Using Gradient Forest to predict climate response and adaptation in Cork Oak

Running title: Climate response and adaptation in Cork Oak

Mathieu Vanhove[1], mathieu.vanhove@gmail.com, orcid: 0000-0002-6090-8796
Francisco Pina-Martins[1], f.pinamartins@gmail.com, orcid: 0000-0003-1836-397X
Ana Cristina Coelho[2], acoelho@ualg.pt, orcid: 0000-0002-2919-5468
Cristina Branquinho[1] , cmbranquinho@fc.ul.pt, orcid: 0000-0001-8294-7924
Augusta Costa[3,4], augusta.costa@iniav.pt, orcid: 0000-0002-0400-8523
Dora Batista[1,5], dccastro@fc.ul.pt, orcid: 0000-0003-4031-8999
Adriana Silva[1], aprsilva@fc.ul.pt, orcid: 0000-0002-5147-835X
Paulo Sousa[1], sousapaulo16@gmail.com, 0000-0002-2333-6493
André Henriques[1], andrefrhenriques@gmail.com, 0000-0003-2800-046X
Isabel Marques[1,6], isabel.ic@gmail.com, 0000-0001-9788-4831
Bouchra Belkadi[7], bbelkadi@gmail.com, 0000-0002-1709-3042
L. Lacey Knowles[8], knowlesl@umich.edu, 0000-0002-6567-4853
Octávio S. Paulo*[1], octavio.paulo@fc.ul.pt, 0000-0001-5408-5212

[1] Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal.
[2] Centro de Electrónica, Optoelectronica e Telecomunicações, Universidade do Algarve, Campus de Gambelas, 8005-239 Faro, Portugal.
[3] Instituto Nacional de Investigação Agrária e Veterinária, Av. da República, Quinta do Marquês, 2780-157 Oeiras, Portugal.
[4] CENSE - Center for Environmental and Sustainability Research. NOVA University of Lisbon, Campus de Caparica, 2829-516 Caparica, Portugal.
[5] Linking Landscape, Environment, Agriculture and Food, Instituto Superior de Agronomia, Universidade de Lisboa,1349-017 Lisboa, Portugal.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/JEB.13765

[6] Forest Research Centre, Instituto Superior de Agronomia, Universidade de Lisboa,1349-017 Lisboa, Portugal.

[7] Faculty of Sciences of Rabat (FSR), LMBM, Avenue Ibn Batouta BP:1014, Rabat Agdal, Morocco.

[8] Department of Ecology and Evolutionary Biology, University of Michigan, 830 N. University Ave., Ann Arbor, MI 48109, USA

Acknowledgements

This work was supported by Project "Keep Pace: Selection of trees keeping pace with fast environmental changes, a science based approach for sustainable XXI century Oak forests", co-funded by the EU ERDF funds, within the Portugal 2020 Partnership Agreement and Programa Operacional Regional Algarve, and by national funds through Fundação para a Ciência e Tecnologia (FCT) Portugal (ALG-01-0145-FEDER-029263 / PTDC/ASP/SIL/29263/2017). Field and laboratory work, and bioinformatics platform were supported by grant numbers SOBREIRO/0036/2009 (under the framework of the Cork Oak ESTs Consortium) and grant PTDC/AGR-GPL/104966/2008. FPM was supported by grant PhD. SFRH/BD/51411/2011, under the PhD program of Biology and Ecology of Global Changes.

*Corresponding author:
Octávio S. Paulo
octavio.paulo@fc.ul.pt
Tel: 00 351 217500614 direct
Tel: 00 351 217500000 ext22359
Fax: 00 351 217500028

Authors Contribution:

MV conducted the data analysis with FPM and OSP, all authors participate in the writing up and reviewing of the manuscript. OSP and FPM design the original sampling design.

The authors have no conflict of interest to declare

Article type : Special Issue

Using Gradient Forest to predict climate response and adaptation in Cork Oak

Abstract

Climate change is impacting locally adapted species such as the keystone tree species cork oak (Quercus suber L.). Quantifying the importance of environmental variables in explaining the species distribution can help build resilient populations in restoration projects and design forest management strategies. Using landscape genomics, we investigated the population structure and ecological adaptation this tree species across the Mediterranean Basin. We applied genotyping by sequencing and derived 2,583 single nucleotide polymorphism markers genotyped from 81 individuals across 17 sites in the studied region. We implemented an approach based on the nearest neighbor haplotype "coancestry" and uncovered a weak population structure along an east-west climatic gradient across the Mediterranean region. We identified genomic regions potentially involved in local adaptation and predicted differences in the genetic composition across the landscape under current and future climates. Variants associated with temperature and precipitation variables were detected and we applied a nonlinear multivariate association method, gradient forest, to project these gene–environment relationships across space. The model allowed the identification of geographic areas within the western Mediterranean region most sensitive to climate change: southwestern Iberia and northern Morocco. Our findings provide a preliminary assessment toward a potential management strategy for the conservation of cork oak in the Mediterranean Basin.

Key words: climate change, local adaptation, landscape genomics, Gradient Forest, Quercus suber L.

Introduction

The adverse effects of climate change on European forests will largely depend on the capacity of trees to tolerate temperature and precipitation changes. In the Mediterranean Basin, higher temperatures and increased aridity are predicted to have serious consequences on species composition (Dukes et al., 2005; Petit et al., 2005) with anticipated latitude and elevation shifts in response to these changes (Benito et al., 2014). Local adaptation will determine the fate of these species, providing insight into the magnitude and location of the potential effects of climate change, knowledge that will help to mitigate these future effects. Scientists have raised concerns regarding the ability of trees to cope with climate change (Lindner et al., 2010; Alberto et al., 2013; Sork et al., 2013). Variations are observed in trees phenotypic (Príncipe et al., 2019) and genetic features across the landscape and their response will depend on the genetic architecture of traits associated with the response to current climate conditions. However, the long lifespan of tree species questions their ability to mitigate the effect of a changing climate (Hughes et al., 2008; Kremer et al., 2014). Finding the most appropriate way to protect trees is central as their future may directly impact the global carbon cycle and the rate of climate change due to the importance of forest ecosystem (Sork et al., 2013).

Forest tree populations are the result of natural demographic and selective processes where gene flow and natural selection are shaping spatial genetic patterns and driving phenotypic variations (Sork, 2016). Using common garden experiments, several studies emphasized the importance of species phenology in mediating adaptation to climate (Savolainen et al., 2007; Alberto et al., 2013). They highlighted variations in tree resistance to cold, drought or the ability of trees to grow under various conditions and aimed to improve forest management practices (Bower & Aitken, 2008; Aitken & Bemmels, 2016). However, these types of experiments are long and costly and biologists turned to genetics to gain knowledge about spatial patterns of adaptation. The development of landscape genomics has shown that gene-environment correlation can mirror phenotypic correlations with environmental gradients (Coop et al., 2010).

In naturally evolving environments, gene flow occurs more often within populations living in close distance to each other. Genomic data have been able to elucidate patterns of Isolation

by Distance (IBD) and to examine spatial relationships across the landscape. With the advances in high-throughput sequencing, researchers have started to examine patterns of Isolation by Environment (IBE). The field of landscape genomics emerged as a framework to study interactions between adaptive processes in natural populations and environmental heterogeneity (Schoville et al., 2011; Sork et al., 2013; Neale et al., 2017). Thousands of genetic markers can be investigated in the light of georeferenced samples to gain insights on evolutionary processes using approaches referred as Environmental Association Analysis (EAA) (Rellstab et al., 2015; Ćalić et al., 2016) or genetic–environment association (GEA) (Lotterhos & Whitlock, 2015). By correlating genomic data and environmental variables, it became possible to identify environmental and genomic factors driving local adaptation (Hoban et al., 2016).

EAA approaches allow to better characterize target species and can help toward a better management of seeds for revegetation purposes. Forest restoration has traditionally been restricted to the use of local seeds (Broadhurst et al., 2008). However, reduced seed sourcing can lead to a limited gene pool which may result in inbreeding depression for future generations especially when ecosystems are under stress or when population size is shrinking. Using integrative population genetic and ecological modelling, it appears now possible to guide seed choices in a process known as Assisted Gene Flow (Aitken & Whitlock, 2013). Obtaining seeds from other geographical areas may improve forest management strategies by increasing genomic and phenotypic diversity (Williams et al., 2014; Supple et al., 2018). These practices may help current populations mitigate the adverse effects of climate change (Prober et al., 2015).

