

RESEARCH ARTICLE

Open Access

Ecological, genetic and evolutionary drivers of regional genetic differentiation in *Arabidopsis thaliana*



Antonio R. Castilla^{1,2}, Belén Méndez-Vigo³, Arnald Marcer^{4,5}, Joaquín Martínez-Minaya⁶, David Conesa⁷, F. Xavier Picó^{2*} and Carlos Alonso-Blanco³

Abstract

Background: Disentangling the drivers of genetic differentiation is one of the cornerstones in evolution. This is because genetic diversity, and the way in which it is partitioned within and among populations across space, is an important asset for the ability of populations to adapt and persist in changing environments. We tested three major hypotheses accounting for genetic differentiation— isolation-by-distance (IBD), isolation-by-environment (IBE) and isolation-by-resistance (IBR)—in the annual plant *Arabidopsis thaliana* across the Iberian Peninsula, the region with the largest genomic diversity. To that end, we sampled, genotyped with genome-wide SNPs, and analyzed 1772 individuals from 278 populations distributed across the Iberian Peninsula.

Results: IBD, and to a lesser extent IBE, were the most important drivers of genetic differentiation in *A. thaliana*. In other words, dispersal limitation, genetic drift, and to a lesser extent local adaptation to environmental gradients, accounted for the within- and among-population distribution of genetic diversity. Analyses applied to the four Iberian genetic clusters, which represent the joint outcome of the long demographic and adaptive history of the species in the region, showed similar results except for one cluster, in which IBR (a function of landscape heterogeneity) was the most important driver of genetic differentiation. Using spatial hierarchical Bayesian models, we found that precipitation seasonality and topsoil pH chiefly accounted for the geographic distribution of genetic diversity in Iberian *A. thaliana*.

Conclusions: Overall, the interplay between the influence of precipitation seasonality on genetic diversity and the effect of restricted dispersal and genetic drift on genetic differentiation emerges as the major forces underlying the evolutionary trajectory of Iberian *A. thaliana*.

Keywords: Genetic diversity, Genetic structure, Iberian Peninsula, Nested maximum-likelihood population effect models, Precipitation seasonality, Spatial hierarchical Bayesian models

Background

Genetic diversity is an important asset for the ability of populations to adapt and persist in changing environments [1–8]. The spatio-temporal changes in genetic diversity constantly taking place in any population—regardless of the

causes, pace and the phenotypic effects of such changes—constitute the raw material upon which natural selection eventually acts [9–11]. At any spatial scale, genetic diversity typically becomes unevenly distributed across space [12], as genetic diversity is determined by how genetic diversity is partitioned within and among populations across the distribution. In other words, the spatial distribution of genetic diversity depends on the extent of genetic differentiation among populations whatever the sources of such

* Correspondence: xpico@ebd.csic.es

²Departamento de Ecología Integrativa, Estación Biológica de Doñana (EBD), Consejo Superior de Investigaciones Científicas (CSIC), Sevilla, Spain
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

differentiation. Nonetheless, it must be noted that genetic differentiation is a spatially-explicit phenomenon, as genetic differentiation strictly depends on the genetic relationship that a given population has with its neighbors near and far.

The inherent spatial nature of genetic differentiation defines the theoretical and methodological framework of three models, which are not mutually exclusive, testing the major drivers of genetic differentiation: isolation-by-distance (IBD hereafter), isolation-by-environment (IBE hereafter) and isolation-by-resistance (IBR hereafter) models. In the classical IBD, genetic differentiation among populations exhibits a positive relationship with geographic distance [13–18]. In this case, dispersal limitation and genetic drift determine the greater genetic differentiation at larger distances. In fact, limited dispersal constrains gene flow among populations, which is not able to counteract the effect of genetic drift within populations [16–18]. Many types of organisms exhibit IBD [19], probably because unrestricted gene flow hardly occurs in nature. Nevertheless, quantifying the contribution of limited dispersal and genetic drift to genetic differentiation is not a straightforward task, as we largely ignore the actual extent of dispersal, the effective population sizes conditioning genetic drift, and the effects of historical factors shaping the genetic relationships among populations [20].

In contrast, the IBE model deals with the effects of environmental differences on genetic differentiation. IBE posits that gene exchange is strongest among populations located in similar environments, which would be mediated by environmental heterogeneity, the extent of local adaptation and spatial variation in gene flow across space [18, 21, 22]. Thus, genetic differentiation among populations increases with their environmental differentiation, independently of their geographic distance. IBE can arise due to multiple factors, such as biased dispersal due to preferences for particular environments, natural selection against maladapted immigrants, sexual selection against immigrants when they exhibit divergence in mating choices or sexual signals, and natural selection against hybrids when they show reduced fitness relative to non-hybrids [22].

Finally, the IBR model takes environmental heterogeneity across landscapes into account as a modulator of gene flow and its effects on genetic differentiation [23]. The IBR model predicts a positive relationship between genetic differentiation and resistance distance among populations [24–26]. The resistance distance between population pairs is a concept inspired in circuit theory, which considers the landscape features reducing the probability of dispersal and gene flow among populations [24–27]. Under IBR, the spatial structure of habitat suitability is of paramount importance to determine the least cost path between population pairs optimizing their connectivity and thus minimizing their resistance distance (24 and references

therein). Given that the resistance distance between population pairs is a function of geographic distance, and that presence-background models estimate habitat suitability using environmental predictors, IBR inevitably conflates IBD and IBE [22].

Here, we tested these three hypotheses to identify the major drivers of genetic differentiation in the annual plant *Arabidopsis thaliana* across the Iberian Peninsula. This is the region of the species' distribution harboring the largest genomic diversity [28–30]. In addition, Iberian *A. thaliana* occurs in a wide array of natural environments practically across the whole region, spanning from seaside to sub-alpine locations [31–33]. These two features are likely the result of *A. thaliana*'s history in the Iberian Peninsula [29, 31, 34], where the species long survived by developing adaptations to ample environmental heterogeneity over dramatic climatic oscillations. The occurrence of relict populations with an African origin [29, 34] also supports such history of Iberian *A. thaliana*. In fact, habitat suitability of relict populations has been associated to more stable vegetation dynamics since the Last Glacial Maximum and during the Holocene in the Iberian Peninsula [35]. Overall, the geographic ubiquity, the large amount of genetic diversity, the broad variety of habitats occupied, and the long evolutionary history make Iberian *A. thaliana* an appropriate study system to disentangle the drivers of genetic differentiation at a regional scale.

Given the fact that a small sample size seriously reduces power and accuracy of spatial analyses [36], IBD, IBE and IBR were tested using 278 Iberian *A. thaliana* populations collected over a decade. About six individuals per population, totaling 1772 individuals, were genotyped with genome-wide, putatively neutral SNPs to estimate genetic diversity, differentiation and structure. We hypothesized that IBD and IBE largely accounted for genetic differentiation for two reasons. First, IBD is a common result in *A. thaliana* genetic studies—including the Iberian Peninsula [31–33]—due to dispersal limitation and high self-fertilization rates. Second, Iberian *A. thaliana* shows significant adaptive variation in fitness-related traits across environmental gradients [30, 32, 37, 38]. However, we ignore the importance of IBR and the joint contribution of IBD, IBE and IBR to the genetic differentiation of Iberian *A. thaliana* populations. For the sake of completeness, we also examined the geographical distribution of genetic diversity in Iberian *A. thaliana*. To this end, we employed a spatial hierarchical Bayesian model to identify regional hot and cold spots of genetic diversity and their potential environmental predictors. Overall, we stress the importance of identifying the drivers of genetic differentiation among populations, but also of pinpointing the forces that determine the amount of genetic diversity within populations, to understand the evolutionary dynamics of any organism.

Results

Genetic diversity

We genotyped 240 genome-wide SNPs in 1172 individuals collected from 278 populations (Fig. 1a) to determine the genetic diversity of *A. thaliana* in the Iberian Peninsula. In this set of populations, the percentage of polymorphic loci (PL) ranged between 0 and 56.7% (mean \pm SD = $24.4 \pm 17.7\%$), the mean number of observed alleles per locus (n_a) between 0.95 and 1.56 (mean \pm SD = 1.23 ± 0.18), and mean gene diversity (H_S) between 0 and 0.224 (mean \pm SD = 0.09 ± 0.07). H_S exhibited a bimodal distribution clearly differentiated in two groups of populations. On the one hand, 66 populations had very low H_S values (54 of 0 and 12 between 0.001 and 0.009; Fig. 1c). On the other hand, the remaining 212 populations had H_S values distributed between 0.015 and 0.224 (Fig. 1c). The 66 populations with no genetic diversity did not show geographic or environmental bias, as they occurred scattered across the region (Fig. 1a) and the altitude gradient (Fig. 1b). Average (\pm SD) genetic differentiation, given by pairwise F_{ST} values, including all populations was of 0.644 ± 0.189 ($F_{ST} = 0.538 \pm 0.150$ without the 66 populations with no genetic diversity).

