In situ genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster*)

Katharina B. Budde[1], Myriam Heuertz[1], Ana Hernández-Serrano[2], Juli G. Pausas[2], Giovanni G. Vendramin[3], Miguel Verdú[2] and Santiago C. González-Martínez[1]

[1]Department of Forest Ecology and Genetics, INIA Forest Research Centre, 28040, Madrid, Spain; [2]Centro de Investigaciones sobre Desertificación (CIDE-CSIC/UV/GV), 46113, Moncada, Valencia, Spain; [3]Plant Genetics Institute, National Research Council, 50019, Sesto Fiorentino, Florence, Italy

Author for correspondence:
Santiago C. González-Martínez
Tel: +34 913471499
Email: santiago@inia.es

Summary

- Wildfire is a major ecological driver of plant evolution. Understanding the genetic basis of plant adaptation to wildfire is crucial, because impending climate change will involve fire regime changes worldwide. We studied the molecular genetic basis of serotiny, a fire-related trait, in Mediterranean maritime pine using association genetics.
- A single nucleotide polymorphism (SNP) set was used to identify genotype : phenotype associations *in situ* in an unstructured natural population of maritime pine (eastern Iberian Peninsula) under a mixed-effects model framework. RR-BLUP was used to build predictive models for serotiny in this region. Model prediction power outside the focal region was tested using independent range-wide serotiny data.
- Seventeen SNPs were potentially associated with serotiny, explaining approximately 29% of the trait phenotypic variation in the eastern Iberian Peninsula. Similar prediction power was found for nearby geographical regions from the same maternal lineage, but not for other genetic lineages.
- Association genetics for ecologically relevant traits evaluated *in situ* is an attractive approach for forest trees provided that traits are under strong genetic control and populations are unstructured, with large phenotypic variability. This will help to extend the research focus to ecological keystone non-model species in their natural environments, where polymorphisms acquired their adaptive value.

Introduction

Wildfires have a long history in shaping natural ecosystems (Pausas & Keeley, 2009), and are a characteristic feature in many regions of the Mediterranean Basin (as reviewed by Pausas *et al.*, 2008). Understanding the genetic basis of plant adaptation to wildfire is especially important, because impending climate change will involve fire regime changes world-wide (Mouillot & Field, 2005; Krawchuk *et al.*, 2009). In the Mediterranean Basin, forest fire frequency and size increased significantly during the last century (Piñol *et al.*, 1998; Pausas, 2004; Pausas & Fernández-Muñoz, 2012), and are expected to increase further in the near future (Mouillot *et al.*, 2002). These new fire regimes may fall outside of the historic variability, creating new selective pressures on plant populations.

Plant populations confronted with new environments, including new fire regimes, will migrate, adapt locally or go extinct (Aitken *et al.*, 2008). As migration rates for most plant species appear to be insufficient to track the rapid environmental shifts predicted from climate change models (Malcolm *et al.*, 2002; McLachlan *et al.*, 2005), long-term persistence will mostly rely on *in situ* adaptation (Hoffmann & Sgrò, 2011). Typically, tree populations harbor high genetic diversity on which selection can act. They also have, generally, little genetic structure because of outcrossed mating systems, high gene flow and large population sizes (Petit & Hampe, 2006). High fecundity and strong selection in early life stages (Le Corre & Kremer, 2003) enable pronounced local adaptation despite strong gene flow (Kawecki & Ebert, 2004). This process can result in divergent natural phenotypes within populations (e.g. Pausas *et al.*, 2012 for flammability). Genetic differentiation at loci underlying these traits is also expected (Howe *et al.*, 2003; Kremer & Le Corre, 2012), which can be explored using association genetic approaches (Neale & Savolainen, 2004).

Fire is a strong selective driver, and there is an emerging view that fire shapes the intraspecific variability of multiple traits (e.g. bark thickness, mature height, self-pruning, age to maturity, serotiny, longevity, flammability) and generates phenotypic variability among plant populations (Keeley *et al.*, 2011; Moreira *et al.*, 2012; Pausas & Schwilk, 2012; Pausas *et al.*, 2012). In particular, multiple phenotypic traits are selected for by stand-replacing crown fire regimes in pines, such as thin bark, absence of self-pruning, early maturity and the presence of serotinous cones (Keeley & Zedler, 1998; Tapias *et al.*, 2001; Keeley *et al.*, 2011;

Keeley, 2012), which results in correlated evolution of fire-related life-history traits in these taxa (Schwilk & Ackerly, 2001; He *et al.*, 2012). In this study, we used serotiny, estimated as the proportion of serotinous cones, as an indicator for multi-trait fire phenotypes in maritime pine (*Pinus pinaster* Aiton), a species living in Mediterranean fire-prone regions. Serotiny refers to the persistence of closed mature cones in the tree canopy until seed release is triggered by high temperatures, such as those that accompany crown fires (Lamont *et al.*, 1991). A simple genetic control (one locus with two alleles) was proposed for serotiny because a fast selection-driven response to fire was detected after only one generation in different pine species (Teich, 1970; Perry & Lotan, 1979). However, more recent evidence points towards a polygenic quantitative trait (e.g. Pike *et al.*, 2010; Parchman *et al.*, 2012). Serotiny is a highly heritable trait in pine species (Perry & Lotan, 1979; Wymore *et al.*, 2011), with narrow-sense heritability (h^2) of 0.20–0.67 (Pike *et al.*, 2010; J. Climent, pers. comm.). Serotiny also shows large phenotypic variation within and among species and populations of Mediterranean pines (e.g. Tapias *et al.*, 2004; He *et al.*, 2012).

Fire-related traits, such as bark thickness, self-pruning and serotiny, are difficult to evaluate in common gardens because they are expressed late in long-lived species. Moreover, natural selection is relaxed when germinating and growing seedlings in optimal glasshouse conditions before trial establishment. An alternative approach, given the high heritability of traits such as serotiny, is to phenotype directly in unstructured natural populations for association studies (e.g. Parchman *et al.*, 2012). Allelic effects are then assessed in exactly the same environment in which they give a selective advantage, without any bias caused by plant manipulation.

Association genetics has been successful in detecting allele effects for adaptive traits in plants (Atwell *et al.*, 2010; Strange *et al.*, 2011), including in some conifers (e.g. González-Martínez *et al.*, 2007; Eckert *et al.*, 2009; Holliday *et al.*, 2010; Cumbie *et al.*, 2011; Westbrook *et al.*, 2013). Recent reports of strong linkage disequilibrium (LD) in non-coding regions of conifer genomes (e.g. in *Cryptomeria japonica*; Moritsuka *et al.*, 2012) have improved previous expectations (see Neale & Savolainen, 2004) of identifying regions of the genome associated with traits of ecological interest using candidate gene approaches in this group of species, despite their large genome sizes (25 227 Mbp in pine; Plant DNA C-values Database, release 5.0, March 2012, http://data.kew.org/cvalues). As a drawback, higher LD could make it more difficult to identify the actual causal polymorphisms within these regions. The use of candidate genes for adaptive traits should also allow the construction of predictive models for adaptive phenotypes with lower genotyping effort. Indeed, marker densities from 2–3 to 10–20 markers/cM, depending on training population size, are considered to be necessary to achieve reasonable accuracy in phenotypic predictive models based solely on random genome sampling and background LD with phenotypic traits (Grattapaglia & Resende, 2011; Resende *et al.*, 2012). However, models based on candidate genes were able to achieve similar prediction power with a much smaller number of molecular markers. For instance, Holliday *et al.* (2012) were able to

explain *c.* 28–34% of the phenotypic variance in predictive models for bud set and cold injury based on a set of only 20 loci, albeit carefully selected from expressional candidate genes (Holliday *et al.*, 2010).

