Research Article

**An ecological history of the relict genetic lineage of
*Arabidopsis thaliana***

**Bernardo Toledo¹, Arnald Marcer^{2,3}, Belén Méndez-Vigo⁴, Carlos Alonso-
Blanco⁴ and F. Xavier Picó^{1*}**

¹Departamento de Ecología Integrativa, Estación Biológica de Doñana (EBD), Consejo
Superior de Investigaciones Científicas (CSIC), 41092 Sevilla, Spain

²CREAF, E08193 Bellaterra (Cerdanyola del Vallès), Catalonia, Spain

³Universitat Autònoma de Barcelona, E08193 Bellaterra (Cerdanyola del Vallès), Catalonia,
Spain

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

Running title: Ecological history of relict *Arabidopsis*

*Corresponding author's e-mail address: xpico@ebd.csic.es

Abstract

The combination of extensive population sampling with whole-genome sequencing in the annual plant *Arabidopsis thaliana* has recently allowed the identification of a genetically differentiated relict lineage. The most important nuclei of relict *A. thaliana* is found in the Iberian Peninsula and North Africa, although relict accessions have also been found scattered across Atlantic oceanic islands, eastern Mediterranean Basin, South and East Africa and China. We hypothesised that relict *A. thaliana* long survived and adapted to a wide array of environments in the Iberian Peninsula and North Africa due to the fact that the region repeatedly acted as glacial refugia. Given the lack of evidence to support this hypothesis, we addressed this issue by studying the ecological history of Iberian and North African relict *A. thaliana*. To this end, we analysed the relationship between the current habitat suitability of relict *A. thaliana*, estimated with species distribution models, and the vegetation dynamics in the region over the last millennia using pollen fossil data from sediment cores and reconstructions of past distribution ranges of the most important tree species occurring in the region. Overall, our results indicated that a higher current habitat suitability of relict *A. thaliana* was correlated with more stable vegetation dynamics since the Last Glacial Maximum and during the Holocene. Given that relict *A. thaliana* is known to harbour genetic variation specific to its relict nature, we also evaluated the differentiation between relict and non-relict *A. thaliana* accessions at the functional genetic level for the known flowering genes, *CRY2* and *TSF*. Genomic surveys detected the existence of haplogroups of these genes occurring at very high frequency only among relict accessions. Overall, our study reinforced the relict character of this lineage thought to be at the base of the species' early history. We stress the need to sample in depth other geographic areas harbouring relict *A. thaliana* and conduct further functional genetic analyses between relict and non-relict accessions to keep disentangling the evolutionary trajectory of this annual plant.

Key words: *Arabidopsis thaliana*; flowering genes; paleobotanical data; remote sensing data; species distribution models; relict genetic lineage; whole-genome sequencing.

Introduction

The main episodes in the evolutionary history of any organism become registered in the geographic distribution of populations as well as imprinted in the genome. However, such evolutionary trajectories cannot be read in a clear-cut manner because the effects of major demographic (e.g. expansions, bottlenecks) and evolutionary (e.g. adaptive evolution, hybridization) events on distribution range and genomic signatures tend to accumulate over time in a confounding manner. For example, the genomic composition of populations may be partially wiped out by demographic turnovers, which may take place repeatedly over millennia (Lee et al., 2017). The only way to overcome this problem lies in the combination of extensive population sampling across the distribution range combined with whole-genome deep sequencing. In plants, outstanding examples of this approach exist for some model systems, such as the annual plant *Arabidopsis thaliana*. At present, the currently available resources of *A. thaliana* comprise thousands of accessions from all over the distribution range (Weigel, 2012; Provart et al., 2016) and more than 1,200 whole-genome sequences (The 1001 Genomes Consortium, 2016; Durvasula et al., 2017; Zou et al., 2017; Fulgione et al., 2018; Hsu et al., 2019), including some from herbarium specimens (Weiß et al., 2016; Durvasula et al., 2017).

Such long-term collaborative investment has recently paid off in a significant finding that is challenging the classical model of the evolutionary history of *A. thaliana* (Fulgione and Hancock, 2018). Based on the extent of genomic differentiation, it has been concluded that some Eurasian, African and Asian accessions represent divergent relict lineages, whereas the

vast majority of Eurasian ones likely derived from a great expansion of a single lineage (The 1001 Genomes Consortium, 2016; Durvasula et al., 2017; Lee et al., 2017; Tsuchimatsu et al., 2017; Zeng et al., 2017; Fulgione et al., 2018; Hsu et al., 2019). Such divergent relict lineages clearly emerge in a phylogenomic tree using *A. lyrata* as the outgroup, the closest relative of *A. thaliana* (Durvasula et al., 2017; Lee et al., 2017; Tsuchimatsu et al., 2017; Zeng et al., 2017; Fulgione et al., 2018; Hsu et al., 2019). Genomic, epigenomic and transcriptomic analyses have indicated that differentiation between relict and non-relict *A. thaliana* spans most of the genome, including a larger number of transposon elements-like methylated genes and approximately 20% of genes with differential expression between relict and non-relict accessions (The 1001 Genomes Consortium, 2016; Kawakatsu et al., 2016). In addition, some degree of genomic hybridization has also been detected in contact zones between relict and non-relict *A. thaliana*, such as the Iberian Peninsula (Durvasula et al., 2017) and China (Hsu et al., 2019), suggesting a complex demographic history that remains to be disentangled.

Thus, relict lineages, in particular the African ones, probably represent the deepest history of *A. thaliana*. In short, the origin of the species is dated in 1 Mya – 500 kya and the split into the relict and non-relict lineages in 120 – 90 kya, which corresponds to the Abbassia Pluvial that was a warm, humid period in Africa (Durvasula et al., 2017). Interestingly, historical population sizes in African relict *A. thaliana* appear to have been larger than those in Eurasia (Fulgione and Hancock, 2018). On top of that, African relict *A. thaliana* harbours genetic variation that was lost or never present in Eurasia, which overall confers the relict character to this genetic lineage. For example, genomic analysis of the self-incompatibility locus (S-locus) showed that only relict accessions from northern Africa carry all known major haplogroups, suggesting that the selfing ability of *A. thaliana* evolved in this region from multiple independent inactivation alleles (Tsuchimatsu et al., 2017; Durvasula et al., 2017). Furthermore, some flowering time genes have also revealed differentiation of relict accessions

for the frequency of natural polymorphisms affecting gene function, such as a very low frequency of early loss-of-function alleles of *FRIGIDA* and multiple functional polymorphisms at *FLC*, *FT*, *FLM*, *PHYB*, *PHYC* and *LHY* genes (Mendez-Vigo et al., 2011; Brennan et al., 2014; Liu et al., 2014; Lutz et al., 2017; de Montaigu and Coupland, 2017; James et al., 2018). In addition, other genes have been detected by phenotypic and environmental genome wide-association (GWA) analyses, like the flowering promoters *TSF* and *SOC1*, the transcription factors *LUG* and *SLK1* involved in flower development, or the stress regulator *EIN2* (The 1001 Genomes Consortium, 2016; Lee et al. 2017; Tabas-Madrid et al., 2018).

Overall, an out-of-Africa scenario as a plausible demographic model for *A. thaliana* history is replacing the classical model of the Central Asian origin of the species, which implied an expansion into Europe and Asia and leaving Africa out from any interpretation (Durvasula et al., 2017; Fulgione and Hancock, 2018; but see Lee et al., 2017; Hsu et al., 2019). So far, the most important nuclei of relict *A. thaliana* can be found in the Iberian Peninsula and Morocco, which was detected first as a highly differentiated genetic cluster with genome-wide nuclear polymorphisms (Picó et al., 2008; Brennan et al., 2014; Marcet et al., 2016) and later confirmed with whole-genome sequencing (The 1001 Genomes Consortium, 2016; Durvasula et al., 2017). Interestingly, other pools of relict *A. thaliana* populations have also been identified in geographic areas as distant as Madeira (Fulgione et al., 2018), western and southern Africa (Durvasula et al., 2017), Tibet (Zeng et al., 2017), the Chinese province of Yunnan (Hsu et al., 2019), and previously known accessions scattered across Atlantic oceanic islands (i.e. Cape Verde and the Canary Islands) and the Mediterranean Basin (i.e. Sicily, Algeria and Lebanon; The 1001 Genomes Consortium, 2016; Durvasula et al., 2017; Hsu et al., 2019).

However, an important piece of information is totally lacking to figure out how relict *A. thaliana* has been able to endure to the present day: the ecological history of the relict lineage of *A. thaliana*. We only know that relict *A. thaliana* populations occur in a diverse array of natural environments with a limited impact of human activities (Marcer et al., 2016; The 1001 Genomes Consortium, 2016). This means that relict *A. thaliana* behaves more like an annual plant belonging to different natural plant communities (see further details on the biology of the species in Montesinos et al., 2009; Picó, 2012; Manzano-Piedras et al., 2014; Gómez et al., 2018) rather than a weed able to colonise humanised environments from croplands to urban spots. Given the tight association of relict *A. thaliana* with natural environments and its genomic differentiation, we hypothesise that the relict lineage of *A. thaliana* in the Iberian Peninsula and Morocco has experienced its own evolutionary trajectory through long-term survival and adaptation to multiple environments across the region. The main reason we suggest this hypothesis has to do with the occurrence of Iberian and Moroccan relict *A. thaliana* in well-known glacial refugia for animals and plants in southern Europe and northern Africa (Cowling et al., 1996; Comes and Kadereit, 1998; Hewitt, 1999; Gomez and Lund, 2006; Hughes et al., 2006; Médail and Diadema, 2009). The region turned into a major biodiversity hotspot of the Mediterranean Basin (e.g. the Baetic–Rifan hotspot; Médail and Quézel, 1997), where several plant families have experienced intense local diversification and speciation (Rodríguez-Sánchez et al., 2010; Lavergne et al., 2013; Molina-Venegas et al., 2017) and co-existing with old, pre-Mediterranean plant genera that still remain in the area (Herrera, 1992).

Here, we sought for evidences from the past to disentangle the ecological history of relict *A. thaliana* across its major distribution range in southwest Mediterranean Basin. Given that small annual plants barely leave any trace in the fossil record, we addressed this issue in a different manner. First of all, we modelled the current potential distribution range of the relict

lineage of *A. thaliana* with species distribution models (SDM) combining climate, remote sensing and soil data and using all known populations identified as relict so far in the Iberian Peninsula and Morocco. We used the resulting distribution range map to explore the association between current habitat suitability of relict *A. thaliana* and vegetation changes since the Last Glacial Maximum (LGM). To do that, we estimated species turnover and changes in arboreal vegetation from fossil pollen data as well as reconstructions of potential distribution ranges of the most important tree species occurring across the region over broad climatic periods since the LGM. Our hypothesis was that relict *A. thaliana* has long occurred in the Iberian Peninsula and Morocco. Given the current tight association between the occurrence of relict *A. thaliana* in natural environments and the inability of the species to behave like a weed (Marcer et al., 2016), we expected to find that current relict *A. thaliana* was more related to environments that experienced more stability over time. Taking advantage of the vast genomic resources available in *A. thaliana*, we also searched for genes with putative functional polymorphisms characteristic of the relict lineages and with a low frequency in non-relict lineages, which would also represent additional evidence of the long evolutionary history of relict *A. thaliana* (Tsuchimatsu et al., 2017; Durvasula et al., 2017). We discussed the results emphasising how relict *A. thaliana* endured past climatic changes and stressing the fact that relict *A. thaliana* is a lineage that continues its own evolutionary trajectory.

