GENETIC BASIS OF ADAPTATION IN *ARABIDOPSIS THALIANA*: LOCAL ADAPTATION AT THE SEED DORMANCY QTL *DOG1*

Ilkka Kronholm,[1,2] F. Xavier Picó,[3] Carlos Alonso-Blanco,[4] Jérôme Goudet,[5] and Juliette de Meaux[1,6,7]

[1] *Department Genetics and Plant Breeding, Max Planck Institute for Plant Breeding Research, DE-50829 Cologne, Germany*

[3] *Departamento de Ecología Integrativa, Estación Biológica de Doñana (EBD), Consejo Superior de Investigaciones Científicas (CSIC), Av. Américo Vespucio s/n, 41092 Sevilla, Spain*

[4] *Departamento de Genética Molecular de Plantas, Centro Nacional de Biotecnología (CNB), Consejo Superior de Investigaciones Científicas (CSIC), Madrid 28049, Spain*

[5] *Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland*

[6] *Institute of Evolution and Biodiversity, University of Münster, DE-48149 Münster, Germany*

[7] *E-mail: juliette.de.meaux@uni-muenster.de*

Local adaptation provides an opportunity to study the genetic basis of adaptation and investigate the allelic architecture of adaptive genes. We study *DELAY OF GERMINATION 1* (*DOG1*), a gene controlling natural variation in seed dormancy in *Arabidopsis thaliana* and investigate evolution of dormancy in 41 populations distributed in four regions separated by natural barriers. Using F_{ST} and Q_{ST} comparisons, we compare variation at *DOG1* with neutral markers and quantitative variation in seed dormancy. Patterns of genetic differentiation among populations suggest that the gene *DOG1* contributes to local adaptation. Although Q_{ST} for seed dormancy is not different from F_{ST} for neutral markers, a correlation with variation in summer precipitation supports that seed dormancy is adaptive. We characterize dormancy variation in several F_2-populations and show that a series of functionally distinct alleles segregate at the *DOG1* locus. Theoretical models have shown that the number and effect of alleles segregatin at quantitative trait loci (QTL) have important consequences for adaptation. Our results provide support to models postulating a large number of alleles at quantitative trait loci involved in adaptation.

KEY WORDS: ***DELAY OF GERMINATION 1* (*DOG1*), F_{ST}, population genetics, Q_{ST}.**

The genetic basis of local adaptation is one of the fundamental questions in evolutionary biology. Local adaptation occurs if selection is strong enough relative to gene flow and favors different phenotypes in different populations (Kawecki and Ebert 2004).

There is a long history of research into local adaptation, especially in sessile organisms such as plants (Turesson 1922, 1925; Clausen et al. 1941). A recent meta-analysis of local adaptation in plants revealed that it is rather common but not universal, with large populations being more often locally adapted than smaller ones (Leimu and Fischer 2008).

In theoretical models, the fitness effects of new beneficial mutations fixing in a single population are expected to follow

[2] Current address: Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, EH9 3JT Edinburgh, United Kingdom.

an exponential distribution (Orr 1998). This prediction has been supported at least qualitatively by QTL-mapping experiments in multicellular organisms (Orr 2005). In the case of multiple populations adapting to distinct optima, the distribution of fitness effects of fixed beneficial mutations is no longer strictly exponential, because alleles can migrate among populations (Griswold 2006). Few QTLs of moderate to large effect on a single adaptive trait have been predicted to explain most of the phenotypic differences between locally adapted populations (Griswold 2006). Overall, however, there is a dearth of data concerning the allelic diversity segregating at QTLs controlling adaptive traits in natural populations. Are QTLs often biallelic or multiallelic? Simulations have shown that the allelic architecture can have substantial effects on adaptation (Yeaman and Guillaume 2009). Are the same QTLs involved in local adaptation throughout the species range, or are there enough potential loci that QTLs involved in adaptation are chosen randomly by selection each time?

We study the model organism *Arabidopsis thaliana* (Brassicaceae) because in this species the genetic basis of adaptation can be studied in great detail (Mitchell-Olds and Schmitt 2006). *Arabidopsis thaliana* is an annual plant that is capable of self-fertilization and outcrossing. We focus on *DELAY OF GERMINATION 1* (*DOG1*), the first cloned locus controlling quantitative variation in seed dormancy in *A. thaliana* (Bentsink et al. 2006). *DOG1* provides a unique opportunity to study allelic diversity at an adaptive locus. Indeed, it was found to colocalize with QTLs for germination timing and fitness in the field (Huang et al. 2010). The timing of germination influences not only seedling survival but also the expression of other life-history traits later in the plant's life cycle (Evans and Cabin 1995; Donohue 2002; Wilczek et al. 2009). Overall evidence that it is under strong selection is compelling in *A. thaliana* (Griffith et al. 2004; Donohue et al. 2005b) and many other plant species (Marks and Prince 1981; Kalisz 1986; Biere 1991; Gross and Smith 1991). The timing of germination is determined to a large extent by the duration and strength of seed dormancy, a physiological process preventing the seed to germinate in the presence of permissive conditions for growth (Finch-Savage and Leubner-Metzger 2006). In *A. thaliana,* germination cannot be induced until dormancy has been released by a process called after ripening (Finch-Savage and Leubner-Metzger 2006). Broad genetic variation is present within *A. thaliana* for the length of after-ripening requirement (Evans and Ratcliffe 1972; Alonso-Blanco et al. 2003), but its variation throughout the species range has not been described. In addition, nucleotide variation segregating at the *DOG1* locus has not been studied and its relevance for local adaptation has not been examined.

To investigate whether a gene is involved in local adaptation, a comparative analysis of genetic divergence across various loci measured by F_{ST} can be useful (reviewed in Holsinger and Weir 2009). Demographic processes are expected to influence allele frequencies and phenotypic diversity, masking the action of geographically heterogeneous selection. Yet, the effect of demography is expected to be the same for the whole genome. Natural selection, by contrast, is predicted to have the greatest influence on allele frequencies of the loci under selection (Charlesworth et al. 1997; Beaumont 2005). At the phenotypic level, genetic divergence in quantitative traits can also be quantified using Q_{ST}, a measure analogous to F_{ST} (Spitze 1993). If divergence is greater at quantitative traits than at neutral markers ($Q_{ST} > F_{ST}$), it is possible to make inferences about the action of geographically heterogeneous selection and local adaptation (Lande 1992; Whitlock 1999). To maximize our chance to hit the spatial scale at which local adaptation can be detected, we used a sample of *A. thaliana* populations collected in four broad regions separated by natural barriers. Genetic variation in this sample has been characterized for single nucleotide polymorphism (SNP) and microsatellite markers distributed genome wide by Kronholm et al. (2010).

We report here a comparative analysis of phenotypic and genetic variation and address the following questions: (1) Is there a signature of local adaptation in *DOG1*? (2) Do we see two major functional alleles segregating at *DOG1* or an allelic series with weaker and stronger alleles? (3) Can we identify the ecological forces driving adaptation at the seed dormancy locus *DOG1*?

Methods

PLANT MATERIAL

For the population genetic study, we used 289 individuals collected in 41 populations and described in Kronholm et al. (2010). Populations are grouped in four geographically separated regions, Spain (7 populations, 70 genotypes), France (15 populations, 109 genotypes), Norway (13 populations, 64 genotypes), and Central Asia (Fig. 1). For candidate gene association (see below), 57 individuals collected by M. Koornneef (MPI, Cologne) in and around Wageningen, the Netherlands, were added to increase statistical power (346 genotypes in total), but not included in the analysis of spatial patterns of variation.

GENOTYPING

All individuals were genotyped for 20 microsatellite markers and 137 SNP markers by Kronholm et al. (2010). These 157 markers were considered to be "neutral markers" because they are distributed genome wide. Based on preliminary results, the first exon of *DOG1* appeared to harbor the greatest number of polymorphisms (M. Debieu, unpubl. results). Thus, we designed primers to amplify and sequence the first exon of *DOG1,* which is 393 bp long. The primers used were D1E1—5^t-AAA CAC AAA CAC GCA AAC CA and i1re—5^t-GCC GCA CCG TAC TGA CTA CC. PCR and Sanger sequencing was performed using standard protocols, the sequencing primer was i1re. When needed PCR

