & Schmitt 2006, Shindo *et al.* 2007). En particulier, alors qu'*A. thaliana* est décrite comme une espèce pionnière souvent trouvée dans des milieux pauvres ou perturbés, rarement en compétition avec d'autres espèces, des études récentes et des observations sur le terrain semblent prouver le contraire. En effet, lors de diverses prospections de populations naturelles d'*A. thaliana* en France réalisées par notre équipe depuis 2009, nous avons pu observer *A. thaliana* dans des milieux très compétitifs (**Figure 13**).

Figure 13: Différentes populations d'*Arabidopsis thaliana* dans des milieux compétitifs de la région Midi-Pyrénées. Les flèches rouges indiquent *A. thaliana*.

Cette diversité d'habitats s'observe même à une échelle géographique de l'ordre de quelques kilomètres (Brachi *et al.* 2013). En accord avec ces observations, une variation génétique importante a été observée pour de nombreux traits phénotypiques (morphologique, phénologique, physiologique, etc. Atwell *et al.* 2010) à différentes échelles

géographiques, voire au sein de populations naturelles (Brachi *et al.* 2013, Huard-Chauveau *et al.* 2013).

Ainsi, (i) les connaissances sur le développement, la génétique et la physiologie d'*A. thaliana*, (ii) la diversité des habitats rencontrés par *A. thaliana*, (iii) la disponibilité de ressources génétiques, et (iv) le développement des technologies NGS couplé au développement de méthodes d'analyses statistiques puissantes, font d'*A. thaliana* une espèce de choix pour aborder des questions en écologie et en biologie évolutive (Koornneef *et al.* 2004, Mitchell-Olds & Schmitt 2006).

E. Plan de la thèse

En adoptant des approches de génomique environnementale et de génomique écologique, je me suis intéressée lors de ma thèse à l'étude du potentiel adaptatif de populations naturelles d'*A. thaliana*, reposant notamment sur l'identification des bases génétiques associées à son adaptation à différents agents sélectifs. Tout au long de ma thèse, un réalisme écologique a été gardé en considérant plusieurs échelles géographiques (régionale, habitat, locale), plusieurs pressions de sélection (abiotique et biotique) et en travaillant avec plusieurs individus par population afin de prendre en compte la variation génétique au sein des populations naturelles.

Dans leurs habitats naturels, les espèces sont soumises simultanément à de nombreuses pressions de sélection abiotique et biotique. Pour comprendre l'adaptation des populations face aux changements globaux, il est important d'identifier ces pressions de sélection et de comparer leur importance relative avant d'identifier les bases génétiques de l'adaptation. Dans un premier chapitre, je me suis donc intéressée à identifier ces facteurs écologiques susceptibles d'être reliés à un proxy de fitness (i.e. production de graines) et aux stratégies reproductives d'*A. thaliana* dans des populations naturelles. J'ai notamment porté mon attention sur la caractérisation biotique des populations naturelles. Par ailleurs, des études précédentes au sein de l'équipe ayant montré à une échelle régionale (i) une variation écologique très importante (climat, sol, intensité de la compétition interspécifique) et (ii) un effet de l'histoire démographique limitée (diminution du taux de faux positifs et des

problèmes d'hétérogénéité génétique et allélique) lors d'analyses GWA mapping (Brachi *et al.* 2013, Baron *et al.* 2015), mes travaux sur l'identification des agents sélectifs potentiels ont porté sur un nouveau jeu de 168 populations naturelles localisées dans la région Midi-Pyrénées. Ce chapitre vise à répondre à trois grandes questions : (i) Quelle est la variation naturelle *in situ* d'un proxy de fitness (i.e. production de graines) et des stratégies reproductives chez *A. thaliana*?, (ii) Quelle est la variation des agents sélectifs potentiels et à quelles échelles spatiales varient-ils?, et (iii) Quels sont les agents sélectifs potentiels agissant sur *A. thaliana*?

Dans un second chapitre, en me basant sur des données de séquençage génomique des 168 populations naturelles obtenues par une approche Pool-Seq, j'ai effectué des analyses d'association génome-environnement (GEA) pour identifier les bases génétiques d'*A. thaliana* associées à des variables climatiques et à des descripteurs des communautés végétales. Ce chapitre s'attachera à répondre à plusieurs questions : (i) Quelles sont les régions génomiques associées à ces agents sélectifs potentiels?, (ii) Ces régions génomiques présentent-elles aussi des traces d'adaptation locale, suggérant que nous avons bien identifié des gènes adaptatifs ?, et (iii) Les régions génomiques identifiées sont-elles communes entre différents agents sélectifs ?

Enfin, dans un troisième chapitre, je me suis intéressée à étudier la dynamique adaptative d'une population locale d'*A. thaliana* dans un milieu spatialement hétérogène. Pour mener à bien cette étude, j'ai utilisée la population bourguignonne TOU-A (i) située dans un environnement hétérogène au niveau du sol et de la compétition interspécifique et (ii) pour laquelle une augmentation de température de 1°C a été observée depuis la fin des années 1980. Par ailleurs, les graines de 195 accessions naturelles ont été récoltées en 2002 (n = 80) et 2010 (n = 115). Une expérience de résurrection et le séquençage génomique individuel des 195 accessions m'ont permis d'aborder les questions suivantes : (i) Observe-t-on une évolution phénotypique adaptative d'*A. thaliana* en moins de 8 générations ? (ii) Cette évolution phénotypique est-elle dépendante des conditions abiotiques et biotiques ? et (ii) Quelle est l'architecture génétique d'*A. thaliana* sous-jacente à cette évolution phénotypique adaptative ?

Chapitre 1

Identification des pressions de sélection potentielles agissant sur *A. thaliana* à une échelle régionale

A. Introduction

Afin de comprendre l'adaptation d'une espèce face aux changements globaux, il est essentiel dans un premier temps d'identifier les pressions de sélection pouvant agir simultanément sur cette espèce et de comparer leur importance relative. Pour mener à bien ce projet chez *A. thaliana*, j'ai adopté une approche de type 'association fitness – variation écologique' pour identifier et comprendre quelles pressions de sélection peuvent potentiellement agir sur cette espèce.

Dans ce chapitre, je me suis focalisée sur 168 populations naturelles d'*A. thaliana* qui ont été caractérisées d'un point de vue phénotypique et écologique (**Figure 1**). En introduction de ce chapitre, je présente rapidement les différentes variables phénotypiques et écologiques utilisées. Les variables écologiques sont décrites plus amplement dans les manuscrits présentés dans le chapitre 2 ou en annexe de ma thèse, sauf pour les variables édaphiques dont les analyses n'ont pu être inclues dans ma thèse par faute de temps.

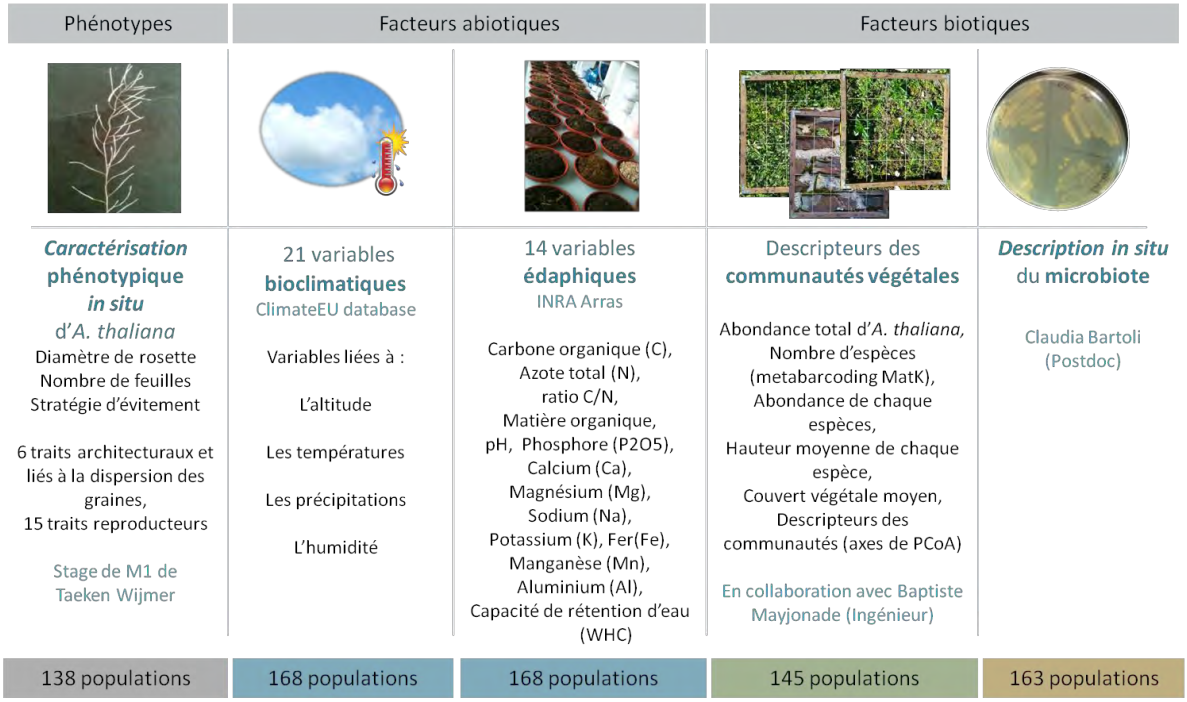


Figure 1: Caractérisations phénotypiques et écologiques de populations naturelles d'*A. thaliana* en Midi-Pyrénées.

Identification de 168 populations naturelles d'A. thaliana

Dans la région Midi-Pyrénées, le cycle de vie d'*A. thaliana* est le suivant : germination fin octobre – début novembre, passage de l'hiver sous forme de rosette, début de floraison fin février – fin mars et production de graines entre fin avril et mi-juin. Au printemps 2014, j'ai identifié avec Fabrice Roux 233 populations naturelles d'*A. thaliana* dans la région Midi-Pyrénées (Sud-Ouest de la France). En accord avec de précédents travaux effectués sur des populations naturelles d'*A. thaliana* dans la péninsule Ibérique (Picó 2012), une forte dynamique des populations a été observée entre le printemps 2014 et l'automne 2015, période à laquelle j'ai commencé les caractérisations phénotypiques et écologiques. En effet, 27,9% des populations identifiées au printemps 2014 ne contenaient plus (ou très peu) d'individus à l'automne 2014. Les principaux facteurs pouvant expliquer cette forte diminution de la taille de population sont une forte attaque par les herbivores, l'application d'herbicides, le fauchage, des travaux publiques ou bien encore un glissement de terrain. Ainsi, lors de ma thèse, j'ai travaillé avec 168 populations (**Figure 2**). Ces populations sont en moyenne distantes de 100.7 km (min = 0 km, max=265.2 km, médiane = 93.7 km), réparties dans tous les départements de la région, sauf dans le département du Lot dont le sol est très calcaire, c'est-à-dire un sol peu propice à la croissance d'*A. thaliana* qui est une espèce calcifuge. Les 168 populations peuvent être classées selon 4 grands types d'habitat (mur, sol nu, pelouse et prairie) avec des degrés de perturbations et d'interactions avec les espèces végétales très différents (**Figure 2**).

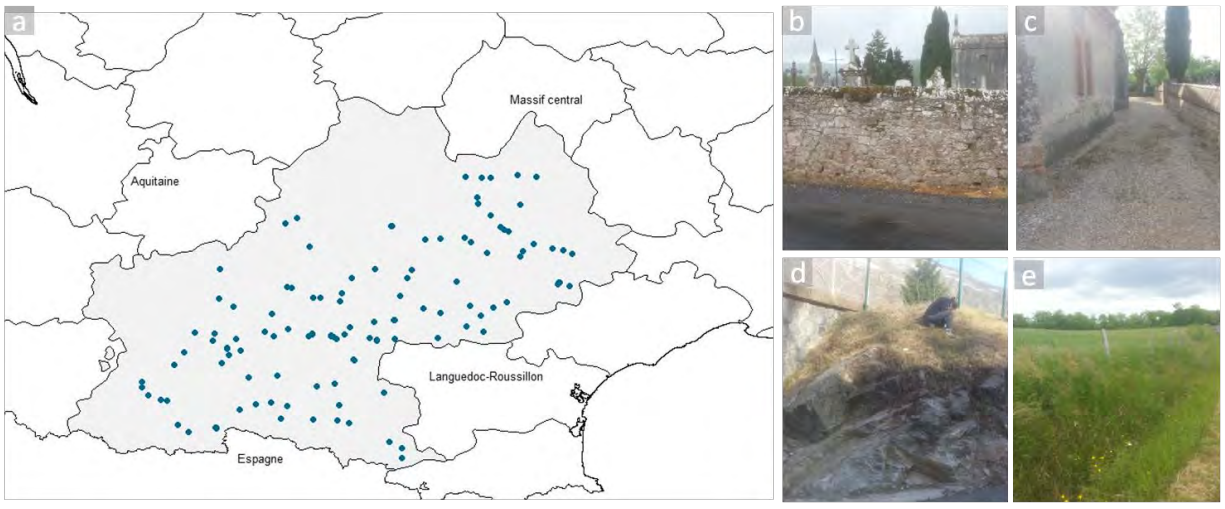


Figure 2: Cartographie des 168 populations en régions Midi-Pyrénées (a) réparties en 4 types d'habitats : mur (b), sol nu (c), pelouse (d) et prairie (e).

De par sa géographie entre les Pyrénées au Sud, le Massif central au Nord-Est, et une grande influence de l'océan Atlantique par l'ouest, la région Midi-Pyrénées est une région où trois climats différents se rencontrent. En effet majoritairement sous influence océanique, la région Midi-Pyrénées est contrastée par (i) un courant montagnard au niveau du Massif Central au Nord-Est et au niveau des Pyrénées au Sud et (ii) un climat Méditerranéen provenant du Sud-Est (influence du vent d'Autan).

Caractérisation phénotypique

Pour identifier les variables écologiques pouvant agir comme pressions de sélection sur *A. thaliana* dans la région Midi-Pyrénées, la production totale de graines a été utilisée comme proxy de la valeur sélective (*fitness*). Chez *A. thaliana*, une même production de graines pouvant être atteinte par différentes combinaisons phénotypiques (Roux *et al.* 2016), 23 traits phénotypiques supplémentaires ont été mesurés pour caractériser ces stratégies phénotypiques (**Figure 3**). Ces traits sont liés à (i) l'acquisition des ressources (diamètre de la rosette), (ii) la dispersion des graines (hauteur maximale de la plante, nombre de branches…) (Wender *et al.* 2005), (iii) des stratégies d'allocation des ressources entre nombre de fruits et nombre de graines par fruit (Reboud *et al.* 2004), ou bien encore (iv) des stratégies d'allocation des ressources entre types de branche (Reboud *et al.* 2004). Lors d'un stage de M1 que j'ai co-encadré avec Fabrice Roux, Taeken Wijmer a phénotypé 21 des 24 traits phénotypiques sur une moyenne de dix plantes par population échantillonnées à la fin du printemps 2015, période correspondant à la fin du cycle de vie d'*A. thaliana*. Certaines populations ayant subi un fauchage précoce ou une application d'herbicides, seules 138 populations ont pu être phénotypées représentant un nombre total de 1 386 plantes. Cette caractérisation phénotypique a été complétée par des mesures de diamètre de la rosette et du nombre de feuilles de la rosette réalisées directement sur le terrain durant l'hiver 2015 (février-mars) dans les 138 populations. Le ratio 'hauteur maximal de la plante / diamètre de la rosette en hiver' a aussi été calculé permettant d'avoir une information sur les stratégies d'évitement de la compétition (Baron *et al.* 2015). Les 24 traits phénotypiques mesurés pour cette étude sont illustrés en **Figure 3**.

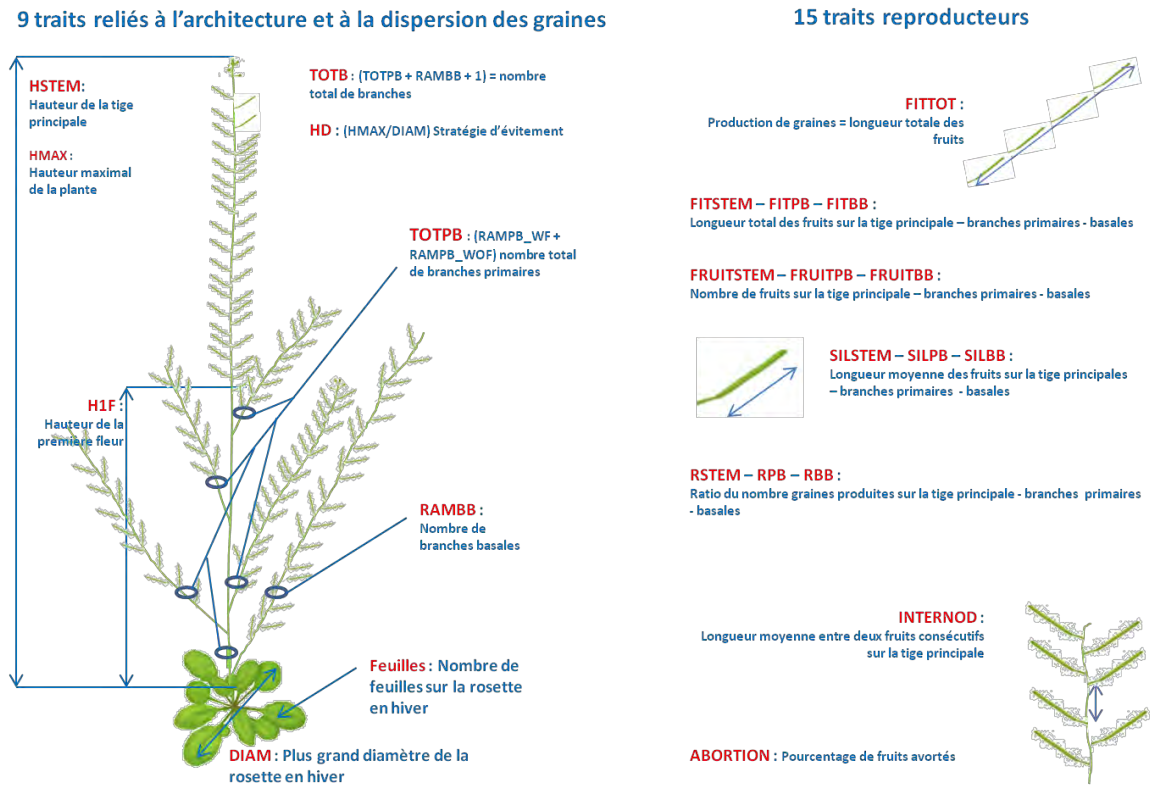


Figure 3: Représentation des 24 traits phénotypiques mesurés sur *A. thaliana* dans 138 populations de la région Midi-Pyrénées.

Caractérisation écologique

Lors de ma thèse, j'ai caractérisé les 168 populations à la fois pour des variables abiotiques comme le climat et le sol, mais aussi pour des variables biotiques comme les communautés végétales et les communautés microbiennes. L'étude des communautés microbiennes constituait le projet de recherche de Claudia Bartoli lors de son post-doctorat dans notre équipe.

1. Climat

Pour caractériser les 168 populations au niveau climatique, j'ai utilisé la base de données européennes ClimatEU (ClimateEU, v4.63 software package disponible sur http://tinyurl.com/ClimateEU) qui, contrairement à la base de données WorldClim (*http://www.worldclim.org/*), m'a permis d'avoir accès à des données annuelles plus

récentes (jusqu'à 2013), tout en ayant une résolution spatiale fine (~1.25 arcmin, ~600 m). Ainsi, j'ai récupéré les données climatiques annuelles de 2003 à 2013 à partir desquelles j'ai calculé une moyenne pour chacune des populations. Par ailleurs, l'altitude de chacune des populations a été récupérée à partir des coordonnées GPS *via* le site internet *www.coordonnees-gps.fr*. Au final, j'ai utilisé un total de 21 variables climatiques liées à l'altitude, aux températures, aux précipitations et à l'humidité (**Tableau 1**; Frachon *et al.* soumis, chapitre 2).

Variable	Description	source (résolution)
altitude	Altitude (m)	www.coordonnees-gps.fr
MAT	Température moyenne annuelle (°C)	ClimateEU (~ 600m)
MWMT	Température moyenne du mois le plus chaud (°C)	ClimateEU (~ 600m)
MCMT	Température moyenne du mois le plus froid (°C)	ClimateEU (~ 600m)
TD	Différence de température entre MWMT et MCMT, ou continentalité (°C)	ClimateEU (~ 600m)
MAP	Précipitation moyenne annuelle (mm)	ClimateEU (~ 600m)
AHM	Indice 'température annuelle:humidité' (MAT+10)/(MAP/1000))	ClimateEU (~ 600m)
SHM	Indice 'température été:humidité été' ((MWMT)/(précipitation moyenne en été/1000))	ClimateEU (~ 600m)
DD<0	Nombre de jours inférieur à 0°C	ClimateEU (~ 600m)
DD>5	Nombre de jours supérieur à 5°C	ClimateEU (~ 600m)
DD<18	Nombre de jours inférieur à 18°C	ClimateEU (~ 600m)
DD>18	nombre de jours supérieur à 18°C	ClimateEU (~ 600m)
NFFD	Nombre de jours sans gel	ClimateEU (~ 600m)
Tave_wt	Température moyenne en hiver (Dec. (année précédente) - Fev. (°C)	ClimateEU (~ 600m)
Tave_sp	Température moyenne au printemps (Mars - Mai) (°C)	ClimateEU (~ 600m)
Tave_sm	Température moyenne en été (Juin - Août) (°C)	ClimateEU (~ 600m)
Tave_at	Température moyenne en automne (Sept. - Nov.) (°C)	ClimateEU (~ 600m)
PPT_wt	Précipitation en hiver (mm)	ClimateEU (~ 600m)
PPT_sp	Précipitation au printemps (mm)	ClimateEU (~ 600m)
PPT_sm	Précipitation en été (mm)	ClimateEU (~ 600m)
PPT_at	Précipitation en automne (mm)	ClimateEU (~ 600m)

Tableau 1 : Description des 21 variables climatiques utilisées pour caractériser les 168 populations de la région Midi-Pyrénées.

2. Sol

Deux échantillons de sol ont été prélevés dans chacune des 168 populations : le premier à l'automne 2014 et le second à la fin de l'hiver 2015. Le fait d'avoir échantillonné ces échantillons à quelques mois d'intervalle permet de prendre en compte une possible évolution des propriétés chimiques des sols entre les saisons. Ces échantillons ont été séchés à l'étuve (50°C pendant 6 à 10h avec une ventilation de 90%). Puis, des sous-échantillons ont été envoyés au laboratoire d'analyses des sols de l'INRA d'Arras

(http://www6.npc.inra.fr/las) afin que soient mesurées treize variables édaphiques (**Tableau 2**). J'ai également mesuré la capacité de rétention en eau du sol (WHC). Pour cela, j'ai tout d'abord pesé 2 échantillons de sol sec (P_{sec}) par population en remplissant de sol des pots de 7cm*7cm*9cm (*L***l***H*). Les pots ont ensuite été imbibés d'eau pendant 30 minutes avant d'être rapidement posés sur du papier absorbant afin de retirer l'excédent d'eau, pour être finalement pesés de nouveau (P_{humide}). La capacité de rétention en eau du sol a été calculée suivant l'équation suivante:

$$WHC = \frac{(P_{humide} - P_{sec})}{P_{sec}}$$

La moyenne de ces 14 variables entre mes deux échantillons (automne – hiver) a été utilisée pour mes analyses.

Variable	Description	source
C	Carbone organique (g/kg)	NF ISO 10694 and NF ISO 13878
N	Azote total (g/kg)	NF ISO 10694 and NF ISO 13878
ratio_CN	ratio carbone sur azote	NF ISO 10694 and NF ISO 13878
mo	Matière organique du sol (g/kg)	NF ISO 10694 and NF ISO 13878
pH	pH	NF ISO 10390
Phosphore	Concentration en phosphore (P2O5) (g/kg)	NF ISO 11263
Calcium	Concentration en calcium (cmol+ / kg)	méthode cobaltihexamine (ICP-AES/EAF)(NF ISO 23470)
Magnesium	Concentration en magnesium (cmol+ / kg)	méthode cobaltihexamine (ICP-AES/EAF)(NF ISO 23470)
Sodium	Concentration en sodium (cmol+ / kg)	méthode cobaltihexamine (ICP-AES/EAF)(NF ISO 23470)
Potassium	Concentration en potassium (cmol+ / kg)	méthode cobaltihexamine (ICP-AES/EAF)(NF ISO 23470)
Fer	Concentration en fer (cmol+ / kg)	méthode cobaltihexamine (ICP-AES/EAF)(NF ISO 23470)
Manganèse	Concentration en manganèse (cmol+ / kg)	méthode cobaltihexamine (ICP-AES/EAF)(NF ISO 23470)
Aluminium	Concentration en aluminium (cmol+ / kg)	méthode cobaltihexamine (ICP-AES/EAF)(NF ISO 23470)
WHC	Capactié de rétention en eau du sol (mL/g)	(poids sol humide - poids sol sec)/poids sol sec

Tableau 2: Description des 14 variables édaphiques utilisées pour caractériser les 168 populations de la région Midi-Pyrénées.

3. Communautés végétales

Au printemps 2015 (mai - juin), nous avons effectué une caractérisation des communautés végétales associées aux populations naturelles d'*A. thaliana*. Pour les même raisons qu'évoquées précédemment (fauchage précoce et application d'herbicide) pour la caractérisation phénotypique des populations, les communautés végétales associées à *A. thaliana* ont seulement pu être caractérisées pour 145 des 168 populations.

Pour caractériser les communautés végétales, deux quadrats de 25 carrés de 10*10cm ont été posés aléatoirement au sein de chaque population, tout en respectant la représentativité de la communauté végétale associée à *A. thaliana* (**Figure 4**).

Figure 4: Photographies de quadrats dans quatre habitats différents : (a) mur, Camarès (Aveyron), (b) sol nu, Auzeville (Haute-Garonne), (c) pelouse, Monferran-Savès (Gers) et (d) prairie, Villecomtal (Aveyron).

Dans chaque quadrat, plusieurs mesures ont été réalisées : (i) nombre d'espèces morphologiquement différentes (richesse spécifique), (ii) nombre d'individus par espèce (abondance totale), (iii) hauteur moyenne de la communauté végétale calculée à partir de la hauteur maximale moyenne de chaque espèce (estimée à partir de 3 individus) pondérée par l'abondance relative de chaque espèce, (iv) recouvrement total du quadrat (toutes espèces confondues), et (v) abondance totale d'*A. thaliana*. Dans chaque population, un échantillon de tissu a été prélevé sur le terrain pour chaque espèce identifiée comme

morphologiquement différente, puis congelé à -80°C en vue d'une extraction d'ADN pour identifier les espèces par une approche metabarcoding. En effet, bien qu'une identification basée sur des caractères morphologiques ait été envisagée dans un premier temps, une approche par metabarcoding a été adoptée car de nombreuses espèces co-existantes avec *A. thaliana* se trouvaient au stade de plantule ou n'avaient pas encore produit de fleurs ou de fruits (organe utilisé pour l'identification de certaines espèces), rendant laborieuse l'identification à partir de caractères morphologiques. L'identification des espèces par une approche metabarcoding a été effectuée à partir du marqueur chloroplastique *matK* qui permet une résolution taxonomique au niveau de l'espèce (Barthet & Hilu 2007, Yan *et al.* 2015). J'ai réalisée cette identification taxonomique en collaboration avec Baptiste Mayjonade, ingénieur d'études dans notre équipe (Frachon *et al.* en préparation, chapitre 2). Avec cette approche, 97% des 2 233 échantillons prélevés sur le terrain ont pu être classés en 244 OTUs (Operational Taxonomic Units) végétaux. Après avoir aligné les séquences *matK* obtenues sur les bases de données du NCBI, 84.6% de ces échantillons ont pu être identifiés au niveau de l'espèce. Des Analyses en Coordonnées Principales (PCoA) ont été réalisées sur l'abondance des 44 OTUs les plus prévalents (i.e. les OTUs présents dans au moins 11 populations) : les trois premiers axes ont alors été utilisés pour décrire la composition des communautés végétales (Frachon *et al.* en préparation, chapitre 2).

4. Communautés microbiennes

La caractérisation des communautés microbiennes a été effectuée lors du projet de recherche de Claudia Bartoli qui a effectué un post-doc de deux ans dans l'équipe. Deux types de communautés ont été caractérisés : les communautés bactériennes et les communautés fongiques.

Les communautés bactériennes (ou bien encore appelées le microbiote bactérien) ont été caractérisées par une approche de métagénomique basée sur le gène de ménage *gyrB* (Barret *et al.* 2015) qui, contrairement au marqueur génétique 16S largement étudié pour ce genre d'études, est en simple copie dans les génomes bactériens et permet une résolution taxonomique au niveau de l'espèce, voire de la sous-espèce (Barret *et al.* 2015). Grâce à cette résolution taxonomique et en nous appuyant sur une liste de 199 espèces

bactériennes phytopathogènes (Bull *et al.* 2010, Bull *et al.* 2012, Bull *et al.* 2014), nous avons pu identifier les OTUs bactériens potentiellement pathogène, nous permettant de décrire un pathobiote potentiel (Bartoli *et al.* en révision, Annexe 1). En collaboration avec Baptiste Mayjonade, les communautés fongiques (ou bien encore appelées le microbiote fongique) ont elles-aussi été caractérisées par une approche de métagénomique, mais cette fois-ci basée sur le marqueur génétique 18S (Bartoli *et al.* données non publiées).

Les communautés bactériennes et fongiques ont été caractérisées au niveau foliaire (rosette) et racinaire à la fin de l'automne 2014 (novembre-décembre) et à la fin de l'hiver 2015 (février-mars). Les plantes étant trop petites à la fin de l'automne dans de nombreuses populations, seulement 84 populations ont pu être échantillonnées à cette période. A la fin de l'hiver, l'échantillonnage s'est effectué dans 163 populations. Mon travail dans ce projet à consister à prélever avec Fabrice Roux les racines et les rosettes d'environ 336 plantes d'*A. thaliana* à la fin de l'automne 2014 et d'environ 636 plantes à la fin de l'hiver 2015.

Dans ce chapitre, j'ai utilisé les descripteurs des communautés microbiennes suivants qui ont estimés pour les 163 populations échantillonnées à la fin de l'hiver 2015:

- α-diversité (indice de Shannon) du microbiote bactérien, du pathobiote potentiel bactérien et du microbiote fongique.
- A partir d'Analyses en Coordonnées Principales (PCoA) réalisées individuellement sur les matrices d'abondance des OTUs les plus fréquents du microbiote bactérien, du pathobiote potentiel bactérien et du microbiote fongique, les compositions du microbiote bactérien, du pathobiote potentiel bactérien et du microbiote fongique ont été décrites selon les deux premiers axes de PCoA.

Manuscrit

“The putative selective agents acting on *Arabidopsis thaliana* depend on the type of habitat”

Léa Frachon, Claudia Bartoli, Baptiste Mayjonade, Johanna Schmitt, Sharon Strauss and Fabrice Roux

B. Manuscrit en préparation (titre provisoire): The putative selective agents acting on *Arabidopsis thaliana* depend on the type of habitat

Léa Frachon, Claudia Bartoli, Baptiste Mayjonade, Johanna Schmitt, Sharon Strauss & Fabrice Roux

NB : les analyses statistiques de type 'sparse Partial Least Square Regression' (sPLSR) (voir-ci-dessous) ont été effectuées lors d'un séjour de deux mois (juillet et août 2016) dans le Department of Ecology and Evolution de l'Université de Davis, en collaboration avec Johanna Schmitt et Sharon Strauss.

Introduction

Les changements globaux tels que le changement climatique, les changements d'utilisation des sols et l'arrivée d'espèces invasives ont d'ores et déjà un impact sur la biodiversité, et notamment sur la diversité des communautés végétales qui apparaissent plus vulnérables du fait d'une dispersion en moyenne plus limitée chez les espèces végétales que chez les espèces animales (Sala *et al.* 2000, Tylianakis *et al.* 2008). Ces variations des composantes écologiques imposent ainsi de nouvelles pressions de sélection agissant sur les espèces, qui n'auront d'autre choix que de s'adapter pour survivre. Comprendre comment les plantes peuvent s'adapter à des environnements altérés par les changements globaux est donc de première importance. Sur le court terme, les plantes peuvent s'acclimater à des changements de conditions environnementales *via* la plasticité phénotypique, en développant et exprimant des valeurs phénotypiques particulières en réponse à des conditions environnementales locales (Hansen *et al.* 2012). Une réponse sur le long terme

passe par la sélection génétique amenant à l'adaptation. La capacité des populations naturelles à répondre aux changements globaux dépendra notamment de leur niveau de diversité génétique. La variation génétique étant un indicateur primordial du potentiel adaptatif, il est donc essentiel d'identifier les bases génétiques de l'adaptation et d'en identifier les signatures de sélection associées (Bergelson & Roux 2010). Les bases génétiques identifiées permettront de mieux prédire le potentiel adaptatif d'une espèce (aussi bien à une échelle régionale qu'à une échelle locale) et d'affiner les modèles de variation temporelle de la biodiversité et du fonctionnement des écosystèmes (Schiffers *et al.* 2014).

L'identification des bases génétiques de l'adaptation et des signatures de sélections associées est actuellement abordée selon deux approches principales, dont l'utilisation est récemment devenue très populaire grâce à l'arrivée des technologies de Nouvelle Génération de Séquençage (NGS). La première approche est liée à la génomique des populations et repose sur une identification le long du génome de régions génomiques présentant des traces de sélection moléculaire. Cette approche est dite en aveugle car elle ne repose sur aucun *a priori* sur l'identité des gènes sous sélection. Cependant, il reste généralement difficile d'associer les gènes identifiés sous sélection à un trait phénotypique donné ainsi qu'à une pression de sélection potentielle. La seconde approche est liée à la génétique quantitative et consiste en un scan génomique permettant d'associer la variation phénotypique naturelle de traits supposés adaptatifs à des polymorphismes génétiques. Le développement récent de méthodes statistiques en Genome Wide Association (GWA) mapping permet à l'heure actuelle de cartographier finement des régions génomiques (Quantitative Trait Loci, QTLs) associées à la variation phénotypique naturelle (Bergelson & Roux 2010), et ainsi proposer des gènes candidats pour l'adaptation. Cependant, (i)

l'utilisation de cette approche reste encore l'apanage de quelques espèces modèles (Atwell *et al.* 2010, Bergelson & Roux 2010, Fournier-Level *et al.* 2011, MacKay *et al.* 2012), et (ii) l'identité des pressions de sélection agissant sur les traits phénotypiques supposés adaptatifs est souvent suggérée mais rarement testée (Vignieri *et al.* 2010, Brachi *et al.* 2012, Linnen *et al.* 2013). Une alternative dérivée de cette seconde approche repose sur l'identification des régions génomiques associées à la variation d'un facteur écologique (Hancock *et al.* 2011, Rellstab *et al.* 2015, Manel *et al.* 2016). Du fait de la disponibilité d'un nombre important de bases de données, cette approche a été principalement employée pour détecter des gènes associés à des variables climatiques (Bay *et al.* 2017).

Cependant, dans leurs milieux naturels, les espèces sont soumises simultanément à de nombreuses pressions de sélection abiotiques et biotiques (Roux & Bergelson 2016). Afin d'identifier les bases génétiques de l'adaptation et d'en étudier les signatures de sélection, il convient donc d'identifier les combinaisons de facteurs écologiques agissant simultanément comme pressions de sélection sur une espèce et de comparer leur importance relative. Par ailleurs, identifier le grain spatial de ces pressions de sélection (ce grain étant certainement dépendant de l'identité des pressions de sélection) permettrait de déterminer l'échelle géographique à laquelle nous devons nous placer pour identifier les bases génétiques de l'adaptation (Bergelson & Roux 2010).

La plante annuelle principalement autogame *Arabidopsis thaliana* est une espèce porte-drapeau pour identifier les bases génétiques de l'adaptation. Cette espèce apparait comme un choix judicieux pour des études en génomique écologique car elle est incontournable en génétique fonctionnelle et est aussi une espèce modèle en écologie évolutive (Mitchell-Olds & Schmitt 2006, Gaut 2012, Krämer 2015). On retrouve cette espèce

dans des milieux très contrastés que ce soit en termes de facteurs abiotiques ou biotiques (Jakob *et al.* 2002, Bodenhausen *et al.* 2013, Brachi *et al.* 2013, Agler *et al.* 2016, Roux & Bergelson 2016). Par ailleurs, *A. thaliana* présente une variation phénotypique naturelle importante pour de nombreux traits adaptatifs tels que la date de floraison, la dormance des graines, la résistance aux pathogènes, l'architecture et la dispersion des graines reliées à la hauteur de la plante ou le nombre de branches (Wender *et al.* 2005, Koornneef *et al.* 2004, Kronholm *et al.* 2012, Debieu *et al.* 2013, Huard-Chauveau *et al.* 2013, Karasov *et al .* 2014, Roux & Bergelson 2016). A une échelle mondiale, différentes approches ont permis de démontrer une adaptation d'*A. thaliana* au climat (Fournier-Level *et al.* 2011, Hancock *et al.* 2011, Lasky *et al.* 2012) et à l'herbivorie (Züst *et al.* 2012, Brachi *et al.* 2015). A une échelle géographique plus restreinte (i.e. au sein de 4 régions géographiques françaises ; Bretagne, Bourgogne, Languedoc et Nord-Pas-de-Calais), il a été suggéré que certains facteurs édaphiques et les interactions plante-plante pouvaient être des pressions de sélection agissant sur la phénologie aussi importantes que les facteurs climatiques (Brachi *et al.*2013). Cependant, trois limitations peuvent être mises en avant dans cette étude. Premièrement, seulement une dizaine de populations ont été échantillonnées au sein de chaque région, limitant ainsi la puissance statistique quant à l'identification des agents sélectifs potentiels. Deuxièmement, la description des facteurs biotiques s'est cantonnée à estimer un degré d'interaction interspécifique (i.e. degré de coexistence d'*A. thaliana* avec des herbacées ou des graminées) sans tenir compte de la diversité et de la composition des communautés végétales. Par ailleurs, malgré les nombreuses études sur le microbiote chez *A. thaliana* (Bulgarelli *et al.* 2012, Lundberg *et al.* 2012, Horton *et al.* 2014, Bai *et al.* 2015), l'avantage adaptatif conféré par le microbiote chez *A. thaliana* a été rarement testé (Haney *et al.* 2015), notamment en conditions écologiquement réalistes. Troisièmement, les mesures

phénotypiques se sont focalisées sur la phénologie et ont été effectuées en conditions contrôlées de serre, ce qui les rend donc peu représentatives du comportement des populations dans leur milieu naturel.

Le but de cette étude est d'identifier les facteurs écologiques susceptibles d'agir comme pressions de sélection sur *A. thaliana* à une échelle régionale et de comparer leur importance relative. Pour mener à bien ce projet, une étude d'association phénotype – écologie a été réalisée à partir de 138 populations naturelles d'*A. thaliana* localisées dans la région Midi-Pyrénées et caractérisées au niveau écologique à la fois pour des facteurs abiotiques (climat et sol) et pour des facteurs biotiques (diversité et composition des communautés végétales et des microbiotes bactériens et fongiques). Les 138 populations ont aussi été caractérisées pour le nombre total moyen de graines produites par une plante, la production de graines étant utilisée comme proxy de *fitness* chez *A. thaliana* (Roux *et al.* 2004). Différentes stratégies phénotypiques pouvant amener à la même production de graines (Roux *et al.* 2016), je me suis aussi intéressée à identifier des relations significatives entre des facteurs écologiques et les stratégies phénotypiques.

Matériel et Méthodes

Matériel végétal et caractérisation écologique

Cette étude est basée sur 138 populations naturelles distribuées dans la région Midi-Pyrénées et pouvant être classées en quatre grands types d'habitats : mur (n = 11), sol nu (n = 55), pelouse (n = 43) et prairie (n = 29) (**Figure 1**).

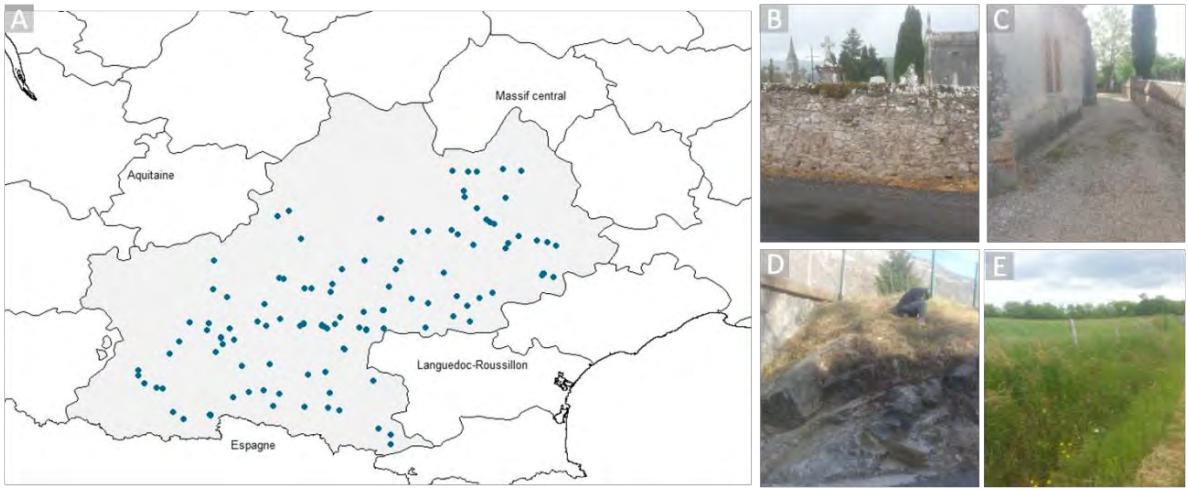


Figure 1: Distribution des 138 populations dans la région Midi-Pyrénées (a). Ces populations peuvent être classées en quatre grands types d'habitats : (b) mur : Camarès (Aveyron), (c) sol nu : Saint-Angel (Tarn), (d) pelouse : Merens-les-Vals (Ariège), et (e) prairie : Villecomtal (Aveyron).

Ces 138 populations ont été préalablement décrites pour 60 variables écologiques (**Table 1**): (i) 21 variables climatiques liés à l'altitude, à la température, à l'humidité et aux précipitations (Frachon *et al*. soumis, chapitre 2), (ii) 14 variables édaphiques (introduction chapitre 1), (iii) 7 descripteurs des communautés végétales (Frachon *et al.* en préparation, chapitre 2), (iv) 6 descripteurs des communautés fongiques (introduction chapitre 1) et 12 descripteurs des communautés microbiennes (Bartoli *et al.* en révision, Annexe 1).

Variable	Description	source (résolution)
altitude	Altitude (m)	www.coordonnees-gps.fr
MAT	Température moyenne annuelle (°C)	ClimateEU (~ 600m)
MWMT	Température moyenne du mois le plus chaud (°C)	ClimateEU (~ 600m)
MCMT	Température moyenne du mois le plus froid (°C)	ClimateEU (~ 600m)
TD	Différence de température entre MWMT et MCMT, ou continentalité (°C)	ClimateEU (~ 600m)
MAP	Précipitation moyenne annuelle (mm)	ClimateEU (~ 600m)
AHM	Indice 'température annuelle:humidité' (MAT+10)/(MAP/1000))	ClimateEU (~ 600m)
SHM	Indice 'température été:humidité été' ((MWMT)/(précipitation moyenne en été/1000))	ClimateEU (~ 600m)
DD<0	Nombre de jours inférieur à 0°C	ClimateEU (~ 600m)
DD>5	Nombre de jours supérieur à 5°C	ClimateEU (~ 600m)
DD<18	Nombre de jours inférieur à 18°C	ClimateEU (~ 600m)
DD>18	nombre de jours supérieur à 18°C	ClimateEU (~ 600m)
NFFD	Nombre de jours sans gel	ClimateEU (~ 600m)
Tave_wt	Température moyenne en hiver (Dec. (année précédente) - Fev. (°C)	ClimateEU (~ 600m)
Tave_sp	Température moyenne au printemps (Mars - Mai) (°C)	ClimateEU (~ 600m)
Tave_sm	Température moyenne en été (Juin - Août) (°C)	ClimateEU (~ 600m)
Tave_at	Température moyenne en automne (Sept. - Nov.) (°C)	ClimateEU (~ 600m)
PPT_wt	Précipitation en hiver (mm)	ClimateEU (~ 600m)
PPT_sp	Précipitation au printemps (mm)	ClimateEU (~ 600m)
PPT_sm	Précipitation en été (mm)	ClimateEU (~ 600m)
PPT_at	Précipitation en automne (mm)	ClimateEU (~ 600m)
C	Carbone organique (g/kg)	NF ISO 10694 and NF ISO 13878
N	Azote total (g/kg)	NF ISO 10694 and NF ISO 13878
ratio_CN	ratio carbone sur azote	NF ISO 10694 and NF ISO 13878
mo	Matière organique du sol (g/kg)	NF ISO 10694 and NF ISO 13878
pH	pH	NF ISO 10390
Phosphore	Concentration en phosphore (P2O5) (g/kg)	NF ISO 11263
Calcium	Concentration en calcium (cmol+ / kg)	méthode cobaltihexamine (ICP-AES/EAF)(NF ISO 23470)
Magnesium	Concentration en magnesium (cmol+ / kg)	méthode cobaltihexamine (ICP-AES/EAF)(NF ISO 23470)
Sodium	Concentration en sodium (cmol+ / kg)	méthode cobaltihexamine (ICP-AES/EAF)(NF ISO 23470)
Potassium	Concentration en potassium (cmol+ / kg)	méthode cobaltihexamine (ICP-AES/EAF)(NF ISO 23470)
Fer	Concentration en fer (cmol+ / kg)	méthode cobaltihexamine (ICP-AES/EAF)(NF ISO 23470)
Manganèse	Concentration en manganèse (cmol+ / kg)	méthode cobaltihexamine (ICP-AES/EAF)(NF ISO 23470)
Aluminium	Concentration en aluminium (cmol+ / kg)	méthode cobaltihexamine (ICP-AES/EAF)(NF ISO 23470)
WHC	Capactié de rétention en eau du sol (mL/g)	(poids sol humide - poids sol sec)/poids sol sec
Richesse	Richesse spécifique de la population	nombre d'OTU végétaux
Shannon_commV	α diversité de Shannon des communautés végétales	Somme(abondance relative * ln (abondance relative))
Hauteur	Hauteur moyenne des communautés	Exprimé en cm
Recouvrement	Couverture végétale des communautés	Exprimé en pourcentage de recouvrement
PCoA1_commV	Axe 1 de la PCoA basée sur l'abondance des 44 OTUs les plus prévalents	PCoA sur matrice de distances des abondances (méthode Jaccard)
PCoA2_commV	Axe 2 de la PCoA basée sur l'abondance des 44 OTUs les plus prévalents	PCoA sur matrice de distances des abondances (méthode Jaccard)
PCoA3_commV	Axe 3 de la PCoA basée sur l'abondance des 44 OTUs les plus prévalents	PCoA sur matrice de distances des abondances (méthode Jaccard)
Fungi_Shannon_Leaf	α diversité de Shannon des champignons microbiens au niveau de la rosette	Somme(abondance relative * ln (abondance relative))
Fungi_PCOA1_Leaf	Axe 1 de la PCoA basée sur la présence-absence des champignons microbiens au niveau de la rosette	PCoA sur matrice de distances de Bray-Curtis
Fungi_PCOA2_Leaf	Axe 2 de la PCoA basée sur la présence-absence des champignons microbiens au niveau de la rosette	PCoA sur matrice de distances de Bray-Curtis
Fungi_Shannon_Root	α diversité de Shannon des champignons microbiens au niveau de la racine	Somme(abondance relative * ln (abondance relative))
Fungi_PCOA1_Root	Axe 1 de la PCoA basée sur la présence-absence des champignons microbiens au niveau de la racine	PCoA sur matrice de distances de Bray-Curtis
Fungi_PCOA2_Root	Axe 2 de la PCoA basée sur la présence-absence des champignons microbiens au niveau de la racine	PCoA sur matrice de distances de Bray-Curtis
Bacteria_Shannon_Leaf	α diversité de Shannon des bactéries au niveau de la rosette	Somme(abondance relative * ln (abondance relative))
Bacteria_PCOA1_Leaf	Axe 1 de la PCoA basée sur la présence-absence des bactéries présentes au niveau de la rosette	PCoA sur matrice de distances de Bray-Curtis
Bacteria_PCOA2_Leaf	Axe 2 de la PCoA basée sur la présence-absence des bactéries présentes au niveau de la rosette	PCoA sur matrice de distances de Bray-Curtis
Bacteria_Shannon_Patho_Leaf	α diversité de Shannon des pathogènes au niveau de la rosette	Somme(abondance relative * ln (abondance relative))
Bacteria_PCOA1_Patho_Leaf	Axe 1 de la PCoA basée sur la présence-absence des pathogènes présentes au niveau de la rosette	PCoA sur matrice de distances de Bray-Curtis
Bacteria_PCOA2_Patho_Leaf	Axe 2 de la PCoA basée sur la présence-absence des pathogènes présentes au niveau de la rosette	PCoA sur matrice de distances de Bray-Curtis
Bacteria_Shannon_Root	α diversité de Shannon des bactéries au niveau de la racine	Somme(abondance relative * ln (abondance relative))
Bacteria_PCOA1_Root	Axe 1 de la PCoA basée sur la présence-absence des bactéries présentes au niveau de la racine	PCoA sur matrice de distances de Bray-Curtis
Bacteria_PCOA2_Root	Axe 2 de la PCoA basée sur la présence-absence des bactéries présentes au niveau de la racine	PCoA sur matrice de distances de Bray-Curtis
Bacteria_Shannon_Patho_Root	α diversité de Shannon despathogènes au niveau de la racine	Somme(abondance relative * ln (abondance relative))
Bacteria_PCOA1_Patho_Root	Axe 1 de la PCoA basée sur la présence-absence des pathogènes présentes au niveau de la racine	PCoA sur matrice de distances de Bray-Curtis
Bacteria_PCOA2_Patho_Root	Axe 2 de la PCoA basée sur la présence-absence des pathogènes présentes au niveau de la racine	PCoA sur matrice de distances de Bray-Curtis

Table 1: Description des 60 variables écologiques : climat (n=21), sol (n=14), communautés végétales (n=7), communautés fongiques (n=6) et communautés bactériennes (n=12).

Caractérisation phénotypique *in situ*

Vingt-quatre traits phénotypiques liés (i) à l'acquisition des ressources, l'architecture et la dispersion des graines (n = 9) et (ii) à la fécondité (n = 15) ont été mesurés **(Table S1, Figure 3 de l'introduction de ce chapitre**). A la fin de l'hiver 2015, le plus grand diamètre de la rosette ainsi que le nombre de feuilles de la rosette ont été mesurées entre 2 et 6 plantes (min = 2, médiane = 6, moyenne = 5.5, max = 6) dans chacune des 138 populations. Au printemps 2015, une dizaine de plantes ont été récoltées dans chacune des 138 populations (min = 2, moyenne = 9.9, médiane = 11, max = 13). Ainsi, un total de 1 386 plantes ont été phénotypés pour 21 traits phénotypiques décrits comme étant adaptatifs chez *A. thaliana* (Reboud *et al.* 2004, Wender *et al.* 2005, Roux *et al.* 2016). Le nombre total de graines produites par une plante étant un bon proxy de la fécondité chez les espèces annuelles principalement autogames comme *A. thaliana* et le nombre de graines dans un fruit étant fortement corrélée à la longueur du fruit (Roux *et al.* 2004), la production totale de graines d'une plante a donc été estimée en cumulant la longueur de tous ses fruits (FITTOT). Enfin, un estimateur de la stratégie d'évitement de la compétition (ratio HD) a été défini comme le ratio 'hauteur maximale de la plante / diamètre de la rosette à la fin de l'hiver'. Comme la hauteur maximale mesurée au printemps et le diamètre de la rosette mesuré à la fin de l'hiver ont été mesurées sur des plantes différentes, le ratio HD par population a été calculé à partir des moyennes estimées par population (Best Linear Unbiased Predictors, BLUPs, voir ci-dessous) de la hauteur maximale et du diamètre de la rosette.

Analyses statistiques

Afin d'explorer la variation naturelle de tous les traits phénotypiques (sauf le ratio HD, voir ci-dessous) entre les 138 populations naturelles d'*A. thaliana*, un modèle linéaire

mixte a été utilisé sous le logiciel SAS (PROC MIXED procédure sous SAS9.3, SAS Institute Inc., Cary, North Carolina, USA) :

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

Avec « Y » : la variable expliquée *i.e.* un des 23 traits phénotypiques étudiés, « μ_{trait} » : la moyenne phénotypique globale, l'effet aléatoire « population » correspond aux différences observées entre les 138 populations naturelles d'*A. thaliana*, et « ε » correspond à la variance résiduelle. Les variables phénotypiques ont été préalablement transformées *via* une transformation Box-Cox (Box & Cox 1964) pour satisfaire l'hypothèse de normalité. La part de variance phénotypique expliquée par l'effet « population » a été estimée en faisant tourner le modèle (1) avec la procédure VARCOMP sous le logiciel SAS. L'effet « population » étant considéré comme aléatoire, les moyennes par population ont été estimées par calcul des BLUPs avec le logiciel SAS (sans transformation des traits, PROC MIXED). Toutes les analyses suivantes ont été réalisées à partir de ces BLUPs.

Afin de tester si les traits phénotypiques étaient différents entre les 4 habitats, un modèle linéaire a été utilisé sous le logiciel SAS (PROC MIXED procédure sous SAS9.3, SAS Institute Inc., Cary, North Carolina, USA) :

$$Y_i = \mu_{trait} + habitat_i + \varepsilon_i \qquad (2)$$

Où « Y » est la variable expliquée *i.e.* les moyennes par population des 24 traits phénotypiques (i.e. BLUPs), « μ_{trait} » la moyenne phénotypique globale, l'effet fixe « habitat » correspond aux différences observées entre les quatre habitats (mur, sol nu, pelouse et prairie), et « ε » correspond à la variance résiduelle.

Relations entre la production totale de graines et les stratégies phénotypiques

Afin de visualiser la relation entre la production totale de graines (FITTOT) et les autres traits phénotypiques mesurés, une Analyse en Composantes Principales (ACP) normée (i.e. réalisée sur une matrice de corrélations) a été réalisée sur tous les traits phénotypiques à l'exception de la production totale de graines (package ade4, fonction dudi.pca(),*R* : Copyright © 2013. The R Foundation for Statistical Computing,version 0.99.891 ; Dray & Dufour 2007). Les BLUPs de ces 23 traits phénotypiques utilisés dans l'ACP ont été préalablement transformés par une transformation Box-Cox (Box & Cox 1964).

Estimation du degré d'homogénéité de la répartition spatiale des populations et détermination de l'échelle spatiale de variation des traits phénotypiques et des variables écologiques

Une analyse des « Principal Coordinates of Neighbour Matrices » (**PCNM**) a été réalisée à partir des coordonnées GPS des 138 populations (package vegan, fonction pcnm(), environnement *R*, Oksanen *et al.* 2016). Cette analyse permet de décomposer l'espace spatial en plusieurs variables ou composantes PCNM représentant différents grains de l'environnement (Borcard & Legendre 2002). Les premières composantes PCNM définissent un grain spatial large, alors que les dernières composantes PCNM définissent un grain plus fin (Ramette & Tiedje 2007, Borcard *et al.* 2004). Afin de déterminer à quelles échelles spatiales varient les 24 variables phénotypiques et les 60 variables écologiques, une régression linéaire multiple a été réalisée pour chaque trait ou variable (préalablement transformé *via* une transformation Box-Cox) sur toutes les composantes PCNM. Pour chaque trait phénotypique ou chaque variable écologique, une correction pour tests multiples a été

effectuée selon la méthode FDR (False Discovery Rate) (fonction p.adjust, méthode fdr dans l'environnement R).

Identification des variables écologiques reliées à la variation phénotypique d'*A. thaliana*

Afin d'analyser simultanément un grand nombre de variables écologiques et de traits phénotypiques, souvent très corrélés entre eux, j'ai adoptée la méthode de **sparse Partial Least Square Régression** (**sPLSR**) (Carrascal *et al.* 2009, Lê Cao *et al.* 2011 ; package mixOmics, fonction spls() dans l'environnement R). Cette analyse permet de créer de nouvelles composantes (ou variables latentes) maximisant la covariance entre des variables écologiques et des traits phénotypiques (Shariari *et al.* 2015, Mevik & Wehrens, 2007). Il est alors possible d'extraire pour les différentes composantes le poids de chaque variable (écologique et phénotypique), permettant ainsi de comparer les variables entre elles et de les classer suivant leur importance. Pour tous les traits phénotypiques et toutes les variables écologiques, des transformations Box-Cox (Box & Cox 1964) ont été réalisées à partir des BLUPs permettant d'obtenir une distribution normale des données, ceci afin d'améliorer la puissance des analyses de sPLSR (Wold *et al.* 2001). Cette transformation Box-Cox attribuant la même valeur à toutes les populations pour PPT_sp et DD_18 (deux variables climatiques, **Table 1**), ces deux variables ont donc été utilisées sans transformation. Les traits phénotypiques ainsi que les variables écologiques n'ont pas tous la même unité de mesure. Ainsi, tous les traits phénotypiques et toutes les variables écologiques ont été centrés-réduits avant de réaliser les analyses de sPLSR.

Pour déterminer le nombre de composantes nécessaires pour expliquer au mieux la covariance entre traits phénotypiques et variables écologiques, une validation par une méthode du Lasso (loo) a été réalisée. Le *Root Mean Square Error of Prediction* (RMSEP) a

alors été calculé et représenté graphiquement permettant de valider le nombre de composantes à conserver (Maestre 2004, Lê Cao *et al.* 2008). Ce choix visuel reste assez subjectif, mais pour le moment aucune autre méthode robuste ne semble avoir été développée (Mevik & Wehrens 2007). Cependant, il est à noter que le choix du nombre de composantes n'influe pas sur leur calcul dans l'analyse sPLSR. Dans cette étude, nous avons choisi de garder 8 variables écologiques et 8 traits phénotypiques dans le calcul de leur poids (loadings) sur la première composante. Ce choix n'affecte pas la manière de calculer les composantes, mais seulement le poids attribué à chaque trait phénotypique ou à chaque variable écologique. Les analyses sPLSR ont été réalisées en considérant (i) la production totale de graines et les 60 variables écologiques ou (ii) les 23 traits sous-jacents à la production total de graines et les 60 variables écologiques. Par ailleurs, les analyses PLSR ont été réalisées à la fois sur l'ensemble des populations et sur les populations de chacun des quatre habitats.

Résultats

Distribution spatiale des 138 populations étudiées

Afin d'estimer le degré d'homogénéité de répartition spatiale des populations, une analyse PCNM a été réalisée sur les coordonnées GPS des 138 populations. Nous avons trouvé 74 PCNMs, suggérant une distribution spatiale très homogène des populations (**Figure S1**). Par la suite, ces PCNMs ont été arbitrairement divisées en 3 catégories : une échelle large (PCNMs de 1 à 24), une échelle intermédiaire (PCNMS de 25 à 50) et une échelle fine (PCNMS de 51 à 74).

Variation naturelle des traits phénotypiques

A l'exception du ratio du nombre de graines produites sur les branches basales (RBB), les traits phénotypiques étaient significativement très différents entre les 138 populations, avec un effet « population » expliquant en moyenne 34.1% de la variance phénotypique (médiane = 33.2%, min = 12.4%, max = 53.6%) (**Table S2**). La production totale de graines d'une plante (estimée par la longueur totale des fruits, FITTOT) peut varier d'un facteur de 18 entre les 138 populations (moyenne = 327.3 mm, min = 89.6 mm, max = 1615.8 mm) (**Figure 2A**). Plus de 44% de la variance de la production totale de graines est expliquée par un effet « population » (**Table S2**).

Une variation significative entre les 4 habitats a uniquement été observée pour les deux traits d'acquisition de ressources (diamètre et nombre de feuilles de la rosette) ainsi que pour la stratégie d'évitement de la compétition (HD) (**Table S3**). Les plantes ont un diamètre et un nombre de feuilles de la rosette significativement plus petits dans l'habitat 'prairie' que dans les habitats 'sol nu' et 'pelouse' (**Figure S2**). La stratégie d'évitement de la compétition est significativement plus prononcée dans l'habitat 'prairie' que dans les habitats 'sol nu' et 'pelouse' (**Figure S2**). Pour ces trois traits, les plantes de l'habitat 'mur' ont un comportement intermédiaire entre l'habitat 'prairie' et les habitats 'sol nu' et 'pelouse' (**Figure S2**).

Afin de visualiser les stratégies phénotypiques (i.e. combinaisons entre les traits phénotypiques) sous-jacentes à la production totale de graines, une ACP a été effectuée sur tous les traits phénotypiques à l'exception de FITTOT. Les deux premiers axes de l'ACP expliquent 52.66 % de la variance phénotypique observée. Alors que le premier axe d'ACP (40.48%) est notamment corrélé aux traits de fécondité, d'architecture et de dispersion des

graines mesurés sur la tige principale et les branches primaires, le deuxième axe d'ACP (12.18%) est plutôt corrélé aux deux traits d'acquisition des ressources (diamètre et nombre de feuilles de la rosette) et à des traits de fécondité et d'architecture mesurés sur les branches primaires (**Figure 2B**).

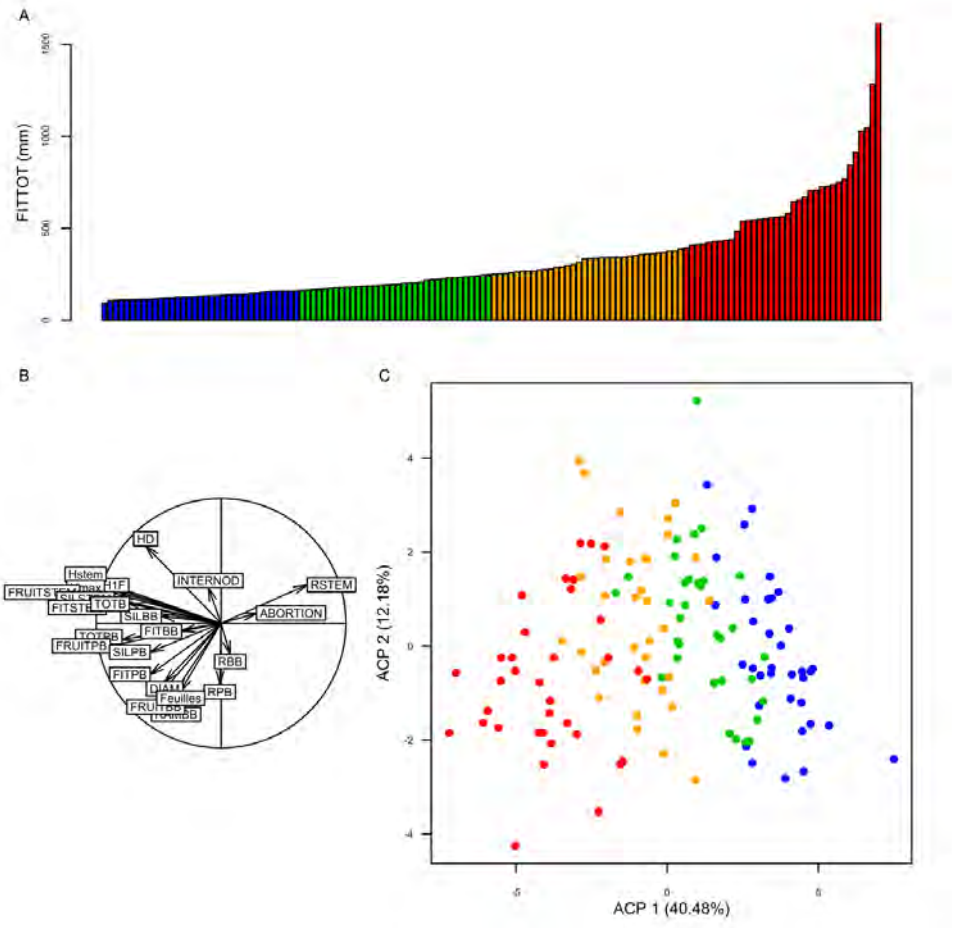


Figure 2 : Variation de la production totale de graines par plante et des stratégies phénotypiques sous-jacentes. (A) Distribution de la production totale de graines par plante (FITTOT) dans les 138 populations. Les couleurs représentent les quartiles de production totale de graines par plante. (B) Cercle de corrélations entre tous les traits phénotypiques à l'exception de FITTTOT. Les axes 1 et 2 expliquent respectivement 40.48% et 12.18% de la variance phénotypique observée. (C) Position des 138 populations naturelles dans l'espace phénotypique déterminé par les axes 1 et 2 de l'ACP. Les couleurs représentent les quatre quartiles de production totale de graines par plante (voir Figure 2A).

Lorsque l'on superpose les valeurs de production totale de graines (FITTOT) sur l'espace phénotypique déterminé par les deux premiers axes de l'ACP, seule la variation de

l'axe 1 est significativement reliée à la variation naturelle de la production de graines (coefficient de corrélation de Pearson = -0.96, *P* = <2.2.10^{-}16 ; axe 2 : coefficient de corrélation de Pearson = -0.091, *P* = 0.2901). La variation le long de l'axe 2 est quant à elle reliée à la variation des stratégies phénotypiques pour une même gamme de variation de la production totale de graines (**Figure 2C**). Ainsi, une même production totale de graines peut être atteinte aussi bien par une stratégie d'évitement de la compétition (HD) que par une stratégie combinant acquisition des ressources et allocation des ressources au niveau des branches basales.

Comme illustré pour la production totale de graines (FITTOT) (**Figure 3A**), aucun patron clair de distribution spatiale de la variation phénotypique n'a été identifié (**Figure S3**), suggérant une variation phénotypique non prédictible à une échelle très fine.

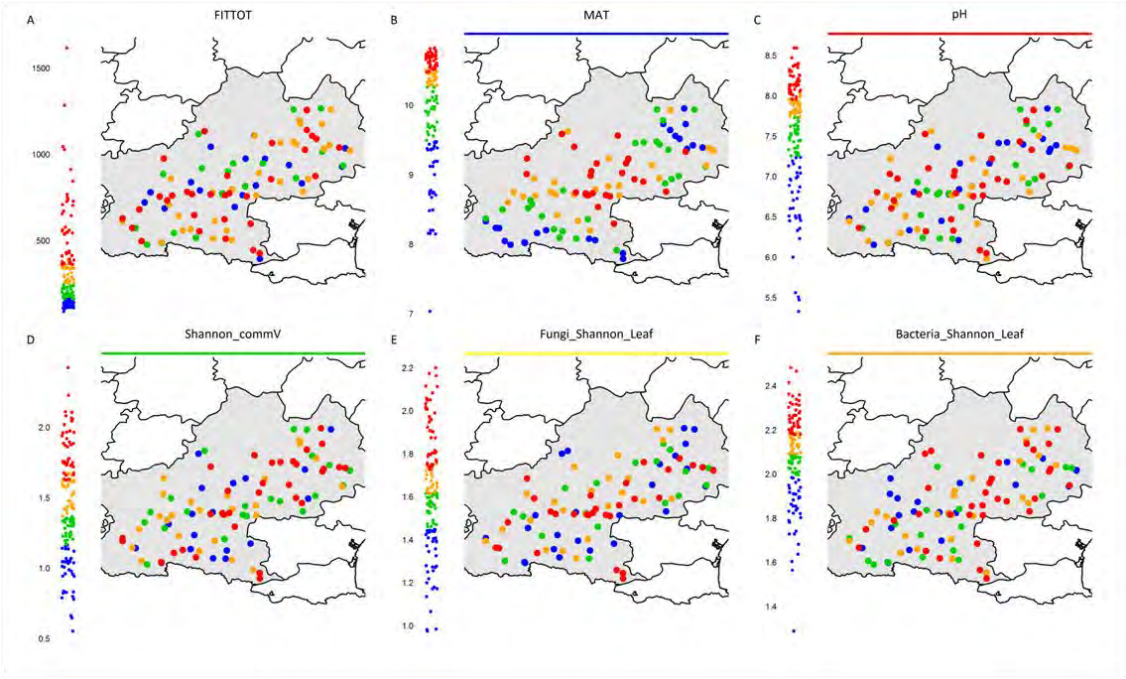


Figure 3: Cartographie (A) de la production totale de graines (FITTOT) et de différents types de variables écologiques : (B) climatique (MAT), (C) édaphique (pH), (D) descripteurs des communautés végétales (α-diversité - Shannon), (E) descripteurs des communautés fongiques (α-diversité - Shannon) et (F) descripteurs des communautés bactériennes (α-diversité - Shannon). La distribution des variables est représentée à gauche de chaque carte. Les points bleus, verts, jaunes et rouges représentent ¼ des populations déterminées par les quartiles.

Echelle(s) spatiale(s) de la variation écologique

Une régression linéaire multiple de chaque variable écologique sur les 74 PCNMs a permis d'identifier les échelles spatiales de variation des facteurs écologiques. Les échelles spatiales auxquelles varient les variables écologiques sont très différentes selon les catégories de variable (**Figure 4**). Le climat varie fortement et principalement à des échelles géographiques larges (**Figure 4**). Par exemple, alors qu'une faible température moyenne annuelle (MAT, **Figure 3B**) est observée dans les montagnes des Pyrénées (sud-est) et de l'Aveyron (nord-est), la température annuelle moyenne est bien plus élevée au centre de la région, notamment autour de la région Toulousaine. En revanche, le degré de significativité des associations entre les autres variables écologiques (à l'exception de la composition du microbiote bactérien dans les racines) et les PCNMs est en général plus faible (**Figure 4**). Par ailleurs, les associations significatives 'écologie-PCNMs' concernent un nombre de PCNMs beaucoup moins important que celui observé pour les variables climatiques. Bien que les variables édaphiques, les descripteurs des communautés végétales et les descripteurs du microbiote fongique soient préférentiellement associés à des échelles géographiques larges (**Figure 4**), aucun pattern géographique n'est évident (**Figures 3C-E**). Les descripteurs du microbiote bactérien varient à de nombreuses échelles géographiques (**Figure 4**), mais là encore aucun pattern géographique n'est évident (**Figure 3F**).

Figure 4 : Régression multiple linéaire des 60 variables écologiques sur les 74 PCNMs. Pour chaque variable écologique, une correction pour tests multiples a été effectuée selon la méthode FDR. La significativité des coefficients de régression entre un trait phénotypique et une PCNM est illustrée par un gradient de bleu (blanc : non significatif, bleu clair * $P < 0.05$, bleu intermédiaire ** $P < 0.01$, bleu foncé *** $P < 0.001$). Les 74 PCNMs ont été arbitrairement divisées en 3 catégories séparées dans le tableau par des lignes verticales: une échelle large (PCNMs de 1 à 24), une échelle intermédiaire (PCNMS de 25 à 50) et une échelle fine (PCNMS de 51 à 74). Chaque variable écologique est décrite dans le **Tableau 1**.

Agents sélectifs potentiels agissant simultanément sur *A. thaliana*

Afin d'identifier les agents sélectifs agissant potentiellement sur *A. thaliana*, les jeux de données phénotypiques et écologiques ont été analysés par la méthode de sPLSR. Ce type d'analyses permet de créer de nouvelles composantes maximisant la covariance entre ces deux jeux de données, et d'en sortir les principaux traits phénotypiques et les principales variables écologiques sous-jacents à ces composantes.

Que l'on considère l'ensemble des populations ou bien chacun des 4 habitats séparément, la variance de la production totale de graines (FITTOT) est reliée en moyenne à 12% de la variance écologique correspondant principalement à des descripteurs des communautés végétales, notamment la richesse spécifique et la hauteur moyenne de la communauté végétale (**Figure 5A**). Une variable édaphique et deux descripteurs des communautés microbiennes sont aussi reliées à la variance de la production totale de graines, mais uniquement dans certains habitats (sodium pour les habitats 'mur' et 'sol nu', composition du microbiote fongique dans les feuilles pour l'habitat 'mur', composition du pathobiote bactérien pour l'habitat 'prairie') (**Figure 5A**). Il est à noter qu'aucune variable climatique n'a été identifiée comme associée à la variance de la production totale de graines.

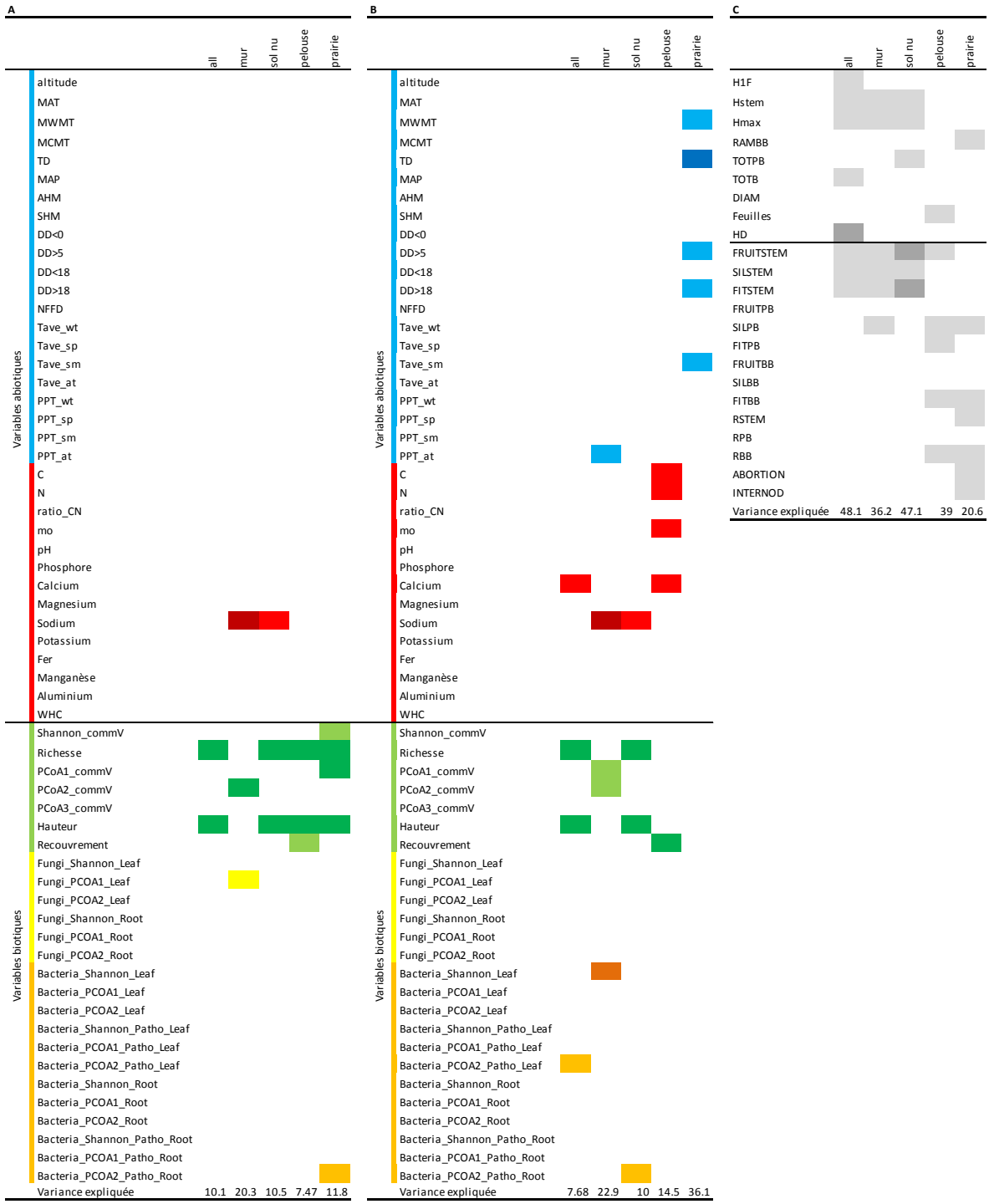


Figure 5: Résultats de la régression partielle des moindres carrées (sPLSR) sur l'ensemble des populations et pour chacun des 4 habitats. (A) Identification des combinaisons de variables écologiques associées à la variance de la production totale de graines. (B) et (C) Covariance entre les combinaisons de variables écologiques et les combinaisons de traits phénotypiques (à l'exception de FITTOT). Les poids (loadings) des variables écologiques et des traits phénotypiques pour la première composante sont indiqués avec une couleur claire lorsqu'ils sont supérieurs à 0.2 et avec une couleur foncée lorsqu'ils sont supérieur à 0.5. La variance phénotypique ou écologique associée à la première composante est indiquée en bas de chaque tableau.

A l'opposé de ce qui a été observé pour la production totale de graines, les résultats issus de la sPLSR sur les stratégies phénotypiques sont très contrastés entre les 4 habitats (**Figures 5B et 5C**). Pour les habitats 'mur' et 'sol nu', une combinaison de variables écologiques appartenant aux 4 grandes catégories écologiques utilisées dans cette étude (climat, sol, communautés végétales et communautés microbiennes) covarie avec une combinaison de traits phénotypiques principalement associés à la tige principale (i.e. hauteur maximale, nombre de fruits et longueur moyenne d'un fruit). Pour les habitats 'pelouse' et 'prairie', les combinaisons de traits phénotypiques concernent principalement les branches primaires (longueur des fruits) et les branches basales (nombre de branches, production de graines et allocation des ressources) et covarient avec une combinaison de variables écologiques appartenant principalement à une seule grande catégorie écologique, c'es-à-dire le sol pour l'habitat 'pelouse' et le climat pour l'habitat 'prairie'.

Discussion

Une des priorités pour comprendre et prédire les trajectoires évolutives des espèces végétales face aux changements globaux passe par l'identification des bases génétiques de l'adaptation (Bergelson & Roux 2010), impliquant en amont de connaître l'identité des agents sélectifs et leur importance relative.

Des grains superposés de l'environnement amènent à une mosaïque écologique complexe à une petite échelle géographique

Définir à quelle échelle spatiale travailler pour identifier les bases génétiques de l'adaptation est essentiel mais reste complexe car cela dépend principalement de l'échelle spatiale de variation des agents sélectifs (Bergelson & Roux 2010, Brachi *et al.* 2013). En

effet, comme observé dans cette étude, les individus sont soumis de manière simultanée à de nombreuses pressions de sélection dont les échelles spatiales de variation peuvent être très contrastées. Alors que le climat varie selon différentes échelles spatiales dans la région Midi-Pyrénées, aucun pattern clair de distribution géographique n'a été mis en évidence pour la majorité des autres variables écologiques. Ces observations suggèrent des conditions écologiques très contrastées entre les populations (notamment pour les facteurs biotiques), sur de courtes distances géographiques.

Cette superposition de pressions de sélection ayant des grains différents peut ainsi entraîner un conflit quant à la stratégie phénotypique à adopter par les plantes dans certaines populations. En effet, en présence d'un agent sélectif avec un grain spatial large, la même stratégie phénotypique pourra être sélectionnée dans deux populations A et B géographiquement proches. Par contre, en présence d'un agent sélectif avec un grain spatial fin, la même stratégie phénotypique pourra être sélectionnée dans la population A mais une autre stratégie phénotypique sera favorisée dans la population B. A notre connaissance, l'impact de différents grains spatiaux superposés sur l'architecture génétique de l'adaptation a été rarement (pour ne pas dire jamais) modélisé mais nécessiterait d'être abordé étant donné la mosaïque écologique complexe observée pour une espèce végétale comme *A. thaliana*.

Les communautés végétales agissent potentiellement comme les principaux agents sélectifs sur *A. thaliana*

Bien qu'*A. thaliana* soit décrite comme une espèce pionnière, les principales variables écologiques associées à la variation de la production totale de graines dans la région Midi-Pyrénées correspondent à des descripteurs des communautés végétales,

quelque soit le type d'habitat considéré. Ces résultats sont en accord avec de précédentes études réalisées au sein de l'équipe où (i) le degré d'interaction interspécifique avec des herbacées ou des graminées a été suggéré comme une pression de sélection agissant sur la phénologie aussi importante que le climat au sein de 4 autres régions géographiques françaises (Brachi *et al.* 2013) et (ii) la production moyenne de graines d'une famille RIL d'*A. thaliana* était négativement corrélée à l'intensité de la compétition avec le pâturin annuel *Poa annua* (Brachi *et al.* 2012).

Dans notre étude, les deux principaux descripteurs des communautés végétales associés à la variation de la production totale de graines sont la richesse spécifique et la hauteur moyenne de la communauté. Alors que la hauteur moyenne de la communauté végétale peut être indicatrice d'un niveau de compétition pour la lumière, la richesse spécifique peut être indicatrice du nombre de niches écologiques occupées dans une communauté végétale et donc, entre autres, d'un niveau de compétition pour les nutriments. Pour les habitats 'mur' et 'prairie', la composition des communautés végétales a aussi été identifiée comme associée à la variation de la production totale de graines, suggérant des phénomènes de spécialisation d'*A. thaliana* à des combinaisons d'espèces végétales. Cette hypothèse fait écho à des travaux où des phénomènes de spécialisation biotique ont été identifiés au sein d'une population locale d'*A. thaliana* localisée en Bourgogne. En effet, certaines accessions produisaient plus de graines en présence qu'en absence d'une espèce compétitrice et l'identité des accessions spécialisées étaient spécifique de l'identité de l'espèce compétitrice (Baron *et al.* 2015).

L'identité des variables écologiques agissant potentiellement comme pressions de sélection sur les stratégies reproductives est dépendante de l'habitat

Etant une espèce avec un régime de reproduction fortement autogame (~98% en moyenne) (Platt *et al.* 2010), le nombre de fruits ou la production totale de graines ont été largement utilisés comme proxy de *fitness* pour étudier l'adaptation locale chez *A. thaliana* dans des conditions écologiquement réalistes (Tian *et al.* 2003, Weinig *et al.* 2003a, Weinig *et al.* 2003b, Weinig *et al.* 2003c, Korves *et al.* 2007, Huang *et al.* 2010, Fournier-Level *et al.* 2011, Hancock *et al.* 2011, Agren *et al.* 2013, Karasov *et al.* 2014, Wilczek *et al.* 2014, Baron *et al.* 2015, Brachi *et al.* 2015, Kerwin *et al.* 2015, Roux *et al.* 2016, Hu *et al.* 2017). Dans notre étude, la même quantité de graines produites par une plante peut être associée à différentes combinaisons phénotypiques, reflétant principalement différentes stratégies d'allocation des ressources entre les traits de fécondité (nombre de fruits *vs* nombre de graines par fruit) (Reboud *et al.* 2004) ainsi que différentes stratégies de dispersion des graines (Wender *et al.* 2005) ou d'évitement de la compétition (Baron *et al.* 2015). Ces résultats renforcent la nécessité de considérer les stratégies reproductives dans l'étude de l'adaptation chez *A. thaliana* (Roux *et al.* 2016).

De manière intéressante, contrairement à la production totale de graines, l'identité des agents sélectifs potentiels est très dépendante de l'habitat considéré, et les combinaisons de traits phénotypiques le sont aussi. Ainsi, dans des habitats régulièrement perturbés par des interventions humaines (i.e. mur et sol nu), la stratégie pour les plantes d'*A. thaliana* serait d'allouer principalement les ressources dans les traits de fécondité associés à la tige principale (i.e. dans le méristème apical dominant). Cette stratégie permettrait d'assurer une production de descendants en présence de perturbations variées

(climat, sol, communautés végétales et microbiote ou bien encore fauchage, application d'herbicide) et non prédictibles d'une année sur l'autre. La variation de la quantité de ressources allouées à la tige principale entre les populations serait ainsi reliée à la différence du degré de perturbation au sein des habitats 'mur' et 'sol nu'. Par contre, dans les habitats plus stables comme les pelouses et les prairies, la majorité des agents sélectifs potentiels ne concernent qu'une seule grande catégorie écologique (sol ou climat) et les traits affectés concernent principalement des traits de fécondité associés aux branches primaires et basales (i.e. allocation des ressources dans les méristèmes secondaires). Pour l'instant, il nous est difficile d'émettre des hypothèses quant aux liens directs (ou indirects) qui puissent exister entre la variation de la quantité de ressources allouées aux branches primaires et basales et la variation du sol ou du climat. Et de manière générale, toutes les relations significatives entre variation phénotypique et variation écologique que nous avons pu identifier dans cette étude ne sont que suggestives du rôle des facteurs écologiques identifiés sur la sélection des traits phénotypiques étudiés. Des expériences en conditions contrôlées où l'agent sélectif potentiel peut être manipulé sont clairement nécessaires pour valider les associations phénotype-écologie identifiées dans cette étude. Néanmoins, le fait que l'identité des agents sélectifs potentiels soit dépendante de l'habitat considéré suggère que l'identification des bases génétiques de l'adaptation doit être effectuée au sein de chacun de ces habitats.

Nécessité de prendre en compte la dynamique des facteurs écologiques au cours du cycle de vie d'*A. thaliana* ?

Bien que plus de 60 variables écologiques aient été mesurées sur les 138 populations, y compris des variables biotiques innovantes comme le microbiote dont la caractérisation

est dorénavant rendue accessible par les technologies NGS, un large nombre de facteurs écologiques n'ont pas été mesurés. C'est notamment le cas de la communauté d'herbivores dont l'impact sur la production totale de graines peut être très importante chez *A. thaliana* (observations personnelles, Brachi *et al.* 2015). Parmi les autres facteurs écologiques, nous pouvons citer les communautés d'oomycètes pathogènes (dont la caractérisation par une approche métagénomique basée sur le marqueur génétique ITS reste encore difficile ; Agler *et al.* 2016), les variations micro-climatiques (à l'opposé des moyennes de données climatiques sur 10 ans utilisées dans cette étude) ou bien encore la topographie du milieu, le vent, les micro-éléments et la micro-faune du sol. Par ailleurs, il ne faut pas oublier que la majorité des variables biotiques mesurées dans cette étude peuvent être très variables au cours du cycle de vie d'*A. thaliana*, comme cela a été observé pour les microbiotes bactériens et fongiques (Bartoli *et al.* en révision, Annexe 1). Dans cette étude, nous avons considéré des descripteurs des communautés biotiques mesurées uniquement au printemps. Il serait intéressant de tester (i) si des descripteurs des communautés biotiques mesurées à d'autres saisons sont eux-aussi reliés à la variance de la production totale de graines et (ii) si l'effet des descripteurs mesurés au début du cycle de vie d'*A. thaliana* conditionnent l'effet des descripteurs mesurés à la fin du cycle de vie d'*A. thaliana*.

Références

Agler MT, Ruhe J, Kroll S, Morhenn C, Kim ST, Weigel D, Kemen EM (2016) Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biology* **14**: e1002352

Agren J, Oakley CG, McKay JK, Lovell JT, Schemske DW (2013) Genetic mapping of adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 21077–21082

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JDG, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**: 627–631

Bai Y, Müller DB, Srinivas G, Garrido-Oter R, Potthoff E, Rott M, Dombrowski N, Münch PC, Spaepen S, Remus-Emsermann M, Hüttel B, McHardy AC, Vorholt JA, Schulze-Lefert P (2015) Functional overlap of the Arabidopsis leaf and root microbiota. *Nature* **528**: 364-369

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. *Functional Ecology* **29**: 1361–1370

Bartoli C, Frachon L, Barret M, Rigal M, Zanchetta C, Bouchez O, Carrere S, Roux F (2017) In situ relationships between microbiota and potential pathobiota in *Arabidopsis thaliana*. *eLife* **En révision**

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. *The American naturalist* **189**: 463-473

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

Bodenhausen N, Horton MW, Bergelson J (2013) Bacterial communities associated with the leaves and the roots of *Arabidopsis thaliana*. *PLoS one* **8**: e56329

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* **153**: 51–68

Borcard D, Legendre P, Avois-Jacquet C, Tuomisto H (2004) Dissecting the spatial structure of ecological data. *Ecology* **85**: 1826–1832

Box GEP, Cox DR (1964) An analysis of transformations. *Journal of the Royal Statistical Society* **26**: 211-252

Brachi B, Aimé C, Glorieux C, Cuguen J, Roux F (2012) Adaptive value of phenological traits in stressful environments: predictions based on seed production and laboratory natural selection. *PLoS one* **7**: e32069

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*. *Molecular ecology* **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*. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 4032–4037

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

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

Debieu M, Tang C, Stich B, Sikosek T, Effgen S, Josephs E, Schmitt J, Nordborg M, Koornneef M, de Meaux J (2013) Co-variation between seed dormancy, growth rate and flowering time changes with latitude in *Arabidopsis thaliana*. *PLoS one* **8**: e61075

Dray S, Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* **22**: 1-20

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**: 86-89

Frachon L, Mayjonade B, Bartoli C, Hautekeete , Roux F (**En préparation**) Adaptation to plant communities across the genome of *Arabidopsis thaliana*.

Frachon L, Bartoli C, Carrère S, Bouchez O, Chaubet A, Gautier M, Roby D, Roux F (2017) A genomic map of adaptation to local climate in *Arabidopsis thaliana*. *New Phytologist* (**Soumis)**

Gaut B (2012) *Arabidopsis thaliana* as a model for the genetics of local adaptation. *Nature Genetics* **44**: 115-116

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

Haney CH, Samuel BS, Bush J, Ausubel FM (2015) Associations with rhizosphere bacteria can confer an adaptive advantage to plants. *Nature plants* **1**: 15051

Hansen MM, Olivieri I, Waller DM, Nielsen EE, THE GeM WORKING GROUP (2012) Monitoring adaptive genetic responses to environmental change. *Molecular ecology* **21**: 1311–1329

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. *Nature communications* **5**: 5320

Hu J, Lei L, de Meaux J (2017) Temporal fitness fluctuations in experimental *Arabidopsis thaliana* populations. *BioRxiv*

Huang X, Schmitt J, Dorn L, Griffith C, Effgen S, Takao S, Koornneef M, Donohue K (2010) The earliest stages of adaptation in an experimental plant population: strong selection on QTLs for seed dormancy. *Molecular Ecology* **19**: 1335–1351

Huard-Chauveau C, Perchepied L, Debieu M, Rivas S, Kroj T, Kars I, Bergelson J, Roux F, Roby D (2013) An atypical kinase under balancing selection confers broad-spectrum disease resistance in Arabidopsis. *PLoS Genetics* **9**: e1003766

Jakob K, Goss EM, Araki H, Van T, Kreitman M, Bergelson J (2002) *Pseudomonas viridiflava* and *P. syringae*—natural pathogens of Arabidopsis thaliana. *The American Phytopathological Society* **15**: 1195-1203

Karasov TL, Kniskern JM, Gao L, DeYoung BJ, Ding J, Dubiella U, Lastra RO, Nallu S, Roux F, Innes RW, Barrett LG, Hudson RR, Bergelson J (2014) The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature* **512**: 436-440

Kerwin R, Feusier J, Corwin J, Rubin M, Lin C, Muok A, Larson B, Li B, Joseph B, Francisco M, Copeland D, Weinig C, Kliebenstein DJ (2015) Natural genetic variation in *Arabidopsis thaliana* defense metabolism genes modulates field fitness. *eLife* **4**: e05604

Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annual Review of Plant Biology* **55**: 141–172

Korves TM, Schmid KJ, Caicedo AL, Mays C, Stinchcombe JR, Purugganan MD, Schmitt J (2007) Fitness effects associated with the major flowering time gene FRIGIDA in *Arabidopsis thaliana* in the field. *The American Naturalist* **169**: E142-E157

Krämer UTE (2015) Planting molecular functions in an ecological context with *Arabidopsis thaliana*. eL*ife* **4**: e06100.

Kronholm I, Picó FX, Alonso-Blanco C, Goudet J, de Meaux J (2012) Genetic basis of adaptation in *Arabidopsis thaliana*: local adaptation at the seed dormancy QTL DOG1. *Evolution* **66**: 2287-2302

Lasky JR, Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH (2012) Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Molecular ecology* **21**: 5512–5529

Lê Cao KA, Rossouw D, Robert-Granié C, Besse P (2008) A sparse PLS for variable selection when integrating omics data. *Statistical Applications in Genetics and Molecular Biology* **7**: 35

Lê Cao KA, Boitard S, Besse P (2011) Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* **12**: 253

Linnen CR, Poh YP, Peterson BK, Barrett RDH, Larson JG, Jensen JD, Hoekstra HE (2013) Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science* **339**: 1312-1316

Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, Tremblay J, Engelbrektson A, Kunin V, del Rio TG, Edgar RC, Eickhorst T, Ley RE, Hugenholtz P, Green Tringe S, Dangl JL (2012) Defining the core *Arabidopsis thaliana* root microbiome. *Nature* **488**: 86-90

Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, Richardson MF, Anholt RRH, Barron M, Bess C, Blankenburg KP, Carbone MA, Castellano D, Chaboub L, Duncan L, Harris Z, Javaid M, Jayaseelan JC, Jhangiani SN, Jordan KW, Lara F, Lawrence F, Lee SL, Librado P, Linheiro RS, Lyman RF, Mackey AJ, Munidasa M, Muzny DM, Nazareth L, Newsham I, Perales L, Pu LL, Qu C, Ramia M, Reid JG, Rollmann SM, Rozas J, Saada N, Turlapati L, Worley KC, Wu YQ, Yamamoto A, Zhu Y, Bergman CM, Thornton KR, Mittelman D, Gibbs RA (2012) The *Drosophila melanogaster* genetic reference panel. *Nature* **482**: 173-178

Maestre FT (2004) On the importance of patch attributes, environmental factors and past human impacts as determinants of perennial plant species richness and diversity in Mediterranean semiarid steppes. *Diversity and Distribution* **10**: 21-29

Manel S, Perrier C, Pratlong M, Abi-Rached I, Paganini J, Pontarotti P, Aurelle D (2016) Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Molecular ecology* **25**: 170–184

Mevik BH, Wehrens R (2007) The pls package: principal component and partial least squares regression in R. *Journal of Statistical Software* **18**: 1-23

Mitchell-Olds T, Schmitt J (2006) Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. *Nature* **441**: 947-952

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.

Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, Agren J, Bossdorf O, Byers D, Donohue K Dunning M et al. (2010) The scale of population structure in *Arabidopsis thaliana*. *PLoS Genetics* **6**: e1000843

Ramette A, Tiedje M (2007) Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 2761–2766

Reboud X, Le Corre V, Scarcelli N, Roux F, David JL, Bataillon T, Camilleri C, Brunel D, McKhann H (2004) Natural variation among accessions of *Arabidopsis thaliana*: beyond the flowering date, what morphological traits are relevant to study adaptation? *Plant adaptation: molecular genetics and ecology.* Ottawa, CAN: NRC Research Press. Eds. Cronk QCB, Whitton J, Ree RH, Taylor IEP: pages 135-142

Rellstab C, Gugerli F,Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular ecology* **24**: 4348–4370

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. *Current topics in developmental biology* **119**: 111–156

Roux F, Gasquez J, Reboud X (2004) The dominance of the herbicide resistance cost in several *Arabidopsis thaliana* mutant lines. *Genetics* **166**: 449-460

Roux F, Touzet P, Cuguen J, Le Corre V (2016) How to be early flowering: an evolutionary perspective? *TRENDS in Plant Science* **11**: 375-381

Sala OE, Chapin III FS, Armesto JJ, Berlow E, Bloomfield J, Dirzo R, Huber-Sanwald E, Huenneke LF, Jackson RB, Kinzig A, Leemans R, Lodge DM, Mooney HA, Oesterheld M, LeRoy Poff N, Sykes MT, Walker BH, Walker M, Wall DH (2000) Global biodiversity scenarios for the year 2100. *Science* **287**: 1770-1774

Schiffers K, Schurr FM, Travis JMJ, Duputié A, Eckhart VM, Lavergne S, McInerny G, Moore KA, Pearman PB, Thuiller W, Wüest RO, Holt RD (2014) Landscape structure and genetic architecture jointly impact rates of niche evolution. *Ecography* **37**: 1218–1229

Shahriari S, Faria S, Gonçalves AM (2015) Variable selection methods in high-dimensional regression—a simulation study. *Communications in Statistics—Simulation and Computation* **44**: 2548–2561

Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J (2003) Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423**: 74-77

Tylianakis JM, Didham RK, Bascompte J, Wardle DA (2008) Global change and species interactions in terrestrial ecosystems. *Ecology Letters* **11**: 1351–1363

Vignieri SN, Larson JG, Hoekstra HE (2010) The selective advantage of crypsis in mice. *Evolution* **64**: 2153–2158

Weinig C, Dorn LA, Kane NC, German ZM, Halldorsdottir SS, Ungerer MC, Toyonaga Y, Mackay TFC, Purugganan MD, Schmitt J (2003) Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. *Genetics* **165**: 321–329

Weinig C, Stinchcombe JR, Schmitt J (2003) Evolutionary genetics of resistance and tolerance to natural herbivory in *Arabidopsis thaliana*. *Evolution* **57**: 1270-1280

Weinig C, Stinchcombe JR, Schmitt J (2003) QTL architecture of resistance and tolerance traits in *Arabidopsis thaliana* in natural environments. *Molecular ecology* **12**: 1153–1163

Wender NJ, Polisetty CR, Donohue K (2005) Density-dependent processes influencing the evolutionary dynamics of dispersal: a functional analysis of seed dispersal in *Arabidopsis thaliana* (Brassicaceae). *American Journal of Botany* **92**: 960–971

Wilczek AM, Cooper MD, Korves TM, Schmitt J (2014) Lagging adaptation to warming climate in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 7906–7913

Wold S, Sjöstöma M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **58**: 109–130

Züst , Heichinger C, Grossniklaus U, Harrington R, Kliebenstein DJ, Turnbull LA (2012) Natural enemies drive geographic variation in plant defenses. *Science* **338**: 116-11

Conclusion

C. Conclusion

Dans une précédente étude menée au sein de l'équipe, 49 populations d'*A. thaliana* collectées dans 4 régions françaises (Bretagne, Bourgogne, Languedoc et Nord-Pas-de-Calais) avaient été caractérisées au niveau phénologique et au niveau écologique (Brachi *et al.* 2013). Il avait été suggéré que les interactions plante-plante et le sol pouvaient être des agents sélectifs aussi importants que le climat à un niveau régional. Cependant, ces résultats étaient basés (i) sur un petit nombre de populations par région, limitant ainsi la puissance des analyses statistiques et (ii) sur un faible nombre de facteurs biotiques.

L'étude effectuée dans la région Midi-Pyrénées est ainsi très complémentaire et a permis (i) de confirmer l'importance de considérer une large gamme de facteurs écologiques sans oublier les facteurs biotiques qui sont au moins tout aussi importants que les facteurs abiotiques, (ii) de mettre en avant la nécessité de s'intéresser non seulement à un proxy majeur de la *fitness* (i.e. la fécondité) mais aussi aux stratégies reproductives sous-jacentes dont les agents sélectifs potentiels sont largement dépendants du type d'habitat considéré, et (iii) de souligner l'intérêt de travailler à différentes échelles spatiales où des combinaisons de facteurs écologiques auront un impact différent sur les populations.

Il est donc essentiel dans le futur de replacer les études de génomique dans un contexte écologiquement réaliste si l'on veut comprendre et prédire l'adaptation des espèces végétales et animales aux changements globaux futurs. Cela implique notamment de passer plus de temps sur le terrain pour caractériser écologiquement des populations naturelles du point de vue abiotique (climat, microclimat, sol, topographie, vent, etc.) mais également biotique (communautés végétales, microbiennes, fongiques, de mousses, d'herbivores, de faune et microfaune, etc.).

C'est dans ce contexte de réalisme écologique que s'inscrit mon deuxième chapitre de thèse où je me suis attachée à comprendre l'architecture génétique de l'adaptation d'*A. thaliana* à différentes catégories écologiques.

Chapitre 2

Identification des bases génétiques associées aux agents sélectifs potentiels et étude de leurs signatures de sélection

A. Introduction

Dans le chapitre précédent, nous avons vu (i) que de nombreuses variables écologiques pouvaient potentiellement agir comme pressions de sélection sur *A. thaliana* dans la région Midi-Pyrénées, et (ii) que ces variables écologiques constituaient une mosaïque écologique complexe observée à une échelle spatiale fine. Afin de progresser dans notre compréhension de l'adaptation d'*A. thaliana* dans la région Midi-Pyrénées, j'ai voulu identifier les bases génétiques associées aux variables écologiques en effectuant des analyses de type GEA à partir de données de fréquences alléliques obtenues le long du génome pour chacune des 168 populations naturelles étudiées. A partir d'environ 16 plantes échantillonnées par population à la fin de l'hiver 2015, ces fréquences alléliques ont été obtenues à partir d'une stratégie de séquençage génomique de type Pool-Seq.

Pour identifier les bases génétiques associées aux variables écologiques, j'ai utilisé une méthode Bayésienne développée par Mathieu Gautier (laboratoire CBGP, INRA Montpellier) avec qui j'ai collaboré durant ma thèse. Elaborée à partir de la méthode BayEnv2, cette méthode BayPass permet une meilleure estimation de la matrice de covariance populationnelle, condition indispensable pour prendre en compte correctement dans les tests d'association génome-environnement les effets de l'histoire démographique d'*A. thaliana* dans la région Midi-Pyrénées (Gautier *et al.* 2015). Au-delà de la simple identification de régions génomiques associées à des variables écologiques, je me suis aussi intéressée (i) à tester si les régions génomiques identifiées présentaient aussi des traces d'adaptation locale, (ii) à identifier les principales fonctions biologiques sous-jacentes à l'adaptation, et (iii) à mettre en lien les gènes candidats avec l'identité des variables écologiques étudiées.

Ce chapitre vise donc à répondre à plusieurs questions : (i) quelles sont les régions génomiques associées à ces agents sélectifs potentiels? (ii) ces régions génomiques présentent-elles aussi des traces d'adaptation locale, suggérant que nous avons bien identifié des gènes adaptatifs ?, et (iii) les régions génomiques identifiées sont-elles communes entre différents agents sélectifs ?

Dans ce chapitre, je présente sous forme de deux manuscrits les résultats obtenus (i) pour les variables climatiques pour lesquelles les bases génétiques sous-jacentes ont été

largement étudiées à une échelle continentale (Hancock *et al.* 2011) ou régionale (Lasky *et al.* 2012), permettant ainsi non seulement de tester la puissance de la méthode BayPass sur notre jeu de données génomiques mais aussi de tester si les fonctions biologiques impliquées dans l'adaptation au climat sont similaires entre différentes échelles géographiques, et (ii) pour les descripteurs des communautés végétales qui ont été trouvés comme fortement associés à la variation de la production totale de graines entre les 168 populations naturelles. Les analyses de type GEA ont aussi été réalisées sur les variables édaphiques. Malheureusement, par faute de temps, je n'ai pas eu le temps de les intégrer dans ce manuscrit. Les analyses de type GEA ont aussi débuté pour le microbiote bactérien et le microbiote fongique, mais ces travaux font partie intégrante du projet de post-doc de Claudia Bartoli (ancienne post-doc de l'équipe).

NB : dans ce chapitre, mon travail a consisté (i) à échantillonner en moyenne 16 plantes par population, (ii) à caractériser les communautés végétales en collaboration avec Fabrice Roux, Baptiste Mayjonade (IE dans notre équipe) et Nina Hautekèete (MCF Université de Lille), (iii) à effectuer toutes les analyses de type GEA en collaboration ave Mathieu Gautier (laboratoire CBGP, INRA Montpellier), (iv) à effectuer toutes les analyses de détection de traces de sélection et d'enrichissement en processus biologiques, et (v) à identifier les gènes candidats en collaboration avec Dominque Roby (DR CNRS au LIPM) et Fabrice Roux. Les extractions d'ADN ont été effectuées par Claudia Bartoli et Fabrice Roux. Les analyses bioinformatiques pour estimer les fréquences alléliques le long du génome dans chaque population ont été réalisées par Sébastien Carrere (IR plate-forme de bioinformatique du LIPM) et Claudia Bartoli.

Manuscrit

“A genomic map of adaptation to local climate in *Arabidopsis thaliana*”

Léa Frachon, Claudia Bartoli, Sébastien Carrère, Olivier Bouchez, Adeline Chaubet, Mathieu Gautier, Dominique Roby and Fabrice Roux

Le manuscrit est en révision à New Phytologist

B. Manuscrit: "A genomic map of adaptation to local climate in *Arabidopsis thaliana*"

Léa Frachon,[¶,1] Claudia Bartoli,[¶,1] Sébastien Carrère,[1] Olivier Bouchez,[2] Adeline Chaubet[2] Mathieu Gautier,[3] Dominique Roby[1] and Fabrice Roux[*,1]

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

[2] INRA, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France

[3] INRA, UMR CBGP (Centre Biologie pour la Gestion des Populations), Campus International de Baillarguet, Montferrier-sur-Lez & IBC (Institut de Biologie Computationnelle), Montpellier, France

[¶] These authors contributed equally to this work.

Corresponding author: Fabrice Roux, e-mail: fabrice.roux@toulouse.inra.fr, phone number: +33 (0)5 61 28 55 57

Word count in the abstract: 188 words

Word count for the main body of text: 6 483 words

Introduction: 1 284 words

Materials and Methods: 2 480 words

Results: 1 346 words

Discussion: 1 356 words

Acknowledgments: 17 words

Number of figures: 5 (Figures 2, 3, 4 and 5 should be published in colour)

Number of tables: 3

Summary

- Understanding the genetic bases underlying climate adaptation is a key element to predict the potential of species to face climate warming. Although substantial climate variation is observed at a regional or local scale, most genomic maps of climate adaptation have been established at a broader geographical scale.
- Here, by using a Pool-Seq approach and a Bayesian hierarchical model, we performed a genome-environment association (GEA) analysis to investigate the genetic basis of adaptation to 21 climate variables in 168 natural populations of *Arabidopsis thaliana* distributed in south-west of France.
- Climate variation among the 168 populations represented ~23.5% of climate variation among 426 European locations where *A. thaliana* inhabits. We identified neat peaks of association, with the most significantly associated SNPs being significantly enriched in likely functional variants and in the extreme tail of spatial differentiation among populations. In addition, genes involved in transcriptional mechanisms, including epigenetic mechanisms, were overrepresented in the plant functions associated with local adaptation.
- Climate adaptation is an important driver of genomic variation in *A. thaliana* at a small geographical scale and mainly involves genome-wide changes in fundamental mechanisms of gene regulation.

Key words: *Arabidopsis thaliana*, Bayesian hierarchical model, climate change, genome-environment association analysis, local adaptation, Pool-Seq, spatial grain.

Chapitre 2

Introduction

In the context of the contemporary environmental change, a major goal in evolutionary ecology is to understand and predict the ability of a given species to persist in the presence of novel climate conditions (Bay *et al.*, 2017). A lack of response of species to selection due to climate change would cause an erosion of biodiversity by disrupting ecosystems sustainably (Pecl *et al.*, 2017). Overall, species can adopt three non-exclusive responses to face the altered and fluctuating climate conditions (Hansen *et al.*, 2012; Pecl *et al.*, 2017). Firstly, species can migrate to track current climate spatial shifts. This response can however be limited for (i) long-distance dispersal organisms because of the presence of multiple anthropogenic barriers such as landscape fragmentation, agriculture and urbanization (Ewers & Didham, 2006), and (ii) organisms with restricted dispersal as for example sessile plants lacking dispersal mechanism (i.e. barochorous species) or disperser reward (Wang *et al.* 2016).

Secondly, organisms can rapidly acclimate to novel climate conditions *via* phenotypic plasticity, a process defined as the capacity of a single genotype to exhibit a range of phenotypes in response to variation in the environment where the genotype inhabits and evolves (Price *et al.*, 2003; Ghalambor *et al.*, 2007; Valladares *et al.*, 2014). Theoretical models predict that adaptive plasticity can help natural populations to reach a new phenotypic optimum (Lande, 2009; Chevin *et al.*, 2010). Despite its theoretical benefits, adaptive phenotypic plasticity is not as frequent as expected in nature because it can be constrained by diverse costs and limits (initially reviewed in DeWitt *et al.*, 1998). For example, one of the main limits concerns the unreliability of environmental cues, leading to non-adaptive or mal-adaptive plastic responses (van Kleunen & Fischer, 2005). In the context of climate change, such unreliable cues can correspond to extreme climate events that fall outside the range of historic climate conditions encountered by natural populations (Orlowsky & Seneviratne, 2012).

Thirdly, over a longer term, organisms can adapt to novel climate conditions *via* genetic selection, provided that there is sufficient standing genetic variation or new genetic variation arising from either *de novo* mutations or immigration of climate-adapted alleles from nearby populations (Hoffmann & Sgrò, 2011; Bay *et al.*, 2017). Predicting the response of species to climate change therefore requires the identification of the genetic bases associated with climate variation. Two major approaches can be used to describe the genomic

architecture (number of genes, allelic effects, locations across the genome) underlying climate adaptation. Few Genome-Wide Association mapping studies (GWAS) reported the genomic architecture associated with phenotypic traits potentially related to climate adaptation such as thermal sensitivity (Li *et al.,* 2014). On the other hand, based on the assumption that each population is adapted to local conditions, the most exploited approach corresponds to genome-environment association (GEA) analyses, in which a genome scan is performed to identify significant associations between genetic polymorphisms and environmental variables (Yoder *et al.,* 2014; Abebe *et al.,* 2015; Lasky *et al.,* 2015; Rellstab *et al.,* 2015; Hoban *et al.,* 2016; Manel *et al.,* 2016). Due to publicly available gridded estimates of climate and the development of next-generation sequencing (NGS) technologies, the number of GEA analyses performed on climate variables rapidly increased in the last few years, thereby revealing that climate adaptation is likely to be highly polygenic (Bay *et al.,* 2017). Most of the GEA analyses on climate were performed at large spatial scales (i.e. from several hundred to several thousand kilometers). However, substantial climate variation can also be observed at smaller spatial scales (from several tens of meters to several tens of kilometers), leading for example to sharp climate gradients in mountains (Manel *et al.,* 2010; Kubota *et al.,* 2015) or a mosaic of climatically optimal and suboptimal sites within the reach of gene flow among populations (Pluess *et al.,* 2016). The complementarity of performing GEA analyses from continental to local geographical scales should shed light on the genetic bases underlying coarse-grained and fine-grained climate variation (Manel *et al.,* 2010), which in turn would increase the reliability of predictions of response to climate change. In addition, as previously advised for adaptive phenotypic traits (Bergelson & Roux, 2010), working at a small geographical scale should reduce the limitations of GEA analyses often observed when working at larger geographical scales such as the confounding background produced by population structure, rare alleles and allelic heterogeneity. Finally, a fine-grained spatial scale is much more coherent with the mean distance of species migration (few km per decade; Chen *et al.,* 2011).

Arabidopsis thaliana, the flagship species of plant genomics, is a widely distributed annual selfing species found in a large range of climate environments across its native range in Eurasia (Hoffmann, 2002). Reciprocal transplants performed at the European scale revealed that climate gradients likely play a major role in local adaptation of *A. thaliana* (Fournier-Level *et al.,* 2011; Ågren & Schemske, 2012; Ågren *et al.,* 2013; Wilczek *et al.,* 2014).

Furthermore, among plant species, *A. thaliana* pioneered the identification of climate-adaptive genetic loci at the genome-wide scale. A GEA analysis based on 948 Eurasian accessions succeeded to establish a genomic map of local adaptation to climate variation (after controlling for population structure), which in turn successfully predicted the relative fitness of a subset of accessions grown together in a common garden in the north of France (Hancock *et al.*, 2011). In a complementary approach, by measuring lifetime fitness of a diverse set of accessions grown in four climatically contrasted European sites, Fournier-Level *et al.* (2011) found that local adaptation to climate was mainly driven by Quantitative Trait Loci (QTLs) that vary in magnitude rather than direction across environments. This prevalence of independent loci underlying local adaptation suggested a flexible genomic architecture that can facilitate responses to climate change.

Local adaptation to climate still explains a substantial portion of genomic variation of *A. thaliana* at a regional scale (i.e. several hundred km; Lasky *et al.*, 2012). In addition, strong climate-phenotype associations were identified at a local scale (from several tens of meters to several tens of kilometers), either along sharp altitudinal gradients (Montesinos-Navarro *et al.*, 2011; Luo *et al.*, 2015; Günther *et al.*, 2016) or in a mosaic of climatically contrasted sites (Brachi *et al.*, 2013). However, although a candidate gene approach was used to start the identification of climate-adaptive genetic loci at the regional scale (i.e. Iberian Peninsula; Méndez-Vigo *et al.*, 2011; Vidigal *et al.*, 2016), studies reporting the genomic architecture of adaptation to various climate variables at a fine spatial grain are still scarce in *A. thaliana*.

In this study, we aimed to establish a genomic map of adaptation to local climate in *A. thaliana*. We focused on 168 natural populations of *A. thaliana* distributed homogeneously in the south-west of France, a geographical region under the influence of three contrasted climates (i.e. oceanic climate, Mediterranean climate and mountain climate). By using a Bayesian hierarchical model that control for confounding by population structure (Gautier, 2015), we conducted a GEA analysis between 1,638,649 SNPs and 21 climate variables. Because most *A. thaliana* natural populations located in France are genetically diverse (Le Corre, 2005; Platt *et al.*, 2010; Brachi *et al.*, 2013), we obtained a representative picture of within-population genetic variation across the genome by adopting a Pool-Seq approach. We then searched for genome-wide signatures of selection on the SNPs the most associated with climate variation (i.e. top SNPs). Following Hancock *et al.* (2011) and Brachi *et al.* (2015), we therefore tested whether those top SNPs were enriched for non-synonymous variants and in the extreme tail of a genome-wide spatial differentiation scan. We finally performed an

enrichment analysis to identify the biological processes involved in the adaptation of *A. thaliana* to local climate and discussed the function of the main candidate genes.

Materials and Methods

Plant material

A field prospection in May 2014 allowed the identification of 233 *A. thaliana* natural populations in the Midi-Pyrénées region (South-West of France). In agreement with an important population turnover of natural populations observed in *A. thaliana* (Picó, 2012), individuals were present in only 168 populations (~72.1%) in late winter 2015 when the sampling campaign was performed. The average distance among the 168 populations was 100.6 km (median = 93.4 km, max = 265.2 km, min = 0 km, 1st quartile = 57.3 km, 3rd quartile = 137.2 km).

Climate characterization

The 168 geo-localized populations were characterized for 20 biologically meaningful climate variables retrieved from the ClimateEU database. Climate data has been generated with the ClimateEU v4.63 software package (available at *http://tinyurl.com/ClimateEU)* based on methodology described by Hamann *et al.* (2013). The grid resolution of the 20 climate variables (~1.25 arcmin, ~600 m) was smaller than the average distance among populations. Climate data were averaged across the 2003-2013 annual data. In addition, altitude was obtained from http://www.gps-coordinates.net.

In order to compare the level of climate variation among the 168 populations of the Midi-Pyrénées region with the level of climate variation among natural populations of *A. thaliana* located in France and Europe, we first extracted data for the 20 biologically meaningful climate variables (ClimateEU database) as well as altitude (http://www.gpsvisualizer.com/elevation) for 472 locations where *A. thaliana* have been collected and geo-localized in Europe (including 46 locations in France; Hancock *et al.*, 2011) and 49 natural populations geo-localized in four climatically contrasted regions of France (continental, semi-oceanic, oceanic and Mediterranean; Brachi *et al.*, 2013). To visualize the climatic space encountered by *A. thaliana* at different geographical scales, we then conducted

a principal component analysis (PCA) based on the 689 locations using the ade4 1.7-6 version package in the *R* environment (Chessel *et al.,* 2004; Dray *et al.,* 2017). Finally, the percentage of climatic variation in Europe observed among locations at sub-geographical scales was calculated by dividing the extent of variation observed on the two first Principal Components (PCs) at the French and Midi-Pyrénées scales by the extent of variation observed at the European scale.

Variable	Description	source	Grid resolution
altitude	altitude (m)	www.coordonnees-gps.fr	-
MAT	mean annual temperature (°C)	ClimateEU	1.25 arcmin
MWMT	mean warmest month temperature (°C)	ClimateEU	1.25 arcmin
MCMT	mean coldest month temperature (°C)	ClimateEU	1.25 arcmin
TD	temperature difference between MWMT and MCMT, or continentality (°C)	ClimateEU	1.25 arcmin
MAP	mean annual precipitation (mm)	ClimateEU	1.25 arcmin
AHM	annual heat:moisture index (MAT+10)/(MAP/1000))	ClimateEU	1.25 arcmin
SHM	summer heat:moisture index ((MWMT)/(MSP/1000))	ClimateEU	1.25 arcmin
DD<0	degree-days below 0°C, chilling degree-days	ClimateEU	1.25 arcmin
DD>5	degree-days above 5°C, growing degree-days	ClimateEU	1.25 arcmin
DD<18	degree-days below 18°C, heating degree-days	ClimateEU	1.25 arcmin
DD>18	degree-days above 18°C, cooling degree-days	ClimateEU	1.25 arcmin
NFFD	the number of frost-free days	ClimateEU	1.25 arcmin
Tave_wt	winter (Dec.(prev. yr) - Feb.) mean temperature (°C)	ClimateEU	1.25 arcmin
Tave_sp	spring (Mar. - May) mean temperature (°C)	ClimateEU	1.25 arcmin
Tave_sm	summer (Jun. - Aug.) mean temperature (°C)	ClimateEU	1.25 arcmin
Tave_at	autumn (Sep. - Nov.) mean temperature (°C)	ClimateEU	1.25 arcmin
PPT_wt	winter precipitation (mm)	ClimateEU	1.25 arcmin
PPT_sp	spring precipitation (mm)	ClimateEU	1.25 arcmin
PPT_sm	summer precipitation (mm)	ClimateEU	1.25 arcmin
PPT_at	autumn precipitation (mm)	ClimateEU	1.25 arcmin

Table 1 List of the 21 climate variables used in this study.

Spatial grains of climatic variation

To estimate the spatial scales of variation of each climate variable, a spectral decomposition of the spatial relationships among the 168 populations was first modeled with Principal Coordinates of Neighbor Matrices (PCNM), by running the pcnm() function implemented in the vegan package (R package version 2.3-5; Oksanen *et al.*, 2016) using the Euclidean distance matrix based on the GPS coordinates for the 168 populations. This analysis allows decomposing the spatial structure among the sites under study into orthogonal PCNM components corresponding to successive spatial grains (Borcard & Legendre, 2002). The first PCNM components define a large spatial grain, while the last PCNM components correspond to finer grains (Ramette & Tiedje, 2007; Borcard *et al.*, 2004). All PCNM components were then used as explanatory variables in a multiple linear regression on each

climate variable. To account for multiple testing, a Benjamin-Hochberg procedure was performed for each climate variable across all PCNM components to control for a false discovery rate (FDR) at a nominal level of 5% (Benjamini & Hochberg, 1995).

Genomic characterization and data filtering

For each population, a mean number of 16.5 plants (total number = 2,776 plants, median = 17 plants , max = 17 plants, min = 5 plants, Supporting Information Table S1) were collected randomly in late February – early March 2015 and brought back to a cold frame greenhouse (no additional light or heating). In April 2015, leaf tissue was collected from approximately 16 plants per population for a total of 2,574 plants (min = 5 plants, max = 16 plants, mean = 15.32 plants, median = 16 plants, Supporting Information Table S1). More precisely, a portion of a rosette leaf for each plant was placed in 96-well Qiagen S-block plates containing a 3 mm bead in each well and samples were stored at -80°C. Prior to DNA extraction, plates were put 30 sec in liquid nitrogen and samples were then crushed by using Mixer Mill MM 300 Retsch ® with 1 min of vibration at a frequency of 30 vibrations/s. Genomic DNA from the 2,574 plants was extracted as described in Brachi *et al.* (2013) and total DNA for each individual extraction was quantified by using a Quant-iT™ PicoGreen® dsDNA Assay Kit with a QPCR ABI7900 machine. Individuals from each population were then used to constitute an equimolar pool and from 50bp to 500bp fragments were produced for each pool by using Covaris M220 Focused-ultrasonicator™. Produced fragments were analyzed with Agilent 2100 Bioanalyzer with a DNA 7500 chip and purified with Agencourt® AMPure® XP paramagnetic beads by following manufacturer instructions protocol. Illumina indexes were added by PCR amplification with the following cycling program: 1 min at 94°C, followed by 12 cycles of 1 min at 94°C, 1 min at 65°C and 1 min at 72°C, followed by a final elongation of 10 min at 72°C. After this step, PCR products were purified as described before. Samples were sequenced at the Get-Plage platform (Toulouse) on an Illumina Hiseq 3000 sequencer. Raw data for each population used in this study are available at the NCBI Sequence Read Archive (SRA) (http://ncbi.nlm.nih.gov/sra) through the study accession SRP103198.

Raw reads were mapped on the reference genome Col-0 with glint tool (version 1.0.rc8.779) (Faraut & Courcelle, unpublished software) by using the following parameters: *glint mappe* --no-lc-filtering --best-score --mmis 5 --lmin 80 --step 2. The mapped reads were

filtered for proper pairs with SAMtools (v0.01.19; Li *et al.*, 2009) (*samtools view -f 0x02*). A semi-stringent SNPCalling across the genome was then performed for each population with SAMtools mpileup2snp (Li *et al.*, 2009) and VarScan mpileup2snp (Koboldt *et al.*, 2012) softwares by using as parameters a minimum coverage (minimum read depth at a position to make a call) of 5 reads and a minimum variant allele frequency threshold of 0.00001. SNP-Pooling was then performed to obtain polymorphic sites across the whole 168 samples and SNP-Calling was inferred on the whole polymorphic sites as described above (VarScan mpileup2cns; min coverage = 1) and bi-allelic positions were filtered. The mean and the median coverage to a unique position in the reference genome was ~26.3x and ~24.5x, respectively (min = 11.60x, max = 48.69x).

After bioinformatics analysis, data consisted of allele read counts (for both the reference and alternate alleles) for a total of 4,781,661 SNPs in the 168 populations. This data set was further filtered using custom scripts. First, SNPs without mapped reads in at least 8 populations were removed (number of remaining SNPs = 3,798,406). Second, for each population, we calculated the relative coverage of each SNP as the ratio of its coverage to the median coverage (computed over all the SNPs in the corresponding population). Because multiple gene copies in the 168 populations can map to a unique gene copy in the reference genome Col-0, we removed SNPs with a mean relative coverage across the 168 populations above 1.5 (number of remaining SNPs = 3,260,041). In addition, we removed SNPs with a standard deviation of allele frequency across the 168 populations below 0.004 (number of remaining SNPs = 3,248,168). Third, because genomic regions present in Col-0 can be absent in most of the 168 populations or genomic regions present in most of the 168 populations can be absent in Col-0, we removed SNPs with a mean relative coverage across the 168 populations below 0.5 (number of remaining SNPs = 3,172,313). Fourth, we removed SNPs that were monomorphic in more than 90% of the populations, leading to a final data set consisting of read counts for 1,638,649 SNPs in the 168 populations.

Genome-environment analysis

Based on the 1,638,649 SNPs, whole genome scans for adaptive differentiation and association with climate variables were performed with BayPass 2.1 (Gautier, 2015). Accommodating Pool-Seq data, the underlying Bayesian hierarchical models explicitly

account for the scaled covariance matrix of population allele frequencies (**Ω**) which make the analyses robust to complex demographic histories.

Here, capitalizing on the large number of available SNPs, we adopted a sub-sampling procedure to estimate **Ω**. This consisted in dividing the full data set into 32 sub-data sets, each containing 3.125% of the 1,638,649 SNPs (ca., 51,000 SNPs taken every 32th rank across the genome), that were further analyzed in parallel under the core model using default options for the Markov Chain Monte Carlo (MCMC) algorithm (except -npilot 15 -pilotlength 500 -burnin 2500). Pairwise comparisons of the 32 resulting covariance matrices confirmed all estimates were consistent with highly correlated elements, while the pairwise FMD distances (Förstner & Moonen, 2003) had a narrow range of variation (from 2.04 to 2.24) with a mean value equal to 2.15.

These analyses carried out under the core model also provided estimate of the XtX measure of differentiation for all the SNPs (combined over sub-data sets). For a given SNP, the XtX is defined as the variance of the standardized population allele frequencies, i.e. rescaled using **Ω** and across population allele frequencies (Günther & Coop, 2013; Gautier, 2015). This allows for a robust identification of overly differentiated SNPs by correcting for the genome-wide effects of confounding demographic evolutionary forces such as genetic drift and gene flow.

As a matter of expedience, given the close similarity of the **Ω** estimates obtained on the 32 sub-data sets, we retained for further analyses the matrix $\widehat{\mathbf{\Omega}_1}$ (obtained on the first sub-data set) as an estimate of the scaled covariance matrix of population allele frequencies. To evaluate the spatial scale of genomic variation, we first performed a singular value decomposition (SVD) of $\widehat{\mathbf{\Omega}_1}$. The coordinates of the resulting Principal Components (PC) were then regressed against latitude and longitude according to the following formula (PROC GLM procedure in SAS 9.3 SAS Institute Inc., Cary, North Carolina, USA):

$$\text{PC coordinates}_{ij} \sim \text{latitude}_i + \text{longitude}_j + \text{latitude}_i * \text{longitude}_j + \varepsilon_{ij} \quad (1)$$

Finally, genome-wide analysis of association with climate covariables were carried out under the AUX model (-auxmodel) parameterized with $\widehat{\mathbf{\Omega}_1}$. The support for association of each SNP with each covariable *k* (i.e., a non null regression coefficient $\widehat{\beta_{ik}}$ between SNP i allele frequencies and a covariable k) was evaluated by computing Bayes Factor (BF) measured in deciban units (dB), a BF>15 being classically considered as strong evidence for

association (Gautier, 2015). Note that the AUX model allows to explicitly accounting for multiple testing issue by integrating over (and estimating) the unknown proportion of SNPs actually associated with a given covariable. Here, 23 covariables (21 climate variables as well as the two first PC's) were considered separately and standardized prior to analyses using the scalecov option. In practice, BF and the associated regression coefficients $\widehat{\beta}_t$ between SNP allele frequencies and climate variation were estimated for each SNP by analyzing in parallel the 32 sub-data sets described above (but with the same matrix **Ω**) using default options (except -npilot 15 -pilotlength 500 -burnin 2500) for MCMC.

Enrichment across annotation categories of variants for climatic associations

To test whether different categories of genetic variants were enriched for those SNPs that were the most significantly associated with climate variation (i.e. with the highest BF), we first annotated all the SNPs (n = 1,638,649 SNPs) by using the SnpEff program (Cingolani *et al.*, 2012). In this study, we considered six categories of genic variants representing 98.5% of all the SNPs tested genome-wide, i.e. intragenic variant (mainly transposable elements, 11.5%), intergenic variant (49.1%), UTR variant (5.6%), intron variant (14.9%), synonymous variant (9.1%) and replacement variant (8.2%). We then tested whether SNPs in the 0.5% upper tail of the BF distribution of each climate variable were over-represented or under-represented in each of the six genetic variant categories. For each category of genetic variant, we used the following equation:

$$FE_{SNPeff} = \frac{s_a/s}{S_a/S} \qquad (2)$$

where s is the number of BF in the 0.5% upper tail of the BF distribution, s_a is the number of SNPs in the 0.5% upper tail of the BF distribution that also belonged to the genetic variant category, S is the total number of annotated SNPs tested genome-wide and S_a is the number of annotated SNPs tested genome-wide that also belonged to the genic variant category. Based on a methodology previously described in Hancock *et al.* (2011) that takes into account original Linkage Disequilibrium (LD) patterns among SNPs in the 0.5% upper tail of the BF distribution, statistical significance of enrichment for each category of genetic variants was assessed by running 1000 null permutations.

Enrichment in signatures of selection for climatic associations

For each climate variable, we tested whether SNPs with the highest BF were over-represented in the extreme tail of the XtX distribution according to the methodology described in Brachi *et al*., (2015):

$$FE_{XtX} = \frac{n_a/n}{N_a/N} \qquad (3)$$

where *n* is the number of XtX in the 0.5% upper tail of the XtX distribution, n_a is the number of SNPs in the 0.5% upper tail of the BF distribution that were also in the 0.5% upper tail of the XtX distribution, *N* is the total number of SNPs tested genome-wide and N_a is the number of SNPs in the 0.5% upper tail of the BF distribution. Statistical significance of enrichment was assessed by running 10,000 null permutations based on the methodology described in Hancock *et al*. (2011).

Enrichment in biological processes for climatic associations

To determine which biological processes were enriched with climate adaptation, we tested whether SNPs in the 0.01% upper tail of the BF distribution of each climate variable were over-represented in each of the 736 Gene Ontology Biological Processes obtained from the GOslim set (The Gene Ontology Consortium, 2008). Statistical significance of enrichment was assessed for the 0.01% upper tail of the BF distribution as described above in accordance to Hancock *et al.* (2011). For each significantly enriched biological process, the identity of the genes containing SNPs in the 0.01% upper tail of the BF distribution was retrieved.

Identification of candidate genes associated with climate variation

A four-step procedure was adopted to identify candidate genes associated with climate variation. First, we selected the 50 SNPs with the highest BF for each of the 21 climate variables as well as for each of the two first PCs, leading to a total of 1,150 SNPs. Second, the 1,150 top SNPs were pruned to obtain a list of 835 unique top SNPs. Third, candidate regions were defined on the genomic regions supported by at least three top SNPs successively separated by less than 10kb. This step led to the identification of 61 candidate regions

containing on average 5.3 SNPs (median = 4 SNPs, min = 3 SNPs, max = 29 SNPs) and with a mean length of 10.7 kb (median = 8.6 kb, min = 0.3 kb, max = 41.6 kb). Finally, using the TAIR 10 database (*https://www.arabidopsis.org/*), we retrieved all the annotated genes located within or overlapping with the 61 candidate regions, leading to the identification of 187 annotated genes.

Results

Climate variation and associated spatial grains

Here, we focused on 168 *A. thaliana* natural populations distributed in the Midi-Pyrénées region located in the south-west of France (Fig. 1). We identified 82 Principal Coordinates of Neighbor Matrices (PCNM) components, suggesting a relatively homogeneous spatial distribution of the 168 populations across the sampling area (Fig. 1).

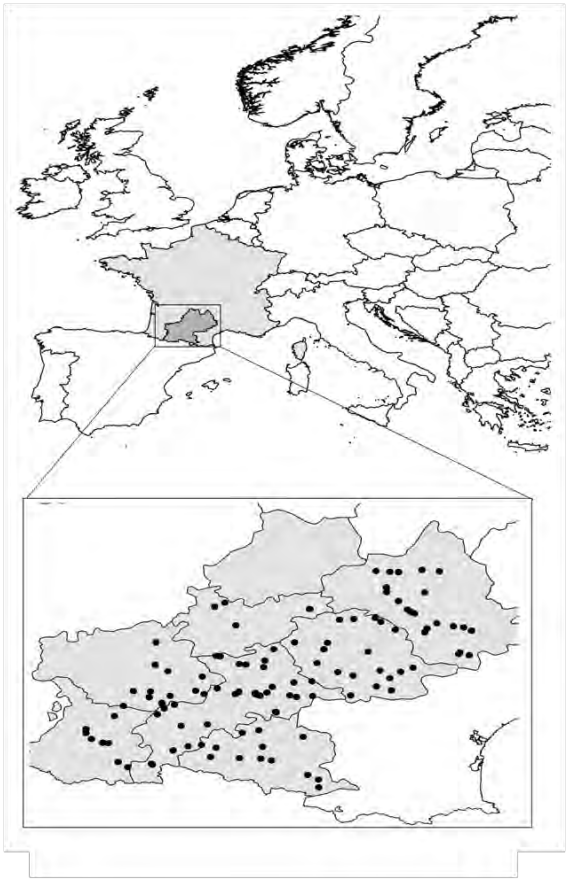


Fig. 1 Distribution of the 168 *A. thaliana* natural populations across the Midi-Pyrénées region (south-west of France). Grey zone represents total area of metropolitan France. Black dots represent locations inhabited by *A. thaliana* in the Midy-Pyrénées region.

The 168 populations were characterized for 21 climate variables with a grid resolution smaller than the average distance among populations (i.e. 100.6 km, SD = 56.0 km). Despite the restricted size of our sampling area (~8.2% of total area of metropolitan France, Fig. 1), climate variation among the 168 populations represented ~23.5% of climate variation among 426 European locations inhabited by *A. thaliana* (Hancock *et al.*, 2011; Fig. 2, Supporting Information Fig. S1). Notably, after removing one extreme altitude value (i.e. 2916 m in a location in Austria), the altitude range observed among the 168 populations (84 m to 1129 m) represented 56.1% of the altitude range observed among the European locations (-2 m to 1862 m) (Supporting Information Fig. S1). In addition, the climate space of the Midi-Pyrénées region largely differed from the climate space of the other French regions inhabited by *A. thaliana* (Fig. 2c).

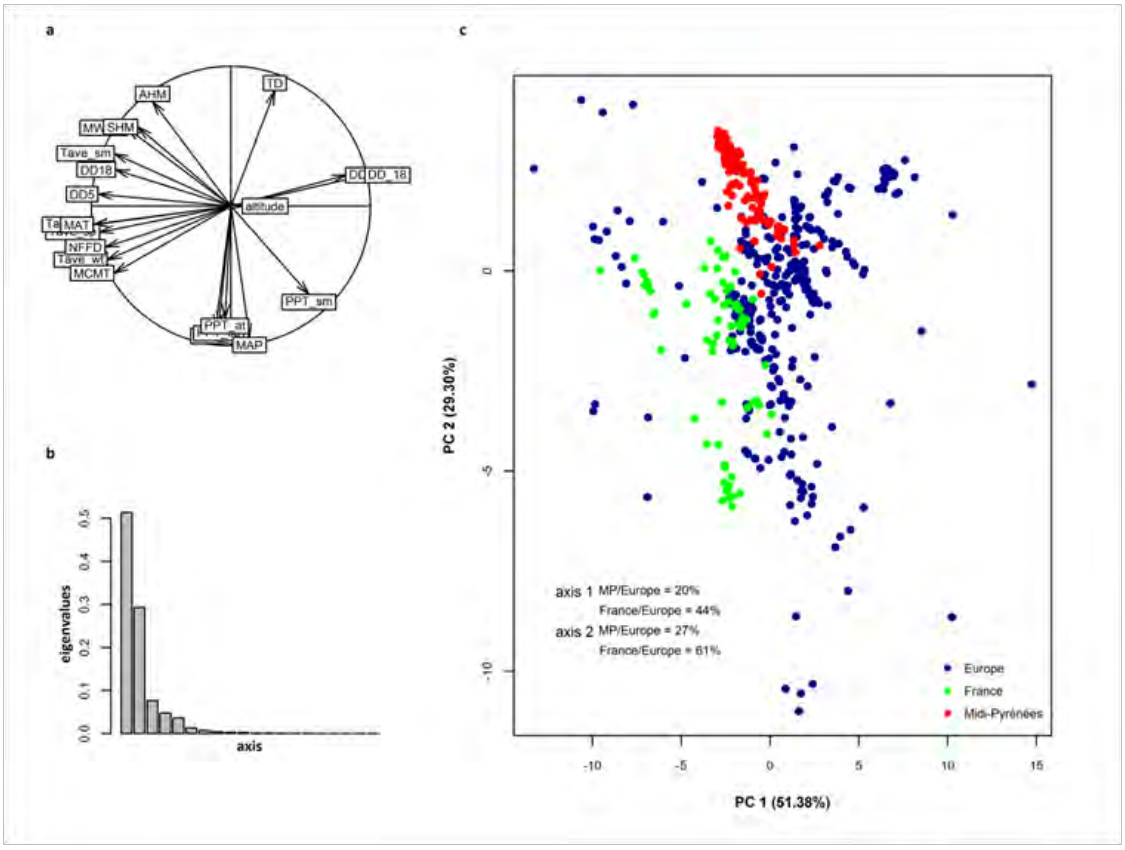


Fig. 2 Climate variation among natural populations of *A. thaliana* collected at different geographical scales. (**a**) Factor loading plot resulting from a principal component analysis. Factor 1 and factor 2 explained 51.38% and 29.30% of total climate variation. See Table 1 for a description of the climate variables. (**b**) Distribution of eigenvalues. (**c**) Position of the 168 populations of the Midi-Pyrénées region in the European and French climatic space of *A. thaliana*. Blue dots represent European locations without considering locations in France (n = 426), green dots represent French locations without considering locations in the Midi-Pyrénées region (n = 95), red dots represent locations in the Midi-Pyrénées region (n = 168). 'France/Europe' indicates the percentage of climatic variance in Europe observed among the locations in France whereas 'MP/Europe'

indicates the percentage of climatic variance in Europe observed among the locations in the Midi-Pyrénées region.

In the Midi-Pyrénées region, all the pairwise correlations among the 21 climate variables were highly significant (mean absolute Spearman's *rho* = 0.810, Supporting Information Fig. S2). Accordingly, more than 90% of climate variation was captured by the two first axes of a PCA based on the 168 populations (Supporting Information Fig. S3). Two patterns of spatial variation were observed for the 21 climate variables (Supporting Information Fig. S4). Altitude and the 15 temperature-related variables were mainly associated with the first PCNM components (Supporting Information Fig. S4), indicating a coarse-grained spatial variation. On the other hand, coarse-grained to very fine-grained spatial variations was observed for the five precipitation-related variables (Supporting Information Fig. S4).

Altogether, these results indicated the presence of contrasted climates, even at a short geographical distance, among the locations inhabited by *A. thaliana* in the Midi-Pyrénées region.

Genome-climate associations

Using Pool-Seq data, we estimated within-population allele frequencies across the genome for a final number of 1,638,649 SNPs (i.e. one SNP every 72 bp). Based on singular value decomposition (SVD) of the population covariance-variance matrix **Ω**, we found that 96.4% of the genomic variation observed in the Midi-Pyrénées region was explained by the first principal component (Supporting Information Fig. S5), supporting strong population subdivision already observed in other French regions (Le Corre, 2005; Brachi *et al.*, 2013). In addition, a weak geographic pattern along a south-west/north-east axis was observed for genomic variation (latitude: t value = 4.734, $P = 4.73 \times 10^{-6}$, longitude: t value = 4.417, $P = 1.81 \times 10^{-5}$, latitude x longitude: t value = -4.428, $P = 1.73 \times 10^{-5}$, adjusted R^2 = 10.5%; Supporting Information Fig. S5).

To identify the genomic regions associated with climate variation, we then performed a genome-wide scan for association with the 21 climate variables and the two first climate PCs, using a Bayesian hierarchical model that includes a population covariance matrix accounting for the neutral covariance structure across population allele frequencies. For each

climate variable, we estimated the regression coefficients between SNP allele frequencies and climate variation (β_i) and evaluated the support for association (non null β_i) of the association between a given SNP and a climate variable with a Bayes factor (BF). We identified very neat peaks of association for most climate variables (Fig. 3, Supporting Information Fig. S6). Accordingly, as illustrated for altitude, the number of degree-days below 0°C and winter precipitations, standardized allele frequencies variation of the most significant SNPs strongly overlapped with climate variation (Fig. 4).

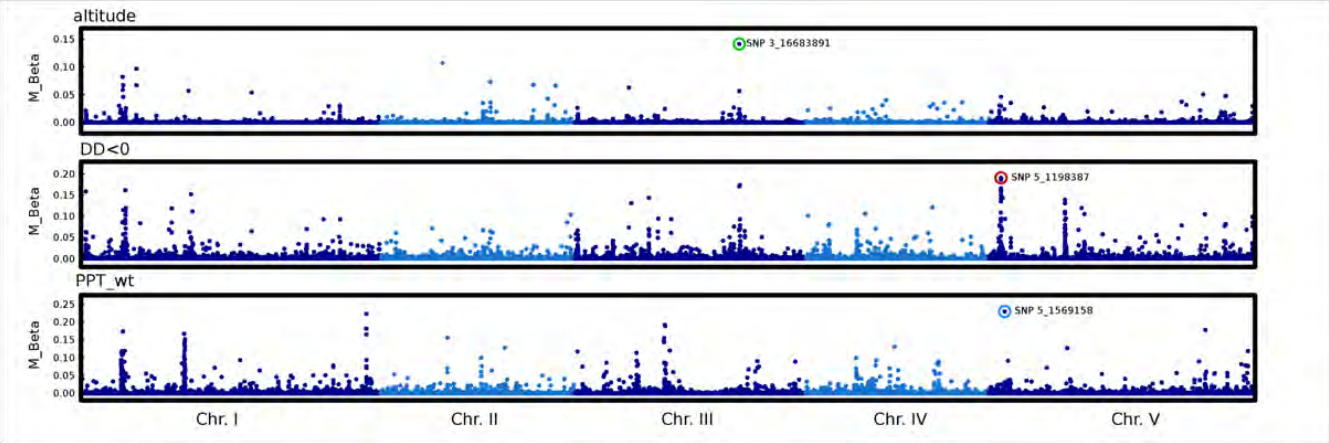


Fig. 3 Manhattan plots of the genome-environment association results for three climate variables. The x-axis indicates the position along each chromosome. The five chromosomes are presented in a row along the x-axis in different degrees of blue. The y-axis indicates the posterior mean of the regression coefficient β_i (M_Beta value) estimated by the AUX model implemented in the program BayPass. The three most associated SNPs are circled with a different color**.** 'DD<0' degree-days below 0°C, 'PPT_wt' mean winter precipitation.

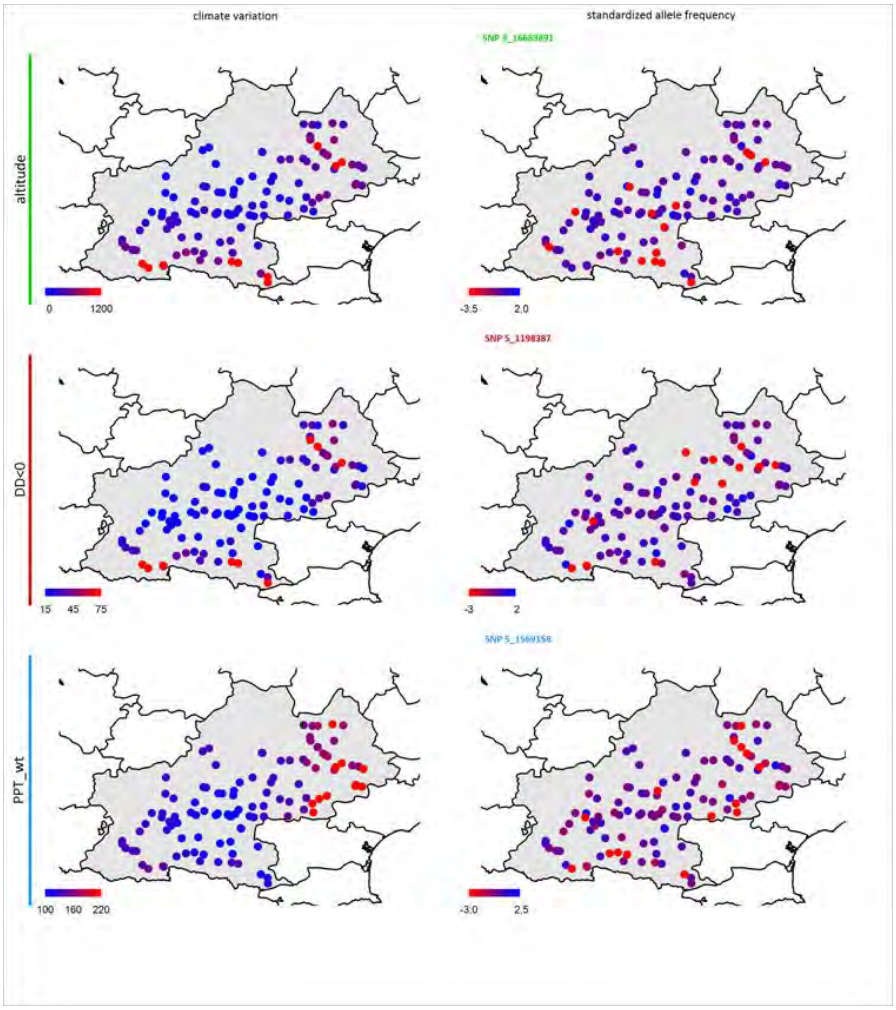


Fig. 4 Map illustrating the geographic variation of three climate variables and the standardized allele frequencies of the SNPs the most associated with these climate variables. 'DD<0' degree-days below 0°C, 'PPT_wt' mean winter precipitation.

To test whether the genomic architecture associated with climate variation differs among the 23 climate variables (21 single climate variables and the two first PCs), we assessed the degree of climatic pleiotropy among the SNPs the most associated with climate variation. We therefore retrieved the 50 SNPs with the highest BF for each of the 23 climate variables. Among the 835 resulting unique top SNPs, 82.8% were associated with a single climate variable whereas the remaining SNPs were associated with up to 12 climate variables (mean = 3.20, median = 2). These results suggest that while different genomic architectures underlay the response of *A. thaliana* to single climate variables, a substantial number of pleiotropic loci (i.e. loci associated with variation in several climate variables) can contribute to the response to a combination of climate variables.

Testing for signatures of selection

To test whether our results detect true signals of adaptation, we first tested whether the top SNPs (i.e. the 0.5% upper tail of the BF distribution) were differentially enriched for six categories of genetic variants (intragenic, intergenic, UTR, intron, synonymous and replacement). For 19 out of the 23 climate variables, we found that likely functional replacement SNPs show strong and significant enrichments (mean = 1.17) to the detriment of putative neutrally evolving intragenic SNPs (mean = 0.60) (Table 2a). Interestingly, as illustrated for mean annual temperature and mean annual precipitations (Fig. 5), fold-enrichment values increased according to the degree of neutral evolution expected for the six categories of genetic variants.

a	classes of sites												b	XtX	
	intragenic		intergenic		UTR		intron		synonymous		replacement				
climate variables	Fold value	*P*	Fold value	*P*	Fold value	*P*	Fold value	*P*	Fold value	*P*	Fold value	*P*	climate variables	Fold value	*P*
altitude	0.61	**	0.94	*	1.24	*	1.15	ns	1.20	**	1.23	**	altitude	2.46	***
MAT	0.62	***	0.97	ns	1.10	ns	1.15	*	1.18	*	1.21	***	MAT	2.12	***
MWMT	0.61	***	0.95	*	1.19	*	1.19	***	1.18	**	1.14	*	MWMT	2.02	***
MCMT	0.70	***	0.97	*	1.13	*	1.11	*	1.13	**	1.19	***	MCMT	1.70	***
TD	0.52	***	0.98	ns	1.23	*	1.16	ns	1.14	ns	1.21	**	TD	1.77	***
AHM	0.54	**	0.98	ns	1.08	ns	1.18	ns	1.14	ns	1.23	*	AHM	2.16	**
SHM	0.59	*	1.02	ns	1.10	ns	1.03	ns	1.14	ns	1.17	ns	SHM	2.61	**
DD<0	0.69	**	0.97	ns	1.20	*	1.06	ns	1.15	*	1.20	**	DD<0	2.17	***
DD>5	0.61	***	0.97	ns	1.18	*	1.13	*	1.18	**	1.13	*	DD>5	2.23	***
DD<18	0.64	***	0.97	ns	1.02	ns	1.14	*	1.22	***	1.16	**	DD<18	2.48	***
DD>18	0.61	***	0.99	ns	1.13	ns	1.16	**	1.14	*	1.10	ns	DD>18	2.31	***
NFFD	0.66	***	0.97	ns	1.19	*	1.07	ns	1.21	***	1.15	**	NFFD	1.80	***
Tave_wt	0.65	***	0.97	*	1.07	ns	1.15	**	1.20	***	1.15	**	Tave_wt	2.02	***
Tave_sp	0.64	**	0.99	ns	1.16	ns	1.04	ns	1.25	***	1.14	ns	Tave_sp	2.29	***
Tave_sm	0.60	***	0.96	*	1.10	ns	1.18	**	1.19	***	1.17	**	Tave_sm	1.83	***
Tave_at	0.62	***	0.96	*	1.19	*	1.16	*	1.21	***	1.14	*	Tave_at	2.14	***
MAP	0.58	***	0.95	*	1.05	ns	1.18	*	1.27	***	1.20	**	MAP	1.82	**
PPT_wt	0.59	***	1.00	ns	1.14	ns	1.10	ns	1.09	ns	1.18	**	PPT_wt	1.18	ns
PPT_sp	0.51	**	1.08	*	1.04	ns	1.06	ns	1.00	ns	1.06	ns	PPT_sp	3.14	***
PPT_sm	0.53	***	0.95	*	1.37	***	1.17	*	1.20	*	1.17	*	PPT_sm	2.49	***
PPT_at	0.62	***	0.99	ns	1.08	ns	1.13	*	1.14	*	1.14	*	PPT_at	2.22	***
PCA Axis1	0.55	***	0.95	*	1.21	*	1.14	ns	1.27	***	1.24	**	PCA Axis1	2.47	***
PCA Axis2	0.61	**	1.01	ns	1.09	ns	1.08	ns	1.06	ns	1.22	**	PCA Axis2	1.32	ns
mean	0.60		0.98		1.14		1.13		1.17		1.17			2.12	

Table 2 Enrichment of different classes of sites (a) and of genome-wide spatial differentiation (XtX) (b) in the 0.5% tail of the BF_{mc} distribution of each individual climate variable. See Table 1 for a description of the climate variables.

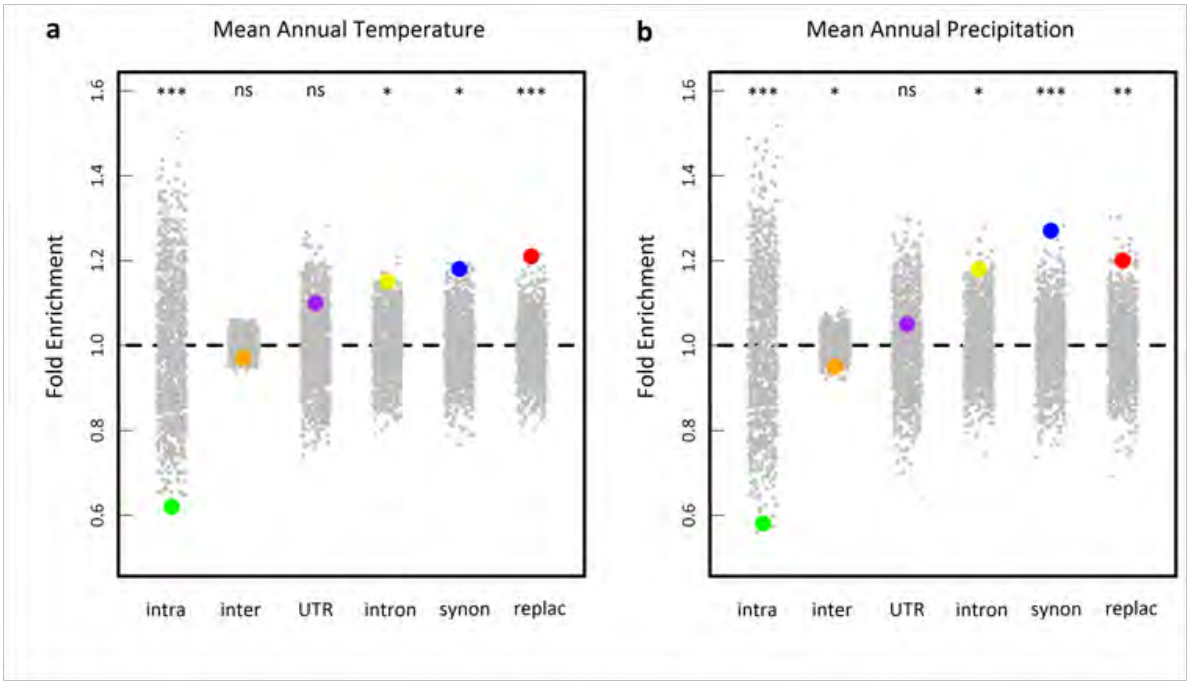


Fig. 5 Enrichment analysis of different classes of sites in the 0.5% upper tail of the BF_{mc} distribution of two climate variables. Enrichments shown are relative to the proportion of each class of SNPs in the genome overall. Large dots represent the enrichment values for each class of sites. Grey dots represent 1000 null permutations of site categories. The horizontal dashed line shows the expected enrichment under the null hypothesis of no enrichment. Enrichments that are significant relative to permutations are denoted by asterisks. ^{ns} non-significant, * $0.05 > P > 0.01$, ** $0.01 > P > 0.001$, *** $P < 0.001$. intra: intragenic SNPs (i.e. SNPs in transposable element gene), inter: intergenic SNPs, UTR: SNPs in 5' and 3' untranslated transcribed regions, intron: intronic SNPs, synon: synonymous SNPs, replac: replacement SNPs (amino-acid changing SNPs and stop codon gained SNPs).

We then tested whether the top SNPs (i.e. the 0.5% upper tail of the BF distribution) significantly overlapped with the 0.5% upper tail of the XtX distribution (i.e. the 0.5% most overly differentiated SNPs among populations). For all climate variables (with the exception of summer precipitation and the second PC), climate-related SNPs were significantly enriched in the 0.5% upper tail of the XtX distribution, with a fold-enrichment ranging from 1.70 for mean coldest month temperature to 3.14 for spring precipitation (Table 2b).

Enrichment in biological processes

Distinct biological processes were significantly associated with a given climate variable (Table 3). For example, processes involved in the responses to abiotic stresses such as heat, sulfate starvation, the abscisic acid (ABA) pathway - a plant hormone pathway

regulating plant growth and response to abiotic stress - and other general biological processes (protein ubiquitination, intracellular signaling, transmembrane transport or gene silencing) were found to be associated with adaptation to altitude. Interestingly, regulation of gene transcription as well as epigenetic mechanisms, both important for cellular regulation, were found as markers for adaptation to temperature. Proteolysis, likely in relationship with membrane protein stability, was also found as a major process for temperature adaptation. For the variables related to precipitations, different developmental processes were identified for precipitation variables, especially when seasons were taken into consideration. More specifically, growth-related processes were identified for the mean annual precipitation variable. On the other hand, different processes either related to photosynthesis and light perception, or to development and the auxin transport pathway, were found to be associated with winter and spring precipitations, respectively. Interestingly, the list of the genes underlying these enriched biological processes (Supporting Information Table S2) revealed a large number of putative or true transcription factors.

		Temperature													Precipitation									
GO term	altitude	MAT	MWMT	MCMT	TD	DD < 0	DD > 5	DD < 18	DD > 18	NFFD	Tave_wt	Tave_sp	Tave_sm	Tave_at	MAP	PPT_wt	PPT_sp	PPT_sm	PPT_at	AHM	SHM	PCA axis1	PCA axis2	No. Climate factors
Protein ubiquitination	3.75											3.17												2
Intracellular signal transduction	4.77																							1
Transmembrane transport	3.95				6.65																			2
Gene silencing by miRNA	9.57												8.14											2
Response to heat	4.89																							1
Cellular response to sulfate starvation	22.59																							1
Abscisic acid-activated signaling pathway	4.28																							1
Nitrate assimilation	11.51		16.83			15.16																		3
Regulation of timing of transition from vegetative to reproductive phase	13.31											13.22												2
Regulation of transcription, DNA-templated	1.92	1.99	2.26			2.16				2.24		1.85										2.19		7
Transcription, DNA-templated												1.84												1
Chromatin silencing						18.63				16.31	15.96													3
DNA methylation				14.42		19.51	14.87	15.15		20.17	17.30													6
Production of siRNA involved in RNA interference						24.48				25.36	21.33													3
RNA-directed DNA methylation						40.38				35.57	35.75													3
Regulation of transcription from RNA polymerase II promoter				12.16																				1
Methylation																					3.82			1
Proteolysis			4.11			4.06	3.11	3.40																4
Proteolysis involved in cellular protein catabolic process		14.27	19.11			13.19	10.32	13.42	10.07	13.12			12.13	13.02								16.20		10
Aging					15.76																8.96			2
Negative regulation of cell proliferation					66.99																			1
Translational initiation					11.02																			1
Pentose-phosphate shunt					31.27																			1
Zinc II ion transport					45.39																			1
Phospholipid metabolic process																							63.82	1
response to water deprivation							3.85																	1
Protein to phosphorylation												5.03												1
Glucosinolate biosynthetic process									11.07					11.34										2
Steroid metabolic process				24.72							24.58													2
Response to hypoxia												10.23												1
Regulation of pollen tube growth										18.80														1
Oligopeptide transport			10.23																					1
Nitrate transport			25.97																					1
Response to brassinosteroid						13.81																		1
Phloem or xylem histogenesis						14.64																		1
Embryo sac development																				7.05				1
Vegetative to reproductive phase transition of meristem			9.70				8.61		10.50					10.55	6.89							8.73		6
RNA splicing															4.80									1
Root development																					4.01			1
Plant-type cell wall organization															6.78									1
Seedling development															26.29									1
Unidimensional cell growth															6.48									1
Growth															14.51					14.55				2
Regulation of organ growth																				45.16				1
Positive regulation of GTPase activity																				38.51				1
Lignin biosynthetic process																				7.70				1
Photosynthesis																9.15								1
Photosynthesis, light harvesting in photosystem I																29.01								1
Negative regulation of flower development																13.37								1
Photoperiodism, flowering																9.67								1
Protein-chromophore linkage																13.21								1
Photomorphogenesis																	8.00							1
Response to nematode																	5.46							1
Auxin polar transport																	7.87			9.22				2
Auxin efflux																	20.00							1
Acropetal auxin transport																	20.39							1
Basipetal auxin transport																	11.87							1
Positive gravitropism																	13.53							1
Stamen development																	15.83							1
Regulation of cell size																	22.72							1
Response to blue light																	8.18							1
Response to red or farred light																	12.82							1
Negative regulation of catalytic activity																	8.01	13.41						2
rRNA methylation																		39.37						1
Hydrogen peroxide catabolic process																		6.09						1
Amino acid transport																		7.49						1
No. Biological processes	10	2	7	3	6	10	5	3	3	7	5	6	2	3	6	5	12	4	0	6	3	3	1	

Table 3 Enrichment of biological processes in the 0.01% tail of the BF_{mc} distribution of each climate variable as well as for the two first axes of the PCA. Only enrichment values with a significant *p*-value < 0.01 are reported (orange squares: ** $P < 0.01$, red squares *** $P < 0.001$). The significance of enrichment was tested against a null distribution using 1000 null permutations. GO term: black, general cellular processes; red, transcriptional and epigenetic processes; orange, proteolysis processes; green, developmental processes; blue: light perception related processes; pink, auxin related pathways; purple, response to stress.

Identification of candidate genes

To identify candidate genes associated with climate variation, we retrieved all the annotated genes located within or overlapping with 61 candidate regions (each supported by at least three top SNPs) (Supporting Information Table S3). By considering a list of 187 candidate genes, a literature survey identified different functions encoded by these genes that could be classified in four main categories: (i) a large number of proteins (n = 41) involved in the regulation of gene expression and/or chromatin accessibility, including many transcriptional factors regulating developmental processes (such as *MYB37*, *TFL1*, *MYB36*, *DTF2*) and abiotic stress response (*WRKY60*, *RAV1*, *DDF1*); (ii) genes involved in developmental processes (n = 16) (such as root (*RSH3*) and trichome development (*ARP3*), seed-seedling transition (*ATHB13*)) and auxin and gibberellin responsive genes (*SAUR38*, *IAA9*, *ARR1*); (iii) genes involved in abiotic (or biotic) stress response (n = 11), including genes involved in cell death regulation (*BAG5*, *metacaspase 8*, *ATG8E*, *MAC5C*); and finally, (iv) ring-type E3 ubiquitin ligases (n = 8) which are known to be required for, or modulate, multiple aspects of biotic and abiotic stress responses (Callis, 2014). Other candidate genes correspond to diverse general functions or to unknown functions.

Discussion

Climate adaptation in a patchy climate environment

In agreement with the influence of three contrasted climates (i.e. oceanic climate, mountain climate and Mediterranean climate) in the south-west of France, up to 27% of climate variation across European locations inhabited by *A. thaliana* was observed in the Midi-Pyrénées region. The presence of mountains in the south and north-east in our sampling area likely explained the steep temperature gradients observed in this study. On the other hand, we observed a mosaic of contrasted precipitation regimes. Therefore, the different spatial grains between temperature and precipitations lead to rugged climate landscapes over very short geographical scales, which in turn better match with the small distance of seed and pollen dispersal expected for a barochorous and selfing plant species such as *A. thaliana*.

In comparison with studies performed at larger geographical scales (Hancock *et al.*, 2011; Lasky *et al.*, 2012), we identified very neat peaks of association with local climate.

Such a pattern is similar to a previous GWAS in *A. thaliana* reporting that the significance level of association peaks for phenotypic traits potentially related to climate adaptation (such as flowering time) was stronger based on regional or local accessions than worldwide or European accessions (Brachi *et al.*, 2013). In our study, two non-exclusive hypotheses can be suggested to explain the relationship between the significance level of association peaks and geographic scale. First, because natural populations of *A. thaliana* have been long considered to have a low genetic variability (likely due its selfing rate close to 98%; Platt *et al.*, 2010), most studies performed at the European or regional scale have been performed using a single accession per population. However, recent studies challenged this view, and many natural populations have been described to be highly genetically variable at both neutral SNPs and polymorphisms associated within natural phenotypic variation (Le Corre, 2005; Picó *et al.*, 2008; Lundemo *et al.*, 2009; Montesinos *et al.*, 2009; Platt *et al.*, 2010; Bomblies *et al.*, 2010; Kronholm *et al.*, 2012; Samis *et al.*, 2012; Brachi *et al.*, 2013; Karasov *et al.*, 2014; Luo *et al.*, 2015). Because pool sequencing has been demonstrated a cost-effective method to infer demography and to identify genetic markers underlying local adaptation in several plant and animal species (Schlötterer *et al.*, 2014), we obtained a representative picture of within-population genetic variation across the genome by sequencing pools of ~16 individuals from each population. Second, in agreement with previous studies reporting that global effects of the demographic evolutionary forces in *A. thaliana* should be limited at a small geographical scale (Nordborg *et al.*, 2005; Platt *et al.*, 2010), we observed a weak geographic pattern of genomic variation among the 168 natural populations. Such a pattern likely alleviated the limitations of GEA analyses often observed at larger geographical scales, which are confounding background produced by population structure, rare alleles and allelic heterogeneity.

Importantly, following methodologies previously developed in *A. thaliana* to identify environment-adaptive genetic loci at the European scale (i.e. climate and herbivore resistance; Hancock *et al.*, 2011; Lasky *et al.*, 2012; Brachi *et al.*, 2015), we found that the SNPs the most associated with climate were significantly enriched in likely functional variants (i.e. non-synonymous variants) and in the extreme tail of spatial differentiation among populations. These clear signatures of selection suggest that climate is an important driver of adaptive genomic variation in *A. thaliana* at a local scale. Although studies reporting the identification of climate adaptive genetic loci at a small geographical scales are still scarce

(Manel *et al.*, 2010; Kubota *et al.*, 2015; Günther *et al.*, 2016; Pluess *et al.*, 2016), there is mounting genomic evidence that microgeographic adaptation to climate is more widespread than is commonly assumed.

Overrepresentation of genes involved in transcriptional mechanisms in the plant functions involved in local adaptation

Using two complementary approaches consisting in the determination of biological processes that were enriched with climate adaptation and the identification of candidate genes associated with climate variation, we were able to show that (i) plant functions involved in local adaptation differ among the climate variables, in particular between altitude, temperature related variables and seasonal precipitations, and (ii) regulation of gene transcription and epigenetic mechanisms are key actors of climate adaptation. Transcription factors (TFs), DNA/chromatin modifying proteins and epigenetic mechanisms were the major functions uncovered by our study. These findings are in agreement with (i) previous studies reporting strong correlations between genome-wide DNA methylation and climate in *A. thaliana* at a worldwide scale (Kawakatsu *et al.*, 2016; Keller *et al.*, 2016), and with (ii) global gene expression studies demonstrating that plant response to environmental constraints relies on a number of distinct transcriptional responses operating spatially, temporally and in combination with other signals like hormones (Coolen *et al.*, 2016). WRKY60, identified here in relation with precipitations in spring and autumn, was shown to control plant response to abscisic acid and abiotic stress through a highly interacting regulatory network shared with WRKY18 and WRKY40 (Chen *et al.*, 2010; Liu *et al.*, 2012). These TFs modulate gene expression in plant stress responses by acting as either transcription activators or repressors. Another interesting candidate gene *HSFA2*, which orchestrates transcriptional dynamics after heat stress in *A. thaliana* (Lämke *et al.*, 2016), acts also as a key component of the Heat Stress Transcription Factors signaling network involved in responses to various environmental stress (Nishizawa-Yokoi *et al.*, 2011). Its implication in the protection against oxidative stress (Zhang *et al.*, 2009) can explain these multiple functions and its identification in relation to the variable altitude. The DDF1 transcriptional activator, found here in relation to temperature variables, upregulates expression of a gibberellin-deactivating gene, GA2ox7, under high-salinity stress in Arabidopsis (Magome *et al.*, 2008) and has been involved in tolerance to

cold, drought, and heat stresses in *A. thaliana* (Kang *et al*., 2011). These examples illustrate the importance of gene expression control by TFs in adaptation to climate variables. Interestingly, some transposable elements (known to be related with gene regulation) have also been identified as candidate genes in this study, in good agreement with recent work demonstrating that the composition and activity of the Arabidopsis mobilome vary greatly among accessions (Quadrana *et al.,* 2016) and that remarkably, loci controlling adaptive responses to the environment are the most frequent transposition targets observed.

Finally, as a consequence of gene expression control and/or through other mechanisms, developmental processes were found to be associated with distinct climate variables in this study. ARR1 (identified for precipitations in spring) is a good example because it has been demonstrated to be involved in low temperature-mediated inhibition of root in conjunction with cytokinin signaling (Zhu *et al.,* 2015), but also in response to drought tolerance (Nguyen *et al.,* 2016). In the same line, ELF6 (Early Flowering 6) modulates flowering time by regulating specific gene expression through interaction with brassinosteroid pathway-specific transcription factors such as BES1 (Yu *et al.,* 2008). Interestingly, this later regulator is related to altitude and temperature variables.

Overall, our study shows an overrepresentation of genes involved in transcriptional mechanisms, including potential epigenetic processes. We also found that developmental processes were associated with different climate variables in conjunction with hormonal pathways. Taken together, these observations suggest that adaptation to local climate in *A. thaliana* mainly involves genome-wide changes in fundamental mechanisms of gene regulation, and other regulatory processes including hormones.

Conclusions

In agreement with the increasing number of phenotypic studies reporting microgeographic adaptation (Richardson *et al.,* 2014), locale climate is an important driver of adaptive genomic variation in *A. thaliana*. This result reinforces the need to choose mapping populations according to the spatial scale of ecological variation at which species are adapted (Bergelson & Roux, 2010). In addition, the identification of climate-adaptive genetic loci at a local scale highlights the importance to include within-population genetic diversity in

ecological niche models for projecting potential species distributional shifts (Valladares *et al.* 2014). The overrepresentation of genes involved in transcriptional mechanisms in the plant functions associated with climate variation is a common pattern at different geographical scales in *A. thaliana*. The candidate genes identified in this study undoubtedly constitute key candidate genes for functional analysis, thereby providing an exciting opportunity to dissect the molecular bases of climate adaptation.

Acknowledgements

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 contribution

L.F., C.B. and F.R. planned and designed the research. L.F. and F.R. conducted fieldwork. L.F. performed climate database searches. C.B. and F.R. performed DNA extraction. C.B., O.B and A.C. generated the sequencing data. C.B. and S.B. performed the bioinformatics analysis. L.F and M.G. performed the genome-environment association analysis. L.F., D.R. and F.R. performed the enrichment analyses and the identification of candidate genes. L.F., C.B., M.G., D.R. and F.R. wrote the manuscript. All authors contributed to the revisions.

References

Abebe TD, Naz AA, Léon J. 2015. Landscape genomics reveal signatures of local adaptation in barley (*Hordeum vulgare* L.). *Frontiers in Plant Science* **6**: 813

Å**gren J, Schemske DW. 2012.** Reciprocal transplants demonstrate strong adaptive differentiation of the model organism *Arabidopsis thaliana* in its native range. *New Phytologist* **194**: 1112–1122

Å**gren J, Oakley CG, McKay JK, Lovell JT, Schemske DW. 2013.** Genetic mapping of adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 21077–21082

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. *The American naturalist* **189**: 463-473

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**: 289-300

Bergelson J, Roux F. 2010. Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nature Reviews Genetics* **11**: 867–879

Bomblies K, Yant L, Laitinen RA, Kim ST, Hollister JD, Warthmann N, Fitz J, Weigel D. 2010. Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *PLoS Genetics* **6**: e1000890

Borcard D, Legendre P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* **153**: 51–68

Borcard D, Legendre P, Avois-Jacquet C, Tuomisto H. 2004. Dissecting the spatial structure of ecological data. *Ecology* **85**: 1826–1832

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*. *Molecular ecology* **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*. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 4032–4037

Callis J. 2014. The ubiquitination machinery of the ubiquitin system. The Arabidopsis book. American Society of Plant Biologists. e0174. doi: 10.1199/tab.0174

Chen IC, Hill JK, Ohlemüller R, Roy DB, Thomas CD. 2011. Rapid range shifts of species associated with high levels of climate warming. *Science* **333**: 1024-1026

Chen H, Lai Z, Junwei Shi1, Xiao Y, Chen Z, Xu X. 2010. Roles of arabidopsis WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress. *BMC Plant Biology* **10**: 281

Chessel D, Dufour AB, Thioulouse J. 2004. The ade4 package-I- One-table methods. R news 4: 5-10

Chevin LM, Lande R, Mace GM. 2010. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS Biology* **8**: e1000357

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide

polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**: 80-92

Coolen S, Proietti S, Hickman R, Davila Olivas NH, Huang PP, VanVerk MC, Van Pelt JA, Wittenberg AHJ, De Vos M, Prins M *et al.* 2016. Transcriptome dynamics of Arabidopsis during sequential biotic and abiotic stresses. *The Plant Journal* **86:** 249-267.

DeWitt TJ, Sih A, Wilson DS. 1998. Costs and limits of phenotypic plasticity. *Trends in Ecology & Evolution* **13**: 77-81

Dray S, Dufour AB, Thioulouse J. 2017. Analysis of ecological data: exploratory and Euclidean methods of environmental sciences. R package version 1.7-6.

Ewers RM, Didham RK. 2006. Confounding factors in the detection of species responses to habitat fragmentation. *Biological Reviews* **81**: 117-142

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

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**: 86-89

Gautier M. 2015. Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics* **201**: 1555–1579

Ghalambor C, McKay JK, Carroll SP, Reznick DN. 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology* **21**: 394–407

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

Günther T, Lampei C, Barilar I, Schmid KJ. 2016. Genomic and phenotypic differentiation of *Arabidopsis thaliana* along altitudinal gradients in the North Italian Alps. *Molecular Ecology* **25**: 3574–3592

Hamann A, Wang T, Spittlehouse DL, Murdock TQ. 2013. A comprehensive, high-resolution database of historical and projected climate surfaces for western North America. *Bulletin of the American Meteorological Society* **94**: 1307-1309

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

Hansen MM, Olivieri I, Waller DM, Nielsen EE, THE GeM WORKING GROUP. 2012. Monitoring adaptive genetic responses to environmental change. *Molecular Ecology* **21**: 1311–1329

Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, Poss ML, Reed LK, Storfer A, Whitlock MC. 2016. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist* **188**: *379–397*

Hoffmann AA, Sgrò CM. 2011. Climate change and evolutionary adaptation. *Nature* **470**: 479-485

Hoffmann MH. 2002. Biogeography of Arabidopsis thaliana (L.) Heynh. (Brassicaceae). *Journal of Biogeography* **29**: 125-134

Kang HG, Kima J, Kimb B, Jeong H, Choia SH, Kima EK, Lee HY, Lim PO. 2011. Overexpression of FTL1/DDF1, an AP2 transcription factor, enhances tolerance to cold, drought, and heat stresses in *Arabidopsis thaliana*. *Plant Science* **180**: 634–641

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

Kawakatsu T, Huang SSC, Jupe F, Saski E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y *et al.* 2016. Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* **166**: 492-505.

Keller, TE, Lasky JR, Yi SV. 2016. The multivariate association between genomewide DNA methylation and climate across the range of *Arabidopsis thaliana*. *Molecular Ecology* **25**: 1823-1837.

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 Research* **22**: 568–576

Kronholm I, Picó FX, Alonso-Blanco C, Goudet J, de Meaux J. 2012. Genetic basis of adaptation in *Arabidopsis thaliana*: local adaptation at the seed dormancy QTL *DOG1*. *Evolution* **66**: 2287-2302

Kubota S, Iwasaki T, Hanada K, Nagano AJ, Fujiyama A, Toyoda A, Sugano S, Suzuki Y, Hikosaka K, Ito M *et al.* 2015. A genome scan for genes underlying microgeographic-scale local adaptation in a wild Arabidopsis species. *PLoS Genetics* **11**: e1005488

Lämke J, Brzezinka K, Bäurle I. 2016. HSFA2 orchestrates transcriptional dynamics after heat stress in *Arabidopsis thaliana*. *Transcription* **7**:111-4

Lande R. 2009. Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of Evolutionary Biology* **22**: 1435-1446

Lasky JR, Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH. 2012. Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Molecular Ecology* **21**: 5512–5529

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. *Science Advances* **1**: e1400218

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

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079

Li Y, Cheng R, Spokas KA, Palmer AA, Borevitz JO. 2014. Genetic variation for life history sensitivity to seasonal warming in Arabidopsis thaliana. *Genetics* **196**: 569–577

Liu ZQ, Yan L, Wu Z, Mei C, Lu K, Yu YT, Liang S, Zhang XF, Wang XF, Zhang DP. 2012. Cooperation of three WRKY-domain transcription factors WRKY18, WRKY40, and WRKY60 in repressing two ABA responsive genes *ABI4* and *ABI5* in Arabidopsis. *Journal of Experimental Botany* **63**: 6371–6392

Lundemo S, Falahati-Anbaran M, Stenøien HK. 2009. Seed banks cause elevated generation times and effective population sizes of *Arabidopsis thaliana* in northern Europe. *Molecular Ecology* **18**: 2798–2811

Luo Y, Widmer A, Karrenberg S. 2015. The roles of genetic drift and natural selection in quantitative trait divergence along an altitudinal gradient in *Arabidopsis thaliana*. *Heredity* **114**: 220–228

Magome H, Yamaguchi S, Hanada A, Kamiya Y, Oda K. 2008. The DDF1 transcriptional activator upregulates expression of a gibberellin-deactivating gene, *GA2ox7*, under high-salinity stress in Arabidopsis. *Plant Journal* **56**: 613-26

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R. 2010. Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular Ecology* **19**: 3824–3835

Manel S, Perrier C, Pratlong M, Abi-Rached I, Paganini J, Pontarotti P, Aurelle D. 2016. Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Molecular Ecology* **25**: 170–184

Méndez-Vigo B, Picó FX, Ramiro M, Martinez-Zapater JM, Alonso-Blanco C. 2011. Altitudinal and climatic adaptation is mediated by flowering traits and *FRI, FLC,* and *PHYC* genes in Arabidopsis. *Plant Physiology* **157**: 1942–1955

Montesinos A, Tonsor SJ, Alonso-Blanco C, Picó FX. 2009. Demographic and genetic patterns of variation among populations of *Arabidopsis thaliana* from contrasting native environments. *PLoS One* **4**: e7213

Montesinos-Navarro A, Wig J, Picó FX, Tonsor SJ. 2011. *Arabidopsis thaliana* populations show clinal variation in a climatic gradient associated with altitude. *New Phytologist* **189**: 282–294

Nishizawa-Yokoi A, Nosaka R, Hayashi H, Tainaka H, Maruta T, Tamoi M, Ikeda M, Ohme-Takagi M, Yoshimura K, Yabuta Y, Shigeoka S. 2011. HsfA1d and HsfA1e involved in the transcriptional regulation of HsfA2 function as key regulators for the Hsf signaling network in response to environmental stress. *Plant Cell Physiology* **52**: 933-45

Nguyen KH, Van Hab C, Nishiyama R, Watanabe Y, Leyva-González MA, Fujitad Y, Tran UT, Li W, Tanaka M, Seki M, Schallerg GE, Herrera-Estrellah L, Phan Tran LS. 2016. Arabidopsis type B cytokinin response regulators ARR1, ARR10, and ARR12 negatively regulate plant responses to drought. *Proceedings of the National Academy of Science USA* **113:** 3090–3095

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R et al. 2005. The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biology* **3**: e196

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens HH, Wagner H. 2016. vegan: Community Ecology Package. R package version 2.3-5.

Orlowsky B, Seneviratne SI. 2012. Global changes in extreme events: regional and seasonal dimension. *Climate change* **110**: 669–696

Pecl GT Araújo MB, Bell JD, Blanchard J, Bonebrake TC, Chen IC, Clark TD, Colwell RK, Danielsen F, Evengård B et al. 2017. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* **355**: eaai9214

Picó FX. 2012. Demographic fate of *Arabidopsis thaliana* cohorts of autumn- and spring-germinated plants along an altitudinal gradient. *Journal of Ecology* **100**: 1009–1018

Picó FX, Mendez-Vigo B, Martinez-Zapater JM, Alonso-Blanco C. 2008. Natural genetic variation of *Arabidopsis thaliana* is geographically structured in the Iberian Peninsula. *Genetics* **180**: 1009–1021

Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, Agren J, Bossdorf O, Byers D, Donohue K Dunning M et al. 2010. The scale of population structure in *Arabidopsis thaliana*. *PLoS Genetics* **6**: e1000843

Pluess AR, Frank A, Heiri C, Lalagüe H, Vendramin GG, Oddou-Muratorio S. 2016. Genome–environment association study suggests local adaptation to climate at the regional scale in *Fagus sylvatica*. *New Phytologist* **210**: 589–601

Price TD, Qvarnström A, Irwin DE. 2003. The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society B* **270**: 1433–1440

Quadrana L, Bortolini Silveira A, Mayhew GF, LeBlanc C, Martienssen RA, Jeddeloh JA, Colot V. 2016. The *Arabidopsis thaliana* mobilome and its impact at the species level. *eLife* **5**: 15716

Ramette A, Tiedje M. 2007. Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 2761–2766

Rellstab C, Gugerli F,Eckert AJ, Hancock AM, Holderegger R. 2015. A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology* **24**: 4348–4370

Richardson JL, Urban MC, Bolnick DI, Skelly DK. 2014. Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology & Evolution* **29**: 165-176

Samis KE, Murren CJ, Bossdorf O, Donohue K, Fenster CB, Malmberg RL, Purugganan MD, Stinchcombe JR. 2012. Longitudinal trends in climate drive flowering time clines in North American *Arabidopsis thaliana*. *Ecology and Evolution* **2**: 1162–1180

Schlötterer C, Tobler R, Kofler R, Nolte V. 2014. Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics* **15**: 749-763

The Gene Ontology Consortium. 2008. The Gene Ontology project in 2008. *Nucleic Acids Research* **36**: D440-D444

Valladares F, Matesanz S Guilhaumon F, Araujo MB, Balaguer L, Benito-Garzon M, Cornwell W, Gianoli E, van Kleunen M, Naya DE *et al*. 2014. The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. *Ecology Letters* **17**: 1351–1364

van Kleunen M, Fischer M. 2005. Constraints on the evolution of adaptive phenotypic plasticity in plants. *New Phytologist* **166**: 49–60

Vidigal DS, Marques ACSS, Willems LAJ, Buijs G, Méndez-Vigo B, Hilhorst HWM, Bentsink L, Picó FX, Alonso-Blanco C. 2016. Altitudinal and climatic associations of seed dormancy and flowering traits evidence adaptation of annual life cycle timing in *Arabidopsis thaliana*. *Plant, Cell & Environment* **39**: 1737-1748

Wang Z, Wang L, Liu Z, Li Y, Liu Q, Liu B. 2016. Phylogeny, seed trait, and ecological correlates of seed germination at the community level in a degraded sandy grassland. *Frontiers in Plant Science* **7**: 1-10

Wilczek AM, Cooper MD, Korves TM, Schmitt J. 2014. Lagging adaptation to warming climate in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 7906–7913

Yoder JB, Stanton-Geddes J, Zhou P, Briskine R, Young ND, Tiffin P. 2014. Genomic signature of adaptation to climate in *Medicago truncatula*. *Genetics* **196**: 1263-1275

Yu X1, Li L, Li L, Guo M, Chory J, Yin Y. 2008. Modulation of brassinosteroid-regulated gene expression by Jumonji domain-containing proteins ELF6 and REF6 in Arabidopsis. *Proceedings of the National Academy of Sciences USA*. **105**: 7618-23

Zhang L, Li Y, Xing D, Gao C. 2009. Characterizationof mitochondrial dynamics and subcellular localization of ROS reveal that HsfA2 alleviates oxidative damage caused by heat stress in Arabidopsis. *Journal of Experimental Botany* **60**: 2073-91

Zhu J, Zhang KX, Wang WS, Gong W, Liu WC, Chen HG, Xu HH, Lu YT. 2015. Low temperature inhibits root growth by reducing auxin accumulation *via* ARR1/12. *Plant Journal* **56**:613-26

Supplementary information

A genomic map of adaptation to local climate in *Arabidopsis thaliana*

Supporting Information

A genomic map of adaptation to local climate

in *Arabidopsis thaliana*

Léa Frachon,[¶,1] Claudia Bartoli,[¶,1] Sébastien Carrère,[1] Mathieu Gautier,[2] Dominique Roby[1] and Fabrice Roux[*,1]

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

[2] INRA, UMR CBGP (Centre Biologie pour la Gestion des Populations), Campus International de Baillarguet, Montferrier-sur-Lez and IBC (Institut de Biologie Computationnelle), Montpellier, France

[¶] These authors contributed equally to this work.

[*] To whom correspondence should be addressed E-mail: fabrice.roux@inra.fr

3 Supplementary Tables

6 Supplementary Figures

Table S1. Names and GPS coordinates (expressed in decimal degrees) of the 168 populations.

Population	Town	Latitude	Longitude	Altitude	No. of collected plants[1]	No. of pooled plants[2]
AMBR-A	Ambres	43.7332	1.8239	136	17	16
ANGE-A	Saint Angel, Salvagnac	43.9120	1.6566	169	17	16
ANGE-B	Saint Angel, Salvagnac	43.9121	1.6569	168	17	16
AULO-A	Aulon	43.1906	0.8158	339	17	15
AURE-B	Aureville	43.4780	1.4522	197	17	16
AUZE-A	Auzeville	43.5278	1.4916	200	17	16
AXLE-A	Ax les Thermes	42.7242	1.8340	734	17	10
AXLE-B	Ax les Thermes	42.7246	1.8335	733	17	16
BACC-B	Baccarets (Cintegabelle)	43.3122	1.5152	201	17	16
BACC-C	Baccarets (Cintegabelle)	43.3119	1.5155	201	8	8
BACC-D	Baccarets (Cintegabelle)	43.3119	1.5156	201	9	8
BACC-E	Baccarets (Cintegabelle)	43.3119	1.5157	201	8	8
BACC-F	Baccarets (Cintegabelle)	43.3119	1.5155	201	9	8
BAGNB-A	Bagnères de Bigore	43.0757	0.1518	532	17	16
BAGNB-B	Bagnères de Bigore	43.0765	0.1515	530	17	16
BANI-A	Banios	43.0429	0.2347	488	5	5
BANI-B	Banios	43.0436	0.2343	487	17	16
BANI-C	Banios	43.0436	0.2343	487	17	16
BARA-B	Baraqueville	44.2697	2.4263	756	17	16
BARA-C	Baraqueville	44.2708	2.4276	756	17	16
BARC-A	Barcugnan	43.3620	0.3877	218	17	16
BARR-A	Barry le Cas (Caylus)	44.2024	1.7675	174	17	14
BAZI-A	Baziège	43.4536	1.6207	166	17	16
BELC-A	Belcastel	44.3875	2.3361	415	17	16
BELC-B	Belcastel	44.3875	2.3368	412	17	16
BELC-C	Belcastel	44.3892	2.3366	453	17	15
BELL-A	Belleserre	43.7903	1.1065	229	17	16
BERNA-A	Bernac-dessus	43.1622	0.1114	394	17	16
BESS-A	Bessuéjouls	44.5264	2.7301	352	17	16
BOULO-A	Boulogne-sur-Gesse	43.2891	0.6398	339	17	16
BROU-A	Brousse-le-château	43.9993	2.6217	281	17	16
BROU-B	Brousse-le-château	44.0331	2.6387	678	17	14
BROU-C	Brousse-le-château	44.0333	2.6387	678	17	16
BULA-A	Bulan	43.0398	0.2773	491	17	16
BULE-B	Buleix (Soulan)	42.9106	1.2481	525	17	12
CAMA-C	Camarès	43.8249	2.8817	397	17	16
CAMA-D	Camarès	43.8237	2.8810	393	17	15
CAMA-E	Camarès	43.8249	2.8817	397	17	16
CAPE-A	Lacappelle - Ségalar	44.1085	1.9902	322	17	16
CARL-A	Carla-bayle	43.1511	1.3923	385	17	16
CASS-A	Cassagne-Begontes	44.1765	2.5182	616	17	16
CAST-A	Castelginset	43.6985	1.4279	126	17	16
CASTI-A	Castillon en Cousserans	42.9205	1.0341	574	17	16
CAZA-B	Cazaux-Fréchet	42.8315	0.4201	1140	17	16
CEPE-A	Cepet	43.7552	1.4360	125	17	16
CERN-A	Saint-Rome-de-Cernon	44.0119	2.9665	406	17	15
CERN-B	Saint-Rome-de-Cernon	44.0147	2.9679	407	17	16

[1] number of plants collected randomly in each population, [2] number of plants used for DNA extraction in the Pool-Seq approach.

Table S1 (continued)

Population	Town	Latitude	Longitude	Altitude	No. of collected plants[1]	No. of pooled plants[2]
CHEI-A	Chein-dessus	43.0137	0.8671	603	17	16
CIER-A	Cier sur Luchon	42.8533	0.6020	594	17	16
CIER-B	Cier de Luchon	42.8600	0.6004	667	17	16
CIER-C	Cier de Luchon	42.8602	0.6011	656	17	16
CINT-A	Cintegabelle	43.3055	1.5204	204	17	16
CINT-B	Cintegabelle	43.3056	1.5207	204	17	16
CLAR-A	Saint Clar-de-Rivière	43.4648	1.2190	218	17	16
CLAR-B	Saint Clar-de-Rivière	43.4653	1.2186	218	17	16
CLAR-C	Saint Clar-de-Rivière	43.4641	1.2180	218	17	16
COLO-A	Colombiès	44.3469	2.3402	670	17	16
COLO-B	Colombiès	44.3477	2.3397	664	17	16
COLO-C	Colombiès	44.3481	2.3397	664	17	16
COMT-A	Villecomtal	44.5407	2.6022	558	17	16
CRAN-A	Cransac (Aubin)	44.5298	2.2605	318	17	13
DAMI-A	Damiatte	43.6545	1.9776	143	8	8
DAMI-B	Damiatte	43.6545	1.9776	143	13	12
DAMI-C	Damiatte	43.6545	1.9776	143	13	12
DECA-A	Châteaude Cas (Espinas)	44.1999	1.7719	234	17	16
DIEU-A	Ville-Dieu-du-temple	44.0598	1.2210	91	17	14
ESPE-B	Esperausses	43.6933	2.5346	573	17	16
FAYA-A	Fayet	43.8021	2.9517	457	17	16
FERR-A	Ferrières	43.6577	2.4437	541	17	16
GAIL-A	Gaillac	43.9089	1.9006	144	17	16
GAIL-B	Gaillac	43.9090	1.9011	144	17	16
GREZ-A	Grézian	42.8769	0.3497	740	17	16
JACO-A	Jacoy (Boussenac)	42.9058	1.4065	990	17	10
JACO-C	Jacoy (Boussenac)	42.9058	1.4065	990	17	16
JULI-A	Saint Julien de Malmont (St Cyprien de Dourdou)	44.5226	2.3635	407	17	16
JUZE-A	Juzes	43.4488	1.7905	235	17	16
JUZET-A	Juzet d'Izaut	42.9777	0.7564	571	17	16
JUZET-B	Juzet d'Izaut	42.9774	0.7555	576	17	16
JUZET-C	Juzet d'Izaut	42.9774	0.7555	576	17	16
LABA-A	Labarthe-sur-Lèze	43.4516	1.4005	165	17	16
LABA-B	Labarthe-sur-Lèze	43.4509	1.4012	165	17	16
LABA-C	Labarthe-sur-Lèze	43.4515	1.3994	165	17	16
LABA-D	Labarthe-sur-Lèze	43.4580	1.3811	172	17	16
LABAS-A	La bastide de Sérou	43.0084	1.4200	397	17	16
LABAS-B	La bastide de Sérou	43.0087	1.4201	396	17	16
LABR-A	Labruguière	43.5312	2.2626	206	17	16
LACR-A	Lacraste (Montgauch)	42.9999	1.0757	497	17	16
LACR-C	Lacraste (Montgauch)	43.0002	1.0756	493	17	16
LAGR-A	Lagraulhet St Nicolas	43.7953	1.0738	264	17	16
LAMA-A	Lamasquère	43.4874	1.2436	182	17	16
LAMA-B	Lamasquère	43.4797	1.2416	182	17	16
LANT-B	Lanta	43.5649	1.6524	246	17	16
LANT-C	Lanta	43.5648	1.6520	246	13	12

Table S1 (continued)

Population	Town	Latitude	Longitude	Altitude	No. of collected plants[1]	No. of pooled plants[2]
LANT-D	Lanta	43.5648	1.6520	246	13	12
LAUZ-A	Lauzerte	44.2561	1.1405	176	17	15
LECT-A	Lectoure	43.9117	0.6297	85	17	16
LECT-B	Lectoure	43.9117	0.6297	85	17	16
LESP-A	Les pujols	43.0942	1.7200	290	17	16
LOUB-A	Loubens-Lauragais	43.5743	1.7860	220	13	12
LOUB-B	Loubens-Lauragais	43.5746	1.7857	217	12	11
LUNA-A	Lunax	43.3397	0.6898	283	17	16
LUZE-A	Luzenac (Garanou)	42.7647	1.7530	584	17	16
LUZE-B	Luzenac (Garanou)	42.7644	1.7536	584	17	16
LUZE-D	Luzenac (Garanou)	42.7644	1.7536	584	17	16
LUZE-E	Luzenac (Garanou)	42.7647	1.7530	584	17	16
MARS-A	Glaciane (Marsans)	43.6625	0.7183	196	17	16
MARS-B	Glaciane (Marsans)	43.6625	0.7183	196	17	16
MART-A	Martres Tolosane	43.2021	1.0110	266	17	16
MASS-A	Masseube	43.4375	0.5793	204	17	16
MAZA-A	Mazamet	43.4978	2.3754	242	17	16
MEDA-A	Saint Medard	43.4905	0.4614	180	17	16
MERE-A	Merens-les-Vals	42.6566	1.8362	1080	17	16
MERE-B	Merens-les-Vals	42.6565	1.8362	1080	17	16
MERV-A	Merville	43.7204	1.2968	156	17	16
MERV-B	Merville	43.7251	1.2476	123	17	16
MONB-A	Monblanc	43.4653	0.9863	231	17	16
MONE-A	Monestiès	44.1154	2.0947	405	17	16
MONF-A	Monferran-Savès	43.6163	0.9724	209	17	16
MONT-A	Montans	43.8522	1.8743	156	17	16
MONT-B	Montans	43.8527	1.8735	156	17	16
MONTB-A	Montbrun bocage	43.1305	1.2699	281	17	16
MONTG-B	Montgaillard	43.1273	0.1107	437	17	16
MONTG-D	Montgaillard	43.1277	0.1106	437	17	16
MONTI-A	Montiès	43.3894	0.6728	313	17	16
MONTI-B	Montiès	43.3839	0.6726	313	17	16
MONTI-D	Montiès	43.3839	0.6726	313	17	16
MONTM-A	Montmajou (Cier de Luchon)	42.8616	0.5959	792	17	16
MONTM-B	Montmajou (Cier de Luchon)	42.8612	0.5969	774	17	16
MOUL-A	Moularès	44.0898	2.2961	428	17	15
NAUV-A	Nauviale	44.5208	2.4274	259	17	16
NAUV-B	Nauviale	44.5204	2.4271	257	17	16
NAUV-C	Nauviale	44.5204	2.4272	259	17	16
NAYR-A	Le Nayrac (Cassagnes-Bégontes)	44.1614	2.5447	631	17	16
NAZA-A	Saint-Pierre-de-Najac (Miramont de Quercy)	44.2203	1.0650	93	17	16
PAMP-A	Pampelonne	44.1249	2.2555	317	17	16
PAMP-B	Pampelonne	44.1249	2.2552	317	17	16
PANA-C	Villefrance de Panat	44.0789	2.7111	767	17	16
PASD-B	Pas du loup (Camarès)	43.8118	2.8717	603	17	16

Table S1 (continued)

Population	Town	Latitude	Longitude	Altitude	No. of collected plants[1]	No. of pooled plants[2]
PREI-A	Preignan	43.7179	0.6233	120	17	16
PUYM-B	Puymaurin	43.3729	0.7657	278	17	16
RADE-A	Sainte Radegonde	44.3452	2.6208	622	17	11
RAYR-A	Rayret (Cassagne-Begontes)	44.1960	2.4932	555	17	16
RAYR-B	Rayret (Cassagne-Begontes)	44.1960	2.4931	555	17	16
REAL-A	Réalmont	43.8317	2.2016	239	17	16
ROME-A	Saint-Rome-du-tarn	44.0416	2.9096	446	17	16
ROQU-B	Roquecourbe	43.6679	2.2902	261	17	16
SALE-A	Saleich	43.0250	0.9660	413	17	16
SALV-A	Saint-Salvy-de-la-Balme	43.6026	2.3634	473	17	15
SAMA-A	Samatan	43.4943	0.9239	179	17	16
SAUB-A	Saubens	43.4649	1.3651	164	17	16
SAUB-B	Saubens	43.4741	1.3642	164	17	16
SAUB-C	Saubens	43.4756	1.3676	164	17	16
SAUR-A	Saurat	42.8898	1.4852	1023	17	16
SEIS-A	Seissan	43.4873	0.5888	193	17	16
SIMO-A	Simorre	43.4494	0.7346	204	17	16
SORE-A	Sorèze	43.4526	2.0725	293	17	16
TARN-C	Villemur-sur-Tarn	43.8533	1.5020	99	17	16
THOM-A	Saint Thomas	43.5140	1.0829	322	17	16
VALE-A	Valence d'Albiegeois	44.0223	2.4034	459	17	15
VICT-B	Saint Victor et Melvieu	44.0522	2.8340	629	17	16
VICT-C	Saint Victor et Melvieu	44.0522	2.8340	629	17	16
VIEL-A	Vielmur sur Agout	43.6238	2.0896	155	17	16
VILLA-A	Villate	43.4582	1.3810	169	17	16
VILLE-A	Villenouvelle	43.4398	1.6710	188	17	16
VILLE-B	Villenouvelle	43.4403	1.6696	187	17	16
VILLE-C	Villenouvelle	43.4400	1.6701	188	17	16
VILLE-D	Villenouvelle	43.4397	1.6707	188	17	16
VILLEM-A	Villembits	43.2738	0.3212	326	17	16

Table S2 List of the genes underlying the enriched biological processes for each climate variable.