Maritime pine (*P. pinaster*) is an iconic Mediterranean conifer that forms large forests in the western Mediterranean Basin. Forest fires appear to be one of the main drivers that have shaped its life history (see Tapias *et al.*, 2004; Keeley, 2012; and references therein) and large differences in fire-related traits are found among populations, such as for serotiny (from zero in Portugal to 73% in Algeria; Tapias *et al.*, 2004; see also Fig. 1). Molecular marker studies have found three completely isolated maternal lineages (based on mitochondrial DNA, mtDNA) in this species: a western lineage (most of the Iberian Peninsula, Atlantic France and Punta Cires in northern Morocco), an eastern lineage (Catalonia in the northeastern Iberian Peninsula, southeastern France, Corsica, Italy, Tunisia and Algeria) and one endemic to Morocco (Burban & Petit, 2003; see also Bucci *et al.*, 2007), as well as several distinct gene pools (based on nuclear markers) within these broad geographical regions (Salvador *et al.*, 2000; Eveno *et al.*, 2008; Santos-del-Blanco *et al.*, 2012). A long history of population isolation, in particular among populations from the different (non-overlapping) maternal lineages, could have resulted in both parallel and lineage-specific adaptations in fire-related traits, as has been shown for other main drivers of tree adaptation (e.g. Prunier *et al.*, 2012 for climate adaptation in *Picea mariana*).

In this study, we used a 384-plex single nucleotide polymorphism (SNP) array (251 successfully scored and polymorphic SNPs) enriched for well-known candidate genes for adaptive traits in forest trees to identify markers potentially associated with serotiny in maritime pine. The study was conducted in an unstructured natural population (Supporting Information Fig. S1; see also Bucci *et al.*, 2007; Santos-del-Blanco *et al.*, 2012) with high phenotypic variability for serotiny (0–100% serotinous cones per tree, average proportion of serotinous cones of 36.29% and standard deviation of 23.36%; Fig. S2; see also Tapias *et al.*, 2004), which is representative of the eastern Iberian maritime pine range. We then used the subset of SNPs potentially associated with serotiny to build a predictive model for fire phenotypes in the sampled region and tested the model accuracy by cross-validation. Finally, to evaluate the utility of the model outside of the geographical range for which it was constructed, we genotyped the same SNPs and tested the model prediction power across range-wide populations, including the three distinct maternal (mtDNA) lineages recognized in maritime pine.

The focus on well-known candidate genes for adaptive traits in pine, the high phenotypic variability in fire phenotypes (including serotiny) in eastern Iberian maritime pine stands, the fact that serotiny is heritable and gauges a multi-trait fire syndrome and the complete lack of population structure in the study region enabled us to successfully assess *in situ* phenotype : genotype associations. These potentially associated SNPs provide insights into a variety of candidate genes that could underlie fire phenotypes in Mediterranean pines and constitute the basis to construct predictive models for fire-related traits of major ecological importance.
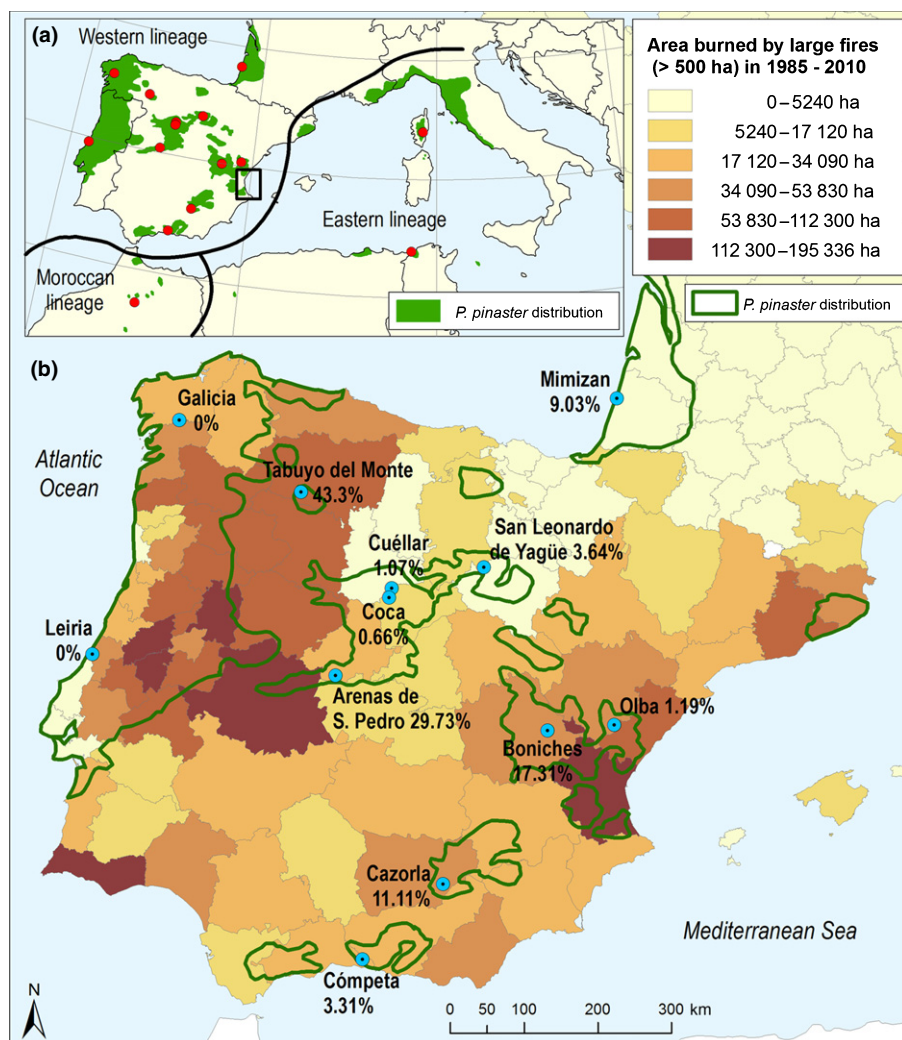
Fig. 1 (a) Focal region in eastern Spain (black box) and range-wide sampling, showing the distribution of the three maternal (mitochondrial DNA, mtDNA) lineages (black lines) known in *Pinus pinaster* (Burban & Petit, 2003). (b) Serotiny estimates for populations within the western maternal lineage (data retrieved from Tapias *et al.*, 2004) superimposed on provincial data of burned area by large fires (> 500 ha, from 1985 to 2010; European Commission, 2010). Maritime pine range is shown in green in the two maps.

Materials and Methods

Plant material

The sampling included a total of 509 maritime pine trees (Fig. 1): (1) 199 individuals collected from three stands in eastern Spain (coordinates: Calderona, 39.75, −0.50; Sinarcas, 39.79, −1.20; Eslida, 39.88, −0.30; each represented by 66–67 trees) which show high within-stand variability for fire phenotypes (0–100% serotinous cones per tree, average proportion of serotinous cones of 36.29% and standard deviation of 23.36%; Fig. S2); this dataset was used to identify loci potentially associated with serotiny and to construct SNP-based predictive models for fire phenotypes in this region; (2) 310 individuals from 15 range-wide populations (average of *c.* 21 trees per population) including the three distinct maternal (mtDNA) lineages known in maritime pine (see Table S1); this dataset was used to test for model prediction power outside the focal region. Needles were collected from the 509 (199 + 310) individuals and desiccated using silica gel. Genomic DNA was isolated using the Invisorb® DNA Plant HTS 96 Kit/C kit (Invitek GmbH, Berlin, Germany).

Nuclear microsatellites (simple sequence repeats, SSRs)

Eleven nuclear microsatellites (SSRs) were used to test for population genetic structure (which may bias genetic association approaches) within the sampling region in eastern Spain ($n = 199$; see also Bucci *et al.*, 2007; Santos-del-Blanco *et al.*, 2012; microsatellite data were deposited in the Dryad repository with doi: 10.5061/dryad.1p2s5/3). Primer sequences were obtained from Mariette *et al.* (2001) (*ITPH4516* and *FRPP94*), Chagné *et al.* (2004) (*rpTest11*, *Ctg4363*, *NZPR1078* and *NZPR413*), Guevara *et al.* (2005) (*A6F03*), Steinitz *et al.* (2011) (*pEST2669*) and F. Sebastiani and G. G. Vendramin (pers. comm., June 2011) (*epi3*, *epi5* and *gPp14*). Forward primers were 5′ end-labeled with fluorochromes (HEX, FAM, VIC or PET) and amplified using the Qiagen Multiplex PCR Kit (Qiagen, Venlo, the Netherlands) following the manufacturer's instructions. Amplified allele fragments were separated using an ABI 3730 genetic analyzer (Applied Biosystems, Carlsbad, CA, USA) and their sizes were determined with reference to the Gene-Scan −500 LIZ® Size Standard (Applied Biosystems) using GeneMapper software version 4.0 (Applied Biosystems).

