High rate of adaptive evolution in two widespread European pines

Delphine Grivet, Komlan Avia, Aleksia Vaattovaara, Andrew J Eckert, B Neale, Outi Savolainen, Santiago González-Martínez

High rate of adaptive evolution in two widespread European pines

Delphine Grivet[1,2]*, Komlan Avia[3,4,5]*, Aleksia Vaattovaara[3], Andrew J. Eckert[6], David B. Neale[7], Outi Savolainen[3], Santiago C. González-Martínez[1,2,8]

[1]Department of Forest Ecology & Genetics, Forest Research Centre, INIA-CIFOR, Carretera de la Coruña km 7.5, ES-28040 Madrid, Spain.

[2]Sustainable Forest Management Research Institute, INIA - University of Valladolid, Avda. Madrid 44, 34004 Palencia, Spain.

[3]Department of Ecology and Genetics and Biocenter Oulu, University of Oulu, FIN-90014 Oulu, Finland.

[4]Algal Genetics Group, UMR 8227, CNRS, Sorbonne Universités, UPMC, Station Biologique Roscoff, CS 90074 29688 Roscoff, France.

[5]UMI 3614 Evolutionary Biology and Ecology of Algae, CNRS, Sorbonne Universités, UPMC, Pontificia Universidad Católica de Chile, Universidad Austral de Chile, Station Biologique Roscoff, CS 90074, 29688 Roscoff, France.

[6]Department of Biology, Virginia Commonwealth University, Richmond, VA 23284, United States.

[7]Department of Plant Sciences, University of California at Davis, Davis, CA 95616, United States.

[8]BIOGECO, INRA, Univ. Bordeaux, 33610 Cestas, France.

*Equal contributors

Keywords: natural selection, adaptation, *Pinus*, Distribution of Fitness Effects, adaptive

evolution

Corresponding author:
Delphine Grivet, Department of Forest Ecology & Genetics, Forest Research Centre, INIA-CIFOR, Carretera de la Coruña km 7.5, ES-28040 Madrid, Spain.
Fax: +34 91 347 6767 Email: dgrivet@inia.es

Running title (45 char including spaces): Adaptive evolution in two European pines

Abstract (≤250 words),

Comparing related organisms with differing ecological requirements and evolutionary histories can shed light on the mechanisms and drivers underlying genetic adaptation. Here, by examining a common set of hundreds of loci, we compare patterns of nucleotide diversity and molecular adaptation of two European conifers (Scots pine and maritime pine) living in contrasted environments and characterized by distinct population genetic structure (low and clinal in Scots pine, high and ecotypic in maritime pine) and demographic histories. We found higher nucleotide diversity in Scots pine than in maritime pine, whereas rates of new adaptive substitutions (ω_a), as estimated from the Distribution of Fitness Effects (DFE), were similar across species, and among the highest found in plants. Sample size and population genetic structure did not appear to have resulted in any significant bias in ω_a. Moreover, the species-specific population contraction-expansion dynamics did not seem to have affected differentially the rate of adaptive substitution in these two pines. Several methodological and biological factors may underlie the unusually high rate of adaptive evolution of Scots pine and maritime pine. By providing two new case studies with contrasting evolutionary history, we contribute to disentangling the multiple factors potentially affecting adaptive evolution in natural plant populations.

Introduction

Understanding of the mechanisms of plant adaptation has been advanced through comparative studies of species differing in demographic history (Slotte *et al.* 2010), effective population size (Gossmann *et al.* 2012; Strasburg *et al.* 2011), genetic structure among populations (Wright & Andolfatto 2008), ecological conditions (Tellier *et al.* 2011), or phylogenetic relationships (Eckert *et al.* 2013a; Grivet *et al.* 2013; Palmé *et al.* 2009). For long-lived species such as most forest trees, unraveling adaptive processes is challenging, as during their long life-cycle, individuals experience different selective pressures that accumulate over time. Genomic research in forest trees, moreover, has been hindered because of large genome sizes and the lack of model species. Nevertheless, recent advances in tree genomics and reanalysis of common garden experiments have fostered a body of literature that provides insights into the mechanisms underlying forest tree adaptation in space and time. Emerging from these studies are the ideas that forest trees are locally adapted to different environmental conditions, especially to temperature, photoperiod, drought, or biotic stress (see examples in Alberto *et al.* 2013a; Savolainen *et al.* 2007), and that they respond to contrasted selection pressure across life stages (Alía *et al.* 2014). Based on this knowledge, it is of prime interest to identify major environmental drivers of adaptation, as well as the genes involved in the process, as they can help forecast the future distribution of these ecologically important species in the face of climate change.

Identifying footprints of natural selection within genomes is complex, as the observed patterns of polymorphism may result from many distinct yet interacting evolutionary forces, including neutral processes such as migration or genetic drift, all acting on variation generated by mutation. Because demography and selection can leave similar patterns within genomes (Biswas & Akey 2006; Excoffier *et al.* 2009; Nielsen 2005), tests

aiming at attributing these patterns to the action of natural selection must take into account the demographic history of the species, as well as the complex interactions between these two processes (Li *et al.* 2012; Schrider & Kern 2016). Tests robust to demography often contrast polymorphism within species with divergence between species. For example, the popular McDonald-Kreitman (MK) test (McDonald & Kreitman 1991) compares the amount of polymorphism to divergence for categories of sites that are expected to evolve differently (e.g., synonymous/silent vs. non-synonymous). From the original MK test, several extensions have been developed. Two of the most popular are the MKPRF test (Bustamante *et al.* 2002, 2003, 2005), which is more powerful but is based on specification of a population genetic model, and the more recent SnIPRE test (Eilertson *et al.* 2012), which is a nonparametric method that does not require estimation of population genetic parameters.

Other methods based on the MK test provide overall estimates of adaptive evolution that can be correlated with population parameters (e.g., effective population size; Gossmann *et al.* 2012) or environmental characteristics (e.g., Lourenço *et al.* 2013) in comparative studies. First, the Distribution of Fitness Effects (DFE) of new mutations at functional sites is estimated from polymorphism data and under a specific demographic scenario. Second, the number of substitutions originating from neutral and slightly deleterious mutations is predicted from the estimated DFE (e.g., the method II of Eyre-Walker & Keightley 2009). Any excess of substitutions (with respect to the neutral expectation) can then be attributed to directional selection, as measured by α (the proportion of adaptive nucleotide substitutions) or ω_a (the relative rate of adaptive substitutions scaled by the rate of neutral substitution) (Bierne & Eyre-Walker 2004; Eyre-Walker & Keightley 2009; Smith & Eyre-Walker 2002). Although α has been extensively used (e.g., (Eckert *et al.* 2013a; Gossmann *et al.* 2010; Slotte *et al.* 2010; Strasburg *et al.* 2011), this estimate is of limited value for

estimating the efficiency of the adaptive process, because α is also influenced by the rate of non-adaptive substitutions (Gossmann *et al.* 2012). The parameter ω_a, which is roughly equivalent to K_a (the number of non-synonymous substitutions per non-synonymous site), is thus more appropriate for comparing adaptive evolution across genomic regions or species (Gossmann *et al.* 2012; Lourenço *et al.* 2013).

In the present study, we assess the effect of distinct demographic histories and selection regimes on inferences of adaptive evolution for two widespread conifer species. Scots pine (*Pinus sylvestris* L.) is widely and continuously distributed in Eurasia, occupying regions that differ greatly in climate (Krakau *et al.* 2013). Its demographic history is characterized by an ancient bottleneck (Kujala & Savolainen 2012; Pyhäjärvi *et al.* 2007) and limited population genetic structure, which is only found along the margins of its wide range (Gullberg 1985; Kujala & Savolainen 2012; Pyhäjärvi *et al.* 2007). The absence of genetic structure shown by molecular markers across much of its range, however, is not indicative of a lack of quantitative trait variation across populations (Andersson & Fedorkov 2004; García-Gil *et al.* 2003; Notivol *et al.* 2007; Oleksyn *et al.* 1998). In particular, strong clines in photoperiod-related traits are often observed across latitudinal gradients in the species (e.g., for timing of growth cessation and budset; see Beck *et al.* 2004; Hurme *et al.* 1997; Mikola 1982; Oleksyn *et al.* 1998).

In contrast to Scots pine, maritime pine (*Pinus pinaster* Ait.) is patchily distributed across the western Mediterranean Basin and the Atlantic regions of Portugal, Spain and France. This conifer grows in warm temperate regions with an oceanic influence on climate (Abad Viñas *et al.* 2016), and is particularly well adapted to dry and fire-prone environments. Its demographic history is characterized by a more recent bottleneck relative to Scots pine (Naydenov *et al.* 2014), and it has a strong genetic structure among populations across its range (e.g., Burban and Petit 2003; Bucci et al. 2007; Jaramillo-

Correa et al. 2015) that is accompanied by morphological and physiological differences (Kremer and Roussel 1986; Alía et al. 1995; Santos del Blanco et al. 2010; Lamy et al. 2011; Corcuera et al. 2012), defining various subspecies and eco-types (Richardson 1998). This genetic structure likely results from post-Pleistocene events (Burban and Petit 2003; Bucci et al. 2007; Naydenov et al. 2014), including adaptation to local climate (Serra-Varela et al. 2015).

Given the large differences between Scots pine and maritime pine with respect to their climatic niches, we focused on genes for general responses to different type of abiotic stress. This is also consistent with previous studies ranking the importance of different fitness-related traits for local adaptation in the species (for Scots pine see Castro *et al.* 2002; Galiano *et al.* 2010; Ryyppö *et al.* 1998; and for maritime pine see Corcuera *et al.* 2011; Gaspar *et al.* 2013; Lamy *et al.* 2014; Ramírez-Valiente & Robledo-Arnuncio 2014). Then, we used complementary methods to get insights into the action of selective forces, both at specific genes and genome-wide, and considered the specific demographic, ecological and historical settings of each species to discuss the possible factors (both methodological and biological) that may explain our results. Our hypotheses are: i) Scots pine has high nucleotide diversity resulting from its large distribution and expected large effective population size; ii) Large effective population size in Scots pine would have also resulted in higher efficiency of selection and thus a higher number of fixed adaptive substitutions (see also Gossmann *et al.* 2012); and iii) the highly fragmented distribution in maritime pine would have resulted in lower levels of overall adaptive evolution. Scots pine and maritime pine, with their contrasting characteristics, allow for exploring how different biological factors may interact with natural selection and adaptive evolution in plants.

Materials and methods

Sampling

Scots pine was sampled across 8-10 populations from its western range (Figure 1 and Table S1) for eight loci from six candidate genes for abiotic stress response and phenology (3-20 individuals per population, and a total of 36-115 individuals per locus), and across a smaller number of populations (7) for the much larger CRSP (Comparative Re-Sequencing in Pinaceae initiative; http://dendrome.ucdavis.edu/NealeLab/crsp/; see Wegrzyn *et al.* 2008) set of 364 loci (1-2 individuals per population, and a total of 6-12 individuals per locus). In a first phase, maritime pine was sampled across 10-12 populations from its full range (Figure 1 and Table S1) for the same eight abiotic stress response and phenology loci as for Scots pine (1-19 individuals per population for a total of 49-100 individuals per locus), and across 11 populations for the CRSP gene dataset (1-3 individuals per population for a total of 6-14 individuals per locus). More details on the gene datasets are presented in the *Sequence data* section. In a second phase, another dataset produced in maritime pine comprising 128 loci common with the CRSP dataset (2-4 individuals per population for a total of 28-36 individuals per locus) was obtained within the Conifer Re-sequencing Initiative for European Conifers (CRIEC, www.evoltree.eu; Figure 1 and Table S1). This dataset with more individuals per population was examined in order to study the effect of sample size and population structure on the DFE-based estimates. Finally, the CRSP and the CRIEC datasets were combined for the 128 loci in common (2-6 individuals per population for a total of 34-49 individuals per locus) to reach a bigger sample size for the statistical analyses ("extended CRSP" dataset). Sequence datasets used for the different analyses are detailed in Figure S1.

Trees were sampled within populations following standard protocols to avoid sampling related trees (i.e., leaving a minimum distance of 50 m between sampled individuals) and without any phenotypic selection.

Sequence data

Sequence alignment and editing

Eight loci from previously studied candidate genes, including two full-length genes (*coL1* and *gia*), were amplified with available primers from different sources (see Table S2). Another 1,600 gene amplicons (hereafter referred as the CRSP gene set) were obtained by resequencing in seven conifers (including Scots pine and maritime pine) some loci originally developed in loblolly pine, *Pinus taeda* (Eckert *et al*. 2013b). For both gene sets, DNA sequences were obtained by direct sequencing from haploid seed megagametophytes. In this way, (i) phase is directly known and does not need to be estimated and (ii) co-amplification of paralogs, a common problem in plant species with large genomes such as conifers, is more easily detected. For the CRSP gene set, loci with at least six successfully sequenced individuals in both Scots pine and maritime pine were accepted for further analyses. This led to 491 loci that were subsequently checked manually and edited with Codon Code Aligner (CodonCode Corporation, Centerville, MA, USA) or Sequencher 4.7 (Gene Codes Corporation, Ann Arbor, MI, USA). Low quality sequences were removed, as well as putative paralogous genes based on phylogenies including the three pines (the two target species together with loblolly pine), as detected by PRANK runs with default parameters (Löytynoja & Goldman 2005). This last quality-filtering step led to a set of 389 loci common to the two target species.