Available online (https://lipm-browsers.toulouse.inra.fr/pub/Frachon2017-PHD/, login: reviewersPHD, password: kryGhehayd4).

Table S3 List of the annotated genes located within or overlapping with 61 candidate regions. 'chromosome' and 'position' stands for the physical positions of the 835 unique top SNPs. Green and yellow blocks delimit the 61 candidate regions, each being supported by at least three top SNPs successively separated by less than 10kb. 'No climate factors': number of climate factors associated with a top SNP. 'Identity of climate factors': see Table 1 for a description of the climate variables. 'Atg number': Atg numbers in red correspond to genes underlying enriched biological process (see Table S2). 'Locus name': white, regulation of gene expression and/or chromatin accessibility; green, genes involved in developmental processes; red, genes involved in abiotic (or biotic) stress response; purple, ring-type E3 ubiquitin ligases.

Available online (https://lipm-browsers.toulouse.inra.fr/pub/Frachon2017-PHD/, login: reviewersPHD, password: kryGhehayd4).

Figure S1. Climatic variation at different geographical scales for all the climate variables (with the exception of altitude). See Table 1 for a description of the climate variables. 'Eur': climate variation among 426 European locations without considering locations in France, 'Fr' climate variation among 95 French locations without considering locations in the Midi-Pyrénées region, 'MP climate variation among 168 locations in the Midi-Pyrénées region.

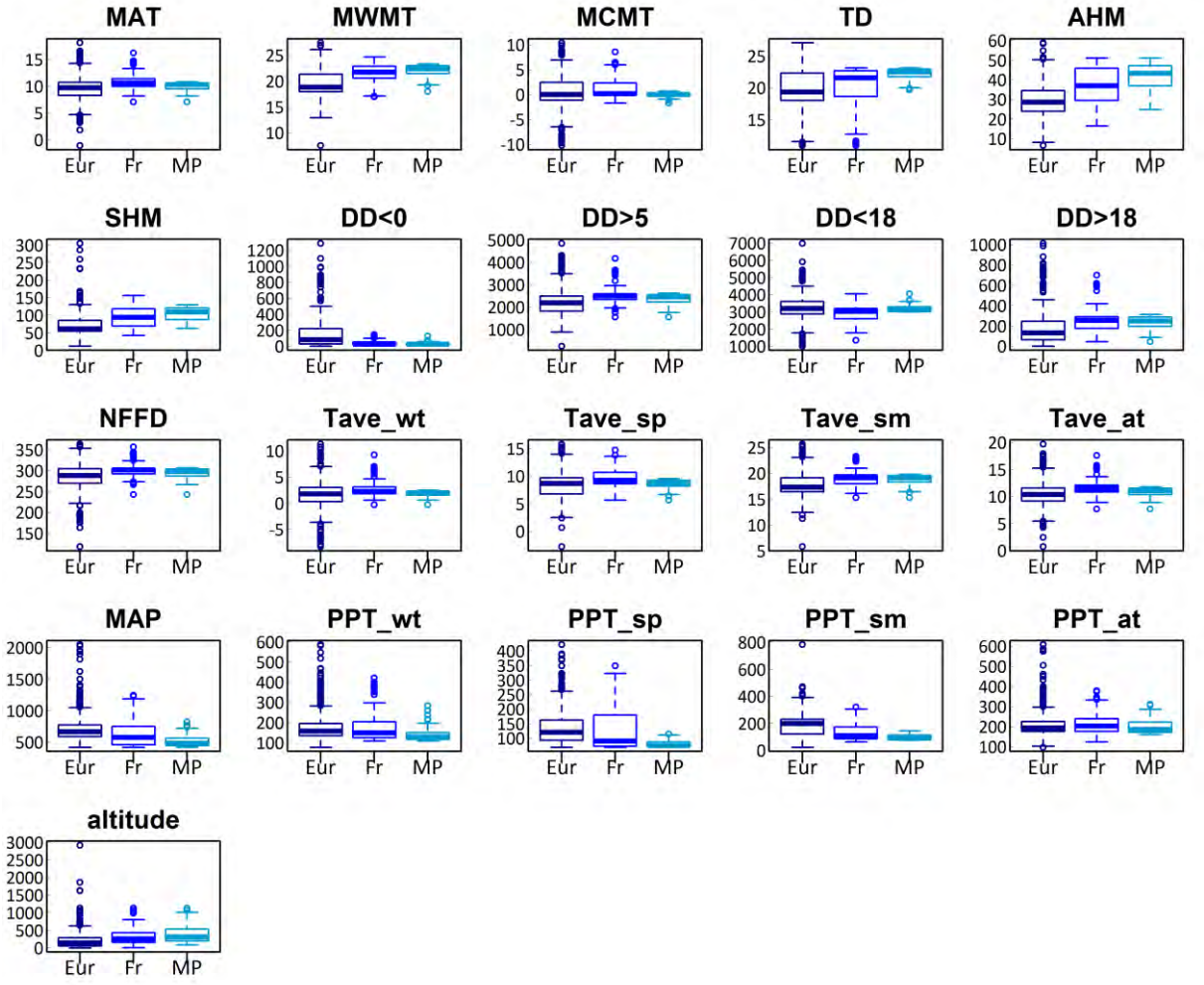


Figure S2. Correlation matrix among the 21 climate variables. Above diagonal: values of Spearman's *rho*. Below diagonal: levels of significance values of Spearman's *rho* after a false discovery rate (FDR) correction at the nominal level of 5%. See Table 1 for a description of the climate variables.

	altitude	MAT	MWMT	MCMT	TD	MAP	AHM	SHM	DD<0	DD>5	DD<18	DD>18	NFFD	Tave_wt	Tave_sp	Tave_sm	Tave_at	PPT_wt	PPT_sp	PPT_sm	PPT_at
altitude		-0.92	-0.87	-0.68	-0.80	0.79	-0.84	-0.88	0.91	-0.94	0.94	-0.89	-0.85	-0.81	-0.97	-0.88	-0.91	0.42	0.85	0.76	0.82
MAT	0		0.97	0.80	0.89	-0.77	0.84	0.89	-0.96	0.99	-0.99	0.98	0.95	0.92	0.97	0.98	1.00	-0.46	-0.76	-0.78	-0.80
MWMT	0	0		0.69	0.96	-0.75	0.81	0.91	-0.91	0.98	-0.94	0.99	0.89	0.84	0.91	1.00	0.97	-0.41	-0.74	-0.79	-0.78
MCMT	0	0	0		0.50	-0.57	0.63	0.56	-0.80	0.75	-0.83	0.71	0.88	0.93	0.78	0.71	0.80	-0.52	-0.52	-0.48	-0.58
TD	0	0	0	5.30E-12		-0.72	0.77	0.90	-0.81	0.92	-0.83	0.95	0.78	0.69	0.82	0.95	0.89	-0.34	-0.71	-0.79	-0.76
MAP	0	0	0	1.18E-15	0		-0.99	-0.83	0.79	-0.78	0.76	-0.79	-0.75	-0.67	-0.79	-0.72	-0.80	0.81	0.80	0.79	0.98
AHM	0	0	0	0	0	0		0.87	-0.85	0.84	-0.83	0.84	0.81	0.74	0.85	0.79	0.85	-0.77	-0.83	-0.82	-0.97
SHM	0	0	0	1.94E-15	0	0	0		-0.87	0.92	-0.87	0.92	0.79	0.73	0.88	0.90	0.90	-0.42	-0.87	-0.94	-0.83
DD<0	0	0	0	0	0	0	0	0		-0.95	0.96	-0.93	-0.93	-0.93	-0.94	-0.91	-0.96	0.50	0.77	0.80	0.80
DD>5	0	0	0	0	0	0	0	0	0		-0.98	0.98	0.93	0.88	0.97	0.99	0.99	-0.43	-0.79	-0.79	-0.81
DD<18	0	0	0	0	0	0	0	0	0	0		-0.95	-0.95	-0.94	-0.99	-0.95	-0.98	0.47	0.78	0.75	0.79
DD>18	0	0	0	0	0	0	0	0	0	0	0		0.91	0.85	0.93	0.99	0.98	-0.47	-0.75	-0.80	-0.81
NFFD	0	0	0	0	0	0	0	0	0	0	0	0		0.97	0.91	0.90	0.96	-0.55	-0.67	-0.68	-0.78
Tave_wt	0	0	0	0	0	0	0	0	0	0	0	0	0		0.89	0.85	0.92	-0.50	-0.64	-0.63	-0.69
Tave_sp	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0.92	0.96	-0.47	-0.83	-0.74	-0.82
Tave_sm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0.97	-0.38	-0.73	-0.77	-0.75
Tave_at	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		-0.50	-0.76	-0.79	-0.82
PPT_wt	1.05E-08	2.93E-10	2.47E-08	6.59E-13	5.50E-06	0	0	2.07E-08	6.70E-12	4.69E-09	2.11E-10	1.48E-10	6.22E-15	5.79E-12	7.83E-11	4.50E-07	5.86E-12		0.45	0.40	0.78
PPT_sp	0	0	0	9.31E-13	0	0	0	0	0	0	0	0	0	0	0	0	0	1.74E-09		0.77	0.79
PPT_sm	0	0	0	3.57E-11	0	0	0	0	0	0	0	0	0	0	0	0	0	1.23E-07	0		0.75
PPT_at	0	0	0	3.17E-16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Figure S3. Climate variation among the 168 natural populations of *A. thaliana* collected in the Midi-Pyrénées region. (**a**) Factor loading plot resulting from a principal component analysis. Factor 1 and factor 2 explained 81.18% and 9.30% of total climate variance in the Midi-Pyrénées region. See Table 1 for a description of the climate variables. (**b**) Position of the 168 populations in the climate space of *A. thaliana* in the Midi-Pyrénées region.

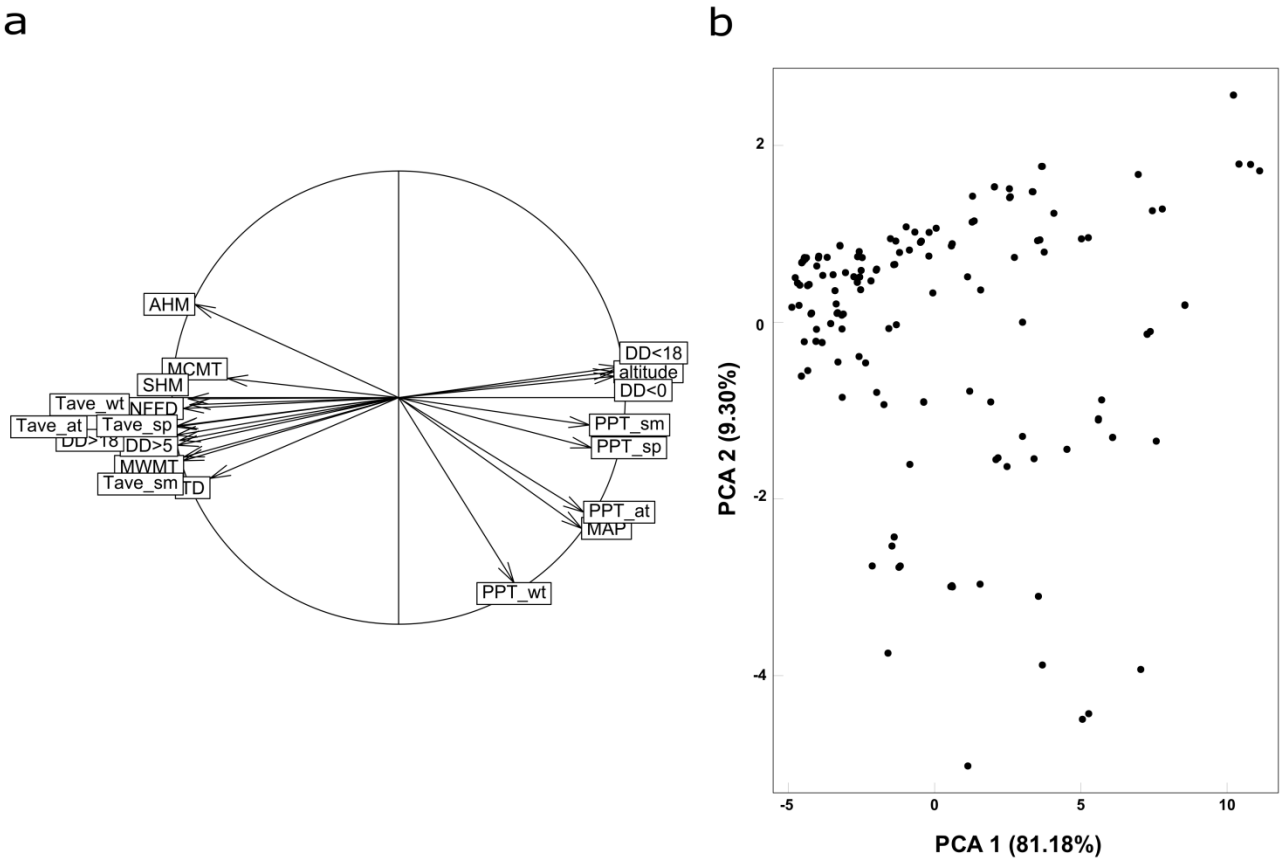


Figure S4. Spatial grains of climate variation based on multiple regressions of each of the 21 climate variables on 82 Principal Coordinates of Neighbour Matrices (PCNM) components. The significant regression coefficients are colored by a blue gradient after a false discovery rate (FDR) correction at the nominal level of 5% (light blue * $P < 0.05$, medium blue ** $P < 0.01$, dark blue *** $P < 0.001$). The PCNM components for which no significant regression coefficient was detected are not represented. The 82 PCNM components were arbitrarily divided according to three spatial scales: large (PCNM 1 to PCNM 27), intermediate (PCNM 28 to PCNM 54) and small (PCNM 55 to PCNM 82), delimited by thicker vertical lines. See Table 1 for a description of the climate variables.

Figure S5. Spatial scale of genomic variation among the 168 populations of *A. thaliana*. (**a**) Singular value decomposition (SVD) run on the covariance-variance matrix obtained with the first sub-data set of 51,208 SNPs. (**b**) and (**c**) Geographic map of the coordinates of the 168 populations on the first and second Principal Components, respectively.

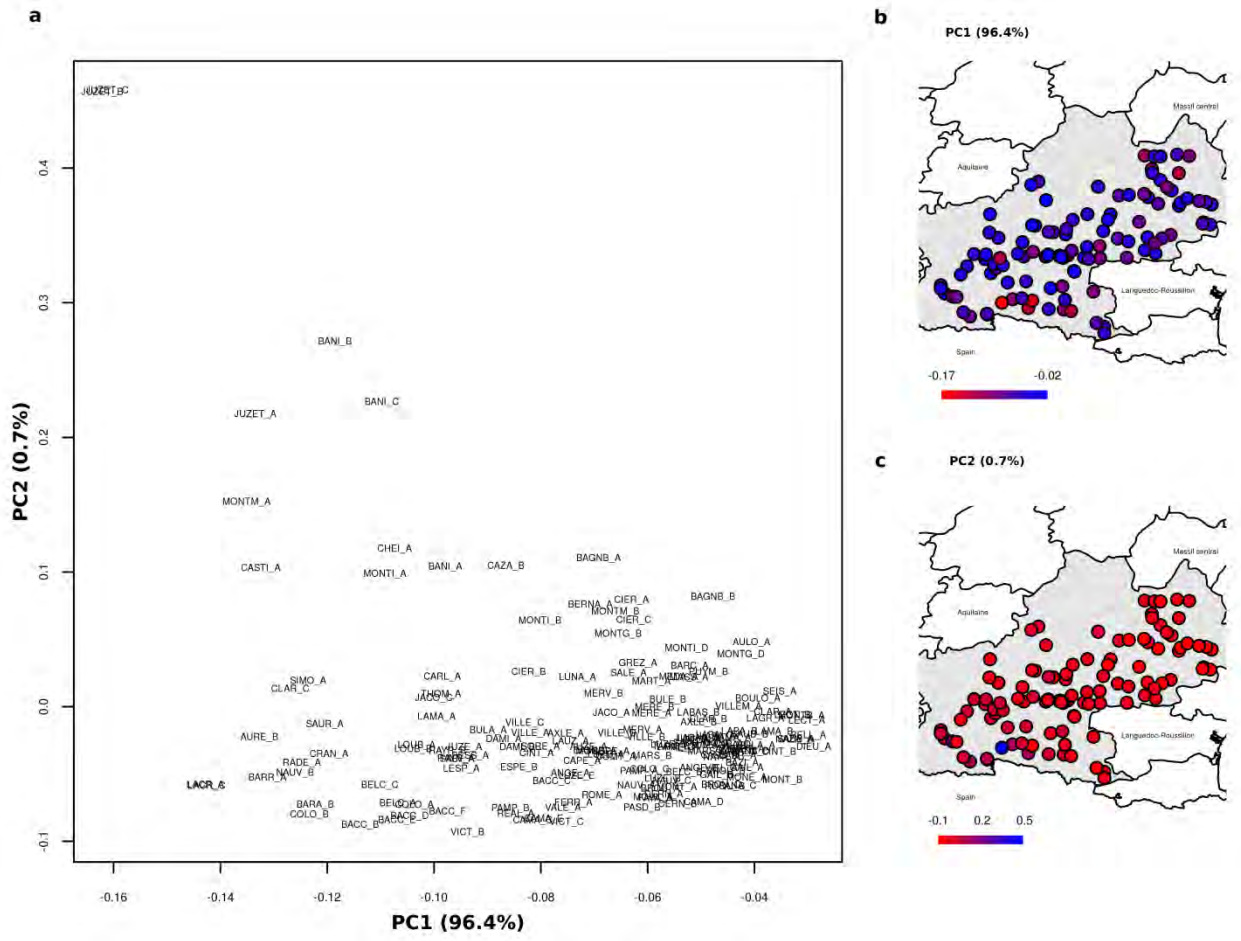


Figure S6. Relationship between the posterior mean of the SNP regression coefficient β_i and the Bayes factor estimates. Right panels: Manhattan plots of the genome-environment association results for the 21 climate variables. The x-axis indicates the position along each chromosome. The five chromosomes are presented in a row along the x-axis in different degrees of blue. The y-axis indicates either the posterior mean of the regression coefficient β_i (M_Beta value) or the Bayes factor (BF_{mc} expressed in deciban units), estimated by the AUX model implemented in the program BayPass. Left panels: Estimates of the Bayes factor as a function of the SNP regression coefficients (β_i) for the 21 climate variables.

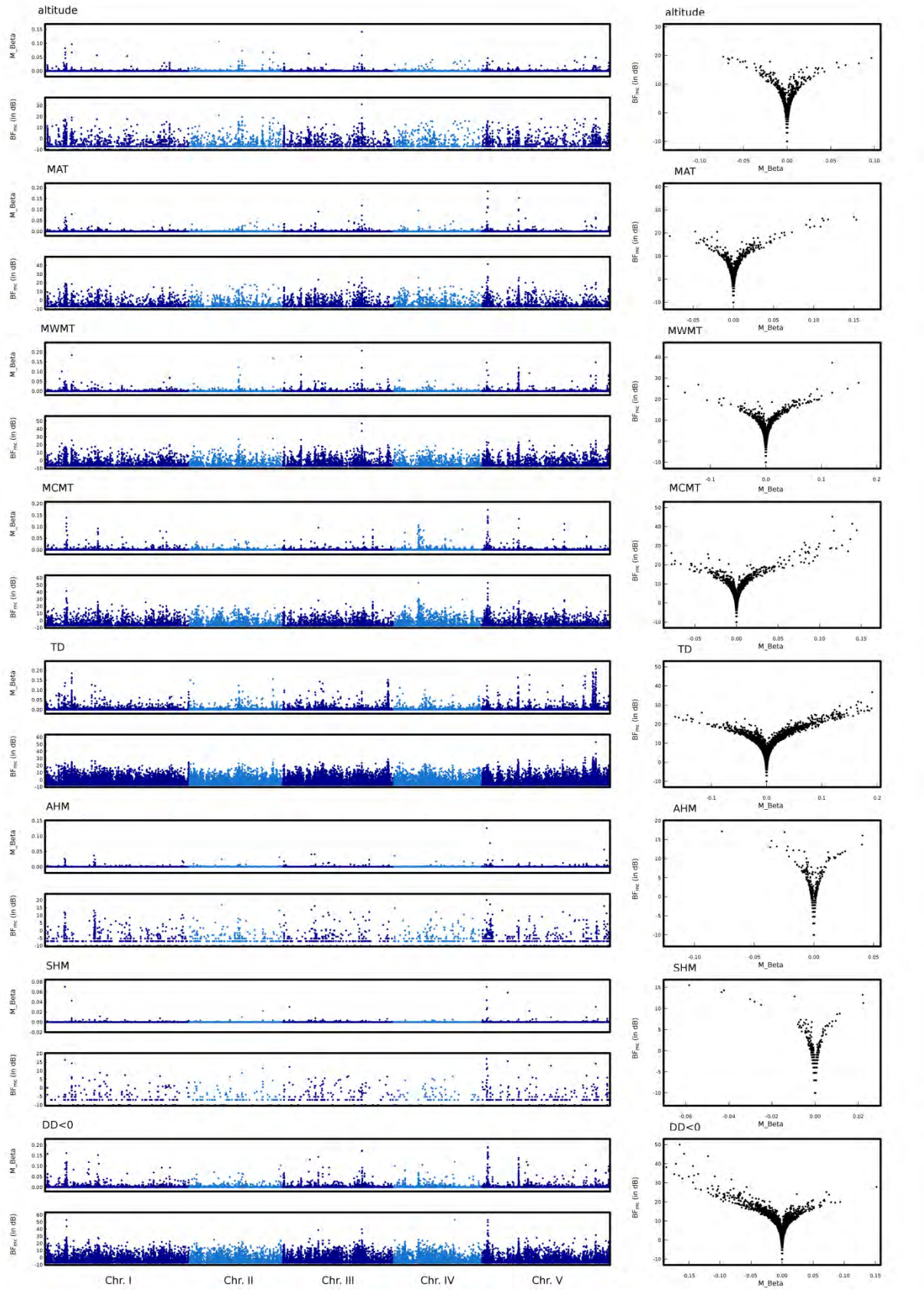

altitude
MAT
MWMT
MCMT
TD
AHM
SHM
DD<0
M_Beta
BF_mc (in dB)
Chr. I
Chr. II
Chr. III
Chr. IV
Chr. V

Figure S6 (continued)

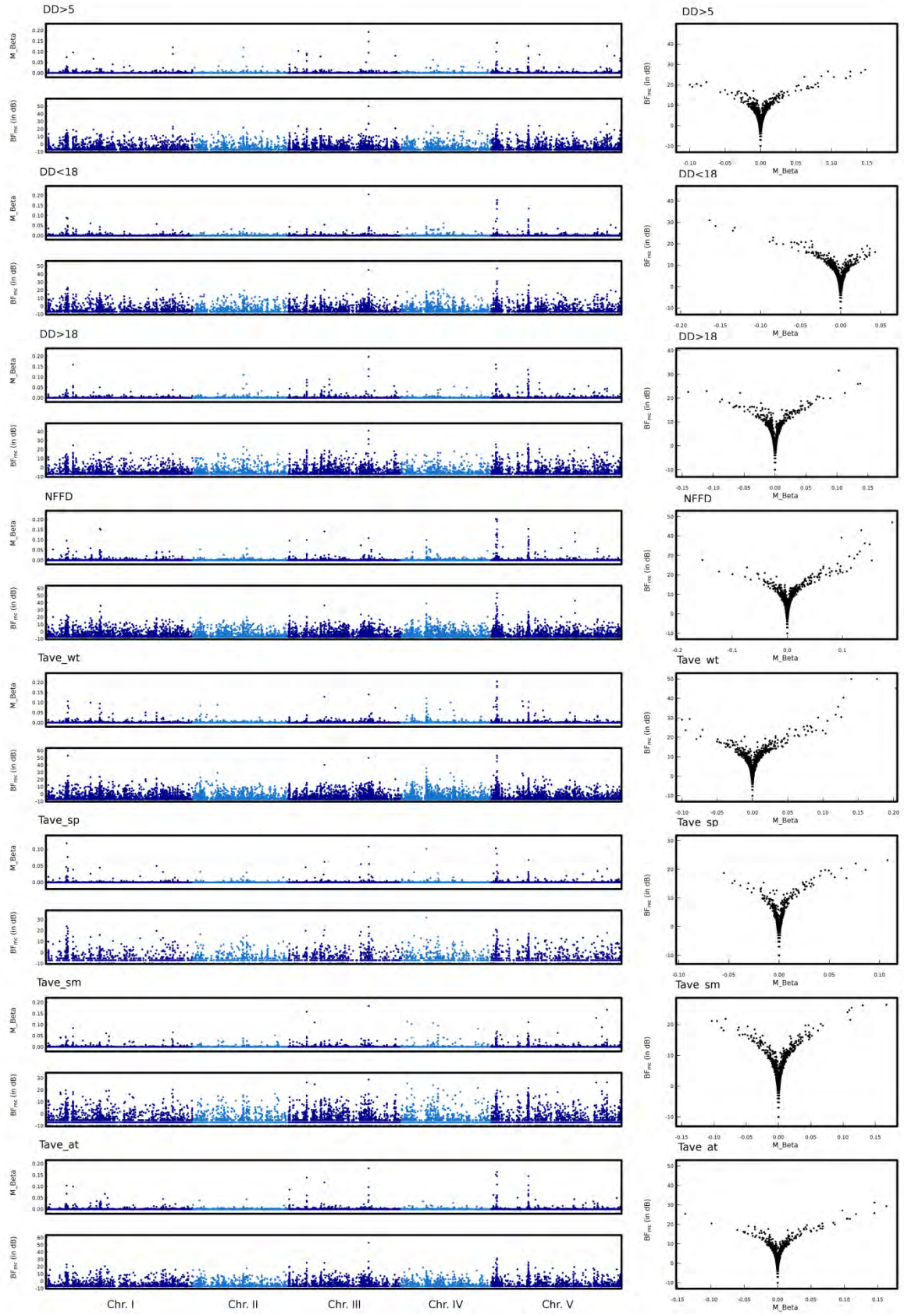
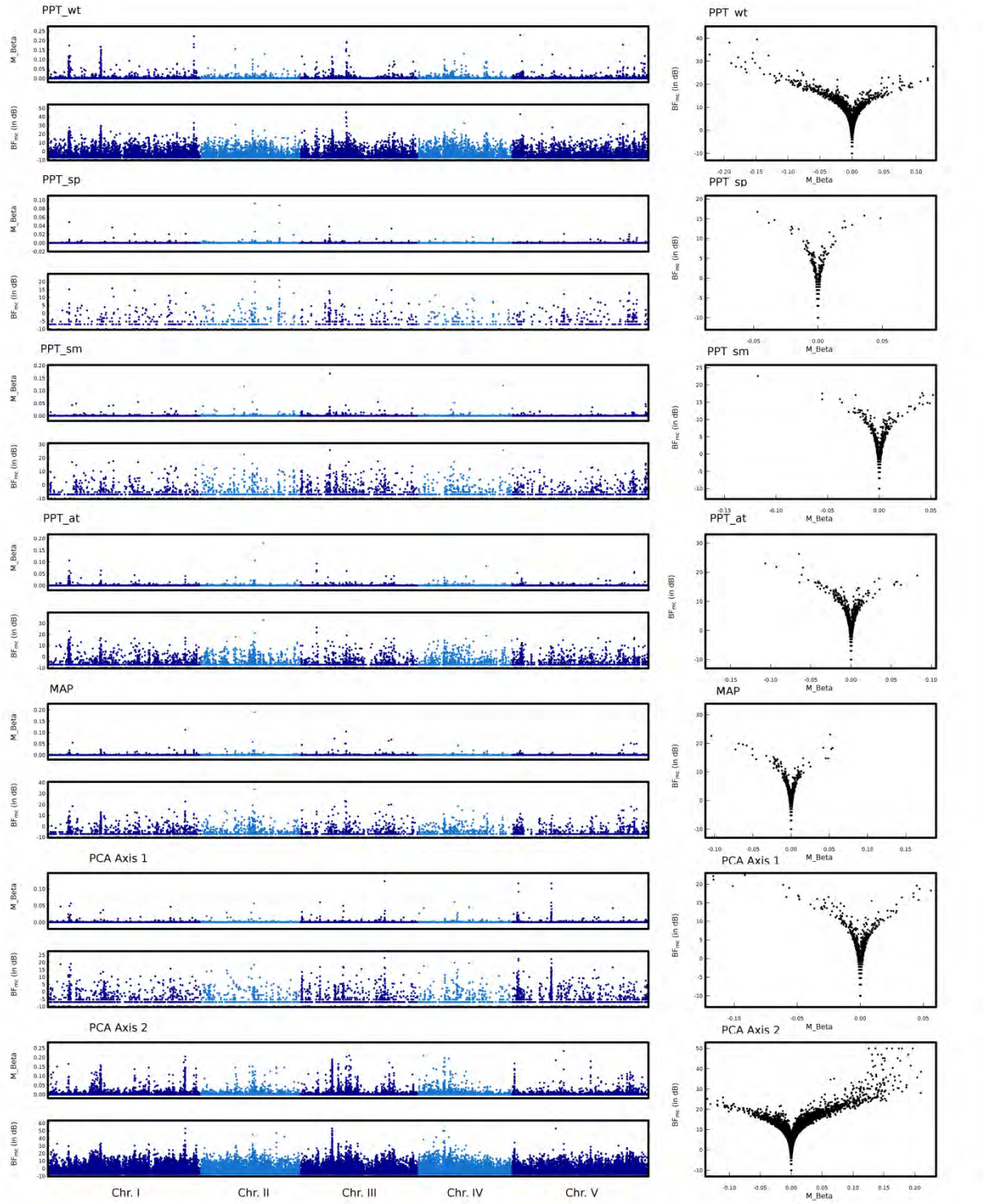


Figure S6 (continued)



Manuscrit

“Adaptation to plant communities across the genome of *Arabidopsis thaliana*”

Frachon L., Mayjonade B., Bartoli C., Hautekeete N.C. & Roux F.

La soumission sera effectuée une fois que le manuscrit « A genomic map of adaptation to local climate in *Arabidopsis thaliana* » sera accepté pour publication

C. Manuscrit: Adaptation to plant communities across the genome of *Arabidopsis thaliana*

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

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

[2] **EEP, Université de Lille 1,** Villeneuve d'Ascq, France

[¶] These authors contributed equally to this work.

[*] To whom correspondence should be addressed E-mail: fabrice.roux@inra.fr

Abstract

Understanding the genetic mechanisms of plant-plant interactions represents a great opportunity to predict the evolutionary trajectories of natural communities and to improve breeding designs of crop mixtures. Yet, studies reporting the identification of genetic variants underlying plant-plant interactions are still scarce and only focused on monospecific interactions, despite the complex and diffuse interactions occurring in a plant assemblage. In this study, based on 145 natural populations of *Arabidopsis thaliana* characterized for plant communities, we conducted a Genome Environment Association Analysis using more than 1.5 million SNPs to finely map genomic regions associated with plant community descriptors, e.g. α-diversity, composition and species abundances. We detected neat peaks of association, associated with the abundance of co-occurring plant species. Still, the genomic architecture of *A. thaliana* largely differed among species belonging to the same botanical family, suggesting a high degree of biotic specialization. In addition, the identification of QTLs for diversity and composition of plant communities highlighted the benefit of exploring diffuse biotic interactions. A substantial fraction of candidate genes was involved either in the shade avoidance syndrome or in nutrient foraging, supporting the hypothesis of inter-specific competition for light and nutrients. In addition, SNPs related to plant community descriptors were significantly enriched in the extreme tail of a genome scan of adaptive spatial differentiation, indicating that the identified candidate genes have been shaped by natural selection. We believe our genomic map of local adaptation to plant communities provides a first step towards our understanding of coevolutionary processes between *A. thaliana* and its plant social network.

Introduction

Studying the genetic mechanisms 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 bases 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 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 & Violle 2015). This challenge calls for a multidisciplinary approach at the interface of community ecology and functional genomics (Whitham *et al.* 2006, Hendry 2013). Still, there is very limited information about the genetic variants underlying plant-plant interactions.

To our knowledge, only two studies reported the identification of Quantitative Trait Loci (QTLs) associated with plant-plant interactions at the interspecific level, i.e. in a heterospecific context (Baron *et al.* 2015, Frachon *et al.* 2017a). Both studies were based on a local genome-wide association mapping population of the annual plant species *Arabidopsis thaliana* located between two permanent meadows dominated by grasses (Frachon *et al.* 2017a). Extensive genetic variation was found for competitive ability in presence of four competitor species (*i.e. Poa annua*, *Stellaria media*, *Trifolium repens* and *Veronica arvensis*) and genomic regions associated with competitive response were highly dependent on the identity of the competitor species (Baron *et al.* 2015). While informative, these studies only focused on monospecific interactions. However, throughout their life cycle, plants can interact simultaneously with a range of plant species, suggesting that the genetics of plant-plant

interactions would be best studied in the context of complex and diffuse interactions occurring in an assemblage (Litrico & Violle 2015, Roux & Bergelson 2016). In addition, it is still unknown whether polymorphic genes involved in plant-plant interactions have been shaped by natural selection.

Here, following the current standards of ecological genomics (Bergelson & Roux 2010), we aimed to establish a genomic map of local adaptation to plant communities in *A. thaliana*. To do so, we first characterized plant communities associated with 145 natural populations of *A. thaliana* located in the south-west of France (Bartoli *et al.* 2017, Frachon *et al.* 2017b). Based on a Bayesian hierarchical model controlling for the genome wide affects of confounding demographic evolutionary forces (Gautier 2015), we then conducted a Genome-Environment Association (GEA) analysis with more than 1.5 million Single Nucleotide Polymorphisms (SNPs) to finely map genomic regions associated with descriptors of plant communities, such as diversity, composition and abundance of plant species. We then explored how natural selection has shaped the loci associated with those descriptors. In particular, we tested whether 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 if specific biological processes were overrepresented among SNPs involved in adaptation to plant communities and discussed the function of candidate genes.

Material and methods

Characterization of plant communities

We focused on 168 natural populations of *A. thaliana* located in the Midi-Pyrénées region and previously characterized for climate (Frachon *et al.* 2017b) and bacterial microbiota (Bartoli *et al.* 2017). Plant communities associated with *A. thaliana* were

characterized during spring (mid-May to mid-June 2015), that is during 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 of 23 natural populations. The wide range of ecological conditions encountered by the remaining 145 populations can be summarized in four broad habitat types: stone wall (n = 13), bare ground (n = 59), grassland (n = 43) and meadow (n = 30) (**Dataset S1**).

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 species present in each quadrat. We then estimated the abundance of each putative 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.* 2017). A herbarium was established by collecting a representative individual of each putative species per population, resulting in 2,233 specimens.

Many specimens were sampled at the seedling stage or without the presence of flower or fruit (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 ***SI text*** (**Fig. S1**). We obtained a

matK sequence (> 500bp) for 97% of the specimens (n = 2166). The *matK* sequences were clustered in 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 S2 and S3**). 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 S1**). For each quadrat, we estimated species richness and Shannon α-diversity by using the functions 'specnumber' and 'diversity' in the package 'vegan' (Oksanen et al. 2016) under 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 (presence in more than 10 populations).

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 were dependent on the type of habitat, we ran the following mixed model based on population BLUP estimates (PROC MIXED procedure in SAS9.3, SAS Institute Inc., Cary, North Carolina, USA):

$$Y_i = \mu_{trait} + habitat_i + \varepsilon_i \qquad (2)$$

where Y_i is one of the five plant community descriptors; 'habitat' accounts for the differences among the four habitat types (stone wall, bare ground, grassland and meadow);'ε' is the residual term. The significance of pairwise comparison was adjusted following a Tukey's studentized range (HSD) test.

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):

$$Y_{ijk} = latitude_i + longitude_j + altitude_k + latitude_i*longitude_j + latitude_i*altitude_k + longitude_j*altitude_k + latitude_i*longitude_j*altitude_k + \varepsilon_{ijk} \qquad (3)$$

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 S1**) (Frachon *et al.* 2017b);'ε' is the residual term.

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.* 2017b). Following Frachon *et al.* (2017b), the matrix of population allele frequencies was trimmed according to four successive criteria (i.e. removing SNPs with missing values in more than seven populations, removing SNPs with a relative coverage depth above 1.5 or below 0.5, removing SNPs with a standard deviation of allele frequency across the populations below 0.004, removing SNPs with the alternative allele present in less than 11 populations), resulting in a final number of 1,519,748 SNPs.

Genome-Environment Association analysis

We performed a Genome-Environment Association (GEA) analysis between 1,519,748 SNPs and 49 plant community descriptors (species richness, Shannon α-diversity, coordinates on three first PCoA axes and abundance of the 44 most prevalent OTUs) based on a Bayesian hierarchical model (i) including a population covariance matrix **Ω** accounting for the neutral covariance structure across population allele frequencies, (ii) adapted to Pool-Seq data and (iii) implemented in the program BayPass (Gautier 2015). In this study, we used the core model to evaluate the association between allele frequencies along the genome and the 49 plant community descriptors. For each SNP, we estimated Bayesian Factor (BF_{is}) and the associated regression coefficient (Beta_is) using an Importance Sampling algorithm (Gautier 2015). 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 corresponded to the average Beta_is and BF_{is} value across the three repeats.

As previously performed in Frachon *et al.* (2017b), 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^{th} rank across the genome). The pairwise FMD distances (Förstner & Moonen 2003) had a narrow range of variation (from 2.02 to 2.27), suggesting consistent estimates among the 32 resulting covariance matrices.

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 BF_{is} and beta_is. We discarded four plant community descriptors (OTU4, OTU7, OTU67 and OTU203) with a correlation coefficient below 0.75.

Enrichment analyses

Enrichments (i) in signatures of selection (XtX statistic similar to a SNP-specific F_{ST} that is corrected for the scaled covariance of population allele frequencies) and (ii) in biological processes (n = 736), for SNPs associated with plant community descriptors were calculated as previously described in Frachon *et al.* (2017b), with the following parameters: 0.1% upper tail of the XtX distribution and 0.1% upper tail of the BF_{is} for the enrichment analysis in signatures of selection, 0.01% upper tail of the BF_{is} for the enrichment analysis in biological processes. For each significantly enriched biological process ($P < 0.001$), we retrieved the identity of all the genes containing SNPs in the 0.01% upper tail of the BF_{is}

values distribution. In this study, enrichment in biological processes was calculated solely for plant community descriptors with a significant enrichment in signatures of selection.

Identification of candidate genes associated with plant community descriptors

For each plant community descriptor, we first selected the 152 SNPs with the highest BF_{is} (i.e. 0.01% of total SNPs). Then, using the TAIR 10 database (*https://www.arabidopsis.org/*), we retrieved all the annotated genes located within a 2kb region around each top SNP, leading to a list of 4,735 unique top genes. Finally, we focused on genes associated with either more than three plant community descriptors or with three plant community descriptors and underlying an enriched biological process, resulting in a final list of 32 pleiotropic candidate genes associated with plant communities.

To test whether the observed frequency distribution of the degree of genetic pleiotropy can deviate from neutral expectation, we first randomly sampled for each plant community descriptor the same number of top genes across the genome and calculated the corresponding frequency distribution of the degree of genetic pleiotropy. This procedure was repeated 1,000 times to generate a null distribution.

Results

Plant communities associated with 145 natural populations of *A. thaliana* located in the Midi-Pyrénées region (**Fig. 1A**) were characterized during spring (mid-May to mid-June 2015) corresponding to the period of seed production of *A. thaliana* in south-west of France. To characterize plant communities, we first established a herbarium by collecting a representative individual of each putative species per population, based on morphological

characteristics, 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 identification, we adopted a metabarcoding approach based on the chloroplast marker *matK* to determine the identity of the species present in the plant communities (**Fig. S1**). 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 (**Dataset S3 and Fig. S1**). In agreement with the deep taxonomic resolution of the marker *matK* (Barthet & Hilu 2007), a large portion of specimens were identified at the species level (84.6%) (**Dataset S3**).

Species richness and Shannon α-diversity largely differed among the 145 populations (**Table S1**), with species richness ranging from 1 to 28 (mean = 12.1, median=12) and Shannon α-diversity ranging from 0.32 to 2.42 (mean = 1.40, median=1.35) (**Fig. 1B and Fig. S2**).

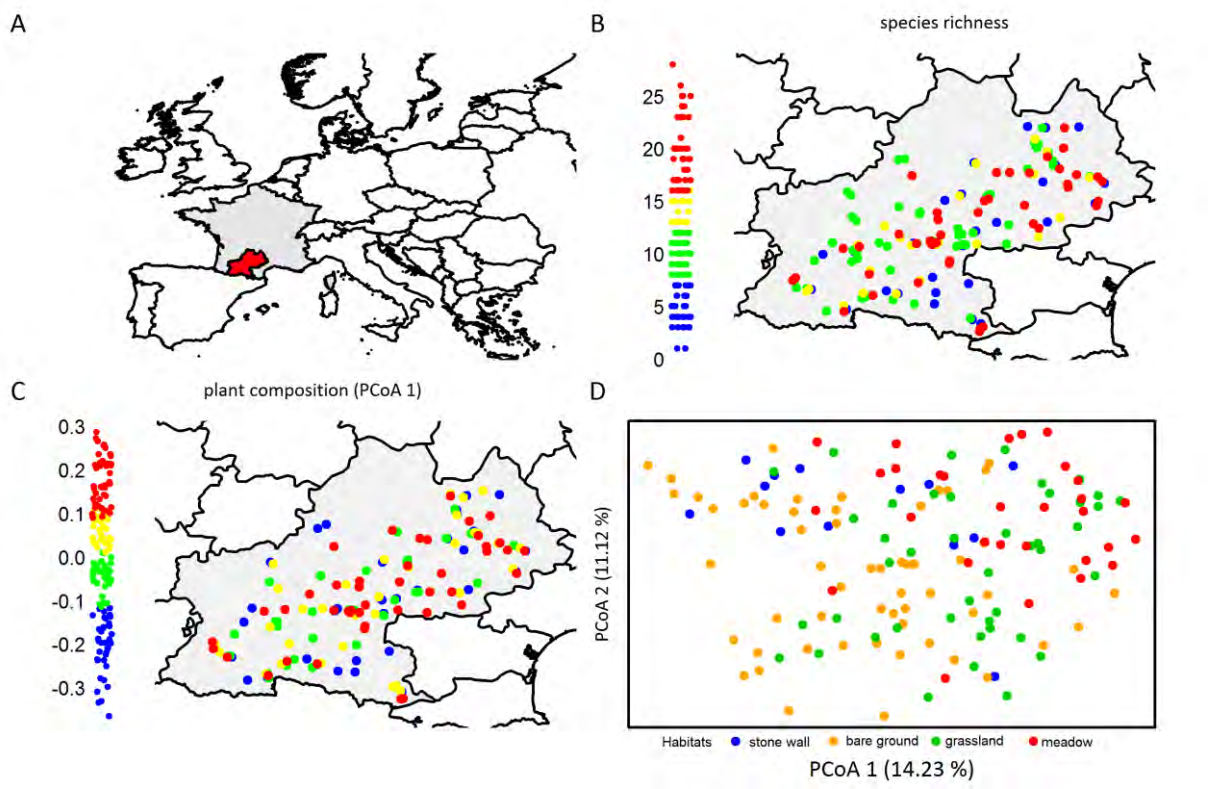


Fig. 1. Species richness and composition of plant communities associated with *A. thaliana*. (*A*) The Midi-Pyrénées region (south-west of France) is colored in red on a European map. (*B*) Geographic variation of species richness. (*C*) Geographic variation of plant community composition represented by the first PCoA axis. (*D*) Effect of the type of habitat on plant community composition represented by the two first PCoA axes. (*B*) and (*C*) The jitter plot on the left of the map illustrates the distribution of species richness and plant community composition. The four colors correspond to the four quartiles.

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 S3**). The first three PCoA axes explained ~34% of the variation in plant composition (**Fig. S3**). Plant composition largely differed among the 145 populations (**Fig. 1C**), with up to 75.9% of variance explained by the 'population' factor (**Table S1**). The first PCoA axis (explaining 14.23% of the total variance) was mainly associated with annual species occurring in bare tilled, fallow or recently abandoned arable lands (EUNIS habitat E1.5, e.g. *Bromus hordeaceus*, *Sonchus oleraceus*, *Veronica persica*) and with perennial species occurring in mesic grasslands (EUNIS habitat E2, e.g. *Cerastium*

fontanum, Crepis biennis, Festuca rubra, Plantago lanceolata) (**Dataset S4**). The second PCoA axis (11.12%) was significantly associated (i) with species occurring in the same habitats (E2 e.g. *Achillea millefolium*, E1.5 e.g. *B. hordeaceus, Trifolium campestre*) on the one side of the axis and (ii) with small annual and pioneer species occurring in grasslands (included in E2, e.g. *Poa annua*, *A. thaliana*) or open habitats like rock debris swards (E1.11, e.g. *Erophila verna*) on the other side (**Dataset S4**). The third PCoA axis (8.61%) was significantly associated with species already described for the first and second PCoA axes, without any clear pattern of habitat or life-cycle (**Dataset S4**). It should be noted that the abundance of *A. thaliana* was significantly, although poorly, correlated with the abundance of three other plant OTUs (i.e. *Sagina apetala,* Spearman's *rho* = 0.25; *Epilobium sp.,* Spearman's *rho* = 0.26; *Holcus lanatus*, Spearman's *rho* = -0.27) (**Dataset S4**).

Plant diversity and plant composition largely differed among the four broad habitat types (i.e. stone wall, bare ground, grassland and meadow) encountered by the 145 populations (**Dataset S1**). Plant diversity was significantly higher in grasslands and meadows than in the stone wall and bare ground habitats (**Table S1**). Large differences between grassland/meadow and stone wall/bare ground habitats were also observed on the first PCoA axis (**Table S1, Fig. 1D, Fig. S3**). In contrast, plant composition on the second PCoA axis was significantly different between the bare ground habitat and the stone wall/meadow habitats; the grassland habitat being intermediate (**Table S1, Fig. 1D, Fig. S3**). No habitat effect was detected for the third PCoA axis (**Table S1, Fig. S3**). In accordance with the differences observed among habitats, plant diversity and plant composition varied at a very fine spatial scale (**Table S1, Fig.1C, Fig. S4**). Altogether, these results indicate contrasted plant assemblages, even at a short fine spatial scale, among the locations inhabited by *A. thaliana* in the Midi-Pyrénées region.

To identify significant associations between genetic polymorphisms and plant community descriptors, we adopted a GEA approach based on within-population allele frequency previously estimated for 1,519,748 SNPs (i.e. one SNP every 78 bp) (Frachon *et al.* 2017b). Because only a weak geographic pattern of genomic variation was observed among *A. thaliana* natural populations in the Midi-Pyrénées region (Frachon *et al.* 2017b), the main drawbacks of GEA analyses often observed at larger geographical scales (i.e. confounding by population structure, rare alleles and allelic heterogeneity) should be limited. We detected neat peaks of association, in particular for the abundance of plant OTUs (**Fig. 2 and Dataset S5**). The majority of QTLs was highly dependent on the identity of plant OTUs (**Fig. 2**). For example, a neat peak of association was detected at the top of chromosome 3 for the abundance of *Convolvulus arvensis* (OTU83, Convolvulaceae) present in 16.5% of the populations, whereas two neat peaks of association were detected at the bottom of chromosomes 3 and 5 for the abundance of *Poa nemoralis* (OTU149, Poaceae) present in 19.3% of the populations. The genomic architecture also largely differed between two species belonging to the same botanical family, and even to the same genus (**Dataset S5**). For example, two neat peaks of association were detected at the top of chromosomes 3 and 5 for the abundance of *Sonchus oleraceus* (OTU27, Asteraceae) present in 34% of the populations, whereas a neat peak of association was detected at the top of chromosome 1 for the abundance of *Helminthotheca echioides* (OTU16, Astereaceae) present in 12% of the populations. Interestingly, we also detected a neat peak of association at the bottom of chromosome 3 associated with the abundance of *A. thaliana*, suggesting a genetic basis of plant-plant interactions at the intraspecific level (**Fig. 2**).

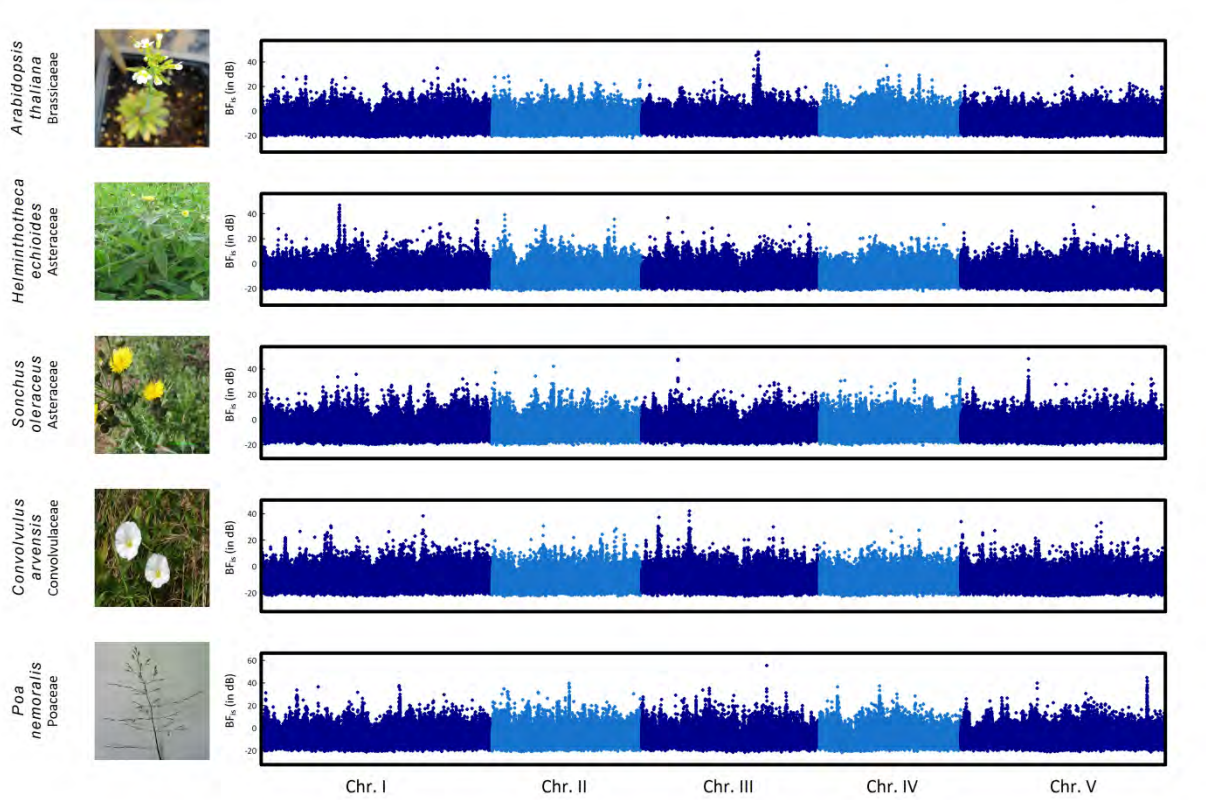


Fig. 2. Manhattan plots of the genome-environment association results for the abundance of five plant species. The x-axis indicates the position along each chromosome. The five chromosomes are presented in a row along the x-axis in different degrees of blue. The y-axis indicates the Bayes factor (BF_{is} expressed in deciban units) estimated by the core model implemented in the program BayPass.

To test whether the SNPs the most associated with plant community descriptors presented genome-wide signatures of selection, we performed a genome-wide scan for spatial differentiation. The 0.1% tail upper tail of the spatial differentiation distribution displayed a significant enrichment (up to 9.97) for SNPs associated with more than half of the plant community descriptors, including species richness, Shannon α-diversity, the first and third PCoA axes and the abundance of 19 OTUs (including *A. thaliana*) belonging to nine botanical families (**Table 1**). This clear signature of selection across the genome suggests that *A. thaliana* is locally adapted to its associated plant communities.

Traits	Species	Familly	Fe	*P*
Species richness	-	-	**3.95**	**
Shannon diversity	-	-	**4.60**	***
PCoA 1	-	-	**5.26**	***
PCoA 2	-	-	1.32	ns
PCoA 3	-	-	**7.89**	***
OTU1	*Conyza canadensis*	Asteraceae	1.32	ns
OTU3	*Crepis biennis*	Asteraceae	1.97	ns
OTU8	*Lactuca serriola*	Asteraceae	**3.29**	*
OTU10	*Achillea millefolium*	Asteraceae	**3.95**	**
OTU15	*Hypochaeris radicata*	Asteraceae	**5.26**	**
OTU16	*Helminthotheca echioides*	Asteraceae	**4.60**	***
OTU18	*Lapsana communis*	Asteraceae	2.63	ns
OTU20	*Senecio vulgaris*	Asteraceae	1.97	ns
OTU27	*Sonchus oleraceus*	Asteraceae	1.32	ns
OTU46	*Valerianella locusta*	Caprifoliaceae	**4.60**	***
OTU49	*Fraxinus excelsior*	Oleaceae	0.00	ns
OTU65	*Plantago lanceolata*	Plantaginaceae	**3.29**	*
OTU71	*Veronica arvensis*	Plantaginaceae	**5.26**	***
OTU72	*Galium mollugo*	Rubiaceae	0.66	ns
OTU78	*Myosotis arvensis*	Boraginaceae	1.97	ns
OTU83	*Convolvulus arvensis*	Convolvulaceae	**9.87**	***
OTU87	*Anagallis arvensis*	Primulaceae	**3.29**	*
OTU88	*Polygonum aviculare*	Polygonaceae	0.66	ns
OTU100	*Sagina apetala*	Caryophyllaceae	**3.29**	*
OTU109	*Arenaria serpyllifolia*	Caryophyllaceae	0.00	ns
OTU113	*Cerastium fontanum*	Caryophyllaceae	**5.26**	***
OTU114	*Papaver rhoeas*	Papaveraceae	1.97	ns
OTU132	*Bromus hordeaceus*	Poaceae	**5.26**	***
OTU136	*Avena sp.*	Poaceae	**7.89**	***
OTU143	*Festuca rubra*	Poaceae	**6.58**	***
OTU145	*Holcus lanatus*	Poaceae	1.97	ns
OTU146	*Dactylis glomerata*	Poaceae	2.63	ns
OTU147	*Catapodium rigidum*	Poaceae	1.97	ns
OTU149	*Poa nemoralis*	Poaceae	**3.95**	**
OTU154	*Poa annua*	Poaceae	**5.92**	***
OTU159	*Aphanes arvensis*	Rosaceae	0.00	ns
OTU179	*Epilobium sp.*	Onagraceae	**6.58**	***
OTU192	*Erophila verna*	Brassicaceae	**3.29**	*
OTU196	*Capsella bursa-pastoris*	Brassicaceae	1.32	ns
OTU198	*Arabidopsis thaliana*	Brassicaceae	**9.87**	***
OTU202	*Cardamine hirsuta*	Brassicaceae	2.63	ns
OTU204	*Geranium sp.*	Geraniaceae	2.63	ns
OTU216	*Medicago lupulina*	Fabaceae	1.32	ns
OTU223	*Vicia sativa*	Fabaceae	1.32	ns
OTU234	*Trifolium campestre*	Fabaceae	2.63	ns

Table 1. Enrichment of genome-wide spatial differentiation (XtX) in the 0.1% tail of the BF_{is} distribution of each plant community descriptor. 'Fe' stands for fold value of enrichment.

To identify candidate genes underlying local adaptation to plant communities, we adopted two non-exclusive approaches. First, we focused on genes associated with more than three plant community descriptors, resulting in a final list of 32 pleiotropic candidate genes. The observed degree of pleiotropy (up to 8 plant community descriptors) followed an L-shaped distribution that significantly deviated from neutral expectation (**Fig. S5, Datasets S6**

and S7), suggesting an over-representation of genes associated with multiple plant species (**Table 2**). Second, we examined which biological processes were overrepresented among the SNPs the most associated with plant community descriptors, focusing on the 23 descriptors for which we observed a significant signature of selection (**Table 1**). We found a significant enrichment in biological processes (up to 276 fold) for the third PCoA axis and the abundance of seven plant OTUs (**Table S2**), leading to the identification of 31 unique candidate genes across 14 biological processes (**Dataset S8**).

Atg number	No of descriptors	Identity of descriptors of plant communities	Locus name	Molecular function
AT1G23380	7	PCoA1-PCoA2-PCoA3-*Cerastium fontanum (Caryophyllaceae)* -*Medicago lupulina (Fabaceae)* -richness-Shannon	KNAT6	Homeodomain transcription factor KNAT6, belonging to class I of KN transcription factor family
AT1G75780	5	*Achillea millefolium (Asteraceae)-Bromus hordeaceus (Poaceae)-Helminthotheca echioides (Asteraceae)-Sonchus oleraceus (Asteraceae)-Convolvulus arvensis (Convolvulaceae)*	TUB1	Beta tubulin gene downregulated by phytochrome A (phyA)-mediated far-red light high-irradiance and the phytochrome B (phyB)-mediated red light high-irradiance responses
AT2G15300	5	PCoA2-*Catapodium rigidum (Poaceae)* -*Veronica arvensis (Plantaginaceae)* -richness-Shannon		Leucine-rich repeat protein kinase family protein
AT2G21140	6	PCoA1-PCoA2-PCoA3-*Medicago lupulina (Fabaceae)* -richness-Shannon	PRP2	Proline-rich protein expressed in expanding leaves, stems, flowers, and siliques.
AT2G21150	6	PCoA1-PCoA2-PCoA3-*Medicago lupulina (Fabaceae)*- richness-Shannon	XCT	Encodes a nuclear localized XAP5 family protein involved in light regulation of the circadian clock and photomorphogenesis.
AT2G24260	3	*Epilobium sp (Onagraceae)-Senecio vulgaris (Asteraceae)-Lactuca serriola (Asteraceae)*	LRL1	Encodes a basic helix-loop-helix (bHLH) protein that regulates root hair development.
AT2G31270	5	PCoA1-*Cerastium fontanum (Caryophyllaceae)-Bromus hordeaceus (Poaceae)* -richness-Shannon	CDT1	Encodes a cyclin-dependent protein kinase. Involved in nuclear DNA replication and plastid division.
AT2G37678	4	PCoA1-*Medicago lupulina (Fabaceae)*- richness-Shannon	FHY1	Positive regulator of photomorphogenesis in far-red light.
AT4G19960	4	PCoA3-*Sagina apetala (Caryophyllaceae)-Holcus lanatus (Poaceae)-Geranium sp (Geraniaceae)*	KT9	Encodes a potassium ion transmembrane transporter.

Table 2. List of pleiotropic candidate genes associated with plant community descriptors.

In both approaches, a substantial fraction of candidate genes are involved in responses to shade (**Table 2 and Dataset S8**), either through signaling pathways of light perception such as *FAR-RED ELONGATED HYPOCOTYL 1* (*FHY1*) (Chen *et al.* 2014) and the bHLH transcription factor *PHYTOCHROME-INTERACTING FACTOR5* (*PIF5*) (Shen *et al.* 2007) or through hormone signaling pathways such as the auxin-responsive gene *SMALL AUXIN*

UPREGULATED67 (*SAUR67*) (Roig-Villanova *et al.* 2007) and the cytochrome P450 *BAS1* (Turk *et al.* 2005). Another major category of candidate genes was related to root development (**Table 2 and Dataset S8**). While the *KNOTTED-like* gene *KNAT6* is required for correct lateral root formation (Dean *et al.* 2004), the bHLH transcription factor *At2g24260* and the histone deacetylase *HDA18* are involved in root hair development and cellular patterning in root epidermis, respectively (Xu *et al.* 2005, Karas *et al.* 2009).

Discussion

By adopting an ecological genomics approach, we established a genomic map of local adaptation to plant communities in *A. thaliana*. Despite its status as a pioneer species, we found that *A. thaliana* can inhabit diverse and contrasted plant assemblages. 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 (Brachi *et al.* 2013) and (ii) the extensive genetic diversity of *A. thaliana* for the response to interspecific competition (Baron *et al.* 2015, Bartelheimer *et al.* 2015).

In agreement with Baron *et al.* (2015), we found that QTLs related to species abundances largely differed among species belonging to different botanical families (i.e. with different growth forms). We further showed that the genomic architecture largely differed among species belonging to the same botanical family (even to the same genus), suggesting a high degree of biotic specialization down to the species level as previously observed in the context of plant-microbe interactions (Roux & Bergelson 2016). Besides studying *A. thaliana* – plant OTU pair, the identification of adaptive QTLs associated with diversity and composition of plant communities highlights the benefit of exploring diffuse biotic interactions, which in turn can help to understand the role of community-wide selection.

Significant genome–environment correlations, however, are only suggestive of the role of ecological factors in shaping adaptive genomic variation. Because diversity and composition of plant communities largely differed among four broad habitat types (i.e. stone wall, bare ground, grassland and meadow), we cannot rule out that the QTLs identified in this study are not related to habitat-specific abiotic conditions. However, three complementary observations challenged this hypothesis. First, a minor fraction of the SNPs the most associated with plant community descriptors were also associated with climate variation (0.15%, Frachon *et al.* 2017b) and edaphic variation (2.68%, Frachon *et al.* unpublished results). Second, no significant difference among habitats was detected for the third PCoA axis for which significant enrichments in signatures of selection and in biological processes were detected. Third, the identity of the candidate genes identified in this study is in line with known molecular mechanisms of plant competition (Pierik *et al.* 2013). As sessile organisms, plants compete for 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. Nevertheless, experiments in controlled conditions are clearly needed to establish a direct link between candidate genes and plant community descriptors.

As a first step, our study reveals the potential to unravel the adaptive genetics underlying biotic interactions within the context of realistic community complexity. Because the diversity and composition of a plant community likely change over the life cycle of *A. thaliana*, studying the playful dynamics of genomic architecture associated with plant community descriptors appears as the next challenge to improve our understanding of coevolutionary processes among interacting plant species.

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 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. and F.R. wrote the manuscript, with contributions from M.B., C.B. 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, Frachon L, Barret M, Rigal M, Zanchetta C, Bouchez O, Carrere S, Roux F (2017) In situ relationships between microbiota and potential pathobiota in *Arabidopsis thaliana*. Elife (in revision).

Bergelson J, Roux F (2010) Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. *Nat 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.

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.

Cuénoud P, Savolainen V, Chatrou LW, Powell M, Grayer REJ, Chase MW (2002) Molecular phylogenetics of Caryophyllales based on nuclear 18s rDNA and plastid RbcL, AtpB, and MatK DNA sequences. *Am J Bot* **89**:132–144.

Dean G, Casson S, Lindsey K (2004) *KNAT6* gene of *Arabidopsis* is expressed in roots and is required for correct lateral root formation. *Plant Mol Biol* **54**:71–84.

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.

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 (2017a) Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time. *Nat Ecol Evol* (in revision).

Frachon L, Bartoli C, Carrère S, Bouchez O, Chaubet A, Gautier M, Roby D, Roux F (2017b) A genomic map of adaptation to local climate in *Arabidopsis thaliana*. *New Phytol* (submitted).

Frachon L, Bartoli C, Roux F. A genome-environment analysis to unravel the genetics of adaptation to edaphic conditions in *Arabidopsis thaliana*. (unpublished data)

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.

Hendry AP (2013) Key questions in the genetics and genomics of eco-evolutionary dynamics. *Heredity* **111**:456–466.

Karas B, Amyot L, Johansen C, Sato S, Tabata S, Kawaguchi M, Szczyglowski K (2009) Conservation of lotus and *Arabidopsis* basic Helix-Loop-Helix proteins reveals new players in root hair development. *Plant Physiol* **151**:1175–1185.

Kim HM, Oh SH, Bhandari GS, Kim CS, Park CW (2014) DNA barcoding of Orchidaceae in Korea. *Mol Ecol Resour* **14**:499–507.

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

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.

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.

Shen Y, Khanna R, Carle CM, Quail PH (2007) Phytochrome induces rapid PIF5 phosphorylation and degradation in response to red-light activation. *Plant Physiol* **145**:1043–1051.

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

Turk EM, Fujioka S, Seto H, Shimada Y, Takatsuto S, Yoshida S, Wang H, Torres QI, Ward JM, Murthy G, Zhang J, Walker JC, Neff MM (2005) *BAS1* and *SOB7* act redundantly to modulate Arabidopsis photomorphogenesis *via* unique brassinosteroid inactivation mechanisms. *Plant J* **42**:23–34.

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.

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.

Data sets

Available online (https://lipm-browsers.toulouse.inra.fr/pub/Frachon2017-PHD/, login: reviewersPHD, password: kryGhehayd4).

Data set 1. Contingency table of abundance of the 244 plant OTUs across the 145 natural populations of *A. thaliana*. 'latitude', 'longitude' and 'latitude' correspond to the GPS coordinates. 'habitat' stands for the four broad habitat types.

Data set 2. Reference sequences for the 244 plant OTUs. The names of the representative specimens are in brackets.

Data set 3. List of 244 OTUs and corresponding family/genus/species name. 'no specimens' stands for the number of specimens belonging to a specific plant OTU. 'no populations' indicates the prevalence of a specific plant OTU among the 145 natural populations of *A. thaliana*.

Data set 4. Correlation among the 49 plant community descriptors. Above-diagonal: Spearman's *rho*. Colors indicates the level of significance of Spearman's *rho* (yellow: $0.05 > P > 0.01$, orange: $0.01 > P > 0.001$, red: $P < 0.001$). Below-diagonal: level of significance of Spearman's *rho* after a false discovery rate (FDR) correction at the nominal level of 5%.

Data set 5. Manhattan plots of the genome-environment association results for the 49 plant community descriptors. The x-axis indicates the position along each chromosome. The five chromosomes are presented in a row along the x-axis in different degrees 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.

Data set 6. List of genes within 2kb of the 152 SNPs with the highest BF_{is} values for each plant community descriptor (species richness, Shannon α-diversity, three first PCoA axes and abundance of 40 plant OTUs).

Data set 7. Identification of pleiotropic genes. 'Atg number': Atg numbers highlighted in red correspond to genes underlying enriched biological processes.

Data set 8. List of the genes underlying enriched biological processes for the third PCoA axis and the abundance of seven plant OTUs.

Supplementary information

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

Supplementary text

Procedure of the metabarcoding approach

We sampled 20-50mg of fresh tissue for all the 2,233 specimens in 96-well plates and samples were then freeze-dried. After grinding with the TissueLyser II (Qiagen), DNA was extracted according to a modified protocol as described in Brachi *et al.* (2013). *matK* was amplified using either universal primers (Cuénoud *et al.* 2002, Kim *et al.* 2014) or genus specific primers designed in this study (**Table S3**). ITS2 gene was amplified using the pair of primers S2F-S3R (Chen *et al.* 2010). PCR amplification was performed in a 20 µl reaction mixture containing 4µl of 5x GoTaq Reaction Buffer (Promega), 0.4 µl of PCR Nucleotide Mix 10mM each (Promega), 0.4µl of each forward and reverse primer at 10µM (Eurofins), 0.2µl GoTaq G2 DNA Polymerase 5u/µl (Promega), 1µl of DMSO, 1µl of BSA 20mg/ml (NEB), 10.6µl of water molecular biology grade and 2µl of template DNA. PCR cycling conditions were: initial denaturation of 2 min at 95°C following by 40 cycles of 30 s at 95°C for denaturation, 30 s at the appropriate Tm (**Table S3**) for annealing, 40 s at 72°C for extension followed by a final extension of 5 min at 72°C. Excess of dNTPs and primers were cleaned up using an ExoSAP treatment. Eight µl of the following enzyme mix were added to the 20 µl of PCR reaction mixture: 0.02 µl of Exonuclease I (20U/µL) (New England Biolabs), 0.2 µl rAPid Alkaline Phosphatase (1U/µL) (Sigma-Aldrich) and 7.78 µl of water molecular biology grade. Then samples were incubated at 37°C for 30 min and 95°C for 5 min. PCR products were visualized on 1% agarose gels and samples showing suitably bright bands were sent to Eurofins (Germany) for Sanger DNA sequencing using forward or reverse PCR primers (**Table S3**) on an ABI3730X. Sequences extremities were trimmed with the Sequencing Analysis v6.0 software (Applied Biosystems) using a 25 bp window, segments with >2 bp showing QV<20 were removed.

Fig. S1. Identification of the species/genus for the 2,233 plant specimens.

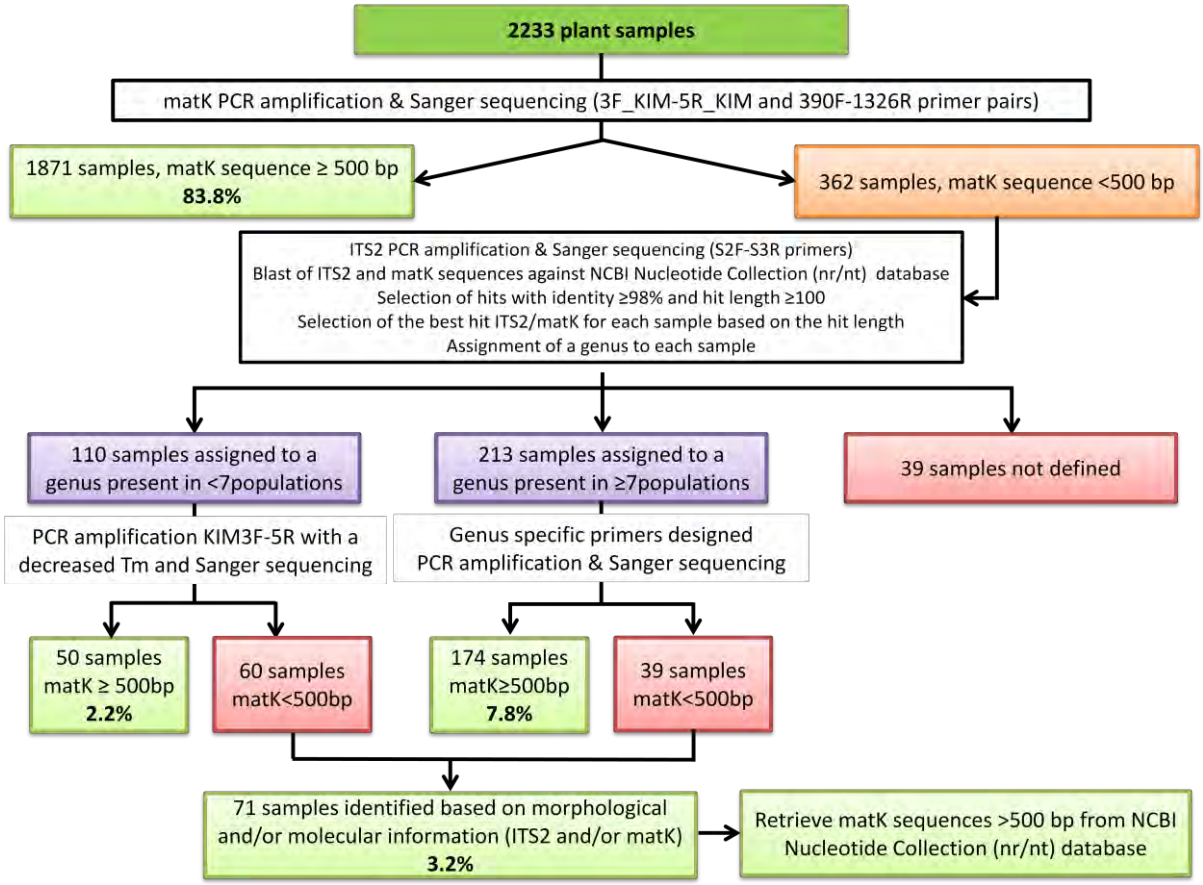


Fig. S2. Geographic variation of Shannon α-diversity. The jitter plot on the left of the map illustrates the distribution of Shannon α-diversity. The four colors correspond to the four quartiles.

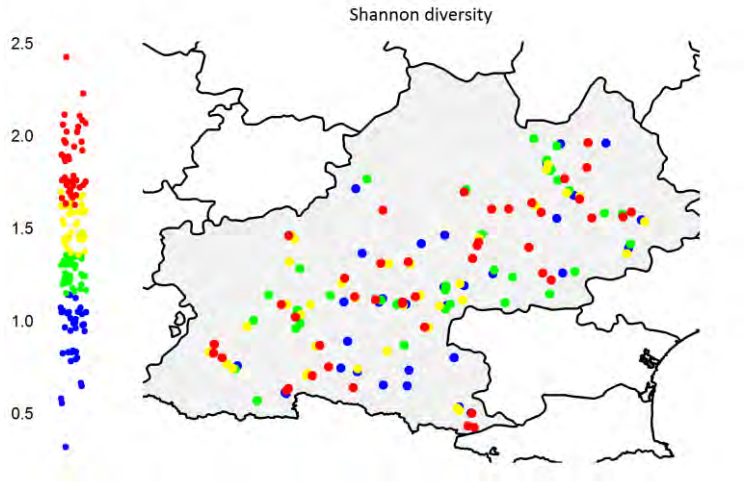


Fig. S3. Plant community composition among the 145 populations. (*A*), (*B*) and (*C*) Position of the 145 populations in the space defined by the three first PCoA axes. The four colors depicted the four habitat types (blue: stone wall, orange: bare ground, green: grassland, red: meadow). (*D*), (*E*) and (*F*) Habitat effect on plant community composition defined by the three first PCoA axes. Different letters indicate different groups according to the habitat after adjustment with a Tukey's studentized range (HSD) test.

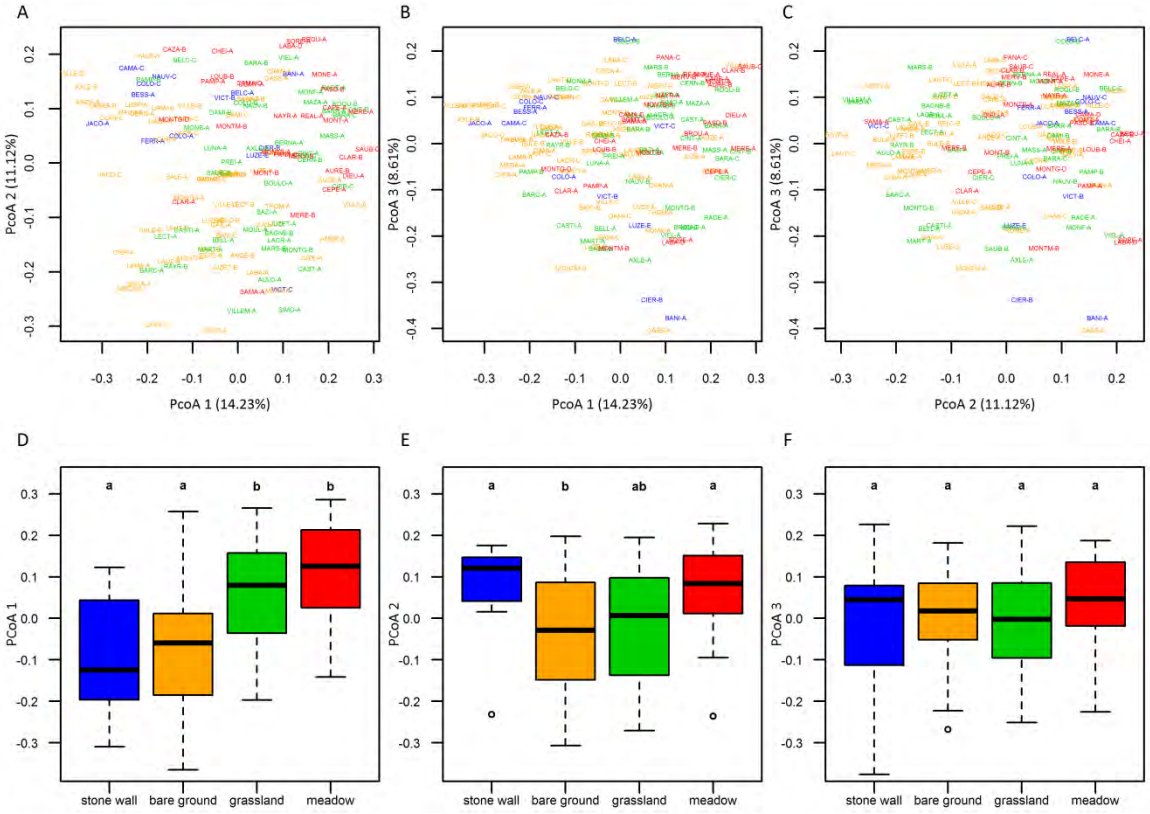


Fig. S4. Geographic variation of plant community composition represented by the coordinates along the three first PCoA axes. The jitter plot on the left of the map illustrates the distribution of the coordinates along the three first PCoA axes. The four colors correspond to the four quartiles.

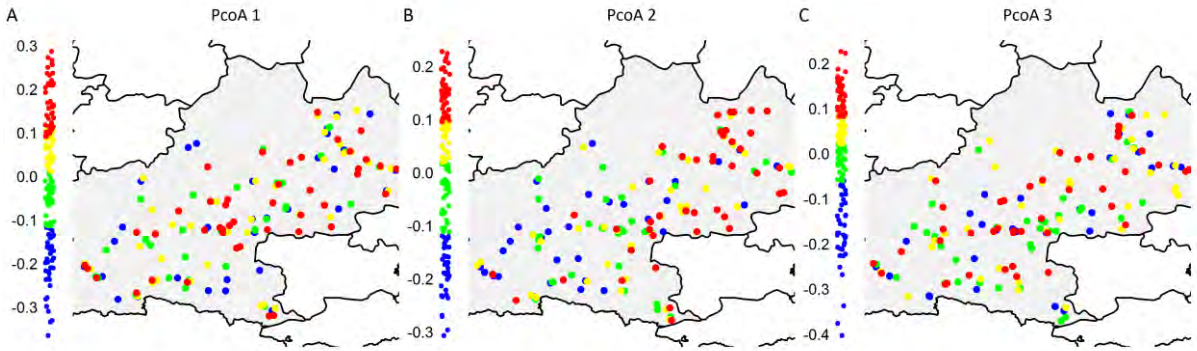


Fig. S5. Observed frequency distribution of the degree of genetic pleiotropy (red dots). Grey dots represent 1000 null permutations across the genome.

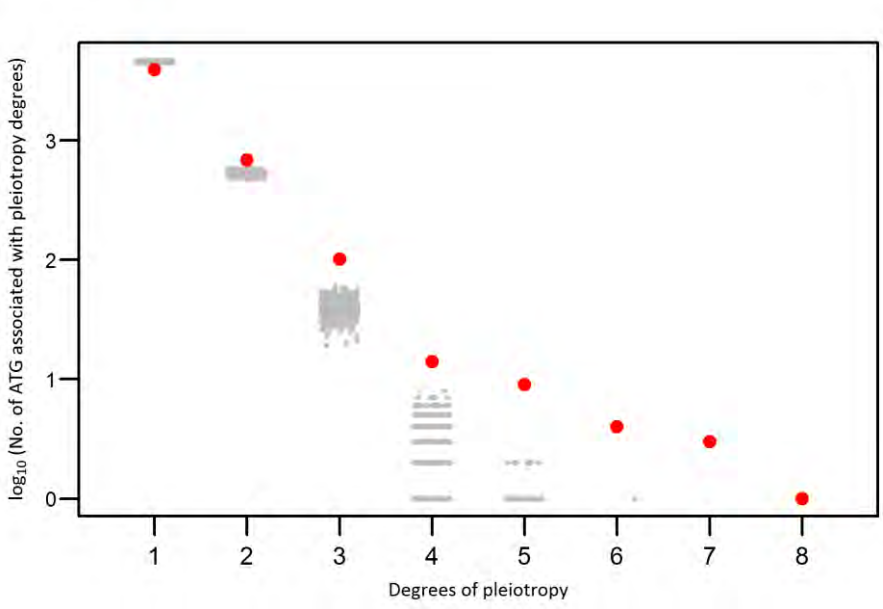


Table S1. Diversity and composition of plant communities. (*A*) Differences of plant community descriptors among the 145 populations. (*B*) Relationships between plant community descriptors and geographical coordinates. (*C*) Differences of plant community descriptors among the four habitat types (i.e. stone wall, bare ground, grassland, meadow).

A. Population effect

Plant descriptors	*Z* value	*P*	Variance explained
Species richness	7.41	***	78.18%
Shannon diversity	6.41	***	63.02%
PCoA1	7.07	***	73.26%
PCoA2	6.94	***	71.04%
PCoA3	7.24	***	75.89%

B. Geographic patterns

	Species richness		Shannon diversity		PCoA1		PCoA2		PCoA3	
Terms	*t* value	*P*	*t* value	*P*	*t* value	*P*	*t* value	*P*	*t* value	*P*
latitude	0.15	*ns*	0.32	*ns*	0.05	*ns*	0.63	*ns*	0.84	*ns*
longitude	0.03	*ns*	-0.12	*ns*	0.29	*ns*	0.63	*ns*	0.39	*ns*
latitude*longitude	-0.05	*ns*	0.1	*ns*	-0.3	*ns*	-0.58	*ns*	-0.39	*ns*
altitude	1.17	*ns*	0.74	*ns*	0.6	*ns*	-0.56	*ns*	-1.34	*ns*
latitude*altitude	-1.19	*ns*	-0.75	*ns*	-0.61	*ns*	0.59	*ns*	1.34	*ns*
longitude*altitude	-1.03	*ns*	-0.54	*ns*	-0.7	*ns*	0.32	*ns*	1.29	*ns*
latitude*longitude*altitude	1.05	*ns*	0.55	*ns*	0.71	*ns*	-0.36	*ns*	-1.29	*ns*

C. Habitat effect

			Tukey test			
Plant descriptors	*F* value	*P*	stone wall	bare ground	grassland	meadow
Species richness	27.67	***	a	a	b	c
Shannon diversity	20.75	***	a	a	b	b
PCoA1	17.13	***	a	a	b	b
PCoA2	7.32	***	a	b	ab	a
PCoA3	0.93	*ns*	a	a	a	a

Table S2. Enrichment of biological processes in the 0.01% tail of the BF_{is} distribution of for plant community descriptors with a significant enrichment in signatures of selection. Only enrichment values with a significant *p*-value < 0.001 are reported. The significance of enrichment was tested against a null distribution using 1000 null permutations. The plant community descriptors with a significant enrichment in signatures of selection but with no significant enrichment in biological processes are not reported in the table.

GO_term	**PCoA 3**	**OTU8** *Lactuca serriola*	**OTU10** *Achillea millefolium*	**OTU16** *Helminthotheca echioides*	**OTU113** *Cerastium fontanum*	**OTU132** *Bromus hordeaceus*	**OTU143** *Festuca rubra*	**OTU149** *Poa nemoralis*
response to brassinosteroid	13.47							
root hair elongation		60.62						
peroxisome fission			55.28					
response to auxin				11.62				
multi cellular organismal development				8.30				
auxin-activated signaling pathway				18.29				
auxin polar transport				55.18				
regulation of organ growth				276.04				
stamen filament development				88.14				
regulation of auxin mediated signaling pathway							51.17	
mismatch repair					27.19			
primary shoot apical meristem specification					34.82			
photorespiration						38.29		
histone deacetylation								51.85
No. Biological processes	1	1	1	6	2	1	1	1

Table S3. MatK and ITS2 primers used in this study. Primers used for Sanger sequencing are highlighted in yellow.

Type	Primer Name	Sequence	Tm
Universal	matK-3F_KIM	CGTACAGTACTTTTGTGTTTACGAG	50°C - 48°C
	matK-1R_KIM	ACCCAGTCCATCTGGAAATCTTGGTTC	
Universal	matk-390F	CGATCTATTCATTCAATATTTC	50°C
	matk-1326R	TCTAGCACACGAAAGTCGAAGT	
Universal	IT2_S2F	ATGCGATACTTGGTGTGAAT	56°C
	ITS2_S3R	GACGCTTCTCCAGACTACAAT	
Veronica genus	veronica_fwd	AAATTCTTCCAATAATCCGG	50°C
	veronica_rv	ATGTTTCTTCTTTGCATTTATTACG	
Geranium genus	geranium_fwd	CGTACAGTMCTTTTGTGTTTACGAG	50°C
	geranium_rv	TGAAGCTCTTCGTTATTGGG	
Cerastium genus	ceratium_fwd	GTTTACGAGCCAAAGTTCTAGC	50°C
	ceratium_rv	CGTCTTTGCATCTATTACGATTC	
Cardamine genus	cardamine_fwd	matK-3F_KIM	50°C
	cardamine_rv	TTCAAACCCTACGTTACCG	
Trifolium genus	trifolium_fwd	CGTACAGTACTTTTGTGTTTACAAGC	50°C
	trifolium_rv	TTCAAATCCTTCGATACTGG	
Vicia genus	vicia_fw	GTACAGTACTTTTGTGTTTACAAGCC	50°C
	vicia_rv	AAATCCTTCGATACTGGGTG	
Hypericum genus	Hypericum_fwd	TTTATGAGCAAAGGTTTTCAGAC	50°C
	Hypericum_rv	TCYTCTTTGCATTTATTACGACTC	
Saxifraga genus	saxifraga_fwd	matK-3F_KIM	50°C
	saxifraga_rv	CTGGTTGAAAGACGCCTC	
Myosotis genus	myosotis_fwd	CTTTTGTGTTTCCGAGCC	50°C
	myosotis_rv	GTTAGATATACTAATACCCCACCCTG	
Ranuculus genus	ranuculus_fwd	GTACASTACTTTTGTGTTTACGTGC	50°C
	ranuculus_rv	GTTGGATACAAGATGCCTCC	
Avena genus	avena_fwd	CGTGCTTTTATGTTTACGAGC	50°C
	avena_rv	TTCTTCAATACCGTATACAAGACG	
Sedum genus	sedum_fwd	AGTGCTTTTGTGTTTACGAGC	50°C
	sedum_rv	GCTACTGGGTGAAAGATGC	
Epibolium genus	epibolium_fwd	RCTTTTATGTTTACGGGCC	50°C
	epibolium_rv	ACCCTTCGATACTGGGTG	
Sonchus genus	sonchus_fwd	GTACAGTACTTTTATGCTTACGAGC	50°C
	sonchus_rv	GGAAATCTTGGTTCAGGC	

Conclusion

D. Conclusion

Afin de progresser dans notre compréhension de l'adaptation d'*A. thaliana* au sein de la région Midi-Pyrénées, j'ai effectué des analyses de type GEA pour identifier les bases génétiques associées à des variables climatiques et à des descripteurs des communautés végétales. A ma connaissance, c'est la première fois qu'une étude de GEA a été réalisée sur des facteurs biotiques.

Pour effectuer les analyses de type GEA, j'ai utilisée une méthode Bayésienne implémentée dans le logiciel BayPass. Cette méthode s'est avérée très puissante pour cartographier finement des régions génomiques associées à des variables écologiques. Il faut cependant mentionner que cette méthode est très consommatrice en temps. En effet, les analyses de type GEA pour un sous-jeu de données génomiques (1/32ème de la totalité des SNPs) peuvent prendre jusqu'à 4 jours pour une seule variable écologique, sans compter l'attente sur la liste des jobs sur les serveurs de calcul. Ainsi, il reste difficile d'imaginer pour l'instant la mise en place des méthodes de re-échantillonnage de type *jackknife* ou *bootstrapping* qui permettrait pourtant d'affiner les résultats issus des analyses de type GEA.

En combinant les analyses de type GEA avec des scans génomiques de différenciation génétique au niveau spatial et des tests d'enrichissement basés sur les différentes catégories de variants génétiques (non-synonymes, synonymes, introniques, intergéniques….), j'ai pu mettre en évidence au sein de la région Midi-Pyrénées que les variables climatiques et les descripteurs des communautés végétales sont des moteurs importants de la variation génomique adaptative d'*A. thaliana*. Par ailleurs, en accord avec les résultats obtenus dans le chapitre 1 sur les relations entre variation de la production totale de graines et variation écologique, les signatures génomiques d'adaptation locale apparaissent plus fortes pour les descripteurs des communautés végétales que pour les variables climatiques. Cependant, il ne faut pas oublier que la relation entre variation phénotypique et variation climatique semble fortement dépendre du type d'habitat considéré. Une prochaine étape serait donc de refaire tourner les analyses de type GEA au sein de chacun des 4 habitats (mur, sol nu, pelouse et prairie). L'attendu pour l'habitat 'prairie' serait donc que les signatures génomiques d'adaptation locale soient plus fortes pour les variables climatiques que pour n'importe quelle autre variable écologique mesurée durant la thèse.

En accord avec le faible nombre de régions génomiques identifiées comme communes entre le climat et les communautés végétales, les fonctions biologiques sous-jacentes à l'adaptation à ces deux grandes catégories écologiques sont largement différentes. Comme précédemment observé à des échelles géographiques plus larges (Kawakatsu *et al.* 2016, Keller *et al.* 2016), un excès de gènes impliqués dans des mécanismes de régulation transcriptionnelle a été détecté pour l'adaptation au climat à une échelle locale. D'un autre côté, de nombreux gènes sous-jacents à l'adaptation aux communautés végétales semblent impliqués dans les voies de signalisation de la perception de la lumière ou des hormones. Cette différence de fonctions biologiques et des gènes sous-jacents suggère que l'évolution d'*A. thaliana* à des changements climatiques ne sera pas contrainte par l'évolution d'*A. thaliana* à des modifications des communautés végétales. Les analyses de type GEA effectuées sur les variables édaphiques et les analyses de type GEA en cours sur le microbiote bactérien et sur le microbiote fongique devraient prochainement compléter le tableau des fonctions biologiques impliquées dans l'adaptation d'*A. thaliana* à de multiples pressions de sélection dans la région Midi-Pyrénées.

Pour résumer, les résultats obtenus dans ce chapitre ont permis de mettre en évidence un fort potentiel adaptatif d'*A. thaliana* au sein de la région Midi-Pyrénées pour répondre rapidement à des modifications abiotiques et biotiques. Pour tester ce potentiel adaptatif, il est complémentaire d'étudier la dynamique adaptative de populations naturelles, notamment sur une courte échelle de temps étant donné la rapidité à laquelle s'effectue les changements globaux.

D. Conclusion

Afin de progresser dans notre compréhension de l'adaptation d'*A. thaliana* au sein de la région Midi-Pyrénées, j'ai effectué des analyses de type GEA pour identifier les bases génétiques associées à des variables climatiques et à des descripteurs des communautés végétales. A ma connaissance, c'est la première fois qu'une étude de GEA a été réalisée sur des facteurs biotiques.

Pour effectuer les analyses de type GEA, j'ai utilisée une méthode Bayésienne implémentée dans le logiciel BayPass. Cette méthode s'est avérée très puissante pour cartographier finement des régions génomiques associées à des variables écologiques. Il faut cependant mentionner que cette méthode est très consommatrice en temps. En effet, les analyses de type GEA pour un sous-jeu de données génomiques (1/32ème de la totalité des SNPs) peuvent prendre jusqu'à 4 jours pour une seule variable écologique, sans compter l'attente sur la liste des jobs sur les serveurs de calcul. Ainsi, il reste difficile d'imaginer pour l'instant la mise en place des méthodes de re-échantillonnage de type *jackknife* ou *bootstrapping* qui permettrait pourtant d'affiner les résultats issus des analyses de type GEA.

En combinant les analyses de type GEA avec des scans génomiques de différenciation génétique au niveau spatial et des tests d'enrichissement basés sur les différentes catégories de variants génétiques (non-synonymes, synonymes, introniques, intergéniques….), j'ai pu mettre en évidence au sein de la région Midi-Pyrénées que les variables climatiques et les descripteurs des communautés végétales sont des moteurs importants de la variation génomique adaptative d'*A. thaliana*. Par ailleurs, en accord avec les résultats obtenus dans le chapitre 1 sur les relations entre variation de la production totale de graines et variation écologique, les signatures génomiques d'adaptation locale apparaissent plus fortes pour les descripteurs des communautés végétales que pour les variables climatiques. Cependant, il ne faut pas oublier que la relation entre variation phénotypique et variation climatique semble fortement dépendre du type d'habitat considéré. Une prochaine étape serait donc de refaire tourner les analyses de type GEA au sein de chacun des 4 habitats (mur, sol nu, pelouse et prairie). L'attendu pour l'habitat 'prairie' serait donc que les signatures génomiques d'adaptation locale soient plus fortes pour les variables climatiques que pour n'importe quelle autre variable écologique mesurée durant la thèse.

