

Temporal variation in genetic diversity and effective population size of Mediterranean and subalpine *Arabidopsis thaliana* populations

NASR H. GOMAA,* ALICIA MONTESINOS-NAVARRO,† CARLOS ALONSO-BLANCO‡ and F. XAVIER PICÓ†

*Department of Botany, Faculty of Science, Beni-Suef University, Beni-Suef, Egypt, †Departamento de Ecología Integrativa, Estación Biológica de Doñana (EBD), Consejo Superior de Investigaciones Científicas (CSIC), 41092 Sevilla, Spain,

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

Abstract

Currently, there exists a limited knowledge on the extent of temporal variation in population genetic parameters of natural populations. Here, we study the extent of temporal variation in population genetics by genotyping 151 genome-wide SNP markers polymorphic in 466 individuals collected from nine populations of the annual plant *Arabidopsis thaliana* during 4 years. Populations are located along an altitudinal climatic gradient from Mediterranean to subalpine environments in NE Spain, which has been shown to influence key demographic attributes and life cycle adaptations. Genetically, *A. thaliana* populations were more variable across space than over time. Common multilocus genotypes were detected several years in the same population, whereas low-frequency multilocus genotypes appeared only 1 year. High-elevation populations were genetically poorer and more variable over time than low-elevation populations, which might be caused by a higher overall demographic instability at higher altitudes. Estimated effective population sizes were low but also showed a significant decreasing trend with increasing altitude, suggesting a deeper impact of genetic drift at high-elevation populations. In comparison with single-year samplings, repeated genotyping over time captured substantially higher amount of genetic variation contained in *A. thaliana* populations. Furthermore, repeated genotyping of populations provided novel information on the genetic properties of *A. thaliana* populations and allowed hypothesizing on their underlying mechanisms. Therefore, including temporal genotyping programmes into traditional population genetic studies can significantly increase our understanding of the dynamics of natural populations.

Keywords: altitudinal gradient, genetic differentiation, natural populations, population ecology, population genetics, SNP markers

Received 17 December 2010; revision received 2 June 2011; accepted 14 June 2011

Introduction

Numerous biotic and abiotic factors acting in concert with key demographic parameters determine population size and the extent of its fluctuations over time. The demographic consequences of temporal variation in

population size are well known: increasing temporal variation in vital rates increases extinction risks (Goodman 1987; Menges 2000). In contrast, the effects of such variation on the genetic properties of populations are less understood. However, this is important because it is widely accepted that the interaction between demographic (e.g. reduction in population size) and genetic factors (e.g. loss of genetic diversity, increasing inbreeding rates) is crucial to accurately assess population

Correspondence: F. Xavier Picó, Fax: +34 95 462 1125; E-mail: xpico@ebd.csic.es

extinction rates (Lande 1988; Lynch *et al.* 1995; Frankham 2005; Picó *et al.* 2009). Genetic monitoring, a term defined as the quantification of temporal changes in population genetic metrics obtained using molecular markers (Luikart *et al.* 1999; Schwartz *et al.* 2006), can provide relevant insights into the extent of temporal genetic variation in natural populations.

Molecular markers are commonly used to make inference on some important demographic events, such as population bottlenecks or migration rates (Cornuet & Luikart 1996; Luikart *et al.* 1999; Keller *et al.* 2001; Avise 2004). Traditionally, genetic data obtained from different populations sampled at a single time point suffice to make such demographic inference based on gene and homozygote frequencies (Luikart *et al.* 1999; Richards *et al.* 2003; Zhang *et al.* 2004). Nevertheless, it remains to be seen to what extent short-term transient population dynamics affect the genetic composition of populations and the demographic interpretation of genetic data. Populations may experience dramatic fluctuations in size from year to year, such fluctuations being more pronounced as the organism's lifespan becomes shorter (García *et al.* 2008). Populations can be sampled in good or poor years, that is, in years with high or low number of individuals, which can strongly determine gene frequencies. On top of that, the number of individuals sampled and genotyped normally represents a very small proportion of individuals present in the population so that sampling strategies are also important to obtain reliable population genetic estimates. Hence, the particularities of the sampling year and the features of the sampling scheme might affect our ability to evaluate demographic patterns from genetic data.

Repeated genotyping of natural populations over time has classically been carried out to obtain genetic data from two time periods. These data have been used to estimate the effective population size (N_e) based on changes in gene frequencies over two generations (Krimbas & Tsakas 1971; Schaffer *et al.* 1977; Williamson & Slatkin 1999; Siol *et al.* 2007; Waples & Yokota 2007; Lundemo *et al.* 2009; Osborne *et al.* 2010). N_e has been defined as the number of breeding individuals of an idealized Wright–Fisher population, i.e. a panmictic population with a binomial distribution of the number of successful offspring per parent (Fisher 1930; Wright 1931) that would exhibit the same genetic properties as the population of study. The theoretical estimation of N_e based on temporal variation in gene frequency has extensively been developed (Krimbas & Tsakas 1971; Nei & Tajima 1981; Pollack 1983; Waples 1989; Caballero 1994; Williamson & Slatkin 1999; Anderson *et al.* 2000; Wang 2001; Berthier *et al.* 2002; Wang *et al.* 2010). Estimates of N_e are important for ecology, genetics, evolutionary biology and conservation biology (Wang 2005;

Schwartz *et al.* 2006; Antao *et al.* 2010) because they help to predict changes in neutral genetic variation or fixation of deleterious alleles among other genetic properties (see Wang 2005). In the last few years, the interest for assessing the extent of temporal changes in genetic variation in natural populations has notoriously increased (Richards *et al.* 2003; Schwartz *et al.* 2006; Mhemmed *et al.* 2008; Honnay *et al.* 2009; Lundemo *et al.* 2009; Palstra *et al.* 2009; Antao *et al.* 2010; Bomblies *et al.* 2010).

The main goal of this study is to generate and analyse temporal series of genetic data to evaluate their potential to better understand the ecological (e.g. population dynamics) and genetic (e.g. genetic drift) processes affecting an organism's population biology. In particular, this study spans up to 4 years and is conducted in nine populations of the annual plant *Arabidopsis thaliana* distributed along an altitudinal climatic gradient in NE Spain. This gradient ranges from low-elevation coastal Mediterranean sites to high-elevation montane subalpine locations. It has recently been shown that high-elevation *A. thaliana* populations exhibit higher plant mortality rates in winter and stronger secondary seed dormancy in spring than low-elevation populations (Montesinos *et al.* 2009). Interestingly, plants from different populations along the altitudinal climatic gradient exhibit relevant life cycle adaptations, including vegetative growth, biomass allocation, phenology and fecundity (Montesinos-Navarro *et al.* 2011), which might ensure the long-term viability of populations under contrasting environmental conditions. Therefore, assuming adaptation to local environmental conditions, temporal changes in the genetic composition of populations are expected to be mainly accounted for by temporal variation in demographic processes.

The assessment of temporal patterns of variation in genetic diversity of populations along an altitudinal climatic pattern can provide important clues to improve our knowledge about the dynamics, and their underlying mechanisms, of *A. thaliana* natural populations. Here, we specifically aim (i) to evaluate the extent of temporal variation in standard genetic parameters, (ii) to quantify the relationship between temporal variation in genetic data and altitude and (iii) to estimate N_e based on temporal changes in gene frequency. The results are also discussed on the basis of their implications for the continuous development of population genetics as a discipline.

Materials and methods

Study species and populations

Arabidopsis thaliana (L.) Heyhn. (Brassicaceae) is an annual plant with a persistent seed bank (Baskin &

Baskin 1972, 1983; Lundemo *et al.* 2009; Montesinos *et al.* 2009). The species is self-compatible and self-fertile. Average outcrossing rates of populations have been reported to vary between 0.3% and 2.5% (Abbott & Gomes 1989; Bergelson *et al.* 1998; Bakker *et al.* 2006; Picó *et al.* 2008; Bomblies *et al.* 2010; Platt *et al.* 2010), although values as high as 14.5% have also been recorded (Bomblies *et al.* 2010). The species' native range includes Europe and the Central Asian mountains (Hoffmann 2002; Platt *et al.* 2010). *Arabidopsis thaliana* is considered a poor competitor and mainly occurs in disturbed areas as well as openings of natural vegetation (Picó *et al.* 2008; Lundemo *et al.* 2009; Montesinos *et al.* 2009; Bomblies *et al.* 2010). A population of *A. thaliana* is typically represented by an assemblage of stands differing in area, number of plants per stand and distance among stands (Montesinos *et al.* 2009; Bomblies *et al.* 2010).