Materials and Methods

Relict source populations

We used our Iberian and Moroccan collection of relict *A. thaliana* populations (39 Iberian and 20 Moroccan accessions; Fig. 1, 2 and Supplementary Table S1) previously identified with sets of genome-wide neutral single nucleotide polymorphisms (Picó et al., 2008; Brennan et

al., 2014; Marcer et al., 2016) and later confirmed with genome-scale sequencing (The 1001 Genomes Consortium, 2016, Durvasula et al., 2017, Fulgione et al., 2018, Tabas-Madrid et al., 2018). Although the southwest Mediterranean Basin is the area with the most important distribution of relict *A. thaliana*, we also compiled the coordinates of all *A. thaliana* populations categorised as relicts from published articles conducting large-scale genome sequencing of *A. thaliana* accessions (Fig. 1 and Supplementary Table S1) in order to illustrate the geographic extension occupied by relict *A. thaliana*. Thus, there were nine additional populations from Africa (including four Moroccan accessions, four Tanzanian, four South African and one Algerian; Durvasula et al., 2017), five from Madeira (including 11 accessions; Fulgione et al., 2018), two from Italy (The 1001 Genomes Consortium, 2016, Durvasula et al., 2017), one from China (including two accessions; Hsu et al., 2019), one from Lebanon (Durvasula et al., 2017), one from Tibet (Zeng et al., 2017), one from the Canary Islands (Hsu et al., 2019), and one from Cape Verde (Hsu et al., 2019).

Given the scarcity of relict populations located out of the western Mediterranean distribution, final analyses were based on 63 southwest Mediterranean populations, including 39 from the Iberian Peninsula, 23 from Morocco and one from Algeria (Fig. 2 and Supplementary Table S1). There is some confusion in the literature around the exact position of the Algerian population included in genome-scale sequencing studies. To deal with this shortcoming, we used the approximate geographical position of a population found in Algeria decades ago by Quézel and Barbero (1989).

Species distribution models

We modelled the current potential distribution range of relict *A. thaliana* to identify the environmental and ecological requirements of this genetic lineage in southwest Mediterranean Basin. The study area to model the potential distribution range of relict *A. thaliana* was

defined by the known distribution of relict populations across the region accumulated in over 15 years of field campaigns and molecular work. Thus, the study area comprised the whole Iberian Peninsula, bordered by the Pyrenees to the north and by the central and western Maghreb to the south, including vast regions of Morocco, Algeria and Tunisia. Nevertheless, to avoid arbitrary decisions to delimit the study area in North Africa, somewhere between the Tell Atlas in Algeria, the Atlas mountains in Morocco and the Sahara desert, its southern limit was defined by the area affected by the so-called greening of the Sahara during the Mid-Holocene, 6 kya (years before present), as reported by paleoclimatological reconstructions indicating the occurrence in North Africa of grassland and scrubland vegetation in wetter environmental conditions (de Noblet-Ducoudré et al., 2000, Compton, 2011). In our opinion, this represents a reasonable distinction between potentially suitable habitat for *A. thaliana* and the inhospitable Sahara desert in an appropriate time frame for this study. Overall, the study area ranged between 28.0° N – 43.8° N and 11.2° E – 11.7° W, totalling 1,389,229 km², of which 583,123 km² corresponded to the Iberian Peninsula and 806,106 km² to North Africa (Fig. 2).

We selected a set of climate, habitat, soil and ecophysiological predictors from different publicly available data sources at a resolution of 1 km². It must be noted that we checked for co-linearity among all predictors to be below |0.8|, as measured by the Pearson's correlation coefficient and using the whole set of data fed to the distribution model (i.e. background points plus relict occurrences). First of all, a total of six of 19 bioclimatic variables were selected from WorldClim v.2 (Fick and Hijmans, 2017): BIO1 (annual mean temperature), BIO3 (isothermality), BIO4 (temperature seasonality), BIO12 (annual precipitation), BIO15 (precipitation seasonality) and BIO18 (precipitation of the warmest quarter).

Vegetation data were represented by two variables: the percentage of crops and the percentage of non-arboreal vegetation. The percentage of crops referred to agricultural lands that reduce the occurrence of relict *A. thaliana* (Marcer et al., 2016). The percentage of crops was obtained from MODIS/Terra and Aqua Combined Land Cover (MCD12Q1) v.6 data product (Friedl and Sulla-Menashe, 2015) combining croplands and croplands/natural vegetation mosaics using the global land cover classes of the International Geosphere-Biosphere Programme (IGBP). We obtained a layer of percentage of crops by assigning the most frequent value for each grid cell (1 km²) for the period 2003 – 2016.

The percentage of non-arboreal vegetation was also obtained following the same procedure but using the MODIS/Terra Vegetation Continuous Fields (MOD44B) v.6 data product (Dimiceli et al., 2015). The percentage of non-arboreal vegetation was chosen because of the affinity of *A. thaliana* for forest openings and scrublands (Montesinos et al., 2009; Gómez et al., 2018). In this case, however, we used the mean value for each grid cell, instead of the most frequent value, to generate the single layer for this variable between 2003 and 2016, as this is a continuous variable.

Next, we obtained the Leaf Area Index (LAI), an ecophysiological trait, from the MODIS/Terra Leaf Area Index (MOD15A2H) v.6 data product (Myneni et al., 2015). Similarly, we used mean values for each grid cell for the period 2003 – 2016. LAI is broadly defined as the amount of leaf area (m²) in a canopy per unit ground area (m²). For this reason, it is an important structural parameter of the vegetation as it is directly and proportionally related to major parameters, such as photosynthetic primary production and evapotranspiration. LAI was highly correlated with other important ecophysiological variables, e.g. Net Photosynthesis, Gross Primary Production or Enhanced Vegetation Index ($r > 0.85$ in all cases), and for this reason we only used LAI to represent the ecophysiological traits in our set of relict *A. thaliana* populations. When generating our layers from MODIS,

we only considered those pixels labelled as highest quality in their respective quality control layers. Only 0.56 % of pixels were excluded for this reason.

Finally, a single topsoil pH layer was generated by merging data provided by the Geographical Database from Eurasia v.4 (<https://esdac.jrc.ec.europa.eu>) and the Harmonized World Soil Database v.1.2 (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2012). We included this variable because it is an important factor accounting for the presence of *A. thaliana* in the Iberian Peninsula, particularly making *A. thaliana* increasingly scarce as topsoil pH becomes more basic (Marcer et. al., 2016).

We used the software Maxent v.3.3.3k (Phillips et al., 2006, Elith et al., 2011) to model the potential distribution of the genetic lineage of *A. thaliana* in the Iberian Peninsula and the Maghreb. The model was fitted using the default parameters given by Maxent. We applied the 10-fold cross-validation area-under-the-curve (AUC) metric to evaluate prediction performance. Finally, we used all occurrence points to fit the final model and project it to the whole study area. All data preparation and modelling was carried out using the R computing environment (R Core Team, 2017) using several packages (i.e. broom, corrgram, dismo, dplyr, ggplot2, jsonlite, magrittr, ncf, raster, rasterVis, readr, readxl, rgdal, rJava, sp, spdep, stringr, and tidyr) and the Geospatial Data Abstraction Library (GDAL/OGR contributors, 2018). Given that the Iberian Peninsula and the Maghreb are two separate sub-regions, we tested for differences between regions for all potential distribution range predictors with one-way ANOVA with SPSS v.25 statistical software (IBM, Chicago, IL, USA).

Fossil pollen data

As *A. thaliana* does not appear in the fossil pollen record, we used indirect means to indentify the historical factors related to past vegetation dynamics accounting for the current potential distribution range of relict *A. thaliana* across the study region. To this end, we quantified the

vegetation changes occurred over several millennia from paleobotanical records from all over the Iberian Peninsula and the Maghreb. Overall, we gathered data from 80 paleobotanical sites reporting fossil pollen data (Fig. 2 and Supplementary Table S2). We obtained the database of paleobotanical records from two sources. First, we used pollen records publicly available in the European Pollen Database (EPD), resulting in 50 sites from Spain, five from Tunisia, three from Algeria, three from Portugal and two from Morocco. Second, we conducted a bibliographic search on recent published articles reporting fossil pollen data from the region that were not available in the EPD. We contacted the corresponding authors and requested for published raw data in the same format than that posted on the EPD (i.e. pollen counts per species, sediment sample depth and calibrated age). Eventually, we added to our database eight sites from Spain (Gil-Romera et al., 2008, Burjachs et al., 2016; González-Sampériz et al., 2017; Jones et al., 2018; López-Sáez et al., 2018), four from Morocco (Campbell et al., 2017; Cheddadi et al., 2017), three from Andorra (Beaulieu et al., 2005; Miras et al., 2010, 2015) and two from Portugal (Santos and Fernández-Goñi, 2003; Fletcher et al., 2007).

Overall, the 80 paleobotanical sites were scattered across the Iberian Peninsula and the Maghreb, from sea side to montane environments (mean altitude $\pm$ SE = 884.0 ± 72.1 m.a.s.l., range = 0 – 2,406 m) and intermingled with relict *A. thaliana* populations (Fig. 2; mean $\pm$ SE distance between relict populations and paleobotanical sites = 652.6 ± 390.7 km; range = 4.0 – 1,686.7 km). Important for our study, data from paleobotanical sites included a calibrated chronology mostly expressed in calendar years before 1950 (cal yr BP). Our dataset contains sediment cores that were calibrated by different methods and age-depth models. However, the assessment of the value of these methodologies is clearly beyond our reach. The length of the age-depth series were highly variable among paleobotanical sites, ranging from a low of 228

cal yr BP to a high of 44,370 cal yr BP (mean $\pm$ SE = 10,061.2 $\pm$ 986.3 cal yr BP;
Supplementary Table S2).

Given the heterogeneity of paleobotanical data, in terms of depth, calibrated age, continuity of the chronology and number of paleobotanical sites, we divided the chronologies in three broad climatic periods (Blockley et al., 2012; González-Sampériz et al., 2017): present (up to 2 kya), Holocene (11.7 – 2 kya) and a final period combining the Younger Dryas and the Allerød (14.0 – 11.7 kya). Each one of these periods is characterised by extensive climatic variability. The present includes the Medieval Climatic Anomaly (900–1300 AD), the Little Ice Age (1300–1850 AD) and the increasing warming since the Industrial Revolution. The Holocene (11.7 – 2 kya) is typically divided into various periods during which the vegetation experienced several important changes across the study region by virtue of climatic variation and human activities that modified the environment (e.g. fire and agriculture). Finally, the last period, including the Younger Dryas and the Allerød (14.0 – 11.7 kya), combined two distinctive periods, such as the Younger Dryas (12.9 – 11.7 kya) characterised by cold glacial conditions, and the Allerød (14.0 – 12.9 kya) characterised by warmer and moister environmental conditions. Beyond 14 kya, the number of paleobotanical sites was too low to attempt any solid statistical analysis.