The first approaches used to model species adaptation to its environment were based on species distribution models (SDM) which rely mostly on species (or subspecies) presence data. These models were not fit to account for the intraspecific variation due to local adaptation (Fitzpatrick & Keller, 2015). In their paper, Fitzpatrick and Keller (2015) demonstrated how to apply community-level modelling approaches to map turnover of allele frequencies along environmental gradients. They used Gradient Forest (GF) (Ellis et al., 2012), a regression tree-based method and Generalized Dissimilarity Modelling (GDM) (Ferrier et al., 2007), a distance-based modelling approach. These methods have promising applications as they can quantify the role of spatial and environmental variables in structuring genetic variations which allow to describe non-linear changes along environmental gradients, and thus overcome the limitation of traditional genotype-environment associations. Several studies successfully applied these methods on tree species (Gugger et al., 2018; Martins et al.,

2018; Supple et al., 2018; Ingvarsson & Bernhardsson, 2020), being able to describe association between genetic structure and environmental variables. For instance, Gugger et al., (2018) generated a dataset of over 11,000 single nucleotide polymorphisms (SNPs) from 311 Acacia koa trees in Hawaii. They predicted using future climatic scenarios that changes in rainfall patterns may result in "genetic offset" (sensu Fitzpatrick & Keller, 2015) where trees may no longer be genetically adapted to fit their environment. These methods appear as a useful tool to guide reforestation strategies through the selection of tree populations better equipped to face climate change.

In the present study, we aimed at investigating the potential adaptive response of cork oak (Quercus suber L.) to climate change. The species is present throughout the western Mediterranean region, where it holds high economic importance and a vast ecological significance in sustaining terrestrial biodiversity and other regulating ecosystem services (Benito Garzón et al., 2008). The current distribution of cork oak ranges from the Atlantic coasts of North Africa and the Iberian Peninsula to the southern regions of Italy (Fig. 1). The evolutionary origin of the cork oak is thought to have been occurred in the western Mediterranean region, where the species was able to persist throughout climate oscillations of the Quaternary period (2.6 Ma) (Magri et al., 2007). In the Iberian Peninsula and Morocco, palynological evidence hints at a long-term persistence of cork oak dating from the last glacial period (Magri et al., 2007). Several study cases have attempted to model the species distribution across various timelines using Ecological Niche Modelling (ENM) (Vessella et al., 2015, 2017; Correia et al., 2018). However, these models assume the uniformity of climate response below the species level and do not account for the multidimensionality of genomic variation.

Here, we seek to understand the ecological drivers of adaptation in cork oak by reanalyzing a previously published dataset which revealed lack of significant population structure (Pina-Martins et al., 2019). We expanded previous investigations (Costa et al., 2011; Modesto et al., 2014; Pina-Martins et al., 2019) by implementing a population structure analysis based on the nearest neighbor haplotype "coancestry" and by combining population genomics and gene-environment associations to further study the local adaptation of these tree keystone species. Moreover we implemented an innovative nonlinear multivariate approach, gradient forest, to identify areas at risk of climate maladaptation and predict genetic changes required to keep pace with a changing environment. This identification of the spatial regions where gene–environment relationships will be most disrupted, the 'genetic offset', is of critical importance for a knowledge based adaptive management of this economic important species.

Methods

Sampling and Genomic data processing

The individual samples used in the study were previously collected and genotyped by us using 95 unrelated cork oak individuals from 17 locations across the species distribution (Table S1) (Pina-Martins et al., 2019). Briefly, DNA was extracted from grounded leaves using "innuPREP Plant DNA" kit (Analytik Jena AG) and the DNA samples were outsourced to the "Genomic Diversity Facility" at the University of Cornell for "Genotyping by sequencing" (GBS) (Elshire et al., 2011). DNA was digested using the EcoT22I restriction enzyme and sequencing was performed on an Illumina HiSeq 2000 platform. Raw reads data in FASTQ format were processed using STACKS 2.41 (Catchen et al., 2013). Samples with 40% of missing data were removed. The "populations" parameters were adjusted to retained one SNP per "stack", with a minimum percentage of individuals in and across populations of 0.8 required for a locus to be processed (populations –r 0.8 –R 0.8 –min-maf 0.01 --write_single_snp). Previous work has shown the lack of population structure in cork oak and since our populations contained few individuals (five individuals in average), various minimum allele frequencies were tested (0.0, 0.01 and 0.03) to observe how they affect summary statistics intending to retain some private alleles in each population.

Population Structure and Summary Statistics

For a pilot approach, we ran the Stacks pipeline with the same parameters but varying the MAF between 0.00, 0.01 and 0.03 (Table S3). Based on this test, a minor allele frequency <0.01 was used for subsequent analyses. For the analysis of population structure and summary statistics we use two datasets: (a) the adaptive-SNP dataset and (b) the "neutral" dataset. The latter exclude SNPs that were identified as significant climate-associated SNPs

by the Environmental Association Analysis (see below), to assess the effect of putative "non neutral" SNPs on the overall pattern of populations differentiation and structure. The Bulgarian population is known to be introduced, probably from Iberia, consequently those samples were included in the population structure analysis but excluded from the Mantel test and any further analyses.

To explore gene flow and pattern of isolation by distance on the genetic structure of cork oak, we calculated pairwise F_{ST} in R 3.3.0 (R Core Team, 2016) using the "hierfstat" package (Goudet, 2005). Mean nucleotide diversity (π) and expected heterozygosity (H) were estimated for each population using STACKS.

To analyze the population structure, we used fineRADStructure v.0.3.1 (Malinsky et al., 2018).FineRADstructure is a model-based Bayesian clustering approach that groups together individuals with high levels of shared coancestry. The high resolution population structure inference is based on this coancestry matrix, which is used to cluster individuals according to the similarity of their RAD haplotypes. This improved capacity is the result of combination of the new RADpainter with fineSTRUCUTRE. The former is simple method of finding the closest relatives for each allele and summed up into the coancestry matrix. Each individual is considered to either being a donor or a recipient of DNA fragments. The coancestry matrix then records the inferred recombination events between each donor and recipient prior to coalescing with another genome. RADpainter was designed to take full advantage of RAD data sets (see Malinsky et al. 2018 for further details on the calculations of the coancestry matrix). The latter is a Markov chain Monte Carlo (MCMC) clustering algorithm. The optimal population structure is obtained by exploring the space of population configurations and a proposed population configuration is accepted with a probability derived from the ratio of the likelihood with the previous configuration, a likelihood that in turn depends on the terms of the scaled coancestry matrix. Based on the final output, we can infer the number of clusters, quantify ancestry sources in each group, and also explore relationships between groups (Lawson et al., 2012; Malinsky et al., 2018).Additionally, population structure was visualized on a principal component analysis (PCA) with the "adegenet" package (Jombart, 2008).

Mantel tests assess the association between genetic and geographic distance and detect spatial autocorrelation (Mantel, 1967). Genetic variation was calculated as the Bray-Curtis distances between loci. The geographic distances consisted of the Euclidean distances between sampling localities. Mantel tests were performed using the vegan package (Goslee & Urban, 2007) in R with 10,000 permutations.

Genetic–Environment Association and Outlier Detection

Outlier detection was performed in BAYESCAN v2.1 (Foll & Gaggiotti, 2008) using 100,000 iterations with a burning of 50,000 steps, and a thinning interval size of 10. The latter has been recognized as the most effective population differentiation method (De Mita et al., 2013; Lotterhos & Whitlock, 2014). To minimize the chance of false positives due to multiple testing, we applied a False Discovery Rate (FDR) criterion of 0.05 (Benjamini & Hochberg, 1995). Q-values were calculated in R 3.3.0 (R Core Team, 2016) using the "qvalue" package (Storey et al., 2015). Since cork oak might exhibit a weak pattern of isolation by distance, this F_{ST} outlier analysis provides credible candidate SNPs resulting from spatially divergent selection pressures.

Redundancy analysis (RDA) was performed in R using the package "vegan" (Oksanen et al., 2013). RDA is an ordination method which examines the variations of how a set of variables is explained by another set. In this study, RDA is used to investigate how much of the genetic variation is attributable to either climate or spatial distances, versus how much can be explained by their joint effect. Gene ontology (GO) terms were investigated and summarized using Blast2GO (Conesa & Götz, 2008) to identify genes harboring putatively selected SNPs that might play a role in adaptation.

