Common microgeographic selection patterns revealed in four European conifers

Scotti Ivan (Orcid ID: 0000-0002-8951-2680)

ODDOU-MURATORIO Sylvie (Orcid ID: 0000-0003-2374-8313)

Grivet Delphine (Orcid ID: 0000-0001-8168-4456)

Lefèvre François (Orcid ID: 0000-0003-2242-7251)

CUBRY Philippe (Orcid ID: 0000-0003-1561-8949)

Common microgeographic selection patterns revealed in four European conifers

Ivan SCOTTI[1]§, Hadrien LALAGÜE[2], Sylvie ODDOU-MURATORIO[1,*], Caroline SCOTTI-SAINTAGNE[1], Rose RUIZ DANIELS[3], Delphine GRIVET[4], François LEFEVRE[1], Philippe CUBRY[5], Bruno FADY[1], Santiago C. GONZÁLEZ-MARTÍNEZ[6], Anne ROIG[1], Isabelle LESUR-KUPIN[6], Francesca BAGNOLI[7], Vanina GUERIN[8], Christophe PLOMION[6], Philippe ROZENBERG[8], Giovanni G. VENDRAMIN[7]

[1] INRAE, URFM, Avignon (France)

[2] UMR EcoFoG, AgroParisTech, CIRAD, CNRS, INRAE, Université des Antilles, Université de Guyane, Campus Agronomique, Kourou (France)

[3] The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian, (United Kingdom)

[4] CSIC, INIA, CIFOR, Madrid (Spain)

[5] DIADE, Univ Montpellier, CIRAD, IRD, Montpellier (France)

[6] Univ. Bordeaux, INRAE, BIOGECO, F-33610 Cestas, France

[7] CNR, IBBR, Sesto Fiorentino (Italy)

[8] INRAE, BIOFORA, Orléans (France)

[*] present affiliation: ECOBIOP Université de Pau et des Pays de l'Adour, E2S UPPA, Saint-Pée-sur-Nivelle (France)

§ Corresponding Author; e-mail ivan.scotti@inrae.fr

Abstract

Microgeographic adaptation occurs when the effects of directional selection persist despite gene flow. Traits and genetic loci under selection can then show adaptive divergence, against the backdrop of little differentiation at other traits or loci. How common such events are and how strong selection is that underlies them, remain open questions.

Here, we discovered and analysed microgeographic patterns of genomic divergence in four European and Mediterranean conifers with widely differing life-history traits and ecological requirements (*Abies alba* MIll., *Cedrus atlantica* (Endl.) Manetti, *Pinus halepensis* Mill. and *Pinus pinaster* Aiton) by screening pairs geographically close forest stands sampled along steep ecological gradients. We inferred patterns of genomic divergence by applying a combination of divergence outlier detection methods, demographic modelling, Approximate Bayesian Computation inferences, and genomic annotation to genomic data.

Surprisingly for such small geographical scales, we showed that selection is strong in all species but generally affects different loci in each. A clear signature of selection was systematically detected on a fraction of the genome, in the order of 0.1% to 1% of the loci depending on the species. The novel modelling method we designed for estimating selection coefficients showed that the microgeographic selection coefficient scaled by population size (Ns) was 2-30.

Our results convincingly suggest that selection maintains within-population diversity at microgeographic scales in spatially heterogeneous environments. Such genetic diversity is likely to be a major reservoir of adaptive potential, helping populations to adapt under fluctuating environmental conditions.

Introduction

Many populations of long-lived, outcrossing species harbour large amounts of genetic variability (Hamrick & Godt, 1990; but see Vendramin et al. (2008) and Jaramillo-Correa et al. (2020) for contrasting cases). Most of this variability is likely to be effectively neutral and to be maintained because of large effective population sizes. However, some of these large populations expand over spatially heterogeneous environments. Under such conditions, microgeographic divergence (i.e., genetic and/or phenotypic divergence occurring within gene dispersal distance, Richardson et al., 2014) may occur if selection is strong enough to overcome the homogenising effects of gene flow (Bulmer, 1972). There is increasing interest in the effect of microgeographic selection, as environmental conditions can be extremely variable at the local spatial scale. This heterogeneity can have a sizable selective effect on within-population divergence (Langin et al., 2015), and call for processes whereby divergence occurs or has occurred in the apparent absence of barriers to gene flow ("strong-selection-with-high-gene-flow" scenario, Sexton et al., 2014). This kind of pattern was observed repeatedly at different taxonomic scales. At one end of the spectrum lie cases of within-population divergence for multiple plant species (in *Anthoxantum odoratum* populations submitted to heavy metal pollution, Antonovics, 1968, 2006; in *Agrostis tenuis* under lead pollution, Bradshaw, 1960; in *Arabidopsis lyrata* on serpentine soils, Turner et al., 2010); at the other end of the spectrum lie the rare proven cases of sympatric speciation (in *Howea* palms on oceanic islands; Babik et al., 2009; Savolainen et al., 2006). In particular, multiple studies in forest tree species - from tropical to temperate and alpine environments - provide evidence in favour of microgeographic adaptation under high gene flow (Audigeos et al., 2013; Brousseau et al., 2013, 2015, 2016; Eckert et al., 2015; Gauzere et al. 2020; Lobo et al., 2018; Mosca et al., 2016; Roschanski et al., 2016; Ruiz Daniels et al. 2019; for a review see Lind et al., 2017). This leads on

to the following questions: (1) How common is microgeographic adaptive genetic divergence? (2) How strong is microgeographic selection? (3) What is the genome-wide architecture of the response to microgeographic selection? (4) How much overlap is there among species in the loci undergoing divergent selection?

We addressed these issues using four conifer trees of the Pinaceae: the Alpine *Abies alba* Mill., the mountain Mediterranean *Cedrus atlantica* (Endl.) Manetti, the thermophilic Mediterranean *Pinus halepensis* Mill. and Atlantic *Pinus pinaster* Aiton as model systems. Conifers are long lived and genetically diverse. The species we chose display large, continuous populations. While showing different and contrasted ecological requirements, biogeographic coverage and demographic histories, they are often found locally across steep ecological gradients (e.g., precipitation for *P. pinaster* and *P. halepensis*, temperature for *A. alba* and *P. halepensis*, and soil composition and water availability for *C. atlantica*. The four selected species are significantly affected by the climate changes and it is expected that their adaptation will become increasingly challenged being unable to efficiently adjust to the warmer climates predicted for the next few years.

This work relies on very general expectations on local adaptation of conifer tree populations experiencing contrasted environmental conditions at microgeographic scale. Conifer trees in general (with few exceptions) form large populations, which harbour large amounts of within-population genetic diversity, as previously stated, and they are outcrossing species with long distance gene flow capacities (Kremer et al. 2012). Therefore, sampling sites with highly contrasted environments, we assume that: (i) selection is not constrained by a lack of genetic diversity within each site, (ii) the risk of false signal of within-site selection due to neutral divergence is very low, (iii) if microgeographic selection exists, we are in appropriate conditions to detect it, or, in other words, if we do not detect microgeographic selection in those sites it has little chance to be detected

elsewhere. We indeed chose pairs of populations that represent extremes of otherwise continuous environmental gradients (probably representing a bundle of environmental variables), and treated them as qualitatively different populations; we contend that, in such conditions, divergent selection is strong enough to partially overcome gene flow, irrespective of the nature of the environmental contrast under scrutiny; indeed, our argument draws generality from the choice of analysing a priori multiple types of environmental contrasts, instead of focusing on the variation of a particular (set of) environmental variable(s). We expect that our study will provide a general view of the molecular basis of adaptation to strong, but geographically short, environmental contrasts, in particular because: (1) the use of the same sequence capture approach for all species should help sampling portions of genomes with a similar structure and similar levels of diversity (Yeaman et al. 2016); (2) beyond specific loci involved in adaptation, comparing estimates of strength of selection in different species and under varying environmental constraints and varying demographic history (but similar local patterns of low neutral divergence) will provide valuable insight on the general way selection shapes diversity within otherwise continuous populations. However, we also expect to find very limited overlap in the loci that respond to divergent selection across species, because (i) the environmental contrasts vary among species, and thus a priori target different genes and characters; (ii) considering the complex genetic control of traits involved in adaptation to climatic drivers, many loci with small effects are expected to be under selection (omnigenic model, Boyle et al. 2017) and this genetic diversity on which selection operates is also conditioned by the demographic history; and (iii) there is evidence suggesting that adaptation to identical environments may involve separate genes and a certain number of possible paths (e.g., Frachon et al. 2018; Tenaillon et al. 2012).

To investigate and compare patterns of divergence by adaptation, we screened genome-wide

polymorphisms and looked for loci that showed excess divergence between closely related stands, relative to the genomic background determined by demography and gene flow. Such loci are indicative of adaptive divergence; we expect that, if microgeographic adaptation is a strong enough force, some fraction of the genome will show adaptive divergence between stands, irrespective of the species and the type of environmental contrast under scrutiny. Our results point to strong selection maintaining divergence at multiple, independent target loci, notwithstanding high gene flow.

Materials and Methods

Species and study sites

Four conifer species of the Pinaceae family, having diverse geographical distributions and ecological niches, were used in this study. The continentally distributed *Abies alba* Mill., is adapted to mid-elevation, wet, cold-temperate climates. The Mediterranean and Atlantic Coast *Pinus pinaster* Ait., is adapted to a range of wet to dry climates at low to medium elevations. The Mediterranean *Pinus halepensis* Mill., is adapted to low-altitude, hot and dry climates, and the North-African *Cedrus atlantica* (Endl.) Manetti is adapted to high-altitude, moderately dry subtropical climates.

For *A. alba*, *P. halepensis* and *P. pinaster*, we sampled wild populations in their native range. For *C. atlantica*, we sampled populations that resulted from the natural expansion of genetically diverse founder populations introduced outside of their native range (sampled two generations after introduction) to investigate potential microgeographic selection in a very short-term. Sampling was carried out in a way that provided multiple pairs of ecologically contrasted, but geographically close, stands. We define each pair of geographically close stands as a *site*. The stands of a given site were chosen as representing the extremes of an otherwise continuous environmental gradient, in general associated to an elevational cline. Notice though that such clines most likely represent a bundle of environmental variables. Because the goal of our study was to identify divergence between contrasting stands, and not to study the effect of a given environmental variable on genetic divergence, we interpret the contrasts in a qualitative, binary way (high/low, dry/wet, …). The study hypothesis is that sufficiently strong contrasts can generate adaptive divergence, no matter what generates the contrast, and therefore it acquires generality from the fact that our sites are a mix of potentially multiple types of contrasts. Three sites were sampled for *A. alba*, and two sites for *P.*

halepensis, P. pinaster and *C. atlantica*; i.e. a total of 18 stands (see Figure 1 for locations and Table 1 for coordinates and further details on site description). Stands within sites were less than one kilometre apart, except for the Spanish *P. halepensis* stands, which were around 100 km apart, for which no neutral genetic differentiation was detected (Budde et al. 2017; Ruiz-Daniels et al. 2019, and see Results section for confirmation) and were thus considered as a *site*. In this nested design, we searched for genomic signatures of microgeographic differentiation between stands belonging to the same site.

We sampled stands with strongly different demographic patterns: (i) likely to have been stable on the long term for *A. alba*, as these stands belong to a biogeographic sector with long-standing presence of the species (Scotti-Saintagne et al., 2021) (ii) recently introduced from a genetically diverse seed source and potentially having adapted to new conditions over few generations for *C. atlantica* (Lefèvre et al. 2004), and (iii) belonging to pioneer species, and therefore having likely undergone repeated colonisation events (*P. halepensis* (Ruiz Daniels et al. 2019, Olsson et al. 2021, Vendramin et al. 2021); *P. pinaster* (Bucci et al. 2007)). To take these diverse stories into account in our analyses, we estimated historical demographic transitions and used them in population models to control for their effects on genetic diversity patterns as explained below. By taking a multi-species approach and replicating study sites within species, we were able to generalise our conclusions about the strength and extent of selective forces shaping landscape-level genetic structure. The wide variation of demographic (and likely selection) histories represented in our study species is a strong asset to assess the effect of such factors on our capacity to detect and estimate selection.

We studied pairs of forest stands identified within continuous populations (or closely related ones) to minimise stochastic genome-wide divergence. Indeed, gene flow and recent shared ancestry

minimise background divergence by drift, and thus maximise the power of detecting loci under selection while reducing False Discovery Rate (FDR; Lotterhos & Whitlock, 2015; Nadeau & Jiggins, 2010). Low overall divergence and a hierarchical island-model sampling scheme allowed us to avoid the classical pitfalls of the adaptive interpretation of genetic divergence patterns, such as the confounding effects of large inter-locus variance in (neutral) population differentiation and of population structure (Hoban et al., 2016). Additionally, large population sizes should decrease the risk that background natural selection causes divergence at neutral loci (Charlesworth et al., 1997).

Sampling and genomic data production

Needles from twenty-five mature, emergent individuals (that is, trees with crown reaching or above the canopy), located at least 20 m apart, were sampled in each stand. Supplementary Methods M1 provide details on DNA isolation.

DNA samples were submitted to sequence capture using a subset of the sequence capture probes designed for *Pinus taeda* by RapidGenomics (Gainesville (FL), USA) (Supplementary Table T1; see Supplementary Methods M1 for details on probe choice). Sequence capture was carried out by RapidGenomics (Gainesville (FL), USA) following Peñalba et al. (2014). After assembly, variant calling was applied with the following parameters: minimum coverage = 20; minimum average quality = 30, minimum variant frequency = 0.3, minimum frequency for homozygote call = 0.7, maximum missing data = 20%; see Supplementary Methods M1 for full details on sequence capture, assembly, SNP calling, and filtering. Further filters were applied based on the biological properties of variants and their distribution, using the *ad hoc* R (R Development Core Team, 2008) *sieve()* script (Supplementary Methods M2). The *sieve()* script takes as input a VCF file and filters out contigs based on SNP density (variants per base) and heterozygosity thresholds; it also allows

the removal of targeted individuals, and returns genotype tables; conversion from vcf to genotype tables was performed by using the vcfR R package v. 1.12.0 (Knaus & Grünwald, 2017). Contigs with variant density > 0.05 (i.e. having more than one variant every 20 bases) and carrying only heterozygous genotypes were excluded, to remove contigs potentially containing paralogs. Filtering with *sieve()* was completed by the removal of monomorphic loci with the *ad-hoc monoRemove()* R script (Supplementary Methods M2).

Data analyses

SNP data were analysed with the aim of identifying loci with significant divergence patterns between stands from the same site, i.e. signatures of putative microgeographic adaptive divergence, and to evaluate the strength and kind of selective pressure that could produce such patterns. We proceeded through the following steps: (*a*) assess genome-wide and contig-level diversity and divergence patterns; (*b*) estimate species demographic parameters to be used for subsequent simulations; (c) apply multiple independent methods for the detection of within-site single-locus divergence outliers and dissect trends of stand-level allele frequencies; (*d*) apply evolutionary-modelling methods to estimate the strength of selection at outlier loci; (*e*) functionally annotate outlier-containing contigs and outlier SNPs. Details of the methods used for each analysis are provided in the main text (for new developments) or as Supplementary Methods (for standard methods). Steps *(c) to (e)* allowed us to characterise in detail the putative adaptive processes underlying divergence patterns for each locus.

(a) diversity analyses

Diversity was assessed by computing Nei's molecular diversity π (Nei, 1987) per contig, taking into account both variant and invariant sites in each contig, within each stand. Calculations were carried

out in R with a custom script (Supplementary Methods M2).

Within-stand observed and expected heterozygosity and heterozygote excess/deficit were computed with the basic.stats() function, and hierarchical *F*-statistics was computed using the pairwise.WCfst() function, both from R package hierfstat (Goudet, 2005; de Meeûs & Goudet, 2007), which implements the hierarchical framework developed by Yang (1998), using all loci.

(b) estimation of stand demographic parameters

We inferred demographic parameters for the stands of each species by applying the Stairway Plot 2 approach (Liu & Fu, 2020) to our data. This method uses the Site Frequency Spectrum (SFS) in a coalescent-based framework to construct past evolution in effective size through a multi-epoch model. For each stand of the four studied species, we computed the SFS and inferred its demographic history using default parameters of the software. We considered a generation time of 25 years (Jia et al., 2018; Hrivnák et al., 2017) and a mutation rate of 10-9 (Willyard et al. 2007) for all species to calibrate the output. The central estimators of effective population size were then used to set the parameters for subsequent simulations, both in (c), to estimate type I and type II error rates, and in (d), to estimate selection parameters. For the *C. atlantica* stands, which result from recent introduction of unknown origin(s), the reconstruction of past demography is likely to be biassed because we ignore the exact sources of the starting seedlings, but we consider it as representing the mean demography of the whole species in its area of origin.

Notice that, because of the scattered type of population sampling we have used, these demographic inferences are not meant to represent the whole species, and must be considered as representing the particular stands under study. The demographic inferences used here are instrumental to the modelling of selective processes in these particular stands(see below).

(c) detection of divergence outliers and trends in allele frequencies

We used two methods aiming at identifying loci with disproportionate allele frequency differentiation among stands within sites (BAMOVA and pcadapt). We view these two methods as complementary, as they have contrasting assumptions, and can therefore help capture divergence signals producing different types of genetic differentiation patterns. The BAMOVA algorithm (Gompert & Buerkle, 2011) is a Bayesian method that rests on pre-defined population groups and a combination of single-locus and whole-genotype information, and was used to compute single-locus and genome-wide estimates of hierarchical Φ-statistics (Φ_{ST}: global differentiation among stands; Φ_{CT}: differentiation among sites; Φ_{SC}: differentiation between stands within sites;).