Overall, we found 1613 non-redundant multilocus genotypes (N_H) in the 1772 individuals, which showed an average proportion of allelic differences between haplotype pairs of 0.30 ± 0.05 . We only found five non-redundant multilocus genotypes in different populations. In

particular, two (separated by 0.6 km) and three populations (separated by 7.6–12.8 km) included the same non-redundant multilocus genotypes.

To assess whether populations experienced major or sudden disturbances that could affect their genetic diversity and differentiation during the last decades, we retrieved temporal series of publicly available orthophotographs from all populations. The analyses of the habitats interpreted from digitalized orthophotographs indicated that land-used changes and major disturbances (wildfires, landslides or the development of large infrastructures) over the time considered did not affect landscapes. In particular, the average change per habitat type between year intervals was concentrated around zero (Fig. 1d), indicating that habitats barely changed over time. We ignore, however, whether pathogen or diseases affected *A. thaliana* populations, which could influence genetic diversity patterns. We have never observed noticeable catastrophic events caused by pests and diseases over 15-plus years of experience sampling natural *A. thaliana* populations (C. Alonso-Blanco and F. X Picó, pers. obs.).

Geographic distribution of genetic diversity

To determine the geographic and environmental distribution of H_S in *A. thaliana* populations, we used a spatial hierarchical Bayesian modeling to explain H_S as a function of environmental variables. The model

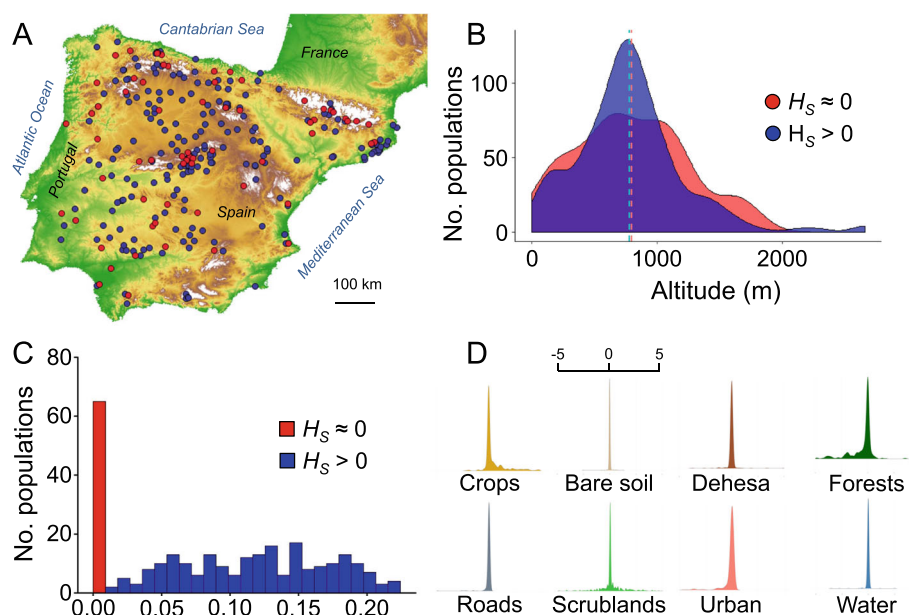


Fig. 1 Distribution of populations, genetic diversity and temporal habitat changes of Iberian *Arabidopsis thaliana* populations. **a** Geographic distribution of the 278 *A. thaliana* populations of study across the Iberian Peninsula. Red and blue dots represent populations with zero and non-zero genetic diversity (H_S) values, respectively. **b** Frequency distribution of populations with zero and non-zero H_S values as a function of altitude. Mean altitude for both groups of populations is almost identical, as indicated by dashed lines. **c** Frequency distribution of populations with zero and non-zero H_S values. **d** Frequency distribution of the average percentage change between year intervals for each habitat type. Data from digitalized orthophotographs available from each population. For the sake of clarity, only one X-axis is shown, indicating the accumulation of populations with average percentage changes around zero. The map of Fig. 1a was obtained from the National Center for Geographic Information (CNIG) of Spain

considered simultaneously the populations with zero ($N = 66$ populations) and non-zero ($N = 212$ populations) H_S values. The best fitting model indicated that the occurrence probability of populations with non-zero H_S values (populations with genetic diversity) was concentrated towards the center and eastern areas of the Iberian Peninsula (Fig. 2a). In contrast, northern, western and southern peripheral areas had higher odds for populations lacking genetic diversity (Fig. 2a). In addition, genetic diversity of *A. thaliana* was unevenly distributed across the Iberian Peninsula, with different nuclei with high diversity in large central and northern areas, as well as in a limited area in NE Spain (Fig. 2b). We did not detect any anomaly in the distribution of the mean effects of the spatial component and its uncertainty (Fig. S2) that made us suspect that the spatial effects were affecting the results.

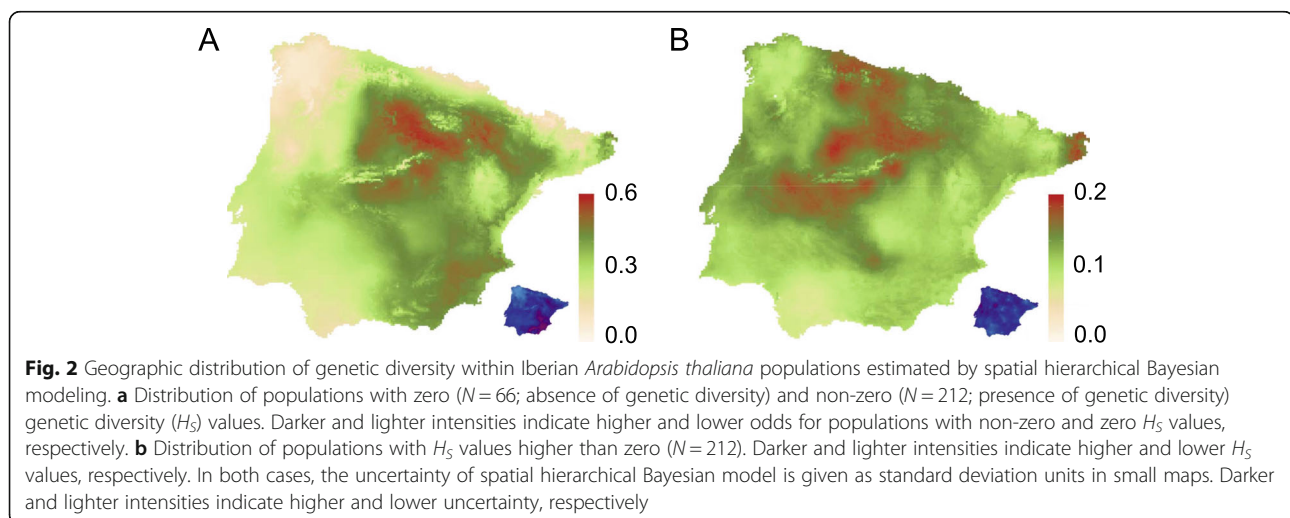
The environmental predictors with more influence on the spatial distribution of H_S were precipitation seasonality (BIO15), precipitation of the warmest quarter (BIO18) and topsoil pH. The contributions of precipitation variables were negative for the occurrence of populations with no genetic diversity, as well as for the genetic diversity of populations with non-zero H_S values (Table 1). Overall, these results indicated that areas with higher precipitation seasonality, and to a lesser extent higher precipitation in the warmest quarter, tended to harbor populations with lower genetic diversity. In the Iberian Peninsula, higher and lower precipitation seasonality characterizes xeric and mesic environments, respectively. The contribution of topsoil pH was positive for the occurrence probability of populations with zero or non-zero H_S values, meaning that populations with non-zero H_S values tended to occur in areas with basic soils. However, in the basic areas of the Iberian Peninsula (E and SE Spain) the species is very rare (Fig. 1 and

S1), and the low number of populations there with non-zero H_S values could be introducing some bias in the model. In contrast, the contribution of topsoil pH was negative for genetic diversity, indicating that populations with higher genetic diversity tended to occur in acidic soils, which is consistent with the preference of the species for this sort of soils (C. Alonso-Blanco and F.X. Picó, pers. obs.).

Genetic structure

We analyzed the structure of the 278 *A. thaliana* populations in different genetic clusters with the 1613 non-redundant multilocus genotypes using Bayesian (Fig. 3a) and ordination (Fig. 3b) methods. Both analyses detected four genetic clusters, three of them showing a strong geographic structure (Fig. 3c): the northwestern cluster 1 (NW-C1 hereafter), the northeastern cluster 2 (NE-C2 hereafter), and the southwestern cluster 4 (SW-C4 hereafter; Fig. 3b). In contrast, cluster 3, which corresponded to the group previously described as the relict *A. thaliana* lineage (relict-C3 hereafter), occurred scattered across the Iberian Peninsula (Fig. 3c). These results were fully consistent with previous studies on Iberian *A. thaliana* based on just one accession per population [31, 32, 34, 39].