384-plex SNP assay and genotyping

Genotyping ($n = 509$) was performed with Illumina VeraCode® technology for a 384-plex SNP Oligo Pool Assay (OPA) (design file provided as Table S2; SNP data deposited in the Dryad repository with doi: 10.5061/dryad.1p2s5/2). This OPA is based on a subsample of the 1536-plex SNP assay developed for *P. pinaster* by Chancerel *et al.* (2011), including polymorphisms from drought stress response candidate genes associated with climate variables in Mediterranean (*P. pinaster* and *P. halepensis*) and American (*P. taeda*) pines, as reported in Grivet *et al.* (2011) and Eckert *et al.* (2010a,b), respectively. The OPA also included 50 and 68 expressional candidate genes for stress response in maritime pine (Perdiguero *et al.*, 2013) and loblolly pine (Lorenz *et al.*, 2011), as well as a wide representation of functional candidate genes for biotic and abiotic stress responses, physical and chemical wood properties, phenology and growth in maritime pine (e.g. Pot *et al.*, 2005; Eveno *et al.*, 2008; Grivet *et al.*, 2011; Lepoittevin *et al.*, 2012; J. P. Jaramillo-Correa, pers. comm.) and other conifers (e.g. González-Martínez *et al.*, 2007, 2008; Eckert *et al.*, 2010a,b; Mosca *et al.*, 2012a,b). For example, Lepoittevin *et al.* (2012) found that SNPs *hdz31-2268* and *m1027* in this OPA were strongly associated with variation in growth and wood cellulose content, respectively, and Grivet *et al.* (2011) and J. P. Jaramillo-Correa (pers. comm.) found the allele frequency of several SNPs in this OPA (including *m705*, *m1196* and *m1211*, see the Results and Discussion sections) to be significantly associated with climatic variables (particularly with temperature variables) at regional and range-wide spatial scales.

Phenotypic measurements

Serotiny was estimated for 199 individuals in three natural stands in eastern Spain (as described previously; phenotypic data deposited in the Dryad repository with doi: 10.5061/dryad.1p2s5/1). Dominated trees and trees with a diameter of < 10 cm were avoided. For each sampled tree, serotinous (closed) and non-serotinous (open or partially open) cones were counted using binoculars on two pairs of opposite branches belonging to the upper and second third of the canopy, respectively. Cones in the main trunk were also counted, as they are abundant in this species. The serotiny level for each tree was then estimated as the number of closed cones (those remaining closed after maturation) with respect to the total number of cones (open and closed). Because changes in humidity can close open cones, serotiny was assessed during hot spring and summer days (i.e. the dry season).

Range-wide serotiny data (population means for the 15 SNP-genotyped populations; as described previously) were retrieved from Tapias *et al.* (2004). The serotiny level was estimated here by counting all closed and open cones. Population means were based on 32 individuals per population (480 observations).

Population genetic structure and kinship

Population genetic structure within the sampling region in eastern Spain ($n = 199$) was assessed using 11 nuclear microsatellites and the Bayesian clustering method implemented in STRUCTURE 2.2 (Pritchard *et al.*, 2000). Ten runs were performed for each number of clusters, $K = 1$ to $K = 5$, with a burn-in length of 50 000 and a run length of 500 000 iterations, and using an admixture model with correlated allele frequencies.

Pairwise kinship was estimated using all available markers (i.e. 11 SSRs and 251 successfully genotyped and polymorphic SNPs) to approximate the covariance matrix among the individuals used in mixed-effects linear models (MLMs, to be described), following Yu *et al.* (2006). The kinship estimator of Loiselle *et al.* (1995), as implemented in SPAGeDi 1.3 (Hardy & Vekemans, 2002), and the skewness of the pairwise kinship distribution were computed. Deviation of this distribution from normal expectations centered on mean zero, as evaluated by D'Agostino's skewness test, implies significant family structure within the population.

Identification of marker loci potentially associated with fire phenotypes

Single-locus approach The identification of SNPs with significant single-locus allelic effects on fire phenotypes followed a two-step approach. First, a preliminary selection of SNPs was based on MLMs (see Yu *et al.*, 2006), fitted independently for each SNP marker, as implemented in Tassel 3.0 (Bradbury *et al.*, 2007). The covariance matrix among individuals for the MLMs was approximated using all available markers (SSRs and SNPs), as explained previously. Negative kinship values were set to zero following Yu *et al.* (2006). We considered three alternative genetic models accounting for additive allele effects, over-dominance and allele dominance. A false discovery rate (FDR) approach (Storey, 2002; Storey & Tibshirani, 2003) was used to estimate the proportion of true null hypotheses, π_0 (i.e. $1 - \pi_0$ indicates the expected proportion of significant associations), among all tests for each genetic model. Second, for marker loci potentially associated with fire phenotypes and minor allele frequency (MAF) > 0.10, a Bayesian mixed-effect association approach (Bayesian Association with Missing Data, BAMD; Gopal *et al.*, 2009; Quesada *et al.*, 2010; Li *et al.*, 2012) in R v.2.13.1 (R Development Core Team, 2008) was used to estimate single-locus allelic effects under the three genetic models. Mean allelic effects (γ) and 95% confidence intervals were obtained from the distribution of the last 20 000 iterations (50 000 in total). Only those SNPs with confidence intervals not overlapping zero were considered to be potentially associated with the trait.

Multi-locus approach A stepwise mixed model strategy was used to further identify SNPs potentially associated with fire phenotypes. This multi-locus approach combined a multiple regression selection strategy, together with the mixed model, following Segura *et al.* (2012). Briefly, the most significant SNPs, based on genetic variance estimates from a mixed model using restricted maximum likelihood (REML), were included, one by one, as cofactor in the mixed model at each step. Then, the Bayesian Information Criterion (BIC) was used to select the best model. Only common SNPs (i.e. with MAF > 0.10) were considered in these analyses.

SNP-based predictive model for serotiny at a local scale

Common (MAF > 0.10) SNPs potentially associated (either alone or in combination with other SNPs, see the Results section) with fire phenotypes in the eastern Iberian range of maritime pine (17 loci) were employed to construct predictive models for serotiny in this region using ridge regression in a mixed-effects modeling framework (RR-BLUP). SNP-based breeding values for serotiny (i.e. numerical predictions of the relative genetic merit of each tree based exclusively on its genotype) were obtained using the rrBLUP R-software package (Endelman, 2011; Endelman & Jannink, 2012). This 'penalized' regression technique is commonly used to circumvent the classical 'large p, small n problem' (Johnstone & Titterington, 2009), although it was originally designed to deal with predictors' collinearity (Hoerl & Kennard, 1970). Despite the typically high effective population size in conifers, RR-BLUP has been shown to produce models that have moderate to high accuracy (0.17–0.51, as estimated by the Pearson correlation between de-regressed breeding values from quantitative genetic analyses and SNP-based breeding values) for a wide range of phenotypic traits (h^2 of 0.07–0.45) in loblolly pine (Resende et al., 2012). In our study case, the use of preselected SNPs potentially linked to the trait (or to causal SNPs, if they are not the causal SNPs themselves) is expected to improve the predictive value of the models (see Westbrook et al., 2013). Six-fold cross-validation was used to evaluate model accuracy in the sampled region (eastern Spain). Model accuracy was estimated by the Pearson's correlation coefficient r between observed serotiny and SNP-based breeding values, and the corresponding adjusted R^2.

Model prediction power at wide geographical scales

The recent literature has shown that SNP loci can have either local or wide-range adaptive value (e.g. Hancock et al., 2011; Prunier et al., 2012). If the same polymorphisms underlie fire-related traits in different regions, predictive models for serotiny based on candidate genes could still be valid outside the focal region. Conversely, different genetic lineages may have undergone independent (i.e. lineage-specific) adaptive processes, restricting the predictive value of the model to the local scale. Moreover, differences in LD, for example those caused by contrasted demographic history (e.g. Heuertz et al., 2006), or in the strength of selection, could also affect predictive model accuracy in distinct geographical regions. To test these alternative hypotheses, the RR-BLUP model was used to predict fire phenotypes for 310 individuals from 15 range-wide populations representing the three distinct maternal (mtDNA) lineages recognized in maritime pine (western, eastern and Moroccan lineages; Burban & Petit, 2003). Serotiny phenotypes (averages by population) were retrieved from Tapias et al. (2004) for the same populations. It should be noted that this is a fully independent dataset to that used for model construction. Significant correlations between population means of estimated breeding values based on SNPs and observed serotiny for distinct maritime pine gene pools were tested using Kendall's τ non-parametric rank correlation coefficient.