En accord avec le faible nombre de régions génomiques identifiées comme communes entre le climat et les communautés végétales, les fonctions biologiques sous-jacentes à l'adaptation à ces deux grandes catégories écologiques sont largement différentes. Comme précédemment observé à des échelles géographiques plus larges (Kawakatsu *et al.* 2016, Keller *et al.* 2016), un excès de gènes impliqués dans des mécanismes de régulation transcriptionnelle a été détecté pour l'adaptation au climat à une échelle locale. D'un autre côté, de nombreux gènes sous-jacents à l'adaptation aux communautés végétales semblent impliqués dans les voies de signalisation de la perception de la lumière ou des hormones. Cette différence de fonctions biologiques et des gènes sous-jacents suggère que l'évolution d'*A. thaliana* à des changements climatiques ne sera pas contrainte par l'évolution d'*A. thaliana* à des modifications des communautés végétales. Les analyses de type GEA effectuées sur les variables édaphiques et les analyses de type GEA en cours sur le microbiote bactérien et sur le microbiote fongique devraient prochainement compléter le tableau des fonctions biologiques impliquées dans l'adaptation d'*A. thaliana* à de multiples pressions de sélection dans la région Midi-Pyrénées.

Pour résumer, les résultats obtenus dans ce chapitre ont permis de mettre en évidence un fort potentiel adaptatif d'*A. thaliana* au sein de la région Midi-Pyrénées pour répondre rapidement à des modifications abiotiques et biotiques. Pour tester ce potentiel adaptatif, il est complémentaire d'étudier la dynamique adaptative de populations naturelles, notamment sur une courte échelle de temps étant donné la rapidité à laquelle s'effectue les changements globaux.

Chapitre 3

Evolution phénotypique et génomique d'une population naturelle d'*A. thaliana* dans un habitat spatialement hétérogène

A. Introduction

Etudier la dynamique adaptative des populations naturelles sur une courte échelle de temps apparaît primordial si l'on souhaite prédire la persistance des espèces face aux changements globaux. Pour étudier l'évolution phénotypique et génomique d'*A. thaliana* sur quelques générations, je me suis focalisée sur la population TOU-A localisée au sud du Morvan, sous une clôture électrique de 300 mètres délimitant deux prairies permanentes. Décrite dans des études précédentes comme étant très polymorphe aussi bien d'un point de vue génétique (Platt *et al.* 2010, Horton *et al.* 2012) que d'un point de vue phénotypique (phénologie : Brachi *et al.* 2013 ; résistance quantitative à des bactéries phytopathogènes : Huard-Chauveau *et al.* 2013, Debieu *et al.* 2015 ; réponse à la compétition interspécifique : Baron *et al.* 2015), la population TOU-A apparaît donc adaptée pour réaliser un suivi phénotypique et génomique. Par ailleurs, comme précédemment observé pour de nombreuses populations naturelles de la région Midi-Pyrénées, cette population est située sur un milieu très hétérogène tant au niveau du sol que des interactions plante-plante. Finalement, une augmentation de 1°C de la température annuelle moyenne a été observée dans cette population sur les 30 dernières années.

A partir d'un transect établi le long de la clôture électrique, Fabrice Roux a récolté les graines de 80 accessions en 2002 et de 115 accessions en 2010. Pour étudier l'évolution phénotypique de cette population dans un contexte écologiquement réaliste, une expérience de phénotypage de 29 traits phénotypiques sur presque 6000 plantes a été mise en place *in situ* avant mon arrivée en thèse. Pour tenir compte de l'hétérogénéité abiotique et biotique rencontrée dans cette population, les 195 accessions ont été phénotypées dans six micro-habitats différents, correspondant à la combinaison de trois types de sol et de la présence/absence du pâturin annuel (*Poa annua*), espèce fréquemment retrouvée dans la communauté végétale associée à la population TOU-A (Baron *et al.* 2015). Pour étudier l'évolution génomique de la population TOU-A, le génome des 195 accessions a été séquencé suivant la technologie Illumina. Ces données génomiques ont non seulement permis de réaliser des analyses de GWA mapping mais aussi d'effectuer un scan génomique des traces de sélection temporelle.

Ce chapitre vise donc à répondre à plusieurs questions : (i) observe-t-on une évolution phénotypique en moins de 8 générations? (ii) l'identité des traits phénotypiques qui évoluent et les vitesses d'évolution phénotypique sont-elles dépendantes des conditions abiotiques et/ou biotiques ?, et (iii) quelle est l'architecture génétique sous-jacente à l'évolution phénotypique ? Et plus particulièrement, quelle est l'importance du degré de pléiotropie génétique sur la dynamique adaptative de la population TOU-A ? En effet, il est prédit que les polymorphismes avec des degrés de pléiotropie intermédiaires seraient favorisés par la sélection naturelle, permettant ainsi d'atteindre un optimum phénotypique plus rapidement (Wang *et al.* 2010). Cependant, ces attendus théoriques n'ont jamais été validés expérimentalement.

NB : dans ce chapitre, mon travail a consisté (i) à phénotyper 23 traits phénotypiques sur environ 4000 plantes d'*A. thaliana* (les 2000 autres plantes ont été phénotypées par Cédric Glorieux (TR laboratoire EEP, Université de Lille)), (ii) à effectuer toutes les analyses statistiques pour étudier la variation phénotypique naturelle au sein de la population locale TOU-A, (iii) à estimer les vitesse d'évolution phénotypique, (iv) à extraire l'ADN des 195 accessions en collaboration avec Fabrice Roux, et (v) à effectuer les analyses GWA mapping en collaboration avec Cyril Libourel (doctorant au sein de l'équipe). Les cinq traits phénologiques et la survie ont été mesurées pendant l'expérience *in situ* par Fabrice Roux. Les analyses bioinformatiques pour identifier les SNPs le long du génome ont été réalisées par Sébastien Carrere (IR plate-forme de bioinformatique du LIPM). Les analyses concernant la pléiotropie ont été réalisées par Cyril Libourel. Le scan génomique de différenciation génétique au niveau temporel a été effectué par Miguel Navascuès et Renaud Vitalis (laboratoire CBGP, INRA Montpellier) ainsi que par Laurène Gay (laboratoire AGAP, INRA Montpellier).

Manuscrit

“Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time”

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.

Le manuscrit a été re-soumis à Nature Ecology and Evolution

B. Manuscrit: Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time

Léa Frachon,[1¶] Cyril Libourel,[1¶] Romain Villoutreix,[2] Sébastien Carrère,[1] Cédric Glorieux,[2] Carine Huard-Chauveau,[1] Miguel Navascués,[3,4] Laurène Gay,[5] Renaud Vitalis,[3,4] Etienne Baron,[2] Laurent Amsellem,[2] Olivier Bouchez,[6,7] Marie Vidal,[6,8] Valérie Le Corre,[9] Dominique Roby,[1] Joy Bergelson,[10] Fabrice Roux[1,2]*

Affiliations :

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

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

[3] INRA, UMR CBGP, F-34988 Montferrier-sur-Lez, France

[4] Institut de Biologie Computationnelle, F- 34095 Montpellier, France

[5] UMR AGAP, INRA, Montpellier, France

[6] INRA, GeT-PlaGe, Genotoul, Castanet-Tolosan, France

[7] GenPhySE, Université de Toulouse, INRA, INPT, INP-ENVT, Castanet Tolosan, France

[8] INRA, UAR1209, Castanet-Tolosan, France

[9] INRA, UMR1347 Agroécologie, France,

[10] Department of Ecology and Evolution, University of Chicago, Chicago, IL USA

[¶] These authors contributed equally to this work.

* To whom correspondence should be addressed. E-mail: fabrice.roux@toulouse.inra.fr

Rapid phenotypic evolution of quantitative traits can occur in natural populations on a timescale of decades or even years[1], but little is known about its underlying genetic architecture[2]. Theoretical investigations have revealed that genes with intermediate pleiotropy will, under certain conditions, drive adaptive evolution[3-4] but these predictions have rarely been tested, especially under ecologically realistic conditions. Here, we performed a resurrection experiment to compare the evolution of multiple traits across six *in situ* micro-habitats within a natural population of the plant *Arabidopsis thaliana*. We then used Genome Wide Association mapping to identify the SNPs associated with evolved and unevolved traits in each of these sites. Finally, a genome-wide analysis of temporal genetic differentiation allowed us to test for selection acting on these SNPs. Phenotypic evolution was consistent across all micro-habitats but GWAS revealed largely distinct genetic bases among sites. Adaptive evolutionary change was largely driven by rare QTLs with intermediate degrees of pleiotropy under strong selection; this pleiotropy was synergistic with the per-trait effect size of a SNP increasing with the degree of pleiotropy. In addition to these rare pleiotropic QTLs, weak selection was detected for frequent small micro-habitat-specific QTLs that shape single traits. In this French population, *A. thaliana* likely responded to both local warming and increased competition, in part mediated by central regulators of flowering time and circadian rhythm such as FLOWERING LOCUS C and TWIN SISTER OF FT. This genetic architecture, which includes both synergistic pleiotropic QTLs and distinct QTLs within particular micro-habitats, enables rapid phenotypic evolution while still maintaining genetic variation in wild populations.

Contemporary and rapid phenotypic evolution has been observed in many natural populations of plant and animal species[1,5], especially during invasion[6] and in response to both global climate change[7] and toxic pollution[8]. Although a handful of studies have identified the genetic architecture of contemporary adaptive evolution of qualitative traits (such as industrial melanism)[9] or single quantitative traits (such as herbicide detoxification in weeds or heavy-metal tolerance)[10,11], the genetic architecture of a suite of quantitative traits experiencing contemporary adaptive evolution remains largely unexplored.

Theoretical studies predict that the number and effect sizes of QTLs underlying multi-trait adaptive evolution depends, in part, on the magnitude of pleiotropy[3,4,12]. Based on Fisher's geometric model, in which every mutation potentially affects all traits, the rate of adaptation of a QTL should decrease with its degree of pleiotropy[4]. This results from the increased probability of antagonistic effects of a mutation when more traits are impacted. However; in contrast to the assumptions of the geometric model, laboratory studies have found an L-shaped distribution of the degree of pleiotropy such that most mutations affect only a small subset of traits[3,12]; this restricted pleiotropy should diminish the 'cost of complexity'. Of additional importance is the relationship between the degree of pleiotropy and the per-trait effect size of a mutation (termed pleiotropic scaling)[3,12]. Most theoretical models assume that the per-trait effect size of a mutation decreases (invariant total effect model) or remains constant (Euclidean superposition model) with the degree of pleiotropy[4]. However, laboratory studies have found synergistic pleiotropy in which the per-trait effect size of a mutation increases with the number of traits affected by that mutation[3]. The combination of restricted and synergistic pleiotropy leads to the prediction that

polymorphisms with intermediate degrees of pleiotropy, while rare, should have the highest rate of adaptive evolution[3,4]. This prediction is yet to be tested empirically.

In its more general sense, pleiotropy refers to the shared impact of SNPs. This can include the effect of a SNP on (i) alternative phenotypic traits in one environment, (ii) a single phenotypic trait among environments, or (iii) alternative traits in multiple environments. Because wild populations evolve in complex abiotic and biotic environments, an exploration of the role of pleiotropy therefore requires consideration of the role of spatial environmental heterogeneity. In particular, when the same SNPs are favored in distinct micro-habitats, then the suite of selective effects may combine to drive rapid adaptive evolution whereas competing demands on a SNP across micro-sites could inhibit adaptive evolution.

In this study, we aimed to describe the genetic architecture underlying rapid phenotypic evolution of multiple quantitative traits of the annual plant *A. thaliana in situ*. More specifically, we aimed to test whether intermediate degrees of synergistic pleiotropy drive contemporary evolution of *A. thaliana* within a local population evolving in a spatially abiotic and biotic heterogeneous environment.

RESULTS AND DISCUSSION

Our study focused on the local population TOU-A (East of France; **Supplementary Fig. 1**) that experienced an increase in mean annual temperature of more than 1°C over the last 30 years (**Supplementary Fig. 2**). The site occupancy by *A. thaliana* additionally increased between 2002 and 2007 and remained stable thereafter (**Supplementary Fig. 1**). Seeds of 80 and 115 individual plants (hereafter named accessions) were collected in 2002 and 2010, respectively. Previous studies conducted on accessions collected in 2002 showed that this population has an estimated outcrossing rate of 6%[13] and is highly diverse at both genetic (based on genotyping at 149 SNPs) and phenotypic levels[13-16]. In addition, the TOU-A population presents fine-scale spatial variation for a broad range of soil characteristics and is located between two permanent meadows dominated by grasses (**Supplementary Figs. 1 and 3**).

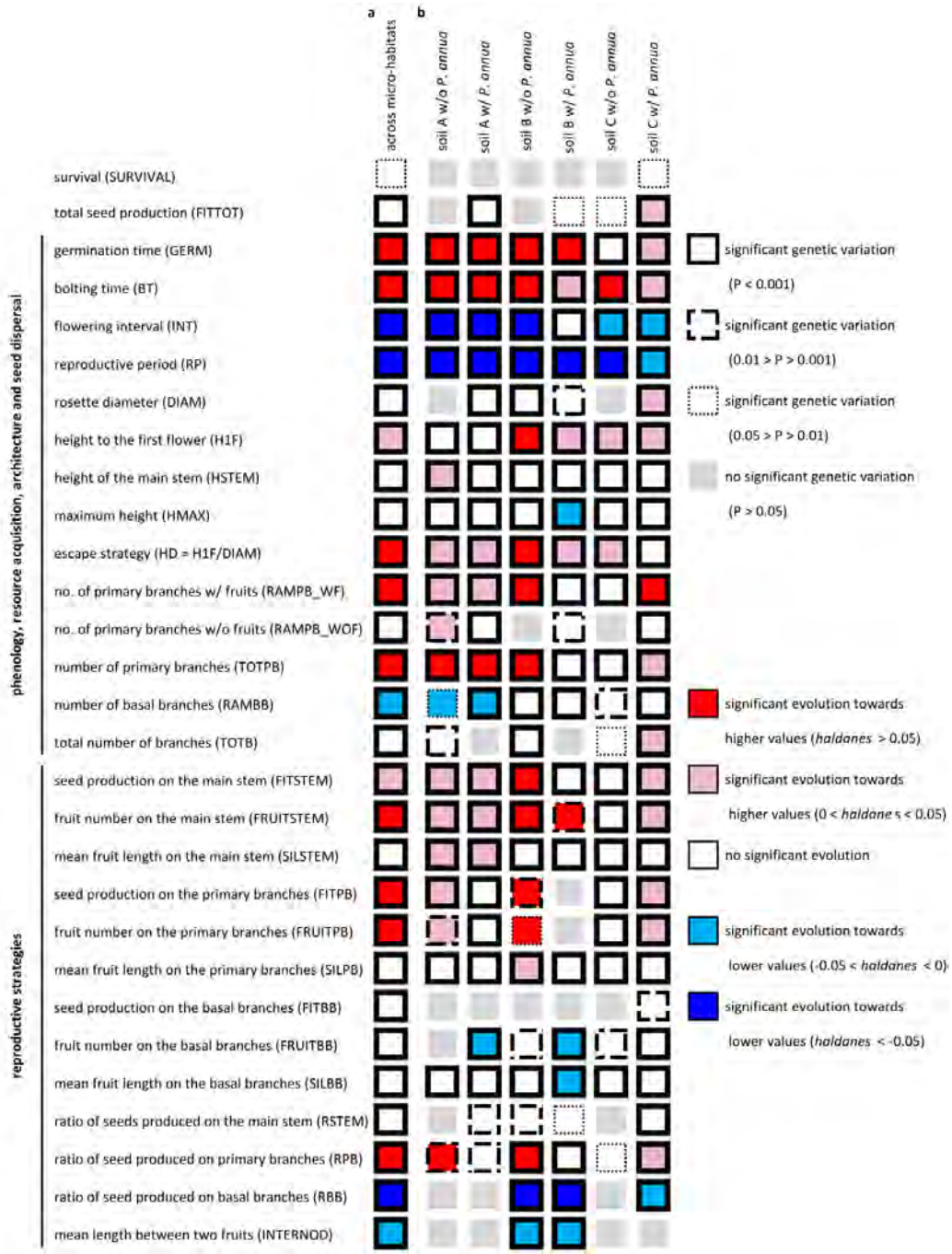


Figure 1 | Genetic variation among accessions and phenotypic evolution between 2002 and 2010. (**a**) Across the six micro-habitats. Genetic variation was detected for the 29 measured phenotypic traits. (**b**) Within each 'soil x competition' micro-habitat. The letters A, B and C stand for the three types of soil. 'w/o *P. annua*' and 'w/*P. annua*' correspond to the absence and presence of *P. annua*, respectively. The number of genetically variable traits varied between 21 (soil A in absence of *P. annua*) and 28 (soil C in presence of *P. annua*). The percentage of evolved genetically variable traits varied between 22.7% (soil C in absence of *P. annua*) and 76.2% (soil A in absence of *P. annua*). Each genetically variable trait (white and colored squares) in a given *in situ* experimental condition was defined as an eco-phenotype (n = 144). The rates of evolution are expressed in *haldanes* (a metric that scales the magnitude of change by incorporating trait standard deviations).

A resurrection experiment revealed rapid phenotypic evolution.

To identify phenotypic traits exhibiting evolutionary change within eight years, we established a resurrection experiment in which the 195 accessions collected in 2002 and 2010 were grown under common environmental conditions. This design enabled us to differentiate plastic from genetic responses[17]. After homogenizing for maternal effects, the 195 accessions were grown *in situ* in six representative micro-habitats, consisting of three contrasting soil types crossed with the presence or absence of the bluegrass *Poa annua*, a species frequently associated with *A. thaliana*[16] (**Supplementary Fig. 1**). A total of 5,850 plants were scored for 29 traits related to phenology, resource acquisition, shoot architecture, seed dispersal, fecundity, reproductive strategy and survival[18]. Interestingly, although no evolutionary change was observed for average total seed production across the six micro-habitats, we detected significant genetic evolution for 16 out of the 28 remaining traits (**Fig. 1a, Supplementary Table 1**). For example, we found a significant mean delay of 6.1 days for bolting time and a significant mean increase of ~7% in the number of fruits produced on the main stem (**Fig. 2a**). These results demonstrate that constant seed numbers can be maintained through evolution of flexible life-history and individual reproductive traits. A comparison of our results with the rates of evolution in other plant species[19] suggests a moderate rate of mean phenotypic evolution in the TOU-A population (**Fig. 2a**).

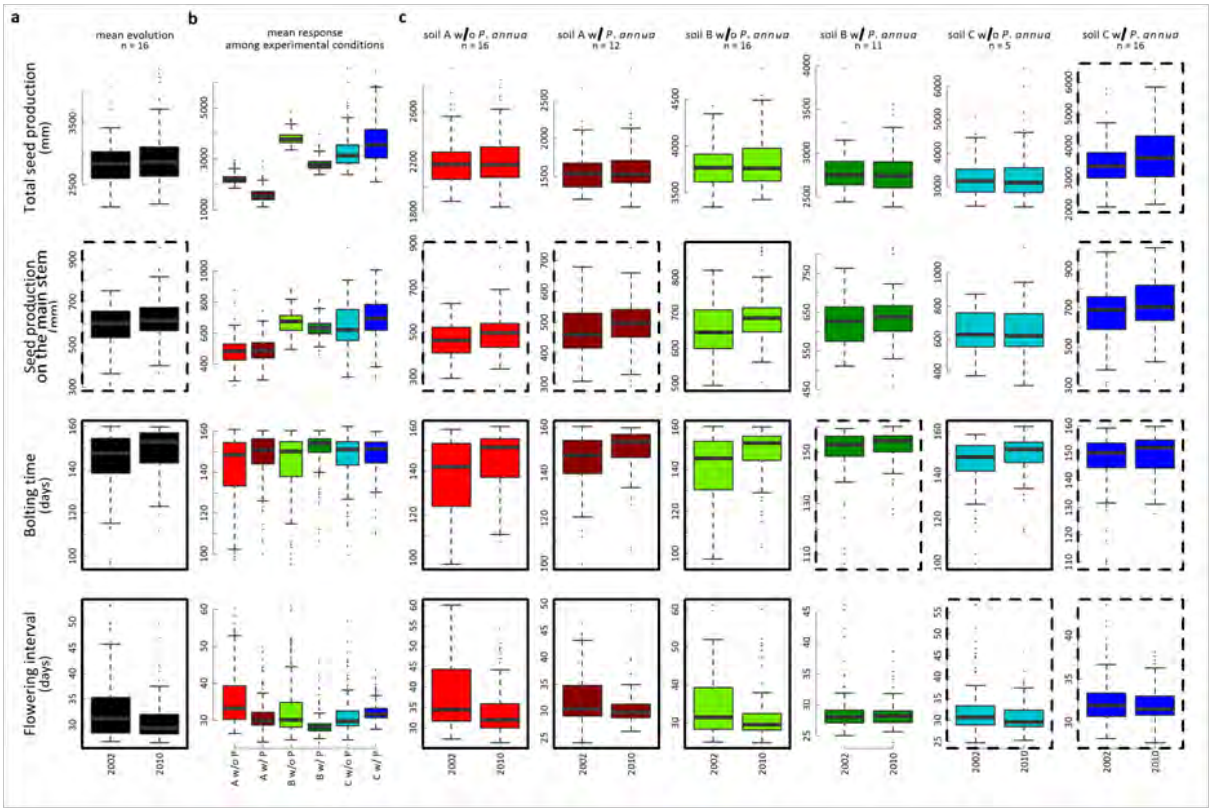


Figure 2 | Phenotypic changes in the TOU-A population over 8 generations. (**a**) Mean phenotypic evolution across the six micro-habitats. The total number of seeds produced can be maintained through evolution of phenological (bolting time and flowering interval) and individual reproductive (seed production on the main stem) traits. (**b**) Comparison among the six *in situ* 'soil x competition' micro-habitats. Average values of the phenotypes differed substantially among the six micro-habitats. (**c**) Evolution within each *in situ* micro-habitat. 'n' indicates the number of evolved phenotypic traits (**Fig. 1**). The identity of genetically variable traits that evolved between 2002 and 2008 depended on the micro-habitat. Each box plot is based on the genotypic values (BLUPs) of the TOU-A accessions (year 2002: n = 80, year 2010: n = 115). (**b**) and (**c**) The letters A, B and C stand for the three types of soil. 'w/o *P. annua*' and 'w/*P. annua*' correspond to the absence and presence of *P. annua*, respectively. (**a**) and (**c**): solid and dashed boxes indicate significant evolution with absolute *haldanes* > 0.05 and with absolute *haldanes* < 0.05, respectively (**Fig. 1**).

To confirm that the mean phenotypic change we observed was not the result of immigration from other phenotypically diverse populations[20], we sequenced the genomes of the 195 accessions collected in 2002 and 2010 (~25x coverage). We detected 1,902,592 Single Nucleotide Polymorphisms, only 5.6 times less than observed in a panel of 1135 worldwide accessions[21]. In addition, the TOU-A population appears strongly genetically isolated from other local populations sampled within 1km (**Fig. 3a**), confirming the negligible role of immigration in the observed phenotypic change.

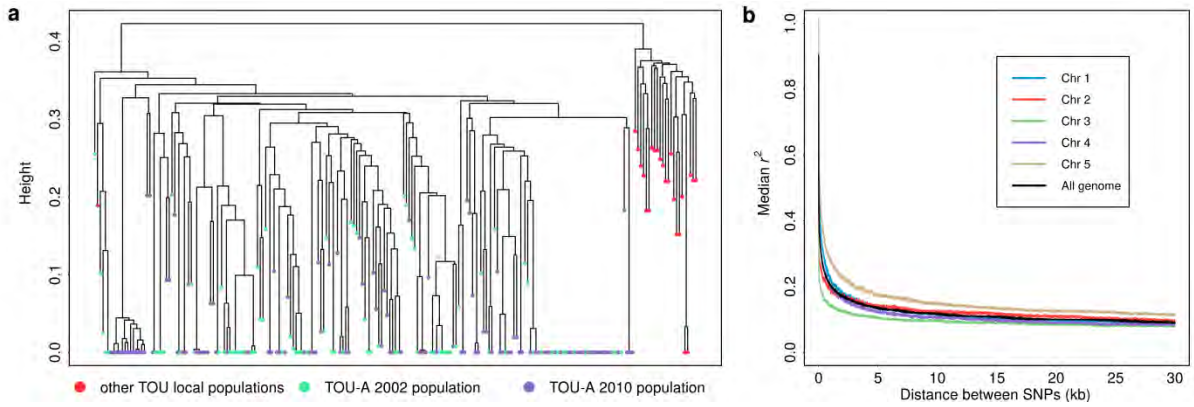


Figure 3 | Genomic patterns of the TOU-A population. (**a**) Hierarchical clustering analysis of the 195 TOU-A accessions and 24 accessions from 10 populations located within 1 km of the TOU-A population. (**b**) Decay of linkage disequilibrium (r^2) with physical distance over the five chromosomes of *A. thaliana*.

Similar phenotypic evolution associated with strong genotype-by-environment interactions.

We dissected the phenotypic evolution within each micro-habitat to test whether local abiotic and biotic growing conditions affect the genotype-phenotype relationships in the TOU-A population. Across the 29 traits measured in the six micro-habitats, 144 of these 174 eco-phenotypes displayed significant genetic variance (**Fig. 1b**), with broad-sense

heritability estimates ranging from 0.20 to 0.87 (mean H^2 = 0.57, median H^2 = 0.60; **Supplementary Table 2**). Average values of the phenotypes differed substantially among the six micro-habitats (**Fig. 2b, Supplementary Table 1**). The proportions (ranging from 22.7% to 76.2%) and identities of genetically variable traits that evolved in our eight-year timespan also depended on the micro-habitat (**Figs. 1b and 2c**). These results highlight the need to consider fine-scale environmental conditions to obtain an accurate picture of the diversity of micro-evolutionary phenotypic processes occurring within a population.

Although each trait that evolved was consistent in its direction in all micro-habitats (**Fig. 1b**), we observed highly significant changes in the ranking of accessions among micro-habitats that resulted from genotype-by-environment interactions for most traits (**Supplementary Table 1**). For example, increased allocation of reproduction to the main stem was consistently observed but different accessions most strongly manifested this allocation pattern among micro-habitats (**Supplementary Fig. 4**). These results are in accordance with previous studies revealing genotype-by-environment interactions for plant fitness-related traits at the scale of a few meters[22,23]. However, the existence of genotype-by-environment interactions does not clarify the extent of pleiotropy governing phenotypes in alternative micro-habitats: phenotypic evolution toward the same optimum may be driven by loci harboring alleles differing in the magnitude of allelic effects across micro-habitats and/or by distinct genetic bases in different micro-habitats[24].

Pleiotropy is restricted and synergistic

To characterize the genetics underlying these environmentally dependent genotype-phenotype relationships, we used GWA mapping to determine the genetic architecture, the

magnitude of pleiotropy and the extent of pleiotropic scaling. The TOU-A population is well-suited for GWA mapping because it is phenotypically diverse and linkage disequilibrium (LD) decays to r^2 = 0.5 within 18 base pairs on average (**Fig. 3b**). In agreement with limited LD, we observed an L-shaped distribution of the size of LD blocks, with a median size of 780bp (mean size = 5.5kb) (**Supplementary Fig. 5**). To verify our ability to finely map genomic regions associated with phenotypic variation, we first tested for the presence of significant associations of known functional polymorphisms. We successfully identified three known functional genes conferring either qualitative or quantitative resistance against bacterial pathogens when the 195 TOU-A accessions were infected under controlled conditions. In two of the three cases, the most highly associated SNP (hereafter named top SNP) was located within the gene (*RPS2* and *RKS1*)[15,25] and in the third case it was located 15 bp away (*RPM1*)[26] (**Supplementary Fig. 6**).

To further assess the efficacy of GWAS mapping in the TOU-A population, we followed the methodology used in Brachi *et al.* (2010)[27] to calculate enrichments for *a priori* candidate genes for bolting time in the six *in situ* micro-habitats (**Fig. 1**). Because bolting time is a quantitative trait for which the genetic network has been extensively studied, it is well suited for calculating enrichments for *a priori* candidate genes. Similar to previous results for a field trial utilizing 197 worldwide accessions[27], the enrichment ratio quickly dropped with the number of top SNPs in five out of the six micro-habitats, demonstrating that candidate genes were overrepresented among top-ranking SNPs (**Fig. 4a, Supplementary Fig. 7**).

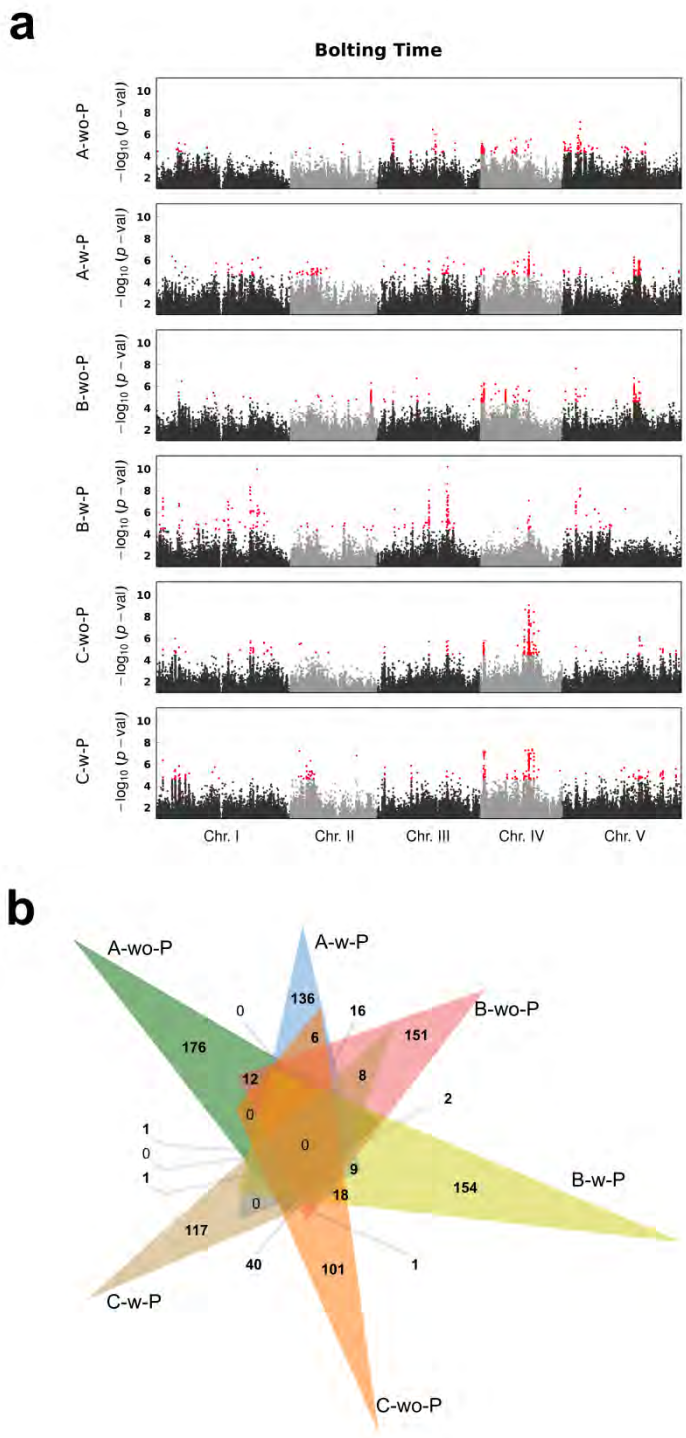


Figure 4 | Identification of genomic regions associated with bolting time variation in the TOU-A population. (**a**) Manhattan plots of mapping results for each of the six *in situ* 'soil x competition' treatments. The *x*-axis indicates the physical position along the chromosome. The *y*-axis indicates the -$\log_{10}$ *p*-values using the EMMAX method. MARF > 7%. For each experimental condition, the 200 top SNPs are highlighted in red. (**b**) Venn diagram partitioning the bolting time SNPs detected among the lists of 200 top SNPs for each *in situ* 'soil x competition' treatment. Genetic bases underlying bolting time are largely distinct across micro-habitats.

Here, we illustrate the impacts of genetic architecture, magnitude of pleiotropy and pleiotropic scaling when considering the 200 top SNPs (0.01% of the total number of SNPs) for each of the 144 eco-phenotypes. Although we observed significant enrichment for up to the 500 SNPs, focus on the 200 top SNPs is conservative in defining pleiotropy and increases the fraction of true positives. Our choice of threshold does not matter: our biological conclusions are robust to successive cutoffs of top SNPs within the range of 50-500 SNPs, and to three successive cutoffs in terms of the significance of SNPs ($-\log_{10}$ *p*-value > 6, $-\log_{10}$ *p*-value > 5, $-\log_{10}$ *p*-value > 4; chosen based on van Rooijen *et al.* 2015, Thoen *et al.* 2016, Kooke *et al.* 2017)[28-30].

We first compared the genetic architecture among micro-habitats for GWA results from each of the 144 heritable eco-phenotypes (**Supplementary Fig. 8**). The number of genes located within 2kb of the 200 top SNPs ranged from 45 (fruit number on basal branches in soil B with *P. annua*) to 141 (maximum height scored in soil B without *P. annua*) (mean = 105 genes, median = 108 genes; **Supplementary Fig. 9**). For a given phenotypic trait, the numbers of associated genes sometimes varied widely across micro-habitats, even when broad-sense heritabilities were similar (**Fig. 4a, Supplementary Fig. 9, Supplementary Table 2**).

The extent of pleiotropy for each top SNP was determined by calculating an effective number of eco-phenotypes, N_{eff}, sharing a given top SNP according to Pavlicev *et al.* (2009)[31]. This statistic corrects for correlations among eco-phenotypes to produce a measure of pleiotropy that is not inflated. In agreement with previous laboratory observations on yeast, nematode and mouse[3], we found that N_{eff} follows an L-shaped distribution (**Fig. 5a**). More than 78% of top SNPs impacted a single trait in a single micro-habitat, indicating that genetic

bases are largely distinct across micro-habitats (**Supplementary Fig. 10 and 11**), as illustrated for bolting time (**Fig. 4b**). This pattern of restricted pleiotropy in our study is more consistent with the notion of modular pleiotropy (with genes being organized into structured networks) than universal pleiotropy in Fisher's geometric model (i.e. each gene affects every trait)[3,4].

We found that the total effect size of a top SNP, calculated by either the Manhattan distance (T_M) or the Euclidean distance (T_E), increased with N_{eff} faster than linearly ($T_M = c^* {N_{eff}}^d$, d = 1.226 ± 0.003; $T_E = a^* Neff^b$, b = 0.724 ± 0.0035; **Fig. 5b, Supplementary Fig. 11 and 12, Supplementary Tables 3 and 4**). This contrasts with most theoretical models, which typically assume that the per-trait effect size of a mutation decreases (d = 0.5 or b = 0, invariant total effect model) or remains constant (d = 1 or b = 0.5, Euclidean superposition model) with the degree of pleiotropy[4]. While previously observed in controlled laboratory conditions[3], our study reveals that such a pattern of synergistic pleiotropy can also extend to phenotypes scored in ecological realistic conditions.

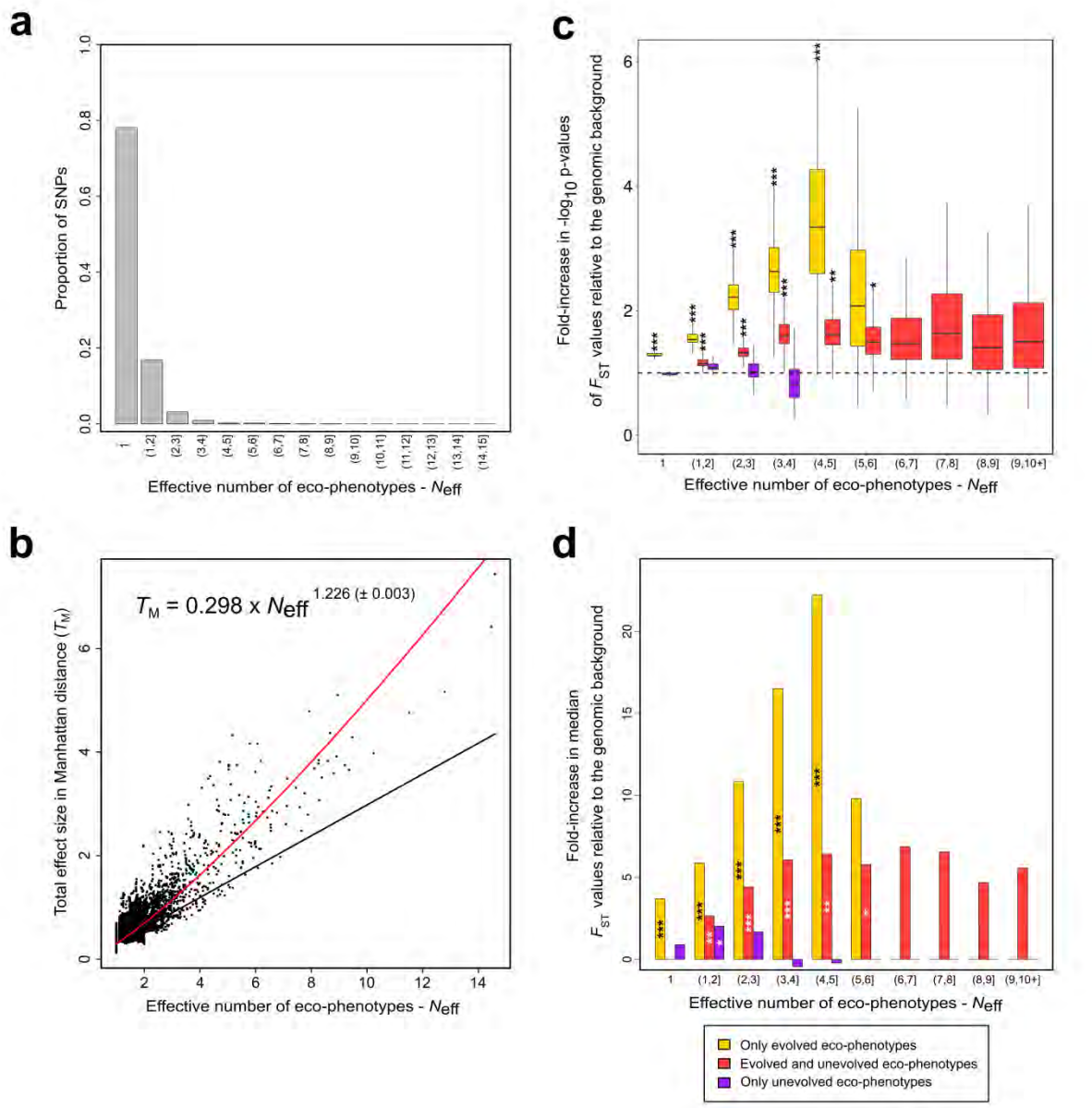


Figure 5 | Genetic architecture underlying *in situ* phenotypic evolution in the TOU-A population when considering a threshold of 200 top SNPs. (**a**) Frequency distribution of the effective number of eco-phenotypes affected by a SNP (N_{eff}, accounting for the correlations between eco-phenotypes)[31] among the 21,268 unique top SNPs. (**b**) Regression of total effect size T_M (total effect size by the Manhattan distance) on N_{eff}. The formula corresponds to the pleiotropic scaling relationship $\boldsymbol{T_M = c^* N_{eff}^{\ d}}$. A scaling component *d* exceeding 1 indicates that the mean per-trait effect size of a given top SNP increased with N_{eff}[3,4]. Solid red line: fitted relationship between T_M and N_{eff}, solid black line: linear dependence (d = 1). (**c**) Fold-increase in median $-\log_{10}$ (*p*-values) of neutrality tests based on temporal differentiation for SNPs that hit only evolved eco-phenotypes, only unevolved eco-phenotypes or both types of eco-phenotypes, according to different classes of effective number of eco-phenotypes. The dashed line corresponds to a fold-increase of 1, i.e. no increase in median significance of neutrality tests based on temporal differentiation. (**d**) Fold-increase in median F_{ST} values for SNPs that hit only evolved eco-phenotypes, only unevolved eco-phenotypes or both types of eco-phenotypes, according to different classes of N_{eff} (median F_{ST} across the genome = 0.00293). Significance against a null distribution obtained by bootstrapping: *0.05 > *P* > 0.01, **0.01 > *P* > 0.001, ***P < 0.001, absence of symbols: non-significant.

Intermediate degrees of synergistic pleiotropy drive adaptive evolution.

According to theoretical predictions[3,4], the combination of an L-shape distribution of N_{eff} and synergistic pleiotropy should lead polymorphisms with intermediate degrees of pleiotropy, while rare, to experience the highest rates of adaptive evolution. A genome-wide scan for selection based on temporal differentiation (F_{ST}) (**Supplementary Fig. 13**) revealed a signature of selection for top SNPs associated with evolved eco-phenotypes, but not for top SNPs associated with unevolved eco-phenotypes; top SNPs jointly associated with evolved and unevolved eco-phenotypes revealed an intermediate signature of selection (**Fig. 5c, Supplementary Fig. 11**). Because this temporal differentiation is tested against changes in the genomic background, this result rejects the hypothesis of selectively neutral evolution for evolved eco-phenotypes. When focusing attention on top SNPs associated with evolved eco-phenotypes, we found that single-trait micro-habitat-specific SNPs were under weak selection while SNPs exhibiting an intermediate degree of pleiotropy revealed the largest fold-increase of median temporal F_{ST} values (**Fig. 5d, Supplementary Fig. 11**). This pattern is strengthened when considering only the top SNPs for evolved phenotypes that have a polarity of effects in line with the direction of phenotypic evolution (~75.4% of the total number of top SNPs associated with evolved eco-phenotypes; **Supplementary Fig. 14**). Altogether, these results confirm that the evolved multi-trait combinations identified *in situ* are under selection.

As previously highlighted for the patterns of restricted pleiotropy and synergistic pleiotropy, the relationships between degree of pleiotropy and signatures of selection were robust to different number of top SNPs and thresholds of significance (within the range considered; **Supplementary Fig. 11**).

Identity of candidate genes under directional selection.

The most pleiotropic genes underlying adaptive evolution in the TOU-A population, were determined by retrieving all genes associated with 11 or more evolved eco-phenotypes. Among the 14 candidate genes (**Supplementary Table 5**), was the floral integrator *TWIN SISTER OF FT* (*TSF*), which was associated with bolting time (three microhabitats), flowering interval (one micro-habitat), the length of reproductive period (three micro-habitats), the number of primary branches (one micro-habitat) and the escape strategy (three micro-habitats). Interestingly, based on a panel of 948 worldwide accessions of *A. thaliana*, *TSF* has been found to be significantly associated with climate variation (i.e. number of consecutive cold days)[32], suggesting that *TSF* may play a major role in the adaptation of *A. thaliana* to climate at different geographical scales.

We additionally tested for biological processes that were enriched in the extreme tail of our genome-wide temporal differentiation scan (**Supplementary Table 6**). In total, 24 biological processes were enriched, 15 of which were supported by genes associated with phenotypic traits measured in this study (**Supplementary Table 6**). Enrichment for vernalization response was supported by *VERNALIZATION2* (*VRN2*) associated with six eco-phenotypes including two proxies of fitness (i.e. survival and seed production, **Supplementary Table 6**). We also detected many related, enriched functions such as stamen development, pollen maturation and callose deposition (**Supplementary Table 6**), which are consistent with the simultaneous evolution of fecundity traits observed in this study (**Fig. 1**). For instance, the candidate gene *POWDERY MILDEW RESISTANT 4* is traditionally regarded as a defense response to wounding and pathogens due to its role in reinforcing the cell wall, although it is also essential for pollen viability and cell division[33]. In this study, *POWDERY*

MILDEW RESISTANT 4 was associated with two fecundity traits: mean fruit length on primary branches (in soil A without *P. annua*) and the number of fruits on the main stem (in soil C with *P. annua*; **Supplementary Table 6**). The simultaneous evolution of fecundity traits suggests an adaptive strategy of short-lived semelparous species like *A. thaliana* in crowded environments, where plants tend to escape competition[16,34]. In agreement with this hypothesis, we observed an evolution of the escape strategy trait in five out of six micro-habitats (**Fig. 1**).

The remaining nine enriched biological processes were supported by genes that were not associated with any measured phenotype. This is not surprising in that we missed the entire seed and seedling stage, and did not capture the entire suite of biotic and abiotic factors that can impact selection over time. Among these candidate genes was the MADS-box transcription factor *FLOWERING LOCUS C* (*FLC*) that, in agreement with the recent local warming experienced by the TOU-A population, supported the strong enrichment detected for vernalization response, response to temperature stimulus and regulation of circadian rhythm (**Supplementary Table 6**). *FLC* is a well-known pleiotropic gene[35] that affects many traits that we did not measure (such as vernalization response, water use efficiency and regulation of seed dormancy by maternal temperature)[36-39], suggesting that one or more of these traits may have undergone contemporary and rapid phenotypic evolution in the TOU-A population.

It is interesting to note that we identified two central regulators of flowering time and circadian clock in our set of candidate pleiotropic genes, *i.e. FLC* and *TSF*. In two *Brassica rapa* populations that evolved rapidly following drought in Southern California[40], rapid evolution was in part mediated by a homologue of *SUPPRESSOR OF OVEREXPRESSION OF*

CONSTANS 1 (*SOC1*), a target of *FLC*-mediated transcriptional repression[41], suggesting that central regulators of flowering time and circadian clock play a major role in the response to global warming.

CONCLUSION

Our ecological genomic comparison of plants separated by eight generations revealed rapid multi-trait adaptive evolution that was similar among six micro-habitats, but largely mediated by different genes. The strong genotype-by-environment interactions highlight the importance of considering fine-scale ecological variation. By limiting the erosion of standing genetic variation, this micro-habitat dependent genetic architecture should allow populations like TOU-A to continue to respond to future environmental changes.

In addition, the combination of GWAS and an *in situ* resurrection experiment validated the prediction that polymorphisms with intermediate degrees of pleiotropy, while rare, should have the highest rate of adaptive evolution. This result reinforces the importance of simultaneous evolution of multiple traits in shaping the genomic adaptive trajectory of natural populations. On-going resurrection projects in plants[42] and long-term population surveys of wild animals[43] represent an exciting opportunity to test whether restricted pleiotropy combined with synergistic pleiotropy also underlies contemporary and rapid adaptive evolution in other plant and animal species.

ACKNOWLEDGEMENTS

This work was funded by the Région Midi-Pyrénées (CLIMARES project), the INRA Santé des Plantes et Environnement department (RESURRECTION project), the INRA-ACCAF metaprogram (SELFADAPT project), the LABEX TULIP (ANR-10-LABX-41, ANR-11-IDEX-0002-02) and the National Institute of Health.

AUTHOR CONTRIBUTIONS

F.R. supervised the project. F.R. conceived of and designed the experiments. E.B., L.A., Ro.V and F.R. conducted the *in situ* experiment. L.F., C.G., C.H.C. and F.R. measured the phenotypic traits. L.F. and F.R. analyzed the phenotypic traits. O.B. and M.V. generated the sequencing data. S.B. and C.L. performed the bioinformatics analyses. L.F., C.L. and F.R. performed the GWA mapping. L.F., C.L., D.R. and F.R. performed and analyzed the enrichment tests. M.N., L.G. and Re.V. developed a methodology in selfing species to perform a genome-wide scan for selection based on temporal differentiation. V.L.C and J.B guided the analysis of phenotypic and genomic data. F.R. and J.B. wrote the manuscript, with contributions from L.F., C.L, Ro.V., M.N., L.G., Re.V. and D.R. All authors contributed to the revisions.

REFERENCES

1. Franks, S.J., Weber, J.J. & Aitken, S.N. Evolutionary and plastic responses to climate changes in terrestrial plant populations. *Evol. Appl.* 7, 123-139 (2014).
2. Bay, A.B., Rose, N., Barrett, R., Bernatchez, L., Ghalambor, C.K., Lasky, J.R., Brem, R.B., Palumbi, S.R. & Ralph, P. Predicting responses to contemporary environmental change using evolutionary response architectures. *Am. Nat.* **189** (2017).
3. Wang, Z., Liao, B.-Y. & Zhang, J. Genomic patterns of pleiotropy and the evolution of complexity. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18034-18039 (2010).
4. Wagner, G.P. & Zhang, J. The pleiotropic structure of the genotype – phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.* **12,** 204-213 (2011).
5. DeLong, J.P., Forbes, V.E., Galic, N., Gibert, J.P., Laport, R.G., Phillips, J.S. & Vavra J.M. How fast is fast? Eco-evolutionary dynamics and rates of change in populations and phenotypes. *Ecol. Evol.* **6,** 573-581 (2016).
6. Buswell, J.M., Moles, A.T. & Hartley, S. Is rapid evolution common in introduced plant species. *J. Ecol.* **99,** 214-224 (2011).
7. Franks, S.J., Sim, S. & Weis, A.E. Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *Proc. Natl. Acad. Sci. U.S.A.* **104**: 1278-1282 (2007).
8. Reid, N.M., Proestou, D.A., Clark, B.W., Warren, W.C., Colbourne, J.K., Shaw, J.R., Karchner, S.I., Hahn, M.E., Nacci, D., Oleksiak, M.F., Crawford, D.L. & Whitehead, A. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science* **354,** 1305-1308.

9. van't Hof, A.E., Edmonds, N., Dalikova, M., Marec, F. & Saccheri, I.J. Industrial melanism in British peppered moths has a singular and recent mutational origin. *Science* **332,** 958-960.

10. Hanikenne, M., Talke, I.N., Haydon, M.J., Lanz, C., Nolte, A., Motte, P., Kroymann, J., Weigel, D. & Krämer, U. Evolution of metal hyperaccumulation required *cis*-regulatory changes and triplication of *HMA4*. *Nature* **453,** 391-395.

11. Délye, C., Jasieniuk, M. & Le Corre V. Deciphering the evolution of herbicide resistance in weeds. *Trends Genet.* **29,** 649-658.

12. Wagner, G.P. *et al.* Pleiotropic scaling of gene effects and the 'cost of complexity'. *Nature* **452,** 470-472 (2008).

13. Platt, A. *et al.* The scale of population structure in *Arabidopsis thaliana*. *PLoS Genet.* **6**, e1000843 (2010).

14. Brachi, B. *et al.* Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in *Arabidopsis thaliana*. *Mol. Ecol.* **22**, 4222-4240 (2013).

15. Huard-Chauveau, C. *et al.* An atypical kinase under balancing selection confers broad-spectrum disease resistance in Arabidopsis. *PLoS Genet.* **9**, e1003766 (2013).

16. Baron, E., Richirt, J., Villoutreix, R., Amsellem, L. & Roux, F. The genetics of intra- and interspecific competitive response and effect in a local population of an annual plant species. *Funct. Ecol.* **29**, 1361-1370 (2015).

17. Franks, S.J. *et al.* The resurrection initiative: storing ancestral genotypes to capture evolution in action. *BioScience* **58**, 870-873 (2008).

18. Roux, F., Mary-Huard, T., Barillot, E., Wenes, E., Botran, L., Durand, S., Villoutreix, R., Martin-Magniette, M.-L., Camilleri, C. & Budar, F. Cytonuclear interactions affect

adaptive phenotypic traits of the annual plant *Arabidopsis thaliana* in ecologically realistic conditions. *Proc. Natl. Acad. Sci. U.S.A.* **113**: 3687-3692 (2016).

19. Bone, E. & Farres, A. Trends and rates of microevolution in plants. *Genetica* **112-113**, 165-182 (2001).
20. Hansen, M.M., Olivieri, I., Waller, D.M., Nielsen, E.E. & The GeM working group. Monitoring adaptive genetic responses to environmental change. *Mol. Ecol.* **21**, 1311-1329 (2012).
21. The 1001 Genomes Consortium. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166**, 1-11 (2016).
22. Kalisz, S. Variable selection on the timing of germination in *Collinsia verna* (Scrophulariaceae). *Evolution* **40,** 479-491 (1986).
23. Stratton, D.A. Spatial scale of variation in fitness of *Erigeron annuus*. *Am. Nat.* **146,** 608-324 (1995).
24. Des Marais, D.L., Hernandez, K.M. & Juenger, T.E. Interaction and plasticity : exploring genomic responses of plant to the abiotic environment. *Annu. Rev. Ecol. Evol. Syst.* **44**, 5-29 (2013).
25. Bent, A.F., Kunkel, B.N., Dahlbeck, D., brown, K.L., Schmidt, R., Giraudat, J., Leung, J. & Staskawicz, B.J. *RPS2* of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. *Science* **265,** 1856-1860 (1994).
26. Grant, M.R., Godirad, L., Straube, E., Ashfield, T., Lewald, J., Sattler, A., Innes, R.W & Dangl J.L. Structure of the Arabidopsis *RPM1* gene enabling dual specificity disease resistance. *Science* **269,** 843-846 (1995).
27. Brachi, B. *et al.* Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet.* **6:**e1000940 (2010).

28. Van Rooijen, R., Aarts, M.G.M. & Harbinson, J. Natural genetic variation for acclimation of photosynthetic light use efficiency to growth irradiance in Arabidopsis. *Plant Physiol.* **167,** 1412-1429 (2015).

29. Kooke, R., Kruijer, W., Bours, R., Becker, F., Kuhn, A., van de Geest, H., Buntjer, J., Doeswijk, T., Guerra, J., Bouwmeester, H., Vreugdenhil, D. & Keurentjes, J.B. Genome-wide association mapping and genomic prediction elucidate the genetic architecture of morphological traits in Arabidopsis. *Plant Physiol.* **170,** 2187-2203 (2016).

30. Thoen, M.P.M. *et al.* Genetic architecture of plant stress resistance: multi-trait genome-wide association mapping. *New Phytol.* **213,** 1346-1362 (2016).

31. Pavlicev, M., Cheverud, J.M. & Wagner, G.P. Measuring morphological integration using eigenvalues variance. *Evol. Biol.* **36,** 157-170 (2009).

32. Hancock, A.M. *et al.* Adaptation to climate across the *Arabidopsis thaliana* genome. Science **334**, 83-86 (2011).

33. Ellinger, D. & Voigt, C.A. Callose biosynthesis in Arabidopsis with a focus on pathogen response: what we have learned with the last decade. *Annals of Botany* **114**, 1349-1358 (2014).

34. Bonser, S.P. High reproduction efficiency as an adaptive strategy in competitive environments. *Funct. Ecol.* **27**, 876-885 (2013).

35. Deng, W. *et al. FLOWERING LOCUS C* (*FLC*) regulates development pathways throughout the life cycle of *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 6680-6685 (2011).

36. McKay, J.K., Richards, H. & T. Mitchell-Olds. Genetics of drought adaptation in *Arabidopsis thaliana*: I. Pleiotropy contributes to genetic correlations among ecological traits. *Mol. Ecol.* **12,** 1137-1151 (2013).

37. Li, P., Filiault, D., Box, M.S., Kerdaffrec, E., van Oosterhout, C., Wilczek, A.M., Schmitt, J., McMullan, M., Bergelson, J., Nordborg, M. & Dean, C. Multiple *FLC* haplotypes defined by independent *cis*-regulatory variation underpin life history diversity in *Arabidopsis thaliana*. *Genes Dev.* **28,** 1635-1640 (2014).

38. Blair, L., Auge, G. & Donohue, K. Effect of *FLOWERING LOCUS C* on seed germination depends on dormancy. *Funct. Plant Biol.* **44,** 493-506 (2017).

39. Auge, G., Blair, L.K., Neville, H. & Donohue, K. Maternal vernalization and vernalization-pathway genes influence seed germination. *New Phytol.* doi: 10.1111/nph.14520 (2017).

40. Franks, S.J., Kane, N.C., O'Hara, N.B., Tittes, S. & Rest, J.S. Rapid genome-wide evolution in *Brassica rapa* populations following drought revealed by sequencing of ancestral and descendant gene pools. *Mol. Ecol.* **25,** 3622-3631 (2016).

41. Salathia, N. *et al. FLOWERING LOCUS C* - dependent and – independent regulation of the circadian clock by the autonomous and vernalization pathways. *BMC Plant Biol.* **6,** 10 (2006).

42. Etterson, J.R. *et al.* Project Basline : an unprecedented resoruce to study plant evolution across space and time. *Am. J. Bot.* **103,** 164-173 (2016).

43. Kruuk, L.E.B., Garant, D. & Charmantier, A. The study of quantitative genetics in wild populations. Pages 1-15 *In Quantitative genetics in wild populations*. Edited by A. Charmantier, D. Garant and L.E.B. Kruuk. Oxford University Press, Oxford, U.K. pp 1-15 (2014).

METHODS

Plant material. The population TOU-A is located under a 350m electric fence separating two permanent meadows experiencing cycles of periodic grazing by cattle in the village of Toulon-sur-Arroux (France, Burgundy, N 46°38'57.302'', E 4°7'16.892''). Seeds from individual plants were collected in 2002 (TOU-A-2002, n = 80) and 2010 (TOU-A-2010, n = 115) according to a sampling scheme allowing us to take into account the density of *A. thaliana* plants along a 350m transect (**Supplementary Fig. 1**). Differences in maternal effects among the 195 accessions collected in 2002 and 2010 were reduced by growing one plant per family under controlled greenhouse conditions, for one generation (16-h photoperiod, 20°C).

Ecological characterization. Eighty-three soil samples collected along the 350m transect were characterized for 14 edaphic factors[14]: pH, maximal water holding capacity (WHC), total nitrogen content (N), organic carbon content (C), C/N ratio, soil organic matter content (SOM), concentrations of P_2O_5, K, Ca, Mg, Mn, Al, Na and Fe. Climate data was generated with the ClimateEU v4.63 software package[44].

Phenotypic characterization. An experiment of 5,850 plants was set up at the local site of the TOU-A population. The 195 accessions collected in 2002 and 2010 were grown in six representative 'soil x competition' micro-habitats. Each of these micro-habitats was organized in five blocks. Each of the five blocks corresponded to an independent randomization of 195 plants with one replicate per accession collected in 2002 and 2010. Seeds were sown in late September to mimic the main natural germination cohort observed in the TOU-A population (**Supplementary Fig. 1**). Each plant was scored for a total of 29 phenotypic traits chosen to characterize the life history of *A. thaliana* including the timing of

offspring production or seed dispersal, or because they are involved in the response to competition and/or are good estimators of life-time fitness and reproductive strategies[18].

Phenotypic analyses, natural variation, phenotypic evolution and evolutionary rates. We explored natural variation of all phenotypic traits using the following statistical mixed model:

$$Y_{ijklm} = \mu_{trait} + block_i(soil_j * comp_k) + soil_j + comp_k + soil_j * comp_k + year_l + soil_j * year_l + comp_k * year_l + soil_j * comp_k * year_l + accession_m(year_l)) + accession_m(year_l)) * soil_j + accession_m(year_l)) * comp_k + accession_m(year_l)) * soil_j * comp_k + \varepsilon_{ijklm} \quad (1)$$

In this model, '*Y*' is one of the 29 phenotypic traits, 'μ' is the overall phenotypic mean; 'block' accounts for differences between the five experimental blocks within each type of 'soil * absence/presence of *P. annua*' experimental combination; 'soil' corresponds to the effects of the three types of soil; 'comp' measure the effect of the presence of *P. annua*; 'year' corresponds to effect of the two sampling years 2002 and 2010; 'accession' measures the effect of accessions within year; interaction terms involving the 'accession' term account for genetic variation in reaction norms of accessions between the three types of soil and the absence or presence of *P. annua*; and 'ε' is the residual term.

All factors were treated as fixed effects, except 'accession' that was treated as a random effect. For fixed effects, terms were tested over their appropriate denominators for calculating *F*-values. Significance of the random effects was determined by likelihood ratio tests of model with and without these effects. When necessary, raw data were either log transformed or Box-Cox transformed to satisfy the normality and equal variance

assumptions of linear regression. A correction for the number of tests was performed for each modeled effect to control the False Discovery Rate (FDR) at a nominal level of 5%.

Inference was performed using ReML estimation, using the PROC MIXED procedure in SAS 9.3 (SAS Institute Inc., Cary, North Carolina, USA) for all traits with the exception of SURVIVAL, which was analyzed using the PROC GLIMMIX procedure in SAS 9.3.

For all traits, Best Linear Unbiased Predictions (BLUPs) were obtained for each accession in each of the six experimental conditions, using the PROC MIXED procedure in SAS 9.3 (SAS Institute Inc., Cary, North Carolina, USA):

$$Y_{imc} = \mu_{trait} + block_i + accession_m + \varepsilon_{im} \quad (2)$$

For each trait, significant genetic variation among the accessions was detected by testing the significance of the 'accession' term in equation (2). A correction for the number of tests was performed for the modeled 'accession' effect (across the 29 traits within each of the six experimental conditions) to control the FDR at a nominal level of 5%. Because *A. thaliana* is a highly selfing species[13], BLUPs correspond to the genotypic values of accessions.

In each of the six experimental conditions, rates of evolutionary change based on genotypic values of accessions were calculated in *haldanes* (h_g) for all eco-phenotypes with significant genetic variation among the 195 accessions collected in 2002 and 2010. *haldanes* is a metric that scales the magnitude of change by incorporating trait standard deviations[45,46]. h_g values were calculated between 2002 and 2010, as:

$$h_g = \frac{(x_2/s_p) - (x_1/s_p)}{g} \qquad (3)$$

where '*x*' corresponds to the mean genotypic value at year 1 (TOU-A population collected in 2002) and year 2 (TOU-A population collected in 2010), 's_p' is the standard deviation of the genotypic values of the trait pooled across the two years, and '*g*' is the number of generations. Because only one germination cohort was observed every year between 2002 and 2010 (i.e. fall germination cohort), only one generation per year was considered in the calculation of *haldanes* values. For a given trait, 95% confidence intervals were estimated based on the distribution of 1000 *haldanes* values obtained by bootstrapping 1000 random samplings with replacement of genetic values within each year. A *haldanes* value was considered significantly different from zero if its 95% confidence intervals did not overlap zero.

Sequencing and polymorphism detection. DNA-seq experiments were performed on an Illumina HiSeq2500 using a paired-end read length of 2x100 pb with the Illumina TruSeq SBS v3 Reagent Kits. Raw reads of each of the 195 accessions were mapped onto the TAIR10 *A. thaliana* reference genome Col-0 with a maximum of 5 mismatches on at least 80 nucleotides. A semi-stringent SNPCalling across the genome was then performed for each accession with SAMtools mpileup (v0.01019)[47] and VarScan (v2.3)[48] with the parameters corresponding to a theoretical sequencing coverage of 30X and the search for homozygous sites.

Patterns of linkage disequilibrium and geographic structure. Considering only SNPs with a Minor Allele Relative Frequency (MARF) > 0.07, the LD extent within 30kb-windows on each chromosome were estimated using *VCFtools*[49]. LD blocks across the genome were identified

in the PLINK environment using the following parameters --blocks no-pheno-req --maf 0.07 --blocks-max-kb 200, leading to the identification of 19,607 blocks with at least two SNPs (mean number of SNPs per block = 47.6, median number of SNPs per block = 12, mean block length = 5.5kb, median block length = 0.78kb). To position the TOU-A population within the French geographic structure, we retrieved the positions of the 214,051 SNPs genotyped on 24 accessions within 10 populations located within 1km of the TOU-A population[50] across the genomes of the TOU-A population. Clustering genotype analysis was performed using the packages gdsfmt and SNPRelate in the *R* environment[51], using the snpgdspLD pruning command with the following parameters *ld.threshold=0.8 slide.max.bp=500 maf=0.07*, leaving us with 90,883 SNPs.

Genome-Wide Association mapping and MARF threshold. GWA mapping was run using a mixed-model approach implemented in the software EMMAX (Efficient Mixed-Model Association eXpedited)[52]. This model includes a genetic kinship matrix as a covariate to control for population structure.

Because of bias due to rare alleles[27,52,53], we estimated a MARF threshold above which the *p*-value distribution is not dependent on the MARF. We plotted the 99% quantile of the *p*-value distribution of all 144 eco-phenotypes (i.e. 'micro-habitat x trait' combinations) displaying significant genetic variance (**Fig. 1**) along 50 MARF values (with an increment of 0.01 from 0.01 to 0.5). A locally-weighted polynomial regression indicated that *p*-value distributions were dependent on MARF value. From visual inspection, we considered a threshold of MARF value > 0.07, which resulted in a total number of 981,617 SNPs for the following analyses (**Supplementary Fig. 15**).

Enrichment for *a priori* candidate genes. To determine the threshold number of top SNPs (i.e. SNPs with the highest associations) above which additional top SNPs would behave like

the rest of the genome, we calculated enrichments for *a priori* candidate genes for natural genetic variation of bolting time observed in the six *in situ* experimental conditions (**Fig. 1**). Based on an algorithm described in Brachi *et al.*(2010)[27] and a list of 328 candidate genes for bolting time[14], enrichment was calculated for progressively fewer selective sets of top SNPs within a 20Kb window of an *a priori* candidate gene. For each set of top SNPs, a null distribution of enrichment was computed to determine a 95% confidence interval[27].

Degree of pleiotropy and pleiotropic scaling. Each trait displaying significant genetic variance in a given *in situ* micro-habitat was considered an "eco-phenotype". The degree of pleiotropy of a given top SNP was defined as the number of eco-phenotypes that shared this top SNP. To account for the correlations between eco-phenotypes that can overestimate the degree of pleiotropy, we followed Wagner *et al.* (2008)[12] by estimating for each top SNP an effective number of eco-phenotypes as $N_{\text{eff}} = N - \text{var}(\lambda)$ where $\text{var}(\lambda)$ is the variance of the eigenvalues of the error-corrected matrix.

The allelic effects were calculated using the mixed model implemented in the software EMMAX after fitting the pairwise genetic kinship effect[52]. Because different units were used to measure the 29 traits scored in this study, we calculated a standardized allelic effect for each eco-phenotype affected by a top SNP according to Wagner *et al.* (2008)[12]. The standardized effect on eco-phenotype *i*, denoted by A_i, is half the difference in genotypic means between the two homozygous genotypes. The total size of the phenotypic effects of a top SNP was then calculated by the Manhattan distance[54] $T_{\text{M}} = \sum_{i=1}^{n} |A_i|$, where n is the degree of pleiotropy and A_i is the standardized allelic effect[3,4,12]. The pleiotropic scaling relationship between the total effect size and the effective number of eco-phenotypes was calculated as $T_{\text{M}} = c^* {N_{\text{eff}}}^d$.

The pleiotropic scaling relationship between the total effect size and the effective number of eco-phenotypes was also calculated as $T_E = a^* Neff^b$, with T_E corresponding to the Euclidean distance and calculated as $T_E = \sqrt{\sum_{i=1}^{n} A_i^2}$, where n is the degree of pleiotropy and A_i is the standardized allelic effect.

The degree of pleiotropy and the pleiotropic scaling relationship were calculated for (i) five threshold number of top SNPs (i.e. 50 SNPs, 100 SNPs, 200 SNPs, 300 SNPs and 500 SNPs) and (ii) three thresholds of significance ($-\log_{10}$ *p*-value > 6, $-\log_{10}$ *p*-value > 5, $-\log_{10}$ *p*-value > 4). To avoid pseudo-replication due to the presence of several top SNPs in a given LD block (n = 19,607 blocks with at least two SNPs), the pleiotropic scaling was also calculated for each threshold number of top SNPs and each threshold of significance, (i) by considering the mean value of T_M (or T_E) and N_{eff} per LD block containing top SNPs and (ii) by randomly sampling one top SNP per LD block (this step was repeated 1,000 times).

Genome-wide scan for selection based on temporal differentiation. In the following, we outline a procedure inspired by Goldringer & Bataillon (2004)[55] to test for the homogeneity of differentiation across SNP markers between two temporal samples. If all SNP markers are selectively neutral, they should provide estimates of temporal differentiation drawn from the same distribution, which depends on the strength of genetic drift in the population (and therefore on its effective size). In contrast, if some marker loci are targeted by selection (or if they are in linkage disequilibrium with selected variants), then some heterogeneity in locus-specific measures of temporal differentiation should be observed. This is due to selection that will tend to drive measures of differentiation to values greater (or smaller) than expected under drift alone. The rationale of our approach is therefore to identify those SNPs that show outstanding differentiation, compared to neutral expectation.

We measure temporal differentiation between sample pairs using F_{ST}. Although the F_C statistic[56] was used in Goldringer & Bataillon (2004)[55], estimators of F_{ST} have better statistical properties in terms of bias and variance, and multilocus estimates have been precisely defined and thoroughly evaluated[57].

Using a multilocus estimate of F_{ST} from the pair of temporal samples, we infer the effective size of the population. Because the 195 *A. thaliana* accessions are considered highly homozygous across the genome, heterozygous sites were discarded (see above) and the data therefore consist of haploid genotypes. We considered a single haploid population of constant size N_e, which has been sampled at generation 0, and τ generations later. Generations do not overlap. New mutations arise at a rate μ, and follow the infinite allele model (IAM). Following Skoglund *et al.* (2014)[58], the pairwise parameter F_{ST} between the two samples can be read:

$$F_{ST} = \frac{1 - e^{-\theta T/2}}{1 + \theta - e^{-\theta T/2}}$$

where $T \equiv \tau / N_e$ and $\vartheta \equiv 2N_e\mu$. In the low mutation limit (i.e., as $\mu \to 0$):

$$F_{ST} \approx \frac{T}{T+2} = \frac{\tau}{\tau + 4N_e}$$

This suggests that a simple moment-based estimator of effective population size can be derived as:

$$\hat{N}_e = \frac{\tau(1 - \hat{\bar{F}}_{ST})}{4\hat{\bar{F}}_{ST}}$$

where $\hat{\bar{F}}_{ST}$ is a multilocus estimate of the parameter F_{ST}. In what follows, we use the estimator of Weir & Cockerham (1984)[57]; preliminary analyses showed that these estimates of effective size have lower bias and variance than averaged estimates based on single-locus estimates of F_C.

In this study, the pairwise differentiation between the 195 *A. thaliana* accessions samples collected in 2002 and 2010 based on the full set of 1,902,592 SNP markers was: $\hat{\bar{F}}_{ST}$ = 0.0215, which gives an estimate of $\hat{N}_e$ = 182 (measured as a number of gene copies).

For each SNP, we tested the null hypothesis that the locus-specific differentiation measured at this focal marker was only due to genetic drift. For this purpose, we computed the expected distribution of F_{ST} for each SNP, conditional upon the estimated effective size (using the same estimated value for all markers: $\hat{N}_e$ = 182), and the allele frequencies at the focal SNP in the initial sample (i.e. 80 accessions collected in 2002). We simulated individual gene frequency trajectories, as follows:

Suppose that we observe k_0 copies of the reference allele, out of n_0 sampled genes, in the 2002 sample. We assume that these observed counts are drawn from a binomial distribution $B(n_0,\pi_0)$ where π_0 is the (unknown) allele frequency of the reference allele in the population. Assuming a Beta(1,1) prior distribution for π_0 (uniform distribution), and using the Bayes inversion formula, the posterior distribution of π_0 is a Beta(k_0 + 1,n_0 – k_0 + 1). For each marker and for each simulation, we therefore draw the initial allele frequency $\tilde{\pi}_0$ from a Beta(k_0 + 1,n_0 – k_0 + 1). We then draw "pseudo-observed" allele counts using a random draw from $B(n_0, \tilde{\pi}_0)$. This procedure allows accounting for the sampling variance in initial allele frequencies, instead of fixing $\tilde{\pi}_0$ to the observed frequency in the sample, as previously done in Goldringer & Bataillon (2004)[55].

Then, we simulated eight generations of drift, using successive binomial draws with parameters $\hat{N}_e$ = 182 and the allele frequency in the previous generation. In the last generation, a sample of genes is taken as a binomial draw with parameters n_τ (the sample

size in 2010), and $\tilde{\pi}_{\tau}$ (the simulated allele frequency in the last generation).

Last, we computed locus-specific estimates of temporal F_{ST} from the simulated allele counts at the initial and last generation. The whole procedure was repeated at least 10,000 times for each marker (additional simulations were performed for some markers to obtain non-null *p*-values).

Finally, we assigned a *p*-value to each SNP marker, computed as the proportion of simulations giving a locus-specific estimate of F_{ST} larger than or equal to the observed value at the focal SNP. We checked that the distribution of *p*-values was fairly uniform (data not shown).

Note that all SNP markers with a MARF ≤ 0.07 (computed as the overall frequency across the two temporal samples) were discarded from the analysis. There were 981,617 remaining loci (**Supplementary Fig. 7**). To avoid any potential bias, all the distributions of F_{ST} were obtained using only simulated markers with a MARF > 0.07.

Enrichment analysis of top SNPs for signals of selection. Based on the effective number of eco-phenotypes affected by a SNP, we tested whether top SNPs related to evolved eco-phenotypes rejected the hypothesis of selectively neutral evolution more often than top SNPs related to unevolved eco-phenotypes for any given degree of pleiotropy. For each set of top SNPs (i.e. top SNPs that hit only evolved eco-phenotypes, top SNPs that hit only unevolved eco-phenotypes and top SNPs that hit both types of eco-phenotypes), we first computed a fold-increase in median significance of F_{ST} values using the following ratio: $ratio_{significance}$ = median of $-\log_{10}$(*p*-values) of F_{ST} values of *n* top SNPs / median of $-\log_{10}$(*p*-values) of F_{ST} values of *n* SNPs randomly sampled across the genome, where *n* = number of top SNPs. This step was repeated 1,000 times, generating a distribution of fold-increase in

median significance of F_{ST} values of top SNPs. We assigned a *p*-value by computing the proportion of $ratio_{significance}$ smaller or equal to 1. The random sampling was done according to a scheme that results in sets of SNPs that resemble the original set with respect to linkage disequilibrium[32].

We then tested whether the strength of selection differed among the degrees of pleiotropy by computing a fold-increase in median F_{ST} values for each set of top SNPs, using the following ratio: $ratio_{values}$ = median of F_{ST} values of *n* top SNPs / median of F_{ST} values of all SNPs. This step was repeated 1,000 times, by randomly sampling the same number *n* of SNPs across the genome. This procedure generated a null distribution of fold-increase in median F_{ST} values. We assigned a *p*-value by comparing $ratio_{values}$ calculated for the set of top SNPs to the quantiles at 95%, 99% and 99.9% of the null distribution.

The enrichment analysis of top SNPs for signals of selection was calculated for (i) five threshold number of top SNPs (i.e. 50 SNPs, 100 SNPs, 200 SNPs, 300 SNPs and 500 SNPs) and (ii) three thresholds of significance ($-\log_{10}$ *p*-value > 6, $-\log_{10}$ *p*-value > 5, $-\log_{10}$ *p*-value > 4).

Identity of candidate genes under directional selection and enrichment in biological processes.

To identify pleiotropic candidate genes associated with the 76 evolved eco-phenotypes, we first selected the 50 SNPs the most associated with each evolved eco-phenotype, leading to a total of 3800 SNPs. We then retrieved all the annotated genes located within a 2kb window on each side of those top SNPs, leading to a final list of 4855 unique candidate genes. We finally focused on genes associated with 11 or more evolved eco-phenotypes.

To determine which biological processes were important for adaptation of the TOU-A population over eight generations, we tested whether SNPs in the 0.1% upper tail of the F_{ST} value distribution were over-represented in each of 736 Gene Ontology Biological Processes from the GOslim set[59]. 10,000 permutations were run to assess significance using the same methodology as described in Hancock *et al.* (2011)[32]. For each significantly enriched biological process, we retrieved the identity of all the genes containing SNPs in the 0.1% upper tail of the F_{ST} values distribution.

Data availability. The raw sequencing data used for this study will be available at the NCBI Sequence Read Archive (http://ncbi.nlm.nih.gov/sra) through the Study accession SRP077483. The phenotypic data that support the findings of this study are available from the authors on a reasonable request. The genomic SNP data files will be archived through the Dryad digital repository upon acceptance for publication.

Code availability. Custom scripts and phenotypic and genomic files used in this study have been archived in a local depository (https://lipm-browsers.toulouse.inra.fr/pub/Frachon2017-NEE/) that can be accessed by the reviewers with the login '**reviewersNEE'** and the password '**FaupKinmyad4**'. All the scripts and data sets will be made available available in the Dryad database upon acceptance of the manuscript. The code for performing genome-wide scan for selection based on temporal differentiation will be made available on the Zenodo database upon acceptance of the manuscript (Vitalis R, Gay L and Navascues M (2016) TempoDiff: a computer program to detect selection from temporal genetic differentiation. INRA. http://dx.doi.org/10.5281/zenodo.375600).

44. Hamann, A., Wang, T., Spittlehouse, D.L. & Murdock, T.Q. A comprehensive, high-resolution database of historical and projected climate surfaces for western North America. *B. Am. Meteorol. Soc.* **94**, 1307 (2013).

45. Hendry, A.P. & Kinnison, M.T. The pace of modern life: measuring rates of contemporary microevolution. *Evolution* **53**, 1637-1653 (1999).

46. Gingerich, P.D. Rates of evolution on the time scale of the evolutionary process. *Genetica* **112-113**, 127-144 (2001).

47. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

48. Koboldt, D.C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568-576 (2012).

49. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).

50. Horton, M.W. *et al.* Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions for the RegMap panel. *Nat. Genet.* **44**, 212-216 (2012).

51. Zheng, X. *et al.* A High-performance computing toolset for relatedness and Principal Component Analysis of SNP data. *Bioinformatics* **28,** 326-3328 (2012).

52. Kang, H.M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348-354 (2010).

53. Atwell, S. *et al.* Genome-wide association study of 107 phenotypes in a common set of *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627-631 (2010).

54. Hermisson J. & McGregor A.P. Pleiotropic scaling and QTL data. *Nature* **456**, E3-E4 (2008).

55. Goldringer, I. & Bataillon, T. On the distribution of temporal variations in allele frequency consequences for the estimation of effective population size and the detection of loci undergoing selection. *Genetics* **168**, 563-568 (2004).

56. Waples, R.S. A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**, 379-391 (1989).

57. Weir, B.S. & Cockerham, C.C. Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370 (1984).

58. Skoglund, P., Sjödin, P., Skoglund, T., Lascoux, M. & Jakobsson, M. Investigating population history using temporal genetic differentiation. *Mol. Biol. Evol.* **31**, 2516-2527 (2014).

59. The Gene Ontology Consortium. The Gene Ontology project in 2008. *Nucleic Acids Res.* **36**, D440-D444 (2008).

Supplementary information

Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time

Supplementary Information

Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time

Léa Frachon, Cyril Libourel, RomainVilloutreix, Sébastien Carrère, Cédric Glorieux, Carine Huard-Chauveau, Miguel Navascués, Laurène Gay, Renaud Vitalis, Etienne Baron, Laurent Amsellem, Olivier Bouchez, Marie Vidal, Valérie Le Corre, Dominique Roby, Joy Bergelson, Fabrice Roux*

*To whom correspondence should be addressed. E-mail: fabrice.roux@inra.fr

This file includes:

Supplementary Information: text

Supplementary Figures 1-15

Supplementary Tables 1-6

SUPPLEMENTARY INFORMATION

Plant material

In this study, we focused on the population TOU-A located under a 350m electric fence separating two permanent meadows experiencing cycles of periodic grazing by cattle (**Supplementary Fig. 1A**) in the village of Toulon-sur-Arroux (Burgundy, East of France, N 46°38'57.302'', E 4°7'16.892''). Seeds from individual plants were collected in 2002 (TOU-A1), 2007 (TOU-A5) and 2010 (TOU-A6) according to a sampling scheme allowing us to take into account the density of *A. thaliana* plants along the transect: (1) from the starting point of the transect (**Supplementary Fig. 1A**), walk along the transect until a plant is found and collect seeds from this plant, (2) if this plant is at the beginning of a patch, then collect seeds from plants located every 50 cm along this patch, (3) else, walk along the transect until a new plant is found and collect seeds from this plant. According to this sampling scheme, seeds of 80, 115 and 115 individual plants were collected in 2002, 2007 and 2010, respectively (**Supplementary Fig. 1**). Seeds collected from those 310 individual plants constitute seed families, hereafter named accessions. Given the outcrossing rate of ~6% observed in the TOU-A population[1], the 310 accessions were considered as relatively homozygous across the genome.

Seeds from the 80 accessions collected in 2002 were grown individually in a controlled greenhouse at The University of Chicago (USA) and seeds for each TOU-A1 accession collected. The analysis of these 80 accessions genotyped at 149 SNPs gave an estimate of selfing rate of ~94%[1].

Differences in the maternal effects between the 310 accessions were reduced by growing one plant of each family for one generation under controlled greenhouse conditions (16-h photoperiod, 20°C) in early 2011 at the University of Lille 1. For this purpose, we planted seeds produced at The University of Chicago for accessions from the TOU-A1 population, and seeds collected in the field for accessions from the TOU-A5 and TOU-A6 populations. For the purpose of this study, we only used seeds from the 80 accessions collected in 2002 and from the 115 accessions collected in 2010.

Ecological characterization

Climate characterization

Data for the mean annual temperature, the mean warmest month temperature, the mean coldest month temperature, the sum of degree-days above 5°C, the sum of degree-days below 0°C and the mean annual precipitation were retrieved from 1970 to 2013. Climate data was generated with the ClimateEU v4.63 softwarepackage, available at http://tinyurl.com/ClimateEU, based on methodology described in Hamann *et al.* (2013)[2].

Soil characterization

A sample of the 5-cm upper soil layer was collected at 83 positions scattered along the transect in 2010 (**Supplementary Fig. 1**). These samples were air-dried in the greenhouse (20-22°C), and then stored at room temperature. As described in Brachi *et al.*(2013)[3], each

soil sample was characterized for 14 edaphic factors: pH, maximal water holding capacity (WHC), total nitrogen content (N), organic carbon content (C), C/N ratio, soil organic matter content (SOM), concentrations of P_2O_5, K, Ca, Mg, Mn, Al, Na and Fe. Iron concentration (Fe) was excluded from further analyses due to a lack of variation among the 83 samples. In order to reduce multicollinearity, the set of remaining 13 edaphic variables was pruned based on the pairwise Spearman correlations of the variables, so that no two variables had a Spearman *rho* greater than 0.8. In cases where variables were strongly inter-correlated, we selected the one with the most obvious link to the ecology of *A. thaliana*. The final set of 10 edaphic variables considered in this study was N, C/N ratio, pH, WHC, P_2O_5, K, Mg, Mn, Na and Al.

To visualize the edaphic space of the TOU-A population, we conducted a principal component analysis (PCA) based on the 83 values of the 10 edaphic traits (*R* package ade4)[4].

Phenotypic characterization

Experimental design

An experiment of 5850 plants was set up at the local site of the TOU-A population. The experimental design and the experimental conditions are illustrated on **Supplementary Fig. 1**. Based on the edaphic space (**Supplementary Fig. 3**), we defined three contrasting edaphic areas under the electric fence, hereafter named soil types A, B and C. In late August 2012, a 12.3-m² (4.4m * 2.8) plot was delimited by an electric fence for protection against cattle in each soil type. In each plot, one subplot of 2.88-m² (4.8m * 0.6m, experimental condition without the presence of *P. annua*, see below) and one subplot of 3.36-m² (4.8m * 0.7m, experimental condition with the presence of *P. annua*, see below) were arranged at 80-cm spacing. In late August 2012, each subplot was manually weeded and tilled for the 10-cm upper soil layer. The 24th of September 2012, subplots were surrounded by green plastic covers for weed control. To mimic the main natural germination cohort observed in the TOU-A population in late September 2012 (**Supplementary Fig. 1**), seeds were sown on the 24th of September 2012 for the experimental conditions 'soil A without *P. annua*', ''soil A with *P. annua*' and 'soil B without *P. annua*', and on the 25th of September 2012 for the experimental conditions 'soil B with *P. annua*', ''soil C without *P. annua*' and 'soil C with *P. annua*'. Each of the six *in situ* experimental conditions was organized in five blocks, each one being represented by 3 arrays of 66 individual wells (Ø4 cm, vol. ~38 cm^3) (TEKU, JP 3050/66). Across the five blocks, the 15 arrays were stuck some on the others and organized according to a grid of 15 columns and one line. To buffer against possible border effects in the experimental conditions with *P. annua*, the 15 arrays were surrounded by one row of wells sown with both *P. annua* and *A. thaliana* (accession TOU-A6-69 collected in 2010). All the wells were first filled with 3 cm of the respective native soil, then with an additional 1cm of the respective native soil that was oven dried for two days at 65°C. The oven dried native soil prevented germination from the seed bank, whereas the 3-cm native soil allowed the colonization of the oven dried native soil by native microbiota.

In each of the six *in situ* experimental conditions, each of the five blocks corresponded to an independent randomization of 195 plants with one replicate per accession collected in 2002 and 2010. In each block, the remaining three wells were left empty. Five seeds of *A. thaliana* were sown in each well. For the three *in situ* experimental conditions with *P. annua*, a mean number of five seeds of *P. annua* were additionally sown in each well. Seeds for *P. annua* were ordered to the company Herbiseeds (http://www.herbiseed.com/home.aspx). After sowing, arrays were directly transported *in situ* and slightly buried in their dedicated soil types. Arrays were covered for 10 days with an agricultural fleece that allowed the seeds to be exposed to rain and sunlight while preventing them from disturbance by rain drops.

Germination date was monitored daily for 10 days (see below). Seeds germinated in more than 97.74 % of the wells. Wells were thinned to one seedling of *A. thaliana* and/or one seedling of *P. annua* between 18 and 22 days after sowing. During the course of the experiment (late September 2012 – late June 2013), plants were protected from herbivory by slugs as described in Brachi *et al.* (2010)[5].

Measured traits

Each plant was scored for a total of 29 phenotypic traits related to phenology (n = 4), resource acquisition (n = 1), architecture and seed dispersal (n = 9), fecundity (n = 14) and survival (n = 1). These traits were chosen to characterize the life history of *A. thaliana* including the timing of offspring production or seed dispersal[3,6-8], or because they are involved in the response to competition[9,10], and/or are good estimators of life-time fitness and reproductive strategies[7,11-14]. Most of these traits have been fully described in Roux *et al.* (2016)[14]:

- *Phenology*: Germination time (GERM) was measured as the number of days between sowing and the emergence of the first seedling (opening of both cotyledons). Bolting time (BT), flowering interval (INT) and the reproductive period (RP) were scored as the interval between germination date and bolting date (inflorescence distinguishable from the leaves at a size < 5 mm), between bolting date and flowering date (appearance of the first open flower) and between flowering date and date of maturation of the last fruit, respectively.
- *Resource acquisition*: At the start of flowering, the maximum diameter of the rosette measured to the nearest millimeter was used as a proxy for plant size (DIAM).
- *Architecture and seed dispersal*: After maturation of the last fruit, the above-ground portion was harvested and stored at room temperature. Plants were later phenotyped for the following architectural and seed dispersal related traits: height from soil to the first fruit on the main stem (H1F, in mm), height of the main stem (HSTEM, in mm), maximum height (HMAX, in mm), number of primary branches on the main stem with fruits (RAMPB_WF) or without fruits (RAMPB_WOF), total number of primary branches (TOTPB), total number of basal branches (RAMBB) and total number of branches (TOTB = TOTPB + RAMBB). We also evaluated a response strategy to competition (ratio HD = H1F / DIAM)[9].

- *Fecundity*: Because the number of seeds in a fruit is highly correlated with fruit length[11], total seed production was approximated by total fruit length (FITTOT, in mm). Seed production is a good proxy for fecundity in a highly selfing annual species like *A. thaliana*. FITTOT was obtained by adding the fruit length produced on the main stem (FITSTEM, in mm), the primary branches on the main stem (FITPB, mm) and the basal branches (FITBB, in mm). These estimates of fruit length were obtained by counting the number of fertilized fruits produced on each type of branches (FRUITSTEM, FRUITPB and FRUITBB) and multiplying these counts by an estimate of their corresponding fruit (or silique) length (SILSTEM, SILPB and SILBB, in mm), estimated as the average of three haphazardly selected representative fruits. We also calculated three ratios corresponding to the percentage of seeds produced by one branch type as a function of the total amount of seed produced: RSTEM = FITSTEM / FITTOT, RPB = FITPB / FITTOT and RBB = FITBB / FITTOT. Finally, we estimated the average length between two fruits on the main stem (INTERNOD = (HSTEM – H1F) / (FRUITSTEM - 1); in mm).
- *Survival*: All plants that germinated but did not survive were counted as dead (SURVIVAL = 0). Harvested plants were counted as alive (SURVIVAL = 1).

Genomic characterization

DNA extraction, libraries preparation and genome sequencing

Genomic DNA for the 195 accessions collected in 2002 and 2010 was extracted as described in Brachi *et al.* (2013)[3]. DNAseq was performed at the GeT-PlaGe core facility (INRA Toulouse). DNA-seq libraries were prepared according to Illumina's protocol using the Illumina TruSeq Nano LT Kit. Briefly, DNA was fragmented by sonication on a covaris M220, size selection was performed using CLEANNA CleanPCR beads and adaptators were ligated for sequencing. Library quality was assessed using an Advanced Analytical Fragment Analyser and libraries were quantified by QPCR using the Kapa Library Quantification Kit. DNA-seq experiments were performed on an Illumina HiSeq2500 using a paired-end read length of 2x100 pb with the Illumina TruSeq SBS v3 Reagent Kits. Each PCR product with tag-sequence was first quantified using PicoGreen® dsDNA Quantitation Reagent. Then a mix was made depending on these quantities in order to obtain an equimolar pool.

Mapping and SNP calling

Raw reads of each of the 195 accessions were mapped onto the *A. thaliana* reference genome Col-0 (genome size: 119Mb, TAIR10, https://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation/annotation_data.jsp) using glint software (1.0.rc8; Faraut & Courcelle, unpublished software) with the following parameters: a maximum of 5 mismatches on at least 80 nucleotides and keep alignments with the best score (*glint mappe --no-lc-filtering --best-score --mmis 5 --lmin 80 --step 2*). The mapped reads were filtered for proper pairs with SAMtools (v0.01.19)[15] (*samtools view -f 0x02*). The mean and the median coverage to a unique position in the reference genome was ~25.5x and ~24.5x, respectively.

A stringent SNPCalling across the genome was then performed for each accession with SAMtools mpileup (v0.01019)[15] and VarScan (v2.3)[16] with the parameters corresponding to a theoretical sequencing coverage of 30X and the search for homozygous sites (*samtools mpileup -B ; VarScan mpileup2snp --min-coverage 5 --min-reads2 4 --min-avg-qual 30 --min-var-freq 0.97 --p-value 0.01*). Due to the relatively high selfing rate observed in *A. thaliana* and the generation(s) of selfing performed in greenhouse conditions (see the subsection 'Plant material'), the frequency of heterozygous sites should be low; those sites were not considered in this study in order to avoid paralogs. All polymorphic sites were then identified among the 195 accessions. Finally, a SNP calling based on all accessions was performed on all polymorphic sites to differentiate null values from the reference value. Sites with more than 50% missing values were discarded from the set of polymorphic sites.

Testing whether the mean Linkage Disequilibrium extent in the TOU-A population is short enough for fine mapping of genomic regions associated with natural phenotypic variation

The presence of significant associations at loci known to be involved in well described phenotypes provides a proof-of-concept for the power of conducting GWAS in a given mapping population. To estimate the power of fine mapping in the TOU-A population, we focused (i) on the *R* genes *RPM1* and *RPS2* responsible for the hypersensitive cell death response (HR) against the engineered bacterial strain of *Pseudomonas syringae* DC3000 expressing either *AvrRpm1* (DC3000::AvrRpm1) or *AvrRpt2* (DC3000::AvrRpt2), respectively, and (ii) on the atypical kinase *RKS1* conferring quantitative broad-spectrum resistance against the vascular bacterial pathogen *Xanthomonas campestris* pv. *campestris* (reviewed in Roux & Bergelson (2016)[17]). The 195 accessions collected in 2002 and 2010 were grown, inoculated and phenotyped for (i) qualitative resistance against DC3000::AvrRpm1 (leaf collapse scored at 6hpi) and DC3000::AvrRpt2 (leaf collapse scored at 1dpi) as described in Vailleau *et al.* (2002)[18], and (ii) quantitative resistance against the strain *Xcc568* (disease index scored using a scale from 0 to 4 at 10dpi) as described in Huard-Chauveau *et al.* (2013)[19]. Given the broad-sense heritability values close to one observed for qualitative resistance[20], four leaves of a single plant were inoculated for each accession. For quantitative resistance against *Xcc568*, a randomized complete block design was set up with two blocks, each being an independent randomization of one replicate per accession. In the latter case, the following general linear model was used to analyze disease index (GLM procedure in SAS9.1, SAS Institute Inc., Cary, North Carolina, USA):

$$\text{disease index}_{ij} = \mu + \text{block}_i + \text{accession}_j + \varepsilon_{ij}$$

where 'μ' is the overall mean; 'block' accounts for differences among the two experimental blocks; 'accession' corresponds to the 195 natural accessions; and 'ε' is the residual term. Normality of the residuals was not improved by transformation of the data. Least-square mean (LSmean) was obtained for each natural accession

GWA mapping was run using a mixed-model approach implemented in the software EMMAX (Efficient Mixed-Model Association eXpedited)[21]. This model includes a genetic kinship matrix as a covariate to control for population structure.GWA mapping was based on (i) raw means for qualitative resistance against DC3000::AvrRpm1 and DC3000::AvrRpt2, and (ii) LSmeans for quantitative resistance against *Xcc568*.

References to Supplementary Information

1. Platt, A. *et al.* The scale of population structure in *Arabidopsis thaliana*. *PLoS Genet.* **6**, e1000843 (2010).
2. Hamann, A., Wang**,** T., Spittlehouse, D.L. & Murdock, T.Q. A comprehensive, high-resolution database of historical and projected climate surfaces for western North America. B. Am. Meteorol. Soc. **94**, 1307 (2013).
3. Brachi, B. *et al.* Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in *Arabidopsis thaliana*. *Mol. Ecol.* **22**, 4222-4240 (2013).
4. Chessel, C., Dufour, A.B. & Thioulouse, J. The ade4 package – I – One-table methods. *R News* **4**, 5 (2004).
5. Brachi, B. *et al.* Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet.* 6:e1000940 (2010).
6. Weinig, C. *et al.* Novel loci control variation in reproductive timing in *Arabidopsis thaliana* in natural environments. *Genetics* **162**, 1875-1884 (2002).
7. Reboud, C. *et al.* Natural variation among accessions of *Arabidopsis thaliana*: beyond the flowering date, what morphological traits are relevant to study adaptation? In Plant adaptation: molecular biology and ecology. Edited by Q. C. Cronk, J. Whitton and I. Taylor. NRC Research Press, Ottawa, Canada. pp 135-142 (2004).
8. Wender, N.J., Polisetty, C.R. & Donohue, K. Density-dependent processes influencing the evolutionary dynamics of dispersal: a functional analysis of seed dispersal in *Arabidopsis thaliana* (Brassicaceae). *Am. J. Bot.* **92**, 960-971 (2005).
9. Baron, E., Richirt, J., Villoutreix, R., Amsellem, L. & Roux, F. The genetics of intra- and interspecific competitive response and effect in a local population of an annual plant species. *Funct. Ecol.* **29**, 1361-1370 (2015).
10. Brachi, B., Aimé, C., Glorieux, C., Cuguen, J. & Roux, F. Adaptive value of phenological traits in stressful environments: predictions based on seed production and Laboratory Natural Selection. *PLoS One* **7**, e32069 (2012).
11. Roux, F., Gasquez, J. & Reboud, X. The dominance of the herbicide resistant cost in several *Arabidopsis thaliana* mutant lines. *Genetics* **166**, 449-460 (2004).
12. Roux, F., Giancola, S., Durand, S. & Reboud, X. Building of an experimental cline with *Arabidopsis thaliana* to estimate herbicide fitness cost. *Genetics* **173**,1023-1031 (2006).
13. Bac-Molenaar, J.A. *et al.* Genome-wide association mapping of fertility reduction upon heat stress reveals developmental stage-specific QTLs in *Arabidopsis thaliana*. *The Plant Cell* **27**, 1857-1874 (2015).
14. Roux, F. *et al.* Cytonuclear interactions affect adaptive phenotypic traits of the annual plant *Arabidopsis thaliana* in ecologically realistic conditions. *Proc. Natl. Acad. Sci. U.S.A.* **113**: 3687-3692 (2016).
15. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
16. Koboldt, D.C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568-576 (2012).

