MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 4222–4240 doi: 10.1111/mec.12396

Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in *Arabidopsis thaliana*

BENJAMIN BRACHI,*†[1] ROMAIN VILLOUTREIX,*[1] NATHALIE FAURE,* NINA HAUTEKÈETE,* YVES PIQUOT,* MAXIME PAUWELS,* DOMINIQUE ROBY,‡§ JOËL CUGUEN,* JOY BERGELSON† and FABRICE ROUX*

*Laboratoire Génétique et Evolution des Populations Végétales, UMR CNRS 8198, Université des Sciences et Technologies de Lille – Lille 1, F-59655 Villeneuve d'Ascq Cedex, France, †Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA, ‡INRA, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR441, F-31326 Castanet-Tolosan, France, §CNRS, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR2594, F-31326 Castanet-Tolosan, France

Abstract

Despite the increasing number of genomic tools, identifying the genetics underlying adaptive complex traits remains challenging in the model species *Arabidopsis thaliana*. This is due, at least in part, to the lack of data on the geographical scale of adaptive phenotypic variation. The aims of this study were (i) to tease apart the historical roles of adaptive and nonselective processes in shaping phenological variation in *A. thaliana* in France and (ii) to gain insights into the spatial scale of adaptive variation by identifying the putative selective agents responsible for this selection. Forty-nine natural stands from four climatically contrasted French regions were characterized (i) phenologically for six traits, (ii) genetically using 135 SNP markers and (iii) ecologically for 42 variables. Up to 63% of phenological variation could be explained by neutral genetic diversity. The remaining phenological variation displayed stronger associations with ecological variation within regions than among regions, suggesting the importance of local selective agents in shaping adaptive phenological variation. Although climatic conditions have often been suggested as the main selective agents acting on phenology in *A. thaliana*, both edaphic conditions and interspecific competition appear to be strong selective agents in some regions. In a first attempt to identify the genetics of phenological variation at different geographical scales, we phenotyped worldwide accessions and local polymorphic populations from the French RegMap in a genome-wide association (GWA) mapping study. The genomic regions associated with phenological variation depended upon the geographical scale considered, stressing the need to account for the scale of adaptive phenotypic variation when choosing accession panels for GWAS.

Keywords: adaptation, *Arabidopsis thaliana*, ecology, flowering time, selection

Received 1 October 2012; revision received 21 May 2013; accepted 23 May 2013

Introduction

Numerous phenotypic traits display extensive variation among natural populations. Many evolutionary forces may jointly shape this variation, including genetic drift, migration, demographic history and natural selection (Belotte *et al.* 2003; Kawecki & Ebert 2004; Olson-Manning *et al.* 2012). The extent to which this natural variation is adaptive is still an open question (Mitchell-Olds *et al.* 2007).

Characterization of how selection has shaped natural variation in complex traits requires that at least two key

Correspondence: Fabrice Roux, Fax: +33 (0)3 20 43 69 79; E-mail: fabrice.roux@univ-lille1.fr

[1]These authors contributed equally to this work.

issues be addressed. The first challenge is to estimate the part played by adaptive processes in shaping phenotypic natural variation and the geographical scale at which adaptation takes place. To date, much of the work on how selection shapes complex traits has focused on local adaptation. Two complementary approaches have been applied. Reciprocal transplants address how genotypes perform in their population of origin relative to more distant populations (Leimu & Fischer 2008; Hereford 2009). Fitness superiority of local individuals at each site demonstrates local adaptation (following the 'local vs foreign' criterion, Kawecki & Ebert 2004). In the second approach, selection gradients are used to quantify the effects of trait variation on an estimator of fitness (Lande & Arnold 1983; Munguía-Rosas *et al.* 2011). Because it is time-consuming to measure fitness, both approaches require that adaptation be studied in a restricted number of natural populations, making generalizations difficult (but see Belotte *et al.* 2003; Laine 2005; Becker *et al.* 2006). One indirect method is to study the relationships between ecological and phenotypic variations (Linhart & Grant 1996; Merilä *et al.* 2001). While correlational analyses such as these do not allow definitive conclusions, this approach facilitates the inclusion of more individuals and more populations and therefore enables study of adaptation at different geographical scales (Conner 2010). Correlations with environmental variables, if they are driven by local adaptation, integrate multiple generations of selection and multiple components of fitness (Merilä *et al.* 2001).

The second challenge is to study the genetic architecture and identify the genetic basis of adaptation as a first step towards reconstructing the adaptive walk followed by a natural population towards the local optimum phenotype (Hermisson & Pennings 2005; Orr 2005; Kopp & Hermisson 2007, 2009). Genome-wide association (GWA) mapping provides a powerful tool to start tackling this question. The power of GWA mapping to identify the genetics underlying phenotypic variation observed at broad geographical scales has been demonstrated in multiple plant studies, but so have its limitations (Atwell *et al.* 2010; Brachi *et al.* 2010; Huang *et al.* 2010). First, strong confounding by population structure introduces false positives and false negatives (Brachi *et al.* 2010). Second, rare alleles, potentially important in local adaptation processes, are difficult to detect (Atwell *et al.* 2010). Third, because the same phenotypic value may be caused by different alleles of the same gene among natural populations, allelic heterogeneity may hinder the detection of genomic regions associated with phenotypic natural variation (Bergelson & Roux 2010). Traditional linkage mapping based on crosses among populations may solve these issues but has the disadvantage of coarse resolution (Bergelson & Roux 2010). An alternative approach would be to define regional or local mapping panels that match the geographical scale at which adaptive phenotypic variation is observed. As a consequence, confounding by population structure could be greatly reduced and causal genes with rare alleles or allelic heterogeneity at the species-wide scale could become detectable (Bergelson & Roux 2010; Brachi *et al.* 2011; Horton *et al.* 2012).

Phenological traits are excellent candidates for studying local adaptation because they display extensive variation in many plant species (Rathcke & Lacey 1985) and have often been found to be adaptive. In a meta-analysis of selection gradients, Munguía-Rosas *et al.* (2011) found the strength of selection on phenology to vary with latitude, as expected if climatic variation acts as a selective agent. Supporting evidence is provided by other studies that similarly reveal significant latitudinal gradients (Van Dijk *et al.* 1997; Stinchcombe *et al.* 2004; Wagmann *et al.* 2012) or a response of phenological traits to global climate change (Franks *et al.* 2007; Inouye 2008). However, climatic factors may not be the only selective agents acting on phenology. Many studies have found evidence for local selection by pollinators or herbivores (Parachnowitsch & Caruso 2008; Sandring & Ågren 2009). Taken together, there is overwhelming evidence that phenological traits are adaptive and that relevant selective agents can act globally or locally.

Arabidopsis thaliana, the flagship species of plant genomics, is a widely distributed annual selfing species found in diverse habitats (Mitchell-Olds & Schmitt 2006) that displays tremendous variation in phenological traits (Donohue 2005; Atwell *et al.* 2010; Brachi *et al.* 2010). Despite its emerging status as a model species in evolutionary ecology (Fournier-Level *et al.* 2011; Hancock *et al.* 2011; Gaut 2012), little is known about the geographical scale of phenological adaptation, that is, the selective pressures acting on its phenology, in natural populations (Méndez-Vigo *et al.* 2011; Montesinos-Navarro *et al.* 2011). The relationship between flowering time (FT) and latitude (Stinchcombe *et al.* 2004; Brachi *et al.* 2010), and between FT and several climatic variables (e.g. the number of consecutive frost-free days, maximum temperature in the warmest month and photosynthetically active radiation, Hancock *et al.* 2011), suggests that climate may impose selection shaping FT variation. A large proportion of phenological variation, however, remains unexplained at the continental scale, suggesting that phenological variation at smaller geographical scales results from either nonselective processes or local selective agents.

Here, we investigated the geographical scale of adaptive phenological variation and its underlying genetics in *A. thaliana*. In the first part of the study, (i) we

estimated the portions of natural phenological variation in *A. thaliana* that could be the result of nonselective or adaptive processes and (ii) we investigated the geographical scale of adaptive variation by identifying the putative selective agents responsible for this selection. To achieve these goals, we used a hierarchical sampling design to collect 800 individuals from 49 natural stands located in four climatically contrasted regions of France. We characterized the selfed progeny of those individuals: phenologically for six traits spanning the annual plant life cycle, genetically using 135 neutral SNPs and ecologically for 42 variables. We found that neutral genetic diversity could explain up to 63% of phenotypic variation for some phenological traits. The remaining variation was associated with many ecological factors, including edaphic variation and competition. Relationships with ecological variables were stronger within regions than considering all regions together, suggesting a prevalence of local adaptation in shaping the natural variation of phenological traits in our study. We made all plant material used in this study publicly available in order to allow follow-up studies (http://publiclines.versailles.inra.fr/).

In the second part of the study, we aimed at testing whether GWA mapping studies were more successful in identifying the genetic basis of phenological variation with a mapping panel of accessions selected to match the geographical scale of adaptive variation observed within the 49 stands. To do so, we performed a GWA study using samples of *A. thaliana* collected from different geographical scales using both worldwide accessions and accessions from the French RegMap panel (Horton *et al.* 2012). Consistent with the scale of adaptive phenological variation observed within the 49 stands, GWA mapping revealed strong signals of association at the population and regional scales that were often located in different genomic regions than those detected at the worldwide scale.

Materials and methods

Plant material

Forty-nine stands of *A. thaliana* were collected from early March to late April 2009 in 42 locations from four regions of France that have contrasting climates (Brittany: oceanic, 11 stands; Burgundy: continental, 11 stands; Languedoc: Mediterranean, 16 stands; and north of France: semi-oceanic, 11 stands; Table S1, Supporting information). We defined stands as a single patch of plants growing in relatively homogeneous ecological conditions. The average within-region distance among stands was 33.1 km (SD = 19.6 km). The pairwise minimal distances among regions ranged from 292 km (Burgundy–Languedoc) to 758 km (Brittany–Languedoc).

In each stand, between 10 and 30 plants were collected randomly and brought back to a cold frame greenhouse (no additional light or heating). Seeds were collected from individual plants to constitute seed families. To reduce maternal effects, families were grown for one generation from June to December 2009 under controlled greenhouse conditions (16-h photoperiod, 20 °C) at the University of Lille 1. For the phenotyping experiment, 20 families were randomly chosen from each stand when possible. If <20 families were available in a given stand, all families were phenotyped (Table S1, Supporting information). In all, each French region was represented by 200 families. Seeds of the 800 families will be publicly available from the Centre de Ressources Biologiques (CRB, INRA Versailles, France, http://publiclines.versailles.inra.fr/).

