

Molecular Proxies for Climate Maladaptation in a Long-Lived Tree (*Pinus pinaster* Aiton, Pinaceae)

Juan-Pablo Jaramillo-Correa,^{*,†} Isabel Rodríguez-Quilón,^{*} Delphine Grivet,^{*} Camille Lepoittevin,^{*,§} Federico Sebastiani,^{**} Myriam Heuertz,^{*,*§} Pauline H. Garnier-Géré,^{*,§} Ricardo Alía,^{*} Christophe Plomion,^{*,§} Giovanni G. Vendramin,^{**} and Santiago C. González-Martínez^{*,††,1}

^{*}Department of Ecology and Genetics, Forest Research Centre, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, E-28040 Madrid, Spain, [†]Department of Evolutionary Ecology, Institute of Ecology, Universidad Nacional Autónoma de México, AP 70-275 México City, Mexico, [‡]Institut National de la Recherche Agronomique, Unité Mixte de Recherche 1202 (Biodiversité Gènes Ecosystèmes; Biogeco), F-33610 Cestas, France, [§]University of Bordeaux, Unité Mixte de Recherche 1202 Biogeco, F-33170 Talence, France, ^{**}Institute of Biosciences and Bioresources, National Research Council, I-50019 Sesto Fiorentino, Florence, Italy, and ^{††}Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland

ABSTRACT Understanding adaptive genetic responses to climate change is a main challenge for preserving biological diversity. Successful predictive models for climate-driven range shifts of species depend on the integration of information on adaptation, including that derived from genomic studies. Long-lived forest trees can experience substantial environmental change across generations, which results in a much more prominent adaptation lag than in annual species. Here, we show that candidate-gene SNPs (single nucleotide polymorphisms) can be used as predictors of maladaptation to climate in maritime pine (*Pinus pinaster* Aiton), an outcrossing long-lived keystone tree. A set of 18 SNPs potentially associated with climate, 5 of them involving amino acid-changing variants, were retained after performing logistic regression, latent factor mixed models, and Bayesian analyses of SNP–climate correlations. These relationships identified temperature as an important adaptive driver in maritime pine and highlighted that selective forces are operating differentially in geographically discrete gene pools. The frequency of the locally advantageous alleles at these selected loci was strongly correlated with survival in a common garden under extreme (hot and dry) climate conditions, which suggests that candidate-gene SNPs can be used to forecast the likely destiny of natural forest ecosystems under climate change scenarios. Differential levels of forest decline are anticipated for distinct maritime pine gene pools. Geographically defined molecular proxies for climate adaptation will thus critically enhance the predictive power of range-shift models and help establish mitigation measures for long-lived keystone forest trees in the face of impending climate change.

KEYWORDS climate adaptation; environmental associations; genetic lineages; single nucleotide polymorphisms; fitness estimates

PAST and present climate changes are major drivers of species displacements and range-size variation (Hughes 2000; Franks and Hoffmann 2012). Current predictions indicate that the impact of climate change will intensify over the next 20–100 years (Loarie *et al.* 2009; Bellard *et al.* 2012), with concomitant phenotypic and genetic effects on wild populations (Gamache and Payette 2004; Franks and Hoffmann 2012; Alberto *et al.* 2013a). The capability of species to respond to such alterations will rely on phenotypic plasticity,

potential for *in situ* adaptation, and/or migration to more suitable habitats (Aitken *et al.* 2008). While phenotypic plasticity and migration might be insufficient to cope with these changes (McLachlan *et al.* 2005; Malcom *et al.* 2011; Zhu *et al.* 2011), successful *in situ* adaptation will depend on the amount of standing genetic variation and the rate at which new alleles arise, are maintained, and/or get to fixation within populations (Hancock *et al.* 2011). Thus, our ability to detect present adaptive polymorphisms and to integrate them in predictive models of future maladaptation might be decisive to ensure the persistence of natural populations under climate change, particularly for keystone taxa (Franks and Hoffmann 2012; Kremer *et al.* 2012).

“Maladaptation” to climate refers to a decrease of the mean population fitness produced by a mismatch between the optimal and realized mean genotype frequencies, which may result from the inability to adjust to rapidly changing

Copyright © 2015 by the Genetics Society of America

doi: 10.1534/genetics.114.173252

Manuscript received July 2, 2014; accepted for publication December 20, 2014; published Early Online December 29, 2014.

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.173252/-/DC1>.

Nuclear microsatellite and SNP genotypes from this article have been deposited with the Dryad Digital Repository under DOI 10.5061/dryad.fb436.

¹Corresponding author: Department of Ecology and Genetics, Forest Research Centre, INIA-CIFOR, E-28040 Madrid, Spain. E-mail: santiago@inia.es

climates (Lynch and Lande 1993; Kremer *et al.* 2012). At the species range scale, such adaptation lags can generate mosaics of selected alleles and increase population differentiation at selected loci, depending on the species' life-history traits and the geographic extent at which selective pressures are operating (Savolainen *et al.* 2007; Hancock *et al.* 2011). For example, in *Arabidopsis thaliana*, a model annual selfing plant, new advantageous mutations associated with fitness and local climate occur in such mosaics, together with more common and widely distributed favorable variants (Hancock *et al.* 2011). On the other hand, evidence suggests that in widespread outcrossing species with long generation times, such as forest trees, local adaptations mostly develop from alleles already present in the gene pools (*i.e.*, from standing genetic variation), which often results in the establishment of gene clines along environmental gradients (Savolainen *et al.* 2007; Eckert *et al.* 2010a; Chen *et al.* 2012; Prunier *et al.* 2012; Alberto *et al.* 2013b). Compared to well-studied selfing annuals, forest trees have to face highly variable selection pressures that result from environmental changes over long periods of time. It follows then that the fitness consequences of molecular maladaptation to climate have to be explored in such long-lived outcrossing taxa.

Many characteristics of forest trees, particularly conifers, make them ideal systems for studying climate (mal)adaptation. They are distributed in recurrently shifting geographic ranges, whose changes can be traced back in time through paleobotanical or phylogeographic studies (Petit and Hampe 2006), they exhibit large and significant differences in adaptive phenotypic traits in common garden experiments (Rehfeldt *et al.* 1999), and they bear low levels of proximal linkage disequilibrium (that decays within gene limits for most species; Brown *et al.* 2004; Heuertz *et al.* 2006). This last feature facilitates the identification of polymorphisms associated with adaptive traits (Neale and Savolainen 2004; González-Martínez *et al.* 2007, 2008), although it also implies that a large number of markers is needed to saturate the genome and adequately capture the genomic signals of adaptation. From this point of view, candidate-gene approaches are particularly attractive as they provide direct links with gene functions and allow targeting potential adaptive traits, gene networks, and/or selection drivers without the need for scanning the whole genome (Neale and Savolainen 2004; Neale and Ingvarsson 2008).

Here, we aimed to identify genetic polymorphisms related to climate adaptation in maritime pine (*Pinus pinaster* Ait.), a long-lived outcrossing forest tree, and used a common garden evaluated for survival under extreme (hot and dry) climatic conditions to validate SNP–climate associations and demonstrate the utility of these genetic markers to predict maladaptation to future climates. Maritime pine is an economically important conifer that forms large populations in southwestern Europe, inhabiting both wet-coastal and seasonally dry continental forests, and that exhibits a strong population structure associated with past climate and demographic changes (Bucci *et al.* 2007; Santos-Del-Blanco *et al.* 2012). Once this phylogeographic structure was accounted for to avoid spurious allele-frequency clines generated by historical factors, we

observed that mean and extreme summer and winter temperatures and winter rainfall were important ecological drivers for adaptation in this species. The spatial distribution of the selected alleles suggested that adaptive forces operate on the standing genetic variation at different scales and that they depend on the climatic heterogeneity experienced by each gene pool. Reduced average population fitness was observed in the common garden as the frequency of locally advantageous alleles diminished. This suggested that contrasting levels of future forest decline may occur in distinct maritime pine gene pools confronted with new (*i.e.*, more arid) climates.

Materials and Methods

Sampling, SNP genotyping, and scoring

Needles were collected from 772 individuals distributed in 36 natural populations across the maritime pine range. Total DNA was extracted with the Invisorb DNA Plant HTS 96 kit following the manufacturer's instructions. Samples were genotyped with two oligo pool assays (OPA) that included 2646 and 384 SNPs by using the Illumina Infinium and Illumina VeraCode platforms, respectively. The first assay contained randomly chosen SNPs detected *in silico* from transcriptome sequence data and are expected to represent neutral variants (hereafter, control SNPs; see Chancerel *et al.* 2013 for more details). Briefly, SNPs were identified from >600,000 sequences obtained from a set of 17 cDNA libraries of different tissues without any prior experimental treatment and were integrated into a unigene set previously used for linkage mapping (Chancerel *et al.* 2013).

The second OPA comprised 384 SNPs distributed in 221 candidate genes (Supporting Information, File S2; see also Budde *et al.* 2014) that included drought-stress candidates identified in Mediterranean (*P. pinaster* and *P. halepensis*) and American (*P. taeda*) pines (Eveno *et al.* 2008; Eckert *et al.* 2010a,b; Grivet *et al.* 2011) and genes overexpressed under abiotic stress (Lorenz *et al.* 2011; Perdiguero *et al.* 2013) and/or that have shown some evidence of adaptive value in different conifers (*e.g.*, González-Martínez *et al.* 2007, 2008; Eveno *et al.* 2008; Eckert *et al.* 2010a,b; Grivet *et al.* 2011; Lepoittevin *et al.* 2012; Mosca *et al.* 2012). SNPs were selected after resequencing these genes in a discovery panel that covered the entire natural range of *P. pinaster* (see Chancerel *et al.* 2011; Budde *et al.* 2014).

For both OPAs, loci in Hardy–Weinberg equilibrium, bearing minor allele frequencies (MAF) >5% and call rates >0.8 (average of 0.97) were retained (1745 and 266 control and candidate SNPs, respectively). Scoring was performed with BeadStudio v. 3.2 (Illumina Inc.) after manual adjustment of genotype clusters. To ensure that genotypes were correctly called, three high-quality DNAs included as positive controls in each genotyped plate were doubled checked for all SNPs.

Microsatellite genotyping and phylogeographic analyses

To corroborate that historical and demographic processes were adequately accounted for, 12 nuclear microsatellites (SSRs) distributed in 8 of the 12 linkage groups of the species (Chancerel *et al.* 2011; de-Miguel *et al.* 2012) were genotyped following Santos-Del-Blanco *et al.* (2012) and references therein. However, three of them (*epi6*, *ctg275*, *NZPR544*) were excluded from further analyses after detecting high frequencies of null alleles, which led to significant deviations from Hardy–Weinberg expectations in more than half of the populations (see Table S1). Genetic structure for these markers and for the control SNPs was estimated independently with a principal components analysis (PCA), performed in *R* v. 3.0.0 (R Development Core Team 2013) and with the Bayesian approach available in Structure v. 2.3.3 (Pritchard *et al.* 2000). Clustering with Structure was determined with an admixture model on correlated allele frequencies and a burn-in of 100,000 steps followed by 1,000,000 iterations. The number of clusters (K) was set from 1 to 15, and 10 runs were performed for each K . Similarity across runs with the same K was calculated with Clumpp (Jakobsson and Rosenberg 2007), and the most plausible number of clusters was determined following Evanno *et al.* (2005). Both the Structure membership coefficients (Q -matrix) and the first six PC scores of each individual for each independent data set were kept as explanatory variables for the environmental association analyses below. The six PC scores were retained in both cases (SSRs and control SNPs) because they each accounted for >5% of the total genetic variation.

Climate data

Summary climate data for the years 1950–2000 were retrieved for 32 variables from Worldclim (Hijmans *et al.* 2005) and a regional climatic model (Gonzalo 2007) for the 12 non-Spanish and the 24 Spanish populations, respectively. Climate variables included monthly mean, highest, and lowest temperatures and mean precipitation (Table S2). Gonzalo's (2007) model was favored for climate data in Spain because it considers a much denser network of meteorological stations than Worldclim, which is known to underperform in this region. Two independent PCAs, again performed in *R* v. 3.0.0 (R Development Core Team 2013), were used to summarize the climate variation of each population for the summer (June to September) and winter (December to March) seasons. Population scores for the first three PCs of each season (Table S2), which explained >95% of the total variance, were kept for further analyses (see below).

SNP–climate associations

Significant SNP–climate associations were identified by combining three different approaches. First, candidate SNP-allele frequencies were correlated with the climate PCs with multivariate logistic regressions (*mlr*) using independently the control SNP- and the SSR-PCs as covariates to account for historical and demographic processes that could have generated allele-frequency clines in the absence of selection (Grivet

et al. 2011; De-Mita *et al.* 2013). Identical associations were obtained when using the Structure membership coefficients (Q -matrix) as covariates (data not shown). These analyses were performed in *R* v. 3.0.0 (R Development Core Team 2013) for each separate candidate SNP by using the *glm* function, the Akaike Information Criterion for model selection, and the Benjamini–Hochberg procedure to control for false discovery rate (FDR; Benjamini 2010).

Second, another set of SNP–climate correlations was performed with the Gibbs sampler and the latent factor mixed models available in the software LFMM (Frichot *et al.* 2013). Following the population clustering results above (Structure and PCA), the number of latent factors (k) was set to six, as they should capture most of the underlying population structure (see Results). Then, 10 runs of one million sweeps were performed after discarding 100,000 iterations as burn-in. Significance was assessed by using the same FDR procedure as in the *mlr* analysis. Preliminary runs made with higher values of k yielded identical results, and tests performed with fewer latent factors resulted in correlations with lower statistical support and less convergence across independent runs (data not shown). Only those significant SNP–climate correlations that overlapped between the *mlr* and LFMM approaches (including all LFMM runs) were retained.

Third, a covariance matrix was built based on the control SNP data set (1745 SNPs) and used as a null model to further test for candidate SNP–climate correlations with the geographical Bayesian association analysis in Bayenv 2.0 (Coop *et al.* 2009; Günther and Coop 2013), a method showing great promise in comparative performance evaluations (Lotterhos and Whitlock 2014). To counterbalance the possibility that any gene-derived marker can be potentially involved in adaptation and can therefore be suboptimal to capture the population structure in the covariance matrix, a series of analyses was performed to validate its suitability. First, the covariance matrices produced by three independent Bayenv 2.0 runs were compared to each other to verify convergence to similar results. Second, these matrices were transformed to correlation matrices with the *cov2cor* function in *R* v. 3.0.0 and correlated to pairwise- F_{ST} matrices generated in Arlequin v. 3.5 (Excoffier and Lischer 2010) from this and the SSR data sets using Pearson's coefficient in *R* v. 3.0.0. Third, the Q -matrices produced in the Structure clustering analyses for the two putatively neutral data sets (control SNPs and SSRs) were similarly correlated in *R* v. 3.0.0.