Since we used about six individuals per population, we could classify populations as homogeneous or heterogeneous based on the assignment of individuals to a single or multiple genetic clusters, respectively. Most populations (230 of 278) were homogeneous, with 118, 44, 35 and 33 belonging to genetic clusters NW-C1, NE-C2, relict-C3 and SW-C4, respectively (Fig. 3c). The remaining 48 heterogeneous populations were distributed across the region, with an accumulation of them in central and NE Spain (Fig. 3c). As expected, the most abundant genetic clusters, NW-C1 and NE-C2, exhibited



a higher number of heterogeneous populations, because 46 of 48 heterogeneous populations were included in these genetic clusters.

Drivers of genetic differentiation

We applied nested maximum-likelihood population effect models (NMLPE) to evaluate simultaneously the effects of the geographic (IBD), environmental (IBE) and resistance (IBR) variation on the genetic differentiation among Iberian *A. thaliana* populations. IBD and IBE accounted for genetic differentiation at the entire Iberian Peninsula scale (Table 2A and Fig. 4a). In addition, we dissected IBE into three environmental PCs, which substantially contributed to the IBE (Table 2A and Fig. 4b). These PCs explained 29.3, 23.8, and 22.2% of the environmental variance, respectively, which accounted for different environmental gradients across the Iberian Peninsula. In particular, PC1 represented a gradient of temperature and precipitation, pinpointing the negative relationship between the two variables in the Iberian Peninsula. PC2 only depicted the gradient of temperature across the region. Finally, PC3 illustrated a gradient of precipitation and pH in which soils with lower pH receive higher precipitation in the Iberian Peninsula.

We found that the single model including IBD and IBE was ranked as the top model for three genetic clusters (NW-C1, NE-C2, and relict-C3) with no models differing less than 2 in AIC (Table 2B and Fig. 4a). Although both IBD and IBE had significant effects on genetic differentiation, size effects were greater for IBD than for IBE (Fig. 4a). In addition, the three PC axes made substantial contributions to the IBE exhibited by the populations for the same three genetic clusters (Table 2B and Fig. 4b). For SW-C4, a single model including IBR was ranked as the top model (Table 2B and

Fig. 4a). When environmental PCs were included separately in the analyses, we found two top-ranked models exhibiting $\Delta AIC < 2$ (Table 2) for SW-C4. However, model averaging discarded a relevant role of PC2 on the genetic differentiation among SW-C4 populations (Fig. 4b).

Discussion

Genetic differentiation is a dynamic process because any population is constantly under the effect of ecological, genetic and evolutionary forces modifying the amount of genetic diversity within and among populations. Disentangling such forces accounting for genetic differentiation, which underlies major evolutionary processes from local adaptation to speciation, has long promoted the development of a strong theoretical and methodological framework since practically the birth of the Modern Synthesis. Here, we addressed this question by testing three hypotheses (IBD, IBE and IBR) accounting for regional-scale genetic differentiation in Iberian *A. thaliana*.

The main result of our study, based on nested maximum-likelihood population effect models (NMLPE), indicated that genetic differentiation in *A. thaliana* was mostly accounted for by IBD, and to a lesser extent by IBE (Fig. 4a). In other words, dispersal limitation, genetic drift, and to a lesser extent local adaptation, determined the distribution of *A. thaliana*'s genetic diversity within and among populations across its Iberian range. This result extends previous studies showing a marked IBD in *A. thaliana* in the Iberian Peninsula [30–33] and elsewhere [28, 34, 40–48]. It is widely accepted that IBD in *A. thaliana* results from the joint effects of dispersal limitation, self-fertilization, local adaptation and demographic history. Our study contributed to disentangle the effect of some of these factors,

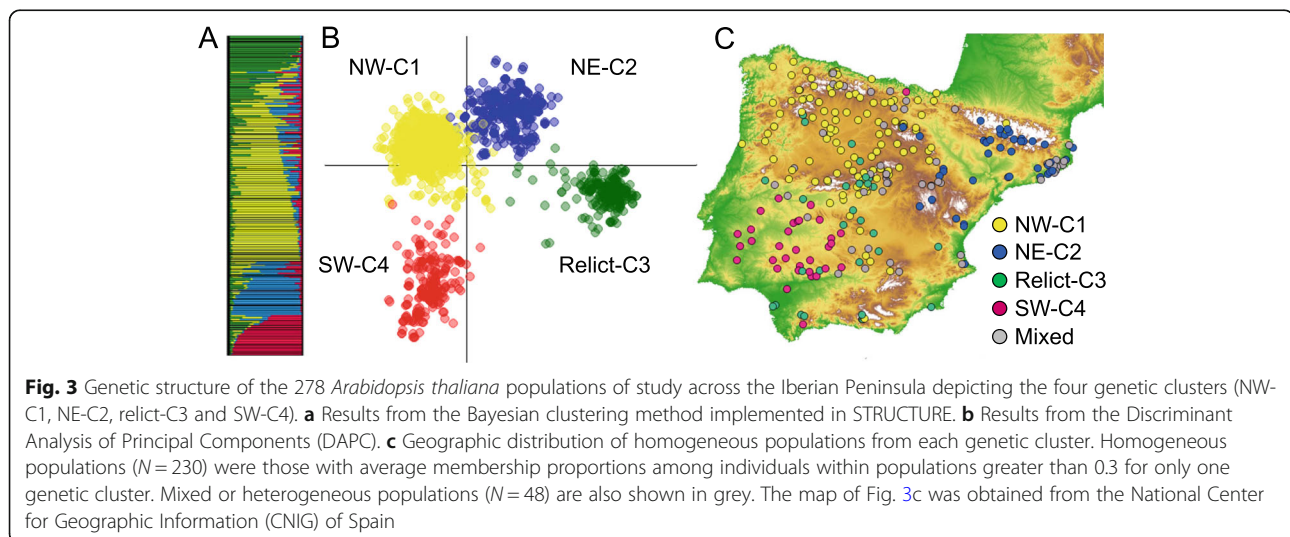


Table 1 Coefficients of the best spatial hierarchical Bayesian model for the geographic distribution of genetic diversity of Iberian *Arabidopsis thaliana*

Variables	$H_s \approx 0$		$H_s > 0$	
	β	SD	β	SD
Intercept	2.054	1.776	0.305	0.604
BIO15	-0.025	0.017	-0.023	0.006
BIO18	-0.009	0.004	-0.004	0.001
Topsoil pH	0.176	0.200	-0.187	0.067

Entries ($\beta \pm$ SD) are given for the degenerate distribution with point mass and zero (binary data including populations with H_s values ≈ 0) and a conditional-to-presence continuous process (continuous data including populations with non-zero H_s values). The best model had a WAIC value of -365.43 and a LCPO value of -0.31. BIO15 is the precipitation seasonality and BIO18 is the precipitation of the warmest quarter

suggesting that dispersal limitation and genetic drift are probably more efficient than local adaptation in shaping genetic differentiation patterns in *A. thaliana*. We found additional support to this conclusion in the low number of populations located at short distances sharing the same multilocus genotypes, stressing the limited natural *A. thaliana*'s dispersal ability [41, 47–49].

In contrast, IBR did not account for genetic differentiation at a regional scale (Fig. 4a), suggesting that landscape heterogeneity was not relevant for genetic differentiation in Iberian *A. thaliana*. We believe that the lack of IBR found across the Iberian Peninsula is likely due to the species' cosmopolitan habit and a

Table 2 Summary statistics for the top-ranked NMLPE models evaluating the effect of isolation-by-distance (IBD), isolation-by-environment (IBE) and isolation-by-resistance (IBR) on the genetic differentiation among Iberian *Arabidopsis thaliana* populations

Cluster	Predictors	AIC	Δ AIC	Weight
A – Model: IBD + IBE + IBR				
IP	IBD + IBE	-128,716.6	0.00	1.00
NW-C1	IBD + IBE	-42,836.0	0.00	1.00
NE-C2	IBD + IBE	-10,406.3	0.00	1.00
Relict-C3	IBD + IBE	-2938.5	0.00	1.00
SW-C4	IBR	-3753.4	0.00	0.79
B – Model: IBD + PC1 + PC2 + PC3 + IBR				
IP	IBD + PC1 + PC2 + PC3	-128,694.0	0.00	1.00
NW-C1	IBD + PC1 + PC2 + PC3	-42,856.2	0.00	1.00
NE-C2	IBD + PC1 + PC2 + PC3	-10,380.3	0.00	0.83
Relict-C3	IBD + PC1 + PC2 + PC3	-2936.6	0.00	0.89
SW-C4	IBR	-3753.4	0.00	0.41
	IBR + PC2	-3751.6	1.84	0.17

Entries are given for the entire Iberian Peninsula (IP) and the four genetic clusters detected using all populations available ($N = 278$). The two models include IBD, IBE and IBR, and IBE given the principal analysis components (PCs). Akaike information criterion (AIC), Δ AIC and model weight are reported

