Does phylogeographical structure relate to climatic niche divergence? A test using maritime pine (Pinus pinaster Ait.)

M. J. Serra-Varela1,2,3, D. Grivet1, L. Vincenot3,4,5, O. Broennimann6, J. Gonzalo-Jiménez1,2* and N. E. Zimmermann7

ABSTRACT

Aim To disentangle the effects of environmental and geographical processes driving phylogenetic distances among clades of maritime pine (Pinus pinaster). To assess the implications for conservation management of combining molecular information with species distribution models (SDMs; which predict species distribution based on known occurrence records and on environmental variables).

Location Western Mediterranean Basin and European Atlantic coast.

Methods We undertook two cluster analyses for eight genetically defined pine clades based on climatic niche and genetic similarities. We assessed niche similarity by means of a principal component analysis and Schoener’s D metric. To calculate genetic similarity, we used the unweighted pair group method with arithmetic mean based on Nei’s distance using 266 single nucleotide polymorphisms. We then assessed the contribution of environmental and geographical distances to phylogenetic distance by means of Mantel regression with variance partitioning. Finally, we compared the projection obtained from SDMs fitted from the species level (SDMsp) and composed from the eight clade-level models (SDMcm).

Results Genetically and environmentally defined clusters were identical. Environmental and geographical distances explained 12.6% of the phylogenetic distance variation and, overall, geographical and environmental overlap among clades was low. Large differences were detected between SDMsp and SDMcm (57.75% of disagreement in the areas predicted as suitable).

Main conclusions The genetic structure within the maritime pine subspecies complex is primarily a consequence of its demographic history, as seen by the high proportion of unexplained variation in phylogenetic distances. Nevertheless, our results highlight the contribution of local environmental adaptation in shaping the lower-order, phylogeographical distribution patterns and spatial genetic structure of maritime pine: (1) genetically and environmentally defined clusters are consistent, and (2) environment, rather than geography, explained a higher proportion of variation in phylogenetic distance. SDMs, key tools in conservation management, better characterize the fundamental niche of the species when they include molecular information.

Keywords Climate change, conservation biology, conservation, genetics, infraspecies, niche conservatism, SDM, species distribution model.
to different conclusions in the literature (Warren et al., 2008). A further question is the degree of overlap between genetic relatedness and environmental niche similarity and, more specifically, how phylogenetic distance among species is correlated to, or explained by, environmental or geographical distances.

These issues relate to major forces of speciation and can be transferred to the infraspecific level, i.e. to the evolutionary significant units (ESUs) that characterize a species (Moritz, 1994). ESUs define sets of populations within a species that are historically distinct and likely to have a distinct evolutionary potential: see Bucci et al. (2007), Pearman et al. (2010) and D’Amen et al. (2013) for examples with maritime pine, herpetofauna and African mammals, respectively. Species evolutionary units play a pivotal role in interpreting a species’ history, i.e. colonization, migration or adaptation events (Rodríguez-Sánchez et al., 2010). If different ESUs follow very distinct evolutionary trajectories, speciation can ultimately occur.

Species distribution models (SDMs), when combined with phylogenetic information, allow us to test niche conservatism versus dynamism (Pearman et al., 2008) and to answer the questions we have posed above. While SDMs enable us to characterize environmental niches and to map the areas potentially occupied by a species, phylogenies inform us about evolutionary relatedness among taxa.

SDMs calibrate associations between environmental variables and known species occurrence records to identify the environmental conditions necessary for the viability of a population (Guisan & Thuiller, 2005). This association can be expressed either by correlative or by mechanistic approaches, and models can be based on just abiotic predictors or may also include biotic ones. A common procedure when comparing the environments of two sister species is to first calibrate SDMs for both species separately, and then to evaluate the overlap in their predicted niche distribution (Warren et al., 2008).

Despite the proliferation of SDMs during the last decade, only a limited number of studies have integrated infraspecific information based on morphologically defined subspecies or clades (e.g. Pearman et al., 2010; Oney et al., 2013), and even fewer have done so with information based on genetically defined groups. This is surprising, given that the importance of combining genetic information with environmental variables in a spatially explicit framework has been pointed out by many authors (Thuiller et al., 2008; Thomassen et al., 2010). Integrating molecular information into SDMs seems especially important when projecting these models under future environmental conditions, as genetically defined subspecies or clades have been found to diverge in their niche characteristics when fitted under current climate conditions (Pearman et al., 2010; D’Amen et al., 2013). These findings suggest local adaptation and niche differentiation and, therefore, a different response to climatic changes.