Results

The 384-plex SNP OPA conversion rate was relatively high (c. 66%; Gentrain, GC50 and GC10 scores of 0.822, 0.819 and 0.801, respectively), and we were able to obtain high-quality genotypes for 251 polymorphic SNP loci in all 509 trees. Successfully genotyped SNPs included numerous loci with suggested adaptive value in previous pine studies (see the Materials and Methods section).

Population genetic structure and kinship

No genetic structure was detected in any of the STRUCTURE runs (Fig. S1). Average pairwise kinship in the sample was close to zero ($-2.858E-05$), as expected. However, the pairwise kinship distribution was skewed towards positive kinship values (D'Agostino's skewness test: skew = 0.208, $z = 7.777$, $P = 7.387E-15$; Fig. S3), thus indicating certain pairs of related trees in the population, which supports the use of mixed-model approaches for genetic association.

Identification of marker loci potentially associated with fire phenotypes

Single-locus MLMs identified 26 SNPs under the additive and allele dominance genetic models that were potentially associated ($P < 0.05$) with serotiny (see Table S3; no significant SNPs were found for the over-dominance model), including one locus with $Q < 0.10$ (m692; see Fig. 2). The inflation factor λ was 1.11–1.29, depending on the genetic model, which indicates moderate inflation of P values. Nevertheless, the FDR approach resulted in $\pi_0 = 0.7568$ (additive model) and $\pi_0 = 0.7532$ (allele dominance model), which suggests that at least some associations are not false positives. Twelve SNPs with MAF > 0.10 had a significant (at $\alpha = 0.05$) allelic effect on serotiny phenotypes, as estimated by BAMD, under the different genetic models tested (two-tailed test; Table 1 and Fig. 2). Two of these loci, m15 and m816, best fitted the over-dominance model, whereas the other ten (m594, m692, m696, m698, m705, m912, m955, m974, m1194 and m1196) best fitted the additive or allele dominance models, in agreement with the MLM results. The best stepwise mixed model combined the effects of 11 SNPs (Fig. 3). This set included six SNPs (m594, m692, m698, m974, m1194 and m1196) of the 12 with significant single-locus allelic effects reported above, and five additional SNPs (m289, m817, m959, m1211 and m1414; see Table S4).

SNP-based predictive model for serotiny at a local scale

Seventeen common SNPs (MAF > 0.10) potentially associated with fire phenotypes in the eastern Iberian Peninsula (12 SNPs from single-locus MLM/BAMD analysis and five additional SNPs detected in the stepwise mixed model) were used to construct a phenotypic predictive model based on RR-BLUP. The model explained 29.15% (Pearson's product-moment correlation coefficient r of 0.556) of the phenotypic variation in serotiny in

Fig. 2 Genotypic effects (box plots) of eight exemplary common single nucleotide polymorphisms (SNPs) (minor allele frequency (MAF) > 0.10) showing significant association with observed serotiny in *Pinus pinaster*, as identified by a two-step approach based on mixed-effects linear models (MLMs) and Bayesian association analysis (Bayesian Association with Missing Data, BAMD). Marker codes follow Chancerel *et al.* (2011). For the box plots, the box indicates the interquartile range and the band inside it the median; the whiskers extend to the furthest data point within 1.5 times the length of the box; outliers are depicted with circles.
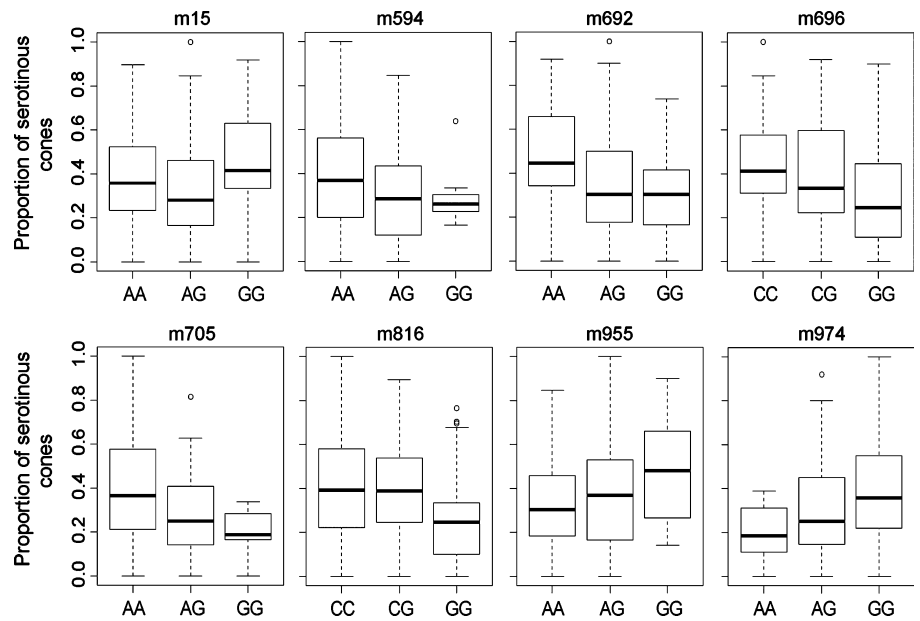
Table 1 Significant marker effects of common single nucleotide polymorphisms (SNPs) (minor allele frequency (MAF) > 0.10) on serotiny in *Pinus pinaster*, as identified by a two-step approach based on mixed-effects linear models (MLMs) and Bayesian genetic association (Bayesian Association with Missing Data, BAMD)

SNP	Annotation	SNP motif	Site annotation[a]	LG	MAF	N	Genetic model	Marker effects Mixed linear model F	P	R²	Bayesian model Mean γ (95% CIs)
m15	Defectively organized tributaries 2 (DOT2)	T/C	nc		0.3706	197	A	6.052	0.003	0.062	−0.2467 (−0.4653, −0.0282)
m594	Pyrophosphate-energized vacuolar membrane proton (AVP)	T/C	syn	8[b,c]	0.1878	197	D	6.915	0.009	0.035	−0.3678 (−0.6264, −0.1105)
m692	Unknown	A/G	unk	3[b]	0.4133	196	D	12.932	4.0E-04	0.067	−0.4696 (−0.7964, −0.1426)
m696	Arabinogalactan-like protein (AGP)	C/G	nc	3[b,c]	0.4031	196	D	5.722	0.018	0.029	−0.3206 (−0.5773, −0.0658)
m698	Nascent polypeptide-associated complex subunit alpha-like protein (NAC-alpha)	T/C	syn		0.2864	199	A	ns	ns	ns	0.2938 (0.1130, 0.4751)
m705	Carotenoid cleavage dioxygenase (CCD)	A/G	nc		0.1231	195	D	6.737	0.010	0.034	−0.3083 (−0.5993, −0.0197)
m816	Receptor protein kinase clavata1 (CLV1)	C/G	syn	1[b,c]	0.4924	197	O	ns	ns	ns	0.3391 (0.1080, 0.5706)
m912	Peroxidase 72-like (PER72)	A/T	non-syn		0.3795	195	D	5.791	0.017	0.031	0.2601 (0.0205, 0.4992)
m955	Unknown	A/G	unk	3[b,c]	0.2475	198	A	3.049	0.050	0.031	0.2294 (0.0469, 0.4160)
m974	1-Aminocyclopropane-1-carboxylate synthase (ACC)	A/G	syn	11[c]	0.1231	199	D	5.387	0.021	0.027	0.2946 (0.0137, 0.5731)
m1194	Cell division-related protein	C/G	syn		0.1439	198	A	3.714	0.026	0.038	0.2501 (0.0275, 0.4744)
m1196	Peptidyl-prolyl cis–trans isomerase (PPI)	A/C	syn		0.3266	199	A	ns	ns	ns	−0.1994 (−0.3871, −0.0086)

