Divergent selection in a Mediterranean pine on local spatial scales

Katharina B. Budde1 | Christian Rellstab2 | Myriam Heuertz3 | Felix Gugerli2 | Tom Hanika4,5 | Miguel Verdú6 | Juli G. Pausas6 | Santiago C. González-Martínez3

1Department of Forest Genetics, University of Goettingen, Goettingen, Germany; 2Swiss Federal Research Institute WSL, Birmensdorf, Switzerland; 3BIOGECO, INRAE, University of Bordeaux, Cestas, France; 4Institute of Computer Science, University of Hildesheim, Hildesheim, Germany; 5Berlin School of Library and Information Science, Humboldt-Universität zu Berlin, Berlin, Germany and 6Centro de Investigaciones sobre Desertificación (CIDE-CSIC/UV/GV), Moncada, Valencia, Spain

Correspondence
Katharina B. Budde
Email: k.budde@uni-goettingen.de

Funding information
European Commission, Grant/Award Number: EU2008-03713; European Science Foundation; Spanish Ministry of Science and Innovation, Grant/Award Number: CGL2008-05289-C02-01/02 and CGL2012-39938-C02-00/01; University of Brodeaux

Handling Editor: Stephen Bonser

Abstract

1. The effects of selection on an organism’s genome are hard to detect on small spatial scales, as gene flow can swamp signatures of local adaptation. Therefore, most genome scans to detect signatures of environmental selection are performed on large spatial scales; however, divergent selection on the local scale (e.g. between contrasting soil conditions) has also been demonstrated, in particular for herbaceous plants.

2. Here, we hypothesised that in topographically complex landscapes, microenvironment variability is strong enough to leave a selective footprint in the genomes of long-lived organisms. To test this, we investigated paired south- versus north-facing Pinus pinaster stands on the local scale, with trees growing in close vicinity (≤820 m distance between paired south- and north-facing stands), in a Mediterranean mountain area. While trees on north-facing slopes experience less radiation, trees on south-facing slopes suffer from especially harsh conditions, particularly during the dry summer season.

3. Two outlier analyses consistently revealed five putatively adaptive loci (out of 4034), in candidate genes two of which encoded non-synonymous substitutions. Additionally, one locus showed consistent allele frequency differences in all three stand pairs indicating divergent selection despite high gene flow on the local scale. Permutation tests demonstrated that our findings were robust.

4. Functional annotation of these candidate genes revealed biological functions related to abiotic stress response, such as water availability, in other plant species.

5. Synthesis. Our study highlights how divergent selection in heterogeneous microenvironments shapes and maintains the functional genetic variation within populations of long-lived forest tree species, being the first to focus on adaptive genetic divergence between south- and north-facing slopes within continuous...
forest stands. This is especially relevant in the current context of climate change, as this variation is at the base of plant population responses to future climate.

**KEYWORDS**
candidate genes, $F_{ST}$ outlier, local adaptation, paired sampling design, *Pinus pinaster*, south-versus north-facing slopes

## 1 | INTRODUCTION

Spatially heterogeneous environments exert divergent selection pressures and can contribute to maintaining high levels of adaptive genetic variation within populations (Richardson et al., 2014). However, understanding under which circumstances, selection is acting and especially on which spatial scale divergence can be detected remains poorly understood. Studying local adaptation in forest tree species is an important endeavour especially under current climate change (Aitken et al., 2008; Kremers et al., 2012). Trees are the dominant species in forest ecosystems and provide important ecosystem services, such as habitat for associated species or local climate regulation and carbon storage (Gamfeldt et al., 2013; Mori, 2017). They are long-lived organisms and thus, are exposed to diverse biotic and abiotic conditions throughout their lifetime (Petit & Hampe, 2006). Nevertheless, trees commonly exhibit local adaptation (Leites & Benito Garzón, 2023). Rapidly progressing anthropogenic climate change is currently challenging the persistence of tree populations (Allen et al., 2015) and their longevity and long generation time makes them particularly vulnerable to fast environmental change.

Numerous studies have already revealed loci potentially involved in environmental adaptation. However, most of these studies have been conducted on regional to continental scales (Holliday et al., 2012; Martins et al., 2018; Postolache et al., 2021; Stöltig et al., 2015), as gene flow on small spatial scales (i.e. tens to hundreds of metres) can blur the migration–selection equilibrium maintaining local adaptation. Divergent selection on the local scale has often been observed in herbaceous plant species (e.g. to toxic soil conditions; Antonovics, 1970; Jain & Bradshaw, 1966; Turner et al., 2010). There is increasing evidence that plants exhibit adaptive divergence on very small spatial scales, that is within tens of metres of distance in herbaceous species and hundreds of metres of distance in some woody species (reviewed in Linhart & Grant, 1996; Scotti et al., 2016). Recent studies have started to address the factors shaping local adaptation on the microenvironmental scale in long-lived tree species (Budde et al., 2017; Dauphin et al., 2021; Mosca et al., 2016; Scotti et al., 2023; von Takach et al., 2021). Several temperate and tropical tree species show adaptation to microenvironmental conditions (Mosca et al., 2016; Schmitt et al., 2021, 2022; Scotti et al., 2023). *Eperua falcata* (Fabaceae), for example, shows divergent selection between groups of individuals growing in seasonally flooded bottomlands and adjacent groups growing on dry *terra firme* soils (Audigeos et al., 2013; Brousseau et al., 2021). Also, Gauzere et al. (2020) found evidence for divergent selection acting on growth and phenology traits along an elevational gradient of European beech (*Fagus sylvatica*), despite high gene flow.