Annotation

Gene annotation for all loci was obtained from homology with loblolly pine EST contigs and the NCBI reference protein database using Geneious version 6.1 (Biomatters, Auckland, New Zealand). All 389 genes had high sequence similarity with loblolly pine genes, suggesting low copy-number genes and orthology across the different species. The biological function of the genes was determined based on their homology with the *Arabidopsis thaliana* and/or *Pinus taeda* protein database with an E-value threshold of 10^{-10} (see Table S3) and the Protein Knowledgebase – UniProtKB (http://www.uniprot.org/uniprot/). Of the 389 genes, 364 could be annotated, leading to a total of 372 annotated loci in the two species when adding the eight previously-studied candidate gene loci. Loblolly pine was used as outgroup when needed.

Target and reference genes

The 372 loci under study, as well as their annotations, are publicly available (NCBI GenBank: MF385275-MF385581 and MF385585-MF397901). From these 372 loci, 64 were selected as target genes, including the eight previously-studied candidate gene loci (see Table S2 for details and references) and 56 other annotated loci selected from the CRSP dataset (see above). All 64 target genes had highly confident annotation associated to adaptive traits: 48 loci related to abiotic stress responses (mostly cold, heat, drought, salt and other oxidative stresses); and 16 loci related to photosynthesis and photosystem (9 loci), and phenology (7 loci). Only target genes were tested for footprints of selection.

The reference genes consisted of all other loci (308) from the CRSP dataset, including loci with unknown function or loci for which there is no evidence of involvement in forest tree adaptation. Because of implementation limitations and statistical power requirements (i.e., the minimum number of required segregating sites) in the MKPRF test (see below), reference genes were combined for this analysis into compound loci. All loci

with a known position on a genetic map (loblolly pine, Eckert *et al.* 2010; maritime pine, Chancerel *et al.* 2011; and our own unpublished genetic map of Scots pine) were grouped according to the linkage groups forming 28 compound loci (see Table S4). To balance the number of loci used in the MKPRF test (64 target versus 28 compound reference loci), we used also 36 other reference loci selected at random among those reference genes not included in the compound loci to also reach 64 loci for the reference group in which the statistical power was mainly driven by the compound loci (higher number of segregating sites).

Nucleotide diversity, genetic divergence and overall patterns of polymorphism

Nucleotide diversity π (Nei & Li 1979) and Watterson's θ_w (Watterson 1975) were calculated for all sites, as well as separately for silent, synonymous and non-synonymous sites, in Scots pine and maritime pine. For each species, the divergence from loblolly pine was characterized by the number of synonymous substitutions per synonymous site (K_s) and of non-synonymous substitutions per non-synonymous site (K_a), both with Jukes-Cantor correction, and their ratio (K_a/K_s), as well as the statistics of shared and fixed segregating sites requested by some of the neutrality tests (see below).

To identify overall differences in patterns of polymorphism across species, different statistics were computed based on the observed Site Frequency Spectrum (SFS): Tajima's *D* (Tajima 1989), which evaluates the difference between low- and intermediate-frequency variants; Zeng et al.'s *E* (Zeng *et al.* 2006), which evaluates the difference between low- and high-frequency variants; and the normalized Fay and Wu's *H* (H_n; Zeng *et al.* 2006), which evaluates the difference between intermediate- and high-frequency variants. These statistics were computed, separately, for the 64 target loci and the 308 reference loci, and compared with those under a standard neutral equilibrium model (constant population size).

All nucleotide diversity statistics were computed using MANVa and mstatspop (https://bioinformatics.cragenomica.es/numgenomics/people/sebas/software/software.html), and, when needed, the SDMTools package in R was used to correct for sample size variation across loci, by weighting means and standard deviations.

Neutrality tests robust to demography

To identify potential genes under selection, we first used the SnIPRE approach (Eilertson *et al.* 2012), which considers polymorphism and divergence data from synonymous/silent and non-synonymous sites under a Poisson Random Effect model. This method is based on the McDonald-Kreitman test, and it can reliably identify genes under weak and strong negative as well as positive selection, without requiring the specification of a population genetic model. We considered both the selection (specific selection effect of a gene relative to neutrality) and constraint (proportion of non-synonymous mutations that are non-lethal, thus having effects on counts) effects provided by the program. The selection effect is useful to identify selection on mildly deleterious and advantageous mutations, while the constraint effect is useful to identify strong negative or purifying selection (Eilertson *et al.* 2012). Thus, significant effects on sequence data are classified as being neutral, negative or positive.

Second, we used the McDonald-Kreitman Poisson Random Fields (MKPRF) test (Bustamante *et al.* 2002, 2005), which implements a Markov chain Monte Carlo (MCMC) Bayesian approach to estimate parameters of the Poisson Random Field (PRF) model (Sawyer & Hartl 1992), allowing to compare groups of genes for selection signatures. MKPRF not only identifies non-neutrally evolving loci but it also estimates the associated strength of selection ($\gamma=2N_e s$). The MKPRF analysis was run using the *mkprf* program (kindly provided by Carlos D. Bustamante and Adam Boyko), with the following

parameters: 10 independent MCMC chains per run and 10,000 samples from each chain drawn on every 10 steps, after burn-in of 1,000 steps. Default values were used for prior distributions and other MCMC parameters. Convergence of the MCMC runs was checked in the MCMC output files generated by the program.

Distribution of fitness effects (DFE) and adaptive evolution

We used a method based on the DFE to estimate the overall rates of adaptive substitutions (ω_a): the method II of Eyre-Walker and Keightley (2009) as implemented in the DoFE 3.1 program. This method accounts for demographic changes that can affect the shape of the SFS by comparing the observed folded SFS at neutral sites to the folded SFS expected from neutral mutations in a stationary population at equilibrium and assumes that demography has the same proportional effect on the SFS of selected sites.

Input files were created using Python and Perl scripts kindly provided by Jaakko Tyrmi and Adrian Schneider, respectively (available upon request). To build the input files for the method II of Eyre-Walker and Keightley (2009), we computed the statistics for non-synonymous sites and synonymous sites, at the intraspecific level (for Scots pine and maritime pine, respectively) and at the interspecific level (using loblolly pine as outgroup for each target pine) using the same number of alleles per species and dataset: 305 loci with 10 alleles each for Scots pine and 291 loci with 11 alleles for maritime pine for the CRSP datasets; 126 loci with 23 alleles each for the CRIEC dataset. The other input statistics were then computed with MANVa (see above) using the same reduced datasets.

Loblolly pine was used as outgroup for both Scots pine and maritime pine and, thus, DFE-based estimates of adaptive evolution for these two pine species were done along a partially shared branch. In phylogenetically close species, shared divergence with the outgroup may result in differences in ω_a mainly due to differences in within-species

polymorphism, i.e., estimates can be biased if the nucleotide divergence between the species is low relative to within species variation (Keightley & Eyre-Walker 2012). To exclude this potential source of bias, shared divergence between loblolly pine and Scots pine/maritime pine was estimated using two approaches. First, one, three or five sequences from each locus and species were randomly sampled and concatenated, leading to three aligned fasta files. Aligned fasta files were subsequently used to build unrooted phylogenetic trees using PhyML (Guindon & Gascuel 2003). The analyses were run with default parameters, using the General Time Reversible (GTR) model (Tavaré 1986), and for all sites, and both 0-fold and 4-fold degenerate sites. The proportion of shared divergence was estimated then as the fraction of the sum of branch lengths due to the common branch leading to loblolly pine (i.e., the common branch leading to loblolly pine divided by the total sum of branch lengths). Second, fixed differences between loblolly pine and both Scots pine and maritime pine were counted, and shared fixed differences were directly obtained for all sites as well as for 0-fold and 4-fold degenerate sites.

Results

Nucleotide diversity, genetic divergence and overall patterns of polymorphism

The two pines showed a significant difference (non-overlapping 95% confidence intervals) in nucleotide diversity, with Scots pine ($\theta_{w\text{-syn}}$=0.00687, 95% CI: 0.00594-0.00781) displaying 1.64 times the nucleotide diversity of maritime pine ($\theta_{w\text{-syn}}$=0.00419, 95% CI: 0.00362-0.00476) (Table 1). In both species, the overall K_a, K_s, and K_a/K_s ratios (relative to loblolly pine) were similar, with overlapping 95% confidence intervals (data not shown), indicating equal divergence from loblolly pine (K_s of 0.0323 for Scots pine and of 0.0375

for maritime pine) and similar rates of evolutionary constraint (Table 1). Scots pine and maritime pine only shared about 2-4% of their synonymous polymorphic sites (depending on the species used as outgroup).

Statistics based on the SFS (Fay & Wu 2000; Tajima 1989; Zeng *et al.* 2006) revealed different polymorphism patterns in the two pines (Table 2). Scots pine displayed an excess of low-frequency variants compared to intermediate-frequency (Tajima's *D*) and high-frequency (Zeng et al.'s *E*) variants, but no differences between intermediate- and high-frequency (Fay and Wu's H_n) variants. In contrast, maritime pine displayed no difference of intermediate- with respect to low-frequency variants (Tajima's *D*), and an excess of high-frequency variants compared to low-frequency (Zeng et al.'s *E*) and intermediate-frequency (Fay and Wu's H_n) variants. Thus, overall, Scots pine displayed an excess of low-frequency variants, while maritime pine showed an excess of high-frequency variants. This pattern is present in both target and reference loci suggesting that demographic processes (and not selective processes) underlie the observed differences between the two pines. The impact of population structure on the statistics is reflected especially by Fay and Wu's H_n, which is more sensitive to population subdivision (Zeng *et al.* 2006), and that deviates more strongly from the standard neutral model in the species with stronger population structure (i.e., maritime pine).

Neutrality tests robust to demography

The robust-to-demography SnIPRE method (Empirical Bayes and Bayesian SnIPRE tests) suggested four and three target genes as being under selection in Scots pine (3 positive and 1 negative selection events) and maritime pine (2 positive and 1 negative selection events), respectively (Table 3 and Table S5). Overall, the 64 target loci had a positive average selection effect in Scots pine, while in maritime pine such effect was not observed (Figure

S2). All genes found under selection with the MKPRF test but one for Scots pine (five and two for Scots pine and maritime pine, respectively) were among those revealed by the SnIPRE method (Table 3). As expected, a higher proportion of loci from the previously-studied candidate gene dataset (see *Material and Methods*) was found under selection (50% and 25% based on the two methods in Scots pine and maritime pine, respectively) compared to the CRSP dataset (0.27% and 0.55%, respectively). Two of these genes were common to both pines using the two methods, with the same pattern of negative (*dhn1*) and positive (*coL1*) selection.

Distribution of fitness effects (DFE) and adaptive evolution

The inferred DFE using DoFE was similar for the two pines, with most of the mutations being strongly deleterious and subject to purifying selection (Figure 2). However, there were also some differences in the DFE spectrum. Scots pine had a significantly higher proportion of deleterious mutations for the three classes with the least deleterious mutations, while the opposite pattern was observed at the other extreme of the DFE spectrum with about 77% of mutation being highly deleterious for Scots pine and about 85% for maritime pine. Increasing the sample size for maritime pine (the "extended CRSP" dataset, see *Sampling* section) resulted in higher proportion of the three classes with the least deleterious mutations but still in lower proportion compared to Scots pine, while the pattern was the opposite for highly deleterious mutations (Figure 2). In terms of adaptive substitutions, both species displayed a relative rate significantly different from zero (Scots pine ω_a = 0.1156 vs. maritime pine ω_a = 0.1535), with no significant differences between them (overlapping 95% CIs, see Figure 2). Increasing the sampling size for maritime pine ("extended CRSP" dataset) led to a slightly smaller ω_a (0.1447), not different from that estimated with the CRSP datasets for Scots pine and maritime pine (Figure 2). Estimating ω_a with a simple

and non-parametric method $\omega_a = \frac{d_N - d_S(P_N/P_S)}{d_S}$ (Gossmann *et al.* 2010; Kousathanas *et al.* 2014) gave ω_a = -0.0740 with 95% CIs (-0.2430; 0.0871) for Scots pine and ω_a = 0.0976 with 95% CIs (-0.0778; 0.2867) for maritime pine. These estimates confirmed the trend of a higher ω_a in maritime pine than in Scots pine, albeit differences were still not significant (i.e., overlapping CIs).

Shared divergence between loblolly pine and Scots pine/maritime pine was estimated using two approaches. First, a three-species unrooted phylogenetic tree was built using 168,534 bp of concatenated sequences, which showed only 43% of shared divergence due to the common branch from loblolly pine to Scots pine/maritime pine (Figure S3). A similar estimate was obtained for 0-fold and 4-fold degenerate sites (43 to 44% of shared divergence). Furthermore, we obtained the same estimates with one, three or five sequences randomly-sampled from each locus and species. Second, counts of fixed differences between loblolly pine and Scots pine/maritime pine showed 629 (all sites), 152 (0-fold degenerate sites) and 106 (4-fold degenerate sites) mutations specific to the Scots pine lineage, while these numbers were 960, 196 and 185 mutations, respectively, for maritime pine (Table S6).