Previous studies investigating ecological differences between genetically based infraspecific clades have focused mainly on animal subspecies (e.g. Pease et al., 2009; D’Amen et al., 2013). In the case of plants, Benito-Garzón et al. (2011) integrated infraspecific information to calibrate models of tree survival for maritime and Scots pine using adaptive traits quantified from provenance trials. Similarly, O’Neill et al. (2008) modelled the distribution and productivity of lodgepole pine using different species provenances tested in multisite trials. Both studies highlighted the importance of considering the different populations of a species when modelling its responses to climate change, as future predicted distributions varied largely depending on the populations modelled.

None of the previous studies has fully explored the level of environmental space overlap, or whether it is related to the (phylo)genetic similarities among the clades evaluated. Such information is important for determining whether niche adaptation, that can potentially lead to speciation, is a significant force behind the spatial and environmental structure of infraspecific clades. Alternatively, niche conservatism could lead to more conserved ecological niches than would be suggested from genetic or spatial distance between clades. Such information is ultimately relevant for designing conservation plans that integrate locally adapted populations as essential components within the context of climate change.

To sum up, to further understand the phylogenetic, environmental and spatial relationships across the whole range of a species, there is a need for studies integrating infraspecific molecular and ecological information in a phylogeographical framework. Our study, using maritime pine (Pinus pinaster Ait.) as a model species, helps to fill this gap, as we calibrate SDMs and combine genetic, environmental and geographical information. This Mediterranean pine is ecologically versatile, growing on a variety of substrates (from siliceous to calcareous), across a wide range of elevations (from sea level to 600 m, and up to 2000 m in its southern range in Morocco) and under different climates (semi-arid to humid). Its strong ecological differentiation and patchy distribution has led to the definition of several varieties (Barbéro et al., 1998). In addition, both adaptive phenotypic traits (Alia et al., 1997; Sierra-de-Grado et al., 2008) and molecular resources from all three genomes, namely, mitochondrial, chloroplast and nuclear (Burban & Petit, 2003; Bucci et al., 2007; Santos-del-Blanco et al., 2012; Jaramillo-Correa et al., 2015), analysed across the full distribution range, show spatially structured patterns. In combination, these characteristics suggest that maritime pine populations may be adapted to distinct niches, making this conifer an excellent model for studying local adaptation and differentiation at the infraspecific level.

The present study aims at answering three main questions: (1) does niche overlap in maritime pine correlate with phylogenetic relatedness at the clade level; (2) is the phylogenetic distance among clades better explained by environmental or by geographical distance and (3) do infraspecific SDMs representing distinct evolutionary units provide better fits and spatial projections than species-level SDMs? To answer these questions, we first fitted and validated ensemble SDMs (i.e. sets of several models) for each genetic clade, and for the species fitted as a whole. We also analysed two predicted distributions of the species, one for the species fitted from species-level data and one composed from the clade-level models. Second, we estimated the niche position of each clade along principal components.
analysis (PCA) axes based on environmental data, and compared genetic, environmental and geographical distances. Finally, we assessed the differences between the species-level and the infraspecific clade-level SDMs regarding predictive performance.

METHODS

Molecular data

The combination of mitochondrial, chloroplast and nuclear (simple sequence repeats, SSRs, and single nucleotide polymorphisms, SNPs) molecular markers from previous studies allowed us to define eight genetic clades for the full distribution range of maritime pine: Atlantic Iberian Peninsula (G1), eastern populations (G2), Atlantic France (G3), Morocco (G4), eastern (G5), central (G6) and southern (G7) Spain, and Tunisia (G8) (Fig. 1) (see Appendix S1 in Supporting Information for a detailed explanation). Three marginal populations (hatched symbols in Fig. 1) were unclassified, as we considered it inappropriate to assign them to any of the eight defined clusters due to insufficient field-sampled molecular information. Another isolated population (Fuencaliente, FCN, Spain), which was genotyped just for SNPs, was excluded from the present study.