Additionally, we detected local adaptation by testing for associations between SNPs allelic frequencies changes and climatic gradients. We tested the nineteen Bioclimatic (BIO) WordClim variables V2.0 (Fick & Hijmans, 2017) at 30 arc-seconds (~1 km) resolution using a Latent Factor Mixed Model implemented in LFMM 1.3 (Frichot et al., 2013). Information for each sample was extracted in R (v 3.1.1) using the "raster" (Hijmans & van Etten, 2012) and "dismo" (Hijmans et al., 2012) packages. LFMM is a Bayesian approach used to detect selection in landscape genomics. Briefly, the method investigates the influence of population structure on allele frequencies by introducing unobserved variables as latent factors (K) (Stucki et al., 2017) to detect signatures of local adaptation while accounting for population structure. K-values ranging from one to eight, and three independent repetitions for each K, were run using the Bayesian clustering method described in the LFMM manual. The method resulted in choosing a K = 2 (Fig. S1). We performed three independent LFMM runs using 10,000 iterations and burn-in of 5,000 using the LEA package in R. The |z|-scores were averaged to strengthen the genetic-environment association and a FDR of 5% was used to

uncover significant SNPs (Frichot & François, 2015). Adjusted p-values (q) were estimated using a genomic inflation factor (λ) procedure (Devlin & Roeder, 1999). A visual examination of histograms of adjusted p-values was performed to assess the confounding effect of the population structure as recommended in the LFMM manual (Fig. S2).

Gradient forest analysis

We modelled current and future patterns of genetic variation using a gradient forest (GF) approach. GF analysis was implemented using "gradientForest" (Breiman, 2001) in R. The method is a nonparametric, machine-learning regression tree approach which allows mapping patterns of turnover in biological composition using nonlinear functions of an environmental gradient. We conducted an initial GF analysis on the nineteen Bioclimatic (BIO) WordClim variables to assess the relative importance of each predicator variable using weighted R^2 values (split importance - Ellis et al., (2012)) (Fig. S3). Split importance measures the amount of variation explained, appearing high along the gradient where allelic frequency change is large. After the initial model, we removed eleven variables (BIO1, BIO2, BIO3, BIO4, BIO5, BIO10 , BIO11, BIO15, BIO16, BIO17 and BIO18) due to high correlation (Pearson's correlation coefficient $|r| > 0.8$) and lower explanatory power than the other remaining variables (Table S2). The GF turnover function describes the magnitude of turnover in genetic distance along the gradient while considering all the other variables constant (Fitzpatrick & Keller, 2015). As a result, in our final modeling approach, we tested for environmental correlation using eight variables, four temperature variables (BIO6, minimum temperature of coldest month; BIO7, temperature annual range; BIO8, mean temperature in the wettest quarter; BIO9, mean temperature in the driest quarter) and four precipitation variables (BIO12, annual precipitation; BIO13, precipitation of the wettest month, BIO14, precipitation of the driest month; BIO19 precipitation of the coldest quarter). These variables were tested for signature of local adaptation. Spatial variables were defined using principal coordinates of neighborhood matrices (PCNMs) or Moran's eigenvector maps (MEM) based on the geographic coordinates (Dray et al., 2006) using the pcnm function in "vegan". We modeled climatic and spatial drivers of genomic variation using GF methods on two distinct SNP sets: (a) the neutral-SNP dataset and (b) the significant climate-associated SNPs. To visualize the GF results, the transformed environmental variables were reduced into multivariate synthetic variables using principal component analysis (PCA) and the first three principal components (PCs) were assigned to a red-green-blue color palette. A Procrustes superimposition (Gower,

1971) on the PCAs was applied to compare mapped genetic composition for the neutral-SNP and the adaptive SNP datasets as described in Martins et al. (2018). The Procrustes rotation compared the PCAs generated in the two models and estimated the differences between them. The Procrustes residuals represent the absolute distance in genetic composition between SNPs datasets at each location.

Finally, we used GF to estimate the genetic offset under future climate. To estimate vulnerability under climate change, Bioclim variables for future climate were obtained for the year 2070 under the RCP emission scenario 8.5 using the general circulation model (GCM): Community Climate System Model version 4 (CCSM4) (Gent et al., 2011). The results from the GF analyses were used to predict genetic change ("genetic offset"). The genetic offset is a predictive measure to identify the spatial regions where gene-environment relationships will be the most disrupted between current and future climates (Fitzpatrick & Keller, 2015). For each grid cell, euclidian distances between current and future genetic composition were calculated and serves as the metric for genetic offset (Ellis et al., 2012). The future predictions inform on how much genetic composition across the landscape needs to change so that current gene–environment relationships are maintained.

Results

Genetic diversity and Population Structure

The total dataset comprised 81 samples as 15 samples were removed due to missing data and 2,583 SNPs were uncovered, with 80% representation in both samples and populations. Depth of coverage for each SNPs averaged 30.8 (SD=13.37). The second dataset ended up with 2,318 "neutral" SNPs after removing the 265 climate-associated SNPs by the Environmental Association Analysis (Table S3).

The average genetic differentiation across loci and populations was F_{ST} = 0.044 (neutral dataset F_{ST} = 0.044). The nucleotide diversity (π) averaged 0.180 and was similar across the 17 populations ranging from 0.160 to 0.206, while the average gene diversity was H_E = 0.160, SD = 0.0037 (Table S4). Pairwise F_{ST} ranged from -0.0023 between Tuscany and Algeria to

0.0977 between the southwest and southeast of France (Table S5). Despite the low overall population differentiation, the fineRADstructure analysis revealed signs of population structure and three defined groups of populations were identified for both the full dataset and the "neutral" (Fig. 2 and S4, Table 1). Tree distribution followed an east-west gradient. Group 1 contained 29 western samples (93.5%, F_{ST} = 0.0212, neutral dataset) with only two samples coming from the East (Var21 and Pug5) (F_{ST} = 0.0379 neutral dataset). Group 2 contained samples from Corsica (n=3) and the Landes region in France (n=4) as well as two introduced Bulgarian samples. Significant genetic exchange was observed among the third group in comparison with the other two. This group was mainly composed of individuals from eastern sampled populations (97.4%, one outlier: Cat3), namely: Tunisia, Algeria, Tuscany, Sicily, Kenitra and Apulia which accounted for 70.1% (n=86) of the private alleles uncovered (Table S4). The Apulia population specifically displayed the highest number of private alleles (n=30). These results were corroborated by the PCA even though only a weak pattern of geographic structure could be observed (Fig. S5a and b). The nucleotide diversity, measured for the full dataset, but with similar results for the "neutral" dataset, seems to be higher among the eastern group than among its western counterpart (π_{EAST} = 0.188; π_{WEST} =0.173, $p < 0.001$), with lower population differentiation in the west (West: F_{ST} = 0.0212, East: F_{ST} = 0.0379, $p < 0.001$) and with a higher number of private alleles (n_{EAST} = 85; n_{WEST} = 20) in the eastern group. Outlier detection analysis for the full dataset revealed 11 outliers with a high F_{ST} indicating divergent selection (Fig. 3) and no outlier shows a signal of balancing or purifying selection (low F_{ST}). Gene ontology associated with one of these SNPs (SNP_14085) identified a gene encoding for a squalene monooxygenase-like protein.

Genetic–Environment Association

The environment seemed to play an important role in the current distribution of cork oak. Redundancy analysis (RDA) found that 72% of cork oak genetic distribution could be explained purely by climate and 18% by spatial distance. The samples showed signs of isolation by distance (Mantel r statistic = 0.282, $p < 0.01$). Landscape genomics analyses revealed 265 SNPs that were significantly associated with the climatic variables tested. A total of 249 SNPs was associated with the four temperature variables, whereas only 45 SNPs were linked to the precipitation variables. Thirty-four SNPs were associated with multiple environmental variables including SNP_7403 and SNP_35704 which were associated with

five variables (BIO07, BIO09, BIO12, BIO13, BIO14 and BIO19) (Table S6). Functional annotation revealed that 11 loci containing an associated SNP matched genes with known functions, including a gene encoding for a DNA-binding transcriptional regulator DhaR for SNP_187039, which was associated with three temperature variables (BIO6, BIO7) and two precipitation variables (BIO13, BIO19) (Table 1). Also, SNP_6044 was correlated with three precipitation variables (BIO12, BIO13, BIO19), and the harboring loci is an ortholog of a gene encoding for a mannosylglycerate hydrolase. Additionally, two SNPs included in genes fragments that are part of two DNA polymerase processivity factors (SNP_143637; SNP_226326) were found to be associated with the mean temperature of the wettest quarter (BIO8).