For each species, 1 million iterations were run (with sampling every 100 iterations), with a "known haplotypes" likelihood model, a random-walk MCMC algorithm for genome-wide parameters α and β, adjustment for the variance of proposal distribution for α and β equal to 0.2, an independence-chain MCMC algorithm for proposing haplotype frequency vectors, haplotype vector variance adjustment parameter equal to 1, a uniform hyper-prior [0.5; 105] for α and β, correlation of bi-variate normal distribution proposals for α and β equal to 0.8, haplotype frequency proposal distribution adjustment parameter equal to 0.00001. Convergence was reached within the first quarter of each run. After discarding the first half of the iterations, posterior densities were estimated on the remaining values.

Outlier loci (i.e. loci with exceedingly high values of Φ_{SC}) were identified based on Bayes Factors (BF; Kass and Raftery 1995; Makowski et al. 2019). The computation of BF requires (i) the estimation of the probability density for the null hypothesis distribution at the mode; (ii) the estimation of the probability density for the alternative hypothesis at the mode of the probability density for the null hypothesis distribution; (iii) the computation of the ratio of (ii) to (i), that is, the BF.

To obtain the null distribution (i.e. the probability density distribution for Φ_{SC} for a random locus) we randomly sampled 50,000 Φ_{SC} values from all loci from all steps of the Markov chain produced by BAMOVA (after discarding the burn-in), and computed the probability density distribution from this set of values. We then computed the position of the distribution's mode and proceeded with points (i) through (iii) described above. We considered loci with BF > 30 (corresponding to the middle of the "strong" evidence class in Kass and Raftery 1995) as outliers.

The second method used for the detection of F_{ST} outliers is implemented in the R package pcadapt v. 4.3.3 (Luu et al., 2017). It ascertains population structure using principal component analysis (PCA) and computes associations between Principal Components (PCs) and locus variation for each SNP. The rationale behind pcadapt is that loci under divergent selection present atypical association with PCs. In this study, we used the first version of pcadapt (v1.0; Duforet-Frebourg et al., 2014). To run pcadapt, missing data were first imputed. Then, the number of PCs to be retained (K) was chosen by running a PCA with a large number of PCs (15 in our case). The best K was chosen based on the rate of decrease in the cumulative variance explained by each PC (i.e. the scree plot). This choice was confirmed by plotting individuals on the first two PCs (i.e. the score plot). After these preliminary steps, PCA was run setting the number of PCs to the optimal K value. Alleles with minor allele frequency (MAF) <5% were removed from the analysis. The communality statistic was used to detect outlier SNPs. An FDR of 0.05 was applied to avoid false positives, by using the R package qvalue (Storey et al., 2015).

Overlap of lists of outliers obtained with the two methods was visualised using the R package upsetR v. 1.4.0 (Conway et al., 2017). The trends in allele frequencies at outlier SNPs were compared between same-species sites, to check whether they were consistent among sites; that is, we checked whether, when the frequency of an allele increased (or decreased) from one

environmental level to the other between the two stands of a site, it also increased (or decreased) between the two same levels at the other site or sites; changes in the same direction are interpreted as suggestive of selection operating the same way at all sites.

To estimate the Type I and Type II error rates in outlier detection, we used the QUANTINEMO software (Neuenschwander et al., 2008) to simulate different evolutionary scenarios with the same population design as in our empirical data, i.e. two sites with two stands in contrasted environments within each site (see Supplementary Methods M3 for detailed methods), following the demographic models estimated at point (b) above (see results). The retained simulated sets of stands (Supplementary methods M3) were submitted to BAMOVA and pcadapt analyses for the detection of outliers, as described above.

In addition to the simulation approach above, we also applied a permutation approach to estimate the potential for the identification of artefactual outliers with our BAMOVA-based outlier detection method. Individuals from one of the data sets (*A. alba*) were permuted across stands, and BAMOVA and our *ad-hoc* BAMOVA outlier detection method were applied to these permuted stand samples.

For outlier loci, stand-level allele frequencies were inspected, to check whether trends of allele frequency variation were consistent among sites within each species (i.e. whether allele frequencies changed in the in the same direction in all sites between stands displaying different environmental conditions: low altitude → high altitude (*A. alba*), good site → bad site (*C. atlantica*), wet → dry (*P. halepensis*), sunny → shady *(P. pinaster*)).

(d) ABC modelling to estimate selection

We develop here an estimation method that rests on Bayesian estimators of divergence statistics. Based on the posterior distributions of genome-wide and locus-level Φ_{SC}, and on drift-migration equilibrium assumptions for neutral loci, we modelled the intensity of divergent selection within each site, i.e. between two stands adapted to divergent habitats. For this, we applied a combination of forward simulations, performed based on the demographic models estimated following the method at point (b) above (see results), using GENOMEPOP2 v. 2.7.6 (Carvajal-Rodríguez, 2008), and Approximate Bayesian Computation (ABC; Beaumont et al., 2002; see Supplementary Methods M4 for details of the ABC procedure). The 'standard' ABC method described in Supplementary Methods M4 was modified to optimise the analysis of our data: BAMOVA analyses provide true Bayesian, posterior probability density distributions of locus-level observed summary statistics (OSS) (namely, Φ_{SC}), instead of point estimates. To take into account the information conveyed by those posterior distributions, we ran 100 independent ABC analyses for each BAMOVA outlier from each species, by randomly drawing from the posterior Φ_{SC} distribution, then cumulated the values retained in a "*hyperposterior"* distribution. "Hyperposterior" probability distributions were obtained for the selection coefficients scaled to population sizes, Ns, for each BAMOVA outlier (Ns is actually the ratio of a selection intensity term (s) to a drift intensity term ($1/N$)). We argue that when a distribution or a confidence interval on OSS is available, this method allows for more accurate posterior estimates on parameters and more conservative (i.e. larger) credible intervals (as OSS are left to vary, and thus the retained parameter values are drawn from a larger, fuzzier hypervolume) than when point estimates are used. See Supplementary Methods M4 for ABC cross-validation analyses.

(e) Annotation

A BLASTX search was performed between the consensus sequences of each species and a protein database made from the *Pinus taeda* transcriptome v1.01 (https://pinerefseq.faculty.ucdavis.edu/), the NCBI plant ref_seq (release of June 2017) and the Uniprot/Swissprot (release of July 2017). Only the best hit was conserved for the annotation using INTERPROSCAN v5.25 and the HAMAP, PFAM, PIRSF, PRINTS, PRODOM, PROSITEPATTEN, PROSITEPROFILE, and TIGRFAM databases. Only the annotation with the highest e-value was reported. The Blastx output was converted to gff format using the script blast_to_gff.py available at https://github.com/alvaralmstedt/py_scripts/blob/master/blast_to_gff.py. The gff containing the ORF and the delimitation of the coding DNA sequences (CDS) was used to identify the functional significance of SNP mutations using SNPEFF v4.3 (Cingolani et al., 2012). This information was added to the vcf files.

Orthology among contigs, obtained for the four species, and with the original P. taeda probes was analysed using Orthofinder (Emms and Kelly, 2015), by running orthology search on the five data sets (four sets of contigs and one set of probes) together. We used the results both to identify the numbers of probes having retrieved sequences in each species and to identify orthologous contigs in pairs of species.

From the whole dataset of the annotated genes, enrichment among the annotated outliers was performed using the R package TOPGO v2.36.0 (Alexa and Rahnenfuhrer, 2019). The parent-child algorithm, which takes into account the hierarchical relationship between Gene Ontology (GO) terms, was used with a node size of 10. The result of the GO enrichment analysis was used for semantic clustering using REVIGO (http://revigo.irb.hr/) with default settings in order to identify non-redundant sub-sets of GO terms (Supek et al., 2011). The hierarchical data obtained were plotted using CirGO package (Kuznetsova et al., 2019) which allow a better representation of the

GO terms categories and sub-categories.

Results

Assembly and variant calling

The raw data (see Supplementary Table T2 for sample-level quality assessment) were assembled in 11,613 contigs for *A. alba*, 11,659 for *C. atlantica*, 8,266 for *P. halepensis,* and 10,664 for *P. pinaster*. Contig consensus sequences for all species are reported in Supplementary Results R1-R4, and the distributions of contig lengths and levels of polymorphism per species in Supplementary Results R5. Overall, the data set used for the annotation contained 59,430 SNPs within 12,676 contigs. After the filtering with the R function *sieve*(), the final data set contained the following numbers of SNPs / contigs: 7,139 / 1,367 for *A. alba*, 8,462 / 1,531 for *C. atlantica*, 7,114 / 2,814 for *P. halepensis*, and 8,354 / 3,419 for *P. pinaster* (Supplementary Table T3). See Supplementary Results R5 for the distribution of the variants over contigs and the distribution of heterozygosity values.

(a) Basic diversity estimates

Per-contig, within-stand Nei's π varied between 0 and 0.0199 (median: 0.0010) for *A. alba*, between 0 and 0.0139 (median: 0.0008) for *C. atlantica*, between 0 and 0.0243 (median: 0.0016) for *P. halepensis*, and between 0 and 0.0231 (median: 0.0019) for *P. pinaster* (Supplementary Results R6).

Hierarchical *F*-statistics are reported in Table 2(a). Genome-wide site and stand divergence was smaller than 0.01 for all species (Table 2(b); the divergence between *P. halepensis* stands SpainAlz and SpainMon, which are about 100 km apart, is in the range of other, geographically closer stand pairs); between-stands, within-site divergence was stronger than between-sites divergence for *P. pinaster* (contrary to expectation under a simple isolation-by-distance pattern), and as large as site divergence for *A. alba*. The distributions of within-stand, per-locus expected heterozygosity values

showed a moderate excess in all species, although with quite large variation among loci (Supplementary Results R6). Extent of heterozygote excess was correlated to minor allele frequency (MAF; Supplementary Results R6), with more negative inbreeding coefficient values for higher MAFs, suggesting that departure from Hardy-Weinberg equilibrium may be caused by some technical bias; heterozygote excess in high-throughput sequence data has been previously reported as partially, but not entirely, caused by paralogous alleles (Gayral et al., 2013), or as an effect of small sample sizes (Liu and Caselli, 2019). Notice, however, that we applied filters to remove paralogs.

(b) Inference of demographic parameters

For each species and stand, historical demographic parameters were estimated from the full set of SNP data. In each case, the estimated demographic history appeared to be consistent among stands of the same species. For two species, *A. alba* and *C. atlantica*, we detected a stable effective population size through time (at least for the last 100,000 years). For the remaining two, *P. halepensis and P. pinaster*, we detected population contraction that reduced it about 100-fold over time. The stairway plots for each species are reported in Figure 2.

(c) Detection of divergence outliers, trends in allele frequencies

BAMOVA

The BAMOVA framework allowed the estimation of posterior distributions for genome-wide (Supplementary Figure F1) and single-locus values of Φ-statistics. Genome-wide within-site (ΦSC) posterior distributions peaked at values ranging between 0.0216 and 0.0308 (Table 2(c)). Notice that the ranking of between-stands (within site) and among-sites divergence, as obtained with BAMOVA and hierfstat (Table 2(a,c)), is the same for the two pines, but differs for *C. atlantica*

(larger between-sites divergence in hierfstat, larger between-stands divergence in BAMOVA), and to a lesser degree in *A. alba* (about the same divergence in hierfstat, larger between-stands divergence in BAMOVA). Moreover, *F*-statistic estimates from hierfstat were in general about an order of magnitude smaller than Φ-statistics obtained in BAMOVA.

There were 37 within-site divergence outlier SNPs in *A. alba*, 22 in *C. atlantica*, 126 in *P. halepensis*, and 147 in *P. pinaster* (Supplementary Table T5(b)). Posterior probability distributions at outlier loci were largely shifted to the right relative to background levels of divergence (Figure 3(a); see Figure 3(b) for the distribution of Bayes Factors), with ΦSC mode values ranging between 0.0462 and 0.2832, approximately 1.5- to 10.5-fold larger than the mode of genome-wide Φ_{SC} for each species (Supplementary Table T5(b)).

Pcadapt

The most likely number of groups was three for *A. alba* and two for the other species; these structures were retained for the subsequent outlier search, which recovered 104, 18, 74 and 7 outliers for *A. alba*, *C. atlantica*, *P. halepensis* and *P. pinaster,* respectively (Figure 4, Supplementary Tables T4 and T5).

Synthesis

Supplementary Table T5 summarises the numbers of outliers found for each species and analysis. Over the four analyses, 137 outliers were found for *A. alba* (1.9% of all SNP analysed), 40 for *C. atlantica* (0.5%), 151 for *P. halepensis* (2.1%), and 196 for *P. pinaster* (2.3%). Of these outliers, four, none, three, and four were identified by both methods, respectively for *A. alba*, *C. atlantica*, *P. halepensis*, and *P. pinaster*. See Supplementary Figure F2 and Supplementary Results R7 for the analysis of overlap of outliers across analyses and for estimation of error rates in outlier detection

tests.

We checked whether, at outlier loci, allele frequency variation was of the same sign between sites within each species: that is, we checked whether, when one allele increased in allele frequency between the two environmental conditions (or decreased, for the alternative allele) at a given site, it did the same at the other site (or sites, for *A. alba*). Allele frequencies at outlier loci did not systematically show consistent trends across sites within species: in *A. alba*, 10 outliers out of 137 showed variation of the same sign between "high" and "low" stands at the three sites; in *C. atlantica*, 16 out of 40 outliers had variation of the same sign between "good site" and "bad site" stands at the two sites; in *P. halepensis*, 107 out of 196 outliers had variation of the same sign between "dry" and "moist" stands at the two sites; and finally, in *P. pinaster*, 106 out of 151 outliers had variation of the same sign between "shady" and "sunny" stands at the two sites (Figure 4, Supplementary Figure F3).

(d) Estimation of selection

Strength of disruptive selection was assessed by inspecting the "hyperposterior" probability distributions of the *Ns* composite parameter as obtained in ABC analysis on all BAMOVA outlier loci. Hyperposterior distributions were clearly informative for *A. alba* and *C. atlantica* and less so for *P. halepensis*; for *P. pinaster*, posteriors are very close to priors, raising doubts on the informativeness of the posterior (Supplementary Figure F4). For *A. alba*, the value of *Ns* at the mode varied between 0.17 and 45.33; for *C. atlantica*, between 0.29 and 49.01; for *P. halepensis* between 0.57 and 77.61; for *P. pinaster* between 0.23 and 77.98 (Supplementary Table T6(a); see Supplementary Table T6(b) for credible intervals of individual outlier loci). The distribution of modes of scaled selection coefficients varied across species (Figure 6), with the mode of modes

varying from 28.52 for *A. alba*, to 8.73 for *C. atlantica*, to around 2 for the two pine species. Yet as remarked above, the posteriors for the latter two species, and in particular for *P. pinaster*, are likely to be uninformative; therefore, the estimates for these two species are to be taken with caution. ABC cross-validation analyses showed that our procedure to estimate *Ns* was effective and unbiased, with prediction error varying between 0.12 and 0.23 and $R^2 \geq 0.5$ for regressions between true and estimated parameters (see Supplementary Results R8).

(e) Annotation

Orthology among contigs of different species was tested by matching their sequences to *P. taeda* probe sequences by blast. 4,367 probes (out of 13,535) correctly matched at least one contig for *A. alba*, 5,130 for *C. atlantica*, 6,597 for *P. halepensis*, and 5,558 for *P. pinaster* (Supplementary Table T1). The fact that more probes retrieved sequences in pines than in non-pine species suggests that there is some identification bias, with probes from *P. taeda* more easily capturing DNA fragments from closely-related pine species. In this view, probes that captured DNA fragments in non-pine species would correspond to more conserved genes, with the consequence that non-pine data would be enriched for conserved regions than pine data, thus introducing a bias. To test whether probes capturing fragments in non-pines corresponded to more conserved, and therefore less polymorphic, DNA fragments, we proceeded as follows. We split the probe list into those which only capture sequences in pines ("pine-specific") and those which capture sequence at least in one non-pine ("non pine-specific"); next, we identified, in the two pines, the contigs captured by the two groups of probes (637 and 2177 for *P. halepensis*; 578 and 2841 for *P. pinaster*). And finally, for each stand, we compared π values between sets of contigs derived from pine-specific and non-pine specific probes. The results were mixed: Aleppo pine comparisons were all non-

significant (Wilcoxon tests: Italy-H, W = 431482, p-value = 0.4377; Italy-L, W = 421698, p-value = 0.9366; Spain-Alz, W = 421935, p-value = 0.9232; Spain-Mon, W = 415192, p-value = 0.7014), while maritime pine comparisons were all marginally significant (Wilcoxon tests: 3-H, W = 573180, p-value = 0.01357; 3-S, W = 573083, p-value = 0.01378; 4-H, W = 574778, p-value = 0.01046; 4-S, W = 569114, p-value = 0.02545). So, there may be some effect of identification bias on patterns of genetic diversity, but we cannot conclude whether the higher levels of diversity observed in the two pines are related to this. However, in *P. pinaster*, "pine-specific" contigs are 4- to 5-fold fewer than the non-pine-specific ones, and therefore their impact on genome-wide diversity patterns may be relatively minor.

Orthology groups and the functional annotation of contigs for all species are reported in Supplementary Table T7 and Supplementary table T8, respectively.