This study was conducted in nine natural populations from NE Spain. Populations were located along an altitudinal climatic gradient ranging from low-elevation Mediterranean coastal sites (population codes: BAR, Barcelona; HOR, Hortsavinyà; COC, Cap de Creus; POB, Poblet; MUR, Mura; Table 1) to high-elevation subalpine montane sites (VDM, Vilanova de Meià; ALE, Albet; BIS, Bisaurri; VIE, Vielha; Table 1). Coastal and montane populations were separated by 24–96 and 29–209 km, respectively. Distance from the sea, which strongly influences climate in the Iberian Peninsula, for coastal and montane populations ranged 4–33 and 104–173 km, respectively. See Montesinos *et al.* (2009) for

environmental and ecological descriptions of each population.

Study area boundaries, plant density and seed collection

We selected an area within each population that ranged from a low of 64 m² to a high of 3540 m² (Table 1). The main criteria to set the boundaries of these areas were high plant density and isolation from other occupied areas because of natural barriers (e.g. thick woody vegetation, creeks). Some areas were clearly delimited with plants clumped in stands, and these areas tended to be small. In larger areas, plants occurred scattered across the whole site. Area did not correlate with any of the genetic parameters of study ($P > 0.37$ in all cases), so this variable was not taken into account in the analyses.

We obtained an estimation of plant density as the number of reproductive individuals in spring along a 4-m permanent transect laid down for demographic monitoring at each population (see details in Montesinos *et al.* 2009). Transects were placed across a representative stand within each area where *A. thaliana* was abundant. The characteristics of each area eventually determined the location of transects. For small areas, this task was straightforward because stands were clearly delimited. As the area became larger, stands were more diffuse and decisions on transect location more difficult to make.

Seeds required for this study were exclusively collected from plants growing within the whole delimited

Table 1 Geographic information and sampling area of *Arabidopsis thaliana* populations (ranked by altitude)

Population	Coordinates	Altitude (m.a.s.l.)	Area (m ²)	Samples 2006	Samples 2007	Samples 2008	Samples 2009
BAR	41.43N, 2.13E	429	1206	13	19	14	15
HOR	41.67N, 2.62E	431	1134	15	20	19	14
COC	42.31N, 3.20E	519	374	–	19	–	19
POB	41.35N, 1.03E	656	238	–	20	21	21
MUR	41.68N, 2.00E	836	330	–	20	13	15
VDM	42.04N, 1.01E	975	64	20	–	19	–
ALE	42.41N, 1.32E	1225	3540	16	19	21	–
BIS	42.49N, 0.54E	1450	540	–	13	16	13
VIE	42.63N, 0.76E	1620	275	14	20	18	–
Controls				0	18	21	26
Total				78	168	162	123

Seed collection was strictly conducted every sampling year within the same delimited area. The final number of individuals ($N = 466$) successfully genotyped with 151 SNPs is indicated for each population and year. Years with no plants within the sampling area or with a low number of individuals to be included in the analyses are represented by dashes. The number of individuals genotyped twice (controls; $N = 65$) to calculate the genotyping error and the total number of individuals genotyped per year are also indicated.

BAR, Barcelona; HOR, Hortsavinyà; COC, Cap de Creus; POB, Poblet; MUR, Mura; VDM, Vilanova de Meià; ALE, Albet; BIS, Bisaurri; VIE, Vielha.

area. Years with no plants in the study area or with a very low number of plants were excluded from the analyses (see Table 1). In May/June of each of the four study years (2006–2009), we collected seed from 20 to 25 haphazardly chosen individual plants within delimited areas at each population. Individuals sampled were not totally removed from the area after seed collection to minimize the alteration in the natural dynamics of populations: some individuals already dispersed seeds from the first fruits produced or still bore unripe fruits that could disperse further seeds. Seeds were stored in cellophane bags at room temperature in darkness.

DNA isolation and SNP genotyping

Every autumn, approximately 3 months after seed collection, plants were grown from seed to isolate DNA. Seeds were stratified in moist Petri dishes at 4 °C in darkness for 4 days in an incubator (Radiber, S.A., Barcelona, Spain). A total of 20–40 seeds per individual were sown on trays with standard soil mixture in an air-conditioned greenhouse (22 °C day, 15 °C night, 16-h day length and constant high moisture) from the Institute for Plant Biochemistry and Photosynthesis (IBVF-CSIC, Sevilla, Spain). Five weeks after sowing, we pooled leaf tissue from 6 to 10 sister plants for each individual. DNA was then isolated following a previously described protocol (Bernartzky & Tanksley 1986) without mercaptoethanol.

Those individuals that did not germinate, had failed genotyping or had missing values > 25% were discarded from the analyses. Individuals (Table 1) were genotyped with 239 presumably neutral nuclear SNP loci using the SNPlex technique (Applied Biosystems, Foster City, CA, USA) through the CEGEN Genotyping Service (<http://www.cegen.org>). On average, there were about 48 SNPs per chromosome (range = 42–55 markers) located at approximately 0.5 Mb from each other (range = 0.11 Kb–1.82 Mb). SNPs were divided into five sets of 47/48 markers. Two of the sets contained SNPs that are polymorphic in Central Europe (Schmid *et al.* 2006; Picó *et al.* 2008). Another two sets included polymorphisms selected from accessions from the Iberian Peninsula (Picó *et al.* 2008). The remaining set of SNPs was randomly selected from a set of frequent polymorphisms in a worldwide collection exhibiting a minor allele frequency between 25% and 30% (Warthmann *et al.* 2007). Previous studies indicated low ascertainment bias when comparing sets of polymorphic SNPs in Central Europe and in the Iberian Peninsula (Picó *et al.* 2008; Montesinos *et al.* 2009). Therefore, all SNPs were analysed simultaneously.

We eventually took into account a total of 466 individuals genotyped with 151 SNPs that were polymor-

phic in at least one population and year and with missing values < 25%. On average, all selected SNPs exhibited $2.52 \pm 4.50\%$ missing values (range = 0.00–24.50%). A total of 65 individuals (c. 14% of the samples; Table 1) were genotyped twice to calculate the genotyping error (equal to 0.74%).

Genetic data analysis

Overall, our data set including all individuals and years showed an observed heterozygosity of 0.002. However, the SNPlex technique fails to detect some heterozygous loci (Picó *et al.* 2008), so we did not report changes in observed heterozygosity. For each population and year, we estimated mean number of alleles per locus (n_a), mean allelic richness per locus (R_S) and mean gene diversity (H_S) using FSTAT v.2.9.3 (Goudet 1995). We calculated the percentage of polymorphic loci (PL), the total and year-specific number of nonredundant multilocus genotypes (N_H), the per cent number of markers that differed among all pairs of nonredundant multilocus genotypes and the proportion of linkage disequilibrium (LD) as the percentage of pairs of loci showing significant LD in relation to the total number of pairwise combinations (P_D ; Stenøien & Sæstad 1999). The number of pairs of loci showing significant LD ($P < 0.05$) was also calculated with FSTAT v.2.9.3 (Goudet 1995) using a set of PL that excluded loci segregating as singleton alleles (as in Picó *et al.* 2008). Finally, for each population, we estimated the percentage of multilocus genotypes that was observed with 1, 2 or 3 years of data to assess the amount of genetic variation that can be recorded with a different number of study years.

We assessed genetic relationships among multilocus genotypes by Bayesian means using the model-based clustering algorithm implemented in Structure v.2.2 (Pritchard *et al.* 2000; Falush *et al.* 2003). Briefly, multilocus genotypes were analysed with a haploid setting, using the linkage model with correlated allele frequencies, and running the algorithm with 50 000 MCMC (Markov chain Monte Carlo) iterations of burn-in length and 20 000 after-burning repetitions for parameter estimations. To estimate the K number of ancestral genetic populations and the ancestry membership proportions of each individual to these clusters, the algorithm was run 10 times for each K value from $K = 2$ to $K = 15$. Differences between the data likelihood of successive K values were tested using the nonparametric Wilcoxon test for two related samples. The final K value was estimated as the largest K value with significantly higher likelihood than that from $K - 1$ runs (two-sided $P < 0.005$). A similarity coefficient between runs of 0.99 was estimated using the symmetric similarity coefficient (Nordborg *et al.* 2005). This parameter and the average

matrix of ancestry membership proportions of the 10 runs (Q matrix) were computed using CLUMPP v.1.1 (Jakobsson & Rosenberg 2007). We also analysed the genetic relationships among multilocus genotypes using an alternative method: UPGMA clustering analyses with the program MEGA v.4 (Tamura *et al.* 2007) based on the per cent number of markers that differed among all pairs of multilocus genotypes.