17. Roux, F. & Bergelson, J. 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 (2016).
18. Vailleau, F. *et al.* A R2R3-MYB gene, *AtMYB30*, acts as a positive regulator of the hypersensitive cell death program in plants in response to pathogen attack. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10179-10184 (2002).
19. Huard-Chauveau, C. *et al.* An atypical kinase under balancing selection confers broad-spectrum disease resistance in Arabidopsis. *PLoS Genet.* **9**, e1003766 (2013).
20. Atwell, S. *et al.* Genome-wide association study of 107 phenotypes in a common set of *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627-631 (2010).
21. Kang, H.M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348-354 (2010).

Figure S1 | General picture of the TOU-A population. (**a**) Photograph showing the habitat type. The population is located under a 350m electric fence separating two permanent meadows. (**b**) Position of plants for which seeds have been collected in 2002, 2007 and 2010. (**c**) Position of soil samples collected in 2010. The letters A, B and C indicate the three edaphic areas (i.e. soil types) in which the *in situ* experiment has been performed (see **Supplementary Fig. 3**). (**d**) Tillage of the 10-cm upper soil layer in late August 2012 and protection from cattle by electric fences. (**e**) Soil cover with green plastic for weed control in late September 2012. (**f**) Observed natural germination flushes in late September 2012.

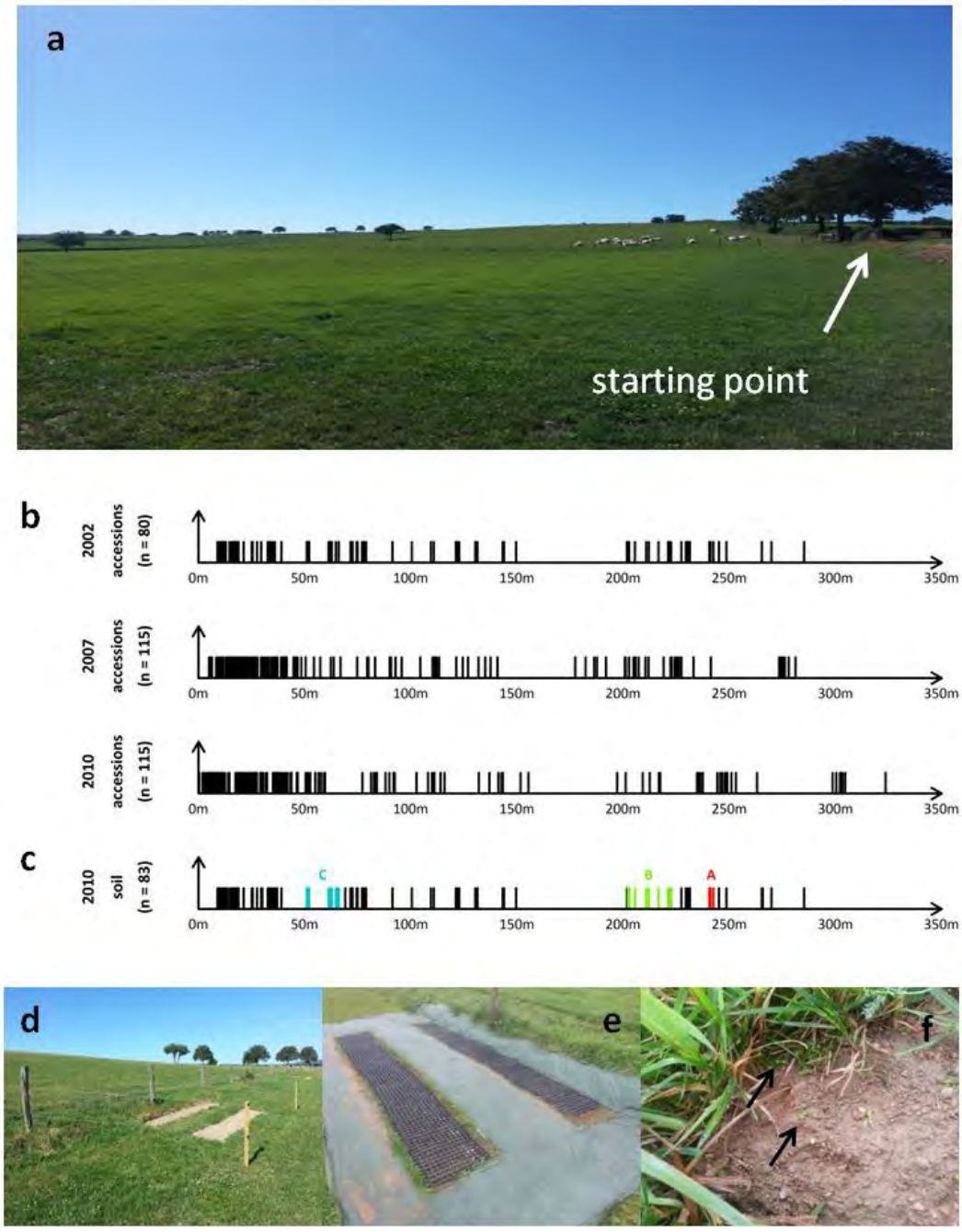


Figure S2 | Climate change since 1970 in the locality of the TOU-A population. Blue dots indicate the three sampling years (2002, 2007 and 2010). The green dot indicates the year of the *in situ* experiment. Red lines correspond to the mean of the last five consecutive years. A significant change over time was detected for the mean annual temperature (Spearman's *rho* = 0.63, P = 5.5 x 10^{-6}), the mean warmest month temperature (Spearman's *rho* = 0.35, P = 0.019) and the sum of degree-days above 5°C (Spearman's *rho* = 0.69, P = 7.1 x 10^{-7}), but not for the mean coldest month temperature (Spearman's *rho* = 0.-0.026, P = 0.865), the mean annual precipitation (Spearman's *rho* = 0.025, P = 0.869) and the sum of degree-days below 0°C (Spearman's *rho* = -0.090, P = 0.560).

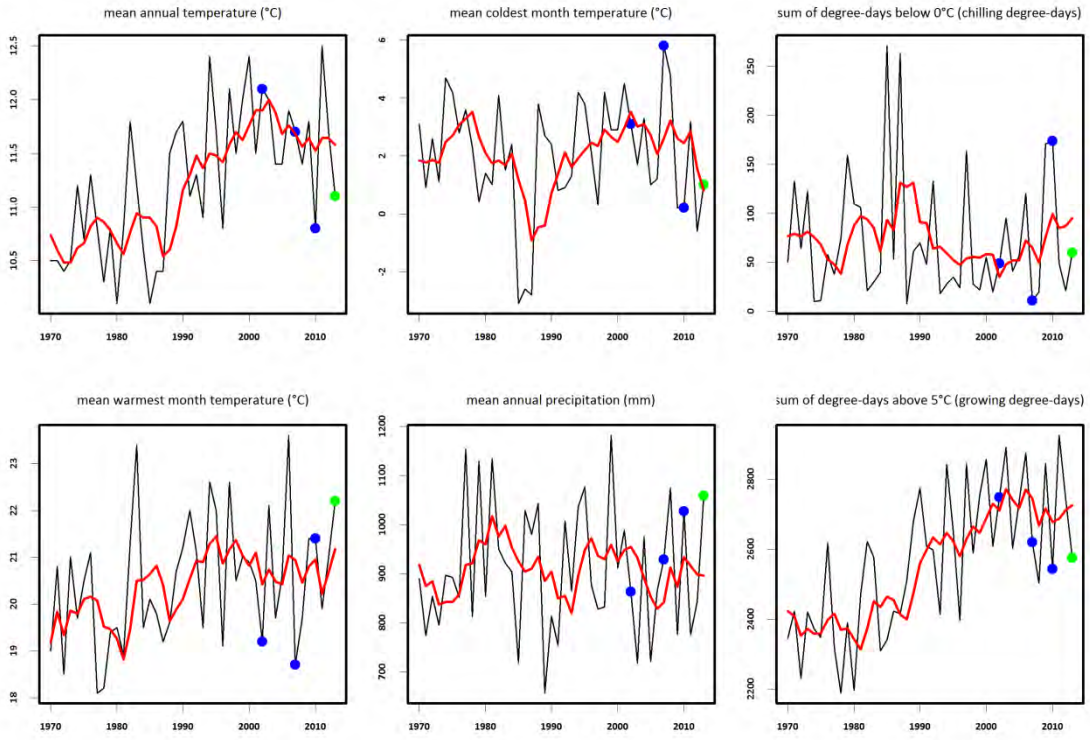


Figure S3 | Edaphic variation in the TOU-A population. (**a**) Factor loading plot resulting from principal components analysis. Factor 1 and factor 2 explained 31.54% and 24.22% of total soil variance. Maximum water holding capacity (WHC), content of total nitrogen (N), organic carbon / total nitrogen ratio (C.N), concentrations of P_2O_5, K, Mg, Mn, Al and Na. (**b**) Distribution of eigenvalues against the ranked component number. (**c**) Position of the 83 soil samples in the 'Factor1 – Factor 2' edaphic space. Red, green and blue dots correspond to the soil samples located in three soil areas 'soil A', 'soil B' and 'soil C', respectively.

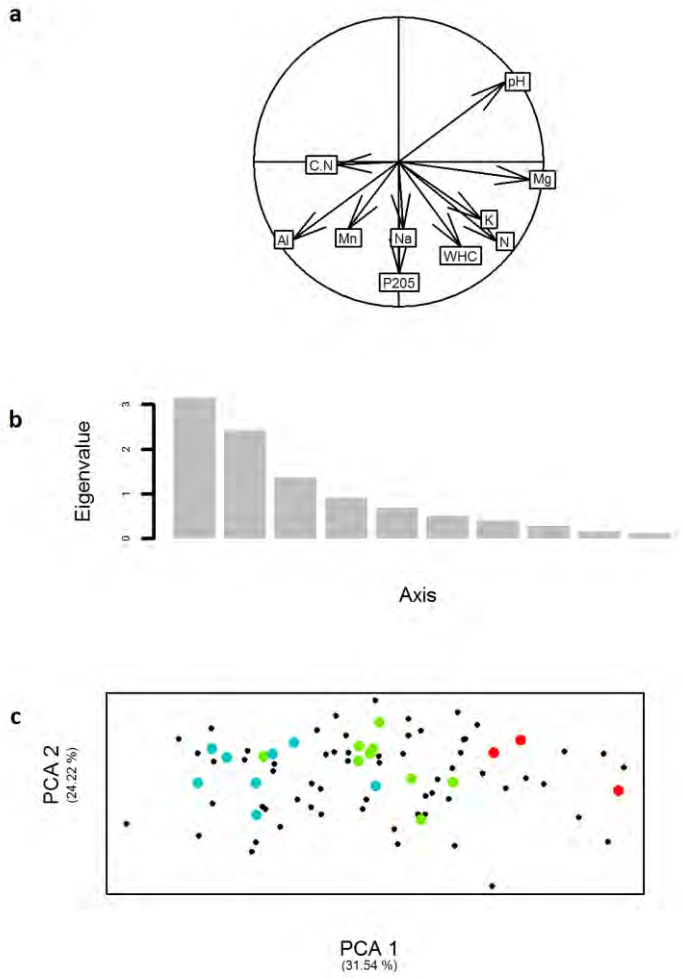


Figure S4 | Illustration of the genotype-by-environment interactions across the six *in situ* 'soil x competition' micro-habitats. (**a**) Genetic variation for reaction norms of bolting time. (**b**) Genetic variation for reaction norms of seed production on the main stem. Solid red lines: reaction norms of the 80 accessions collected in 2002, dashed blue lines: reaction norms of the 115 accessions collected in 2010.

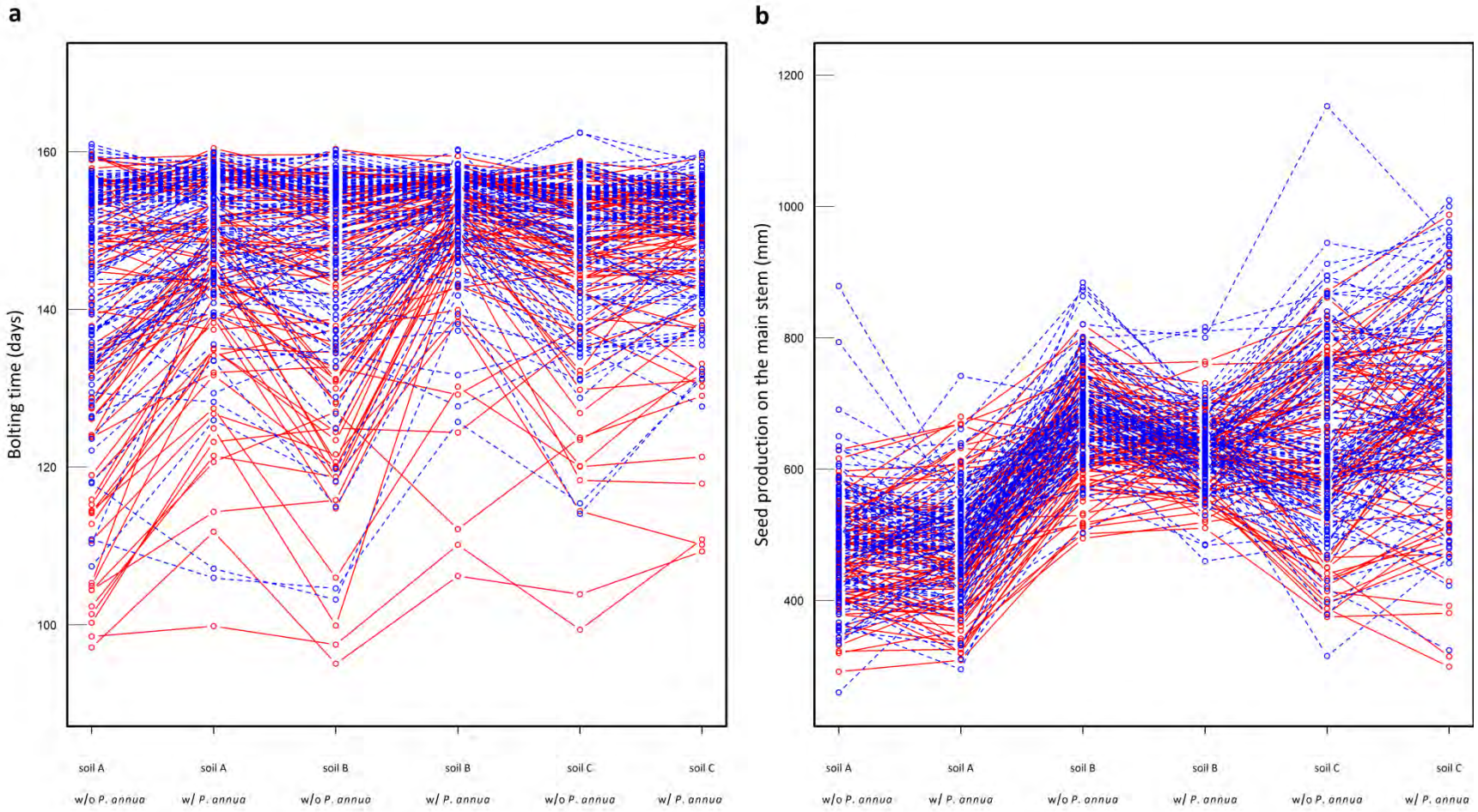


Figure S5 | Distribution of the size of LD blocks in the TOU-A population.

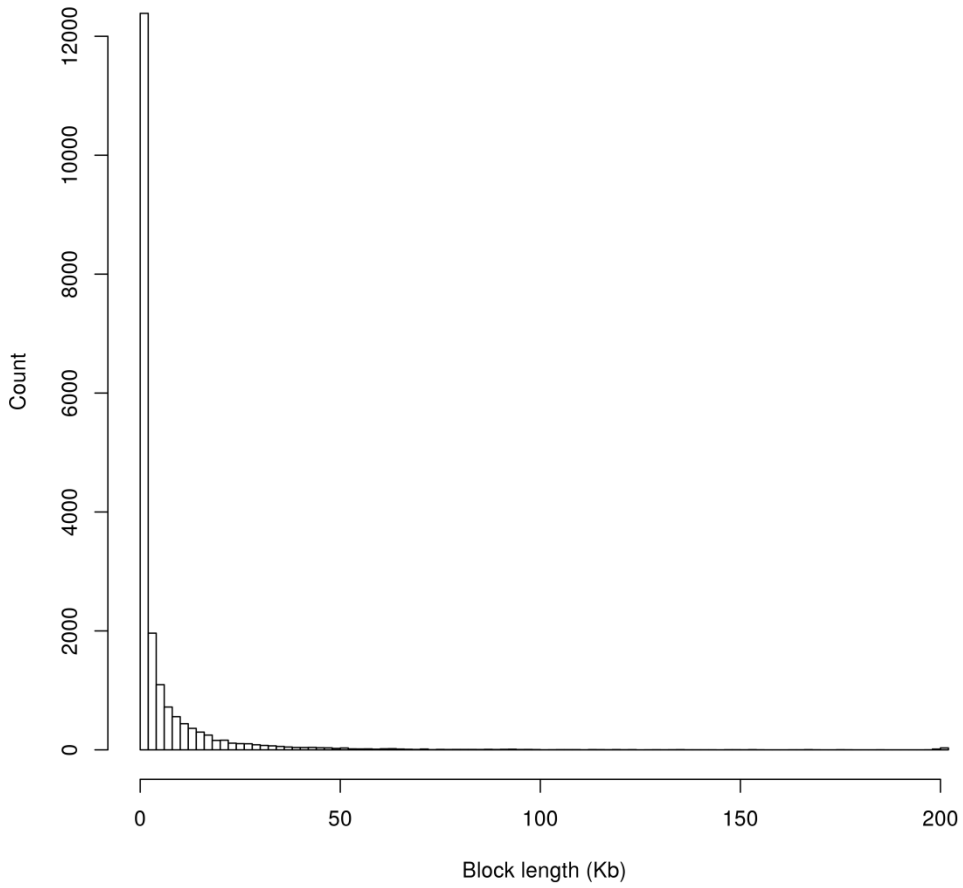


Figure S6 | GWA analysis of hypersensitive response to the bacterial elicitors *AvrRpm1* (a) and *AvrRpt2* (b) and quantitative resistance to *Xanthomonas campestris* pv. *campestris* strain *Xcc568* (c). The top SNPs are located 15bp from *RESISTANCE TO PSEUSOMONAS SYRINGAE PV MACULICOLA* (*RPM1*), within *RESISTANT TO PSEUDOMONAS SYRINGAE 2* (*RPS2*) and within *RESISTANCE RELATED KINASE 1* (*RKS1*). The *x*-axis indicates the physical position along the chromosome. The *y*-axis indicates the $-\log^{10}$ *p*-values of phenotype-SNP associations using the EMMAX method. MARF > 7%.

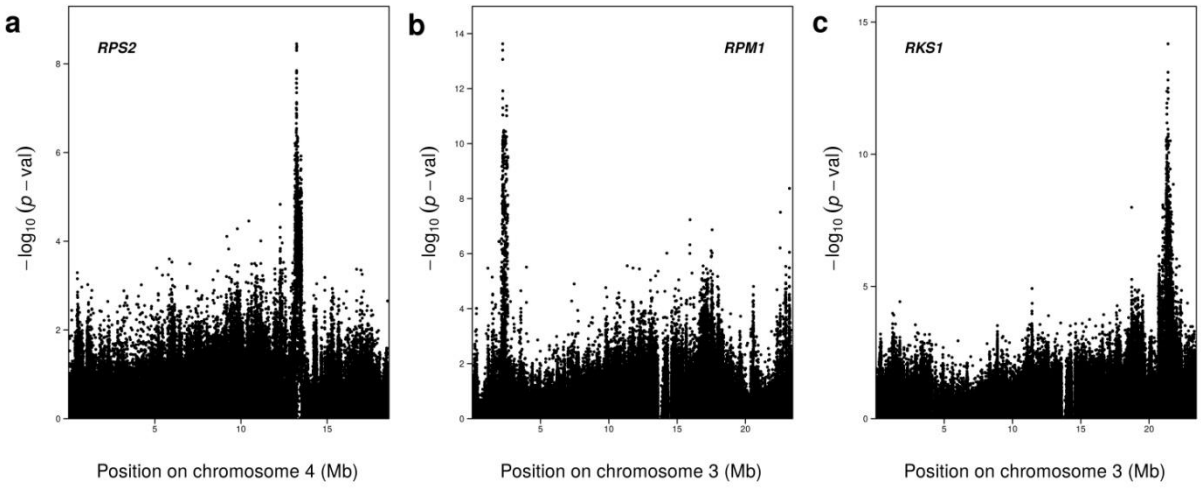


Figure S7 | Enrichment ratios in flowering time candidate genes for the six *in situ* 'soil x competition' micro-habitats (i.e. three soils A, B and C x absence or presence of *P. annua*), as a function of the number of top SNPs chosen in the GWA mapping results for bolting time using the EMMAX method. The corresponding 95% confidence intervals from the null distributions are represented by the green area.

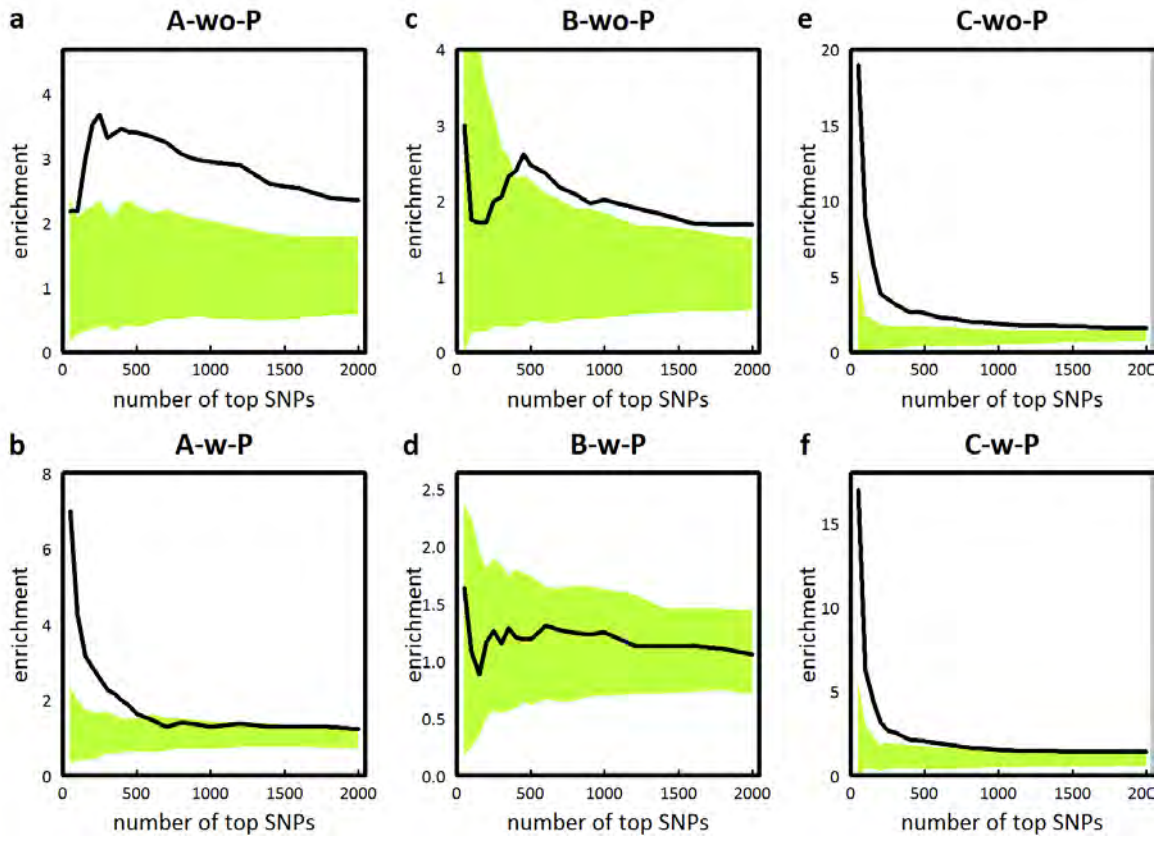


Figure S8 | Identification of genomic regions associated with the 144 heritable eco-phenotypes in the TOU-A population. The *x*-axis indicates the physical position along the chromosome. The *y*-axis indicates the -log^{10} *p*-values using the EMMAX method. MARF > 7%. On each Manhattan plot, the 100 and 200 top SNPs are highlighted in blue and red, respectively.

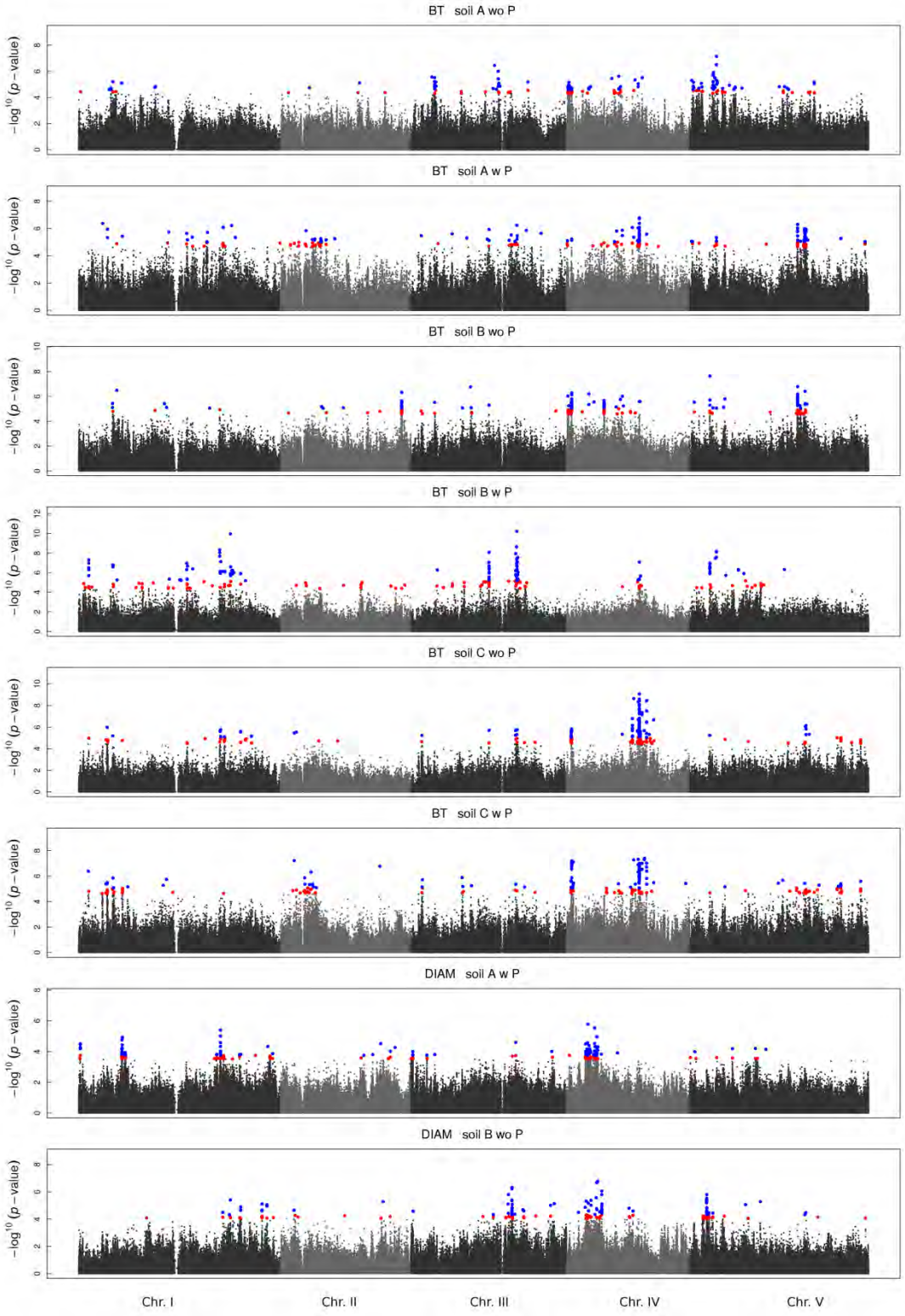


Figure S8 (continued)

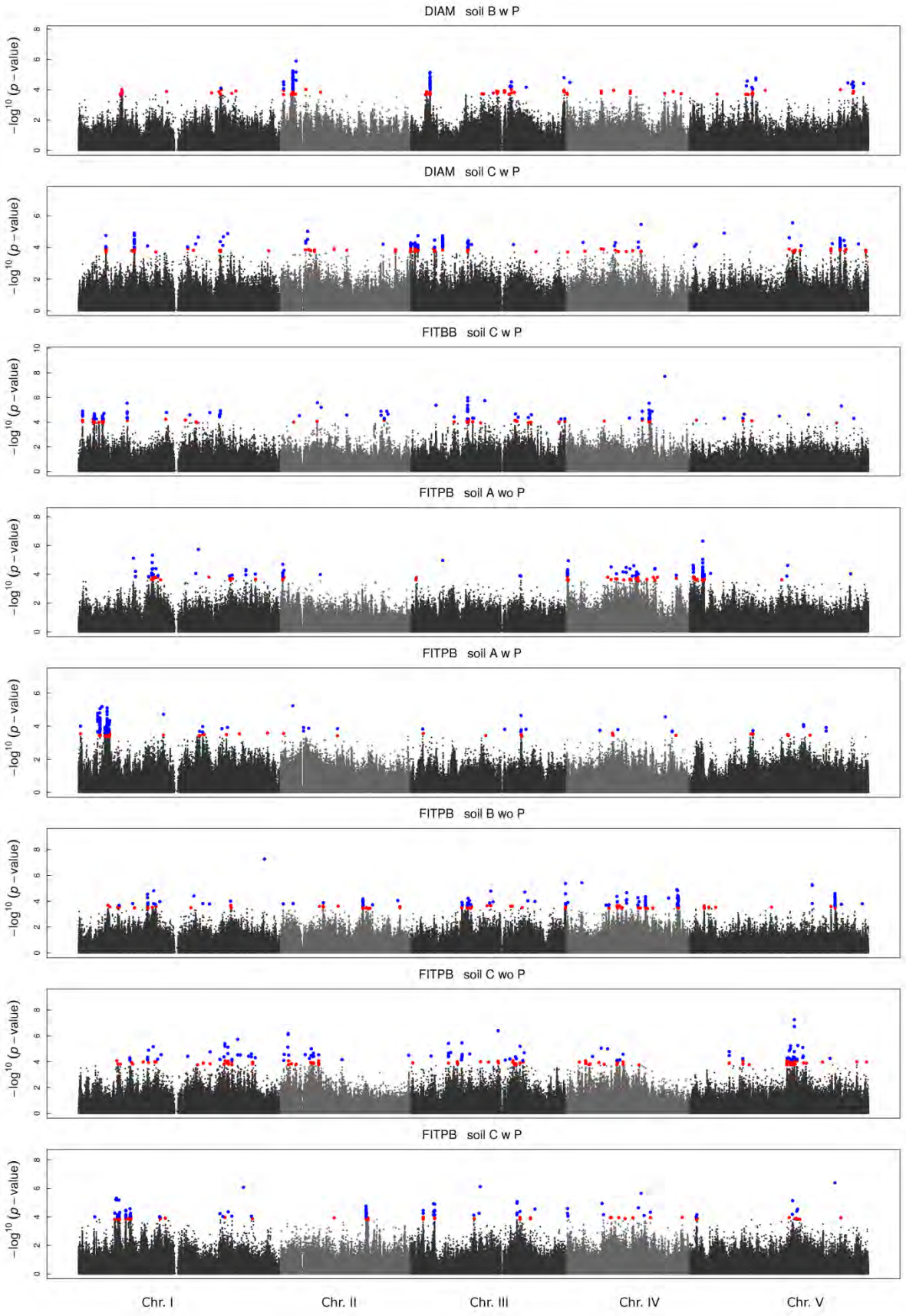


Figure S8 (continued)

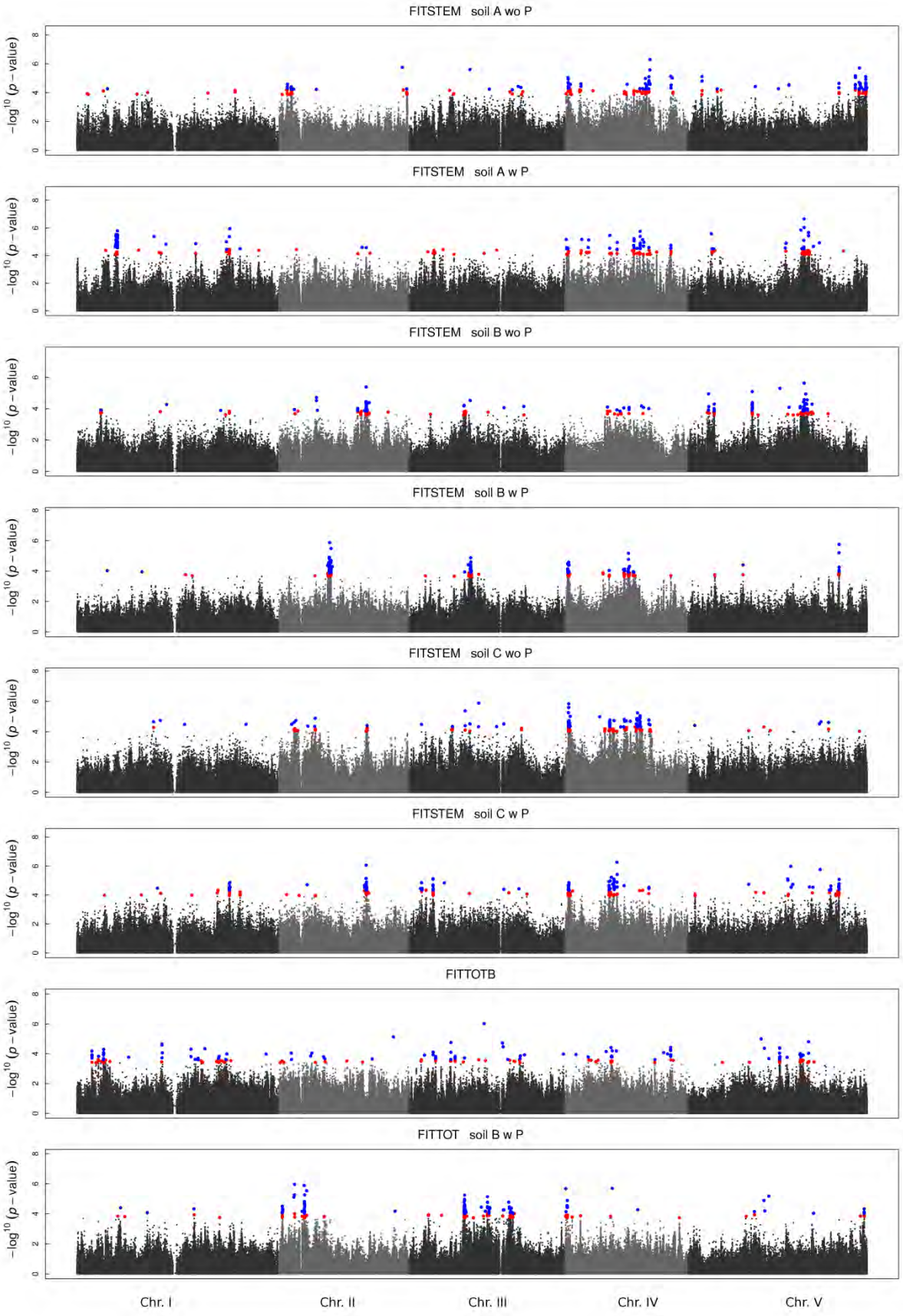


Figure S8 (continued)

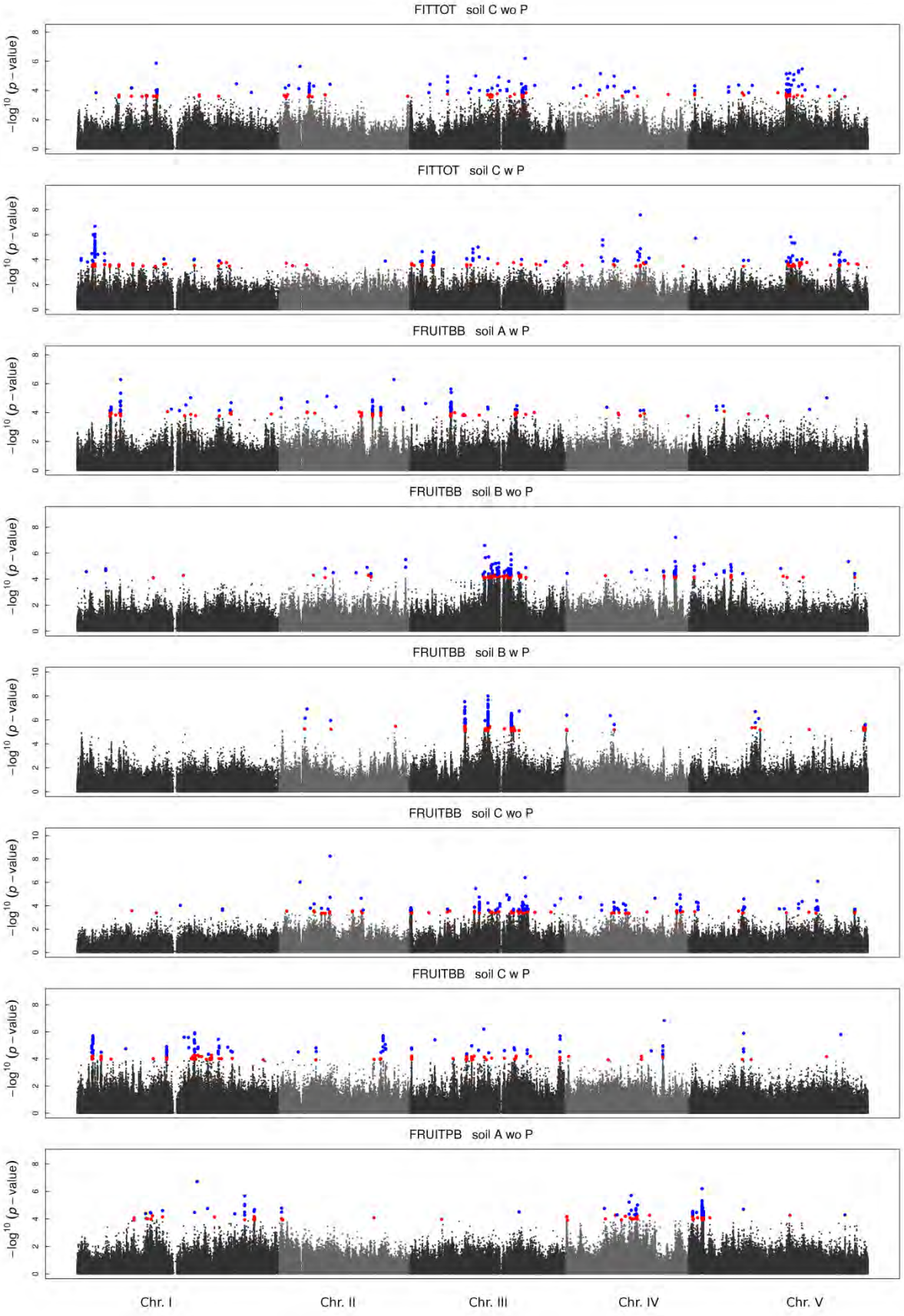


Figure S8 (continued)

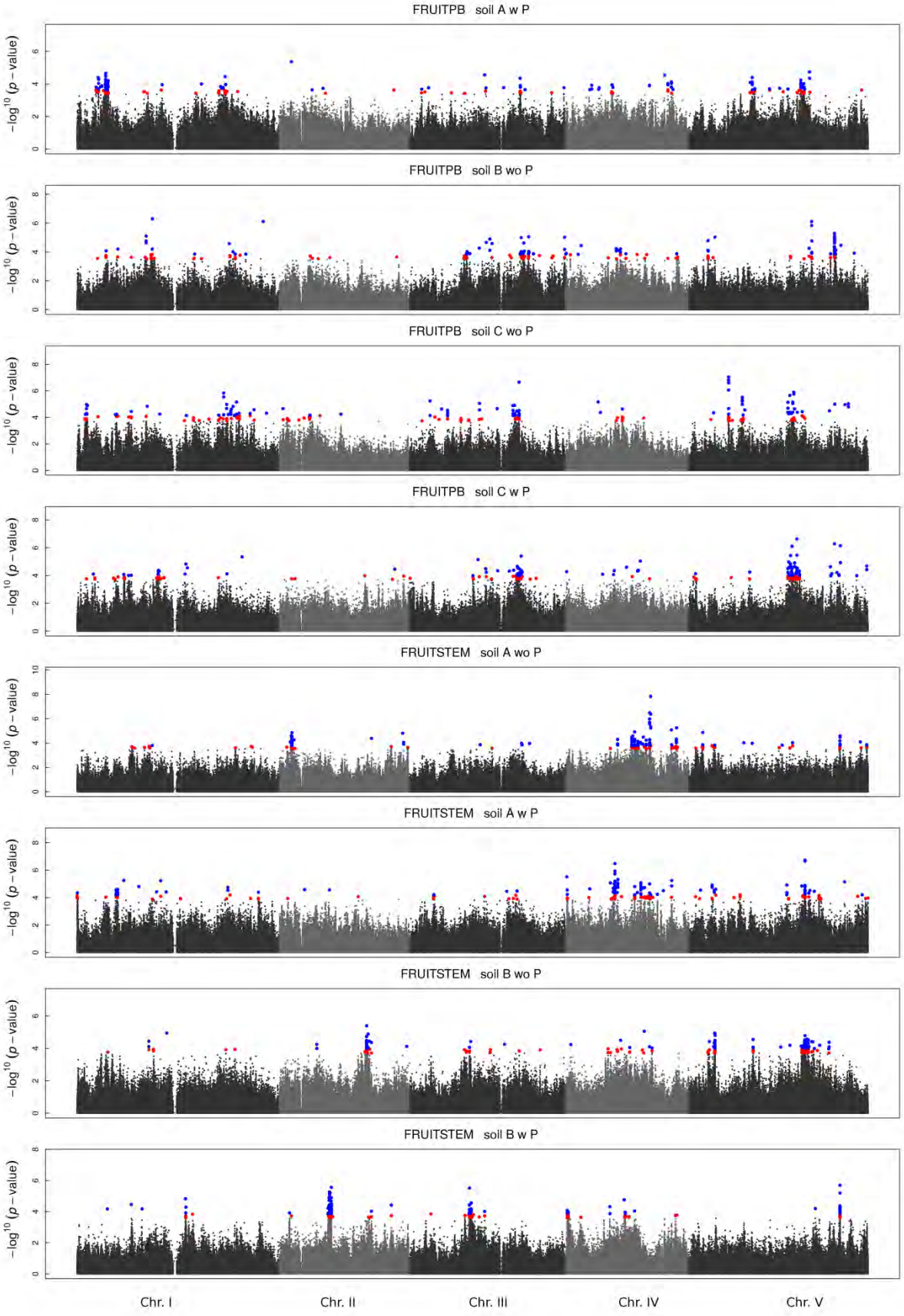


Figure S8 (continued)

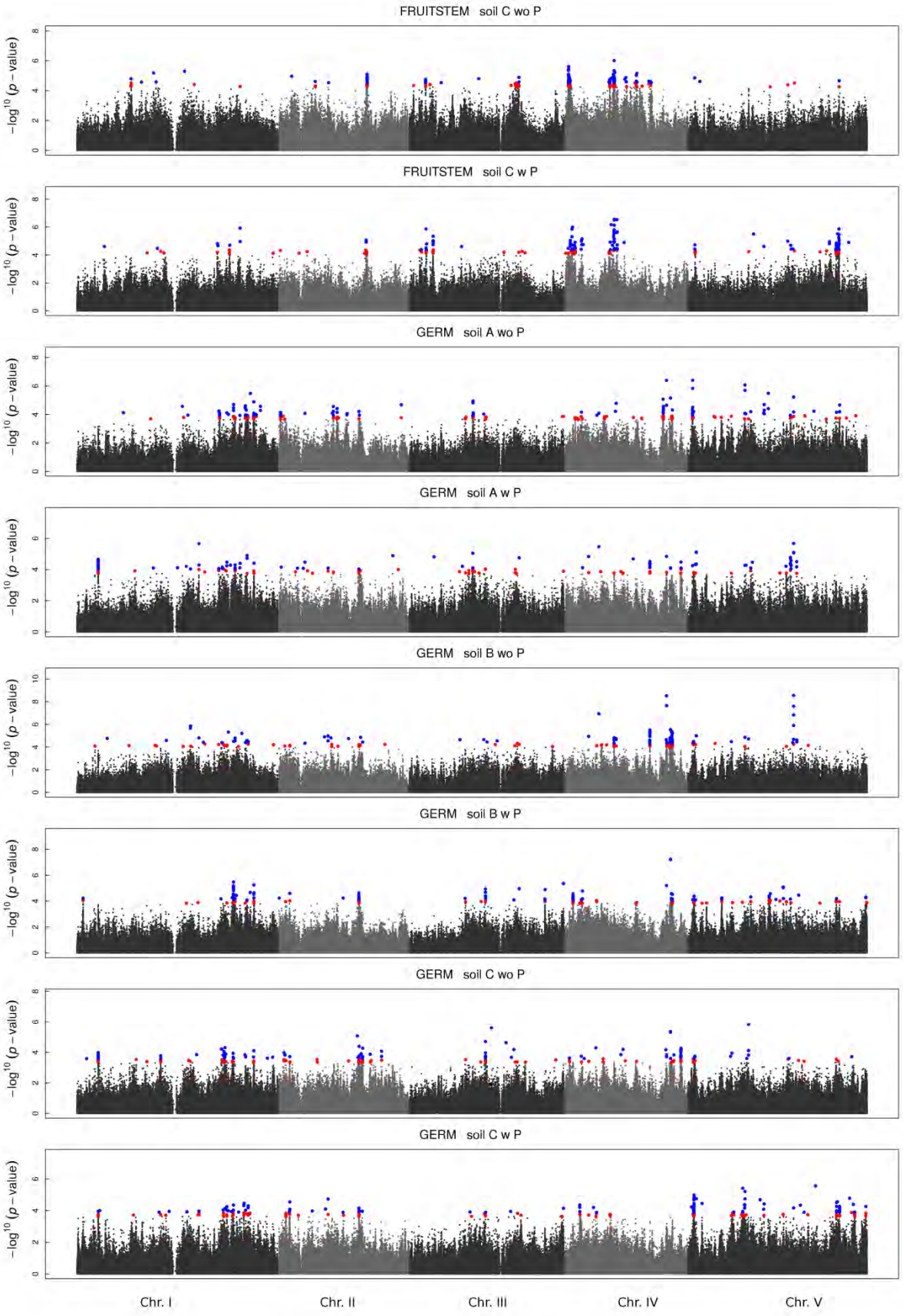


Figure S8 (continued)

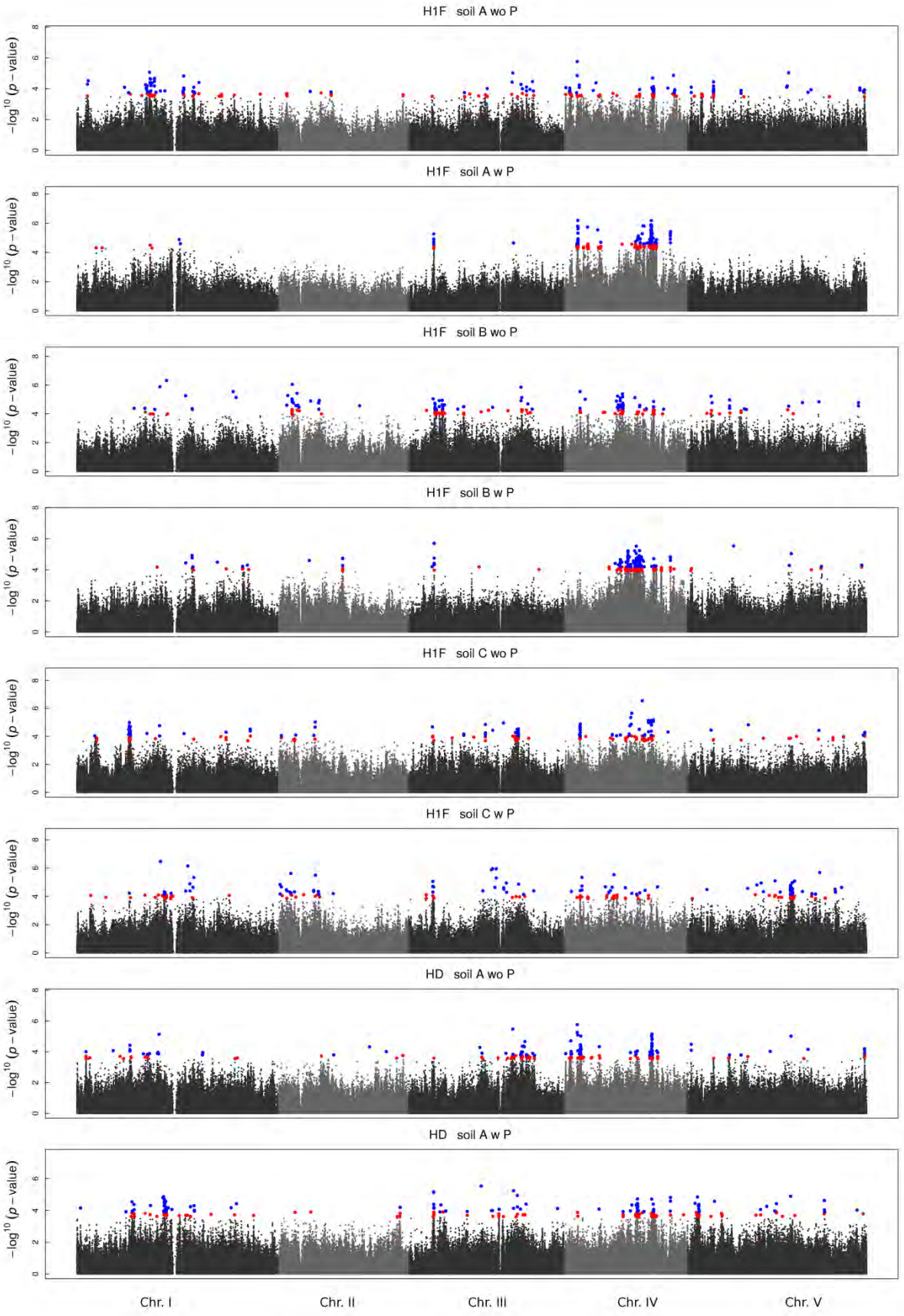


Figure S8 (continued)

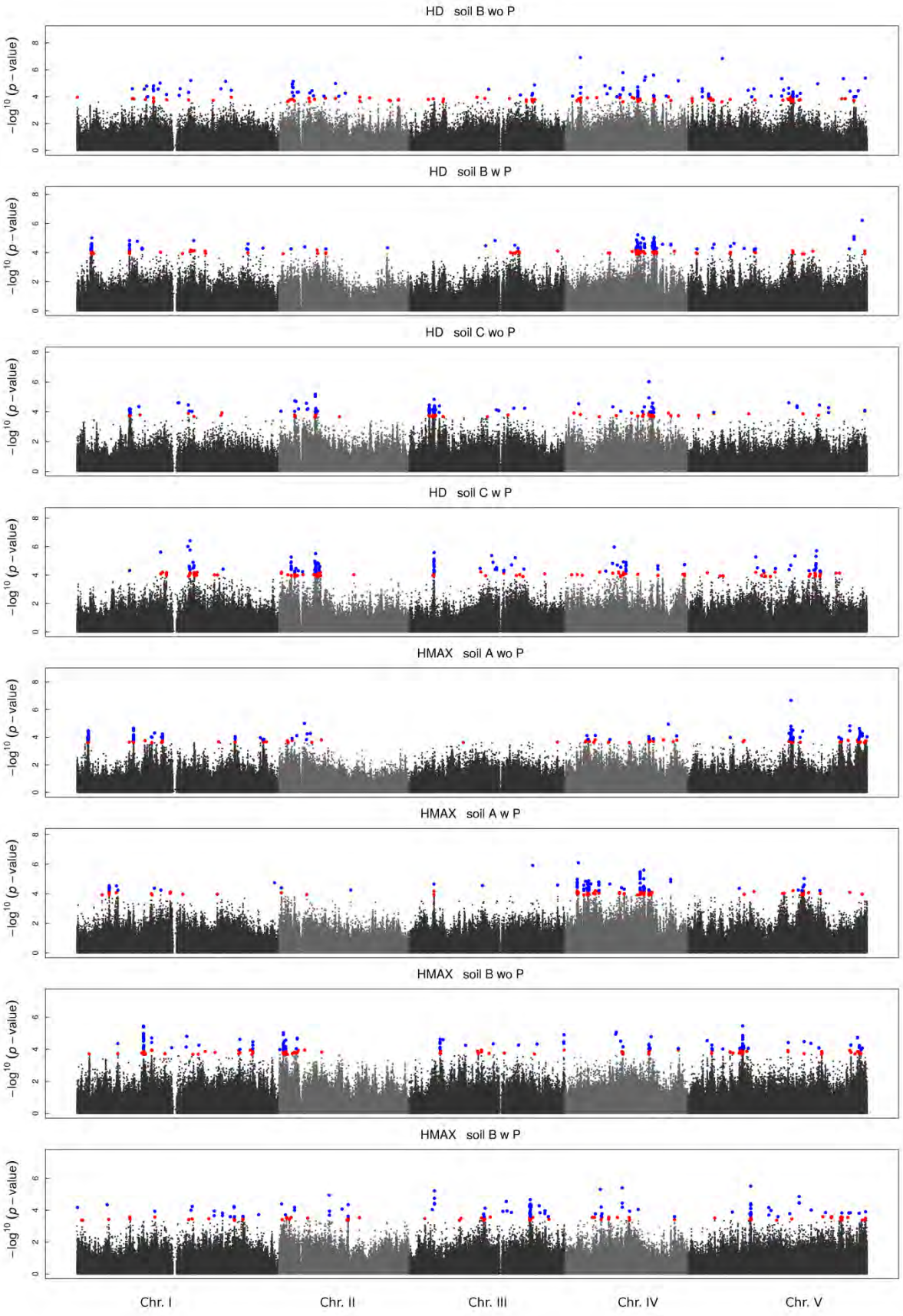


Figure S8 (continued)

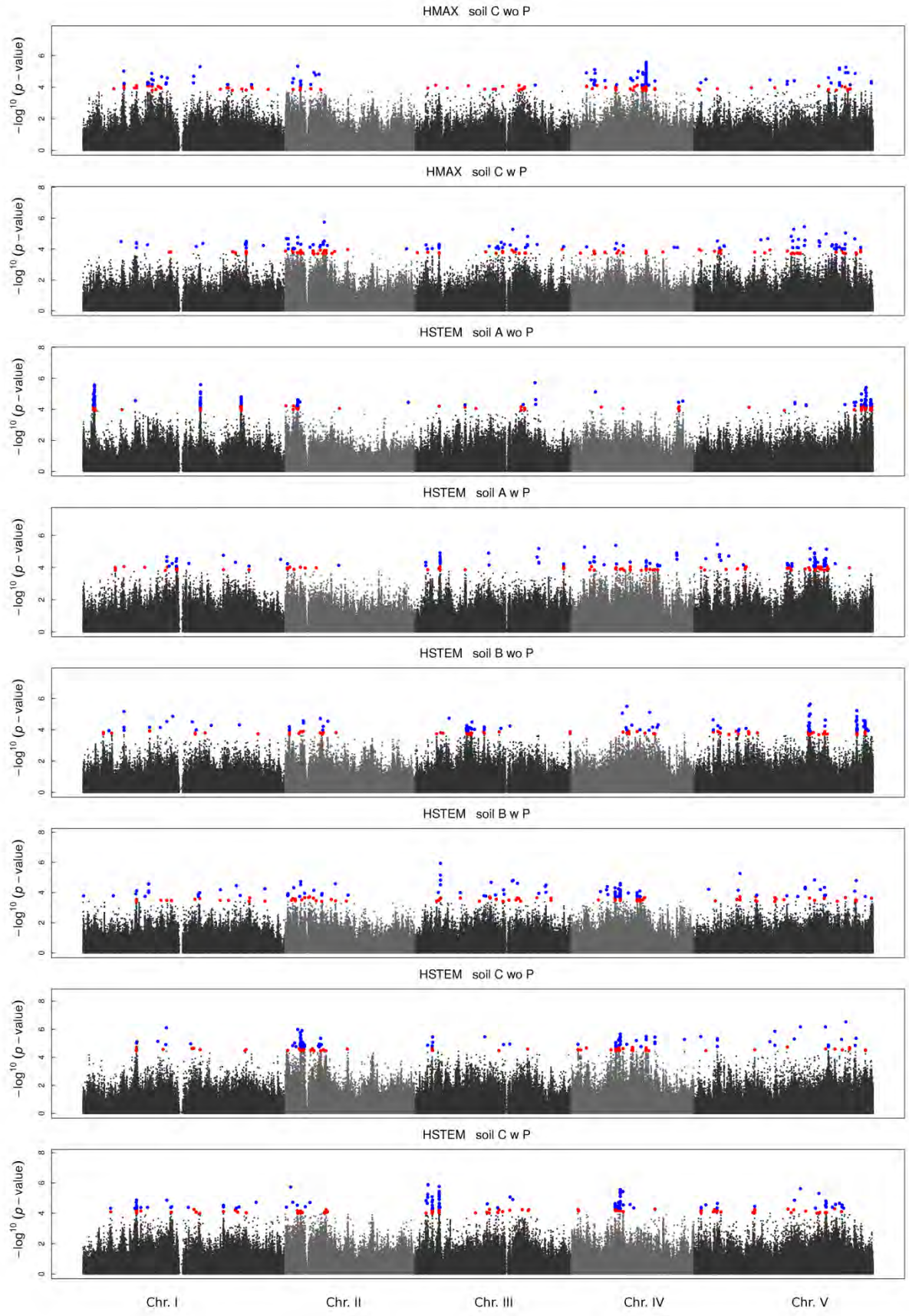


Figure S8 (continued)

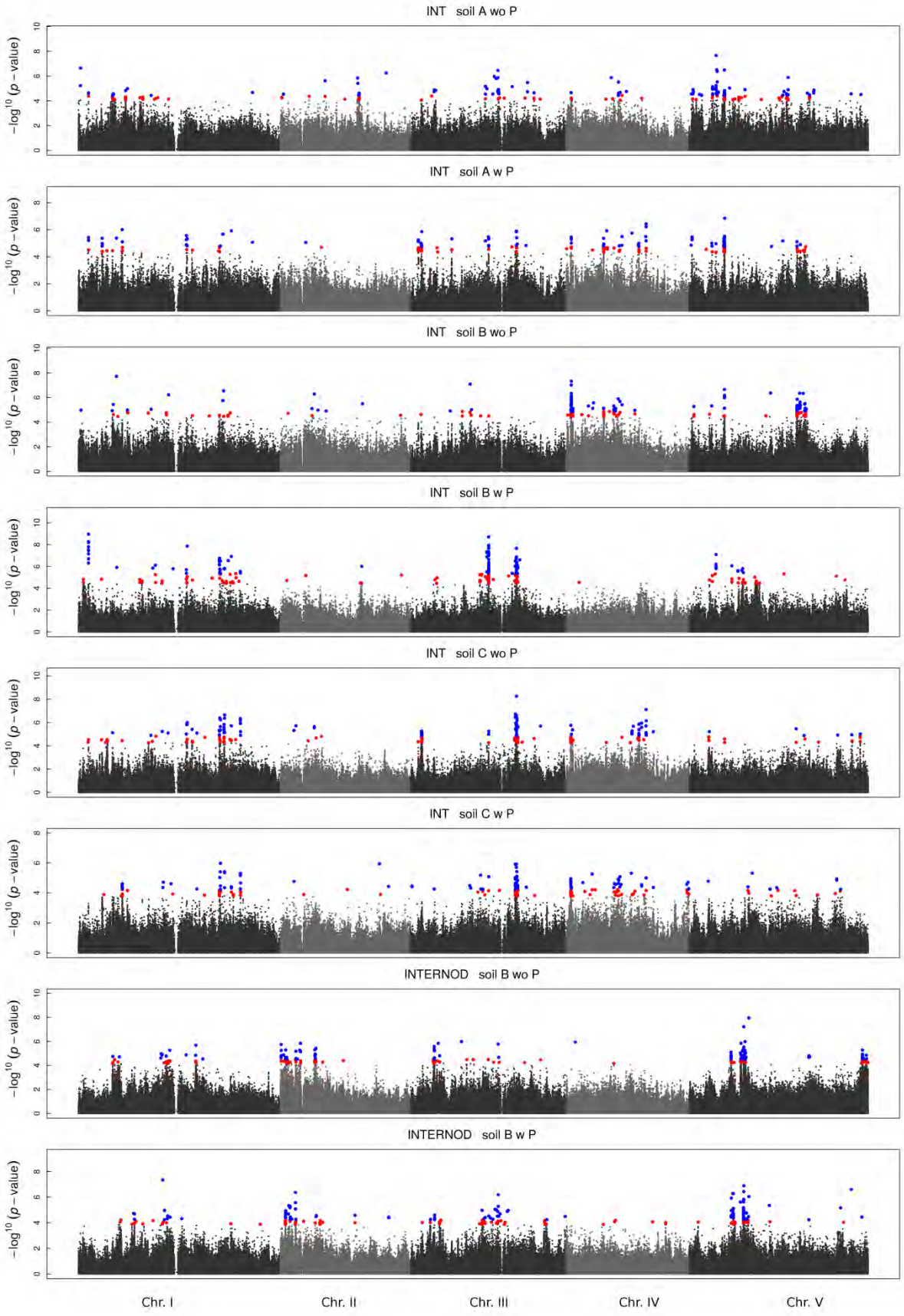


Figure S8 (continued)

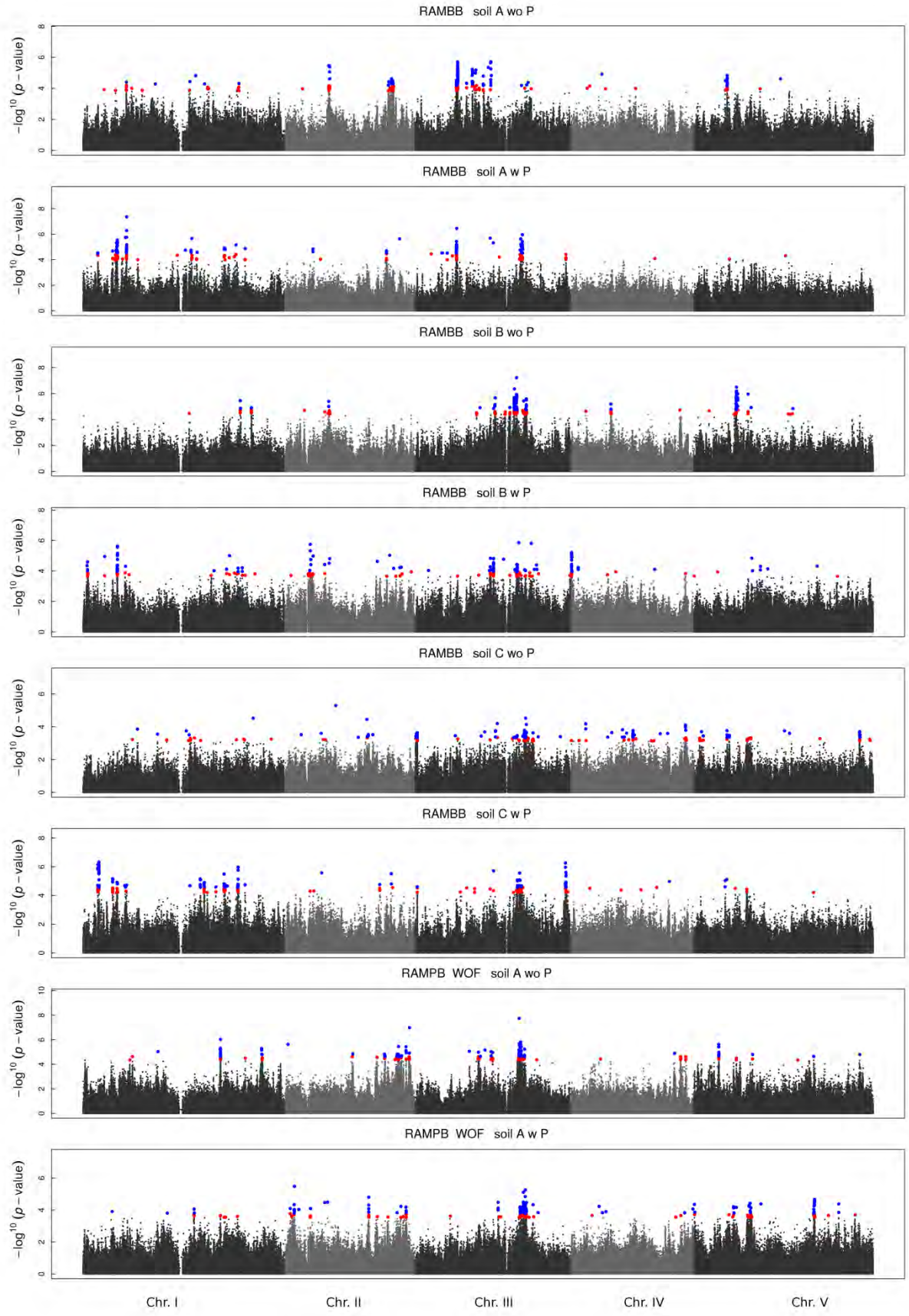


Figure S8 (continued)

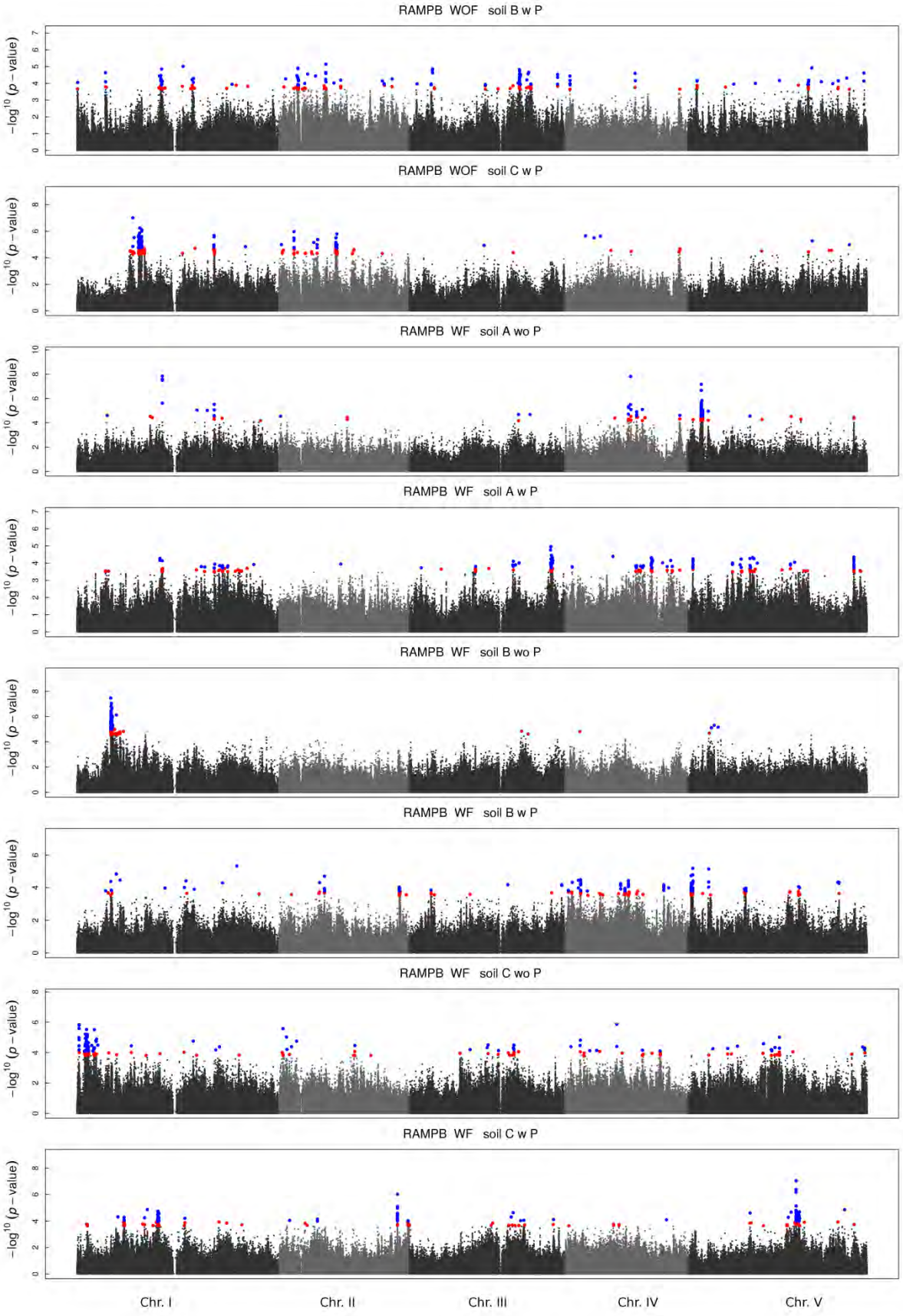


Figure S8 (continued)

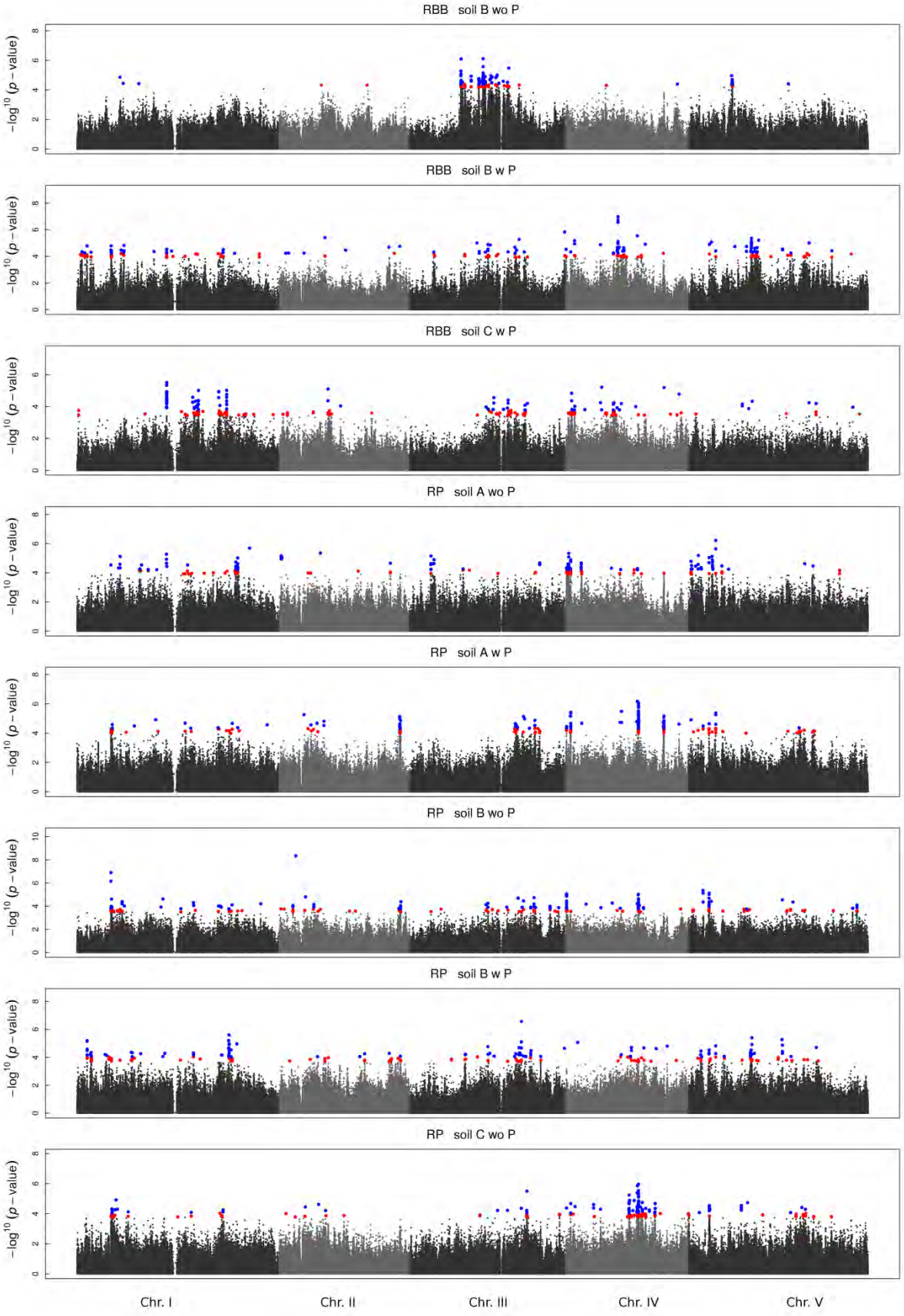


Figure S8 (continued)

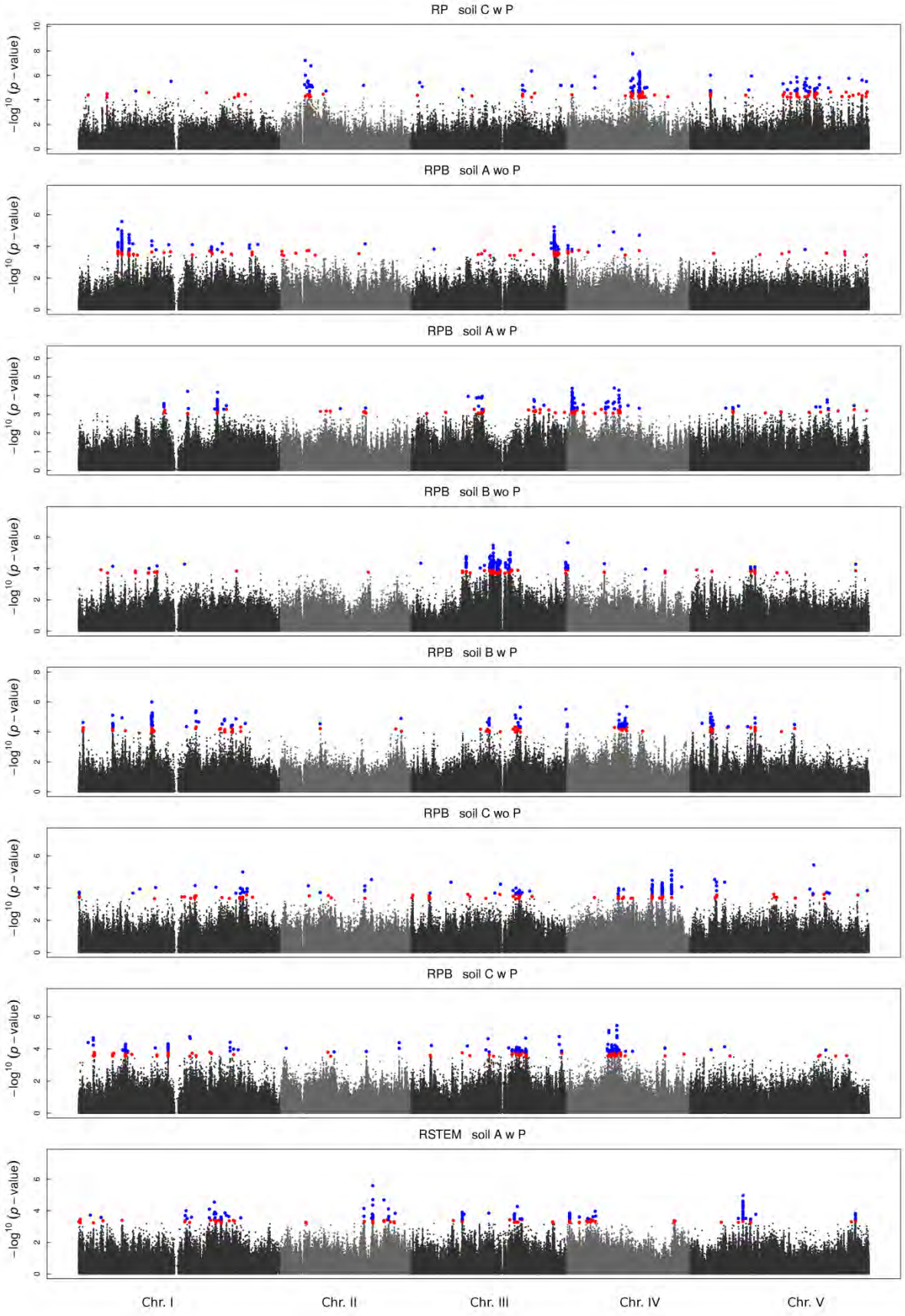


Figure S8 (continued)

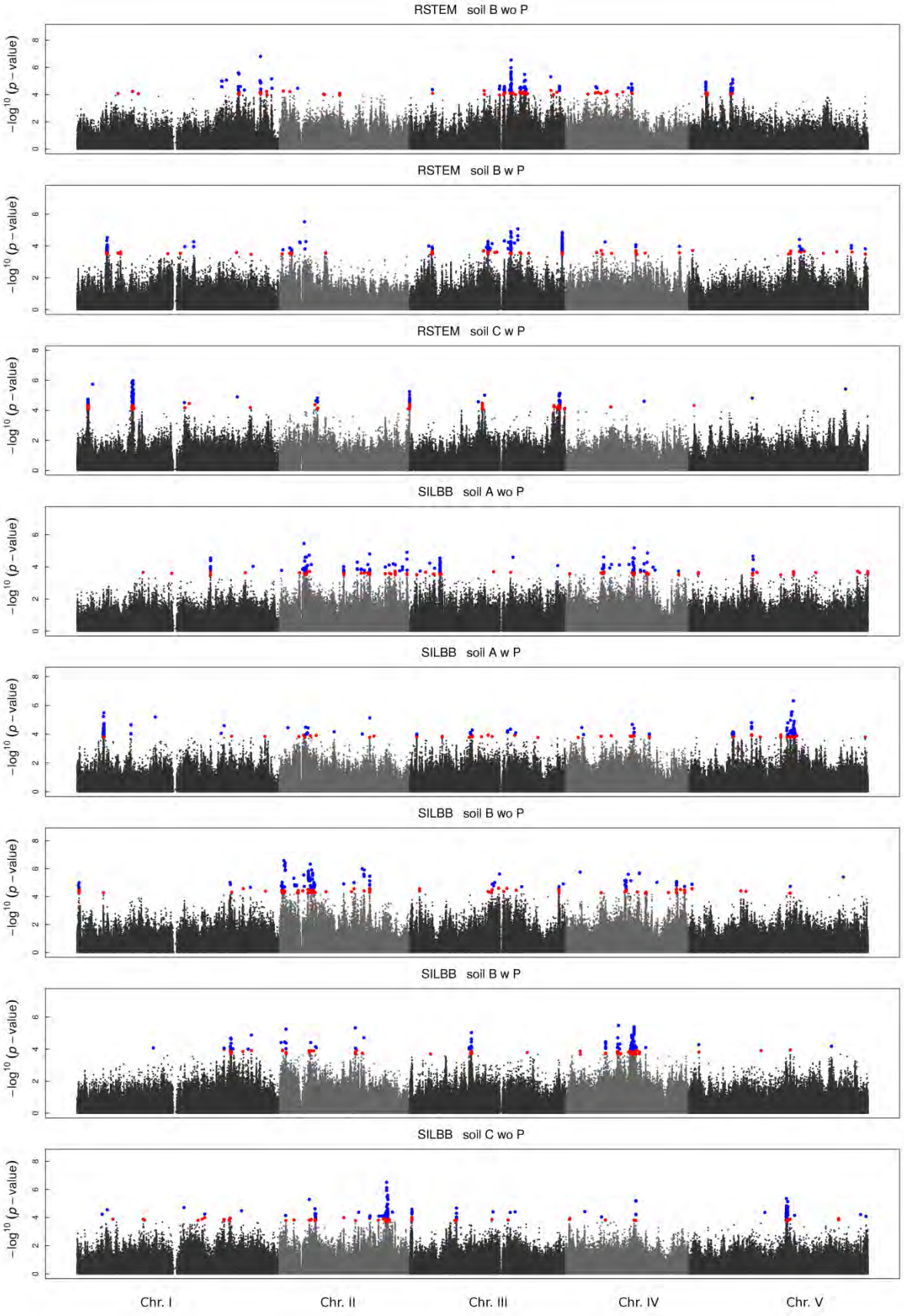


Figure S8 (continued)

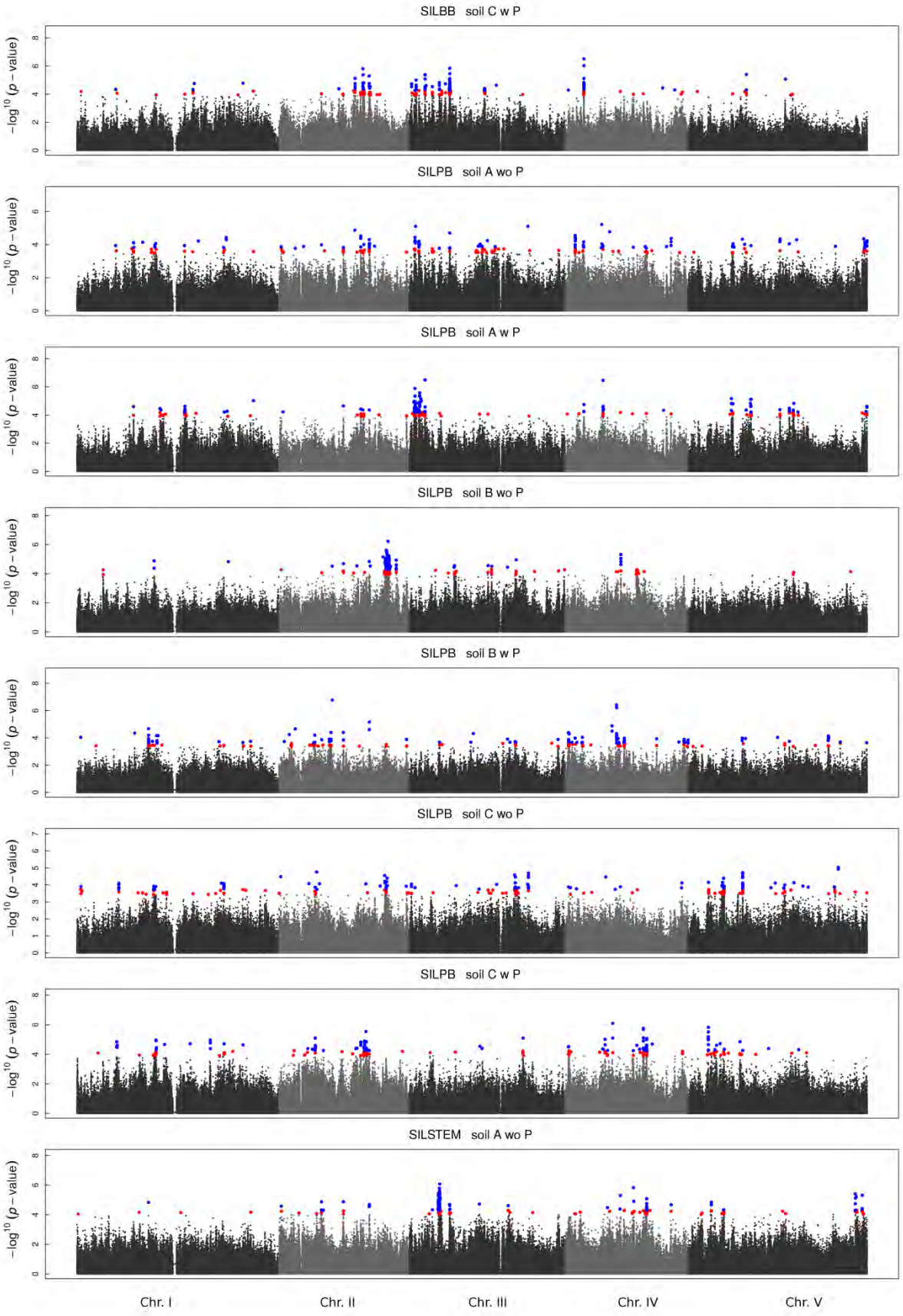


Figure S8 (continued)

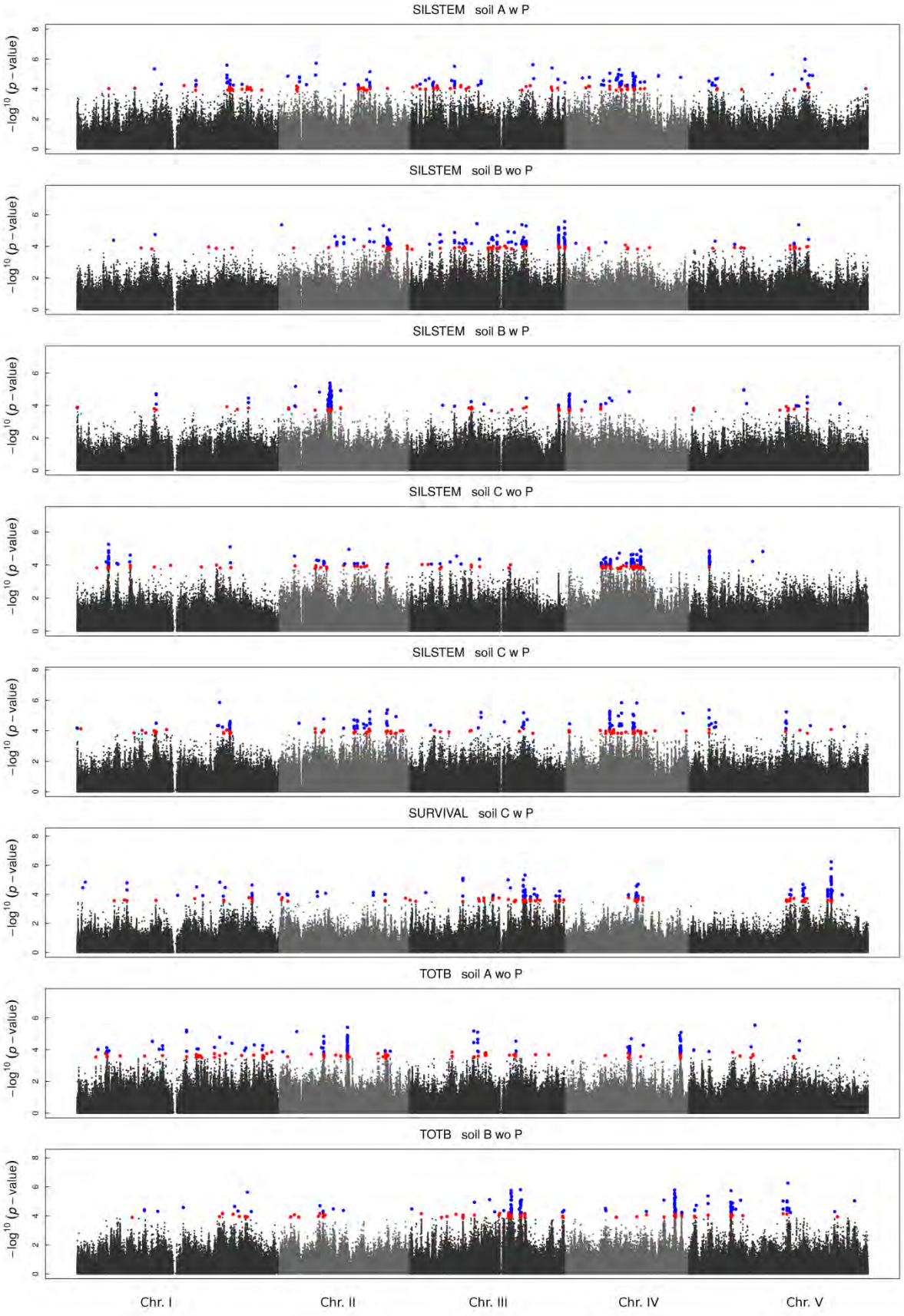


Figure S8 (continued)

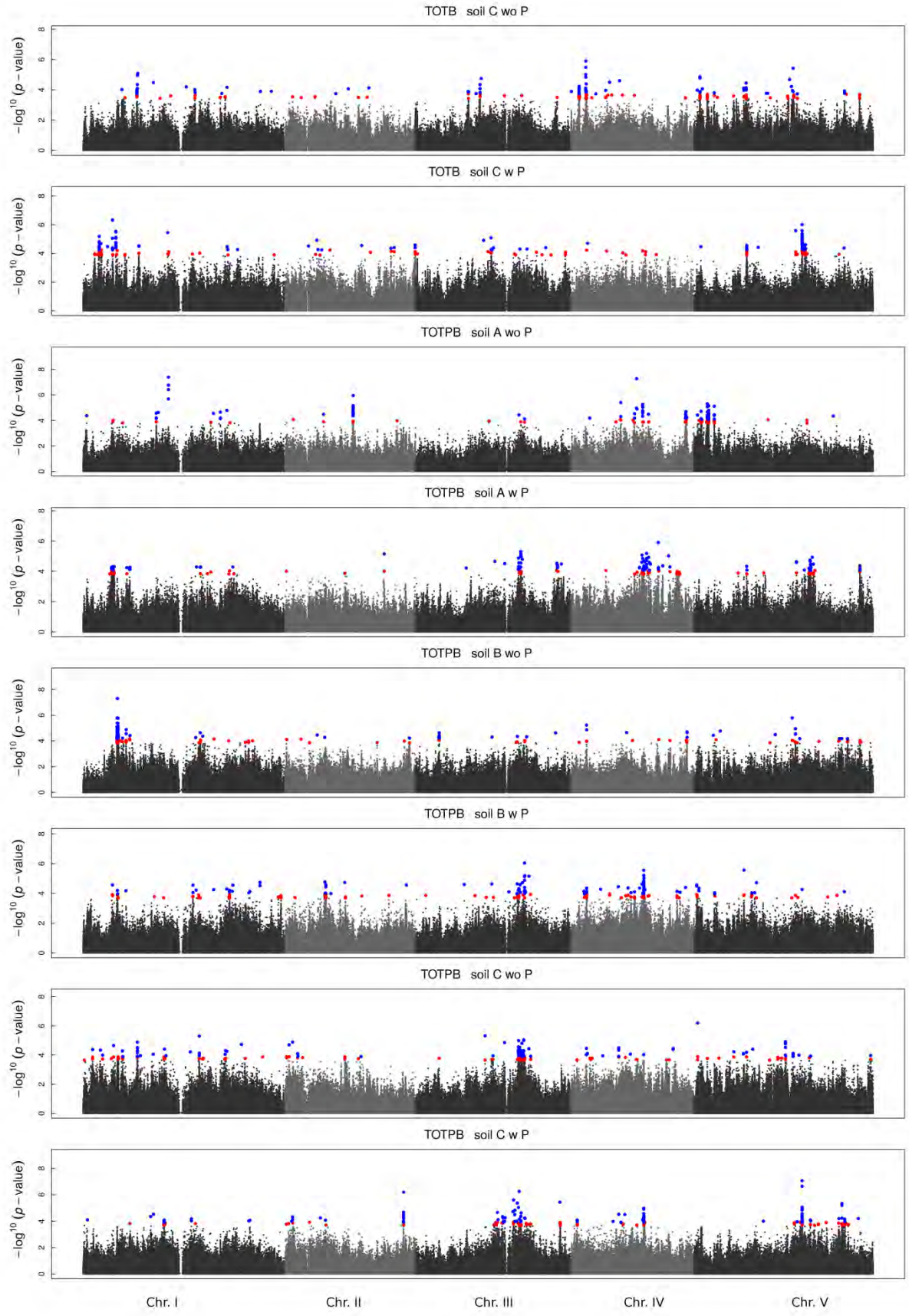


Figure S9 | Number of genes represented in the top 200 SNPs for each of the 144 heritable eco-phenotypes. Genes have been retrieved in a 1kb window size on each side of each top SNP.

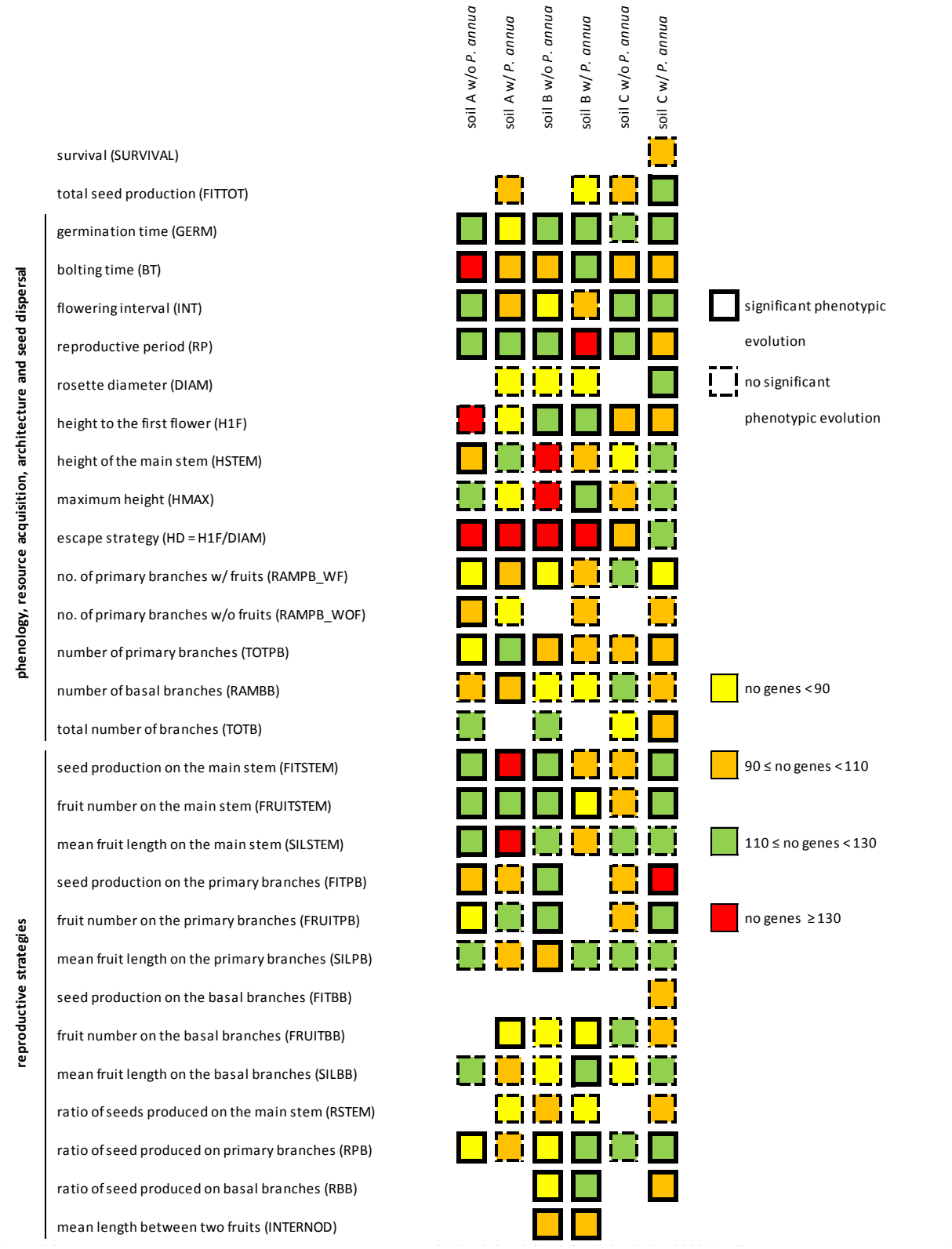


Figure S10 | Non proportional Venn diagram presenting the partitioning of top SNPs associated with the 144 heritable eco-phenotypes between the six *in situ* 'soil x competition' micro-habitats. (i.e. three soils A, B and C x absence or presence of *P. annua*). Numbers in brackets indicate the number of eco-phenotypes for each *in situ* 'soil x competition' micro-habitat.