After these validation steps (see File S1), a Bayes factor (BF) describing the deviation between the null demographic model (*i.e.*, the covariance matrix) and the alternative one (the candidate SNP–climate correlations) was estimated for each SNP. Model convergence was assured by performing three independent runs of 100,000 iterations and a burn-in of 10,000 chains with random seeds. Statistical support was assessed by Spearman's rank correlation (ρ) tests for those associations exhibiting unusually high BF (Eckert *et al.* 2010b; De La Torre *et al.* 2014). This was complemented by comparing these unusually high BFs to those observed for

SNPs that did not show any associations with the environment in any of the two previous methods (*mlr* and LFMM; see Figure S1). Only those candidate SNP–climate associations with a BF >10 (corresponding to strong support according to Jeffreys' scale) and a p -value of 0.25 or higher were conserved for validation and fitness prediction in the common garden (see below).

All three sets of analyses were repeated at the regional scale for the Iberian Mediterranean and Atlantic regions, where sampling was more intensive in terms of number of populations and individuals. The Mediterranean region of the Iberian Peninsula is a climatically heterogeneous area characterized by a high seasonality with very dry summers, in which maritime pine forms scattered populations near the coast, along altitudinal gradients in different mountain systems, and in the central plateau (Alía *et al.* 1996). The Atlantic region of the Iberian Peninsula is climatically more homogeneous, with wet-temperate climate and low seasonality. In this region, *P. pinaster* exhibits large continuous populations, partly due to plantations of local origin (Alía *et al.* 1996).

Genetic diversity and spatial structure

Expected and observed heterozygosity (H_E and H_O) and standardized genetic differentiation (G'_{ST}) were calculated with Smogd (Crawford 2010) for each population and gene pool (*i.e.*, Iberian Mediterranean and Atlantic regions) by using separately the SSRs, the candidate SNPs associated with climate, and the control SNPs. The existence of spatial autocorrelation was surveyed for the last two sets of markers by using the spatial structure analysis available in SAM v. 4.0 (Rangel *et al.* 2010). Briefly, a relative Moran I index (I/I_{max}) was determined for each marker by using a Gabriel connectivity matrix with symmetric distance classes. Significance was assessed with 999 permutations and the Bonferroni criterion for multiple testing.

Fitness predictions

Survival data for 19 representative populations were obtained from a common garden established in February 2004 in northeastern Spain (Cálcena; see Figure 1). The climate in this region is arid (average temperature 11.6°, annual rainfall 502 mm, summer rainfall 101 mm, for the period 1975–2008) and characteristic of the drier extreme of maritime pine's climatic breadth, which provides an ideal setting in which to test for differences in fitness due to climate maladaptation as expected under climate change (aridification and warming) in the core natural distribution of the species. The trial was designed as a nested structure of families within populations. Briefly, 1-year-old seedlings from 520 families were planted at a spacing of 2 m × 3 m in an α -lattice incomplete block design with 3 replications of 65 blocks, 8 families per incomplete block, and 4 plants per family plot (total number of 3 × 65 × 8 × 4 = 6240 plants). Survival was estimated as the ratio of seedlings alive per experimental unit when the plants were 5 years old. The climate in the test site during these 5 years was representative of the regional average (average temperature 11.8°, annual rainfall 533 mm, summer rainfall 99 mm; compare with

data for the period 1975–2008 above), with no particular extreme events in any year that might have affected tree survival.

Empirical best linear unbiased predictors (EBLUPs) for survival were computed for each factor (populations and families within population) after arcsine transformation (adequate for percentage data) using a mixed model with Gaussian error distribution that considered both factors as random. Climate data for the test site were used to determine the expected locally selected alleles for each of the 18 SNPs associated with climate (see *Results*) using the logistic regressions fitted above. The average frequency of locally advantageous alleles at each source population was then calculated and correlated with the EBLUPs for survival using Pearson's (r) and (nonparametric) Kendall's (τ) coefficients and associated significance tests. Then, given that SNPs associated with climate may have heterogeneous allelic effects on fitness, which would render the use of average frequencies from different loci inadequate, single-locus regressions with survival were also obtained for each of the 18 candidate markers to climate adaptation using standard linear regression (*lm* function) in R v. 3.0.0.

To demonstrate that significant correlations were not due to population structure, we first performed a resampling procedure as follows. A series of 1000 random samples of 18 SNPs with frequencies matching those of the retained candidates were obtained from the control SNP data set. Then, correlation coefficients with survival were computed for each 18-SNP sample as described above, and the probability of having equal or higher correlation coefficients than the ones obtained with the retained SNPs was estimated by comparing this value with the distribution of random correlation coefficients. Second, we computed a new PCA on the complete set of candidate SNPs tested for adaptation to climate (266 SNPs) and a new set of correlations with survival were obtained for the 18 SNPs that best explained population subdivision in this PCA.

Results

A total of 1745 control (*i.e.*, putatively neutral) SNPs and 9 nuclear SSRs were retained and reliably genotyped for 772 individuals from 36 maritime pine populations and used to infer background phylogeographic structure (Figure 1). The first six principal components (PCs) accounted for >67 and 61% of the total genetic variation, for the SNP and SSR data sets, respectively, with each individual component explaining at least 5% of the total variance. The most plausible number of genetic groups (following Evanno *et al.* 2005) produced by the Bayesian clustering analyses (Pritchard *et al.* 2000) was equal to six and coincided with the genetic structure identified by the PCs (Figure S2). These groups, hereafter called “gene pools,” coincided with those previously reported from independent chloroplast and nuclear SSR data sets for maritime pine (Bucci *et al.* 2007; Santos-Del-Blanco *et al.* 2012). Maritime pine gene pools were centered in Morocco,

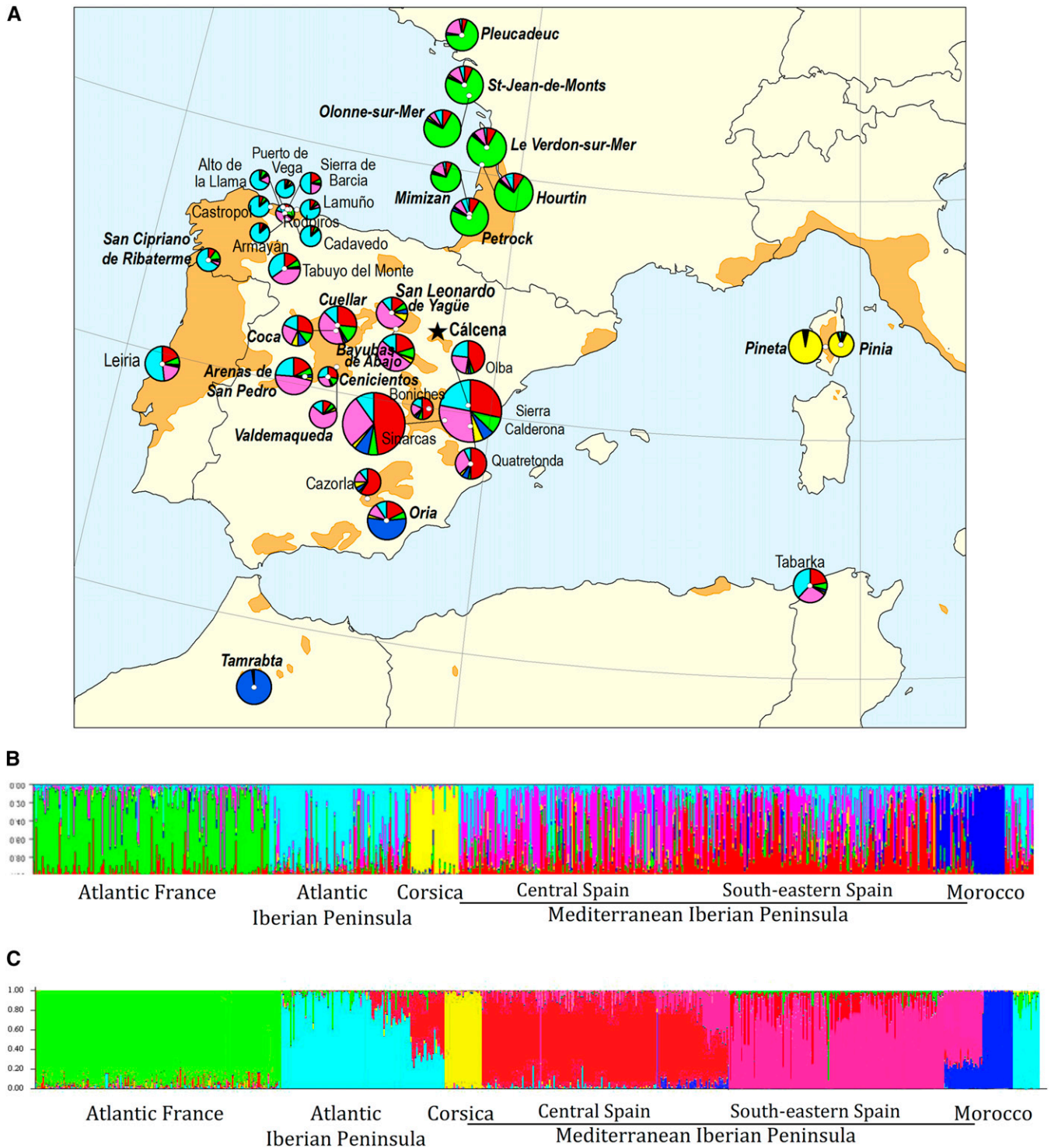


Figure 1 (A) Geographic distribution of the six gene pools obtained from (B) nine nuclear SSRs in 36 natural populations of maritime pine. (C) Genetic partition is also shown for 1745 control (*i.e.*, putatively neutral) SNPs. The shading in A denotes the species natural range, and the star indicates the location of the common garden used to evaluate fitness (Cálçena, Spain). Populations included in the common garden are in boldface italics.

Corsica, the Atlantic coast of France, and the Atlantic and Mediterranean (southeastern and central) regions of the Iberian Peninsula (Figure 1 and Figure S2); they are probably the result of the expansion of as many glacial refugia (Bucci *et al.* 2007; Santos-Del-Blanco *et al.* 2012).

Climate records (32 variables) for each population were reduced to three PCs for the summer (June to September) and winter months (December to March), explaining 96.3 and 95.9% of the total climatic variation for each season, respectively (Table S2 and Figure S3). The three winter PCs

were mainly loaded by the lowest and mean temperatures (PC-Winter1), the mean precipitation (PC-Winter2), and the highest temperature (PC-Winter3), while the summer axes mostly corresponded to the mean precipitation (PC-Summer1), and the lowest (PC-Summer2) and mean (PC-Summer3) temperatures, respectively. Population PC scores were tested for correlation (after correcting for phylogeographic structure) with the genotypes of the 266 candidate SNPs that were retained from the second OPA. At the range-wide scale, the *mlr* revealed 29 significant allele frequency–climate correlations for 24 SNPs, while the LFMM produced 49 significant correlations for 41 SNPs. In total, 21 correlations for 18 SNPs overlapped for both methods, including those for five nonsynonymous polymorphisms; all of these correlations were also significant in the Bayesian environmental association analysis (Coop *et al.* 2009) (Table 1 and Table S3).

The 18 retained SNPs were located in 17 genes involved in growth, synthesis of secondary metabolites, membrane transport, and abiotic stress response, among others (Table 1). Sixteen of them were correlated with axes loaded by temperature (PC-Winter1, PC-Winter3, or PC-Summer2), while only five showed significant correlations with components associated with precipitation (*e.g.*, PC-Winter2 or PC-Summer1; Table 1, Table S3, and Figure 2). Interestingly, four of the amplicons containing these SNPs (*2_1014_02*, *0_4032_02*, *2_3919_01*, and *CL813*; see functional annotations in Table 1) were overexpressed in maritime pine under experimental drought treatments (Perdiguerro *et al.* 2013), including the two examples given in Figure 2.

Average population differentiation (*e.g.*, G'_{ST}) was higher for the candidate SNPs associated with climate than for the control (*i.e.*, putatively neutral) ones or the nuclear SSRs (Tukey's post-hoc test, $P < 0.01$; see Figure 3A). Likewise, H_O and H_E were higher for these potentially adaptive loci when compared to the control markers (Tukey's post-hoc test, $P < 0.005$; Figure 3B), while relative Moran's I indices were also consistently significant for SNPs associated with climate (mean $I/I_{max} = 0.51$). These last values indicated a range-wide spatial autocorrelation for the alleles of these loci, which were more clumped, albeit widely distributed, than those of the control SNPs (mean $I/I_{max} = 0.19$).

Analyses were repeated at the regional scale for the Iberian Atlantic and Mediterranean regions, which represent independent gene pools currently experiencing contrasting climates (Bucci *et al.* 2007; Gonzalo 2007; Santos-Del-Blanco *et al.* 2012; see also Figure 1). In the more arid and climatically heterogeneous Iberian Mediterranean region, four SNPs (*m8*, *m80*, *m657*, and *m1309*) exhibited significant correlations with climate (Figure S4), while no associations were detected across the more humid and climatically homogeneous Iberian Atlantic region. Moreover, the 18 SNPs previously retained in the range-wide analyses displayed higher population differentiation than that of the other SNPs or SSRs within the first gene pool (Tukey's post-hoc test, $P < 0.01$), while in the second one all types of markers showed virtually the same population differentiation (Figure 3,

E and F). Likewise, genetic diversity was higher in the Mediterranean region for the SNPs associated with climate (Tukey's post-hoc test, $P < 0.02$; Figure 3D), while similar genetic diversity levels were found across regions for the control SNPs (H_E of 0.198 and 0.224, respectively). Alleles of the SNPs associated with climate (when variable) were also more clumped in the Mediterranean region than in the Atlantic one (mean $I/I_{max} = 0.46$ vs. 0.16).

Seedling survival after 5 years in a common garden test under the extreme climatic conditions of northeastern Spain was low and varied greatly across the 19 source populations surveyed (Figure 1 and Figure 4). Such a low (and variable) survival was somehow expected, as the common garden was located at the drier extreme of maritime pine's climatic breadth. Climate data for the common garden were used to determine the expected locally selected alleles for the 18 retained and potentially adaptive SNPs using the fitted logistic regressions above. The average frequency of these locally advantageous alleles in the source populations was positively and significantly correlated to survival in the common garden when using both parametric (Pearson's $r = 0.58$, $P = 0.0093$) and nonparametric (Kendall's $\tau = 0.44$, $P = 0.0082$) methods (Figure 4). Single-locus regressions provided additional support for these predictions, with 14 candidate SNPs (out of 18) each explaining $>5\%$ of the genetic variance for survival (Figure 4 and Table S4). Furthermore, only 0.2% ($P = 0.002$) of the 1000 random sets of 18 SNPs resampled from the control SNP data set (matched by allele frequency) had equal or higher correlation coefficients with survival than the one obtained with the SNPs that showed environmental associations. Finally, the 18 SNPs that best explained population subdivision for each of the four PCs explaining $>5\%$ of the variance in a PCA (accounting for $\sim 60\%$ of the total variance) did not produce any significant correlation (Table S5). Altogether, these tests suggest that the correlation between selected markers and climate maladaptation was not caused by chance or population structure.