A set of 184 worldwide accessions and a set of 210 French accessions corresponding to the French RegMap (Horton *et al.* 2012) were also included in the experiment (Table S1, Supporting information). In order to (i) have enough power to run GWA mapping analyses at small geographical scales and (ii) use accessions collected in the same geographical region than the 49 stands, French accessions were over-represented. For the purpose of our study, we excluded from our statistical analyses the natural accessions that were likely to be contaminants, that is, accessions for which geographical origin is suspicious (Anastasio *et al.* 2011), leaving us with a set of 352 accessions (Table S1, Supporting information). From these 352 accessions, we designed six sets of accessions corresponding to different geographical scales: WORLD (n = 167), EUROPE (n = 143), FRANCE (n = 203), BURGUNDY (regional scale, n = 121), and MIB and TOU (two local populations in Burgundy; n = 52 and n = 69, respectively). The WORLD, EUROPE and FRANCE sets shared 18 French accessions that are representative of the French diversity (Atwell *et al.* 2010).

Phenological characterization

A greenhouse experiment of 4928 plants (French families, worldwide accessions and accessions from the French RegMap) was set up at the University of Lille (North, France) in January 2010 using a split-plot design arranged as a randomized complete block design (RCBD) with two seasonal germination cohorts nested within two experimental blocks. The two germination cohorts were grown to mimic the two germination flushes found in natural stands, that is, fall and spring (Donohue 2005; Picó 2012). Each block was represented by 19 arrays of 66 individual wells (Ø4 cm, vol.

~38 cm^3) (TEKU, JP 3050/66) filled with damp standard culture soil (Huminsubstrat N3, Neuhaus). Each block corresponded to 1232 plants with one replicate per family (n = 800), one replicate per worldwide accession (n = 184), one replicate per French accession (n = 210) and a control worldwide accession Bg-2 placed in the same two positions within each array (n = 38 = 19 trays * 2 replicates). Those control plants allowed correcting for micro-environmental variation within blocks. At least five seeds were placed in each well. To promote germination, seeds were stratified for 7 days at 4 °C. After the 7-day stratification treatment, plants were grown at 20 °C under natural light supplemented by artificial light to provide a 16-hr photoperiod. Germination date was monitored daily in each well. *A. thaliana* seeds that had not germinated 13 days after sowing were replaced by extra seedlings of the same family or accession from other blocks. Seedlings were thinned to one per well 14 days after the stratification treatment, keeping the first germinated seedling. Plants in the spring germination cohort underwent the same greenhouse conditions. A winter treatment was simulated for plants in the fall germination cohort. Twenty-two days after the stratification treatment, plants in the fall germination cohort were grown in winter conditions (4 °C and 12-h photoperiod) for 3 weeks and then moved back to greenhouse conditions (20 °C and 16-h photoperiod). During the whole growing period, arrays within blocks were rotated every day to minimize potential effects of uneven lighting across the growth room. Plants were monitored every 2 or 3 days for bolting date (inflorescence distinguishable from the leaves at a size < 5 mm), flowering date (appearance of the first open flower), date of senescence of the last flower on the main stem and date of maturation of the last fruit on the main stem. A period of 3 weeks, corresponding to the length of the winter treatment, was subtracted for plants in the fall germination cohort. We measured six phenological traits spanning the life cycle of *A. thaliana*, as previously described in the study by Brachi *et al.* (2012). Bolting time (BT) was measured as the number of days between germination and bolting. The flowering interval (INT) was measured as the difference between bolting and flowering dates. FT was measured as the number of days between germination and flowering. Flowering period (FP) was measured as the number of days between the onset of flowering and the senescence of the last flower. The reproductive period (RP) was measured as the number of days between the onset of flowering and maturation of the last fruit on the main stem. The flowering-to-reproductive ratio (FRR) was calculated as the ratio between flowering (FP) and reproductive period ratio (RP). FRR may indicate a trade-off between seed number and seed quality as plants that spend a smaller fraction of their RP engaged in flowering may spend relatively more time filling and maturing seeds (Brachi *et al.* 2012). Plants that had not bolted 100 days after sowing (i.e. 3.3%) were assigned a bolting date value of 100. INT, FT, FP, RP and FRR were therefore not available for these plants.

DNA extraction and genotyping of SNP markers

After one generation of multiplication, seeds from the 800 French families were sown in March 2010 in arrays of 66 individual wells (Ø4 cm, vol. ~38 cm^3) (TEKU, JP 3050/66) filled with damp standard culture soil (Huminsubstrat N3, Neuhaus). After a 7-day stratification treatment at 4 °C, plants were grown at 20 °C and under natural light supplemented by artificial light to provide a 16-h photoperiod. Seedlings were thinned to one or two per well 14 days after the stratification treatment. Four weeks after the stratification treatment, plants were cut and oven-dried for 2 days at 65 °C. We used a NucleoSpin_96 Plant Kit (Macherey-Nagel) to extract and purify total DNA from 10 mg dried leaf tissue. All DNA samples were adjusted to 5 ng/μL. A total of 765 French families were genotyped for a set of 149 SNPs spread across the genome (Clark *et al.* 2007), already used to describe the scale of population structure in *A. thaliana* (Platt *et al.* 2010a) and genetic variation in natural stands of *A. thaliana* in Germany (Bomblies *et al.* 2010). The remaining 35 French families that were not genotyped correspond to those for which plant tissue was not available. The genotyping of 149 SNPs was performed by University of Chicago DNA sequencing facility (Chicago, IL, USA) using the Sequenom MassArray system.

Following Platt *et al.* (2010a), seven French families were removed due to an excess of missing genotype calls (>50 of 149). Fourteen SNP assays were also removed due to an excess of missing genotypes or heterozygous calls (>25% of families).

Ecological characterization of stands

The wide range of ecological conditions sampled in each region can be summarized in three broad habitat types: grassland (n = 23), hoed land (n = 16) and meadow (n = 10; Table S1, Supporting information). Each stand was characterized for climatic and edaphic variables, as well as for plant–plant interactions (Table 1). To obtain biologically meaningful climatic variables (Hijmans *et al.* 2005), we chose 25 climatic variables with a grid resolution smaller than the average distance among stands within regions (i.e. 33.1 km, SD = 19.6 km). Nineteen bioclimatic variables were obtained from the WorldClim database (www.worldclim.org) and extracted using

Table 1 Summary of 42 ecological variables gathered for this study

Variable	Description	Source or method	Grid resolution
Alt	Altitude (*m*)	—	—
RH winter	Relative humidity in winter (%)	NCAR/NCEP	5 amin (~9 km)
RH spring	Relative humidity in spring (%)	NCAR/NCEP	5 amin (~9 km)
RH summer	Relative humidity in summer (%)	NCAR/NCEP	5 amin (~9 km)
RH fall	Relative humidity in fall (%)	NCAR/NCEP	5 amin (~9 km)
Aridity	Aridity (mm/day)	FAO GeoNetwork	0.17°× 0.17° grid (~20 km)
Bio 1	Annual mean temperature (°C × 10)	WorldClim	30 asec (~1 km)
Bio 2	Mean diurnal range (mean of monthly temperature range (maximum temperature—minimum temperature))	WorldClim	30 asec (~1 km)
Bio 3	Isothermality (Bio 2/Bio 7)(*100)	WorldClim	30 asec (~1 km)
Bio 4	Temperature seasonality (standard deviation * 100)	WorldClim	30 asec (~1 km)
Bio 5	Maximum temperature of warmest month (°C × 10)	WorldClim	30 asec (~1 km)
Bio 6	Minimum temperature of coldest month (°C × 10)	WorldClim	30 asec (~1 km)
Bio 7	Temperature annual range (Bio 5–Bio 6)	WorldClim	30 asec (~1 km)
Bio 8	Mean temperature of wettest quarter (°C × 10)	WorldClim	30 asec (~1 km)
Bio 9	Mean temperature of driest quarter (°C × 10)	WorldClim	30 asec (~1 km)
Bio 10	Mean temperature of warmest quarter (°C × 10)	WorldClim	30 asec (~1 km)
Bio 11	Mean temperature of coldest quarter (°C × 10)	WorldClim	30 asec (~1 km)
Bio 12	Annual precipitation (mm)	WorldClim	30 asec (~1 km)
Bio 13	Precipitation of wettest month (mm)	WorldClim	30 asec (~1 km)
Bio 14	Precipitation of driest month (mm)	WorldClim	30 asec (~1 km)
Bio 15	Precipitation seasonality (coefficient of variation)	WorldClim	30 asec (~1 km)
Bio 16	Precipitation of wettest quarter (mm)	WorldClim	30 asec (~1 km)
Bio 17	Precipitation of driest quarter (mm)	WorldClim	30 asec (~1 km)
Bio 18	Precipitation of warmest quarter (mm)	WorldClim	30 asec (~1 km)
Bio 19	Precipitation of coldest quarter (mm)	WorldClim	30 asec (~1 km)
OC	Organic carbon (g/kg)	NF ISO 10694 and NF ISO 13878 standards	—
N	Total nitrogen (g/kg)	NF ISO 10694 and NF ISO 13878 standards	—
C/N	Carbon/nitrogen ratio	NF ISO 10694 and NF ISO 13878 standards	—
SOM	Soil organic matter (g/kg)	NF ISO 10694 and NF ISO 13878 standards	—
P2O5	Phosphorus (P_2O_5) (g/kg)	Olsen method (ISO 11263 standard)	—
Ca	Exchangeable calcium (cmol+ per kg)	Cobaltihexammine method (ICP-AES, INRA method)	—
Mg	Exchangeable magnesium (cmol+ per kg)	Cobaltihexammine method (ICP-AES, INRA method)	—
Na	Exchangeable sodium (cmol+ per kg)	Cobaltihexammine method (ICP-AES, INRA method)	—
K	Exchangeable potassium (cmol+ per kg)	Cobaltihexammine method (ICP-AES, INRA method)	—
Fe	Exchangeable potassium (cmol+ per kg)	Cobaltihexammine method (ICP-AES, INRA method)	—
Mn	Exchangeable iron (cmol+ per kg)	Cobaltihexammine method (ICP-AES, INRA method)	—
Al	Exchangeable aluminium (cmol+ per kg)	Cobaltihexammine method (ICP-AES, INRA method)	—
WHC	Soil water holding capacity (mL/g)	Granier and Tardieu (1999)	—
pH	pH	NF ISO 10390 standard	—
Herb	Interspecific competition with herbs which are not grasses	50 × 50 cm quadrat	—
Grass	Interspecific competition with grasses	50 × 50 cm quadrat	—
Thal	Intraspecific competition	50 × 50 cm quadrat	—

DIVA-GIS software (www.diva-gis.org). These 19 bioclimatic variables were derived from monthly temperature and rainfall values, based on averages calculated for the 1960–1990 period with a spatial resolution of 1 km². Six additional climatic variables corresponding to relative humidity for each of the four seasons, altitude and aridity were obtained as described in the study by Hancock *et al.* (2011).