Given the imbalance of our data set as a result of years and populations with no plants (Table 1), we could not analyse the partition of the genetic variance among years, among populations and within populations with a unique hierarchical analysis of molecular variance (AMOVA). Consequently, we first estimated for each population, genetic differentiation among years, and then for each year, genetic differentiation among and within populations. AMOVAS were conducted using the program Arlequin v.3.1 (Excoffier *et al.* 2005). F_{ST} statistics (Weir & Cockerham 1984) and their significance were calculated from 1000 permutations. Isolation by distance was analysed for each year as the relationship between genetic distance and geographic distance using the Isolation-By-Distance Web Service v.3.15 (Jensen *et al.* 2005). Genetic distance (D) was calculated as the Slatkin's linearized F_{ST} , $D = F_{ST}/[1 - F_{ST}]$ (Slatkin 1995). Mantel tests (Mantel 1967) were conducted with log-transformed genetic and geographic distances and their significance estimated with 1000 permutations.

Two estimates of effective population size (N_e) were computed with the software NEESTIMATOR v.1.3 (Peel *et al.* 2004) using moment-based (Pollack 1983; Waples 1989) and likelihood-based (Berthier *et al.* 2002) approaches. Both methods estimate N_e examining allele frequency variation across generations. For each population, we used genetic data from the first and last year of study spanning 2–3 years. The computation of N_e estimates can only be performed using PL between the two sampling years. The moment-based N_e requires the estimation of the standardized variance in allele frequencies between initial and final generations, and the 95% confidence intervals are computed using a chi-square approximation (Waples 1989). In contrast, the likelihood-based N_e takes into account the gene genealogy described by the standard coalescent model (Berthier *et al.* 2002 and references therein). Simulations to compute likelihood-based N_e were performed five times to verify the consistency of the simulations. Likelihood-based N_e values were represented by the average of the five runs because output simulations were very similar (results not shown).

The relationship between plant density estimates from permanent transects and genetic diversity parameters (n_a , R_S , H_S , PL, N_H and the percentage of markers that differed among multilocus genotypes) was analysed by computing Pearson's correlation coefficients for

those populations with at least 3 years of demographic and genetic data (BAR, HOR, POB, MUR and BIS in 2007, 2008 and 2009). We computed mean values across years and the coefficients of variation among years (CV) for all genetic diversity parameters (n_a , R_S , H_S , PL, N_H and the percentage of markers that differed among multilocus genotypes), which were regressed against altitude using linear regression models. We also explored the relationship between altitude and F_{ST} values among years and the two estimates of N_e by means of linear regression models. We tested the effect of year of occurrence (year-specific multilocus genotypes in 2006, 2007, 2008, 2009 and multilocus genotypes observed in more than a year) on the total number of individuals per multilocus genotype observed across years using ANOVA and Student–Newman–Keuls *post hoc* tests. We did not transform any parameter in the analyses because the visual inspection of residuals indicated that the assumptions of parametric analyses (e.g. homoscedasticity) were met. All linear regressions were performed using SPSS v.17 statistical software (SPSS Inc., Chicago, IL, USA).

Sample size effects and subsampling technique

The number of individuals genotyped per population and year varied between 13 and 21 individuals (Table 1). To test whether differences in sample size biased the observed patterns of variation and their relationship with altitude, we carried out the following subsampling approach. For each population and year, we created 100 random subsamples of 13 individuals each, the minimum sample size observed. Subsequent input files were analysed with FSTAT to obtain new estimates of PL, n_a , H_S , R_S and among-year genetic differentiation F_{ST} . For each population and year, we calculated mean (\pm SD) of genetic parameters for each set of 100 runs. Observed and simulated data were highly consistent (Fig. S1 and Table S1, Supporting information), and previous analyses with both sets of data showed the same patterns of variation with altitude (results not shown). On top of that, 86 of 108 of genetic parameters for observed data fell within the SD of those for simulated data (Table S1, Supporting information). Overall, these results indicate that bias because of variation in sample size is low in this data set. Results are based on observed data.

Results

Plant density within transects

Yearly demographic surveys of permanent transects laid down in representative stands from spring 2007 to

spring 2010 indicated that plant density estimates declined over time. The BIS population was the only one with a different pattern as this stand reached the largest size in 2009 (Fig. 1). Based on a subset of populations with coupled demographic and genetic data, temporal variation in plant density was not significantly correlated with temporal variation in genetic diversity parameters in any population ($P > 0.07$ in all cases).

Multilocus genotypes

A total of 142 different multilocus genotypes were detected among the 466 individuals genotyped in all

populations and years (Table S2, Supporting information). No identical multilocus genotype was found among populations in any year. The total number of multilocus genotypes recorded in each population over the study years varied between three and 35 (Table 2). Thirty-six multilocus genotypes (25%) were found in different years, while the remaining 106 multilocus genotypes (75%) were detected specifically in a single year and population. The number of total and year-specific multilocus genotypes varied considerably across years in all populations (Table 2). On average, one and 2 years of data captured up to 49.1% and 73.8% of all multilocus genotypes observed over the entire study periods, respectively (Table 3).

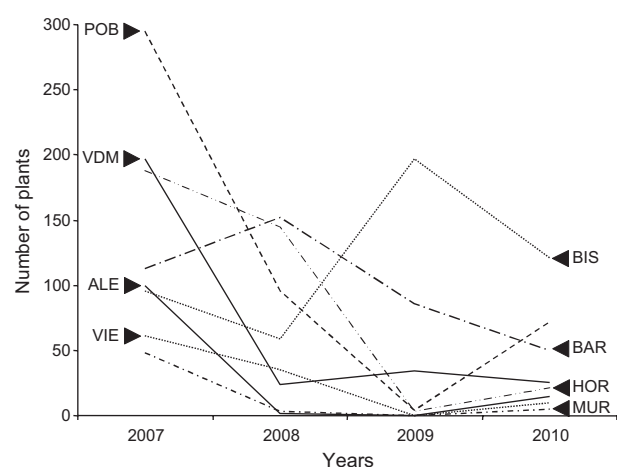


Fig. 1 Number of reproductive *Arabidopsis thaliana* plants recorded in spring at each population within 4-m permanent transects. No plant density estimates were available for the COC population.

Table 3 Mean (\pm SD) percentage of multilocus genotypes observed as a function of the number of study years for each *Arabidopsis thaliana* population (ranked by altitude)

Population	1 year	2 years	3 years	4 years
BAR	30.0 \pm 10.3	55.7 \pm 10.9	81.0 \pm 10.8	100.0
HOR	32.9 \pm 21.2	57.0 \pm 21.0	87.7 \pm 11.0	100.0
COC	58.7 \pm 15.4	100.0	–	–
POB	49.0 \pm 12.2	80.4 \pm 9.0	100.0	–
MUR	41.7 \pm 21.9	75.0 \pm 12.5	100.0	–
VDM	64.3 \pm 10.1	100.0	–	–
ALE	55.6 \pm 21.0	86.1 \pm 4.1	100.0	–
BIS	66.7 \pm 33.3	88.9 \pm 19.2	100.0	–
VIE	43.3 \pm 20.8	73.3 \pm 15.3	100.0	–

For each population, the maximum number of study years corresponds to 100% of the multilocus genotypes observed over the entire study period.

BAR, Barcelona; HOR, Hortsavinyà; COC, Cap de Creus; POB, Poblet; MUR, Mura; VDM, Vilanova de Meià; ALE, Albet; BIS, Bisaurri; VIE, Vielha.

Table 2 Total and private (round brackets) number of multilocus genotypes (N_H) for each *Arabidopsis thaliana* population (ranked by altitude) and sampling year

Population	N_H 2006	N_H 2007	N_H 2008	N_H 2009	N_H total	Difference (%)
BAR	7 (5)	14 (10)	8 (4)	13 (11)	35	14.8 \pm 6.3
HOR	4 (3)	12 (9)	6 (4)	3 (0)	19	18.6 \pm 7.5
COC	–	16 (12)	–	11 (7)	23	26.0 \pm 6.4
POB	–	9 (5)	10 (3)	6 (2)	17	21.8 \pm 5.3
MUR	–	10 (6)	7 (4)	3 (2)	16	14.9 \pm 5.3
VDM	4 (2)	–	5 (3)	–	7	15.9 \pm 4.2
ALE	4 (1)	9 (2)	7 (2)	–	12	15.8 \pm 5.7
BIS	–	1 (0)	3 (1)	2 (0)	3	9.5 \pm 7.7
VIE	2 (1)	5 (3)	6 (4)	–	10	12.9 \pm 10.9

The total number of multilocus genotypes for each population detected over years is also indicated. The mean (\pm SD) percentage of markers that differed among all pairs of nonredundant multilocus genotypes found in each population across years is also given. BAR, Barcelona; HOR, Hortsavinyà; COC, Cap de Creus; POB, Poblet; MUR, Mura; VDM, Vilanova de Meià; ALE, Albet; BIS, Bisaurri; VIE, Vielha.

