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

J. P. Jaramillo-Correa,*,§ I. Rodríguez-Quilón,*, D. Grivet,*, C. Lepoittevin,†,‡ F. Sebastiani,**

M. Heuertz,*,†,‡ P.H. Garnier-Géré,†,‡ R. Alía,*, C. Plomion,†,‡ G. G. Vendramin,** S. C. González-Martínez*§§

*Department of Ecology and Genetics, Forest Research Centre, INIA-CIFOR, E-28040 Madrid, Spain

§Department of Evolutionary Ecology, Institute of Ecology, UNAM, AP 70-275, México, D.F., Mexico

†INRA, UMR1202 BIOGECO, F-33610 Cestas, France

‡University of Bordeaux, UMR1202 BIOGECO, F-33170 Talence, France

**Institute of Biosciences and Bioresources, National Research Council, I-50019 Sesto Fiorentino (Firenze), Italy

§§Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland


Running title: Proxies for climate maladaptation in pine

Copyright 2014.
Keywords: climate adaptation, environmental associations, genetic lineages, Single Nucleotide Polymorphisms, fitness estimates

To whom correspondence should be addressed:
Santiago C. González-Martínez
Department of Ecology and Genetics,
Forest Research Centre, INIA-CIFOR,
E-28040 Madrid, Spain
Fax: +34 913476767
E-mail: santiago@inia.es
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, five 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 establishing mitigation measures for long-lived keystone forest trees in the face of impending climate change.
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 to 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 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 palaeobotanical 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 south-western Europe, inhabiting both wet-coastal and seasonally-dry continental forests, and which 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 2,646 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 over 600,000 sequences obtained from a set of 17 cDNA libraries of different tissues without any prior experimental treatment, and which 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 (File S2 in Supplementary Information; 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 over-expressed under abiotic stress (Lorenz et al. 2011; Perdiguerro et al. 2013) and/or which 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 re-sequencing 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) above 5% and call rates above 0.8 (average of 0.97) were retained (1,745 and 266 control and candidate SNPs, respectively). Scoring was performed with BeadStudio ver. 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 double-checked for all SNPs.

SSR genotyping and phylogeographic analyses: To corroborate that historical and demographic processes were adequately accounted for, twelve nuclear microsatellites (SSRs) distributed in eight of the twelve 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 ver. 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 one to fifteen, 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 dataset 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 more than 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 more than 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 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, ten 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 dataset (1,745 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 20...
201201) To counterbalance the possibility that any gene-derived marker can be potentially 202involved in adaptation and can therefore be suboptimal to capture the population structure in the 203covariance matrix, a series of analyses was performed to validate its suitability. First, the 204covariance matrices produced by three independent BAYENV 2.0 runs were compared to each 205other to verify convergence to similar results. Second, these matrices were transformed to 206correlation matrices with the cov2cor function in R v 3.0.0 and correlated to pairwise-$F_{ST}$ 207matrices generated in Arlequin ver. 3.5 (Excoffier et al. 2010) from this and the SSR datasets 208using Pearson’s coefficient in R v 3.0.0. Third, the $Q$-matrices produced in the STRUCTURE 209clustering analyses for the two putatively neutral datasets (control SNPs and SSRs) were 210similarly correlated in R v 3.0.0.

211After these validation steps (see File S1 in Supplementary Information), a Bayes factor (BF) 212describing the deviation between the null demographic model (i.e. the covariance matrix) and the 213alternative one (the candidate SNP-climate correlations) was estimated for each SNP. Model 214convergence was assured by performing three independent runs of 100,000 iterations and a burn- 215in of 10,000 chains with random seeds. Statistical support was assessed by Spearman’s rank 216correlation ($\rho$) tests for those associations exhibiting unusually high BF (Eckert et al. 2010b; De 217La Torre et al. 2014). This was complemented by comparing these unusually high BFs to those 218observed for SNPs that did not show any associations with the environment in any of the two 219previous methods (mlr and LFMM; see Figure S1). Only those candidate SNP-climate 220associations with a BF above 10 (corresponding to strong support according to Jeffreys’ scale) 221and a $\rho$–value of 0.25 or higher were conserved for validation and fitness prediction in the 222common garden (see below).