A sample of the upper soil layer was collected in each stand. Soil samples were transferred to the greenhouse and air-dried. Soil samples were then stored in the laboratory at room temperature. Each stand was characterized for 14 edaphic factors (Table 1). Mean pH per stand was calculated using two soil subsamples. Maximal water holding capacity of two samples of 200 cm^3 per stand was measured as the amount of water held in soil after excess water had been drained away. The twelve other soil properties (content of total N and organic C, C/N ratio, content of organic matter, concentrations of P_2O_5, K, Ca, Mg, Mn, Al, Na and Fe) were assessed at INRA Arras (France, www.lille.inra.fr/las).

To estimate intra- and interspecific competition intensities, a 50 × 50 cm quadrat divided into 25 smaller squares (10 × 10 cm) was established in a representative area of each stand. Intraspecific competition was calculated as the density of intraspecific competitors based on the presence/absence of *A. thaliana* in each 10 × 10 cm square. Two interspecific competition indices were estimated as the presence/absence in each 10 × 10 cm square of either grasses or other herbs.

Data analysis

Phenological variation and spatial scale. Based on the 49 stands, we studied phenological variation in France by using the following statistical model according to our split-plot design:

$$\begin{aligned} Y_{ijklmc} = {} & \mu_{\text{trait}} + \text{block}_i + \text{cohort}_j + \text{block}_i \times \text{cohort}_j \\ & + \text{region}_k + \text{stand}_l(\text{region}_k) \\ & + \text{family}_m(\text{stands}_l(\text{region}_k)) + \text{cohort}_j \times \text{region}_k \\ & + \text{cohort}_j \times \text{stands}_l(\text{region}_k) \\ & + \text{cohort}_j \times \text{family}_m(\text{stands}_l(\text{region}_k)) + \text{covBg2}_c \\ & + \varepsilon_{ijklmc} \end{aligned} \quad (1)$$

In this model, '*Y*' is one of the six phenological traits; 'μ' is the overall mean; 'block' accounts for differences in micro-environment among the two experimental blocks; 'cohort' corresponds to the two germination cohorts grown; 'region', 'stand' and 'family' measure the effect of three spatial scales in France, that is, regions, stands within regions and families within stands; interaction terms involving the 'cohort' factor account for genetic variation in reaction norms among regions, among stands within regions and among families within stands; $\text{cov}_{\text{Bg-2}}$ is a covariate accounting for array effects within blocks (phenotypic mean of the two Bg-2 replicates per array was used as a covariate); and 'ε' is the residual term. All factors were treated as fixed effects, except for 'family' that was treated as a random effect. For calculating *F*-values, terms were tested over their appropriate denominators. Given the split-plot design used in this study, the variance associated with 'block × cohort' was for example used as the error term for testing the 'block' and 'cohort' effects. Raw data were Box–Cox-transformed to satisfy the normality and equal variance assumptions of linear regression. Model fitting was conducted using the PROC MIXED procedure in SAS 9.3 (SAS Institute Inc., Cary, NC, USA).

Within each cohort, the model described above was used excluding the terms involving 'cohort_j'. The contribution of each factor to the total phenotypic variance was estimated using variance component analysis, treating all factors as random effects and excluding $\text{cov}_{\text{Bg-2}}$. Model fitting was conducted using the PROC VARCOMP procedure in SAS 9.3 (REML method; SAS Institute Inc.).

Within each cohort, a within-stand coefficient of genetic variation CV was calculated for each phenological trait based on variance components estimated by REML.

Phenological variation scored on the worldwide accessions and the natural accessions from the French RegMap was analysed with the following model treating all the factors as fixed effects, except for 'accession' that was treated as a random effect (PROC MIXED procedure in SAS 9.3):

$$\begin{aligned} Y_{ijnc} = {} & \mu_{\text{trait}} + \text{block}_i + \text{cohort}_j + \text{block}_i \times \text{cohort}_j \\ & + \text{accession}_n + \text{cohort}_j \times \text{accessions}_n + \text{covBg2}_c \\ & + \varepsilon_{ijnc} \end{aligned} \quad (2)$$

For calculating *F*-values, terms were tested over their appropriate denominators. Raw data were Box–Cox-transformed to satisfy the normality and equal variance assumptions of linear regression.

For each cohort, best linear unbiased predictions (BLUPs) were obtained for each genotype (i.e. natural accessions and families) using the PROC MIXED procedure in SAS 9.3 (SAS Institute Inc.):

$$Y_{igc} = \mu_{\text{trait}} + \text{block}_i + \text{genotype}_g + \text{covBg2}_c + \varepsilon_{igc} \quad (3)$$

Genetic diversity and spatial scale of genetic variation. Six genetic diversity parameters were estimated for each of the 49 stands. Mean gene diversity (H_S), percentage of

polymorphic loci (*PL*), mean number of observed alleles per locus (n_a) and mean allelic richness per locus (R_S) were estimated using FSTAT v.2.9.3 (Goudet 1995). Number of haplogroups (*HG*) and number of private haplogroups (*PHG*) were estimated as described in the study by Platt *et al.* (2010a), with a 0.01 per-site genotyping error rate and a 0.05 *P*-value threshold to exclude an additional French family from a haplogroup. Only the French families with a level of heterozygosity below 1% (n = 710) were considered for the estimation of *HG* and *PHG*. A multivariate analysis of variance (MANOVA) was used to test the effect of either French region or habitat type on the six genetic diversity parameters simultaneously (PROC MANOVA in SAS 9.3; SAS Institute Inc.). Raw data for *HG* and *PHG* were log-transformed to satisfy the normality and equal variance assumptions of MANOVA.

Spatial partitioning of the total genetic variation was estimated for all loci using the R package 'hierfstat' (Goudet 1995; de Meeûs & Goudet 2007). Hierarchical *F*-statistics and variance components were calculated for the 'region', 'stands' and 'individual' spatial levels. Ten thousand bootstraps were performed to estimate confidence intervals for the *F*-statistics. Within region, an among-population *F*-statistic (Weir & Cockerham 1984) was calculated by averaging over the 135 SNPs F_{ST} values obtained using the R package 'pegas' (Paradis 2010).

Spatial scale of ecological variation. For each ecological factor, the contribution of regions to the total ecological variance was estimated by variance component analysis (PROC VARCOMP procedure in SAS 9.3), according to the following statistical model:

$$Y_i = \text{region}_i + \varepsilon_i \quad (4)$$

where Y stands for the different ecological variables and 'region' corresponds to the four French regions. Raw data for the three competition indices were arc-sin transformed to satisfy the normality and equal variance assumptions of linear regression. Relationships among the 42 ecological factors were assessed with both Pearson and Spearman correlation coefficients.

Untangling the portions of phenotypic variation resulting from adaptive and nonselective processes. To estimate the portion of phenological variation that could be explained by neutral genetic variation, a principal component analysis (PCA) was first performed on the 135 SNPs, recoded as binary haploid SNP genotypes (0 and 1 corresponding to the common and rare allele at each SNP, respectively). Principal components (PCs) with eigenvalues ≥ 1 were then included in the following linear model:

$$Y_i = \mu_{\text{trait}} + \text{PC1}_i + \ldots + \text{PCN}_i + \varepsilon_i \quad (5)$$

where Y is a vector of phenological BLUPs, μ_{trait} is the overall phenological trait mean, PC1…PCN are the genetic principal components included in the model and ε_i are the residuals of the model. The residuals were considered to represent the phenological variation minus the effect of nonselective processes. The adjusted r-square obtained for each trait can be seen as an upper limit for the amount of phenotypic variation explained by nonselective processes.

Estimating the geographical scale of adaptive variation through the identification of the putative selective agents on phenology. Partial least square regression (PLSR) was used to identify the ecological factors potentially acting as selective pressures on phenological traits in France. PLSR identifies combinations of ecological variables that, unlike PCA, maximize the variation explained in a response variable (Geladi & Kowalski 1986; Carrascal *et al.* 2009), such as seed dormancy (Wagmann *et al.* 2012) or our phenological traits in this study. PLSR was chosen for two reasons. First, the number of explanatory variables (i.e. ecological factors) is not limited by the number of observations (i.e. stands). Second, PLSR can be carried out on nonindependent explanatory variables.

Using the R package 'pls' (Mevik & Wehrens 2007), PLSR was performed twice. In the first analysis, we investigated the relationships between standardized ecological factors and phenological median per stand calculated from standardized BLUPs. In the second analysis, we investigated the relationship between standardized ecological factors and phenological median per stand calculated from standardized residuals obtained from eqn 5. Standardization of BLUPs and residuals allowed us to compare regression coefficients obtained from PLSR before or after accounting for nonselective processes, respectively. In both analyses, leave-one-out cross-validation was used to determine the optimal number of components to be included in the model. Significance of regression coefficients was tested by approximate *t*-tests based on jackknife variance estimates.

GWA mapping at different geographical scales. All 352 natural accessions corresponding to the WORLD, EUROPE, FRANCE, BURGUNDY, MIB and TOU sets have been genotyped for 214 051 SNPs evenly spaced across the genome (Horton *et al.* 2012, http://bergelson.uchicago.edu/regmap-data/). To fine-map genomic regions associated with natural phenological variation in each set, we first ran a Wilcoxon rank-sum test on the association between phenotypes and genotypes for each marker (Atwell *et al.* 2010). GWA mapping was then run using a mixed-model approach implemented in the software

EFFICIENT MIXED-MODEL ASSOCIATION EXPEDITED (EMMAX, Kang *et al.* 2010). To control for population structure, this model includes a genetic kinship matrix K accounting for genome-wide patterns of relatedness among the accessions (i.e. identity-by-state). These analyses were based on BLUPs obtained by the statistical model described in eqn (3). Because of bias due to rare alleles, we only considered SNPs with minor allele relative frequency (MARF) > 10% (Brachi *et al.* 2010; Kang *et al.* 2010) at the geographical scale considered.

Candidate genes close to highly associated SNPs were identified among a list of 282 a priori FT candidate genes described in the study by Brachi *et al.* (2010). This list was enriched with FT candidate genes from the following website (Max Planck Institute for plant breeding research, http://www.mpipz.mpg.de/14637/Arabidopsis_flowering_genes), resulting in a list of 328 candidate genes in total (Table S2, Supporting information).

Results

Natural diversity of A. thaliana in France: phenological, genetic and ecological characterization of 49 stands

Phenological variation: comparison with a set of worldwide natural accessions. The distribution of BT at large geographical scales, that is, WORLD and EUROPE scales, was bimodal with an apparent excess of early and late bolting accessions (Fig. 1). Apparent excess of late bolting accessions in the fall cohort might result, in part, to an unsatisfied vernalization requirement in Scandinavian accessions (Shindo *et al.* 2005). Interestingly, in the spring cohort, the natural variation observed for BT in the local population TOU (n = 69) almost covered the range of natural variation for BT at the worldwide scale. The distributions of BT in the four French regions also differed from its distribution in worldwide accessions for both germination cohorts (Fig. 1). Noteworthy was the excess of intermediate values of BT in the Languedoc region compared with worldwide accessions. Trends similar to those for BT were observed for FT (Fig. S1, Supporting information).