On average, multilocus genotypes observed in multiple years included an average of 8.9 individuals (range = 2–38), whereas year-specific multilocus genotypes included 1.4 individuals (range = 1–6). The number of individuals per year-specific multilocus genotype observed in each single year did not differ from one another but they all significantly differed from the number of individuals per multilocus genotype observed in more than a year ($F_{4,139} = 18.9$, $P < 0.0001$). High-frequency multilocus genotypes were detected in more than a year, whereas low-frequency ones appeared in a single year (Table S2, Supporting information). On average, the mean number of markers that differed among all multilocus genotypes observed in each population was 16.7% (Table 2).

Altitude was significantly negatively correlated with the mean number of multilocus genotypes observed across years (Fig. 2), but it was not significantly correlated with the coefficient of variation among years for the mean number of multilocus genotypes ($P = 0.70$; Fig. 2). In addition, altitude was significantly negatively correlated with the mean number of differences among multilocus genotypes observed across years (Fig. 2) and significantly positively correlated with the coefficient of variation for this same variable (Fig. 2).

Genetic diversity, differentiation and structure

Percentage of PL, mean number of alleles per locus (n_a), mean allelic richness per locus (R_S), mean gene diversity (H_S) and the proportion of LD (P_D) showed substantial variation among populations and years. The ranges of observed values for these parameters across populations and years were as follows: PL = 0.0–79.5%, $n_a = 1.0$ –1.8 alleles per locus, $H_S = 0.00$ –0.28 and $R_S = 1.00$ –1.76 (Table S1, Supporting information). On average, the proportion of LD estimated as P_D was 50.2% (range = 7.2–100.0%), which was in agreement with P_D estimates from other *Arabidopsis thaliana* populations based on SNPs (Picó *et al.* 2008; Lundemo *et al.* 2009).

Mean values of PL, n_a , H_S and R_S computed for each population across years were all significantly negatively correlated with altitude (Fig. 3). CV among years and altitude were significantly positively correlated for PL and H_S , marginally positively correlated for R_S and not significantly correlated for n_a ($P = 0.19$; Fig. 3).

Average F_{ST} estimates among populations for each year varied between 0.47 and 0.79 ($P < 0.0001$ in all cases; Table 4A). All population-pair F_{ST} values were significant for each year (Table 4A). The average estimates of significant among-year genetic differentiation

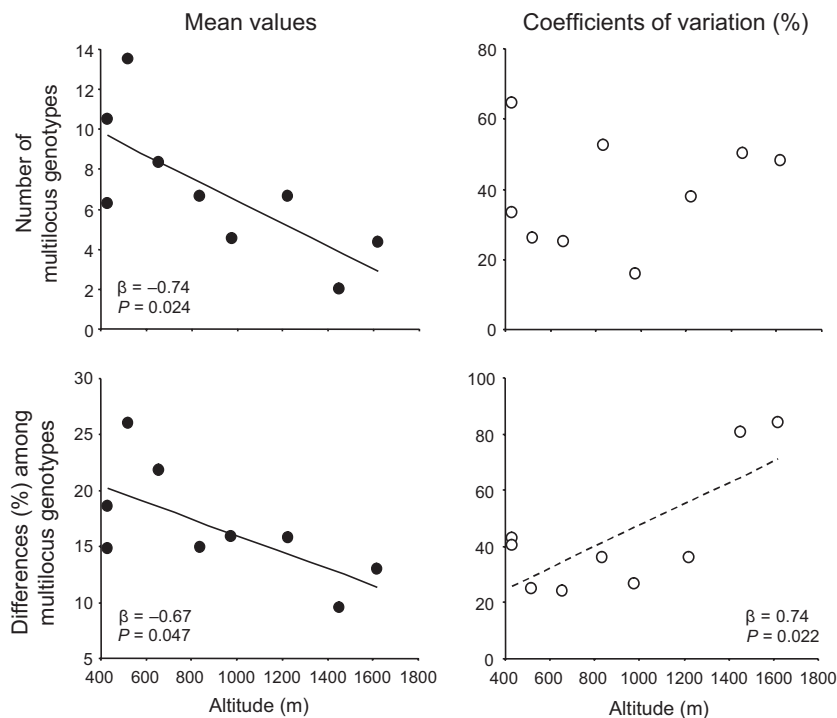


Fig. 2 Relationship between altitude and mean number of multilocus genotypes, per cent differences among multilocus genotypes and their coefficients of variation among years observed in each *Arabidopsis thaliana* population. When significant, regression lines, regression coefficients and their corresponding P -values are also indicated.

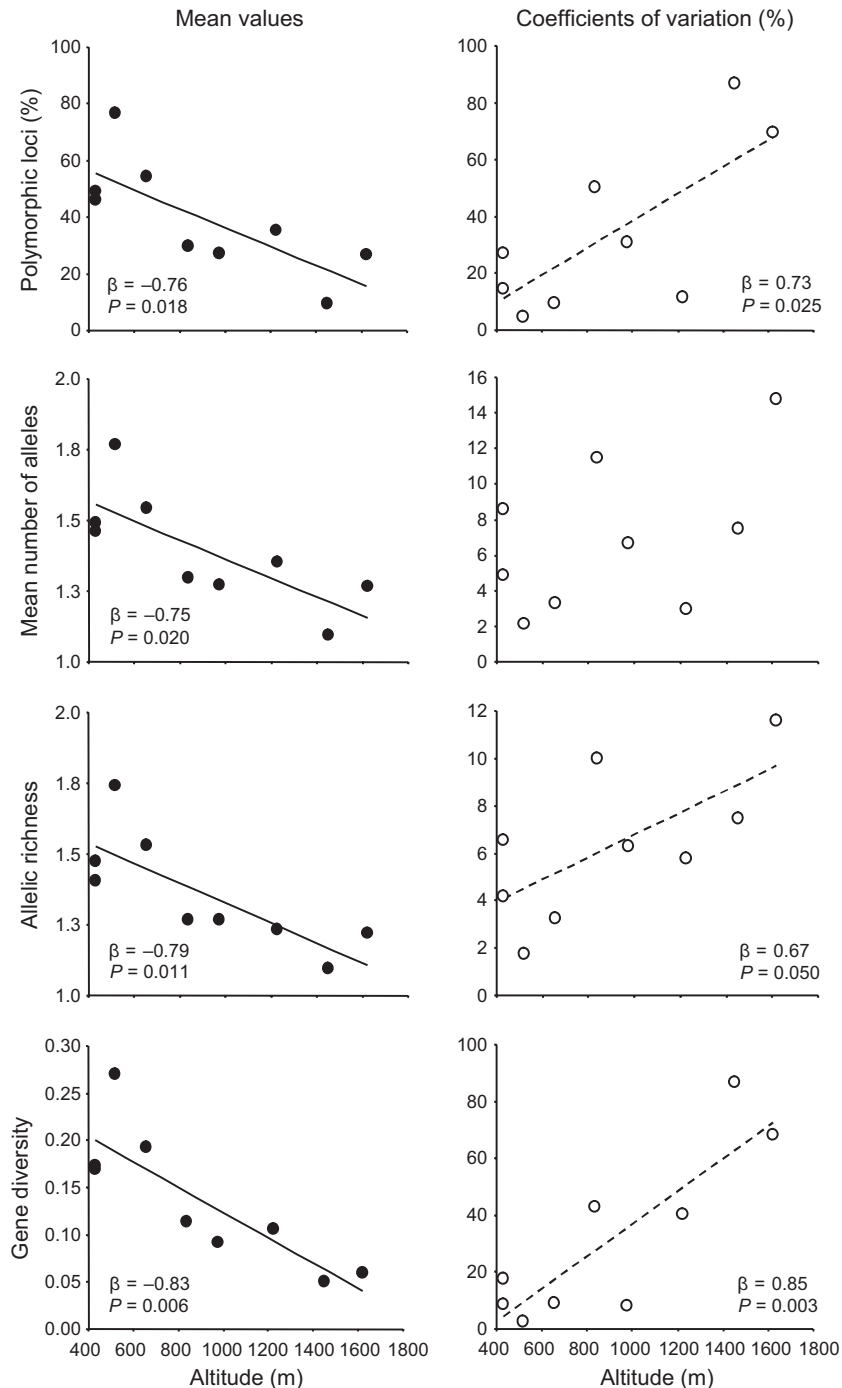


Fig. 3 Relationship between altitude and mean values or coefficients of variation among years for different genetic parameters of *Arabidopsis thaliana* populations. When significant, regression lines, regression coefficients and their corresponding P -values are also indicated.