Discussion

The utility of candidate-gene approaches in long-lived species with large genomes

In this study, we showed that the frequency of locally advantageous alleles at SNPs from carefully selected candidate genes can be used as predictors of climate maladaptation. This approach should be particularly appealing for outcrossing long-lived species like forest trees, for which establishing common gardens is expensive and time consuming. In addition, for taxa with extremely large genomes, such as conifers (Birol *et al.* 2013; Nystedt *et al.* 2013; Neale *et al.* 2014), the use of a candidate gene strategy seems an interesting and feasible cost-efficient alternative to genome-wide selection (*e.g.*, Westbrook *et al.* 2013), for which millions of markers evenly distributed across the genome might be necessary to have a good predictive power (*e.g.*, Grattapaglia and Resende 2011; Desta and Ortiz 2014). In this context, the adequate preselection of candidate

Table 1 Identity, annotation, and summary estimates of genetic diversity, population differentiation, and spatial structure for 18 SNPs exhibiting significant association with climate principal components (climate PCs)

EST contig or amplicon	<i>A. thaliana</i> locus	E-value	Annotation	LG ^a	SNP name (type of change) ^b	Accession number ^c	Climate PC	H _o	H _E	G' _{ST}	I / I _{max} ^d
0_11649	At2g29550	2.4 E-85	β-Tubulin	12F / 11	m783 (A/G) ^{syn}	836187519	PC-Winter1	0.34	0.33	0.23	0.47**
0_4032_02	At1g29340	2.8 E-24	U-box containing protein (protein kinase)	—/—	m8 (A/T) ^{syn}	325995579	PC-Winter1	0.34	0.33	0.37	0.47**
0_9486_01	NA	NA	Unknown protein	—/—	m1513 (A/C) ^{nonsyn}	836187520	PC-Summer2	0.23	0.25	0.37	0.47**
2_1014_02	At4g36990	2.5 E-07	Heat-stress transcription factor	—/7	m1250 (A/G) ^{NA}	836187522	PC-Winter2	0.41	0.42	0.20	0.40**
2_3919_01	At3g53720	8.1 E-11	Cation/H(+) antiporter	—/—	m80 (A/G) ^{nonsyn}	836187523	PC-Summer2	0.34	0.34	0.30	0.40**
2_4183_01	At1g08960	4.7 E-52	Cation exchanger	—/10	m1309 (A/G) ^{nonsyn}	836187524	PC-Winter1	0.18	0.22	0.52	0.62***
2_9542_01	At5g51690	8.3 E-54	1-Aminocyclopropane-1-carboxylate synthase	—/11	m100 (A/G) ^{syn}	836187525	PC-Winter3	0.30	0.31	0.27	0.51***
4cPt_C	NA	NA	4-Coumarate-CoA ligase	—/7	m973 (A/G) ^{syn}	836187526	PC-Winter2	0.36	0.38	0.25	0.41**
CL813	At5g20830	5.6 E-18	Sucrose-UDP glucosyltransferase	10M/10	m137 (A/G) ^{syn}	836187527	PC-Summer1	0.35	0.36	0.22	0.37**
Cobra2	At3g02210	1.2 E-19	Cobra-like protein	—/—	m607 (A/C) ^{noncod}	325995573	PC-Winter3	0.41	0.41	0.16	0.50***
CT11487	NA	NA	Unknown protein	—/1	m1156 (A/G) ^{noncod}	184956809	PC-Winter1	0.26	0.25	0.18	0.34**
CT2134	At5g48480	2.8 E-18	Glyoxalase I-like protein	7M/—	m1211 (C/G) ^{NA}	184975277	PC-Winter1	0.31	0.34	0.30	0.52***
CT2331	At5g14040	0.0	Mitochondrial phosphate transporter	—/—	m646 (C/G) ^{nonsyn}	184975314	PC-Winter1	0.30	0.32	0.22	0.38**
CT2334	At1g14320	4.9 E-159	60S ribosomal protein	5F	m657 (A/C) ^{nonsyn}	184975344	PC-Winter1	0.42	0.40	0.17	0.46**
CT3538	NA	0.0	Chlorophyll a/b binding protein	3F/3	m658 (A/G) ^{syn}	184975347	PC-Summer2	0.34	0.34	0.31	0.56***
CT384	At2g29960	9.2 E-89	Peptidyl-prolyl cis-trans isomerase	—/—	m685 (A/G) ^{syn}	184975431	PC-Winter2	0.34	0.32	0.22	0.64***
CT4109	At2g40590	2.6 E-30	40S ribosomal protein (predicted)	—/—	m1196 (A/C) ^{syn}	184975463	PC-Summer3	0.43	0.39	0.33	0.67***
					m705 (A/G) ^{noncod}	184975489	PC-Summer2	0.29	0.33	0.27	0.52***

Only those loci displaying significant correlations at $P < 0.001$ and with Bayes factors > 10 are shown (see also Table S3).

^a Linkage groups retrieved from two independent maritime pine mapping populations: Chanceret *et al.* (2011) and de-Miguel *et al.* (2012).

^b Abbreviations: syn, synonymous; nonsyn, nonsynonymous; noncod, noncoding; NA, unknown/data unavailable. The locally selected alleles at the common garden used to evaluate fitness, judging from logistic models, appear in boldface type.

^c Accession numbers as provided by NCBI's dbSNP.

^d Levels of significance: ** $p < 0.01$; *** $p < 0.005$.

m80 (Heat stress transcription factor)

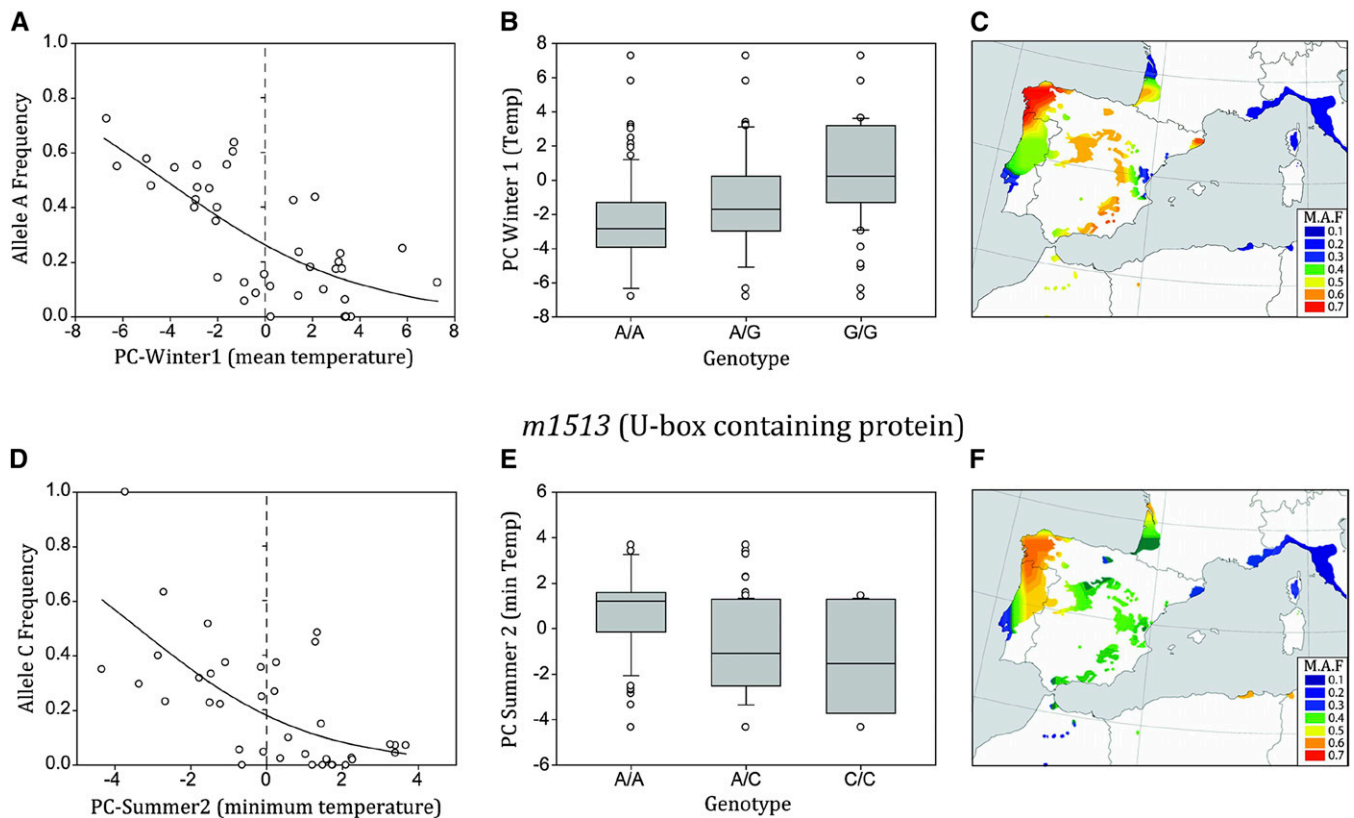


Figure 2 Examples of allele–climate associations for maritime pine populations at the range-wide scale. (A and D) Scatter plots, (B and E) box plots, and (C and F) distribution of minor allele frequencies (MAF) are shown for 2 of the 18 potentially adaptive SNPs. Lines within the scatter plots indicate clines of allele frequencies under a logistic regression model. Boxes denote the interquartile range and horizontal lines within boxes represent genotypic means.

genes becomes a fundamental step toward capturing a large part of the adaptive/phenotypic variance that needs to be used to perform such predictions.

Herein, we performed the preselection of genes by pinpointing candidates from previous population and functional studies. These genes included drought-stress-response candidates detected by population genetics (e.g., Eveno *et al.* 2008; Grivet *et al.* 2011; Mosca *et al.* 2012; see also Budde *et al.* 2014) or expression studies (Lorenz *et al.* 2011; Perdiguero *et al.* 2013), as well as variants previously associated to wood properties, cold hardiness, and growth in maritime pine (Pot *et al.* 2005; Eveno *et al.* 2008; Lepoittevin *et al.* 2012) and other conifers (González-Martínez *et al.* 2007, 2008; Eckert *et al.* 2010a,b). Many of these genes belong to gene families that have been associated with adaptive responses in other plants. For instance, amplicons *CL813* and *CT2134* contain genes directly involved in osmotic adjustments that are related to the metabolism of carbohydrates in both maritime pine and *Arabidopsis* (Seki *et al.* 2002; Pot *et al.* 2005), while contig *O_11649* encodes for a β -tubulin, a family of genes whose expression changes under low temperatures in different plants (Chu *et al.* 1993; Seki *et al.* 2002; Perdiguero *et al.* 2013).

This preselection of genes allowed us to survey many of the potential targets of selection and climate adaptation in

maritime pine, but it may also have led to ascertainment bias and to limiting our power to identify selection drivers in this species (see also Morin *et al.* 2004; Namroud *et al.* 2008). For instance, by selecting candidates from studies not only in maritime pine but also in other conifers, we assumed that adaptive processes mostly occur by convergent evolution in the same set of genes, while there is evidence suggesting that species adaptation to identical environments may involve separate genes and a certain number of possible paths (e.g., Tenaillon *et al.* 2012). For example, only 2 of the 18 genes surveyed in latitudinal clines of two Eurasian boreal spruces showed identical associations with bud phenology (Chen *et al.* 2012, 2014), while only 3 outlier genes related to climate were shared between two sympatric *Picea* species in eastern Canada (Prunier *et al.* 2011). Similar results have been reported for adaptation to freshwater in sticklebacks (Deagle *et al.* 2012) and to high elevation in humans (Bigham *et al.* 2010). Further bias may originate from excluding noncoding regions, micro-RNAs, or copy-number and presence-absence variants (including transposon insertions), which have been directly associated with adaptive responses in many taxa, including conifers (e.g., González *et al.* 2008; Yakovlev *et al.* 2010; Fischer *et al.* 2011; Hanikenne *et al.* 2013). While genome-wide approaches are appropriate to capture most of this variability,

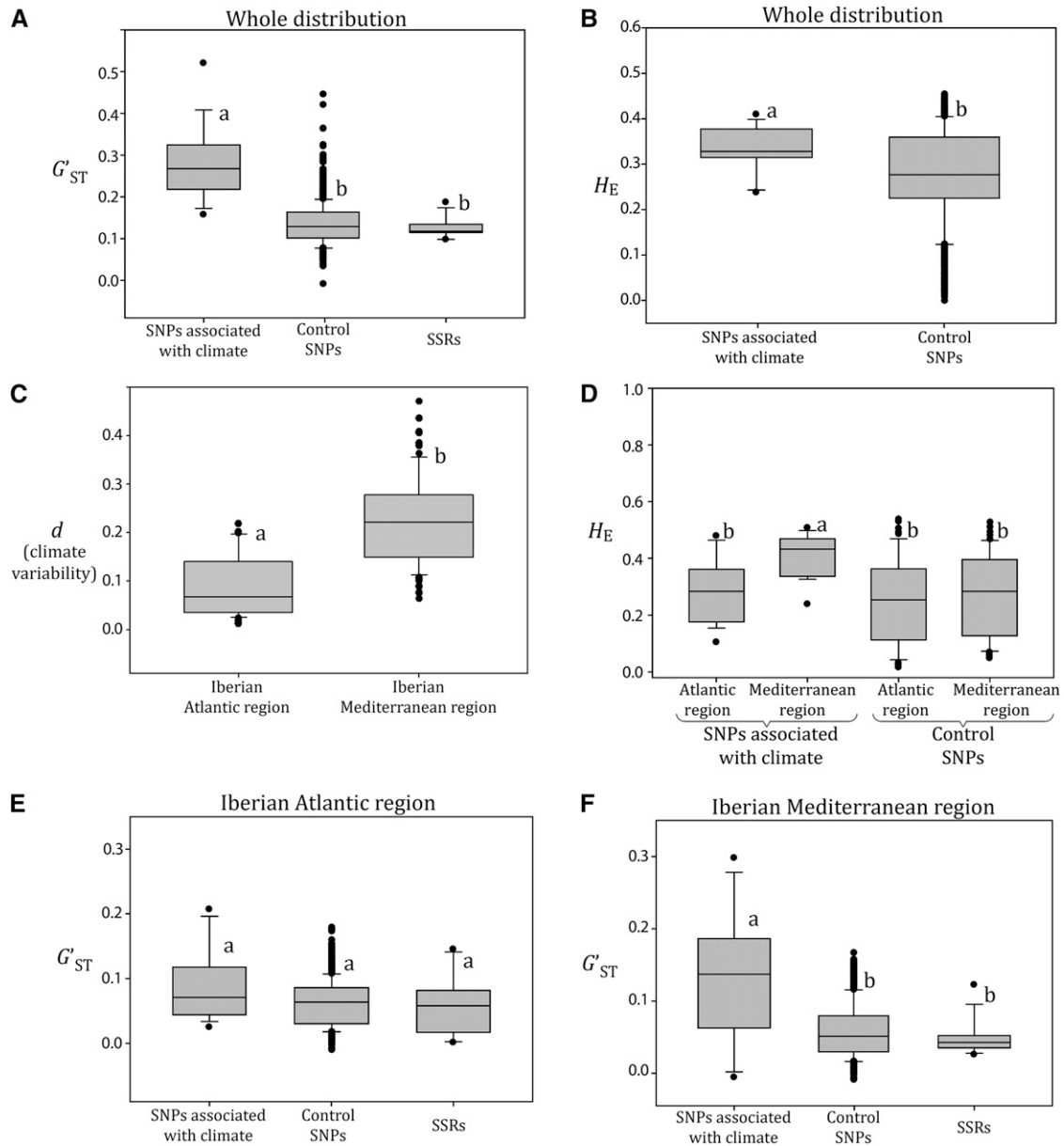


Figure 3 (A) Range-wide and (E and F) regional population differentiation (G'_{ST}) for nuclear SSRs, control SNPs (*i.e.*, putatively neutral), and SNPs associated with climate across natural populations of maritime pine. (B) Overall genetic diversity (H_E) for SNPs associated with climate and control SNPs; (D) genetic diversity (H_E) by SNP type in the different regions. (C) Euclidean distances (d) on climate variability at the regional scale are also shown. Boxes denote the interquartile range, whiskers the 95% confidence intervals, and horizontal lines within boxes represent genotypic (or climatic) means.

having a satisfactory representation of the whole segregating genomic diversity is still a challenge for nonmodel species with large genomes, such as conifers (*e.g.*, Birol *et al.* 2013; Nystedt *et al.* 2013; see also Tiffin and Ross-Ibarra 2014).