Overall, there were 10,803 SNPs (27%) in non-coding regions, and 30,362 SNPs (73%) in coding DNA sequences (CDS); among the latter, 19,324 (64%) variants were non synonymous, and 11,038 (36%) variants were synonymous (Supplementary Table T9(a,d)). Outlier SNPs were represented by 205 SNPs (39%) in non-coding regions and 319(61%) SNPs in CDS; among the latter, 166 (52%) were non-synonymous, and 153 (48%) synonymous (Supplementary Table T9(b,c,d)). The distribution of variants in outliers and in the non-outlier set differed: with all species taken together, and for all types of variants, the outlier SNPs were proportionally more likely localized in non-coding regions ($\chi^2 = 44.8$, 1 d.f., p-value < 0.001) , and in the CDS, they were more likely synonymous ($\chi^2 = 18.3$, 1 d.f., P-value < 0.001) (Figure 7(a)). When the comparison was restricted to the non-synonymous SNP in CDS, there were proportionally slightly more SNPs with a low impact on the amino-acid change (i. e. variants that did not change the class (charged, plor, hydrophobic) of the aminoacid) in the outlier set than in the non-outlier set ($\chi^2 = 5.48$, 1 d.f., p-

value = 0.019) (Figure 7(b)). Few outliers were carried by contigs of different species, which belonged to the same orthology group. Respectively, one outlier of *A. alba* was found in a contig in the same orthology group as a contig of *P. halepensis* carrying three outliers; one outlier of *C. atlantica* was in a contig that matched a contig of *P. halepensis* carrying three outliers, and another *C. atlantica* outlier was in a contig matching another *P. halepensis* contig carrying four outliers; and finally, two *P. halepensis* outliers were in two contigs, that matched two *P. pinaster* contigs carrying one outlier each. However, none corresponded to the same position between contigs of different species.

GO term enrichment

Twelve GO-terms were enriched among the set of outlier contigs detected across the four species (see Supplementary Table T10). The most numerous GO-terms belong the molecular function class and may be categorised into three main activities: binding activity, catalytic activity and cytoskeletal motor activity (Supplementary Figure F5).

Discussion

We have detected similar patterns of divergence between environmentally contrasted forest stands for all the four species we have investigated. In particular all species show signal of selection at microgeographic scale, and display similar values and posterior distribution of population-scaled selection coefficients (Ns) at BAMOVA outlier loci. Taken together these results allow us to answer the four questions asked in the Introduction:

(1) *How common is microgeographic adaptive genetic divergence?* We have found signatures of genetic differentiation, putatively attributable to adaptive divergence, in all the four species we have studied. Given that they represent a variety of species, population histories, and ecological contexts, it seems highly likely that adaptive divergence is a pervasive phenomenon, at least in conifer tree species.

(2) *How strong is microgeographic selection?* Based on modelling and on our implementation of ABC analyses, we obtained central estimates of scaled selection coefficients at outlier loci between close to 0.1 and close to 80 (suggestive of strong selection). The scatter of individual-locus Ns values is large, and the central value of the distribution of all loci (the *mode of modes*) differed greatly across species, being about one order of magnitude larger in *A. alba* than in the two pines, with *C. atlantica* showing intermediate values. Thus, it would seem that the impact of selection is not equally distributed over loci, yet the distributions of values reveal our better capacity to detect loci undergoing strong selection than undergoing weak selection. For comparison, if the highest levels of selection estimated here were applied, under the form of hard selection (selective sweep), to a large, isolated diploid population of size N, the advantageous allele would be driven to fixation in a few tens of generations (Kimura, 1980). This provides a reference of how strong selection can be at the microgeographic scale for single loci. The differences among species are likely driven by

the observed differences in genome-wide background diversity structure, as captured by our analysis of historical demography, as the two pines show historically smaller effective sample sizes and historical bottlenecks, unlike fir and cedar (see also Budde et al., 2017; Olsson et al., 2021). It is worth noting that the recently introduced (and therefore recently diverged) *C. atlantica* populations show as large estimates of selection as longer established populations, suggesting that selection operated indeed not only over short distances, but also at short time scales (Saleh et al., 2022).

(3) *What is the genome-wide architecture of the response to microgeographic selection?* In all studied cases, we detected between 1.3% and 3.5% of the loci as being putatively under selection; while this is in absolute terms a relatively small portion of the genome, one has to consider that we developed a rather robust detection strategy and many loci remained undetected, in particular those under weak selection or involved in epistatic interactions. Furthermore, even though we have looked at as many (partially or entirely different) contrasts as there are stand pairs (and sites), and even though those contrasts may indeed represent a bundle of covarying factors, they presumably only represent a subset of the variety of selective agents that populations undergo. This suggests that, in general, relevant portions of the genome are indeed involved in response to selection. Recent theory on the role of genome size suggests that large genomes, as conifers', have more potential for adaptive variation in non-coding regions than small genomes (Mei et al., 2018) and that the difference in mutational target can affect the expected dynamics of adaptation (Höllinger et al., 2019). Assuming that mutation rates are more or less uniform across species, species with larger genomes will be subject to more mutations per genome in each generation. Considering that variation in gene number across plant species is not substantial, most of the additional mutational input in larger genomes is expected to occur outside of coding sequence (Mei et al., 2018). Our results showed that outliers were enriched in synonymous, non-coding variants and amino acid

changes with low structural impact. This suggests that structural variation may not be the main target of selection at microgeographic scale, or that polymorphisms at non-synonymous sites mostly undergo background selection and are maintained by drift, without contributing to local adaptation (and rather being a component of genetic load; González-Martínez et al., 2017). However, an analysis of variance (not shown) testing for differences in the estimated intensity of selection among variant classes did not detect any significant effect. Despite the known caveats of the GO enrichment analysis (Gaudet & Dessimoz, 2017), this approach remains an efficient tool for summarising functions among large quantities of genes. The GO terms enriched in this study were also enriched in previous plant studies in response to biotic and abiotic variables. Remarkably, one study on a tree species, *Populus trichocarpa* (Evans et al., 2014), found four GO terms (hydrolase activity acting on acid anhydrides, GO:0016817; nucleobase-containing compound kinase activity, GO:0019205; organic cyclic compound binding, GO:0097159 and heterocyclic compound binding, GO:1901363), out of the twelve reported here, as being associated with adaptive traits (mostly, growth rates and leaf phenology) . Organic cyclic compound binding and heterocyclic compound binding were also found in several studies investigating plant responses to abiotic stresses (Zhou et al., 2016; Liu et al., 2017; Alves da Silva et al., 2019; Safdarian et al., 2019, Zhang et al. 2020).

(4) *How much overlap is there among species in the loci undergoing divergent selection?* Only seventeen outliers are located in orthologous sequences shared by pairs of species (but do not correspond to the same positions). Within species, trends in the variation of allele frequencies were only partially consistent between sites, which suggests a genetic background impact on allele effects due to complex epistatic interactions. The absolute numbers of shared detected loci putatively under selection is very small. This is consistent with the choice of species and sites, which deliberately aimed at diversifying the types of environmental contrasts and, consequently, the genetic bases of

adaptive response.

In summary, we showed that the effects of microgeographic selection are likely pervasive; that even very recent microgeographic selection leaves a detectable genomic signature; that selection apparently involves a small fraction of the genome (even if it may be several times larger than the fraction we have detected, due to lack of power); that selection can be intense at many loci; and that it targets a variety of genomic regions, depending on the species and the type of environmental contrast. One consequence of this finding is that microgeographic divergence appears to contribute to maintaining genetic diversity in populations through the existence of multiple optima (Delph & Kelly, 2014), in a kind of self-sustained dynamic process. This finding reinforces the evidence of the strength of microgeographic adaptation since, in a hierarchical environmental heterogeneity pattern, when many QTL are involved in an adaptive trait, microgeographic adaptation is expected to rely mainly on the covariances among QTL and strong selection intensity is required to affect individual QTL-frequencies as we studied here (Cubry et al, 2022). The estimated values of scaled selection partially match theoretical expectations for the intensity of selection needed to maintain divergence with gene flow, as derived by Yeaman & Whitlock (2011) and Yeaman & Otto (2011), and to which we can attempt a simple comparison as follows. Let us consider a genome-wide divergence between stands of $F_{ST} = 0.003$, similar to the values we observed, leading to an approximate estimate of $Nm = 83$ (applying island-model equations). Let us also consider a scaled selection coefficient around $Ns = 20$, $Ns = 8$, and $Ns = 0.2$, close to the middle of the distribution of the most-likely values respectively for *A. alba*, *C. atlantica*, and the two pines. For such Ns values, the ratio $Nm/Ns = m/s$ would be respectively around 4, 10, and 40. If we plot the straight lines corresponding to $m = 4s$ and $m = 10s$ onto figure 1(b) of Yeaman & Otto (2011; reproduced here with permission as Supplementary Figure F6), the values are close to the threshold for the

maintenance of polymorphism in large populations. The situation is different for m = 40s (corresponding to the two pines), for which the values would be far from the conditions of maintenance of polymorphism (and all the more so because estimated population sizes are small; yet estimates of Ns can be considerably larger than the central values and be compatible with polymorphism). The fact that the two pines do not match the theoretical expectation contrasts with the fact that they showed as many divergence outliers as fir and cedar, and with previous results showing local adaptation in the very same stands (Ruiz-Daniels et al., 2019). It is possible that the peculiarities of population dynamics of pioneer species makes it difficult to correctly estimate selection parameters (see below).

The demographic analyses presented here focused on the particular stands (sites) under scrutiny, and were instrumental to the identification of signatures of selection while taking into account the demographic background, and are not therefore meant to represent species-level processes. However, some trends worth noticing emerge: *A. alba* and *C. atlantica* stands were stable over time, possibly reflecting the presence of glacial refugia in the sampled areas: in Southern France for *A. alba*, were the presence of a secondary refugium has been hypothesised (Liepelt et al. 2009); and in the area of origin of the planted populations - Algeria - for *C. atlantica*, which was a refugial zone (Terrab et al. 2008), while the contraction signal observed for *P. halepensis* confirms the results obtained by Grivet et al. (2009) for the Western part of the species range; on the contrary, the decline observed for *P. pinaster* stands, located in Eastern Spain, is possibly caused by local changes in environmental conditions, including the effect of anthropic action over the course of the Holocene; indeed, the local population contraction observed for *P. pinaster* disagrees with the range-level population expansion detected by Bucci et al. (2007).

Some limitations of the methods may affect the interpretation of our results on the identification of

divergence outliers and on the estimation of selection intensity. For instance, the two outlier search methods rest on different assumptions and, consequently, may detect different types of differentiation patterns (Rellstab et al., 2015), thus explaining the little overlap among methods (and supporting the idea of keeping the union of all outlier sets, and not their intersection). In addition to this, if the fraction of the genome undergoing selection is small, as observed here, then, even with satisfactory Type I error rates and intermediate Type II error rates, there will be many false positives among the outliers, because the number of true negatives far exceed those of true positives. It may be harder to detect loci undergoing weak selection, which may explain why we apparently detect few loci with weak estimated selection coefficients. Differences in the biology and demography of the four species may, in addition, drive differences in selection intensity estimates. Pioneer species like the two pines may undergo fluctuations both in population sizes and selection regime, offsetting stands from potential optima and therefore blurring divergence and local adaptation, reducing the potential to detect selection. Biases due to the sequence capture method we used may introduce differences in patterns of genetic diversity between species. Finally, particularly for *C. atlantica*, where fewer outliers were detected than for the three remaining species, it is possible that the small number of generations elapsed since the start of the divergence in these artificial stands was insufficient for sizeable divergence to arise, even under moderately strong selection.

To conclude, our study provides support to the view that microgeographic selection is common in conifer tree populations, suggesting that adaptation to habitat variation can take place at very small spatial scales, rapidly. Our results represent an important step towards the identification of processes (i.e. functions involved in adaptation and population-genetic processes allowing

maintenance of polymorphism), such as strong natural selection, high adaptive differentiation, and strength and stability of selection that may allow tree populations to adapt to spatially varying environmental conditions, and thus maintain their adaptive potential – a much-needed resource in times of an unpredictably changing climate.

Acknowledgements

This work was supported by the French national research agency (ANR) under the FLAG project (ANR-12-ADAP-0007-01), as well as by the French Ministry of Ecology's GICC GRAAL project (10-MCGOT-GICC-8-CVS-139), by the European Union-funded TOPWOOD project (MSCA-RISE-2014-645654-TOPWOOD), and by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 676876 (GenTree). We wish to thank the INRAE-UEFM team for sampling fir and cedar, and three anonymous reviewers for their comments. The genomic analyses were performed by RapidGenomics, Gainseville, FL.

References

Alexa, A., & Rahnenfuhrer, J. (2019). topGO: Enrichment Analysis for Gene Ontology. R package version 2.36.0

Alves da Silva, D., Tsai, S. M., Chiorato, A. F., da Silva Andrade, S. C., de Fatima Esteves, J. A., Recchia, G. H., & Morais Carbonell, S. A. (2019). Analysis of the common bean (*Phaseolus vulgaris* L.) transcriptome regarding efficiency of phosphorus use. *PLoS One, 14*, 1-27. https://doi.org/10.1371/journal.pone.0210428

Antonovics, J. (1968). Evolution in closely adjacent plant populations VI. Manifold effects of gene flow. *Heredity, 23*, 508-524. https://doi.org/10.1038/hdy.1968.70

Antonovics, J. (2006). Evolution in closely adjacent plant populations X: long-term persistence of prereproductive isolation at a mine boundary. *Heredity, 97*, 33-7. https://doi.org/10.1038/sj.hdy.6800835

Audigeos, D., Brousseau, L., Traissac, S., Scotti-Saintagne, C., & Scotti, I. (2013). Molecular divergence in tropical tree populations occupying environmental mosaics. *Journal of Evolutionary Biology, 26*, 529-544. https://doi.org/10.1111/jeb.12069

Babik, W., Butlin, R.K., Baker, W.J., Papadopulos, A.S.T., Boulesteix, M., Anstett, M-C, Lexer, C., Hutton, I., & Savolainen, V. (2009). How sympatric is speciation in the Howea palms of Lord Howe Island? *Molecular Ecology, 18*, 3629-3638. https://doi.org/10.1111/j.1365-294X.2009.04306.x

Beaumont, M. A., Zhang, W., & Balding, D.J. (2002). Approximate Bayesian Computation in population genetics. *Genetics, 162*, 2025-2035. https://doi.org/10.1093/genetics/162.4.2025

Boyle, E. A., Li, Y. I., & Pritchard J. K. (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell, 169*, 1177- 1186. https://doi.org/10.1016/j.cell.2017.05.038

Bradshaw, A. D. (1960). Population differentiation in *Agrostis tenuis* Sibth. III. Populations in varied environments. *New Phytologist, 59*, 92-103. https://doi.org/10.1111/j.1469-8137.1960.tb06206.x

Brousseau, L., Bonal, D., Cigna, J., & Scotti, I. (2013). Highly local environmental variability promotes intrapopulation divergence of quantitative traits: An example from tropical rain forest trees. *Annals of Botany, 112*, 1169-1179. https://doi.org/10.1093/aob/mct176

Brousseau, L., Foll, M., Scotti-Saintagne, C., & Scotti, I. (2015). Neutral and adaptive drivers of microgeographic genetic divergence within continuous populations: The case of the neotropical tree *Eperua falcata* (aubl.). *PLoS ONE, 10(4)*, e0127134. https://doi.org/10.1371/journal.pone.0127134

Brousseau, L., Postolache, D,. Lascoux, M., Drouzas, A. D., Källman, T., Leonarduzzi, C., Liepelt, S., Piotti, A., Popescu, F., Roschanski, A. M., Zhelev, P., Fady, B., & Vendramin, G. G. (2016). Local adaptation in European firs assessed through extensive sampling across altitudinal gradients in Southern Europe. *PLoS ONE, 11*, e0158216. https://doi.org/10.1371/journal.pone.0158216

Bucci, G., González-Martínez, S. C., le Provost, G., Plomion, C., Ribeiro, M. M., Sebastiani, F., Alía, R., & Vendramin, G. G. (2007). Range-wide phylogeography and gene zones in *Pinus*

pinaster Ait. revealed by chloroplast microsatellite markers. *Molecular Ecology*, *16*, 2137-2153. https://doi.org/10.1111/j.1365-294X.2007.03275.x

Budde, K. B., González-Martínez, S. C., Navascués, M., Burgarella, C., Mosca, E., Lorenzo, Z., Zabal-Aguirre, M., Vendramin, G. G., Verdú, M., Pausas, J. G., & Heuertz, M. (2017**).** Increased fire frequency promotes stronger spatial genetic structure and natural selection at regional and local scales in *Pinus halepensis* Mill. *Annals of Botany, 119*, 1061-1072. https://doi.org/10.1093/aob/mcw286

Bulmer, M.G. (1972). The genetic variability of polygenic characters under optimizing selection, mutation and drift. *Genetical Research, 19*, 17-25. https://doi:10.1017/S0016672300014221

Carvajal-Rodríguez, A. (2008). GENOMEPOP: a program to simulate genomes in populations. *BMC Bioinformatics, 9*, 223. https://doi.org/10.1186/1471-2105-9-223

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research, 70*, 155-174. https:// doi:10.1017/S0016672397002954

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., & Ruden, D.M. (2012**).** A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3.". *Fly, 6*, 80-92. https://doi.org/10.4161/fly.19695

Conway, J. R., Lex, A., & Gehlenborg, A. N. (2017)**.** UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics, 33*, 2938-2940. https://doi.org/10.1093/bioinformatics/btx364

Cubry, P., Oddou-Muratorio, S., Scotti, I., & Lefèvre, F. (2022). Interactions between micro-environment, selection and genetic architecture drive multiscale adaptation in a simulation experiment. *Journal of Evolutionary Biology*, *35*, 451-466. https://doi.org/10.1111/jeb.13988

De Meeûs, T., & Goudet, J. (2007)**.** A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. *Infection, Genetics and Evolution, 7*, 731-735. https://doi.org/10.1016/j.meegid.2007.07.005

Delph, L. F., & Kelly, J. K. (2014). On the importance of balancing selection in plants. *New Phytologist, 201*, 45-56. https://doi.org/10.1111/nph.12441

Duforet-Frebourg, N., Bazin, E., & Blum, M. G. B. (2014). Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Molecular Biology and Evolution, 31*, 2483-2495. https://doi.org/10.1093/molbev/msu182

Eckert, A.J., Maloney, P.E., Vogler, D.R., Jensen, C.E., Mix, A.D., & Neale, D.B. (2015). Local adaptation at fine spatial scales: an example from sugar pine (*Pinus lambertiana*, Pinaceae). *Tree Genetics & Genomes, 11*, 42. https://doi.org/10.1007/s11295-015-0863-0

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*, 1-14.