For each climatic period and paleobotanical site, we estimated the species turnover using the metric developed by MacArthur and Wilson (1963) and modified by Diamond (1969). Species turnover was calculated as the proportion of species that differs between two time points (i.e. the sum of species gained and lost divided by the total species observed in both time periods). To do that, we used time windows of 100 years within each climatic period from which we computed mean and SD species turnover. We tested narrower and wider time windows and results were quite consistent, although 100-year time windows captured better the high heterogeneity of fossil pollen data. Across paleobotanical sites, the

mean ($\pm$ SD) number of 100-year time windows was 13.1 ± 5.9 (range = 1 – 23) for the present, 30.5 ± 22.1 (range = 1 – 93) for the Holocene, and 7.4 ± 5.0 (range = 1 – 17) for the Younger Dryas and the Allerød.

For each climatic period, we also estimated the mean and SD of the percentage of arboreal pollen with and without pollen from the genus *Pinus*, which can be quite dominant across the region, using the same 100-year time windows. The percentage of arboreal pollen was estimated including pollen data from trees and shrubs. The drop or increase of arboreal pollen over time provides hints on the temporal change in vegetation structure and dynamics. We did not consider species abundance as the interpretation of this parameter is always controversial requiring mathematical approaches to reconstruct pollen diversity at each paleobotanical site (Weng et al., 2006).

For each climatic period, we calculated the correlation between current habitat suitability of relict *A. thaliana* estimated with Maxent (mean and SD) and (i) the percentage of arboreal pollen with and without pollen from the genus *Pinus* (mean and SD across 100-year time windows), and (ii) species turnover (mean and SD across 100-year time windows). Habitat suitability was estimated for a circular area (5 km radius) around the geographic position of each paleobotanical site. We sought to figure out whether the habitat suitability of relict *A. thaliana* in present times was related to the vegetation dynamics registered in paleobotanical sites at a local scale for the three broad climatic periods. We repeated the analyses with other radii (10, 20 and 50 km around each paleobotanical site) and results were quite consistent although larger areas obviously tended to blur the patterns. Given the spatially-explicit nature of variables, we conducted Pearson's correlation with and without correcting for the spatial autocorrelation of variables (Dutilleul's modified *t*-test; Dutilleul et al., 1993) using SAM software v.4.0 (Rangel et al., 2010). Spatial autocorrelation may result in several statistical flaws, such as violation of the assumption of independent error or bias in the

estimation of correlation coefficients and variances (Dutilleul et al., 1993). The comparison of results with and without the correction for spatial autocorrelation allows the assessment of the extent of the problem caused by spatial autocorrelation of data.

Distribution maps of dominant Iberian tree species over time

We explored the relationship between the potential distribution range of the most important tree species with that of relict *A. thaliana*. Tree species can be considered as representative taxa of their plant communities. Thus, we used them to pinpoint those plant communities more or less related to the current potential occurrence of relict *A. thaliana*. To this end, we compared the current potential distribution range of relict *A. thaliana* with the current potential distribution range of the most important tree species occurring in the Iberian Peninsula using data published elsewhere (Benito-Garzón et al., 2007). The rasterized maps (1 km² grid format) were kindly provided by Marta Benito-Garzón including the present potential distributions and two past potential distributions at two particular points in time: the Last Glacial Maximum (LGM; 21 kya) and the Mid-Holocene (6 kya). Past potential distributions were used to infer the dynamics of such plant communities over the last millennia. We excluded the Maghreb from these analyses as we did not have similar data from that region.

In short, the present distributions of 17 tree species were determined from occurrences reported in the Spanish Forest Map and the Portuguese Inventário Florestal Nacional. Bioclimatic envelope models were used to estimate potential distributions using (i) Agroclimatic Characterization of Spain's Provinces and the Portuguese AGRIBASE database for present distributions and (ii), the Palaeoclimate Modelling Intercomparison Project for the LGM and Mid-Holocene distributions (see Benito-Garzón et al., 2007). Overall, the 17 trees considered in this study included six Mediterranean oaks and pines (*Quercus suber*, *Q. ilex*

subsp. *ilex*, *Q. ilex* subsp. *ballota*, *Pinus pinea*, *P. pinaster* and *P. halepensis*), four sub-Mediterranean trees (*Juniperus thurifera*, *P. nigra* subsp. *salzmanii*, *Q. faginea* subsp. *faginea* and *Q. pyrenaica*), four temperate broadleaved trees (*Quercus robur*, *Q. petraea*, *Castanea sativa* and *Fagus sylvatica*), and three montane conifers (*Pinus sylvestris*, *P. uncinata* and *Abies alba*).

First of all, we calculated the current niche overlap between the potential distribution of relict *A. thaliana* and each tree species in the present time using the Schoener's *D* statistic for niche overlap (Schoener, 1968; Warren et al., 2008). For each tree species, we also calculate the Schoener's *D* for different combinations of climatic periods, such as between present and LGM, between present and Mid-Holocene, and between LGM and Mid-Holocene, which was used to infer the stability of plant communities represented by each tree species over time. The *D* metric has been widely used for its simplicity and interpretability as it ranges from 0 (niche models have no overlap) to 1 (niche models are identical). In our case, the *D* metric considers the relationship between probabilities of relict *A. thaliana* and each tree species, and between probabilities of each tree species across climatic periods. We used the *dismo* R package (Hijmans et al., 2017) to compute Schoener's *D* values. We analysed the relationship of niche overlaps between relict *A. thaliana* and each tree species with the stability of tree distribution ranges over climatic periods by performing Pearson's correlations with SPSS v.25 statistical software (IBM, Chicago, IL, USA).

Genetics of flowering time in relict *A. thaliana*

Given the large genetic differentiation between relict and non-relict genomes, it can be expected that genetic studies of relict accessions will provide natural functional polymorphisms that otherwise cannot be found. To illustrate this expectation, we described two examples of putative functional polymorphisms that are segregating mainly within the

relict group, but showing very low frequency in non-relict accessions. We focused on the two flowering time genes, *CRY2* (locus: At1g04400) and *TSF* (locus: At4g20370), because they are the main genes in which potential functional polymorphisms were previously studied specifically in relict accessions (El-Assal et al., 2001; Lee et al., 2017; Tabas-Madrid et al., 2018). Genomic sequences were obtained for 1,216 genome sequences available from world-wide accessions (The 1001 Genomes Consortium, 2016; Durvasula et al. 2017). Single nucleotide polymorphisms (SNPs) were extracted from these sequences and filtered to remove singletons, providing 142 and 110 SNPs for *CRY2* and *TSF*, respectively. The genetic relationships among relict and non-relict accessions were determined from non-singletons SNPs by neighbor-joining (NJ) analysis using MEGA v.7 (Tamura et al., 2011) and applying 1,000 bootstrap permutations to determine statistical significances.

Results

Potential distribution range of relict *A. thaliana*

We used climate, habitat, soil and ecophysiological data to model the current potential distribution range of the relict lineage of *A. thaliana* in the Iberian Peninsula and the Maghreb with Maxent (Fig. 3). The resulting model exhibited a good performance based on the AUC test values ($AUC \pm SD = 0.926 \pm 0.033$). The percent contribution of environmental variables to the fit of the model, given by their percent permutation importance (Table 1), was clearly dominated by the Leaf Area Index (LAI; percent contribution = 53.5 %) and annual mean temperature (BIO1; 10.5 %). The rest of variables included in the model had lower percent contributions (range = 1.4 – 7.3 %; Table 1). The LAI values obtained for relict *A. thaliana* populations (range = 0.24 – 2.00 m² m⁻²) were in the range of those characterising non-arboreal vegetation, such as sub-desertic areas, grasslands and scrublands (Asner et al., 2003). Annual mean temperature ranged between 4.64 and 17.98 °C, which is not surprising given

the range of altitudes at which relict *A. thaliana* can be found in SW Mediterranean Basin. Although their percent contribution to the fit of the model was low, the results indicated that relict *A. thaliana* prefers non-arboreal vegetation (58.0 % of non-arboreal vegetation across populations; Table 1) and avoids croplands (0.1 % of croplands across populations; Table 1).

Populations from the Iberian Peninsula and the Maghreb did not vary for the two most important variables accounting for the potential distribution range of relict *A. thaliana*, i.e. LAI and BIO1 (P -value > 0.13 in both cases; Table 1). In contrast, they differed for precipitation of the warmest quarter (BIO18; $F_{1,62} = 27.99$, P -value < 0.001), precipitation seasonality (BIO15; $F_{1,62} = 10.94$, P -value $= 0.002$) and topsoil pH ($F_{1,62} = 7.36$, P -value $= 0.009$). On average, Iberian relict populations had more precipitation in the warmest quarters, lower precipitation seasonality and more acidic soils than the Maghreb relict populations (Table 1).

Vegetation changes from fossil pollen data

We correlated the current habitat suitability of relict *A. thaliana* across the study region with variables accounting for vegetation changes over different climatic periods. In the present (up to 2 kya), none of the vegetation change variables estimated (e.g. mean and SD percentage of arboreal pollen with and without the genus *Pinus* and species turnover) significantly correlated with mean or SD of habitat suitability of relict *A. thaliana* around the paleobotanical sites ($N = 68$; P -values > 0.130 and > 0.081 in all cases with and without the correction for spatial autocorrelation of data, respectively). In contrast, we found significant relationships during the Holocene (11.7 – 2 kya). In particular, mean habitat suitability of relict *A. thaliana* was significantly negatively correlated with the SD of the percentage of arboreal pollen including and excluding the genus *Pinus* (Fig. 4A and 4B). We also detected significant relationships deeper in the past with the combined climatic period of the Younger

Dryas and the Allerød (14.0 – 11.7 kya). In this case, the SD of species turnover was significantly negatively correlated with mean and SD of habitat suitability of relict *A. thaliana* (Fig. 4C and 4D). Except in one case (Fig. 4A), the significance of the relationship held up with and without the correction for spatial autocorrelation of data, stressing the robustness of the patterns found. Overall, our results indicated that sites with lower temporal variability, as expressed by SD, in the vegetation over different climatic periods are related with higher occurrence probabilities of relict *A. thaliana* today.