Discussion

The comparison of the same sets of loci between Scots pine and maritime pine, two related species that occupy different ecological niches and are characterized by different evolutionary histories, allows for exploring how different biological factors may interact with natural selection and adaptive evolution in plants. We discuss how our results are in agreement with each species evolutionary history, as well as how the comparison between

401 species helps to better understand the evolutionary forces responsible for the observed
402 molecular pattern.
403
404 *Nucleotide diversity*
405 This study presents the first large set of loci sequenced in a range-wide sample of two
406 important European conifers. The primers used in this study were transferred from a related
407 New World species, loblolly pine, and thus may suffer from ascertainment bias due to
408 enrichment with low-diversity conserved genes. To test this hypothesis, we compared a set
409 of 364 loblolly pine loci orthologous to those used in this study with 1,000 sets of 364 loci
410 randomly selected with replacement from a larger set of ca. 6,000 loci available in this
411 species (see Eckert *et al.* 2013b). This comparison indicated a significant bias towards
412 lower nucleotide diversity in our gene set (i.e., non-overlapping 95% confidence intervals),
413 but only when considering statistics for all sites (see Table S7). This suggests that
414 comparisons based on synonymous/silent sites or even non-synonymous sites in our study
415 are subject to only limited bias. Moreover, because the ascertainment bias would affect both
416 species in the same way, it would not prevent a comparative analysis between the two pines,
417 which is the main focus of this study.
418 The sequencing of the same 372 loci in both pines revealed significantly higher
419 nucleotide diversity in the widespread and continuously distributed Scots pine than in the
420 narrower and more patchily distributed maritime pine. Only a limited number of genes has
421 been sequenced for each pine (and rarely the same genes across species) in previous studies,
422 making interspecific comparisons difficult. Watterson's nucleotide diversity for silent sites
423 (compare with Table S8) in Scots pine was 0.00525 for 16 loci (Pyhäjärvi *et al.* 2007), and
424 0.00620 for 11 loci (Kujala & Savolainen 2012); while that of maritime pine was 0.00824
425 for 11 loci (Eveno *et al.* 2008), but was only 0.00280 for six other loci (Grivet *et al.* 2011)

To test whether using a smaller gene set could lead to biased nucleotide diversity estimates, we subsampled different number of loci from the total 372 in our study, and compared nucleotide diversity for these subsamples. Subsampling for 10, 50 or 100 loci did not affect average nucleotide diversity estimates (see confidence intervals in Table S9), suggesting that the number of genes used did not necessarily lead to bias. Nevertheless, subsets of specific genes may indeed have substantially different levels of nucleotide diversity. For example, the set of six previously-studied candidates genes (see *Material and methods*) showed higher nucleotide diversity (Scots pine $\theta_{w\text{-}silent}$=0.00895; maritime pine $\theta_{w\text{-}silent}$=0.00664) than the 372-locus average (Table S8). These results show that nucleotide diversity can only be properly compared across species when using a large set of (preferably) common genes, as this estimate can be very variable across small specific gene sets.

Insights into demographic history

Patterns within the SFS for each species (as summarized by statistics such as Tajima's *D*, Zeng et al.'s *E*, and Fay and Wu's H_n) can provide insights into demographic history (see references in Gravel *et al.* 2011). In Scots pine and maritime pine, they reflected relatively well what is known for each pine species. Scots pine has likely been through a very ancient (~1-2 Ma) and severe (shrinking populations to about 1% of present time population size) bottleneck (Pyhäjärvi *et al.* 2007). In more recent times, the species would have recolonized a vast territory in northern, central and Eastern Europe, reaching northern Fennoscandia some 10,000 to 7,000 years ago (see Pyhäjärvi *et al.* 2007, 2008; Savolainen *et al.* 2011; Kujala *et al.* 2012; Cheddadi *et al.* 2006; Naydenov *et al.* 2007). This wide-range colonization process would have led to a weak population genetic structure in Scots pine (except especially at the southern margins) (Cheddadi *et al.* 2006; Dvornyk *et al.* 2002;

García-Gil *et al.* 2003; Karhu *et al.* 1996; Pyhäjärvi *et al.* 2007). The excess of low-frequency variants in this species could have originated during repeated long-range expansions and would correspond (mostly) to relatively new mutations (but likely before the most recent ice age). In contrast, maritime pine has likely survived in multiple glacial refugia (Bucci *et al.* 2007; Burban & Petit 2003), from which it would have recolonized its current range around the Last Glacial Maximum (~20,000 years ago; Naydenov *et al.* 2014). Its spatially limited expansion, combined with population fragmentation, would have led to distinct and regionally restricted gene pools (Bucci *et al.* 2007; Burban & Petit 2003; Jaramillo-Correa *et al.* 2015), which would have been relatively stable across time, leading to the accumulation of high-frequency variants when considering full-range SFS patterns. Because of the very distinct demographic histories of the two pines, we used methods integrating demography to characterize their pattern of adaptive evolution at the molecular level and to test loci for signatures of positive and negative selection.

Genes under selection

Three dehydrins were found to be under selection in the two pines. Previous studies looking at various members of this family, and using different approaches (neutrality tests, F_{ST}-outlier detection, allele frequency-environment correlations), also showed the action of natural selection on dehydrin genes: *dhn1*, *dhn3* and *dhn9* in Scots pine (Palmé *et al.* 2009; Wachowiak *et al.* 2009); *dhn1* (Eveno *et al.* 2008), *dhn2* and *dhn5* (Grivet *et al.* 2011) in maritime pine. In this study, two new dehydrin genes (*dhn2* and *dhn5*) were identified as possible targets of positive selection in Scots pine, while our results confirmed the adaptive role of *dhn1* in maritime pine. Gene expression of dehydrins in water stress experiments pointed to their involvement in drought resistance in maritime pine (Perdiguero *et al.* 2012; Velasco-Conde *et al.* 2012). They have also been shown to be involved in wounding, cold,

476 and drought stress response in white spruce and loblolly pine (Lorenz *et al.* 2011; Richard
477 *et al.* 2000; Watkinson *et al.* 2003). Altogether these studies point to the pivotal role of
478 dehydrins in conifer adaptive response to abiotic stress, over the short- (F_{ST}) and long-term
479 (SnIPRE) timescale, emphasizing their importance.

480 An interesting case is that of locus *0_4042_01* which showed a signal of positive
481 selection in maritime pine (16 fixed non-synonymous mutations compared to only three in
482 Scots pine). Locus *0_4042_01* shows high similarity (E-value of 2×10^{-21}) with *glutathione*
483 *S-transferase* (*gst*) in *Arabidopsis thaliana*, an enzyme involved in secondary metabolism
484 response to the processes of detoxification and stress response to cold (Goulas *et al.* 2006),
485 salt (Jiang *et al.* 2007), and pathogens (Jones *et al.* 2006). In the Chinese pine, *Pinus*
486 *tabuliformis*, five residues within *gst* were found under positive selection, four of them
487 involved in the enzyme activity and specificity (Lan *et al.* 2013). Interestingly, one of these
488 residues codes for a different amino acid in maritime pine (a proline) compared to loblolly
489 pine (alanine), while in Scots pine most individuals maintain the putative ancestral form.
490 This amino acid (positions 31-33) is located very close to the catalytically active G-site in
491 the spatial conformation of the protein, and the substitution of this residue could cause
492 structural changes in the GSH binding pocket (where the conjugation of intracellular
493 glutathione to a wide variety of molecules occurs; Lan *et al.* 2013).

494 Finally, *constans-like 1* (*coL1*), which was found under positive selection in both
495 pines, codes for a putative transcription factor suggested to affect flower development in
496 *Arabidopsis* (Ledger *et al.* 2001). In trees, homologues to *constans* are involved in bud
497 development (Alberto *et al.* 2013b; Ruttink *et al.* 2007), as well as in photoperiodic control
498 of shoot elongation (Holefors *et al.* 2009). Evidence of selection on *constans*-like genes,
499 moreover, has been found in poplar (Chen *et al.* 2014; Ma *et al.* 2010; Ruttink *et al.* 2007;

Smith *et al.* 2004; Wei *et al.* 2013), spruce (Holliday *et al.* 2010), oak (Lind-Riehl *et al.* 2014), and the perennial *Arabidopsis lyrata* (Mattila *et al.* 2016).

Distribution of fitness effects (DFE) and adaptive evolution

Scots pine and maritime pine case studies provided new insights on conifer adaptive evolution and the evolutionary forces shaping conifer genomes. The DFE was similar between the two pines and to other plants (e.g., Eckert *et al.* 2013a; Gossmann *et al.* 2010), as well as to various other organisms (see references in Eyre-Walker & Keightley 2007), with most of the mutations being strongly deleterious and subject to purifying selection. Compared to other conifers (and plants) however (see Eckert *et al.* 2013a, and Hodgins *et al.* 2016), both targeted pines presented an atypical pattern with a lower proportion of slightly deleterious mutations and a higher proportion of highly deleterious mutations. The relative rates of new adaptive substitutions (ω_a) for Scots pine (0.1156) and maritime pine (0.1535) were within the range found in other species (between -0.14 and 0.31; Gossmann *et al.* 2012), although at the upper range limit for plants (Gossmann *et al.* 2010, 2012) as well as higher by a factor over two than those estimated with the same methodology in other pines (ω_a= from -0.0477 to 0.0325 for 11 species of soft pines in Eckert *et al.* 2013a; ω_a= 0.0592 for lodgepole pine in Hodgins *et al.* 2016).

Several outcomes emerge from these results. First, sampling intensity may bias the estimate of DFE and ω_a, as illustrated by the "extended CRSP dataset" that led to weaker differences between the two pines. Second, albeit very distinct in terms of their evolutionary histories (i.e., demographic history, population structure, and effective population size), the two pines present similar rates of adaptive evolution. This suggests that other factors may also govern the efficacy of selection across these taxa (see below). Third, it is noteworthy to highlight the high efficiency of natural selection at purging highly deleterious mutations as

well as the high rate of positive selection in both pines, in regards to plants in general and more specifically to other conifers studied so far. Plants tend to have low rates of adaptive evolution, linked to contracting populations and a high level of population structure (Gossmann *et al.* 2010). Since pines present in general large effective population size, being also relatively undomesticated species, it is expected that they display also high rates of adaptive evolution. However, this trend was not found in previous studies including pines, suggesting that other factors, still largely unknown, may also be relevant (Chen *et al.* 2017).

The reliability of our estimates may depend on methodological factors, among them the choice of the loci under study, the outgroup species, or the analytical method, as well as the sampling intensity and distribution (Eckert *et al.* 2013a; Phifer-Rixey *et al.* 2012; Städler *et al.* 2009). More specifically, (i) because we used mostly primers transferred from another species (loblolly pine), gene sets in our study may be more conserved and thus undergo less adaptive evolution (Bachtrog 2008; Eckert *et al.* 2013a; Gossmann *et al.* 2010); notice that this would make even more remarkable the high ω_a found in both pines. (ii) Different genic regions may be under different evolutionary constraints (e.g., Hodgins *et al.* 2016), and therefore lead to different estimates of adaptive evolution. We computed DFE and ω_a for silent sites (intron + untranslated regions, UTR) and found lower rates of positive selection (data not shown) although not significantly different from those estimated with synonymous sites. (iii) Recombination and mutation rates may differ among selected loci and thus can directly affect the proportion of segregating sites (Bachtrog 2008). (iv) The level of divergence of the species of interest with the outgroup may reveal different proportion of segregating sites, i.e., the closer the outgroup the fewer differences will be detected (Gossmann *et al.* 2010; Strasburg *et al.* 2011). In our study, low phylogenetic distance and partial sharing by the targeted species of the branch conducting to the outgroup could have overestimated ω_a (Keightley & Eyre-Walker 2012). However, phylogenetic

analyses indicated less than 50% of shared divergence between loblolly pine and each of the targeted species suggesting that observed patterns were not only due to within-species polymorphism. Furthermore, counts of fixed differences between the outgroup and each of the targeted species pointed to sufficient level of lineage specific mutations in both Scots pine and maritime pine as to correctly estimate ω_a. (v) The choice of the methodology used to infer adaptive evolution (Eckert *et al.* 2013a; Welch 2006) may give different output. By selecting the same loci and outgroup, as well as the same methodology for the two pines and the main studies in other organisms, we attempted to control for these factors and ensured that the estimates (although probably conservative, see point (i)) were comparable across species. Finally, both sampling intensity and distribution may have affected our estimate of adaptive evolution rate. (vi) To test effects of sampling intensity, in terms of both individuals and loci, first, the CRSP dataset (less samples per population but more loci) was compared with the "extended CRSP" dataset (more samples per population but fewer loci) in maritime pine and, second, simulated datasets with different number of loci were compared. Neither approach suggested any effect of sampling intensity on our estimates. (vii) Sampling scheme influences what aspects of the history are emphasized (see Städler *et al.* 2009). Scots pine populations were sampled from both the main range (Finland, Sweden and Poland) and the margins (Spain, Italy and UK), representing fairly well the species evolutionary history overall. Nevertheless, there may still be a bias in Scots pine estimates, as the CRSP loci were not sampled across the full species range (in particular the eastern-northern range was poorly sampled). Thus, some apparently fixed non-synonymous mutations may show polymorphism in the unsampled range, resulting in an upward bias. In maritime pine, however, this bias should be minimal (if any) as populations were sampled across its full range, considering all gene pools known in the species. In addition, population genetic structure does not seem to have affected the

proportion of adaptive substitutions, with estimates being similar between the weakly structured Scots pine and the highly structured maritime pine, and still higher than published estimates for other species. Finally, other factors may constrain selective forces, as suggested by comparative studies and theoretical work, and may have influenced the estimates of adaptive evolution in the two pines (Galtier 2016; Lanfear *et al.* 2014): genetic draft and background selection (Messer & Petrov 2013; Peischl *et al.* 2013), environmental heterogeneity (Tellier *et al.* 2011), phenotypic space dimensionality (i.e., fitness landscape) and rate of environmental change (Gillespie 2001; Lourenço *et al.* 2013). These factors are challenging to tease apart.