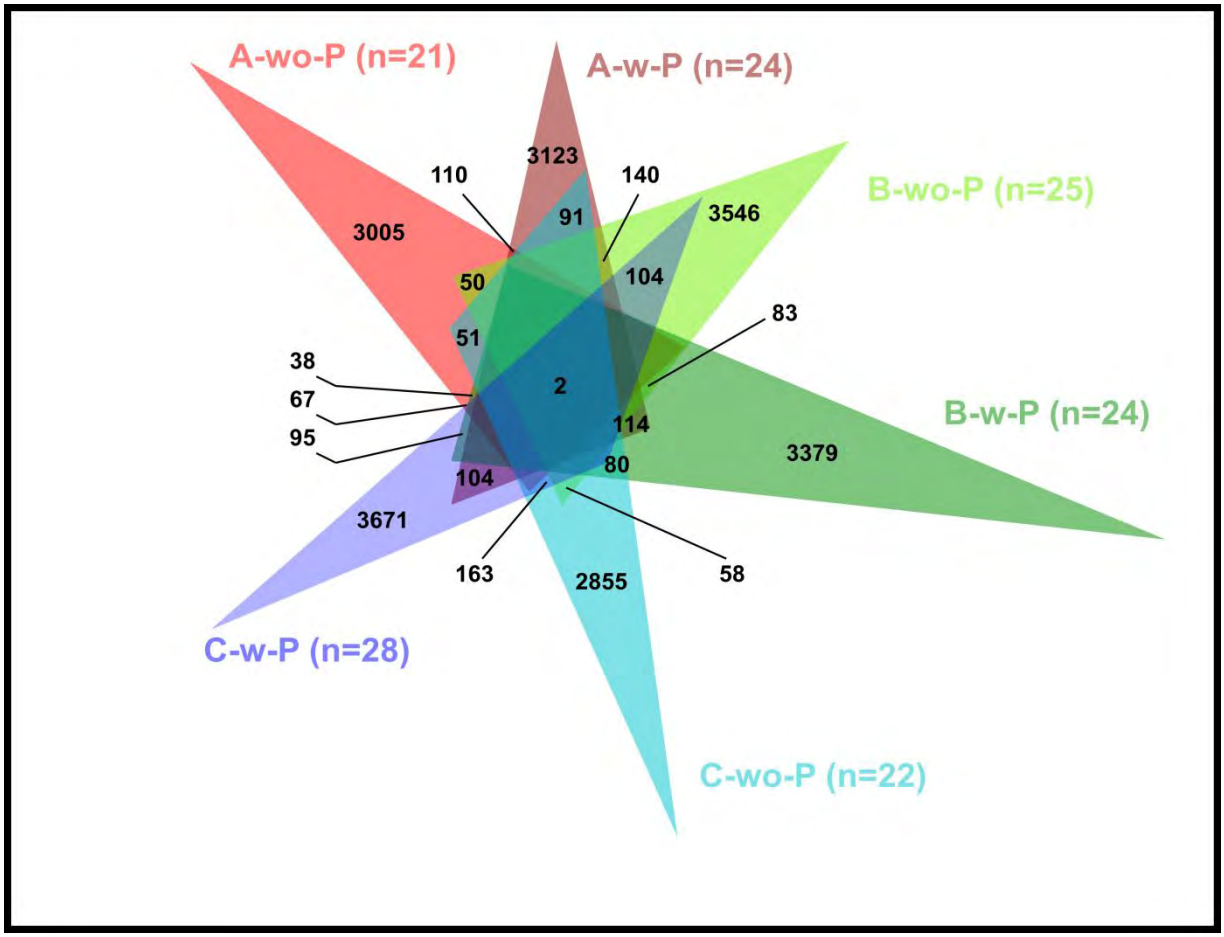


Figure S11A | Degree of pleiotropy and pleiotropic scaling in the TOU-A population when considering a threshold of 50, 100, 300 and 500 top SNPs. (**Top panels**) Frequency distribution of the effective number of eco-phenotypes affected by a SNP (N_{eff}, accounting for the correlations between eco-phenotypes) among the 21,268 unique top SNPs. (**Bottom panels**) Regression of total effect size T_M (total effect size by the Manhattan distance) on N_{eff}. The formula corresponds to the pleiotropic scaling relationship $\boldsymbol{T_M = c^* N_{eff}^{\,d}}$. A scaling component d exceeding 1 indicates that the mean per-trait effect size of a given top SNP increased with N_{eff}[8]. Solid red line: fitted relationship between T_M and N_{eff}, solid black line: linear dependence ($d = 1$).

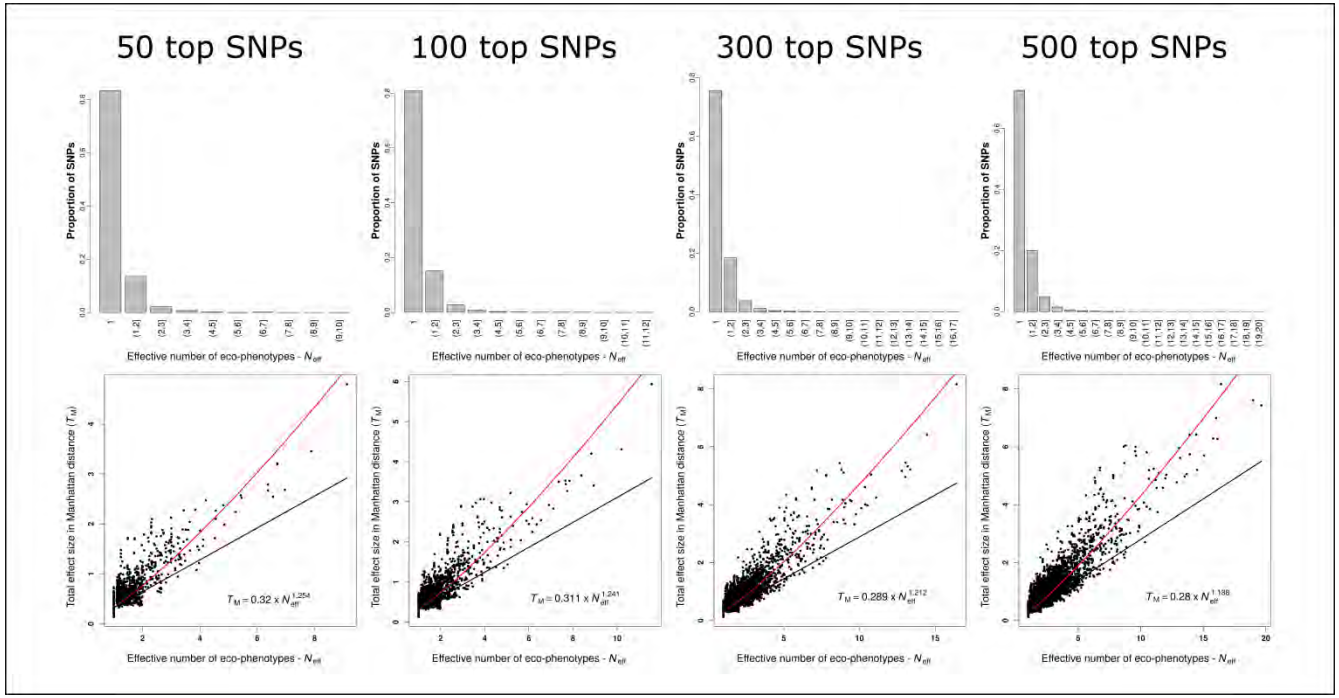


Figure S11B | Significance and strength of selection in the TOU-A population when considering a threshold of 50, 100, 300 and 500 top SNPs. (**Top panels**) Fold-increase in median $-\log_{10}$ (*p*-values) of neutrality tests based on temporal differentiation for SNPs that hit only evolved eco-phenotypes, only unevolved eco-phenotypes or both types of eco-phenotypes, according to different classes of effective number of eco-phenotypes. The dashed line corresponds to a fold-increase of 1, i.e. no increase in median significance of neutrality tests based on temporal differentiation. (**Bottom panels**) Fold-increase in median F_{ST} values for SNPs that hit only evolved eco-phenotypes, only unevolved eco-phenotypes or both types of eco-phenotypes, according to different classes of N_{eff} (median F_{ST} across the genome = 0.00293). Significance against a null distribution obtained by bootstrapping: *0.05 > *P* > 0.01, **0.01 > *P* > 0.001, ***$P < 0.001$, ns: non-significant.

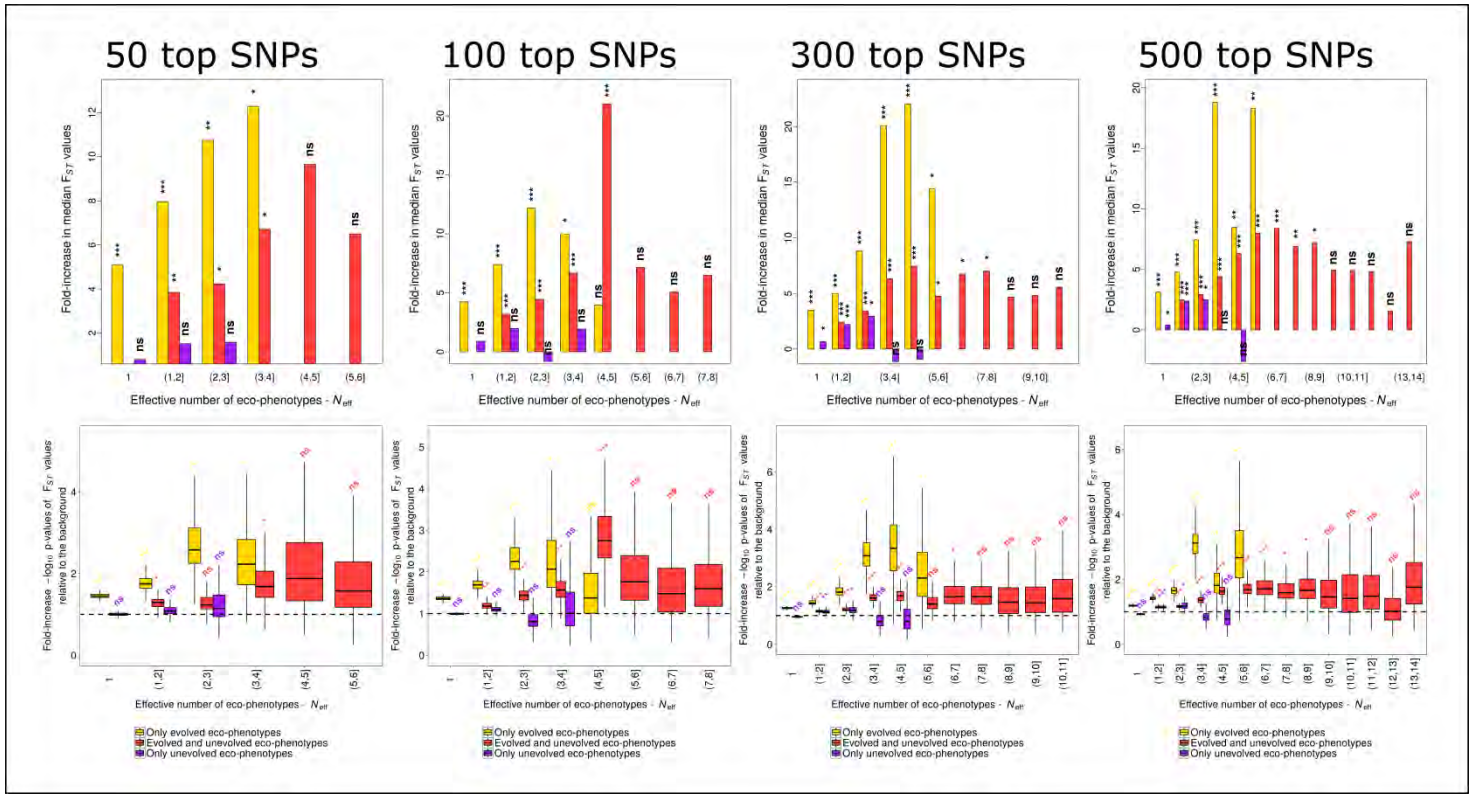


Figure S11C | Degree of pleiotropy and pleiotropic scaling in the TOU-A population when considering SNPs with a $-\log_{10}$ *p*-value above 6, 5 and 4. (**Top panels**) Frequency distribution of the effective number of eco-phenotypes affected by a SNP (N_{eff}, accounting for the correlations between eco-phenotypes) among the 21,268 unique top SNPs. (**Bottom panels**) Regression of total effect size T_M (total effect size by the Manhattan distance) on N_{eff}. The formula corresponds to the pleiotropic scaling relationship $\boldsymbol{T_M = c^* N_{eff}^{\,d}}$. A scaling component d exceeding 1 indicates that the mean per-trait effect size of a given top SNP increased with N_{eff}[8]. Solid red line: fitted relationship between T_M and N_{eff}, solid black line: linear dependence ($d = 1$).

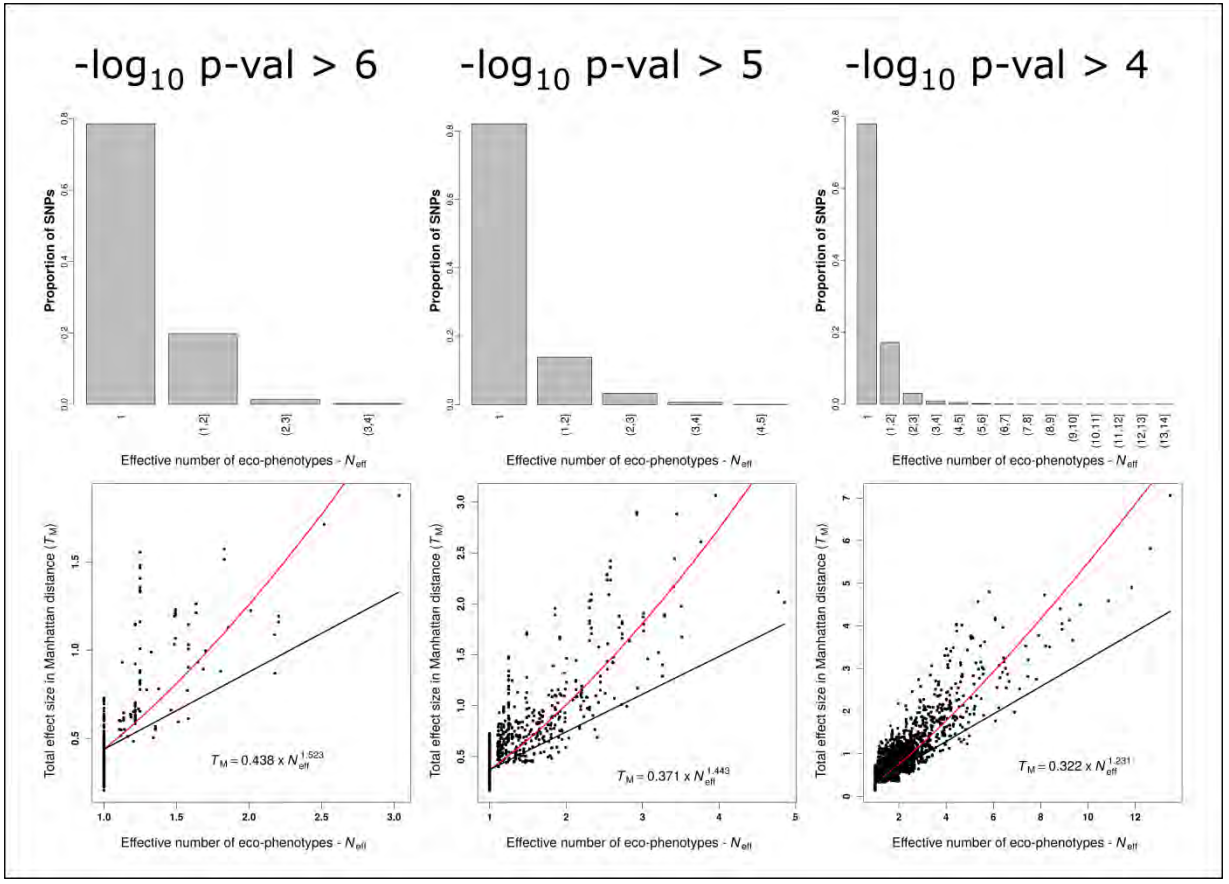


Figure S11D | Significance and strength of selection in the TOU-A population when SNPs with a $-\log_{10}$ *p*-value above 6, 5 and 4. (**Top panels**) Fold-increase in median $-\log_{10}$ (*p*-values) of neutrality tests based on temporal differentiation for SNPs that hit only evolved eco-phenotypes, only unevolved eco-phenotypes or both types of eco-phenotypes, according to different classes of effective number of eco-phenotypes. The dashed line corresponds to a fold-increase of 1, i.e. no increase in median significance of neutrality tests based on temporal differentiation. (**Bottom panels**) Fold-increase in median F_{ST} values for SNPs that hit only evolved eco-phenotypes, only unevolved eco-phenotypes or both types of eco-phenotypes, according to different classes of N_{eff} (median F_{ST} across the genome = 0.00293). Significance against a null distribution obtained by bootstrapping: *0.05 > *P* > 0.01, **0.01 > *P* > 0.001, ***$P < 0.001$, ns: non-significant.

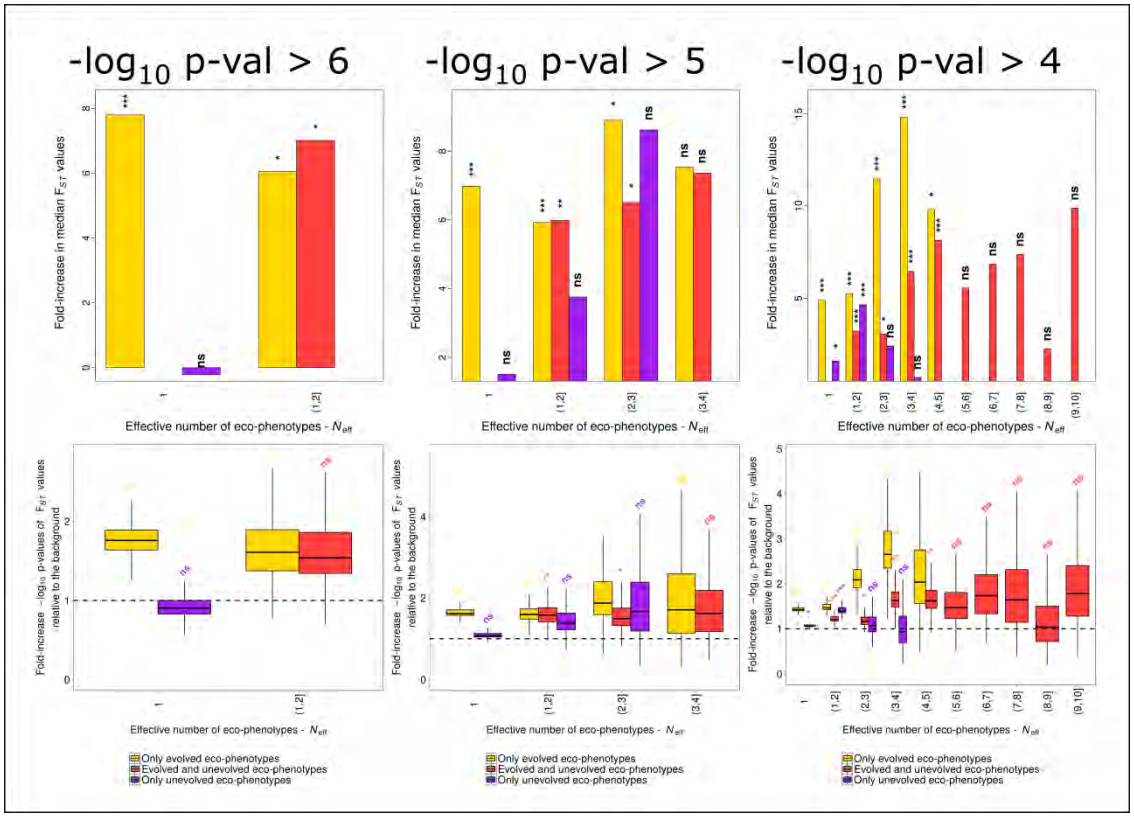


Figure S12 | Scaling relationships between total phenotypic effect size of the 200 top SNPs and the number of eco-phenotypes (*N*, left panels) or the effective number of eco-phenotypes (N_{eff}, right panels). The pleiotropic scaling relationship was calculated as (i) $T_M = c^* {N_{eff}}^d$, with T_M corresponding to the Manhattan distance (bottom panels) and (ii) $T_E = a^* Neff^b$, with T_E corresponding to the Euclidean distance (top panels).

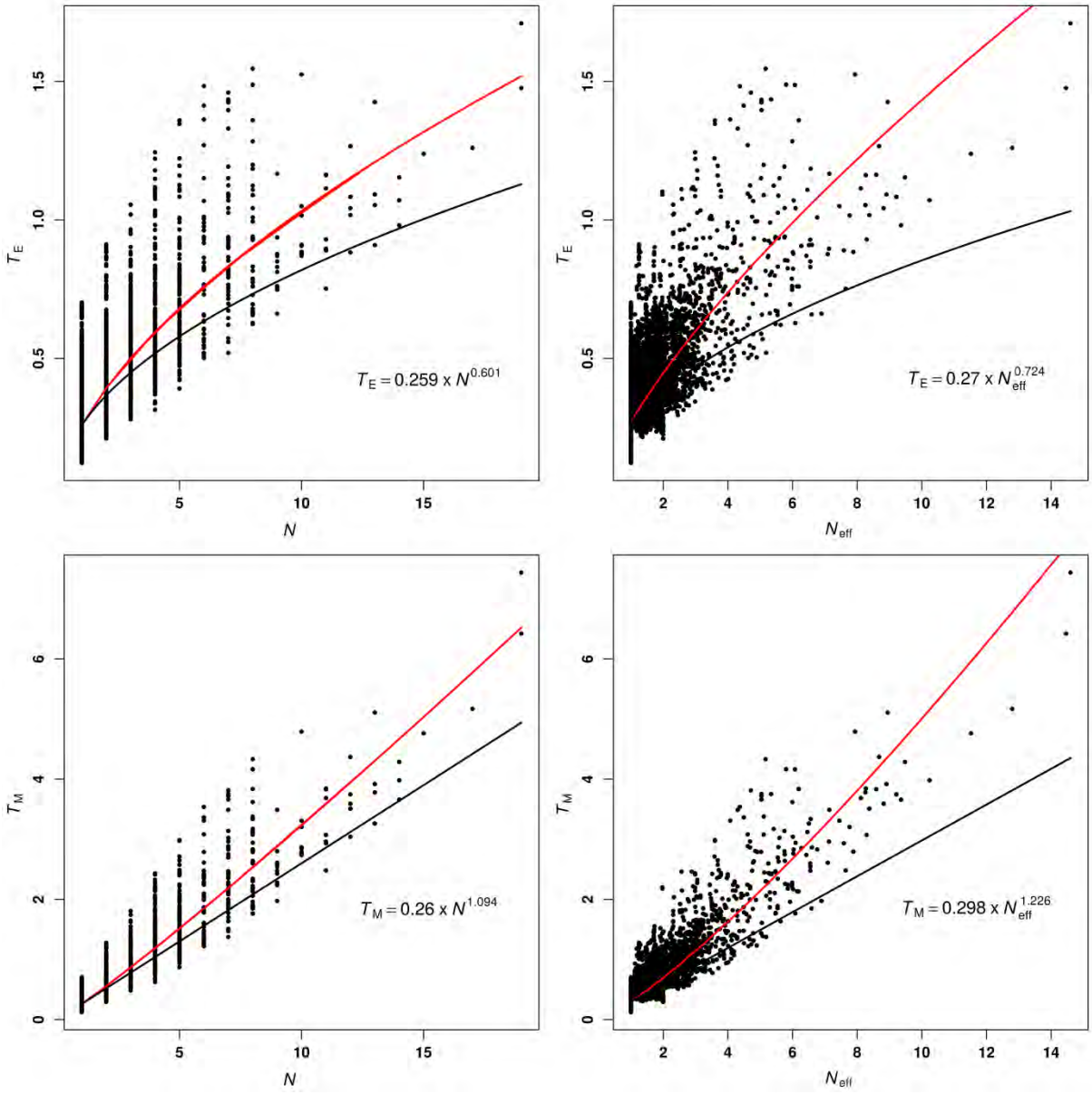


Figure S13 | Genome-wide scan for selection based on temporal differentiation. (**a**) Manhattan plot of F_{ST} at each SNP marker (dots) along the *A. thaliana* genome. The blue dashed line corresponds to the 0.1% upper tail of the F_{ST} value distribution (n = 982). Median F_{ST} across the genome = 0.00293. (**b**) -$\log_{10}$ (*p*-value) of the simulation-based test of the null hypothesis that the locus-specific differentiation measured at each SNP is only due to genetic drift. Only SNP markers with MARF > 7% are considered.

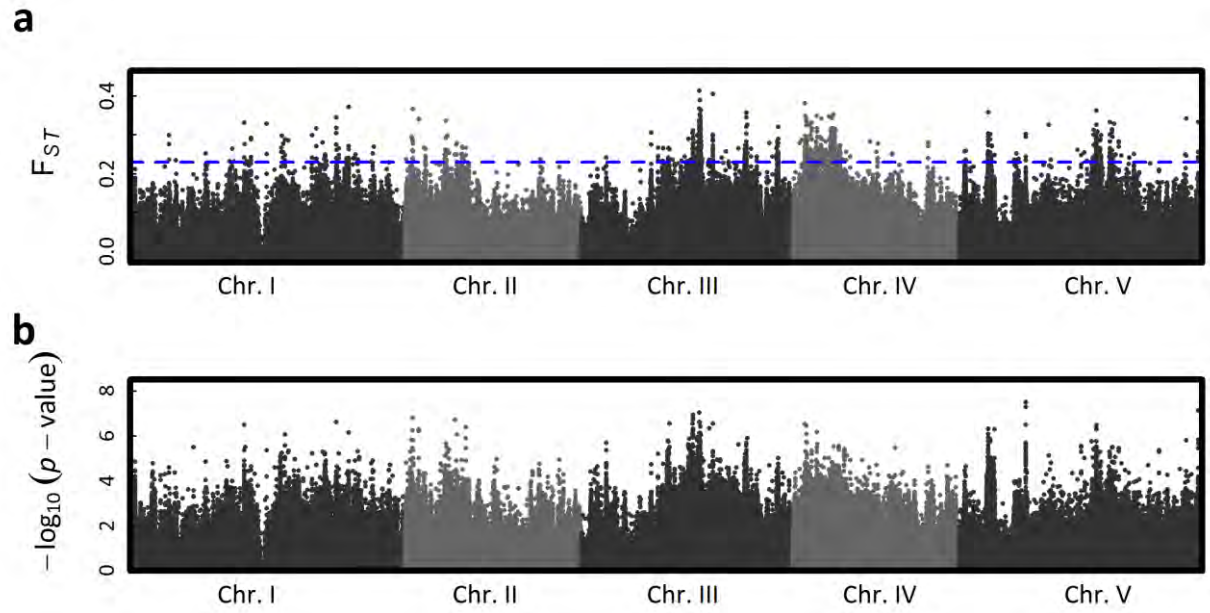


Figure S14 | Polarity of effects. (**a**) Proportion of top SNPs associated with evolved eco-phenotypes with a polarity of effects in line with the direction of phenotypic evolution, according to different classes of N_{eff}. (**b**) Effect of polarity effects on the fold-increase in median F_{ST} values for SNPs that hit only evolved eco-phenotypes, according to different classes of N_{eff} (median F_{ST} across the genome = 0.00293). Significance against a null distribution obtained by bootstrapping: *0.05 > P > 0.01, **0.01 > P > 0.001, ***P < 0.001, absence of symbols: non-significant. Due to the small number of SNPs with an effective number of eco-phenotypes above 4, those SNPs were grouped for testing the significance of fold-increase in median F_{ST} values.

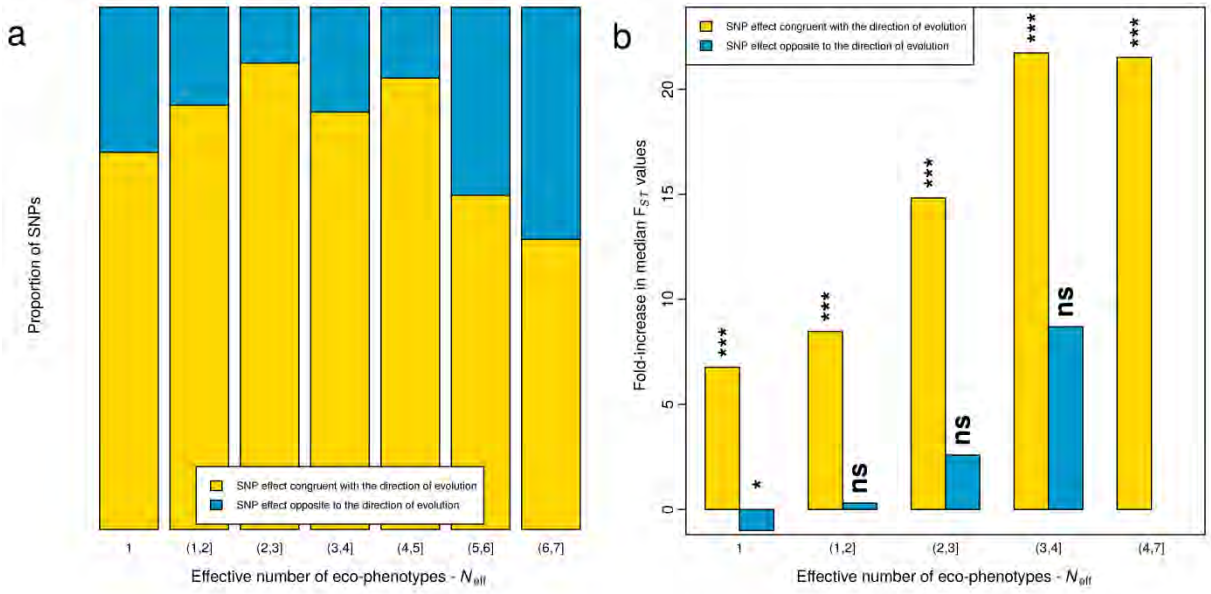


Figure S15 | The distribution dependence of *p*-value distribution on minor allele relative frequency (MARF) for EMMAX across the 144 eco-phenotypes (see Fig. 1). For a given MARF value, each point corresponds to the quantile at 99% of the *p*-value distribution of one of the 144 heritable eco-phenotypes. A locally-weighted polynomial regression is illustrated by a red solid line. A MARF threshold above 7% is depicted by a dashed blue line.

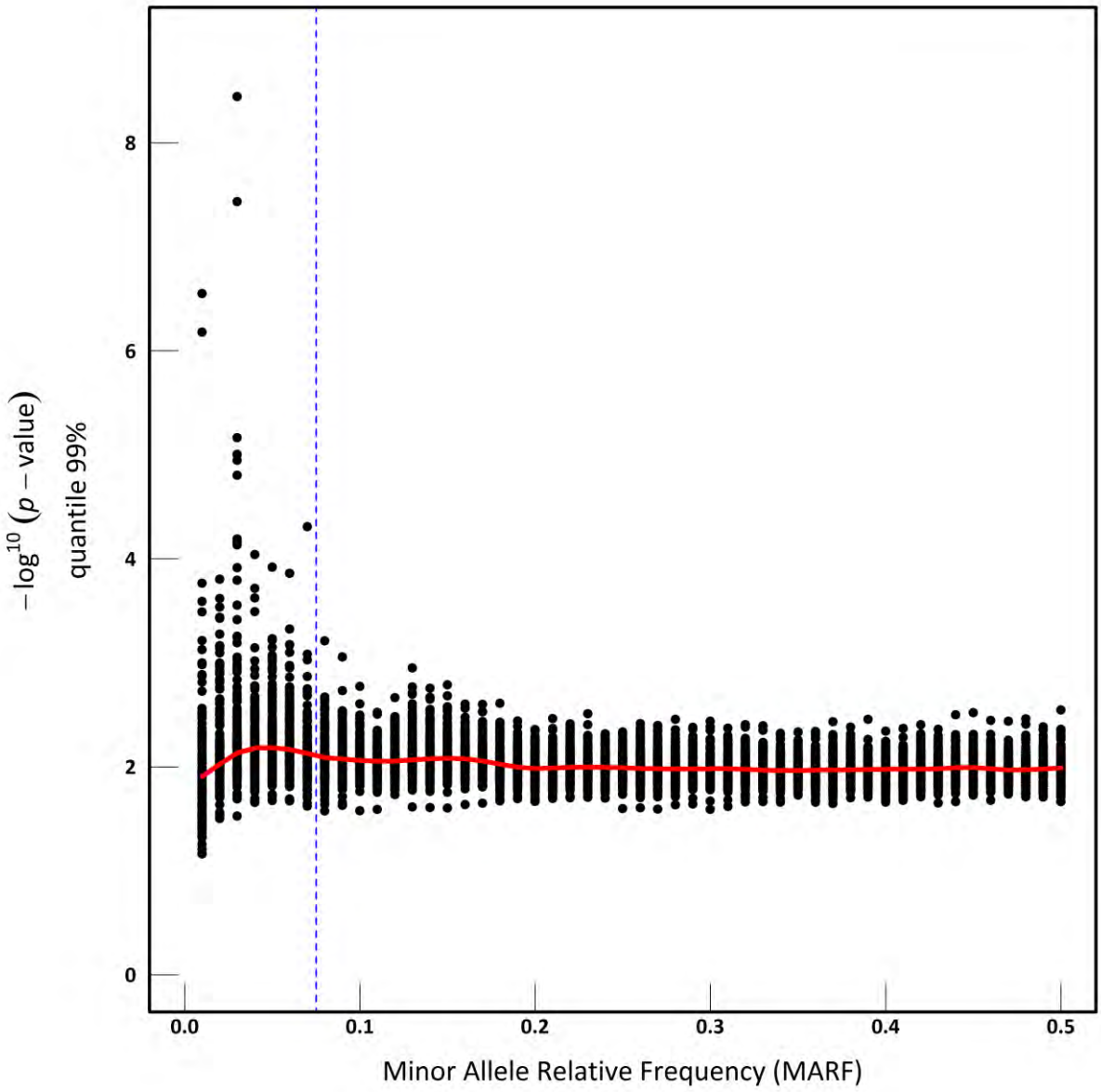


Table S1 | Phenotypic variation of 195 accessions sampled in 2002 and 2010 and scored across six *in situ* 'soil x competition' micro-habitats.

	Model terms§																					
	block (soil*comp)		soil		comp		soil*comp		year		soil*year		comp*year		soil*comp* year		acc (year)		acc(year)* soil		acc(year)* comp		acc(year)* soil*comp	
Traits ǂ	*F*	*P*	*F*	*P*	*F*	*P*	*F*	*P*	*F*	*P*	*F*	*P*	*F*	*P*	*F*	*P*	*LRT*	*P*	*LRT*	*P*	*LRT*	*P*	*LRT*	*P*
GERM	5.40	***	53.44	***	104.47	***	64.64	***	8.61	**	10.13	***	0.51	ns	0.08	ns	299.1	***	0.0	ns	0.0	ns	7.9	*
BT	3.90	***	23.57	***	120.85	***	25.07	***	13.46	**	11.13	***	22.40	***	4.54	***ns***	280.7	***	5.0	*	13.5	**	5.8	***ns***
INT	1.65	*	29.20	***	66.50	***	42.51	***	13.22	**	7.93	**	16.85	***	2.39	ns	140.4	***	0.6	ns	16.1	**	4.1	***ns***
RP	8.24	***	132.02	***	45.37	***	20.95	***	19.18	***	3.65	***ns***	12.60	**	1.62	ns	287.4	***	17.3	***	2.7	ns	0.0	ns
DIAM	5.28	***	75.04	***	57.82	***	46.57	***	0.16	ns	5.34	*	0.11	ns	0.12	ns	40.1	***	2.8	ns	0.0	ns	0.2	ns
H1F	2.41	***	177.58	***	31.60	***	86.96	***	7.29	*	1.41	ns	0.64	ns	2.37	ns	125.5	***	10.7	**	0.8	ns	1.3	ns
HSTEM	4.01	***	342.68	***	14.99	***	55.05	***	0.71	ns	0.08	ns	1.51	ns	2.41	ns	178.0	***	49.9	***	0.2	ns	0.0	ns
HMAX	6.63	***	584.30	***	4.24	***ns***	84.33	***	0.39	ns	0.88	ns	0.17	ns	2.72	ns	162.5	***	43.4	***	0.2	ns	0.0	ns
HD	1.82	*	77.84	***	99.75	***	27.52	***	8.96	**	2.19	ns	2.90	ns	0.89	ns	175.8	***	0.0	ns	0.0	ns	8.9	*
RAMPB_WF	2.73	***	34.20	***	16.26	***	37.44	***	13.43	**	0.20	ns	0.19	ns	3.42	***ns***	47.8	***	6.7	*	7.1	***ns***	0.0	ns
RAMPB_WOF	1.43	ns	2.01	ns	3.89	ns	2.58	ns	0.41	ns	1.88	ns	4.45	***ns***	0.36	ns	53.8	***	1.3	ns	0.0	ns	0.0	ns
TOTPB	2.43	***	53.77	***	13.13	***	48.04	***	12.57	**	1.92	ns	3.18	ns	6.41	*	118.9	***	5.3	*	4.8	***ns***	0.0	ns
RAMBB	2.90	***	9.94	***	120.87	***	14.61	***	3.28	ns	3.00	ns	0.84	ns	0.12	ns	68.2	***	1.9	ns	1.3	ns	0.1	ns
TOTB	3.48	***	42.53	***	113.24	***	49.52	***	1.95	ns	0.35	ns	0.15	ns	1.57	ns	42.3	***	2.1	ns	3.3	ns	0.0	ns
FITTOT	5.07	***	201.80	***	28.37	***	35.21	***	0.29	ns	0.31	ns	0.22	ns	1.61	ns	20.5	***	13.1	**	0.0	ns	0.0	ns
FITSTEM	3.23	***	161.17	***	0.79	ns	14.78	***	7.09	*	0.14	ns	0.46	ns	4.16	***ns***	119.6	***	44.5	***	0.1	ns	0.0	ns
FRUITSTEM	3.01	***	83.92	***	0.49	ns	9.86	***	8.97	**	0.09	ns	0.00	ns	3.18	***ns***	85.6	***	47.9	***	0.1	ns	0.1	ns
SILSTEM	3.76	***	434.03	***	101.02	***	31.50	***	1.78	ns	2.39	ns	2.52	ns	1.96	ns	256.6	***	20.8	***	3.0	ns	0.0	ns
FITPB	3.24	***	197.91	***	0.19	ns	42.05	***	9.71	**	0.99	ns	0.78	ns	8.94	**	25.0	***	4.8	*	0.0	ns	8.1	*
FRUITPB	3.49	***	158.53	***	1.19	ns	45.88	***	10.64	**	0.72	ns	0.01	ns	6.29	*	22.7	***	12.7	**	0.0	ns	2.0	ns
SILPB	3.86	***	450.20	***	33.62	***	25.65	***	0.34	ns	1.11	ns	1.42	ns	4.17	***ns***	211.2	***	8.1	*	0.0	ns	0.0	ns
FITBB	1.82	*	20.38	***	36.04	***	2.59	ns	0.21	ns	2.51	ns	0.23	ns	0.17	ns	7.9	**	1.7	ns	0.0	ns	0.1	ns
FRUITBB	2.81	***	24.36	***	95.10	***	8.08	***	2.16	ns	2.18	ns	0.56	ns	0.14	ns	41.4	***	8.3	*	0.0	ns	0.7	ns
SILBB	2.48	***	148.46	***	12.90	***	16.34	***	0.03	ns	1.91	ns	0.14	ns	2.66	ns	149.6	***	6.0	*	0.0	ns	0.0	ns
RSTEM	3.12	***	76.90	***	34.61	***	19.77	***	0.97	ns	0.09	ns	1.87	ns	0.17	ns	25.9	***	5.7	*	0.3	ns	0.0	ns
RPB	1.61	*	67.83	***	42.64	***	5.22	**	7.69	*	0.56	ns	1.43	ns	4.39	***ns***	55.9	***	6.9	*	0.0	ns	0.2	ns
RBB	1.73	*	19.98	***	27.71	***	0.17	ns	15.53	***	1.14	ns	0.91	ns	0.51	ns	6.0	***ns***	3.3	ns	0.9	ns	0.2	ns
INTERNOD	1.21	ns	32.07	***	1.77	ns	1.10	ns	3.42	ns	0.03	ns	0.00	ns	1.79	ns	2.7	ns	2.0	ns	0.0	ns	0.0	ns
SURVIVAL	39.31	***	57.15	***	0.06	ns	47.30	***	Inf	***	Inf	***	Inf	***	Inf	**	1.0	ns	3.5	ns	0.0	ns	ne	ne

*0.05 > *P* > 0.01, **0.01 > *P* > 0.001, ***$P < 0.001$. ns: non-significant, *ns* : significant before a false discovery rate (FDR) correction at the nominal level of 5%, ne: not estimated.

ǂ All traits were measured quantitatively with the exception of survival which is a binary trait. § Each trait was modeled separately using a mixed model. Model random terms were tested with likelihood ratio tests (LRT) of models with and without these effects. A correction for the number of tests was performed for each modeled effect (*i.e.* per column) to control the FDR at a nominal level of 5%.

Table S2 | Broad-sense heritability values (H^2) of the 174 eco-phenotypes scored across six *in situ* 'soil x competition' micro-habitats. *P*: bold values indicate significant broad-sense heritability estimates after a false discovery rate (FDR) correction at the nominal level of 5%.

Ecophenotype	H^2	*P*
BT_A_wo_P	0.868	**0.00E+00**
BT_A_w_P	0.847	**0.00E+00**
BT_B_wo_P	0.864	**0.00E+00**
BT_B_w_P	0.827	**0.00E+00**
BT_C_wo_P	0.864	**0.00E+00**
BT_C_w_P	0.843	**0.00E+00**
DIAM_A_wo_P	0.084	6.27E-01
DIAM_A_w_P	0.480	**4.50E-08**
DIAM_B_wo_P	0.329	**1.01E-03**
DIAM_B_w_P	0.303	**4.90E-03**
DIAM_C_wo_P	0.174	1.38E-01
DIAM_C_w_P	0.470	**5.40E-08**
FITBB_A_wo_P	0.261	2.64E-01
FITBB_A_w_P	0.005	1.00E+00
FITBB_B_wo_P	0.256	1.69E-01
FITBB_B_w_P	0.260	1.60E-01
FITBB_C_wo_P	0.323	8.22E-02
FITBB_C_w_P	0.451	**2.21E-03**
FITPB_A_wo_P	0.442	**3.14E-04**
FITPB_A_w_P	0.398	**3.84E-04**
FITPB_B_wo_P	0.341	**2.06E-03**
FITPB_B_w_P	0.174	1.93E-01
FITPB_C_wo_P	0.490	**1.05E-06**
FITPB_C_w_P	0.602	**1.51E-12**
FITSTEM_A_wo_P	0.626	**2.07E-11**
FITSTEM_A_w_P	0.644	**3.58E-16**
FITSTEM_B_wo_P	0.495	**3.65E-08**
FITSTEM_B_w_P	0.419	**2.80E-05**
FITSTEM_C_wo_P	0.709	**0.00E+00**
FITSTEM_C_w_P	0.716	**0.00E+00**
FITTOT_A_wo_P	0.230	7.29E-02
FITTOT_A_w_P	0.399	**6.65E-04**
FITTOT_B_wo_P	0.170	1.86E-01
FITTOT_B_w_P	0.202	**2.89E-02**
FITTOT_C_wo_P	0.418	**2.63E-02**
FITTOT_C_w_P	0.566	**8.90E-07**
FRUITBB_A_wo_P	0.256	5.33E-02
FRUITBB_A_w_P	0.342	**6.65E-04**
FRUITBB_B_wo_P	0.334	**2.20E-03**
FRUITBB_B_w_P	0.400	**5.20E-05**
FRUITBB_C_wo_P	0.303	**6.05E-03**
FRUITBB_C_w_P	0.635	**0.00E+00**
FRUITPB_A_wo_P	0.326	**8.95E-03**
FRUITPB_A_w_P	0.401	**4.18E-05**
FRUITPB_B_wo_P	0.252	**2.21E-02**
FRUITPB_B_w_P	0.147	2.45E-01
FRUITPB_C_wo_P	0.447	**2.94E-06**
FRUITPB_C_w_P	0.591	**7.08E-15**

Ecophenotype	H^2	*P*
FRUITSTEM_A_wo_P	0.515	**1.53E-07**
FRUITSTEM_A_w_P	0.601	**2.61E-14**
FRUITSTEM_B_wo_P	0.370	**3.27E-04**
FRUITSTEM_B_w_P	0.323	**3.51E-03**
FRUITSTEM_C_wo_P	0.676	**0.00E+00**
FRUITSTEM_C_w_P	0.749	**0.00E+00**
GERM_A_wo_P	0.827	**0.00E+00**
GERM_A_w_P	0.796	**0.00E+00**
GERM_B_wo_P	0.781	**0.00E+00**
GERM_B_w_P	0.773	**0.00E+00**
GERM_C_wo_P	0.659	**0.00E+00**
GERM_C_w_P	0.738	**0.00E+00**
H1F_A_wo_P	0.536	**6.72E-09**
H1F_A_w_P	0.567	**4.27E-12**
H1F_B_wo_P	0.705	**0.00E+00**
H1F_B_w_P	0.541	**2.66E-09**
H1F_C_wo_P	0.574	**6.09E-12**
H1F_C_w_P	0.695	**0.00E+00**
HD_A_wo_P	0.518	**1.10E-07**
HD_A_w_P	0.673	**0.00E+00**
HD_B_wo_P	0.728	**0.00E+00**
HD_B_w_P	0.666	**0.00E+00**
HD_C_wo_P	0.529	**1.80E-09**
HD_C_w_P	0.714	**0.00E+00**
HMAX_A_wo_P	0.607	**2.19E-12**
HMAX_A_w_P	0.627	**0.00E+00**
HMAX_B_wo_P	0.625	**0.00E+00**
HMAX_B_w_P	0.574	**1.30E-11**
HMAX_C_wo_P	0.725	**0.00E+00**
HMAX_C_w_P	0.720	**0.00E+00**
HSTEM_A_wo_P	0.615	**2.30E-12**
HSTEM_A_w_P	0.640	**0.00E+00**
HSTEM_B_wo_P	0.620	**0.00E+00**
HSTEM_B_w_P	0.614	**1.25E-14**
HSTEM_C_wo_P	0.748	**0.00E+00**
HSTEM_C_w_P	0.761	**0.00E+00**
INT_A_wo_P	0.755	**0.00E+00**
INT_A_w_P	0.641	**0.00E+00**
INT_B_wo_P	0.771	**0.00E+00**
INT_B_w_P	0.623	**0.00E+00**
INT_C_wo_P	0.720	**0.00E+00**
INT_C_w_P	0.513	**4.52E-10**
INTERNOD_A_wo_P	0.026	1.00E+00
INTERNOD_A_w_P	0.103	5.66E-01
INTERNOD_B_wo_P	0.595	**1.73E-14**
INTERNOD_B_w_P	0.472	**9.37E-07**
INTERNOD_C_wo_P	0.105	4.71E-01
INTERNOD_C_w_P	0.160	1.00E+00

Table S2 (continued)

Ecophenotype	H^2	P
RAMBB_A_wo_P	0.316	**1.19E-02**
RAMBB_A_w_P	0.343	**6.65E-04**
RAMBB_B_wo_P	0.464	**9.28E-07**
RAMBB_B_w_P	0.388	**1.73E-04**
RAMBB_C_wo_P	0.344	**1.25E-03**
RAMBB_C_w_P	0.669	**0.00E+00**
RAMPB_WOF_A_wo_P	0.314	**8.95E-03**
RAMPB_WOF_A_w_P	0.555	**1.17E-11**
RAMPB_WOF_B_wo_P	0.182	1.54E-01
RAMPB_WOF_B_w_P	0.305	**5.83E-03**
RAMPB_WOF_C_wo_P	0.152	2.14E-01
RAMPB_WOF_C_w_P	0.405	**1.95E-05**
RAMPB_WF_A_wo_P	0.493	**8.70E-07**
RAMPB_WF_A_w_P	0.438	**2.94E-06**
RAMPB_WF_B_wo_P	0.484	**1.18E-07**
RAMPB_WF_B_w_P	0.398	**2.14E-04**
RAMPB_WF_C_wo_P	0.349	**9.08E-04**
RAMPB_WF_C_w_P	0.497	**1.09E-08**
RBB_A_wo_P	0.398	1.95E-01
RBB_A_w_P	0.440	1.05E-01
RBB_B_wo_P	0.520	**3.27E-04**
RBB_B_w_P	0.607	**1.90E-06**
RBB_C_wo_P	0.299	1.60E-01
RBB_C_w_P	0.664	**3.17E-05**
RP_A_wo_P	0.801	**0.00E+00**
RP_A_w_P	0.827	**0.00E+00**
RP_B_wo_P	0.798	**0.00E+00**
RP_B_w_P	0.742	**0.00E+00**
RP_C_wo_P	0.842	**0.00E+00**
RP_C_w_P	0.817	**0.00E+00**
RPB_A_wo_P	0.403	**2.81E-03**
RPB_A_w_P	0.345	**2.23E-03**
RPB_B_wo_P	0.555	**4.52E-10**
RPB_B_w_P	0.430	**1.65E-04**
RPB_C_wo_P	0.255	**1.56E-02**
RPB_C_w_P	0.505	**1.09E-08**
RSTEM_A_wo_P	0.253	7.42E-02
RSTEM_A_w_P	0.369	**1.43E-03**
RSTEM_B_wo_P	0.355	**5.97E-03**
RSTEM_B_w_P	0.293	**1.61E-02**
RSTEM_C_wo_P	0.074	1.00E+00
RSTEM_C_w_P	0.405	**3.47E-04**
SILBB_A_wo_P	0.677	**2.55E-06**
SILBB_A_w_P	0.681	**3.64E-04**
SILBB_B_wo_P	0.803	**0.00E+00**
SILBB_B_w_P	0.718	**2.63E-07**
SILBB_C_wo_P	0.713	**1.40E-11**
SILBB_C_w_P	0.729	**2.45E-10**

Ecophenotype	H^2	P
SILPB_A_wo_P	0.638	**3.52E-12**
SILPB_A_w_P	0.604	**9.79E-12**
SILPB_B_wo_P	0.779	**0.00E+00**
SILPB_B_w_P	0.702	**0.00E+00**
SILPB_C_wo_P	0.635	**8.41E-14**
SILPB_C_w_P	0.725	**0.00E+00**
SILSTEM_A_wo_P	0.774	**0.00E+00**
SILSTEM_A_w_P	0.781	**0.00E+00**
SILSTEM_B_wo_P	0.797	**0.00E+00**
SILSTEM_B_w_P	0.797	**0.00E+00**
SILSTEM_C_wo_P	0.743	**0.00E+00**
SILSTEM_C_w_P	0.755	**0.00E+00**
SURVIVAL_A_wo_P	0.022	8.72E-01
SURVIVAL_A_w_P	0.000	1.00E+00
SURVIVAL_B_wo_P	0.135	2.02E-01
SURVIVAL_B_w_P	0.113	2.90E-01
SURVIVAL_C_wo_P	0.000	1.00E+00
SURVIVAL_C_w_P	0.227	**2.29E-02**
TOTB_A_wo_P	0.359	**2.53E-03**
TOTB_A_w_P	0.216	6.31E-02
TOTB_B_wo_P	0.372	**4.52E-04**
TOTB_B_w_P	0.175	1.75E-01
TOTB_C_wo_P	0.254	**2.65E-02**
TOTB_C_w_P	0.542	**6.87E-11**
TOTPB_A_wo_P	0.565	**7.37E-10**
TOTPB_A_w_P	0.498	**3.19E-08**
TOTPB_B_wo_P	0.668	**0.00E+00**
TOTPB_B_w_P	0.580	**1.30E-11**
TOTPB_C_wo_P	0.525	**6.60E-10**
TOTPB_C_w_P	0.618	**4.60E-16**

Table S3 | Manhattan distance: scaling relationships between total phenotypic effect size of SNPs with the highest association and the effective number of eco-phenotypes (N_{eff}). The pleiotropic scaling relationship between the total effect size and the effective number of eco-phenotypes was calculated as $T_M = c^* {N_{eff}}^d$, with T_M corresponding to the Manhattan distance and calculated as $T_M = \sum_{i=1}^{n} |A_i|$, where n is the degree of pleiotropy and A_i is the standardized allelic effect. To avoid pseudo-replication due to the presence of several top SNPs in a given LD block, the pleiotropic scaling was also calculated for each threshold number of top SNPs and each threshold of significance, (i) by considering the mean value of the total effect size and N_{eff} per LD block containing top SNPs ('Mean per block' column) and (ii) by randomly sampling one top SNP per LD block (this step was repeated 1,000 times) ('Random' column).

T_M	Threshold	Total SNPs	Unique SNPs	% pleiotropic SNPs	All unique SNPs		Mean per block		Random	
					c	d	c	d	c	d
number of top SNPs	50 SNPs	7200	5728	16.69	0.317	1.255	0.317	1.3	0.32 (0.315 - 0.324)	1.268 (1.224 - 1.323)
	100 SNPs	14400	11100	19.05	0.308	1.242	0.309	1.275	0.311 (0.307 - 0.316)	1.253 (1.214 - 1.294)
	200 SNPs	28800	21268	21.86	0.294	1.228	0.295	1.255	0.296 (0.292 - 0.3)	1.243 (1.208 - 1.274)
	300 SNPs	43200	30854	24.4	0.286	1.215	0.289	1.223	0.289 (0.285 - 0.293)	1.217 (1.187 - 1.249)
	500 SNPs	72000	48851	27.64	0.277	1.19	0.283	1.178	0.282 (0.278 - 0.287)	1.181 (1.152 - 1.204)
$-\log_{10} p$-value	> 6	538	424	21.46	0.433	1.545	0.423	1.799	0.425 (0.416 - 0.438)	1.736 (1.503 - 1.92)
	> 5	3165	2457	17.91	0.366	1.446	0.361	1.51	0.362 (0.35 - 0.372)	1.49 (1.382 - 1.637)
	> 4	22822	16720	22.06	0.319	1.232	0.318	1.267	0.32 (0.314 - 0.326)	1.241 (1.197 - 1.293)

Table S4 | Euclidean distance: scaling relationships between total phenotypic effect size of SNPs with the highest association and the effective number of eco-phenotypes (N_{eff}). The pleiotropic scaling relationship between the total effect size and the effective number of eco-phenotypes was calculated as $T_E = a^* Neff^b$, with T_E corresponding to the Euclidean distance and calculated as $T_E = \sqrt{\sum_{i=1}^{n} A_i^2}$, where n is the degree of pleiotropy and A_i is the standardized allelic effect. To avoid pseudo-replication due to the presence of several top SNPs in a given LD block, the pleiotropic scaling was also calculated for each threshold number of top SNPs and each threshold of significance, (i) by considering the mean value of the total effect size and N_{eff} per LD block containing top SNPs ('Mean per block' column) and (ii) by randomly sampling one top SNP per LD block (this step was repeated 1,000 times) ('Random' column).

T_E	Threshold	Total SNPs	Unique SNPs	% pleiotropic SNPs	All unique SNPs		Mean per block		Random	
					a	b	a	b	a	b
number of top SNPs	50	7200	5728	16.69	0.292	0.739	0.295	0.766	0.296 (0.294 - 0.298)	0.757 (0.718 - 0.803)
	100	14400	11100	19.05	0.28	0.743	0.283	0.771	0.284 (0.282 - 0.286)	0.764 (0.729 - 0.797)
	200	28800	21268	21.86	0.267	0.726	0.268	0.751	0.269 (0.267 - 0.271)	0.747 (0.722 - 0.772)
	300	43200	30854	24.4	0.26	0.712	0.26	0.728	0.261 (0.259 - 0.263)	0.73 (0.705 - 0.755)
	500	72000	48851	27.64	0.25	0.692	0.25	0.697	0.252 (0.25 - 0.253)	0.705 (0.682 - 0.725)
$-\log_{10} p$-value	> 6	538	424	21.46	0.398	0.919	0.393	1.158	0.395 (0.39 - 0.401)	1.104 (0.884 - 1.285)
	> 5	3165	2457	17.91	0.34	0.838	0.335	0.917	0.335 (0.331 - 0.339)	0.906 (0.835 - 0.99)
	> 4	22822	16720	22.06	0.287	0.727	0.288	0.743	0.29 (0.288 - 0.292)	0.74 (0.706 - 0.776)

Table S5 | List of candidate genes associated with 11 or more evolved eco-phenotypes.

Atg number	no eco-phenotypes	Locus name	Molecular function
AT4G01820	17	ABCB3	member of MDR subfamily
AT4G01830	11	PGP5	P-glycoprotein 5 (PGP5)
AT4G14660	12	NRPE7	Non-catalytic subunit specific to DNA-directed RNA polymerase V
AT4G18350	12	NCED2	Encodes 9-*cis* -epoxycarotenoid dioxygenase, a key enzyme in the biosynthesis of abscisic acid.
AT4G19960	24	AtKUP/HAK/KT9	Encodes a potassium ion transmembrane transporter.
AT4G20325	12		unknown
AT4G20330	11		Transcription initiation factor TFIIE, beta subunit
AT4G20340	13		Transcription factor TFIIE, alpha subunit
AT4G20350	18		oxidoreductases
AT4G20362	15	SORF6	Potential natural antisense gene, locus overlaps with AT4G20360
AT4G20370	11	TSF	Encodes a floral inducer that is a homolog of FT.
AT4G24520	12	ATR1	Encodes a cyp450 reductase likely to be involved in phenylpropanoid metabolism.
AT5G12430	14	TPR16	Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins
AT5G43430	13	ETFBETA	Encodes the electron transfer flavoprotein ETF beta, a putative subunit of the mitochondrial electron transfer flavoprotein complex

Table S6 | Enrichment of biological process in the 0.1% tail of the F_{ST} values.

Biological process	Enrichment	*P* value	Atg number	Locus name	Molecular function	Associated eco-phenotypes[1]
vernalization response	22	**	AT5G10140	*FLC*	MADS-box protein	
			AT4G16845	*VRN2*	nuclear-localized zinc finger protein	H1F_B_w_P, RSTEM_B_wo_P, SURVIVAL_C_w_P, DIAM_B_wo_P, H1F_C_wo_P, SILBB_B_w_P, FITTOT_C_wo_P
regulation of circadian rhythm	21	**	AT5G10140	*FLC*	MADS-box protein	
response to temperature stimulus	21	**	AT5G10140	*FLC*	MADS-box protein	
negative regulation of flower development	21	*	AT5G10140	*FLC*	MADS-box protein	
regulation of cell shape	17	*	AT3G59100	*GLUCAN SYNTHASE-LIKE 11*	protein similar to callose synthase	
			AT4G03550	*POWDERY MILDEW RESISTANT 4*	callose synthase	FRUITSTEM_C_w_P, SILPB_A_wo_P
beta-D-glucan biosynthetic process	17	*	AT3G59100	*GLUCAN SYNTHASE-LIKE 11*	protein similar to callose synthase	
			AT4G03550	*POWDERY MILDEW RESISTANT 4*	callose synthase	FRUITSTEM_C_w_P, SILPB_A_wo_P
pollen tube development	15	*	AT4G05450	*MFDX1*	mitochondrial ferredoxin 1	FRUITSTEM_C_w_P, SILPB_A_wo_P
electron transport chain	15	*	AT4G05450	*MFDX1*	mitochondrial ferredoxin 1	FRUITSTEM_C_w_P, SILPB_A_wo_P
polar nucleus fusion	14	*	AT4G05440	*EMBRYO SAC DEVELOPMENT ARREST 35*	unknown	FRUITSTEM_C_w_P, SILPB_A_wo_P
			AT5G42020	*BIP2*	luminal binding protein	DIAM_C_w_P, TOT_B_C_w_P, RAMPB_WF_C_w_P
stamen development	14	*	AT4G03190	*AFB1*	F box protein belonging to the TIR1 subfamily	
			AT5G41700	*UBIQUITIN CONJUGATING ENZYME 8*	one of the polypeptides that constitute the ubiquitin-conjugating enzyme E2	FITTOT_C_w_P, FRUITPB_C_w_P, RSTEM_B_w_P, SILPB_C_w_P, INT_B_wo_P, SILSTEM_B_wo_P, RAMPB_WF_C_w_P
defense response by callose deposition in cell wall	14	*	AT4G03550	*POWDERY MILDEW RESISTANT 4*	callose synthase	FRUITSTEM_C_w_P, SILPB_A_wo_P
salicylic acid mediated signaling pathway	14	*	AT4G03550	*POWDERY MILDEW RESISTANT 4*	callose synthase	FRUITSTEM_C_w_P, SILPB_A_wo_P
defense response signaling pathway, resistance gene-dependent	14	*	AT4G03550	*POWDERY MILDEW RESISTANT 4*	callose synthase	FRUITSTEM_C_w_P, SILPB_A_wo_P
cell cycle arrest	13	*	AT4G05440	*EMBRYO SAC DEVELOPMENT ARREST 35*	unknown	FRUITSTEM_C_w_P, SILPB_A_wo_P
calcium-mediated signaling	12	*	AT4G03560	*TPC1*	depolarization-activated Ca(2+) channel	
trehalose biosynthetic process	12	*	AT5G10100	*TPPI*	haloacid dehalogenase-like hydrolase (HAD) superfamily protein	FITPB_A_wo_P, FRUITPB_A_wo_P
calcium ion transmembrane transport	12	*	AT4G03560	*TPC1*	depolarization-activated Ca(2+) channel	
calcium ion transport	12	*	AT4G03560	*TPC1*	depolarization-activated Ca(2+) channel	
regulation of salicylic acid mediated signaling pathw	10	*	AT4G03440		Ankyrin repeat family protein	FRUITSTEM_C_w_P
			AT4G03460		Ankyrin repeat family protein	GERM_A_wo_P, SILPB_B_w_P, SILPB_A_wo_P, SILSTEM_B_wo_P SILPB_B_w_P, SILPB_A_wo_P,
			AT4G03470		Ankyrin repeat family protein	SILSTEM_B_wo_P
			AT4G03500		Ankyrin repeat family protein	FRUITSTEM_C_w_P
cellular response to salicylic acid stimulus	10	*	AT4G03440		Ankyrin repeat family protein	FRUITSTEM_C_w_P
			AT4G03450		Ankyrin repeat family protein	FRUITSTEM_C_w_P, GERM_A_wo_P
			AT4G03460		Ankyrin repeat family protein	GERM_A_wo_P, SILPB_B_w_P, SILPB_A_wo_P, SILSTEM_B_wo_P SILPB_B_w_P, SILPB_A_wo_P,
			AT4G03470		Ankyrin repeat family protein	SILSTEM_B_wo_P
			AT4G03500		Ankyrin repeat family protein	FRUITSTEM_C_w_P
photosynthetic electron transport chain	10	*	AT4G03280	*PGR1*	Encodes the Rieske FeS center of cytochrome b6f complex	FRUITSTEM_C_w_P
developmental growth	7	*	AT4G03190	*AFB1*	F box protein belonging to the TIR1 subfamily	
pollen maturation	7	*	AT4G03190	*AFB1*	F box protein belonging to the TIR1 subfamily	
regulation of auxin mediated signaling pathway	5	*	AT3G59060	*PIF5*	novel Myc-related bHLH transcription factor	

*0.05 > *P* > 0.01, **0.01 > *P* > 0.001. The significance of enrichment was tested against a null distribution using 10,000 permutations.

[1] The letters A, B and C stand for the three types of soil. 'wo_P' and 'w_P' correspond to the absence and presence of *P. annua*, respectively.

Conclusion

C. Conclusion

Dans ce chapitre, le principal objectif était d'étudier la dynamique adaptative d'une population locale d'*A. thaliana* face au réchauffement climatique. En combinant une expérience de résurrection dans des conditions écologiquement réalistes avec des analyses de GWA mapping et un scan génomique de traces de sélection temporelle, nous avons pu mettre en évidence qu'une population locale d'*A. thaliana* pouvait rapidement atteindre un nouvel optimum phénotypique *via* une architecture génétique originale combinant (i) de rares QTLs avec des degrés de pléiotropie intermédiaires et fortement sélectionnés et (ii) de très nombreux QTLs spécifiques d'un micro-habitat et faiblement sélectionnés. Une telle architecture génétique aurait pu difficilement être mise en évidence en réalisant l'expérience de phénotypage en conditions contrôlée de serre ou bien en ne considérant qu'un seul micro-habitat au sein de la population TOU-A. Par ailleurs, il faut souligner que nous n'avons considéré dans cette étude qu'une faible fraction des micro-habitats rencontrés par *A. thaliana* dans la population TOU-A. Ceci est notamment le cas pour les interactions biotiques. En effet, nous avons seulement étudié les interactions avec *Poa annua*, mais de nombreuses autres espèces interagissent avec *A. thaliana* dans cette population. A partir d'une étude de GWA mapping réalisée à partir de 48 accessions de la population TOU-A, une précédente étude menée au sein de l'équipe a mis en évidence une architecture génétique très dépendante de l'identité de l'espèce compétitrice (Baron *et al.* 2015). Cette étude met donc en avant l'importance de considérer l'hétérogénéité des facteurs abiotiques et biotiques à une échelle micro-géographique, et par conséquent des interactions génotype x environnement, dans l'étude de la dynamique adaptative des populations naturelles.

Pour résumer, la très forte flexibilité de l'architecture génétique mise en évidence dans cette étude pourrait permettre d'expliquer les observations régulières du maintien de la variation génétique dans les populations naturelles malgré la présence de la sélection naturelle. Pour vérifier la généralité de nos résultats, la prochaine étape serait d'étendre les études de résurrection couplées à des analyses génomiques à d'autres populations d'*A. thaliana* et à d'autres espèces.

Conclusion générale

Conclusion générale

L'objectif principal de ma thèse était de comprendre l'architecture génétique de l'adaptation d'*A. thaliana* en considérant la complexité des agents sélectifs auxquels cette espèce est confrontée de manière simultanée dans les habitats naturels. Pour cela, j'ai abordé ce projet par des approches en génomique écologique très complémentaires et reposant sur diverses méthodes statistiques: (i) des analyses sparse Partial Least Square Regression (sPLSR) afin d'identifier les agents sélectifs potentiels, (ii) des analyses d'association génome-environnement (GEA) pour identifier les bases génétiques associées aux agents sélectifs potentiels, (iii) des analyses de GWA mapping afin d'identifier des régions génomiques associés à la variation phénotypique, et (iv) des scans génomiques pour identifier les traces de sélection aux niveaux spatial et temporel (court terme) (**Figure 1**).

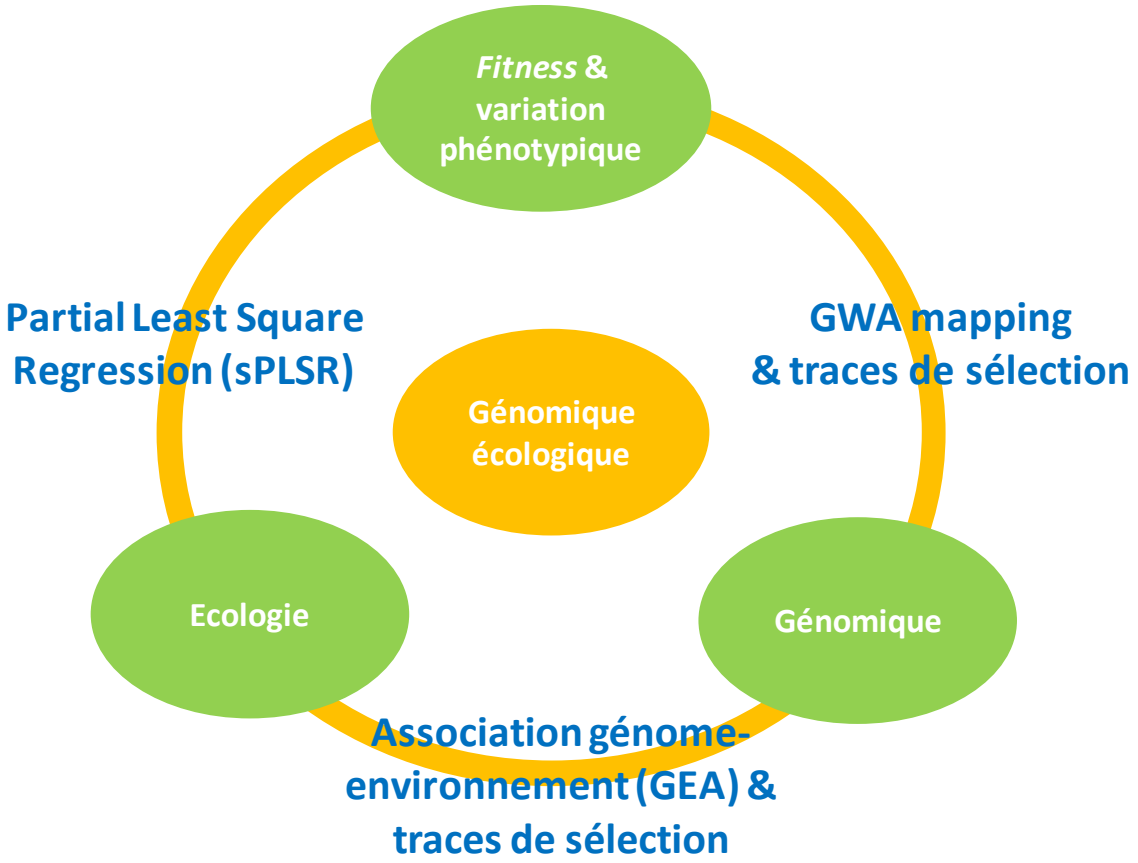


Figure 1 : Combinaison de 3 approches en génomique écologique permettant de comprendre l'architecture génétique de l'adaptation chez *A. thaliana*.

Importance de considérer les stratégies reproductives dans la définition de la fitness

A. thaliana étant une espèce avec un régime de reproduction largement autogame (Platt *et al.* 2010), la production de graines est considérée comme un bon proxy de la *fitness* dans de nombreuses études. Or dans le premier chapitre, en accord avec une précédente étude (Roux *et al.* 2016), nous avons mis en avant qu'une même production de graines pouvait être associée à différentes stratégies reproductives liées à l'acquisition des ressources, l'allocation des ressources entre les traits individuels de fécondité (nombre de fruits *vs* nombre de graines par fruit, nombre de graines produites par type de branche) et l'architecture de la plante déterminant le pattern de dispersion des graines. Par exemple, chez *A. thaliana*, il a été démontré expérimentalement que la distance moyenne de dispersion des graines correspond à la hauteur de la plante (Wender *et al.* 2005). Ainsi, la stratégie de dispersion des graines pourrait dépendre du grain spatial de l'environnement au sein d'une population. Si le grain spatial de l'environnement est grossier, une stratégie d'allocation des ressources dans les traits reproducteurs de la tige principale sera favorisée afin de disperser les graines de manière homogène et sur une plus longue distance (**Figure 2**). Un tel grain augmenterait donc la probabilité que la majorité des graines produites par une plante mère tombe dans un environnement identique à celui rencontré par la plante mère, tout en limitant la compétition entre plantes sœurs à la génération suivante.

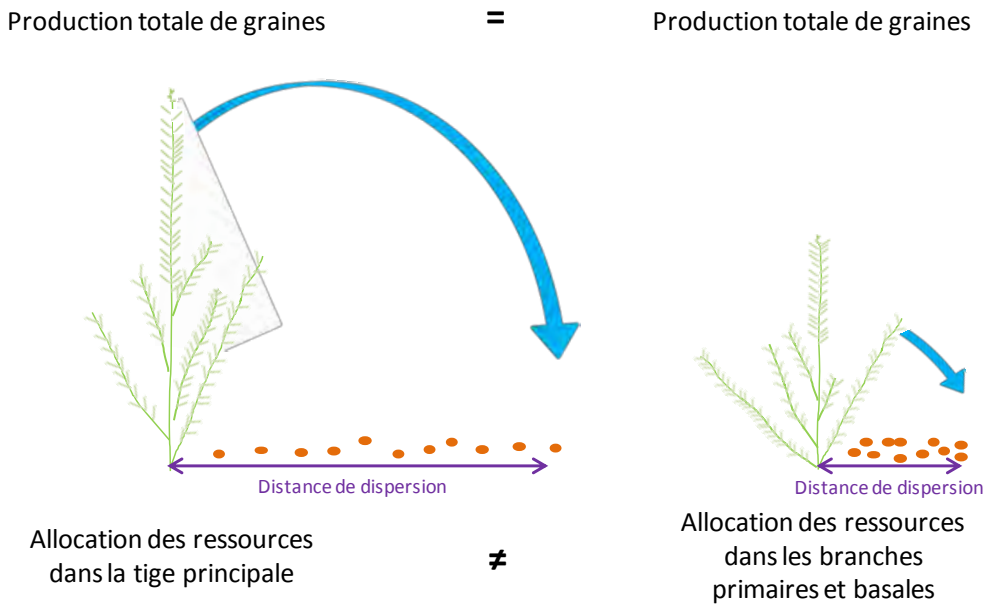


Figure 2: Différents patterns de dispersion des graines entre deux génotypes qui produisent la même quantité de graines, selon un grain spatial de l'environnement grossier (à gauche) ou fin (à droite).

A l'opposé, si le grain spatial de l'environnement est fin, une stratégie d'allocation des ressources vers les traits reproducteurs des branches primaires et des branches basales sera favorisée afin de disperser les graines sur une courte distance (**Figure 2**). Ces prédictions restent néanmoins basées sur des expériences réalisées dans des milieux ouverts (Wender *et al.* 2005). Il serait intéressant de compléter des études en tenant compte de la présence d'espèces compétitrices.

Pour comprendre comment *A. thaliana* peut s'adapter à un nouvel environnement, il semble donc important d'ajouter la notion de stratégies reproductives à la production totale de graines. L'importance de prendre en compte ces stratégies dans la définition de la *fitness* a tout d'abord été confirmée par l'étude des relations entre variation écologique et variation phénotypique dans le premier chapitre. Ainsi, nous avons mis en évidence que les variations de stratégie reproductive pouvaient être fortement associées à des variables écologiques, mais que l'identité de ces agents sélectifs potentiels était très dépendante de l'habitat considéré.

L'importance de tenir compte des stratégies reproductives dans la définition de la *fitness* a aussi été confirmée dans l'étude de résurrection du troisième chapitre. Lors de cette étude menée sur une population locale d'*A. thaliana*, nous avons vu qu'il était indispensable de considérer de nombreux traits phénotypiques dans la compréhension de l'évolution phénotypique d'une population. En effet, bien qu'aucun changement évolutif n'ait été observé pour la production totale de graines, une évolution génétique significative a été observée pour de nombreux autres traits phénotypiques liés aux stratégies reproductives. La flexibilité des traits d'histoire de vie et des traits reproducteurs individuels permettrait ainsi d'assurer une stabilité démographique de la population.

Prendre en compte le réalisme écologique des populations passe par considérer le maillage complexe des agents sélectifs

Les individus sont soumis simultanément à de nombreuses pressions de sélections abiotiques (climat, sol, etc.) et biotiques (communautés végétales et microbiennes, pathogènes, herbivore, etc.), dont les échelles spatiales de variation peuvent être très

contrastées. Cette diversité d'agents sélectifs agissant à différentes échelles spatiales crée non seulement une mosaïque écologique très complexe (**Figure 3**), mais peut aussi entraîner des conflits pour la plante quant à la stratégie à adopter pour y répondre. Par exemple, en présence du réchauffement climatique observé à une large échelle géographique, il est attendu une accélération de la phénologie chez les espèces végétales. Cependant, cette nouvel optimum phénologique peut se trouver en opposition avec des optimums locaux de tardivité de floraison déterminés par d'autres variables écologiques, comme une forte teneur en azote dans le sol où une longue période d'accumulation des ressources serait favorisée. Ainsi, il pourrait être difficile pour une population de répondre à des variables écologiques qui favorisent différentes stratégies phénotypiques, ralentissant ou empêchant alors l'évolution de cette population vers un nouvel optimum phénotypique. Les changements globaux (changements climatiques, utilisation des sols, espèces invasives, arrivées de nouvelles maladies, etc.) vont entraîner de nouvelles pressions de sélections sur les populations à différentes échelles spatiales, pouvant déstabiliser les combinaisons écologiques préexistantes. Il est donc essentiel de comprendre l'adaptation des populations en considérant un maximum d'agents sélectifs à une échelle spatiale réduite en complément des variables climatiques déjà très étudiées à de larges échelles géographiques (Walther *et al.* 2010).

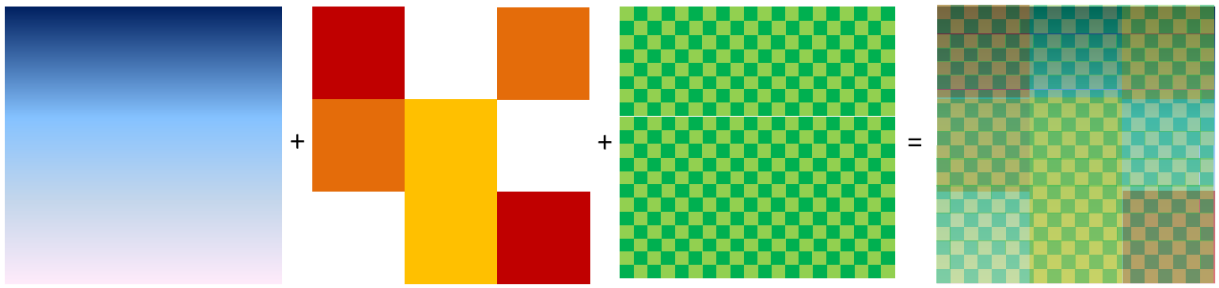


Figure 3 : Maillage individuel de trois agents sélectifs entraînant une mosaïque écologique complexe.

Afin d'étudier la complexité de la mosaïque écologique rencontrée par *A. thaliana* dans la région Midi-Pyrénées, 168 populations naturelles ont été caractérisées aussi bien d'un point de vue abiotique avec le climat et le sol, que biotique avec les communautés végétales et (de manière innovante) les communautés microbiennes. Alors que le climat

varie à une échelle géographique large, aucun pattern géographique évident n'a pu être mis en évidence pour les variables édaphiques et les variables biotiques, suggérant que les milieux où poussent *A. thaliana* sont très contrastés et non-prédictibles, même sur une courte distance géographique (i.e. quelques dizaines de mètres). L'observation d'une mosaïque écologique complexe à une fine échelle spatiale a été confirmée par la description écologique de la population locale TOU-A où les propriétés chimiques du sol peuvent être très variables sur une échelle de seulement quelques mètres. Comme nous avons pu l'observer, que ce soit à une échelle régionale ou au sein des habitats (chapitre 1) ou bien à l'intérieur d'une population locale (chapitre 3), les stratégies phénotypiques sous sélection vont fortement dépendre de la combinaison d'un ensemble de variables écologiques.

Par conséquent, pour comprendre l'architecture génétique de l'adaptation d'*A. thaliana*, il apparaît primordial d'identifier non plus seulement les bases génétiques spécifiques à un trait phénotypique donné ou une variable écologique donnée, mais aussi les bases génétiques associées aux stratégies phénotypiques ou à des combinaisons de variables écologiques.

Architecture génétique de l'adaptation chez A. thaliana

Pour identifier les bases génétiques de l'adaptation à des agents sélectifs, j'ai utilisé une approche de type GEA (chapitre 2). D'un autre coté, pour identifier les bases génétiques associées à des traits phénotypiques, j'ai utilisé une approche de GWA mapping (chapitre 3). Dans les deux cas, ces approches ont été couplées à des scans génomiques d'identification des traces de sélection le long du génome, permettant ainsi de valider le statut adaptatif des régions génomiques identifiées par GEA ou GWA mapping.

Dans un premier temps, j'ai identifié les bases génétiques de l'adaptation d'*A. thaliana* au climat et aux communautés végétales à une échelle régionale. J'ai ainsi pu montrer que les variations climatiques étaient des moteurs de la variation génomique adaptative d'*A. thaliana* quelque soit l'échelle géographique considérée (Hancock *et al*. 2011, Lasky *et al.* 2012). Cependant, comme la variation de la production totale de graines était plus fortement associée à la variation de descripteurs des communautés végétales qu'à

la variation climatique (chapitre 1), il était donc essentiel d'identifier les bases génétiques de l'adaptation aux communautés végétales dans la région Midi-Pyrénées. Comme pour le climat, les analyses combinées de GEA et de différenciation génétique entre populations ont révélé une forte adaptation locale d'*A. thaliana* aux communautés végétales. De plus, allant dans le sens des résultats obtenus dans le premier chapitre, les signatures d'adaptation locale étaient plus fortes pour les régions génomiques associées aux communautés végétales que pour les régions génomiques associées au climat (chapitre 2). Ce résultat met en avant l'importance de considérer les communautés végétales, et de manière générale les facteurs biotiques dans la compréhension de l'adaptation des plantes aux changements globaux. Par ailleurs, à ma connaissance, c'est la première étude basée sur une approche de type GEA montrant qu'une espèce végétale peut être localement adaptée à la fois spécifiquement à certaines espèces végétales mais également de manière générale à l'ensemble de la communauté végétale (diversité spécifique et composition d'espèces), soulignant l'importance de considérer les interactions diffuses au sein des communautés si l'on souhaite prédire le potentiel adaptatif d'une espèce (Roux & Bergelson 2016). L'importance de considérer les interactions plante-plante est soutenue (i) par les résultats obtenus dans notre étude de résurrection où l'évolution phénotypique d'*A. thaliana* était différente en présence ou en absence de *Poa annua* (chapitre 3), mais également (ii) par les travaux de Baron *et al.* (2015) montrant une architecture génétique de la réponse à la compétition qui était très dépendante de l'identité de l'espèce compétitrice.

En combinant une expérience de résurrection dans des conditions écologiquement réalistes avec des analyses de GWA mapping, nous avons identifié une architecture génétique originale sous-jacente à l'évolution d'une population locale vers un nouvel optimum phénotypique. En effet, l'architecture génétique des 29 traits phénotypiques mesurés était très variable entre les six micro-habitats testés, avec plus de 78% des SNPs les plus associés aux variations phénotypiques qui étaient spécifiques à un micro-habitat et aussi faiblement sous sélection. En complément de cette architecture, nous avons trouvé quelques rares SNPs avec un degré de pléiotropie intermédiaire (liées à 3-5 combinaisons 'trait phénotypique – micro-habitat') fortement sous sélection favorisant des combinaisons phénotypiques optimales. Cette architecture génétique flexible permettrait « d'allumer » rapidement des régions génomiques pour répondre à des conditions écologiques

spécifiques, tout en sélectionnant fortement des régions génomiques pléiotropes permettant d'atteindre rapidement un nouvel optimum phénotypique. Ainsi, cette architecture génétique combinant (i) de rares QTLs avec des degrés de pléiotropie intermédiaires et fortement sélectionnés et (ii) de très nombreux QTLs spécifiques d'un micro-habitat et faiblement sélectionnés, permettrait le maintien d'une diversité génétique dans les populations, et donc d'un potentiel adaptatif face à de futurs changements environnementaux.

Est-ce que les signatures de sélection dépendent aussi du niveau de degré de pléiotropie écologique ? Cela reste une question ouverte. A partir des résultats obtenus avec les analyses de type GEA, les SNPs les plus associés à une catégorie de variables écologiques étaient rarement les mêmes que les SNPs les plus associés à une autre catégorie de variables écologiques. Par exemple, seulement 0.15% des SNPs les plus corrélés à la variation des descripteurs des communautés végétales était aussi fortement corrélés aux variations climatiques dans la région Midi-Pyrénées. Cependant, au sein d'une catégorie de variables écologiques comme les descripteurs des communautés végétales, j'ai mis en évidence un excès significatif de gènes pléiotropes (i.e. de gènes associés à l'abondance de plusieurs espèces végétales). Une fois les analyses GEA réalisées sur l'ensemble des variables écologiques mesurées sur les 168 populations (climat, sol, descripteurs des communautés végétales et microbiote), il serait intéressant (i) d'étudier la distribution du degré de pléiotropie écologique en considérant soit toutes les variables écologiques soit uniquement les variables d'une catégorie donnée, et (ii) de tester si les patrons de différenciation génétique au niveau spatial diffèrent entre les degrés de pléiotropie écologique.

Perspectives

Bien que les études mises en place lors de ma thèse ont permis de commencer à comprendre l'adaptation d'*A. thaliana* dans la région Midi-Pyrénées et potentiellement l'évolution de populations naturelles face aux changements globaux, il est essentiel dans le futur d'approfondir ces études (i) en travaillant au sein des habitats, (ii) en considérant les échelles temporelles de variation des facteurs écologiques abordés durant la thèse, et (iii) en mettant en place des études comparatives avec d'autres espèces végétales.

(i) Travailler au sein des habitats

Nous avons vu que l'identité des agents sélectifs pouvait être très différente selon le type d'habitat considéré, notamment en ce qui concerne les stratégies reproductives. Dans un premier temps, les analyses de type GEA ont été réalisées à une échelle régionale. Il pourrait être maintenant intéressant de les réaliser également au sein de chaque habitat. De manière similaire aux résultats de GWA mapping obtenus entre micro-habitats, nous pouvons nous attendre à ce que les régions génomiques associées aux facteurs écologiques soient très différentes selon l'habitat considéré. On pourrait ainsi imaginer que dans les prairies où le climat apparaît comme une pression de sélection relativement forte par rapport aux autres catégories de variables écologiques, les signatures d'adaptation locale au climat soient beaucoup plus prononcées que dans les autres habitats où l'importance du climat est moindre.

(ii) Considérer les échelles temporelles de variation des facteurs écologiques

Pendant de ma thèse, je me suis certes intéressée à de nombreuses variables écologiques. Mais les variables climatiques sont représentées par des moyennes climatiques calculées sur une période de 10 ans et les variables biotiques ont été caractérisées à un instant donné du cycle de vie d'*A. thaliana* et sur une seule génération. Or, la fréquence et l'intensité des agents sélectifs peuvent être très variables au cours du cycle de vie d'un individu et/ou entre les générations (Siepielski *et al.* 2009). Une des perspectives majeures serait de considérer la variation temporelle des agents sélectifs dans l'étude de l'adaptation d'*A. thaliana*. Par exemple, dans le cadre du changement climatique, les variations intra-annuelle et inter-annuelle ont été décrites comme une pression de sélection majeure (Mearns *et al*. 1997). Comme nous le voyons sur la **Figure 4**, la température annuelle moyenne et les précipitations annuelles moyennes sont très variables d'une année sur l'autre dans la région Midi-Pyrénées. Il serait donc intéressant de réaliser des analyses de type GEA sur les coefficients de variation climatiques intra- et interannuelles. Nous pourrions alors tester (i) si les régions génomiques associées à la variation climatique temporelle présentent aussi des traces d'adaptation locale, et (ii) si des régions génomiques associées à la variation climatique temporelle sont communes à des régions génomiques associées à la variation climatique spatiale.

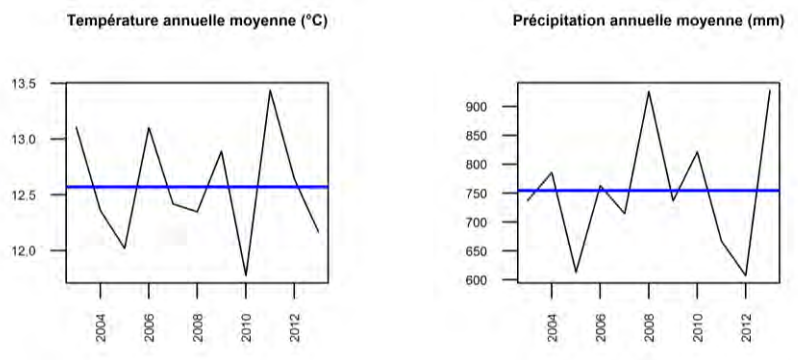


Figure 4: Variation climatique de 2003 à 2013 pour la température annuelle moyenne et les précipitations annuelles moyennes. Les courbes noires représentent la variation de ces variables observée au cours des années alors que la courbe bleue indique la moyenne de ces mêmes variables entre 2003 et 2013.

Pour les facteurs biotiques, une variation saisonnière des descripteurs des microbiotes a d'ores et déjà été observée au sein des 168 populations naturelles d'*A. thaliana* (Bartoli *et al.* en révision, Annexe 1). Par exemple, l'α-diversité de Shannon pour le microbiote bactérien et le pathobiote bactérien peut très vite changer entre le début du cycle de vie d'*A. thaliana* (i.e. automne) et sa sortie de l'hiver sous forme de rosette, et le degré de changement entre les deux saisons est très variable entre les populations (**Figure 5**). Il serait donc intéressant d'étudier la flexibilité de l'architecture génétique sous-jacente à la variation des descripteurs des microbiotes au cours du cycle de vie d'*A. thaliana*.

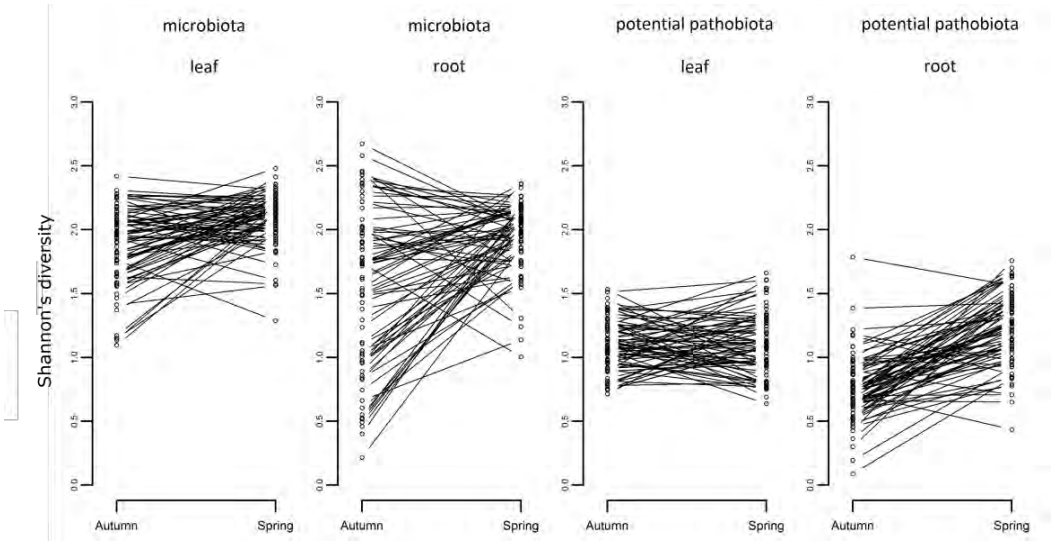


Figure 5: Variation entre les populations de l'évolution de l'α-diversité de Shannon entre l'automne et la fin de l'hiver, pour le microbiote bactérien et le pathobiote bactérien potentiel mesurés dans les feuilles et dans les racines. Chaque point et chaque ligne correspondent à une des populations naturelles d'*A.thaliana* localisées dans la région Midi-Pyrénées (Bartoli *et al.* en révision, Annexe 1).

Dans la région Midi-Pyrénées, j'ai aussi pu observer que la diversité et la composition des communautés végétales pouvaient être très variables au cours du cycle de vie d'*A. thaliana* (**Figure 6**).

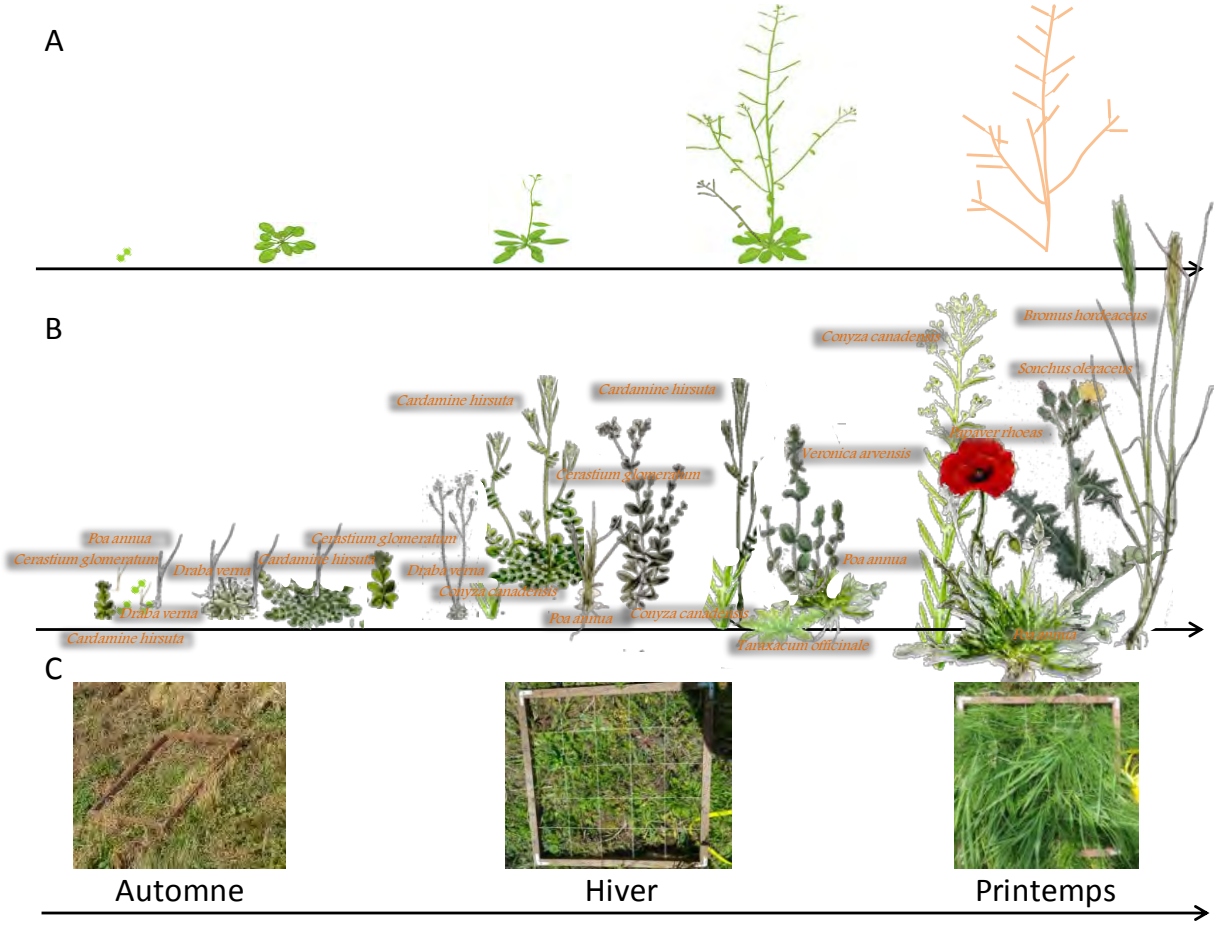


Figure 6: Dynamique des communautés végétales associées à *A. thaliana*. (A) Cycle de vie d'*A. thaliana* de l'automne au printemps. (B) Dynamique de certaines espèces pouvant être associées à *A. thaliana* au cours de son cycle de vie. (C) Exemple de la dynamique d'une communauté végétale associée à *A. thaliana* au cours de son cycle de vie (population de Montbrun-Bocage, Haute-Garonne).