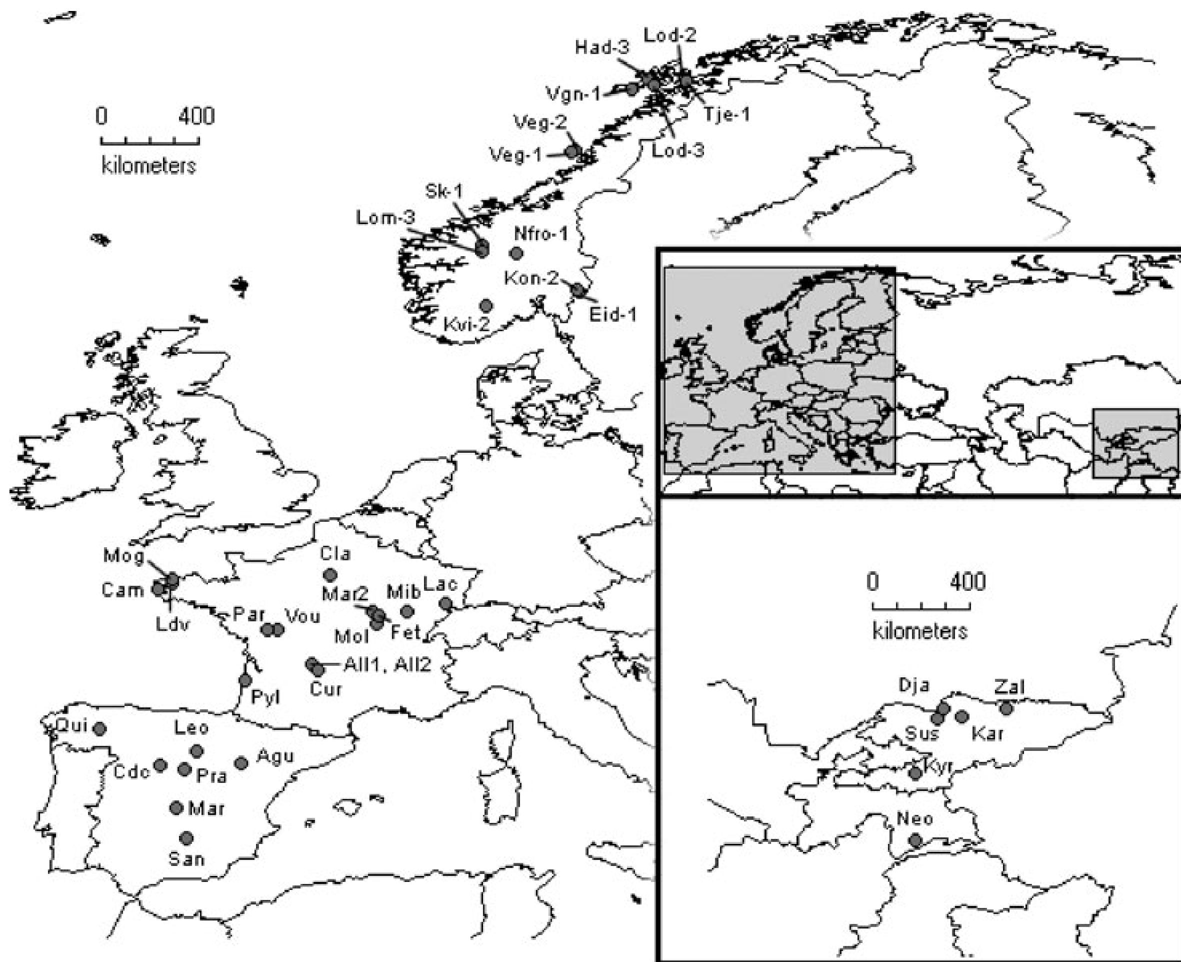

Figure 1. **Map of the populations used in this study. Inset shows the Central Asian populations.**

products were cloned to a pCR®4-TOPO Vector using TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA, USA) using manufacturers instructions. Electropherograms were inspected for errors and sequences could be aligned unambiguously using BioEdit 7.0.5.3 (Hall 1999). Sequences have been deposited to GenBank with accession numbers HQ128719-HQ129004. Some genotypes had a large length polymorphism in the first intron of *DOG1*. The primers ee1f—5′-CGA CGG CTA CGA ATC TTC AG and i1re (see above) were used to amplify this polymorphism. The presence or absence of this insertion was scored by resolving the PCR products on a 3% agarose gel.

We genotyped additional SNP markers distributed in the vicinity of *DOG1* on the *A. thaliana* chromosome 5. For this, we first designed primers to amplify and sequence 9 loci around *DOG1* to discover SNP markers. Primer sequences and positions are presented in Table S1. The SNP discovery panel was composed of 16 different genotypes from different regions, two from Wageningen, four from Spain, three from France, three from Norway, and four from Central Asia. From each of the nine sequenced fragments, 1–4 SNPs were chosen for genotyping. Pyrosequencing assays were designed with the Assay Design Software 1.0.6 (Qiagen, Hilden, Germany). The SNPs were genotyped using pyrosequencing (Fakhrai-Rad et al. 2002), with the PSQ 96MA Pyrosequencing system (Qiagen, Hilden, Germany). Primer sequences for the SNP markers are given in Table S2. For biotinylated primers, we used the universal primer method of Aydin et al. (2006). A universal sequence was added to the 5′ end of the specific primers. In the PCR reaction four primers were used, the specific primers and the universal primers with the appropriate universal primer labeled with biotin (Table S2). For some assays the four-primer reaction did not work efficiently, so two separate PCR reactions had to be performed.

POPULATION GENETICS ANALYSES

All statistical analyses were done using the statistical computing language R (R Development Core Team 2006) unless otherwise stated. Measurements of genetic diversity, Nei's gene diversity (Hs) and allelic richness (AR) were calculated using FSTAT 2.9.3 (Goudet 2001) for the microsatellite markers. AR is a measure of the number of alleles that corrects for sample size differences between populations. To compare genetic diversity between groups

of populations, a permutation test, which permutes over populations, implemented in FSTAT was used.

We estimated F_{ST} for the 137 SNP markers following Weir and Cockerham (1984) as in Kronholm et al. (2010). For loci with more than two alleles such as microsatellites and *DOG1,* we used ϕ_{ST}, which takes distances between alleles into account (Michalakis and Excoffier 1996; Excoffier 2007). ϕ_{ST} was implemented via R-scripts written by IK. ϕ_{ST} is not correlated with heterozygosity for multiallelic loci in contrast to Wright's F_{ST} estimate (Kronholm et al. 2010). ϕ_{ST} is in fact the best estimator of comparative analysis of F_{ST} across markers of different types because it corrects for differences in mutation rate (or heterozygosity) between loci (Slatkin 1995; Kronholm et al. 2010; Whitlock 2011). To compare ϕ_{ST} value of *DOG1* or Q_{ST} of dormancy to neutral markers, we used the empirical distribution of microsatellites and SNPs, 157 markers in total, and compared the F_{ST} or Q_{ST} values of interest to the quantiles of this distribution.

We constructed a haplotype network of the first exon of *DOG1* using TCS version 1.21 (Clement et al. 2000). TCS implements a maximum parsimony method to infer the evolutionary relationships between the haplotypes. In analyses that required an outgroup, we used the first exon sequence of *DOG1* from *A. lyrata*. Sequence diversity indices were calculated using DnaSP version 4.10.4 (Rozas et al. 2003). DnaSP was also used to estimate the minimum number of recombination events in exon 1 of *DOG1*.

PHENOTYPING AND QUANTIFICATION OF SEED DORMANCY

For the common garden experiment all lines were first multiplied by selfing in the greenhouse under the same environmental conditions to remove any possible maternal effects. All plants were grown in the same climatized greenhouse set at +20°C during the day and +18°C during the night. Natural light was supplemented with lamps to reach a photoperiod of 16 h of light when necessary. Plants were grown in a soil mixture (70% peat, 20% sand, and 10% clay) in 6 cm diameter round pots with one plant in each pot. The common garden experiment was started in the fall of 2007. Three plants (replicates) from the selfed progeny of each genotype (346 in total) were grown in a randomized block design (1038 plants in total). Seventeen plants died before flowering, and this resulted in complete loss of phenotypic information for three genotypes. Because the maternal environment can affect seed dormancy (Munir et al. 2001; Donohue et al. 2005a), seeds for all genotypes should mature in similar environmental conditions and thus flower simultaneously. To synchronize flowering time, we planted the genotypes in three different groups. The seeds were water imbibed and stratified (cold treatment at +4°C) in the dark for 4 days to induce germination. Thereafter, they were potted and moved to the greenhouse. After 14 days, the plants were vernalized for 28 days, in a climate chamber at +4°C, under short days (8 h light) and then moved back to the greenhouse. Due to shifted planting of very late flowering genotypes and 4-week rosette vernalization, we were able to synchronize flowering so that most seeds matured during March–April 2008. See Supporting information and Figure S1 for details. Ripening of the siliques (fruits in Brassicaceae) was assessed visually by observing a color change from green to brown. *Arabidopsis thaliana* produces siliques over a long period of time, and these were harvested when there were enough ripened siliques on the plant (usually siliques were harvested from the main stem). After ripening occurred in room temperature and seeds were stored in paper bags. On the day the seeds were harvested, the germination experiment was started.

To measure seed dormancy, we measured the ability of the seeds to germinate in a time course experiment performed for each seed batch (replicate) following Alonso-Blanco et al. (2003). For each time point, a sample of approximately 50–100 seeds were sown on a small petri dish, with filter paper. Water (700 μl) was added to imbibe seeds. Then the petri dishes were transferred to a growth cabinet with a temperature of +25°C during the day (12 h light period) and +20°C during the night. After 1 week, the number of germinated and dormant seeds was counted using a stereomicroscope. Seeds were scored as germinated when the root tip had protruded the seed coat. For each seed batch germination tests were performed immediately after harvest (0 weeks) and then subsequently 1, 2, 4, 8, 16, 24, 32, 40, and 52 weeks after harvest. When a seed batch was germinating at 100% in two consecutive tests, it was considered to have lost dormancy. The germination experiment was stopped after 52 weeks. A viability test was performed for seed batches that had not reached 100% germination following Cadman et al. (2006) (see Supporting information). We found that these seeds were still viable.

To quantify seed dormancy for a given replicate, we followed Alonso-Blanco et al. (2003). We fitted a binomial regression through the germination data for each replicate, using a logit link function (Venables and Ripley 2002). From the fitted function, we calculated the time for which the probability of germination is 0.25, 0.5, or 0.75, referred as D25, D50, and D75. This is a measure of the time of dry storage required to reach a given probability of germination (weeks of dry storage, WODS). This transformation is particularly well suited for time course experiments measuring variations in proportions (Crawley 2005). Three estimates were used, to capture different aspects of dormancy. D25 is a time point at which early germinants appear, and D75 is a time point at which germination is nearly completed. We also used a linear model to quantify seed dormancy and got very similar results (see Supporting information for details).