Niche overlap with tree species

The niche overlap between the current potential distribution range of relict *A. thaliana* and the current potential distribution range of 17 tree species in the Iberian Peninsula (Supplementary Figure S1) varied widely. Overall, the mean $\pm$ SD of Schoener's *D* values between relict *A. thaliana* and the 17 tree species was 0.26 ± 0.13 . The tree species with the higher niche overlap with relict *A. thaliana* were *Quercus ilex* subsp. *ballota* ($D = 0.499$) and *Q. pyrenaica* ($D = 0.455$), whilst the tree species with the lower niche overlap were *Quercus ilex* subsp. *ilex* (Schoener's $D = 0.107$) and *Pinus uncinata* ($D = 0.052$) (see all values in Supplementary Table S3). Considering all 17 niche overlaps between relict *A. thaliana* and tree species in the present, we found that tree species with higher stability in potential distribution ranges since the LGM tended to have a higher niche overlap with the potential distribution range of relict *A. thaliana* in the present (Fig. 5A). The same result was found for stability between the LGM and the Mid-Holocene (Fig. 5B) but not for stability since the Mid-Holocene ($N = 17$, $r = -0.32$, $P = 0.21$). These results suggested how a higher stability over time of major plant communities accounts for the presence today of relict *A. thaliana* in the Iberian Peninsula.

Genetics of flowering time in relict *A. thaliana*

We analysed genomic sequences of *A. thaliana* to quantify functional genetic variation characterising relict accessions in two flowering time genes: the blue light photoreceptor *CRY2* and the florigen encoding gene *TSF*. On the one hand, we found two major *CRY2* haplogroups (haplogroup A and B) that were highly differentiated by eight missense mutations previously found to be associated with flowering time variation (Olsen et al., 2004). In addition, the relict accession Cvi, originally collected in Cape Verde (Lobin et al., 1983), was demonstrated to carry a specific early flowering allele of *CRY2* caused by the Met/Val367 missense mutation (El-Assal et al., 2001). Analysis of *CRY2* sequences in currently available genomes showed that Cvi allele is absent in the rest of relict and non-relict genomes. However, this analysis identified a new haplogroup (haplogroup C; Brennan et al., 2014), which was found at high frequency in Moroccan relicts (40%) as well as in the Algerian and Lebanon relicts (Fig. 6A and 6B). This haplogroup was differentiated from the rest of accessions by three amino acids, and it appears at very low frequency in non-relict accessions from the rest of world (<1%).

On the other hand, analysis of *TSF* sequences from the available genomes showed that most non-relict accessions from all over the world (97%) carried the major haplogroup (haplogroup A; Fig. 6C and 6D). By contrast, 54% of the relict accessions had the so-called haplogroup B differentiated by multiple non-coding polymorphisms, which might affect *TSF* regulation. Although the functional effect of the missense and non-coding mutations differentiating *CRY2*-C and *TSF*-B need to be demonstrated, these haplogroups highlighted the potential usefulness of relict accessions to unveil novel functional genetic diversity.

Discussion

The collaborative efforts of a large and multidisciplinary research community merging extensive field sampling and whole-genome deep sequencing represent an outstanding asset

to unravel the evolutionary history of the annual plant *Arabidopsis thaliana*. The uninterrupted work of the community for decades has recently paid off by providing new insights into the early history of *A. thaliana* indicating the existence of relict lineages with a wide geographic distribution spanning up to three continents (The 1001 Genomes Consortium, 2016; Durvasula et al., 2017; Tsuchimatsu et al., 2017; Zeng et al., 2017; Fulgione et al., 2018; Hsu et al., 2019). However, the ecological history of relict *A. thaliana* is of paramount importance to complete the picture provided by whole-genome sequencing. This study attempted to reconstruct the ecological history of relict *A. thaliana* by seeking for indirect evidences of the recent past accounting for the occurrence of relict *A. thaliana* in southwest Mediterranean Basin, the most important nuclei of known relict populations. Our hypothesis was that relict *A. thaliana* has long survived and adapted across the region, as inferred for other annual plant species in the Iberian Peninsula (Fernández-Mazuecos and Vargas, 2013; Bobo-Pinilla et al., 2018), probably in areas that experienced more stability over time. In fact, it is quite remarkable the occurrence of some relict *A. thaliana* populations growing in or very close to populations of well-known relict plant species that long endured confined to mountain refugia in Spain and northern Africa (Carrión et al., 2010), such as relict forests of silver birch (*Betula alba*) in central Spain, relict firs (*Abies pinsapo*, *A. marocana* and *A. tazaotana*) in southwest Spain and the Rif, the Sictus tree (*Tetraclinis articulata*) in southeast Spain, or the Atlas cedar (*Cedrus atlantica*) in the Atlas Mountains. Nevertheless, such coincidences, although stimulating and inspiring for this work, required of additional support to become more than mere observations.

To test our hypothesis, we first modelled the current potential distribution range of relict *A. thaliana* in Morocco and the Iberian Peninsula, which can also be interpreted in terms of current habitat suitability across the region. Our results were consistent with a previous model for relict *A. thaliana* in the Iberian Peninsula (Marcer et al., 2016), even after the

inclusion of some new elements, such as the extension of the study area to North Africa and the use of new vegetation parameters from remote sensing data. Overall, we confirm a known ecological attribute of relict *A. thaliana*, which is its preference for natural habitats and probably its difficulty to colonise humanised environments (Marcer et al., 2016; The 1001 Genomes Consortium, 2016). As a matter of fact, recent genomic analyses gave rise to the hypothesis that admixture of locally adapted relict *A. thaliana* with non-relict lineages helped the latter to colonise multiple natural and humanised habitats through rapid demographic expansion until dominating most of the species' current distribution range (The 1001 Genomes Consortium, 2016; Lee et al., 2017; Hsu et al., 2019), which is what we observe today.

Next, we used the current habitat suitability of relict *A. thaliana* in Morocco and the Iberian Peninsula, which showed small differences in some minor predictors of habitat suitability but not in the most relevant ones (Table 1), to explore its relationship with estimates of past vegetation dynamics over the last millennia. The analysis of fossil pollen data from sediment cores located across the study region indicated that paleobotanical sites with higher current habitat suitability for relict *A. thaliana* exhibited lower temporal variation in vegetation dynamics for different broad climatic periods (Fig. 4), as measured by species turnover and by the proportion of arboreal pollen. Despite the complexity of fossil pollen data, even in the interpretation of vegetation dynamics over millennia in the region (Carrión et al., 2010), the existence of such relationships suggests that relict *A. thaliana* required of environments characterised by higher stability over time to make it through the present times. This conclusion lines up with the preference of known relict *A. thaliana* populations for natural environments in which they have long survived. It is worth emphasising that a previous analysis comparing historical climatic conditions between relict and non-relict *A. thaliana* populations suggested that temperature changed less for relict than for non-relict *A.*

thaliana since the LGM (The 1001 Genomes Consortium, 2016). Although such analysis was carried out by comparing climatic stability of only 26 relict versus 1,109 non-relict accessions worldwide, without taking the spatial autocorrelation of climatic data into account, and assuming that all populations (relict and non-relict) remained in the same location since the LGM, which is a huge assumption for an annual plant, our results might provide some support to that suggestion if we assume that vegetation changes have chiefly been mediated by variation in temperature.

On top of the indirect evidence reported in this study on the relationship between current habitat suitability of relict *A. thaliana* and vegetation dynamics over the last millennia, we provided additional support to our original hypothesis by exploring the niche overlap between relict *A. thaliana* and some tree species in the Iberian Peninsula that represent whole plant communities. Our results allowed the identification of those plant communities with a higher (e.g. *Quercus ilex* subsp. *ballota* and *Q. pyrenaica* widely distributed in central Iberian Peninsula) and lower (*Quercus ilex* subsp. *ilex* and *Pinus uncinata* restricted to northeast Iberian Peninsula) affinity with relict *A. thaliana* in the present times. Interestingly, niche overlap between relict *A. thaliana* and tree species was significantly positively correlated with tree niche stability since the LGM (Fig. 5), suggesting that relict *A. thaliana* in the present times is tightly associated to plant communities whose distribution experienced less variation over millennia. This result is pointing in the same direction than the previous one and reinforces the view that the persistence of relict *A. thaliana* through time depended to a large extent on the stability of the plant communities where it occurred. It must be emphasised that the dynamics of non-relict *A. thaliana* is thought to be completely different, as it is hypothesised that non-relict lineages have experienced a demographic expansion colonising multiple environments (Durvasula et al., 2017; Fulgione and Hancock, 2018; Lee et al., 2017; Hsu et al., 2019). Given the wide array of environments where relict *A. thaliana* can be found

in Morocco and the Iberian Peninsula in the present times, which is also supported by the important niche overlaps detected in our study between relict *A. thaliana* and several tree species (Supplementary Figure S1), it is plausible to suggest that the fate of relict populations was tightly determined by the fate of plant communities in which the plant was able to occur. This highlights the long evolutionary history of relict *A. thaliana*, which would have had time to adapt to very contrasting environments since the split of the species into relict and non-relict lineages.

Finally, our survey of *A. thaliana* genome sequences available from world-wide accessions yielded additional support to the strong genomic differentiation between relict and non-relict *A. thaliana* at the functional level. This is important because relict *A. thaliana* is expected to harbour specific genetic variation resulting from its long evolutionary history (Tsuchimatsu et al., 2017; Durvasula et al., 2017). In particular, we identified highly differentiated haplogroups of *CRY2* and *TSF* genes, which are present at high frequency among relict accessions but nearly absent in non-relict genomes (Fig. 6). As shown with the two genes analysed here, it can be expected that genetic studies of relict accessions will provide natural functional polymorphisms that otherwise cannot be detected. Thus, we stress the need to step up our efforts to sample in depth the rest of African, Middle East and Asian regions where relict accessions have been detected to increase sample size in these regions and to further investigate the genetic differentiation of relict *A. thaliana* from a functional perspective. This will provide hints on the genetic mechanisms underlying the long evolutionary history of relict *A. thaliana*. In fact, the evolutionary history of relict *A. thaliana* still hides many fascinating secrets that remain to be discovered, such as the genetic relationships among relicts from distant geographic locations, the extent of admixture between relict and non-relict accessions and its evolutionary consequences, and the amount and structure of functional genetic variation that makes relict *A. thaliana* so particular.

594

595 **Data**

596 Data available from the Dryad Digital Repository: XXX.

597

598 **Sources of Funding**

599 FXP and AM were funded by grant CGL2016-77720-P (AEI/FEDER, UE). BT acknowledges
600 the contract BES-2017-081803 (AEI/FEDER, UE). AM acknowledges the Agency for
601 Management of University and Research Grants of the Generalitat de Catalunya (2014-SGR-
602 913).

603

604 **Contributions by the Authors**

605 FXP conceived and planned the research. BT, AM, BMV, CAB and FXP acquired and
606 analysed the data. FXP wrote the first draft of the manuscript and all authors contributed to
607 the final version of it.