Bayesian mean allelic effects (γ) and 95% confidence intervals (CIs) were obtained from the distribution of the last 20 000 iterations in BAMD (for details, see Li *et al.*, 2012). Allelic effects are provided for the genetic model (A, additive; O, over-dominance; D, allele dominance) with higher effect on fire phenotype. Marker names and linkage groups (LG) as reported in Chancerel *et al.* (2011) and De Miguel *et al.* (2012); ns, not significant for that particular genetic model.
[a]Site annotation: nc, non-coding (untranslated regions or introns); non-syn, non-synonymous; syn, synonymous; unk, unknown.
[b]LG from Chancerel *et al.* (2011).
[c]LG from De Miguel *et al.* (2012).
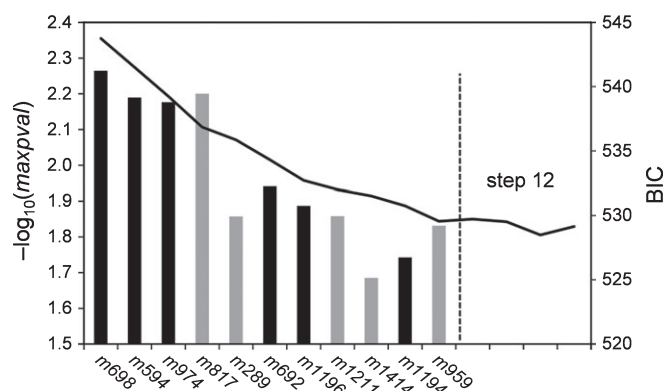
Fig. 3 Single nucleotide polymorphisms (SNPs) selected by the best stepwise mixed model in *Pinus pinaster* (step 12), as evaluated by the Bayesian Information Criterion (BIC); *maxpval*, P values for the best SNP introduced in each step. SNPs with black bars were also identified by the single-locus approach.

Fig. 4 Correlation of predicted breeding values based on ridge regression in a mixed-effects modeling framework (RR-BLUP) and observed levels of serotiny (standardized) in the eastern Iberian *Pinus pinaster* range. A linear trend is also shown.

this region (Fig. 4). Cross-validation showed a highly significant correlation between the SNP-based breeding values obtained from RR-BLUP and observed serotiny (Pearson's r of 0.429–0.632; P values of 0.00008–0.01270), with adjusted R^2 of 0.158–0.378. SNPs were not in LD, and even those obtained from the same gene (*m816* and *m817*) showed low allelic correlation values ($r^2 = 0.262$; see also LD plot in Fig. S4). Ten mapped SNPs belonged to five different linkage groups (four SNPs to LG1, three SNPs to LG3 and one SNP each to LGs 5, 8 and 11; Chancerel *et al.*, 2011; De Miguel *et al.*, 2012; see also Tables 1, S4). Thus, most significant associations represent (or are linked to) distinct causal SNPs.

Model prediction power at wide geographical scales

The predictive model for serotiny developed for the eastern Iberian Peninsula (as shown previously) had variable success outside the focal population depending on the geographical region. At range-wide scale, correlations between observed and predicted serotiny at the population level were still significant within the western maternal lineage of maritime pine (Kendall's $\tau = 0.44$, $z = 1.9933$, $P = 0.046$), but not when the eastern (Corsica and Tunisia) and Moroccan lineages were also considered (Kendall's $\tau = 0.13$, $z = 0.6937$, $P = 0.488$). This result is not surprising, considering that maternal lineages are completely isolated in maritime pine (Burban & Petit, 2003), which would facilitate lineage-specific adaptations, and that remarkable differences across geographical regions have been described for phenotypic and morphological traits in this species (Scott, 1962; Resch, 1974). Within the western maternal lineage, correlation between predicted and observed serotiny was higher in geographical regions closer to the focal population, from *c.* 30% in nearby populations from eastern and southern Spain to *c.* 20% in further away central Spain and Atlantic regions of maritime pine. In these models, low levels of serotiny were associated with negative breeding values, with the Atlantic and central Spain regions having more negative values (average of −0.228 and −0.036, respectively), and
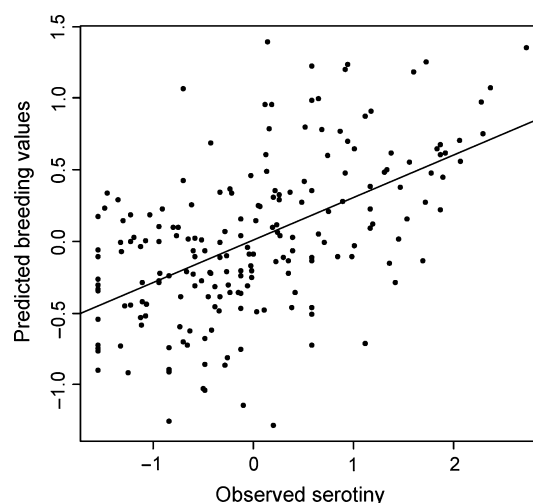
eastern and southern Spain (which are known for the higher incidence of forest fires; see Fig. 1) having more positive ones (average of 0.354 and 0.063, respectively).

Discussion

In this article, we have provided a case study for genetic association of an ecologically relevant trait (serotiny) evaluated *in situ* in natural forest tree populations. As demonstrated here, *in situ* genetic association can be achieved when the trait under study: (1) has large phenotypic variability within a region that lacks population genetic structure; (2) is under strong genetic control (i.e., heritability is high); and (3) can be accurately quantified in a large number of individuals. This approach is especially suitable for forest trees that generally form large random-mating unstructured natural populations with relatively high nucleotide diversity (Neale & Savolainen, 2004; Neale & Ingvarsson, 2008) and for which several heritable adaptive traits (e.g. female reproduction, Santos-del-Blanco, 2013; wood density, Cornelius, 1994; cold hardiness, Howe *et al.*, 2003 and references therein) can be readily evaluated in a large number of individuals.

Targeting traits that represent 'ecological syndrome' phenotypes (i.e. involving several correlated traits) that have evolved as a response to the same selective pressure (Reich *et al.*, 2003), such as serotiny for fire phenotypes, increases the chances of finding associated marker variation, even with relatively low genotyping effort. The interpretation of genetic associations for these complex phenotypes, however, can be obscured by genetic correlations among traits. In these cases, functional annotation of potentially associated loci can help to elucidate the specific traits involved and their underlying genetic architecture (to be described for serotiny). Genetic dissection of 'ecological syndrome' phenotypes, although challenging, is a promising field, as many of the most relevant ecological adaptations involve syndromes rather than single traits (e.g. pollination syndromes, Fenster *et al.*, 2004; plant defense

syndromes, Agrawal & Fishbein, 2006; Mediterranean plant syndromes, Verdú & Pausas, 2013).

In maritime pine, we identified 17 loci potentially associated with serotiny which together explained *c.* 29% of the phenotypic variation found in natural eastern Iberian populations of the species. Similar levels of explained variance have been reported in association studies (based on clonal banks or common gardens) in other conifers (e.g. *c.* 20% for wood properties in *Pinus taeda*, González-Martínez *et al.*, 2007; 17% for cold hardiness in *Pseudotsuga menziesii*, Eckert *et al.* 2009; *c.* 34% for bud set and *c.* 28% for cold hardiness in *Picea sitchensis*, Holliday *et al.*, 2010). Model accuracy within the target region (estimated using Pearson's correlation coefficient) was also similar to that found in a predictive model for oleoresin flow in loblolly pine (0.51–0.62 vs 0.43–0.63 in our study). In addition, Parchman *et al.* (2012) identified 11 SNPs that explained 50% of the variance for serotiny using a genotyping-by-sequencing (GBS) approach in *Pinus contorta*. One advantage of GBS approaches is the high marker density covering the whole genome, albeit site annotations and gene functions are normally unknown (in non-model species). Genome-wide approaches, such as GBS, are also able to reveal unexpected functional associations that would normally not have been considered in more targeted candidate gene studies. Nevertheless, Westbrook *et al.* (2013) found that 20–30 significantly associated SNPs had the same predictive value as the full dataset altogether (4854 SNPs) for oleoresin flow in loblolly pine, supporting the idea that a few, well-selected loci could have as much predictive power as genome-wide datasets.