223All three sets of analyses were repeated at the regional scale for the Iberian Mediterranean and 224Atlantic 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 climates 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 v4.0 (Rangel et al. 2010). Briefly, a relative Moran’s *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 north-eastern Spain (Càlcena; see Figure 1). The climate in this region is arid (average temperature: 11.6°C, annual rainfall: 502mm, summer rainfall: 101mm, for the period 1975-2008) and characteristic of the drier extreme of maritime pine’s climatic breadth, which provides an ideal setting 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, one year old seedlings from 520 families were planted at a spacing of 2m × 3m 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 = 6,240 plants). Survival was estimated as the ratio of seedlings alive per experimental unit when the plants were five years old. The climate in the test site during these five years was representative of the regional average (average temperature: 11.8°C, annual rainfall: 533mm, summer rainfall: 99mm; 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 (non-parametric) Kendall’s (r) 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 (\textit{lm} function) in R.

To demonstrate that significant correlations were not due to population structure, we first performed a re-sampling procedure as follows. A series of 1,000 random samples of 18 SNPs with frequencies matching those of the retained candidates were obtained from the control SNP
dataset. 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 1,745 control (i.e. putatively neutral) SNPs and nine 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 over 67% and 61% of the total genetic variation, for the SNP and SSR datasets, respectively, with each individual component explaining at least 5% of the total variance. The most plausible number of genetic groups (following Evano 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 datasets for maritime pine (Bucci et al. 2007; Santos-del-Blanco et al. 2012). Maritime pine gene pools were centered in Morocco, Corsica, the Atlantic coast of France, and the Atlantic and Mediterranean (south-eastern and central) regions of the Iberian Peninsula (Figures 1 and 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, 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 multivariate logistic regressions (mlr) revealed 29 significant allele frequency-climate correlations for 24 SNPs, while the latent factor mixed models (LFMM) produced 49 significant correlations for 41 SNPs. In total, 21 correlations for 18 SNPs overlapped for both methods, including those for five non-synonymous polymorphisms; all of these correlations were also significant in the Bayesian environmental association analysis (Coop et al. 2009) (Tables 1 and 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; Tables 1 and S3; 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 over-expressed in maritime pine under experimental drought treatments (Perdiguero 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, observed and expected heterozygosities ($H_O$ 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 ($m_8$, $m_80$, $m_657$ and $m_1309$) 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 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 3E 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 five years in a common garden test under the extreme climatic
conditions of north-eastern Spain was low, and varied greatly across the 19 source populations surveyed (Figures 1 and 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 non-parametric (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 more than 5% of the genetic variance for survival (Figure 4 and Table S4). Furthermore, only 0.2% ($P=0.002$) of the 1,000 random sets of 18 SNPs re-sampled from the control SNP dataset (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 Principal Components explaining over 5% of the variance in a PCA (accounting for ~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 pre-selection of candidate genes becomes a fundamental step to capture a large part of the adaptive/phenotypic variance that needs to be used to perform such predictions.

Herein, we performed the pre-selection of genes by pinpointing candidates from previous population and functional studies. These genes included drought stress response candidates detected by genetic association (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 to 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 0_11649 encodes for a ß–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 pre-selection 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 two 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 three 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 non-coding regions, microRNAs, or copy-number and presence-absence variants (including transposon insertions), which have been directly associated to 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, having a satisfactory representation of the whole segregating genomic diversity is still a challenge for non-model 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
ecollogically-relevant variants (ideally under multiple environments). It can also provide a much