Distributions of the flowering interval (INT) showed very small differences among geographical scales, with within-region (and even within-population) diversity spanning worldwide diversity (Fig. S1, Supporting information). Flowering and reproductive periods (FP and RP) displayed patterns similar to INT except for the Languedoc region. In the spring cohort, Languedoc families exhibited an excess of short flowering and reproductive periods. In contrast, in the fall cohort, an excess of long flowering periods was observed for Languedoc (Figs 1 and S1). In both cohorts, Languedoc families exhibited extensive flowering-to-reproductive ratios (FRR) natural variation in comparison with worldwide accessions and with the other three French regions (Fig. 1).

The interaction between the 352 natural accessions and germination cohort was highly significant for BT and FT and nonsignificant for the other traits (Table 2). In the French populations, the interaction between germination cohort and either region or stand within region was highly significant for all traits (except for FT and FRR at the regional scale; Table 3). Contrasting responses to winter treatment among phenological traits

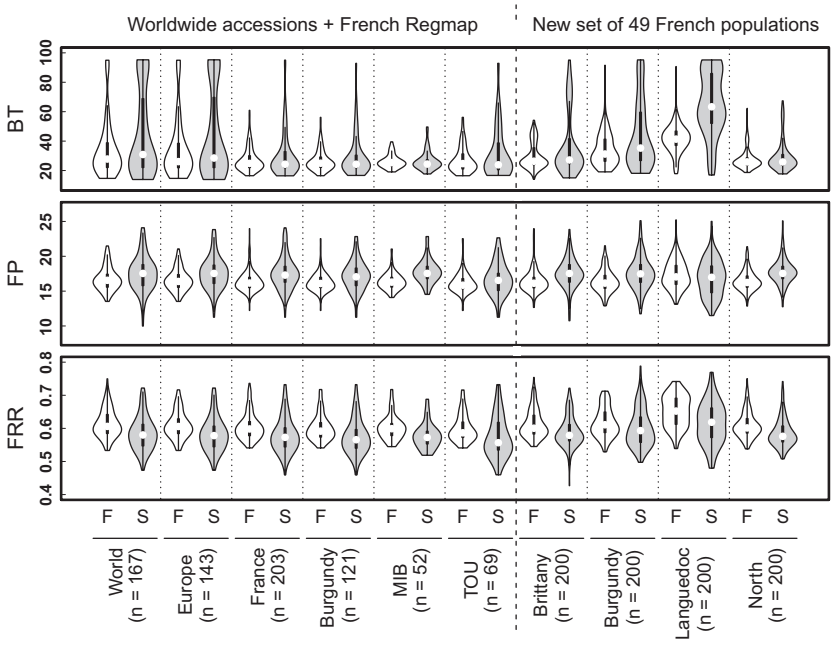


Fig. 1 Violin plots of natural variation for bolting time (BT), flowering period (FP) and flowering-to-reproductive ratio (FRR) for all accessions (Worldwide, French RegMap and new set of 49 French populations). BT and FP are expressed in days. '*F*' and '*S*' denote fall and spring germination cohorts, respectively. Violin plots are based on best linear unbiased predictions calculated for each of the 352 natural accessions or 800 French families. Violin plots are a combination of a boxplot (the white dot represents the median, the black solid box around it the range from the first to the third quartile and the thin black lines above and below the box extent to the 1.5 times the inner quartile ranges) and a rotated kernel density plots (Hintze & Nelson 1998; Adler 2005).

Table 2 Phenological variation among worldwide accessions and natural accessions from the French RegMap

Traits	BT		INT		FT		FP		RP		FRR	
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*
Block	1.38	0.4477	2.51	0.1133	0.94	0.5095	3.00	0.0836	0.00	0.9985	1.02	0.4955
Cohort	3.49	0.0622	1.23	0.2685	9.44	**0.0022**	0.39	0.5301	9.68	**0.0020**	0.41	0.5247
Accession	739.20	**0.0001**	69.10	**0.0001**	566.00	**0.0001**	35.50	**0.0001**	34.20	**0.0001**	65.70	**0.0001**
*Cohort*Accession*	12.30	**0.0005**	0.00	1.0000	23.60	**0.0001**	2.20	0.1380	0.60	0.4386	0.00	1.0000
Control Bg-2	8.68	**0.0002**	2.33	0.0976	14.42	**0.0001**	0.74	0.4757	5.74	**0.0033**	0.72	0.4852

BT, bolting time; FP, flowering period; FRR, flowering-to-reproductive ratio; FT, flowering time; INT, flowering interval; RP, reproductive period.
Model random terms were tested with likelihood ratio tests of models with and without these effects. Random effects are in italic. Bold *P*-values indicate significant effect after Bonferroni correction. Because the variance associated with the model term 'block* cohort' is the correct error term for testing the 'block' and 'cohort' effects, *F* value and *P*-value were not reported for the 'block* cohort' term.

Table 3 Phenological variation of the 49 French stands

Traits	BT		INT		FT		FP		RP		FRR	
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*
Block	2.54	0.3529	0.99	0.3189	9.73	0.1975	0.01	0.9325	0.15	0.7637	0.69	0.5583
Cohort	0.15	0.6952	0.32	0.5727	3.19	0.0743	0.73	0.3938	14.24	**0.0002**	0.63	0.4307
Region	158.12	**0.0001**	6.90	**0.0001**	116.07	**0.0001**	3.49	0.0153	9.31	**0.0001**	14.89	**0.0001**
Stand(Region)	26.33	**0.0001**	4.01	**0.0001**	26.30	**0.0001**	6.51	**0.0001**	5.16	**0.0001**	7.52	**0.0001**
Family(Stand(Region))	397.50	**0.0001**	11.60	**0.0007**	384.90	**0.0001**	0.00	1.0000	17.80	**0.0001**	1.23	0.1213
Cohort*Region	9.49	**0.0001**	4.09	**0.0067**	2.72	0.0432	10.93	**0.0001**	14.85	**0.0001**	0.74	0.5259
Cohort*Stand(Region)	3.28	**0.0001**	2.35	**0.0001**	2.87	**0.0001**	4.59	**0.0001**	4.14	**0.0001**	1.91	**0.0006**
Cohort(Family(Stand(Region)))*	0.00	1.0000	1.70	0.1923	0.00	1.0000	17.70	**0.0001**	0.00	1.0000	1.13	1.0000
Control Bg-2	5.14	**0.0060**	0.73	0.4839	6.99	**0.0010**	1.13	0.3225	8.64	**0.0002**	2.02	0.1336

BT, bolting time; FP, flowering period; FRR, flowering-to-reproductive ratio; FT, flowering time; INT, flowering interval; RP, reproductive period.
Model random terms were tested with likelihood ratio tests of models with and without these effects. Random effects are in italic. Bold *P*-values indicate significant effect after Bonferroni correction. Because the variance associated with the model term 'block* cohort' is the correct error term for testing the 'block' and 'cohort' effects, *F* value and *P*-value were not reported for the 'block* cohort' term.

can be observed at the scale of stands. For example, while the winter treatment decreased BT in the BAU and BRI stands from the North, it increased FP in BAU and decreased FP in BRI (Fig. 2). The interaction between germination cohort and the families was only significant for FP, suggesting similar responses to winter treatment among all families within a specific stand for the other five traits (Table 3).

In the four French regions, the partitioning of phenological variation was similar in the two germination cohorts (Tables S3 and S4, Supporting information). Variation in BT was partitioned among regions (fall: 26.3%, spring: 28.7%), among stands within regions (fall: 37.1%, spring: 37.6%) and among families within stands (fall: 19.0%, spring, 19.0%; Fig. 2). Extensive within-region variation was observed in Burgundy and Languedoc, whereas Brittany and North regions were less variable, displaying mainly early bolting families (Fig. 2). Results similar to BT were found for FT (Tables S3 and S4, Supporting information). For INT, FP, RP and FRR, a large fraction of phenotypic variation (from 50.3% to 86.7%) remained unexplained, suggesting high levels of phenotypic plasticity in those traits to uncontrolled micro-environmental variation in the greenhouse. For these phenological traits, the remaining phenotypic variation was mostly observed at local scales, that is, within stands and families. Stands from Burgundy and Languedoc were significantly more diverse than stands from Brittany and North for BT, INT and FT in the fall cohort and for BT, FT and RP in