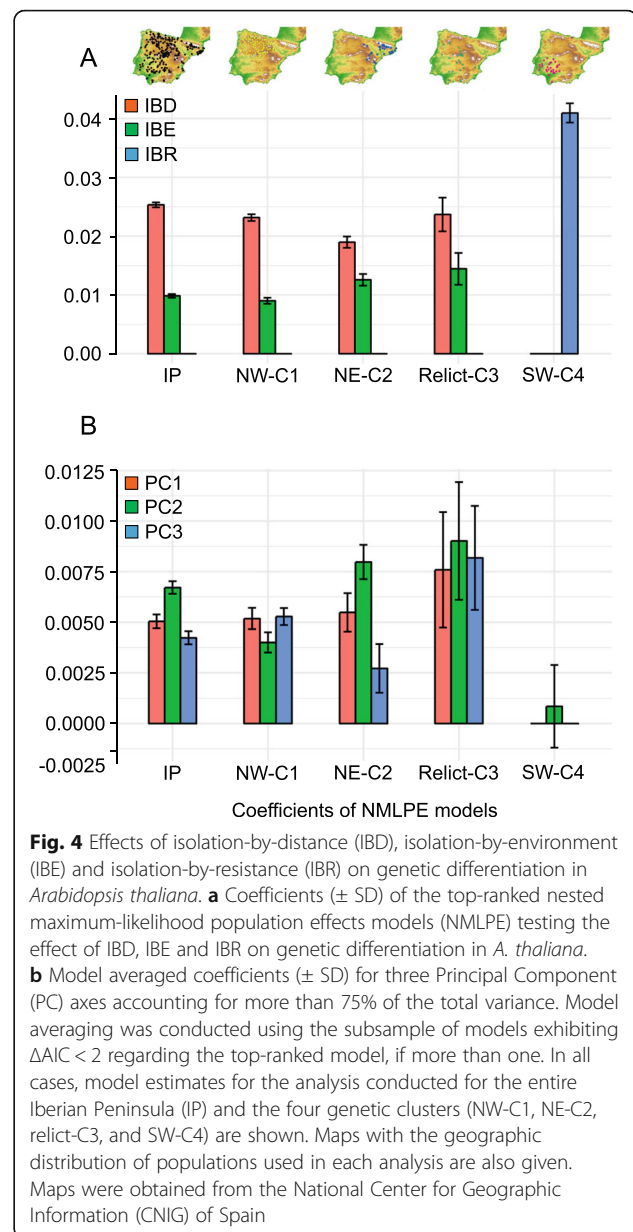


Fig. 4 Effects of isolation-by-distance (IBD), isolation-by-environment (IBE) and isolation-by-resistance (IBR) on genetic differentiation in *Arabidopsis thaliana*. **a** Coefficients (\pm SD) of the top-ranked nested maximum-likelihood population effects models (NMLPE) testing the effect of IBD, IBE and IBR on genetic differentiation in *A. thaliana*. **b** Model averaged coefficients (\pm SD) for three Principal Component (PC) axes accounting for more than 75% of the total variance. Model averaging was conducted using the subsample of models exhibiting Δ AIC < 2 regarding the top-ranked model, if more than one. In all cases, model estimates for the analysis conducted for the entire Iberian Peninsula (IP) and the four genetic clusters (NW-C1, NE-C2, relict-C3, and SW-C4) are shown. Maps with the geographic distribution of populations used in each analysis are also given. Maps were obtained from the National Center for Geographic Information (CNIG) of Spain

distribution without major discontinuities (Fig. S1). Thus, most population pairs are connected by relatively high habitat suitability, diminishing resistance distances. The demographic history, characterized by the presence of the relict African lineage and other non-relict Eurasian lineages [28–31, 34, 35, 50], and the well-known ability to adjust flowering time and seed dormancy to contrasting environments [38, 51–55], both account for the cosmopolitan habit of Iberian *A. thaliana*.

We also explored the effect of IBD, IBE and IBR for each of the four genetic clusters, which represent the joint outcome of the long demographic and adaptive history of the species in the region [30–32, 34, 39, 56]. Overall, NW-C1, NE-C2 and relict-C3 exhibited the same patterns than

those at the regional scale, i.e. IBD and IBE mostly accounted for genetic differentiation within each cluster (Fig. 4a). However, it is worth noting that the importance of IBE was not the same for these three clusters, since NE-C2 and relict-C3 exhibited higher IBE than NW-C1 (Fig. 4a), stressing probably the higher influence of local adaptation for genetic differentiation in these clusters. In fact, NE-C2 is a cluster with pronounced altitudinal gradients in NE Spain (from maritime to sub-alpine environments) along which the species is known to have adapted by adjusting various life-cycle and physiological traits [57–61]. In the case of relict-C3, the long evolutionary history of this lineage has provided the means to survive and adapt to diverse Iberian environments over the last millennia [35]. On the contrary, that was not the case for SW-C4, whose genetic differentiation was accounted for by IBR (Fig. 4a). Whereas SW-C4 is the genetic cluster with the most restricted geographic distribution (Fig. 3c), its populations are located across an area showing dramatic changes in habitat suitability (Fig. S1). Most likely, this particularity accounted for the enormous weight of IBR in this cluster.

Finally, we also conducted NMLPE using principal component (PC) axes to pinpoint the major environmental predictors underlying IBE. We found that the three PCs depicting gradients of temperature and/or precipitation across the region made substantial contributions to IBE in NW-C1, NE-C2 and relict-C3 Iberian clusters (Fig. 4b), probably due to the similar contributions of the three PCs to the environmental variation among these populations. In addition, as IBE informs on the influence of local adaptation on genetic differentiation, our results reinforce the view of the strong effects of environmental gradients on local adaptation in Iberian *A. thaliana*. For example, previous studies showed that geographic variation in fitness-related traits in Iberian *A. thaliana* (flowering time and seed dormancy) strongly co-varied as a function of temperature and to a lesser extent of precipitation [30, 32, 33, 38, 55, 62, 63]. In particular, *A. thaliana* adjusts its life cycle by advancing flowering time and increasing seed dormancy as the environment becomes warmer, drier and more seasonal [38, 55, 63].

As genetic diversity and its within- and among-population distribution patterns represents the raw material upon which genetic differentiation is estimated, we discuss some results dealing with genetic diversity of Iberian *A. thaliana* that are worth considering. For example, analyses of the genetic diversity in this large number of populations showed that 24% of them had no or very low genetic diversity (Fig. 1a–c). This is not exceptional as other authors also detected populations with practically no genetic diversity in *A. thaliana* [40, 41, 49, 64, 65]. Strong founder effects [66] and low migration rates [41] could well account for this result. Furthermore, small *A.*

thaliana populations rather isolated and locally adapted to particular environments can have very low genetic diversity [67], a scenario that could also apply to some of our populations. Finally, sampling bias might also affect genetic diversity [68], particularly because some populations were sampled late in the reproductive season.

Whatever the probable combination of factors accounting for the existence of *A. thaliana* populations with no or very low genetic diversity, the spatial hierarchical Bayesian model allowed the analysis of the geographic distribution of genetic diversity in Iberian *A. thaliana* (Fig. 2), particularly to detect the environmental predictors of the geographic distribution of genetic diversity. Precipitation, and not temperature, emerged as the major environmental factor accounting for the split of populations with and without genetic diversity, as well as for genetic diversity (Table 1). This result is in agreement with other studies on the distribution of genetic diversity in plants, which indicated a trend for a higher role of precipitation variables over those of temperature [69–74]. In addition, precipitation seasonality, which determines climatic variation in the Iberian Peninsula spanning from drier Mediterranean to more humid Atlantic climates, was the most important bioclimatic variable for the geographic distribution of genetic diversity in *A. thaliana* (Table 1). Previous studies also found that precipitation seasonality was a good predictor of individual fitness [33] and distribution of the four genetic clusters [39]. Albeit precipitation seasonality may be hard to handle as a fixed factor, further experiments are needed to better understand how *A. thaliana* responds to this variable, which might be a major evolutionary force in this species.

Conclusions

This study dissected the complexity of geographic, environmental and evolutionary factors contributing to the genetic differentiation among *A. thaliana* populations, while illustrating the power of dense collections to disentangle complex biological questions [8, 28, 30, 49, 63, 75, 76]. Beyond genetic differentiation patterns, we also identified the environmental predictors accounting for genetic diversity within populations. Clearly, the processes driving among-population genetic differentiation may not be the same than those determining the amount of within-population genetic diversity. However, we need to deal with the latter to understand the former. In this sense, recent resurrection studies showed that natural *A. thaliana* populations exhibited substantial changes in their genetic composition in just a decade [77, 78]. The extent of temporal genetic variation in *A. thaliana* represents a reminder that populations are not static, and that the geographic distribution of genetic diversity within and among-populations is changing too. Repeated genomic scans over time on the same populations will provide real-time insights into the intensity and pace of

genetic variation within populations, which will increase our understanding of the evolutionary processes shaping genetic differentiation in plant populations.