Identifying the genes and gene variants that confer local adaptation, that is, higher fitness in the local habitat/environment, is of great interest in ecology and evolution. It is especially relevant under climate change to sustainably conserve and manage forest genetic resources. The detection and validation of candidate loci potentially under selection, however, remains challenging. Experimental functional validation is not attainable in non-model species, especially in trees with their longevity and long generation times. Previous studies showed that many approaches to detect loci under selection can be prone to false positives (e.g. de Villemereuil et al., 2014; Lotterhos & Whitlock, 2015) and that the identified genomic signatures of selection might not always be observed in other locations with similar environmental conditions (Buehler et al., 2014; Korves et al., 2007; Rellstab et al., 2017). Therefore, combining several analytical approaches is recommended to reduce false positives (Rellstab et al., 2015). Additionally, an appropriate sampling design can increase the power to detect loci involved in local adaptation. A replicated, paired design, comprising sampling sites with contrasting environmental conditions, seems promising for the detection of loci displaying weak signatures of selection. A simulation study by Lotterhos and Whitlock (2015) showed that sampling pairs of nearby populations (i.e. at gene flow distance) with contrasting environmental conditions increase the probability of detecting true-positive outlier loci compared to gradient or random sampling designs.

In Mediterranean ecosystems, water availability is one of the most important factors driving selection, and plant species are typically well adapted to summer dry conditions (Nardini et al., 2014; Ramírez-Valiente et al., 2022). Still, considerable microenvironmental variation can be observed especially in topographically complex landscapes, such as Mediterranean mountain systems in southern Europe. In such systems, equator-facing (i.e. south-facing in the northern hemisphere) slopes receive higher solar radiation flux density, leading to higher evapotranspiration rates and higher daily maximum temperatures during summer drought periods (Ivanov et al., 2008). Therefore, these slopes show a significantly different composition, structure and density of plant communities as compared to pole-facing (i.e. north-facing in the northern hemisphere) slopes (Cantlon, 1953; Pigott & Pigott, 1993; Sternberg & Shoshany, 2001).

In this study, we tested whether, in topographically complex Mediterranean forests, microclimate variability, potentially due to differences in water availability, is strong enough to leave a selective
footprint on long-lived trees despite the massive gene flow expected in continuous forest populations. We used a robust sampling design with replicated pairs of stands in close vicinity within a natural population of maritime pine (Pinus pinaster Aiton, Pinaceae) to identify consistent genetic signatures of divergent selection between xeric (south-facing slope) and mesic (north-facing slope) conditions. Finally, an extensive resampling procedure was developed to assess the robustness of the identification of outlier loci.

2 | MATERIALS AND METHODS

2.1 | Study species and sample collection

Maritime pine is a monoecious conifer species growing in the western part of the Mediterranean basin and along the Atlantic coast in south-western Europe. It is a wind-pollinated species with wind-dispersed seeds. Pollen flow is therefore wide ranging, following highly leptokurtic dispersal kernels with average dispersal distances of 78–174 m and frequent long-distance dispersal events (De-Lucas et al., 2008). Gene flow via seeds is more restricted (average of 26.53 m; González-Martínez et al., 2002), but post-dispersal processes, such as higher seedling mortality close to the mother plant (i.e. the Janzen–Connell effect; Connell, 1971; Janzen, 1970), and microenvironmental variation affecting survival at early life stages can substantially increase effective dispersal distances (González-Martínez et al., 2006).

For this study, we sampled P. pinaster in six locations, comprising three pairs of stands with contrasting microenvironmental conditions, in a natural forest near Esilda in Sierra de Espadán, eastern Spain (Figure 1). No special licences or permits were required for the fieldwork and P. pinaster sample collection. This region is characterised by a warm and dry Mediterranean climate. The eastern Spain P. pinaster forests constitute a single gene pool with low genetic differentiation among stands (Budde et al., 2014; Jaramillo-Correa et al., 2015). We therefore consider all six sample stands as a single population. Stand-replacing crown fire events are common in the region (Pausas & Paula, 2012). Under these conditions, regeneration is mostly driven by fire events leading to even-aged cohorts (Hernandez-Serrano et al., 2013). The sampling was performed in a hilly environment and each stand pair consisted of trees from one mainly south- and one mainly north-exposed slope. In detail, one stand pair consisted of one stand on a south-facing slope (with mainly southern aspect) and another stand with trees from a nearby shady valley along a (mostly) north-exposed stream (S1/N1). The other two stand pairs (S2/N2 and S3/N3) were not crossed by any stream and grew on mostly south- (dry and warm) and north-facing slopes (more humid and cooler). For simplicity, we will refer to this sampling design as three pairs of south- and north-facing slopes. In each of the six stands, we haphazardly sampled 25 trees with DBH (diameter at breast height) > 16 cm and separated by a minimum distance of 10 m, making a total of 150 trees (Figure 1; Table 1). The straight line distance between samples ranged from 3.7 to 10 km between stand pairs and from 10 to 820 m between trees within pairs. All trees were georeferenced using a Garmin Oregon 550t (Garmin, Wichita, USA), height was assessed using a Digital hypsometer Forester Vertex (Haglöf, Långsele, Sweden) and DBH was measured with a tree calliper.

2.2 | DNA extraction and genotyping

Needles were collected from the 150 trees and desiccated using silica gel. Genomic DNA was isolated using the Invisorb® DNA Plant
TABLE 1  Paired stand sampling of south- (S) and north-facing (N) slopes for *Pinus pinaster* in the Sierra de Espadán (eastern Spain), and genetic diversity estimates based on 5024 single-nucleotide polymorphisms.