Gradient Forest analysis and Genetic offset under future climate

The GF model was run initially on the neutral dataset of 2,583 SNPs and then on the 265 adaptive SNPs dataset found in the LFMM analysis. The GF models that explained better the variation was the adaptive SNPs model (mean R^2 = 14.0%) compared to the neutral-SNPs model (mean R^2 = 12.4%). The spatial location appeared to be the strongest predicators (Fig. 4, Fig. S6). After summing variables importance, all PCNMs variables explained 58% of the model variation and environmental variables explained 42% in both neutral-SNPs and adaptive-SNPs models. These results indicate that spatial variables had the strongest influence on the turnover in allele frequency across the distribution of cork oak. The contribution of environmental variables differed between the two models in which temperature annual range (BIO7) and mean temperature of the driest month (BIO9) appeared with highest R^2 weighted importance among climatic variables in the adaptive SNPs model. When inspecting the cumulative importance for the neutral-SNPs and adaptive SNPs models (Fig. 4), we observed that in the adaptive model, the turnover of allele frequencies occurs much earlier for the variables temperature annual range (BIO7) and precipitation of the driest month (BIO14). For the temperature annual range variable, a major change in allelic frequencies is observed at 21°C and a similar change can be observed for areas where precipitation in the driest month are below 10 mm.

The two GF models gave similar results when mapped onto the ecological niche of cork oak (Fig. S6). Briefly, the genetic importance values resulting from the GF models were transformed into multivariable synthetic variables using PCA and these predictions converted to a red-green-blue color scale using the first three axes of the PCA (see Methods). Different patterns of allele frequencies turnover were observed (Fig. 5a) with potentially unmapped levels of genetic diversity present in the northeast of the Iberian Peninsula and in the Landes region of France. The procrustes superimposition was performed to identify regions where selection could be stronger in order to prioritize areas to be sampled in future studies. The procrustes residuals measure the absolute distance in genetic composition between the neutral-SNP and adaptive SNP datasets. Warmer colors on Figure 5b represent the largest difference in adaptive variation compared to the overall genetic composition. These areas include the Portuguese and Moroccan coasts, Catalonia and northern Italy.

The "genetic offset" allows inferring how much the genetic composition across the landscape is required to change in order to preserve the current gene-environment relationships (see Methods). Most of the regions seem to be affected by this changing environment especially Southwest Portugal and Northern Morocco (Fig. 6). Portugal hinterland, northern Italy, Corsica or the Var regions appear as areas where the genetic offset is minimal.

Discussion

Genetic diversity and Population Structure

The present study uncovered 2,583 SNPs, identified some levels of local adaptation and cork oak individuals could be assigned to three groups of populations (Fig. 1). The previous study used a stringent MAF choice of 0.03 yielding a dataset of 1,996 SNPs and marginally differences between K1 and K2 (Pina-Martins et al., 2019). Given the relative low sample size of each population (~ 4.63), we retained a minimum allele frequency (MAF=0.01) which allowed to detect a reasonable number of private alleles (mean = 7.1 per population, Table S3). A MAF value of 0.03 led to only one private allele. However, varying the MAF value did

not result in changes in population structure or in the choice of an inflation factor when assessing population structure for LFMM analysis (K = 2). The results from the "neutral" dataset for F_{ST} and Mantel test were similar to the ones of the full dataset.

The summary statistics obtained from the Stacks pipeline differed from previous analyses performed using the ipyrad pipeline. Variations in summary statistics and results from independent null expectations (IBD or expected transition-to-transversion ratio T_S/T_V) among RAD-seq pipelines have been previously reported (Shafer et al., 2017). In the current analysis, a pattern of isolation by distance was observed (mantel test = 0.281, $P < 0.001$) reinforcing the role of local adaptation in shaping the structure of cork oak.

The overall F_{ST} value of 0.0444 appeared similar to the one in previous study (F_{ST} = 0.0541) (Pina-Martins et al., 2019), with high genetic diversity (H_E = 0.160) which seems to indicate some overall degree of differentiation with some historical gene flow.

To further investigate population structure, a fineRADstructure analysis was performed. The software offers a high resolution based on the nearest-neighbor relationships (first coalescence) between haplotypes and allows the identification of substructure within populations (Malinsky et al., 2018). fineRADstructure inferred the presence of three clades which displayed an east-west pattern. In comparison with other approaches used for population inference, fineRADstructure offers an improved insight on the Quercus suber L. population structure. When using STRUCTURE-like approaches, the choice of model complexity (K = 2 – Fig S1) is based on the rate of decrease in the value the Bayesian Information Criterion (BIC). This approach explores clustering between sampling locations but remains limited to the choice of model complexity. On the other hand, fineRADstructure implements a model-based Bayesian clustering approach that groups together individuals using an inferred coancestry matrix based on patterns of haplotype similarity. This new efficient way of capturing information on population structure, was developed for RADseq data, a genome-wide dense markers and proven to be robust to missing RAD alleles. These combinations of characteristics make fineRADstructure particularly suitable for our type of data and consequently could better explain the resolution obtained when compared with methods such as STRUCTURE. Similar results to ours, where fineRADstructure provided a better resolution than the optimal number of clusters based on STRUCTURE-like methods, and coinciding with suboptimal number of STRUCTURE-like clusters, are starting to appear in the literature (Dincă et al., 2019; Balao et al., 2020). The structure pattern evidenced here had already been identified, to some extent in the original STRUCTURE-like analysis, but concluded it was an incomplete separation between eastern and western groups (Pina-Martins

et al., 2019). The use of Principal Component Analysis (Fig. S5) was also unable to clearly partition the described genetic diversity although substantial overlap was observed with fineRADstructure. In both cases, Quercus suber in the western part of the Mediterranean (Group 1 – Fig 3) appeared less genetically diverse than its Eastern counterpart (Group 3). However, fineRADstructure was able to capture some signal of local structure for a third population in Corsica and the Landes (Group 2). In terms of biological insights, fineRADstructure offers a major improvement over other methods by providing evidence into what appears to be a more genetically diverse eastern population. This result highlights that the western population probably resulted from postglacial recolonization history (Lumaret et al., 2005; Magri et al., 2007). Previous studies using chloroplast DNA identified five different haplotypes within the species, two haplotypes in the west and three in the east (Magri et al., 2007). The authors argued that the presence of cork oak in the Iberian peninsula, consistent with fossil records, might have an early Cenozoic origin (De Carvalho, 1958; Losa Quintana, 1978). The present work suggests a complicated network of relationships within eastern locations suggesting that the species might have originated in this region. This hypothesis appears coherent with previous studies (Bellarosa et al., 2005; Lumaret et al., 2005) which stipulated that cork oak originated in the eastern part of its current range before expanding westward in the Mediterranean. Further testing using different species-models and approximate Bayesian computation (ABC) could contribute to understand historical range shifts (Bemmels et al., 2016).

In addition to isolation by distance, the study revealed higher levels of nucleotide diversity among the eastern group of populations with a higher number of private alleles and high pairwise F_{ST}. The average among western populations (F_{ST} = 0.0212) was lower than in the east (F_{ST} = 0.0379) indicating a higher gene flow in the west. A third clade (group 2 in Fig. 1 and Fig 2) emerged composed of samples from Corsica and the Landes region which appeared as the most differentiated of the studied populations (F_{ST} = 0.061). However additional sampling efforts are required to draw any definitive conclusions on the evolutionary history of the species. Eleven SNPs were detected as outliers, representing a credible set of candidate loci under divergent selection. However, demographic events might lead to false-positive especially when using a fractional genome sequencing strategy (De Mita et al., 2013; Lotterhos & Whitlock, 2014), limiting which inferences can be drawn from this small number of loci.