Methods

Source populations

We sampled 278 natural populations of the annual plant *Arabidopsis thaliana* (L.) Heyhn. (Brassicaceae) across the entire Iberian Peninsula ($\sim 800 \times 700 \text{ km}^2$; $36.00\text{--}43.48^\circ\text{N}$, $3.19^\circ\text{E}\text{--}9.30^\circ\text{W}$; Fig. 1a) during the decade of the 2000s. C. Alonso-Blanco and F.X. Picó identified and collected all the material. No voucher specimens of this material were deposited in a publicly available herbarium. For each population, we recorded its geographic coordinates and altitude using a GPS (Garmin International, Inc., Olathe, KS, USA). We extracted environmental data from publicly available repositories, such as WorldClim v.2 [79]; <https://www.worldclim.org/>; accessed 6 June 2019), The CORINE Land Cover 2000 (<https://land.copernicus.eu/pan-european/corine-land-cover>; accessed 6 June 2019), and The Soil Geographical Database from Eurasia v.4 (<https://esdac.jrc.ec.europa.eu/tags/soil-geographical-database-eur-asia>; accessed 6 June 2019).

Geographic distance among populations varied between 1 and 1059 km (Fig. 1a) and altitude between 1 and 2662 m.a.s.l. (Fig. 1b). For this set of populations, annual mean temperature varied between 5.3 and 18.4 °C (mean \pm SD = 12.3 ± 2.7 °C), annual total precipitation between 216.2 and 1778.8 mm (mean \pm SD = 760.9 ± 289.5 mm), percentage of natural vegetation (i.e. forests, scrublands or grasslands) between 0 and 100% (mean \pm SD = $61.3 \pm 35.7\%$), and soil pH between 3.6 and 7.5 (mean \pm SD = 5.7 ± 0.8).

Natural *A. thaliana* populations are made of patches of individuals widely differing in size and density, which is important when it comes to design a sampling scheme to study the spatio-temporal distribution of genetic diversity in *A. thaliana* [49, 57, 65, 77]. For each population and whenever possible, we collected seeds from several individuals from different patches (separated 1–20 m from each other) representing well each study population. Every sampling year and a few months after sampling, we multiplied field-collected seeds by the single seed descent method, in a glasshouse at the Centro Nacional de Biotecnología (CNB-CSIC) in Madrid. Bulk seeds were stored in dry, dark conditions in cellophane bags at room temperature, storing conditions that can preserve *A. thaliana* seeds for years.

In this study, each population included about six individuals, ranging between four and seven. Given the importance to work with a dense collection of natural populations, we selected this number as a compromise between the number of populations and the number of individuals per population that could be handled to extract the genetic patterns of interest. This choice was also based on a previous large-scale study exploring the within- and

among-population partitioning of genetic diversity in *A. thaliana*, which yielded interpretable results with four individuals per population [40]. Besides, other studies using more individuals per population and different markers, albeit at smaller geographical scales, also found populations with no or very low genetic diversity (see below), indicating that is not unusual in *A. thaliana* [40, 41, 49, 64, 65, 77].

We collected all seeds from wild populations. *Arabidopsis thaliana* is a common plant species not categorized as protected or endangered in any species list of the Convention on the Trade in Endangered Species of Wild Fauna and Flora. We carried out field sampling in locations where no permission was required, except at Doñana National Park (permission issued by Estación Biológica de Doñana) and Sierra de Grazalema Natural Park (permission issued by Red Andaluza de Jardines Botánicos de la Consejería de Medio Ambiente y Ordenación del Territorio de la Junta de Andalucía). Based on the Royal Decree of the Spanish legislation (Real Decreto 124/2017, de 24 de febrero; <https://www.boe.es/eli/es/rd/2017/02/24/124>), the genetic resources included in this study fall within the definition of “exclusively for taxonomic purposes” as they were used for scientific, educational or non-commercial purposes.

Historical factors

Recent historical changes in land use or large-scale perturbations may dramatically affect the amount of genetic diversity in plant populations. Such changes can be a misleading factor when attempting to evaluate the joint effects of IBD, IBE and IBR on genetic differentiation. This is because IBD, IBE and IBR do not contemplate perturbations producing sudden or erratic changes in genetic diversity that eventually may mask the patterns of interest. To make sure that historical factors were not an issue in this study, we estimated the temporal landscape dynamics of all study populations. We examined all aerial orthophotographs available for each population (spanning between 1945 and 2016), which were retrieved from different public regional administrations in Spain [77] with a tool specifically developed for this purpose. We excluded up to 15 populations because aerial orthophotographs were not available. The final number of aerial orthophotographs was highly variable per population, ranging between 4 and 24 (mean \pm SE = 8.6 ± 0.3 aerial orthophotographs) and covering between 9 and 71 years (37.1 ± 1.7 years between the first and the last orthophotograph).

To quantify temporal landscape dynamics, a circular area (500 m radius) around the GPS coordinate was divided into a regular grid of 80 squares (100 m side) for each aerial orthophotograph available per population. Based on vegetation and land use, each square was categorized as forests, scrublands, dehesa (i.e. agro-silvicultural

ecosystems based on Mediterranean oak woodlands), bare soil, crops, urban, infrastructures, and water, when one of these categories occupied at least more than half of the square. Categories were assigned to each square by digitalizing manually all squares from all aerial orthophotographs available ($N = 2252$ orthophotographs) with QGIS v.3.4 [80]. For each habitat type and population, we estimated the average change between year intervals available for each population to quantify landscape changes over time.

SNP genotyping

A total of 1772 *A. thaliana* individuals from 278 populations were genotyped with 245 presumably neutral nuclear SNPs using the SNPlex technique (Applied Biosystems, Foster City, CA, USA) through the CEGEN Genotyping Service (www.usc.es/cegen/). These genome-wide SNPs are frequent polymorphisms in Central Europe, the Iberian Peninsula, and worldwide collections, which altogether minimized ascertainment bias [31, 32, 34]. On average, there were about 49 SNPs per chromosome (range = 45–53 SNPs) located at approximately 0.5 Mb from each other (range = 0.11 Kb – 1.82 Mb). Four SNPs had percentages of missing data above 25% and one was monomorphic. We discarded these five SNPs from the analyses maintaining 240 SNPs.

Genetic analyses

For each *A. thaliana* population, we calculated the percentage of polymorphic loci (PL), the mean number of observed alleles per locus (n_a) and mean gene diversity (H_S) using FSTAT v.2.9.3 [81]. In addition, we computed the percentage of differences among all pairs of non-redundant multilocus genotypes. We used the ‘pairwise.WCfst’ function implemented in the R package *hierfstat* v.0.04–22 [82] to estimate genetic differentiation as pairwise F_{ST} values according to Weir and Cockerham [68], which represents the dependent variable to test IBD, IBE and IBR (see below).

The genetic structure of *A. thaliana* populations across the Iberian Peninsula was examined using two different methods: the Bayesian clustering method implemented in STRUCTURE v.2.3.3 [83, 84] and the ordination method represented by the Discriminant Analysis of Principal Components (DAPC) [85] available in the R package *adegenet* v.2.1.1 [85]. In the case of STRUCTURE, model settings included haploid non-redundant multilocus genotypes, correlated allele frequencies between populations and a linkage model. We identified the number of identical multilocus genotypes using the ‘mlg.filter’ function implemented in *adegenet*. Each run consisted of 50,000 burn-in MCMC iterations and 100,000 MCMC afterburning repetitions for parameter estimation. To determine the K number of ancestral populations and the

ancestry membership proportions of each accession in each population, we ran the algorithm 20 times for each defined number of groups (K value) from 1 to 10. The number of distinct genetic clusters was determined by evaluating the differences between the data likelihood for successive K values. The largest K value with significantly higher likelihood than that of K-1 runs (two-sided $P < 0.005$; Wilcoxon tests for related samples) gave the final K number. This was supported by a high similarity among the ancestry membership matrices from different runs of the same K value ($H' = 0.99$). We used CLUMPP v.1 [86] to calculate the average symmetric similarity coefficient H' among runs and the average matrix of ancestry membership proportions, derived from the 10 runs having the highest likelihood. STRUCTURE simulations were conducted at the The Supercomputing Center of Galicia (CESGA; <http://www.cesga.es/>).

For the DAPC analysis, we assessed the number of clusters using the ‘find.cluster’ function in *adegenet*, which runs successive K-means clustering with an increasing number of clusters to determine the best supported number of genetic clusters using the Bayesian Information Criterion (BIC). The K value with the lowest BIC value represents the optimal number of clusters, although BIC values may keep decreasing after the true K value in case of genetic clines and hierarchical structure [87]. Therefore, the rate of decrease in BIC values was visually examined to identify values of K, after which BIC values only decreased in a subtle manner [87]. The ‘dapc’ function was used with the final K value, retaining the axes of the Principal Component Analysis.

Geographic distribution of genetic diversity

We modeled the effects of environmental heterogeneity on the spatial distribution of genetic diversity by using a modified version of a spatial hierarchical Bayesian model recently used to model the effects of warming on the Iberian *A. thaliana*’s distribution [56]. The aim of this model was to visualize hot and cold spots of genetic diversity across the Iberian Peninsula and to identify their environmental predictors.