Genotype means were estimated using a linear model $y_{ijk} = \mu + g_i + b_j + e_{ikj}$, where y_{ijk} is the phenotypic observation of the kth replicate of the ith genotype in block j, μ is the overall

mean, g_i is the genotypic effect of the ith genotype, b_j is the block effect for the jth block, and e_{ikj} is the residual. Genotypic means are obtained from the term $\mu + g_i$ and thus possible block effects are subtracted from the genotype means. In general, block effects were absent or very small and do not affect any biological conclusions of this study (see Supporting information). To investigate differences between populations and regions we used a linear model $y_{ijk} = \mu + r_i + p_{ij} + e_{ijk}$, where y_{ijk} is the mean phenotype of the kth genotype in the jth population within the ith region, μ is the overall mean, r_i is the effect of the ith region, p_{ij} is the effect of the jth population nested within the ith region, and e_{ijk} is the residual. The estimation of heritabilities and Q_{ST} is described at the end of the section Methods.

ASSOCIATION BETWEEN *DOG1* AND SEED DORMANCY

We tested whether genetic variation in *DOG1* is associated with phenotypic variation in seed dormancy. F_{ST} between these *A. thaliana* populations is usually high (Kronholm et al. 2010). To avoid spurious marker–phenotype associations that arise when some alleles are associated with certain populations, population structure has to be corrected for. We performed an association test using mixed model association following Yu et al. (2006) using the PK_T method of Stich et al. (2008) to control for population structure and kinship of individuals within populations. Thus related genotypes are accounted for. The SNP markers were used in determining the optimal value of T for the kinship matrix (Table S3). Detailed description of the model is given in Supporting information. Correcting for population structure is important in our sample, without a correction many spurious associations would be observed (Figure S2).

To increase statistical power to detect significant association between *DOG1* alleles and seed dormancy, we included a sample of accessions from Wageningen. This increased the sample size to 346 genotypes. We also tested for association within the different geographic regions. In this way, associations may be revealed that are masked on the larger sample by the segregation of distinct haplotypes with similar function. After determining the optimal T value, the association test for *DOG1* using the mixed model was done using the program TASSEL 2.0.1 (Bradbury et al. 2007). Sequence haplotypes of the first exon of *DOG1* were used as different alleles in the association study. Because there were multiple tests done due to multiple alleles, we corrected for multiple testing using the Bonferroni–Holm correction (Holm 1979).

LINKAGE ANALYSIS BETWEEN *DOG1* AND SEED DORMANCY IN F_2 POPULATIONS

To confirm some of the candidate gene associations, we constructed F_2 populations where alleles that had significant associations to dormancy were segregating. The crosses were: All2-1 (haplotype 1) × Fet-6 (haplotype 5), both from France, (size of F_2-population $N = 133$); Cam-4 (haplotype 15) × Fet-6 (haplotype 5), both from France ($N = 126$); Cam-4 (haplotype 15) × All2-1 (haplotype 1), both from France ($N = 145$); Kon-2-2 (haplotype 19) × Fet-6 (haplotype 5), from Norway and France, respectively ($N = 121$); Kon-2-2 (haplotype 19) × Nfro-1-4 (haplotype 18), both from Norway ($N = 122$). F_1 individuals were allowed to self to produce F_2 seeds. Leaves were collected from F_2 individuals for DNA extraction after flowering. To genotype *DOG1* in the F_2 populations, pyrosequencing assays were designed for *DOG1* SNP markers distinguishing segregating haplotypes. Primers for these assays are given in Table S2. Dormancy was measured in F_3 seeds collected in a common garden experiment similar to the one described above (see Supporting information) and associated with the *DOG1* genotype of the corresponding F_2 individual. For this, the F_2 populations were analyzed using a linear model $y_{ij} = \mu + g_i + e_{ij}$, where y_{ij} is the phenotypic observation of the jth line in genotypic class i, g_i is the effect of the ith *DOG1* genotypic class, and e_{ij} is the residual. Following Lynch and Walsh (1998), we denote genotypic values of the genotypes D_1D_1, D_1D_2, and D_2D_2 as 0, $(1 + k)a$, and $2a$, respectively. Taking the estimates of the different genotypes from the linear model, the effect of allele D_2 is obtained from $a = (D_2D_2 - D_1D_1) / 2$ and the dominance coefficient from $k = ((D_1D_2 - D_1D_1) / a) - 1$.

ESTIMATION OF HERITABILITIES AND Q_{ST}

Broad sense heritability, which measures the proportion of observed variation that is genetic variation, was estimated as $H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_E^2}$. Because *A. thaliana* is predominantly self-fertilizing, genetic variance components can be estimated in a straightforward manner from our common garden experiment. Dominance variation is not defined because all lines are homozygous. Assuming complete selfing, variation between replicates within genotypes allows estimating σ_E^2, the environmental variance component, and variation between genotypes allows estimating σ_G^2, the genetic variance component. Q_{ST} measures how quantitative genetic variation is partitioned between populations, and was estimated as $Q_{ST} = \frac{\sigma_{GB}^2}{\sigma_{GB}^2 + \sigma_{GW}^2}$ (Bonnin et al. 1996), where σ_{GB}^2 is genetic variation between populations and σ_{GW}^2 is genetic variation within populations. We used two different methods to estimate the variance components: a linear mixed effects model in R, from which variance components were estimated using REML (Venables and Ripley 2002) or a Bayesian method of estimating variance components implemented in WinBUGS 1.4.3 (Lunn et al. 2000). The model itself stays the same for these two methods, only the method of estimating variance components differs. For heritability, variance components were estimated from a model $y_{ijk} = \mu + b_i + g_j + e_{ijk}$, where y_{ijk} is the phenotypic observation of the kth replicate of the jth genotype in the ith block, b_i is the block effect

for the ith block, g_j is the genotypic effect for the jth genotype. Blocks were included as fixed effects and genotypes as random effects. For Q_{ST} this model was extended such that $y_{ijkl} = \mu + b_i + p_j + g_{jk} + e_{ijkl}$, where p_j is the population effect and other terms are the same as in the previous model, for block i, population j, genotype k nested within population, and replicate l nested within genotype. Blocks are included as fixed effects and population and genotype are random effects. Specification of the WinBUGS models was done following O'Hara and Merilä (2005). Details of WinBUGS model specification and priors used are in the Supporting information. Pairwise Q_{ST} between populations was estimated using REML, while Q_{ST} over several populations was estimated using WinBUGS, to obtain an interval estimate for Q_{ST}.

COVARIATION OF DORMANCY AND DOG1 VARIATION WITH CLIMATIC VARIABLES

To find possible causes for selection, we examined if trait values of the populations are related to any environmental variables. We used the program DIVA-GIS 5.2.0.2 (Hijmans et al. 2001) in combination with the 2.5 arc-minute resolution current global climate environmental data (Hijmans et al. 2005), available at www.worldclim.org. We extracted 10 climatic parameters for our populations: latitude, altitude, annual mean temperature, temperature seasonality, mean temperatures of the warmest or coldest quarters, annual precipitation, precipitation seasonality, and precipitations over the warmest or the coldest quarters. These data were an average of the conditions in the past 50 years. Thereafter, we built a linear model that explains variation in plant traits by climatic conditions. Population means were used in this analysis (see Supporting information for details). After identifying that summer precipitation is correlated with dormancy (see results), we calculated a pairwise matrix of absolute differences between populations for this variable. A matrix of pairwise F_{ST} values of *DOG1* or neutral markers between populations were correlated to a matrix of environmental distances. Mantel tests were used to assess the statistical significance of the correlations, implemented in R-package *vegan* (Oksanen et al. 2007).

Results

POPULATION GENETICS OF *DOG1*

Twenty-two haplotypes could be defined for *DOG1* on the basis of the sequence of its first exon and a large insertion in the first intron (Table S4). A summary of haplotype frequencies by region is presented in Table 1. In total, 11 haplotypes were present in the Spanish populations, six in the French and Norwegian populations, and three in the Central Asian populations. Different haplotypes were at high frequency in different regions. In

Table 1. **Summary of *DOG1* haplotype frequencies in different regions.**

Region	Haplotype 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Spain	0.06	–	–	–	0.22	0.06	0.01	0.04	0.2	0.13	0.06	–	–	0.16	0.04	–	0.01	–	–	–	–	–
France	0.52	–	0.06	0.04	0.04	–	–	–	–	–	–	–	–	–	0.3	0.01	–	–	–	–	0.03	–
Norway	–	0.23	–	–	–	–	–	–	–	–	–	0.08	0.09	–	–	–	–	0.25	0.33	0.02	–	–
Central Asia	–	–	–	0.39	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.43	0.17
Overall	0.21	0.05	0.02	0.08	0.07	0.01	0.003	0.01	0.05	0.03	0.01	0.02	0.02	0.04	0.12	0.004	0.004	0.06	0.07	0.004	0.08	0.03