<table>
<thead>
<tr>
<th>ID</th>
<th>Lat.</th>
<th>Long.</th>
<th>Aspect [°]</th>
<th>Elevation [m a.s.l.]</th>
<th>Ns</th>
<th>Height [m (SD)]</th>
<th>Poly. Loci</th>
<th>H_e (SE)</th>
<th>F_IS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>39.865</td>
<td>-0.298</td>
<td>185</td>
<td>632–666</td>
<td>25</td>
<td>7.300 (1.524)</td>
<td>4739</td>
<td>0.336 (0.003)</td>
<td>-0.018</td>
</tr>
<tr>
<td>N1</td>
<td>39.866</td>
<td>-0.298</td>
<td>297</td>
<td>645–737</td>
<td>25</td>
<td>8.732 (1.334)</td>
<td>4777</td>
<td>0.330 (0.003)</td>
<td>0.008</td>
</tr>
<tr>
<td>Pair 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>39.895</td>
<td>-0.353</td>
<td>159</td>
<td>719–763</td>
<td>25</td>
<td>8.716 (1.763)</td>
<td>4778</td>
<td>0.328 (0.003)</td>
<td>0.013</td>
</tr>
<tr>
<td>N2</td>
<td>39.895</td>
<td>-0.351</td>
<td>35</td>
<td>655–730</td>
<td>25</td>
<td>12.496 (1.900)</td>
<td>4803</td>
<td>0.337 (0.003)</td>
<td>-0.004</td>
</tr>
<tr>
<td>Pair 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>39.917</td>
<td>-0.397</td>
<td>181</td>
<td>655–728</td>
<td>25</td>
<td>8.716 (1.763)</td>
<td>4745</td>
<td>0.332 (0.003)</td>
<td>-0.005</td>
</tr>
<tr>
<td>N3</td>
<td>39.913</td>
<td>-0.389</td>
<td>340</td>
<td>696–731</td>
<td>25</td>
<td>11.104 (1.981)</td>
<td>4753</td>
<td>0.326 (0.003)</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Abbreviations: Aspect, average aspect in degrees; Elevation, elevation in metres above sea level; F_IS, fixation index; H_e, expected heterozygosity; Height, tree height in metres with standard deviation; ID, identifier for each stand (see Figure 1); Latitude, latitude in decimal degrees; Longitude, longitude in decimal degrees; Ns, number of samples; poly. loci, number of polymorphic loci; SE, standard error.

HTS 96 Kit/C kit (Invitae GmbH, Berlin, Germany) following the manufacturer’s instructions.

An *Illumina Infinium* SNP (single-nucleotide polymorphism) array (Illumina, Inc., San Diego, USA), developed by Plomion et al. (2016), was used for genotyping. This array is enriched in SNPs from genes that showed signatures of natural selection in previous studies (Eveno et al., 2008; Grivet et al., 2011; Jaramillo-Correa et al., 2015) or differential expression under biotic and abiotic stress (Plomion et al., 2016) in maritime pine. After removing SNPs with uncertain scoring based on visual inspection using *GenomeStudio Genotyping Module* v1.0 (Illumina, Inc.) and monomorphic loci, we kept 5024 high-quality SNPs, of which 4034 had a minor allele frequency (MAF) > 0.1. The amount of missing genotype data per stand was very low (maximum of 1%). This data set has recently been used to characterise the effective population size in Sierra de Espadán, as part of a meta-analysis (Budde & González-Martínez, 2022; Santos del Blanco et al., 2022).

2.3  Data analyses

As a preliminary step, we tested the hypothesis that south- and north-facing slopes were subjected to distinct selection pressures by looking at phenotypic differences in height and DBH between the stands in each pair, using two-sample Student’s *t*-tests in R v. 4.1.2 (R Core Team, 2021). Then, based on the SNP data, we estimated genetic diversity parameters, such as the number of polymorphic loci, the observed and expected heterozygosity and the fixation index using the R package *hierfstat* (Goudet, 2005). These genetic diversity statistics were computed to test whether different selection pressures in south- and north-facing slopes resulted in different levels of gene diversity. After this, we tested for genetic differentiation between the sampled stands by estimating pairwise *F*<sub>ST</sub> between all stands, overall *F*<sub>ST</sub> (*F*<sub>ST,global</sub>) and comparing all south- versus all north-facing slopes (*F*<sub>ST,S/N</sub>) (Weir & Cockerham, 1984), using the complete SNP data set and comparing with neutral expectations from 1000 permutations. To visualise the neutral population genetic structure inherent to our data, we also performed a principal component analysis (PCA) using the function *dudi.pca* implemented in the R package ade4 (Dray & Dufour, 2007) and a supervised (i.e. defining each stand as a group) discriminant analysis of principal components (DAPC) using the *dapc* function in the R package adegenet (Jombart, 2008) based on all SNP markers. The PCA reflects the largest total variance in the data while the DAPC maximises the differentiation between sampled stands and enables the visualisation even of weak population genetic structure (Jombart et al., 2010). Additionally, we assessed the fine-scale spatial genetic structure (FSGS), that is, the decay of kinship with geographic distance between pairs of individuals, within each of the three stand pairs. The two stands of each pair were analysed together in order to increase sample size and the maximum distance among trees (i.e. to obtain distances over 150 m between pairs of trees), two critical sampling requirements for FSGS analyses (see e.g. De-Lucas et al., 2009).