We used a priori neutral markers to define the eight clades, i.e. primarily influenced by demographic processes and not by adaptation. While adaptive selection, especially the SNPs, cannot be completely excluded (see Jaramillo-Correa et al., 2015), the fact that each independent set of molecular markers points to the same overall genetic structure strongly supports our selection.

To assess the genetic relatedness among the clusters, Nei’s genetic distances (Nei, 1978) were calculated for 266 SNPs (Jaramillo-Correa et al., 2015) and used to build a phylogenetic cladogram with unweighted pair group method with arithmetic mean (UPGMA), with 10,000 bootstrap replications using Poptree2 (Takezaki et al., 2010).

![Figure 1](image_url) (a) Distribution of genetic clades (G1–G8) along the native distribution range of maritime pine. Hatched areas correspond to non-classified clades. (b) Unweighted pair group method with arithmetic mean (UPGMA) dendrogram based on pairwise population Nei’s distances on 266 SNPs in 45 populations, 10,000 bootstrap replications. Only bootstrap support values greater than or equal to 50% are displayed.
Species data

Two different sources of species occurrence were combined to represent the complete native range of maritime pine: (1) the Tree Species Distribution for Europe (TSDE; Köhle & Seufert, 2001) from the Joint Research Centre’s AFOLU data portal (ftp://mars.jrc.ec.europa.eu/AFoludata/Public/D566/), and (2) the EUFORGEN database from the European forest genetic resources programme (http://www.euforgen.org/distribution _maps.html). TSDE maps tree species percentage occupancy in Europe at a 1-km grid level, but it does not differentiate between native and planted populations. EUFORGEN delivers a distribution shape that, while spatially less accurate, includes the whole species range within Europe and northern Africa, and specifically excludes planted populations. Therefore, by filtering TSDE occurrences with EUFORGEN, we obtained a good approximation of the native range of maritime pine. Finally, we split the species’ distribution range into eight subranges based on our eight, geographically distinct, clades (see Appendix S1 for details on maritime pine clades).

A set of presence and pseudo-absence records was created for the species and for each of the genetically defined infraspecific clades. A total of 128,653 presences were used for the calibration of the model at the species level, and from 59,109 (G1) to 3997 (G8) at the clades level. Pseudo-absences (643,265 records for all the models) were randomly selected within the study area (see Fig. 1b) within those cells where TSDE reported 0% occupancy (see Appendix S2 for more details).

Bioclimatic data

We considered 23 bioclimatic variables representative of the period 1950–2000 for the analysis. Nineteen of them (BIO1–BIO19) were downloaded from the current bioclimatic variables available in WorldClim (Hijmans et al., 2005). Four new variables were created following Zimmermann et al. (2007) to better characterize water availability: summer and spring potential evapotranspiration (ETPTsummer, ETPTspring) and summer and spring moisture index (MINDsummer, MINDspring) (see Appendix S3 for a detailed explanation of bioclimatic data).

After screening for correlation and variance inflation, we selected four climate variables as predictors for our analyses: BIO3 (isothermality), BIO11 (mean temperature of the coldest quarter), BIO13 (precipitation of the wettest month) and MINDsummer (see Appendix S3 for a detailed explanation of variable selection).

Species distribution models

We used five different statistical methods to calibrate the models individually, namely general linear models (GLMs; McCullagh & Nelder, 1989), generalized additive models (GAMs; Hastie & Tibshirani, 1990), random forest (RF; Breiman, 2001), classification tree analysis (CTA; Breiman et al., 1984) and MaxEnt (Phillips et al., 2006). We built 45 individual algorithm models, five for each of the eight genetic clades, and five for the species level. All models were processed in biomod (Thuiller et al., 2009) using the package ‘biomod2’ in the R statistical software environment (R Development Core Team, 2013), which was also used for all additional statistical analyses. Model performance was assessed by means of the true skill statistic (TSS), the area under the receiver operating characteristic curve (AUC) and the H-measure (see Appendix S2 for more information on algorithm specifications and model performance assessment).