Conclusion

By analyzing a common set of 372 gene loci, we detected specific patterns of molecular evolution and adaptation in two widespread European conifers. First, as expected, nucleotide diversity was higher in the continuously distributed Scots pine than in the patchily distributed maritime pine. Second, by using methods that incorporate demographic effects, we detected an unexpected high relative rate of adaptive substitutions in both pines, and in particular in maritime pine. Although we cannot fully discard methodological caveats, these high rates of adaptive evolution do not seem to be correlated with population genetic structure nor demographic histories that differ between the two pines. Altogether, our results suggest that more than one factor may be responsible of the high rate of adaptive evolution found in these two emblematic pine species, with several factors being difficult to tease apart.

Acknowledgements

We are grateful to Carmen García-Barriga (laboratory), Mario Zabal-Aguirre (sequence editing), Kirsten Eilertson (SnIPRE), Jaakko Tyrmi (Python script), Adrian Schneider (method for inferring the DFE), Sebastián E. Ramos-Onsins (MANVa) and Athanasios Kousathanas (DFE) for valuable information exchanges, as well as to Adam Eyre-Walker and various anonymous reviewers for constructive comments. This work was supported by the EU EVOLTREE Network of Excellence, AdapCon project (CGL2011-30182-C02-01) from the Spanish Ministry of Economy and Competiveness, TipTree project (BiodivERsA-ERANET) and EcoGenPin project (IdEx Bordeaux - Chaires d'installation 2015). D.G. acknowledges support from the Spanish Ministry of Economy and Competiveness through a 'Ramón y Cajal' fellowship.

References

Abad Viñas RA, Caudullo G, Oliveira S, de Rigo D (2016) *Pinus pinaster* in Europe: distribution, habitat, usage and threats. In: *European Atlas of Forest Tree Species*, p. e012d059+. Publ. Off. EU Luxembourg.

Alberto FJ, Aitken SN, Alía R, González-Martínez SC, Hänninen H, Kremer A, . . . Whetten R (2013a) Potential for evolutionary responses to climate change–evidence from tree populations. *Global Change Biology* **19**, 1645-1661.

Alberto FJ, Derory J, Boury C, Frigerio J-M, Zimmermann NE, Kremer A (2013b) Imprints of natural selection along environmental gradients in phenology-related genes of *Quercus petraea*. *Genetics* **195**, 495-512.

Alía R, Chambel R, Notivol E, Climent J, González-Martínez SC (2014) Environment-dependent microevolution in a Mediterranean pine (*Pinus pinaster* Aiton). *BMC Evolutionary Biology* **14**, 200.

Alía R, Gil L, Pardos J (1995) Performance of 43 Pinus pinaster Ait. provenances on 5 locations in central Spain. *Silvae Genetica* **44**, 75-80.

Andersson B, Fedorkov A (2004) Longitudinal differences in Scots pine frost hardiness. *Silvae Genetica*.

Bachtrog D (2008) Similar rates of protein adaptation in *Drosophila miranda* and *D. melanogaster*, two species with different current effective population sizes. *BMC Evolutionary Biology* **8**, 334.

Beck EH, Heim R, Hansen J (2004) Plant resistance to cold stress: Mechanisms and environmental signals triggering frost hardening and dehardening. *Journal of Biosciences* **29**, 449-459.

Bierne N, Eyre-Walker A (2004) The genomic rate of adaptive amino acid substitution in Drosophila. *Molecular Biology And Evolution* **21**, 1350-1360.

Biswas S, Akey JM (2006) Genomic insights into positive selection. *Trends Genet.* **22**, 437-445.

Bucci G, González-Martínez SC, Le Provost G, Plomion C, RIBEIRO M, Sebastiani F, . . . Vendramin GG (2007) Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Molecular Ecology* **16**, 2137-2153.

Burban C, Petit RJ (2003) Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology* **12**, 1487-1495.

Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, . . . Clark AG (2005) Natural selection on protein-coding genes in the human genome. *Nature* **437**, 1153-1157.

Bustamante CD, Nielsen R, Hartl DL (2003) Maximum likelihood and Bayesian methods for estimating the distribution of selective effects among classes of mutations using DNA polymorphism data. *Theoretical Population Biology* **63**, 91-103.

Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD, Hartl DL (2002) The cost of inbreeding in Arabidopsis. *Nature* **416**, 531-534.

Castro J, Zamora R, Hódar JA, Gómez JM (2002) Use of shrubs as nurse plants: a new technique for reforestation in Mediterranean mountains. *Restoration ecology* **10**, 297-305.

Chambel MR, Climent J, Alia R (2007) Divergence among species and populations of Mediterranean pines in biomass allocation of seedlings grown under two watering regimes. *Annals of Forest Science* **64**, 87-97.

660 Chancerel E, Lepoittevin C, Le Provost G, Lin YC, Jaramillo-Correa JP, Eckert AJ, . . .
661 Plomion C (2011) Development and implementation of a highly-multiplexed
662 SNP array for genetic mapping in maritime pine and comparative mapping with
663 loblolly pine. *BMC Genomics* **12**.
664 Cheddadi R, Vendramin GG, Litt T, François L, Kageyama M, Lorentz S, . . . Jost A
665 (2006) Imprints of glacial refugia in the modern genetic diversity of *Pinus*
666 *sylvestris*. *Global Ecology and Biogeography* **15**, 271-282.
667 Chen J, Glémin S, Lascoux M (2017) Genetic diversity and the efficacy of purifying
668 selection across plant and animal species. *Molecular Biology And Evolution* **34**,
669 1417-1428.
670 Chen Z, Cao Z, Yu Z (2014) Identification and characterization of differentially
671 expressed genes during incompatible interaction between the foliar rust
672 *Melampsora larici-populina* and poplar. *Genet Mol Res* **13**, 2082-2093.
673 Corcuera L, Gil-Pelegrin E, Notivol E (2011) Intraspecific variation in *Pinus pinaster*
674 PSII photochemical efficiency in response to winter stress and freezing
675 temperatures. *PLoS ONE* **6**, e28772.
676 Corcuera L, Gil-Pelegrín E, Notivol E (2012) Differences in hydraulic architecture
677 between mesic and xeric Pinus pinaster populations at the seedling stage. *Tree*
678 *Physiology*, **32**, 1442-1457.
679 Dvornyk V, Sirviö A, Mikkonen M, Savolainen O (2002) Low nucleotide diversity at
680 the pal1 locus in the widely distributed *Pinus sylvestsris*. *Mol. Biol. Evol.* **19**,
681 179-199.
682 Eckert AJ, Bower AD, González-Martínez SC, Wegrzyn JL, Coop G, Neale DB (2010)
683 Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae).
684 *Molecular Ecology* **19**, 3789-3805.
685 Eckert AJ, Bower AD, Jermstad KD, Wegrzyn JL, Knaus BJ, Syring JV, Neale DB
686 (2013a) Multilocus analyses reveal little evidence for lineage-wide adaptive
687 evolution within major clades of soft pines (Pinus subgenus Strobus). *Molecular*
688 *Ecology* **22**, 5635-5650.
689 Eckert AJ, Wegrzyn JL, Liechty JD, Lee JM, Cumbie WP, Davis JM, . . . Quesada T
690 (2013b) The evolutionary genetics of the genes underlying phenotypic
691 associations for loblolly pine (*Pinus taeda*, Pinaceae). *Genetics* **195**, 1353-1372.
692 Eilertson KE, Booth JG, Bustamante CD (2012) SnIPRE: Selection Inference Using a
693 Poisson Random Effects Model. *Plos Computational Biology* **8**.
694 Eveno E, Collada C, Guevara MA, Leger V, Soto A, Diaz L, . . . Garnier-Gere PH
695 (2008) "Contrasting Patterns of Selection at *Pinus pinaster* Ait. Drought Stress
696 Candidate Genes as Revealed by Genetic Differentiation Analyses". *Mol Biol*
697 *Evol* **25**, 417-437.
698 Excoffier L, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions.
699 *Annual Review of Ecology, Evolution, and Systematics* **40**, 481-501.
700 Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new
701 mutations. *Nature reviews. Genetics* **8**, 610.
702 Eyre-Walker A, Keightley PD (2009) Estimating the rate of adaptive molecular
703 evolution in the presence of slightly deleterious mutations and population size
704 change. *Molecular Biology And Evolution* **26**, 2097-2108.
705 Fay JC, Wu C-I (2000) Hitchhiking Under Positive Darwinian Selection. *Genetics* **155**,
706 1405-1413.
707 Galiano L, Martínez-Vilalta J, Lloret F (2010) Drought-induced multifactor decline of
708 Scots pine in the Pyrenees and potential vegetation change by the expansion of
709 co-occurring oak species. *Ecosystems* **13**, 978-991.

Galtier N (2016) Adaptive protein evolution in animals and the effective population size hypothesis. *Plos Genetics* **12**, e1005774.
García-Gil MR, Mikkonen M, Savolainen O (2003) Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. *Mol. Ecol.* **12**, 1195-1206.
Gaspar MJ, Velasco T, Feito I, Alía R, Majada J (2013) Genetic variation of drought tolerance in *Pinus pinaster* at three hierarchical levels: a comparison of induced osmotic stress and field testing. *PLoS ONE* **8**, e79094.
Gillespie JH (2001) Is the population size of a species relevant to its evolution? *Evolution* **55**, 2161-2169.
Gonzalez-Martinez SC, Alia R, Gil L (2002) Population genetic structure in a Mediterranean pine (*Pinus pinaster* Ait.): a comparison of allozyme markers and quantitative traits. *Heredity* **89**, 199-206.
Gossmann T, Keightley P, Eyre-Walker A (2012) The Effect of Variation in the Effective Population Size on the Rate of Adaptive Molecular Evolution in Eukaryotes. *Genome Biology and Evolution* **4**, 658-667.
Gossmann TI, Song B-H, Windsor AJ, Mitchell-Olds T, Dixon CJ, Kapralov MV, . . . Eyre-Walker A (2010) Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Mol Biol Evol* **27**, 1822-1832.
Goulas E, Schubert M, Kieselbach T, Kleczkowski LA, Gardestrom P, Schroder W, Hurry V (2006) The chloroplast lumen and stromal proteomes of *Arabidopsis thaliana* show differential sensitivity to short- and long-term exposure to low temperature. *Plant Journal* **47**, 720-734.
Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, . . . Altshuler DL (2011) Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences* **108**, 11983-11988.
Grivet D, Climent J, Zabal-Aguirre M, Neale DB, Vendramin GG, González-Martínez SC (2013) Adaptive evolution of Mediterranean pines. *Molecular phylogenetics and evolution* **68**, 555-566.
Grivet D, Sebastiani F, Alía R, Bataillon T, Torre S, Zabal-Aguirre M, . . . González-Martínez SC (2011) Molecular Footprints of Local Adaptation in Two Mediterranean Conifers. *Molecular Biology And Evolution* **28**, 101-116.
Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**, 696-704.
Gullberg U (1985) Allozyme variation in Scots pine (*Pinus sylvestris* L.) in Sweden. *Silvae. Genet.* **34**, 193-201.
Hodgins KA, Yeaman S, Nurkowski KA, Rieseberg LH, Aitken SN (2016) Expression divergence is correlated with sequence evolution but not positive selection in conifers. *Molecular Biology And Evolution* **33**, 1502-1516.
Holefors A, Opseth L, Rosnes AKR, Ripel L, Snipen L, Fossdal CG, Olsen JE (2009) Identification of PaCOL1 and PaCOL2, two CONSTANS-like genes showing decreased transcript levels preceding short day induced growth cessation in Norway spruce. *Plant Physiology and Biochemistry* **47**, 105-115.
Holliday JA, Ritland K, Aitken SN (2010) Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytologist* **188**, 501-514.
Hurme P, Repo T, Savolainen O, Paakkonen T (1997) Climatic adaptation of bud set and frost hardiness in Scots pine (*Pinus sylvestris*). *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* **27**, 716-723.