First, we estimated the pairwise Loiselle kinship coefficient between individuals (Loiselle et al., 1995) in SPAGEDi v. 1.5d (Hardy & Vekemans, 2002). The Loiselle kinship coefficient is most suitable for codominant markers and particularly for loci with few alleles, such as SNPs (Vekemans & Hardy, 2004). It is also widely used in studies assessing FSGS in tree populations and hence enables comparisons between studies (Gonçalves et al., 2022; Vekemans & Hardy, 2004).

The average kinship coefficient per distance class was regressed against the logarithm of spatial distances and significance was assessed based on 10,000 permutations of individual locations. The strength of FSGS was estimated as $Sp = b/(1 - F_{IS})$, where $b$ is the regression slope and $F_{IS}$ is the average kinship coefficient in the first distance class (Vekemans & Hardy, 2004).

To detect loci potentially under selection in trees growing on slopes with contrasting aspects (south/north) in a hierarchically structured population (Excoffier et al., 2009), we used two hierarchical *F*<sub>ST</sub> outlier detection approaches. These take into account the paired sampling design: one implemented in Arlequin v 3.5.2 (Excoffier &
Lischer, 2010) and the other in BayeScanH, which is especially suitable for small sample sizes (Foll et al., 2014). We used the data set based on MAF > 0.1, comprising 4034 SNPs for these analyses and suitable input data formats were generated using PGDspider (Lischer & Excoffier, 2012). ARLECORE, the command line version of Arlequin was employed, to estimate \( F_{ST} \) outlier loci for divergence between stands (south- versus north-facing) within pairs based for all 4034 SNPs with MAF > 0.1 and 20,000 simulations. To identify outlier loci with BayeScanH, we used default parameters with an odds prior of 10. We tested two models, one with the same selection pressure acting between contrasting slopes in the three stand pairs and another one with three independent selection pressures on contrasting slopes within the three pairs. Finally, using the R-script ‘paired_GEA.R’ (included in Hanika et al., 2023), we tested if any of the candidate SNPs identified with Arlequin/ARLECORE or BayeScanH showed consistent patterns in allele frequencies between the paired stands in all replicates. For this, we checked whether the differences in allele frequency had the same sign in all pairs, that is, whether the allele frequency in all north-facing slopes was consistently lower or higher than the allele frequency in all south-facing slopes (denoted ‘strict sign’). Then, using the same script, we ran single-locus linear mixed models using the function lme implemented in the package nlme (Pinheiro et al., 2014), with allele frequency as response variable, slope aspect (categorical, south/ north) as fixed effect and pair as random factor.

We combined several methods, namely hierarchical \( F_{ST} \) outlier detection models, implemented in Arlequin/ARLECORE and BayeScanH, and the paired GEA tests to detect loci potentially under selection. Combining different approaches has been recommended to avoid the detection of false positives (Rellstab et al., 2015). However, to assess formally the robustness of our approach, we randomised the input data set by permuting the assignment of trees to south- and north-facing slopes within the three sample pairs (Figure S4.1; Hanika et al., 2023). The randomised data sets were analysed using the same methods with the same model parameters as the observed data. The results were stored in a relational database. We then estimated empirical \( p \)-values of the observed values based on the distribution of significant outlier loci detected in about 1000 randomised data sets. In the same way, empirical \( p \)-values were computed for the overlap of outlier loci detected by several approaches. These \( p \)-values can be understood as the probability that the number of significant outlier loci detected in a randomised data set is equal to or above the observed value (see also Rellstab et al., 2020; Table S4.1).

The sequences flanking SNPs identified as loci potentially under selection in our observed data, and associated annotation, were retrieved from Plomion et al., (2016). These sequences were newly blasted against the NCBI nucleotide database to check for new functional annotations (Table S3.2).

3 RESULTS

Tree height was consistently lower in south- than in north-facing slopes and the difference was significant in two out of the three stand pairs (Figure S1.1; Table S1.1), while no significant difference was detected for DBH in any stand pair. Of the total of 5024 SNPs, between 4745 and 4803 were found within single stands, and these numbers were consistently slightly lower on south- compared to north-facing slopes of the same sample pair (Table 1). Expected and observed heterozygosity (not shown) were very similar in all six study stands with values around 0.337, resulting in fixation indices close to zero. Pairwise genetic differentiation (\( F_{ST} \)) between stands based on all 5024 SNP markers was weak, ranging from 0.004 to 0.033, but highly significantly above zero (\( p < 0.001 \); Table S2.1). Global \( F_{ST} \) over all six stands (\( F_{ST,global} \)) was 0.021 (\( p < 0.001 \)) and pairwise \( F_{ST} \) between all south- versus all north-facing slopes (\( F_{ST,N/S} \)) was 0.005 (\( p < 0.001 \)). The overall population genetic structure was thus weak albeit significant. Moreover, the DAPC clearly depicted the hierarchical population structure due to the paired sampling design, with weak but slightly stronger genetic differentiation among than within stand pairs (i.e. between south- and north-facing stands; Figure 2). The hierarchical population structure was also visible but less evident in the PCA plot (Figure S2.1). Overall, these analyses suggested a neutral hierarchical population genetic structure that needed to be accounted for in SNP-outlier detection. Finally, FSGS was also significant, showing isolation by distance, in all stand pairs and strongest in pair N3/S3 (Figure S2.2).