Identifying SNP–climate associations and potential selection drivers

Complementary statistical analyses were used herein for detecting genotype–climate associations, which should be adequate to minimize the detection of false positives (De-Mita *et al.* 2013). Additional support for the adaptive role of the retained associations (involving 18 SNPs) was obtained from

a common garden experiment under extreme climate. This strategy (*i.e.*, environmental associations plus common gardens to evaluate fitness) should enrich traditional association studies that provide links between genotype and phenotype (mostly in the form of candidate genes lists), by identifying some of the genotype–environment relationships involved in adaptation (*i.e.*, the selection drivers) and by estimating the fitness effects of ecologically relevant variants (ideally under multiple environments). It can also provide a much needed validation step (Ioannidis *et al.* 2009; König 2011).

Most of the SNPs associated with climate in maritime pine were related to PC axes loaded by temperature, during

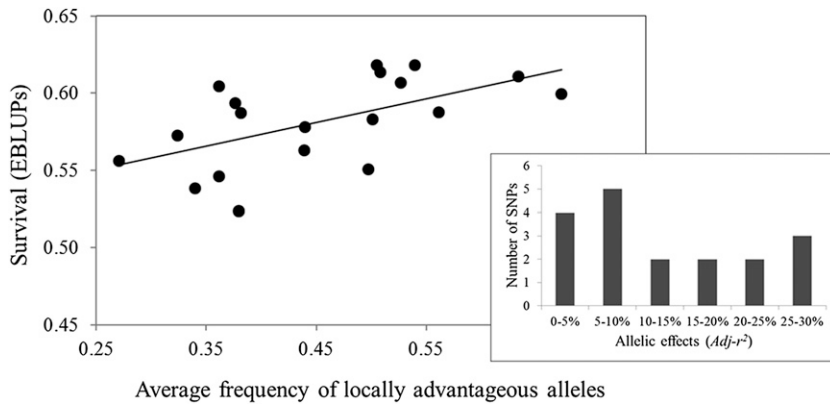


Figure 4 Scatter plot of average frequency of locally advantageous alleles in the common garden (Cálcena, Spain) at each source population and EBLUPs (empirical best linear unbiased predictors) for survival, a proxy for fitness. The inset shows the distribution of allelic effects for the 18 SNPs associated with climate.

both the winter and the summer months, and, to a lesser extent, by winter precipitation. Adaptation to these elements are critical for assuring survival in plants (Condit *et al.* 1995), which supports the view that maladaptation to increasing temperature and drought can become an important source of vulnerability for forests under climate change (Carnicer *et al.* 2011; Choat *et al.* 2012). The overrepresentation of correlations with winter temperatures in our analyses (53% of the total; Table 1) might seem counterintuitive for a Mediterranean conifer that should be more affected by drought stress during the summer, according to modeled distributions (Benito-Garzón *et al.* 2011). However, the genetic control of growth cessation and cold hardiness is capital for the survival of many plants, including forest trees (Tanino *et al.* 2010; Cooke *et al.* 2012). Moreover, given that both freezing and drought trigger analogous physiological responses in conifers (Bigras and Colombo 2001; Blödner *et al.* 2005), the same genes are expected to control (at least partially) responses to both types of stress (Urano *et al.* 2010). Indeed, four of the amplicons whose SNPs were related to winter temperatures were also overexpressed under drought conditions in maritime pine (Perdiguero *et al.* 2013), while four of these SNPs were also associated with summer PCs, including a nonsynonymous variant located on a gene coding for a putative heat-stress transcription factor (*m80*; Table 1 and Figure 2). This is consistent with results from *Arabidopsis thaliana*, where 30 genes were upregulated by both cold temperatures and drought, including four amplicons coding for heat-shock proteins (Seki *et al.* 2002). Other genes associated with environment that are worth highlighting include the above-mentioned *CL813*, *CT2134*, and *O_11649*, which were correlated with winter PCs (Table 1; Seki *et al.* 2002; Pot *et al.* 2005), and *4cl-Pt_c*. This last gene belongs to the *4cl* family, which is involved in the phenylpropanoid metabolism and the biosynthesis of lignin (Yun *et al.* 2005; González-Martínez *et al.* 2007; Wagner *et al.* 2009) and has previously exhibited environmental associations at the haplotype level in maritime pine and other Mediterranean pine species (Grivet *et al.* 2011).

It must be noted, however, that although promising, SNP–environment correlations alone only hint at the real drivers of selection and local adaptation. For instance, in maritime pine, any other factor covarying with extreme temperatures and low

precipitation should exhibit similar correlations with the retained candidate SNPs. Indeed, it can be argued that these correlations might be biased if the used climate variables are spatially structured or partially match the genetic structure of the species. For example, two of the clusters detected for maritime pine visually coincide with the regions of highest average precipitation (northern Spain and Atlantic France), while the southeastern Spain gene pool is mostly located in a region with overly high winter temperatures. Indeed, when nonlinear models were fitted to explain each climatic PC with either geography (latitude and longitude) or genetic structure (the PCs derived from the SSR and control SNP data sets), significant correlations were observed for PC-Winter3 ($r^2 = 0.85$ for geography and $r^2 = 0.6$ for the genetic PCs) and PC-Summer2 ($r^2 = 0.39$ for geography and $r^2 = 0.7$ for the genetic PCs; Table S6), two components loaded by extreme temperatures (Table S2). Although these results suggest that caution should be taken when interpreting genetic associations with these two particular PCs, most of the correlations detected herein (67%) involved climate components that showed no evidence of being spatially structured (Table 1 and Table S6), which lends support to the view that the selected SNPs are effectively associated with adaptation to climate.

Geographic extent of SNP–climate associations and modes of adaptation

Alleles from potentially adaptive SNPs were both widely distributed and locally clumped, while the control (*i.e.*, putatively neutral) ones showed a strong population structure at the range-wide scale and a lower aggregation at the local level (Table 1, Figure 1, and Figure 2). Such patterns, together with the particular biological traits of forest trees, suggest that selection could be acting on the standing genetic variation of maritime pine and at relatively small geographical scales (see Eckert *et al.* 2012 for a similar case in *P. contorta*). Forest trees are characterized by large population sizes, extensive gene flow, and long generation times, which favor the long-term maintenance of ancestral polymorphisms (Bouillé and Bousquet 2005; Petit and Hampe 2006); indeed, selection on standing genetic variation seems to be the rule in these taxa (Savolainen *et al.* 2011; Chen *et al.* 2012; Eckert *et al.* 2012; Prunier *et al.* 2012; Alberto *et al.* 2013a,b). Such attributes

would also allow tree populations to better respond to heterogeneous local selection (Kremer *et al.* 2012), leading to contrasted regional adaptive patterns, such as observed herein for the Atlantic and Mediterranean regions of the Iberian Peninsula. The genetic differentiation and diversity of the potentially adaptive SNPs reflect the environmental variability of these areas (Figure 3C), which supports the view that there should be a correlation between the within-population genetic variation at adaptive traits and the environmental heterogeneity at the regional scale (Yeaman and Jarvis 2006; Kremer *et al.* 2012). However, it must be noted that most of the statistical approaches currently used to identify loci related to local adaptation (including those employed herein) assume that their alleles exhibit antagonistic pleiotropy (*e.g.*, De Mita *et al.* 2013; see review in Tiffin and Ross-Ibarra 2014), that is, that their variants are beneficial in particular regions of the species range and deleterious elsewhere. Nevertheless, empirical studies suggest that many alleles contributing to local adaptation in particular environments are effectively neutral in other regions (*e.g.*, Anderson *et al.* 2011; Fournier-Level *et al.* 2011). Such conditional neutrality could be an alternative explanation for the differences observed herein between the Iberian Atlantic and Mediterranean regions.

Other than the action of selection on standing genetic variation, it can be argued that selection on newly arisen variants could have also played a role in modeling the patterns described above. When a new advantageous mutation appears and hard selective forces drive it to high frequency within a few generations, its geographic distribution is expected to be narrower than that of the alleles already present in the species gene pool (Hancock *et al.* 2011). Nonetheless, given enough time and relatively stable populations, extensive gene flow, such as the one of conifers, can rapidly spread this variant until matching the patterns of more ancient alleles (Kremer *et al.* 2012). Thus, to address the contribution of *de novo* adaptive variants, the time scale and intensity of the selective forces that are currently operating in maritime pine should be estimated. Such estimations are out of the scope of this study, but taking into account the selection intensities previously inferred for putatively adaptive SNPs in other conifers (*i.e.*, $s = 0.01 - 0.04$; Prunier *et al.* 2011; Eckert *et al.* 2012), an adaptive allele that arose just after the last glacial maximum (some 10,000 YBP) should currently have a frequency < 0.03 , assuming an average generation time of 50 years. These rough calculations imply that, because of the use of markers with MAF > 0.05 , our strategy should be adequate to account only for far more ancient mutations or for those that have been favored by much stronger selection coefficients. Such a case was recently reported for Norway spruce, where a new variant in the *PaFTL2* gene appeared and spread across most of Fennoscandia in < 6500 years, until reaching frequencies > 0.6 in some modern populations in northern Sweden (Chen *et al.* 2012). However, even in this extreme case, the new advantageous allele did still have a local distribution (*i.e.*, only Fennoscandia), which is something that was

not observed herein for the SNPs associated with climate in maritime pine.

Spatially heterogeneous patterns of adaptation can also be accounted for by demographic history (*i.e.*, the fixation of lineage-specific polymorphisms by drift), as proposed for boreal black spruce (Prunier *et al.* 2012). However, in contrast to this conifer, overall levels of neutral genetic diversity in Mediterranean maritime pine were not different across regions (Figure 3D), thus suggesting similar local demographic histories, likely modeled by past bottlenecks and local expansions (Grivet *et al.* 2011). Furthermore, despite boreal and Mediterranean conifers bearing equally high levels of pollen-driven gene flow (O'Connell *et al.* 2006; de-Lucas *et al.* 2008), they have differential capacities (higher in boreal conifers) for quick range-shifts following environmental changes (McLeod and MacDonald 1997; Rubiales *et al.* 2010; Santos-Del-Blanco *et al.* 2012). Thus, it would not be surprising to observe selection to play contrasting roles on adaptive alleles in these species (Kuparinen *et al.* 2010). Several theoretical and empirical works (reviewed by Kremer *et al.* 2012) have shown that fixation of adaptive polymorphisms can be facilitated under rapid dispersal scenarios, such as for boreal conifers (McLeod and MacDonald 1997; Savolainen *et al.* 2011; Chen *et al.* 2012; Prunier *et al.* 2012), while more local expansions, such as those of Mediterranean taxa (Bucci *et al.* 2007; Rubiales *et al.* 2010; Grivet *et al.* 2011), should result in allele-frequency divergence for adaptive variants, particularly under contrasting environments and by assuming either an antagonistic-pleiotropic or a conditionally neutral model (see above).

Species evolving under this local-expansion scenario should also be more sensitive to maladaptation when facing rapid environmental changes, as suggested herein. Indeed, when grown in a common garden at the drier extreme of their climatic breadth, maritime pine populations bearing low frequency of locally advantageous alleles at potentially adaptive SNPs showed increased mortality (*i.e.*, lower fitness), suggesting that different gene pools should exhibit contrasting responses to climate change (see also Benito-Garzón *et al.* 2011). Strong selective regimes, such as the one observed in our common garden, can also foster adaptation and ultimately result in population persistence when acting concomitantly with the high reproductive capacity and population growth of forest trees (Lynch and Lande 1993; Kuparinen *et al.* 2010; Savolainen *et al.* 2011). However, it is uncertain whether a species such as maritime pine, which has a fragmented distribution and is already at its ecological limit in large parts of its southern range, can cope with environments that will quickly become more arid in the near future (Loarie *et al.* 2009; Bellard *et al.* 2012; Alberto *et al.* 2013a).

Integrating molecular predictors into range-shift models

Incorporating fitness proxies to range-shift models based on occurrence data can substantially change their predictions (*e.g.*, Kearney *et al.* 2009; Benito-Garzón *et al.* 2011; see also Hoffmann and Sgrò 2011). The results of the present study suggest that population genomic information (*e.g.*, the frequency

of beneficial alleles for a particular site identified through environmental association) could be a good proxy for fitness and to be integrated into range-shift models. Moreover, breeding predictors could be estimated from adaptive marker genotypes for each particular population/gene pool under different climate change scenarios, and used to guide future reforestation programs. A marker-based approach might also simplify the prediction of evolutionary trajectories by considering the underlying changes in allele frequency of locally selected alleles and their covariance. Within this framework, new markers from genome-wide studies could be readily incorporated, thus allowing improved predictive models. The sequencing of the first conifer reference genomes is a promising step in this direction (Birol *et al.* 2013; Nystedt *et al.* 2013; Neale *et al.* 2014).