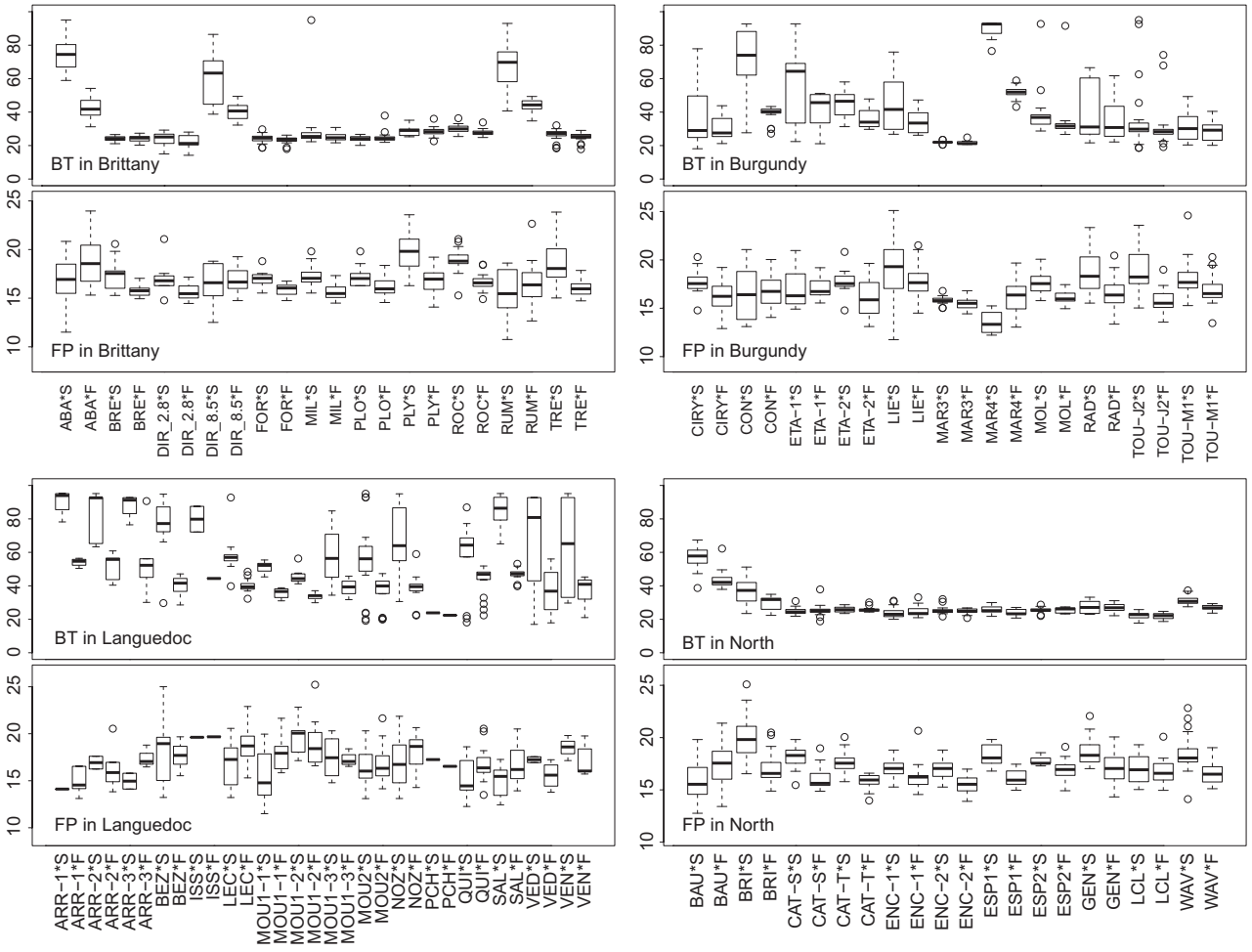


Fig. 2 Box-and-whiskers plots of diversity in bolting time (BT) and flowering period (FP) variation (both expressed in days) for each stand studied in each French region. Code names for each stand are given in Table S1 (Supporting information), and '*F*' and '*S*' denote fall and spring germination cohorts, respectively. On boxplots, the horizontal bold black line represents the median, the black hollow box around it the range from the first to the third quartile. The dashed black lines above and below the box extent to the 1.5 time the inner quartile ranges. The data points outside this interval are plotted as black circles.

the spring cohort (Fig. 2, Table S5, Supporting information).

Neutral genetic variation. A total of 758 French families were successfully genotyped for this set of 135 SNPs (Table S1, Supporting information). Extensive variation was observed across the 49 French stands for mean gene diversity (0–0.332) and percentage of *PL* (0–88.2%) (Table S6, Supporting information). A significant 'region' effect was detected on the six genetic diversity parameters tested simultaneously (Wilks' lambda F value = 2.26, P = 0.0051). No significant 'region' effect was detected for mean gene diversity, the percentage of *PL*, the mean number of observed alleles per locus, the mean allelic richness per locus and the number of *HG* (Table S6, Supporting information). A significant 'region' effect was detected for the number of *PHG* per stand (F = 5.50, P = 0.0026), with stands from Burgundy having on average more *PHG* than stands from Brittany, Languedoc and North after a Tukey's studentized range (HSD) test (Table S6, Supporting information). No significant 'habitat type' effect was detected on the six genetic diversity parameters tested simultaneously (Wilks' lambda F value = 0.74, P = 0.71). No genetic diversity parameter was significantly associated with either altitude or the number of families genotyped per stand (data not shown). Noteworthy was the significant, positive correlation observed between the coefficient of genetic variation in BT (or FT) and mean gene diversity (fall cohort: BT Spearman's rho = 0.506 P = 0.0003, FT Spearman's rho = 0.499 P = 0.0005; spring cohort: BT Spearman's rho = 0.445 P = 0.0017, FT Spearman's rho = 0.411 P = 0.0056).

A hierarchical analysis of molecular variation (AMOVA) revealed that genetic variation was partitioned among regions (9.1%), among stands within

regions (55.4%) and among families within stands (33.7%). A strong stand subdivision was observed within each region, with fixation index F_{ST} (from 0.48 in Burgundy to 0.61 for Brittany) similar to values reported in a previous study on other French stands (Le Corre 2005).

Ecological variation. Climate variables appeared more strongly correlated with each other than with edaphic variables and competition indices (Fig. S2, Supporting information). Both edaphic variables and competition indices were only weakly intercorrelated (Fig. S2, Supporting information). The variance within bioclimatic variables appeared to be mainly partitioned among regions, the 'region' effect explaining on average 90.3% of the bioclimatic variance in France (Fig. S3, Supporting information). In contrast, the variance of edaphic variables and competition indices was mostly partitioned among stands within regions, the 'region' effect explaining on average only 22.5% and 5.2% of the edaphic and competition variance in France, respectively (Fig. S3, Supporting information).

Investigation of the geographical scale of adaptive phenological variation

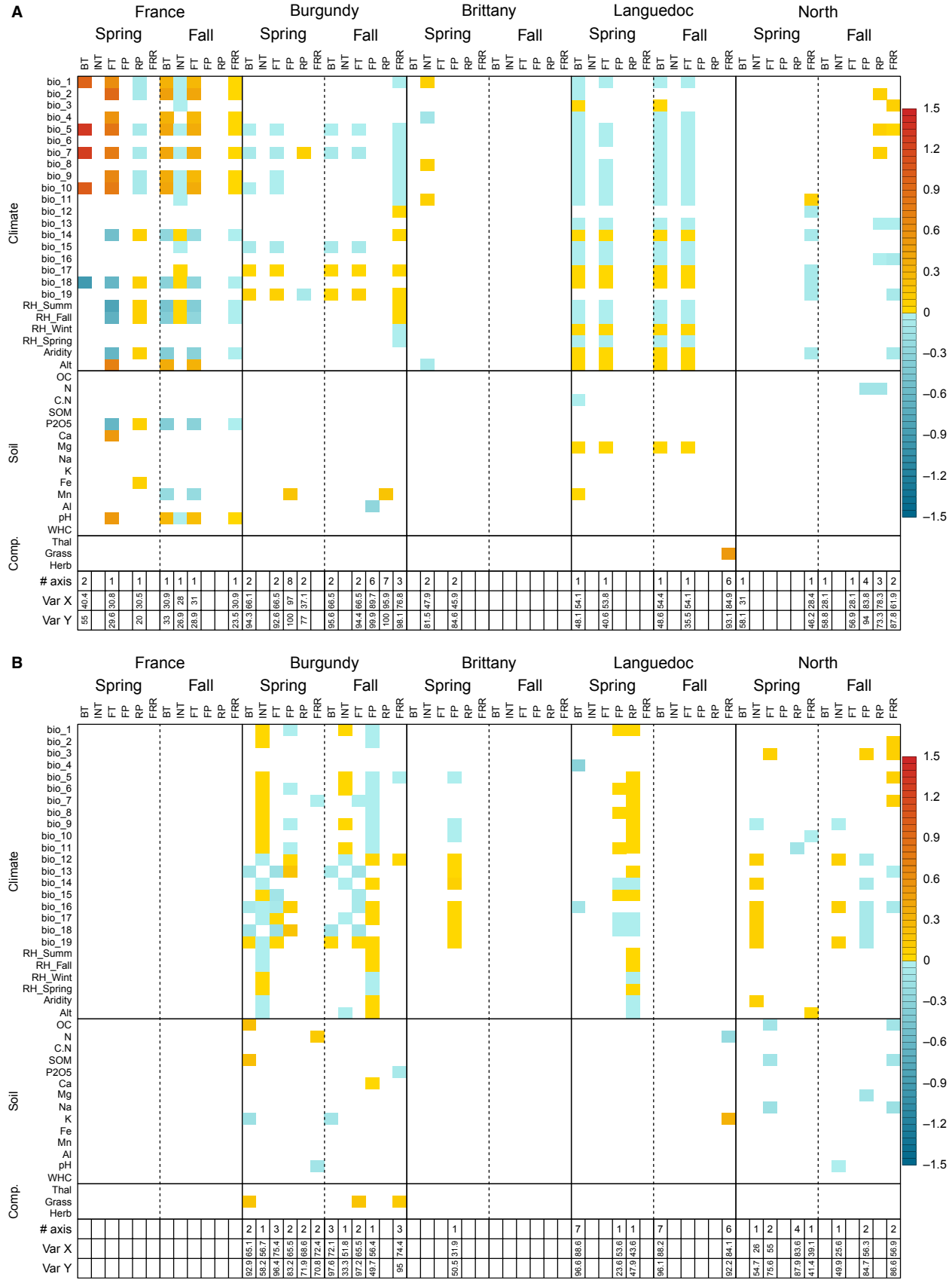
The percentage of phenological variation explained by genetic principal components with an eigenvalue ≥ 1 ($n = 31$) ranged from ~52% (BT and FT) to 8% (RP) in the fall cohort (mean = 27.2%) and from ~63% (BT and FT) to 10% (INT) in the spring cohort (mean = 35.6%).

Five main features characterized the relationships between phenological and ecological variation. First, nonselective processes may have generated spurious relationships between ecological factors and phenological variation. At the scale of France, all significant phenology–ecology relationships disappeared when neutral genetic similarities among families were accounted for (Fig. 3). In contrast, within regions, correcting for neutral genetic similarities revealed new significant phenology–ecology relationships (Fig. 3). Second, despite the small number of stands sampled within each region, the percentage of phenological variation explained by PLSR components was generally higher at the within region than at the broader scale of the France (Fig. 3). Third, at the within-region scale, edaphic factors and competition indices showed relationships with phenological traits that were as strong as those observed for climatic factors. For example, significant phenology–ecology relationships were detected between the density of grasses and BT (spring cohort), FT and FRR (fall cohort) in Burgundy (Fig. 3). Fourth, phenological traits generally displayed significant relationships with a particular ecological variable in only one or two of the four regions. For example, in the spring cohort, FT variation was significantly associated with climate variation in Burgundy and North, but not in Brittany and Languedoc (Fig. 3). Finally, the phenology–ecology relationships depended on the germination cohort season. In the North, significant relationships were detected between FP and climate variation only in the fall cohort. FP and RP displayed a significant relationship with climate only in the spring cohort in Languedoc.

Identification of the genomic regions associated with phenological variation from worldwide scale to local population scale

For BT and FT, an excess of low *P*-values due to confounding by population structure was found at the worldwide and European scales (Fig. S4, Supporting information). This excess of low *P*-values decreased at the French and regional (i.e. Burgundy) scale and was almost eliminated at the local scales. For INT, FP and FRR, the effect of population structure was mainly observed at the French and regional scales and was almost undetectable at the local scales (Fig. S4, Supporting information). No excess of low *P*-values due to confounding by population structure was detected for RP at any geographical scale. For each trait, cohort and geographical scale, the excess of low *P*-values detected from the Wilcoxon rank-sum analyses was eliminated from a mixed-model approach that takes genetic similarity among accessions into account (Fig. S4, Supporting information).