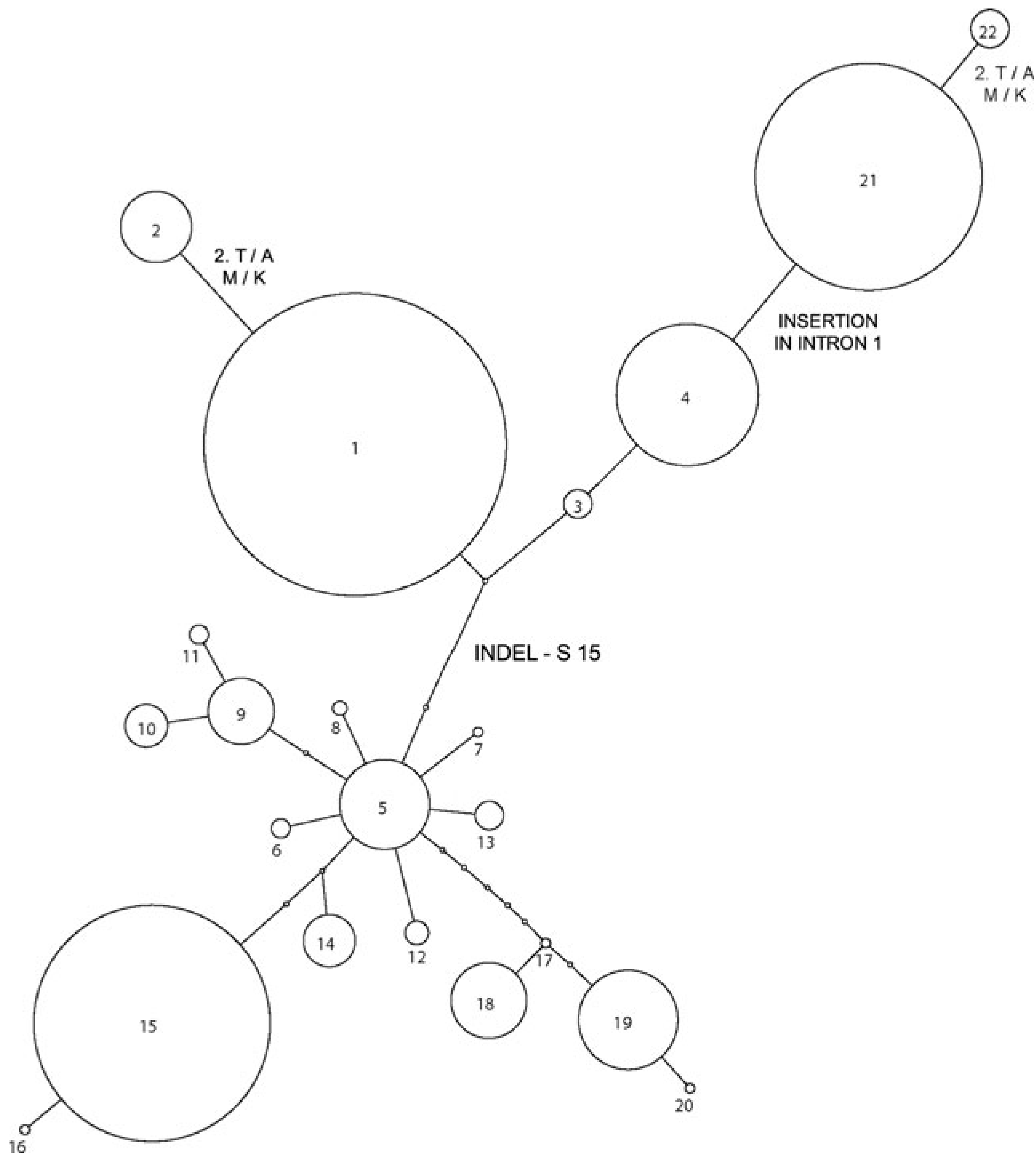

Figure 2. **Haplotype network of *DOG1*. Each node represents a single mutation; the radius of the circle is proportional to the frequency of that haplotype. The sample was the population sample and some accessions from Wageningen.**

Spain, haplotypes 5, 9, 10, and 14 were at moderate frequencies, while other haplotypes present in Spain were at low frequencies. In France there were two predominant haplotypes, 1 and 15. In Norway three haplotypes, 2, 18, and 19 were at high frequencies. Finally, in the Central Asian populations haplotypes 4 and 21 were at nearly equal high frequencies and 22, also was at moderate frequency.

DOG1 haplotype diversity (*Hd*) was 0.87 in Spain, 0.62 in France, 0.77 in Norway, and 0.64 in Central Asia. For microsatellite markers AR was 2.269, 1.720, 1.245, and 1.383 for the Spanish, French, Norwegian, and Central Asian populations, respectively. Except when comparing the Central Asian populations to those of Norway and France, differences in AR were significant ($P < 0.05$, 1000 permutations). Only one recombination event could be detected between haplotypes 2 and 22, which have a mutation at position 2 (see Supporting information for details). These two haplotypes segregate at low frequency in our sample and are found in different regions (Table 1).

The haplotype network of *DOG1* is presented in Figure 2. The *A. lyrata* outgroup cannot be joined to the network with 95% confidence. The Spanish haplotypes are mostly found in the central part of the network, while haplotypes from other regions occupy the peripheral parts of the network. The closely related haplotypes 18 and 19, which are found only in Norway and at high frequency, are connected to haplotype 5 by a long branch. Haplotypes 15 and 1 that are common in France are not closely related to each other, unlike haplotypes 4, 21, and 22 which are common in the Central Asian populations. The common

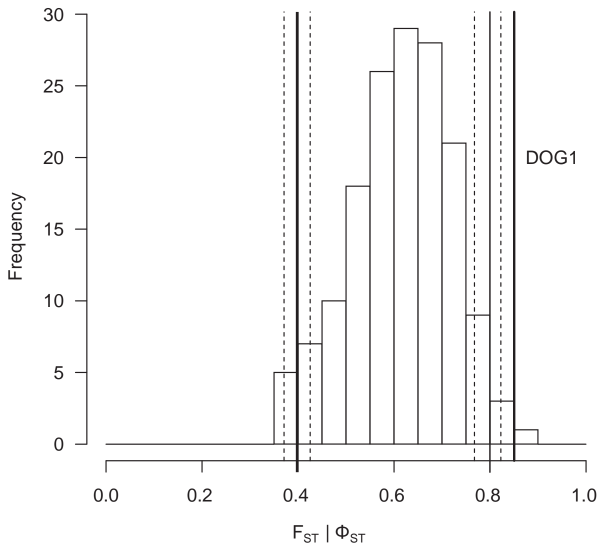

Figure 3. **Genetic differentiation in *DOG1* was compared to the F_{ST} distribution of 157 neutral markers, microsatellites and SNP markers. The histogram is the distribution of F_{ST} for neutral markers. Solid line is the 8_{ST} of *DOG1*, dashed lines denote the quantiles of the neutral distribution. Values on x-axis are F_{ST} values for SNP markers, 8_{ST} values for microsatellites and *DOG1* haplotypes.**

haplotypes in France are present in Spain at low frequencies (Table 1, Fig. 2).

While the restricted geographic distribution of *DOG1* haplotypes reveals the possibility that there is local adaptation, this could be a result of drift and restricted gene flow. Therefore, we tested whether genetic differentiation in *DOG1* is higher than expected by chance alone. 8_{ST} for *DOG1* was 0.8502 for all 35 European populations and 0.8769 when the Central Asian populations were included. These values lie in the tails of the neutral marker distribution (Fig. 3). The probability of observing equal or greater values was 0.0064 for the European populations and 0.0127 when all populations were included. When considering only the Spanish and the French populations or the French and the Norwegian populations, 8_{ST} for *DOG1* was 0.7432 and 0.9094, respectively. In both of these cases, *DOG1* lies at the tail of the neutral marker distribution and there are only two markers with higher F_{ST} values. Within each of the geographic regions 8_{ST} values for *DOG1* are 0.4810, 0.7421, 0.9459, and 0.9559 for Spain, France, Norway, and Central Asia, respectively. However, in these cases *DOG1* does not have a different value from the F_{ST} of neutral markers.

If *DOG1* is under selection, population genetic theory predicts that there should be a peak of F_{ST} at the position of *DOG1,* when genetic divergence is viewed along the chromosome (Charlesworth et al. 1997). We tested this by using SNP haplotypes around the position of *DOG1*. It is clear that 8_{ST} peaks at the position of *DOG1* and then decreases to levels expected from the neutral markers (Fig. 4). Within population genetic variance is included in F_{ST} measurements. This could be a problem if there are different amounts of within population genetic variance in different chromosomal regions, due to differences in the amount of crossing over, for example (Charlesworth et al. 1997). Therefore, we also calculated the between-population heterozygosity ($H_T - H_S$). The results show that there is also a peak for between-population heterozygosity at the position of *DOG1* (Fig. 4). This suggests that the high genetic differentiation is specific to *DOG1.*

GENETIC VARIATION IN SEED DORMANCY

Heritability values for seed dormancy are presented in Table 2. The heritability, calculated over all genotypes in the population

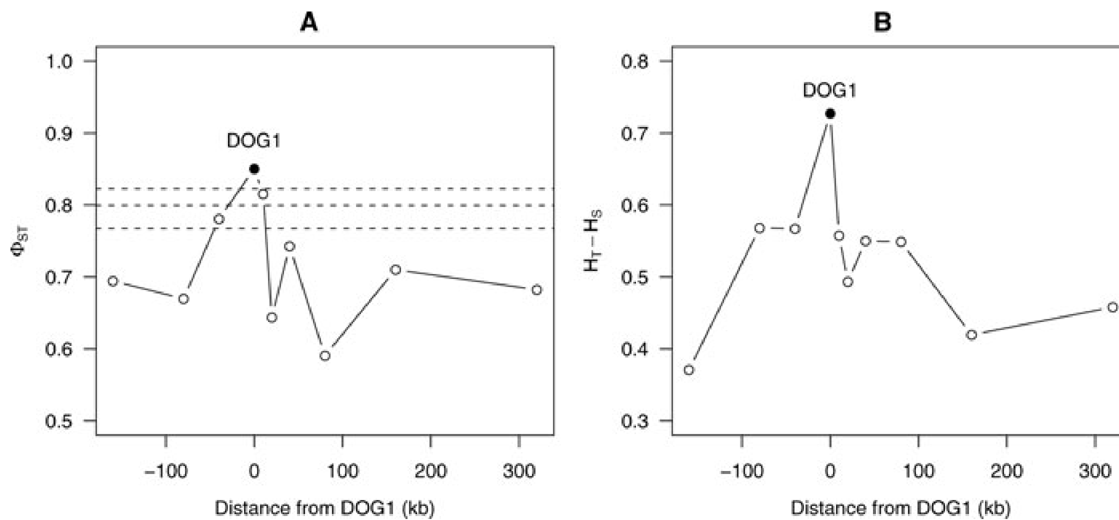

Figure 4. **Genetic differentiation along chromosome V at the position of *DOG1*. In panel (A), 8_{ST} along the chromosome. The dashed lines are the upper quantiles of the neutral F_{ST} distribution. (B) $H_T - H_S$ along the chromosome.**