Several reviews in the past have compared the advantages of using different statistical methods (e.g. Elith et al., 2006; Bedia et al., 2011), but today using just one model or the best among several is considered a less powerful approach than using several models in an ensemble (Araújo & New, 2007; Grenouillet et al., 2011). Thus, we averaged the predictions from the five statistical methods we had used to create nine ensemble models, one for each genetic clade and one for the species level.

The large number of occurrence records available permitted a random division of each dataset (corresponding to the species and to each clade) into two equally sized subsets for training and evaluating. Both subsets maintained the initial proportion between presence and pseudo-absence records. Evaluation was performed on 20% of the set-aside evaluation records (see Appendix S2 for more details).

Probabilistic model outputs based on current climatic conditions were converted to binary presence–absence maps by defining thresholds that optimized TSS values. Thereby, nine different distribution maps were created, eight corresponding to the different genetically defined clade models (G1–G8 projections) and one to the species-level model (referred to as ‘SDMsP’).

Niche overlap

Niche overlap was determined in geographical and environmental space. In geographical space, projections G1 to G8 were used to create a single map (‘overlapping clade projection’) highlighting the overlap in bioclimatically suitable areas among the different clades. Values could range from zero (i.e. indicating that bioclimatic conditions are unsuitable for any clade) to eight (i.e. indicating cells that are suitable for all genetic clades). Then, geographical overlap was quantified for every pair of clades as the percentage suitable for both clades among the total area suitable for any of the two clades.

To measure niche overlap in environmental space we applied the methodology proposed by Broennimann et al. (2012). This recent technique corrects for spatial autocorrelation in the environment, thus reducing the possibility of confounding effects between spatial patterns and ecological processes (Warren et al., 2014). The method uses Schoener’s D metric (Schoener, 1970) as a measure of environmental niche overlap, and includes a statistical framework to test for niche similarity and equivalency, as proposed by Warren et al. (2008). The idea is to create a multivariate environmental grid, based on the first two axes of a PCA summarizing all the selected environmental variables, in which each cell represents a unique combination of the environmental conditions available in the study area. Then, Gaussian
kernel density functions are used to estimate the occupancy of each cell \((z_i)\), and the \(D\) metric is calculated based on the different \(z_i\) values obtained. Finally, equivalency and similarity tests in both directions were performed. To create the PCA axes, we used the same four bioclimatic variables that were used to calibrate the SDMs across all study-area pixels (PCA-env in Broennimann et al., 2012). The median PCA score obtained for the presence records of each genetic clade was used as an indicator of its position along the PCA axes, and the variance among the group as an estimator of niche breadth in each axis.

**Constructing SDMs for species and molecular-based clades**

In order to create a composite model for the whole species based on the sum of all molecular-based clades, projections G1 to G8 were superimposed. We refer to this result as the ‘composite clade model’ or SDMcm, as opposed to the ‘species-level model’ or SDMsp. It differs from the ‘overlapping clade model’ in that it only separates areas that are climatically suitable and unsuitable for the species. SDMcm and SDMsp projections were compared regarding: (1) concordance in predicting a climatically suitable area, and (2) agreement in predicting a climatically unsuitable area.

**Analysing phylogenetic, environmental and geographical distances**

We calculated phylogenetic distances among clades using the ‘distTips’ command in the ‘adephylo’ package. Environmental distance was calculated as the Euclidean distance between the clades along the PCA axes. When calculating geographical distance, we considered the centre of gravity of the genetic-clade distributions as their representative position. We based the measure on a least-cost path analysis, considering Mediterranean water resistance as 100 times higher than that of land, to avoid easy dispersal across the Mediterranean Sea. Values in distance matrices are not independent, so we used Mantel regression as it corrects partial matrix value dependence. We used the ‘multimat’ command in the ‘phytools’ package to explain phylogenetic distance among clades by means of environmental and geographical distances, and we set the number of permutations to \(n = 9999\). Mantel correlations with 9999 permutations were run with the ‘manteltest’ command in the ‘ade4’ package. We used variation partitioning (Borcard et al., 1992; Meier et al., 2010) to single out the individual contributions of environmental and geographical distances in explaining phylogenetic relatedness.