F_{ST} for each population ranged between 0.13 and 0.26 ($P < 0.04$ in all cases; Table 4B). F_{ST} values among years for VIE, BAR and COC populations were not significant (Table 4B). The proportion of significant population-pair F_{ST} values among years for each population varied between 66.7% and 100.0% (Table 4B). Mean F_{ST} values

computed across years were not significantly correlated with altitude ($P = 0.47$). We did not detect significant correlations between genetic and geographic distances among populations for any year ($P > 0.09$ in all cases).

A total of six genetic clusters were inferred with Structure when using all 142 multilocus genotypes

Table 4 Mean F_{ST} and range of pairwise F_{ST} values for each (A) sampling year and (B) population (ranked by altitude)

(A)				
Year	Number of populations	Mean F_{ST}	Range of pairwise F_{ST}	% Significant pairwise F_{ST}
2006	5	0.792***	0.562–0.858	100.0
2007	8	0.471***	0.204–0.810	100.0
2008	8	0.561***	0.403–0.689	100.0
2009	6	0.480***	0.245–0.779	100.0
(B)				
Population	Number of years	Mean F_{ST}	Range of pairwise F_{ST}	% Significant pairwise F_{ST}
BAR	4	0.066 ns	0.006–0.265	83.3
HOR	4	0.126**	0.068–0.197	100.0
COC	2	0.053 ns	–	–
POB	3	0.145**	0.104–0.211	100.0
MUR	3	0.260***	0.107–0.498	100.0
VDM	2	0.192**	–	–
ALE	3	0.259***	0.041–0.403	100.0
BIS	3	0.264*	0.034–0.470	66.7
VIE	3	0.033 ns	0.015–0.112	66.7

The number of populations and years to compute F_{ST} values is given. The proportion of significant pairwise F_{ST} values is also indicated ($P < 0.022$ in all significant pairwise F_{ST} comparisons).

BAR, Barcelona; HOR, Hortsavinyà; COC, Cap de Creus; POB, Poblet; MUR, Mura; VDM, Vilanova de Meià; ALE, Albet; BIS, Bisaurri; VIE, Vielha.

Significance: *** $P < 0.0001$; ** $P < 0.001$; * $P < 0.05$; ns, nonsignificant.

based on 151 SNPs (Fig. 4). The majority of multilocus genotypes exhibited very high membership proportions in the ancestral genetic clusters (Fig. 4). All low-elevation populations (BAR, HOR, COC, POB and MUR) eventually grouped into different clusters, whereas the four high-elevation populations (VDM, ALE, BIS and VIE) grouped into a single cluster (Fig. 4). Clustering analysis based on the per cent number of markers that differed among all pairs of multilocus genotypes yielded consistent results with those obtained with Structure (Fig. S2, Supporting information).

Effective population size

Both moment and likelihood effective population size (N_e) estimators were highly consistent among populations, although likelihood N_e estimators tended to be higher than moment N_e estimators (Table 5). Moment N_e estimators ranged between one and 12 individuals, whereas likelihood N_e estimators varied between one and 34 individuals (Table 5). VIE exhibited the highest N_e value, probably as a result of the notorious increase in the number of PL between the two sampling years used to compute N_e while keeping low genetic diversity

values (Table S1, Supporting information). When excluding N_e estimators of the VIE population from the analyses, moment N_e estimators were significantly negatively correlated with altitude (Fig. 5), whereas likelihood N_e estimators showed a marginal negative correlation with altitude ($P = 0.07$).

Discussion

Temporal variation in population genetics

A previous genetic study conducted in 2007 on the nine *Arabidopsis thaliana* populations used here revealed that montane populations were genetically less diverse than coastal populations (Montesinos *et al.* 2009). Overall, this conclusion holds when considering genetic data from multiple years generated in this study. This result suggests that 1 year of data might suffice to obtain a realistic snapshot of the genetic composition of *A. thaliana* populations. Other studies also concluded that the lack of temporal replicates may not necessarily represent a major drawback in genetic studies (Tessier & Bernatchez 1999), although further empirical evidence on different study systems and environments is needed to

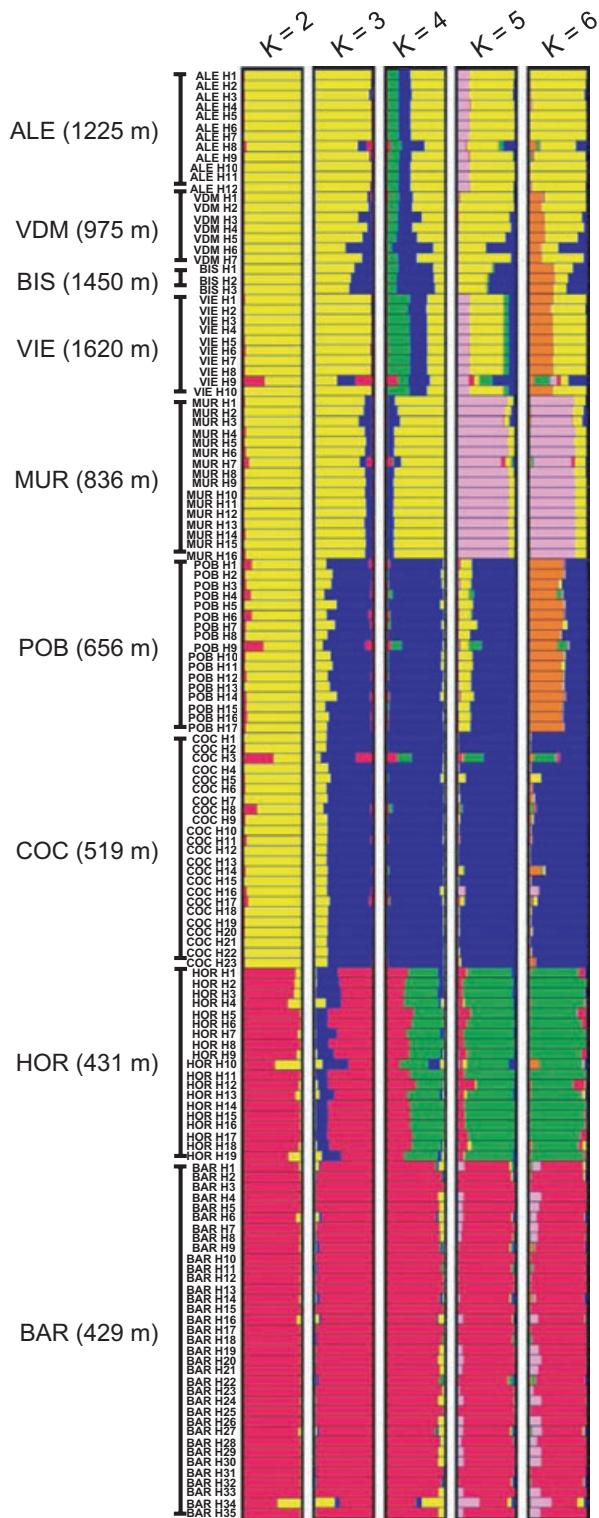


Fig. 4 Genetic structure of *Arabidopsis thaliana* detected across all populations and years estimated with Structure. Each multilocus genotype from each population is depicted as a horizontal line divided into segments representing the estimated membership proportions in the ancestral genetic cluster (K) fitted in the model. Population names and altitude are given.

assess the generalization of this result. However, it is clear from our results that increasing the number of study years increases the potential to detect more genetic variation in *A. thaliana* populations because rare multilocus genotypes mostly appear in a single year. On top of that, the added value of repeated genotyping over time lies in the fact that temporal data provide information not only on mean patterns, but on the extent of variation across years. Assessing the implications of temporal variation in genetic data represents a relevant but uncommon contribution to standard population genetic studies.