608

609 **Conflicts of Interest**

610 None declared.

611

612 **Acknowledgements**

613 We thank the following researchers for sharing their published fossil pollen data with us (by
614 alphabetical order): Michelle Barbier-Leydet, Quico Burjachs, Rachid Cheddadi, William
615 Fletcher, Graciela Gil-Romera, Penélope González-Sampériz, Samantha Jones, José Antonio
616 López-Sáez, Yannick Miras and Luísa Santos-Fidalgo. We are particularly grateful to
617 Penélope González-Sampériz for her help to interpret pollen fossil data. We also thank María
618 Fernanda Sánchez-Goñi, Yolanda Carrión, Albert Pélachs, Donatella Magri, Paloma

Uzquiano-Ollero, Nathalie Nebout, Castor Muñoz-Sobrino, and Mercedes García for their willingness to share the data although we did not use it eventually. We are indebted to Marta Benito-Garzón for preparing all distribution maps of the 17 tress species used in this study. We are grateful to María Begoña García, Juan Arroyo and Johan Ehrlén for the invitation to participate in this special issue.

Literature Cited

- Asner, G.P., Scurlock, J.M.O., Hicke, J.A., 2003. Global synthesis of leaf area index observations: implications for ecological and remote sensing studies. *Global Ecol. Biogeogr.* 12, 191-205.
- Benito-Garzón, M., Sánchez de Dios, R., Sáinz-Ollero, H., 2007. Predictive modelling of tree species distributions on the Iberian Peninsula during the Last Glacial Maximum and Mid-Holocene. *Ecography* 30, 120-134.
- Blockley, S.P.E., Lane, C.S., Hardiman, M., Rasmussen, S.O., Seierstad, I.K., Steffensen, J.P., Svensson, A., Lotter, A.F., Turney, C.S.M., Ramsey, C.B., INTIMATE members, 2012. Synchronisation of palaeoenvironmental records over the last 60,000 years, and an extended INTIMATE event stratigraphy to 48,000 b2k. *Quaternary Sci. Rev.* 36, 2-10.
- Bobo-Pinilla, J., Peñas de Giles, J., López-González, N., Mediavilla, S., Martínez-Ortega, M.M., 2018. Phylogeography of an endangered disjunct herb: long-distance dispersal, refugia and colonization routes. *AoB PLANTS* 10, ply047.
- Brennan, A.C., Méndez-Vigo, B., Haddioui, A., Martínez-Zapater, J.M., Picó, F.X., Alonso-Blanco, C., 2014. 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.* 14, 17.

643 Burjachs, F., Jones, S.E., Giralt, S., Fernández-López de Pablo, J., 2016. Lateglacial to Early
 644 Holocene recursive aridity events in the SE Mediterranean Iberian Peninsula: The Salines
 645 Campbell, J.F.E., Fletcher, W.J., Joannin, S., Hughes, P.D., Rhanem, M., Zielhofer, C., 2017.
 646 Environmental drivers of Holocene forest development in the Middle Atlas, Morocco.
 647 Front. Ecol. Evol. 5, 113.
 648 Carrión, J.S., Fernández, S., González-Sampériz, P., Gil-Romera, G., Badal, E., Carrión-
 649 Marco, Y., López-Merino, L., López-Sáez, J.A., Fierro, E., Burjachs, F., 2010. Expected
 650 trends and surprises in the Lateglacial and Holocene vegetation history of the Iberian
 651 Peninsula and Balearic Islands. Rev. Palaeobot. Palyno. 162, 458-475.
 652 Cheddadi, R., Henrot, A.J., François, L., Boyer, F., Bush, M., Carré, M., Coissac, E., De
 653 Oliveira, P.E., Ficetola, F., Hambuckers, A., Huang, K., Lézine, A.M., Nourelbait, M.,
 654 Rhoujjati, A., Taberlet, P., Sarmiento, F., Abel-Schaad, D., Alba-Sánchez, F., Zheng, Z.,
 655 2017. Microrefugia, climate change, and conservation of *Cedrus atlantica* in the Rif
 656 mountains, Morocco. Front. Ecol. Evol. 5, 114.
 657 Comes, H.P., Kadereit, J.W., 1998. The effect of Quaternary climatic changes on plant
 658 distribution and evolution. Trends Plant Sci. 3, 432-438.
 659 Compton, J.S., 2011. Pleistocene sea-level fluctuations and human evolution on the southern
 660 coastal plain of South Africa. Quaternary Sci. Rev. 30, 506-527.
 661 Cowling, R.M., Rundel, P.W., Lamont, B.B., Arroyo, M.K., Arianoutsou M., 1996. Plant
 662 diversity in Mediterranean-climate regions. Trends Ecol. Evol. 11, 362-366.
 663 De Beaulieu, J.L., Miras, Y., Andrieu-Ponel, V., Guiter, F., 2005. Vegetation dynamics in
 664 north-western Mediterranean regions: Instability of the Mediterranean bioclimate. Plant
 665 Biosyst. 139, 114-126.

666 de Montaigu, A., Coupland, G., 2017. The timing of *GIGANTEA* expression during day/night
 667 cycles varies with the geographical origin of *Arabidopsis* accessions. *Plant Signal Behav.*
 668 12, e1342026.

669 de Noblet-Ducoudré, N., Claussen, M., Prentice, C., 2000. Mid-Holocene greening of the
 670 Sahara: first results of the GAIM 6000 year BP experiment with two asynchronously
 671 coupled atmosphere/biome models. *Clim. Dynam.* 16, 643-659.

672 Diamond, J.M., 1969. Avifaunal equilibria and species turnover rates on the Channel Islands
 673 of California. *P. Natl. Acad. Sci. USA* 64, 57-63.

674 [dataset] Dimiceli, C., Carroll, M., Sohlberg, R., Kim, D.H., Kelly, M., Townshend, J.R.G.,
 675 2015. MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m
 676 SIN Grid V006. NASA EOSDIS Land Processes DAAC. doi:
 677 10.5067/MODIS/MOD44B.006.

678 Durvasula, A., Fulgione, A., Gutaker, R.M., Alacakaptan, S.I., Flood, P.J., Neto, C.,
 679 Tsuchimatsu, T., Burbano, H.A., Picó, F.X., Alonso-Blanco, C., Hancock, A.M., 2017.
 680 African genomes illuminate the early history and transition to selfing in *Arabidopsis*
 681 *thaliana*. *P. Natl. Acad. Sci. USA* 114, 5213-5218.

682 Dutilleul, P., Clifford, P., Richardson, S., Hemon, D., 1993. Modifying the *t* test for assessing
 683 the correlation between two spatial processes. *Biometrics* 49, 305-314.

684 El-Assal, S., Alonso-Blanco, C., Peeters, A.J., Raz, V., Koornneef, M., 2001. A QTL for
 685 flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. *Nat Genet.* 29, 435-440.

686 Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical
 687 explanation of MaxEnt for ecologists. *Diversity Distrib.* 17, 43-57.

688 [dataset] FAO/IIASA/ISRIC/ISSCAS/JRC, 2012. Harmonized World Soil Database (version
 689 1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

690 Fernández-Mazuecos, M., Vargas, P., 2013. Congruence between distribution modelling and
 691 phylogeographical analyses reveals Quaternary survival of a toadflax species (*Linaria*
 692 *elegans*) in oceanic climate areas of a mountain ring range. *New Phytol.* 198, 1274-1289.
 693 Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces
 694 for global land areas. *Int. J. Climatol.* 37, 4302-4315.
 695 Fletcher, W.J., Boski, T., Moura, D., 2007. Palynological evidence for environmental and
 696 climatic change in the lower Guadiana valley, Portugal, during the last 13 000 years.
 697 *Holocene* 17, 481-494.
 698 [dataset] Friedl, M., Sulla-Menashe, D., 2015. MCD12Q1 MODIS/Terra+Aqua Land Cover
 699 Type Yearly L3 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC.
 700 doi: 10.5067/MODIS/MCD12Q1.006.
 701 Fulgione, A., Hancock, A.M., 2018. Archaic lineages broaden our view on the history of
 702 *Arabidopsis thaliana*. *New Phytol.* 219, 1194-1198.
 703 Fulgione, A., Koornneef, M., Roux, F., Hermisson, J., Hancock, A.M., 2018. Madeiran
 704 *Arabidopsis thaliana* reveals ancient long-range colonization and clarifies demography in
 705 Eurasia. *Mol. Biol. Evol.* 35, 564-574.
 706 GDAL/OGR contributors, 2018. GDAL/OGR Geospatial Data Abstraction software Library.
 707 Open Source Geospatial Foundation. URL <http://gdal.org> (accessed 15 December 2018).
 708 Gil-Romera, G., García-Antón, M., Calleja, J.A., 2008. The late Holocene palaeoecological
 709 sequence of Serranía de las Villuercas (southern Meseta, western Spain). *Veget. Hist.*
 710 *Archaeobot.* 17, 653-666.
 711 Gomez, A., Lunt, D., 2006. Refugia within refugia: patterns of phylogeographic concordance
 712 in the Iberian Peninsula, in: Weiss, S., Ferrand, N. (Eds.), *Phylogeography of Southern*
 713 *European Refugia*. Springer, Dordrecht, The Netherlands, pp. 155-188.

714 Gómez, R., Méndez-Vigo, B., Marcer, A., Alonso-Blanco, C., Picó, F.X., 2018. Quantifying
715 temporal change in plant population attributes: insights from a resurrection approach. *AoB*
716 *PLANTS* 10, ply063.

717 González-Sampériz, P., Aranbarri, J., Pérez-Sanz, A., Gil-Romera, G., Moreno, A., Leunda,
718 M., Sevilla-Callejo, M., Corella, J.P., Morellón, M., Oliva, B., Valero-Garcés, B., 2017.
719 Environmental and climate change in the southern Central Pyrenees since the Last Glacial
720 Maximum: A view from the lake records. *Catena* 149, 668-688.

721 Herrera, C.M., 1992. Historical effects and sorting processes as explanations for
722 contemporary ecological patterns: character syndromes in Mediterranean woody plants.
723 *Am. Nat.* 140, 421-446.

724 Hewitt, G.M., 1999. Post-glacial re-colonization of European biota. *Biol. J. Linn. Soc.* 68, 87-
725 112.

726 Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2017. Package 'dismo': Species distribution
727 modeling. <https://CRAN.R-project.org/package=dismo>.

728 Hsu, C.W., Lo, C.Y., Lee, C.R., 2019. On the postglacial spread of human commensal
729 *Arabidopsis thaliana*: journey to the East. *Plant Physiol.* *in press*.

730 Hughes, P.D., Woodward, J.C., Gibbard, P.L., 2006. Quaternary glacial history of the
731 Mediterranean mountains. *Prog. Phys. Geog.* 30, 334-364.

732 James, A.B., Sullivan, S., Nimmo, H.G., 2018. Global spatial analysis of *Arabidopsis* natural
733 variants implicates 5'UTR splicing of *LATE ELONGATED HYPOCOTYL* in responses to
734 temperature. *Plant Cell Environ.* 41, 1524-1538.