760 Jaramillo-Correa J-P, Rodríguez-Quilón I, Grivet D, Lepoittevin C, Sebastiani F,
761 Heuertz M, . . . Vendramin GG (2015) Molecular proxies for climate
762 maladaptation in a long-lived tree (*Pinus pinaster* Aiton, Pinaceae). *Genetics*
763 **199**, 793-807.
764 Jiang Y, Yang B, Harris NS, Deyholos MK (2007) Comparative proteomic analysis of
765 NaCl stress-responsive proteins in Arabidopsis roots. *Journal Of Experimental*
766 *Botany* **58**, 3591-3607.
767 Jones AM, Thomas V, Bennett MH, Mansfield J, Grant M (2006) Modifications to the
768 Arabidopsis defense proteome occur prior to significant transcriptional change
769 in response to inoculation with *Pseudomonas syringae*. *Plant Physiology* **142**,
770 1603-1620.
771 Karhu A, Hurme P, Karjalainen M, Karvonen P, Kärkkäinen K, Neale D, Savolainen O
772 (1996) Do molecular markers reflect patterns of differentiation in adaptive traits
773 of conifers? *TAG Theoretical and Applied Genetics* **93**, 215-221.
774 Keightley PD, Eyre-Walker A (2012) Estimating the rate of adaptive molecular
775 evolution when the evolutionary divergence between species is small. *Journal*
776 *Of Molecular Evolution* **74**, 61-68.
777 Kousathanas A, Halligan DL, Keightley PD (2014) Faster-X adaptive protein evolution
778 in house mice. *Genetics* **196**, 1131-1143.
779 Krakau U-K, Liesebach M, Aronen T, Lelu-Walter M-A, Schneck V (2013) Scots pine
780 (*Pinus sylvestris* L.). In: *Forest tree breeding in Europe*, pp. 267-323. Springer.
781 Kremer A, Roussel G (1986) Subdivision of shoot growth of maritime pine (*Pinus*
782 *pinaster* Ait.). Geographic variation of morphogenetical and phenological
783 components [phyllotaxis, polycyclism, plastochron lenght time, provenances]
784 **43**.
785 Kujala S, Savolainen O (2012) Sequence variation patterns along a latitudinal cline in
786 Scots pine (*Pinus sylvestris*): signs of clinal adaptation? *Tree Genetics &*
787 *Genomes* **8**, 1451-1467.
788 Lamy J-B, Bouffier L, Burlett R, Plomion C, Cochard H, Delzon S (2011) Uniform
789 selection as a primary force reducing population genetic differentiation of
790 cavitation resistance across a species range. *PLoS ONE*, **6**, e23476.
791 Lamy JB, Delzon S, Bouche PS, Alia R, Vendramin GG, Cochard H, Plomion C (2014)
792 Limited genetic variability and phenotypic plasticity detected for cavitation
793 resistance in a Mediterranean pine. *New Phytologist* **201**, 874-886.
794 Lan T, Wang X-R, Zeng Q-Y (2013) Structural and functional evolution of positively
795 selected sites in pine glutathione S-transferase enzyme family. *Journal of*
796 *Biological Chemistry* **288**, 24441-24451.
797 Lanfear R, Kokko H, Eyre-Walker A (2014) Population size and the rate of evolution.
798 *Trends in Ecology & Evolution* **29**, 33-41.
799 Ledger S, Strayer C, Ashton F, Kay SA, Putterill J (2001) Analysis of the function of
800 two circadian-regulated CONSTANS- LIKE genes. *The Plant Journal* **26**, 15-
801 22.
802 Li J, Li H, Jakobsson M, Li SEN, SjÖDin PER, Lascoux M (2012) Joint analysis of
803 demography and selection in population genetics: where do we stand and where
804 could we go? *Molecular Ecology* **21**, 28-44.
805 Lind-Riehl JF, Sullivan AR, Gailing O (2014) Evidence for selection on a
806 CONSTANS-like gene between two red oak species. *Annals Of Botany* **113**,
807 967-975.

Lorenz WW, Alba R, Yu Y-S, Bordeaux JM, Simões M, Dean JF (2011) Microarray analysis and scale-free gene networks identify candidate regulators in drought-stressed roots of loblolly pine (*P. taeda* L.). *BMC Genomics* **12**, 264.

Lourenço JM, Glémin S, Galtier N (2013) The rate of molecular adaptation in a changing environment. *Molecular Biology And Evolution* **30**, 1292-1301.

Löytynoja A, Goldman N (2005) An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **102**, 10557-10562.

Ma X-F, Hall D, Onge KRS, Jansson S, Ingvarsson PK (2010) Genetic differentiation, clinal variation and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. *Genetics* **186**, 1033-1044.

Mattila TM, Aalto EA, Toivainen T, Niittyvuopio A, Piltonen S, Kuittinen H, Savolainen O (2016) Selection for population-specific adaptation shaped patterns of variation in the photoperiod pathway genes in *Arabidopsis lyrata* during post-glacial colonization. *Molecular Ecology* **25**, 581-597.

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351**, 652-654.

Messer PW, Petrov DA (2013) Frequent adaptation and the McDonald–Kreitman test. *Proceedings of the National Academy of Sciences* **110**, 8615-8620.

Mikola J (1982) Bud-set phenology as an indicator of climatic adaptation of Scots pine in Finland. *Silva Fennica* **16**, 178-184.

Naydenov K, Senneville S, Beaulieu J, Tremblay F, Bousquet J (2007) Glacial vicariance in Eurasia: mitochondrial DNA evidence from Scots pine for a complex heritage involving genetically distinct refugia at mid-northern latitudes and in Asia Minor. *BMC Evolutionary Biology*, **7**, 233 doi:10.1186/1471-2148-7-233.

Naydenov KD, Alexandrov A, Matevski V, Vasilevski K, Naydenov MK, Gyuleva V, . . . Kamary S (2014) Range-wide genetic structure of maritime pine predates the last glacial maximum: evidence from nuclear DNA. *Hereditas* **151**, 1-13.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences* **76**, 5269-5273.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics* **39**, 197-218.

Notivol E, Garcia-Gil M, Alia R, Savolainen O (2007) Genetic variation of growth rhythm traits in the limits of a latitudinal cline in Scots pine. *Canadian Journal of Forest Research* **37**, 540-551.

Oleksyn J, Tjoelker MG, Reich PB (1998) Adaptation to changing environment in Scots pine populations across a latitudinal gradient. *Silva Fennica* **32**, 129-140.

Palmé A, Pyhäjärvi T, Wachowiak W, Savolainen O (2009) Selection on nuclear genes in a *Pinus* phylogeny. *Molecular Biology And Evolution* **26**, 893-905.

Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L (2013) On the accumulation of deleterious mutations during range expansions. *Molecular Ecology* **22**, 5972-5982.

Perdiguero P, Barbero MC, Cervera MT, Soto Á, Collada C (2012) Novel conserved segments are associated with differential expression patterns for Pinaceae dehydrins. *Planta* **236**, 1863-1874.

Petit RJ, Bahrman N, Baradat P (1995) Comparison of genetic differentiation in maritime pine (*Pinus pinaster* Ait.) estimated using isozyme, total protein and terpenic loci. *Heredity* **75**, 382-389.

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology* **14**, 689-701.
Phifer-Rixey M, Bonhomme F, Boursot P, Churchill GA, Piálek J, Tucker PK, Nachman MW (2012) Adaptive evolution and effective population size in wild house mice. *Molecular Biology And Evolution* **29**, 2949-2955.
Pyhäjärvi T, García-Gil MR, Knürr T, Mikkonen M, Wachowiak W, Savolainen O (2007) Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. *Genetics* **177**, 1713-1724.
Pyhäjärvi T, Salmela MJ, Savolainen O (2008) Colonization routes of *Pinus sylvestris* inferred from distribution of mitochondrial DNA variation. *Tree Genetics & Genomes* **4**, 247-254.
Ramírez-Valiente JA, Robledo-Arnuncio JJ (2014) Adaptive consequences of human-mediated introgression for indigenous tree species: the case of a relict *Pinus pinaster* population. *Tree Physiology* **34**, 1376-1387.
Richard S, Morency M-J, Drevet C, Jouanin L, Séguin A (2000) Isolation and characterization of a dehydrin gene from white spruce induced upon wounding, drought and cold stresses. *Plant Molecular Biology* **43**, 1-10.
Richardson DM (2000) *Ecology and biogeography of Pinus* Cambridge University Press.
Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm J, . . . Rohde A (2007) A molecular timetable for apical bud formation and dormancy induction in poplar. *The Plant Cell Online* **19**, 2370-2390.
Ryyppö A, Iivonen S, Rikala R, Sutinen ML, Vapaavuori E (1998) Responses of Scots pine seedlings to low root zone temperature in spring. *Physiologia Plantarum* **102**, 503-512.
Santos-del-Blanco L, Notivol E, Zas R, Chambel MR, Majada J, Climent J (2010) Variation of early reproductive allocation in multi-site genetic trials of maritime pine and Aleppo pine. *Forest Systems*, **19**, 381-392.
Savolainen O, Kujala ST, Sokol C, Pyhajarvi T, Avia K, Knurr T, . . . Hicks S (2011) Adaptive potential of northernmost tree populations to climate change, with emphasis on Scots pine (*Pinus sylvestris* L.). *J Hered* **102**, 526-536.
Savolainen O, Pyhäjärvi T, Knürr T (2007) Gene Flow and Local Adaptation in Trees. *Annual Review of Ecology, Evolution, and Systematics* **38**, 595-619.
Sawyer SA, Hartl DL (1992) Population genetics of polymorphism and divergence. *Genetics* **132**, 1161-1176.
Schrider DR, Kern AD (2016) S/HIC: Robust identification of soft and hard sweeps using machine learning. *Plos Genetics* **12**, e1005928.
Serra-Varela MJ, Grivet D, Vincenot L, Broennimann O, Gonzalo-Jiménez J, Zimmermann NE (2015) Does phylogeographical structure relate to climatic niche divergence? A test using maritime pine (*Pinus pinaster* Ait.). *Global Ecology and Geography*, **24**, 1302-1313.
Slotte T, Foxe JP, Hazzouri KM, Wright SI (2010) Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size. *Mol Biol Evol* **27**, 1813-1821.
Smith CM, Rodriguez-Buey M, Karlsson J, Campbell MM (2004) The response of the poplar transcriptome to wounding and subsequent infection by a viral pathogen. *New Phytologist* **164**, 123-136.
Smith NG, Eyre-Walker A (2002) Adaptive protein evolution in Drosophila. *Nature* **415**, 1022-1024.

Städler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P (2009) The Impact of Sampling Schemes on the Site Frequency Spectrum in Nonequilibrium Subdivided Populations. *Genetics* **182**, 205-216.
Strasburg JL, Kane NC, Raduski AR, Bonin A, Michelmore R, Rieseberg LH (2011) Effective Population Size Is Positively Correlated with Levels of Adaptive Divergence among Annual Sunflowers. *Molecular Biology And Evolution* **28**, 1569-1580.
Taberlet P, Fumagalli L, Wust-Saucy A, Cosson J (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**, 453-464.
Tajima F (1989) Statistical methods for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.
Tavaré S (1986) Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. In: *Lectures on Mathematics in the Life Sciences Volume 17* (ed. Miura RM), pp. 57-86. American Mathematical Society, Providence, RI, USA.
Tellier A, Fischer I, Merino C, Xia H, Camus-Kulandaivelu L, Städler T, Stephan W (2011) Fitness effects of derived deleterious mutations in four closely related wild tomato species with spatial structure. *Heredity* **107**, 189.
Velasco-Conde T, Yakovlev I, Majada J, Aranda I, Johnsen Ã (2012) Dehydrins in maritime pine (*Pinus pinaster*) and their expression related to drought stress response. *Tree Genetics & Genomes* **8**, 957-973.
Vendramin GG, Anzidei M, Madaghiele A, Bucci G (1998) Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites. *Theoretical and Applied Genetics* **97**, 456-463.
Wachowiak W, Balk P, Savolainen O (2009) Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in Scots pine (*Pinus sylvestris* L.). *Tree Genetics & Genomes* **5**, 117-132.
Watkinson JI, Sioson AA, Vasquez-Robinet C, Shukla M, Kumar D, Ellis M, . . . Watson LT (2003) Photosynthetic acclimation is reflected in specific patterns of gene expression in drought-stressed loblolly pine. *Plant Physiology* **133**, 1702-1716.
Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**, 256-276.
Wegrzyn JL, Lee JM, Tearse BR, Neale DB (2008) TreeGenes: a forest tree genome database. *International journal of plant genomics* **2008**.
Wei H, Yordanov YS, Georgieva T, Li X, Busov V (2013) Nitrogen deprivation promotes Populus root growth through global transcriptome reprogramming and activation of hierarchical genetic networks. *New Phytologist* **200**, 483-497.
Welch JJ (2006) Estimating the genomewide rate of adaptive protein evolution in Drosophila. *Genetics* **173**, 821-837.
Wright SI, Andolfatto P (2008) The impact of natural selection on the genome: emerging patterns in Drosophila and Arabidopsis. *Annual Review of Ecology, Evolution, and Systematics* **39**.
Zeng K, Fu Y-X, Shi S, Wu C-I (2006) Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* **174**, 1431-1439.