Fig. 3 Identification of the putative selective agents acting on phenology in *A. thaliana* in France. (A) Results from the Partial least square regression (PLSR), without control for neutral genetic variation. (B) Results of the PLSR obtained while accounting for neutral genetic variation. Each column corresponds to a combination of scale/region, germination cohort and phenological trait. The first 42 lines correspond to the ecological variables used in this study to characterize the 49 stands, separated in three categories: climate, soil and competition (Comp.). For each line and column, the coloured squares indicate significant regression coefficients (*P*-value < 0.05) and the colours represent the strength of the regression coefficient estimated between phenological variation and ecological variation. The last three lines correspond to the number of PLSR components retained after cross-validation ('axis'), the percentage of ecological variation explained by PLSR components ('Var *X*') and the percentage of phenological variation explained by PLSR components ('Var *Y*'). Because analyses were performed on standardized data, regression coefficients are comparable between panels A and B. See Table 1 for description of the ecological variables.



As illustrated by BT in the spring cohort, the identity of the genomic regions associated with natural variation depends on the geographical scale considered (Fig. 4 and Table S7, Supporting information). While a neat peak of association was detected in the vicinity of *DELAY OF GERMINATION 1* (*DOG1*) on chromosome 5 at the worldwide and European scales, GWA mapping revealed a unique and strong peak of association centred on the FT gene *FRIGIDA* (*FRI*) on chromosome 4 at the French, regional and local scales (Figs 4 and S5). In the TOU local population, the highest associated SNP is located within *FRI* (Fig. S5, Supporting information). In the fall cohort, association peaks centred on *FRI* were also found at the French and smaller geographical scales. In contrast, six and five new association peaks were detected at the Bonferroni threshold at the worldwide and European scales, respectively. Two of these new peaks correspond to the circadian gene *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) on chromosome 2 and *DWARF IN LIGHT 2* (*DFL2*) on chromosome 4. No FT candidate gene was found in a ~30-kb genomic region of chromosome 5 strongly associated with BT in the world and in Europe (i.e. region 6623–6652 kb on chromosome 5, Fig. S5, Supporting information). GWA mapping also revealed an association peak after the winter treatment that is only present in the local TOU population. This association peak located 120 kb upstream of the strongest association peak detected at the worldwide and European scales in the fall cohort contains no FT candidate genes (Fig. S5, Supporting information).

For FT, GWA mapping revealed the same genomic regions as for BT at the French and smaller geographical scales (Fig. S4, Supporting information). In contrast, no shared association peak was detected between BT and FT at the worldwide and European scales (Fig. S4, Supporting information). Two and nine association peaks were detected for FT in the fall cohort at the Bonferroni threshold at the worldwide and European scales, respectively. Only one association peak with no FT candidate gene is shared between these two scales. Two association peaks for FT in the fall cohort at the European scale colocalizes with *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 4* (*SPL4*) and *SPA1-RELATED 4* (*SPA4*). In the fall cohort, *SPL4* and *SPA4* were also found in the vicinity of strong association peaks detected for INT at both the worldwide and European scales (Fig. S4, Supporting information). The strongest SNP ($P = 4.02 \times 10^{-10}$) for INT in the fall cohort resides in the gene *AT1G52880*, a homolog of the petunia gene *NAM* that is involved in the development of the shoot. No association peak was detected at the Bonferroni threshold for INT either in the fall cohort at the French (or smaller geographical) scales or in the spring cohort.

Few association peaks have been detected for post-flowering traits. No association peak was detected for RP, whereas only three association peaks were detected at the Bonferroni threshold for FP across the two cohorts and the six geographical scales (Fig. S4, Supporting information). For FRR, GWA mapping revealed a unique peak of association in the region centred on *FRI* at the French, regional and local scales in the spring cohort, but only at the French and regional scales in the fall cohort (Fig. S4, Supporting information). No association peak was detected for FRR at the worldwide and European scales.

A list of FT candidate genes in the vicinity of 20 kb of the 100 most associated SNPs for each 'trait × cohort × geographical scale' combination is available in Table S7 (Supporting information).

Discussion

Extensive natural phenological variation at the local scale

Our phenological characterization of 49 French stands, worldwide accessions and accessions from the French RegMap revealed a relatively continuous distribution of bolting and flowering time, with an excess of intermediate values from populations collected in the South-East France. This stands in contrast to the distribution observed in current worldwide collections that show a bimodal distribution encompassing two major types in controlled growth conditions, that is, summer and winter annuals (Shindo *et al.* 2007). This phenological discrepancy suggests that thorough characterization of new populations should improve our picture of the natural variation observed in *A. thaliana*.

Our study also revealed highly polymorphic local populations at the phenological level, with some populations (like TOU) almost spanning the range of natural variation observed at the worldwide scale. For this reason, single accessions cannot reliably be used to characterize the phenotypes or genotypes present at ecologically distinct sites. At least three hypotheses can explain the high phenological polymorphism observed in natural populations of *A. thaliana* within France. First, nonselective processes may have played a major role in shaping the phenological variation at the within-population scale. This hypothesis is consistent with the patterns observed for BT and FT, which showed significant, positive relationships between phenological diversity and neutral genetic diversity. Second, phenological variation at the within-population scale may reflect adaptation to a fine-grained environmental variation, as previously observed in other plant species (interspecific competition: Turkington & Harper 1979a,b; herbivory: Schemske 1984). Third, the strength and direction of

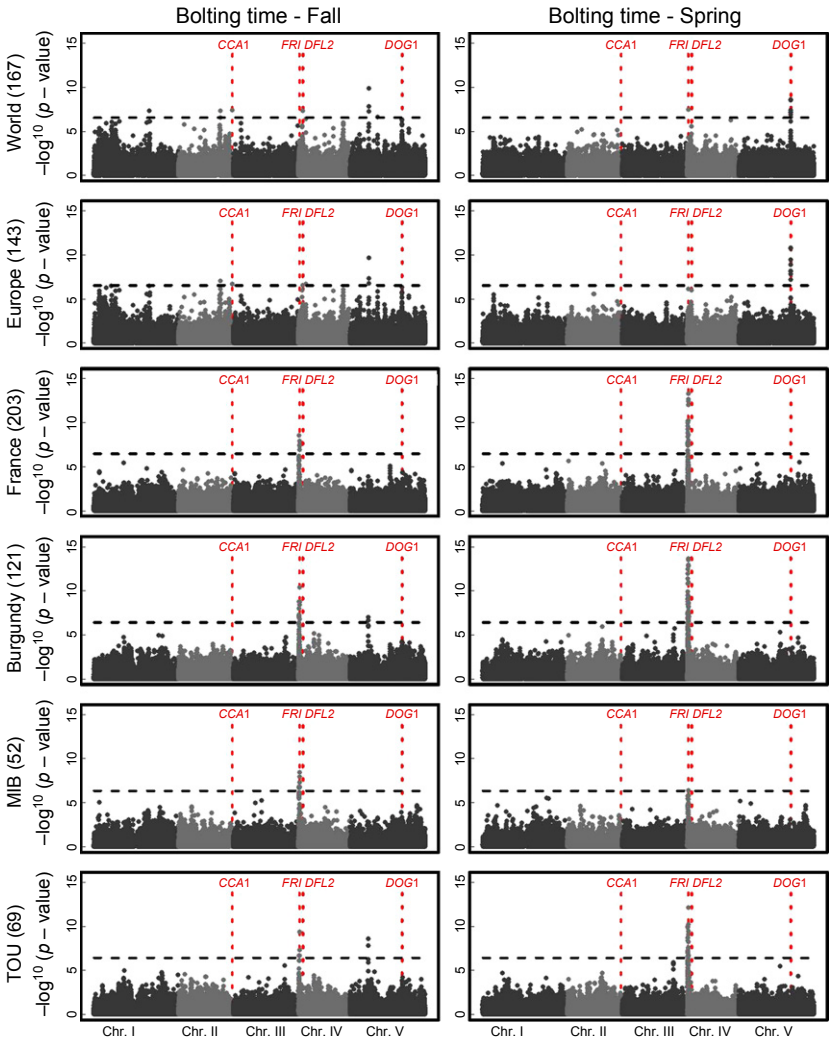


Fig. 4 Manhattan plots of the genome-wide association mapping (EMMAX) results for bolting time in each cohort and at different geographical scales. The *x*-axis indicates the position along each chromosome. The five chromosomes are presented in a row along the *x*-axis in different degrees of grey. The *y*-axis indicates the $-\log^{10}$ *p*-values using the EMMAX method. Minor allele relative frequency > 10%. The dashed line denotes the Bonferroni threshold. Vertical dotted lines indicate the physical position of the flowering time genes *CCA1*, *FRI*, *DFL2* and *DOG1*.

selection in natural populations may vary considerably over time (Siepielski *et al.* 2009). Although temporal dynamics of phenotypic selection may lead to the selection of plastic genotypes (Stomp *et al.* 2008), the cost of plasticity may promote the coexistence of specialist genotypes in the same population (Dewitt *et al.* 1998).

Geographical scale of adaptive phenological variation

The need to account for effects of nonselective processes is increasingly considered in the study of phenology–ecology relationships (Keller & Taylor 2008; Keller *et al.* 2009; Chun *et al.* 2011; Hancock *et al.* 2011; Méndez-Vigo *et al.* 2011; Kooyers & Olsen 2012; Lee & Mitchell-Olds 2012). In this study, phenology–ecology relationships were strongly affected by neutral genetic diversity at the scale of France, suggesting that phenology–ecology relationships could be spurious at this geographical scale. Interestingly, controlling for neutral genetic diversity revealed new phenology–ecology relationships at the regional scale. This is a similar phenomenon to the GWA peaks that become evident only after population structure is controlled in mixed-model association mapping (Kang *et al.* 2008). Here, correction for nonselective processes may reduce the phenotypic variance within populations or within types of environment, revealing otherwise nonsignificant relationships.

Similar to studies on the identification of genomic regions associated with natural phenotypic variation by GWA mapping, we should acknowledge that our approach may produce false negatives after controlling for the effects of nonselective processes (Brachi *et al.* 2010), that is, true phenology–ecology associations that overlap with neutral genetic diversity. Significant phenology–ecology relationships detected without controlling for the effects of nonselective processes can be validated in experiments in controlled conditions (see below).

The ecological characterization of 49 French stands suggests an overlap of environmental grains, with variation for edaphic and intra- and interspecific

competition factors observed at a finer scale than climate variables. This is in agreement with the ecological characterization of natural populations of *A. thaliana* in the Iberian Peninsula (Montesinos-Navarro *et al.* 2011). As in other plant species (Bischoff *et al.* 2006; Becker *et al.* 2008; Manel *et al.* 2010), the selective agents shaping the adaptive population differentiation for phenological traits appear to act at multiple scales in *A. thaliana*, but seem to be stronger at the within-region level.