735 Jones, S.E., Burjachs, F., Ferrer-García, C., Giralt, S., Schulte, L., Fernández-López de Pablo,
736 J., 2018. A multi-proxy approach to understanding complex responses of salt-lake
737 catchments to climate variability and human pressure: A Late Quaternary case study from
738 south-eastern, Spain. *Quaternary Sci. Rev.* 184, 201-223.

739 Kawakatsu, T., Huang, S.C., Jupe, F., Sasaki, E., Schmitz, R.J., Urich, M.A., Castanon, R.,
 740 Nery, J.R., Barragan, C., He, Y., Chen, H., Dubin, M., Lee, C.R., Wang, C., Bemm, F.,
 741 Becker, C., O'Neil, R., O'Malley, R.C., Quarless, D.X., The 1001 Genomes Consortium,
 742 Schork, N.J., Weigel, D., Nordborg, M., Ecker, J.R., 2016. Epigenomic diversity in a
 743 global collection of *Arabidopsis thaliana* accessions. *Cell*, 166, 492-505.
 744 Lavergne, S., Hampe, A., Arroyo, J., 2013. In and out of Africa: how did the Strait of
 745 Gibraltar affect plant species migration and local diversification? *J. Biogeogr.* 40, 24-36.
 746 Lee, C.R., Svardal, H., Farlow, A., Exposito-Alonso, M., Ding, W., Novikova, P., Alonso-
 747 Blanco, C., Weigel, D., Nordborg, M. 2017. On the post-glacial spread of human
 748 commensal *Arabidopsis thaliana*. *Nature Commun.* 8, 14458.
 749 Liu, L., Adrian, J., Pankin, A., Hu, J., Dong, X., von Korff, M., Turck F., 2014. Induced and
 750 natural variation of promoter length modulates the photoperiodic response of
 751 *FLOWERING LOCUS T*. *Nat Commun.* 5, 4558.
 752 Lobin, W., 1983. The occurrence of *Arabidopsis thaliana* in the Cape Verde Islands.
 753 *Arabidopsis Information Service* 20, 119-123.
 754 López-Sáez, J.A., Pérez-Díaz, S., Galop, D., Alba-Sánchez, F., Abel-Schaad, D., 2018. A late
 755 antique vegetation history of the Western Mediterranean in context. *Late Antiq. Archaeo.*
 756 2018, 93-114.
 757 Lutz, U., Nussbaumer, T., Spannagl, M., Diener, J., Mayer, K.F., Schwechheimer, C., 2017.
 758 Natural haplotypes of *FLM* non-coding sequences fine-tune flowering time in ambient
 759 spring temperatures in *Arabidopsis*. *eLife* 6, e22114.
 760 MacArthur, R.H., Wilson, E.O., 1963. An equilibrium theory of insular zoogeography.
 761 *Evolution* 17, 373-387.

762 Manzano-Piedras, E., Marcer, A., Alonso-Blanco, C., Picó, F.X., 2014. Deciphering the
 763 adjustment between environment and life history in annuals: lessons from a
 764 geographically-explicit approach in *Arabidopsis thaliana*. PLoS ONE 9, e87836.
 765 Marcer, A., Méndez-Vigo, B., Alonso-Blanco, C., Picó, F.X., 2016. Tackling intraspecific
 766 genetic structure in distribution models better reflects species geographical range. Ecol.
 767 Evol. 6, 2084-2097.
 768 Médail, F., Quézel, P., 1997. Hot-spots analysis for conservation of plant biodiversity in the
 769 Mediterranean Basin. Ann. Mo. Bot. Gard. 84, 112-127.
 770 Médail, F., Diadema, K., 2009. Glacial refugia influence plant diversity patterns in the
 771 Mediterranean Basin. J. Biogeogr. 36, 1333-1345.
 772 Miras, Y., Ejarque, A., Orengo, H., Mora, S.R., Palet, J.M., Poiraud, A., 2010. Prehistoric
 773 impact on landscape and vegetation at high altitudes: An integrated palaeoecological and
 774 archaeological approach in the eastern Pyrenees (Perafita valley, Andorra). Plant Biosyst.
 775 144, 924-939.
 776 Miras, Y., Ejarque, A., Riera-Mora, S., Orengo, H. & Palet, J.M. 2015. 28. Andorran high
 777 Pyrenees (Perafita Valley, Andorra): Serra Mijtana fen. Grana 54, 313-316.
 778 Molina-Venegas, R., Aparicio, A., Lavergne, S., Arroyo, J., 2017. Climatic and topographical
 779 correlates of plant palaeo- and neoendemism in a Mediterranean biodiversity hotspot.
 780 Ann. Bot. 119, 229-238.
 781 Montesinos, A., Tonsor, S.J., Alonso-Blanco, C., Picó, F.X., 2009. Demographic and genetic
 782 patterns of variation among populations of *Arabidopsis thaliana* from contrasting native
 783 environments. PLoS ONE 4, e7213.
 784 [dataset] Myneni, R., Knyazikhin, Y., Park, T., 2015. MOD15A2H MODIS/Terra Leaf Area
 785 Index/FPAR 8-Day L4 Global 500m SIN Grid V006. NASA EOSDIS Land Processes
 786 DAAC. doi: 10.5067/MODIS/MOD15A2H.006.

787 Olsen, K.M., Halldorsdottir, S.S., Stinchcombe, J.R., Weinig, C., Schmitt, J., Purugganan,
 788 M.D., 2004. Linkage disequilibrium mapping of *Arabidopsis CRY2* flowering time alleles.
 789 Genetics 167, 1361-1369.

790 Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species
 791 geographic distributions. Ecol. Model. 190, 231-259.

792 Picó, F.X., Méndez-Vigo, B., Martínez-Zapater, J.M., Alonso-Blanco, C., 2008. Natural
 793 genetic variation of *Arabidopsis thaliana* is geographically structured in the Iberian
 794 Peninsula. Genetics 180, 1009-1021.

795 Picó, F.X., 2012. Demographic fate of *Arabidopsis thaliana* cohorts of autumn- and spring-
 796 germinated plants along an altitudinal gradient. J. Ecol. 100, 1009-1018.

797 Provart, N.J., Alonso, J., Assmann, S.M., Bergmann, D., Brady, S.M., Brkljacic, J., Browse,
 798 J., Chapple, C., Colot, V., Cutler, S., Dangl, J., Ehrhardt, D., Friesner, J.D., Frommer,
 799 W.B., Grotewold, E., Meyerowitz, E., Nemhauser, J., Nordborg, M., Pikaard, C.,
 800 Shanklin, J., Somerville, C., Stitt, M., Torii, K.U., Waese, J., Wagner, D., McCourt, P.,
 801 2016. 50 years of *Arabidopsis* research: highlights and future directions. New Phytol. 209,
 802 921-944.

803 Quézel, P., Barbero, M., 1989. Les formations à genévriers rampants du Djurdjura (Algérie).
 804 Leur signification écologique, dynamique et syntaxonomique dans une approche globale
 805 des cédraies kabyles. Lazaroa 11, 85-99.

806 R Core Team, 2017. R: A language and environment for statistical computing. R Foundation
 807 for Statistical Computing, Vienna, Austria.

808 Rangel, T.F., Diniz-Filho, J.A.F., Bini, L.M., 2010. SAM: a comprehensive application for
 809 spatial analysis in macroecology. Ecography 33, 46-50.

810 Rodríguez-Sánchez, F., Hampe, A., Jordano, P., Arroyo, J., 2010. Past tree range dynamics in
811 the Iberian Peninsula inferred through phylogeography and palaeodistribution modelling:
812 A review. *Rev. Palaeobot. Palyno.* 162, 507-521.

813 Santos, L., Fernández-Goñi, L.F., 2003. Lateglacial and Holocene environmental changes in
814 Portuguese coastal lagoons 3: vegetation history of the Santo André coastal area.
815 *Holocene* 13, 459-464.

816 Schoener, T.W., 1968. Anolis lizards in Bimini: resource partitioning in a complex fauna.
817 *Ecology* 49, 704-726.

818 Tabas-Madrid, D., Méndez-Vigo, B., Arteaga, N., Marcer, A., Pascual-Montano, A., Weigel,
819 D., Picó, F.X., Alonso-Blanco, C., 2018. Genome-wide signatures of flowering adaptation
820 to climate temperature: Regional analyses in a highly diverse native range of *Arabidopsis*
821 *thaliana*. *Plant Cell Environ.* 41, 1806-1820.

822 The 1001 Genomes Consortium, 2016. 1,135 Genomes reveal the global pattern of
823 polymorphism in *Arabidopsis thaliana*. *Cell* 166, 481-491.

824 Tsuchimatsu, T., Goubet, P.M., Gallina, S., Holl, A.C., Fobis-Loisy, I., Bergès, H., Marande,
825 W., Prat, E., Meng, D., Long, Q., Platzer, A., Nordborg, M., Vekemans, X., Castric, V.
826 2017. Patterns of polymorphism at the self-incompatibility locus in 1,083 *Arabidopsis*
827 *thaliana* genomes. *Mol. Biol. Evol.* 34, 1878-1889.

828 Warren, D.L., Glor, R.E., Turelli, M., 2008. Environmental niche equivalency versus
829 conservatism: quantitative approaches to niche evolution. *Evolution* 62, 2868-2883.

830 Weigel, D., 2012. Natural variation in *Arabidopsis*: from molecular genetics to ecological
831 genomics. *Plant Physiol.* 158, 2-22.

832 Weiß, C.L., Schuenemann, V.J., Devos, J., Shirsekar, G., Reiter, E., Gould, B.A.,
833 Stinchcombe, J.R., Krause, J., Burbano, H.A., 2016. Temporal patterns of damage and

834 decay kinetics of DNA retrieved from plant herbarium specimens. R. Soc. Open Sci. 3,
835 160239.

836 Weng, C., Hooghiemstra, H., Duivenvoorden, J.F., 2006. Challenges in estimating past plant
837 diversity from fossil pollen data: statistical assessment, problems, and possible solutions.
838 Diversity Distrib. 12, 310-318.

839 Zeng, L., Gu, Z., Xu, M., Zhao, N., Zhu, W., Yonezawa, T., Liu, T., Qiong, L., Tersing, T.,
840 Xu, L., Zhang, Y., Xu, R., Sun, N., Huang, Y., Lei, J., Zhang, L., Xie, F., Zhang, F., Gu,
841 H., Gengm, Y., Hasegawa, M., Yang, Z., Crabbe, M.J.C., Chen, F., Zhong, Y., 2017.
842 Discovery of a high-altitude ecotype and ancient lineage of *Arabidopsis thaliana* from
843 Tibet. Sci. Bull. 62, 1628-1630.

844 Zou, Y.P., Hou, X.H., Wu, Q., Chen, J.F., Li, Z.W., Han, T.S., Niu, X.M., Yang, L., Xu,
845 Y.C., Zhang, J., Zhang, F.M., Tan, D., Tian, Z., Gu, H., Guo, Y.L., 2017. Adaptation of
846 *Arabidopsis thaliana* to the Yangtze River basin. Genome Biol. 18, 239.