**RESULTS**

**Genetic clusters**

The phylogenetic reconstruction of the eight genetically defined maritime pine clades based on SNPs (Fig. 1) mainly matched the one obtained from chloroplast markers (Bucci et al., 2007). It grouped the eastern (G5), central (G6) and southern (G7) Spanish populations, while Atlantic Iberian Peninsula (G1), Atlantic France (G3) and Tunisia (G8) were clustered together. These two clusters are connected to the eastern populations (south-eastern France, Italy and Corsica) (G2). Finally, the Moroccan populations (G4) are the most genetically different in relation to all other clades.

These results indicate short genetic distances among G5/G6/G7, and among G1/G3/G8, while G2 is genetically more distant from them and G4 is the most genetically distant clade.

However, Tunisian populations (G8) present atypical behaviour, grouping with western European clades (G1 and G3) when we expected them to group with the eastern one (G2), based on the mitochondrial and chloroplast markers for which distinct sets of individuals were analysed (see Burban & Petit, 2003; Bucci et al., 2007). We suspect a possible mistake in labelling the individuals from Tunisia (G8) used for SNPs and nuclear SSRs (nSSRs) (the same sample may have been used for both markers; see Jaramillo-Correa et al., 2015).

**Species distribution models**

The models calibrated individually for each genetic clade achieved a better performance than the models calibrated at the species level. The five statistical algorithms performed well (all AUC scores above 0.89; see Table S2 in Appendix S2), although RF outcompeted the rest in all cases except for G8, for which GLM and GAM performed slightly better.

Overall, the ensemble models outperformed most individual algorithm models except for RF. Even with that exception, the relationship between maritime pine distribution and its relevant climatic variables is well captured by the ensemble models, which represent a sound method for including the variability originating from different statistical algorithms (Araújo & New, 2007; Grenouillet et al., 2011).

More details on model performance can be found in Appendix S2. All projections from 45 individual algorithms and 9 ensemble models are presented in Appendix S4.

**Niche overlap**

Overall, the overlap values obtained with both geographical and environmental analyses were low, with the exception of a few clades (Table 1). In the area catalogued as suitable territory for maritime pine (681,911 km²), 79.8% was assigned to only one specific genetic clade, indicating a generally limited overlap among clades. Consistently, only 15.9% of the area was suitable for two different genetic clades and 4.3% for three different genetic clades (see Fig. 2); no area was suitable for more than three clades. Most of the territory suitable for three different genetic clades was concentrated in Spain (east and central plateau), mainly corresponding to range areas suitable for clades G5/G6/G7 with a very small contribution for G2/G6/G7. There was also geographical overlap for the three clades G1/G4/G8 in northern Africa.
Clades G5, G6 and G7 located in eastern, central and southern Spain, respectively, showed particularly high niche overlap in both geographical and environmental space compared with the results obtained for the other genetic clades. We also found some overlap in environmental space among G1, G4 and G8, located on the Atlantic side of the Iberian Peninsula, Tunisia and Morocco, respectively. However, in these latter cases the overlap in environmental space did not translate into significant geographical overlap, suggesting that these environmental conditions appear in the study area only rarely.

The first two PCA-env axes explained 86.54% of the total variance (PC1 = 69.12% and PC2 = 17.24%; Fig. 3a), with PC1 being strongly correlated with summer water availability. Niche position and results (Fig. 3b) showed four distinguishable groups along the first two axes: G1/G3/G8, G5/G6/G7, G2 and G4. All equivalency tests were significant (P-value < 0.05), meaning that we could reject the null hypothesis that any pair of clade niches was equivalent. In addition, most similarity tests were not significant (except for G1/G8), so it cannot be assumed that the environmental conditions occupied by the

Table 1 Geographical and environmental overlap between genetic clades (G1–G8).

<table>
<thead>
<tr>
<th>Geographical overlap (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Env. overlap (D metric)</td>
</tr>
<tr>
<td>G1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>G1</td>
</tr>
<tr>
<td>G2</td>
</tr>
<tr>
<td>G3</td>
</tr>
<tr>
<td>G4</td>
</tr>
<tr>
<td>G5</td>
</tr>
<tr>
<td>G6</td>
</tr>
<tr>
<td>G7</td>
</tr>
<tr>
<td>G8</td>
</tr>
</tbody>
</table>

Geographical overlap (above the diagonal) is calculated as the percentage of the total area suitable for any of the two components of the pair that is suitable for both members at the same time. Environmental overlap (below the diagonal) is estimated by the D metric (Schoener, 1970). Values above 0 indicate overlap in geographical or environmental (bioclimatic) space among any pair of the eight clades.