In total, 45 SNPs were located above the 99% confidence interval using the hierarchical island model in Arlequin/ARLECORE and, thus, were considered as significant outliers for genetic differentiation between south- and north-facing slopes (Figure 3; Table S3.1, S3.2). Additionally, 10 loci were identified as significant \( F_{ST} \) outliers by BayeScanH when assuming independent selection pressures for each of the three pairs of stands (Table S3.1). None of these outlier loci was significant in all three sampling pairs in BayeScanH. Moreover, no significant outlier locus was detected when assuming the same selection pressure in all three pairs of stands.
When comparing the two methods, five loci were identified as outliers by both Arlequin/arlecore and BayeScanH (Figure S3.1), and only one additional outlier locus, AL751008_691 detected by Arlequin/arlecore, showed consistent allele frequency differences between south- and north-facing slopes (Figure 4) and a significant effect of the site aspect as indicated by the linear mixed model ($p_{site\ type}=0.002$). Two of these six outlier SNPs showed non-synonymous changes and coded for a putative RNA-binding protein and a V-type proton ATPase catalytic subunit respectively (Table 2).

We generated and analysed about 1000 randomised data sets. The empirical $p$-values estimated from the distribution of the outlier loci detected in the randomised data sets gave clear evidence that it was highly unlikely to observe 45 significant outlier loci in Arlequin/arlecore ($p_{arlecore}<0.001$; Table S4) just by chance. The probability to randomly detect 10 outlier loci in BayeScanH using the model with three independent selection pressures was not so low ($p_{BayeScanH}=0.198$) in the randomised data. However, an overlap of five significant outlier loci between Arlequin/arlecore and BayeScanH was unlikely ($p_{A/B}=0.003$), and one significant outlier locus detected by both Arlequin/arlecore and the paired GEA approach was highly unlikely ($p_{A/GEA}<0.001$) in the randomised data sets.

**4 | DISCUSSION**

The paired sampling design in Sierra de Espadán, contrasting south- and north-facing slopes within a large and continuous *P. pinaster* population, was specifically used to test for microenvironmental adaptation most likely driven by different levels of water availability. Paired sampling in stands with contrasting environments, such as dry versus humid patches, represents a powerful approach to reveal loci under selection (Lotterhos & Whitlock, 2015; Rellstab et al., 2015), because it maximises the potential to detect divergent selection.
while minimising the effect of confounding population structure. Several studies employed the paired sampling design to detect loci under selection (e.g. Bekkevold et al., 2020; Brousseau et al., 2016, 2021; Daniels et al., 2019; Dauphin et al., 2021; Rollstab et al., 2020; Roschanski et al., 2016; Scotti et al., 2023). In conifers, four previous studies revealed loci significantly associated with elevation or other microenvironmental gradients in Abies alba (Brousseau et al., 2016; Roschanski et al., 2016), Pinus halepensis (Daniels et al., 2019) and P. pinaster and Cedrus atlantica (Scotti et al., 2023), but only few loci showed consistent patterns of allele frequency shifts along the replicated stand pairs. These studies tested divergent selection mainly between high and low elevation and between sites at a distance of ≥1000 m and, thus, may have been influenced by environmental variation operating on regional scales. Here, we specifically tested for consistent patterns of divergent selection between stands growing in close vicinity (≤820 m), on south- and north-facing slopes that likely experience extensive gene flow and differences in water availability. To our knowledge, this is the first study in forest trees focusing on the genetic divergence in stands with different aspect within continuous forests.

We first showed a hierarchical population genetic structure despite high gene flow in P. pinaster within one large population in Sierra de Espadán (eastern Spain). These forests constitute a single gene pool (Budde et al., 2014; Jaramillo-Correa et al., 2015); therefore, we would expect a weak fine-scale spatial genetic structure in this wind-pollinated and wind-dispersed species (De-Lucas et al., 2009). However, we found low but significant differentiation among all stands, including between the neighbouring south- and north-facing slopes, clearly depicting the hierarchical population structure. This finding guided us in our choice for hierarchical island models for consistent patterns of divergent selection between stands grow- ing slopes. Clonal common gardens of range-wide P. pinaster showed consistent allele frequency shifts along the replicated stand pairs. These studies tested divergent selection mainly between high and low elevation and between sites at a distance of ≥1000 m and, thus, may have been influenced by environmental variation operating on regional scales. Here, we specifically tested for consistent patterns of divergent selection between stands growing in close vicinity (≤820 m), on south- and north-facing slopes that likely experience extensive gene flow and differences in water availability. To our knowledge, this is the first study in forest trees focusing on the genetic divergence in stands with different aspect within continuous forests.

We first showed a hierarchical population genetic structure despite high gene flow in P. pinaster within one large population in Sierra de Espadán (eastern Spain). These forests constitute a single gene pool (Budde et al., 2014; Jaramillo-Correa et al., 2015); therefore, we would expect a weak fine-scale spatial genetic structure in this wind-pollinated and wind-dispersed species (De-Lucas et al., 2009). However, we found low but significant differentiation among all stands, including between the neighbouring south- and north-facing slopes, clearly depicting the hierarchical population structure. This finding guided us in our choice for hierarchical island models for consistent patterns of divergent selection between stands growing slopes. Clonal common gardens of range-wide P. pinaster showed consistent allele frequency shifts along the replicated stand pairs. These studies tested divergent selection mainly between high and low elevation and between sites at a distance of ≥1000 m and, thus, may have been influenced by environmental variation operating on regional scales. Here, we specifically tested for consistent patterns of divergent selection between stands growing in close vicinity (≤820 m), on south- and north-facing slopes that likely experience extensive gene flow and differences in water availability. To our knowledge, this is the first study in forest trees focusing on the genetic divergence in stands with different aspect within continuous forests.