847

848

Figure legends

Figure 1. Geographic location of all known relict *A. thaliana* populations to date. The number of accessions and the number of populations with coordinates, respectively, are given in parenthesis.

Figure 2. Geographic location of Iberian and African relict *A. thaliana* populations and paleobotanical sites used in this study. Photographs illustrate seven representative environments where relict *A. thaliana* occurs: montane Scots pine forests (Fun), Mediterranean mixed forests (Sca), high altitude grasslands (Vel), Mediterranean scrublands (Gra), Atlas cedar forests (Ket), and Mediterranean holm-oak forests (Ifr). The location of populations onto the map is indicated by numbers.

Figure 3. Current potential distribution range of relict *A. thaliana* across the region of study. The location of the populations used to model habitat suitability is also indicated by dots. White areas and pixels inside the distribution range indicate that the quality of MODIS layers did not reach the highest values, which did not allow the estimation of habitat suitability for these pixels.

Figure 4. Relationship between habitat suitability of relict *A. thaliana* in the present with the variability in percentage of arboreal pollen with and without pollen of the genus *Pinus* (A and B) and variability in species turnover (C and D) for different broad climatic periods. The number of paleobotanical sites (*N*), Pearson's correlation coefficients (*r*) and *P*-values are also given. *P*-values corrected for spatial autocorrelation are given in parenthesis.

Figure 5. Relationship between niche overlap values for relict *A. thaliana* with 17 tree species in the present and stability in plant communities (see text for details) in two broad climatic periods: (A) since the Last Glacial Maximum (LGM) and (B), between the LGM and the mid-Holocene (MH).

Figure 6. Neighbour-joining trees showing the genetic relationships among 1,216 genomic sequences of *CRY2* (A) and *TSF* (C). Bootstrap values larger than 50% are shown. Branches corresponding to partitions reproduced in less than 25% bootstrap replicates are collapsed. Geographical distribution of major *CRY2* (B) and *TSF* (D) haplogroups. Relict and non-relict accessions carrying different haplogroups are shown in different colours as described in the legends of maps.

Supporting Information

The following additional information is available in the online version of this article:

Table S1 Region, accession code, altitude (m.a.s.l.) and coordinates (degrees) of relict *Arabidopsis thaliana* populations known to date (Fig. 1). Unknown altitudes are indicated by n/a. Accessions sharing the same geographic coordinates are given once and the others indicated by dashes.

Table S2 Location, country, maximum calibrated years before present, altitude (m.a.s.l.), coordinates (degrees) and source of the 80 paleobotanical sites from which fossil pollen data were available (Fig. 2).

898 **Table S3** Niche overlap (Schoener's D) between the potential distribution range of relict *A.*
899 *thaliana* and each tree species in the present. Niche overlap values between the potential
900 distribution range of each trees species between the present and the Mid-Holocene (MH),
901 between the present and the Last Glacial Maximum (LGM), and between MH and LGM are
902 also given. All D values were estimated for the Iberian Peninsula. Tree species are ranked in
903 decreasing D values for relict *A. thaliana*.

904

905 **Figure S1** Distribution range maps in the present for each of the 17 tree species used in this
906 study. Data from Benito-Garzón et al. (2007).

907

908 **Table 1** Variable percent permutation importance to the fit of the distribution model for the relict lineage of *Arabidopsis thaliana* in southwest
909 Mediterranean Basin (Fig. 3). Mean and standard deviation (SD) values for each variable included in the model are also given for all populations
910 ($N = 63$) as well as for Iberian ($N = 39$) and Maghreb ($N = 24$) populations. Variables with significant differences between the Iberian Peninsula
911 and the Maghreb (P -value < 0.05 ; one-way ANOVA) are given in bold face. Dimensionless variables have no units.

Variable	Importance (%)	All populations	Iberian Peninsula	Maghreb
Leaf Area Index (LAI)	53.5	$1.00 \pm 0.42 \text{ m}^2 \text{ m}^{-2}$	$1.00 \pm 0.39 \text{ m}^2 \text{ m}^{-2}$	$1.01 \pm 0.46 \text{ m}^2 \text{ m}^{-2}$
Annual mean temperature (BIO1)	10.5	$12.71 \pm 2.74 \text{ }^{\circ}\text{C}$	$12.52 \pm 3.26 \text{ }^{\circ}\text{C}$	$13.03 \pm 1.58 \text{ }^{\circ}\text{C}$
Isothermality (BIO3)	7.3	$0.37 \pm 0.03 \text{ }^{\circ}\text{C}$	$0.37 \pm 0.03 \text{ }^{\circ}\text{C}$	$0.36 \pm 0.02 \text{ }^{\circ}\text{C}$
Precipitation of warmest quarter (BIO18)	6.7	$49.40 \pm 28.00 \text{ mm}$	$61.62 \pm 28.18 \text{ mm}$	$29.54 \pm 11.67 \text{ mm}$
Precipitation seasonality (BIO15)	5.9	50.26 ± 12.82	46.37 ± 11.62	56.58 ± 12.35
Percentage of non-arboreal vegetation	5.8	$58.00 \pm 17.69 \%$	$57.95 \pm 20.61 \%$	$58.08 \pm 11.92 \%$
Annual precipitation (BIO12)	4.4	$594.92 \pm 257.64 \text{ mm}$	$630.41 \pm 275.52 \text{ mm}$	$537.25 \pm 218.89 \text{ mm}$
Topsoil pH	2.5	5.99 ± 0.90	5.76 ± 0.75	6.37 ± 1.01
Temperature seasonality (BIO4)	2.0	6.41 ± 0.56	6.39 ± 0.62	6.42 ± 0.45
Percentage of crops	1.4	$0.07 \pm 0.24 \%$	$0.05 \pm 0.19 \%$	$0.11 \pm 0.31 \%$

1 **Table S1** Region, accession code, altitude (m.a.s.l.) and coordinates (degrees) of relict
2 *Arabidopsis thaliana* populations known to date (Fig. 1). Unknown altitudes are indicated by
3 n/a. Accessions sharing the same geographic coordinates are given once and the others
4 indicated by dashes.

Region	Accession	Altitude	Long(x)	Lat(y)
Algeria	Alg-h1	1950	4.1500	36.4500
Canary Islands	Can-0	n/a	-13.4811	29.2144
Cape Verde	Cvi-0	n/a	-23.6167	15.1111
China	Tibet-0	4200	90.9338	29.6903
China	SRR2204316	n/a	98.6300	28.0200
China	SRR2204703	—	—	—
Iberian Peninsula	Bad-1	720	-7.5548	41.5693
Iberian Peninsula	Bdm-0	2124	-3.9915	40.7842
Iberian Peninsula	Bon-0	1	-6.2898	36.8768
Iberian Peninsula	Can-0	879	-4.5141	39.3884
Iberian Peninsula	Cat-0	733	-3.6870	40.5421
Iberian Peninsula	Cem-0	898	-4.3173	41.1484
Iberian Peninsula	Con-0	756	-5.5991	37.9382
Iberian Peninsula	Don-0	23	-6.3598	36.8323
Iberian Peninsula	Fun-0	1804	-4.0503	40.7872
Iberian Peninsula	Gra-0	1284	-5.3949	36.7678
Iberian Peninsula	Hee-0	787	-3.4268	39.3048
Iberian Peninsula	Her-12	584	-5.7828	39.3948
Iberian Peninsula	Hum-2	924	-3.6889	42.2272
Iberian Peninsula	Las-0	1071	-4.1527	40.5680
Iberian Peninsula	Lso-0	769	-3.1629	38.8577
Iberian Peninsula	Mar-1	1030	-3.9316	39.5813
Iberian Peninsula	Mdc-0	725	-3.5296	38.8743
Iberian Peninsula	Med-0	771	-6.7657	42.4552
Iberian Peninsula	Moj-0	705	-5.2814	36.7567
Iberian Peninsula	Mor-0	803	-6.0732	40.8705
Iberian Peninsula	Nac-0	1420	-3.9939	40.7510
Iberian Peninsula	Nav-0	735	-4.6466	40.4177
Iberian Peninsula	Ped-0	1096	-3.8978	40.7385
Iberian Peninsula	Per-0	472	-1.1216	37.5952
Iberian Peninsula	Pol-1	520	-4.2512	39.8163
Iberian Peninsula	Pub-0	906	-6.6334	42.0567
Iberian Peninsula	Pun-0	826	-4.7757	40.4033
Iberian Peninsula	Rdc-0	800	-4.2866	40.4517
Iberian Peninsula	Sca-0	501	-5.3700	37.9622
Iberian Peninsula	Slc-3	945	-3.7443	38.4706
Iberian Peninsula	Sln-22	646	-3.9948	40.3569
Iberian Peninsula	Sne-0	1650	-3.2625	37.0746
Iberian Peninsula	Toe-0	929	-3.9151	40.5925
Iberian Peninsula	Vdc-0	593	-5.0302	38.0498
Iberian Peninsula	Vel-0	2662	-3.3818	37.0854
Iberian Peninsula	Ven-0	1553	-4.0141	40.7585
Iberian Peninsula	Vim-0	652	-6.5101	41.8771
Iberian Peninsula	Vis-0	244	-6.0455	39.8460
Iberian Peninsula	Zar-0	1250	-4.1948	40.5518

Italy	Etna-2	1800	14.9800	37.6900
Italy	Mr-0	1000-1500	9.6500	44.1500
Lebanon	Qar-8a	1160-1213	35.8400	34.1000
Madeira	A1	1812	-16.9300	32.7400
Madeira	A2	—	—	—
Madeira	A3	—	—	—
Madeira	A4	—	—	—
Madeira	A5	—	—	—
Madeira	A6	1600	-16.9370	32.7450
Madeira	R1	1200	-17.1300	32.7540
Madeira	R2	—	—	—
Madeira	R3	—	—	—
Madeira	R4	1200	-17.1320	32.7580
Madeira	U1	1800	-16.9430	32.7590
Morocco	Agl-0	1530	-5.4486	32.9724
Morocco	Ait-1	2075	-7.8129	31.2366
Morocco	Arb-0	1070	-7.5262	31.4199
Morocco	Azr-0	1726	-5.1791	33.4236
Morocco	Bab-0	1186	-5.0237	35.0426
Morocco	Bba-2	1320	-4.0839	34.0193
Morocco	Bbe-0	1340	-4.8314	34.9952
Morocco	Elh-2	1510	-7.4064	31.4720
Morocco	Elk-1	1230	-6.0150	32.5352
Morocco	Ifr-2	1605	-5.1747	33.5501
Morocco	Ket-10	1614	-4.6661	34.9608
Morocco	Khe-0	1470	-5.5103	32.9274
Morocco	Meh-0	1336	-4.0515	33.9561
Morocco	Oua-0	1825	-6.2753	32.0785
Morocco	Set-0	1430	-7.6736	31.2266
Morocco	Tah-0	1224	-4.2225	34.0529
Morocco	Taz-0	1386	-4.1026	34.0917
Morocco	Til-2	1420	-6.2296	32.0421
Morocco	Tiz-0	1433	-4.0265	33.8723
Morocco	Zin-9	116	-5.4270	35.4528
Morocco	Toufl-1	n/a	-7.4217	31.4687
Morocco	Aitba-1	n/a	-7.4499	31.4840
Morocco	Aitba-2	—	—	—
Morocco	Ita-0	2000	-4.1989	34.0787
South Africa	SA-h1	n/a	19.3750	-34.1250
South Africa	SA-h2	—	—	—
South Africa	SA-h3	—	—	—
South Africa	SA-h4	n/a	19.2820	-33.3990
Tanzania	Tanz-1	2775	36.1200	-2.8739
Tanzania	Tanz-2	—	—	—
Tanzania	TZ-Kilimanjaro	n/a	37.1596	-3.0216
Tanzania	TZ-Mt. Meru	n/a	36.7167	-3.2333

6 **Table S2** Location, country, maximum calibrated years before present, altitude (m.a.s.l.), coordinates (degrees) and source of the 80
7 paleobotanical sites from which fossil pollen data were available (Fig. 2).