By sequencing additional populations using genome-wide markers, we are likely to unveil unmapped levels of genomic diversity and to uncover additional clades in the eastern part of

the range as well as providing insights onto the role of anthropogenic activities in the distribution of this tree species within Spain and Portugal. Cork oak has been widely cultivated for the production of cork in the Iberian Peninsula and this exploitation might explain its current distribution (Carrión et al., 2000), and to some extent, its genetic background in the region.

Genetic–Environment Association

Redundancy analysis showed that climate played a significant role in the distribution of cork oak (72%). To investigate this gene-environment relationship, we deployed a Genetic Environment Association (GEA) method which aimed to characterize empirical patterns of adaptation (Forester et al., 2016). LFMM provides a way to investigate signatures of local adaptation by the detection of high degrees of correlation between polymorphism and environmental variables. This method has proven to be a robust approach (Stucki et al., 2017) even if demographic factors such as IBD tend to bias the analysis (De Mita et al., 2013; Hoban et al., 2016). In the present study, the majority of SNPs associated with climatic variables were correlated with temperature variables and no overlap was observed between the LFMM approach and the outlier detection method.

Cork oak belongs to evergreen oaks and is strictly adapted to the Mediterranean Basin which is very selective in terms of temperature and rainfall (Aronson et al., 2009). In SNPs-association studies, the role of temperature over precipitation has been previously highlighted (Cox et al., 2011; De Kort et al., 2014; Martins et al., 2018) and the present analyses emphasize the role of temperature in shaping cork oak distribution. Similarly to a previous study where only 4.4% of the queried sequences could match a region in the genome (Pina-Martins et al., 2019), only 11 genes (4.15%) were annotated with a protein prediction (Table 1). Of these markers, SNP_12029 was associated with two temperature variables (BIO6, BIO7) and the respective locus match a gene encoding for a quinolinate synthase. The latter is known to be involved in the onset of leaf senescence (Schippers et al., 2008). SNP_6044 was associated with three precipitation variables and the respective gene annotated encodes for mannosylglycerate (MG) hydrolase. MG accumulates in thermophilic bacteria in response to salt or thermal stress. In plants, the ability to hydrolyze MG is important for the plant physiology and hydrolysis (Nobre et al., 2013) and this polymorphism might represent a signal of local adaptation as a response to drought. However, results from environmental associations models must be taken carefully as the impact of population structure on the

findings remains debatable (De Villemereuil et al., 2014; Lotterhos & Whitlock, 2014). Integrating information on phenotypic variations in future studies might be beneficial to identify putative candidate genes.

Gradient Forest analysis and genetic offset under future climate

Geographic variables represented the strongest driver of turnover in allele frequencies over the landscape (Fig. 4a). The adaptive SNPs GF model revealed that temperature annual range (BIO7) was the most important variable explaining changeover of allele frequencies across cork oak distribution (Fig. 4b). The important role of spatial variables might be due to spatial autocorrelation due to isolation by distance but it can also be suggestive of important unmeasured environmental predictors (Martins et al., 2018). The cumulative function showed a steep turnover in allele frequencies occurring in the adaptive SNPs set at 21°C. The turnover could also be observed for the mean temperature in the driest month (BIO9) indicating an adaptation to higher temperatures. In the neutral SNPs model, a steep change in allele frequencies occurred between 20 to 30 mm of rain in the driest month (BIO14), whereas these changeovers occurred around 5 to 10 mm in the adaptive SNP model. These findings might be revealing of a genomic adaptation to drought.

Mapped projections of the GF results led to potential unmapped levels of cork oak genetic diversity present in northern Portugal (Fig. 5) while similarities are expected between the Landes region and Catalonia. The Procrustes superimposition identified regions where adaptation is expected to be more intense. The Procrustes residuals (absolute distance in genetic composition between “full SNP” and “adaptive SNP” datasets for each point location) were high in Catalonia and in Baetic region meaning that these populations are potentially adapting to a changing environment.

From future gene-environment projections, northern Morocco, southwest Portugal and northern Algeria appear as the regions which are expected to experience the largest disruptions. Local populations will require a significant genetic offset to persist in the region. However, the Gradient Forest analysis revealed that trees are expected to find a favorable region in the hinterland of the Iberian Peninsula. These findings corroborate the forecasts obtained from ecological niche modelling performed on the cork Oak (Vessella et al., 2017). In this study, for an ENM under the hardest scenario (CSSM4 RCP 8.5), the model predicted only 30-50% of suitability of its current ecological niche and 16% at the end of the century. It is worth noting that the actual evolutionary response of these populations to climate change

will be more complex than these projections as adaptation is the result of multiple evolutionary processes such as migration, mutation, recombination and the species effective population size (Fitzpatrick & Keller, 2015).

Management implications

The study demonstrates a weak population structure of cork oak populations along an east-west gradient. Environmental association analyses revealed that temperature was more frequently associated with polymorphisms than precipitation. Temperature annual range appeared as the strongest environmental variable shaping genetic variation within cork oak ecological niche. Moreover, the study revealed vulnerable areas of the species distribution that are at risk of climate change (Fig. 6). Our results are in accordance to the ones reported in a previous study based on ENM analyses which revealed, that temperature and precipitation variables are important in the distribution of cork oak (Vessella et al., 2017). However, making use of genetic data, the GF analysis was able to identify the drivers of genomic variation within the species. This nonlinear, multivariate environmental association method may help guide seed selection by identifying the suitable provenance of seeds which would respond more adequately to future climates (Gugger et al., 2018).

In conclusion, the present study provides compelling evidence that a large area of the species distribution in the Mediterranean Basin will experience drastic changes which will require strong adaptation from the local populations. Gradient Forest along with other approaches appear as a useful tool to develop forest management strategies at a faster pace and cheaper costs than traditional approaches (Fitzpatrick & Keller, 2015; Rellstab et al., 2016; Bernatchez et al., 2019). Our results reveal the potential of landscape genomics to identify regions which could benefit from Assisted Gene Flow (Aitken & Whitlock, 2013) such as Southwest Portugal, Baetic region and Northern Morocco but additional data is required as Assisted Gene Flow must be performed with caution (Aitken & Bemmels, 2016). The study allowed to identify areas within Quercus suber's distribution which are most sensitive to climate change. Additional sampling along two parallel transects from northern Spain to Southern Portugal and Morocco is projected to further explore the genetic diversity of cork oak and its adaptation to future climate.

References

Aitken, S.N. & Bemmels, J.B. 2016. Time to get moving: Assisted gene flow of forest trees. Evol. Appl. **9**: 271–290.

Aitken, S.N. & Whitlock, M.C. 2013. Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. Annu. Rev. Ecol. Evol. Syst. **44**: 367–388.

Alberto, F.J., Aitken, S.N., Alía, R., González-Martínez, S.C., Hänninen, H., Kremer, A., et al. 2013. Potential for evolutionary responses to climate change - evidence from tree populations. Glob. Chang. Biol. **19**: 1645–1661.

Aronson, J., Pereira, J.S. & Pausas, J.G. 2009. Cork oak woodlands on the edge: ecology, adaptive management, and restoration. Island Press, Washington.

Balao, F., Lorenzo, M.T., Sánchez-Robles, J.M., Paun, O., García-Castaño, J.L. & Terrab, A. 2020. Early diversification and permeable species boundaries in the Mediterranean firs. Ann. Bot. **125**: 495–507.

Bellarosa, R., Simeone, M.C., Papini, A. & Schirone, B. 2005. Utility of ITS sequence data for phylogenetic reconstruction of Italian Quercus spp. Mol. Phylogenet. Evol. **34**: 355–370.

Bemmels, J.B., Title, P.O., Ortego, J. & Knowles, L.L. 2016. Tests of species-specific models reveal the importance of drought in postglacial range shifts of a Mediterranean-climate tree: insights from integrative distributional, demographic and coalescent modelling and ABC model selection. Mol. Ecol. **25**: 4889–906.

Benito, B.M., Lorite, J., Pérez-Pérez, R., Gómez-Aparicio, L. & Peñas, J. 2014. Forecasting plant range collapse in a mediterranean hotspot: when dispersal uncertainties matter. Divers. Distrib. **20**: 72–83.

Benito Garzón, M., Sánchez de Dios, R. & Sainz Ollero, H. 2008. Effects of climate change on the distribution of Iberian tree species. Appl. Veg. Sci. **11**: 169–178.

Benjamini, Y. & Hochberg, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B **57**: 289–300.