Figure 2 Geographical overlap of the projections for the different genetic clades (G1–G8). The suitability value indicates the number of genetic clades for which the area is climatically suitable. No territory is suitable for more than three genetic clades.
genetic clades were more similar than those expected by chance.

Mantel regressions on environmental and geographical distance matrices explained 12.6% of the variation in phylogenetic distance among the eight clades (Fig. 4). In them, the linear term for environmental distance was barely significant ($P = 0.06$), while the intercept ($P = 0.87$) and the geographical term ($P = 0.74$) were not (see the scatterplots of phylogenetic versus genetic distances and phylogenetic versus geographical distances in Appendix S2). Environmental distance explained c. 60% of the phylogenetic distance, while the contribution of geographical distance and the joint geography–environment distance contribution explained c. 20% each (Fig. 4). This was also reflected in higher and significant Mantel correlations between phylogenetic and environmental distances ($r = 0.32$, $P = 0.01$), as compared to geographical distances ($r = 0.21$, $P = 0.25$).

**Constructing SDMs for species and molecular-based clades**

In the comparison between SDMcm and SDMsp we detected a considerable mismatch between the two types of projections encompassing the species as a whole (Fig. 5). Notably, SDMcm projected habitat suitability for at least one of the molecular-based clades outside the projected geographical range covered by the species model. These areas contain conditions that are modelled as unsuitable from the species-level perspective, yet at least one of the clades found these regions suitable. Only very small areas predicted by the species model were not included within the niche of any of the eight clades.

**DISCUSSION**

**SDMs, climatic niches and phylogeography**

Both the strong niche differentiation and the low values of niche overlap among clades indicate niche evolution at the infraspecific level for maritime pine. Our results are in line with observed morphological differences among populations. They also support the subdivision of maritime pine into two subspecies (subsp. atlantica present in Atlantic plains, and the circum-Mediterranean subsp. mesogeensis) and several geographical varieties: var. renoui in Algeria and Tunisia, var. maghrebiana in Morocco, var. cortensis in Corsica, var. hamiltonii in the Mediterranean and var. atlantica in the Atlantic region (Barbére et al., 1998).

Furthermore, our study reinforces the conclusions of previous ones, in that both the phylogeographical pattern of maritime pine and the between-clade relationships largely relate to the demographic and migration history of the species, as most of the variation in phylogenetic distance cannot be explained by environmental or geographical distance. The existence of several
glacial refugia (in Africa, south-eastern Spain and the Atlantic coast of Portugal), together with their historical isolation, would be responsible for the delineated geographical pattern of the distinct genetic pools (Burban & Petit, 2003; Bucci et al., 2007; Naydenov et al., 2014).

Additional studies point to further evolutionary forces having played a role in shaping the regional structure of maritime pine, in particular selection. González-Martínez et al. (2002) detected higher differentiation in quantitative traits (i.e. probably driven by environmental factors) compared with molecular markers (i.e. a priori neutral) in populations from the western-most range of the species, while Jaramillo-Correa et al. (2015) detected a link between the genetic variation of potentially adaptive SNPs and the environmental heterogeneity at the regional scale. These results suggest that maritime pine populations have adapted to the novel and highly diverse ecological conditions after glaciation (see Cacho et al., 2010, for a description of climatic deglaciation processes), and ultimately to different climatic niches across its current distribution.

The role of adaptation is further confirmed by our results, as the clusters defined in the environmental space (G1/G3/G8, G5/G6/G7, G2 and G4; Fig. 3) match the four basic clusters detected with the phylogenetic analyses (Fig. 1), pointing to an association between the environment and the delineation of the genetic clades. Additionally, environmental distance was shown to contribute more than geographical distance in explaining population divergence in maritime pine. Finally, most of the area suitable for maritime pine (83.6%) has only been projected as suitable for one specific genetic clade, suggesting that each clade has specialized to the specific climatic conditions it occupies (as in D’Amen et al., 2013, for African mammals). To sum up, these results indicate that regional environmental adaptation has contributed to the process of genetic diversification in maritime pine.

