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

Katharina B. Budde1, Myriam Heuertz1, Ana Hernández-Serrano2, Juli G. Pausas2, Giovanni G. Vendramin3, Miguel Verdu2 and Santiago C. González-Martínez1

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

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 expression 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.
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; Sinarca, 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.1p2x5/3). 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 GeneScan –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 expression 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 bdx31-2268 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 Tapia 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, π0 (i.e. 1 − π0 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 $\tau$ 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 $P < 0.10$ ($m692$; see Fig. 2). The inflation factor $\hat{\lambda}$ 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 ($m594$, $m692$, $m697$, $m912$, $m955$ and $m1194$) 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
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).

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

Bayesian mean allelic effects (β) 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.

aSite annotation: nc, non-coding (untranslated regions or introns); non-syn, non-synonymous; syn, synonymous; unk, unknown.
bLG from Chancerel et al. (2011).
cLG from De Miguel et al. (2012).
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 over geographical scales, with the Atlantic and central Spain 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 hardness, 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...
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 hardness in *Pseudotsuga menziesii*, Eckert *et al.*, 2009; c. 34% for bud set and c. 28% for cold hardness 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 (*Avila 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 extend 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.

Acknowledgements
We thank M. Zabal-Aguirre, A. I. de-Lucas, C. Odoíñez, Y. Kurt and J. Chofre for field assistance. Thanks are extended to C. Lepoittevin and P. Garnier-Géré for help during SNP array design, and to J. Fernández for preparing Fig. 1. Data for Fig. 1 were provided by the European Forest Fire Information System – EFFIS (http://effis.jrc.ec.europa.eu) of the European Commission Joint Research Centre. SNP genotyping services were provided by the Spanish ‘Centro Nacional de Genotipado’ (CEGEN-ISCIII, http://www.cegen.org). This study was financed by the following projects from the Spanish government: VAMPIRO (CGL2008-05289-C02-01/02) and TREVOL (CGL2012-39938-C02-00/01), as well as by the EU FP7 project ‘Novel Tree Breeding Strategies’ (NovelTree) and the Italian MIUR project ‘Biodiversitaly’ (RBAP10A2T4). This research was also funded by the ERA-Net BiodivERsA (LinkTree project, EUI2008-03713), with the Spanish Ministry of Economy and Competitiveness as national funder, part of the 2008 BiodivERsA call for research proposals.

References


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.