Our results show that genetic differentiation among years is below 0.26, which is clearly lower than genetic differentiation observed among populations ($F_{ST} > 0.47$; Table 4). Furthermore, all among-population F_{ST} values for each year were significant, whereas six of nine among-year F_{ST} values for each population were significant. Thus, we conclude that there is greater genetic differentiation among populations than among years in this set of *A. thaliana* populations. This conclusion is in broad agreement with other studies based on two consecutive years of genetic data conducted on *A. thaliana* populations in Norway at a regional scale (Lundemo *et al.* 2009) and in Germany at a local scale (Bomblies *et al.* 2010). In fact, processes accounting for spatial (e.g. historical founder effects) and temporal differentiation (e.g. yearly variation in windows of opportunity for recruitment) operate at very different timescales, which chiefly determine the spatial and temporal patterns of variation in genetic differentiation.

It is widely accepted that migration among *A. thaliana* populations is low and that it poorly accounts for within-population genetic variation in *A. thaliana*. Basically, as a result of novel mutations, outcrossing and recombination within populations, *A. thaliana* populations are generally composed of genotypes that are genetically closely related and strongly differentiated from other genotypes detected in other populations (Stenøien *et al.* 2005; Bakker *et al.* 2006; Picó *et al.* 2008; Lundemo *et al.* 2009; Montesinos *et al.* 2009; Bomblies *et al.* 2010). The evaluation of the genetic relationships among all multilocus genotypes found across years also supports this pattern, especially for low-altitude populations (Fig. 4). However, although high-altitude populations do not share any multilocus genotype, they all group together into a single cluster (Fig. 4). This finding suggests that different subregional long-term historical processes have been operating at low- and high-elevation populations. Nevertheless, we cannot discard that the inherent lower genetic diversity found among high-elevation populations reduced the power to detect the genetic structure of these populations.

Table 5 Moment and likelihood estimates ($\pm 95\%$ confidence intervals) of the effective population size (N_e) for each *Arabidopsis thaliana* population (ranked by altitude)

Population	Sampling years	Number individuals	Number loci	Moment N_e estimator	Likelihood N_e estimator
BAR	2006–2009	28	102	4 (2.5–4.8)	5 (4.2–6.2)
HOR	2006–2009	29	88	6 (3.9–8.4)	9 (7.1–11.1)
COC	2007–2009	38	137	7 (4.9–9.4)	7 (5.6–8.4)
POB	2007–2009	41	108	2 (1.5–2.8)	2 (1.8–2.5)
MUR	2007–2009	35	68	2 (1.3–2.7)	1 (0.7–1.2)
VDM	2006–2008	39	62	3 (2.0–4.6)	5 (3.6–6.1)
ALE	2006–2008	37	76	2 (1.1–2.2)	2 (1.6–2.5)
BIS	2007–2009	26	23	1 (0.4–1.4)	2 (1.4–2.7)
VIE	2006–2008	32	75	12 (6.2–28.8)	34 (21.5–61.9)

The two sampling years, the number of individuals and the number of polymorphic SNP loci used to compute N_e are given. Likelihood N_e values are the average of five simulations.

BAR, Barcelona; HOR, Hortsavinyà; COC, Cap de Creus; POB, Poblet; MUR, Mura; VDM, Vilanova de Meià; ALE, Albet; BIS, Bisaurri; VIE, Vielha.

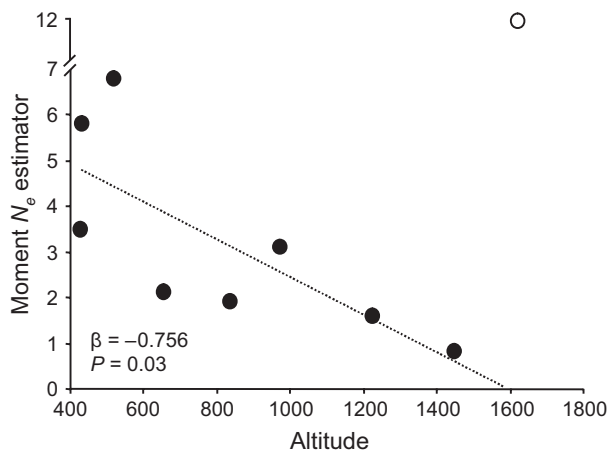


Fig. 5 Relationship between altitude and the moment effective population size (N_e) estimates computed for each *Arabidopsis thaliana* population. The correlation coefficient and its corresponding P -value were computed after excluding the VIE population (hollow dot).

Our study also finds temporal fluctuations in the number of multilocus genotypes in all populations (Table 2). In addition, multilocus genotypes with a high frequency within a population were mostly detected in more than 1 year, but none of the genotypes, regardless of their frequency and year of occurrence, was observed in more than one population. Given that for each population we strictly sampled plants within a restricted area every year and no identical multilocus genotype was found in different populations, all low-frequency multilocus genotypes observed in one single year likely come from the genetic reservoir of the soil seed bank, as suggested to occur in other *A. thaliana* populations by Bomblies *et al.* (2010) and elegantly tested by Lun-

demo *et al.* (2009). In fact, field experiments have estimated that the half-life of *A. thaliana* seeds in the soil seed bank is about 3 years (Montesinos *et al.* 2009), which illustrates the potential of *A. thaliana* seed bank to recruit new individuals, and genotypes, into the population over time.

Clinal variation in population genetic metrics

The *A. thaliana* populations of study are located along an altitudinal gradient that represents a proxy for a climatic gradient in which precipitation increases and temperature decreases with altitude (Montesinos *et al.* 2009). Our results also provide evidence for clinal variation for genetic parameters not only when considering mean values but also temporal variation. For example, the mean number of multilocus genotypes decreases with altitude but the coefficient of variation among years in the number of multilocus genotypes does not vary with altitude (Fig. 2). This result suggests that environmental factors associated with altitude, with important effects on *A. thaliana* demographic (Montesinos *et al.* 2009) and life cycle traits (Montesinos-Navarro *et al.* 2011), influence the total amount of genetic variation encompassed at each population, expressed as the number of multilocus genotypes, but it does not affect the temporal dynamics of appearance and disappearance of such genotypes.

Furthermore, mean differences among multilocus genotypes decrease with altitude (Fig. 2). We hypothesize that the higher similarity among multilocus genotypes at high-elevation populations could be accounted for by different factors that may act in concert: stronger founder effects, higher outcrossing rates or stronger selection in montane sites that overall would reduce differences among multilocus genotypes. In contrast,

the coefficient of variation among years for the mean differences among multilocus genotypes increases with altitude (Fig. 2). However, this relationship is strongly determined by the particularities of two populations: BIS and VIE (Fig. 2). BIS is composed of three multilocus genotypes, two of them only differing in a single marker, and VIE shows a group of four of 10 multilocus genotypes only differing in 1–12 markers (results not shown). This enormously increases the coefficient of variation in the mean differences among multilocus genotypes in these populations.

As mentioned previously, mean values for the main genetic diversity parameters (PL , n_a , R_S and H_S) decrease with altitude. Interestingly, the CV for some of the important parameters (PL and H_S) also increase with altitude. We hypothesize that the lower temporal variation in genetic diversity metrics at low-elevation populations can be the result of an overall higher demographic stability in low-elevation environments. In general, it is accepted that overall environmental stability maintains genetic diversity (Carnaval *et al.* 2009) and that demographic stability and genetic variation may be tightly correlated (Booy *et al.* 2000; Spielman *et al.* 2004; Finn *et al.* 2009). Research underway on the long-term spatiotemporal demography of all *A. thaliana* populations of study (F.X. Picó, unpublished) will provide insights into the extent of demographic stability along the altitudinal climatic gradient to test this hypothesis.

Effective population size

A recently published study based on temporal changes in allele frequency for other *A. thaliana* natural populations reported N_e values ranging between a few hundreds and a few thousands (Lundemo *et al.* 2009). In contrast, our N_e values for the *A. thaliana* populations of study varied between one and 34. Our N_e values were based on biallelic SNP markers, whereas those of Lundemo *et al.* (2009) were estimated using multiallelic microsatellites as well as a different maximum-likelihood approach for estimating θ ($\theta = 4N_e\mu$, where μ is the mutation rate per generation), which overall make comparisons between studies difficult to evaluate. The biallelic nature of SNPs could be one of the main factors accounting for such low N_e values owing to their effect on allelic frequencies. Another reason could be the conservative sampling scheme applied in this study. We only sampled plants within a restricted sampling area at each population and year, so we did not pool genotypes from different areas within each population in different years. Given the high genetic heterogeneity found among closely located *A. thaliana* stands (Bombles *et al.* 2010), pooling genotypes from different areas would increase temporal changes in allele frequencies

and subsequently N_e values. Hence, our N_e estimates refer to the sampling areas within each population. Furthermore, Lundemo *et al.* (2009) estimated an average effective generation time of 3–4 years for *A. thaliana*. If this also applies to our populations of study, we might then be using overlapping generations because the number of years between samples ranges 2–3 in our study, and overlapping generations have been shown to bias N_e (Waples & Yokota 2007; Wang *et al.* 2010). Finally, the selfing habit of *A. thaliana* can also account for the low N_e values found in this study. In the field, individuals can produce dozens or hundreds of selfed seeds each (F.X. Picó, unpublished). Given that *A. thaliana* populations are composed of a few high-frequency genotypes and closely related low-frequency genotypes, the net number of different genotypes contributing to the next generation is likely to be low.