Most of the SNPs associated with climate in maritime pine were related to PC axes loaded by
temperature, both during 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 modelled
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 non-synonymous 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 up-regulated by both
cold temperatures and drought, including four amplicons coding for heat-shock proteins (Seki et
Other genes associated with environment that are worth highlighting include the abovementioned CL813, CT2134 and 0_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 co-varying 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 south-eastern Spain gene-pool is mostly located in a region with overly high winter temperatures. Indeed, when non-linear 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 datasets), 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 (Tables 1
and 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, Figures 1 and 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, in order 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 years BP) should have a frequency below 0.03 nowadays, assuming an average generation time of 50 years. These rough calculations imply that, because of the use of markers with MAF above 0.05, our strategy should only be adequate to account 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 less than 6,500 years, until reaching frequencies above 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 to 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 bear 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 alleles, 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, to integrate 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;
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 only targeted after over 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 (Yakolev 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 has to be surveyed.

In conclusion, we have shown that, while new technology is developed that allows in-depth studies of the huge conifer genomes, carefully-selected candidate genes can still be useful to identify 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.

ACKNOWLEDGEMENTS

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, AI 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-ISCIII (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 ERA-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).

GGV was also funded by a grant of the Italian Ministry (MIUR project ‘Biodiversitalia’, RBAP10A2T4), while JPJ-C, DG and MH were supported by postdoctoral fellowships (“Juan de la Cierva” and “Ramón y Cajal”) from the Spanish Ministry of Science and Innovation. Finally, SCG-M and MH acknowledge the support of Marie Curie Intra European Fellowships within the
7th European Community Framework Programme (FP7-PEOPLE-2012-IEF, project numbers 328146 and 329088, respectively).

LITERATURE CITED


Coop, G., J. K. Pickrell, J. Novembre, S. Kudaravalli, J. Li et al., 2009 The role of geography in

Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard, 2010 Using environmental
correlations to identify loci underlying local adaptation. Genetics 185: 1411-1423.

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


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

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

Deagle, B. E., F. C. Jones, Y. F. Chan YF, D. M. Absher, D. M. Kingsley et al., 2012 Population
(genomics of parallel phenotypic evolution in stickleback across stream-lake ecological


the divergence between parapatric subspecies of lodgepole pine (Pinus contorta, Pinaceae).


Gamache, I., and S. Payette, 2004 Height growth response of tree line black spruce to recent


Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, 2005 Very high resolution


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


Namroud, M. C., J. Beaulieu, N. Juge, J. Laroche, and J. Bousquet, 2008 Scanning the genome


profile of 7000 Arabidopsis genes under drought, cold and high salinity stresses using a full-length cDNA microarray. Plant J. 31: 279-292.