Evans, L., Slavov, G., Rodgers-Melnick, E., Martin, J., Ranjan, P., Muchero, W., Brunner, A. M., Schackwitz, W., Gunter, L., Chen, J. G., Tuskan, G. A., & DiFazio, S. P. (2014). Population

genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nature Genetics*, *46*, 1089-1096. https://doi.org/10.1186/s13059-015-0721-2

Frachon, L., Bartoli, C., Carrère, S., Bouchez, O., Chaubet, A., Gautier, M., Roby, D., & Roux, F. (2018). A genomic map of climate adaptation in *Arabidopsis thaliana* at a micro-geographic scale. *Frontiers in Plant Science, 9*: 967. https://doi: 10.3389/fpls.2018.00967.

Gaudet, P., & Dessimoz, C. (2017). Gene Ontology: Pitfalls, Biases, and Remedies. In: Dessimoz C, Škunca N (eds) The gene ontology handbook. Methods in molecular biology, vol 1446, Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-3743-1_14

Gauzere, J., Klein, E. K., Brendel, O., Davi, H., & Oddou-Muratorio, S. (2020). Microgeographic adaptation and the effect of pollen flow on the adaptive potential of a temperate tree species. *New Phytologist, 227*, 641-653. https://doi.org/10.1111/nph.16537

Gayral, P., Melo-Ferreira, J., Glémin, S., Bierne, N., Carneiro, M., Nabholz, B., Lourenco, J. M., Alves, P. C., Ballenghien, M., Faivre, N., et al. (2013). Reference-Free Population Genomics from Next-Generation Transcriptome Data and the Vertebrate-Invertebrate Gap. *PLOS Genetics, 9*, e1003457. https://doi.org/10.1371/journal.pgen.1003457

Gompert, Z., & Buerkle, C. A. (2011). A hierarchical Bayesian model for next-generation population genomics. *Genetics, 187*, 903-17. https://doi.org/10.1534/genetics.110.124693

González-Martínez, S. C., Ridout, K., & Pannell, J. R. (2017). Range Expansion Compromises Adaptive Evolution in an Outcrossing Plant. *Current Biology, 27*, 2544-2551.e4. https://doi.org/10.1016/j.cub.2017.07.007

Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical *F*-statistics. *Molecular Ecology Notes, 5*, 184-186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Grivet, D., Sebastiani, F., González-Martínez, S. C., & Vendramin, G. G. (2009). Patterns of polymorphism resulting from long-range colonization in the Mediterranean conifer Aleppo pine. *New Phytologist*, *184*, 1016-1028. https://doi.org/10.1111/j.1469-8137.2009.03015.x

Hallingbäck, H. R., Burton, V., Vizcaíno-palomar, N., Trotter, F., Liziniewicz, M., Marchi, M., Berlin, M., Ray, D., & Benito-Garzón, M. (2021). Managing uncertainty in Scots pine range-wide adaptation under climate change. *Frontiers in Ecology and Evolution, 9,* 1-18. https://doi.org/10.3389/fevo.2021.724051

Hamrick, J. L., & Godt, M. J. W. (1990). Allozyme diversity in plant species. In A. H. D. Brown, M.T. Clegg, A. Kahler, & B. Weir (Eds.), *Plant Population Genetics, Breeding, and Genetic Resources* (pp. 43-63). Sinauer Associates Inc.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A., & Whitlock, M. C. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist, 188*, 379-397. https:// doi/10.1086/688018

Höllinger, I., Pennings, P. S., & Hermisson, J. (2019). Polygenic adaptation: From sweeps to subtle frequency shifts. *PLoS Genetics*, *15*(3), 1-26. https://doi.org/10.1371/journal.pgen.1008035

Hrivnák, M., Paule, L., Krajmerová, D., Kulaç, Ş., Şevik, H., Turna, İ., Tvauri, I., & Gömöry, D. (2017). Genetic variation in Tertiary relics: The case of eastern-Mediterranean *Abies* (Pinaceae). *Ecology and Evolution*, *7*, 10018-10030. https://doi.org/10.1002/ece3.3519

Jaramillo-Correa, J. P., Bagnoli, F., Grivet, D., Fady, B., Aravanopoulos, F. A., Vendramin, G. G., & González-Martínez, S. C. (2020). Evolutionary rate and genetic load in an emblematic Mediterranean tree following an ancient and prolonged population collapse. *Molecular Ecology, 29*, 4797-4811. https://doi.org/10.1111/mec.15684

Jia, Y., Zhu, J., Wu, Y., Fan, W. B., Zhao, G. F, & Li, Z. H. (2018) Effects of geological and environmental events on the diversity and genetic divergence of four closely related pines: *Pinus koraiensis*, *P. armandii*, *P. griffithii*, and *P. pumila*. *Frontiers in Plant Science, 9*, 1264. https://doi.org/10.3389/fpls.2018.01264

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* 90: 773-795. https://doi.org/10.1080/01621459.1995.10476572

Kimura, M. (1980**)**. Average time until fixation of a mutant allele in a finite population under continued mutation pressure: Studies by analytical, numerical, and pseudo-sampling methods. *Proceedings of the National Academy of Science of USA, 77*, 522-526. https://doi.org/10.1073%2Fpnas.77.1.522

Knaus, B. J., & Grünwald, N. J. (2017**)**. vcfr : a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources, 17*, 44-53. https://doi.org/10.1111/1755-0998.12549

Kremer, A., Ronce ,O., Robledo-Arnuncio, J. J., Guillaume, F., Bohrer, G., Nathan, R., Bridle, J. R., Gomulkiewicz, R., Klein, E. K., Ritland, K., Kuparinen, A., Gerber, S., & Schueler S., (2012). Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters, 15,* 378-392. https://doi.org/10.1111%2Fj.1461-0248.2012.01746.x

Kuznetsova, I., Lugmayr, A., Siira, S. J., Rackham, O., & Filipovska, A. (2019). CirGO: An alternative circular way of visualising gene ontology terms. *BMC Bioinformatics, 20*(1), 1–7. https://doi.org/10.1186/s12859-019-2671-2

Langin, K. M., Sillett, T. S., Funk, W. C., Morrison, S. A., Desrosiers, M. A., & Ghalambor, C. K. (2015**)**. Islands within an island: Repeated adaptive divergence in a single population. *Evolution, 69*, 653-665. https://doi.org/10.1111/evo.12610

Lefèvre, F., Fady, B., Fallour-Rubio, D., Ghosn, D., & Bariteau M. (2004). Impact of founder population, drift and selection on the genetic diversity of a recently translocated tree population. *Heredity, 93,* 542-550. https://doi.org/10.1038/sj.hdy.6800549

Liepelt, S., Cheddadi, R., de Beaulieu, J.-L., Fady, B., Gömöry, D., Hussendörfer, E., Konnert, M., Litt, T., Longauer, R., Terhürne-Berson, R., & Ziegenhagen, B. (2009). Postglacial range expansion and its genetic imprints in *Abies alba* (Mill.) — A synthesis from palaeobotanic and genetic data. *Review of Palaeobotany and Palynology*, *153*, 139-149. http://dx.doi.org/10.1016/j.revpalbo.2008.07.007

Lind, B. M., Friedline, C. J., Wegrzyn, J. L., Maloney, P. E., Vogler, D. R., Neale, D. B., & Eckert,

A. J. (2017) Water availability drives signatures of local adaptation in whitebark pine (*Pinus albicaulis* Englm.) across fine spatial scales of the Lake Tahoe Basin, USA. *Molecular Ecology, 26*, 3168-3185. https://doi.org/10.1111/mec.14106

Liu, L., & Caselli, R. J. (2019). Unbalanced Sample Size Introduces Spurious Correlations to Genome-Wide Heterozygosity Analyses. *Human Heredity*, 84:197-202. https://doi.org/10.1159/000507576

Liu, X., & Fu, Y. X. (2020). Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biology., 21*, 280. https://doi.org/10.1186/s13059-020-02196-9

Liu, J., Wang, Y., & Li, Q. (2017). Analysis of differentially expressed genes and adaptive mechanisms of *Prunus triloba* Lindl. under alkaline stress. *Hereditas, 154*, 10. https://doi.org/10.1186/s41065-017-0031-7

Lobo, A., Hansen, J. K., Hansen, L. N., & Dahl Kjaer E. (2018). Differences among six woody perennials native to Northern Europe in their level of genetic differentiation and adaptive potential at fine local scale. *Ecology and Evolution, 8*, 2231-2239. https://doi.org/10.1002%2Fece3.3824

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology, 24*, 1031-1046. https://doi.org/10.1111/mec.13100

Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources, 17*, 67-77. https://doi.org/10.1111/1755-0998.12592

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., Lüdecke, D. (2019). Indices of Effect Existence and Significance in the Bayesian Framework. *Frontiers in Psychology, 10,* 2767. https://doi.org/10.3389/fpsyg.2019.02767

Mei, W., Stetter, M. G., Gates, D. J., Stitzer, M. C., & Ross-Ibarra, J. (2018). Adaptation in plant genomes: Bigger is different. *American Journal of Botany, 105(1),* 16-19. https://doi.org/10.1002/ajb2.1002

Mosca, E., Gugerli, F., Eckert, A. J., & Neale, D. B. (2016). Signatures of natural selection on *Pinus cembra* and *P. mugo* along elevational gradients in the Alps. *Tree Genetics & Genomes, 12*, 9. https://www.dora.lib4ri.ch/wsl/islandora/object/wsl%3A5492

Nadeau, N. J., & Jiggins, C. D. (2010). A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends in Genetics, 26*, 484-492. https://doi.org/10.1016/j.tig.2010.08.004

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York Chichester, West Sussex: Columbia University Press. https://doi.org/10.7312/nei-92038

Neuenschwander, S., Hospital, F., Guillaume, F., & Goudet, J. (2008). QUANTINEMO: An individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics, 24*, 1552-1553. https://doi.org/10.1093/bioinformatics/btn219

Olsson, S., Lorenzo, Z., Zabal-Aguirre, M., Piotti, A., Vendramin, G. G., González-Martínez, S. C., & Grivet, D. (2021). Evolutionary history of the mediterranean *Pinus halepensis-brutia* species complex using gene-resequencing and transcriptomic approaches. *Plant Molecular Biology, 106*, 367-380. https://doi.org/10.1007/s11103-021-01155-7

Peñalba, J. V., Smith, L. L., Tonione, M., Sass, C., Hykin, S. M., Skipwith, P. L., McGuire, J., Bowie, R. C. K., & Moritz, C. (2014). Sequence capture using PCR-generated probes: a cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Molecular Ecology Resources, 14*, 1000-1010. https://doi.org/10.1111/1755-0998.12249

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology, 24*, 4348-4370. https://doi.org/10.1111/mec.13322

Richardson, J. L., Urban, M. C., Bolnick, D. I., & Skelly, D. K. (2014). Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology & Evolution, 29*, 165-176. https://doi.org/10.1016/j.tree.2014.01.002

Roschanski, A. M., Csilléry, K., Liepelt, S., Oddou-Muratorio, S., Ziegenhagen, B., Huard, F., Ullrich, K. K., Postolache, D., Vendramin, G. G., & Fady, B. (2016). Evidence of divergent selection for drought and cold tolerance at landscape and local scales in *Abies alba* Mill. in the French Mediterranean Alps. *Molecular Ecology, 25*, 776-794. https://doi.org/10.1111/mec.13516

Ruiz Daniels, R., Taylor, R.S., Martínez, S.C.G., Vendramin, G.G., Fady, B., Oddou-Muratorio, S., Piotti, A., Simioni, G., Grivet, D., & Beaumont, M.A. (2019). Looking for local adaptation: Convergent microevolution in aleppo pine (*Pinus halepensis*). *Genes, 10*: 673. https://doi.org/10.3390%2Fgenes10090673

Safdarian, M., Askari, H., Shariati J. V., & Nematzadeh, G. (2019). Transcriptional responses of wheat roots inoculated with *Arthrobacter nitroguajacolicus* to salt stress. *Scientific Reports, 9*, 1-13. https://doi.org/10.1038/s41598-018-38398-2

Saleh, D., Chen, J., Leplé, J-C., Leroy, T., Truffaut, L., Dencausse, B., Lalanne. C., Labadie, K., Lesur, I., Bert, D., Lagane, F., Morneau, F., Aury, J. M., Plomion, C., Lascoux, M., & Kremer, A. (2022). Genome-wide evolutionary response of European oaks since the Anthropocene. *Evolution Letters*, *6(1),* 4-20. https://doi.org/10.1002/evl3.269

Savolainen, V., Anstett, M-C., Lexer, C., Hutton, I., Clarkson, J. J., Norup, M. V., Powell, M. P., Springate, D., Salamin, N., & Baker, W. J. (2006). Sympatric speciation in palms on an oceanic island. *Nature, 441*, 210-213. https://doi.org/10.1038/nature04566

Scotti-Saintagne, C., Boivin, T., Suez, M., Musch, B., Scotti, I., & Fady, B. (2021). Signature of mid-Pleistocene lineages in the European silver fir (*Abies alba* Mill.) at its geographic distribution margin. *Ecology and Evolution*, *11*, 10984-10999. https://doi.org/10.1002/ece3.7886

Sexton, J. P., Hangartner, S. B. & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution, 68*, 1-15. https://doi.org/10.1111/evo.12258

Storey, J. D., Bass, A. J., Dabney, A., & Robinson, D. (2015). qvalue: Q-value estimation for false discovery rate control. R package version 2.0. http://github.com/jdstorey/qvalue.

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE, 6*(7):e21800. https://doi.org/10.1371/journal.pone.0021800

Tenaillon, O., Rodríguez-Verdugo, A., Gaut, R.L., McDonald, P., Bennett, A.F., Long, A.D., Gaut, B.S. (2012). The molecular diversity of adaptive convergence. *Science, 335,* 457-461. https://doi.org/10.1126/science.1212986

Terrab, A., Hampe, A., Lepais, O., Talavera, S., Vela, E., & Stuessy, T. F. (2008). Phylogeography of North African Atlas cedar (*Cedrus atlantica*, Pinaceae): Combined molecular and fossil data reveal a complex Quaternary history. *American Journal of Botany*, *95*, 1262-1269. https://doi.org/10.3732/ajb.0800010

Turner, T. L., Bourne, E. C., von Wettberg, E. J., Hu, T. T., & Nuzhdin, S.V. (2010). Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics*, *42*(3), 260-263. https://doi.org/10.1038/ng.515

Vendramin, G.G., Fady, B., González-Martínez, S.C., Hu, F.S., Scotti, I., Sebastiani, F., Soto, A., & Petit, R.J. (2008). Genetically depauperate but widespread: the case of an emblematic Mediterranean pine. *Evolution, 62*, 680-688. https://doi.org/10.1111/evo.2008.62.issue-3

Vendramin, G.G., Avanzi, C., González-Martínez, S.C., & Grivet, D. (2021). Population Genetics and Genomics of Aleppo Pine (*Pinus halepensis*). In G. Ne'eman, & Y. Osem (Eds), *Pines and their mixed forest ecosystems in the Mediterranean Basin* (pp. 83-99). Springer Nature Switzerland..

Wang, T., O'Neill, G.A., & Aitken, S.N. (2010). Integrating environmental and genetic effects to predict responses of tree populations to climate. *Ecological Applications, 20*(1), 153-63. https://doi.org/10.1890/08-2257.1

Willyard, A., Syring, J., Gernandt, D. S., Liston, A., & Cronn, R. (2007). Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for Pinus. *Molecular Biology and Evolution*, *24*, 90-101. https://doi.org/10.1093/molbev/msl131

Yang, R-C. (1998). Estimating hierarchical *F*-statistics. *Evolution, 52*, 950-956. https://doi.org/10.1111/j.1558-5646.1998.tb01824.x

Yeaman, S., & Whitlock, M.C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution, 65*, 1897-911. https://doi.org/10.1111/j.1558-5646.2011.01269.x

Yeaman, S., & Otto, S.P. (2011). Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution, 65*, 2123-2129. https://doi.org/10.1111/j.1558-5646.2011.01277.x

Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C., Nurkowski, K. A., Smets, P., Wang, T., Gray, L. K., Liepe, K. J., Hamann, A., Holliday, J. A., Whitlock, M. C., Rieseberg, L. H., & Aitken, S. N. (2016). Convergent local adaptation to climate in distantly related conifers. *Science*, *353*(6306):1431-1433. https://doi.org/10.1126/science.aaf7812

Zhang, X., Zhuang, L., Liu, Y., Yang, Z., & Huang, B. (2020). Protein phosphorylation associated with drought priming-enhanced heat tolerance in a temperate grass species. *Horticulture Research, 7*, 207 . https://doi.org/10.1038/s41438-020-00440-8

Zhou, B., Zhang, L., Ullah, A., Jin, X., Yang, X., & Zhang, X. (2016). Identification of multiple stress responsive genes by sequencing a normalized cDNA library from sea-land cotton (*Gossypium barbadense* L.). *PLoS One, 11*, 1-22. https://doi.org/10.1371/journal.pone.0152927

Data accessibility

Raw data are accessible at the European Nucleotide Archive, Study Id: PRJEB35366 (ERP118402).