The 'fire syndrome' arose in pines at the same time as the genus split into its two main lineages (i.e. subgenus *Pinus* and subgenus *Strobus*; He *et al.*, 2012). Species of the subgenus *Pinus*, such as *P. pinaster*, are typical of fire-prone ecosystems, whereas those in subgenus *Strobus* are mainly adapted to low productivity sites with either low soil nutrient conditions or hot and cold climatic extremes (Millar, 1998; Keeley, 2012). Interestingly, six of the loci potentially associated with serotiny in *P. pinaster* (see annotation in Table S4) are found in genes involved in the water stress response (*m289*, *m696*, *m698*, *m705*, *m912* and *m974*; Schwanz *et al.*, 1996; Zhu, 2002; Jenks & Wood, 2009), including one non-synonymous mutation in a gene coding for a peroxidase (*m912*). Another locus (*m1211*) is associated with winter temperatures at range-wide scales (J. P. Jaramillo-Correa, pers. comm.). Together, they may reflect a correlation of drought with fire phenotypes (as gauged by serotiny), a finding that has also been reported in the fire ecology literature (Pausas & Fernández-Muñoz, 2012). Three other loci (*m15*, *m816* and *m817*) are found in genes involved in cell differentiation and root, shoot and flower development (Ávila *et al.*, 2006; Casson *et al.*, 2009), and could be related to distinct growth habits and maturity age in highly serotinous trees adapted to crown fires (e.g. short trees without self-pruning and with early maturity age). Further loci are found in genes involved in more general functions (e.g. cell division, membrane transport or protein folding) or for which the function is still unknown.

The predictive value for serotiny of SNP-based models constructed in the eastern Iberian Peninsula (i.e. the linkage of the marker loci with the trait) was higher, as expected, in regions geographically closer to the focal population, and lost any predictive power outside the maternal genetic lineage within which the models were constructed. This is in line with recent findings in natural populations of *Arabidopsis thaliana*, the selfing model plant species. For example, Fournier-Level *et al.* (2011) found that alleles associated with higher survival in different sites were locally more abundant than genomic controls, and that different loci underlie the same trait in sites with contrasting environments. Our results are also in agreement with previous research in conifers, which reported lineage-specific adaptations at the molecular level (e.g. Prunier *et al.*, 2012 for climate adaptation in *Picea mariana*). Alternatively, reduction in model prediction power could reflect different levels of LD across regions (for instance, as a result of regional demographic events such as bottlenecks) or variable strength of selection. However, we did not find any evidence of lower LD in populations from the eastern (average r^2 of 0.034 vs. 0.018 in eastern Spain) or Moroccan (average r^2 of 0.052) maternal lineages of maritime pine (see also Fig. S5). Moreover, fire regime (assumed to be related to the strength of selection for serotiny; Gauthier *et al.*, 1996; Keeley & Zedler, 1998; Tapias *et al.*, 2004), although variable, is similar in regions with high and low model prediction power (see Tapias *et al.*, 2004 and Fig. 1). Finally, it is noteworthy to point out that the predictive value of the serotiny model, although limited, still extended to populations located hundreds of kilometers away from the focal population for which it was constructed. Westbrook *et al.* (2013) showed that significantly associated SNPs can be used to construct predictive models that are robust to environmental variation. Thus, a few well-constructed models covering the main gene pools of the species may be sufficient for accurate phenotypic prediction of serotiny in maritime pine.

Comparative studies of closely related species can shed light on adaptive evolutionary processes at higher phylogenetic scales than can within-species population genetics studies. Within the genus *Pinus*, serotiny evolved several times independently (Grotkopp *et al.*, 2004; He *et al.*, 2012), which provides a rich source of comparative data. Although evidence across species is still scarce, the only two available genetic association studies for serotiny in pines were able to explain substantial amounts of the phenotypic variance for this trait, 50% in lodgepole pine, a North American species, and *c.* 29% in the Mediterranean maritime pine (Parchman *et al.*, 2012 and this study, respectively), which makes comparative approaches promising. Furthermore, these two studies provide strong support to reject the previously proposed simple genetic control (one locus with two alleles) for serotiny in pines (Teich, 1970; Perry & Lotan, 1979), as distinct unlinked SNPs were potentially associated with serotiny in both species and, in the case of maritime pine (no mapping data are available for lodgepole pine), they mapped to five different linkage groups (see linkage maps in Chancerel *et al.*, 2011; De Miguel *et al.*, 2012).

Newly available genomic tools and analytical methods, such as association genetics, provide opportunities for a better understanding of the molecular basis of ecological adaptations in non-model species (Stinchcombe & Hoekstra, 2008; Stapley *et al.*, 2010), particularly with regard to climate change (Feder &

Mitchell-Olds, 2003; González-Martínez *et al.*, 2006; Hoffmann & Sgrò, 2011). Fire and climate are closely linked, and adaptive responses to forest fires will acquire increasing importance as climate changes (Pausas, 2004). Thus, current predictive models of range shifts under climate change would benefit from genetic knowledge, such as the spatial distribution of genetic variation for fire-related traits, including serotiny. In addition, predictive models for serotiny, such as that developed here, can help to identify populations and individuals with an expected good response to increased fire recurrence and intensity. Ideally, genetic effects should be studied in the natural environment where they confer adaptive value. So far, genetic association studies have been mostly focused on model organisms and species with economic value, such as major crops and some forest trees, and under controlled environments (e.g. González-Martínez *et al.*, 2007; Holliday *et al.*, 2010). However, it is important to extent this field of research to natural environments and to other ecological keystone species with distinct life-history traits and evolutionary history (Feder & Mitchell-Olds, 2003; González-Martínez *et al.*, 2006; Stinchcombe & Hoekstra, 2008; Stapley *et al.*, 2010). The probability of the detection of ecologically relevant functional markers increases with the strength of the selection drivers. Therefore, populations adapted to extreme environments (Feder & Mitchell-Olds, 2003) or that have undergone rapid environmental change (e.g. during invasion of new areas; Hoffmann & Sgrò, 2011) are ideal for the *in situ* identification of ecologically relevant genetic variation.

References

Agrawal AA, Fishbein M. 2006. Plant defense syndromes. *Ecology* 87: 132–149.

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* 1: 95–111.

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT *et al.* 2010. Genome-wide association study of 107 phenotypes in a common set of *Arabidopsis thaliana* inbred lines. *Nature* 465: 627–631.

Ávila C, Pérez-Rodríguez J, Cánovas FM. 2006. Molecular characterization of a receptor-like protein kinase gene from pine (*Pinus sylvestris* L.). *Planta* 224: 12–19.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633–2635.

Bucci G, González-Martínez SC, Le Provost G, Plomion C, Ribeiro MM, Sebstiani F, Alia R, Vendramin GG. 2007. Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Molecular Ecology* 16: 2137–2153.

Burban C, Petit RJ. 2003. Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology* 12: 1487–1495.

Casson SA, Topping JF, Lindsey K. 2009. MERISTEM-DEFECTIVE, an RS domain protein, is required for the correct meristem patterning and function in *Arabidopsis*. *Plant Journal* 57: 857–869.

Chagné D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigerio JM, Echt C *et al.* 2004. Cross species transferability and mapping of genomic and cDNA SSRs in pines. *Theoretical and Applied Genetics* 109: 1204–1214.

Chancerel E, Lepoittevin C, Le Provost G, Lin YC, Jaramillo-Correa JP, Eckert AJ, Wegrzyn JL, Zelenika D, Boland A, Frigerio JM *et al.* 2011. Development and implementation of a highly-multiplexed SNP array for genetic mapping in maritime pine. *BMC Genomics* 12: 368.

Cornelius J. 1994. Heritabilities and additive genetic coefficients of variation in forest trees. *Canadian Journal of Forest Research* 24: 372–379.

Cumbie WP, Eckert AJ, Wegrzyn J, Whetten R, Neale DB, Goldfarb B. 2011. Association genetics of carbon isotope discrimination, height and foliar nitrogen in a natural population of *Pinus taeda* L. *Heredity* 107: 105–114.

De Miguel M, de María N, Guevara MA, Díaz L, Sáez-Laguna E, Sáchnez-Gómez D, Chancerel E, Aranda I, Collada C, Plomion C *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.

Eckert AJ, Bower AD, González-Martínez SC, Wegrzyn JL, Coop G, Neale DB. 2010a. Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology* 19: 3789–3805.

Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutowsky KV, St. Clair JB, Neale DB. 2009. Association genetics of Coastal Douglas Fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-hardiness related traits. *Genetics* 182: 1289–1302.

Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra R, González-Martínez SC, Neale DB. 2010b. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* 185: 969–982.

Endelman JB. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome* 4: 250–255.

Endelman JB, Jannink JL. 2012. Shrinkage estimation of the realized relationship matrix. *G3: Genes, Genomes, Genetics* 2: 1405–1413.