<table>
<thead>
<tr>
<th>EST Contig or amplicon</th>
<th>A. thaliana Locus</th>
<th>E-Value</th>
<th>Annotation</th>
<th>LG</th>
<th>SNP name (type of change)</th>
<th>Accession number</th>
<th>Climate PC</th>
<th>$H_o$</th>
<th>$H_E$</th>
<th>$G_{ST}$</th>
<th>$I / I_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0_11649</td>
<td>At2g29550</td>
<td>2.4 E-85</td>
<td>Beta tubulin</td>
<td>12F</td>
<td>m783 (A/G) \textsuperscript{syn}</td>
<td>836187519</td>
<td>PC-Winter1</td>
<td>0.34</td>
<td>0.33</td>
<td>0.23</td>
<td>0.47 \textsuperscript{**}</td>
</tr>
<tr>
<td>0_4032_02</td>
<td>At1g29340</td>
<td>2.8 E-24</td>
<td>U-Box containing protein (protein kinase)</td>
<td>11</td>
<td>m8 (A/T) \textsuperscript{syn}</td>
<td>325995579</td>
<td>PC-Winter1</td>
<td>0.34</td>
<td>0.33</td>
<td>0.47</td>
<td>0.47 \textsuperscript{**}</td>
</tr>
<tr>
<td>0_9486_01</td>
<td>NA</td>
<td>NA</td>
<td>Unknown protein</td>
<td>-/-</td>
<td>m1250 (A/G) \textsuperscript{n.n.}</td>
<td>836187522</td>
<td>PC-Winter1</td>
<td>0.41</td>
<td>0.42</td>
<td>0.20</td>
<td>0.40 \textsuperscript{**}</td>
</tr>
<tr>
<td>2_1014_02</td>
<td>At4g36990</td>
<td>2.5 E-07</td>
<td>Heat stress transcription factor</td>
<td>-/7</td>
<td>m1300 (A/G) \textsuperscript{non-syn}</td>
<td>836187523</td>
<td>PC-Summer2</td>
<td>0.34</td>
<td>0.34</td>
<td>0.30</td>
<td>0.40 \textsuperscript{**}</td>
</tr>
<tr>
<td>2_3919_01</td>
<td>At3g53720</td>
<td>8.1 E-11</td>
<td>Cation / H(+) antiporter</td>
<td>-/-</td>
<td>m1309 (A/G) \textsuperscript{non-syn}</td>
<td>836187524</td>
<td>PC-Winter1</td>
<td>0.18</td>
<td>0.22</td>
<td>0.52</td>
<td>0.62 \textsuperscript{***}</td>
</tr>
<tr>
<td>2_4183_01</td>
<td>At1g08960</td>
<td>4.7 E-52</td>
<td>Cation exchanger</td>
<td>-/10</td>
<td>m100 (A/G) \textsuperscript{syn}</td>
<td>836187525</td>
<td>PC-Winter3</td>
<td>0.30</td>
<td>0.31</td>
<td>0.27</td>
<td>0.51 \textsuperscript{***}</td>
</tr>
<tr>
<td>2_9542_01</td>
<td>At5g51690</td>
<td>8.3 E-54</td>
<td>1-aminocyclopropane-1-carboxylate synthase</td>
<td>-/11</td>
<td>m973 (A/G) \textsuperscript{syn}</td>
<td>836187526</td>
<td>PC-Winter2</td>
<td>0.36</td>
<td>0.38</td>
<td>0.25</td>
<td>0.41 \textsuperscript{**}</td>
</tr>
<tr>
<td>4cl-Pt c</td>
<td>NA</td>
<td>NA</td>
<td>4-Coumarate-CoA ligase</td>
<td>-/7</td>
<td>m137 (A/G) \textsuperscript{syn}</td>
<td>836187527</td>
<td>PC-Summer1</td>
<td>0.35</td>
<td>0.36</td>
<td>0.22</td>
<td>0.37 \textsuperscript{**}</td>
</tr>
<tr>
<td>CL813</td>
<td>At5g20830</td>
<td>5.6 E-18</td>
<td>sucrose-UDP glucosyltransferase</td>
<td>10M</td>
<td>m607 (A/C) \textsuperscript{non-cod}</td>
<td>325995573</td>
<td>PC-Winter1</td>
<td>0.41</td>
<td>0.41</td>
<td>0.16</td>
<td>0.50 \textsuperscript{***}</td>
</tr>
<tr>
<td>Cobra2</td>
<td>At3g02210</td>
<td>1.2 E-19</td>
<td>Cobra-like protein</td>
<td>-/-</td>
<td>m1126 (A/G) \textsuperscript{non-cod}</td>
<td>184956809</td>
<td>PC-Winter1</td>
<td>0.26</td>
<td>0.25</td>
<td>0.18</td>
<td>0.34 \textsuperscript{**}</td>
</tr>
<tr>
<td>CT1487</td>
<td>NA</td>
<td>NA</td>
<td>Unknown protein</td>
<td>-/1</td>
<td>m1211 (C/G) \textsuperscript{n.n.}</td>
<td>184975277</td>
<td>PC-Winter1</td>
<td>0.31</td>
<td>0.34</td>
<td>0.30</td>
<td>0.52 \textsuperscript{***}</td>
</tr>
<tr>
<td>CT2134</td>
<td>At5g48480</td>
<td>2.8 E-18</td>
<td>Glyoxalase I-like protein</td>
<td>7M</td>
<td>m646 (C/G) \textsuperscript{non-syn}</td>
<td>184975314</td>
<td>PC-Winter1</td>
<td>0.30</td>
<td>0.32</td>
<td>0.22</td>
<td>0.38 \textsuperscript{**}</td>
</tr>
<tr>
<td>CT2331</td>
<td>At5g14040</td>
<td>0.0</td>
<td>Mitochondrial phosphate</td>
<td>-/-</td>
<td>m657 (A/C) \textsuperscript{non-syn}</td>
<td>184975344</td>
<td>PC-Winter1</td>
<td>0.42</td>
<td>0.40</td>
<td>0.17</td>
<td>0.46 \textsuperscript{**}</td>
</tr>
</tbody>
</table>
Only those loci displaying significant correlations at $P < 0.001$ and with Bayes Factors > 10 are shown (see also Table S3).