VCF files used to run all population analyses are permanently stored at https://entrepot.recherche.data.gouv.fr/ with doi: https://doi.org/10.15454/R5VV8Z.

Author contributions

IS, SOM, DG, FL, BF, SGCM, CP, PR and GGV conceived the study; RRD, DG, FL, PC, AR, FB, VG, and GGV collected tissue samples, obtained DNA samples, produced simulations; IS, HL, SOM, CSS, RRD, DG, FL, PC, ILK, and FB analysed data; all Authors contributed to writing the manuscript.

FIGURES

Figure 1 A: Sampling sites (solid circles) and distribution ranges (highlighted as coloured areas; source: EUFORGEN www.euforgen.org) of *Abies alba* (green), *Cedrus atlantica* (red), *Pinus halepensis* (yellow) and *Pinus pinaster* (light blue); B: geographic localization of the study area.

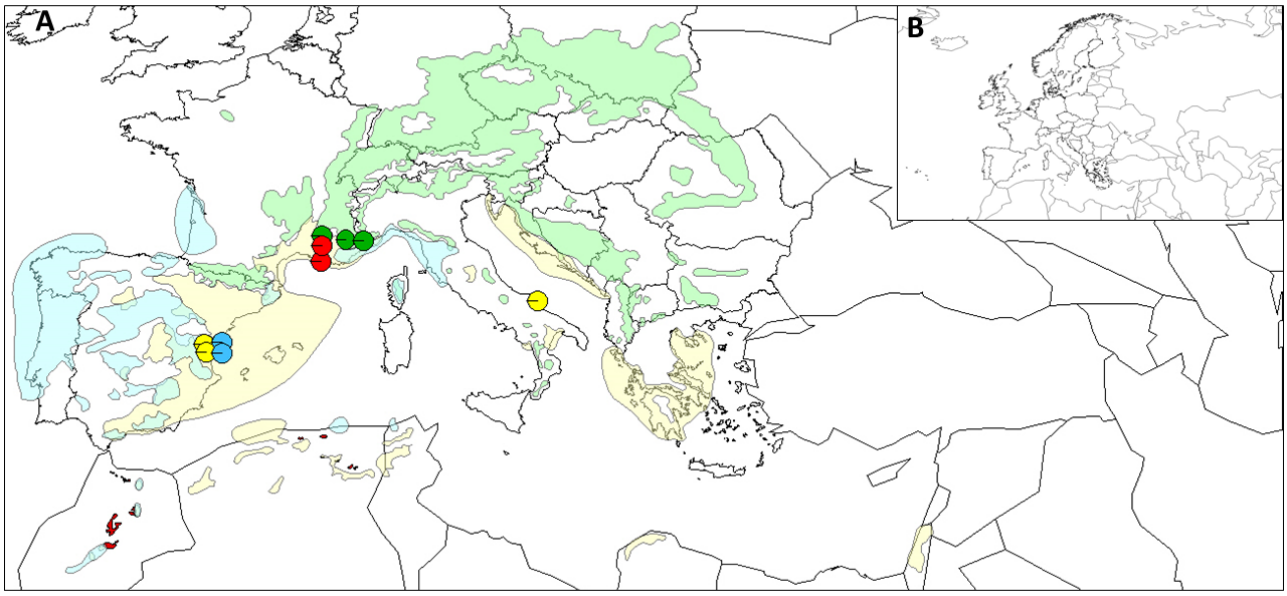
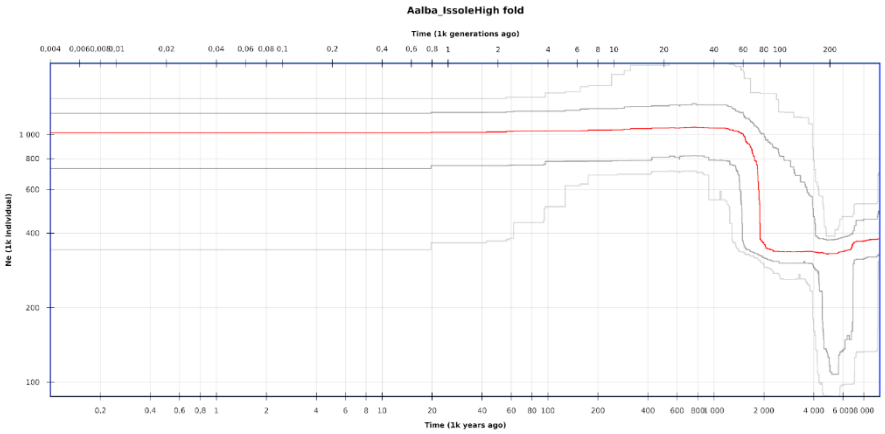
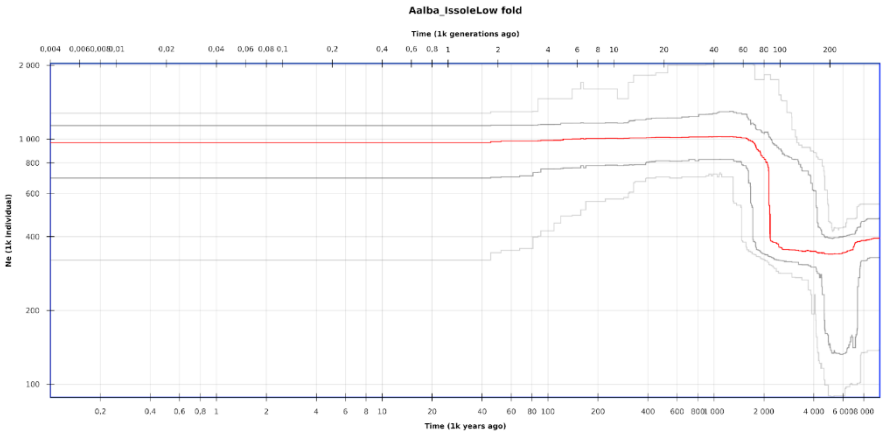
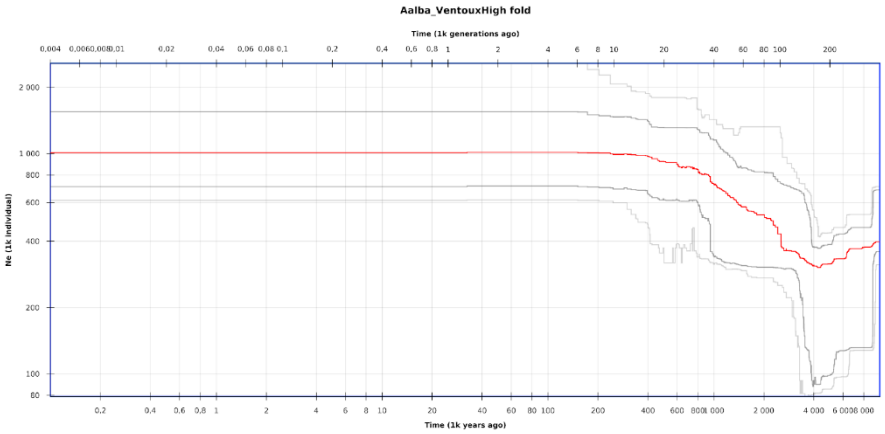


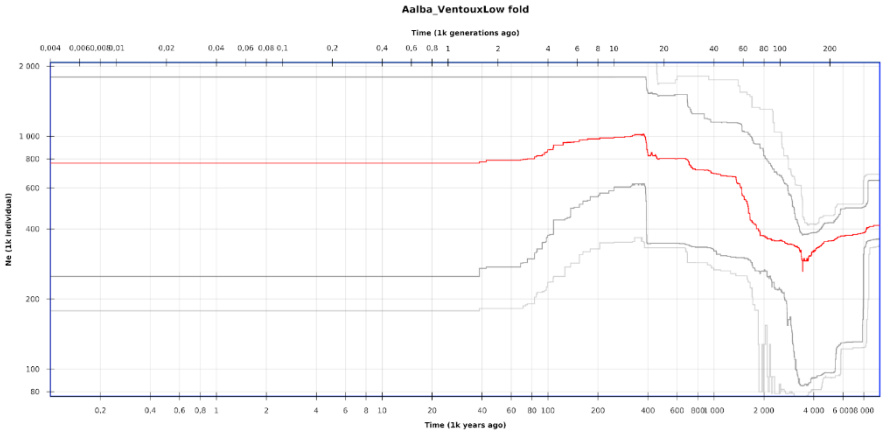
Figure 2. Stairway Plot analysis of the demographic history of stands of (a) *Abies alba*, (b) *Cedrus atlantica*, (c) *Pinus halepensis*, (d) *Pinus pinaster*. Stand names as indicated in plot titles.

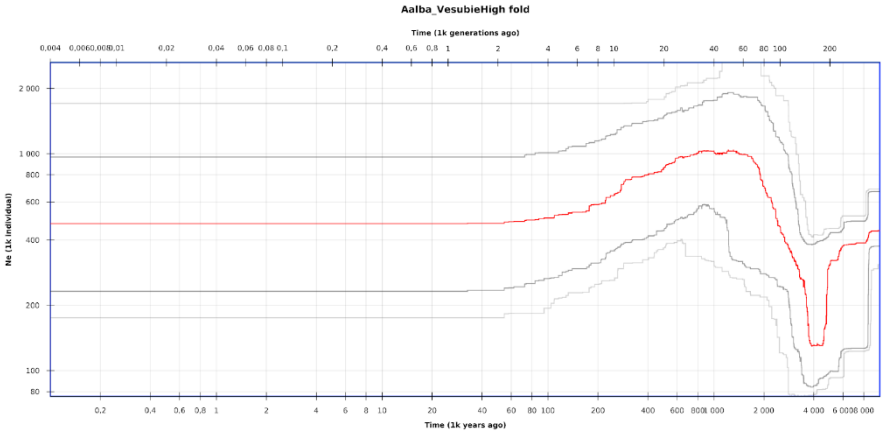
(a)

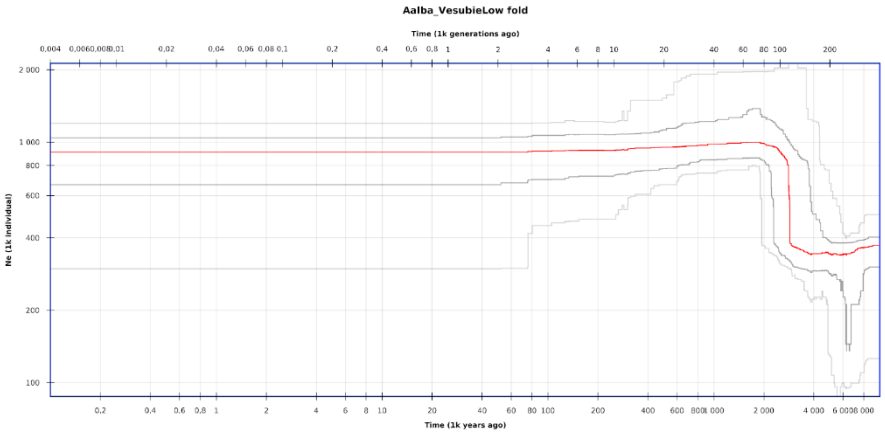




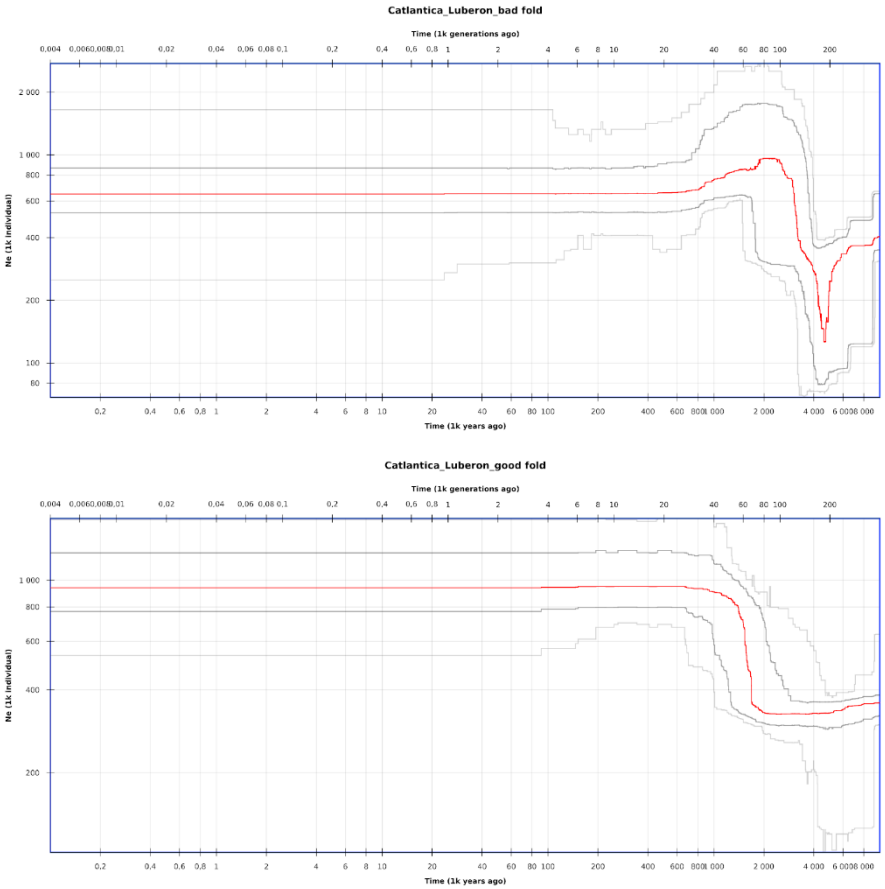


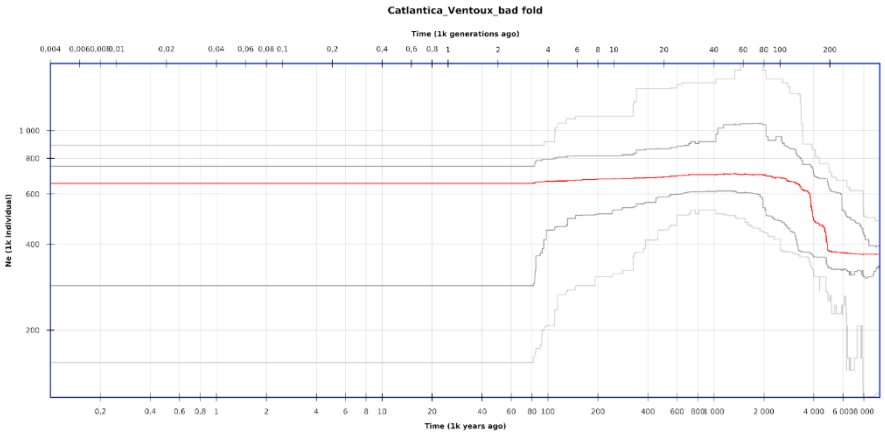

Aalba_VentouxLow fold
Time (1k generations ago)
Ne (1k individual)
Time (1k years ago)

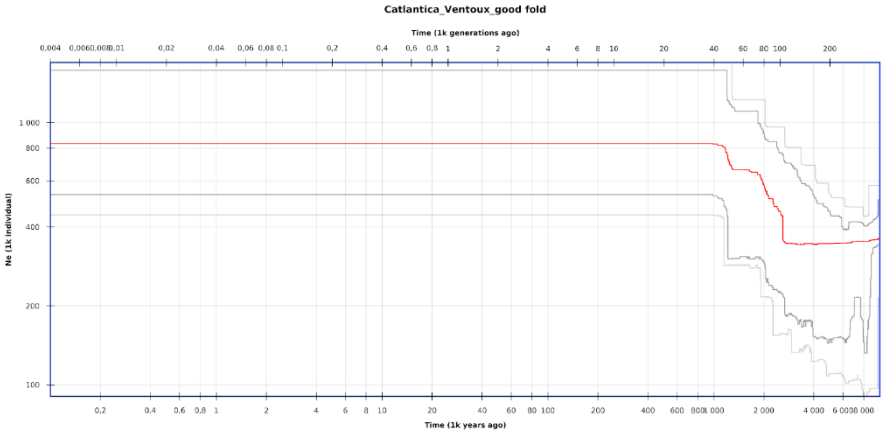

Aalba_VesubieHigh fold
Time (1k generations ago)
Ne (1k individual)
Time (1k years ago)


Aalba_VesubieLow fold
Time (1k generations ago)
Ne (1k individual)
Time (1k years ago)

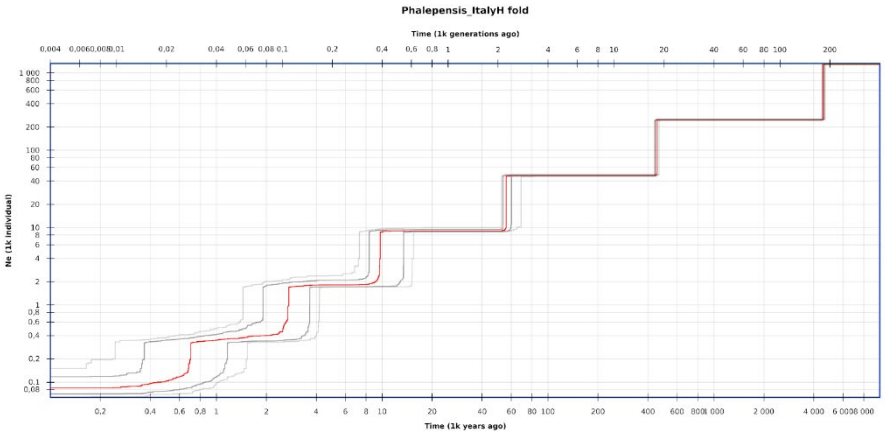
(b)