Table 2. **Heritabilities (H^2) for seed dormancy in different regions, 2.5% and 97.5% denote the limits of the 95% highest posterior density interval. D25, D50, and D75 are seed dormancy measurements, defined as time taken to reach 25%, 50%, or 75% germination, respectively.**

Region	Trait	H^2	2.5%	97.5%
All	D25	0.7829	0.7431	0.8191
	D50	0.8058	0.7694	0.8387
	D75	0.7859	0.7464	0.8218
Spain	D25	0.6000	0.4709	0.7159
	D50	0.6961	0.5877	0.7889
	D75	0.7423	0.6463	0.8229
France	D25	0.6926	0.6002	0.7726
	D50	0.7507	0.673	0.8174
	D75	0.7129	0.6271	0.788
Norway	D25	0.5844	0.4476	0.7072
	D50	0.737	0.6345	0.8238
	D75	0.7922	0.7055	0.8641
Asia	D25	0.8348	0.7508	0.9002
	D50	0.8468	0.7677	0.908
	D75	0.7785	0.671	0.8648

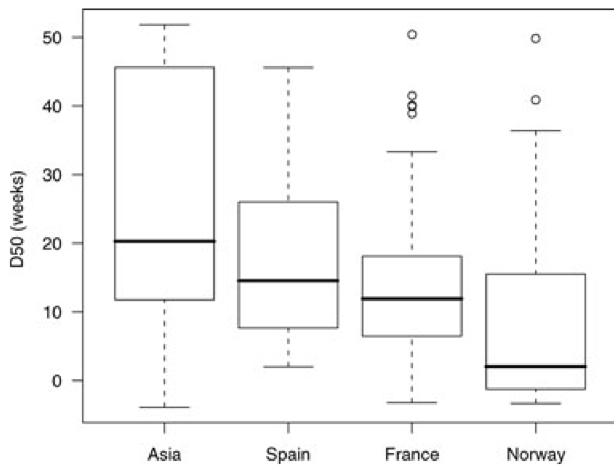

Figure 5. **Seed dormancy (D50, time taken to reach 50% germination) box plots for the four geographic regions. Data are genotype means.**

sample, was around 0.8. The heritability remained high when calculated over genotypes within each of the regions (Table 2) and REML and the Bayesian methods gave nearly identical results. High heritability values show that the observed differences in seed dormancy between the different genotypes were mostly due to genetic variation.

There were significant differences in seed dormancy both between regions, and between populations within regions (Fig. 5, Table 3). The strongest seed dormancy was observed in Central Asian populations, where some genotypes were still dormant after one year of after ripening. Among the European regions, seed

Table 3. **Analysis of variance table for seed dormancy in different regions. Data are genotype means for D50, time taken to reach 50% germination.**

	df	F-value	P-value
Region	3	42.872	$<2.2\times10^{-16}$
Population within region	37	12.830	$<2.2\times10^{-16}$
Residuals	245		

dormancy decreases from Southern to Northern Europe (Fig. 5). However, within all regions, there was a substantial amount of variation with differences among population means being often greater than differences between region means (Table S5). Within each region, there were some populations that had levels of dormancy different from the rest of the populations as well as low genetic variation. This can be an indication of local adaptation (see Supporting information).

ASSOCIATION BETWEEN *DOG1* AND SEED DORMANCY

We tested whether allelic variation in *DOG1* was associated with phenotypic variation in seed dormancy, by performing a candidate gene association study with *DOG1*. First, we tested each *DOG1* haplotype for association with each of the three seed dormancy estimates (D25, D50, and D75) in the whole sample. We also performed an analysis of genetic association within each of the regions (Table 4). Haplotype 4, which is present in French, Dutch, and Central Asian populations, was the most strongly associated allele. It was associated with increased dormancy. Haplotype 4 has the highest marker R^2 values explaining up to 9% of the variance in the French populations. Haplotypes 6, 9, and 10 were weakly associated with dormancy when only the Spanish populations were considered, although they are not significant after correcting for multiple testing (Table 4). Haplotype 13 was weakly associated with an increase of dormancy in the whole sample. Haplotype 15 was associated with decreased dormancy in the French populations. Although, the effect of haplotype 15 is seen only for D25, it explains comparatively large amount of the variance, 5% (Table 4). Haplotypes 18 and 19 were also weakly associated with decreased dormancy in the whole sample. Haplotypes 21 and 22 were both associated with decreased dormancy in the Central Asian populations.

LINKAGE ANALYSIS OF *DOG1* SEED DORMANCY

To confirm some of the associations and to examine allelic effects of different *DOG1* haplotypes, we performed linkage analysis between *DOG1* haplotype and seed dormancy in a set of F_2 populations generated by crossing parents carrying distinct alleles. *DOG1* had Mendelian segregation in all crosses except in the cross Cam-4 × Fet-6 (haplotypes 15 and 5), where segregation

Table 4. **Associations for *DOG1* haplotypes and seed dormancy. Associations have been tested for all three time points for the whole sample and within each region. For multiple testing corrections the Bonferroni–Holm method was used.**

Haplotype	Sample	Trait	P-value	P-adjusted	Effect direction	Marker R^2
2	All	D50	0.014	0.245	Increase	0.006
		D75	0.002	0.042	Increase	0.010
4	All	D50	1.10×10^{-7}	2.42×10^{-6}	Increase	0.028
		D25	2.55×10^{-8}	5.61×10^{-7}	Increase	0.027
		D75	4.69×10^{-6}	1.03×10^{-4}	Increase	0.021
	France	D50	7.05×10^{-4}	0.005	Increase	0.049
		D25	2.35×10^{-5}	1.65×10^{-4}	Increase	0.088
		D75	0.005	0.038	Increase	0.032
6	Spain	D25	0.050	0.450	Increase	0.017
9	Spain	D50	0.045	0.451	Decrease	0.016
		D25	0.023	0.253	Decrease	0.022
10	Spain	D50	0.023	0.255	Decrease	0.020
		D25	0.023	0.253	Decrease	0.022
		D75	0.025	0.270	Decrease	0.019
13	All	D50	0.008	0.160	Increase	0.007
		D25	0.016	0.304	Increase	0.005
		D75	0.010	0.187	Increase	0.007
15	France	D25	0.002	0.011	Decrease	0.050
18	All	D50	0.010	0.184	Decrease	0.007
		D25	0.014	0.288	Decrease	0.005
		D75	0.007	0.124	Decrease	0.008
19	All	D50	0.015	0.252	Decrease	0.006
		D25	0.016	0.304	Decrease	0.005
		D75	0.039	0.620	Decrease	0.005
21	Central Asia	D50	0.001	0.003	Decrease	0.025
		D25	0.005	0.009	Decrease	0.019
		D75	1.84×10^{-4}	5.52×10^{-4}	Decrease	0.039
22	All	D50	1.63×10^{-5}	3.42×10^{-4}	Decrease	0.019
		D25	7.43×10^{-6}	1.56×10^{-4}	Decrease	0.018
		D75	4.35×10^{-4}	0.009	Decrease	0.013
	Central Asia	D50	0.001	0.003	Decrease	0.025
		D25	0.003	0.008	Decrease	0.021
		D75	2.60×10^{-4}	5.52×10^{-4}	Decrease	0.037

was distorted with an excess of homozygous lines ($\chi^2 = 11.5$, df $= 2$, $P = 0.003$).

DOG1 cosegregated with dormancy in all crosses except in the cross between haplotypes 18 and 19 (Table 5), thereby confirming the significant associations reported above. Haplotypes 15 and 1 both decreased dormancy relative to haplotype 5 in F_2 populations. When crossed with each other, F_2 individuals with haplotype 15 had a slightly lower dormancy than those with haplotype 1, in agreement with association results. When haplotype 5 was crossed to haplotype 19, F_2 individuals carrying haplotype 19 had a significantly lower dormancy.

Allelic effects conferred by the different haplotypes were mostly around one week, with up to 2 weeks in the F_2 population in which haplotypes 15 and 5 segregated (Table 5). Dominance coefficients were very close to zero, indicating that *DOG1* alleles behaved almost additively. In general, observed allelic effects were not as large as one could have expected from the phenotypic differences measured for the parents in the common garden experiment. But dormancy levels of the parent lines measured in the F_2 experiment were also lower than in the common garden experiment.

LOCAL ADAPTATION FOR SEED DORMANCY

To test if the observed differences in seed dormancy are adaptive, Q_{ST} for seed dormancy was compared to F_{ST} values from neutral markers. Although some of the observed Q_{ST} values were high, they were never outside the distribution of neutral markers and the confidence intervals around these estimates were large (Table 6). Q_{ST} for dormancy was always higher than 0.7 except in Spain, where Q_{ST} was only 0.38 (Table 6).