We first showed a hierarchical population genetic structure despite high gene flow in P. pinaster within one large population in Sierra de Espadán (eastern Spain). These forests constitute a single gene pool (Budde et al., 2014; Jaramillo-Correa et al., 2015); therefore, we would expect a weak fine-scale spatial genetic structure in this wind-pollinated and wind-dispersed species (De-Lucas et al., 2009). However, we found low but significant differentiation among all stands, including between the neighbouring south- and north-facing slopes, clearly depicting the hierarchical population structure. This finding guided us in our choice for hierarchical island models for consistent patterns of divergent selection between stands growing slopes. Clonal common gardens of range-wide P. pinaster showed consistent allele frequency shifts along the replicated stand pairs. These studies tested divergent selection mainly between high and low elevation and between sites at a distance of ≥1000 m and, thus, may have been influenced by environmental variation operating on regional scales. Here, we specifically tested for consistent patterns of divergent selection between stands growing in close vicinity (≤820 m), on south- and north-facing slopes that likely experience extensive gene flow and differences in water availability. To our knowledge, this is the first study in forest trees focusing on the genetic divergence in stands with different aspect within continuous forests.

We first showed a hierarchical population genetic structure despite high gene flow in P. pinaster within one large population in Sierra de Espadán (eastern Spain). These forests constitute a single gene pool (Budde et al., 2014; Jaramillo-Correa et al., 2015); therefore, we would expect a weak fine-scale spatial genetic structure in this wind-pollinated and wind-dispersed species (De-Lucas et al., 2009). However, we found low but significant differentiation among all stands, including between the neighbouring south- and north-facing slopes, clearly depicting the hierarchical population structure. This finding guided us in our choice for hierarchical island models for consistent patterns of divergent selection between stands growing slopes. Clonal common gardens of range-wide P. pinaster showed consistent allele frequency shifts along the replicated stand pairs. These studies tested divergent selection mainly between high and low elevation and between sites at a distance of ≥1000 m and, thus, may have been influenced by environmental variation operating on regional scales. Here, we specifically tested for consistent patterns of divergent selection between stands growing in close vicinity (≤820 m), on south- and north-facing slopes that likely experience extensive gene flow and differences in water availability. To our knowledge, this is the first study in forest trees focusing on the genetic divergence in stands with different aspect within continuous forests.

We first showed a hierarchical population genetic structure despite high gene flow in P. pinaster within one large population in Sierra de Espadán (eastern Spain). These forests constitute a single gene pool (Budde et al., 2014; Jaramillo-Correa et al., 2015); therefore, we would expect a weak fine-scale spatial genetic structure in this wind-pollinated and wind-dispersed species (De-Lucas et al., 2009). However, we found low but significant differentiation among all stands, including between the neighbouring south- and north-facing slopes, clearly depicting the hierarchical population structure. This finding guided us in our choice for hierarchical island models for consistent patterns of divergent selection between stands growing slopes. Clonal common gardens of range-wide P. pinaster showed consistent allele frequency shifts along the replicated stand pairs. These studies tested divergent selection mainly between high and low elevation and between sites at a distance of ≥1000 m and, thus, may have been influenced by environmental variation operating on regional scales. Here, we specifically tested for consistent patterns of divergent selection between stands growing in close vicinity (≤820 m), on south- and north-facing slopes that likely experience extensive gene flow and differences in water availability. To our knowledge, this is the first study in forest trees focusing on the genetic divergence in stands with different aspect within continuous forests.

We first showed a hierarchical population genetic structure despite high gene flow in P. pinaster within one large population in Sierra de Espadán (eastern Spain). These forests constitute a single gene pool (Budde et al., 2014; Jaramillo-Correa et al., 2015); therefore, we would expect a weak fine-scale spatial genetic structure in this wind-pollinated and wind-dispersed species (De-Lucas et al., 2009). However, we found low but significant differentiation among all stands, including between the neighbouring south- and north-facing slopes, clearly depicting the hierarchical population structure. This finding guided us in our choice for hierarchical island models for consistent patterns of divergent selection between stands growing slopes. Clonal common gardens of range-wide P. pinaster showed consistent allele frequency shifts along the replicated stand pairs. These studies tested divergent selection mainly between high and low elevation and between sites at a distance of ≥1000 m and, thus, may have been influenced by environmental variation operating on regional scales. Here, we specifically tested for consistent patterns of divergent selection between stands growing in close vicinity (≤820 m), on south- and north-facing slopes that likely experience extensive gene flow and differences in water availability. To our knowledge, this is the first study in forest trees focusing on the genetic divergence in stands with different aspect within continuous forests.

We first showed a hierarchical population genetic structure despite high gene flow in P. pinaster within one large population in Sierra de Espadán (eastern Spain). These forests constitute a single gene pool (Budde et al., 2014; Jaramillo-Correa et al., 2015); therefore, we would expect a weak fine-scale spatial genetic structure in this wind-pollinated and wind-dispersed species (De-Lucas et al., 2009). However, we found low but significant differentiation among all stands, including between the neighbouring south- and north-facing slopes, clearly depicting the hierarchical population structure. This finding guided us in our choice for hierarchical island models for consistent patterns of divergent selection between stands growing slopes. Clonal common gardens of range-wide P. pinaster showed consistent allele frequency shifts along the replicated stand pairs. These studies tested divergent selection mainly between high and low elevation and between sites at a distance of ≥1000 m and, thus, may have been influenced by environmental variation operating on regional scales. Here, we specifically tested for consistent patterns of divergent selection between stands growing in close vicinity (≤820 m), on south- and north-facing slopes that likely experience extensive gene flow and differences in water availability. To our knowledge, this is the first study in forest trees focusing on the genetic divergence in stands with different aspect within continuous forests.