Evaluating the clusters in more detail reveals that the clades that are phylogenetically most closely related (G5/G6/G7 and G1/G8; see Fig. 1) show a certain degree of niche overlap, as indicated by their high estimates for environmental overlap (Table 1). Indeed, clades G5/G6/G7 represent a genetic admixture, as shown by the estimated membership coefficients (Q) from the structure clustering analysis (Appendix S1). The lack of major geographical and climatic barriers that would impede gene flow among these populations is reflected by the short genetic distances among them. A similar hypothesis can be drawn for clades G1/G3/G8, although it would be necessary to repeat the sampling for SNPs and nSSR for G8 (see Results). In their case, however, there are some geographical and climatic barriers that may have reduced gene flow and may thus have resulted in larger genetic distances. In this work, we have estimated environmental distances only within a bioclimatic framework. In future studies, it would be informative to include other factors relevant to determining species distribution if these results are to be confirmed. Edaphic properties, for example, are of great importance when studying the patterns of distribution

Figure 5 Comparison between projections of the species-level model (SDMsp not based on genetic clades) and the composite model (SDMcm based on molecular data). We distinguish four possible areas: agreement in presence, agreement in absence, projected presence only for SDMcm and projected presence only for SDMsp.
of vegetation, as they directly influence resource availability. In our study, the lack of edaphic maps capturing the local variability of the soil inhabited by maritime pine prevented us from integrating such valuable information into SDMs.

Despite the general insight gained of specialized niches for the main genetic clades, it is also possible that all the clades share the same fundamental niche, i.e. the abiotic conditions in which a species can survive (Hutchinson, 1957). Dispersal limitations and local biotic interactions may have prevented them from occupying all available environments, which would have translated into different realized niches, i.e. the abiotic and biotic conditions in which a species can survive. The integration of information from common garden experiments with maritime pine (Alía et al., 1997; Correia et al., 2010) could help to further test whether the different genetic clades share the same fundamental niche, or if each of them has maintained differences in its fundamental niche behaviour.

Importance of molecular information in SDMs

Maritime pine occupies an extremely variable range of habitats, (climates types from Atlantic to Mediterranean, and elevations from sea level to 2000 m), and its distribution is restricted by multiple factors (e.g. wildfires, land-use policies prioritizing plantations and farming activities, overexploitation). Considering these constraints, and the fact that this study focused solely on climatic variables, we expected the SDMs to predict a suitable habitat extending beyond the observed distribution of the species. However, SDMsp only predicted a climatically suitable area for the species that was very similar to its actual range, and only when genetic information was included in the models (SDMcmm) was the predicted spatial extent of suitable habitat largely broadened (Fig. 5). Because maritime pine has a broad climatic niche, and because our models do not constrain predictions geographically, the SDMcmm, based on the SDMs for infraspecific clades, seems more realistic than the SDMsp projection. Similar results were obtained by Benito-Garzón et al. (2011), who found that including local adaptation and plasticity in SDMs also enlarged the area suitable for maritime pine in future climate scenarios when compared with classical SDMs.

Potential uses in the management of genetic resources

SDMs relate environmental variables to species distribution, thus providing a key tool for supporting conservation management decisions (e.g. to identify critical habitats, assess the probable future distribution of the species or select reserve zones). Integrating molecular information into SDMs is essential, as genetic diversity is a key component of biodiversity (Convention on Biological Diversity, http://www.cbd.int/convention/text). Our approach, integrating all available genetic species information into SDMs and assessing the relationship among genetic and ecological clusters, can benefit conservation and sustainable use of forest genetic diversity in at least in three essential aspects. 1. EUFORGEN has implemented a network of genetic conservation units based on ecogeographical zonation and the distribution of tree species in a given country (Koskela et al., 2013). Infraspecific SDMs integrating molecular information can improve the definition of conservation units, as they provide spatially explicit information based on the performance of the species and its ability to colonize different ecological niches. In general, such models can help to define general conservation guidelines and to detect critically endangered species, genetic clades or areas (e.g. Thuiller et al., 2011).