Bernatchez, S., Xuereb, A., Laporte, M., Benestan, L., Steeves, R., Laflamme, M., et al. 2019. Seascape genomics of eastern oyster (Crassostrea virginica) along the Atlantic coast of Canada. Evol. Appl. **12**: 587–609.

Bower, A.D. & Aitken, S.N. 2008. Ecological genetics and seed transfer guidelines for Pinus albicaulis (Pinaceae). Am. J. Bot. **95**: 66–76.

Breiman, L. 2001. Random forests. Mach. Learn. **45**: 5–32.

Broadhurst, L.M., Lowe, A., Coates, D.J., Cunningham, S.A., McDonald, M., Vesk, P.A., et al. 2008. Seed supply for broadscale restoration: Maximizing evolutionary potential. Evol. Appl. **1**: 587–597.

Ćalić, I., Bussotti, F., Martínez-García, P.J. & Neale, D.B. 2016. Recent landscape genomics studies in forest trees—what can we believe? Tree Genet. Genomes **12**: 1–7.

Carrión, J.S., Parra, I., Navarro, C. & Munuera, M. 2000. Past distribution and ecology of the cork oak (Quercus suber) in the Iberian Peninsula: A pollen-analytical approach. Divers. Distrib. **6**: 29–44.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. & Cresko, W.A. 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. **22**: 3124–3140.

Conesa, A. & Götz, S. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics 1–12.

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

Correia, R.A., Bugalho, M.N., Franco, A.M.A. & Palmeirim, J.M. 2018. Contribution of spatially explicit models to climate change adaptation and mitigation plans for a priority forest habitat. Mitig. Adapt. Strateg. Glob. Chang. **23**: 371–386.

Costa, J., Miguel, C., Almeida, H., Oliveira, M.M., Matos, J.A., Simões, F., et al. 2011. Genetic divergence in Cork Oak based on cpDNA sequence data. BMC Proc. **5**: P13.

Cox, K., Vanden Broeck, A., Van Calster, H. & Mergeay, J. 2011. Temperature-related natural selection in a wind-pollinated tree across regional and continental scales. Mol. Ecol. **20**: 2724–2738.

De Carvalho, A. 1958. Identificaçao de um possivel fossil de sobreiro (Quercus suber L.) proveniente de solos do mioceno lacustre do Alentejo. Bol Soc Broteriana **32**: 75–79.

De Kort, H., Vandepitte, K., Bruun, H.H., Closset-Kopp, D., Honnay, O. & Mergeay, J. 2014. Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species Alnus glutinosa. Mol. Ecol. **23**: 4709–4721.

De Mita, S., Thuillet, A.C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., et al. 2013. Detecting selection along environmental gradients: Analysis of eight methods and their effectiveness for outbreeding and selfing populations. Mol. Ecol. **22**: 1383–1399.

De Villemereuil, P., Frichot, É., Bazin, É., François, O. & Gaggiotti, O.E. 2014. Genome scan methods against more complex models: When and how much should we trust them?

Mol. Ecol. **23**: 2006–2019.

Devlin, B. & Roeder, K. 1999. Genomic control for association studies. Biometrics **55**: 997–1004.

Dincă, V., Lee, K.M., Vila, R. & Mutanen, M. 2019. The conundrum of species delimitation: a genomic perspective on a mitogenetically super-variable butterfly. Proc. R. Soc. B Biol. Sci. **286**: 20191311.

Dray, S., Legendre, P. & Peres-Neto, P.R. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol. Modell. **196**: 483–493.

Dukes, J.S., Chiariello, N.R., Cleland, E.E., Moore, L.A., Rebecca Shaw, M., Thayer, S., et al. 2005. Responses of grassland production to single and multiple global environmental changes. PLoS Biol. **3**: e319.

Ellis, N., Smith, S.J. & Pitcher, C.R. 2012. Gradient forests: calculating importance gradients on physical predictors. Ecology **93**: 156–168.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One **6**: e19379.

Ferrier, S., Manion, G., Elith, J. & Richardson, K. 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. Divers. Distrib. **13**: 252–264.

Fick, S.E. & Hijmans, R.J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. **37**: 4302–4315.

Fitzpatrick, M.C. & Keller, S.R. 2015. Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. Ecol. Lett. **18**: 1–16.

Foll, M. & Gaggiotti, O. 2008. A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics **180**: 977–993.

Forester, B.R., Jones, M.R., Joost, S., Landguth, E.L. & Lasky, J.R. 2016. Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. Mol. Ecol. **25**: 104–120.

Frichot, E. & François, O. 2015. A short manual for LFMM. version 1.4.

Frichot, E., Schoville, S.D., Bouchard, G. & François, O. 2013. Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. Mol.

Biol. Evol. **30**: 1687–1699.

Gent, P.R., Danabasoglu, G., Donner, L.J., Holland, M.M., Hunke, E.C., Jayne, S.R., et al. 2011. The community climate system model version 4. J. Clim. **24**: 4973–4991.

Goslee, S.C. & Urban, D.L. 2007. The ecodist package for dissimilarity-based analysis of ecological data. J. Stat. Softw. **22**: 1–19.

Goudet, J. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol. Ecol. Notes **5**: 184–186.

Gower, J.C. 1971. Statistical methods of comparing different multivariate analyses of the same data. Math. Archaeol. Hist. Sci. 138–149.

Gugger, P.F., Liang, C.T., Sork, V.L., Hodgskiss, P. & Wright, J.W. 2018. Applying landscape genomic tools to forest management and restoration of Hawaiian koa (Acacia koa) in a changing environment. Evol. Appl. **11**: 231–242.

Hijmans, R.J., Phillips, S., Leathwick, J. & Elith, J. 2012. dismo: Species distribution modeling. R Packag. version 0.7-17.

Hijmans, R.J. & van Etten, J. 2012. raster: Geographic analysis and modeling with raster data. R Packag. version **1**: 9–92.

Hoban, S., Kelley, J.L., Lotterhos, K.E., Antolin, M.F., Bradburd, G., Lowry, D.B., et al. 2016. Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. Am. Nat. **188**: 379–397.

Hughes, A.R., Inouye, B.D., Johnson, M.T.J., Underwood, N. & Vellend, M. 2008. Ecological consequences of genetic diversity. Ecol. Lett. **11**: 609–623.

Ingvarsson, P.K. & Bernhardsson, C. 2020. Genome-wide signatures of environmental adaptation in European aspen (Populus tremula) under current and future climate conditions. Evol. Appl. **13**: 132–142.

Jombart, T. 2008. Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics **24**: 1403–1405.

Kremer, A., Potts, B.M. & Delzon, S. 2014. Genetic divergence in forest trees: understanding the consequences of climate change. Funct. Ecol. **28**: 22–36.

Lawson, D.J., Hellenthal, G., Myers, S. & Falush, D. 2012. Inference of population structure using dense haplotype data. PLoS Genet. **8**.

Lindner, M., Maroschek, M., Netherer, S., Kremer, A., Barbati, A., Garcia-Gonzalo, J., et al. 2010. Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. For. Ecol. Manage. **259**: 698–709.

Losa Quintana, J.M. 1978. Estudio mineralogico y estructural de un fosil vegetal de Galicia.

An. del Inst. Botánico AJ Cavanilles **35**: 235–243.

Lotterhos, K.E. & Whitlock, M.C. 2014. Evaluation of demographic history and neutral parameterization on the performance of F ST outlier tests. Mol. Ecol. **23**: 2178–2192.

Lotterhos, K.E. & Whitlock, M.C. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol. Ecol. **24**: 1031–1046.

Lumaret, R., Tryphon-Dionnet, M., Michaud, H., Sanuy, A., Ipotesi, E., Born, C., et al. 2005. Phylogeographical variation of chloroplast DNA in cork oak (Quercus suber). Ann. Bot. **96**: 853–861.

Magri, D., Fineschi, S., Bellarosa, R., Buonamici, A., Sebastiani, F., Schirone, B., et al. 2007. The distribution of Quercus suber chloroplast haplotypes matches the palaeogeographical history of the western Mediterranean. Mol. Ecol. **16**: 5259–5266.

Malinsky, M., Trucchi, E., Lawson, D.J. & Falush, D. 2018. RADpainter and fineRADstructure: Population Inference from RADseq Data. Mol. Biol. Evol. **35**: 1284–1290.

Mantel, N. 1967. The Detection of Disease Clustering and a Generalized Regression Approach. Cancer Res. **27**: 209–220.