We first showed a hierarchical population genetic structure despite high gene flow in P. pinaster within one large population in Sierra de Espadán (eastern Spain). These forests constitute a single gene pool (Budde et al., 2014; Jaramillo-Correa et al., 2015); therefore, we would expect a weak fine-scale spatial genetic structure in this wind-pollinated and wind-dispersed species (De-Lucas et al., 2009). However, we found low but significant differentiation among all stands, including between the neighbouring south- and north-facing slopes, clearly depicting the hierarchical population structure. This finding guided us in our choice for hierarchical island models for consistent patterns of divergent selection between stands growing slopes. Clonal common gardens of range-wide P. pinaster showed consistent allele frequency shifts along the replicated stand pairs. These studies tested divergent selection mainly between high and low elevation and between sites at a distance of ≥1000 m and, thus, may have been influenced by environmental variation operating on regional scales. Here, we specifically tested for consistent patterns of divergent selection between stands growing in close vicinity (≤820 m), on south- and north-facing slopes that likely experience extensive gene flow and differences in water availability. To our knowledge, this is the first study in forest trees focusing on the genetic divergence in stands with different aspect within continuous forests.
Environmental conditions on south- and north-facing slopes are known to differ strongly, for example, in light and water availability (Bennie et al., 2008). Slopes with different aspects are often characterised by differences in composition, structure and density of plant communities (Cantlon, 1953; Pigott & Pigott, 1993; Sternberg & Shoshany, 2001) and tree species are known to have developed diverse adaptations in response to strong selection pressures in dry environmental conditions (Aranda et al., 2013). Pinus pinaster stands as a suitable study species to test for divergent selection related to the environment. Multisite clonal common gardens comprising range-wide populations already revealed that the species is generally susceptible to drought. Survival was lowest in the common garden sites with the harshest (dry and hot) conditions (Rodríguez-Quilón et al., 2016), and certain alleles at candidate loci associated with climate were connected to a higher probability of survival (Jaramillo-Correa et al., 2015). Here, we showed that contrasting environmental conditions on different slopes, in direct vicinity and in the presence of gene flow, could also shape the distribution of genetic variation in long-lived forest trees, such as P. pinaster.

Outlier loci related to differences in drought intensity and temperature have been found in different pine species on range-wide spatial scales. For example, Eckert et al. (2010), found five outlier loci associated with aridity in Pinus taeda. In natural Pinus albicaulis populations, Lind et al. (2017) also identified water availability as a strong driver of genomic signatures of adaptation. They detected allele frequency changes at candidate genes along a precipitation gradient on the regional scale in the Lake Tahoe Basin, an ecosystem similar to that studied here (i.e. Mediterranean-type mountains). Candidate gene approaches in maritime pine also found various outlier loci related to drought response and precipitation on large spatial scales (Eveno et al., 2008; Grivet et al., 2011; Jaramillo-Correa et al., 2015) and between shady and sunny stands on the microenvironmental scale (Scotti et al., 2023). Our study detected a small number of outlier loci likely related to water availability on south- and north facing slopes in maritime pine on the local scale, that is, within gene flow distance. One of these outlier loci (CT384-490, coding for a non-synonymous change) has been previously associated with winter precipitation on the range-wide scale (Jaramillo-Correa et al., 2015). Four of the six candidate genes, in which loci showing strong evidence of divergent selection were located, were functionally annotated and two of them could be used for non-synonymous changes. Locus BX250086 coded for an oligouridylylate-binding (like) protein and BX251523 for a V-type proton ATPase catalytic subunit. Locus i09683s215pg, which is coding for a non-synonymous change, is located in a gene encoding for a raffinose_syn domain containing protein. Genes annotated with similar functions have been described to be involved in abiotic stress response, such as drought stress, in other plant species (Ambrosone et al., 2012; Dietz et al., 2001; Vinson et al., 2020).

In the last years, reference genomes, even for conifer species with extremely large genomes (>18 Gbp), have been published (Mosca et al., 2019; Nystedt et al., 2013; Wegrzyń et al., 2014); however, the functional annotation of conifer genomes is still limited and a reference genome for P. pinaster is lacking. In this study, we were able to retrieve putative annotations for only four of six candidate genes in which significant outlier loci were found, highlighting the need to complete and improve our knowledge of conifer genomes and their functional annotation. In addition, although we were able to identify some candidate loci under divergent selection on the local scale, only one locus showed consistent differences in allele frequencies in all three stand pairs. This finding agrees with a recent study by Scotti et al. (2023), in which only a small number of outlier loci (0.1%-1% of all loci, depending on the species) showed consistent allele frequency differences between pairs of sites with contrasting conditions. This finding indicated that common signatures of selection are scarce. In BayeScanH, significant results were only obtained when assuming three independent selection pressures, which suggested the probable existence of differences in strength and direction of selection pressures even on very small spatial scales. This is consistent with other studies employing replicated paired sample designs (Brousseau et al., 2016; Daniels et al., 2019; Roschanski et al., 2016; Scotti et al., 2023), highlighting the complexity of selection drivers and the difficulties to identify them in natural experimental settings.