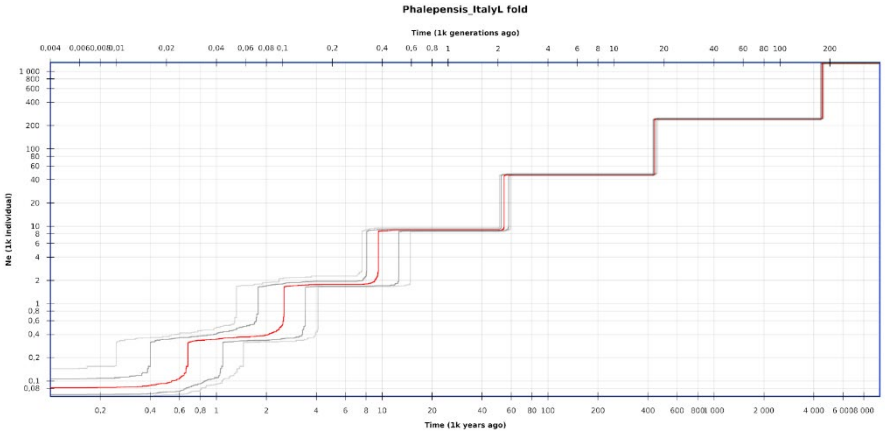

Catlantica_Luberon_bad fold
Time (1k generations ago)
Ne (1k individual)
Time (1k years ago)
Catlantica_Luberon_good fold
Time (1k generations ago)
Ne (1k individual)
Time (1k years ago)

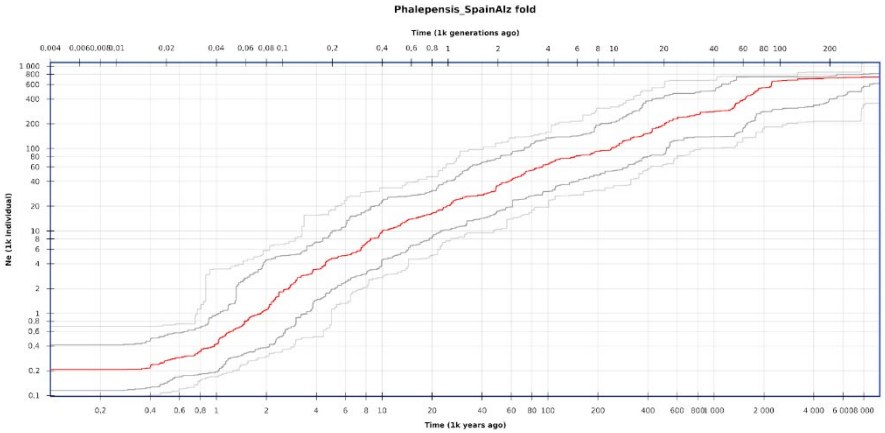

Catlantica_Ventoux_bad fold
Time (1k generations ago)
Ne (1k individual)
Time (1k years ago)

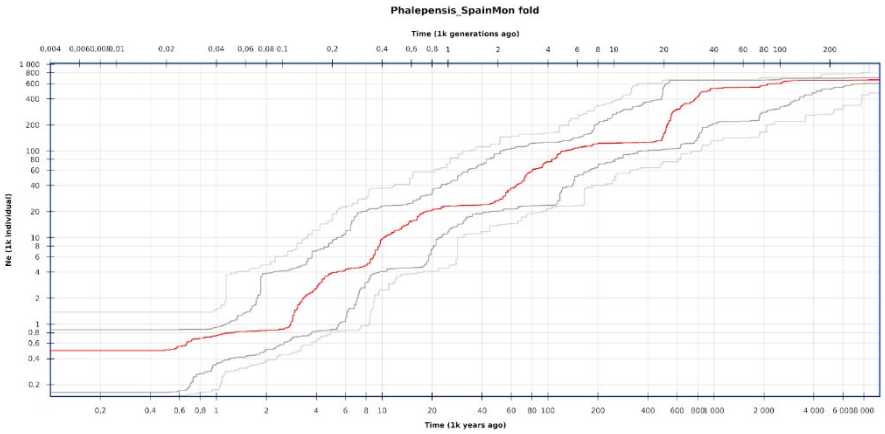

Catlantica_Ventoux_good fold
Time (1k generations ago)
Ne (1k individual)
Time (1k years ago)

(c)

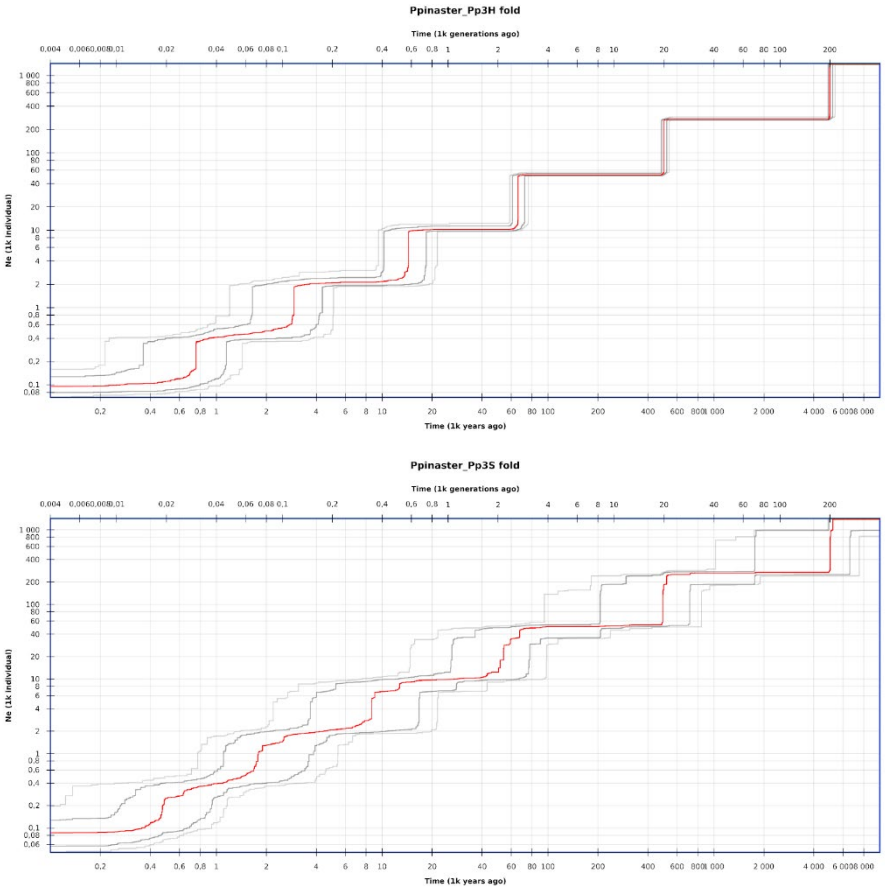

Phalepensis_ItalyH fold
Time (1k generations ago)
Ne (1k individual)
Time (1k years ago)

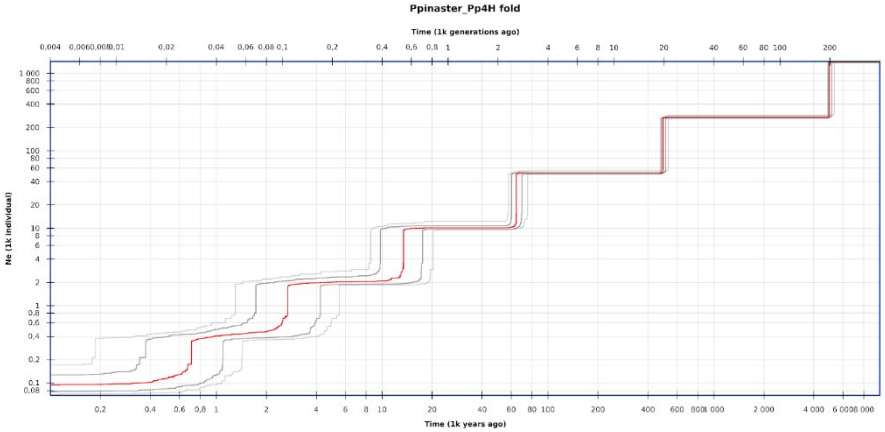

Phalepensis_ItalyL fold
Time (1k generations ago)
Ne (1k individual)
Time (1k years ago)


Phalepensis_SpainAlz fold
Time (1k generations ago)
Ne (1k individual)
Time (1k years ago)


Phalepensis_SpainMon fold
Time (1k generations ago)
Ne (1k individual)
Time (1k years ago)

(d)


Ppinaster_Pp3H fold
Time (1k generations ago)
Ne (1k individual)
Time (1k years ago)
Ppinaster_Pp3S fold
Time (1k generations ago)
Ne (1k individual)
Time (1k years ago)


Ppinaster_Pp4H fold
Time (1k generations ago)
Ne (1k individual)
Time (1k years ago)

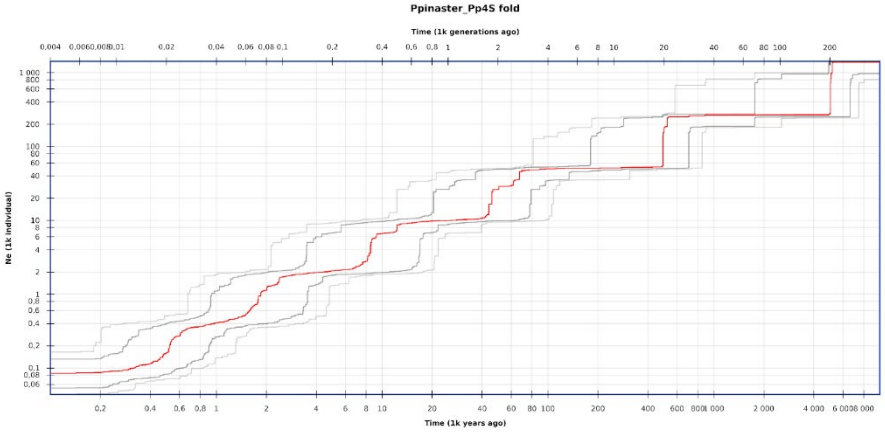
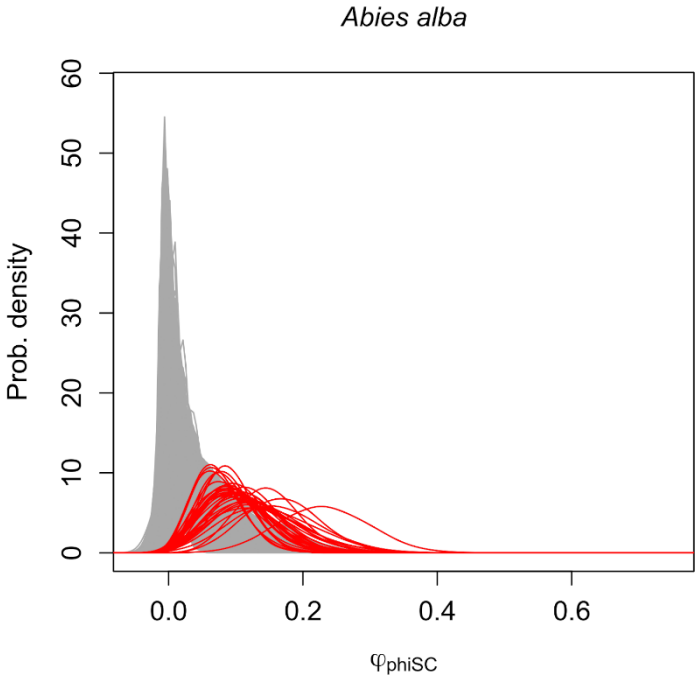
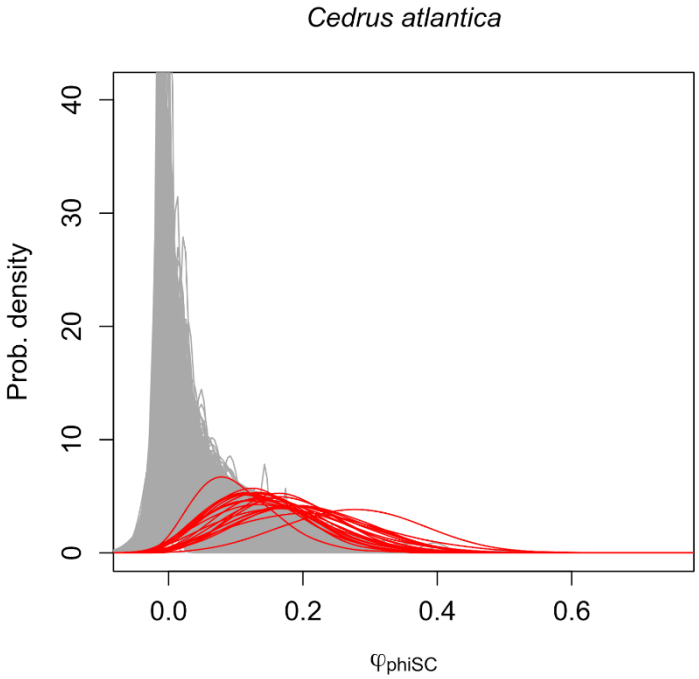
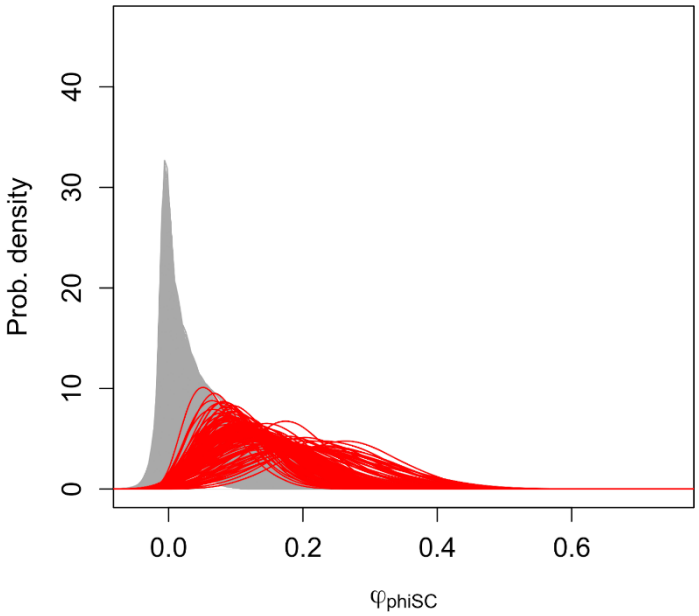

Ppinaster_Pp4S fold
Time (1k generations ago)
Ne (1k individual)
Time (1k years ago)

Figure 3(a) Posterior distribution of background loci (grey curves) and of outlier loci (red curves) as obtained in the BAMOVA analyses. Numbers of outlier loci: *A. alba*, 37; *C. atlantica*, 22; *P. halepensis*, 126; *P. pinaster*, 147.





Pinus halepensis



Pinus pinaster

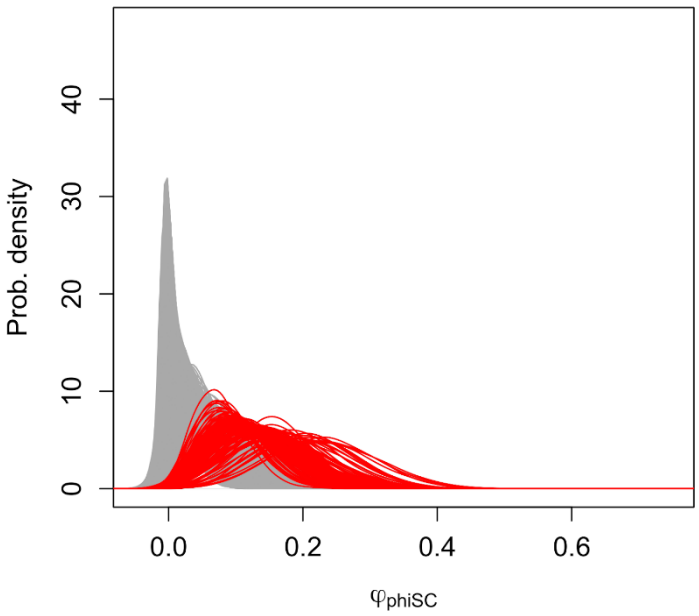
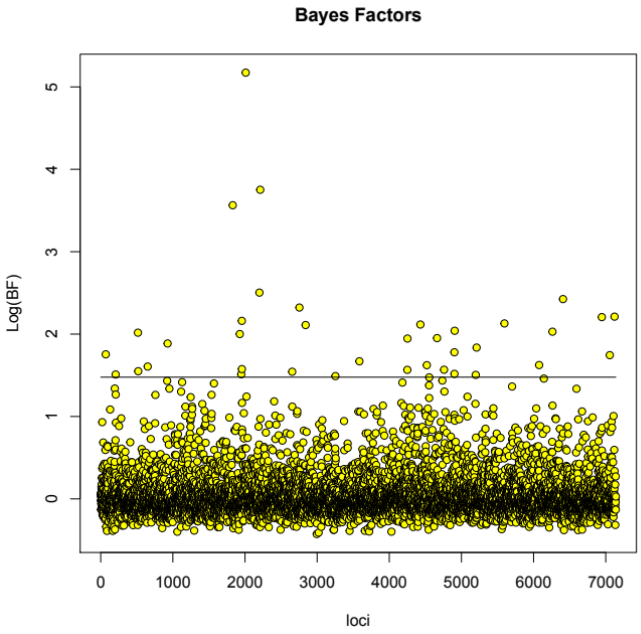
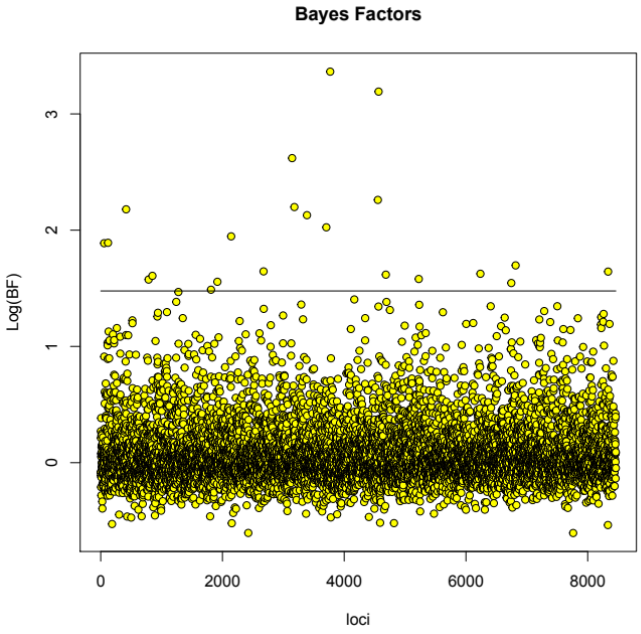


Figure 3(b) distribution of the log10(BF) for each species; horizontal bar shows the threshold to declare an outlier (BF = 30). BF = Bayes Factor. Numbers of outlier loci: *A. alba*, 104; *C. atlantica*, 18; *P. halepensis*, 74; *P. pinaster*, 7.