Table 5. **Cosegregation of seed dormancy and *DOG1* in F_2 populations. D50 difference is the difference in the mean homozygote values for the different haplotypes. The significance of this difference was tested with a post-hoc test (Tukey HSD), corrected for multiple testing. Haplotype on the right in the third column is always the more dormant haplotype.**

Cross	Regions	Haplotypes	N	D50 difference	P-adjusted	R^2	Allelic effect, a	Dominance coefficient, k
All2-1×Fet-6	France	1 and 5	133	−1.40	5.69×10^{-13}	0.35	0.70	−0.30
Cam-4×Fet-6	France	15 and 5	126	−4.37	7.55×10^{-15}	0.54	2.19	−0.15
Cam-4×All2-1	France	15 and 1	145	−1.56	3.49×10^{-5}	0.12	0.78	0.02
Kon-2-2×Fet-6	Norway×France	19 and 5	121	−2.27	4.66×10^{-15}	0.52	1.13	−0.12
Kon-2-2×Nfro-1-4	Norway	18 and 19	122	−0.06	0.119	-	-	-

Table 6. **Q_{ST} values for seed dormancy in different regions. 2.5% and 97.5% denote the limits of the 95% highest posterior density interval for Q_{ST}. 95% F_{ST} indicates the value for the 95% quantile of neutral marker F_{ST}.**

Region	Q_{ST} D50	2.5%	97.5%	95% F_{ST}
All	0.7523	0.6478	0.8421	0.7973
Europe	0.7053	0.5746	0.8184	0.7674
Spain	0.3815	0.1084	0.7301	0.6471
France	0.7237	0.5246	0.8785	0.7857
Norway	0.9237	0.8025	0.9911	1.0000
Central Asia	0.7912	0.5494	0.9523	1.0000

Variation for seed dormancy was also compared to environmental variation. Summer precipitation (precipitation in the warmest quarter of the year) partly explains variation in seed dormancy (Fig. 6), with populations that received more precipitation in the summer being less dormant. In a linear model with dormancy (D25) as a response, summer precipitation was significant

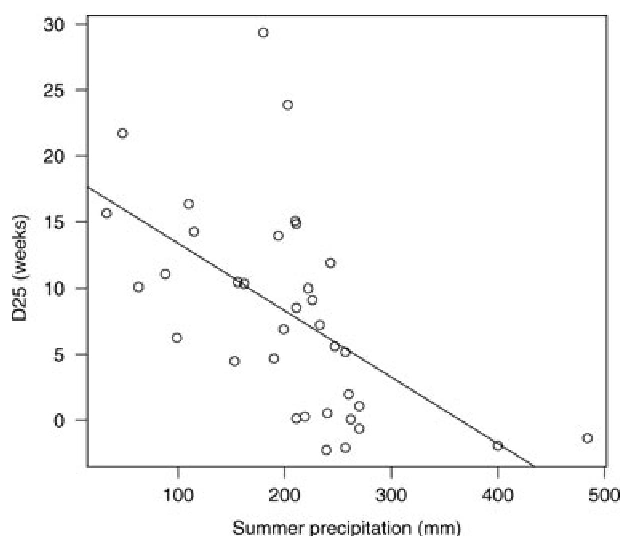

Figure 6. **Relationship between seed dormancy and summer precipitation. Data are population means.**

($F_{1,33}$ = 16.16 and $P = 0.0003$; $R^2 = 0.31$). There were some outlier populations that were quite dormant but received a considerable amount of precipitation (Mog and Sk-1), or were nondormant but received considerably more precipitation than the other populations (Veg-1 and Veg-2). These outliers did not drive the relationship, as excluding them increased the R^2 to 0.41. Setting the small negative values for some nondormant populations to zero had almost no effect. Because the climate of the Central Asian populations is quite different from Western Europe, it makes sense to compare only the European populations, which form a cline, therefore only the European populations were used. However, the relationship remained significant when the Central Asian populations were included ($P = 0.0011$, $R^2 = 0.22$). The effect of summer precipitation was the strongest for D25, but remains significant for D50 ($P = 0.005$, $R^2 = 0.16$). Furthermore, summer precipitation had an effect even when it was included in a model with geographic region as a factor ($P = 0.044$, $R^2 = 0.29$). We also investigated if including population structure, as means of principle component analysis (PCA) components that were used in the association study for each population, had any effect on the model. We included the first two components. In a model with D25, summer precipitation and the two PCA components, only summer precipitation had significant effect on dormancy ($P = 0.007$, $R^2 = 0.27$). The two PCA components were not significant ($P = 0.968$ and $P = 0.741$ for components 1 and 2, respectively). When both latitude and summer precipitation were included in the model, only summer precipitation had a significant effect ($P = 0.034$, R^2 of the full model was 0.30) and the effect of latitude was not significant in this model ($P = 0.502$). In such a dataset many environmental variables are correlated with latitude. However, the effect of summer precipitation seems to be the main factor because it is the only one that remains significant when latitude or temperature are analyzed jointly with summer precipitation (see Supporting information for details). Additionally, the correlation of seed dormancy to summer precipitation is stronger than to any of the nine other climatic parameters we tested (Table S6). Five other phenotypic traits were scored in the same common garden

Table 7. **Correlations between genetic differentiation and geography. Pairwise F_{ST} between populations, for SNPs or *DOG1*, correlated either to absolute differences in summer precipitation. Significance of correlations was tested with the Mantel test, 1000 permutations.**

Region	Precipitation versus SNP F_{ST}	Precipitation *versus DOG1* 8_{ST}
European regions	r=−0.0626	r=0.1229
	p=0.666	p=0.055
Spain and France	r=−0.0785	r=0.1385
	p=0.704	p=0.053
France and Norway	r=0.1299	r=0.1715
	p=0.256	p=0.002

experiment (flowering time, number of basal and lateral branches, plant height at maturity, seed weight) and seed dormancy was the only phenotypic trait correlated with summer precipitation (Table S7).

We also tested for selection on *DOG1* by comparing genetic distances to environmental differences. If genetic divergence between populations increases as an environmental variable changes, a stronger divergence for functional variation than for neutral variation might suggest that selection is operating. Neutral divergence did not correlate with summer precipitation differences between populations, but *DOG1* divergence increased slightly with increased differences in summer precipitation. This relationship is only suggestive for all European regions together or when the Spanish and the French populations were used. However, when the Norwegian and the French populations are compared the correlation is weak but significant (Table 7). This result further suggests that *DOG1* variation in these populations is not neutral.

Discussion

We report here a series of population genetics and functional genetic analyses that collectively bring a strong indication that *DOG1* is subject to local selection in *A. thaliana*, thereby emphasizing the importance of studying local adaptation with an array of approaches. Below, we first show that our results are robust to the major caveats associated with F_{ST}-based approaches and then discuss the possible reasons for the discordant result of the Q_{ST}/F_{ST} analysis. We further review all other lines of evidence that support local adaptation for *DOG1*. Finally, we discuss our results in the light of population genetics models and highlight their implications for our understanding of local adaptation in general.

Our study provides several lines of evidence that consistently support that natural selection has shaped variation at *DOG1*. First, 8_{ST} for *DOG1* was higher than expected from neutral markers (Fig. 3). Estimates of neutral F_{ST} could be in some instances underestimated as a consequence of ascertainment bias in the choice of SNP markers (Clark et al. 2005). However, we believe it is unlikely that SNP ascertainment has a large effect on F_{ST} estimates in our study. The selected SNPs indeed tended to have a high frequency throughout the species range, but we observed in a previous study that F_{ST} estimates were not biased (Kronholm et al. 2010). This is presumably because SNP markers used here were selected from a sample that included genotypes from many different locations. Moreover, the microsatellite markers do not suffer from such a bias, because microsatellites have a high mutation rate. If a microsatellite locus is polymorphic in a panel of genotypes it is likely to be polymorphic in a another set of genotypes. The mean 8_{ST} of the microsatellite markers is nearly equal to the mean F_{ST} of the SNP markers. For the European populations, microsatellite $8_{ST} = 0.660$ and SNP $F_{ST} = 0.621$, this again, suggests that the SNP markers are unlikely to be greatly biased by ascertainment. Importantly, we also investigated whether *DOG1* is under local selection by examining 8_{ST} along the chromosome at the position of *DOG1*. SNP markers along chromosome 5 were discovered from a panel of accessions from all regions used in this study. There was a clear peak in both 8_{ST} and between-population heterozygosity (H_T - H_S) at the position of *DOG1* (Fig. 4). This provides a strong indication that the high 8_{ST} of *DOG1* is likely to have been caused by selection and not by a lower recombination rate in this part of the genome (Charlesworth et al. 1997).

If *DOG1* is under spatially heterogeneous selection, variation at *DOG1* should cause phenotypic variation; else natural selection could not act. *DOG1* is a known QTL, but alleles present in previous QTL mapping populations (Bentsink et al. 2010) are not necessarily representative of natural variation segregating throughout the species range. We therefore conducted an analysis of genetic association between *DOG1* and dormancy. Several *DOG1* alleles were associated with dormancy (Table 4) and these associations were confirmed by analyses of cosegregation between *DOG1* and dormancy in F_2 populations (Table 5). These results are therefore in agreement with the idea that the high F_{ST} observed at *DOG1* was caused by natural selection on dormancy. Importantly, the F_2 populations show at least four functional classes of alleles segregate in the population. Placing the functional differences on the haplotype network of *DOG1* suggest that mutations modifying dormancy have originated several times independently from the haplotype 5, as suggested for haplotypes 1, 15 and the branch leading to haplotypes 18 and 19. In addition, haplotype 4 appears to increase dormancy and haplotype 21, decreases dormancy relative to haplotype 4 from which it is derived. Haplotypes 2 and 22, which may result from a recombination event, were associated with opposite effects on the phenotype, suggesting that recombination can also participate to the generation of novel functional

alleles. However, because all other mutations are in complete linkage disequilibrium, the series of alleles found to associate with dormancy is unlikely to be explained by recombination alone. Given that *DOG1* is a small gene and recombination was found also to be rare along the full *DOG1* sequence (M. Debieu, unpubl. ms.), it appears that functionally different alleles in *DOG*1 have evolved independently, to either increase or decrease dormancy.