Par exemple, les deux espèces de Brassicacées *Cardamine hirsuta* et *Draba verna*, germent en même temps qu'*A. thaliana* à l'automne. Cependant, leur cycle de vie est généralement plus rapide. Ainsi, alors que la présence de *Cardamine hirsuta* a été notée dans 118 populations à la fin de l'hiver, nous l'avons retrouvée uniquement dans 45 populations lors de notre caractérisation des communautés végétales au printemps. Suite à ce constat, j'ai réalisé des analyses de type GEA préliminaires pour tester si les régions génomiques associées à la présence de *C. hirsuta* étaient différentes entre l'hiver et le

printemps. Les profils d'association génome-environnement sont largement différents entre les deux saisons (**Figure 7**), soulignant la nécessité de caractériser les communautés végétales à différentes étapes du cycle de vie d'*A. thaliana* si l'on souhaite obtenir une vision plus complète de l'adaptation d'*A. thaliana* aux communautés végétales. En partant de ce constat, j'ai réalisé avec Baptiste Mayjonade (IE dans notre équipe) trois nouvelles sessions de terrain pour caractériser les communautés végétales d'une 60^{aine} de populations à l'automne 2015, à l'hiver 2016 et au printemps 2016. Cette caractérisation sur différentes saisons permettra ainsi de mieux comprendre l'architecture génétique d'*A. thaliana* associée à l'adaptation locale à différentes espèces mais aussi aux communautés elles-mêmes et à leur dynamique. Un stagiaire de Master 1 Thomas Dussarrat est actuellement en charge de l'identification des espèces *via* l'approche metabarcoding décrite dans le chapitre 2.

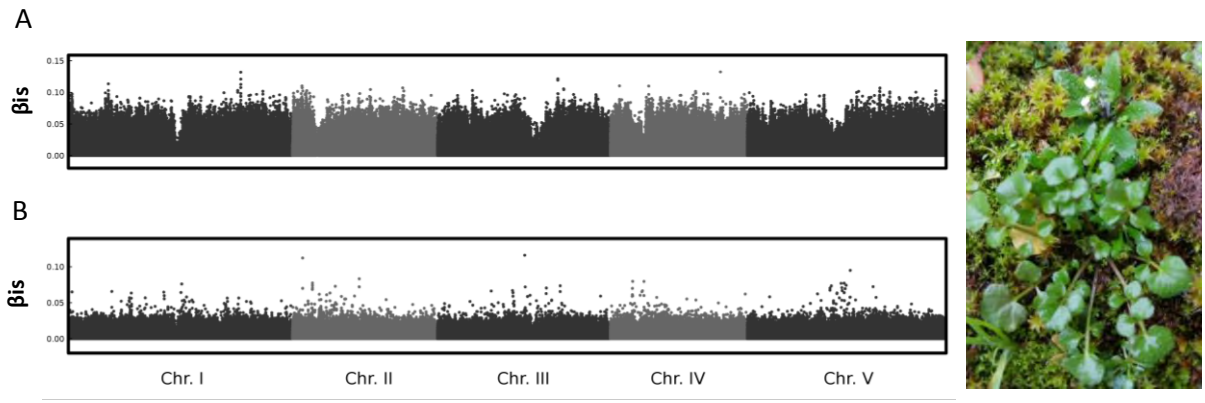


Figure 7: Analyse d'association génome-environnement (GEA) réalisée sur la présence-absence de *Cardamine hirsuta* (A) à la fin de l'hiver 2015 et (B) à la fin du printemps 2015.

(iii) Réaliser des études comparatives

Durant ma thèse, tous les résultats ont été obtenus sur l'espèce modèle *A. thaliana*. Afin de tester la généralité de nos conclusions, il serait intéressant de réaliser les mêmes études sur d'autres espèces végétales co-habitant avec *A. thaliana*. La description des communautés végétales associées aux 168 populations naturelles d'*A. thaliana* en est une première étape. Ainsi, avec Fabrice Roux, nous avons échantillonné environ 10 plantes de *C. hirsuta* dans les 118 populations naturelles d'*A. thaliana* où *C. hirsuta* était présente. Un tel

échantillonnage a aussi été effectué pour *D. verna* dans 54 populations à la fin de l'hiver 2017. Ces études comparatives avec *C. hirsuta* et *D. verna* permettront de répondre à plusieurs questions : Observe-t-on une architecture génétique de l'adaptation aussi flexible ? Les gènes associés aux variables écologiques sont-ils les mêmes entre les trois espèces ? En d'autres termes, l'évolution génétique adaptative est-elle prédictible ? La réponse à ces questions permettrait de débuter l'estimation du potentiel adaptatif des communautés végétales face aux changements globaux.

Bibliographie

Agrawal AA (2001) Phenotypic plasticity in the interactions and evolution of species. *Science* **294**: 321-326

Agren J, Oakley CG, McKay JK, Lovell JT, Schemske DW (2013) Genetic mapping of adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 21077–21082

Aguilar R, Ashworth L, Galetto L, Aizen MA (2006) Plant reproductive susceptibility to habitat fragmentation: review and synthesis through a meta-analysis. *Ecology Letters* **9**: 968–980

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JDG, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**: 627–631

Bandillo N, Raghavan C, Muyco PA, Sevilla MAL, Lobina IT, Dilla-Ermita CJ, Tung CW, McCouch S, Thomson M, Mauleon R, Singh RK, Gregorio G, Redoña E, Leung H (2013) Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* **6**: 1-15

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. *Functional Ecology* **29**: 1361–1370

Barret M, Briand M, Bonneau S, Préveaux A, Valière S, Bouchez O, Hunault G, Simoneau P, Jacques MA (2015) Emergence shapes the structure of the seed microbiota. *Applied and Environmental Microbiology* **81**: 1257-1266

Barthet MM, Hilu KW (2007) Expression of MatK: functional and evolutionary implications. *American Journal of Botany* **94**: 1402–1412

Bartoli C, Roux F (2017) Genome-Wide Association studies in plant pathosystems: towards an ecological genomics approach. Frontiers in Plant Science (**Sous presse**)

Bartoli C, Frachon L, Barret M, Rigal M, Zanchetta C, Bouchez O, Carrere S, Roux F (2017) In situ relationships between microbiota and potential pathobiota in *Arabidopsis thaliana*. e*Life* (**En revision**)

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. *The American naturalist* **189**: 463-473

Bazakos C, Hanemian M, Trontin C, Jiménez-Gomez JM, Loudet O (2017) New strategies and tools in quantitative genetics: How to go from the phenotype to the genotype? A*nnual Review of Plant Biology* **68**: 435-455

Beinart W, Middleton K (2004) Plant transfers in historical perspective: a review article. *Environment and History* **10**: 3-29

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

Bossdorf O, Shuja Z, Banta JA (2009) Genotype and maternal environment affect belowground interactions between *Arabidopsis thaliana* and its competitors. *Oikos* **118**: 15411551

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genetics* **6**: e1000940

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*. *Molecular ecology* **22**: 4222-4240

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Villeda HS, da Silva HS, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of Maize flowering time. *Science* **325**: 714-718

Bull CT, De Boer SH, Denny TP, Firrao G, Fischer-Le Saux M, Saddler GS, Scortichini M, Stead DE, Takikawa Y (2010) Comprehensive list of names of plant pathogenic bacteria, 1980-2007. *Journal of Plant Pathology* **92**: 551-592

Bull CT, De Boer SH, Denny TP, Firrao G, Fischer-Le Saux M, Saddler GS, Scortichini M, Stead DE, Takikawa Y (2012) List of new names of plant pathogenic bacteria (2008-2010). *Journal of Plant Pathology* **94**: 21-27

Bull CT, Coutinho TA, Denny TP, Firrao G, Fischer-Le Saux M, Li X, Saddler GS, Scortichini M, Stead DE, Takikawa Y (2014) List of new names of plant pathogenic bacteria (2011-2012). *Journal of Plant Pathology* **96**: 223-226

Carruthers J, Robin L, Hattingh JP, Kull CA, Rangan H, van Wilgen BW (2011) A native at home and abroad: the history, politics, ethics and aesthetics of acacias. *Diversity and Distributions* **17**: 810–821

Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology* **11**: 215–221

Chapin III FS, Zavaleta ES, Eviner VT, Naylor RL, Vitousek PM, Reynolds HL, Hooper DU, Lavorel S, SalaI OE, Hobbie SE, Mack MC, Díaz S (2000) Consequences of changing biodiversity. *Nature* **405**: 234-242

Charmantier A, McCleery RH, Cole R, Perrins C, Kruuk LEB, Sheldon BC (2008) Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science* **320**: 800-803

Chen IC, Hill JK, Ohlemüller R, Roy DB, Thomas CD (2011) Rapid range shifts of species associated with high levels of climate warming. *Science* **333**: 1024-1026

Chevin LM, Lande R, Mace GM (2010) Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS Biology* **8**: e1000357

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**: 1411–1423

de Villemereuil P, Gaggiotti E (2015) A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution* **6**: 1248–1258 doi:

De Villemereuil P, Frichot E, Bazin E, François O, Gaggiotti OE (2014) Genome scan methods against more complex models: when and how much should we trust them? *Molecular Ecology* **23**: 2006–2019

Debieu M, Tang C, Stich B, Sikosek T, Effgen S, Josephs E, Schmitt J, Nordborg M, Koornneef M, de Meaux J (2013) Co-variation between seed dormancy, growth rate and flowering time changes with latitude in *Arabidopsis thaliana*. *PLoS one* **8**: e61075

DeWitt TJ, Sih A, Wilson DS (1998) Costs and limits of phenotypic plasticity. *Trends in Ecology and Evolution* **13**: 77-81

Ellis EC, Antill EC, Kreft H (2012) All is not loss: Plant biodiversity in the Anthropocene. *PLoS one* **7**: e30535

Elton CS (1958) The ecology of invasions of plants and animals. *Methuen, London*.

Felde AV, Kapfer J, Grytnes JA (2012) Upward shift in elevational plant species ranges in Sikkilsdalen, central Norway. *Ecography* **35**: 922–932

Fischer MC, Rellstab C, Tedder A, Zoller S, Gugerli F, Shimizu KK, Holderegger R, Widmer A (2013) Population genomic footprints of selection and associations with climate in natural populations. *Molecular Ecology* **22**: 5594–5607

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. *Genetics* **180**: 977–993

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**: 86-89

Fracassetti M, Griffin PC, Willi Y (2015) Validation of pooled whole-genome re-sequencing in *Arabidopsis lyrata*. *PLoS one* **10**: e0140462

Frachon L, Mayjonade B, Bartoli C, Hautekeete , Roux F (2017) Adaptation to plant communities across the genome of *Arabidopsis thaliana*. (**En préparation**)

Frachon L, Bartoli C, Carrère S, Bouchez O, Chaubet A, Gautier M, Roby D, Roux F (2017) A genomic map of adaptation to local climate in *Arabidopsis thaliana. New Phytologist* (**Soumis**)

Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution* **30**: 1687–1699

Gaut B (2012) *Arabidopsis thaliana* as a model for the genetics of local adaptation. *Nature Genetics* **44**: 115-116

Gautier M (2015) Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics* **201**: 1555–1579

Gilman SE, Urban MC, Tewksbury J, Gilchrist GW, Holt RD (2010) A framework for community interactions under climate change. *Trends in Ecology and Evolution* **25**: 325-331

Gugerli F, Brandl R, Castagneyrol B, Franc A, Jactel H, Koelewijn HP, Martin F, Peter M, Pritsch K, Schroder H, Smulders MJM Kremer A, Ziegenhagen B, Evoltree Jera3 Contributors (2013) Community genetics in the time of next-generation molecular technologies. *Molecular Ecology* **22**: 3198–3207

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

Guo B, De Faveri J, Sotelo G, Nair A, Merilä J (2015) Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC Biology* **13**: 19

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

Hansen MM, Olivieri I, Waller DM, Nielsen EE, THE GeM WORKING GROUP (2012) Monitoring adaptive genetic responses to environmental change. *Molecular ecology* **21**: 1311–1329

Hautekèete NC, Frachon L, Luczak C, Toussaint B, Van Landuyt W, Van Rossum F, Piquot Y (2014) Habitat type shapes long-term plant biodiversity budgets in two densely populated regions in north-western Europe. *Diversity and Distributions* **21**: 631–642

Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, Poss ML, Reed LK, Storfer A, Whitlock MC (2016) Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American naturalist* **188**: 379–397

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**: 101-104

Hoffmann AA, Sgro CM (2011) Climate change and evolutionary adaptation. *Nature* **470**: 479-485

Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, Muliyati NW, Platt A, Sperone FG, Vilhjálmsson BJ, Nordborg M, Borevitz JO, Bergelson J (2012) Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nature Genetics* **44**: 212–216

Huang BE, George AW, Forrest KL, Kilian A, Hayden MJ, Morell MK, Cavanagh CR (2012) A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnology Journal* **10**: 826–839

Huard-Chauveau C, Perchepied L, Debieu M, Rivas S, Kroj T, Kars I, Bergelson J, Roux F, Roby D (2013) An atypical kinase under balancing selection confers broad-spectrum disease resistance in Arabidopsis. *PLoS Genetics* **9**: e1003766

Jackson ST, Sax DF (2010) Balancing biodiversity in a changing environment: extinction debt, immigration credit and species turnover. *Trends in Ecology and Evolution* **25**: 153-160

Jakob K, Goss EM, Araki H, Van T, Kreitman M, Bergelson J (2002) *Pseudomonas viridiflava* and *P. syringae*—natural pathogens of *Arabidopsis thaliana* . *The American Phytopathological Society* **15**: 1195-1203

Johnson CN (2002) Determinants of loss of mammal species during the Late Quaternary 'megafauna' extinctions: life history and ecology, but not body size. *Proceedings of the Royal Society B* **269**: 2221–2227

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong Sy, Freimer NB, Sabatti C, Eskin E (2010) Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* **42**: 348-354

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

Kawakatsu T, Huang SSC, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y, Chen H, Dubin M, Lee CR, Wang C, Bemm F, Becker C, O'Neil R, O'Malley RC, Quarless DX, The 1001 Genomes Consortium, Schork NJ, Weigel D, Nordborg M, Ecker JR (2016) Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* **166**: 492–505

Keller TE, Lasky JR, Yi SV (2016) The multivariate association between genome-wide DNAmethylation and climate across the range of *Arabidopsis thaliana*. *Molecular Ecology* **25**: 1823–1837

Kelly AE, Goulden ML (2008) Rapid shifts in plant distribution with recent climate change. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 11823–11826

Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annual Review of Plant Biology* **55**: 141–172

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genetics* **5**: e1000551

Lande R (2009) Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of Evolutionary Biology* **22**: 1435-1446

Lasky JR, Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH (2012) Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Molecular ecology* **21**: 5512–5529

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. *Science Advances* **1**: e1400218

López-García JM, Blain HA, Morales JI, Lorenzo C, Bañuls-Cardona S, Cuenca-Bescós G (2013) Small-mammal diversity in Spain during the late Pleistocene to early Holocene: Climate, landscape, and human impact. *Geology* **41**: 267-270

Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular ecology* **24**: 1031–1046

Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, Richardson MF, Anholt RRH, Barron M, Bess C, Blankenburg KP,

Carbone MA, Castellano D, Chaboub L, Duncan L, Harris Z, Javaid M, Jayaseelan JC, Jhangiani SN, Jordan KW, Lara F, Lawrence F, Lee SL, Librado P, Linheiro RS, Lyman RF, Mackey AJ, Munidasa M, Muzny DM, Nazareth L, Newsham I, Perales L, Pu LL, Qu C, Ramia M, Reid JG, Rollmann SM, Rozas J, Saada N, Turlapati L, Worley KC, Wu YQ, Yamamoto A, Zhu Y, Bergman CM, Thornton KR, Mittelman D, Gibbs RA (2012) The *Drosophila melanogaster* genetic reference panel. *Nature* **482**: 173-178

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular ecology* **19**: 3824–3835

Martin PS, Steadman DW (1999) Prehistoric Extinctions on Islands and Continents. *MacPhee RDE*. ed. Extinctions in Near Time: Causes, Contexts, and Consequences. New York, Kluwer Academic/Plenum: pages 17-55

Matesanz S, Valladares F (2014) Ecological and evolutionary responses of Mediterranean plants to global change. *Environmental and Experimental Botany* **103**: 53-67

Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S, Purugganan MD (2007) The extent of linkage disequilibrium in Rice (*Oryza sativa L.*). *Genetics* **177**: 2223–2232

Mearns LO, Rosenzweig C, Goldberg R (1997) Mean and variance change in climate scenarios: methods, agricultural applications, and measures of uncertainty. *Climatic Change* **35**: 367–396

Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M (1998) *Arabidopsis thaliana*: a model plant for genome analysis. *Science* **282**: 662-682

Meyerowitz EM, Somerville CR (2002) The Arabidopsis book.

Millennium Ecocystem Assessment (2005) Ecosystem an Human Well-being, synthesis. *Island Press, Washington D.C.*

Mitchell-Olds T, Schmitt J (2006) Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. *Nature* **441**: 947-952

Moran NA (1992) The evolutionary maintenance of alternative phenotypes. *The American naturalist* **139**: 971–989

Murren CJ, Auld JR, Callahan H, Ghalambor CK, Handelsman CA, Heskel MA, Kingsolver JG, Maclean HJ, Masel J, Maughan H, Pfennig DW, Relyea RA, Seiter S, Snell-Rood E, Steiner UK, Schlichting CD (2015) Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and plasticity. *Heredity* **115**: 293–301

Picó FX (2012) Demographic fate of *Arabidopsis thaliana* cohorts of autumn- and spring-germinated plants along an altitudinal gradient. *Journal of Ecology* **100**: 1009–1018

Pimm SL, Russell GJ, Gittleman JL, Brooks TM (1995) The future of biodiversity. *Science* **269**: 347-350

Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO (2014) The biodiversity of species and their rates of extinction, distribution, and protection. *Science* **344**: 1246752

Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, Agren J, Bossdorf O, Byers D, Donohue K Dunning M et al. (2010) The scale of population structure in *Arabidopsis thaliana*. *PLoS Genetics* **6**: e1000843

Pluess AR, Frank A, Heiri C, Lalagüe H, Vendramin GG, Oddou-Muratorio S (2016) Genome–environment association study suggests local adaptation to climate at the regional scale in *Fagus sylvatica*. *New Phytologist* **210**: 589–601

Price TD, Qvarnström A, Irwin DE (2003) The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the royal society B* **270**: 1433–1440

Ratcliffe FN (1959) The Rabbit in Australia. Biogeography and Ecology in Australia.

Reboud X, Le Corre V, Scarcelli N, Roux F, David JL, Bataillon T, Camilleri C, Brunel D, McKhann H (2004) Natural variation among accessions of *Arabidopsis thaliana*: beyond the flowering date, what morphological traits are relevant to study adaptation? *Plant adaptation: molecular genetics and ecology* Ottawa, CAN : NRC Research Press. Eds. Cronk QCB, Whitton J, Ree RH, Taylor IEP: pages 135-142

Rellstab C, Zoller S, Walthert L, Lesur I, Pluess AR, Graf R, Bodénès C, Sperisen C, Kremer A, Gugerli F (2016) Signatures of local adaptation in candidate genes of oaks (*Quercus spp.*) with respect to present and future climatic conditions. *Molecular ecology* **25**: 5907–5924

Richardson DM, Pysek P (2007) Elton, C.S. 1958: The ecology of invasions by animals and plants. London: Methuen. *Progress in Physical Geography* **31**: 659–666

Richardson JL, Urban MC, Bolnick DI, Skelly DK (2014) Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology and Evolution* **29**: 165-176

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. *Current topics in developmental biology* **119**: 111–156

Roux F, Touzet P, Cuguen J, Le Corre V (2016) How to be early flowering: an evolutionary perspective? *Trends in Plant Science* **11**: 375-381

Sala OE, Chapin III FS, Armesto JJ, Berlow E, Bloomfield J, Dirzo R, Huber-Sanwald E, Huenneke LF, Jackson RB, Kinzig A, Leemans R, Lodge DM, Mooney HA, Oesterheld M,

LeRoy Poff N, Sykes MT, Walker BH, Walker M, Wall DH (2000) Global biodiversity scenarios for the year 2100. *Science* **287**: 1770-1774

Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature reviews Genetics* **14**: 807-820

Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. *Nature reviews Genetics* **15**: 749-763

Shindo C, Bernasconi G, Hardtke CS (2007) Natural genetic variation in Arabidopsis: tools, traits and prospects for evolutionary ecology. *Annals of Botany* **99**: 1043–1054

Siepielski AM, DiBattista JD, Carlson SM (2009) It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecology Letters* **12**: 1261–1276

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

Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, Bennison C, Ball AD, Beckerman AP, Slate J (2010) Adaptation genomics: the next generation. *Trends in Ecology and Evolution* **25**: 705–712

Sultan SE, Spencer HG (2002) Metapopulation structure favors plasticity over local adaptation. *The American naturalist* **160**: 271-283

Sultan SE (2000) Phenotypic plasticity for plant development, function and life history. *Trends in Plant Science* **5**: 537-542

The 1001 Genomes Consortium (2016) 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166**: 481-491

Thuiller W (2007) Climate change and the ecologist. *Nature* **448**: 550-552

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics* **42**: 260-263

Udo N, Tarayre M, Atlan A (2017) Evolution of germination strategy in the invasive species *Ulex europaeus*. *Journal of Plant Ecology* **10**: 375–385

Ungerer MC, Johnson LC, Herman MA (2008) Ecological genomics: understanding gene and genome function in the natural environment. *Heredity* **100**: 178–183

van Kleunen M, Fischer M (2005) Constraints on the evolution of adaptive phenotypic plasticity in plants. *New Phytologist* **166**: 49–60

Van Rossum F, Triest L (2010) Pollen dispersal in an insect-pollinated wet meadow herb along an urban river. *Landscape and Urban Planning* **95**: 201–208

Vitousek PM, D'antonio CM, 1, Loope LL, Rejmánek M, Westbrooks R (1997) Introduced species: a significant component of human-caused global change. *New Zealand Journal of Ecology* **21**: 1-16

Wang Z, Liao BY, Zhang J (2010) Genomic patterns of pleiotropy and the evolution of complexity. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 18034–18039

Waters CN, Zalasiewicz J, Summerhayes C, Barnosky AD, Poirier C, Gałuszka A, Cearreta A, Edgeworth M, Ellis AC, Ellis M, Jeandel C, Leinfelder R, McNeill JR, deB. Richter D, Steffen W, Syvitski J, Vidas D, Wagreich M, Williams M, Zhisheng A, Grinevald J, Odada E, Oreskes N, Wolfe AP (2016) The Anthropocene is functionally and stratigraphically distinct from the Holocene. *Science* **351**: aad2622

Weber MG, Wagner CE, Best RJ, Harmon LJ, Matthews B (2017) Evolution in a community context: on integrating ecological interactions and macroevolution. *Trends in Ecology and Evolution* **32**: 291-304

Weinig C, Ungerer MC, Dorn LA, Kane NC, Toyonaga Y, Halldorsdottir SS, Mackay TFC, Purugganan MD, Schmitt J (2002) Novel loci control variation in reproductive timing in *Arabidopsis thaliana* in natural environments. *Genetics* **162**: 1875–1884

Wender NJ, Polisetty CR, Donohue K (2005) Density-dependent processes influencing the evolutionary dynamics of dispersal: a functional analysis of seed dispersal in *Arabidopsis thaliana* (Brassicaceae). *American Journal of Botany* **92**: 960–971

Wolverton S (2010) The North American Pleistocene overkill hypothesis and the re-wilding debate. *Diversity and Distributions* **16**: 874–876

Yan LJ, Liu J, Möller M, Zhang L, Zhang XM, Li DZ, Gao LM (2015) DNA barcoding of Rhododendron (Ericaceae), the largest Chinese plant genus in biodiversity hotspots of the Himalaya–Hengduan Mountains. *Molecular Ecology Resources* **15**: 932–944

Yoder JB, Stanton-Geddes J, Zhou P, Briskine R, Young ND, Tiffin P (2014) Genomic signature of adaptation to climate in *Medicago truncatula*. *Genetics* **196**: 1263–1275

Annexe 1

24 Hills Road, **P** 01223 855346
Cambridge **W** elifesciences.org
CB2 1JP **T** @elife_sciences

In situ relationships between microbiota and potential pathobiota in Arabidopsis thaliana

Tracking no: 09-01-2017-RA-eLife-24583

Fabrice Roux (CNRS), Claudia Bartoli (INRA), Léa Frachon (INRA), Matthieu Barret (INRA), Mylène Rigal (INRA), Catherine Zanchetta (INRA), Olivier Bouchez (INRA), and Sébastien Carrère (INRA)

Abstract:
A major challenge in plant pathology is to study the relationships between whole microbial communities inhabiting plants (microbiota) and potentially pathogenic microbes (pathobiota). Here, we investigated the *in situ* bacterial microbiota-potential pathobiota relationships in the native habitats of 163 *Arabidopsis thaliana* populations. Seasonal community succession revealed a strong dynamics of both microbiota and potential pathobiota in most populations. While up to 81% of microbiota variation was explained by differences among populations at a very small geographical scale, plant organs were the main source of pathobiota variation. In agreement with the diversity-invasion theoretical relationship, a poorly diverse pathobiota was associated with highly diverse microbiota. In addition, a large fraction of pathobiota composition was explained by season-specific combinations of few microbiota OTUs, suggesting a dynamics in the potential biomarkers controlling pathogens spread. Finally, strain isolation and *in planta* tests confirmed the pathogenicity of one the most abundant species composing the potential pathobiota.

Impact statement:

Competing interests: No competing interests declared

Author contributions:
Fabrice Roux: Conceptualization; Resources; Data curation; Formal analysis; Supervision; Funding acquisition; Validation; Investigation; Visualization; Methodology; Writing—original draft; Project administration; Writing—review and editing Claudia Bartoli: Conceptualization; Resources; Data curation; Formal analysis; Supervision; Investigation; Visualization; Methodology; Writing—original draft; Writing—review and editing Léa Frachon: Conceptualization; Resources; Formal analysis; Visualization; Methodology; Writing—original draft Matthieu Barret: Resources; Data curation; Formal analysis; Investigation; Methodology; Writing—original draft Mylène Rigal: Resources; Formal analysis; Methodology Catherine Zanchetta: Resources; Methodology Olivier Bouchez: Resources; Methodology Sébastien Carrère: Conceptualization; Resources; Data curation; Formal analysis; Methodology

Funding:
ANR-10-LABX-41: Fabrice Roux, Claudia Bartoli, Léa Frachon, Mylène Rigal, Sébastien Carrère; ANR-11-IDEX-0002-02: Fabrice Roux, Claudia Bartoli, Léa Frachon, Mylène Rigal, Sébastien Carrère; CLIMARES: Fabrice Roux, Claudia Bartoli, Léa Frachon; MEM, INRA metabar program: Matthieu Barret The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Datasets:
Datasets Generated: CLIMARES-BACTERIA-ROOTS_AND_LEAVES: Bartoli C, Frachon L, Barret M, Rigal M, Zanchetta C, Bouchez O, Carrère S, Roux F, 2017, https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP096011, SRP096011 Reporting Standards: N/A

Ethics:
Human Subjects: No Animal Subjects: No

Author Affiliation:
Fabrice Roux(LIPM,CNRS,France) Claudia Bartoli(LIPM,INRA,France) Léa Frachon(LIPM,INRA,France) Matthieu Barret(IRHS,INRA,France) Mylène Rigal(LIPM,INRA,France) Catherine Zanchetta(Get-PlaGe,INRA,France) Olivier Bouchez(Get-PlaGe,INRA,France) Sébastien Carrère(LIPM,INRA,France)

Dual-use research: No

Permissions: Have you reproduced or modified any part of an article that has been previously published or submitted to another journal? No

In situ relationships between microbiota and potential pathobiota in *Arabidopsis thaliana*

Claudia Bartoli[1¶], Léa Frachon[1¶], Matthieu Barret[2], Mylène Rigal[1], Catherine Zanchetta[3], Olivier Bouchez[3], Sébastien Carrere[1], Fabrice Roux[1*]

Affiliations:

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

[2] INRA, UMR1345 Institut de Recherches en Horticulture et Semences, SFR4207 QUASAV, -49071 Beaucouzé, France

[5] INRA, GeT-PlaGe, Genotoul, Castanet-Tolosan, France

[¶]These authors contributed equally to this work.

[*] To whom correspondence should be addressed E-mail: fabrice.roux@inra.fr

ABSTRACT

A major challenge in plant pathology is to study the relationships between whole microbial communities inhabiting plants (microbiota) and potentially pathogenic microbes (pathobiota). Here, we investigated the *in situ* bacterial microbiota-potential pathobiota relationships in the native habitats of 163 *Arabidopsis thaliana* populations. Seasonal community succession revealed a strong dynamics of both microbiota and potential pathobiota in most populations. While up to 81% of microbiota variation was explained by differences among populations at a very small geographical scale, plant organs were the main source of pathobiota variation. In agreement with the diversity-invasion theoretical relationship, a poorly diverse pathobiota was associated with highly diverse microbiota. In addition, a large fraction of pathobiota composition was explained by season-specific combinations of few microbiota OTUs, suggesting a dynamics in the potential biomarkers controlling pathogens spread. Finally, strain isolation and *in planta* tests confirmed the pathogenicity of one the most abundant species composing the potential pathobiota.

INTRODUCTION

In the last decade, a conspicuous effort has been made to characterize and understand the role of microorganisms associated with their hosts. Animals and plants are currently considered as holobionts associated with the commensal microorganisms that inhabit and evolve with them (Bordenstein & Theis 2015; Vandenkoornhuyse et al. 2015; Youle et al. 2013). Whether the commensal microbes protect against obligatory or opportunistic pathogens remains an open question. For example, human commensal gut microbiota can protect against the potential overgrowth of indigenous opportunistic pathobionts (defined as pathogenic species that inhabit at low bacterial population size) (Vayssier-Taussat et al. 2014) or pathogenic invaders (Kamada et al. 2013) by niche competition and/or induction of the host immune system (Belkaid & Hand 2014). It is also a common opinion in the medical field that some of the microorganisms colonizing the gut can easily switch in their pathogenic form when environmental conditions drastically change. In light of this, the fate of potential resident pathobionts strongly depends on the multipartite interactions among commensal and pathogenic microorganisms, the host and the environment (Baumler & Sperandio 2016, Vorholt 2012).

Theoretical models developed on human pathogens supported the notion that a highly diversified microbiota has a detrimental effect on pathogen spread into the host, a process called the diversity-invasion relationship (Mallon et al. 2015). Several non-exclusive mechanisms can be advanced to explain such a relationship. An increase in species diversity leads to a decrease of available niches in the host, thereby limiting infection by depleting resources available for pathogens. In addition, some bacterial taxa can be important for the maintenance of the whole microbial diversity. For example, the Lotka-Voltera model supports that some gut bacterial species can be considered as predators. Without these predators, dominant species can rapidly grow and outcompete most of the bacterial entities of the gut

52 microbiota. This decreasing in bacterial richness can in turn promote the pathogen's
53 occurrence within the resident microbial communities (Marino et al. 2014; Mosca et al. 2016).
54 However, few studies have been specifically designed to test this hypothesis (Mallon et al.
55 2015).

56 The recent advances in next generation sequencing coincided with a burst in the
57 medical field of studies aiming to understand how the whole host microbiota can protect
58 against infections. By contrast, plant pathology is still a step away from the advances made in
59 the medical field. As for humans, plant-associated microbes have been shown to play an
60 important role for plant health (and thereby for plant growth and survival). Bacterial entities
61 associated with the rhizosphere and phyllosphere of different plant species can increase
62 directly (e.g. production of microbial compounds) or indirectly (elicitation of plant defense)
63 host resistance to phytopathogenic microorganisms (Haney et al. 2015; Berg et al. 2014;
64 Ritpitakphong et al. 2016; Mendes et al. 2013; Bodenhausen et al. 2014; Santhanam et al.
65 2015). For example, artificial inoculations demonstrated that *Sphingomonas* spp. strains
66 isolated from plants can protect the model annual plant *Arabidopsis thaliana* from infections
67 caused by the causal agent of bacterial spot *Pseudomonas syringae* (Innerebner et al. 2011).
68 These studies support the importance of certain microbes in protecting plants against
69 pathogen infections; however, they mainly focused on interactions between a single
70 commensal species and a single pathogenic species. Therefore, the current challenge in plant
71 pathology is to study the possible relationships between commensal microbes and pathogenic
72 microbes at the community level, i.e. between microbiota (defined here as the whole
73 microbial community inhabiting the plant) and pathobiota (defined here as the complex of
74 microorganisms with the potential to cause disease on the plant host).

75 In the area of ecological genomics, this challenge calls for a microbiota
76 characterization across the range of native habitats encountered by a given plant species

(Roux & Bergelson 2016). In fact, as recently highlighted by Wagner et al. (2016), the ecology of the habitats where both plants and microorganisms co-exist and evolve is a crucial variable to take into account for investigating the intimate relationships between a host and its microbes participating in the holobiont system. In addition, in native habitats, plants are naturally exposed to pathogens whose attacks are influenced by local abiotic/biotic conditions (Bartoli et al. 2016). In this context, *A. thaliana* seems to be a relevant model plant to investigate the microbiota-pathobiota relationships in native habitats. Firstly, *A. thaliana* is found in diverse habitats (Mitchell-Olds & Schmitt 2006; Brachi et al. 2013) and substantial efforts have been recently made to characterize the rhizosphere and phyllosphere bacterial populations (Bulgarelli et al. 2012, Lundberg et al. 2012, Vorholt 2012). Secondly, first attempts by using culture-isolation methods and 16S rRNA gene sequencing showed that 30 OTUs mainly constitute the *A. thaliana* leaves bacterial communities (Jakob et al. 2002). Among these OTUs, *P. syringae*, *Pseudomonas viridiflava* and *Xanthomonas campestris* were the most abundant potential pathogens colonizing *A. thaliana* leaves (Kniskern et al. 2007).

In the present study, we aimed to investigate the *in situ* relationships between bacterial communities and the potential bacterial pathobiota in 163 natural *A. thaliana* populations collected in southwest of France. More precisely we attempted (i) to test whether a poorly diversified pathobiota is associated with a highly diversified microbiota, (ii) to identify microbial compositions that can be prone to pathogen invasion and subsequent persistence, and (iii) to assess the pathogenicity of a highly abundant bacterial species of the potential pathobiota. Because the plant bacterial community can rapidly change over time (Copeland et al. 2015) and can largely differ between plant compartments (Bodenhausen et al. 2013; Coleman-Derr et al. 2016; Wagner et al. 2016), we described microbiota and potential pathobiota in both leaf and root compartments across two seasons within a single life cycle.

RESULTS

Seasonal effect on population sampling

In this study, we focused on 163 natural *A. thaliana* populations identified in May 2014 in the Midi-Pyrénées region (Figure 1, Supplementary Table 1). As described in Brachi et al. (2013), we defined a population as a single group of plants growing in relatively homogeneous ecological conditions. These populations were chosen to maximize the diversity of habitats such as climate, soil type, vegetation type and degree of anthropogenic perturbation. The average distance among populations was 99.9 km (median = 92.6 km, SD = 55.4 km).

In agreement with previous observations in natural populations of *A. thaliana* located in northeast of Spain (Montesinos et al. 2009), the 163 populations studied here strongly differed in their main germination flush in autumn 2014, thereby leading to the observation of different life plant stages among populations. We therefore defined three seasonal groups (Figure 1). The first group, hereafter named 'autumn group', corresponded to 84 populations where most plants had reached the 5-leaf rosette stage during the 23-day sampling period in autumn (mid – November 2014 – early December 2015). The second group, hereafter named 'spring with autumn group', corresponded to 80 populations already sampled in autumn and additionally sampled during a 29-day period in early-spring (mid-February 2015 – mid-March 2015). Four populations sampled during autumn were not collected in early-spring to avoid modifications of their demographic dynamics. The third group, hereafter named 'spring without autumn group', corresponded to 79 populations only sampled during the 29-day period in early-spring. These populations were not sampled in autumn because the life stage of most plants was between 2-cotyledon and 4-leaf. The 'autumn' and 'spring with autumn' groups allowed to test whether the evolution of diversity and composition of bacterial

communities across seasons was dependent on the population considered. On the other hand, the 'spring with autumn' and 'spring without autumn' groups allowed to test whether the diversity and composition of bacterial communities in spring 2015 were affected by germination timing in autumn 2014.

Validation of the *gyrB* marker used for characterization of bacterial communities

For characterization of the *A. thaliana* bacterial fraction, a segment of the *gyrB* housekeeping gene (encoding for the β subunit of the bacterial gyrase) has been used. This molecular marker has a deeper taxonomic resolution (species-level) than other molecular markers designed on the hypervariable regions of the 16S rRNA gene (Barret et al. 2015). Furthermore this single-copy gene limits the overestimation of taxa carrying multiple copies of *rrn* operons. The *gyrB* prevalence was investigated in 32,062 bacterial genomic sequences available in the IMG database v4 (Markowitz et al. 2014) at the time of analysis (10th December 2015). Coding sequences that exclusively belong to the protein family TIGR01059 and KO2470 were defined as GyrB orthologs and retrieved for further analysis (30,627 hits found in 30,175 genomic sequences). The corresponding nucleotide sequences were aligned against a reference *gyrB* alignment (Barret et al. 2015) with the align.seqs () function in mothur (Schloss et al. 2009). Sequences that did not align (102) were discarded and only unique sequences were conserved in the reference alignment (10,427 haplotypes). According to the gANI (Varghese et al. 2015), the *gyrB* marker was highly precise (0.964) and sensitive (0.955) at a genetic distance of 0.02 (98% identity cutoff). In order to assess the potential amplification bias of *gyrB*, we amplified a mock community containing 52 bacterial strains (Supplementary Data Set 1) with both *gyrB* and 16S rRNA V4 region primers (Caporaso et al. 2011). Results showed that 16S rRNA gene and *gyrB* sequences were clustered with an

identity threshold of 97% and 98%, respectively. Based on this clustering, we obtained n = 19 OTUs with the 16 rRNA gene and n = 45 OTUs with *gyrB*. The 52 members of the mock community were all detected with the 16S rRNA gene marker, while three strains were not detected with the *gyrB* segment (Supplementary Data Set 1). Overall, our results suggested a better taxonomic resolution of the *gyrB* but associated with a small cost on bacterial detection.

Diversity of the *A. thaliana* microbiota and potential pathobiota

To avoid a confounding effect of developmental stage with seasons, all plants were collected at the rosette stage. After applying technical reproducibility thresholds, we obtained 18,610,383 total high-quality reads across a total number of 1655 samples, with on average ~10,136 reads per sample. After data trimming (see 'Materials and methods' section), we identified 6,627 non-singleton bacterial OTUs. A large amount of these OTUs were specific to roots or leaves, as only ~8.1% OTUs (n = 540 OTUs) were shared between both plant compartments. However, the relative abundance of OTUs shared between leaf and root samples was 20.2 and 16.0 higher than the relative abundance of leaf and root specific OTUs. This suggests that generalists OTUs are dominant members of the *Arabidopsis thaliana* microbiota.

As commonly observed in other plant species (Bulgarelli et al. 2013; Lundberg et al. 2012; Horton et al. 2014; Coleman-Derr et al. 2016; Wagner et al. 2016), bacterial communities were largely dominated by Proteobacteria (> 80%). At the order level, Burkolderiales (29.3%) and Sphingomonodales (27.9%) were dominant (Figure 2A). In comparison with autumn, samples in spring were enriched for Burlkolderiales ($\chi^2 = 7.93$, $P < 0.01$) and depleted for Sphingomonodales ($\chi^2 = 18.18$, $P < 0.001$) but only in the root

compartment (Figure 2A). No germination timing effect during autumn was observed on the relative abundance of OTUs at the order level for plants collected in spring (Figure 2A).

In order to characterize the potential pathobiota embedded within the microbiota, we established a list of 199 phytopathogenic bacterial species (Supplementary Data Set 2). From this list, 12 bacterial species were identified across all samples. Among them, the three most abundant species representing more than 88% of the whole potential pathobiota were *Janthinobacterium agaricidamnosum* (55.5%), *X. campestris* (17.3%) and *P. viridiflava* (15.3%) (Figure 2B). For each seasonal group, we observed an enrichment for *J. agaricidamnosum* ('autumn', $\chi^2 = 7.39$, $P < 0.01$; 'spring with autumn', $\chi^2 = 43.56$, $P < 0.001$; 'spring without autumn', $\chi^2 = 19.99$, $P < 0.001$) and a depletion for *X. campestris* ('autumn', $\chi^2 = 11.44$, $P < 0.001$; 'spring with autumn', $\chi^2 = 11.84$, $P < 0.001$; 'spring without autumn', $\chi^2 = 6.95$, $P < 0.01$) and *P. viridiflava* ('autumn', $\chi^2 = 0.26$, $P = 0.6075$; 'spring with autumn', $\chi^2 = 19.22$, $P < 0.001$; 'spring without autumn', $\chi^2 = 8.95$, $P < 0.01$) in the roots in comparison with the leaves (Figure 2B). In the leaf compartment, an enrichment for *P. viridiflava* between autumn and spring ($\chi^2 = 9.10$, $P < 0.01$) was associated with a depletion of *X. campestris* ($\chi^2 = 7.57$, $P < 0.01$). In the root compartment, an enrichment for *J. agaricidamnosum* between autumn and spring ($\chi^2 = 15.19$, $P < 0.001$) was associated with a depletion of *X. campestris* ($\chi^2 = 8.08$, $P < 0.01$). Similarly to the microbiota, no germination timing effect during autumn was observed on the relative abundance of the potential pathogenic species for plants collected in spring (Figure 2).

It is noteworthy that the relationship between relative abundance and prevalence across populations was weak for some potential pathogenic bacterial species (Figure 2C). For example, although the total relative abundance of *Pantoea agglomerans*, *P. syringae* and *Sphingomonas melonis* across all samples ranged from 2.1% to 4.6% (Figure 2B), these three species were present on average in more than 57% of the populations (Figure 2C). Visual

198 inspection of original data suggests that this pattern is mainly explained by the presence of
199 few highly infected plants in many populations.

200 After controlling for biological (such as plant age) and technical noises (such as
201 sampling date and total number of observations per sample), we investigated whether richness
202 (i.e. total number of OTUs per sample) and Shannon α-diversity were dependent on the
203 combined effects of season, plant compartment and population. Because statistical results led
204 to similar biological conclusions between richness and Shannon α-diversity (Supplementary
205 Tables 2 - 9), we only present results for the Shannon α-diversity. Across the three seasonal
206 groups, bacterial communities of the microbiota were on average less diverse in roots than in
207 leaves (Figure 3A, Supplementary Tables 2 & 4). For the potential pathobiota, a similar
208 pattern was observed in autumn but not in spring where the potential pathogenic communities
209 were more diverse in roots than in leaves whatever the germination timing in autumn
210 considered (Figure 3B, Supplementary Tables 2 & 4). In addition, the evolution of Shannon
211 α-diversity between autumn and spring was highly dependent on the population considered,
212 both for the microbiota and the potential pathobiota (Figure 4, Supplementary 2). The
213 Shannon α-diversity variance explained by the factor 'season × population' was nonetheless
214 higher for the microbiota (~25.2%) than for the potential pathobiota (~14.1%)
215 (Supplementary Table 3). The strength of this interaction was further significantly dependent
216 on the plant compartment for the microbiota but not for the potential pathobiota
217 (Supplementary Table 2). In particular, the difference among populations for the evolution of
218 Shannon α-diversity between autumn and spring was greater in the root compartment than in
219 the leaf compartment (Figure 4), with the Shannon α-diversity in roots increasing and
220 decreasing between the two seasons in 70.3% and 29.7% of the populations, respectively
221 (Figure 4).

Within each seasonal group, the Shannon α-diversity variance of the microbiota was first explained by differences among populations (~25%), followed by the interaction between plant compartment × population (~11.7%) and differences between leaves and roots (~3.7%) (Figure 5, Supplementary Tables 5-8). A similar pattern was observed for the potential pathobiota, with the exception of the seasonal group 'spring without autumn' for which variation of Shannon α-diversity among populations was similar between the leaf and root compartment (i.e. no Shannon α-diversity variance was explained by the factor 'plant compartment × population') (Figure 5, Supplementary Tables 5-8). At the 'plant compartment × seasonal group' level, the difference among populations for Shannon α-diversity was on average higher for the microbiota than for the potential pathobiota (~36.8% *vs* ~25.7% of variance explained by the factor 'population') (Figure 5, Supplementary Table 9). In addition, Shannon α-diversity was found to vary at a very small geographical scale (i.e. less than 1km; Figure 5).

Finally, whatever the 'seasonal group × plant compartment' considered, we observed a highly significant concave relationship between the Shannon α-diversity of the potential pathobiota and the Shannon α-diversity of the microbiota (Figure 6, Supplementary Tables 10-11). In other words, poorly diversified potential pathobiota were associated either with highly or with poorly diversified microbiota, whereas highly diversified potential pathobiota were found in presence of microbiota with an intermediate level of diversity.

Composition and structure of *A. thaliana* microbiota and potential pathobiota

To identify the main sources explaining between-sample diversity, we run separately a principal coordinates analysis (PCoA) on both microbiota and potential pathobiota β-diversity matrices. For the microbiota, the first two PCoA axes explained 19.9% of the β-diversity (Figure 7). The microbiota was structured according to a pattern of two perpendicular

247 branches and similar patterns were observed for normalized procedures (see Material and
248 Methods section). The variation along the first branch was related to the variation among
249 samples for the abundance of the most abundant OTU (~16.1% and ~25.3% in the leaf and
250 root compartments, respectively) corresponding to the '*Sphingomonas sp.*' OTU (Figure 7).
251 The variation among samples for the abundance of the four next more abundant OTUs
252 (*Burkolderia fungorum*, a species belonging to the family Oxalobacteraceae, *Variovorax sp.*
253 and *Pseudomonas moraviensis*) was related to the variation along the second branch (Figure
254 7). In agreement with the perpendicularity observed between the two branches, we observed a
255 pattern of repulsion between *Sphingomonas sp.* and *Burkolderia fungorum*, i.e. these two
256 OTUs were almost never found together in the same sample (Figure 8A). Within each
257 seasonal group, the variance along the PCoA axes was largely explained by differences
258 among populations (up to ~81%) (Figure 9, Supplementary Tables 5-8). In most populations
259 sampled during spring, *Sphingomonas sp.* was either present or absent in all individuals
260 sampled in a given population. More precisely, in the root compartment, *Sphingomonas sp.*
261 was present (30.8% of the populations) and absent (59.1% of the populations) in all
262 individuals sampled. The remaining populations (10.1%) were constituted by a mix of
263 sampled individuals in which *Sphingomonas sp.* was present or absent. A similar pattern was
264 observed for the leaf compartment with *Sphingomonas sp.* being present (28.2% of the
265 populations) and absent (55.5% of the populations) in all individuals sampled. In addition, the
266 evolution of the β-diversity between autumn and spring was highly dependent to the
267 population considered (Figures 9 and 10, Supplementary Table 2). When considering the
268 populations sampled both in autumn and spring, the variance explained by the factor 'season
269 × population' was 51.5% and 44.0% for the first and second PCoA axes, respectively
270 (Supplementary Table 3). Accordingly, the abundance of *Sphingomonas sp.* dramatically
271 changed between autumn and spring in many populations. In more than half of the

populations sampled in autumn (55% for the leaf compartment and 61.7% for the root compartment), *Sphingomonas sp.* was present in all individuals sampled in autumn but absent in all individuals sampled in spring. In agreement with Shannon α-diversity, no effect of germination timing during autumn was observed on the β-diversity of plants collected in spring (Figure 9, Supplementary Table 4).

For the potential pathobiota, the first two PCoA axes explained 14.9% of the β-diversity (Figure 7). The potential pathobiota was structured according to a pattern of three branches (Figure 7). The variation along two of these branches was related strongly but independently to the variation among samples for the abundances of *P. viridiflava* and *X. campestris* (Figure 7). The third branch was weakly related to the variation among samples for the abundance of *J. agaricidamnosum* (Figure 7). Similarly to the relationship between *Sphingomonas sp.* and *Burkolderia fungorum* observed in the microbiota, *P. viridiflava* and *X. campestris* were almost never found together in the same sample (Figure 8B). Within each seasonal group, the variance along the PCoA axes was first explained by differences between the leaf and root compartments (up to ~31.1%), followed by the differences among populations (~11.4%) (Figures 7, 9 and 10, Supplementary Tables 5-8). Similarly, when considering the populations sampled both in autumn and spring, the β-diversity variance was first explained by differences between the leaf and root compartments (i.e. 19.2% and 11.0% for the first and second PCoA axes, respectively) (Supplementary Tables 2 and 3). The evolution of β-diversity of the potential pathobiota between autumn and spring was also dependent on the population considered but to a lesser extent than what was observed for the microbiota β-diversity (i.e. 12.6% and 1.3% for the first and second PCoA axes, respectively) (Figures9 and 10, Supplementary Tables 2 and 3). In contrast to Shannon α-diversity, the level of differences between the leaf and root compartments for the β-diversity was strongly

296 affected by the germinating timing of the populations in autumn (Figure 9, Supplementary
297 Table 4).

298 At the 'plant compartment × seasonal group' level, the difference among populations
299 for the β-diversity was on average higher for the microbiota than for the potential pathobiota
300 (~74.3% *vs* ~15.0% of the first PCoA axis variance explained by the factor 'population')
301 (Figure 9, Supplementary Table 9, Supplementary Figure 1). Similarly to Shannon α-
302 diversity, β-diversity was found to vary at a very small geographical scale (i.e. less than 1km;
303 Figure 9).

304 In order to study the relationship between the composition of microbiota and potential
305 pathobiota, we run a sparse Partial Least Square Regression (sPLSR) (Carrascal et al. 2009;
306 Cao et al. 2011) by considering only OTUs that were present in at least in 1% of the samples
307 considered. This method was adopted to maximize the covariance between linear
308 combinations of OTUs from the microbiota and linear combinations of OTUs from the
309 potential pathobiota. Across all samples, a large percentage of variation of the potential
310 pathobiota (39.9%) was explained by a combination of OTUs explaining up to 27.2% of the
311 variation of the microbiota (Figure 11). The percentage of variation of the potential pathobiota
312 explained by variation of the microbiota was higher in spring (~68.3%) than in autumn
313 (~36.8%) and, to a lesser extent, in the root compartment (~66.6%) than in the leaf
314 compartment (~48.9%). Variation of the potential pathobiota was explained by different
315 combinations of the microbiota OTUs (between 4 and 7 OTUs) among the six 'season group
316 × plant compartment' combinations (Figure 11). The third most abundant microbiota OTU (a
317 species belonging to the family Oxalobacteraceae) was the only microbiota OTU found
318 among the different microbiota OTU combinations explaining the potential pathobiota.
319 Among the 34 candidate OTUs of the microbiota, a large fraction (~47.1%) corresponds to
320 unclassified OTUs at the order level.

Isolation and *in planta* tests of strains belonging to the *P. syringae* complex to confirm pathogenicity of the potential microbiota

To test whether bacterial species found in the potential pathobiota had pathogenicity abilities, we first attempted to isolate strains belonging to the *P. syringae* complex (where we included both *P. syringae sensu stricto* and *P. viridiflava*) from both leaf and root samples. By using culture plating and a *P. syringae* PCR marker method (Guilbaud*et al*. 2015), we succeed in isolating a total number of 97 strains, all from the leaf compartment (Figure 12, Supplementary Data Set 3). In agreement with the results obtained by community profiling approach (Figure 2B), most of the strains (n = 74) clustered with *P. viridiflava*, and in particular with the phylogroup 7. All the remaining strains belong to *P. syringae sensu stricto*. Among these strains, 10, 6 and 2 strains were placed into the phylogroups 2, 13 and 9 respectively (Figure 12, Supplementary Data Set 3). A single strain was found for the phylogroups 1 and 11. Three strains (0108-Psy-GAIL-BL, 0117-Psy-NAZA-AL and 0097-Psy-BAGNB-BL) were not affiliated to any *P. syringae* phylogroup and they might be considered as new phylogroups (Figure 12, Supplementary Data Set 3). We found more *P. syringae* strains in spring (n=74) than in autumn (n=23) (Figure 12, Supplementary Data Set 3). These results are in accordance to those obtained by metagenomics showing a burst of strains belonging to the *P. syringae* complex during spring in the leaf compartment (Figure 2).

Hypersensitive Response (HR) on tobacco showed that 84 strains displayed positive reactions. Among the HR negative strains, eight belonged to the *P. viridiflava* phylogroup 7, four strains were clustered in the phylogroup 13 and one strain was not affiliated (Supplementary Data Set 3). Four strains (0114-Psy-NAUV-BL and 0124-Psy-SAUB-AL from the *P. viridiflava* phylogroup 7; 0132-Psy-BAZI-AL from phylogroup 2 and 0143-Psy-THOM-AL from phylogroup 13) were tested for *in planta* bacterial growth in the four corresponding local natural populations of *A. thaliana*, each represented by two randomly

346 selected accessions. A highly significant bacterial growth was observed for each strain
347 (Supplementary Table 12), with bacterial concentrations reaching10^6 CFU cm^{-2} 7 days post
348 inoculation (Figure 13A). In addition, eight strains were tested for pathogenicity on each of
349 the eight corresponding local natural populations of *A. thaliana*. Disease symptoms were
350 observed for strains belonging to either *P. viridiflava* (Supplementary Figure 2) or *P. syringae*
351 *sensu stricto* (Supplementary Figure 3). For example, the 0111-Psy-RAYR-BLstrain
352 (phylogroup 1) was highly aggressive on all the accessions tested (Figure 13B). Interestingly,
353 the appearance of symptoms over time was highly dependent on the interactions between the
354 eight strains and the eight corresponding local accessions (Supplementary Table 13),
355 suggesting G x G interactions. Notably, while almost no genetic variation in disease
356 symptoms was observed among *A. thaliana* accessions for each *P. syringae sensu stricto*
357 strain, disease symptoms largely differed among *A. thaliana* accessions for each *P. viridiflava*
358 strain (Supplementary Figures 2 and 3).

359 Taken together, our results demonstrated that most of the strains belonging to the *P.*
360 *syringae* complex are able to induce disease on the host of origin as well on a non-host
361 species (i.e. tobacco), thereby supporting the pathogenicity behavior of strains identified in
362 the potential pathobiota.

363

DISCUSSION

Putting the characterization of bacterial communities in an ecological genomics framework

Characterizing the microbiota and potential pathobiota across a large range of native habitats brought complementary information to laboratory-based studies on the role of bacterial species on *A. thaliana* health. By adopting an ecological genomics approach (Wagner et al. 2016), we found that the diversity and composition of bacterial communities of *A. thaliana* were affected *in situ* by the combined effects of season, plant compartment and population. First, the bacterial communities were highly dynamic between the two seasons investigated in this study. This seasonal succession community is in agreement with a previous study performed on three crop species (common bean, soybean and canola) over a growing season in a field experiment in Northern America (Copeland et al. 2015). Therefore, as advocated by Maignien et al. (2014), we need to develop microbiota studies over the whole plant life cycle (from seed to seed) to obtain an unbiased overview of the *A. thaliana* interacting microbial diversity. In this study, shifts in microbial communities can originate from changes in weather conditions during winter. These changes can have modified the plant physiology resulting in a shift of the relative abundance of the host ecological niches, which in turn affected the competitive relationships among the microbial members with potential direct effects on pathogen invasion. For example, during spring, plants were enriched for Burlkolderiales, an order composing the root core microbiota of *A. thaliana* and the three related species *Arabidopsis lyrata*, *Arabidopsis halleri* and *Cardamina hirsuta* (Schlaeppi et al. 2014). In controlled conditions, *Burkholderia* strains isolated from soil can reduce disease severity on both tomato and soybean plants infected by fungi and oomycete species (Benítez & Gardener 2009). The increasing of Burlkolderiales observed in our study suggests a possible role of this order in protecting *A. thaliana* populations from pathogen attacks

occurring mostly in spring. Besides modifications of plant resources, another explanation relies on climate optima of some bacterial species for an optimal growth in the phyllosphere. In agreement with this hypothesis, we found that the relative abundance of the phytopathogenic bacterium *P. viridiflava* increased between autumn and spring when environmental conditions are known to be more suitable for the spread and epidemics occurrence of this pathogen (Bartoli et al. 2014). However, our results point out the general missing information on the ecology of most microbes inhabiting *A. thaliana*.

Second, the diversity and composition of bacterial communities largely differed among the populations, in particular for the microbiota. More interestingly, these differences occurred at a very small geographic scale. Four non-exclusive hypotheses can be advanced to explain this fine-grained spatial variation. Firstly, large differences among populations can result from regional factors such as limitation in dispersal rates of microbes among populations (Lindström & Langenheder 2012). Secondly, the diversity and composition of bacterial communities in *A. thaliana* in the south-west of France is related to local habitat conditions, such as soil conditions (Bulgarelli et al. 2013; Lundberg et al. 2012) and composition of plant communities (Aleklett et al. 2015; Geremia et al. 2016). Thirdly, as previously demonstrated in the perennial wild mustard *Boecherastricta* (Wagner et al. 2016), plant age can shape the leaf and root microbiota. Local variation in the diversity and composition of bacterial communities can result from different ages of *A. thaliana* among populations due to fine-grained spatial variation in germination timing. At the regional scale, temperature and precipitation regimes have been suggested as selective agents acting on phenological seed-related traits in *A. thaliana*, (Montesinos et al. 2009). The very fine spatial scale in germination timing observed in this study suggests that ecological factors acting at a local scale (such as soil conditions and the stage of vegetative development of associated plant communities) can also act as selective agents. Fourth, although the importance of

genetics in shaping natural variation in bacterial communities is still debated in *A. thaliana* (Horton et al. 2014; Roux & Bergelson 2016), among-population differentiation at genes involved in the molecular dialog with microbes may drive variation at a small spatial scale (Lebeis et al. 2015; Roux & Bergelson 2016). Teasing apart the relative roles of these putative factors will require a thorough complementary ecological, demographic and genomic characterization of the populations (Hacquard et al. 2016).

Third, the plant compartment drastically influenced the composition of the potential pathobiota as well as the relative abundance of the most abundant pathogenic species. Our results are in accordance with previous studies reporting that pathogenic species such as *P. syringae sensu latu* and *X. campestris* evolved specific strategies (e.g. entry by stomata and hydatodes) to infect the leaf compartment in a wide range of crops (Mansfield et al. 2012). On the other hand, we observed that the mostly abundant OTUs of the microbiota were shared between leaves and roots. Similarly, a recent culture-dependent study employing whole-genome sequencing and functional analysis of bacteria associated with both leaves and roots of *A. thaliana* pointed out a clear taxonomy and functional overlap of the bacterial populations inhabiting this plant species (Bai et al. 2016). Altogether, these results are in contrast with previous studies on human microbiota demonstrating a remarkable partitioning of commensals within the human body where each host tissue constitutes a unique microenvironment for the microbial communities (Belkaid & Hand 2014).This discrepancy might originate from the lack of a specialized immune system in plants. In fact, the specialization of human microbiota is mainly related to the specialization of the cells constituting the host immune system. In this concern, unique microbial communities have co-evolved under the control of host specific immunity (Belkaid & Hand 2014). In addition, as previously demonstrated for several plant commensals (Marchetti et al. 2010; Kawaguchi & Minamisawa 2010), the lack of a type three secretion system (the most important infection

machinery in both plant and animal pathogens) could allow to some microbial species to penetrate into the plant roots and rapidly spread over the rest of the plant tissues.

Relationships between microbiota and potential pathobiota

In agreement with theoretical expectations on the diversity-invasion relationships (Mallon et al. 2015), a poorly diversified potential pathobiota was associated with a highly diversified microbiota, suggesting that more diverse bacterial communities better resist invasion by potential pathogenic species than less diverse bacterial communities. This pattern is likely explained by a better exploitation of plant resources by bacterial species from the microbiota, thereby drastically limiting the number of ecological niches available for potential pathogenic bacteria (Eisenhauer et al. 2013). However, we cannot rule out the effects on pathogens of antagonistic (e.g. production of anti-microbial compounds) and predator (e.g. grazing) species whose occurrence increases with the microbiota diversity (Mallon et al. 2015). More surprisingly, we observed that a poorly diversified potential pathobiota was also associated with a poorly diversified microbiota. Several non-exclusive hypotheses can support this pattern. First, it has been observed that neutral drift increases in communities under ecological disturbance. By following this scenario, the establishment of a single pathogenic species is random and it mainly depends on pathogen absolute abundance (Kinnunen et al. 2016). Second, a pathogenic species can deploy its arsenal of virulence proteins for the exclusive invasion of a given plant compartment, depending on the defense genetic basis of the host. Third, both microbiota and potential pathobiota can be poorly diversified because other microbial communities (such as fungal and oomycete communities) exploit most of the resources available in the plant, leading to niche competition between microbial communities (Agler et al. 2016).

Although pathogen-focused network analysis was proposed for investigating the relationships between microbiota and pathogenicity, this method seems to be only suitable for studying monospecific interactions between a single pathogenic species and the other members of the microbial community (Poudel et al. 2016). Here, by applying sPLSR, we identified combinations of OTUs from the microbiota associated with combinations of pathogenic species. In accordance with community succession observed between autumn and spring, those microbial combinations were dependent on the season, suggesting a dynamics in the potential biomarkers controlling pathogen spread. In addition, we found repulsion patterns both within the microbiota and within the potential pathobiota. Two hypotheses can be suggested to explain this pattern. Firstly, the different microbial species cannot coexist due to non-overlapping abiotic (climate, soil…) niches. Secondly, both microbial counterparts compete for the same resources. In the context of pathobiota, co-occurrence in infectious diseases is well studied. However, interspecific competition among pathogenic species is still poorly understood. Experimental evolution on the ϕ6 bacteriophage demonstrated that intraspecific competition for resources can lead to pathogen jump on a new host (Bono et al. 2013). In this view, we can speculate that competition for resources that occur between functional equivalent species (Burke et al. 2011) such as *X. campestris* and *P. viridiflava*, can drive one of these two pathogens to jump to another host during spring when other host plant species around *A. thaliana* populations are more abundant.

Testing the hypotheses underlying the relationships we observed between and within microbiota and potential pathobiota will require the isolation of a large number of strains representative of the microbial communities found in the 163 *A. thaliana* populations. Although time consuming, a first step in strain isolation allowed confirming the pathogenicity of one of the three most abundant microbial species composing the potential pathobiota. Recent studies on *A. thaliana* microbiota reported the added value of using synthetic bacterial

488 communities in controlled conditions for testing relationships among bacterial OTUs (Bai et
489 al. 2015; Lebeis et al. 2015). By following an ecological genomics framework, the next
490 challenge will be to set up synthetic bacterial communities-based experiments with the
491 transferring of sterile plants into more realistic abiotic and biotic conditions.

MATERIAL AND METHODS

Identification of *A. thaliana* populations

A field prospection in May 2014 allowed the identification of 233 natural populations of *A. thaliana* in the Midi-Pyrénées region. In agreement with population dynamics observed in *A. thaliana* (Picó 2012), we observed an important population turnover. During the sampling period in autumn 2014, plants were present in only 72.5% of the populations (n = 169). Due to small population size, no plants were sampled in six of the 169 populations. The remaining 163 populations were sampled during two different seasons, autumn (November/December 2014) and spring (February/March 2015).

To avoid a confounding effect between the sampling date and geographical origin, populations were randomly collected during the sampling periods in autumn 2014 and early-spring 2015.

Generation of the *gyrB* amplicons and sequences

To characterize the bacterial community of the *A. thaliana* microbiota, ~4 individuals per population and per season were *in situ* sampled at the rosette stage resulting in a total number of 1,912 leaf and root samples. Plants were excavated using flame-sterilized spoons and then manipulated with flame-sterilized forceps on a sterilized porcelain plate. Gloves and the porcelain plate were sterilized by using Surface'SafeAnios®. Roots were rinsed into individual tubes of sterilized distilled water to remove all visible rhizosphere. Rosettes were also rinsed into individual tubes of sterilized distilled water to remove eventual traces of dust. Both leaves and roots were then placed into clean autoclaved tubes and immediately stored in dry ice to avoid alterations in the composition of the microbiota during transferring to the

laboratory. Samples were then stored at -80°C prior DNA extraction. For each plant, we recorded the sampling date. In addition, the age of each plant was approximated by measuring the maximum rosette diameter and by counting the number of leaves.

The epiphytic and endophytic bacterial components of either leaf or root samples were not separated. We therefore extracted the total DNA for both epiphytic and endophytic microbes inhabiting leaves and roots. The DNA of each samples was extracted as follows: i) leaves were placed in 96 well plates containing sterilized beads and homogenized for 1 min with 30 vibrations per second in a plate shaker and incubated 30 min in 500 µl of buffer containing 200 mM of Tris-HCl at pH 7.5, 250 mM of NaCl, 25 mM of EDTA and 0.5% SDS, ii) roots were placed in Eppendorf tubes and incubated 10 min in a bath sonicator and treated with the same conditions described above for the leaves, iii) for both leaves and roots phenol/chloroform 25:24:1 pH 8.0 (Sigma Aldrich®) was used for extraction and purification of the DNA, iv) DNA was precipitated with isopropanol and washed with 70% EtOH and eluted in 100 µl of DNA-free water. DNA samples were stored at -20°C prior PCR amplification.

To characterize the bacterial communities, the housekeeping gene gyrase *B* (*gyrB*) was amplified with some modification of the protocol already described in (Barret et al. 2015). Briefly, three tags were added at each 5' and 3' of the original primers to allow the multiplexing of three plates. Primers including Illumina adapter sequences and without internal tags were: Fw (5'-CTTTCCCTACACGACGCTCTTCCGATCTMGNCCNGSNATGTAYATHGG - 3') and Rv (5'-GGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTTACNCCRTGNARDCCDCCNGA - 3'). The internal tags for multiplexing consisted in: TAG1 (Fw - GACTAC, Rv - AAGGCC), TAG2 (Fw - CTGGTT, Rv - GTCAGG), TAG3 (Fw - ACTCGA, Rv - CCTCTT).MTP taq

DNA Polymerase-Sigma Aldrich® was used and the PCR mix was composed by 2.5 µl of Taq Buffer, 0.2 µl of dNTPs 10mM, 1 µl of Fw Primers (10p/mol), 1 µl of Rv Primers (10p/mol), 0.3 µl of Taq polymerase, and 1 µl of 10 fold diluted DNA for a final volume of 25 µl. PCR amplifications were performed by using 95°C for 5 min of initial denaturation followed by 40 cycles with 95°C for 30 sec, 52°C for 1 min and 30 sec, 68°C for 1 min and a final elongation of 68°C for 5 min. Negative controls were also added to investigate whether amplification of bacterial DNA was detected on the water used for both leaves and roots washing, the water used for DNA elution and the water used for PCR mix.

For each sample, PCR amplifications were repeated three times and technical replicates were pooled in a unique PCR plate. Each technical replicate plate was pooled by considering the three tags for multiplexing. PCR products were purified by using Agencourt®AMPure® magnetic beads following manufacturer's instructions and purified amplicons were quantified with Nanodrop and appropriately diluted to obtain an equimolar concentration. Two µl of equimolar PCR purified products were used for a second PCR with the Illumina adaptors. The second PCR amplicons were then purified and quantified as described above to obtain a unique equimolar pool. The latter was quantified by RQ-PCR and then sequenced on a Illumina MiSeq 2X250 v2 (Illumina Inc., San Diego, CA, USA) in the GetPlage Platform (Toulouse, France). MS-102-3003 MiSeq Reagent Kit v3 600 cycle was used for this purpose.

Bioinformatic analysis

After MiSeq sequencing, reads were demultiplexed by using the internal tags sequences added in the first PCR. Results after demultiplexing were ranged in an average value of 46,135 ± 19,820 sequences per sample. Prior further analysis, the negative controls were checked for presence/absence of amplicons by blasting them against the *gyrB* database

565 composed by 30,627 sequences (Barret et al. 2015). Because negative controls showed no
566 trace of entire *gyrB* sequences, they were removed before clustering. Samples showing low
567 sequence quality were also removed before clustering, resulting in a total of 1,903 samples.
568 Clustering and taxonomy assigning of the sequences in OTUs was performed with mothur
569 (Schloss et al. 2009) software with a swarm clustering algorithm (Mahé et al. 2015) by using
570 a clustering threshold (d) = 1. Only the OTUs that were composed by a minimum of 5
571 sequences across all samples were kept for further analysis, resulting in a total of 278,336
572 OTUs. Then, we applied three steps of trimming. First, to control for sampling limitation
573 within each sample, we estimated a Good's coverage score for each sample. Based on the
574 distribution of the Good's coverage score, we decided to consider only the samples with a
575 score more than 0.5. Second, we removed samples with less than 600 reads. Third, only OTUs
576 showing a minimum frequency of 1% in at least one sample were selected by using a home-
577 made perl program. The final data set corresponded to a matrix of 1,655 samples by 6,627
578 OTUs.

579

580 **Analysis of the α and β-diversity and characterization of the potential pathobiota**

581 The final matrix of 1,655 samples by 6,627 OTUs was used to estimate the Shannon
582 diversity and the richness (total number of OTUs per sample) by using the summary.single ()
583 function of the mothur software (Schloss et al. 2009). Indexes of α-diversity were also
584 calculated by sample rarefaction of 300/600/900 iters. Results between non-rarefied and
585 rarefied samples were similar and the complete non-rarefied data set was used for the
586 calculation of microbiota diversity.

587 Because of the presence of many double zeros in the matrix, the β-diversity was
588 studied by first converting the microbiota matrix with a Hellinger transformation (Legendre &

589 Gallagher 2001) by using the *vegan* package (Jari Oksanen et al. 2009) in the *R* environment.
590 Following Ramette (2007), the resulting Hellinger distance matrix was reduced by running a
591 Principal Coordinates Analysis (PCoA) with the ape package (Paradis et al. 2004) in the *R*
592 environment. Because PCoA performed on Hellinger distance matrix based on rarefied data
593 and on Jaccard similarity coefficient matrix distance led to similar patterns of ordination
594 (Supplementary Figure 4), the PCoA coordinates from the non-rarefied Hellinger distance
595 matrix were retrieved and used for statistical analysis described below.

596 Non-metric multidimensional scaling (NMDS) was also run on the Hellinger distance
597 matrix. However, the values of stress were 0.431 and 0.293 for 2D and 3D NMDS ordination
598 space, respectively. This stress values suggest lack of fit between the ranks on the NMDS
599 ordination configuration and the ranks in the original distance matrix (Ramette 2007).

600 To characterize the potential pathobiota in both leaves and roots of the 163 *A. thaliana*
601 populations, a list of the phytopatogenic bacteria was created by using the summary of plant
602 pathogenic bacteria described previously (Bull et al. 2010; Bull et al. 2012; Bull et al. 2014).
603 The pathobiota list (Supplementary Data Set 2) was composed by 199 bacterial species and
604 was filtered on the microbiota matrix showing the taxonomy affiliation of the OTUs. The sub-
605 matrix for the potential pathobiota generated by the filtering was analyzed as described for the
606 microbiota matrix. The α-diversity and β- diversity were calculated as described above for the
607 microbiota matrix.

608

609

610

611

Statistical analysis

Natural variation of microbiota and potential pathobiota

Natural variation for the seven descriptors of microbiota (i.e. richness, α-diversity Shannon index, PCo1 and PCo2) and potential pathobiota microbiota (i.e. α-diversity Shannon index, PCo1 and PCo2) was explored using the following mixed models:

(i) To explore natural variation in populations collected both in autumn and spring, we used the following mixed model:

$$Y_{ijklmno} = \mu_{trait} + season_i + compartment_j + season_i \times compartment_j + population_k + season_i \times population_k + compartment_j \times population_k + season_i \times compartment_j \times population_k + sampling\_date_l(season_i) + diameter_m(season_i) + leaf\_number_n(season_i) + obs_o + \varepsilon_{ijklmno} \quad (1)$$

where '*Y*' is one of the 7 descriptors,'μ' is the overall phenotypic mean; 'season' accounts for differences between autumn and spring; 'compartment' accounts for differences between leaves and roots; 'population' measures the effect of populations; interaction terms involving the 'population' term account for variation among populations in reaction norms across the two seasons and/or the two plant compartments; 'ε' is the residual term. Four terms were added to control for noise that may affect significance of the other model terms. First, 'sampling_date' accounts for the number of days since the first population was collected within each season. Second, because age can shape leaf and root microbiota (Wagner et al. 2016), the two traits 'rosette diameter' and 'leaf number' were used as proxies of plant age. Third, 'obs' corresponds to the total number of observations and accounts for technical noise attributable to sequencing depth.

(ii) To explore natural variation in populations collected in spring, we used the following mixed model:

$$Y_{ijklmno} = \mu_{trait} + w/wo\_Autumn_i + compartment_j + w/wo\_Autumn_i \times compartment_j + population_k(w/wo\_Autumn_i) + compartment_j \times population_k(w/wo\_Autumn_i) + sampling\_date_l + diameter_m + leaf\_number_n + obs_o + \varepsilon_{ijklmno} \quad (2)$$

All the terms are described in model (1) with the exception of 'w/wo_Autumn' accounting for differences between populations collected both in autumn and spring (i.e. 'spring with autumn' seasonal group) and populations collected only in spring (i.e. 'spring without autumn' seasonal group).

(iii) To explore natural variation in populations within each of the three following categories of populations (i.e. 'autumn' populations, 'spring with autumn' populations and 'spring without autumn' populations), we used the following mixed model:

$$Y_{ijklmno} = \mu_{trait} + compartment_j + population_k + compartment_j \times population_k + sampling\_date_l + diameter_m + leaf\_number_n + obs_o + \varepsilon_{ijklmno} \quad (3)$$

(iv) To explore natural variation in populations within each 'plant compartment x categories of populations (i.e. 'autumn' populations, 'spring' populations, 'spring with autumn' populations and 'spring without autumn' populations)' combination, we used the following mixed model:

$$Y_{ijklmno} = \mu_{trait} + population_k + sampling\_date_l + diameter_m + leaf\_number_n + obs_o + \varepsilon_{ijklmno} \quad (4)$$

In these four models, all factors were treated as fixed effects, except 'population' which was treated as a random effect. For fixed effects, terms were tested over their

appropriate denominators for calculating *F*-values. Significance of the random effects was determined by likelihood ratio tests of model with and without these effects. Inference was performed using ReML estimation, using the PROC MIXED procedure in SAS 9.3 (SAS Institute Inc., Cary, North Carolina, USA) for all traits. A Bonferroni correction for the number of tests was performed for each modeled effect at a nominal level of 5%.For the purpose of drawing plots, Best Linear Unbiased Predictions (BLUPs) were obtained for each population by running model (4).

To estimate the percentage of phenotypic variance explained by each classification variable (i.e. 'season', 'compartment', 'population' and interacting terms) in models (1), (3) and (4), noise was first taken into account by performing a first regression of the descriptors against the terms 'sampling_date', 'diameter', 'leaf_number' and 'obs' using the PROC MIXED procedure in SAS 9.3 (SAS Institute Inc., Cary, North Carolina, USA). Then, a second regression including the appropriate classification terms was run on the residuals of the first regression using the PROC VARCOMP procedure in SAS 9.3 (SAS Institute Inc., Cary, North Carolina, USA).

Microbiota – potential pathobiota relationships

In order to study the relationship between microbiota α-diversity and potential pathobiota α-diversity, linear and non-linear regressions were fitted using the 'lm' and 'nls' functions implemented in the *R* environment, respectively. Model selection was based on a difference of three points in Akaike's information criterion (AIC) and Bayesian information criterion (BIC).

In order to study the relationship between microbiota composition and potential pathobiota composition, a sparse Partial Least Square Regression (sPLSR) (Carrascal et al.

682 2009; Lê Cao et al. 2008) was adopted to maximize the covariance between linear
683 combinations of OTUs from microbiota (matrix X) and linear combinations of OTUs from
684 potential pathobiota (matrix Y). sPLSR was run using the mixOmics package implemented in
685 the *R* environment (Lê Cao et al. 2008; Lê Cao et al. 2010). For the microbiota matrix, only
686 OTUs present in at least 1% of the samples were considered. For the potential pathobiota
687 matrix, we considered the seven most abundant OTUs (Figure 2). In addition to the lasso
688 model, sPLSR results were validated by plotting the Root Mean Square Error of Prediction
689 (Maestre 2004; Lê Cao et al. 2008). For the microbiota, we calculated the final loadings for
690 the ten OTUs with the highest initial loadings for each component. Given the small number of
691 OTUs in the potential pathobiota (n = 7), the initial loading of each OTU was kept for each
692 component. Following Carrascal et al. (2009), only OTUs (from the microbiota or the
693 potential pathobiota) with a loading value above 0.2 were considered as significant.

694

695 **Testing for pathogenicity of potential pathobiota**

696 *Isolation and characterization of strains belonging to the* Pseudomonas syringae *complex*

697 Strains belonging to the *P. syringae* complex were isolated to test the pathogenicity of
698 the potential pathobiota identified with the metagenomics approach. As previously described,
699 both *P. syringae sensu stricto* and *P. viridiflava* belong to the same *P. syringae* phylogenetic
700 complex (Berge et al. 2014). For the latter reason, in this study, we considered that the *P.*
701 *viridiflava* strains are part of the same species complex of strains falling under the name of *P.*
702 *syringae*.

703 Two/three plants per populations were collected to isolate strains in the *P. syringae*
704 complex from both leaves and roots. Plants for bacterial isolation were transferred in
705 sterilized bags and placed at 4°C before processing. Roots were cut from the rosette and

706 washed after transferring into the laboratory. Both leaves and roots were placed into sterilized
707 Eppendorf tubes containing 500ul of distilled sterilized water. Leaves were homogenized with
708 a scalpel and roots were placed into a bath sonicator for 10 min to allow the release of the
709 endophytic bacteria. Suspensions from leaves and roots were then diluted and placed on
710 Trypticase Soy Agar (TSA) medium implemented with 100 mg/L of ciclohexemide as
711 described previously (Bartoli et al. 2014). Plates were incubated at 24°C for two days and
712 pure bacterial colonies were tested for the cytochrome *C* oxidase test. All cytochrome *C*
713 oxidase negative colonies were stored at -20°C in 30% of glycerol and screened by PCR
714 amplification for the *P. syringae* marker (Psy) as described in Guilbaud et al. (2016). This
715 marker allows the identification of strains belonging to the *P. syringae* complex with a high
716 sensitivity. Strains positive for the Psy PCR were sequenced for the housekeeping gene citrate
717 synthase (*cts*) as described in Morris et al. (2007). To place the *A. thaliana* strains in the
718 diversity of the *P. syringae* complex, phylogenetic analysis was performed. The *cts* sequences
719 were trimmed and concatenated with DAMBE version 5.1.1 (Xia 2013) and MEGA6 was
720 used to infer the phylogeny by following a maximum likelihood model (Tamura et al. 2013).
721 Sequences for the *cts* genes are available in Supplementary Data Set 3. Reference *P. syringae*
722 sequences previously published (Berge et al. 2014) were also used in the phylogeny to allow
723 the identification of the strains isolated from *A. thaliana*.

724 Information about the strains is available in Supplementary Data Set 3. Strains are
725 stored and maintained in the bacterial collection of the Laboratory of Plant Microbes
726 Interaction (LIPM) of INRA (Toulouse, France) and available under reasonable request to the
727 corresponding author.

728

729

730

Phenotyping and statistical analyses

In order to test for pathogenicity of natural strains belonging to the *Pseudomonas syringae* complex, we first evaluated *in planta* bacterial growth over seven days of two *P. viridiflava* strains (0114-Psy-NAUV-BLand 0124-Psy-SAUB-AL) and two *P. syringae sensu stricto* strains (0132-Psy-BAZI-AL and 0143-Psy-THOM-AL) in the four corresponding local natural populations of *A. thaliana*, each represented by two randomly selected accessions (*At*-NAUV-B-7, *At*-NAUV-B-14, *At*-SAUB-A-3, *At*-SAUB-A-7, *At*-BAZI-A-1, *At*-BAZI-A-2, *At*-THOM-A-3 and *At*-THOM-A-6). A growth chamber experiment of 768 plants was set up at the Toulouse Plant Microbe Phenotyping Platform (TPMP) using a randomized complete block design (RCBD) with six experimental blocks. Each block was represented by 128 plants corresponding to the combination of four strains × eight accessions × four time points of scoring (0, 3, 5 and 7 days post-inoculation). After a 4-day stratification treatment, plants were grown at 22 °C under 90% humidity and artificial light to provide a 9-hr photoperiod as described in Huard-Chauveau et al. (2013). Bacterial infection was conducted on 22-day-old plants using a blunt-ended syringe (*Terumo® SYRINGE 1mL, SS+01T1)*. Three leaves per plant were infiltrated with 50µL of a 10^3 $CFUmL^{-1}$ bacterial solution. Plants were scored for bacterial growth by taking a hole-punch (Ø7 mm) from each infected leafand grinding the leaf discs in 100µL of Milli-Q water to release bacteria with glass bead in a 96-well plate (25 strokes per second, twice 30"). Appropriate serial dilutions, plating and calculation of the number of CFUs per cm² were performed according to (Bartoli et al. (2014). For each plant, the number of CFUs per cm² was obtained by calculating the median between the three leaf discs.

To explore natural variation of *in plant* bacterial growth, we used the following mixed model:

$$\log CFU_{ijklm} = \mu_{trait} + block_i + strain_j + population_k + time_l + strain_j \times population_k + strain_j \times time_l + population_k \times time_l + strain_j \times population_k \times time_l + accession_m(population_k) + strain_j \times time_l \times accession_m(population_k) + strain_j \times time_l \times accession_m(population_k) + accession_m(population_k) + \varepsilon_{ijklm} \quad (5)$$

where 'μ' is the overall phenotypic mean; 'block' accounts for differences in micro-environment among the six experimental blocks; 'strain' measures the effect of the four *Pseudomonas* strains; 'population' accounts for differences among the four *A. thaliana* populations; 'accession' measures the mean effect of accessions within each population; 'time' tests the evolution of bacterial growth over time; interaction terms involving the 'time' term account for variation among strains and populations for bacterial growth over time; 'ε' is the residual term. All factors were treated as fixed effects, except 'accession' which was treated as a random effect.

In a second growth chamber experiment, we evaluated the occurrence of disease and estimated the genetic variation of *A. thaliana* for response to natural *Pseudomonas syrngae* complex strains. For this purpose, we used four strains of *P. viridiflava* (0114-Psy-NAUV-BL, 0124-Psy-SAUB-AL, 0105-Psy-JACO-CLand0106-Psy-RADE-AL), four strains of *P. syringae sensu stricto* (0111-Psy-RAYR-BL, 0117-Psy-NAZA-AL, 0099-Psy-SIMO-AL and 0132-Psy-BAZI-AL) and eight corresponding local natural populations of *A. thaliana*, each represented by one randomly selected accession (*At*-NAUV-B-14, *At*-SAUB-A-3, *At*-JACO-C-5, *At*-RADE-A-6, *At*-RAYR-B-13, *At*-NAZA-A-2, *At*-SIMO-A-15 and *At*-BAZI-A-2) (Supplementary Table 1). A growth chamber experiment with 320 plants was set up at the Toulouse Plant Microbe Phenotyping Platform (TPMP) using a randomized complete block design (RCBD) with five experimental blocks. Each block was represented by 64 plants corresponding to the combination of eight strains × eight accessions. Growth chamber conditions were similar to the *in planta* bacterial growth experiment. Bacterial infection was

conducted on 28-day-old plants using a blunt-ended syringe (*Terumo® SYRINGE 1mL, SS+01T1)*. Three leaves per plant were entirely infiltrated with a 5.10^7 $CFUmL^{-1}$ bacterial solution. Disease symptoms were scored visually 1, 2 and 3 days after inoculation as described in Roux et al. (2010). Each infected leaf received a score from 0 to 1, with 0 corresponding to no symptoms and 0.5 and 1 corresponding to medium and severe symptoms, respectively. These scores categorize the percentage of leaf area infected, as determined by the presence of visible chlorosis, water soaking, or cell death. We averaged the scores for the three infected leaves per plant.

To explore natural variation of disease symptoms, we used the following mixed model:

$$\log CFU_{ijklm} = \mu_{trait} + block_i + strain_j + accession_m + time_l + strain_j \times accession_m + strain_j \times time_l + accession_m \times time_l + strain_j \times accession_m \times time_l + \varepsilon_{ijklm} \quad (6)$$

where ‘μ’ is the overall phenotypic mean; ‘block’ accounts for differences in micro-environment among the five experimental blocks; ‘strain’ measures the effect of the eight *Pseudomonas* strains; ‘accession’ accounts for differences among the eight *A. thaliana* accessions; ‘time’ tests the evolution of disease symptoms over time; interaction terms involving the ‘time’ term account for variation among strains and populations for disease evolution; ‘ε’ is the residual term. All factors were treated as fixed effects, except ‘accession’ which was treated as a random effect.

In models (5) and (6), the significance of terms of fixed and random effects was evaluated as described in the subsection ‘*Natural variation of microbiota and potential pathobiota*’. A Bonferroni correction for the number of tests was performed at a nominal level of 5%.

803 Hypersensitive Reaction (HR) on tobacco was also tested for all *P. syringae sensu*
804 *stricto* and *P. viridiflava* strains by infiltrating 20µL of 10^7 CFU mL^{-1} bacterial suspensions in
805 14-day-old *Nicotiana tabacum* plants. Inoculated tobacco plants were incubated at room
806 temperature for 24 hours before HR scoring (presence/absence). HR results are shown in
807 Supplementary Data Set 3.

808

809 **Data availability**

810 The raw FastQ reads for the metagenomic data were deposited in the Sequence Read
811 Archive (SRA) of NCBI https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP096011 under
812 the study number SRP096011. The trimmed matrix for both microbiota and potential
813 pathobiota are available in Supplementary Data Set 4 and Supplementary Data Set 5
814 respectively. Raw values for both α and β-diversity used for statistical analysis are available
815 in Supplementary Data Set 6.

816

ACKNOWLEDGEMENTS

We are grateful to Paul Schulze-Lefert, Stéphane Hacquard and Dominique Roby for their helpful discussions. We are also grateful to the staff of the LIPM greenhouse for their assistance during the growth chamber experiments. Special thanks are given to Joy Bergeslon and Benjamin Brachi for initial discussions on the project. This work was funded by the Région Midi-Pyrénées (CLIMARES project), the LABEX TULIP (ANR-10-LABX-41, ANR-11-IDEX-0002-02) and the Métaprogramme MEM (INRA, Metabar programme).

REFERENCES

Agler, M.T. et al., 2016. Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biol*, 14, p.e1002352.

Aleklett, K. et al., 2015. Wild plant species growing closely connected in a subalpine meadow host distinct root-associated bacterial communities. *PeerJ*, 3, p.e804.

Bai, Y. et al., 2015. Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature*, 528, pp.364–369.

Barret, M. et al., 2015. Emergence shapes the structure of the seed microbiota. *Appl Environ Microbiol*, 81, pp.1257–1266.

Bartoli, C. et al., 2014. A framework to gage the epidemic potential of plant pathogens in environmental reservoirs: the example of kiwifruit canker. *Mol Plant Pathol*, 16, pp.137-149.

Bartoli, C. et al., 2014. The *Pseudomonas viridiflava* phylogroups in the *P. syringae* complex are characterized by genetic variability and phenotypic plasticity of pathogenicity-related traits. *Environ Microbiol*, 16, pp.2301-2315.

Bartoli, C., Roux, F. & Lamichhane, J.R. 2016. Molecular mechanisms underlying the emergence of bacterial pathogens: an ecological perspective. *Mol. Plant. Pathol.* 16, pp. 860-869.

Baumler, A.J. & Sperandio, V., 2016. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature*, 535, pp.85–93.

Belkaid, Y. & Hand, T.W., 2014. Role of the microbiota in immunity and inflammation. *Cell*, 157, pp.121–141.

Benítez, M.-S. & Gardener, B.B.M., 2009. Linking sequence to function in soil bacteria: sequence-directed isolation of novel bacteria contributing to soilborne plant disease suppression. *Appl Environ Microbiol*, 75, pp.915–924.

Berg, G. et al., 2014. Unraveling the plant microbiome: looking back and future perspectives. *Front Microbiol*, 5, p.148.

Berge, O. et al., 2014. A user's guide to a data base of the diversity of *Pseudomonas syringae* and its application to classifying strains in this phylogenetic complex. *PloS One*, 9, p.e105547.

Bodenhausen, N. et al., 2014. A synthetic community approach reveals plant genotypes affecting the phyllosphere microbiota. *PLoS Genet*, 10, p.e1004283.

Bodenhausen, N., Horton, M.W. & Bergelson, J., 2013. Bacterial communities associated with the leaves and the roots of *Arabidopsis thaliana*. *PloS One*, 8, p.e56329.

Bono, L.M. et al., 2013. Competition and the origins of novelty: experimental evolution of niche-width expansion in a virus. *Biol Lett*, 9, p.20120616.

Bordenstein, S.R. & Theis, K.R., 2015. Host biology in light of the microbiome: Ten principles of holobionts and hologenomes. *PLoS Biol*, 13, pp.1–23.

Brachi, B. et al., 2013. Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in *Arabidopsis thaliana*. *Mol Ecol*, 22, pp.4222–4240.

Bulgarelli, D. et al., 2013. Structure and functions of the bacterial microbiota of plants. *Annu Rev Plant Biol*, 64, pp.807–838.

Bull, C.T. et al., 2010. Comprehensive list of names of plant pathogenic bacteria, 1980–2007. *J Plant Pathol*, 92, pp.551–592.

Bull, C.T. et al., 2012. List of new names of plant pathogenic bacteria (2008-2010). *J Plant Pathol*, 94, pp.21–27.

Bull, C.T. et al., 2014. List of new names of plant pathogenic bacteria (2011-2012). *J Plant Pathol*,, 96, pp.223–226.

Burke, C. et al., 2011. Bacterial community assembly based on functional genes rather than species. *PNAS*, 108, pp.14288–14293.

Cao, K.A.L., Boitard, S. & Besse, P., 2011. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *Bmc Bioinf*, 12, p.16.

Caporaso, J.G. et al., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *PNAS*, 108, pp.4516–4522.

Carrascal, L.M., Galván, I. & Gordo, O., 2009. Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos*, 118, pp.681–690.

Coleman-Derr, D. et al., 2016. Plant compartment and biogeography affect microbiome composition in cultivated and native *Agave* species. *New Phytol*, 209, pp.798–811.

Copeland, J.K. et al., 2015. Seasonal community succession of the phyllosphere microbiome. *Molecular plant-microbe interactions : MPMI*, 28, pp.274–285.

Eisenhauer, N. et al., 2013. Plant diversity effects on soil food webs are stronger than those of elevated CO2 and N deposition in a long-term grassland experiment. *PNAS*, 110, pp.6889–6894.

Geremia, R.A. et al., 2016. Contrasting microbial biogeographical patterns between anthropogenic subalpine grasslands and natural alpine grasslands. *New Phytol*, 209, pp.1196–1207.

Guilbaud, C. et al., 2016. Isolation and identification of *Pseudomonas syringae* facilitated by a PCR targeting the whole *P. syringae* group. *FEMS Microb Ecol*, 92, doi: 10.1093/femsec/fiv146.

Hacquard, S. et al., 2016. Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nature Comm*, 7, p.11362.

Haney, C.H. et al., 2015. Associations with rhizosphere bacteria can confer an adaptive advantage to plants. *Nature Plants*, 1, p.15051.

Horton, M.W. et al., 2014. Genome-wide association study of *Arabidopsis thaliana* leaf microbial community. *Nature Comm*, 5, p.5320.

Huard-Chauveau, C. et al., 2013. An atypical kinase under balancing selection confers broad-spectrum disease resistance in *Arabidopsis*. *PLoS Gen*, 9, p.e1003766.

Innerebner, G., Knief, C. & Vorholt, J.A., 2011. Protection of *Arabidopsis thaliana* against leaf-pathogenic *Pseudomonas syringae* by *Sphingomonas* strains in a controlled model system. *Appl Enviro Microbiol*, 77, pp.3202–3210.

Jakob, K. et al., 2002. *Pseudomonas viridiflava* and *P. syringae*--natural pathogens of Arabidopsis thaliana. *MPMI*, 15, pp.1195–1203.

Jari Oksanen, F. et al., 2009. vegan: community ecology package. R package version 1.16-0.

Kamada, N. et al., 2013. Control of pathogens and pathobionts by the gut microbiota. *Nat Immunol*, 14, pp.685–690.

Kawaguchi, M. & Minamisawa, K., 2010. Plant-microbe communications for symbiosis. *Plant Cell Physiol*, 51, pp.1377–1380.

Kinnunen, M. et al., 2016. A conceptual framework for invasion in microbial communities. *ISME J*, 10, pp.2773–2775.

Kniskern, J.M., Traw, M.B. & Bergelson, J., 2007. Salicylic acid and jasmonic acid signaling defense pathways reduce natural bacterial diversity on *Arabidopsis thaliana*. *MPMI*, 20, pp. 1512-1522.

Lê Cao, K.A. et al., 2008. Sparse PLS: Variable selection when integrating omics data. *Stat Appl Mol Biol*, 7, doi: 10.2202/1544-6115.

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

Lebeis, S.L. et al., 2015. Plant microbiome. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science*, 349, pp.860–864.

Legendre, P. & Gallagher, E.D., 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129, pp.271–280.

Lindström, E.S. & Langenheder, S., 2012. Local and regional factors influencing bacterial community assembly. *Environ Microbiology Rep*, 4, pp.1–9.

Lundberg, D.S. et al., 2012. Defining the core *Arabidopsis thaliana* root microbiome. *Nature*, 488, pp.86–90.

Maestre, F.T., 2004. On the importance of patch attributes, environmental factors and past human impacts as determinants of perennial plant species richness and diversity in mediterranean semiarid steppes. *Diver Distrib*, 10, pp.21–29.

Mahé, F. et al., 2015. Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ*, 3, p.e1420.

Maignien, L. et al., 2014. Ecological succession and stochastic variation in the assembly of *Arabidopsis thaliana* phyllosphere communities. *mBio*, 5, pp.e00682-13.

Mallon, C.A. et al., 2015. Resource pulses can alleviate the biodiversity–invasion relationship in soil microbial communities. *Ecology*, 96, pp.915–926.

939 Mansfield, J. et al., 2012. Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol*
940 *Plant Pathol*, 13, pp.614–629.

941 Marchetti, M. et al., 2010. Experimental evolution of a plant pathogen into a legume
942 symbiont. *PLoS Biol*, 8, p.e1000280.

943 Marino, S. et al., 2014. Mathematical modeling of primary succession of murine intestinal
944 microbiota. *PNAS*, 111, pp.439–44.

945 Markowitz, V.M. et al., 2014. IMG 4 version of the integrated microbial genomes
946 comparative analysis system. *Nucleic Acids Res*, 42, pp.D560-D567.

947 Mendes, R., Garbeva, P. & Raaijmakers, J.M., 2013. The rhizosphere microbiome:
948 significance of plant beneficial, plant pathogenic, and human pathogenic
949 microorganisms. *FEMS Microbiol Rev*, 37, pp.634–663.

950 Mitchell-Olds, T. & Schmitt, J., 2006. Genetic mechanisms and evolutionary significance of
951 natural variation in *Arabidopsis*. *Nature*, 441, pp.947–952.

952 Montesinos, A. et al., 2009. Demographic and genetic patterns of variation among populations
953 of *Arabidopsis thaliana* from contrasting native environments. *PloS one*, 4, p.e7213.

954 Morris, C.E. et al., 2007. Surprising niche for the plant pathogen *Pseudomonas syringae*.
955 *Infect. Genet. Evol*, 7, pp.84–92.

956 Mosca, A., Leclerc, M. & Hugot, J.P., 2016. Gut microbiota diversity and human diseases:
957 should we reintroduce key predators in our ecosystem? *Front Microbiol*, 7, p.455.

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

960 Picó, F.X., 2012. Demographic fate of *Arabidopsis thaliana* cohorts of autumn- and spring-
961 germinated plants along an altitudinal gradient. *J Ecol*, 100, pp.1009–1018.