The use of haplotypes can further improve the accuracy of predictive models that rely solely on the marginal effects of individual loci (Calus *et al.* 2008; Eveno *et al.* 2008; Grivet *et al.* 2011), as they allow incorporating epistatic interactions of alleles (*i.e.*, genetic context), which also affects fitness. This is of particular interest given that beneficial allele combinations and common large-effect variants are the first to be captured after the establishment of new divergent selective forces, as may be the case for impending climate change, while rarer variants and small-effect alleles at individual loci are targeted only after >50 generations (Kremer and Le Corre 2012). In the absence of adequate LD estimates, genetic context might be approximated by using allele frequency covariation, but in species with strong population genetic structure, such as maritime pine, the utility of this approach is limited. In such cases, looking for an enrichment of coupling-phase LD among adaptive alleles might be an alternative, although newly developed approaches may provide the adequate framework to overcome these issues more easily (*e.g.*, Berg and Coop 2014). Finally, the incorporation of epigenetic variation into these predictive models must also be considered. Epigenetic factors have been shown to drive a substantial proportion of phenotypic variation and climate adaptation in Norway spruce (Yakovlev *et al.* 2010), which seems particularly rich in gene families involved in DNA and chromatin methylation (Nystedt *et al.* 2013). Thus, their role in modulating the expression of key adaptive genes in this and other taxa, including the candidates retained herein, still must be surveyed.

In conclusion, we have shown that, while new technology that allows in-depth studies of the huge conifer genomes is developed, carefully selected candidate genes can still be useful in identifying genetic variation underlying adaptation to climate. Adaptive patterns are expected to vary across geographically separate gene pools, as observed for the Iberian Mediterranean and Atlantic ranges of maritime pine. Thus, the success of programs to preserve biological diversity under impending climate change will largely depend on our capacity to identify and understand how adaptive variation in keystone species is distributed and evolves. Fitness experiments under extreme environmental conditions, as the one developed herein, do not only provide much-needed validation to association and outlier-locus studies but are also a first step

to integrate this knowledge into ecological models to foretell the fate of modern populations and species.

Acknowledgments

We thank T. Kawecki for discussing analytical approaches to predict maladaptation based on SNPs using common gardens under extreme environmental conditions. Thanks are extended to J. Majada, M. Zabal-Aguirre, Y. Kurt, C. García-Barriga, A. I. de-Lucas, C. Martínez-Alonso, C. Feito, and E. Hidalgo for field and laboratory assistance, J. Wegrzyn, T. Lang, and J.-M. Frigerio for bioinformatics support, and S. I. Wright and two anonymous reviewers for thoughtful comments on a previous version of this manuscript. CEGEN-ISCI (http://www.cegen.org) provided SNP genotyping services. We acknowledge grants from the European Commission (FP6 NoE EvolTree and FP7 NovelTree Breeding), the Spanish National Research Plan (ClonaPin, RTA2010-00120-C02-01; VaMPiro, CGL2008-05289-C02-01/02; AdapCon, CGL2011-30182-C02-01; and AFFLORA, CGL2012-40129-C02-02), and the European Research Area-Net BiodivERsA (LinkTree project, EUI2008-03713), which included the Spanish Ministry of Economy and Competitiveness as national funder (part of the 2008 BiodivERsA call for research proposals). G.G.V. was funded by a grant from the Italian Ministry (Ministero dell'Istruzione, dell'Università e della Ricerca project Biodiversitalia, RBAP10A2T4), while J.P.J.-C., D.G., and M.H. were supported by postdoctoral fellowships (Juan de la Cierva and Ramón y Cajal) from the Spanish Ministry of Science and Innovation. Finally, S.C.G.-M. and M.H. acknowledge the support of Marie Curie Intra European Fellowships within the 7th European Community Framework Programme (FP7-PEOPLE-2012-IEF, project nos. 328146 and 329088, respectively).

Literature Cited

- Aitken, S. N., S. Yeaman, J. A. Holliday, T. Wang, and S. Curtis-McLane, 2008 Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evol. Appl.* 1: 95–111.
- Alberto, F. J., S. N. Aitken, R. Alía, S. C. González-Martínez, H. Hänninen *et al.*, 2013a Potential for evolutionary responses to climate change: evidence from tree populations. *Glob. Change Biol.* 19: 1645–1661.
- Alberto, F. J., J. Derory, C. Boury, J. M. Frigerio, N. E. Zimmermann *et al.*, 2013b Imprints of natural selection along environmental gradients in phenology-related genes of *Quercus petraea*. *Genetics* 195: 495–512.
- Alía, R., S. Martín, J. de-Miguel, R. Galera, D. Agúndez *et al.*, 1996 *Las regiones de procedencia de Pinus pinaster Aiton*. Organismo Autónomo de Parques Nacionales, Madrid.
- Anderson, J. T., J. H. Willis, and T. Mitchell-Olds, 2011 Evolutionary genetics of plant adaptation. *Trends Genet.* 27: 258–266.
- Bellard, C., C. Bertelsmeier, P. Leadley, W. Thuiller, and F. Courchamp, 2012 Impacts of climate change on the future of biodiversity. *Ecol. Lett.* 15: 365–377.
- Benito Garzón, M., R. Alía, T. M. Robson, and M. A. Zavala, 2011 Intra-specific variability and plasticity influence potential tree species distributions under climate change. *Glob. Ecol. Biogeogr.* 20: 766–778.

- Benjamini, Y., 2010 Discovering the false discovery rate. *J. R. Stat. Soc. B* 72: 405–416.
- Berg, J. J., and G. Coop, 2014 A population genetic signal of polygenic adaptation. *PLoS Genet.* 10: e1004412.
- Bigham, A., M. Bauchet, D. Pinto, X. Mao, J. M. Akey *et al.*, 2010 Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet.* 6: e1001116.
- Bigras, F., and S. J. Colombo, 2001 *Conifer Cold Hardiness*. Kluwer, Dordrecht, The Netherlands.
- Biol, I., A. Raymond, S. D. Jackman, S. Pleasance, R. Coope *et al.*, 2013 Assembling the 20Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics* 29: 1492–1497.
- Blödner, C., T. Skroppa, Ø. Johnsen, and A. Polle, 2005 Freezing tolerance in two Norway spruce (*Picea abies* [L.] Karst.) progenies is physiologically correlated with drought. *J. Plant Physiol.* 162: 549–558.
- Bouillé, M., and J. Bousquet, 2005 Trans-species shared polymorphisms at orthologous gene loci among distant species in the conifer *Picea* (Pinaceae): implications for the long-term maintenance of genetic diversity in trees. *Am. J. Bot.* 92: 63–73.
- Brown, G. R., G. P. Gill, R. J. Kuntz, C. H. Langley, and D. B. Neale, 2004 Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proc. Natl. Acad. Sci. USA* 101: 15255–15260.
- Bucci, G., S. C. González-Martínez, G. Le Provost, C. Plomion, M. M. Ribeiro *et al.*, 2007 Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Mol. Ecol.* 16: 2137–2153.
- Budde, K. B., M. Heuertz, A. Hernández-Serrano, J. G. Pausas, G. G. Vendramin *et al.*, 2014 *In situ* genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster*). *New Phytol.* 201: 230–241.
- Calus, M. P. L., T. H. E. Meuwissen, A. P. W. de Roos, and R. F. Veerkamo, 2008 Accuracy of genomic selection using different methods to define haplotypes. *Genetics* 178: 553–561.
- Carnicer, J., M. Coll, M. Ninyerola, X. Pons, G. Sánchez *et al.*, 2011 Widespread crown condition decline, food web disruption, and amplified tree mortality with increased climate change-type drought. *Proc. Natl. Acad. Sci. USA* 108: 1474–1478.
- Chancerel, E., C. Lepoittevin, G. Le Provost, Y. C. Lin, J. P. Jaramillo-Correa *et al.*, 2011 Development and implementation of a highly-multiplexed SNP array for genetic mapping in maritime pine. *BMC Genomics* 12: 368.
- Chancerel, E., J. B. Lamy, I. Lesur, C. Noirot, C. Klopp *et al.*, 2013 High-density linkage mapping in a pine tree reveals a genomic region associated with inbreeding depression and provides clues to the extent and distribution of meiotic recombination. *BMC Biol.* 11: 50.
- Chen, J., T. Källman, X. Ma, N. Gyllenstrand, G. Zaina *et al.*, 2012 Disentangling the roles of history and local selection in shaping clinal variation of allele frequencies and gene expression in Norway spruce (*Picea abies*). *Genetics* 191: 865–881.
- Chen, J., Y. Tsuda, M. Stocks, T. Källman, N. Xu *et al.*, 2014 Clinal variation at phenology-related genes in spruce: Parallel evolution in *FTL2* and *Gigantea*? *Genetics* 197: 1025–1038.
- Choat, B., S. Jansen, T. J. Brodribb, H. Cochard, S. Delzon *et al.*, 2012 Global convergence in the vulnerability of forests to drought. *Nature* 491: 752–755.
- Chu, B., D. P. Snustad, and J. V. Carter, 1993 Alteration of β -tubulin gene expression during low-temperature exposure in leaves of *Arabidopsis thaliana*. *Plant Physiol.* 103: 371–377.
- Condit, R., S. P. Hubbell, and R. B. Foster, 1995 Mortality rates of 205 neotropical tree and shrub species and the impact of a severe drought. *Ecol. Monogr.* 65: 419–439.
- Cooke, J. E. K., M. E. Eriksson, and O. Junntila, 2012 The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. *Plant Cell Environ.* 35: 1707–1728.
- Coop, G., J. K. Pickrell, J. Novembre, S. Kudaravalli, J. Li *et al.*, 2009 The role of geography in human adaptation. *PLoS Genet.* 5: e1000500.
- Crawford, N. G., 2010 SMOGD: software for the measurement of genetic diversity. *Mol. Ecol. Res.* 10: 556–557.
- De La Torre, A. R., D. R. Roberts, and S. N. Aitken, 2014 Genome-wide admixture and ecological niche modelling reveal the maintenance of species boundaries despite long history of interspecific gene flow. *Mol. Ecol.* 23: 2046–2059.
- de-Lucas, A. I., J. J. Robledo-Arnuncio, E. Hidalgo, and S. C. González-Martínez, 2008 Mating system and pollen gene flow in Mediterranean maritime pine. *Heredity* 100: 390–399.
- de-Miguel, M., N. de-Maria, M. A. Guevara, L. Diaz, E. Sáez-Laguna *et al.*, 2012 Annotated genetic linkage maps of *Pinus pinaster* Ait. from a central Spain population using microsatellite and gene based markers. *BMC Genomics* 13: 527.
- De-Mita, S., A. C. Thuillet, L. Gay, N. Ahmadi, S. Manel *et al.*, 2013 Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol. Ecol.* 22: 1383–1399.
- Deagle, B. E., F. C. Jones, Y. F. Chan, D. M. Absher, D. M. Kingsley *et al.*, 2012 Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proc. R. Soc. B Biol. Sci.* 279: 1277–1286.
- Desta, Z. A., and R. Ortiz, 2014 Genome-selection: genome-wide prediction in plant improvement. *Trends Plant Sci.* 19: 592–601.
- Eckert, A. J., H. Shahi, S. L. Datwyler, and D. B. Neale, 2012 Spatially variable natural selection and the divergence between parapatric subspecies of lodgepole pine (*Pinus contorta*, Pinaceae). *Am. J. Bot.* 99: 1323–1334.
- Eckert, A. J., J. van Heerwaarden, J. L. Wegrzyn, C. D. Nelson, J. Ross-Ibarra *et al.*, 2010a Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* 185: 969–982.
- Eckert, A. J., A. D. Bower, S. C. González-Martínez, J. L. Wegrzyn, G. Coop *et al.*, 2010b Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Mol. Ecol.* 19: 3789–3805.
- Evanno, G., S. Regnaut, and J. Goudet, 2005 Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14: 2611–2620.
- Eveno, E., C. Collada, M. A. Guevara, V. Léger, A. Soto *et al.*, 2008 Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Mol. Biol. Evol.* 25: 417–437.
- Excoffier, L., and H. E. L. Lischer, 2010 Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564–567.
- Fischer, I., L. Camus-Kulandaivelu, F. Allal, and W. Stephan, 2011 Adaptation to drought in two wild tomato species: the evolution of the *Asr* gene family. *New Phytol.* 190: 1032–1044.
- Fournier-Level, A., A. Korte, M. D. Cooper, M. Nordborg, J. Schmitt *et al.*, 2011 A map of local adaptation in *Arabidopsis thaliana*. *Science* 334: 86–89.
- Franks, S. J., and A. A. Hoffmann, 2012 Genetics of climate change adaptation. *Annu. Rev. Genet.* 46: 185–208.
- Frichot, E., S. D. Schoville, G. Bouchard, and O. François, 2013 Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol. Biol. Evol.* 30: 1687–1699.
- Gamache, I., and S. Payette, 2004 Height growth response of tree line black spruce to recent climate warming across the forest-tundra of eastern Canada. *J. Ecol.* 92: 835–845.
- González, J., K. Lenkov, M. Lipatov, J. M. MacPherson, and D. A. Petrov, 2008 High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biol.* 6: e251.
- González-Martínez, S. C., N. C. Wheeler, E. Ersoz, C. D. Nelson, and D. B. Neale, 2007 Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* 175: 399–409.