Martins, K., Gugger, P.F., Llanderal-Mendoza, J., González-Rodríguez, A., Fitz-Gibbon, S.T., Zhao, J.L., et al. 2018. Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of Quercus rugosa. Evol. Appl. **11**: 1842–1858.

Modesto, I.S., Miguel, C., Pina-Martins, F., Glushkova, M., Veloso, M., Paulo, O.S., et al. 2014. Identifying signatures of natural selection in cork oak (Quercus suber L.) genes through SNP analysis. Tree Genet. Genomes **10**: 1645–1660.

Neale, D.B., Martínez-García, P.J., De La Torre, A.R., Montanari, S. & Wei, X.-X. 2017. Novel Insights into Tree Biology and Genome Evolution as Revealed Through Genomics. Annu. Rev. Plant Biol. **68**: 457–483.

Nobre, A., Empadinhas, N., Nobre, M.F., Lourenço, E.C., Maycock, C., Ventura, M.R., et al. 2013. The plant Selaginella moellendorffii possesses enzymes for synthesis and hydrolysis of the compatible solutes mannosylglycerate and glucosylglycerate. Planta **237**: 891–901.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., et al. 2013. Package 'vegan.' Community Ecol. Packag. version **2**: 1–295.

Petit, R.J., Hampe, A. & Cheddadi, R. 2005. Climate changes and tree phylogeography in the Mediterranean. Taxon **54**: 877–885.

Pina-Martins, F., Baptista, J., Pappas, G. & Paulo, O.S. 2019. New insights into adaptation and population structure of cork oak using genotyping by sequencing. Glob. Chang. Biol. **25**: 337–350.

Príncipe, A., Matos, P., Sarris, D., Gaiola, G., do Rosário, L., Correia, O., et al. 2019. In Mediterranean drylands microclimate affects more tree seedlings than adult trees. Ecol. Indic. **106**. Elsevier B.V.

Prober, S.M., Byrne, M., McLean, E.H., Steane, D.A., Potts, B.M., Vaillancourt, R.E., et al. 2015. Climate-adjusted provenancing: A strategy for climate-resilient ecological restoration. Front. Ecol. Evol. **3**: 65.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. version 3.3. 0.

Rellstab, C., Gugerli, F., Eckert, A.J., Hancock, A.M. & Holderegger, R. 2015. A practical guide to environmental association analysis in landscape genomics. Mol. Ecol. **24**: 4348–4370.

Rellstab, C., Zoller, S., Walthert, L., Lesur, I., Pluess, A.R., Graf, R.R., et al. 2016. Signatures of local adaptation in candidate genes of oaks (Quercus spp.) in respect to present and future climatic conditions. Mol. Ecol. **25**: 5907–5924.

Savolainen, O., Pyhäjärvi, T. & Knürr, T. 2007. Gene Flow and Local Adaptation in Trees. Annu. Rev. Ecol. Evol. Syst. **38**: 595–619.

Schippers, J.H.M., Nunes-Nesi, A., Apetrei, R., Hille, J., Fernie, A.R. & Dijkwel, P.P. 2008. The Arabidopsis onset of leaf death5 mutation of quinolinate synthase affects nicotinamide adenine dinucleotide biosynthesis and causes early ageing. Plant Cell **20**: 2909–2925.

Schoville, S.D., Bonin, A., François, O., Lobreaux, S., Melodelima, C. & Manel, S. 2011. Adaptive Genetic Variation on the Landscape: Methods and Cases. Annu. Rev. Ecol. Evol. Syst. **43**: 23–43.

Shafer, A.B.A., Peart, C.R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C.W., et al. 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods Ecol. Evol. **8**: 907–917.

Sork, V.L. 2016. Gene flow and natural selection shape spatial patterns of genes in tree populations: implications for evolutionary processes and applications. Evol. Appl. **9**: 291–310.

Sork, V.L., Aitken, S.N., Dyer, R.J., Eckert, a. J., Legendre, P. & Neale, D.B. 2013. Putting the landscape into the genomics of trees: approaches for understanding local adaptation

and population responses to changing climate. Tree Genet. Genomes **9**: 901–911.

Storey, J., Bass, A., Dabney, A. & Robinson, D. 2015. qvalue: Q-Value estimation for false discovery rate control. R Packag. version 2.10.

Stucki, S., Orozco-terWengel, P., Forester, B.R., Duruz, S., Colli, L., Masembe, C., et al. 2017. High performance computation of landscape genomic models including local indicators of spatial association. Mol. Ecol. Resour. **17**: 1072–1089. Blackwell Publishing Ltd.

Supple, M.A., Bragg, J.G., Broadhurst, L.M., Nicotra, A.B., Byrne, M., Andrew, R.L., et al. 2018. Landscape genomic prediction for restoration of a Eucalyptus foundation species under climate change. Elife **7**: 1–22.

Vessella, F., López-Tirado, J., Simeone, M.C., Schirone, B. & Hidalgo, P.J. 2017. A tree species range in the face of climate change: cork oak as a study case for the Mediterranean biome. Eur. J. For. Res. **136**: 555–569.

Vessella, F., Simeone, M.C. & Schirone, B. 2015. Quercus suber range dynamics by ecological niche modelling: from the Last Interglacial to present time. Quat. Sci. Rev. **119**: 85–93.

Williams, A. V, Nevill, P.G. & Krauss, S.L. 2014. Next generation restoration genetics: applications and opportunities. Trends Plant Sci. **19**: 529–537.

Data access

Raw GBS data are available on NCBI's Sequence Read Archive (SRA) as "BioProject" PRJNA413625.

Tables

Table 1 - Onthology of the SNPs associated with climatic variables

Locus	**Environmental variables**	**Description**
SNP_187039	Bio06, Bio07, Bio13, Bio19	DNA-binding transcriptional regulator DhaR
SNP_6044	Bio12, Bio13, Bio19	Mannosylglycerate hydrolase
SNP_12029	Bio06, Bio07	quinolinate synthase, chloroplastic

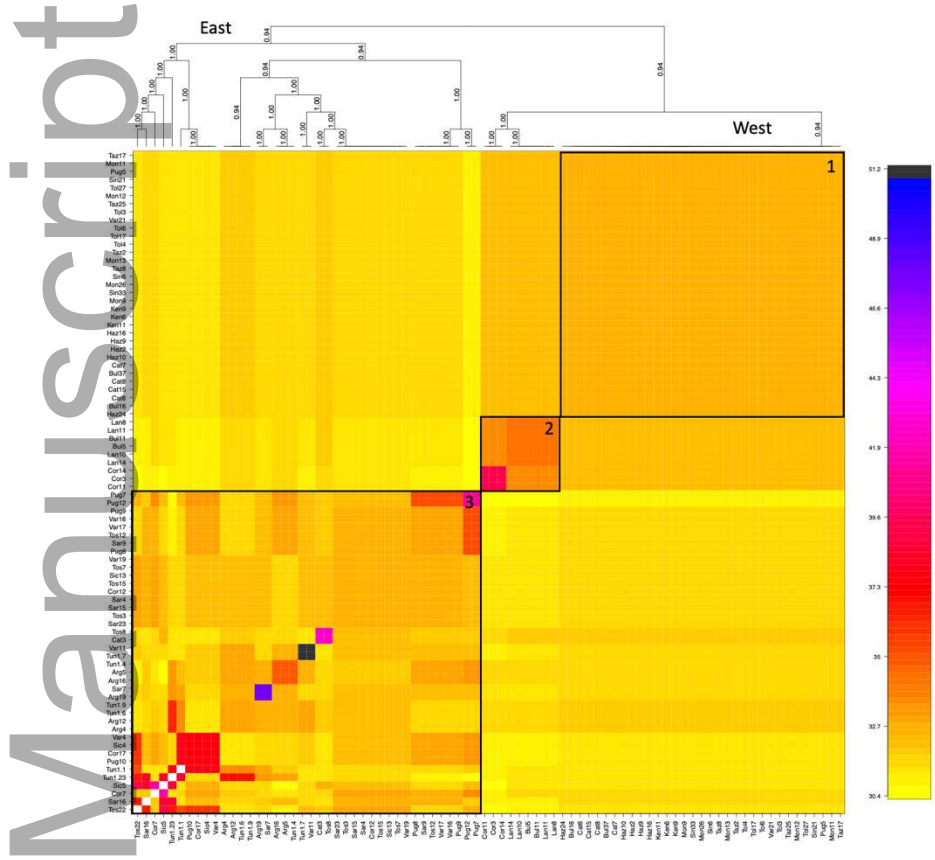
SNP_133831	Bio09	ribose-phosphate diphosphokinase
SNP_32261	Bio09	kinesin-like protein KIN-14L
SNP_3604	Bio09	DNA mismatch repair protein MSH3 isoform X1
SNP_11143	Bio07	Alanine tRNA ligase
SNP_11666	Bio08	ATP-dependent DNA helicase PcrA
SNP_226326	Bio08	DNA polymerase processivity factor
SNP_143637	Bio08	DNA polymerase processivity factor
SNP_245202	Bio08	Plipastatin synthase subunit A

Figures

Figure 1 - Ecological niche, dark green, of Q. suber with each sampling location colored according to the associated and group of populations (blue, black and red)

Figure 2 – FineRADstructure analyses of the cork oak population structure. On the x-axis, each sample is considered as a recipient, and on the y-axis, the sample is considered a donor of genomic regions. The western (1) and eastern (3) group of populations are clearly separated with limited sharing of genomic regions between the two groups of populations.