The significant negative relationship between N_e and altitude reveals a novel pattern that has not previously been described. An assumption of the two approaches applied in this study to estimate N_e is that natural selection, mutation and migration have a very low or null effect. Therefore, the trend for decreasing N_e with increasing altitude can be interpreted in terms of higher random genetic drift occurring at higher altitudes. Such random genetic drift at high-elevation *A. thaliana* populations could be caused by more frequent and sharper demographic fluctuations producing bottlenecks and/or increasing variance in reproductive output of genotypes, which eventually account for lower N_e values (Wang 2005).

Methodological limitations

We want to stress some of the caveats inherent to this sort of empirical studies that might affect some of our results. First of all, although we first checked out the effects of sample size on genetic parameter estimation, the sample size used in this study (13–21 individuals) could still be small compared with the real census size at our delimited study areas. Hence, our potential to observe rare multilocus genotypes in multiple years could be limited, affecting the accuracy in which gene frequencies change over time. In this sense, our suggestion is to increase sample size when assessing the genetic diversity of *A. thaliana* populations to avoid possible problems related to low sampling size relative to high population size. It must be emphasized that repeated genotyping over time could also represent a method to decrease bias because of low sample sizes relative to census sizes given the potential of repeated genotyping to capture a higher amount of genetic variation.

Second, estimating accurately population size turns out to be a difficult task in natural *A. thaliana* populations.

In fact, we failed in our attempt to relate temporal changes in plant density within transects and temporal changes in genetic diversity. The spatial distribution of clumps may easily change from year to year, and our permanent transects could not capture the real dynamics of the whole delimited study area sampled. In other words, representative transects in 2007 when they were laid down could be no longer representative in 2010. On top of that, temporal variability in environmental conditions strongly affects *A. thaliana* performance. This has important implications in the number and spatial position of plants that can be sampled and genotyped every year. This explains why for some years we were not able to obtain genetic data despite the fact that sampling areas bore plants in those years.

Conclusions

Overall, our results support the view that increasing temporal variation in population genetic metrics with increasing altitude is the result of demographic processes: higher plant mortality in winter and higher secondary seed dormancy in spring at high-elevation populations decrease plant numbers (Montesinos *et al.* 2009) that in turn increase demographic stochasticity and genetic drift. Overall, this can be translated into higher temporal variation in the genetic attributes of populations. Although we expect selective pressures to be stronger at high-elevation populations, as a matter of fact, we ignore the nature of all possible selective pressures along the altitudinal climatic gradient and their strength and consistency over time. Clearly, further research is needed to improve our current knowledge on the selective pressures that may be operating in *Arabidopsis thaliana* populations at different native environments.

Assessing the extent of temporal variation in population genetic metrics provides the opportunity to gather additional information on the genetic characteristics of populations, which is quite relevant and meaningful for population biology studies. Traditional demographic studies are entirely based on spatiotemporal variation in demographic vital rates, temporal variation being particularly relevant for long-term demographic assessments (Tuljapurkar & Orzack 1980; Caswell 2001). Based on our results, we conclude that genetic studies should also be based on spatiotemporal variation in genetic parameters. The key point, which will depend on the specific questions to be addressed, is to decide the adequate time lag to sample and genotype populations to maximize the impact of results but also to economize resources. If yearly samplings are not possible because of economic or logistic reasons, the time lag between sampling events can be determined by the estimated organism's generation time to minimize the effects of overlapping genera-

tions and avoid underestimating important population genetic attributes, such as N_e .

The usefulness of repeated genotyping is enhanced when populations of study are selected according to specific designs across a given geographical region. In our case, *A. thaliana* populations were distributed along an altitudinal climatic gradient in NE Spain, which has been reported to affect demographic attributes (Montesinos *et al.* 2009) and life cycle adaptations (Montesinos-Navarro *et al.* 2011). This is especially relevant when genetic data are used to estimate N_e by means of temporal changes in allele frequency. Given that N_e is notoriously difficult to estimate regardless of the methodological approach used for it (Wang 2005), the biological significance of N_e estimates arises from the comparison among different ecological and/or evolutionary scenarios. Such comparative approach provides additional insights into the meaning and implications of N_e estimates.

Acknowledgements

We are grateful to the staff of the laboratory of molecular ecology (LEM) of the EBD-CSIC and the staff of the greenhouse of the IBVF-CSIC for assistance. IBVF-CSIC granted us permission to use the greenhouse. We thank M. Abdelaziz, C. García and J. Muñoz for their insightful comments on earlier versions of the manuscript. Two anonymous reviewers also contributed to improve the manuscript with valuable comments. PF Quintana-Ascencio, E Caballero and SJ Tonsor made suggestions and help in the field at different years. Funding was provided by Ministerio de Ciencia e Innovación of Spain (Grants CGL2006-09792/BOS and CGL2009-07847/BOS) and CSIC (Grant 200630I255) to FXP and Ministry of Higher Education and State of Scientific Research of Egypt (ParOwn Grant 1207 Cycle) to NHG.

References

- Abbott RJ, Gomes MF (1989) Population genetic structure and outcrossing rate in *Arabidopsis thaliana* (L.) Heynh. *Heredity*, **62**, 411–418.
- Anderson EC, Williamson EG, Thompson EA (2000) Monte Carlo evaluation of the likelihood for N_e from temporally spaced samples. *Genetics*, **156**, 2109–2118.
- Antao T, Pérez-Figueroa A, Luikart G (2010) Early detection of population declines: high power of genetic monitoring using effective population size estimators. *Evolutionary Applications*, **4**, 144–154.
- Avise JC (2004) *Molecular Markers, Natural History, and Evolution*, 2nd edn. Sinauer Associates Inc., Sunderland, USA.
- Bakker EG, Stahl EA, Toomajian C, Nordborg M, Kreitman M, Bergelson J (2006) Distribution of genetic variation within and among local populations of *Arabidopsis thaliana* over its species range. *Molecular Ecology*, **15**, 1405–1418.
- Baskin JM, Baskin CC (1972) Ecological life cycle and physiological ecology of seed germination of *Arabidopsis thaliana*. *Canadian Journal of Botany*, **50**, 353–360.