European Commission. 2010 *Forest Fires in Europe 2009*. Luxembourg, Germany: Office for Official Publications of the European Communities.

Eveno E, Collada C, Guevara MA, Léger V, Soto A, Diaz L, Léger P, González-Martínez SC, Cervera MT, Plomion C *et al.* 2008. Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Molecular Biology and Evolution* 25: 417–437.

Feder ME, Mitchell-Olds T. 2003. Evolutionary and ecological functional genomics. *Nature Reviews Genetics* 4: 649–655.

Fenster CB, Armbruster WS, Wilson P, Dudash MR, Thomson JD. 2004. Pollination syndromes and floral specialization. *Annual Review of Ecology, Evolution and Systematics* 35: 375–403.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* 334: 86–89.

Gauthier S, Bergeron Y, Simon JP. 1996. Effects of fire regime on the serotiny level of jack pine. *Journal of Ecology* 84: 539–548.

González-Martínez SC, Huber D, Ersoz E, Davis JM, Neale DB. 2008. Association genetics in *Pinus taeda* L. II. Water use efficiency. *Heredity* 101: 19–26.

González-Martínez SC, Krutovsky KV, Neale DB. 2006. Forest-tree population genomics and adaptive evolution. *New Phytologist* 170: 227–238.

González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB. 2007. Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* 175: 399–409.

Gopal V, Li Z, Casella G. 2009. *BAMD: Bayesian association model for genomic data with missing covariates.* R package version 3.1.16 [WWW document] URL http://CRAN.R-project.org/package=BAMD [accessed 18 January 2013].

Grattapaglia D, Resende MD. 2011. Genomic selection in forest tree breeding. *Tree Genetics & Genomes* 7: 241–255.

Grivet D, Sebastiani F, Alía R, Bataillon T, Torre S, Zabal-Aguirre M, Vendramin GG, González-Martínez SC. 2011. Molecular footprints of local adaptation in two Mediterranean conifers. *Molecular Biology & Evolution* 28: 101–116.

Grotkopp E, Rejmánek M, Sanderson MJ, Rost TL. 2004. Evolution of genome size in pines (Pinus) and its life-history correlates: supertree analyses. *Evolution* 58: 1705–1729.

Guevara MA, Chagné D, Almeida MH, Byrne M, Collada C, Favre JM, Harvengt L, Jeandroz S, Orazio C, Plomion C *et al.* 2005. Isolation and characterization of nuclear microsatellite loci in *Pinus pinaster* Ait. *Molecular Ecology Notes* 5: 57–59.

Hancock AM, Brachi B, Faure N, Horton MW, Sperone FG, Jarymowycz LB, Toomajian C, Roux F, Bergelson JB. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334: 83–86.

Hardy OJ, Vekemans X. 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2: 618–620.

He T, Pausas JG, Belcher CM, Schwilk DW, Lamont BB. 2012. Fire-adapted traits of Pinus arose in the fiery Cretaceous. *New Phytologist* 194: 751–759.

Heuertz M, De Paoli E, Källman T, Larsson H, Jurman I, Morgante M, Lascoux M *et al.* 2006. Multi-locus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics* 174: 2095–2105.

Hoerl AE, Kennard RW. 1970. Ridge regression: biased estimation and applications for non-orthogonal problems. *Technometrics* 12: 55–82.

Hoffmann AA, Sgrò CM. 2011. Climate change and evolutionary adaptation. *Nature* 470: 479–485.

Holliday JA, Ritland K, Aitken SN. 2010. Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytologist* 188: 501–514.

Holliday JA, Wang T, Aitken SN. 2012. Predicting adaptive phenotypes from multi-locus genotypes in Sitka Spruce (*Picea sitchensis*) using Random Forest. *G3: Genes, Genomes, Genetics* 2: 1085–1093.

Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen THH. 2003. From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Canadian Journal of Botany* 81: 1247–1266.

Jenks MA, Wood AJ. 2009. *Genes for plant abiotic stress.* Ames, IA, USA: Wiley-Blackwell.

Johnstone IM, Titterington DM. 2009. Statistical challenges of high-dimensional data. *Philosophical Transactions of the Royal Society A* 367: 4237–4253.

Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. *Ecology Letters* 7: 1225–1241.

Keeley JE. 2012. Ecology and evolution of pine life histories. *Annals of Forest Science* 69: 445–453.

Keeley JE, Pausas JG, Rundel PW, Bond WJ, Bradstock RA. 2011. Fire as an evolutionary pressure shaping plant traits. *Trends in Plant Science* 16: 406–411.

Keeley JE, Zedler PH. 1998. Evolution of life histories in Pinus. In: Richardson DM, ed. *Ecology and biogeography of Pinus.* Cambridge, UK: Cambridge University Press, 219–250.

Krawchuk MA, Moritz MA, Parisien MA, Van Dorn J, Hayhoe K. 2009. Global pyrogeography: the current and future distribution of wildfire. *PLoS ONE* 4: e5102.

Kremer A, Le Corre V. 2012. Decoupling of differentiation between traits and their underlying genes in response to divergent selection. *Heredity* 108: 375–385.

Lamont BB, Le Maitre DC, Cowling RM, Enright NJ. 1991. Canopy seed storage in woody plants. *Botanical Review* 57: 277–317.

Le Corre V, Kremer A. 2003. Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics* 164: 1205–1219.

Lepoittevin C, Harvengt L, Plomion C, Garnier-Géré P. 2012. Association mapping for growth, straightness and wood chemistry traits in the *Pinus pinaster* Aquitaine breeding population. *Tree Genetics & Genomes* 8: 113–126.

Li Z, Vikneswaran G, Li X, Davis JM, Casella G. 2012. Simultaneous SNP identification in association studies with missing data. *Annals of Applied Statistics* 6: 432–456.

Loiselle BA, Sork VL, Nason J, Graham C. 1995. Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany* 82: 1420–1425.

Lorenz WW, Alba R, Yu YS, Bordeaux JM, Simões M, Dean JFD. 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.

Malcolm JR, Markham A, Neilson RP, Garcia M. 2002. Estimated migration rates under scenarios of global climate change. *Journal of Biogeography* 29: 835–849.

Mariette S, Chagné D, Decroocq S, Vendramin GG, Lalanne C, Madur D, Plomion C. 2001. Microsatellite markers for *Pinus pinaster* Ait. *Annals of Forest Science* 58: 201–206.

McLachlan JS, Clark JS, Manos PS. 2005. Molecular indicators of tree migration capacity under rapid climate change. *Ecology* 86: 2088–2098.

Millar CI. 1998. Early evolution of pines. In: Richardson DM, ed. *Ecology and biogeography of Pinus.* Cambridge, UK: Cambridge University Press, 69–91.

Moreira B, Tavsanoglu Ç, Pausas JG. 2012. Local versus regional intraspecific variability in regeneration traits. *Oecologia* 168: 671–677.

Moritsuka E, Hisataka Y, Tamura M, Uchiyama K, Watanabe A, Tsumura Y, Tachida H. 2012. Extended linkage disequilibrium in non-coding regions in a conifer, *Cryptomeria japonica.* *Genetics* 190: 1145–1148.

Mosca E, Eckert AJ, Di Pierro EA, Rocchini D, La Porta N, Belletti P, Neale DB. 2012a. The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Molecular Ecology* 21: 5530–5545.

Mosca E, Eckert AJ, Liechty JD, Wegrzyn JL, La Porta N, Vendramin GG, Neale DB. 2012b. Contrasting patterns of nucleotide diversity for four conifers of Alpine European forests. *Evolutionary Applications* 5: 762–775.

Mouillot F, Field CB. 2005. Fire history and the global carbon budget: a 1 degrees × 1 degrees fire history reconstruction for the 20th century. *Global Change Biology* 11: 398–420.

Mouillot F, Rambal S, Joffre R. 2002. Simulating climate change impacts on fire frequency and vegetation dynamics in a Mediterranean-type ecosystem. *Global Change Biology* 8: 423–437.

Neale DB, Ingvarsson PK. 2008. Population, quantitative and comparative genomics of adaptation in forest trees. *Current Opinion in Plant Biology* 11: 149–155.

Neale DB, Savolainen O. 2004. Association genetics of complex traits in conifers. *Trends in Plant Science* 9: 325–330.