- González-Martínez, S. C., D. Huber, E. Ersoz, J. M. Davis, and D. B. Neale, 2008 Association genetics in *Pinus taeda* L. II. Water use efficiency. *Heredity* 101: 19–26.
- Gonzalo, J., 2007 Phytoclimatic analysis of the Spanish Peninsula: update and geostatistical analysis. PhD Thesis, University of Valladolid, Palencia, Spain.
- Grattapaglia, D., and M. D. V. Resende, 2011 Genome selection in forest tree breeding. *Tree Genet. Genomes* 7: 241–255.
- Grivet, D., F. Sebastiani, R. Alía, and T. Bataillon, S. Torre *et al.*, 2011 Molecular footprints of local adaptation in two Mediterranean conifers. *Mol. Biol. Evol.* 28: 101–116.
- Günther, T., and G. Coop, 2013 Robust identification of local adaptation from allele frequencies. *Genetics* 195: 205–220.
- Hancock, A. M., B. Brachi, N. Faure, M. W. Horton, L. B. Jarymowycz *et al.*, 2011 Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334: 83–86.
- Hanikenne, M., J. Kroymann, A. Tramczynska, M. Bernal, P. Motte *et al.*, 2013 Hard selective sweep and ectopic gene conversion in a gene cluster affording environmental adaptation. *PLoS Genet.* 9: e1003707.
- Heuertz, M., E. De Paoli, T. Källman, H. Larsson, I. Jurman *et al.*, 2006 Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce. [*Picea abies* (L.) Karst] *Genetics* 174: 2095–2105.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, 2005 Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25: 1965–1978.
- Hoffmann, A. A., and C. M. Sgrò, 2011 Climate change and evolutionary adaptation. *Nature* 470: 479–485.
- Hughes, L., 2000 Biological consequences of global warming: Is the signal already apparent? *Trends Ecol. Evol.* 15: 56–61.
- Ioannidis, J. P. A., G. Thomas, and M. J. Daly, 2009 Validating, augmenting and refining genome-wide association signals. *Nat. Rev. Genet.* 10: 318–329.
- Jakobsson, M., and N. A. Rosenberg, 2007 CLUMPP: a software matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801–1806.
- Kearney, M., W. P. Porter, C. Williams, S. Ritchie, and A. A. Hoffmann, 2009 Integrating biophysical models and evolutionary theory to predict climatic impacts on species' ranges: the dengue mosquito *Aedes aegypti* in Australia. *Funct. Ecol.* 23: 528–538.
- König, I. K., 2011 Validation in genetic association studies. *Brief. Bioinform.* 12: 253–258.
- Kremer, A., and V. Le Corre, 2012 Decoupling of differentiation between traits and their underlying genes in response divergent selection. *Heredity* 108: 375–385.
- Kremer, A., O. Ronce, J. J. Robledo-Arnuncio, F. Guillaume, G. Bohrer *et al.*, 2012 Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecol. Lett.* 15: 378–392.
- Kuparinen, A., O. Savolainen, and F. M. Schurr, 2010 Increased mortality can promote evolutionary adaptation of forest trees to climate change. *For. Ecol. Manage.* 259: 1003–1008.
- Lepoittevin, C., L. Harvengt, C. Plomion, and P. H. Garnier-Géré, 2012 Association mapping for growth, straightness and wood chemistry traits in the *Pinus pinaster* Aquitaine breeding population. *Tree Genet. Genomes* 8: 113–126.
- Loarie, S. R., P. B. Duffy, H. Hamilton, G. P. Asner, C. B. Field *et al.*, 2009 The velocity of climate change. *Nature* 462: 1052–1055.
- Lorenz, W. W., R. Alba, Y. S. Yu, J. M. Bordeaux, M. Simões *et al.*, 2011 Microarray analysis and scale-free gene networks identify candidate regulators in drought-stressed roots of loblolly pine (*P. taeda* L.). *BMC Genomics* 12: 264.
- Lotterhos, K. E., and M. C. Whitlock, 2014 Evaluation of demographic history and neutral parameterization on the performance of F_{st} outlier tests. *Mol. Ecol.* 23: 2178–2192.
- Lynch, M., and R. Lande, 1993 Evolution and extinction in response to environmental change, pp. 234–250 in *Biotic Interactions and Global Change*, edited by P. M. Kareiva, J. G. Kingsolver, and R. B. Huey. Sinauer Associates, Sunderland, MA.
- Malcom, J. R., A. Markham, R. P. Neilson, and M. Garaci, 2011 Estimated migration rates under scenarios of global climate change. *J. Biogeogr.* 29: 835–849.
- McLachlan, J. S., J. S. Clark, and P. S. Manos, 2005 Molecular indicators of tree migration capacity under rapid climate change. *Ecology* 86: 2088–2098.
- McLeod, T. K., and G. MacDonald, 1997 Postglacial range expansion and population growth of *Picea mariana*, *P. glauca* and *Pinus banksiana* in the western interior of Canada. *J. Biogeogr.* 24: 865–881.
- Morin, P. A., G. Luikart, and R. K. Wayne, and the SNP workshop Group, 2004 SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.* 19: 208–216.
- Mosca, E., A. J. Eckert, E. A. Di Pierro, D. Rocchini, N. La Porta *et al.*, 2012 The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Mol. Ecol.* 21: 5530–5545.
- Namroud, M. C., J. Beaulieu, N. Juge, J. Laroche, and J. Bousquet, 2008 Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Mol. Ecol.* 17: 3599–3613.
- Neale, D. B., and P. K. Ingvarsson, 2008 Population, quantitative and comparative genomics of adaptation in forest trees. *Curr. Opin. Plant Biol.* 11: 149–155.
- Neale, D. B., and O. Savolainen, 2004 Association genetics of complex traits in conifers. *Trends Plant Sci.* 9: 325–330.
- Neale, D. B., J. L. Wegrzyn, K. A. Stevens, A. V. Zimin, D. Puiu *et al.*, 2014 Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol.* 15: R59.
- Nystedt, B., N. Street, A. Wetterbom, A. Zuccolo, Y. C. Lin *et al.*, 2013 The Norway spruce genome sequence and conifer genome evolution. *Nature* 497: 579–584.
- O'Connell, L. M., A. Mosseler, and O. P. Rajora, 2006 Impact of forest fragmentation on the mating system and genetic diversity of white spruce (*Picea glauca*) at the landscape level. *Heredity* 97: 418–426.
- Perdiguerro, P., M. C. Barbero, M. T. Cervera, C. Collada, and A. Soto, 2013 Molecular responses to water stress in two contrasting Mediterranean pines (*Pinus pinaster* and *Pinus pinea*). *Plant Physiol. Biochem.* 67: 199–208.
- Petit, R. J., and A. Hampe, 2006 Some evolutionary consequences of being a tree. *Annu. Rev. Ecol. Evol. Syst.* 37: 187–214.
- Pot, D., L. McMillan, C. Echt, G. Le Provost, P. H. Garnier-Géré *et al.*, 2005 Nucleotide variation in genes involved in wood formation in two pine species. *New Phytol.* 167: 101–112.
- Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Prunier, J., J. Laroche, J. Beaulieu, and J. Bousquet, 2011 Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Mol. Ecol.* 20: 1702–1716.
- Prunier, J., S. Gerardi, J. Beaulieu, and J. Bousquet, 2012 Parallel and lineage-specific molecular adaptation to climate in boreal black spruce. *Mol. Ecol.* 21: 4270–4286.
- R Development Core Team, 2013 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: <http://www.R-project.org>.
- Rangel, T. F., J. A. F. Diniz-Filho, and L. M. Bini, 2010 SAM: a comprehensive application of spatial analysis in macroecology. *Ecography* 33: 46–50.
- Rehfeldt, G. E., C. C. Ying, D. L. Spittlehouse, and D. A. Hamilton, 1999 Genetic responses to climate in *Pinus contorta*: niche

- breadth, climate change, and reforestation. *Ecol. Monogr.* 69: 375–407.
- Rubiales, J. M., I. García-Amorena, L. Hernández, M. Génova, F. Martínez *et al.*, 2010 Late quaternary dynamics of pinewoods in the Iberian Mountains. *Rev. Palaeobot. Palynol.* 162: 476–491.
- Santos-del-Blanco, L., J. Climent, S. C. González-Martínez, and J. R. Pannell, 2012 Genetic differentiation for size at first reproduction through male vs. female functions in the widespread Mediterranean tree *Pinus pinaster*. *Ann. Bot.* 110: 1449–1460.
- Savolainen, O., T. Pyhäjärvi, and T. Knürr, 2007 Gene flow and local adaptation in trees. *Annu. Rev. Ecol. Evol. Syst.* 38: 595–619.
- Savolainen, O., S. T. Kujala, C. Sokol, T. Pyhäjärvi, K. Avia *et al.*, 2011 Adaptive potential of northernmost tree populations to climate change, with emphasis on Scots pine (*Pinus sylvestris* L.). *J. Hered.* 102: 526–536.
- Seki, M., M. Narusaka, J. Ishida, T. Nanjo, M. Fujita *et al.*, 2002 Monitoring the expression profile of 7000 Arabidopsis genes under drought, cold and high salinity stresses using a full-length cDNA microarray. *Plant J.* 31: 279–292.
- Tanino, K. K., L. Kalcsits, S. Silim, E. Kendall, and G. R. Gray, 2010 Temperature-driven plasticity in growth cessation and dormancy development in deciduous woody plants: a working hypothesis suggesting how molecular and cellular function is affected by temperature during dormancy induction. *Plant Mol. Biol.* 73: 49–65.
- Tenaillon, O., A. Rodríguez-Vergugo, R. L. Gaut, P. McDonald, A. F. Bennett *et al.*, 2012 The molecular diversity of adaptive convergence. *Science* 335: 457–461.
- Tiffin, P., and J. Ross-Ibarra, 2014 Advances and limits of using population genetics to understand local adaptation. *Trends Ecol. Evol.* 29: 673–680.
- Urano, K., Y. Kurihara, M. Seki, and K. Shinozaki, 2010 “Omics” analyses of regulatory networks in plant abiotic stress responses. *Curr. Opin. Plant Biol.* 13: 132–138.
- Wagner, A., L. Donaldson, H. Kim, L. Phillips, H. Flint *et al.*, 2009 Suppression of 4-Coumarate-CoA ligase in the coniferous gymnosperm *Pinus radiata*. *Plant Physiol.* 149: 370–383.
- Westbrook, J. W., M. F. Resende, Jr., P. Munoz, A. R. Walker, J. L. Wegrzyn *et al.*, 2013 Association genetics of oleoresin flow in loblolly pine: discovering genes and predicting phenotype for improved resistance to bark beetles and bioenergy potential. *New Phytol.* 199: 89–100.
- Yakovlev, I. A., C. G. Fossdal, and Ø. Johnsen, 2010 MicroRNAs, the epigenetic memory and climate adaptation in Norway spruce. *New Phytol.* 187: 1154–1169.
- Yeaman, S., and A. Jarvis, 2006 Regional heterogeneity and gene flow maintain variance in a quantitative trait within populations of lodgepole pine. *Proc. R. Soc. B Biol. Sci.* 273: 1587–1593.
- Yun, M. S., W. J. Chen, F. Deng, and Y. Yogo, 2005 Selective growth suppression of five annual plant species by chalcone and naringenin correlates with the total amount of 4-coumarate: coenzyme A ligase. *Weed Biol. Manage.* 9: 27–37.
- Zhu, K., C. W. Woodall, and J. S. Clark, 2011 Failure to migrate: lack of tree range expansion in response to climate change. *Glob. Change Biol.* 18: 1042–1052.

Communicating editor: S. I. Wright

GENETICS

Supporting Information

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.173252/-/DC1>

Molecular Proxies for Climate Maladaptation in a Long-Lived Tree (*Pinus pinaster* Aiton, Pinaceae)

Juan-Pablo Jaramillo-Correa, Isabel Rodríguez-Quilón, Delphine Grivet, Camille Lepoittevin,
Federico Sebastiani, Myriam Heuertz, Pauline H. Garnier-Géré, Ricardo Alía,
Christophe Plomion, Giovanni G. Vendramin, and Santiago C. González-Martínez

Supporting Information

Molecular Proxies for Climate Maladaptation in a Long-Lived Tree

(Pinus pinaster Aiton, Pinaceae)

Juan P. Jaramillo-Correa, Isabel Rodríguez-Quilón, Delphine Grivet, Camille Lepoittevin, Federico Sebastiani, Myriam Heuertz, Pauline Garnier-Géré, Ricardo Alía, Christophe Plomion, Giovanni G. Vendramin, Santiago C. González-Martínez*

*To whom correspondence should be addressed. E-mail: santiago@inia.es

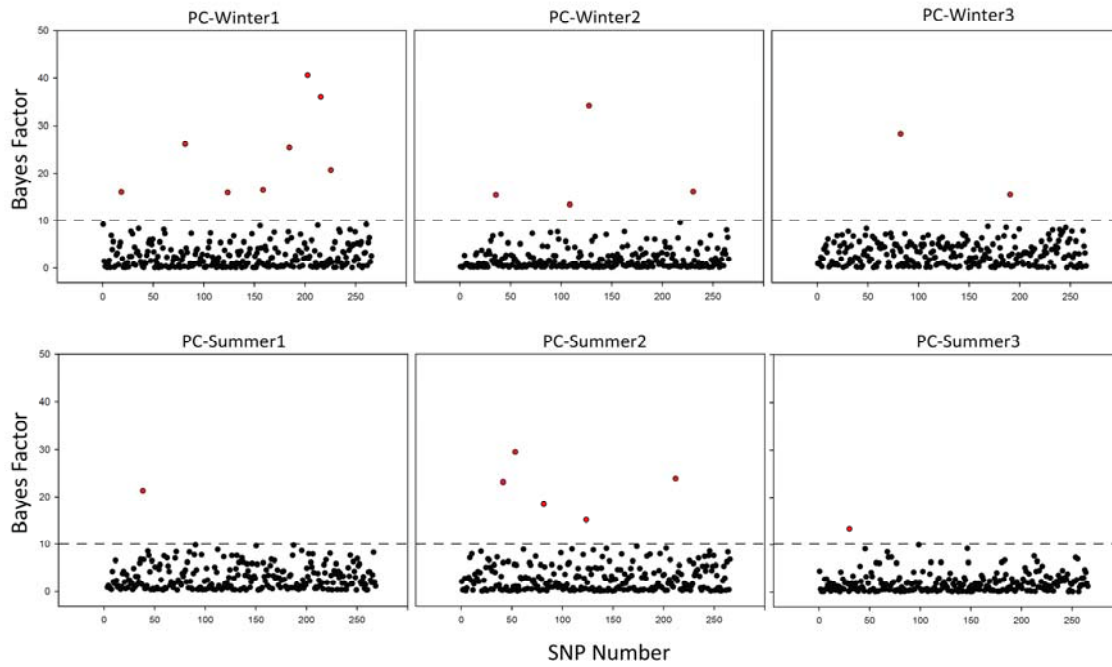


Figure S1 Bayes Factors (BFs) describing the deviation of the genotype-environment associations of 266 putative candidate SNPs from a null demographic model (i.e. a co-variance matrix built with 1,745 control SNPs) obtained from BAYENV 2.0. Red dots are those SNPs above the BF=10 threshold (horizontal discontinuous line), and which were assumed to be significantly associated with climate.

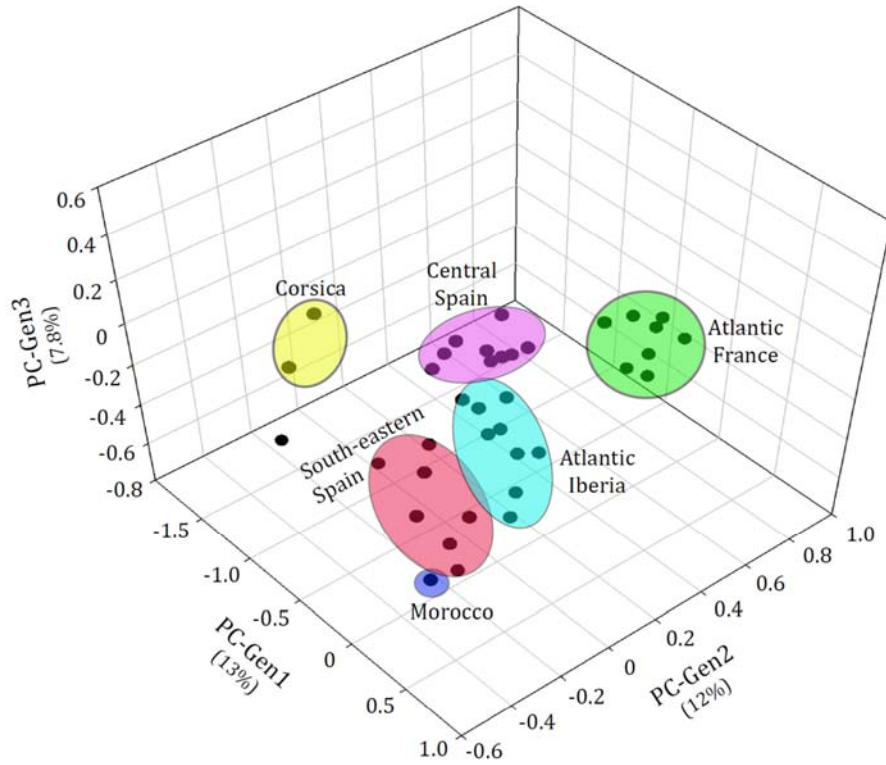


Figure S2 Scatterplot of the population scores for the first three principal components derived from the variation of nine nuclear SSRs on 36 natural stands of maritime pine.