- Baskin JM, Baskin CC (1983) Seasonal changes in the germination responses of buried seeds of *Arabidopsis thaliana* and ecological interpretation. *Botanical Gazette*, **144**, 540–543.
- Bergelson J, Stahl E, Dudek S, Kreitman M (1998) Genetic variation within and among populations of *Arabidopsis thaliana*. *Genetics*, **148**, 1311–1323.
- Bernartzky R, Tanksley S (1986) Genetics of acting-related sequences in tomato. *Theoretical and Applied Genetics*, **72**, 314–324.
- Berthier P, Beaumont MA, Cornuet JM, Luikart G (2002) Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics*, **160**, 741–751.
- Bombliès K, Yant L, Laitinen RA *et al.* (2010) Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *PLoS Genetics*, **6**, e1000890.
- Booy G, Hendriks RJJ, Smulders MJM, Van Groenendael JM, Vosman B (2000) Genetic diversity and the survival of populations. *Plant Biology*, **2**, 379–395.
- Caballero A (1994) Developments in the prediction of effective population size. *Heredity*, **73**, 657–679.
- Carnaval AC, Hickerson MJ, Haddad CFB, Rodrigues MT, Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, **323**, 785–789.
- Caswell H (2001) *Matrix Population Models*, 2nd edn. Sinauer Associates Inc., Sunderland, USA.
- Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.
- Excoffier L, Laval G, Scheneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.
- Finn DS, Bogan MT, Lytle DA (2009) Demographic stability metrics for conservation prioritization of isolated populations. *Conservation Biology*, **23**, 1185–1194.
- Fisher RA (1930) *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford, UK.
- Frankham R (2005) Genetics and extinction. *Biological Conservation*, **126**, 131–140.
- García MB, Picó FX, Ehrlén J (2008) Life span correlates with population dynamics in perennial herbaceous plants. *American Journal of Botany*, **95**, 258–262.
- Goodman D (1987) The demography of chance extinction. In: *Viable Populations for Conservation* (ed. Soulé ME), pp. 11–34. Cambridge University Press, Cambridge, UK.
- Goudet J (1995) FSTAT: a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.
- Hoffmann MH (2002) Biogeography of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae). *Journal of Biogeography*, **29**, 125–134.
- Honnay O, Jacquemyn H, Van Looy K, Vandepitte K, Breyne P (2009) Temporal and spatial genetic variation in a metapopulation of the annual *Erysimum cheiranthoides* on stony river banks. *Journal of Ecology*, **97**, 131–141.
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.
- Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation-by-distance, web service. *BMC Genetics*, **6**, 13.
- Keller LF, Jeffery KJ, Arcese P *et al.* (2001) Immigration and the ephemerality of a natural population bottleneck: evidence from molecular markers. *Proceedings of the Royal Society B: Biological Sciences*, **268**, 1387–1394.
- Krimbas CB, Tsakas S (1971) The genetics of *Daucus oleae* V. Changes of esterase polymorphism in a natural population following insecticide control: selection or drift? *Evolution*, **25**, 454–460.
- Lande R (1988) Genetics and demography in biological conservation. *Science*, **241**, 1455–1460.
- Luikart G, Cornuet JM, Allendorf FW (1999) Temporal changes in allele frequencies provide estimates of population bottleneck size. *Conservation Biology*, **13**, 523–530.
- Lundemo S, Falahati-Anbaran M, Stenøien HK (2009) Seed banks cause elevated generation times and effective population sizes of *Arabidopsis thaliana* in northern Europe. *Molecular Ecology*, **18**, 2798–2811.
- Lynch M, Conery J, Burger R (1995) Mutation accumulation and the extinction of small populations. *American Naturalist*, **146**, 489–518.
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **104**, 209–220.
- Menges ES (2000) Population viability analyses in plants: challenges and opportunities. *Trends in Ecology and Evolution*, **15**, 51–56.
- Mhemmed G, Kamel H, Chedly A (2008) Does habitat fragmentation reduce genetic diversity and subpopulation connectivity? *Ecography*, **31**, 751–756.
- Montesinos A, Tonsor SJ, Alonso-Blanco C, Picó FX (2009) Demographic and genetic patterns of variation among populations of *Arabidopsis thaliana* from contrasting native environments. *PLoS ONE*, **4**, e2713.
- Montesinos-Navarro A, Wig J, Picó FX, Tonsor SJ (2011) *Arabidopsis thaliana* populations show clinal variation in a climatic gradient associated with altitude. *New Phytologist*, **189**, 282–294.
- Nei M, Tajima F (1981) Genetic drift and estimation of effective population-size. *Genetics*, **98**, 625–640.
- Nordborg M, Hu TT, Ishino Y *et al.* (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biology*, **3**, e196.
- Osborne MJ, Davenport SR, Hoagstrom CW, Turner TF (2010) Genetic effective size, N_e , tracks density in a small freshwater cyprinid, Pecos bluntnose shiner (*Notropis simus pecosensis*). *Molecular Ecology*, **19**, 2832–2844.
- Palstra FP, O'Connell MF, Ruzzante DE (2009) Age structure, changing demography and effective population size in Atlantic Salmon (*Salmo salar*). *Genetics*, **182**, 1233–1249.
- Peel D, Ovenden JR, Peel SL (2004) *NeEstimator: Software for Estimating Effective Population Size, Version 1.3*. Queensland Government, Department of Primary Industries and Fisheries, Brisbane, Australia.
- Picó FX, Méndez-Vigo B, Martínez-Zapater JM, Alonso-Blanco C (2008) Natural genetic variation of *Arabidopsis thaliana* is geographically structured in the Iberian Peninsula. *Genetics*, **180**, 1009–1021.
- Picó FX, Quintana-Ascencio PF, Mildén M, Ehrlén J, Pflingsten I (2009) Modelling the effects of genetics and habitat on the demography of a grassland herb. *Basic and Applied Ecology*, **10**, 122–130.

- Platt A, Horton M, Huang YS *et al.* (2010) The scale of population structure in *Arabidopsis thaliana*. *PLoS Genetics*, **6**, e1000843.
- Pollack E (1983) A new method for estimating the effective population size from allele frequency changes. *Genetics*, **104**, 531–548.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Richards CM, Emery SN, McCauley DE (2003) Genetic and demographic dynamics of small populations of *Silene latifolia*. *Heredity*, **90**, 181–186.
- Schaffer HE, Yardley D, Anderson WW (1977) Drift or selection: a statistical test of gene frequency variation over generations. *Genetics*, **87**, 371–379.
- Schmid KJ, Törjék O, Meyer R, Schmutz H, Hoffmann MH, Altmann T (2006) Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. *Theoretical and Applied Genetics*, **112**, 1104–1114.
- Schwartz MK, Luikart G, Waples RS (2006) Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology and Evolution*, **22**, 25–33.
- Siol M, Bonnin I, Olivieri I, Prosperi JM, Ronfort J (2007) Effective population size associated with self-fertilization: lessons from temporal changes in allele frequencies in the selfing annual *Medicago truncatula*. *Journal of Evolutionary Biology*, **20**, 2349–2360.
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.
- Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 15261–15264.
- Stenøien HK, Sæstad SM (1999) Genetic structure in three haploid peat mosses (*Sphagnum*). *Heredity*, **82**, 391–400.
- Stenøien HK, Fenster CB, Tonteri A, Savolainen O (2005) Genetic variability in natural populations of *Arabidopsis thaliana* in northern Europe. *Molecular Ecology*, **14**, 137–148.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.
- Tessier N, Bernatchez L (1999) Stability of population structure and genetic diversity across generations assessed by microsatellites among sympatric populations of landlocked Atlantic salmon (*Salmo salar* L.). *Molecular Ecology*, **8**, 169–179.
- Tuljapurkar SD, Orzack SH (1980) Population dynamics in variable environments. 1. Long-run growth rates and extinction. *Theoretical Population Biology*, **18**, 314–342.
- Wang J (2001) A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genetics Research*, **78**, 243–257.
- Wang J (2005) Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **360**, 1395–1409.
- Wang J, Brekke P, Huchard E, Knapp LA, Cowlshaw G (2010) Estimation of parameters of inbreeding and genetic drift in populations with overlapping generations. *Evolution*, **64**, 1704–1718.
- Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, **121**, 379–391.
- Waples RS, Yokota M (2007) Temporal estimates of effective population size in species with overlapping generations. *Genetics*, **175**, 219–233.
- Warthmann N, Fitz J, Weigel D (2007) MSQT for choosing SNP assays from multiple DNA alignments. *Bioinformatics*, **23**, 2784–2787.
- Weir B, Cockerham C (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Williamson EG, Slatkin M (1999) Using maximum likelihood to estimate population size from temporal changes in allele frequencies. *Genetics*, **152**, 755–761.
- Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.
- Zhang X-S, Wang J, Hill WG (2004) Redistribution of gene frequency and changes of genetic variation following a bottleneck in population size. *Genetics*, **167**, 1475–1492.

N.H.G. is a botanist interested in the population biology of desert plants. A.M-N. recently obtained a PhD on life cycle variation in *Arabidopsis thaliana*. C.A-B. is a genetist specialized in the genetic and molecular bases of natural variation of plant development. F.X.P. is an ecologist interested in population ecology and population genetics of plants.

Data accessibility

Data deposited at Dryad: doi:10.5061/dryad.fv348.

Supporting information

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

Table S1 Mean (\pm SD) values of the percentage of polymorphic loci (PL), mean number of alleles per locus (n_a), allelic richness per locus (R_s), and gene diversity (H_s) for each *A. thaliana* population and year.

Table S2 Number of *A. thaliana* individuals (N) included in each multilocus genotype (N_H) over the whole study period, and number of years (yrs) in which each multilocus genotype was observed.

Fig. S1 Relationship between observed and mean simulated data for the genetic parameters (percent polymorphic loci, mean number of alleles per locus, gene diversity, allelic richness and genetic differentiation among years) computed for each *A. thaliana* population and year.

Fig. S2 UPGMA cluster of *A. thaliana* multilocus genotypes found across populations and years.

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