- **Linkage groups retrieved from two independent maritime pine mapping populations:** Chancerel et al. (2011) / de-Miguel et al. (2012).
- **Abbreviations:** syn: synonymous; non-syn: non-synonymous; non-cod: non-coding; n.a.: unknown / data unavailable.
- The locally selected alleles at the common garden used to evaluate fitness, judging from logistic models, are given in bold.
- **Accession numbers** as provided by NCBI’s dbSNP.
- **Levels of significance:** **, $P < 0.01$; ***, $P < 0.005$.  

<table>
<thead>
<tr>
<th>Locus</th>
<th>Gene</th>
<th>SNP</th>
<th>Correlation</th>
<th>Bayes Factor</th>
<th>Population</th>
<th>Summer PC</th>
<th>Winter PC</th>
<th>Summer PC</th>
<th>Winter PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT2334</td>
<td>At1g14320</td>
<td>m658</td>
<td>$5F$</td>
<td>184975347</td>
<td>PC-Summer2</td>
<td>0.34</td>
<td>0.34</td>
<td>0.31</td>
<td>0.56</td>
</tr>
<tr>
<td>CT3538</td>
<td>NA</td>
<td>m685</td>
<td>$3F$ / 3</td>
<td>184975431</td>
<td>PC-Winter2</td>
<td>0.34</td>
<td>0.32</td>
<td>0.22</td>
<td>0.64</td>
</tr>
<tr>
<td>CT384</td>
<td>At2g29960</td>
<td>m1196</td>
<td>- / -</td>
<td>184975463</td>
<td>PC-Winter2</td>
<td>0.43</td>
<td>0.39</td>
<td>0.33</td>
<td>0.67</td>
</tr>
<tr>
<td>CT4109</td>
<td>At2g40590</td>
<td>m705</td>
<td>- / -</td>
<td>184975489</td>
<td>PC-Summer2</td>
<td>0.29</td>
<td>0.33</td>
<td>0.27</td>
<td>0.52</td>
</tr>
</tbody>
</table>
FIGURE 1 (A) Geographic distribution of the six gene pools obtained from (B) 9 nuclear SSRs in 36 natural populations of maritime pine. (C) Genetic partition is also shown for 1,745 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 (Cálcena, Spain). Populations included in the common garden are in bold italics.
FIGURE 2 Examples of allele-climate associations for maritime pine populations at the range-wide scale. (A & D) Scatter-plots, (B & E) box-plots and (C & F) distribution of minor allele frequencies (MAF), are shown for two 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.
FIGURE 3 (A) Range-wide and (E & 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, and (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.
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. Kendall’s ($\tau$) and Pearson’s ($r$) correlation coefficients and corresponding significance test values are also shown.