We used a Bayesian approach to handle the particularities of genetic diversity data, such as the semi-continuous nature of the variable. In the case of Iberian *A. thaliana*, 66 of 278 populations had very low mean gene diversity (H_S) values ($H_S < 0.009$), whereas the rest of populations exhibited non-zero H_S values (Fig. 1c). To handle such complex distribution of H_S , we developed a Hurdle-beta Bayesian model [88] to separate the zero structure (absence of genetic diversity) from the non-zero structure (presence of genetic diversity) of data. The Hurdle model is defined as a finite mixture of two processes: a degenerate distribution with point mass and zero, which determines the presence or absence of data—in our case H_S

values above or below the value of 0.009, respectively—and a conditional-to-presence continuous process supported by the open interval between the lowest non-zero H_S value and 1. The first process (absence of genetic diversity) was modeled with a Bernoulli distribution, whilst the continuous component (presence of genetic diversity) was modeled with a Beta distribution (see the full development of the model in the Supporting Information section). The predictors of the model were the 19 bioclimatic variables available from WorldClim and topsoil pH available from The Geographical Database from Eurasia. The Watanabe–Akaike information criterion (WAIC) [89]; was computed to determine the best models. The model included a stochastic spatial effect (the spatial term) to remove the spatial autocorrelation of data [56].

Drivers of genetic differentiation

We estimated the ecological, genetic and evolutionary drivers of genetic differentiation of Iberian *A. thaliana* as follows. IBD was based on geographic distance. However, instead of using the Euclidian distance among population pairs, we created a raster by assigning a 0.5 value to all 1-km² pixels [90, 91] of the Iberian Peninsula map. We then calculated pairwise distances, known as geographic resistance distance, among all *A. thaliana* populations employing the new raster. We did not use the Euclidean distance to compute IBD because straight lines between population pairs may introduce some bias when such straight lines traverses natural barriers for the study organism (i.e. the coastline or open sea) [92].

IBE was based on the 19 bioclimatic variables. In addition, we used The CORINE Land Cover to obtain the percentage of human-modified habitat (i.e. urban areas, crops and semi-natural grasslands) within a 500 m radius from the population GPS coordinate [39]. The percentages of human-modified and natural habitat were significantly negatively correlated ($N = 278$; $r = -0.95$; $P < 0.001$; Pearson's correlation) and nearly summed to 100%. Finally, a topsoil pH layer was also used. Combining these 21 environmental variables, we conducted a principal component analysis (PCA) with varimax rotation using the software SPSS v.25 (IBM, Chicago, IL USA). Thus, we obtained the PC scores of the first three principal components (PC) for each population. We then used the PC scores to calculate environmental dissimilarity between populations as Euclidean distances derived with the 'dist' function in R. Furthermore, we also calculated dissimilarity matrices among populations for each PC axis separately, as they were interpretable in terms of environmental gradients.

To compute IBR, we first estimated habitat suitability for *A. thaliana* across the Iberian Peninsula with the maximum-entropy modeling technique implemented in the software Maxent 3.3.3 k [93, 94]. To do that, we used

478 *A. thaliana* populations available to date (collected between 2000 and 2019; Fig. S1), and the same environmental variables used in a previous habitat suitability model [39]. In particular, the model included topsoil pH and eight bioclimatic variables, which were annual mean temperature (BIO1), mean diurnal temperature range (BIO2), isothermality (BIO3), temperature seasonality (BIO4), mean temperature of wettest quarter (BIO8), annual precipitation (BIO12), precipitation seasonality (BIO15) and precipitation of the warmest quarter (BIO18). We applied Maxent using default parameters except for features using the hinge type, making it comparable to a Generalized Additive Model [94]. We obtained consistent results with previous habitat suitability models conducted for Iberian *A. thaliana* with lower sample sizes [39, 56].

Next, we transformed the resulting habitat suitability into resistance values as 1 minus the value of each pixel. Thus, greater pixel values represented lower occurrence probability of *A. thaliana* (higher resistance) and vice-versa. Using circuit theory [24, 25], we examined whether genetic differentiation among *A. thaliana* populations was accounted for by IBR. We used Circuitscape v.4.0 [95] to calculate resistance distance matrices assigning pixel values as resistance values. We calculated resistance distance among all pairs of populations across Iberian Peninsula, as well as among all pairs of populations within each genetic cluster.

To quantify the relative contributions of IBD, IBE and IBR on the genetic differentiation (pairwise F_{ST} values) among *A. thaliana* populations, we fitted maximum-likelihood population effect models (MLPE) [96]. Code implementing the MLPE correlation structure within the R package nlme v. 3.1–143 comes from the corMLPE package. The model used penalized least squares and a residual covariance structure designed to account for the non-independence of pairwise distances. Given the inherent spatial dependence structure in pairwise comparisons between *A. thaliana* populations, all MLPE considered spatial locations. To this end, we used a modification of the MLPE model incorporating the correlation between pairwise measurements due to comparison of populations and spatial locations (nested MLPE or NMLPE) [91]. Since the independent variables (IBD, IBE and IBR) used in NMLPE had very different units, we normalized them by subtracting the mean and dividing by the standard deviation in all analyses. We did not linearize pairwise F_{ST} values. We conducted NMLPE on the complete dataset ($N = 278$ populations) as well as on the subset of those populations exhibiting H_S different from zero ($N = 212$ populations with $H_S > 0.009$). The results were consistent between the two datasets, showing that the 66 populations with no genetic diversity were not biasing the results.

In addition, we ran NMLPE for each genetic cluster to evaluate whether the effects of IBD, IBE and IBR on genetic differentiation differed among genetic clusters. This is because the number of genetic clusters represents the outcome of the demographic and evolutionary history of *A. thaliana* in the region. We considered each population to belong to a specific genetic cluster when the average value of the membership proportions of its individuals was ≥ 0.3 for that specific cluster. Given the marked genetic structure of the study system, threshold values of 0.25 and 0.40 were consistent, but 0.3 provided a clearer classification of individuals into genetic clusters. The majority of populations were assigned to a single cluster ($N = 230$). We assigned populations reaching the minimum average membership proportion for more than one cluster to multiple clusters. Overall, 47 populations were assigned to two clusters, whereas only one population had average membership proportions higher than 0.3 for three clusters.

We also ran NMLPE on the complete dataset and each cluster separately, but replacing IBE by the PC axes. Thus, we used IBD, PC1, PC2, PC3 and IBR as independent variables. We used the Akaike information criterion (AIC) to compare models containing all possible combinations of non-collinear predictions ($r < 0.6$), created with the 'dredge' function from the R package MuMIn v1.4 [97]. In those cases where several models were top-ranked with $\Delta AIC < 2$, we calculated the model averaged parameter estimates (β) with standard errors for the explanatory variables included in all top-ranked models also using MuMIn. We considered effect sizes as significant when 95% confidence intervals did not overlap zero.

Abbreviations

AIC: Akaike information criterion; BIC: Bayesian information criterion; DAPC: Discriminant analysis of principal components; IBD: Isolation-by-distance; IBE: Isolation-by-environment; IBR: Isolation-by-resistance; MCMC: Monte Carlo Markov chain; NMLPE: Nested maximum-likelihood population effect models; SNP: Single nucleotide polymorphism; WAIC: Watanabe–Akaike information criterion

Acknowledgements

We thank Mercedes Ramiro for technical assistance. We are grateful to David Ragel, Marisol Muñoz and Rocío Gómez for digitalizing all orthophotographs used in this study. We are indebted to Joaquín Ortego, Rodolfo Jaffé, Nathaniel Pope, and Bernardo Toledo for their insightful and constructive comments on various aspects of the manuscript during its development. We thank two anonymous reviewers and the editor for constructive comments on this manuscript. We particularly thank Nathaniel Pope for his help with the implementation of NMLPE. The Supercomputing Center of Galicia (CESGA) facilitated the analyses of genetic data. We acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI).

Authors' contributions

CA-B and FXP conceived the idea. BM-V, CA-B and FXP collected seeds, grew plants and generated the SNP data. ARC and AM analyzed the data. AM developed the habitat suitability model. JM-M and DC developed the Bayesian model. FXP wrote the first version of the manuscript and all authors contributed to it.

Funding

This research has been funded by grant BIO2016–75754-P from the Agencia Estatal de Investigación (AEI) of Spain and the Fondo Europeo de Desarrollo Regional (FEDER, UE) to CA-B. The funding body played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

Data deposited in the Dryad repository: <https://doi.org/10.5061/dryad.47d7wm393> [98]. The tool to retrieve and examine aerial orthophotographs deposited in the Zenodo repository: <https://zenodo.org/record/2552031>.

Ethics approval and consent to participate

The species sampled is not endangered or protected.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Centre for Applied Ecology "Prof. Baeta Neves", InBIO, School of Agriculture, University of Lisbon, Lisbon, Portugal. ²Departamento de Ecología Integrativa, Estación Biológica de Doñana (EBD), Consejo Superior de Investigaciones Científicas (CSIC), Sevilla, Spain. ³Departamento de Genética Molecular de Plantas, Centro Nacional de Biotecnología (CNB), Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain. ⁴CREAF, Centre de Recerca Ecològica i Aplicacions Forestals, Bellaterra, E08193 Cerdanyola de Vallès, Catalonia, Spain. ⁵Universitat Autònoma de Barcelona, Bellaterra, E08193 Cerdanyola de Vallès, Catalonia, Spain. ⁶BCAM - Basque Center for Applied Mathematics, Bilbao, Spain. ⁷Departament d'Estadística i Investigació Operativa, Universitat de València, Valencia, Spain.