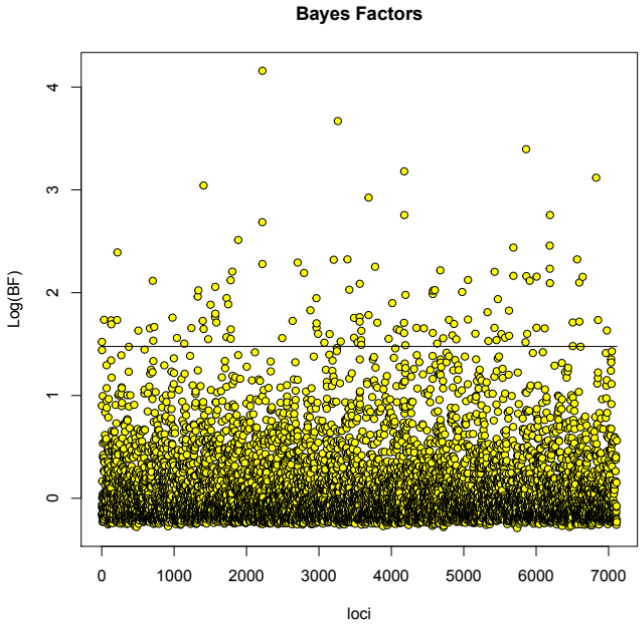
(i) *Abies alba*



(ii) *Cedrus atlantica*



Pinus halepensis



(iii) *Pinus pinaster*

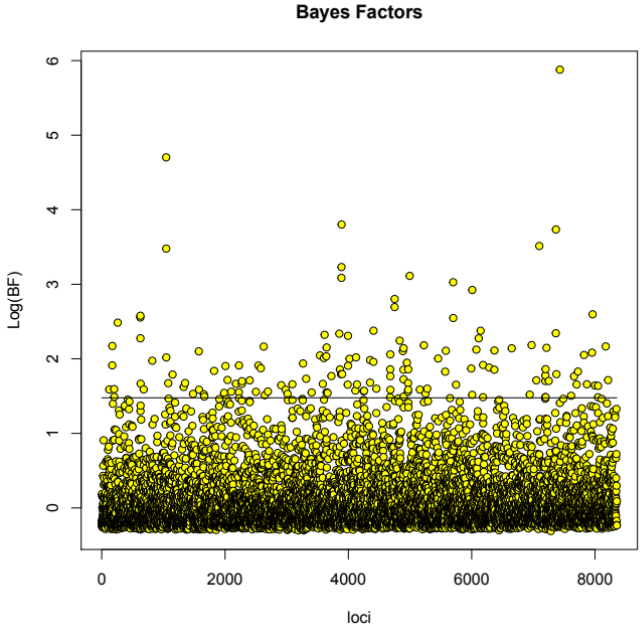
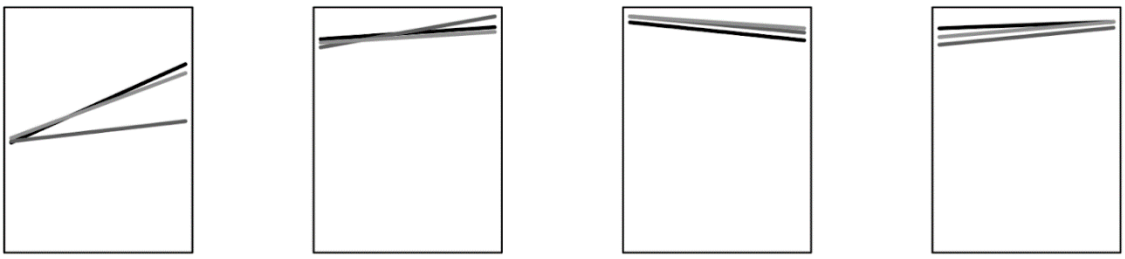
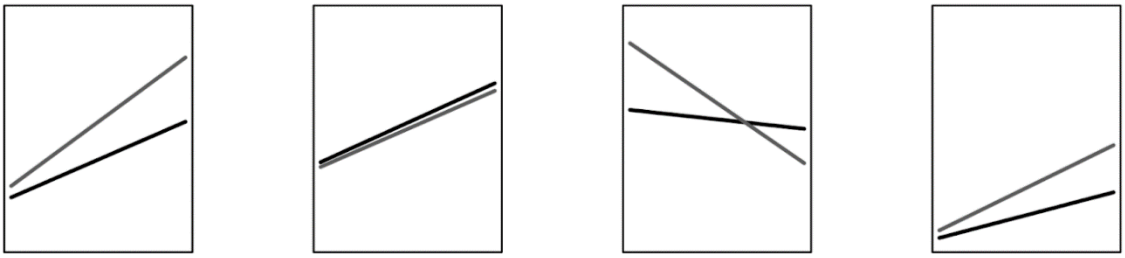


Figure 4. Trends in allele frequencies at the three sites (for *Abies alba*) or the two sites (for all other species). Allele frequency trends are shown for the four SNP showing the largest average within-site, between-stands difference in allele frequency, among those showing variation of the same sign for all sites. In each plot, the name of the SNP is given under the plot; the two stands within each site are represented along the *x*-axis; relative allele frequencies are represented on the *y*-axis; each straight line corresponds to one site. For each species, stands within each site are sorted according to their environmental conditions: *Abies alba*, "high" on the left, "low" on the right; *Cedrus atlantica*, "good site" on the left, "bad site" on the right; *Pinus halepensis*, "humid" on the left, "dry" on the right; *Pinus pinaster*, "shady" on the left, "sunny" on the right. The allele for which allele frequencies are represented was chosen arbitrarily. (a) *A. alba*; (b) *C. atlantica*; (c) *P. halepensis*; (d) *P. pinaster*.

(a)

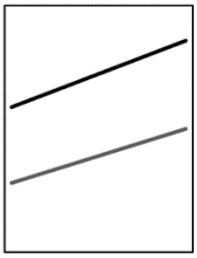


(b)

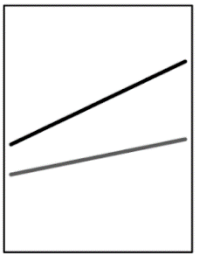


(c)

14972_32



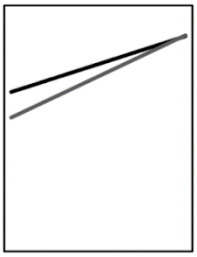
15138_186



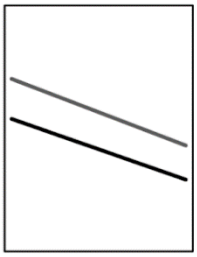
14972_161

16081_190

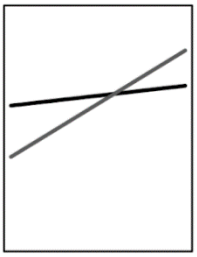
(d)



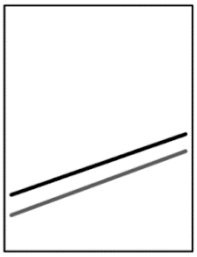
17342_44



12907_243



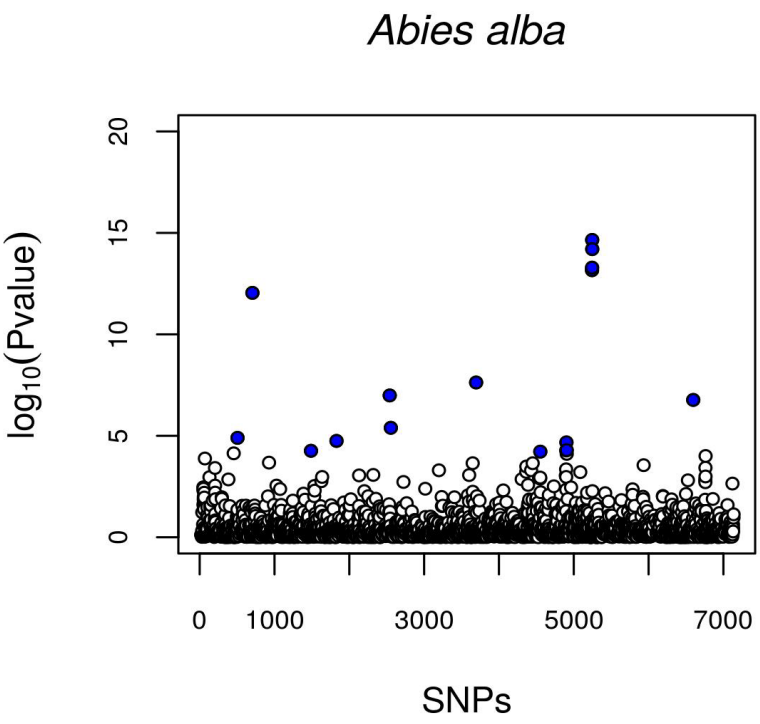
11426_136

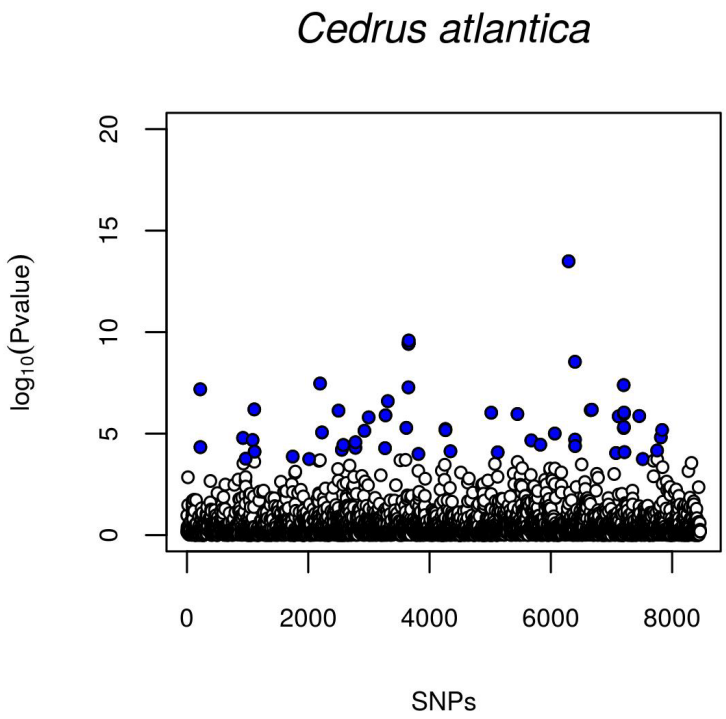


20023_299

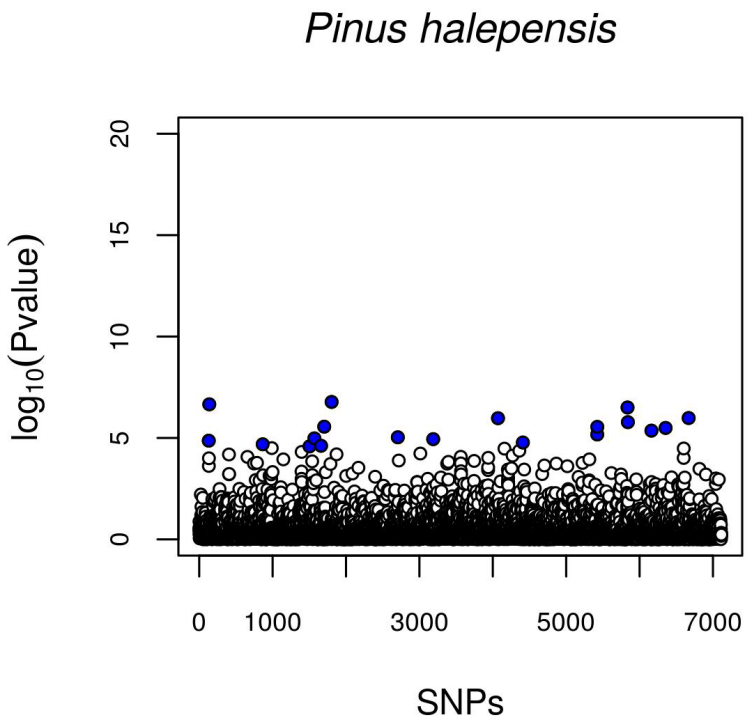
Figure 5. log10(p-values) for all loci as obtained in pcadapt. (a) *Abies alba*; (b) *Cedrus atlantica*; (c) *Pinus halepensis*; (e) *Pinus pinaster*. Filled dots: outliers. Empty dots: non-significant loci.

(a)




Cedrus atlantica
log10(Pvalue)
0
5
10
15
20
0
2000
4000
6000
8000
SNPs

(b)


Pinus halepensis
log10(Pvalue)
0
5
10
15
20
0
1000
3000
5000
7000
SNPs

(c)

(d)

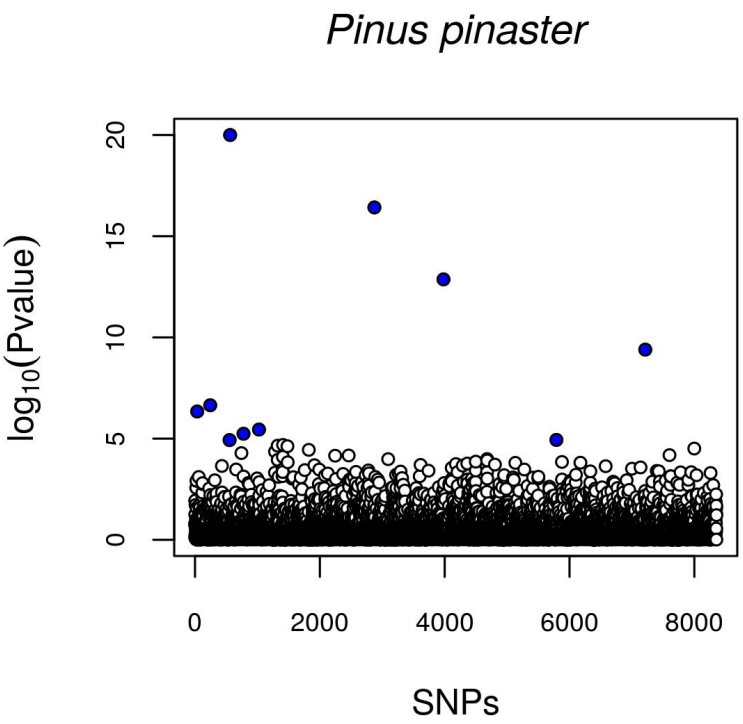

Pinus pinaster
log10(Pvalue)
0
5
10
15
20
0
2000
4000
6000
8000
SNPs

Figure 6. Violin-plots of the distribution of *modes* of the posterior distribution of stand-scaled selection coefficients (*Ns*) at *BAMOVA* outlier loci, species by species, on a logarithmic scale. Species: *A. alba = Abies alba; C. atlantica = Cedrus atlantica; P. halepensis = Pinus halepensis; P. pinaster = Pinus pinaster*. White dots indicate medians.