The apparent lack of relationship between some traits and ecological factors may originate from (i) a lack of statistical power; (ii) reduced phenological or ecological variation in some within-region samples; (iii) uncharacterized ecological factors acting on phenology, especially biotic agents in natural populations like herbivores (Lennartsson *et al.* 1997), pathogen attacks (Roux *et al.* 2010), predispersal seed predators (Elzinga *et al.* 2007) or pollinators (Hoffmann *et al.* 2003); (iv) the absence of relationships between phenological variation and fitness in some locations; (v) fine-scale environmental differentiation within populations and/or (vi) selection for phenotypic plasticity in natural habitats (Pérez de la Vega 1996); and (vii) inaccurate estimates of phenological variation in our greenhouse conditions. It is important to note that our estimation of natural variation in phenology was done under controlled standard greenhouse conditions. Although phenological traits in *A. thaliana* often display strong genotype by environment interactions when tested across greenhouse and common garden conditions (Bergelson & Roux 2010; but see Méndez-Vigo *et al.* 2012), we believe our measures are meaningful as they reflect genetically fixed variation among families.

In a recent study based on an interconnected mapping population of 117 recombinant inbred lines (RILs), the influence of reproductive timing on fitness at the phenotypic level was found to greatly differ among four field sites across the native European range of *A. thaliana* (Fournier-Level *et al.* 2013). It would be worth measuring phenological traits on the 49 French stands of this study in more ecologically realistic conditions to check whether an equivalent amount of phenological variation explained by nonselective processes is consistent with the results of this study and whether adaptive phenological variation estimated in field settings is associated with others ecological factors.

Diversity of selective agents acting on phenology among French accessions

The putative selective agents acting on phenological traits were clearly dependent on the region, the trait considered and the germination cohort; the latter confirming the importance of germination cohort in determining the environmental conditions experienced by plants after germination in *A. thaliana* (Donohue 2002; Donohue *et al.* 2005). The geographical heterogeneity in the relationships between phenological traits and ecological variables is also consistent with results from a recent study, suggesting the selection acts on different traits and loci in different locations across Europe (Fournier-Level *et al.* 2013).

Although rarely mentioned in *A. thaliana*, relationships with edaphic factors and plant–plant interactions were as strong as correlations with climatic factors, which were previously suggested to be selective agents on BT in *A. thaliana* (Stinchcombe *et al.* 2004). Naturally occurring variation in edaphic factors and long-term application of different fertilizer treatments are known to act as selective pressures on the phenology of plants (Snaydon & Davies 1982; Rajakaruna & Bohm 1999; Kittelson & Maron 2001; Antonovics 2006), but have seldom been tested in *A. thaliana*.

Significant phenotype–ecology correlations, however, are only suggestive of the role of ecological factors in selecting on phenology. Experiments in controlled conditions are clearly needed for validation. In a previous study, both estimates of seed production and experimental evolution were used to estimate the adaptive values of phenological traits in two stressful environments, that is, water stress and interspecific competition with grasses (Brachi *et al.* 2012). The relationships of RP and FRR with aridity and interspecific competition with grasses that were detected in this study are congruent with the predictions of our previous study. This congruence suggests that aridity and interspecific competition with grasses are true selective agents acting on phenology in natural populations in France.

Complementarity of GWA mapping studies at different geographical scales

The geographical scale of adaptive phenological variation observed in the 49 stands led us to test whether the identification of the genomic regions associated with phenological variation differed across geographical scales. In a first attempt, combining both worldwide accessions and accessions from the French RegMap indicates that the genomic regions associated with phenological variation appear to depend on the geographical scale considered. Strong association peaks were detected at the continental and local geographical scales. The variants underlying those peaks may reflect adaptation to coarse-grained and fine-grained ecological variation, respectively.

Performing GWA using sets of accessions spanning different geographical scales also proved to be

promising in resolving major limitations of GWA mapping with regard to population structure, rare alleles and allelic heterogeneity. First, as expected from the pattern of isolation by distance observed across the species range of *A. thaliana* (Platt *et al.* 2010a), confounding by population structure was greatly reduced at small geographical scales. This was especially true for phenotypic traits like BT or FT whose natural variation overlaps with population structure at the worldwide scale (Zhao *et al.* 2007). While dual linkage-association mapping has been shown to reduce the rate of false positives and false negatives in GWA studies (Brachi *et al.* 2010), running GWA mapping in regional or local panels may also circumvent confounding by population structure.

Second, association studies often lack power to detect rare variants. However, rare variants at the worldwide scale may be common in local populations, making them easier to detect in association studies. For example, using a regional collection of wild *A. thaliana* genotypes in the Iberian Peninsula, Sánchez-Bermejo *et al.* (2012) recently identified a novel *cis*-regulatory *FLC* polymorphism located only in the northeast of Spain that is associated with an increase in vernalization sensitivity. In our study, BT in the fall cohort revealed a significant association peak located on chromosome 5 that was only detected in the TOU population, suggesting that variants may be so rare that they are only present in one local population.

Third, because the FT gene *FRI* is a classic example of allelic heterogeneity in *A. thaliana* (Atwell *et al.* 2010), we expected to detect this gene when running GWA mapping at geographical scales smaller than the worldwide scale. Despite the description of 13 *FRI* nonfunctional alleles in France (Le Corre *et al.* 2002; Le Corre 2005; Shindo *et al.* 2005), *FRI* was detected using the French mapping panel and local French populations. Two hypotheses can explain our ability to map *FRI* in the French mapping panel despite the presence of numerous *FRI* nonfunctional alleles. First, one of the nonfunctional allele may be much more prevalent than the other nonfunctional alleles. Second, one polymorphism located in *FRI* may be shared by several nonfunctional alleles (Platt *et al.* 2010b).

In this study, we used all the worldwide and French accessions for which both seeds and 214 k SNPs data were available at the beginning of the experiment. The ongoing genomic characterization of the 49 stands will soon reveal whether (i) we obtain similar biological conclusions as with the French RegMap panel and (ii) the genomic regions associated with phenological variation differ across the four French regions and overlap with imprints of selection.

Conclusion

While next-generation sequencing (NGS) technologies will facilitate the identification of causal polymorphisms underlying natural variation of complex traits (Brachi *et al.* 2011), it should not be forgotten that the genetics of adaptation may largely depend on the environmental grain both at the spatial and at the temporal scales (Kopp & Hermisson 2007, 2009; Roux & Reboud 2007; Roux *et al.* 2008). By investigating putative selective pressures acting on phenological traits, we improved our understanding of the geographical scale of adaptive variation in *A. thaliana*. Our results suggest that phenological variation at small geographical scales might be adaptive and that different phenological traits could be under selection in different regions. While GWA mapping is powerful in detecting common genes underlying natural variation at a worldwide scale, it suffers from limitations like confounding by population structure, rare alleles and allelic heterogeneity. In our study, the geographical scale of adaptive variation suggested for phenological traits and the scale at which associations were detected are consistent. Mapping in regional panels of accessions or even in local populations may resolve limitations of GWA studies. Overall our study suggests that mapping panels that span geographical areas over which phenotypic variation appears to be adaptive have the potential to unravel the genetics underlying adaptive complex trait and allow reconstruction of the adaptive walks that natural populations follow towards local phenotypic optima.

Acknowledgements

We are grateful to Eric Imbert and Valérie Le Corre for their assistance in locating Arabidopsis stands. Special thanks are given to Adeline Courseaux, Cédric Glorieux and Cécile Godé for their assistance during the greenhouse experiment and for DNA extraction protocols. We also thank Alexander Platt for sharing his script for estimating the number of *HG* per stand. This study was supported by a PhD fellowship from the French Ministry of Higher Education and Research to B.B. and R.V., an ANR BLANC grant for the QUANTIREX project (NT09_473214), and by a Dropkin fellowship and an NIH grant to J. Bergelson.

References

Adler D (2005) *vioplot: Violin Plot. R Package Version 0.2*. Available from http://CRAN. R-project. org/package= vioplot.

Anastasio AE, Platt A, Horton M *et al.* (2011) Source verification of mis-identified *Arabidopsis thaliana* accessions. *The Plant Journal*, **67**, 554–566.

Antonovics J (2006) Evolution in closely adjacent plant populations X: long-term persistence of prereproductive isolation at a mine boundary. *Heredity*, **97**, 33–37.

Atwell S, Huang YS, Vilhjálmsson BJ *et al.* (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*, **465**, 627–631.

Becker U, Colling G, Dostal P, Jakobsson A, Matthies D (2006) Local adaptation in the monocarpic perennial *Carlin vulgaris* at different spatial scales across Europe. *Oecologia*, **150**, 506–518.

Becker U, Dostal P, Jorritsma-Wienk LD, Matthies D (2008) The spatial scale of adaptive population differentiation in a wide spread, well-dispersed plant species. *Oikos*, **117**, 1865–1873.

Belotte D, Curien JB, Maclean RC, Bell G (2003) An experimental test of local adaptation in soil bacteria. *Evolution*, **57**, 27–36.

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

Bischoff A, Crémieux L, Smilauerova M *et al.* (2006) Detecting local adaptation in widespread grassland species- the importance of the scale and local plant community. *Journal of Ecology*, **94**, 1130–1142.

Bomblies K, Yant L, Laitinen RA *et al.* (2010) Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *PLoS Genetics*, **6**, e1000890.

Brachi B, Faure N, Horton M *et al.* (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genetics*, **6**, e1000940.

Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biology*, **12**, 232.

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.

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

Chun YJ, Le Corre V, Bretagnolle F (2011) Adaptive divergence for a fitness-related trait among invasive *Ambrosia artemisiifolia* populations in France. *Molecular Ecology*, **20**, 1378–1388.

Clark RM, Schweikert G, Toomajian C *et al.* (2007) Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science*, **317**, 338–342.

Conner JK (2010) Natural selection in plants 151 years after the origin: introduction. *International Journal of Plant Sciences*, **171**, 927–929.

Dewitt T, Sih A, Wilson D (1998) Costs and limits of phenotypic plasticity. *Trends in Ecology & Evolution*, **13**, 77–81.

Donohue K (2002) Germination timing influences natural selection on life-history characters in *Arabidopsis thaliana*. *Ecology*, **83**, 1006–1016.

Donohue K (2005) Seeds and seasons: interpreting germination timing in the field. *Seed Science Research*, **15**, 175–187.

Donohue K, Dorn LA, Griffith C *et al.* (2005) The evolutionary ecology of seed germination of *Arabidopsis thaliana*: variable natural selection on germination timing. *Evolution*, **59**, 758–770.

Elzinga JA, Atlan A, Biere A *et al.* (2007) Time after time: flowering phenology and biotic interactions. *Trends in Ecology & Evolution*, **22**, 432–439.

Fournier-Level A, Korte A, Cooper MD *et al.* (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science*, **334**, 86–89.

Fournier-Level A, Wilczek AM, Cooper MD *et al.* (2013) Paths to selection on life history loci in different natural environments across the native range of *Arabidopsis thaliana*. *Molecular Ecology*, doi:10.1111/mec.12285.