Classical comparative analysis of Q_{ST}/F_{ST} estimates of population differentiation could not reject the hypothesis that seed dormancy variation departs from neutral evolution, although this approach has proved successful in a number of other studies (Merilä and Crnokrak 2001; Leinonen et al. 2008). Because Q_{ST} has both high sampling variance and high evolutionary variance (O'Hara and Merilä 2005; Goudet and Büchi 2006; Miller et al. 2008; Whitlock 2008), our result may simply reflect the limited power of this approach (Whitlock 1999; Goudet and Büchi 2006; Goudet and Martin 2007; Miller et al. 2008). Both experimental and theoretical studies have shown that finding evidence for local adaptation is very difficult when neutral F_{ST} is very high (Le Corre and Kremer 2003; Porcher et al. 2006). *Arabidopsis* is highly structured (Nordborg et al. 2005; Pico et al. 2008; Platt et al. 2010) and this is the case for our populations as well (Kronholm et al. 2010). Other studies in *A. thaliana* have also failed to find $Q_{ST} > F_{ST}$ (Kuittinen et al. 1997; Stenoien et al. 2005, but see Banta et al. 2007).

Demographic events can increase the variance of summary statistics such as F_{ST} across the genome, so the possibility that the pattern we observed in *DOG1* is due to chance alone cannot be completely discarded. However, the adaptive relevance of *DOG1* is also supported by independent findings. In a field study conducted at two locations in North America, QTLs for germination timing and fitness colocalized with *DOG1* (Donohue et al. 2005b; Huang et al. 2010). The genotypes used were not local to the field sites, preventing inference of local adaptation, but show that variation in *DOG1* can associate with substantial fitness effects. Here, the analysis of covariation between seed dormancy and the environment brings a novel indication that seed dormancy and *DOG1* are subject to local selective forces. We observed a negative correlation between seed dormancy and the amount of precipitation received in the summer months (Fig. 6). Variation in *DOG1* showed a similar trend in Norway and France (Table 7). Importantly, neutral markers were not correlated with summer precipitation, supporting the hypothesis that differences in dormancy between populations do not reflect only the action of genetic drift (Table 7). This finding also reveals the putative ecological forces acting on *DOG1* evolution. It fits ecological predictions for dormancy: plants can avoid summer drought by not germinating in the spring (Baskin and Baskin 1972; Evans and Ratcliffe 1972; Baskin and Baskin 1983).

The relationship between summer precipitation and dormancy was stronger for D25 than for D50, a result suggesting that summer precipitation is important in determining the time when seeds can begin germination. In *A. thaliana*, the environment is known to influence seed dormancy induction, and can act to prevent early spring or summer germination and favor germination in the fall (Montesinos et al. 2009). In *Digitaria milanjiana*, the amount of total precipitation was related to seed dormancy, although a limited number of populations were studied (Hacker 1984; Hacker et al. 1984). In contrast, germination of chilled seeds of *Artemisia tridentata* correlated with mean January temperature (Meyer and Monsen 1991). Furthermore, a relationship between germination patterns and the environment was found for *Linum perenne* (Meyer and Kitchen 1994) and in several species of *Penstemon* (Meyer et al. 1995). However, correlations between dormancy and environmental factors have not always been found (Schütz and Milberg 1997; Petrů and Tielbörger 2008).

Our results may also have bearings on our understanding of the process of local adaptation in general. By using simulations, Yeaman and Guillaume (2009) showed that a genetic model with multiple alleles per locus, where allelic effects can freely evolve, permitted local adaptation in the presence of stronger gene flow than a model with biallelic loci or with a Gaussian approximation of the phenotype. As QTL effects can be larger this also permits larger selection coefficients for individual loci. Consequently, larger differences can be maintained in the presence of gene flow (Yeaman and Guillaume 2009). By showing that *DOG1* evolution fits better to a model with multiple alleles per locus, our results also find a broader significance, beyond the mere analysis of dormancy evolution. The situation we observe for *DOG1* may be relatively common. A similar pattern has been found in the multiple independent loss-of-function mutations segregating for the gene *FRIGIDA* (Johanson et al. 2000; Le Corre et al. 2002; Le Corre 2005; Toomajian et al. 2006). When migration between populations is low relative to mutation rate, that is $2N_e\mu > 2N_em$, adaptation is predicted to result from the fixation of independent beneficial mutations in different parts of the species range (Pennings and Hermisson 2006). This happens because gene flow is too small relative to mutation rate to allow for the same allele to spread to all populations where it would be beneficial. Therefore, many models in quantitative genetics, for example, Spichtig and Kawecki (2004), do not recapitulate adequately the whole process of local adaptation. Loss-of-function alleles, as in the case of *FRIGIDA*, are likely to arise readily by mutations. At *DOG1* we have not observed any loss-of-function alleles. Yet, mutations seem to frequently generate functional variation at this gene. Studies on natural variation in *A. thaliana* do indeed hold great promise for elucidating the genetic basis of adaptation (Koornneef et al. 2004). Further studies of developmental pathways

controlling adaptive traits will help explain why some genes are involved in adaptive evolution and not others.

ACKNOWLEDGMENTS

We would like to thank U. Tartler and A. O. Fandiño for technical assistance. S. Antoniazza kindly provided an R-script to calculate geographic distances between sets of coordinates. B. O'Hara gave advice on Bayesian models and Q_{ST}. IK and JdM were funded by the Max Planck Society and SFB-680. CAB and FXP were funded by Ministry of Science and Innovation of Spain, project references: BIO2010-15022 and CGL2009-07847/BOS, respectively. We thank three anonymous referees who gave constructive comments that improved the quality of the manuscript.

LITERATURE CITED

Alonso-Blanco, C., L. Bentsink, C. J. Hanhart, H. Blankestijn-de Vries, and M. Koornneef. 2003. Analysis of natural allelic variation at seed dormancy loci of *Arabidopsis thaliana*. Genetics 164:711–729.

Aydin, A., M. R. Toliat, S. Bahring, C. Becker, and P. Nurnberg. 2006. New universal primers facilitate pyrosequencing. Electrophoresis 27:394–397.

Banta, J. A., J. Dole, M. B. Cruzan, and M. Pigliucci. 2007. Evidence of local adaptation to coarse-grained environmental variation in *Arabidopsis thaliana*. Evolution 61:2419–2432.

Baskin, J. M., and C. C. Baskin. 1972. Ecological life cycle and physiological ecology of seed germination of *Arabidopsis thaliana*. Can. J. Bot. 50:353–360.

———. 1983. Seasonal changes in the germination responses of buried seeds of *Arabidopsis thaliana* and ecological interpretation. Bot. Gaz. 144:540–543.

Beaumont, M. A. 2005. Adaptation and speciation: what can F_{ST} tell us. Trends Ecol. Evol. 20:435–440.

Bentsink, L., J. Jowett, C. J. Hanhart, and M. Koornneef. 2006. Cloning of *DOG1*, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 103:17042–17047.

Bentsink, L., J. Hanson, C. J. Hanhart, H. Blankestijn-de Vries, C. Coltrane, P. Keizer, M. El-Lithy, C. Alonso-Blanco, M. T. de Andrés, M. Reymond, *et al*. 2010. Natural variation for seed dormancy in *Arabidopsis* is regulated by additive genetic and molecular pathways. Proc. Natl. Acad. Sci. USA 107:4264–4269.

Biere, A. 1991. Parental effects in *Lychnis flos-cuculi*. II: selection on time of emergence and seedling performance in the field. J. Evol. Biol. 4:467–486.

Bonnin, I., J. M. Prosperi, and I. Olivieri. 1996. Genetic markers and quantitative genetic variation in *Medicago truncatula* (*Leguminosae*): a comparative analysis of population structure. Genetics 143:1795–805.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635.

Cadman, C. S., P. E. Toorop, H. W. Hilhorst, and W. E. Finch-Savage. 2006. Gene expression profiles of *Arabidopsis* Cvi seeds during dormancy cycling indicate a common underlying dormancy control mechanism. Plant J. 46:805–822.

Charlesworth, B., M. Nordborg, and D. Charlesworth. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. 70:155–174.

Clark, A. G., M. J. Hubisz, C. Bustamante, S. Williamson, and R. Nielsen. 2005. Ascertainment bias in studies of human genome-wide polymorphism. Genome Res. 15:1496–1502.

Clausen, J., D. D. Keck, and W. M. Hiesey. 1941. Regional differentiation in plant species. Am. Nat. 75:231–250.

Clement, M., D. Posada, and K. A. Crandall. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9:1657–1659.

Crawley, M. J. 2005. Statistics: an introduction using R. Wiley, New York.

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

Donohue, K., L. Dorn, C. Griffith, E. Kim, A. Aguilera, C. R. Polisetty, and J. Schmitt. 2005a. Environmental and genetic influences on the germination of *Arabidopsis thaliana* in the field. Evolution 59:740–757.

———. 2005b. The evolutionary ecology of seed germination of *Arabidopsis thaliana*: variable natural selection on germination timing. Evolution 59:758–770.

Evans, A. S., and R. J. Cabin. 1995. Can dormancy affect the evolution of post-germination traits? The case of *Lesquerella fendleri*. Ecology 76:344–356.

Evans, J., and D. Ratcliffe. 1972. Variation in 'after-ripening' and its ecological significance. Arabidopsis Information Service 9:3–5.

Excoffier, L. 2007. Analysis of population subdivision. *In* D. J. Balding, M. Bishop, and C. Cannings, eds. Handbook of statistical genetics. Wiley, Hoboken, NJ.

Fakhrai-Rad, H., N. Pourmand, and M. Ronaghi. 2002. Pyrosequencing^{TM}: an accurate detection platform for single nucleotide polymorphisms. Hum. Mutat. 19:479–485.

Finch-Savage, W. E., and G. Leubner-Metzger. 2006. Seed dormancy and the control of germination. New Phytol. 171:501–523.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available at http://www2.unil.ch/popgen/softwares/fstat.htm. Accessed February 17, 2012.