2. The use of genetic diversity in assisted migration programmes needs to be have a solid scientific basis. This approach highlights areas that are bioclimatically suitable for the species but not yet occupied. Thus, this method can help to select the territory and the appropriate genetic clade for extending the distribution of the species, minimizing the risk of failure.

3. Finally, this study contributes to our knowledge about regions of provenance (the basic unit for commercialization of seeds and plants in the European Union), currently already in implementation in many countries (see Alía et al., 1996, for an example with maritime pine based on ecological, phenotypic and genetic similarities).

CONCLUSIONS

Climate change is nowadays guiding the conservation agenda, by helping to establish guidelines that ensure species survival. Forest trees, with their widespread distribution, are already facing a broad range of environmental conditions, which, as in the case of maritime pine, has led to distinct clades that are adapted to specific local climates. Understanding the ecological processes that have shaped the actual spatial structure of genetic clades is of great interest, not only for a better understanding of their adaptation to their native local environment but also to help to predict future behaviour and define conservation guidelines.

Our approach suggests that niche evolution is a major force driving the adaptive evolution of maritime pine clades. The broad spatial genetic structure that is observed today was most likely generated during the species’ colonization history, including ecological and evolutionary processes and their interactions, starting from Ice Age refugia. The concordance among the environmental and genetic clusters, and the fact that the environment explains phylogenetic structure better than geographical distance, point to adaptation as a driver in shaping spatial genetic patterns in maritime pine. One key point remains to be tackled: whether the integration of provenance trials in SDMs would confirm this conclusion, or would rather point to habitat fragmentation or historical isolation as drivers of spatial genetic patterns.

ACKNOWLEDGEMENTS

We are grateful to Ricardo Alía and Santiago C. González-Martínez for the insightful discussion about this paper, and to Achilleas Psomas for his collaboration during the analysis stage. Thanks are extended to Rose Ruiz Daniels for the English
language revision. We thank the COST Action FP1202 for enabling the collaboration between the different groups involved, which has supported the development of this paper. We acknowledge funding from the Spanish National Research Plan (AdapCon CGL2011-30182-C02-01), as well as grant FP7 NovelTree Breeding from the European Commission. M.S.V. was supported by a FPU from the Spanish Ministry of Education, Culture and Sport; D.G. was supported by a Ramón y Cajal Fellowship from the Spanish Ministry of Science and Innovation. We also thank two anonymous referees and two editors for their valuable help in improving the paper. Finally, we thank P. C. Grant-Pérez from Grant & Co. Academic English for revising the final manuscript.

REFERENCES


**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of this article at the publisher’s web-site.

**Appendix S1** Detailed description of genetic clades

**Table S1** Estimated membership coefficients ($Q$) for each individual within each cluster.

**Appendix S2** Detailed methods and specifications for the preparation of presence/pseudo-absence data inputs, algorithm settings for the species distribution models, model performance and distance scatterplots.

**Figure S1** Combination of EUFORGEN and Tree Species Distribution for Europe data bases for *Pinus pinaster* distribution.

**Figure S2** Scatterplots of phylogenetic versus environmental distances and phylogenetic versus geographical distances.
Table S2 Assessment of model performance by means of the true skill statistic, area under the curve and H-measure scores.

Table S3 True skill statistic, H-measure, cutoff, sensitivity and specificity scores of the ensemble model (average of the five algorithms), for the species level model and for the models for the different genetic clades (G1–G8).

Appendix S3 Detailed description of bioclimatic variables and the selection procedure

Appendix S4 Geographical projections of the five algorithms (generalized linear model, general additive model, random forest, classification tree analysis and MaxEnt) and the ensemble model of maritime pine and its genetic clades (G1–G8).

Figure S3–S11 Geographical projection of *P. pinaster* and its eight genetically defined clades using five algorithms.

**BIOSKETCH**

This study represents a collaborative effort aimed at integrating genetic information in ecological niche modelling within the framework of the ADAPCON project. Within this project, we investigate patterns of adaptive variation in mediterranean conifers, integrating genetic variation (neutral and potentially adaptive) in natural populations, phenotypic variation in common garden experiments and ecological niche modelling. This work will be included in the first author’s PhD thesis.

Editor: Linda Beaumont