804 The second population (2) containing samples from Corsica and the Landes region is closely
805 related to the western group. The highest amount of shared genome regions between samples
806 appears in purple and the lowest in yellow.
807



808
809
810
811 Figure 3 – Results for the outlier Fst test based on 17 sampled populations of cork oak
812

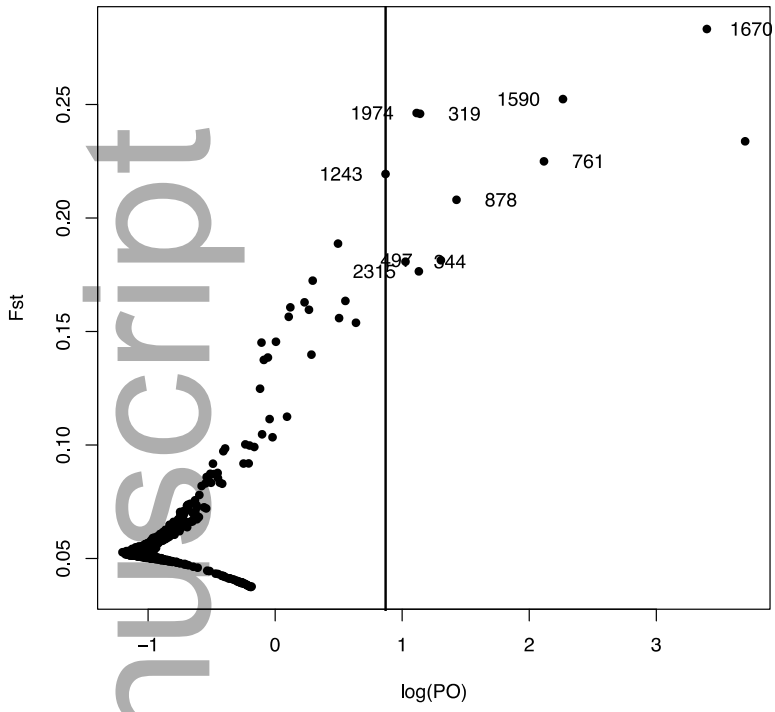


Figure 4 – (a) R^2-weighted importance of environmental and spatial variables for the adaptive SNPs model. (b) Cumulative importance of allelic change along four environmental gradients for the adaptive and neutral SNPs model.

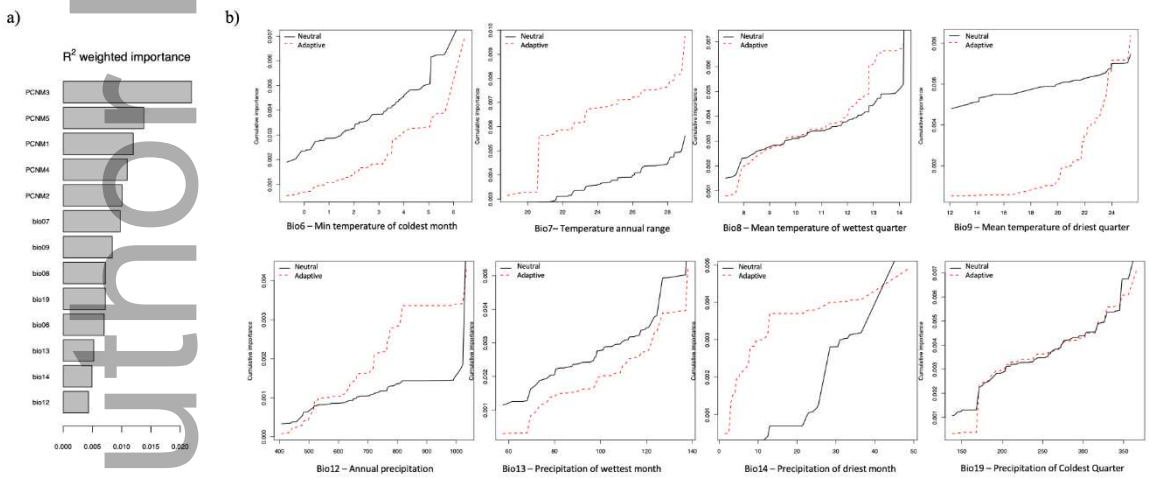
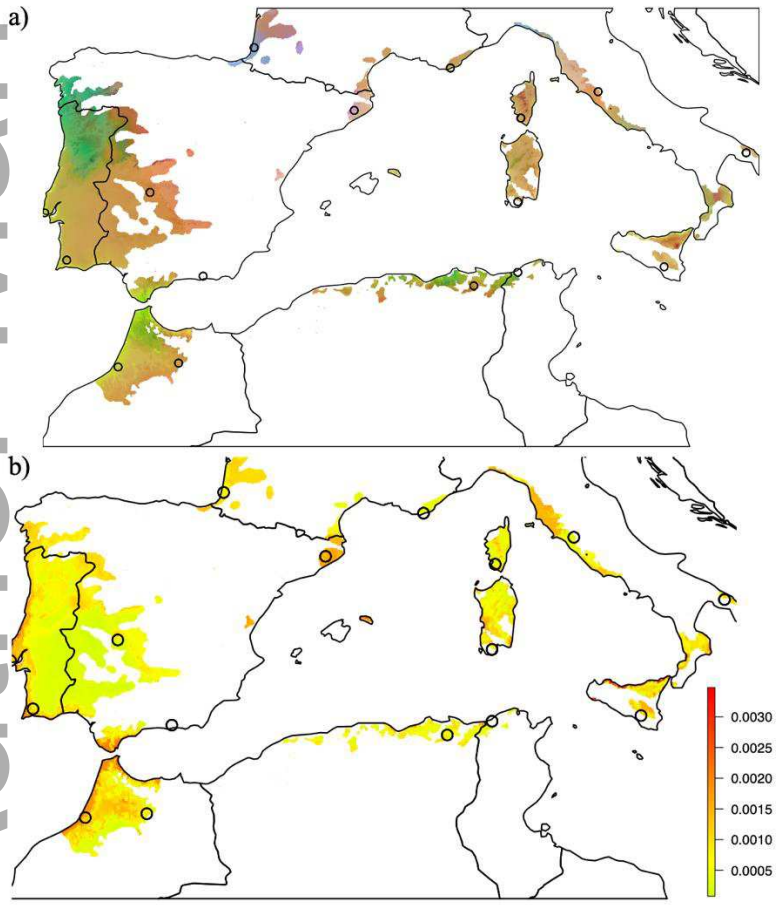


Figure 5 – Maps of predicted turnover in allele frequencies. a) Current landscape patterns of allelic composition predicted under the gradient forest for the adaptive SNPs dataset. A red-green-blue color scale was generated using the first three axes of the principal components of the gradient forest prediction. Regions with similar colors are expected to have similar genetic composition. b) the difference between full SNP and adaptive SNPs datasets based on Procrustes residuals.



836 Figure 6 – Predictive genetic offset (full SNPs dataset) under climate change for 2070.
837 Euclidian distances between current and future climate were calculated for each model and
838 regions with larger euclidian distances are expected to have a larger genetic offset.

839