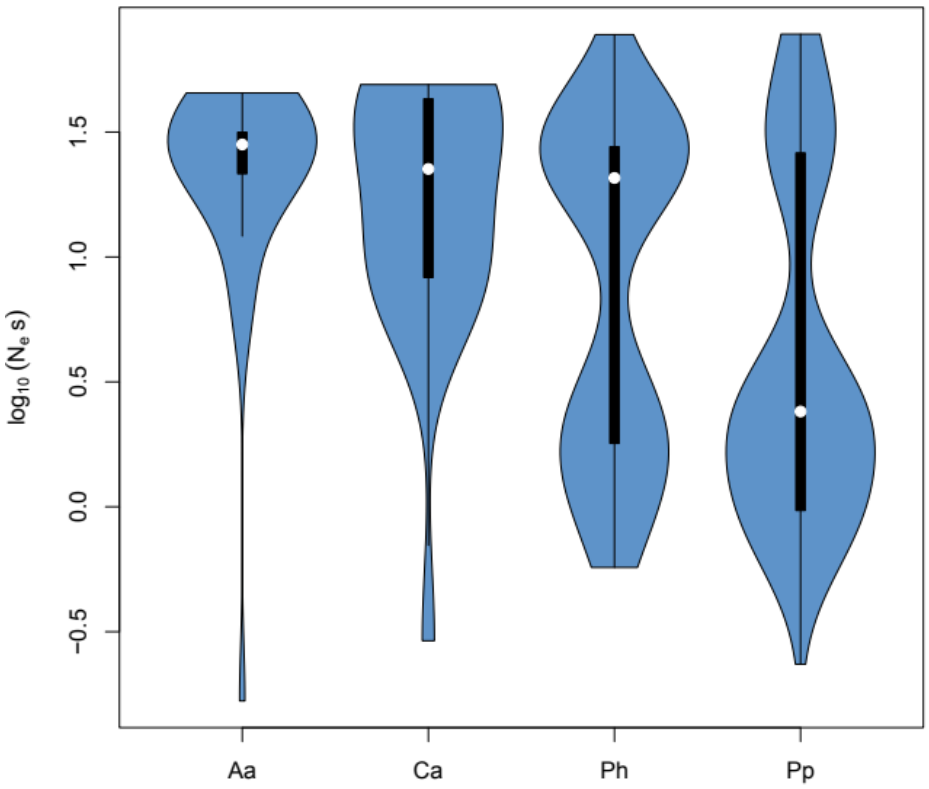
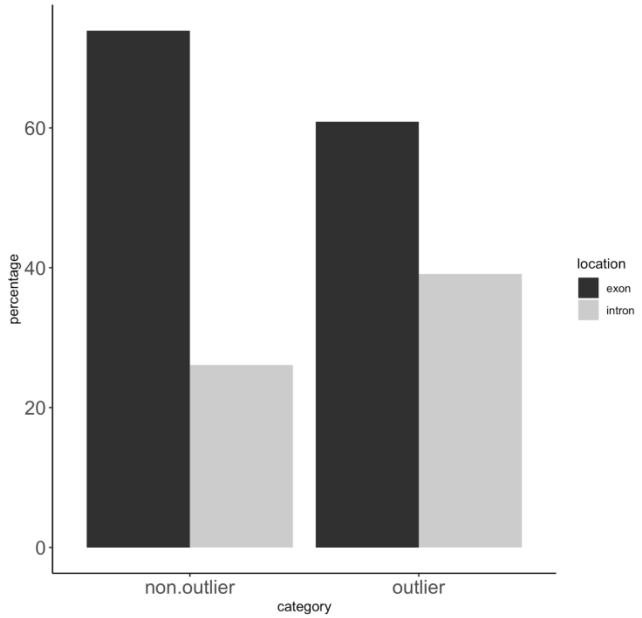
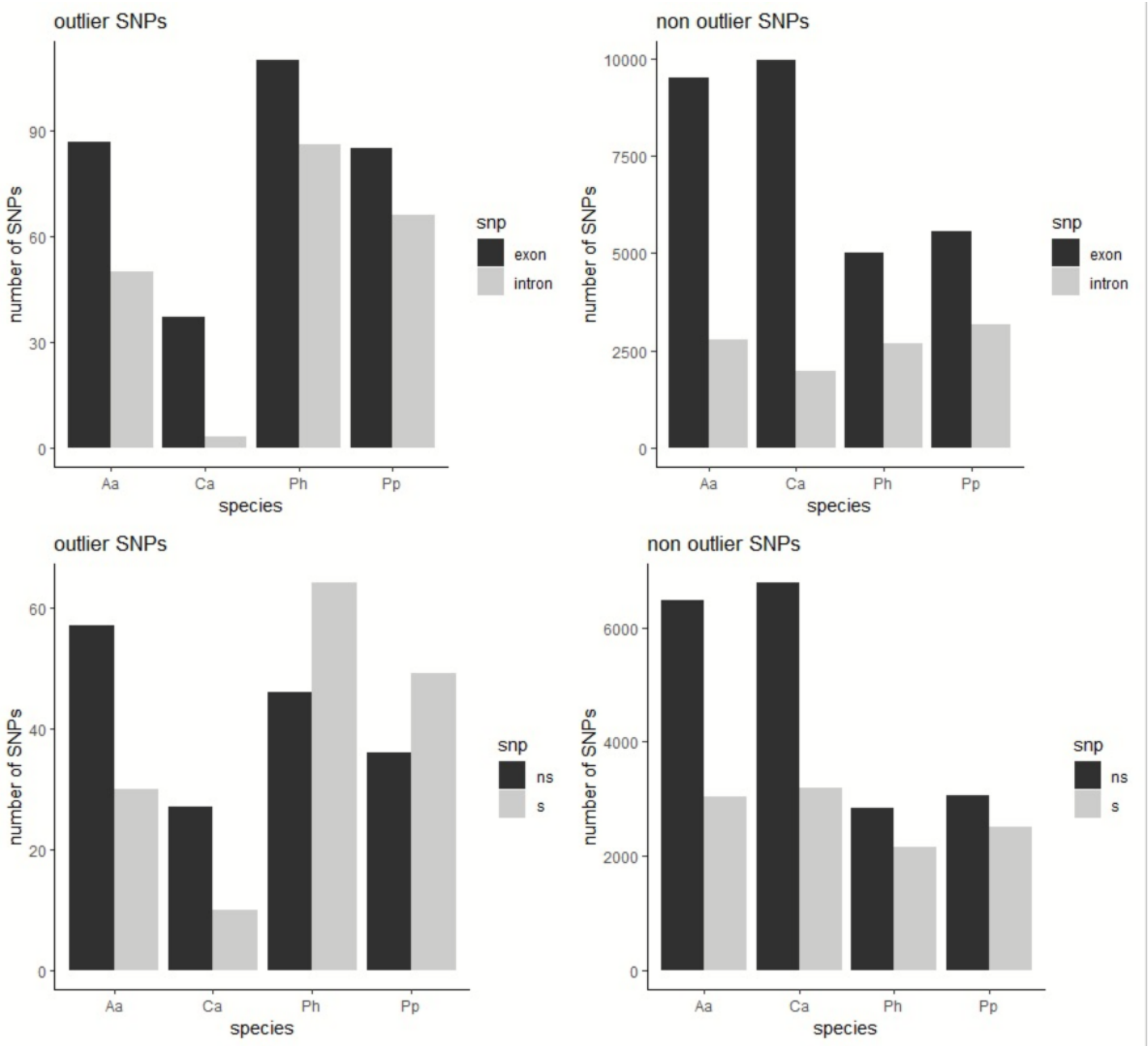


Figure 7. Distribution of outlier and non-outlier SNPs by location in sequence (exon / intron), effect on amino acid identity (synonymous / non-synonymous), and species. (a) Distribution of outlier and non-outlier SNPs in introns and exons for all species; (b) Distribution of outlier and non-outlier SNPs in introns / exons and in synonymous (s) / non-synonymous (ns) positions, by species. Aa = *Abies alba*; Ca = *Cedrus atlantica*; Ph = *Pinus halepensis*; Pp = *Pinus pinaster*.

(a)



(b)


outlier SNPs
number of SNPs
0
30
60
90
Aa
Ca
Ph
Pp
species
snp
exon
intron
non outlier SNPs
number of SNPs
0
2500
5000
7500
10000
Aa
Ca
Ph
Pp
species
snp
exon
intron
outlier SNPs
number of SNPs
0
20
40
60
Aa
Ca
Ph
Pp
species
snp
ns
s
non outlier SNPs
number of SNPs
0
2000
4000
6000
Aa
Ca
Ph
Pp
species
snp
ns
s

TABLES

Table 1 Sampling sites. Contrast = type of contrast between stands within site. Annual Heat Moisture index (AHM) and Summer Heat Moisture Index (SHM) are calculated as follow: AHM: (mean annual temperature+10)/(mean annual precipitation/1000)); SHM: mean warmest month temperature/(mean summer precipitation+1000) (Wang et al., 2010; Hallingbäck et al., 2021).

Species		Site	Distance between stands (km)	Coordinates	Stand names	Contrast
A. alba		Issole	1	44.03N/6.48E	Iss5	high altitude 1550 m; Annual Heat Moisture index: 12.4; Summer Heat Moisture index: 31.1
					Iss2	low altitude 1200 m; Annual Heat Moisture index: 15.2; Summer Heat Moisture index: 38.6
		Ventoux	1	44.17N/5.28E	N4	high altitude 1380 m; Annual Heat Moisture index: 13.9; Summer Heat Moisture index: 37.9
					N2	low altitude 1100 m; Annual Heat Moisture index: 15.3; Summer Heat Moisture index: 42.6
		Vesubie	1	43.98N/7.36E	ves5	high altitude 1500 m; Annual Heat Moisture index: 14.7; Summer Heat Moisture index: 40.1
					ves1	low altitude 1050 m; Annual Heat Moisture index: 18.5; Summer Heat Moisture index: 51.4
C. atlantica		Luberon	1	43.8N/5.21E	LubLaco_C3 071-095	high site index (mean height at 50 years 15.5m)
					LubLaco_C3 096-136	low site index (mean height at 50 years 8.5m)

		Ventoux	1	44N/5.28E	VentRol	high site index (mean height at 70 years 18.1m)
					VentFey	low site index (mean height at 70 years 12.9m)
P. halepensis		Italy	3.5	41.7N/16.0E	Monte S. Angelo	humid / high altitude (500 m asl), Annual Heat Moisture index: 44.6; Summer Heat Moisture index: 127.7
				41.4N/16.02E	Mattinata	dry / low altitude (50 m asl), Annual Heat Moisture index: 67.9; Summer Heat Moisture index: 222.9
		Spain	100	40.05N/0.59W	Montan	humid (precipitation of the driest quarter: 65.8 mm) / low frequency of crown fire, Annual Heat Moisture index: 42.5; Summer Heat Moisture index: 112.9
				39.74N/0.48W	Alzira	dry (precipitation of the driest quarter: 44.2 mm) / high frequency of crown fire, Annual Heat Moisture index: 43.5; Summer Heat Moisture index: 128.4
P. pinaster		Ain-Almedijar	1	39.89N/0.35W	H3	Shady slope, North exposition (aspect: 34.6º)
					S3	Sun-exposed slope (aspect: 159.0º)
		Alcudia de Veo	1	39.91N/0.39W	H4	Shady slope, West exposition in close valley (aspect: 262.0º)

		Algimia de Almonacid			S4	Sun-exposed slope (aspect: 181.0º)

Table 2
(a) Genome-wide estimates of *hierfstat*-computed hierarchical *F*-statistics. *Stand*: between-stands F component; *Site*: among-sites F component; F_{IS}: individual-level F component, averaged over loci.

Species	Stand (between *stands,* within site)	Site (among *sites*)	F_{IS}
Abies alba	0.0034	0.0035	-0.069
Cedrus atlantica	0.0024	0.0073	0.020
Pinus halepensis	0.0036	0.0092	-0.044
Pinus pinaster	0.0041	0.0017	-0.026

(b) stand pairwise, genome-wise F_{ST} point estimates.

A. alba					
	Issole-High	Issole-Low	Ventoux-High	Ventoux-Low	Vesubie-High
Issole-Low	0.015				
Ventoux-High	0.016	0.017			
Ventoux-Low	0.036	0.037	0.046		
Vesubie-High	0.028	0.029	0.036	0.019	
Vesubie-Low	0.016	0.016	0.016	0.041	0.031
C. atlantica					
	Luberon bad_site	Luberon good_site	Ventoux bad_site		
Luberon good_site	0.024				
Ventoux bad_site	0.046	0.054			
Ventoux good_site	0.040	0.045	0.057		
P. halepensis					
	ItalyH	ItalyL	SpainAlz		
ItalyL	0.030				
SpainAlz	0.082	0.129			
SpainMon	0.096	0.143	0.017		
P. pinaster					
	3H	3S	4H		
3S	0.022				
4H	0.024	0.031			
[illegible]	0.025	0.027	0.028		

(c) Summary of the posterior probability density (value of the median; limits of 99% credible intervals) for the *BAMOVA*-estimated within-site (between stands) Φ-statistics, at the genome level and for outlier loci: genome-level Φ_{SC} (i.e. between stands, within site) and Φ_{CT} (i.e. among sites) values for each species. lCI: lower boundary of the 99% credible interval of the posterior distribution; Median: median of the posterior distribution; Mode: mode of the posterior distribution; uCI: upper boundary of the 99% credible interval of the posterior distribution.

	Φ_{SC}	Φ_{CT}

	(between *stands*, within site)				(among *sites*)			
Species	lCI	Median	Mode	uCI	lCI	Median	Mode	uCI
A.alba	0.0205	0.0216	0.0216	0.0227	-0.0036	0.0005	0.0005	0.0013
C.atlantica	0.0275	0.0291	0.0291	0.0307	0.0086	0.0104	0.0104	0.0319
P.halepensis	0.0290	0.0304	0.0304	0.0320	0.0674	0.0712	0.0712	0.0749
P.pinaster	0.0294	0.0307	0.0308	0.0321	-0.0002	0.0009	0.0009	0.0021

Caption to Supplementary Material Items

Supplementary Table T1.

(a) List of the probes used by rapidGenomics (Gainsville, FL) to carry out sequence capture, with the indication of whether the probe matched ("TRUE") or did not match ("FALSE") at least a contig for each species.

Species codes: Aa = *Abies alba*; Ca = *Cedrus atlantica*; Ph = *Pinus halepensis*; Pp = *Pinus pinaster*.

NbSpecies: number of species for which the probe matched at least one contig (i.e., number of "TRUE" values on the row).

(b) probe sequences.

Supplementary Table T2.

Sample-level quality assessment of raw data

Supplementary Table T3.

List of SNPs for each species.

Name format: [contig number]_[variant position]

Supplementary Table T4.

Lists of outliers for all species and all methods.

SNP names are coded as in Supplementary Table T3 ([contig name]_[SNP. Position]).

Supplementary table T5.

Number of outliers detected by each method.

Supplementary Table T6.

(a) summary (minimum, mean, maximum) of mode values for each species; (b) 99% Credible interval (lower boundary, mode, upper boundary) of posterior distributions of the scaled selection coefficient *Ns* for BAMOVA outlier loci. Locus names coded as (*contig name)_(SNP position)*. Prior = prior distribution from simulations.

Supplementary Table T7.

Orthology groups for outlier-containing contigs

Supplementary Table T8.

Functional annotation of outlier-containing contigs

Supplementary Table T9

Summary of the distribution of variants by species and by functional annotation.

(a) all variants; (b) outliers; (c) comparison of numbers of each type of variant between outliers and all loci in CDS; (d) same as (c) but with SnpEff characteristics aggregated and the impact of an amino acid change (i.e. if the amino acid change implies a change of its class: charged, polar or hydrophobic and a change in the termination).

Supplementary Table T10

GO-terms enrichment in outlier-contaning contigs

Supplementary Results R1-R4

Contig consensus sequences

Supplementary Results R5

Background data on assembly properties, SNP distribution and SNP heterozygosity.

(a) *Abies alba*; (b) *Cedrus atlantica*; (c) *Pinus halepensis*; (d) *Pinus pinaster*. For each species: upper left pane, distribution of contig lengths in the assembly; upper right pane, number of variants per contig before final variant filtering; lower left pane, density of variants per base before final variant filtering; lower right pane, heterozygote deficit (F_{IS}) within each stand after final variant filtering.

Supplementary Results R6

1. Table of summary statistics of per-locus normalised π by population and histograms of normalised π by population; 2. Average inbreeding coefficients per species.

Supplementary Results R7

Analysis of overlap of outliers across analyses and estimation of error rates in outlier detection.

Supplementary Results R8

ABC cross-validation.

Supplementary Methods M1

(a) DNA isolation (b) choice of sequence capture probes (c) sequence capture, assembly, and variant calling

Supplementary Methods M2

Scripts used for additional variant filtering.

(a) sieve() script; (b) monoRemove() script; (c) ad-hoc script for calculation of π

Supplementary Methods M3

Simulation of evolutionary scenarios to estimate power and FDR of outlier tests.

Supplementary Methods M4

1.Checks on forward simulations; 2. Detailed ABC methods; 3. Cross validations of ABC estimations

Supplementary Figure F1

Poster distribution of divergence estimates as obtained in BAMOVA. Φ_{ST} = global divergence; Φ_{CT} = divergence among sites; Φ_{SC} = divergence between populations within site. Notice that scales differ among plots.

Supplementary Figure F2

Overlap of outlier lists between the two methods. Horizontal bars represent numbers of outliers obtained with each method, and vertical bars represent the number of shared outliers over methods. The group of methods corresponding to each vertical bar is indicated by the dots connected by the solid line underlying the bar (the first bars on the left of each plot represent the overlap of each method with itself - that is, the non-overlapping outliers).

Supplementary Figure F3

Between-stands trends in allele frequencies at BAMOVA and pcadapt outlier loci, for each species. X-axis: locus names (see Materials & Methods for naming conventions); y-axis: allele frequencies for an arbitrary allele (all frequencies bound between 0 and 1). For each site, the population on the left of the plot is the lower elevation one, and the population on the right is the higher elevation one.

Species and population identity:

(a) *Abies alba*, filled line: Issole; dashed line: Ventoux; dotted line: Vésubie.

(b) *Cedrus atlantica*, filled line: Luberon; dashed line: Ventoux;

(c) *Pinus halepensis*, filled line: Italy; dashed line: Spain;

(d) *Pinus pinaster*, filled line: Ain-Almedijar; dashed line: Algimia de Almonacid.

Supplementary Figure F4

Hyperposterior distributions for the scaled selection coefficient (Ns) for BAMOVA outlier loci (filled lines) against the prior (dotted lines).

Supplementary Figure F5

ReviGo analysis of enriched GO terms.