Received: 13 February 2020 Accepted: 1 June 2020

Published online: 22 June 2020

References

- Barrett RDH, Schluter D. Adaptation from standing genetic variation. *Trends Ecol Evol.* 2008;23:38–44.
- Jump AS, Marchant R, Peñuelas J. Environmental change and the option value of genetic diversity. *Trends Plant Sci.* 2009;14:51–8.
- Hancock AM, Brachi B, Faure N, Horton MW, Jarymowicz LB, Sperone FG, et al. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science.* 2011;334:83–6.
- Wilczek AM, Cooper MD, Korves TM, Schmitt J. Lagging adaptation to warming climate in *Arabidopsis thaliana*. *P Natl Acad Sci USA.* 2014;111:7906–13.
- Matuszewski S, Hermisson J, Kopp M. Catch me if you can: adaptation from standing genetic variation to a moving phenotypic optimum. *Genetics.* 2015;200:1255–74.
- Fournier-Level A, Perry EO, Wang JA, Braun PT, Migneault A, Cooper MD, et al. Predicting the evolutionary dynamics of seasonal adaptation to novel climates in *Arabidopsis thaliana*. *P Natl Acad Sci USA.* 2016;113:E2812–21.
- Exposito-Alonso M, Brennan AC, Alonso-Blanco C, Picó FX. Spatio-temporal variation in fitness responses to contrasting environments in *Arabidopsis thaliana*. *Evolution.* 2018;72:1570–86.
- Exposito-Alonso M. 500 genomes field experiment team, Burbano HA, Bossdorf O, Nielsen R, Weigel D. natural selection on the *Arabidopsis thaliana* genome in present and future climates. *Nature.* 2019;573:126–9.
- Fisher RA. The genetical theory of natural selection. Oxford: Oxford University Press; 1930.
- Orr HA. The genetic theory of adaptation: a brief history. *Nat Rev Genet.* 2005;6:119–27.
- Hughes AL. The origin of adaptive phenotypes. *P Natl Acad Sci USA.* 2008; 105:13193–4.
- Rauch EM, Bar-Yam Y. Theory predicts the uneven distribution of genetic diversity within species. *Nature.* 2004;431:449–52.
- Wright S. Isolation by distance. *Genetics.* 1943;28:114–38.
- Kimura M, Weiss GH. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics.* 1964;49:561–76.

15. Malécot G. Heterozygosity and relationship in regularly subdivided populations. *Theor Pop Biol.* 1975;8:212–41.
16. Slatkin M. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution.* 1993;47:264–79.
17. Rousset F. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics.* 1997;145:1219–28.
18. Sexton JP, Hangartner SB, Hoffmann AA. Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution.* 2014;68:1–15.
19. Jenkins DG, Carey M, Czerniewska J, Fletcher J, Hether T, Jones A, et al. A meta-analysis of isolation by distance: relic or reference standard for landscape genetics? *Ecography.* 2010;33:315–20.
20. Aguilón SM, Fitzpatrick JW, Bowman R, Schoech SJ, Clark AG, Coop G, et al. Deconstructing isolation-by-distance: the genomic consequences of limited dispersal. *PLoS Genet.* 2017;13:e1006911.
21. Wang IJ, Summers K. Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Mol Ecol.* 2010;19:447–58.
22. Wang IJ, Bradburd GS. Isolation by environment. *Mol Ecol.* 2014;23:5649–62.
23. Manel S, Schwartz MK, Luikart G, Taberlet P. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol.* 2003;18:189–97.
24. McRae BH. Isolation by resistance. *Evolution.* 2006;60:1551–61.
25. McRae BH, Beier P. Circuit theory predicts gene flow in plant and animal populations. *P Natl Acad Sci USA.* 2007;104:19885–90.
26. Spear SF, Balkenhol N, Fortin M-J, McRae BH, Scribner K. Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Mol Ecol.* 2010;19:3576–91.
27. Lundgren E, Ralph PL. Are populations like a circuit? Comparing isolation by resistance to a new coalescent-based method. *Mol Ecol Resour.* 2019;19:1388–406.
28. The 1001 Genomes Consortium. 1,135 Genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell.* 2016;166:481–91.
29. Durvasula A, Fulgione A, Gutaker RM, Alacakaptan SI, Flood PJ, Neto C, et al. African genomes illuminate the early history and transition to selfing in *Arabidopsis thaliana*. *P Natl Acad Sci USA.* 2017;114:5213–8.
30. Tabas-Madrid D, Méndez-Vigo B, Arteaga N, Marcer A, Pascual-Montano A, Weigel D, et al. Genome-wide signatures of flowering adaptation to climate temperature: regional analyses in a highly diverse native range of *Arabidopsis thaliana*. *Plant Cell Environ.* 2018;41:1806–20.
31. Picó FX, Méndez-Vigo B, Martínez-Zapater JM, Alonso-Blanco C. Natural genetic variation of *Arabidopsis thaliana* is geographically structured in the Iberian Peninsula. *Genetics.* 2008;180:1009–21.
32. Méndez-Vigo B, Picó FX, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C. Altitudinal and climatic adaptation is mediated by flowering traits and *FR1*, *FLC* and *PHYC* genes in *Arabidopsis*. *Plant Physiol.* 2011;157:1942–55.
33. Manzano-Piedras E, Marcer A, Alonso-Blanco C, Picó FX. Deciphering the adjustment between environment and life history in annuals: lessons from a geographically-explicit approach in *Arabidopsis thaliana*. *PLoS One.* 2014;9:e87836.
34. Brennan AC, Méndez-Vigo B, Haddioui A, Martínez-Zapater JM, Picó FX, Alonso-Blanco C. The genetic structure of *Arabidopsis thaliana* in the South-Western Mediterranean range reveals a shared history between North Africa and southern Europe. *BMC Plant Biol.* 2014;14:17.
35. Toledo B, Marcer A, Méndez-Vigo B, Alonso-Blanco C, Picó FX. An ecological history of the relict genetic lineage of *Arabidopsis thaliana*. *Environ Exp Bot.* 2020;170:103800.
36. Legendre P, Dale MRT, Fortin M-J, Gurevitch J, Hohn M, Myers D. The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography.* 2002;25:601–15.
37. Méndez-Vigo B, Savic M, Ausin I, Ramiro M, Martín B, Picó FX, Alonso-Blanco C. Environmental and genetic interactions reveal *FLOWERING LOCUS C* as a modulator of the natural variation for the plasticity of flowering in *Arabidopsis*. *Plant Cell Environ.* 2016;39:282–94.
38. Vidigal DS, Marques ACSS, Willems LAJ, Buijs G, Méndez-Vigo B, Hilhorst HWM, et al. Altitudinal and climatic associations of seed dormancy and flowering traits evidence adaptation of annual life cycle timing in *Arabidopsis thaliana*. *Plant Cell Environ.* 2016;39:1737–48.
39. Marcer A, Méndez-Vigo B, Alonso-Blanco C, Picó FX. Tackling intraspecific genetic structure in distribution models better reflects species geographical range. *Ecol Evol.* 2016;6:2084–97.
40. Bakker EG, Stahl EA, Toomajan C, Nordborg M, Kreitman M, Bergelson J. Distribution of genetic variation within and among local populations of *Arabidopsis thaliana* over its species range. *Mol Ecol.* 2006;15:1405–18.
41. Falahati-Anbaran M, Lundemo S, Stenøien HK. Seed dispersal in time can counteract the effect of gene flow between natural populations of *Arabidopsis thaliana*. *New Phytol.* 2014;202:1043–54.
42. Sharbel TF, Haubold B, Mitchell-Olds T. Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Mol Ecol.* 2000;9:2109–18.
43. Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajan C, Zheng H, et al. The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* 2005;3:e196.
44. Ostrowski M-F, David J, Santoni S, McKhann H, Reboud X, Le Corre V, et al. Evidence for a large-scale population structure among accessions of *Arabidopsis thaliana*: possible causes and consequences for the distribution of linkage disequilibrium. *Mol Ecol.* 2006;15:1507–17.
45. Beck JB, Schmuths K, Schaal BA. Native range genetic variation in *Arabidopsis thaliana* is strongly geographically structured and reflects Pleistocene glacial dynamics. *Mol Ecol.* 2008;17:902–15.
46. Lundemo S, Falahati-Anbaran M, Stenøien HK. Seed banks cause elevated generation times and effective population sizes of *Arabidopsis thaliana* in northern Europe. *Mol Ecol.* 2009;18:2798–811.
47. Lewandowska-Sabat AM, Fjellheim S, Rogli OA. Extremely low genetic variability and highly structured local populations of *Arabidopsis thaliana* at higher latitudes. *Mol Ecol.* 2010;19:4753–64.
48. Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, et al. The scale of population structure in *Arabidopsis thaliana*. *PLoS Genet.* 2010;6:e1000843.
49. Bombliès K, Yant L, Laitinen RA, Kim ST, Hollister JD, Warthmann N, et al. Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *PLoS Genet.* 2010;6:e1000890.
50. Lee C-R, Svardal H, Farlow A, Exposito-Alonso M, Ding W, Novikova P, et al. On the post-glacial spread of human commensal *Arabidopsis thaliana*. *Nat Commun.* 2017;8:14458.
51. Donohue K. Niche construction through phenological plasticity: life history dynamics and ecological consequences. *New Phytol.* 2005;166:83–92.
52. Debieu M, Tang C, Stich B, Sikosek T, Effgen S, Josephs E, et al. Co-variation between seed dormancy, growth rate and flowering time changes with latitude in *Arabidopsis thaliana*. *PLoS One.* 2013;8:e61075.
53. Springthorpe V, Penfield S. Flowering time and seed dormancy control use external coincidence to generate life history strategy. *eLife.* 2015;4:e05557.
54. Burghardt LT, Edwards BR, Donohue K. Multiple paths to similar germination behavior in *Arabidopsis thaliana*. *New Phytol.* 2016;209:1301–12.
55. Marcer A, Vidigal DS, James PMA, Fortin M-J, Méndez-Vigo B, Hilhorst HWM, et al. Temperature fine-tunes Mediterranean *Arabidopsis thaliana* life-cycle phenology geographically. *Plant Biol.* 2018;20:148–56.
56. Martínez-Minaya J, Conesa D, Fortin M-J, Alonso-Blanco C, Picó FX, Marcer A. A hierarchical Bayesian Beta regression approach to study the effects of geographical genetic structure and spatial autocorrelation on species distribution range shifts. *Mol Ecol Resour.* 2019;19:929–43.
57. Montesinos A, Tonsor SJ, Alonso-Blanco C, Picó FX. Demographic and genetic patterns of variation among populations of *Arabidopsis thaliana* from contrasting native environments. *PLoS One.* 2009;4:e7213.
58. Montesinos-Navarro A, Wig J, Picó FX, Tonsor SJ. *Arabidopsis thaliana* populations show clinal variation in a climatic gradient associated with altitude. *New Phytol.* 2011;189:282–94.
59. Montesinos-Navarro A, Picó FX, Tonsor SJ. Clinal variation in seed traits influencing life cycle timing in *Arabidopsis thaliana*. *Evolution.* 2012;66:3417–31.
60. Wolfe MD, Tonsor SJ. Adaptation to spring heat and drought in northeastern Spanish *Arabidopsis thaliana*. *New Phytol.* 2014;201:323–34.
61. Zhang N, Belsterling B, Raszewski J, Tonsor SJ. Natural populations of *Arabidopsis thaliana* differ in seedling responses to high-temperature stress. *AoB Plants.* 2015;7:plv101.
62. Kronholm I, Picó FX, Alonso-Blanco C, Goudet J, de Meaux J. Genetic basis of adaptation in *Arabidopsis thaliana*: local adaptation at the seed dormancy QTL *DOG1*. *Evolution.* 2012;66:2287–302.
63. Martínez-Berdeja A, Stitzer MC, Taylor MA, Okada M, Ezcurra E, Runcie DE, et al. Functional variants of *DOG1* control seed chilling responses and variation in seasonal life-history strategies in *Arabidopsis thaliana*. *P Natl Acad Sci USA.* 2020;117:2526–34.
64. Stenøien HK, Fenster CB, Tonteri A, Savolainen O. Genetic variability in natural populations of *Arabidopsis thaliana* from northern Europe. *Mol Ecol.* 2005;14:137–48.