Data Accessibility

DNA sequences are deposited in GenBank under accessions MF385275-MF385581 and MF385585-MF397901.

Author Contribution

DG, OS and SCGM conceived the study. AJE, DBN and SCGM designed and produced the sequence data sets. DG and KA analyzed the data and drafted the manuscript. AV and SCGM contributed to analyze the data. All the authors contributed to editing and revising the manuscript.

Tables

Table 1. Summary statistics for 372 common loci sequenced in Scots pine and maritime pine.

	Scots pine	maritime pine
Number of loci	372	372
Total length (bp)	165,048	166,641
Segregating sites	1543	983
Average length (bp)	444	448
$\theta_{w\text{-syn}}$ (stdev)	0.0069 (0.0090)	0.0042 (0.0054)
π_{syn} (stdev)	0.0063 (0.0098)	0.0044 (0.0069)
$\theta_{w\text{-nsyn}}$ (stdev)	0.0015 (0.0027)	0.0008 (0.0014)
π_{nsyn} (stdev)	0.0013 (0.0025)	0.0009 (0.0018)
K_s (stdev)	0.0323 (0.0253)	0.0375 (0.0313)
K_a (stdev)	0.0070 (0.0081)	0.0070 (0.0080)
K_a / K_s (stdev)	0.3988 (1.0366)	0.3335 (0.8254)

bp=base pair
stdev=standard deviation
syn=synonymous
nsyn=nonsynonymous
θ_w: average Watterson's nucleotide diversity (Watterson 1975) per site
π: average Tajima's nucleotide diversity (Tajima 1989) per site
K_s: number of synonymous substitutions per synonymous site with Jukes-Cantor correction, using *P. taeda* as outgroup
K_a: number of non-synonymous substitutions per non-synonymous site with Jukes-Cantor correction, using *P. taeda* as outgroup

Table 2. SFS-based statistics for 64 target and 308 reference loci in Scots pine and maritime pine, with their 95% Confidence Interval in square brackets.

Locus set	Statistic	Scots pine	maritime pine
Target	Tajima's *D*	-0.5230 [-0.8559,-0.1901]	0.1195 [-0.2020,0.4410]
Target	Zeng et al.'s *E*	-0.3596 [-0.6483,-0.0709]	0.5309 [0.1473,0.9145]
Target	Fay and Wu's H_n	-0.0844 [-0.2788,0.1100]	-0.5269 [-0.9309,-0.1229]
Reference	Tajima's *D*	-0.4776 [-0.5831,-0.3720]	0.0106 [-0.1122,0.1335]
Reference	Zeng et al.'s *E*	-0.3376 [-0.4517,-0.2235]	0.2898 [0.1566,0.4230]
Reference	Fay and Wu's H_n	0.0610 [-0.0483,0.1703]	-0.3077 [-0.4587,-0.1567]

Table 3. Neutrality tests and type of selection for target genes in Scots pine and maritime pine. Symbols in parenthesis represent genes related to biotic stress (S) and phenology/photosystem (P).

Locus	Scots pine		maritime pine	
	SnIPRE*	MKPRF	SnIPRE*	MKPRF
dhn1 (S)	negative	negative		negative
dhn2 (S)	positive	positive		
dhn5 (S)	positive	positive		
coL1 (P)	positive	positive	positive	
0_4042_01 (S)			positive	positive
2_9480_01 (S)			negative	
0_12156_02 (P)		positive		

*Only genes in common between the Empirical Bayes and the Bayesian SnIPRE tests are reported. Outputs for each method are presented in full in Table S5.

Figure Legends

Figure1. Species distribution and sampling for Scots pine (blue, black dots) and maritime pine (green, red dots).

Figure 2. Distribution of Fitness Effects of new mutations (DFE) (a), and relative rate of adaptive substitutions ω_a (b) for the CRSP dataset in Scots pine and maritime pine, and for the "extended CRSP" dataset (only maritime pine), using the method II of Eyre-Walker and Keightley (2009), as implemented in DoFE. $N_e s$ denotes the product of the effective population size N_e and the strength of selection s, with $N_e s<1$ corresponding to slightly deleterious mutations and $N_e s>100$ corresponding to highly deleterious mutations. Bars in (a) represent standard errors (with different letters indicating significant differences), while bars in (b) represent 95% Confidence Intervals.

Figure1.

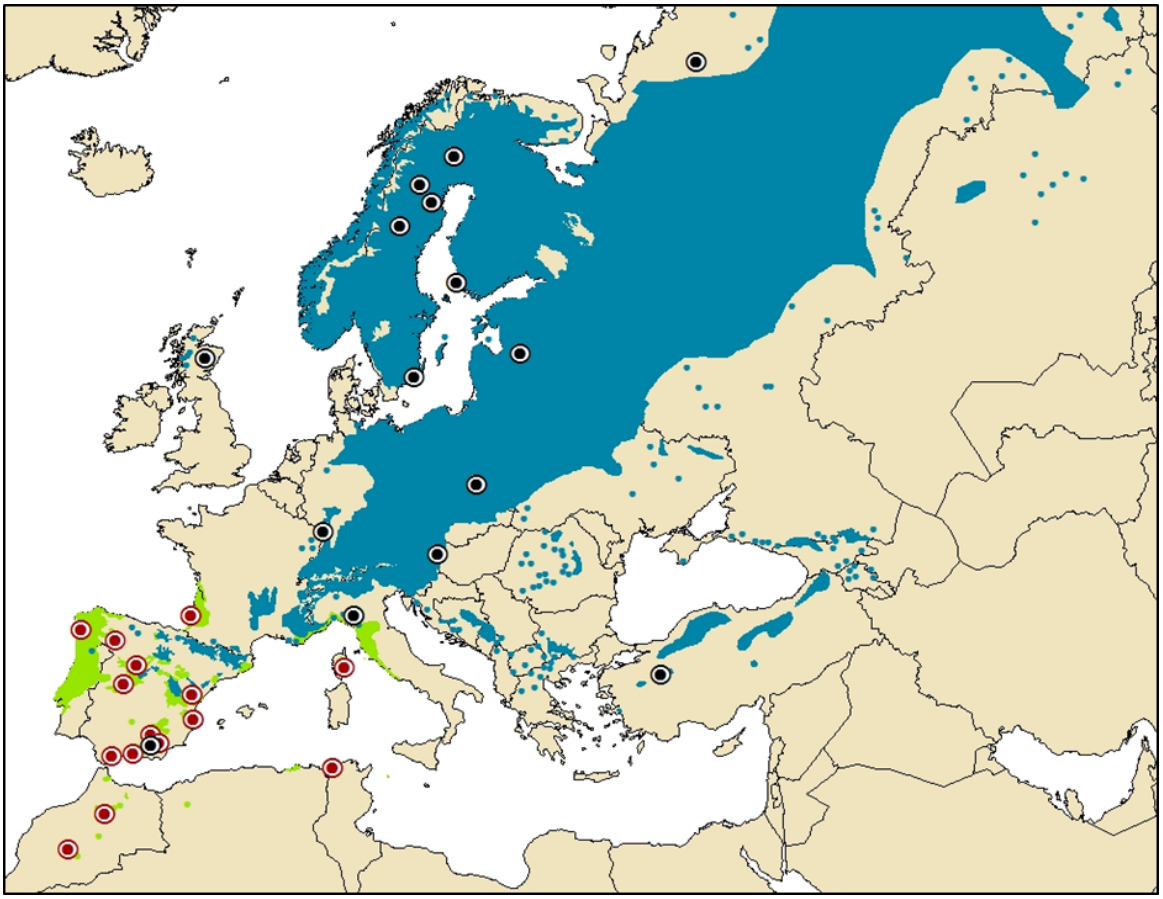

Figure 2.

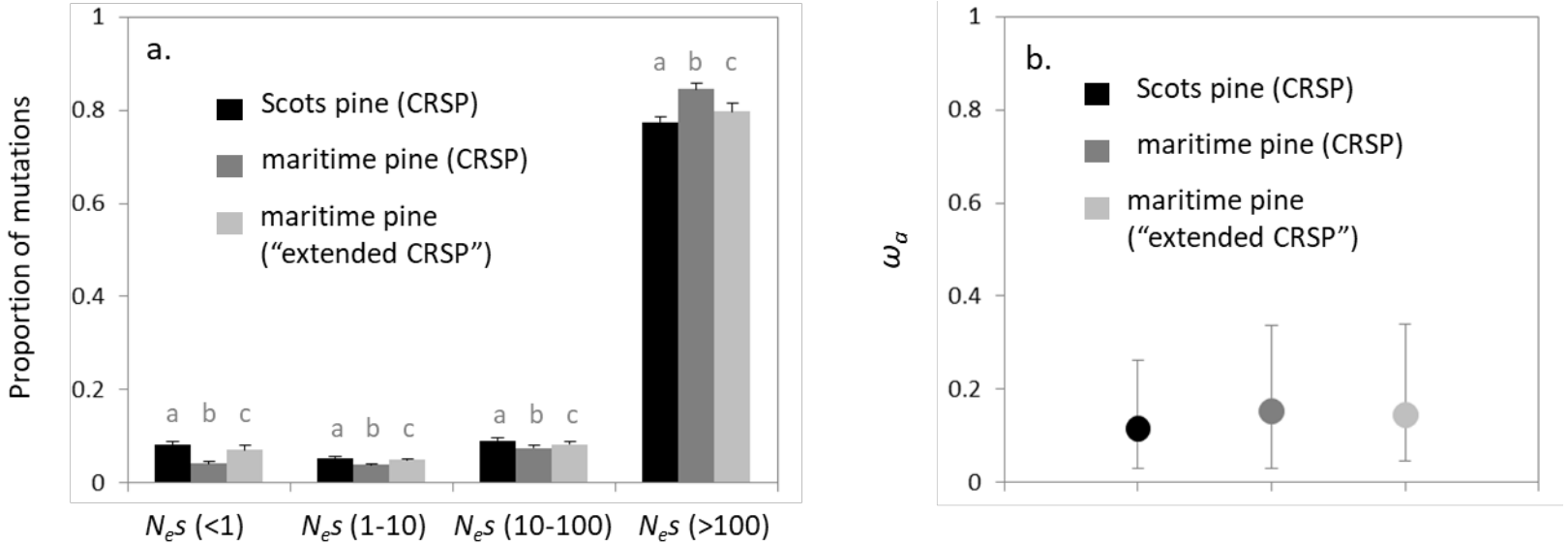
a.
Scots pine (CRSP)
maritime pine (CRSP)
maritime pine ("extended CRSP")
Proportion of mutations
1
0.8
0.6
0.4
0.2
0
a b c
a b c
a b c
a b c
$N_e s$ (<1)
$N_e s$ (1-10)
$N_e s$ (10-100)
$N_e s$ (>100)
b.
Scots pine (CRSP)
maritime pine (CRSP)
maritime pine ("extended CRSP")
ω_a
1
0.8
0.6
0.4
0.2
0

Figure S1. Molecular datasets used for the different analyses. Species abbreviations are as follows: Loblolly pine (Pita), Scots pine (Pisy) and maritime pine (Pipn).