5 | CONCLUSIONS

Our findings are in line with recent studies that identified loci under divergent selection between stands growing in contrasting environmental conditions on the local (regional) scale in long-lived forest trees (Audiego et al., 2013; Brousseau et al., 2021; Csilléry et al., 2014; Gauzere et al., 2020). We extended these studies by revealing consistent patterns of divergent selection on even smaller spatial scales, between stands in direct vicinity (<820 m) growing on slopes with different aspect within continuous forests. Differences in water availability on south- and north-facing slopes could have exerted divergent selection pressures and contributed to maintaining high genetic diversity on the local scale, even in a tree species with efficient gene flow. The increasing number of available genetic markers, also in non-model species, will increase the chances to detect such patterns on local scales. Understanding how microenvironmental heterogeneity shapes and maintains the functional genetic variation is relevant as this local scale variation is at the base of the population response to future climate. Especially with respect to climate change, the knowledge about genetic variation and processes that shape the genetic structure on different geographic scales are of utmost importance to develop suitable forest tree conservation and management strategies. Forest management, for instance, could be used to foster natural standing genetic variation and, hence, in situ evolution (Lefèvre et al., 2014), potentially making the use of assisted gene flow or migration unnecessary.

AUTHOR CONTRIBUTIONS

Santiago C. González-Martínez, Juli G. Pausas, Miguel Verdú and Myriam Heuertz designed the study; Katharina B. Budde performed the study, did most data analyses and wrote the first draft of the manuscript; Felix Gugerli, Christian Rellstab and Tom
Hanika provided scripts and support for data analyses; Santiago C. González-Martínez contributed to manuscript writing. All authors critically read and revised versions of the manuscript.

ACKNOWLEDGEMENTS
We thank Ana Hernández-Serrano, Fernando del Caño, Delphine Grivet, Mario Zabal-Aguirre, Natalia Vizcaino-Palomar, Enrique Sáez-Laguna, Diana Turrión and Miguel Navarro for help in the field and Carmen García-Barriga (INIA-CIFOR) for assistance in the laboratory. The present study was funded by projects from the Spanish Ministry of Science and Innovation (MICINN): VaMPiro (CGL2008-05289-C02-01/02), LinkTree (EU2008-03713, under the ERANet-BiodivERsA call) and TrEvol (CGL2012-39938-C02-00/01) and the IdEx Bordeaux (EcoGenPin, Chaires d’installation 2015). KBB acknowledges an exche grant of the European Science Foundation (ESF) within the framework of ‘Conservation Genomics: Amalgamation of Conservation Genetics and Ecological and Evolutionary Genomics’ (ConGenOmics). We thank the associate editor and two anonymous reviewers for their constructive and insightful comments.

CONFLICT OF INTEREST STATEMENT
The authors have no conflict of interest to declare.

REFERENCES


SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1: Phenotypes.

Figure S1.1: Boxplots of height (a) and diameter at breast height (DBH, b) for 150 Pinus pinaster trees in three stand pairs contrasting south- (S1, S2, S3) and north-facing slopes (N1, N2, N3).
Table S1.1: Two sample t-tests assessing differences in height and diameter at breast height (DBH) of *Pinus pinaster* trees in each pair of south- and north-facing slopes.

Appendix S2: Genetic structure.

Table S2.1: Pairwise differentiation between all six *Pinus pinaster* study stands (comprising south- [S1, S2, S3] and north-facing slopes [N1, N2, N3]) based on 5024 SNPs. All $F_{ST}$ values are significant with $p < 0.001$.

Figure S2.1: Principal component analysis (PCA) based on 5024 single nucleotide polymorphism markers of all *Pinus pinaster* samples from eastern Spain representing three south-facing (S) and three north-facing (N) slopes in a paired sampling design. Each stand is depicted with a different colour and the stand centroid is labelled with the site identifier.

Figure S2.2: Fine-scale spatial genetic structure (FSGS), plotted as average pairwise Loiselle kinship coefficient against the geographic distance between *Pinus pinaster* trees within pairs of south- and north-facing slopes. FSGS was strongest for pair N3/S3. $Sp$, intensity of the FSGS; ***, significance level of regression slope $p < 0.001$.

Appendix S3: $F_{ST}$ outlier detection in south- versus north-facing slopes.

Table S3.1: (provided as additional pdf) Results summary and annotation details of significant single nucleotide polymorphisms detected by the hierarchical models in Arlequin/arlecore and BayeScanH between *Pinus pinaster* stand pairs of south- and north-facing slopes.

Table S3.2: (provided as additional pdf) Annotation details of significant single nucleotide polymorphisms detected by the hierarchical models implemented in Arlequin, BayeScanH and the paired GEA between *Pinus pinaster* stand pairs of south- and north-facing slopes (retrieved from Plomion et al., 2016).

Figure S3.1: Plots showing differences in allele frequency between south- and north-facing slopes in all three *Pinus pinaster* stand pairs (left side: a, c, e, f, g, i) for the five candidate loci jointly identified by Arlequin/arlecore and BayeScanH. Pairwise $F_{ST}$ between stands is also shown (right side: b, d, f, h, j) for each of the five loci.

Appendix S4: Permutation tests.

Figure S4.1: Workflow to assess the robustness of our approach. We generated 1000 randomised data sets by permuting the assignment to south- and north-facing slopes between trees within each sample pair. The data sets were analysed using the same methods as the observed data set, namely $F_{ST}$ outlier detection using hierarchical island models implemented in Arlequin/arlecore and BayeScanH, and additionally consistent differences in allele frequency changes using the paired GEA script. The same parameters as for the observed data were used. The results were written to a standard relational database.

Table S4.1: Overview of significant outlier loci detected in the observed data and the empirical $p$-values based on the distribution of significant outlier loci detected in about 1000 randomised data sets, and the overlap of outlier loci detected by several approaches. These $p$-values can be understood as the probability that the number of significant outlier loci detected in a randomised data set is equal to or above the observed value.