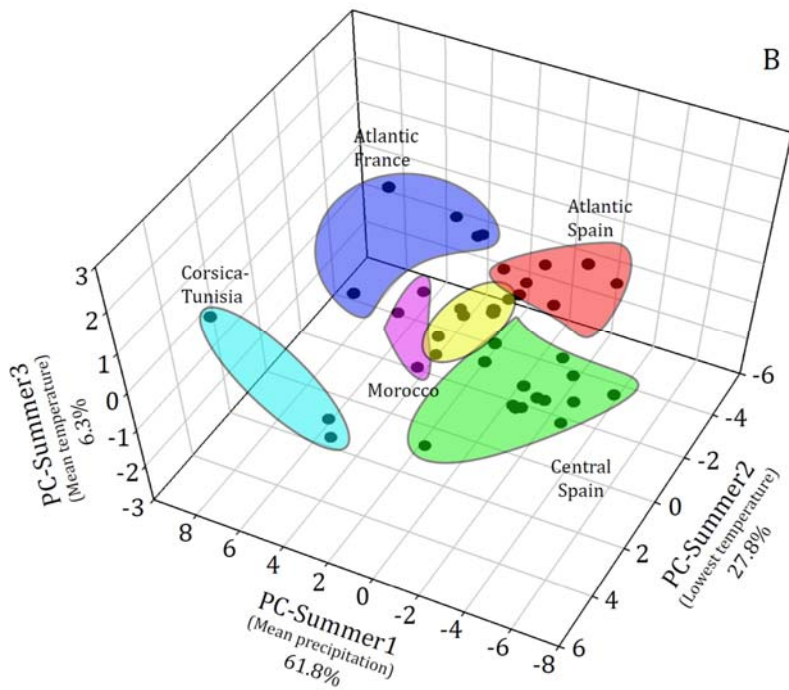
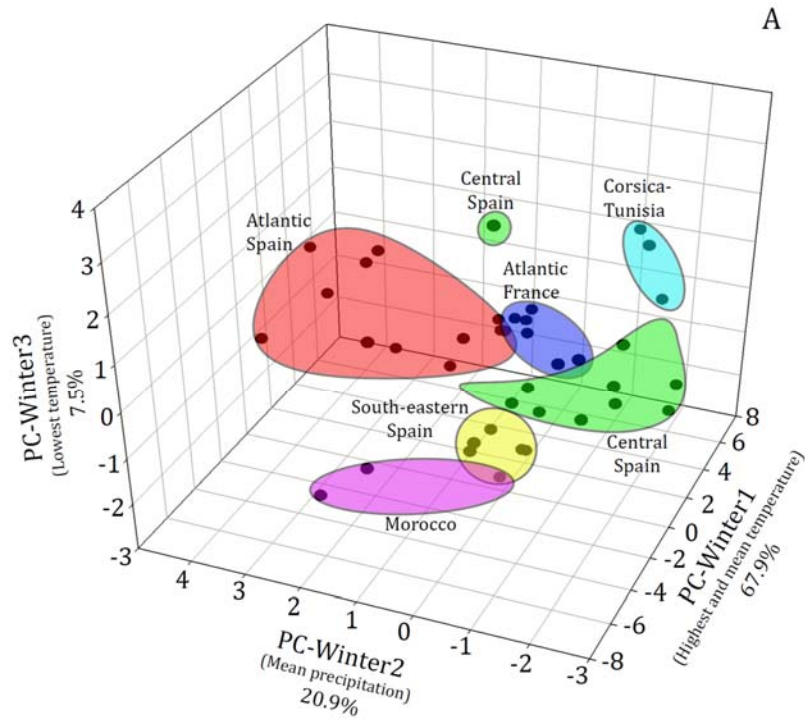


Figure S3 Scatterplot of the population scores for the first three principal components derived from the (A) winter and (B) summer climate variation of 36 natural stands of maritime pine.

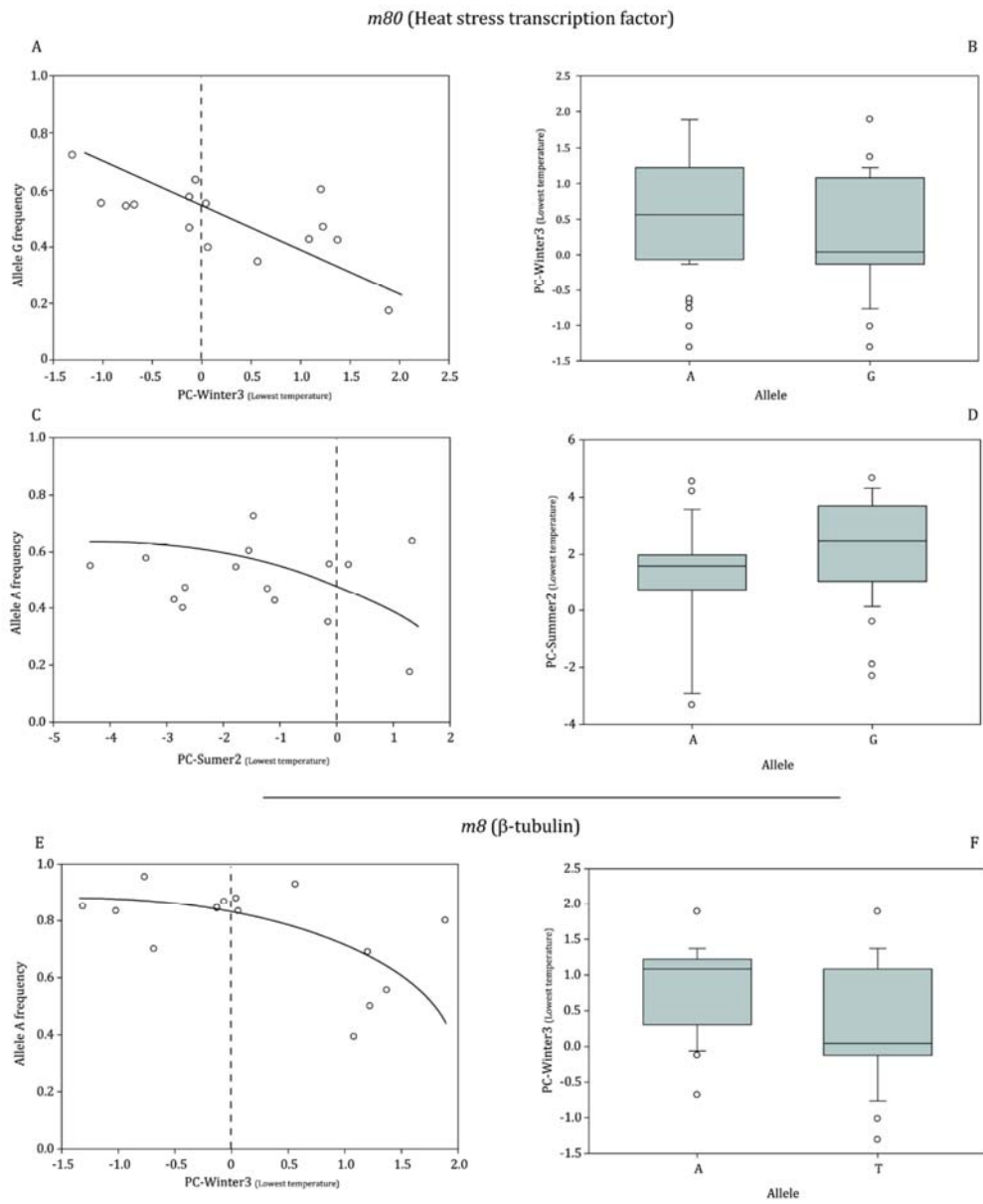


Figure S4 (continued in next page)

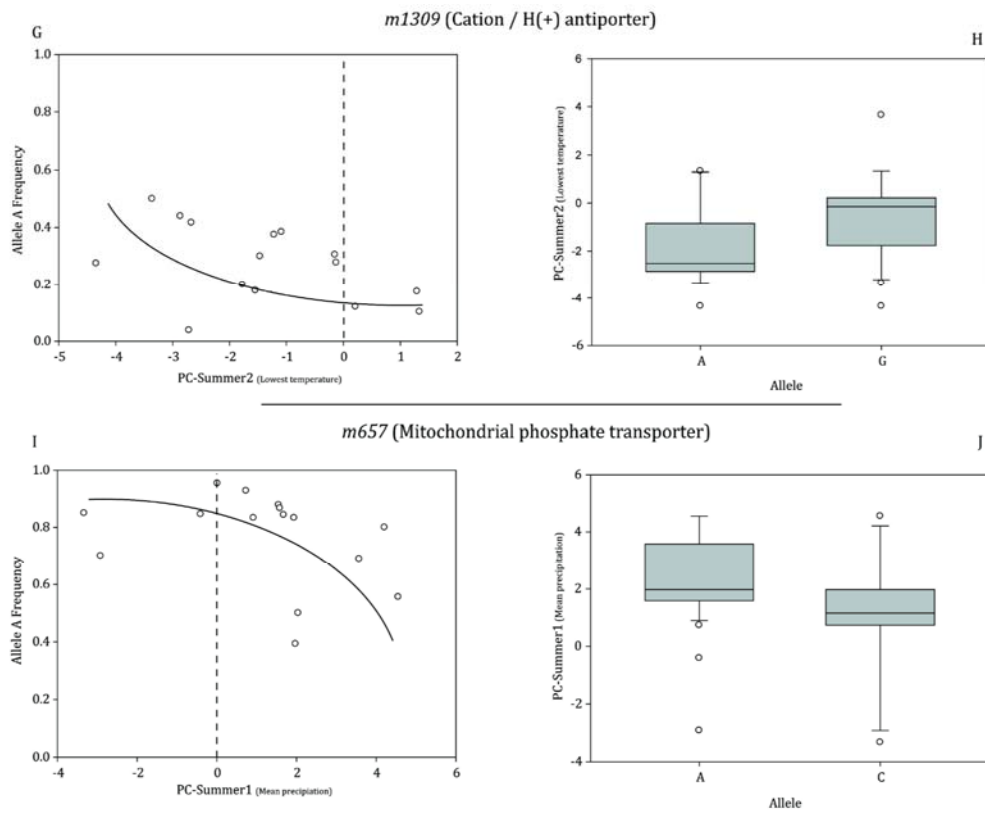


Figure S4 Allele-climate associations obtained for maritime pine populations in Mediterranean Spain. (A, C, E, G, I) Scatterplots and (B, D, F, H, J) box-plots are shown for all SNPs exhibiting significant associations (see main text). Lines within scatterplots indicate clines of allele frequencies under a logistic regression model. Boxes denote the interquartile range and horizontal lines within boxes represent genotypic means.

File S1

Performance of the covariance matrix in null models of Bayesian environmental association using BAYENV 2.0

When performing Bayesian association analyses, it is important to verify that the covariance matrix captures adequately the underlying population structure of the taxon of interest (Coop *et al.* 2009; Günther and Coop 2013). Herein, this matrix was built from variation at 1,745 control SNPs that were assumed to be good representatives of the whole-genome neutral variation of maritime pine (see Materials and Methods in the main text). Control and candidate SNP markers had similar allele frequency distributions, as shown by a Kolmogorov-Smirnov test. In particular, control SNPs did not seem to be enriched by rare or common alleles, which could have indicated different levels of purifying selection (see Nielsen 2005). However, given the absence of factual information concerning their neutrality, some unnoticed non-stochastic bias might be present in the covariance matrix derived from these markers, as virtually any gene-derived SNP is a potential target of selective processes. For instance, if some of these SNPs were involved in the same adaptive processes aimed to be detected, they may have modified the covariance matrix in a way that does not fully represent the underlying neutral population structure, and thus would have reduced the statistical power for environmental associations.

To validate this covariance matrix, three different comparisons were performed. First, the matrices produced by each BAYENV 2.0 run were compared to each other with correlation tests, which were all highly significant (mean $r^2 = 0.943$; $P < 0.0001$), indicating that independent runs were converging to similar results. These covariance matrices were afterwards converted into correlation matrices using the *cov2cor* function in R and compared to a pairwise- F_{ST} matrix derived from this same SNP dataset and to a second matrix derived from SSR markers. Again, the correlation between matrices was extremely high (mean $r^2 = 0.974$; $P < 0.001$ for the SNP- F_{ST} matrix; mean $r^2 = 0.812$; $P < 0.005$ for the SSR one), implying that the covariance matrices were reflecting similar population differentiation values than the F_{ST} statistics. Finally, two independent STRUCTURE analyses were performed for the control-SNP and SSR datasets and their Q -matrices were correlated. For both types of markers, the best partition was that of six genetic groups (Figure 1), which matched perfectly those previously reported in other studies of this species (Bucci *et al.* 2007; Santos-del-Blanco *et al.* 2012). The correlation coefficients (Pearson's r) between the Q -matrices were relatively high and significant (correlation coefficients of 0.662; $P < 0.01$), although the groups obtained from the SNP dataset appeared better resolved (i.e. with less admixed individuals) than those determined with SSRs (see Figure 1). This is probably due to lower homoplasy and higher number of markers in the SNP dataset, which then should be more adequate to capture the underlying population structure of maritime pine than SSRs. This finding strongly supports our choice of the SNP dataset (and not

the SSRs) to compute the covariance matrix in null models for Bayesian association analyses. It also explains why correlation coefficients at the individual level are just moderately high between SNP and SSR Q-matrices.