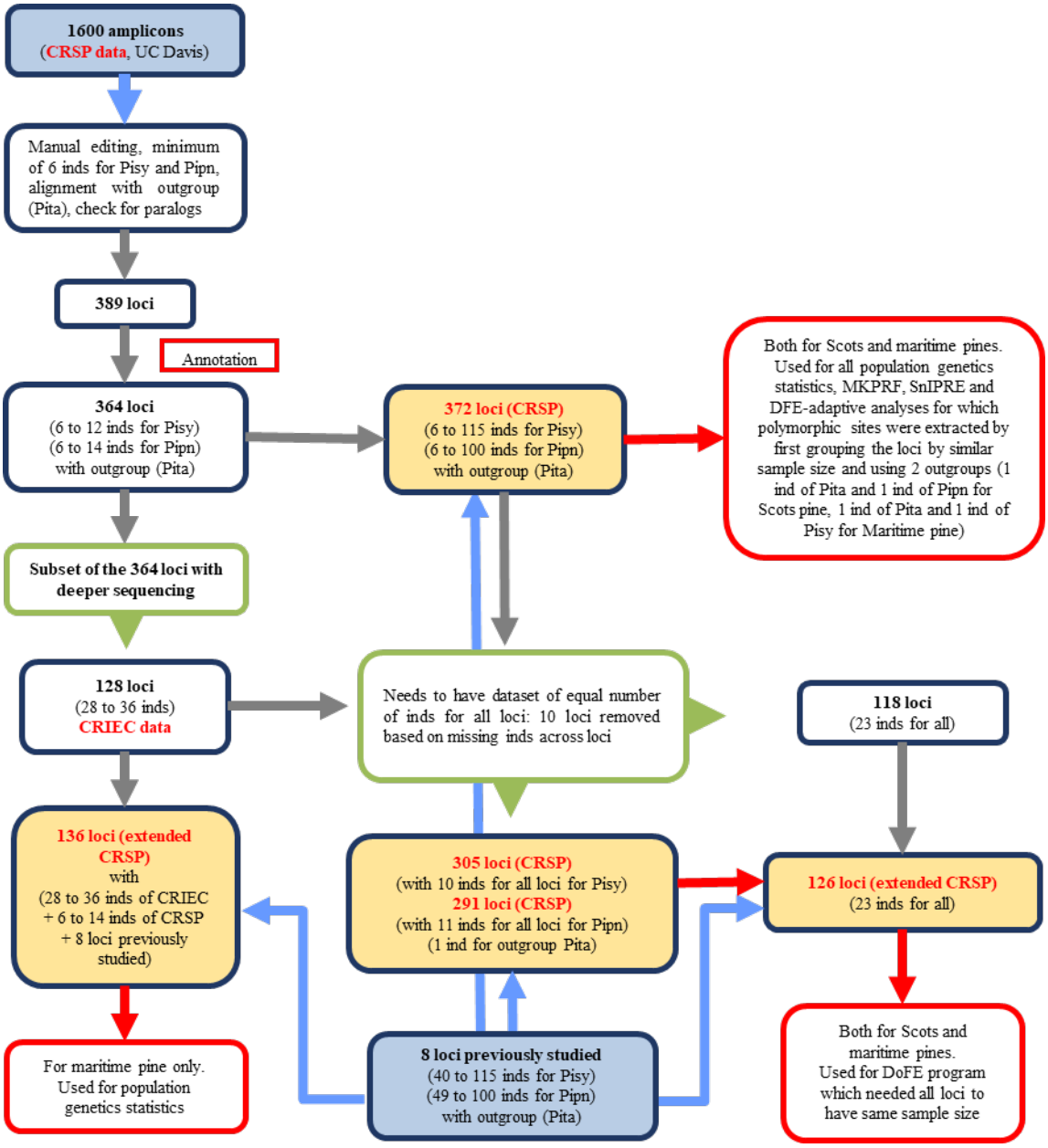

Figure S2. Selection events detected with the SnIPRE approach for the 64 target loci and silent sites in Scots pine and maritime pine. The graphs represent the selection effect (fixation rate of non-synonymous mutations) vs. constraint effect (deleterious mutations).

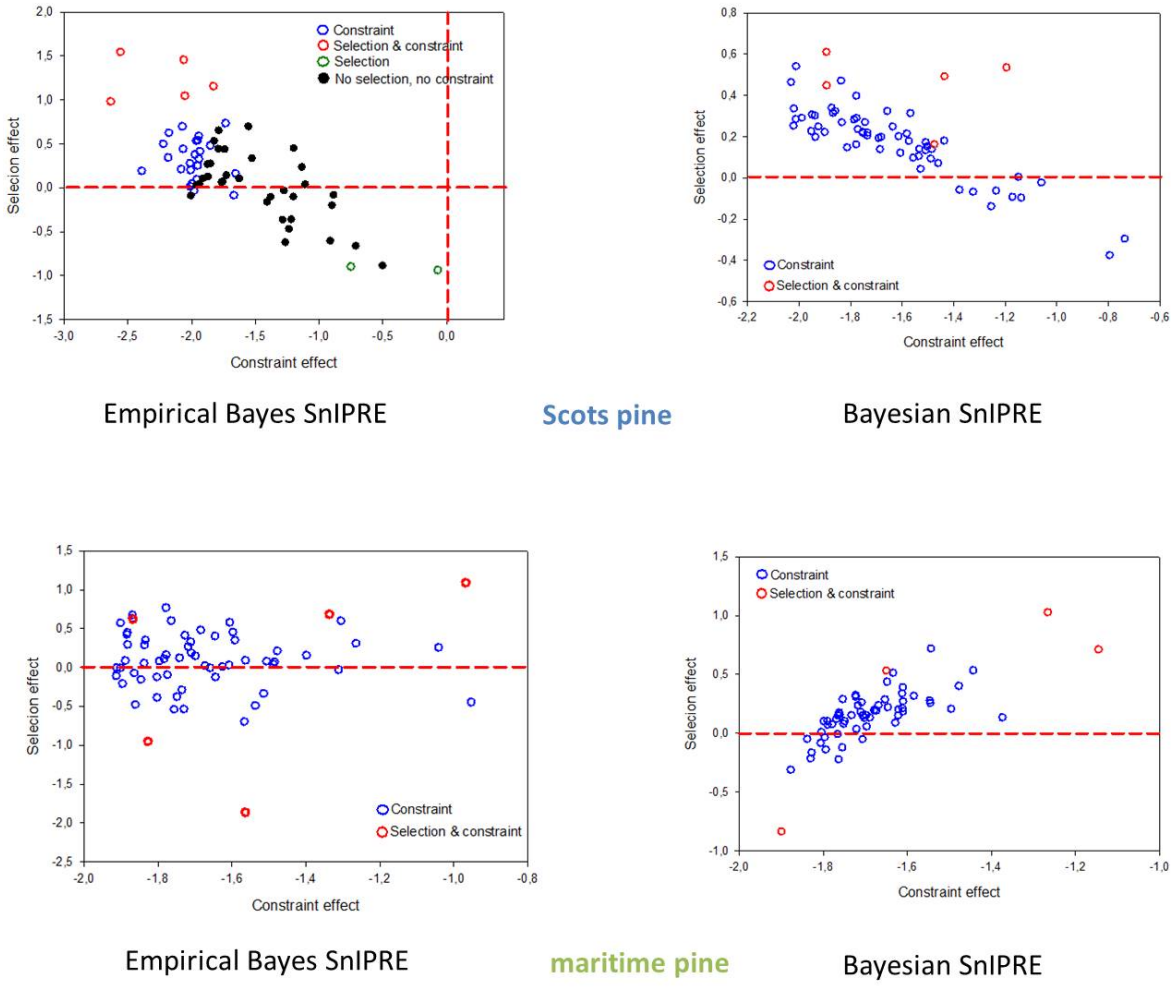

Figure S3. Maximum likelihood three-species unrooted phylogenetic tree to estimate shared divergence between loblolly pine and Scots pine/maritime pine (see main text).

Example of branch length estimates using 168,534 bp of concatenated sequence per species (Pipn_1: one sequence for maritime pine, Pisy_1: one sequence for Scots pine and Pita_1: one sequence for loblolly pine).

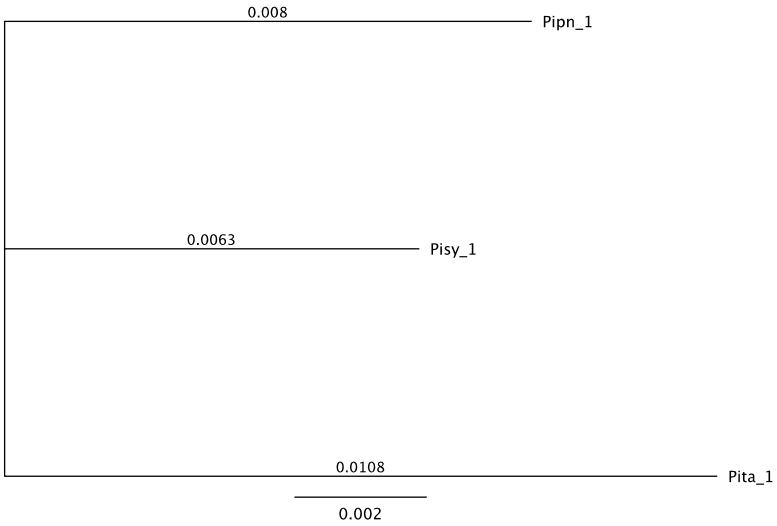

Table S1. Sampling for the two conifer species. Numbers under candidate genes' column correspond to the exact number of samples per population for each of the six previously studied candidate genes. For CRSP and "extended CRSP" datasets the numbers show the maximum sample size per population.

Scots pine

Site	Country	Latitude	Longitude	8 loci (6 genes) from candidate genes						CRSP
				coL1	*dhn1*	*dhn2*	*dhn5*	*gia*	*4cl*	
Kolari	Finland	67.18	24.05	19	5	4	5	19	7	2
Usinsk	Russia	66.08	57.5	7				10	5	
Kaddekielas	Sweden	66.07	19.1	8				9	6	
Northern Sweden	Sweden	65.13	20.23							1
Northern Sweden	Sweden	64.15	16.07							1
Uusikaupunki	Finland	60.87	21.33	10	5	4	5	10	5	
Eastern Scotland	UK	57.05	-3.27	10				10	5	2
Kalsnava	Latvia	56.75	25.88	9				9	4	
Norra Gullabo	Sweden	56.47	15.92	10	5	5	5	10	6	
Radom	Poland	50.68	20.08	20	5	5	5	20	9	2
Haguenau	France	48.85	7.87	10	5	4	5	9	6	
Oberloisdorf	Austria	47.43	16.48		5	5	5			
Parma	Italy	44.62	10.15	10				9	3	2
Kalabak	Turkey	39.45	30.3		5	5	5			
Sierra de Baza	Spain	37.37	-2.83		5	4	5			2
Total (individuals)				113	40	36	40	115	56	12
Total (populations)				10	8	8	8	10	10	7

maritime pine

Site	Country	Latitude	Longitude	8 loci (6 genes) from candidate genes						CRSP	CRIEC
				coL1	*dhn1*	*dhn2*	*dhn5*	*gia*	*4cl*		
Landes	France	na	na							3	
Pleucadec	France	47.78	-2.34		5						
Mimizan (Landes)	France	44.13	-1.3		15	4	2		1		2
Galicia	Spain	na	na							1	
Unknown origin*	France	na	na							1	
Tabuyo del Monte	Spain	42.3	-6.22								2
San Cipriano de Ribarteme	Spain	42.12	-8.36		7						
Pinia (Corsica)	France	42.02	9.46	4	8	9	9	3	7		2
Pineta (Corsica)	France	41.96	9.04		6						
Coca	Spain	41.23	-4.5	13	17	15	13	10	16	2	4
Arenas de San Pedro	Spain	40.19	-5.12	12	19	15		10	13	1	4
Olba	Spain	40.17	-0.62	4		8	7	4	7	1	4
Quatretonda	Spain	38.97	-0.36	3		8	6	4	5	1	4
Cazorla	Spain	37.92	-2.92	9		7	7	9	9	1	4
Oria	Spain	37.52	-2.33	6	6	9	9	1	6	1	4
Tabarka	Tunisia	36.94	8.7	5	10	8	8	4	9	1	2
Cómpeta	Spain	36.85	-3.88			1	4		8		
Estepona	Spain	36.52	-5.12								2
Tamrabta	Morocco	33.6	-5.02	3	7	8	3	2	8	1	2
Sidi Meskour	Morocco	31.47	-6.83	1			1	2	3		
Total (individuals)				60	100	92	77	49	92	14	36
Total (populations)				10	10	11	12	10	12	11	12

* Parent of a QTL mapping progeny with only approximate known origin (France)

Table S2. Description of the six previously studied candidate genes (8 loci).

Locus	Length (bp)[a]	Scots pine	maritime pine
4cl (exon 1, 2-3, 5) [b]	516+360+251 = 1127	this study[3] (accessions MF385276-MF385442)	Grivet et al. (2011)
dhn-1	636	Wachowiak et al. (2009)[d]	Eveno et al. (2008)
dhn-2	713	Wachowiak et al. (2009)	Grivet et al. (2011)[c]
dhn-5	559	Wachowiak et al. (2009)	Grivet et al. (2011)
coL1	3829	Pyhäjärvi et al. (2007); Kujala and Savolainen (2012)	this study[1] (accessions MF385443-MF385502)
gia	1376	Pyhäjärvi et al. (2007); Kujala and Savolainen (2012)	this study[2] (accessions MF385510-MF385558)

[a]Length based on the common alignment between Scots pine and maritime pine, including indels.
[b]Exon_1, exon_2-3, exon_5 correspond to exon_c, exon_a, exon_b in Grivet et al. (2011).
[c]The candidate gene corresponds to *dhn2-Ps*.
[d]*dhn-1* corresponds to *dhn-9* in Wachowiak et al. (2009).
[1] PCR reaction: for primer pairs Copr-promU2/ex1L1 and Copr-ex2U2/3utrL2 the 20 µl mix of reaction contained 0.8 mM dNTP, 0.5 µM of each primer, 1x Phusion HF Buffer (Phusion , Finnzymes), 25 ng DNA and 0.4 unit Taq polymerase (Phusion, Finnzymes); PCR conditions: 1 min at 98 ºC, 35 cycles of 10 sec at 98 ºC, 30 sec at 66 ºC, 30 sec 72 ºC, followed by 10 min at 72 ºC (see primer's specification in Kujala and Savolainen 2012). For primer pairs Copr-109U/941L, Copr-718U/1919L and Copr1820U/1311L the 20 µl mix of reaction contained 0.5 µM of each primer, 1x Phusion Flash PCR Master Mix (Phusion Flash, Finnzymes) and 25 ng DNA. PCR conditions: 1 min at 98 ºC, 35 cycles of 10 sec at 98 ºC, 30 sec at 67 ºC (Copr-109U/941L) or 63 ºC (Copr-718U/1919L and Copr1820U/1311L), 30 sec at 72 ºC, followed by 10 min at 72 ºC (see primer's specification in Kujala and Savolainen 2012).
[2] PCR reaction: the 20 µl mix of reaction contained 0.8 mM dNTP, 0.5 µM of each primer, 1x Phusion HF Buffer (Phusion, Finnzymes), 25 ng DNA and 0.8 unit Taq polymerase (Phusion, Finnzymes). PCR conditions: 1 min at 98 ºC, 35 cycles of 10 sec at 98 ºC, 30 sec at 65 ºC (ex11-U1/3utr-L1) or 66 ºC (ex10-U1/ex11-L2), 1 min at 72 ºC, followed by 10 min at 72 ºC (see primer's specification in Kujala and Savolainen 2012).
[3] PCR conditions are the same as in Kujala and Savolainen (2012), and primers are identical to those described in Grivet et al. (2011).