Parchman TL, Gompertz Z, Mudge J, Schilkey FD, Benkman CW, Buerkle CA. 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology* 21: 2991–3005.

Pausas JG. 2004. Changes in fire and climate in the Eastern Iberian Peninsula (Mediterranean basin). *Climatic Change* 63: 337–350.

Pausas JG, Alessio G, Moreira B, Corcobado G. 2012. Fires enhance flammability in *Ulex parviflorus.* *New Phytologist* 193: 18–23.

Pausas JG, Fernández-Muñoz S. 2012. Fire regime changes in the Western Mediterranean Basin: from fuel-limited to drought-driven fire regime. *Climatic Change* 110: 215–226.

Pausas JG, Keeley JE. 2009. A burning story: the role of fire in the history of life. *BioScience* 59: 593–601.

Pausas JG, Llovet J, Rodrigo A, Vallejo R. 2008. Are wildfires a disaster in the Mediterranean basin? – A review. *International Journal of Wildland Fire* 17: 713–723.

Pausas JG, Schwilk DW. 2012. Fire and plant evolution. *New Phytologist* 193: 301–303.

Perdiguero P, Barbero MC, Cervera MT, Collada C, Soto A. 2013. Molecular responses to water stress in two contrasting Mediterranean pines (*Pinus pinaster* and *Pinus pinea*). *Plant Physiology and Biochemistry* 67: 199–208.

Perry DA, Lotan JE. 1979. A model of fire selection for serotiny in lodgepole pine. *Evolution* 33: 958–968.

Petit JR, Hampe A. 2006. The evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution and Systematics* 37: 187–214.

Pike C, Warren J, David A. 2010. *Results update: 2nd generation jack pine population. Annual report Minnesota tree improvement cooperative.* Minneapolis, MN, USA: University of Minnesota.

Piñol J, Terradas J, Lloret F. 1998. Climate warming, wildfire hazard, and wildfire occurrence in coastal Eastern Spain. *Climatic Change* 38: 345–357.

Pot D, Mcmillan L, Echt C, Le Provost G, Garnier-Géré P, Cato S, Plomion C. 2005. Nucleotide variation in genes involved in wood formation in two pine species. *New Phytologist* 167: 101–112.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multi-locus genotype data. *Genetics* 155: 945–959.

Prunier J, Gérardi S, Laroche J, Beaulieu J, Bousquet J. 2012. Parallel and lineage-specific molecular adaptation to climate in boreal black spruce. *Molecular Ecology* 21: 4270–4286.

Quesada T, Gopal V, Cumbie WP, Eckert AJ, Wegrzyn JL, Neale DB, Goldfarb B, Huber DA, Casella G, Davis JM. 2010. Association mapping of quantitative disease resistance in a natural population of loblolly pine (*Pinus taeda* L.). *Genetics* 186: 677–686.

R Development Core Team. 2008. *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Reich PB, Wright IJ, Caveder-Bares J, Craine JM, Oleksyn J, Westoby M, Walters MB. 2003. The evolution of plant functional variation: traits, spectra, and strategies. *International Journal of Plant Sciences* 164: 143–164.

Resch T. 1974. Essai de distinction des races majeurs de *Pinus pinaster*. *Annales de la Recherche Forestîère au Maroc* 14: 91–102.

Resende MFR Jr, Muñoz P, Resende MDV, Garrick DJ, Fernando RL, Davis JM, Jokela EJ, Martin TA, Peter GF, Kirst M. 2012. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* 190: 1503–1510.

Salvador L, Alía R, Agúndez D, Gil L. 2000. Genetic variation and migration pathways of maritime pine (*Pinus pinaster* Ait.) in the Iberian Peninsula. *Theoretical and Applied Genetics* 100: 89–95.

Santos-del-Blanco L. 2013. *Evolutionary ecology of reproduction in two Mediterranean pine species (*Pinus pinaster* Ait. and P. halepensis Mil.).* PhD thesis, University of Valladolid, Spain.

Santos-del-Blanco L, Climent J, González-Martínez SC, Pannell JR. 2012. Genetic differentiation for size at first reproduction through male versus female functions in the widespread Mediterranean tree *Pinus pinaster*. *Annals of Botany* 110: 1449–1460.

Schwanz P, Picon C, Vivin P, Dreyer E, Guehl JM, Polle A. 1996. Responses of antioxidative systems to drought stress in pedunculate oak and maritime pine as modulated by elevated CO_2. *Plant Physiology* 110: 393–402.

Schwilk DW, Ackerly DD. 2001. Flammability and serotiny as strategies: correlated evolution in pines. *Oikos* 94: 326–336.

Scott CW. 1962. A summary of information on *Pinus pinaster*. *Forestry Abstracts* 23: i–xviii.

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M. 2012. An efficient multi-locus mixed model approach for genome-wide association studies in structured populations. *Nature Genetics* 44: 825–832.

Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, Bennison C, Ball AD, Beckerman AP, Slate J. 2010. Adaptation genomics: the next generation. *Trends in Ecology and Evolution* 25: 705–712.

Steinitz O, Troupin D, Vendramin GG, Nathan R. 2011. Genetic evidence for a Janzen–Connell recruitment pattern in reproductive offspring of *Pinus halepensis* trees. *Molecular Ecology* 20: 4152–4164.

Stinchcombe JR, Hoekstra HE. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100: 158–170.

Storey JD. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B* 64: 479–498.

Storey JD, Tibshirani R. 2003. Statistical significance for genome-wide studies. *Proceedings of the National Academy of Sciences, USA* 100: 9440–9445.

Strange A, Li P, Lister C, Anderson J, Warthmann N, Shindo C, Irwin J, Nordborg M, Dean C. 2011. Major-effect alleles at relatively few loci underlie distinct vernalization and flowering variation in *Arabidopsis thaliana* accessions. *PLoS ONE* 6: e19949.

Tapias R, Climent J, Pardos JA, Gil L. 2004. Life histories of Mediterranean pines. *Plant Ecology* 171: 53–67.

Tapias R, Gil L, Fuentes-Utrilla P, Pardos JA. 2001. Canopy seed banks in Mediterranean pines of south eastern Spain: a comparison between *Pinus halepensis* Mill., *P. pinaster* Ait., *P. nigra* Arn. and *P. pinea* L. *Journal of Ecology* 89: 629–638.

Teich AH. 1970. Cone serotiny and inbreeding in natural population of *Pinus banksiana* and *Pinus contorta*. *Canadian Journal of Botany* 48: 1805–1809.

Verdú M, Pausas JG. 2013. Syndrome-driven diversification in a Mediterranean ecosystem. *Evolution* 67: 1756–1766.

Westbrook JW, Resende MFR, Munoz P, Walker AR, Wegrzyn JL, Nelson CD, Neale DB *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 Phytologist* 199: 89–100.

Wymore AS, Keeley ATH, Yturralde KM, Schroer ML, Propper CR, Whitham TG. 2011. Genes to ecosystems: exploring the frontier of ecology with one of the smallest biological units. *New Phytologist* 191: 19–36.

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB *et al.* 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* 38: 203–208.

Zhu JK. 2002. Salt and drought stress signal transduction in plants. *Annual Review of Plant Biology* 53: 247–273.

Supporting Information

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

Fig. S1 STRUCTURE software bar plots showing ancestry proportions for $K = 2$–5.

Fig. S2 Phenotypic variability (box plot) for serotiny in the eastern Iberian Peninsula.

Fig. S3 Density plot for pairwise kinship, as estimated by SPAGeDi 1.3.

Fig. S4 Linkage disequilibrium (LD) heatmap for single nucleotide polymorphisms (SNPs) associated with serotiny in focal study region.

Fig. S5 Linkage disequilibrium (LD) heatmaps for all polymorphic single nucleotide polymorphisms (SNPs) in four range-wide populations.

Table S1 Details on range-wide populations included in the single nucleotide polymorphism (SNP) genotyping

Table S2 Illumina Oligo Pool Assay (OPA) design file and designability scores provided as Excel file

Table S3 Marker effects of 26 single nucleotide polymorphisms (SNPs) with $P < 0.05$ in the mixed-effects linear models

Table S4 Annotation and mapping information for single nucleotide polymorphisms (SNPs) associated with serotiny

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

About *New Phytologist*

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Trust, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews.

- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <25 days. There are **no page or colour charges** and a PDF version will be provided for each article.

- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.

- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@ornl.gov)

- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**