Goudet, J., and L. Büchi. 2006. The effects of dominance, regular inbreeding and sampling design on Q_{ST}, an estimator of population differentiation for quantitative traits. Genetics 172:1337–1347.

Goudet, J., and G. Martin. 2007. Under neutrality, $Q_{ST} \leq F_{ST}$ when there is dominance in an island model. Genetics 176:1371–1374.

Griffith, C., E. Kim, and K. Donohue. 2004. Life-history variation and adaptation in the historically mobile plant *Arabidopsis thaliana* (*Brassicaceae*) in North America. Am. J. Bot. 91:837–849.

Griswold, C. K. 2006. Gene flow's effect on the genetic architecture of a local adaptation and its consequences for QTL analyses. Heredity 96:445–453.

Gross, K. L., and A. D. Smith. 1991. Seed mass and emergence time effects on performance of *Panicum dichotomiflorum* Michx across environments. Oecologia 87:270–278.

Hacker, J. B. 1984. Genetic variation in seed dormancy in *Digitaria milanjiana* in relation to rainfall at the collection site. J. Appl. Ecol. 21:947–959.

Hacker, J. B., M. H. Andrew, J. G. McIvor, and J. J. Mott. 1984. Evaluation in contrasting climates of dormancy characteristics of seed of *Digitaria milanjiana*. J. Appl. Ecol. 21:961–969.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. 41:95–98.

Hijmans, R. J., L. Guarino, M. Cruz, and E. Rojas. 2001. Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. Plant Genet. Res. Newsl. 127:15–19.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Clim. 25:1965–1978.

Holm, S. 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6:65–70.

Holsinger, K. E., and B. S. Weir. 2009. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. Nat. Rev. Genet. 10:639–650.

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

Johanson, U., J. West, C. Lister, S. Michaels, R. Amasino, and C. Dean. 2000. Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. Science 290: 344–347.

Kalisz, S. 1986. Variable selection on the timing of germination in *Collinsia verna* (*Scrophulariaceae*). Evolution 40:479–491.

Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. Ecol. Lett. 7:1225–1241.

Koornneef, M., C. Alonso-Blanco, and D. Vreugdenhil. 2004. Naturally occurring genetic variation in *Arabidopsis thaliana*. Annu. Rev. Plant Biol. 55:141–172.

Kronholm, I., O. Loudet, and J. de Meaux. 2010. Influence of mutation rate on estimators of genetic differentiation—lessons from *Arabidopsis thaliana*. BMC Genet. 11:33.

Kuittinen, H., A. Mattila, and O. Savolainen. 1997. Genetic variation at marker loci and in quantitative traits in natural populations of *Arabidopsis thaliana*. Heredity 79(Pt. 2):144–152.

Lande, R. 1992. Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. Evolution 46:381–389.

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

Le Corre, V., and A. Kremer. 2003. Genetic variability at neutral markers, quantitative trait land trait in a subdivided population under selection. Genetics 164:1205–1219.

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

Leimu, R., and M. Fischer. 2008. A meta-analysis of local adaptation in plants. PLoS One 3:e4010.

Leinonen, T., R. B. O'Hara, J. M. Cano, and J. Merila. 2008. Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. J. Evol. Biol. 21:1–17.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter. 2000. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10:325–337.

Lynch, M., and B. Walsh. 1998. Genetics and the analysis of quantitative traits. Sinauer Associates, Inc., Sunderland.

Marks, M., and S. Prince. 1981. Influence of germination date on survival and fecundity in wild lettuce *Lactuca serriola*. Oikos 36:326–330.

Merilä, J., and P. Crnokrak. 2001. Comparison of genetic differentiation at marker loci and quantitative traits. J. Evol. Biol. 14:892–903.

Meyer, S. E., and S. G. Kitchen. 1994. life history variation in blue flax (*Linum perenne*: Linaceae): seed germination phenology. Am. J. Bot. 81:528–535.

Meyer, S. E., and S. B. Monsen. 1991. Habitat-correlated variation in mountain big sagebrush (*Artemisia tridentata* ssp. vaseyana) seed germination patterns. Ecology 72:739–742.

Meyer, S. E., S. G. Kitchen, and S. L. Carlson. 1995. Seed germination timing patterns in intermountain Penstemon (Scrophulariaceae). Am. J. Bot. 82:377–389.

Michalakis, Y., and L. Excoffier. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 142:1061–1064.

Miller, J. R., B. P. Wood, and M. B. Hamilton. 2008. F_{ST} and Q_{ST} under neutrality. Genetics 180:1023–1037.

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

Montesinos, A., S. J. Tonsor, C. Alonso-Blanco, and F. X. Pico. 2009. Demographic and genetic patterns of variation among populations of *Arabidopsis thaliana* from contrasting native environments. PLoS One 4:e7213.

Munir, J., L. A. Dorn, K. Donohue, and J. Schmitt. 2001. The effect of maternal photoperiod on seasonal dormancy in *Arabidopsis thaliana* (Brassicaceae). Am. J. Bot. 88:1240–1249.

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

O'Hara, R. B., and J. Merilä. 2005. Bias and precision in Q_{ST} estimates: problems and some solutions. Genetics 171:1331–1339.

Oksanen, J., R. Kindt, P. Legendre, R. B. O'Hara, M. Henry, and M. H. H. Stevens. 2007. Vegan: community ecology package. R package version 1.8-7. Available at http://cran.r-project.org/web/packages. Accessed February 17, 2012.

Orr, H. A. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution 52:935–949.

———. 2005. The genetic theory of adaptation: a brief history. Nat. Rev. Genet. 6:119–127.

Pennings, P. S., and J. Hermisson. 2006. Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. Mol. Biol. Evol. 23:1076–1084.

Petrů, M., and K. Tielbörger. 2008. Germination behaviour of annual plants under changing climatic conditions: separating local and regional environmental effects. Oecologia 155:717–728.

Pico, F. X., B. Mendez-Vigo, J. M. Martinez-Zapater, and C. Alonso-Blanco. 2008. Natural genetic variation of *Arabidopsis thaliana* is geographically structured in the Iberian Peninsula. Genetics 180:1009–1021.

Platt, A., M. Horton, Y. S. Huang, Y. Li, A. E. Anastasio, N. W. Mulyati, J. Agren, O. Bossdorf, D. Byers, K. Donohue, et al. 2010. The scale of population structure in *Arabidopsis thaliana*. PLoS Genet. 6: e1000843.

Porcher, E., T. Giraud, and C. Lavigne. 2006. Genetic differentiation of neutral markers and quantitative traits in predominantly selfing metapopulations: confronting theory and experiments with *Arabidopsis thaliana*. Genet. Res. 87:1–12.

R Development Core Team. 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org.

Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497.

Schütz, W., and P. Milberg. 1997. Seed dormancy in *Carex canescens*: Regional differences and ecological consequences. Oikos 78:420–428.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462.

Spichtig, M., and T. Kawecki. 2004. The maintenance (or not) of polygenic variation by soft selection in heterogeneous environments. Am. Nat. 164:70–84.

Spitze, K. 1993. Population structure in *Daphnia obtusa*: Quantitative genetic and allozymic variation. Genetics 135:367–374.

Stenoien, H. K., C. B. Fenster, A. Tonteri, and O. Savolainen. 2005. Genetic variability in natural populations of *Arabidopsis thaliana* in northern Europe. Mol. Ecol. 14:137–148.

Stich, B., J. Mohring, H.-P. Piepho, M. Heckenberger, E. S. Buckler, and A. E. Melchinger. 2008. Comparison of mixed-model approaches for association mapping. Genetics 178:1745–1754.

Toomajian, C., T. T. Hu, M. J. Aranzana, C. Lister, C. Tang, H. Zheng, K. Zhao, P. Calabrese, C. Dean, and M. Nordborg. 2006. A nonparametric test reveals selection for rapid flowering in the *Arabidopsis* genome. PLoS Biol. 4:e137.

Turesson, G. 1922. The genotypical response of the plant species to the habitat. Hereditas 3:211–350.

———. 1925. The plant species in relation to habitat and climate. Hereditas 6:147–236.

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S. Springer, New York.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of populations structure. Evolution 38:1358–1370

Whitlock, M. C. 1999. Neutral additive genetic variance in a metapopulation. Genet. Res. 74:215–221.

———. 2008. Evolutionary inference from Q_{ST}. Mol. Ecol. 17:1885–1896.

———. 2011. G'_{ST} and D do not replace F_{ST}. Mol. Ecol. 20:1083–1091.

Wilczek, A. M., J. L. Roe, M. C. Knapp, M. D. Cooper, C. Lopez-Gallego, L. J. Martin, C. D. Muir, S. Sim, A. Walker, J. Anderson, et al. 2009. Effects of genetic perturbation on seasonal life history plasticity. Science 323:930–934.

Yeaman, S., and F. Guillaume. 2009. Predicting adaptation under migration load: the role of genetic skew. Evolution 63:2926–2938.

Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki, J. F. Doebley, M. D. McMullen, B. S. Gaut, D. M. Nielsen, J. B. Holland, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203–208.

Associate Editor: J. Kelly

Supporting Information

The following supporting information is available for this article:

Figure S1. Distribution of harvesting dates for the common garden experiment.
Figure S2. An example of population structure correction.
Figure S3. Amino acid alignment of exon 1 of *DOG1*.
Table S1. Primers used to amplify and sequence the fragments for SNP discovery.
Table S2. Primer sequences used in the pyrosequencing assays.
Table S3. Minimum MSD values and their corresponding T values for different traits and samples.
Table S4. Polymorphic positions in the observed *DOG1* haplotypes.
Table S5. Population means and genetic variation within populations for seed dormancy.
Table S6. Relationship between several environmental variables and seed dormancy (D25).
Table S7. Relationship between other phenotypic traits and summer precipitation.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.