962 Poudel, R. et al., 2016. Microbiome networks: a systems framework for identifying candidate
963 microbial assemblages for disease management. *Phytopathology*, 106, pp.1083–1096.

964 Ramette, A., 2007. Multivariate analyses in microbial ecology. *FEMS Microbiol Ecol*, 62,
965 pp.142–160.

966 Ritpitakphong, U. et al., 2016. The microbiome of the leaf surface of *Arabidopsis* protects
967 against a fungal pathogen. *New Phytol*, 210, pp.1033–1043.

968 Roux, F. & Bergelson, J., 2016. Chapter Four – The genetics underlying natural variation in
969 the biotic interactions of *Arabidopsis thaliana* : the challenges of linking evolutionary
970 genetics and community ecology. *Curr Top Dev Biol*. pp. 111–156.

971 Roux, F., Gao, L. & Bergelson, J., 2010. Impact of initial pathogen density on resistance and
972 tolerance in a polymorphic disease resistance gene system in *Arabidopsis thaliana*.
973 *Genetics*, 185, pp.283–291.

974 Santhanam, R. et al., 2015. Native root-associated bacteria rescue a plant from a sudden-wilt
975 disease that emerged during continuous cropping. *PNAS*, 112, pp.E5013-E5020.

976 Schlaeppi, K. et al., 2014. Quantitative divergence of the bacterial root microbiota in

977 Arabidopsis thaliana relatives. *PNAS*, 111, pp.585–592.

978 Schloss, P.D. et al., 2009. Introducing mothur: open-source, platform-independent,
979 community-supported software for describing and comparing microbial communities.
980 *Applied Environ Microbiol*, 75, pp.7537–7541.

981 Tamura, K. et al., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol*
982 *Biol Evol*, 30, pp. 2725-2729.

983 Vandenkoornhuyse, P. et al., 2015. The importance of the microbiome of the plant holobiont.
984 *The New Phytol*, 206, pp.1196–1206.

985 Varghese, N.J. et al., 2015. Microbial species delineation using whole genome sequences.
986 *Nucleic Acids Res*, 43, pp.6761–6771.

987 Vayssier-Taussat, M. et al., 2014. Shifting the paradigm from pathogens to pathobiome: new
988 concepts in the light of meta-omics. *Front Cell Infect Microbiol*, 4, p.29.

989 Vorholt, J.A., 2012. Microbial life in the phyllosphere. *Nat Rev Micro*, 10, pp.828–840.

990 Wagner, M.R. et al., 2016. Host genotype and age shape the leaf and root microbiomes of a
991 wild perennial plant. *Nature Comm*, 7, p.12151.

992 Xia, X., 2013. DAMBE5: a comprehensive software package for data analysis in molecular
993 biology and evolution. *Mol Biol Evol*, 30, pp.1720–1728.

994 Youle, M. et al., 2013. Superorganisms and holobionts. *Microbe Magazine*, 8, pp.152–153.

995

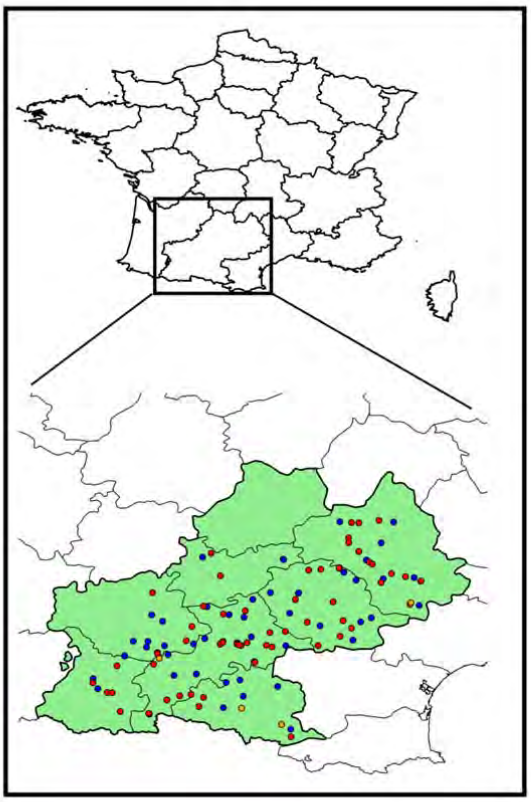


Figure 1. Location of the 163 *A. thaliana* populations across the region Midi-Pyrénées (southwest of France). Blue dots represent the 80 populations collected in both autumn and spring, red dots represent the populations collected in spring only and orange dots represent the 4 populations collected in autumn only. The map clearly shows that the range of sampling across the Midi-Pyrénées region was homogeneous during the two sampling seasons.

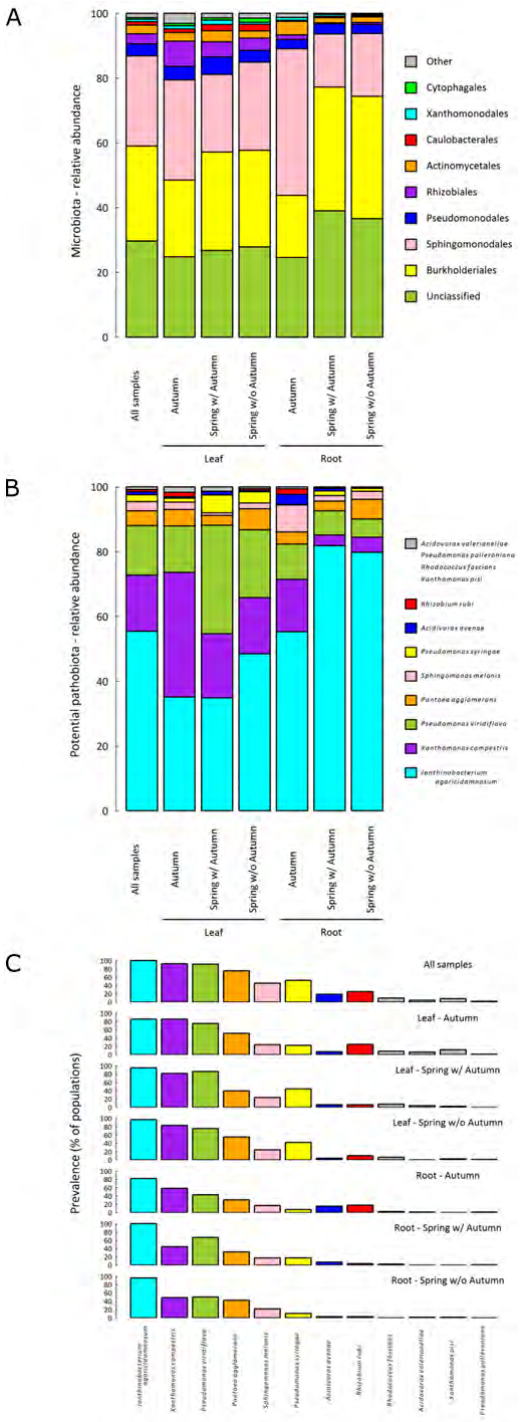


Figure 2. Stacked bar plots of the relative abundances of the major bacterial phyla and bacterial species for both microbiota and potential pathobiota of the *A. thaliana* populations collected in the Midi-Pyrénées region. 'All samples' n = 165; Leaf: 'Autumn' n= 314, 'Spring w/ Autumn' n = 245, 'Spring wo/ Autumn' n = 262; Root: 'Autumn' n= 309, 'Spring w/ Autumn' n = 267, 'Spring wo/ Autumn' n = 258. A) Stacked bar plots representing the relative abundance for the ten most abundant bacterial orders. B) Stacked bar plots representing the relative abundance of the twelve most abundant bacterial species for the potential pathobiota. C) Prevalence (number of populations in which the potential pathogen species was detected) of the twelve bacterial species characterizing the potential pathobiota.

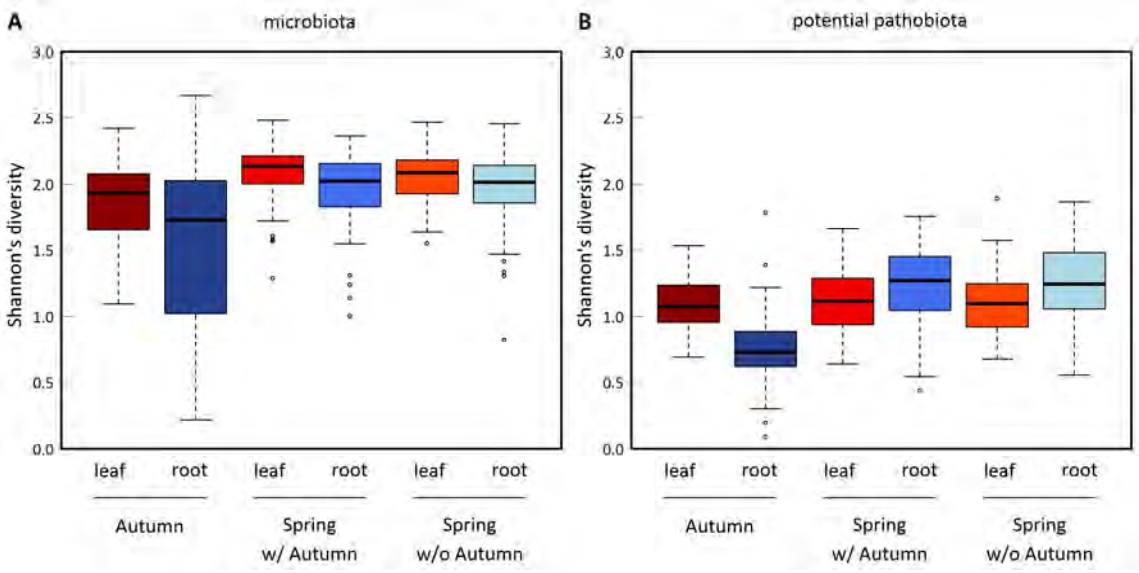
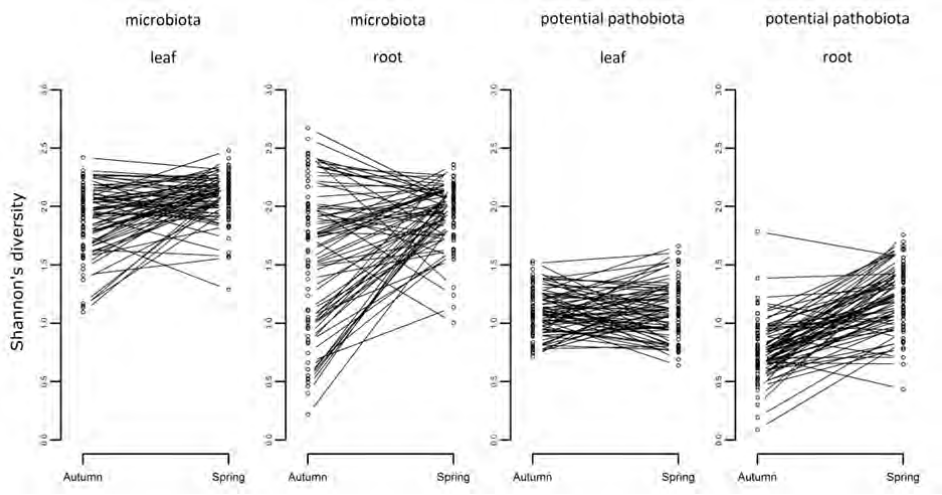


Figure 3. Bar plots representing the seasonal variation of the α-diversity inferred as Shannon's index for both A) microbiota and B) potential pathobiota for the 163 natural populations of *A. thaliana* collected in the Midi-Pyrénées region. Leaf and root samples are represented with a red and blue color scale, respectively. Each bar plot corresponds to distribution of the mean α-diversity (estimated as BLUPs) per population. Microbiota and pathobiota: 'Autumn – Leaf' n= 82, 'Autumn – Root' n = 78, 'Spring w/ Autumn - Leaf' n= 77, 'Spring w/ Autumn - Root' n= 80, 'Spring w/o Autumn - Leaf' n= 79, 'Spring w/o Autumn - Root' n= 79.



1022

1023 **Figure 4**.Variation among populations in the evolution of α-diversity between autumn and spring. Each dot corresponds to
1024 the mean α-diversity (estimated as BLUPs) of a population. 'leaf' n = 75 populations, 'root' n = 74 populations.

1025

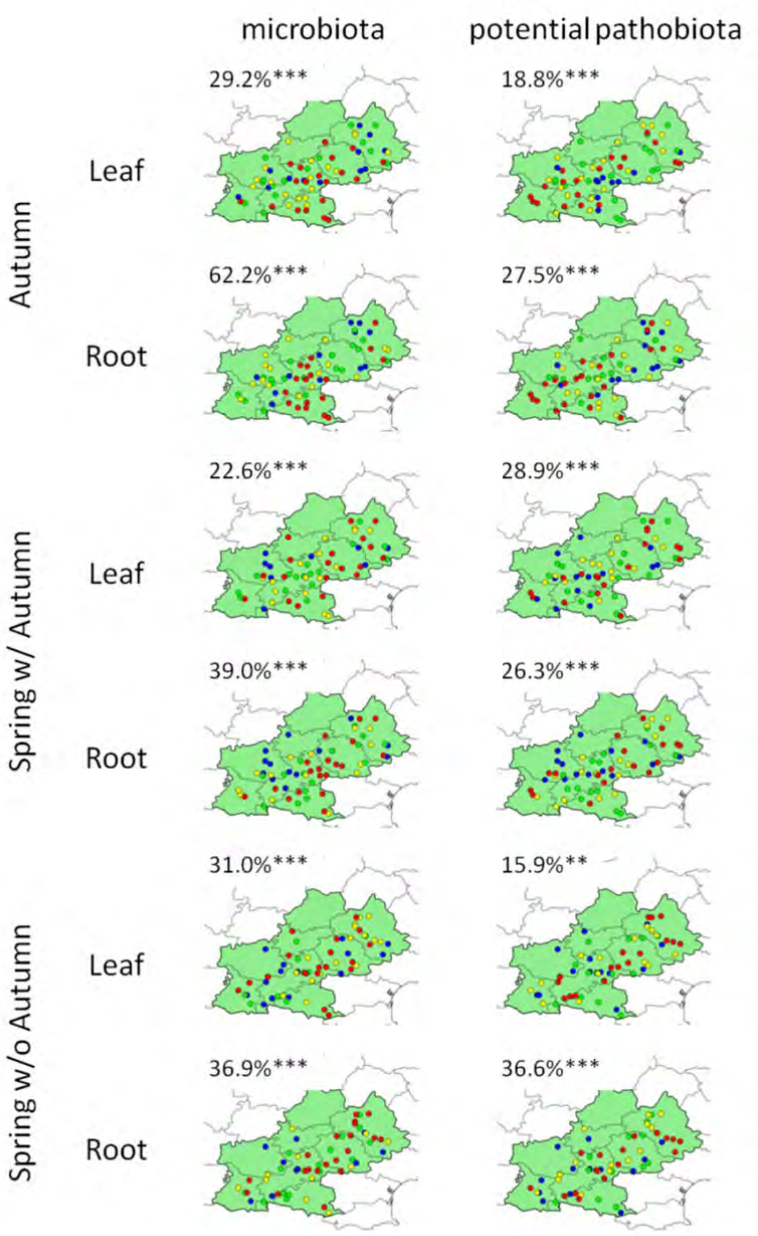


Figure 5. Geographic variation of α-diversity (inferred as Shannon's index) for microbiota and potential pathobiota. For each 'season x plant compartment' combination, blue, green, yellow and red dots correspond to populations from the 1st, 2nd, 3rd and 4th quartiles of the α-diversity (estimated as BLUPs) distribution. Values indicate the percentage of α-diversity variance among populations. Significance after a Bonferroni correction at a nominal level of 5%: ** $0.01 > P > 0.001$,, *** $P < 0.001$. Microbiota and pathobiota: 'Autumn – Leaf' n= 82, 'Autumn – Root' n = 78, 'Spring w/ Autumn - Leaf' n= 77, 'Spring w/ Autumn - Root' n= 80, 'Spring w/o Autumn - Leaf' n= 79, 'Spring w/o Autumn - Root' n= 79.

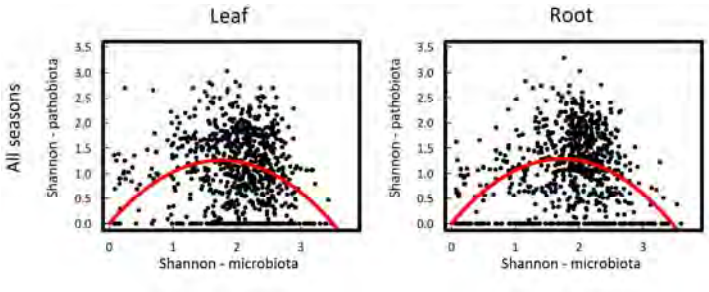
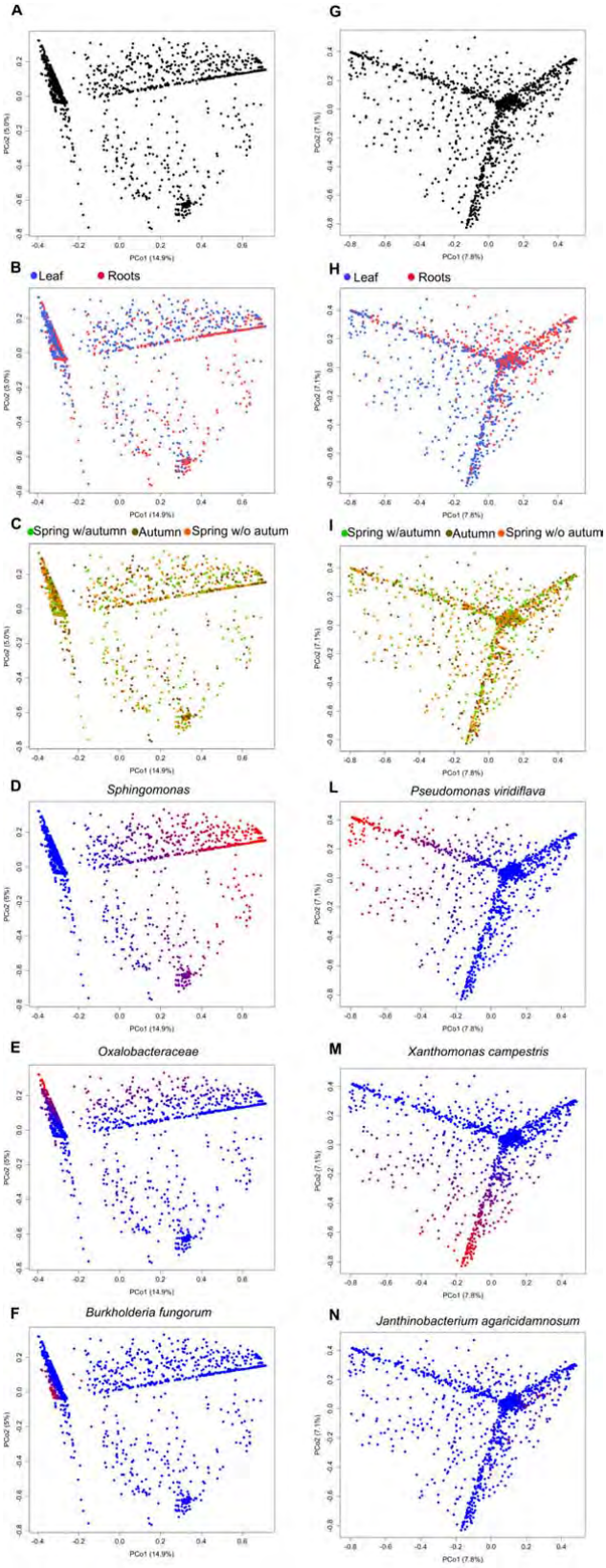


Figure6. Correlation between α-diversity of potential pathobiota and α-diversity of microbiota for the leaf and root compartments considering all samples. α-diversity was inferred as Shannon's index. Leaf: n = 821, root: n = 834. The red lines indicate a significant quadratic relationship, according to the following non-linear model: pathobiota-α-diversity ~ k x microbiota-α-diversity– q x microbiota-α-diversity.



1041

1042 **Figure 7.** Bacterial composition and structure of *A. thaliana* illustrated by Principal coordinates (PCoA) plots of Hellinger
1043 dissimilarity matrices for microbiota (left panels) and potential pathobiota (right panels). N = 1655 samples. **A**) PCoA plot of
1044 microbiota for all samples indicates that bacterial composition is mainly structured along two axes. **B**) Significant difference

of the microbiota composition between leaves and roots (PCo1: $P < 0.001$, PCo2: $P < 0.001$). **C**) No significant difference of the microbiota composition between the three categories of populations, namely 'autumn' populations, 'spring w/ autumn' populations and 'spring w/o autumn' populations. **D**), **E**) and **F**) Relationships between the microbiota composition and the abundance of the three most abundant OTUs, i.e. *Sphingomonas sp.*, *Burkholderia fungorum* and *Oxalobacteraceae unclassified sp.*, respectively. Red-to-blue color gradient indicates high-to-low OTU abundance. While the abundance of *Sphingomonas sp.* mainly drives the composition of the *A. thaliana* microbiota along the first PCoA axis, the abundance of *Oxalobacteraceae unclassified sp.* and *B. fungorum* partially drive the composition of the *A. thaliana* microbiota along the second PCoA axis. **G**) PCoA plot of potential pathobiota for all samples indicates that bacterial composition is mainly structured along three axes. **H)**Significant difference of the potential pathobiota composition between leaves and roots (PCo1: $P < 0.001$, PCo2: $P < 0.001$). **I**) No significant difference of the potential pathobiota composition between the three categories of populations, namely 'autumn' populations, 'spring w/ autumn' populations and 'spring w/o autumn' populations. **L**), **M**) and **N**) Relationships between the potential pathobiota composition and the abundance of the three most abundant potential pathogenic OTUs, i.e. *Janthinobacterium agaricidamnosum*, *Pseudomonas viridiflava* and *Xanthomonas campestris*, respectively. Red-to-blue color gradient indicates high-to-low OTU abundance. The abundance of *P. viridiflava* and *X. campestris* mainly drive the composition of the potential pathobiota along the first and second PCoA axis, respectively. The abundance of *J. agaricidamnosum* partially drives the composition of the potential pathobiota along the second PCoA axis.

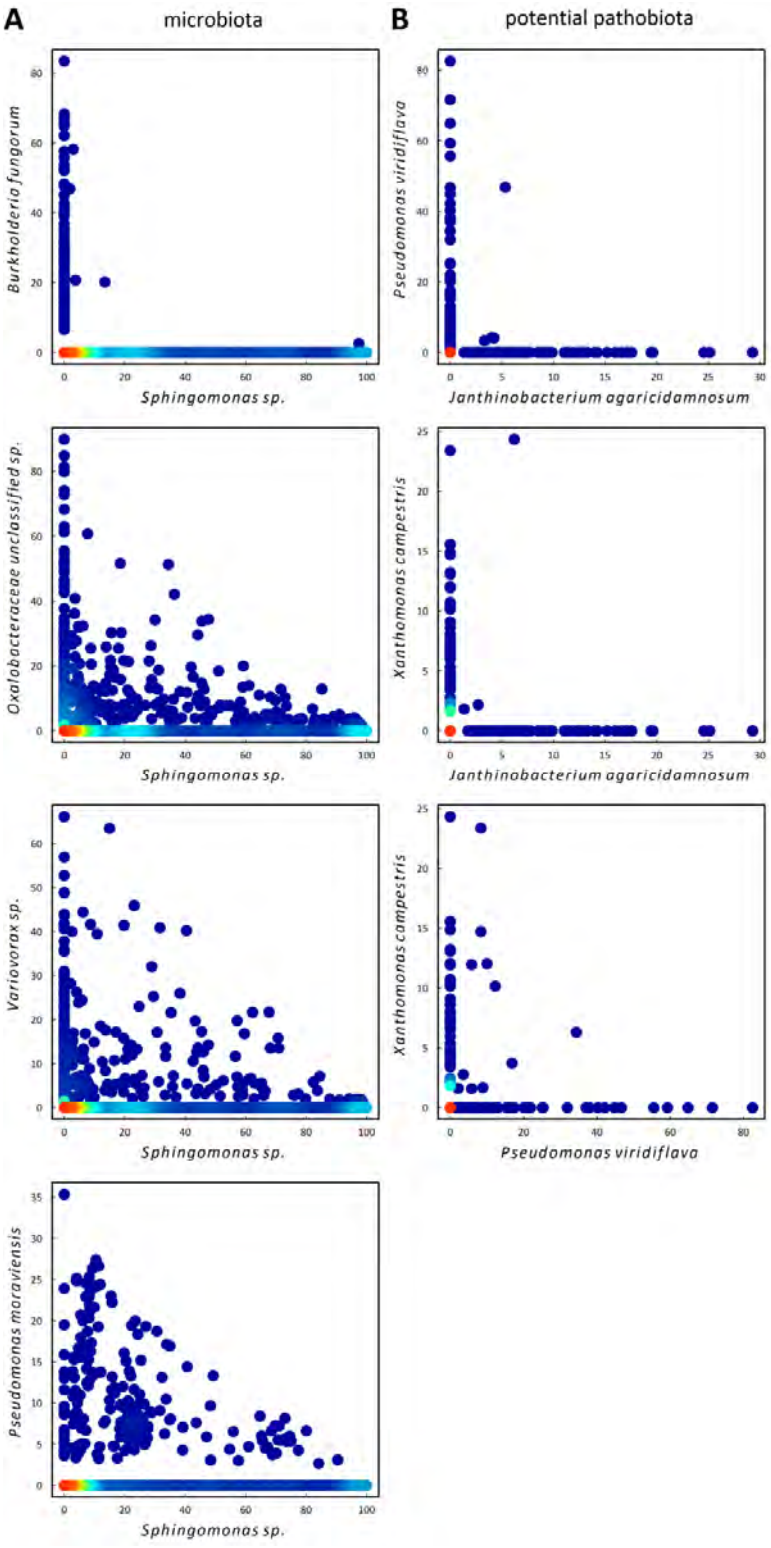
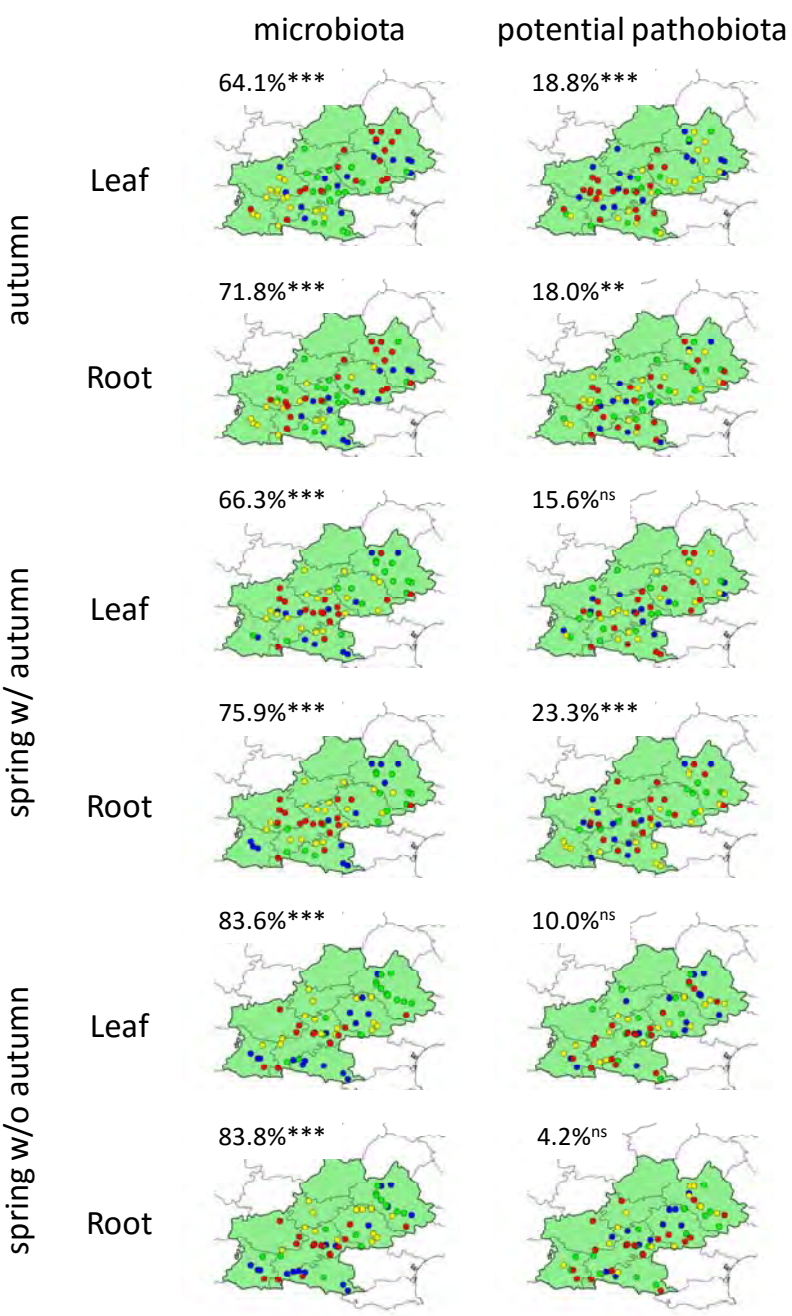


Figure 8. 2D density plots of the relative abundance between the most abundant OTUs. N = 1655 samples. **A**) microbiota with the relative abundance of the 2nd (*Burkholderia fungorum*), 3rd (*Oxalobacteraceae unclassified sp.*), 4th (*Variovorax sp.*) and 5th (*Pseudomonas moraviensis*) most abundant OTUs plotted against the relative abundance of the 1st most abundant OUT (*Sphingomonas sp.*).**B**) potential pathobiota with the relative abundance of the three most abundant potential pathogen species (*Janthinobacteriumagaracidamnosum*, *Pseudomonas viridiflava*and *Xanthomonascampestris*)plotted against each other. Red-to-blue color gradient represents high-to-low density gradient.



1072

1073 **Figure 9.** Geographic variation of β-diversity illustrated by Principal Coordinates (PCoA, first axis) plots for microbiota and
1074 potential pathobiota. For each 'season x plant compartment' combination, blue, green, yellow and red dots correspond to
1075 populations from the 1st, 2nd, 3rd and 4th quartiles of the PCOA1 (estimated as BLUPs) distribution. Values indicate the
1076 percentage of PCOA1 variance among populations. Significance after a Bonferroni correction at a nominal level of 5%: **
1077 $0.01>P>0.001$, *** $P < 0.001$, ns non-significant. Microbiota: 'Autumn – Leaf' n= 82, 'Autumn – Root' n = 82, 'Spring w/
1078 Autumn - Leaf' n= 78, 'Spring w/ Autumn - Root' n= 80, 'Spring w/o Autumn - Leaf' n= 79, 'Spring w/o Autumn - Root' n=
1079 79. Pathobiota: 'Autumn – Leaf' n= 82, 'Autumn – Root' n = 75, 'Spring w/ Autumn - Leaf' n= 77, 'Spring w/ Autumn -
1080 Root' n= 80, 'Spring w/o Autumn - Leaf' n= 79, 'Spring w/o Autumn - Root' n= 78.

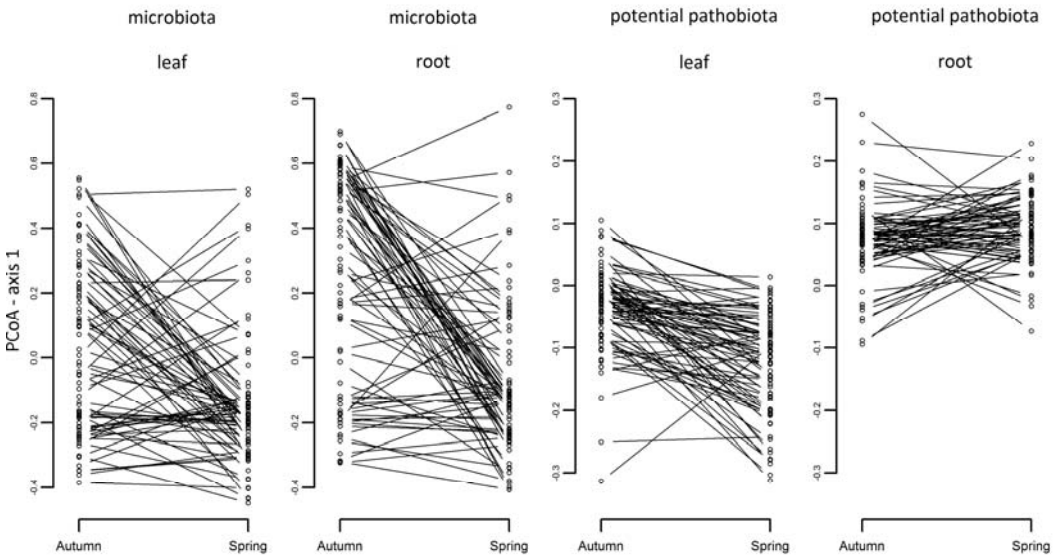
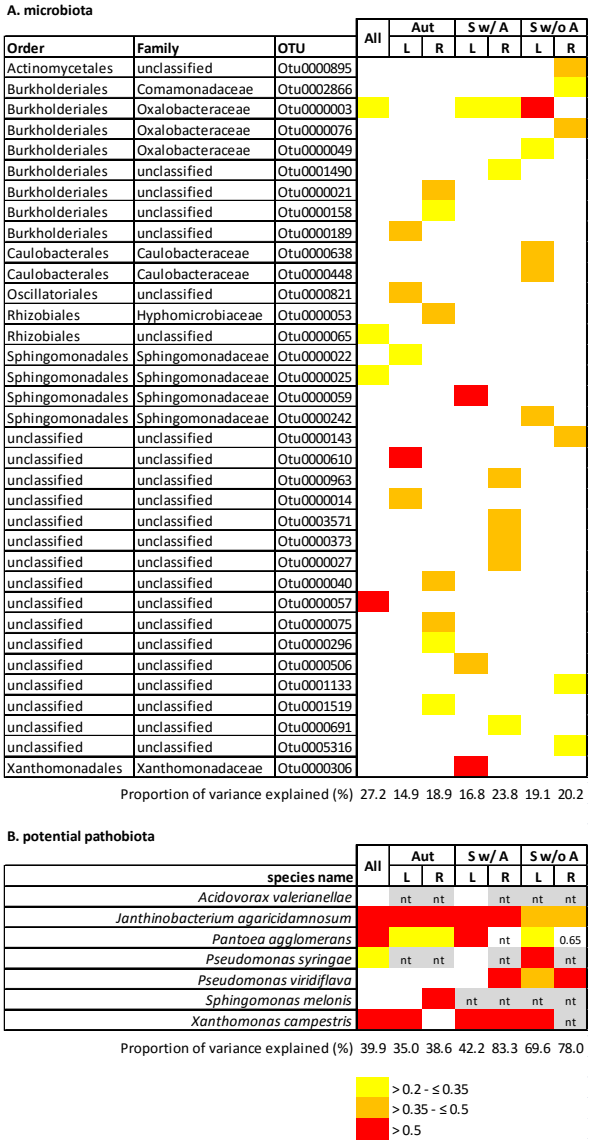


Figure 10. Variation among populations in the evolution of β-diversity (PCoA, first axis) between autumn and spring. Each dot corresponds to the mean β-diversity(estimated as BLUPs) of a population. Microbiota: 'leaf' n = 76 populations, 'root' n = 78 populations. Pathobiota : 'leaf' n = 75 populations, 'root' n = 71 populations.



1088

1089

1090 **Figure 11**.Relationships between microbiota β-diversity and potential pathobiota β-diversity. A sparse Partial Least Square
1091 Regression (sPLSR) (Carrascal et al. 2009, Le Cao et al. 2011) was adopted to maximize the covariance between linear
1092 combinations of OTUs from microbiota (matrix X) and linear combinations of OTUs from potential pathobiota (matrix Y).
1093 sPLSR was run using the mixOmics package implemented in the *R* environment (Le Cao et al. 2008, Le Cao et al. 2011). For
1094 the microbiota matrix, only OTUs with a relative abundance above 1% were considered. For the potential pathobiota matrix,
1095 we considered the seven most abundant OTUs. In addition to the lasso model, sPLSR results were validated by plotting the
1096 Root Mean Square Error of Prediction (Maestre 2004, Le Cao et al. 2008). For the microbiota, we calculated the final
1097 loadings for the ten OTUs with the highest initial loadings for each component. Given the small number of OTUs in the
1098 potential pathobiota (n = 7), the initial loading of each OTU was kept for each component. The color gradient indicates the
1099 strength of the loading values for both microbiota and potential pathobiota. Only OTUs (from the microbiota or the potential
1100 pathobiota) with a loading value above 0.2 were considered as significant. 'All', 'Aut', 'Sw/ A' and 'Sw/o A' stand for all
1101 samples and samples from the three categories of populations (i.e. 'autumn' populations, 'spring w/ autumn' populations and
1102 'spring w/o autumn' populations),'L' and 'R' stand for leaf and root, respectively. 'nt' indicates that the corresponding OTU
1103 was absent in the potential pathobiota. . 'All' n = 165; 'Aut - L' n= 314, 'Sw/ A - L' n = 245, 'Sw/o A - L' n = 262; 'Aut – R'
1104 n= 309, 'Sw/ A - R' n = 267, 'Sw/o A – R' n = 258.

1105

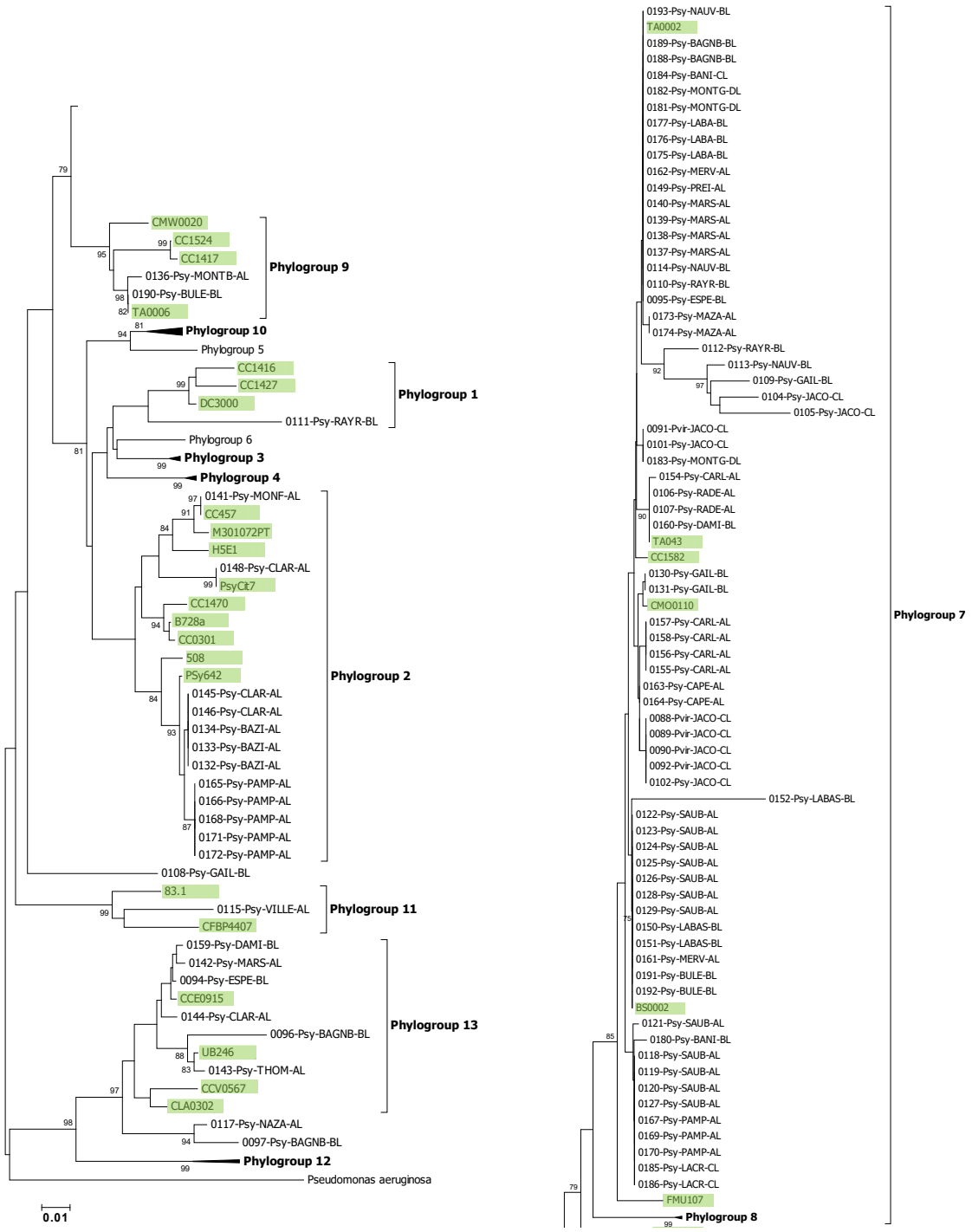


Figure 12. Phylogenetic tree inferred with a Neighbor Joining (NJ) model (3000 bootstrap repetitions) and based on the *cts* sequences (350bp) of the 97 strains isolated from the 163 *A. thaliana* populations. Bootstrap values are shown at each node and names of the strain at each branches. All the strains belong to the *P. syringae* complex and are mainly distributed in the phylogroups 7, 9, 1, 2, 11 and 13. Reference *P. syringae* strains representative of the 13 phylogroups are labeled in green. Phylogroup affiliation was based on the previous work from Berge et al., (2014).

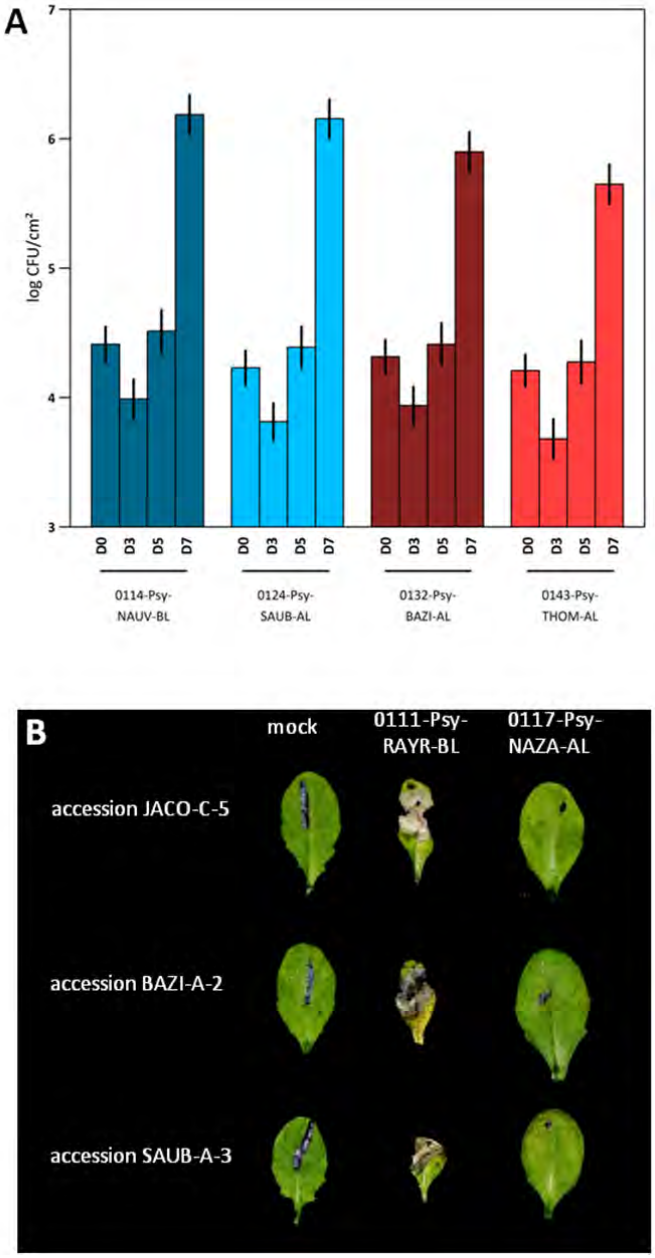


Figure 13. Pathogenicity of *Pseudomonas sp.* strains isolated from the 163 *A. thaliana* populations of the region Midi-Pyrénées. **A**) Mean bacterial growth across eight accessions collected in the region Midi-Pyrénées for two strains of *P. viridiflava* (0114-Psy-NAUV-BL and 0124-Psy-SAUB-AL) and two strains of *P. syringae sensu stricto* (0132-Psy-BAZI-AL and 0143-Psy-THOM-AL).**B**) Illustration of symptoms observed three days after inoculation for two strains of *P. syringae*. Presence and absence of symptoms was observed on each of the eight accessions collected in the region Midi-Pyrénées for the strains 0111-Psy-RAYRB-BL and 0117-Psy-NAZA-AL, respectively. D0, D3, D5 and D7 indicate the number of days after inoculation.

Supplementary Files

Supplementary Table 1. Names and GPS coordinates (expressed in degrees) of the 169 *A. thaliana* populations

Supplementary Table 2. Natural variation for microbiota and potential pathobiota in natural populations of *A. thaliana* collected both in autumn and spring.

Supplementary Table 3. Percentage of variance of microbiota and potential pathobiota explained by the terms 'seasons', 'plant compartment', 'population' and their interactions in natural populations of *A. thaliana* collected both in autumn and spring.

Supplementary Table 4. Natural variation for microbiota and potential pathobiota in all the natural populations of *A. thaliana* collected in spring.

Supplementary Table 5. Natural variation for microbiota and potential pathobiota in the natural populations of *A. thaliana* collected in 'autumn'.

Supplementary Table 6. Natural variation for microbiota and potential pathobiota in the natural populations of *A. thaliana* collected in 'spring with autumn'.

Supplementary Table 7. Natural variation for microbiota and potential pathobiota in the natural populations of *A. thaliana* collected in 'spring without autumn'.

Supplementary Table 8. Percentage of variance of microbiota and pathobiota explained by the terms 'population', 'plant compartment' and their interactions in natural populations of *A. thaliana* collected either in autumn or in spring.

Supplementary Table 9. Natural variation for microbiota and potential pathobiota for each 'season x plant compartment' combination.

Supplementary Table 10. Model selection among one linear model and one non-linear model on the relationship between the α-diversities of microbiota and potential pathobiota.

Supplementary Table 11. Fitting of a non-linear model on the relationship between the α-diversities of microbiota and potential pathobiota.

Supplementary Table 12. *In planta* bacterial growth of four natural *Pseudomonas strains* in four corresponding local natural populations of *A. thaliana*, each represented by two accessions.

Supplementary Table 13. Natural interactions between eight local natural *A. thaliana* accessions and eight corresponding local natural *Pseudomonas syringae* strains for disease symptom evolution.

Supplementary Figure 1. Percentage of variance of the β-diversity among populations.

Supplementary Figure 2. Genetic variation among five accessions from the region Midi-Pyrénées for the response to *P. viridiflava* strains.

Supplementary Figure 3. Heatmap for the symptoms observed three days after inoculation with *P. syringae* strains. The heatmap shows the interactions between eight natural accessions of *A. thaliana* from the region Midi-Pyrénées and eight natural strains belonging to the *P. syringae* complex collected in the same populations than the eight natural accessions.

Supplementary Figure 4. PCoA perfomed on a Hellinger distance matrix based on rarefied data (top panel) and on a Jaccard similarity coefficient matrix distance (bottom panel).

Supplementary Data Set 1. Validation of the *gyrB* gene marker.

Supplementary Data Set 2. List of pathogenic bacteria used to determine the potential pathobiota of the 163 *A. thaliana* populations.

Supplementary Data Set 3. Strains belonging to the *P. syringae* complex isolated from the 163 *A. thaliana* populations for the validation of the potential pathobiota

Supplementary Data Set 4. Abundance matrix obtained after data trimming for the whole microbiota

Supplementary Data Set 5. Presence/Absence matrix obtained after data trimming for the potential pathobiota

Supplementary Data Set 6. Raw values for both α and β diversity used in the statistical analysis

Supplementary figures and tables

In situ relationships between microbiota and potential pathobiota in *Arabidopsis thaliana*

Supplementary Table 1. Names and GPS coordinates (expressed in degrees) of the 169 populations.

Population name	Locality	Latitude	Longitude
AMBR-A	Ambres	43.733229	1.823869
ANGE-A	Saint Angel, Salvagnac	43.911999	1.656649
ANGE-B	Saint Angel, Salvagnac	43.91214	1.656855
AULO-A	Aulon	43.190552	0.815774
AURE-B	Aureville	43.477976	1.452214
AUZE-A	Auzeville	43.527792	1.491628
AXLE-A	Ax les Thermes	42.724197	1.834034
AXLE-B	Ax les Thermes	42.724588	1.833497
BACC-B	Baccarets (Cintegabelle)	43.312225	1.515167
BACC-C	Baccarets (Cintegabelle)	43.311868	1.515459
BACC-D	Baccarets (Cintegabelle)	43.311866	1.515623
BACC-E	Baccarets (Cintegabelle)	43.31187	1.515709
BACC-F	Baccarets (Cintegabelle)	43.311926	1.515463
BAGNB-A	Bagnères de Bigore	43.075729	0.151764
BAGNB-B	Bagnères de Bigore	43.076454	0.151533
BANI-A	Banios	43.042867	0.234732
BANI-B	Banios	43.043644	0.234303
BANI-C	Banios	43.043644	0.234303
BARA-B	Baraqueville	44.269727	2.426322
BARA-C	Baraqueville	44.270842	2.427551
BARC-A	Barcugnan	43.362044	0.387723
BARR-A	Barry le Cas (Caylus)	44.202421	1.767492
BAZI-A	Baziège	43.453602	1.620674
BELC-A	Belcastel	44.387532	2.336117
BELC-B	Belcastel	44.387527	2.336782
BELC-C	Belcastel	44.389212	2.336636
BELL-A	Belleserre	43.790307	1.106456
BERNA-A	Bernac-dessus	43.16215	0.111398
BESS-A	Bessuéjouls	44.526359	2.730092
BOULO-A	Boulogne-sur-Gesse	43.28908	0.639795
BROU-A	Brousse-le-château	43.999349	2.621684
BROU-B	Brousse-le-château	44.033129	2.638672
BROU-C	Brousse-le-château	44.03326	2.638683
BULA-A	Bulan	43.039803	0.277297
BULE-B	Buleix (Soulan)	42.91058	1.248122
CAMA-C	Camarès	43.824878	2.881661
CAMA-D	Camarès	43.823736	2.881003
CAMA-E	Camarès	43.824878	2.881661
CAPE-A	Lacappelle - Ségalar	44.108545	1.990168
CARL-A	Carla-bayle	43.151102	1.3923
CASS-A	Cassagne-Begontes	44.17653	2.518164
CAST-A	Castelginset	43.698534	1.427856
CASTI-A	Castillon en Cousserans	42.920498	1.034063
CAZA-B	Cazaux-Fréchet	42.831484	0.420091
CEPE-A	Cepet	43.755183	1.435978
CERN-A	Saint-Rome-de-Cernon	44.01194	2.966488
CERN-B	Saint-Rome-de-Cernon	44.014684	2.967927
CHEI-A	Chein-dessus	43.013708	0.86707
CIER-A	Cier sur Luchon	42.85332	0.602039

Supplementary Table 1. (continued)

Population name	Locality	Latitude	Longitude
CIER-B	Cier de Luchon	42.859978	0.600413
CIER-C	Cier de Luchon	42.860166	0.601088
CINT-A	Cintegabelle	43.305466	1.520441
CINT-B	Cintegabelle	43.305611	1.520735
CLAR-A	Saint Clar-de-Rivière	43.464776	1.219019
CLAR-B	Saint Clar-de-Rivière	43.465281	1.218577
CLAR-C	Saint Clar-de-Rivière	43.464058	1.21799
COLO-A	Colombiès	44.346915	2.340243
COLO-B	Colombiès	44.34773	2.339715
COLO-C	Colombiès	44.34806	2.339698
COMT-A	Villecomtal	44.540652	2.602245
CRAN-A	Cransac (Aubin)	44.529845	2.260486
DAMI-A	Damiatte	43.654515	1.977636
DAMI-B	Damiatte	43.654515	1.977636
DAMI-C	Damiatte	43.654515	1.977636
DECA-A	Châteaude Cas (Espinas)	44.199896	1.77189
DIEU-A	Ville-Dieu-du-temple	44.059797	1.220975
ESPE-B	Esperausses	43.693335	2.534582
FAYA-A	Fayet	43.8021	2.951709
FERR-A	Ferrières	43.657743	2.44371
GAIL-A	Gaillac	43.908928	1.900574
GAIL-B	Gaillac	43.909032	1.901077
GREZ-A	Grézian	42.876896	0.349714
JACO-A	Jacoy (Boussenac)	42.905839	1.406513
JACO-B	Jacoy (Boussenac)	42.905839	1.406513
JACO-C	Jacoy (Boussenac)	42.905839	1.406513
JULI-A	Saint Julien de Malmont (St Cyprien de Dourdou)	44.522606	2.36351
JUZE-A	Juzes	43.448838	1.79053
JUZET-A	Juzet d'Izaut	42.977713	0.756373
JUZET-B	Juzet d'Izaut	42.977354	0.755498
JUZET-C	Juzet d'Izaut	42.977354	0.755498
LABA-A	Labarthe-sur-Lèze	43.45155	1.400498
LABA-B	Labarthe-sur-Lèze	43.450892	1.40116
LABA-C	Labarthe-sur-Lèze	43.451451	1.39935
LABA-D	Labarthe-sur-Lèze	43.458019	1.381137
LABAS-A	La bastide de Sérou	43.00844	1.420039
LABAS-B	La bastide de Sérou	43.008716	1.420053
LABR-A	Labruguière	43.531185	2.262591
LACR-A	Lacraste (Montgauch)	42.999869	1.075659
LACR-C	Lacraste (Montgauch)	43.000155	1.075624
LAGR-A	Lagraulhet St Nicolas	43.795323	1.073752
LAMA-A	Lamasquère	43.487424	1.243559
LAMA-B	Lamasquère	43.479745	1.241592
LANT-B	Lanta	43.564943	1.65239
LANT-C	Lanta	43.564822	1.65201
LANT-D	Lanta	43.564822	1.65201
LAUZ-A	Lauzerte	44.25608	1.140526
LECT-A	Lectoure	43.911721	0.629745
LECT-B	Lectoure	43.911721	0.629745

Supplementary Table 1. (continued)

Population name	Locality	Latitude	Longitude
LESP-A	Les pujols	43.094237	1.719981
LOUB-A	Loubens-Lauragais	43.574273	1.786038
LOUB-B	Loubens-Lauragais	43.574647	1.785723
LUNA-A	Lunax	43.339706	0.689839
LUZE-A	Luzenac (Garanou)	42.764683	1.752959
LUZE-B	Luzenac (Garanou)	42.764419	1.753595
LUZE-D	Luzenac (Garanou)	42.764419	1.753595
LUZE-E	Luzenac (Garanou)	42.764683	1.752959
MARS-A	Glaciane (Marsans)	43.662542	0.718265
MARS-B	Glaciane (Marsans)	43.662542	0.718265
MART-A	Martres Tolosane	43.202147	1.010976
MASS-A	Masseube	43.437536	0.579271
MAZA-A	Mazamet	43.497754	2.375372
MEDA-A	Saint Medard	43.490485	0.461439
MERE-A	Merens-les-Vals	42.656618	1.836221
MERE-B	Merens-les-Vals	42.656546	1.836175
MERV-A	Merville	43.720426	1.296824
MERV-B	Merville	43.725141	1.247629
MONB-A	Monblanc	43.46529	0.986273
MONE-A	Monestiès	44.115354	2.094725
MONF-A	Monferran-Savès	43.616254	0.972435
MONT-A	Montans	43.852212	1.87432
MONT-B	Montans	43.852723	1.873536
MONTB-A	Montbrun bocage	43.130495	1.269927
MONTG-B	Montgaillard	43.12729	0.110681
MONTG-D	Montgaillard	43.127713	0.110633
MONTI-A	Montiès	43.389383	0.67282
MONTI-B	Montiès	43.3839336	0.67257
MONTI-D	Montiès	43.3839336	0.67257
MONTM-A	Montmajou (Cier de Luchon)	42.86156	0.595943
MONTM-B	Montmajou (Cier de Luchon)	42.861218	0.596869
MOUL-A	Moularès	44.089762	2.296094
NAUV-A	Nauviale	44.520751	2.427404
NAUV-B	Nauviale	44.520418	2.427129
NAUV-C	Nauviale	44.520397	2.42721
NAYR-A	Le Nayrac (Cassagnes-Bégontes)	44.161368	2.544711
NAZA-A	Saint-Pierre-de-Najac (Miramont de Quercy)	44.220329	1.064953
PAMP-A	Pampelonne	44.124864	2.255514
PAMP-B	Pampelonne	44.124876	2.255184
PANA-C	Villefrance de Panat	44.078884	2.711136
PASD-B	Pas du loup (Camarès)	43.811758	2.871661
PREI-A	Preignan	43.717856	0.623298
PUYM-B	Puymaurin	43.372913	0.765694
RADE-A	Sainte Radegonde	44.345163	2.620821
RAYR-A	Rayret (Cassagne-Begontes)	44.196005	2.493157
RAYR-B	Rayret (Cassagne-Begontes)	44.196006	2.493076
REAL-A	Réalmont	43.83165	2.20155
ROME-A	Saint-Rome-du-tarn	44.041553	2.909576
ROQU-B	Roquecourbe	43.667907	2.290214

Supplementary Table 1. (continued)

Population name	Locality	Latitude	Longitude
SALE-A	Saleich	43.024966	0.965965
SALV-A	Saint-Salvy-de-la-Balme	43.602578	2.36338
SAMA-A	Samatan	43.494325	0.92391
SAUB-A	Saubens	43.464914	1.365136
SAUB-B	Saubens	43.474107	1.364175
SAUB-C	Saubens	43.475583	1.367589
SAUR-A	Saurat	42.889844	1.485209
SEIS-A	Seissan	43.487302	0.588798
SIMO-A	Simorre	43.449392	0.734601
SORE-A	Sorèze	43.452628	2.072476
TARN-C	Villemur-sur-Tarn	43.85328	1.502009
THOM-A	Saint Thomas	43.513975	1.082859
VALE-A	Valence d'Albiegeois	44.022296	2.403434
VICT-B	Saint Victor et Melvieu	44.052243	2.834023
VICT-C	Saint Victor et Melvieu	44.052243	2.834023
VIEL-A	Vielmur sur Agout	43.623801	2.089616
VILLA-A	Villate	43.458174	1.380951
VILLE-A	Villenouvelle	43.439784	1.671
VILLE-B	Villenouvelle	43.440342	1.669595
VILLE-C	Villenouvelle	43.440024	1.670111
VILLE-D	Villenouvelle	43.439733	1.670712
VILLEM-A	Villembits	43.273815	0.321238

Supplementary Table 2. Natural variation for microbiota and potential pathobiota in natural populations of *A. thaliana* collected both in Autumn and Spring.

	microbiota								potential pathobiota					
	richness		Shannon		PCo1		PCo2		Shannon		PCo1		PCo2	
Model terms	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*
season	1.9	0.1681	1.4	0.2416	0.7	0.4042	0.4	0.5468	0.0	0.9803	0.8	0.3609	6.5	0.0113
plant compartment	0.8	0.3739	29.1	**<.0001**	67.7	**<.0001**	29.0	**<.0001**	4.5	0.0376	155.5	**<.0001**	62.3	**<.0001**
season*plant compartment	60.0	**<.0001**	5.7	0.0198	21.4	**<.0001**	2.7	0.1080	33.4	**<.0001**	29.7	**<.0001**	5.0	0.0263
population	0.0	1.0000	0.5	0.4795	1.4	0.2367	1.2	0.2733	0.0	1.0000	1.2	0.2733	6.5	0.0108
*season*population*	79.9	**<.0001**	26.6	**<.0001**	109.8	**<.0001**	84.2	**<.0001**	9.7	**0.0018**	18.8	**<.0001**	0.8	0.3711
*plant compartment*population*	1.0	0.3173	0.3	0.5839	0.8	0.3711	0.0	1.0000	2.8	0.0943	0.3	0.5839	1.0	0.3173
*season*plant compartment*population*	4.9	0.0269	7.4	**0.0065**	8.3	**0.0040**	13.2	**0.0003**	1.4	0.2367	0.0	1.0000	0.0	1.0000
sampling date (season)	16.4	**<.0001**	10.2	**<.0001**	12.7	**<.0001**	19.1	**<.0001**	12.5	**<.0001**	3.3	0.0383	1.2	0.3167
diameter(season)	1.0	0.3607	0.3	0.7173	2.4	0.0959	0.2	0.8649	0.6	0.5629	2.3	0.0980	0.0	0.9763
leaf number(season)	3.9	0.0208	0.2	0.8325	1.4	0.2424	0.2	0.8375	1.8	0.1735	2.0	0.1390	0.5	0.6331
obs	1690.3	**<.0001**	0.7	0.4061	4.3	0.0375	25.4	**<.0001**	34.2	**<.0001**	0.5	0.5029	1.8	0.1858

Each trait was modeled separately using a mixed model. Model random terms (in italics) were tested with likelihood ratio tests (LRT) of models with and without these effects. A Bonferroni correction for the number of tests was performed for each modeled effect (i.e., per line) at a nominal level of 5%. Bold values indicate statistically significant results after correction for multiple comparisons. 'obs', total number of observations.

Supplementary Table 3. Percentage of variance of microbiota and potential pathobiota explained by the terms 'seasons', 'plant compartment', 'population' and their interactions in natural populations of *A. thaliana* collected both in Autumn and Spring.

	microbiota				potential pathobiota		
Model terms	richness	Shannon	PCo1	PCo2	Shannon	PCo1	PCo2
season	0.0	0.0	0.0	0.0	0.0	0.0	0.0
plant compartment	0.0	**5.0**	**5.9**	**3.4**	0.0	**19.2**	**11.0**
*season*plant compartment*	**4.7**	*0.9*	**2.1**	0.4	**4.4**	**3.6**	*0.6*
population	0.0	3.5	6.5	7.8	0.0	2.9	*8.2*
*season*population*	**37.7**	**25.2**	**51.5**	**44.0**	**14.1**	**12.6**	1.3
*plant compartment*population*	1.8	1.8	1.5	0.7	6.4	0.8	2.2
*season*plant compartment*population*	*5.5*	**9.8**	**4.7**	**8.1**	4.8	0.0	0.0
error	50.3	53.8	27.8	35.7	70.2	60.9	76.8

Bold values indicate statistically significant results after a Bonferroni correction for multiple comparisons (see Supplementary Table 2). Italic values indicate statistically significant results before a Bonferroni correction for multiple comparisons (see Supplementary Table 2).

Supplementary Table 4. Natural variation for microbiota and potential pathobiota in all the natural populations of *A. thaliana* collected in Spring.

	microbiota								potential pathobiota					
	richness		Shannon		PCo1		PCo2		Shannon		PCo1		PCo2	
Model terms	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*
w/wo_Autumn	0.2	0.6457	0.4	0.5178	1.3	0.2514	2.0	0.1626	0.3	0.5900	2.5	0.1164	3.4	0.0690
plant compartment	13.1	**0.0004**	13.4	**0.0003**	44.2	**<.0001**	48.9	**<.0001**	8.8	**0.0036**	173.1	**<.0001**	44.3	**<.0001**
w/wo_Autumn*plant compartment	7.4	**0.0072**	0.2	0.6447	0.2	0.6664	0.9	0.3391	0.0	0.9628	12.4	**0.0006**	0.8	0.3806
population(w/wo_Autumn)	72.0	**<.0001**	32.5	**<.0001**	244.9	**<.0001**	220.8	**<.0001**	30.7	**<.0001**	19.0	**<.0001**	12.2	**0.0005**
*plant compartment*population(w/wo_Autumn)*	42.0	**<.0001**	9.3	**0.0023**	11.1	**0.0009**	23.0	**<.0001**	3.4	0.0652	0.2	0.6547	0.6	0.4386
sampling date	23.0	**<.0001**	3.3	0.0725	46.2	**<.0001**	51.0	**<.0001**	25.7	**<.0001**	22.4	**<.0001**	0.9	0.3502
diameter	0.4	0.5200	4.1	0.0432	1.7	0.1994	0.0	0.8611	1.4	0.2407	0.8	0.3872	0.0	0.8614
leaf number	0.7	0.4085	3.2	0.0750	4.4	0.0371	0.0	0.9212	0.0	0.8477	2.1	0.1500	0.1	0.7944
obs	1478.5	**<.0001**	2.6	0.1043	26.9	**<.0001**	16.7	**<.0001**	46.1	**<.0001**	0.3	0.5728	2.1	0.1442

Each trait was modeled separately using a mixed model. Model random terms (in italics) were tested with likelihood ratio tests (LRT) of models with and without these effects. A Bonferroni correction for the number of tests was performed for each modeled effect (i.e., per line) at a nominal level of 5%. Bold values indicate statistically significant results after correction for multiple comparisons. 'obs', total number of observations.

Supplementary Table 5. Natural variation for microbiota and potential pathobiota in the natural populations of *A. thaliana* collected in 'Autumn'.

	microbiota								potential pathobiota					
	richness		Shannon		PCo1		PCo2		Shannon		PCo1		PCo2	
Model terms	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*
plant compartment	25.2	**<.0001**	18.5	**<.0001**	60.1	**<.0001**	7.2	0.0091	32.1	**<.0001**	37.1	**<.0001**	50.8	**<.0001**
population	29.2	**<.0001**	21.5	**<.0001**	72.7	**<.0001**	29.6	**<.0001**	8.1	**0.0044**	28.7	**<.0001**	9.1	**0.0026**
*plant compartment*population*	9.0	**0.0027**	21.0	**<.0001**	17.6	**<.0001**	18.9	**<.0001**	7.1	**0.0077**	0.0	1.0000	0.0	1.0000
sampling date	25.3	**<.0001**	12.0	**0.0009**	3.1	0.0818	36.6	**<.0001**	12.3	**0.0007**	0.0	0.9022	2.9	0.0924
diameter	0.0	0.9009	0.3	0.5654	2.2	0.1385	0.2	0.6529	0.2	0.6973	5.8	0.0162	0.0	0.9328
leaf number	1.5	0.2291	0.1	0.8083	1.3	0.2506	0.2	0.6997	1.1	0.2963	0.9	0.3328	0.3	0.5863
obs	975.6	**<.0001**	0.7	0.3996	0.7	0.4010	16.5	**<.0001**	21.0	**<.0001**	2.2	0.1357	1.6	0.2015

Each trait was modeled separately using a mixed model. Model random terms (in italics) were tested with likelihood ratio tests (LRT) of models with and without these effects. A Bonferroni correction for the number of tests was performed for each modeled effect (i.e., per line) at a nominal level of 5%. Bold values indicate statistically significant results after correction for multiple comparisons. 'obs', total number of observations.

Supplementary Table 6. Natural variation for microbiota and potential pathobiota in the natural populations of *A. thaliana* collected in 'Spring with Autumn'.

	microbiota								potential pathobiota					
	richness		Shannon		PCo1		PCo2		Shannon		PCo1		PCo2	
Model terms	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*
plant compartment	25.6	**<.0001**	9.5	**0.0029**	18.2	**<.0001**	26.9	**<.0001**	3.7	0.0574	123.9	**<.0001**	24.8	**<.0001**
population	46.6	**<.0001**	21.0	**<.0001**	98.7	**<.0001**	90.1	**<.0001**	5.8	0.0160	14.0	**0.0002**	5.5	0.0190
*plant compartment*population*	8.1	**0.0044**	1.7	0.1923	7.1	0.0077	10.4	**0.0013**	7.5	**0.0062**	0.3	0.5839	0.9	0.3428
sampling date	12.6	**0.0006**	4.8	0.0313	25.6	**<.0001**	11.2	**0.0013**	10.4	**0.0019**	8.3	**0.0051**	0.0	0.8435
diameter	1.7	0.1920	0.9	0.3477	1.3	0.2474	0.2	0.6262	1.0	0.3114	0.0	0.9528	0.0	0.9540
leaf number	5.1	0.0249	0.5	0.4683	1.2	0.2705	0.1	0.8287	3.0	0.0845	3.0	0.0854	0.0	0.9170
obs	787.2	**<.0001**	0.2	0.6356	10.1	**0.0016**	10.6	**0.0012**	14.4	**0.0002**	0.4	0.5441	0.4	0.5341

Each trait was modeled separately using a mixed model. Model random terms (in italics) were tested with likelihood ratio tests (LRT) of models with and without these effects. A Bonferroni correction for the number of tests was performed for each modeled effect (i.e., per line) at a nominal level of 5%. Bold values indicate statistically significant results after correction for multiple comparisons. 'obs', total number of observations.

Supplementary Table 7. Natural variation for microbiota and potential pathobiota in the natural populations of *A. thaliana* collected in 'Spring without Autumn'.

	microbiota								potential pathobiota					
	richness		Shannon		PCo1		PCo2		Shannon		PCo1		PCo2	
Model terms	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*
plant compartment	0.4	0.5470	5.1	0.0265	27.9	**<.0001**	22.4	**<.0001**	6.6	0.0108	49.2	**<.0001**	17.8	**<.0001**
population	27.9	**<.0001**	12.9	**0.0003**	153.9	**<.0001**	125.3	**<.0001**	27.5	**<.0001**	4.4	0.0359	6.9	0.0086
*plant compartment*population*	39.7	**<.0001**	9.2	**0.0024**	3.6	0.0578	13.2	**0.0003**	0.0	1.0000	0.0	1.0000	0.0	1.0000
sampling date	11.5	**0.0011**	0.2	0.6854	20.4	**<.0001**	44.9	**<.0001**	15.2	**0.0002**	15.0	**0.0002**	2.7	0.1066
diameter	12.5	**0.0004**	5.5	0.0199	0.3	0.5905	1.3	0.2562	0.6	0.4526	3.1	0.0817	0.3	0.5829
leaf number	4.9	0.0267	5.3	0.0219	5.7	0.0171	0.1	0.7734	3.2	0.0753	0.2	0.6786	0.2	0.6494
obs	741.2	**<.0001**	3.5	0.0625	18.1	**<.0001**	6.0	0.0151	35.8	**<.0001**	0.0	0.9756	2.6	0.1076

Each trait was modeled separately using a mixed model. Model random terms (in italics) were tested with likelihood ratio tests (LRT) of models with and without these effects. A Bonferroni correction for the number of tests was performed for each modeled effect (i.e., per line) at a nominal level of 5%. Bold values indicate statistically significant results after correction for multiple comparisons. 'obs', total number of observations.

Supplementary Table 8. Percentage of variance of microbiota and pathobiota explained by the terms 'population', 'plant compartment' and their interactions in natural populations of *A. thaliana* collected either in Autumn or in Spring.

		microbiota				potential pathobiota		
Season	Model terms	richness	Shannon	PCo1	PCo2	Shannon	PCo1	PCo2
Autumn	*population*	**24.4**	**30.3**	**54.3**	**34.1**	**12.3**	**19.0**	**9.2**
	plant compartment	**6.1**	**6.7**	**10.9**	*1.9*	**11.5**	**9.9**	**14.5**
	*plant compartment * population*	**7.4**	**15.5**	**6.8**	**13.5**	**9.1**	0.0	0.0
	error	62.1	47.5	28.0	50.5	67.0	71.1	76.3
Spring w/ Autumn	*population*	**40.2**	**24.2**	**65.8**	**62.7**	*13.8*	**12.8**	*10.2*
	plant compartment	**6.6**	**2.7**	**2.8**	**5.0**	1.3	**31.1**	**9.0**
	*plant compartment * population*	**7.9**	6.2	*4.2*	**5.5**	**14.3**	2.0	4.4
	error	45.3	66.9	27.3	26.8	70.6	54.1	76.5
Spring w/o Autumn	*population*	**37.4**	**20.5**	**81.0**	**69.7**	**25.8**	*6.4*	*10.9*
	plant compartment	0.0	*1.6*	**1.8**	**2.1**	*1.4*	**15.0**	**5.8**
	*plant compartment * population*	**20.1**	**13.3**	1.5	**4.0**	0.0	0.7	0.0
	error	42.5	64.6	15.7	24.1	72.8	77.8	83.3

Bold values indicate statistically significant results after a Bonferroni correction for multiple comparisons (see Supplementary Tables 4, 5 and 6). Italic values indicate statistically significant results before a Bonferroni correction for multiple comparisons (see Supplementary Tables 4, 5 and 6).

Supplementary Table 9. Natural variation for microbiota and potential pathobiota for each 'season x plant compartment' combination.

			microbiota								potential pathobiota					
			richness		Shannon		PCo1		PCo2		Shannon		PCo1		PCo2	
Season	Compartment	Model terms	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*	*F* or *LRT*	*P*
Autumn	root	*population*	43.5	**<.0001**	115.5	**<.0001**	196.6	**<.0001**	109.9	**<.0001**	30.6	**<.0001**	9.3	**0.0023**	3.3	0.0693
		sampling date	27.9	**<.0001**	8.4	**0.0050**	3.4	0.0700	27.0	**<.0001**	22.7	**<.0001**	0.2	0.6765	3.9	0.0526
		diameter	0.1	0.7254	0.7	0.4075	0.5	0.4791	0.0	0.8551	0.7	0.4069	1.6	0.2094	0.0	0.8907
		leaf number	2.6	0.1115	0.1	0.8004	0.1	0.7052	0.3	0.6112	3.1	0.0821	0.0	0.8606	0.1	0.7085
		obs	425.1	**<.0001**	6.8	0.0099	2.2	0.1400	8.7	**0.0034**	19.4	**<.0001**	1.4	0.2451	0.0	0.8430
Autumn	leaf	*population*	70.5	**<.0001**	26.7	**<.0001**	179.2	**<.0001**	48.3	**<.0001**	13.8	**0.0002**	14.7	**0.0001**	6.5	0.0108
		sampling date	12.4	**0.0007**	12.6	**0.0007**	2.9	0.0899	34.5	**<.0001**	1.1	0.3070	0.0	0.8703	0.7	0.3962
		diameter	0.0	0.8351	1.3	0.2654	0.4	0.5072	1.5	0.2156	0.0	0.8775	3.4	0.0683	0.0	0.9926
		leaf number	0.3	0.6027	0.3	0.5998	0.0	0.8446	0.9	0.3405	0.0	0.9318	1.2	0.2745	0.6	0.4598
		obs	524.9	**<.0001**	0.2	0.6857	2.5	0.1188	11.4	**0.0008**	5.8	0.0164	0.4	0.5247	3.0	0.0842
Spring	root	*population*	228.8	**<.0001**	71.9	**<.0001**	425.0	**<.0001**	498.1	**<.0001**	53.9	**<.0001**	12.4	**0.0004**	7.8	**0.0052**
		sampling date	20.2	**<.0001**	3.0	0.0868	47.9	**<.0001**	59.0	**<.0001**	21.6	**<.0001**	29.8	**<.0001**	0.2	0.7014
		diameter	0.9	0.3571	1.0	0.3105	0.1	0.8291	0.8	0.3686	2.4	0.1207	0.7	0.4202	0.0	0.8387
		leaf number	5.8	0.0163	2.5	0.1129	3.3	0.0707	0.0	0.8869	1.0	0.3222	0.8	0.3863	3.6	0.0580
		obs	562.7	**<.0001**	1.4	0.2330	3.2	0.0742	1.3	0.2595	13.1	**0.0003**	0.5	0.4641	0.1	0.8203
Spring	leaf	*population*	142.9	**<.0001**	35.7	**<.0001**	341.0	**<.0001**	318.2	**<.0001**	24.9	**<.0001**	11.9	**0.0006**	8.4	**0.0038**
		sampling date	16.8	**<.0001**	1.7	0.1915	39.6	**<.0001**	37.4	**<.0001**	15.4	**0.0001**	8.7	**0.0038**	0.7	0.4121
		diameter	0.0	0.8950	4.1	0.0430	2.2	0.1430	0.0	0.8601	0.1	0.7329	0.2	0.6481	0.0	0.9786
		leaf number	1.3	0.2468	1.4	0.2349	3.3	0.0718	1.8	0.1773	0.9	0.3336	5.2	0.0232	0.9	0.3526
		obs	707.5	**<.0001**	4.4	0.0356	14.8	**0.0001**	21.7	**<.0001**	18.4	**<.0001**	0.0	0.9483	3.0	0.0866
Spring w/ Autumn	root	*population*	107.6	**<.0001**	39.7	**<.0001**	185.4	**<.0001**	239.5	**<.0001**	19.4	**<.0001**	14.8	**0.0001**	4.1	0.0429
		sampling date	9.4	**0.0029**	3.7	0.0579	28.2	**<.0001**	14.9	**0.0002**	6.8	0.0112	14.9	**0.0002**	1.2	0.2705
		diameter	0.7	0.4027	0.7	0.3988	0.0	0.8609	0.2	0.6240	0.6	0.4237	0.5	0.4691	0.0	0.8442
		leaf number	9.2	**0.0026**	3.9	0.0504	1.5	0.2279	0.0	0.8657	5.0	0.0266	0.4	0.5158	1.1	0.2918
		obs	271.7	**<.0001**	0.6	0.4420	0.8	0.3741	0.9	0.3337	6.5	0.0114	1.0	0.3140	0.1	0.8316
Spring w/ Autumn	leaf	*population*	59.9	**<.0001**	12.8	**0.0003**	128.3	**<.0001**	116.3	**<.0001**	20.3	**<.0001**	6.0	0.0143	6.1	0.0135
		sampling date	12.0	**0.0009**	3.6	0.0632	19.5	**<.0001**	6.4	0.0136	6.5	0.0131	3.0	0.0872	0.1	0.8140
		diameter	0.4	0.5209	1.0	0.3258	2.4	0.1257	1.0	0.3143	0.5	0.4630	0.6	0.4448	0.0	0.8657
		leaf number	0.6	0.4504	0.4	0.5406	0.7	0.3952	0.0	0.9426	0.1	0.7527	5.8	0.0173	0.2	0.6689
		obs	383.4	**<.0001**	1.8	0.1841	4.7	0.0313	12.8	**0.0004**	5.5	0.0202	0.0	0.8569	1.6	0.2019
Spring w/o Autumn	root	*population*	140.1	**<.0001**	30.7	**<.0001**	240.9	**<.0001**	236.4	**<.0001**	33.3	**<.0001**	0.8	0.3711	3.1	0.0783
		sampling date	11.8	**0.0010**	0.4	0.5350	20.3	**<.0001**	49.2	**<.0001**	16.3	**0.0001**	14.6	**0.0003**	3.3	0.0730
		diameter	11.0	**0.0010**	0.3	0.5752	0.1	0.7254	0.5	0.4671	3.7	0.0548	0.1	0.7923	1.7	0.1936
		leaf number	0.5	0.4860	0.0	0.9410	2.1	0.1509	0.1	0.7126	1.4	0.2361	0.7	0.3909	1.4	0.2389
		obs	322.2	**<.0001**	0.8	0.3645	3.3	0.0694	0.3	0.6168	8.1	**0.0047**	0.1	0.7457	0.4	0.5236
Spring w/o Autumn	leaf	*population*	77.6	**<.0001**	29.6	**<.0001**	236.9	**<.0001**	195.6	**<.0001**	7.4	**0.0065**	3.1	0.0783	2.5	0.1138
		sampling date	6.7	0.0118	0.0	0.9764	19.9	**<.0001**	39.4	**<.0001**	7.8	**0.0066**	6.6	0.0125	1.0	0.3111
		diameter	2.4	0.1218	7.6	**0.0064**	0.2	0.7010	1.3	0.2496	0.1	0.7321	3.1	0.0786	0.0	0.9310
		leaf number	6.8	0.0096	9.5	**0.0023**	4.2	0.0417	3.5	0.0613	2.7	0.0999	1.0	0.3091	2.1	0.1523
		obs	342.6	**<.0001**	3.2	0.0760	11.6	**0.0008**	11.0	**0.0011**	12.9	**0.0004**	0.1	0.7934	1.4	0.2313

Each trait was modeled separately using a mixed model. Model random terms (in italics) were tested with likelihood ratio tests (LRT) of models with and without these effects. A Bonferroni correction for the number of tests was performed for each modeled effect (i.e., per line) at a nominal level of 5%. Bold values indicate statistically significant results after correction for multiple comparisons. 'obs', total number of observations.

Supplementary Table 10. Model selection among one linear model and one non-linear model on the relationship between the α-diversities of microbiota and potential pathobiota. For each 'plant compartment x samples' combination, different letters indicate different fits between the two models, with the lowest values indicating a better model fit.

		lm		nlm	
Compartment	Samples	AIC	BIC	AIC	BIC
Leaf	All seasons	1765.01 [a]	1779.10 [a]	1719.28 [b]	1733.37 [b]
	Autumn	638.80 [a]	649.99 [a]	630.37 [b]	641.56 [b]
	Spring w/ Autumn	525.99 [a]	536.45 [a]	500.81 [b]	511.26 [b]
	Spring w/o Autumn	605.06 [a]	615.74 [a]	594.83 [b]	605.51 [b]
Root	All seasons	1760.85 [a]	1774.87 [a]	1625.52 [b]	1639.54 [b]
	Autumn	576.79 [a]	587.64 [a]	538.04 [b]	548.89 [b]
	Spring w/ Autumn	574.49 [a]	585.22 [a]	540.11 [b]	550.83 [b]
	Spring w/o Autumn	544.20 [a]	554.79 [a]	508.06 [b]	518.65 [b]

lm: linear model (pathobiota-Shannon ~ intercept + a*microbiota-Shannon); nlm: non-linear model, quadratic function (pathobiota-Shannon ~ k* microbiota-Shannon – q* microbiota-Shannon). AIC: Akaike's information criterion; BIC Bayesian information criterion.

Supplementary Table 11. Fitting of a non-linear model on the relationship between the α-diversities of microbiota and potential pathobiota.

Compartment	Samples	k	q
Leaf	All seasons	1.44295 ***	0.41021 ***
	Autumn	1.44336 ***	0.40783 ***
	Spring w/ Autumn	1.49777 ***	0.43395 ***
	Spring w/o Autumn	1.38687 ***	0.38764 ***
Root	All seasons	1.49407 ***	0.43215 ***
	Autumn	1.21044 ***	0.36224 ***
	Spring w/ Autumn	1.55035 ***	0.42845 ***
	Spring w/o Autumn	1.66111 ***	0.48925 ***

Non-linear model: quadratic function (pathobiota-Shannon ~ k*microbiota-Shannon – q*microbiota-Shannon).
****P* < 0.001.

Supplementary Table 12. *In planta* bacterial growth of four natural *Pseudomonas* strains in four corresponding local natural populations of *A. thaliana*, each represented by two accessions.

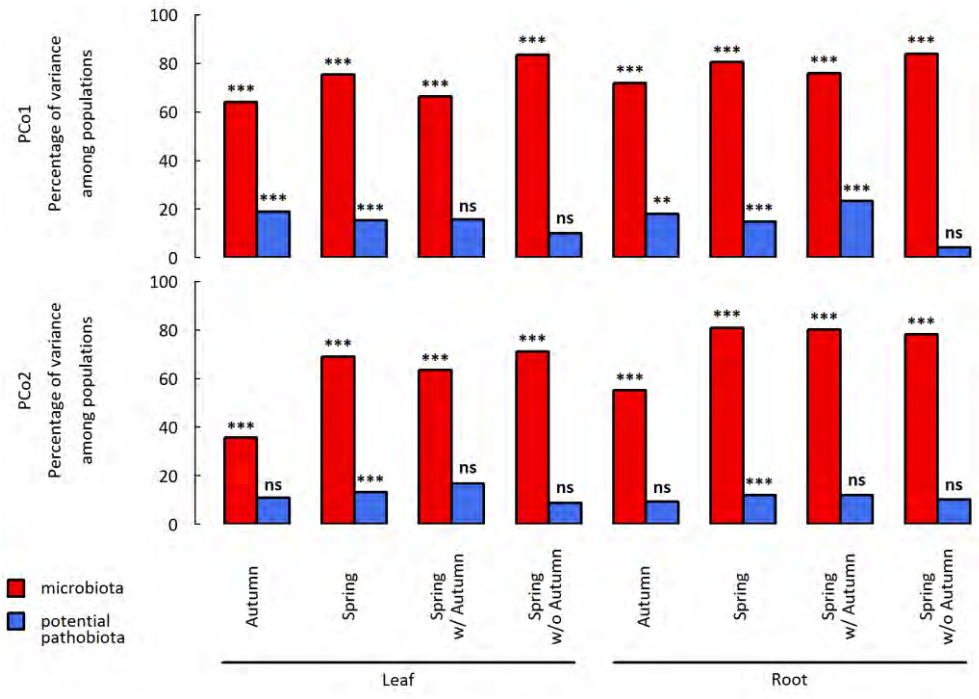
Model terms	***F* or *LRT***	***P***
block	28.95	<.0001
strain	0.56	0.6396
population	1.29	0.3292
time	178.54	<.0001
strain*population	0.21	0.9932
strain*time	1.13	0.3373
time*population	1.62	0.2484
strain*population*time	0.35	0.9579
accession(population)	1.20	0.2733
*strain*accession(population)*	0.00	1.0000
*time*accession(population)*	0.00	1.0000
*strain*time*accession(population)*	0.00	1.0000

In plant bacterial growth ($logCFU.cm^{-2}$) was modeled using a mixed model. Model random terms (in italics) were tested with likelihood ratio tests (LRT) of models with and without these effects. A Bonferroni correction for the number of tests was performed at a nominal level of 5%.

Supplementary Table 13. Natural interactions between eight local natural *A. thaliana* accessions and eight corresponding local natural *Pseudomonas syringae* strains for disease symptom evolution.

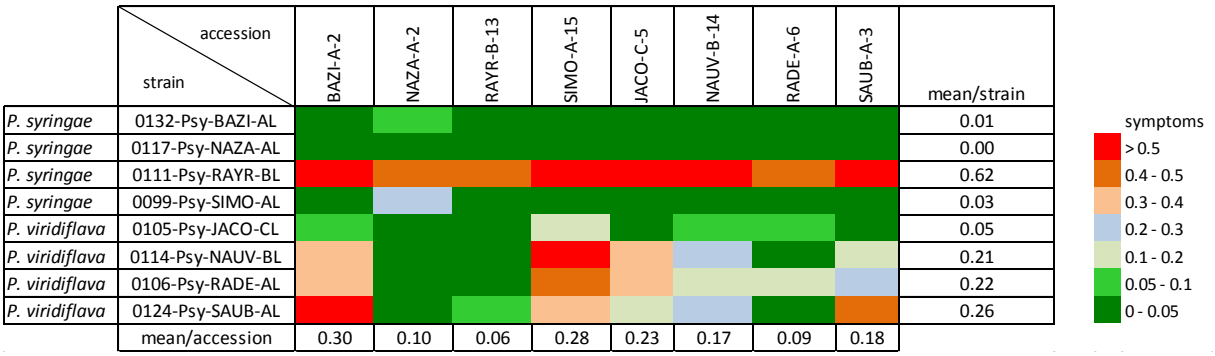
Model terms	***F* or *LRT***	***P***
block	6.90	**<.0001**
strain	1.77	0.0904
time	29.38	**0.0002**
strain*time	13.35	**<.0001**
accession	0.00	1.0000
*strain*accession*	0.00	1.0000
*time*accession*	4.50	0.0339
*strain*time*accession*	22.40	**<.0001**

Disease symptoms were modeled using a mixed model. Model random terms (in italics) were tested with likelihood ratio tests (LRT) of models with and without these effects. A Bonferroni correction for the number of tests was performed at a nominal level of 5%. Bold values indicate statistically significant results after correction for multiple comparisons.

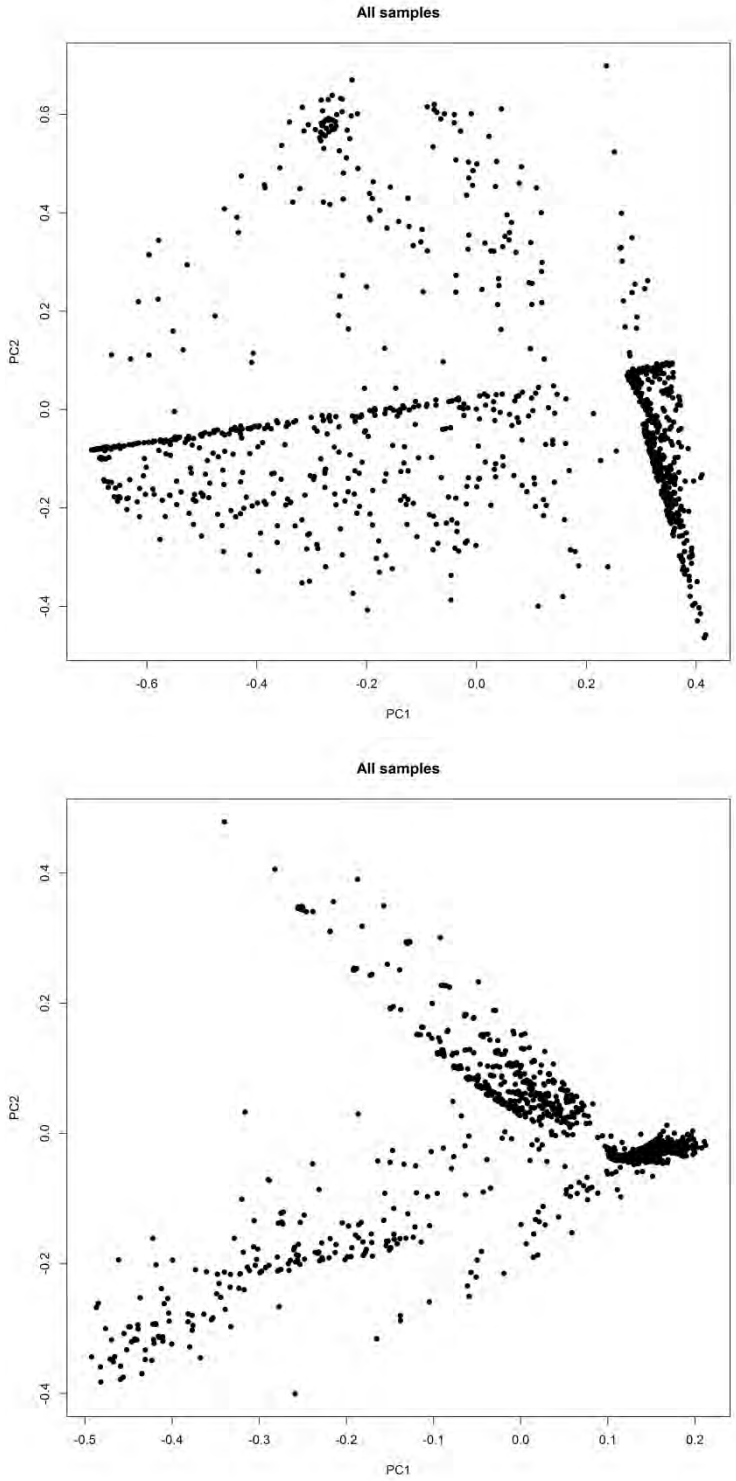


Supplementary Figure 1. Percentage of variance of the β-diversity among populations. Red and blue bars correspond to microbiota and potential pathobiota, respectively. Significance after a Bonferroni correction at a nominal level of 5%: ^{ns} non-significant, ** 0.01> *P* > 0.001, *** *P* < 0.001.

Supplementary Figure 2. Genetic variation among five accessions from the region Midi-Pyrénées for the response to the *P. viridiflava* strain 0124-Psy-SAUB-AL, three days after infiltration. Mock: infiltration with water.



Supplementary Figure 3. Heatmap for the symptoms observed three days after inoculation with *P. syringae* strains. The heatmap shows the interactions between eight natural accessions of *A. thaliana* from the region Midi-Pyrénées and eight natural strains belonging to the *P. syringae* complex collected in the same populations than the eight natural accessions.



Supplementary Figure 4. PCoA perfomed on a Hellinger distance matrix based on rarefied data (top panel) and on a Jaccard similarity coefficient matrix distance (bottom panel).

Génomique écologique de l'adaptation d'*Arabidopsis thaliana* dans un environnement spatialement hétérogène

Dans le contexte des changements globaux, un des enjeux majeurs en génomique écologique est d'estimer le potentiel adaptatif des populations naturelles. Répondre à cet enjeu nécessite 3 étapes: identification des agents sélectifs et de leurs échelles spatiales de variation, identification des bases génétiques de l'adaptation et étude de la dynamique adaptative sur une courte échelle de temps. Durant ma thèse, je me suis intéressé à étudier le potentiel adaptatif de la plante modèle *Arabidopsis thaliana*. A partir de 168 populations naturelles d'*A. thaliana* caractérisées pour 24 traits phénotypiques et 60 facteurs abiotiques (climat, sol) et biotiques (communautés végétales et microbiote), j'ai pu mettre en évidence que les communautés végétales étaient les principaux agents sélectifs associés à la *fitness*. Après avoir séquencé le génome de ces 168 populations (~ 4.8 millions de SNPs), j'ai effectué des analyses de type 'association génome-environnement' couplé à des scans génomiques de différenciation génétique spatiale. Ces analyses ont confirmé l'importance de considérer les interactions plante-plante dans l'étude de l'adaptation chez *A. thaliana.* Afin d'étudier le potentiel adaptatif d'*A. thaliana* sur le court terme dans le contexte d'un réchauffement climatique, j'ai combiné une étude de résurrection *in situ* avec une étude de Genome Wide Association mapping, à partir de 195 accessions locales caractérisées pour 29 traits phénotypiques et pour environ 1.9 million de SNPs. J'ai identifié une architecture originale de l'adaptation vers un nouvel optimum phénotypique combinant (i) de rares QTLs avec des degrés de pléiotropie intermédiaires fortement sélectionnés et (ii) de très nombreux QTLs spécifiques d'un micro-habitat et faiblement sélectionnés. A travers les différents projets abordés pendant ma thèse, j'ai pu suggérer qu'une architecture génétique flexible pouvait permettre à *A. thaliana* de s'adapter rapidement aux changements globaux, tout en maintenant de la diversité génétique au sein des populations naturelles d'*A. thaliana*.

Mots clés: génomique écologique, adaptation locale, agents sélectifs, grain de l'environnement, populations naturelles, *Arabidopsis thaliana*, Association Génome-Environnement, GWA mapping

Ecological genomics of adaptation of *Arabidopsis thaliana* in a spatially heterogeneous environment

In the context of global changes, one of the challenges in ecological genomics is to estimate the adaptive potential of natural populations. Three steps are requested to address this challenge: identification of the selective agents and their associated spatial grains, identification of the genetic bases of adaptation and monitoring the adaptive dynamics of natural population over a short time period. Here, I aimed at studying the adaptive potential of the model plant *Arabidopsis thaliana*. Based on 168 natural populations of *A. thaliana* characterized for 24 phenotypic traits and 60 abiotic (climate, soil) and biotic (plant communities and microbiota) factors, plant communities were found to be the main selective agents. Based on 4.8 million SNPs, I combined Genome Environment Association analysis with genome scans for signatures of selection. I confirmed the importance to consider plant-plant interactions when studying adaptation in *A. thaliana*. To monitor the adaptive dynamics of a natural population in the context of global warming, I combined an *in situ* resurrection study with an approach of GWA mapping based on 195 local accessions characterized for 29 phenotypic traits and 1.9 million SNPs. Adaptive evolutionary changes were largely driven by rare QTLs with intermediate degrees of pleiotropy under strong selection. In addition to these rare pleiotropic QTLs, weak selection was detected for frequent small micro-habitat-specific QTLs that shape single traits. Overall, I suggest that a rapid adaptive phenotypic evolution can be rapidly achieved in *A. thaliana*, while still maintaining genetic variation in natural populations.

Keywords: ecological genomics, local adaptation, selective agents, spatial grain, natural populations, *Arabidopsis thaliana*, Genome-Environment Association, GWA mapping