65. Gomaa NH, Montesinos-Navarro A, Alonso-Blanco C, Picó FX. Temporal variation in genetic diversity and effective population size of Mediterranean and subalpine *Arabidopsis thaliana* populations. *Mol Ecol*. 2011;20:3540–54.
66. Mayr E. Systematics and the origin of species: from the viewpoint of a zoologist. New York: Columbia University Press; 1942.
67. Méndez-Vigo B, Gomaa NH, Alonso-Blanco C, Picó FX. Among- and within-population variation in flowering time of Iberian *Arabidopsis thaliana* estimated in field and glasshouse conditions. *New Phytol*. 2013;197:1332–43.
68. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution*. 1984;38:1358–70.
69. Manel S, Gugerli F, Thuiller W, Alvarez N, Legendre P, Holderegger R, et al. Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Mol Ecol*. 2012;21:3729–38.
70. Cushman SA, Max T, Meneses N, Evans LM, Ferrier S, Honchak B, et al. Landscape genetic connectivity in a riparian foundation tree is jointly driven by climatic gradients and river networks. *Ecol Appl*. 2004;24:1000–14.
71. Mosca E, González-Martínez SC, Neale DB. Environmental versus geographical determinants of genetic structure in two subalpine conifers. *New Phytol*. 2014;201:180–92.
72. Mosca E, Gugerli F, Eckert AJ, Neale DB. Signatures of natural selection on *Pinus cembra* and *P. mugo* along elevational gradients in the Alps. *Tree Genet Genomes*. 2016;12:9.
73. Bothwell HM, Cushman SA, Woolbright SA, Hersch-Green EI, Evans LM, Whitham TG, et al. Conserving threatened riparian ecosystems in the American west: precipitation gradients and river networks drive genetic connectivity and diversity in a foundation riparian tree (*Populus angustifolia*). *Mol Ecol*. 2017;26:5114–32.
74. Cuervo-Alarcon L, Arend M, Müller M, Sperisen C, Finkeldey R, Krutovsky KV. Genetic variation and signatures of natural selection in populations of European beech (*Fagus sylvatica* L.) along precipitation gradients. *Tree Genet Genomes*. 2018;14:84.
75. Lasky JR, Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH. Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Mol Ecol*. 2012;21:5512–29.
76. Frachon L, Bartoli C, Carrère S, Bouchez O, Chaubet A, Gautier M, et al. A genomic map of climate adaptation in *Arabidopsis thaliana* at a micro-geographic scale. *Front Plant Sci*. 2018;9:967.
77. Gómez R, Méndez-Vigo B, Marcer A, Alonso-Blanco C, Picó FX. Quantifying temporal change in plant population attributes: insights from a resurrection approach. *AoB Plants*. 2018;10:ply063.
78. Frachon L, Libourel C, Villoutreix R, Carrère S, Glorieux C, Huard-Chauveau C, et al. Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time. *Nature Ecol Evol*. 2017;1:1551–61.
79. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol*. 2017;37:4302–15.
80. QGIS Development Team. QGIS Geographic information system. Open source geospatial foundation project. 2017. <http://qgis.osgeo.org>. Accessed 15 Mar 19.
81. Goudet J. FSTAT: a computer program to calculate F-statistics. *J Hered*. 1995;86:485–6.
82. Goudet J. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Mol Ecol Notes*. 2005;5:184–6.
83. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–59.
84. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*. 2003;164:1567–87.
85. Jombart T. Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. 2008;24:1403–5.
86. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 2007;23:1801–6.
87. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet*. 2010;11:94.
88. Quiroz ZC, Prates MO, Rue H. A Bayesian approach to estimate the biomass of anchovies off the coast of Perú. *Biometrics*. 2015;71:208–17.
89. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J Mach Learn Res*. 2010;11:3571–94.
90. Castilla AR, Pope N, Jaffé R, Jha S. Elevation, not deforestation, promotes genetic differentiation in a pioneer tropical tree. *PLoS One*. 2016;11:e0156694.
91. Jaffé R, Veiga JC, Pope NS, Lanes ECM, Carvalho CS, Alves R, et al. Landscape genomics to the rescue of a tropical bee threatened by habitat loss and climate change. *Evol Appl*. 2019;12:1164–77.
92. Pereoglou F, Lindenmayer DB, MacGregor C, Ford F, Wood J, Banks SC. Landscape genetics of an early successional specialist in a disturbance-prone environment. *Mol Ecol*. 2013;22:1267–81.
93. Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecol Model*. 2006;190:231–59.
94. Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. A statistical explanation of MaxEnt for ecologists. *Divers Distrib*. 2011;17:43–57.
95. McRae BH, Dickson BG, Keitt TH, Shah VB. Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology*. 2008;89:2712–24.
96. Clarke RT, Rothery P, Raybould AF. Confidence limits for regression relationships between distance matrices: estimating gene flow with distance. *J Agr Biol Envir St*. 2002;7:361–72.
97. Bartoń K. Package 'MuMIn'. 2019. <https://CRAN.R-project.org/package=MuMIn>. Accessed 15 Oct 2019.
98. Picó FX, Castilla AR, Méndez-Vigo B, Marcer A, Martínez-Minaya J, Conesa D, Alonso-Blanco C. Data from: ecological, genetic and evolutionary drivers of regional genetic differentiation in *Arabidopsis thaliana*. Dryad Digital Repository. 2020. <https://doi.org/10.5061/dryad.47d7wm393>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