Franks SJ, Sim S, Weis AE (2007) Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *Proceedings of the National Academy of Sciences*, **104**, 1278–1282.

Gaut B (2012) *Arabidopsis thaliana* as a model for the genetics of local adaptation. *Nature Genetics*, **44**, 115–121.

Geladi P, Kowalski BR (1986) Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, **18**, 5.

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.

Hancock AM, Brachi B, Faure N *et al.* (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science*, **334**, 83–86.

Hereford J (2009) A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist*, **173**, 579–588.

Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, **169**, 2335–2352.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hintze JL, Nelson RD (1998) Violin plots: a box plot-density trace synergism. *The American Statistician*, **52**, 181–184.

Hoffmann MH, Bremer M, Schneider K *et al.* (2003) Flower visitors in a natural population of *Arabidopsis thaliana*. *Plant Biology*, **5**, 491–494.

Horton MW, Hancock AM, Huang YS *et al.* (2012) Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nature Genetics*, **44**, 212–216.

Huang X, Wei X, Sang T *et al.* (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*, **42**, 961–967.

Inouye DW (2008) Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. *Ecology*, **89**, 353–362.

Kang HM, Zaitlen NA, Wade CM *et al.* (2008) Efficient control of population structure in model organism association mapping. *Genetics*, **178**, 1709–1723.

Kang H, Sul J, Service S *et al.* (2010) Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, **42**, 348–402.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Keller S, Taylor D (2008) History, chance and adaptation during biological invasion: separating stochastic phenotypic evolution from response to selection. *Ecology Letters*, **11**, 852–866.

Keller SR, Sowell DR, Neiman M, Wolfe LM, Taylor DR (2009) Adaptation and colonization history affect the evolution of clines in two introduced species. *New Phytologist*, **183**, 678–690.

Kittelson PM, Maron JL (2001) Fine-scale genetically based differentiation of life-history traits in perennial shrub *Lupinus arboreus*. *Evolution*, **55**, 2429–2438.

Kooyers NJ, Olsen KM (2012) Rapid evolution of an adaptive cyanogenesis cline in introduced North American white clover (*Trifolium repens L.*). *Molecular Ecology*, **21**, 2455–2468.

Kopp M, Hermisson J (2007) Adaptation of a quantitative trait to a moving optimum. *Genetics*, **176**, 715–719.

Kopp M, Hermisson J (2009) The genetic basis of phenotypic adaptation i: fixation of beneficial mutations in the moving optimum model. *Genetics*, **182**, 233–249.

Laine AL (2005) Spatial scale of local adaptation in a plant-pathogen metapopulation. *Journal of Evolutionary Biology*, **18**, 930–938.

Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution*, **37**, 1210–1226.

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.

Le Corre V, Roux F, Reboud X (2002) DNA polymorphism at the FRIGIDA gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. *Molecular Biology and Evolution*, **19**, 1261–1271.

Lee CR, Mitchell-Olds T (2012) Environmental adaptation contributes to gene polymorphism across the *Arabidopsis thaliana* genome. *Molecular Biology and Evolution*, **29**, 3721–3728.

Leimu R, Fischer M (2008) A meta-analysis of local adaptation in plants (A Buckling, Ed,). *PLoS One*, **3**, e4010.

Lennartsson T, Tuomi J, Nilsson P (1997) Evidence for an evolutionary history of overcompensation in the grassland biennial *Gentianella campestris* (Gentianaceae). *The American Naturalist*, **149**, 1147–1155.

Linhart YB, Grant MC (1996) Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics*, **23**, 237–277.

Manel S, Poncet B, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular Ecology*, **19**, 3824–3859.

de Meeûs T, Goudet J (2007) A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. *Infection, Genetics and Evolution*, **7**, 731–735.

Méndez-Vigo B, Picó FX, Ramiro M, Martínez-Zapater J, 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.

Méndez-Vigo B, Gomaa NH, Alonso-Blanco C, Xavier Picò F (2012) Among- and within-population variation in flowering time of Iberian *Arabidopsis thaliana* estimated in field and glasshouse conditions. *New Phytologist*, **197**, 1332–1343.

Merilä J, Sheldon BC, Kruuk LE (2001) Explaining stasis: microevolutionary studies in natural populations. *Genetica*, **112-113**, 199–222.

Mevik BH, Wehrens R (2007) The pls package: principal component and partial least squares regression in R. *Journal of Statistical Software*, **18**, 1–25.

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

Mitchell-Olds T, Willis JH, Goldstein DB (2007) Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics*, **8**, 845–856.

Montesinos-Navarro A, Wig J, Pico F, Tonsor S (2011) *Arabidopsis thaliana* populations show clinal variation in a climatic gradient associated with altitude. *The New phytologist*, **189**, 282–294.

Munguía-Rosas MA, Ollerton J, Parra-Tabla V, De-Nova JA (2011) Meta-analysis of phenotypic selection on flowering phenology suggests that early flowering plants are favoured. *Ecology Letters*, **14**, 511–521.

Olson-Manning CF, Wagner MR, Mitchell-Olds T (2012) Adaptive evolution: evaluating empirical support for theoretical predictions. *Nature Reviews Genetics*, **13**, 867–877.

Orr HA (2005) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, **6**, 119–127.

Parachnowitsch AL, Caruso CM (2008) Predispersal seed herbivores, not pollinators, exert selection on floral traits via female fitness. *Ecology*, **89**, 1802–1810.

Paradis E (2010) pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics (Oxford, England)*, **26**, 419–439.

Pérez de la Vega M (1996) Plant genetic adaptedness to climatic and edaphic environment. *Euphytica*, **92**, 27–38.

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.

Platt A, Horton M, Huang YS *et al.* (2010a) The scale of population structure in *Arabidopsis thaliana* (J Novembre, Ed,). *PLoS Genetics*, **6**, e10000843.

Platt A, Vilhjálmsson B, Nordborg M (2010b) Conditions under which genome-wide association studies will be positively misleading. *Genetics*, **186**, 1045–1052.

Rajakaruna N, Bohm BA (1999) The edaphic factors and patterns of variation in *Lasthenia californica* (Asteraceae). *American Journal of Botany*, **86**, 1576–1596.

Rathcke B, Lacey EP (1985) Phenological patterns of terrestrial plants. *Annual Review of Ecology and Systematics*, **16**, 179–214.

Roux F, Reboud X (2007) Herbicide resistance dynamics in a spatially heterogeneous environment. *Crop Protection*, **26**, 335–341.

Roux F, Paris M, Reboud X (2008) Delaying weed adaptation to herbicide by environmental heterogeneity: a simulation approach. *Pest Management Science*, **64**, 16–29.

Roux F, Gao L, Bergelson J (2010) Impact of initial pathogen density on resistance and tolerance in a polymorphic disease resistance gene system in *Arabidopsis thaliana*. *Genetics*, **185**, 283–291.

Sánchez-Bermejo E, Méndez-Vigo B, Picó FX, Martinez-Zapater JM, Alonso-Blanco C (2012) Novel natural alleles at FLC and LVR loci account for enhanced vernalization responses in *Arabidopsis thaliana*. *Plant, Cell & Environment*, **35**, 1672–1684.

Sandring S, Ågren J (2009) Pollinator-mediated selection on floral display and flowering time in the perennial herb *Arabidopsis lyrata*. *Evolution*, **63**, 1292–1300.

Schemske DW (1984) Population structure and local selection in *Impatiens pallida* (Balsaminaceae), a selfing annual. *Evolution*, **38**, 817–832.

Shindo C, Aranzana MJ, Lister C *et al.* (2005) Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of Arabidopsis. *Plant Physiology*, **138**, 1163–1173.

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.

Snaydon RW, Davies TM (1982) Rapid divergence of plant populations in response to recent changes in soil conditions. *Evolution*, **36**, 289–297.

Stinchcombe JR, Weinig C, Ungerer M *et al.* (2004) A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proceedings of the National Academy of Sciences, USA*, **101**, 4712–4717.

Stomp M, van Dijk M, van Overzee HT *et al.* (2008) The timescale of phenotypic plasticity and its impact on competition in fluctuating environments. *The American Naturalist*, **172**, 169–185.

Turkington R, Harper JL (1979a) The growth, distribution and neighbour relationships of *Trifolium Repens* in a permanent pasture: II. Inter-and intra-specific contact. *Journal of Ecology*, **67**, 219–230.

Turkington R, Harper JL (1979b) The growth, distribution and neighbour relationships of *Trifolium Repens* in a permanent pasture: IV. Fine-scale biotic differentiation. *Journal of Ecology*, **67**, 245–254.

Van Dijk H, Boudry P, McCombie H, Vernet P (1997) Flowering time in wild beet (*Beta vulgariss* sp. maritima) along a latitudinal cline. *Acta Oecologica*, **18**, 47–60.

Wagmann K, Hautekèete NC, Piquot Y *et al.* (2012) Seed dormancy distribution: explanatory ecological factors. *Annals of Botany*, **110**, 1205–1219.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Zhao K, Nordborg M, Marjoram P (2007) Genome-wide association mapping using mixed-models: application toGAW15Problem 3. *BMC Proceedings*, **1**(Suppl 1), S164.

B.B., J.C. and F.R conceived the study; B.B., R.V., N.F., D.R. and F.R. developed experimental materials; B.B., R.V. and F.R. performed the experiments; B.B., R.V. N.H., Y.P. M.P., J.B. and F.R. designed the analyses; B.B., R.V. and F.R. performed the analyses; B.B., R.V., J.B. and F.R. wrote the manuscript, revised by all authors.

Data accessibility

Phenotypic data, ecological variables and genotypes are available in the Dryad database: doi:10.5061/dryad.07s25.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Plant material.

Table S2 List of the 328 a priori candidate genes for flowering time.

Table S3 Structure of phenological variation in France in the Fall cohort.

Table S4 Structure of phenological variation in France in the Spring cohort.

Table S5 Genetic coefficient of phenological variation for BT, INT, FT, FP, RP and FRR for each region and cohort.

Table S6 Mean values for genetic diversity parameters of the 49 French stands.

Table S7 List of flowering time candidate genes close to the 100 most associated SNPs for each 'trait × cohort × geographical scale' combination.

Fig. S1 Violin plots of natural variation for the interval between bolting and flowering (INT), flowering time (FT) and reproductive period (RP).

Fig. S2 Pairwise matrix of similarity between ecological variables.

Fig. S3 Box-and-whiskers plots of ecological variation in French stands.

Fig. S4 Manhattan plots of the genome-wide association mapping (EMMAX) results and quantile-quantile plot of *P*-values (negative logarithm, Wilcoxon and EMMAX) for BT, INT, FT, FP, RP and FRR in each cohort and at different geographic scales.

Fig. S5 Close up of the association peaks at the beginning of chromosome 5 (left) and around *FRIGIDA* (right) at three different geographical scales for bolting time.