LITERATURE CITED

- Bucci, G., S. C. González-Martínez, G. Le Provost, C. Plomion, M. M. Ribeiro *et al.*, 2007 Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Mol. Ecol.* **16**: 2137-2153.
- Coop, G., J. K. Pickrell, J. Novembre, S. Kudaravalli, J. Li *et al.*, 2009 The role of geography in human adaptation. *PloS Genetics* **5**: e1000500.
- Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard, 2010 Using environmental correlations to identify loci under local adaptation. *Genetics* **185**: 1411-1423.
- Günther, T., and G. Coop, 2013 Robust identification of local adaptation from allele frequencies. *Genetics* **195**: 205-220.
- Nielsen, R., 2005 Molecular signatures of natural selection. *Annu. Rev. Genet.* **39**:197-218.
- Santos-del-Blanco, L., J. Climent, S. C. González-Martínez, and J. R. Pannell, 2012 Genetic differentiation for size at first reproduction through male versus female functions in the widespread Mediterranean tree *Pinus pinaster*. *Ann. Bot.* **110**: 1449-1460.

File S2

Design file for the 384-SNP (Single Nucleotide Polymorphism) candidate gene OPA used to genotype 36 populations of maritime pine (*Pinus pinaster*)

File S2 is available for download as an Excel file at

<http://www.genetics.org/lookup/supp/doi:10.1534/genetics.114.173252/-/DC1>

Table S1 Mean and standard errors for the number of individuals genotyped (N), the observed number of alleles (N_a), the effective number of alleles (N_e), the observed (H_o) and expected (H_e) heterozygosity, and the consanguinity index (F) estimated for 36 populations of maritime pine (*Pinus pinaster*) with 12 microsatellite loci. The number of populations deviating from Hardy-Weinberg Equilibrium (HWE) at $P < 0.01$ is shown in the last column. Shaded loci were removed from further analyses.

Locus	N		N_a		N_e		H_o		H_e		F		HWE ^a
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	
<i>A6F03</i>	19.436	2.092	4.667	0.189	2.854	0.115	0.568	0.021	0.622	0.019	0.075	0.028	4
<i>3pet</i>	10.846	1.503	5.923	0.707	3.884	0.459	0.568	0.060	0.572	0.059	0.006	0.021	5
<i>NZPR1078</i>	19.231	2.001	3.154	0.086	2.366	0.073	0.609	0.023	0.560	0.015	-0.091	0.033	1
<i>NZPR544</i>	19.410	2.059	2.359	0.107	1.881	0.046	0.316	0.025	0.457	0.013	0.311	0.049	12 (0.312)
<i>ctg4363</i>	20.487	2.125	4.154	0.174	2.701	0.082	0.650	0.024	0.617	0.011	-0.053	0.035	3
<i>rptest1</i>	19.923	2.142	3.692	0.133	2.770	0.115	0.646	0.024	0.612	0.018	-0.057	0.027	1
<i>ctg275</i>	20.385	2.142	7.949	0.380	4.090	0.229	0.649	0.027	0.721	0.018	0.097	0.033	12 (0.191)
<i>2669</i>	19.795	2.052	4.026	0.231	1.985	0.099	0.448	0.034	0.442	0.030	-0.013	0.031	5
<i>epi5</i>	19.077	2.090	3.128	0.192	1.605	0.071	0.274	0.026	0.333	0.027	0.133	0.047	2
<i>epi6</i>	15.974	1.325	3.974	0.239	2.499	0.157	0.298	0.031	0.553	0.031	0.458	0.041	19 (0.283)
<i>gPp14</i>	19.513	2.048	2.923	0.129	1.605	0.075	0.315	0.028	0.329	0.028	0.016	0.035	3
<i>epi3</i>	18.538	2.072	5.154	0.274	2.953	0.153	0.621	0.029	0.634	0.015	0.027	0.034	4

^aAverage null-allele frequency for discarded loci in populations deviating from HWE is given between parentheses, as estimated with MICRO-CHECKER software (vanOosterhout *et al.* 2004).

Literature cited: vanOosterhout, C., W. F. Hutchinson, D. P. M. Willis, and P. Shipley, 2004 Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**: 535-538.

Table S2 Percentage of the variance explained (% Variance), factor loads, and population scores for the top-three principal components summarizing winter (December to March) and summer (June to September) climatic variation in 36 populations of maritime pine (*Pinus pinaster*).

	Winter			Summer		
	PC1	PC2	PC3	PC1	PC2	PC3
% Variance	67.861	20.981	7.462	61.752	27.838	6.299
<i>Factor</i>						
Mean Temperature - December	0.966	-0.132	-0.002	---	---	---
Mean Temperature - January	0.980	-0.148	-0.087	---	---	---
Mean Temperature - February	0.982	-0.175	0.015	---	---	---
Mean Temperature - March	0.949	-0.235	0.020	---	---	---
Highest Temperature - December	0.927	-0.063	0.291	---	---	---
Highest Temperature - January	0.869	-0.090	0.468	---	---	---
Highest Temperature - February	0.717	-0.198	0.580	---	---	---
Highest Temperature - March	0.941	-0.089	0.209	---	---	---
Lowest Temperature - December	0.911	-0.197	-0.352	---	---	---
Lowest Temperature - January	0.927	-0.206	-0.306	---	---	---
Lowest Temperature - February	0.914	-0.212	-0.314	---	---	---
Lowest Temperature - March	0.921	-0.191	-0.325	---	---	---
Mean Precipitation - December	0.527	0.832	-0.018	---	---	---
Mean Precipitation - January	0.489	0.862	-0.007	---	---	---
Mean Precipitation - February	0.428	0.843	0.004	---	---	---
Mean Precipitation - March	0.514	0.820	-0.067	---	---	---
Mean Temperature - June	---	---	---	0.882	0.356	0.200
Mean Temperature - July	---	---	---	0.973	-0.050	0.138
Mean Temperature - August	---	---	---	0.979	0.055	0.126
Mean Temperature - September	---	---	---	0.816	0.511	0.104
Highest Temperature - June	---	---	---	0.879	-0.316	0.305
Highest Temperature - July	---	---	---	0.803	-0.546	0.208
Highest Temperature - August	---	---	---	0.842	-0.489	0.213
Highest Temperature - September	---	---	---	0.893	-0.238	0.292
Lowest Temperature - June	---	---	---	0.414	0.887	-0.022
Lowest Temperature - July	---	---	---	0.590	0.780	-0.069
Lowest Temperature - August	---	---	---	0.576	0.795	-0.078
Lowest Temperature - September	---	---	---	0.346	0.918	-0.112
Mean Precipitation - June	---	---	---	-0.855	0.083	0.439
Mean Precipitation - July	---	---	---	-0.877	0.281	0.302
Mean Precipitation - August	---	---	---	-0.789	0.442	0.374

Mean Precipitation - September		---	---	---	-0.718	0.464	0.455
<i>Population</i>	Country						
Hourtin	France	-0.020	-0.390	-0.538	-1.760	1.482	0.943
Leverdon	France	-0.376	-0.670	-0.896	-1.447	1.593	0.488
Mimizan	France	1.440	0.315	0.108	-2.198	2.250	2.205
Olonne sur Mer	France	-0.863	-0.540	-1.626	-2.395	1.209	-0.676
Petrocq	France	1.414	0.380	0.113	-2.228	2.269	2.284
Pleucadec	France	-1.990	-0.300	-1.927	-3.886	0.368	-1.370
St-Jean des Monts	France	-0.866	-0.471	-1.626	-2.629	1.027	-0.800
Alto de la Llama	Spain	0.250	1.066	-0.491	-3.580	-0.714	-0.810
Armayán	Spain	0.235	0.818	0.169	-3.134	-0.640	-0.383
Cadavedo	Spain	3.377	0.502	-0.196	-3.718	1.747	0.129
Sierra de Barcia	Spain	3.123	0.753	-0.209	-3.821	1.446	-0.100
Castropol	Spain	2.476	0.102	-0.184	-4.465	0.576	-0.543
Lamuño	Spain	3.643	0.342	-0.277	-3.111	2.093	-0.049
Puerto de Vega	Spain	3.384	0.199	0.069	-3.149	1.732	0.303
Rodoiros	Spain	2.126	2.486	-0.984	-5.861	0.260	-0.260
San Cipriano de Ribaterme	Spain	1.915	3.717	0.025	-2.063	-1.495	0.296
Leiria	Portugal	5.809	-1.742	0.065	1.149	1.582	-2.328
Pineta	France (Corsica)	3.006	-2.410	-1.465	3.803	3.400	-1.500
Pinia	France (Corsica)	3.188	-2.479	-0.904	3.947	3.266	-1.108
Tabuyo del Monte	Spain	-6.793	1.849	-1.309	-3.328	-1.461	1.535
Arenas de San Pedro	Spain	1.213	2.568	1.372	4.554	-1.084	2.135
Valdemaqueda	Spain	-2.344	-0.548	-0.126	1.681	-1.214	-0.339
Cenicientos	Spain	-1.699	-0.071	-1.018	1.941	-0.123	-0.996
Coca	Spain	-2.842	-1.725	1.224	2.042	-2.668	0.641
Cuellar	Spain	-2.915	-1.405	1.082	1.975	-2.861	0.590
Bayubas de Abajo	Spain	-5.089	-0.448	-0.127	-0.406	-3.358	0.008
San Leonardo	Spain	-6.337	1.121	-0.683	-2.916	-4.342	-0.540
Boniches	Spain	-3.909	-0.146	-0.766	0.014	-1.770	-0.092
Olba	Spain	-2.068	-2.192	0.565	0.736	-0.146	0.257
Sinarcas	Spain	-2.963	-1.836	0.041	1.554	0.215	-0.242
Sierra Calderona	Spain	-1.294	-2.118	-0.064	1.584	1.339	-0.410
Quatretonda	Spain	3.256	-1.819	1.889	4.210	1.295	0.944
Cazorla	Spain	-2.020	4.007	0.061	0.922	-2.711	0.002
Oria	Spain	-1.341	-2.193	1.204	3.574	-1.543	-0.429
Tamrabta	Morocco	-4.910	1.378	1.428	0.604	-3.725	-0.828
Tabarka	Tunisia	7.272	-1.045	1.043	8.234	3.671	0.930

Table S3 Strength of genotype-environment correlations (Akaike Information Criterion in multivariate logistic models –*mlr*, *P*-values in latent factor mixed models, and Bayes Factors in Bayesian environmental associations with BAYENV 2.0) for 18 maritime pine SNPs associated with climate.

SNP name (number; see Fig. S1)	Climate PC	<i>mlr</i>	LFMM	BAYENV ^a	
		AIC	<i>P</i>	BF1	BF2
<i>m783</i> (227)	PC-Winter1	21.6	1.28×10^{-19}	36.81	39.64
<i>m8</i> (89)	PC-Winter1	9.3	5.95×10^{-19}	26.43	25.62
<i>m1513</i> (53)	PC-Summer2	15.4	1.29×10^{-15}	29.41	33.84
<i>m1250</i> (142)	PC-Winter2	29.7	9.09×10^{-20}	34.46	38.03
<i>m80</i> (218)	PC-Summer2	7.3	2.18×10^{-23}	22.31	21.04
<i>m80</i> (218)	PC-Winter1	12.4	8.95×10^{-25}	40.84	38.89
<i>m1309</i> (176)	PC-Winter1	10.1	5.69×10^{-21}	25.37	24.12
<i>m100</i> (93)	PC-Winter3	13.6	8.21×10^{-11}	29.13	27.48
<i>m100</i> (93)	PC-Summer2	9.6	9.96×10^{-13}	18.46	15.84
<i>m973</i> (118)	PC-Winter2	10.2	7.65×10^{-11}	12.33	12.98
<i>m137</i> (46)	PC-Summer1	7.3	8.73×10^{-15}	20.31	18.44
<i>m607</i> (194)	PC-Winter3	18.5	4.12×10^{-15}	16.78	18.44
<i>m1156</i> (131)	PC-Winter1	16.4	1.24×10^{-11}	14.16	13.96
<i>m1211</i> (159)	PC-Winter1	22.8	2.34×10^{-24}	15.26	14.01
<i>m646</i> (19)	PC-Winter1	11.6	3.93×10^{-21}	17.29	15.52
<i>m657</i> (238)	PC-Winter1	15.9	2.21×10^{-23}	20.11	18.34
<i>m658</i> (47)	PC-Summer2	32.1	5.45×10^{-19}	21.91	22.43
<i>m685</i> (41)	PC-Winter2	16.0	7.52×10^{-11}	14.32	12.97
<i>m685</i> (41)	PC-Summer3	19.8	3.78×10^{-16}	12.58	11.86
<i>m1196</i> (238)	PC-Winter2	24.3	2.33×10^{-14}	14.87	16.12
<i>m705</i> (127)	PC-Summer2	13.6	3.18×10^{-15}	14.31	15.96

^aResults are shown for two independent BAYENV 2.0 runs made with a covariance matrix built using 1,745 control SNPs (see Materials and Methods for more details).

Table S4 Single-locus regressions of advantageous-allele frequency with survival estimated under extreme climate conditions for each of 18 SNPs associated with climate in maritime pine (*Pinus pinaster*).

SNP name	Advantageous allele	DF	<i>F</i>	<i>P</i>	<i>Adj-r</i> ²
<i>m783</i>	A	17	4.037	0.061	0.144
<i>m8</i>	A	17	5.773	0.028	0.210
<i>m1513</i>	C	17	5.182	0.036	0.188
<i>m1250</i>	A	17	0.095	0.762	-0.053
<i>m80</i>	A	17	7.178	0.016	0.256
<i>m1309</i>	G	17	7.865	0.012	0.276
<i>m100</i>	A	17	3.929	0.064	0.140
<i>m973</i>	G	17	1.976	0.178	0.051
<i>m137</i>	G	17	0.133	0.720	-0.051
<i>m607</i>	C	17	2.083	0.167	0.057
<i>m1156</i>	G	17	0.085	0.774	-0.054
<i>m1211</i>	C	17	8.380	0.010	0.291
<i>m646</i>	G	17	0.802	0.383	-0.011
<i>m657</i>	A	17	5.891	0.027	0.214
<i>m658</i>	G	17	0.825	0.376	0.046
<i>m685</i>	A	17	2.484	0.133	0.076
<i>m1196</i>	C	17	2.284	0.149	0.067
<i>m705</i>	A	17	5.485	0.032	0.200

Table S5 Correlation between the 18 SNPs that best explained each Principal Component (PC) in a PCA including all 266 candidate-gene SNPs and survival estimates obtained in a common garden under extreme (dry and hot) climate; ns: not significant.

Principal Components	% variance explained	Pearson's correlation coefficient (<i>r</i>)
PC1	24.30	0.28 (ns)
PC2	15.55	0.27 (ns)
PC3	10.23	-0.41 (ns)
PC4	9.16	0.26 (ns)

Table S6 Correlation coefficients for the non-linear models that best fitted the climate principal components (see Table S2) of maritime pine (*Pinus pinaster*) populations with their geographic location (latitude and longitude) and population structure loads (control SNP and SSR datasets). ** $P < 0.01$

Climate PC	Geography (latitude - longitude)	Population Structure
PC-Winter1	0.176	0.179
PC-Winter2	0.140	0.153
PC-Winter3	0.849**	0.593**
PC-Summer1	0.196	0.181
PC-Summer2	0.392**	0.706**
PC-Summer3	0.110	0.114