Location	Country	Max cal BP	Altitude	Long(x)	Lat(y)	Source
Aanasser	Morocco	3550	1342	-4.9925	35.0176	Rachid Cheddadi
Almenara de Adaja	Spain	2747	784	-4.6681	41.1919	European Pollen Database
Alsa	Spain	4310	560	-4.0167	43.1178	European Pollen Database
Antas	Spain	8855	0	-1.8236	37.2083	European Pollen Database
Arroyo de las Cárcavas	Spain	2234	1300	-4.0314	40.8419	European Pollen Database
Arroyo de Navalacarreta	Spain	643	1250	-4.0311	40.8522	European Pollen Database
Arroyo de Valdeconejos	Spain	580	1380	-4.0597	40.8622	European Pollen Database
Atxuri	Spain	6437	500	-1.5500	43.2500	European Pollen Database
Ayoó de Vidriales	Spain	18468	780	-6.0703	42.1261	European Pollen Database
Bab El Karn	Morocco	9163	1178	-5.2070	35.0225	Rachid Cheddadi
Bajondillo	Spain	43000	20	-4.4964	36.6197	European Pollen Database
Basa de la Mora	Spain	10089	1914	0.3167	42.5333	Penélope González-Sampériz
Bosc dels Estanyons	Andorra	11899	2180	1.6289	42.4803	Yannick Miras
Bourdim	Algeria	3057	13	8.2539	36.8033	European Pollen Database
Canaleja	Spain	10241	1029	-2.4500	40.9000	European Pollen Database
Charco da Candieira	Portugal	12792	1409	-7.5764	40.3417	European Pollen Database
CM5 Beliche	Portugal	13092	25	-7.4500	37.2667	Will Fletcher
Cueto de Avellanosa	Spain	6218	1320	-4.3642	43.1167	European Pollen Database
Culazón	Spain	3811	592	-4.4892	43.2328	European Pollen Database
Dar Fatma	Túnez	44370	780	8.7667	36.8167	European Pollen Database
Delta del Río Besós	Spain	902	0	2.2483	41.3803	European Pollen Database
Djebel El Ghorra	Túnez	6726	1203	8.3947	36.5975	European Pollen Database
El Maíllo	Spain	10553	1100	-6.2097	40.5467	European Pollen Database
El Payo	Spain	3850	984	-6.7711	40.2533	European Pollen Database
El Pirulejo	Spain	15000	660	-4.1950	37.4389	European Pollen Database
El Portalet	Spain	32977	1802	-0.3978	42.8000	Penélope González-Sampériz
Fuente de la Leche	Spain	3000	1382	-5.0569	40.3468	José Antonio López-Sáez
Fuente del Pino Blanco	Spain	599	1343	-4.9776	40.2363	José Antonio López-Sáez
Garaat El-Ouez	Algeria	19706	45	8.3333	36.8183	European Pollen Database
Garganta del Mesto	Spain	3760	1000	-5.3333	39.5000	Graciela Gil-Romera

Hoya del Castillo	Spain	9048	260	-0.5000	41.2500	European Pollen Database
Ifri Oudadane	Morocco	10860	9	-3.2583	35.2128	European Pollen Database
La Molina	Spain	4538	650	-6.3272	43.3811	European Pollen Database
Lago de Ajo	Spain	23035	1570	-6.1500	43.0500	European Pollen Database
Lago de Banyoles	Spain	30559	173	2.7500	42.1333	European Pollen Database
Lago de Sanabria	Spain	12821	1050	-6.7333	42.1000	European Pollen Database
Lago Estanya	Spain	20578	670	0.5333	42.0333	Penélope González-Sampériz
Lago Las Pardillas	Spain	10604	1870	-3.0453	42.0456	European Pollen Database
Lagoa Comprida 2	Portugal	12444	1650	-7.6361	40.3628	European Pollen Database
Lagoa Travessa II	Portugal	4343	3	-8.7725	38.3044	European Pollen Database
Laguna de la Roya	Spain	14559	1608	-6.7667	42.2167	European Pollen Database
Laguna Guallar	Spain	9095	336	-0.2167	41.4000	European Pollen Database
Laguna Mesagosa	Spain	18649	1600	-2.8122	41.9653	European Pollen Database
Laguna Negra	Spain	17628	1760	-2.8483	42.0003	European Pollen Database
Laguna Salada Chiprana	Spain	6523	150	-0.1667	41.2333	European Pollen Database
Lake Sidi Ali	Morocco	12000	2080	-5.0000	33.0500	Will Fletcher
Lanzahíta	Spain	2434	588	-4.9358	40.2222	European Pollen Database
Majen Ben H'mida	Túnez	11638	330	9.0833	37.1333	European Pollen Database
Majen El Orbi	Túnez	5585	200	9.0833	37.1500	European Pollen Database
Majen El Orbi	Túnez	5850	200	9.0833	37.1500	European Pollen Database
M'Had	Morocco	5741	754	-5.4387	35.1284	Rachid Cheddadi
Monte Areo	Spain	11812	200	-5.7689	43.5289	European Pollen Database
Montes do Buio Cuadramón	Spain	12319	700	-7.5347	43.4742	European Pollen Database
Navarrés	Spain	30936	225	-0.6833	39.1000	European Pollen Database
Oumm El-Khaled	Algeria	5899	970	7.6000	35.0833	European Pollen Database
Patateros	Spain	2497	700	-4.6742	39.5972	European Pollen Database
Pedrido	Spain	5019	770	-7.0697	43.4594	European Pollen Database
Peña Negra	Spain	4077	1258	-5.7922	40.3347	European Pollen Database
Pico del Sertal	Spain	4590	940	-4.4361	43.2156	European Pollen Database
Planell de Perafita	Andorra	10158	2240	1.5669	42.4792	Yannick Miras
Posidonia Lligat	Spain	1312	0	3.2911	42.2922	European Pollen Database
PRD-4	Spain	12011	260	-8.5167	42.5333	European Pollen Database
Puerto de Belate	Spain	7650	847	-2.0500	43.0333	European Pollen Database
Puerto de las Estacas de Trueba	Spain	5460	1160	-3.7006	43.1214	European Pollen Database
Puerto de los Tornos	Spain	7235	920	-3.4333	43.1500	European Pollen Database

Puerto de Serranillos	Spain	1953	1700	-4.9342	40.3072	European Pollen Database
Quintanar de la Sierra	Spain	16125	1470	-3.0167	42.0333	European Pollen Database
Roquetas de Mar	Spain	5950	0	-2.5889	36.7944	European Pollen Database
Salada Pequeña	Spain	4167	357	-0.2167	41.0333	European Pollen Database
Saldropo	Spain	9098	625	-2.7167	43.0500	European Pollen Database
Salines	Spain	13543	475	-0.8883	38.5006	Francesc Burjachs
San Rafael	Spain	17319	0	-2.6014	36.7736	European Pollen Database
Santo André	Portugal	14160	38	-8.4700	38.5000	Luisa Santos Fidalgo
Serra Mitjana	Andorra	1522	2406	1.5828	42.4675	Yannick Miras
Tigalmamine	Morocco	11974	1626	-5.3500	32.9000	European Pollen Database
Tubilla del Lago	Spain	7458	900	-3.5725	41.8083	European Pollen Database
Turbera de la Panera Cabras	Spain	228	1648	-5.7581	40.1658	European Pollen Database
Valdeyernos	Spain	3397	850	-4.0964	39.4411	European Pollen Database
Villena	Spain	10920	550	-0.9199	38.6297	Samantha Jones
Xan de Llamas	Spain	4944	1500	-6.3214	42.3042	European Pollen Database

9 **Table S3** Niche overlap (Schoener's *D*) between the potential distribution range of relict *A. thaliana* and each tree species in the present. Niche
10 overlap values between the potential distribution range of each trees species between the present and the Mid-Holocene (MH), between the
11 present and the Last Glacial Maximum (LGM), and between MH and LGM are also given. All *D* values were estimated for the Iberian Peninsula.
12 Tree species are ranked in decreasing *D* values for relict *A. thaliana*.

Tree species	Relict <i>A. thaliana</i>	Present – MH	Present – LGM	MH – LGM
<i>Quercus ilex</i> subsp. <i>ballota</i>	0.499	0.791	0.630	0.714
<i>Quercus pyrenaica</i>	0.455	0.808	0.421	0.446
<i>Pinus pinaster</i>	0.373	0.811	0.532	0.625
<i>Pinus pinea</i>	0.367	0.771	0.551	0.624
<i>Castanea sativa</i>	0.328	0.849	0.376	0.394
<i>Quercus faginea</i>	0.317	0.786	0.490	0.524
<i>Quercus suber</i>	0.310	0.785	0.342	0.388
<i>Pinus nigra</i>	0.292	0.758	0.384	0.435
<i>Juniperus thurifera</i>	0.267	0.724	0.364	0.410
<i>Pinus halepensis</i>	0.258	0.747	0.493	0.591
<i>Pinus sylvestris</i>	0.195	0.804	0.352	0.363
<i>Quercus robur</i>	0.169	0.868	0.432	0.448
<i>Quercus petraea</i>	0.162	0.855	0.435	0.433
<i>Abies alba</i>	0.150	0.831	0.632	0.622
<i>Fagus sylvatica</i>	0.116	0.848	0.339	0.328
<i>Quercus ilex</i> subsp. <i>ilex</i>	0.107	0.784	0.296	0.308
<i>Pinus uncinata</i>	0.052	0.819	0.197	0.178

13

14 **Figure S1** Current potential distribution ranges of 17 tree species in the Iberian Peninsula.
 15 Data from Benito-Garzón et al. (2007).

16

17

Quercus ilex
spp. ilex

Quercus petraea

Quercus pyrenaica

Quercus robur

Quercus suber

18

19