Table S3. 64 target genes related to stress responses (S) and phenology/photosynthesis (P).

Locus	Function	E-value *Pinus* EST	E-value *A. thaliana*	Length (bp) [a]
4cl (exon 1) [b]	4-coumarate_CoA ligase (S)	0	7.00E-49	515
4cl (exon 2-3) [b]	4-coumarate_CoA ligase (S)	0	1.00E-24	359
4cl (exon 5) [b]	4-coumarate_CoA ligase (S)	0	1.00E-19	249
dhn-1 [c]	Dehydrin (S)	0	5.00E-07	601
dhn-2 [d]	Dehydrin (S)	0	5.00E-03	573
dhn-5	Dehydrin (S)	0	6.00E-04	386
0_16976_02	3-ketoacyl-CoA synthase 6 (S)	na	1.00E-41	402
0_18745_02	Mitogen-activated protein kinase 4 (S)	2.00E-089	8.00E-17	685
0_2070_01	Heat stress transcription factor B-2b (S)	na	7.00E-11	411
0_3790_01	Phospholipase D alpha 1 (S)	3.00E-120	5.00E-35	512
0_4032_02	ARM repeat superfamily protein (S)	na	3.00E-28	390
0_4042_01	Glutathione S-transferase TAU 8 (S)	1.00E-169	2.00E-21	476
0_4285_01	Amino acid permease 3 (S)	0	3.00E-27	495
0_6683_01	Salt-inducible zinc finger 1 (S)	na	6.00E-10	459
0_6878_01	F-box leucine-rich repeat family protein MAX2 (S)	7.00E-047	5.00E-62	435
0_768_02	Putative protein kinase (S)	1.00E-109	1.00E-38	481
0_9082_01	Putative beta-1,3-endoglucanase (S)	0	2.00E-32	421
0_9524_02	U-box domain-containing protein 41 (S)	na	3.00E-25	449
0_990_01	Putative calcium-binding protein CML25 (S)	na	7.00E-30	399
2_1582_02	DNAJ heat shock protein-like protein (S)	0	9.00E-56	457
2_2931_01	Ethylene-responsive transcription factor RAP2.4 (S)	0	7.00E-33	454
2_3319_01	Autophagy-related protein 18D (S)	na	2.00E-13	325
2_3726_02	DNAJ heat shock protein-like protein (S)	na	1.00E-28	450

2_6731_01	F-box protein GID2 (S)	na	4.00E-14	430
CL1524Contig1_03	Histidinol dehydrogenase (S)	5.00E -053	3.00E-17	439
CL1536Contig1_03	Mannose-1-phosphate guanylyltransferase (S)	5.00E -105	6.00E-37	207
CL2332Contig1_01	Calcium-dependent protein kinase 6 (S)	6.00E -057	3.00E-19	426
CL263Contig2_03	RNA-binding protein 47C' (S)	7.00E -047	6.00E-10	444
CL3771Contig1_04	Ubiquitin-conjugating enzyme E2 32 (S)	9.00E -091	3.00E-11	446
UMN_2399_01	U-box domain-containing protein 13 (S)	0	8.00E-27	437
UMN_5272_01	6-phosphogluconate dehydrogenase, decarboxylating 3 (S)	na	3.00E-83	444
UMN_CL132Contig1_03	Malate dehydrogenase (S)	7.00E -026	3.00E-62	319
2_9480_01	Malate dehydrogenase (S)	na	2.00E-63	420
2_1014_01	Heat stress transcription factor B-1 (S)	1.00E -059	2.00E-06	431
0_1123_01	Heat shock protein 70B (S)	0	1.00E-38	404
0_11591_01	Protein auxin RESPONSE 4 (S)	0	5.00E-33	419
0_11649_01	Tubulin beta-8 chain (S)	6.00E -128	4.00E-84	572
0_11649_03	Tubulin beta-2/beta-3 chain (S)	0	2.00E-72	344
0_11684_01	Coronatine-insensitive protein 1 (S)	0	7.00E-37	491
0_12117_01	Adenine nucleotide alpha hydrolase-like protein (S)	0	2.00E-15	414
0_12896_01	F-box protein SKIP2 (S)	2.00E -136	5.00E-55	434
0_143_01	Peroxidase 15 (S)	na	8.00E-22	434
0_17010_02	Putative UDP-glucose 6-dehydrogenase 1 (S)	0	1.00E-76	375
0_4588_01	Aldehyde dehydrogenase 2B4 (S)	na	1.00E-14	218
CL1029Contig1_01	Putative galactinol--sucrose galactosyltransferase 2 (S)	3.00E -085	3.00E-13	427
CL305Contig1_05	Dihydrolipoyl dehydrogenase 1 (S)	na	6.00E-44	302
CL3795Contig1_01	Amino acid dehydrogenase family protein (S)	na	1.00E-21	545
0_7921_01	Glucose and ribitol dehydrogenase homolog 1 (S)	na	2.00E-17	336
0_15991_01	E3 ubiquitin-protein ligase COP1 (P)	na	3.00E-17	317
coL1	Zinc finger protein Constans-like 3 (P)	0	8,00E-35	3797
gia	Gigantea protein (P)	0	1,00E-27	1287
0_12156_01	Inactive leucine-rich repeat receptor-like protein kinase	0	4.00E-47	431

	CORYNE (P)			
0_12156_02	inactive leucine-rich repeat receptor-like protein kinase CORYNE (P)	0	7.00E-37	450
0_16400_01	Protein UNUSUAL FLORAL ORGANS (P)	0	1.00E-49	437
UMN_3408_01	histone-binding protein RBBP4 (P)	0	2.00E-26	433
0_3723_01	STRUBBELIG-receptor family 3 (P)	2.00E -072	3.00E-16	572
0_7454_01	Probable serine/threonine-protein kinase (P)	na	9.00E-17	454
0_8850_02	Photosystem I P700 chlorophyll a apoprotein (P)	na	1.00E-52	320
2_6995_01	Phosphoenolpyruvate carboxylase 4 (P)	0	3.00E-61	407
UMN_3561_02	Photosystem II 47 kDa protein (P)	0	3.00E-41	352
UMN_5101_03	Cytochrome b6/f complex subunit V (P)	na	2.00E-13	402
UMN_6852_02	Cytochrome f (P)	0	1.00E-37	397
UMN_6924_03	Photosystem II 47 kDa protein (P)	na	2.00E-42	321
CL1430Contig1_06	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha 1 (P)	1.00E -056	4.00E-19	619

[a]Length based on the common alignment between Scots pine and maritime pine, including indels.
[b]*exon-1*, *exon-2-3*, *exon-5* correspond to *exon-c*, *exon-a*, *exon-b* respectively in Grivet et al. (2011).
[c]*dhn-1* corresponds to *dhn-9* in Wachowiak et al. (2009).
[d]*dhn-2* corresponds to *dhn2-Ps* in Grivet et al. (2011).

Table S4. 28 compound reference loci obtained using linkage map information.

Compound reference locus	Individual reference locus
LG1_a	0_18018_01, CL3054Contig1_01, UMN_1023_01
LG1_b	0_15329_01, 0_17206_01, 0_17607_02
LG2_a	2_9455_01, 0_8531_01, 0_9091_01, UMN_5867_01
LG2_b	0_13929_02, 0_16732_01, 0_2217_01, 2_9087_01
LG3_a	0_18261_01, 0_7001_01, UMN_3006_01, UMN_3444_01, UMN_927_01
LG3_b	0_17082_01, 0_5575_01, 0_9448_01, 2_1528_01
LG3_c	0_4756_01, 2_6618_01, 0_11270_01, 0_12683_01, 0_846_01, CL814Contig1_06
LG4_a	0_16068_01, 0_9444_01, CL4511Contig1_02
LG4_b	0_13383_01, 0_7171_01, 2_3591_03, CL1045Contig1_01, CL1238Contig1_01, UMN_1037_01
LG5_a	0_10453_01, 0_6465_01, CL3037Contig1_06, CL3758Contig1_05, 0_18350_01
LG5_b	0_10054_01, 2_3941_01, 2_5724_02, 2_9603_01, CL415Contig1_04
LG5_c	UMN_4904_01, 0_236_01, 2_2936_01, CL4342Contig1_01, UMN_801_01
LG6_a	0_9383_01, 2_7725_01, 2_8852_01, CL1004Contig1_08, CL4432Contig1_04
LG6_b	0_11980_01, 0_12929_02, 0_8359_01, 0_8844_01, 0_9329_02, 2_5064_01, CL544Contig1_03
LG7_a	0_10667_02, 2_5996_01, CL1848Contig1_01, CL572Contig1_02
LG7_b	0_1659_02, 0_18470_01, 0_2078_01, 2_5636_01, CL4470Contig1_01
LG7_c	0_14976_01, 0_4105_01, 0_4394_01, 2_6491_01, 2_9291_02
LG8_a	0_10267_01, 0_14221_01, 2_3947_01, CL1455Contig1_07
LG8_b	0_17127_01, 0_6999_01, 2_2960_02, CL1698Contig1_01, CL3539Contig1_01
LG9_a	0_17143_02, 2_10236_01, CL1694Contig1_04, UMN_6426_02
LG9_b	2_5099_01, 2_7852_01, 2_9930_01
LG9_c	0_13841_01, 0_16459_01, 2_684_01, 2_974_01
LG10_a	0_12021_01, 0_12978_02, 2_4724_01, 2_6130_01
LG10_b	0_13484_01, 0_16860_01, 2_6052_01, UMN_5833_01
LG11_a	0_12190_02, 0_16009_01, 0_17247_02, 2_7918_01, CL2472Contig1_01
LG11_b	0_16889_02, 0_2433_01, 0_5204_01, CL4023Contig1_01

LG12_a	CL905Contig2_01, 0_16169_01, 0_2885_01, 0_3261_01
LG12_b	0_9922_01, 0_11090_01, 1_5675_01, 2_10212_01, UMN_2174_01

Table S5. Loci found under selection with the Empirical Bayes SnIPRE and the Bayesian SnIPRE methods (for silent sites).

Locus	Scots pine		maritime pine	
	Empirical Bayes SnIPRE	Bayesian SnIPRE	Empirical Bayes SnIPRE	Bayesian SnIPRE
dhn1 (S)	negative	negative		positive
dhn2 (S)	positive	positive		
dhn5 (S)	positive	positive		
coL1 (P)	positive	positive	positive	positive
4cl_exon1 (S)	positive			
0_4042_01 (S)			positive	positive
2_9480_01 (S)			negative	negative
0_143_01		positive		
0_3723_01	negative			
0_4032_02			positive	
0_4588_01		positive		
0_9082_01	positive			
0_11684_01		positive		
0_11649_01			negative	
CL1430Contig1_06		positive		
Total positive	5	7	3	3
Total negative	3	2	2	1
Total	8	9	5	4

Table S6. Count of fixed differences between loblolly pine and Scots pine/maritime pine. “Originated in one species” refers to mutations that arose only in Scots pine or in maritime pine.

	Scots pine			maritime pine		
	All sites	0-fold sites	4-fold sites	All sites	0-fold sites	4-fold sites
Fixed sites	1868	402	333	2199	446	412
Shared sites	1239	250	227	1239	250	227
Originated in one species	629	152	106	960	196	185

Table S7. Bootstrapping procedure to compare nucleotide diversity estimates in loci transferred from loblolly pine to European pines (364 CRSP loci, *N*=14) with those randomly resampled from loblolly pine genome (about 6,000 loci available). Numbers in bold indicate significant departure from random set of loci.

	S total	π total	θ_w total	π syn	θ_w syn	π nsyn	θ_w nsyn
Transferred loci set	3.36	**0.00213823**	**0.00253591**	0.00480557	0.00560486	0.00093046	0.00112435
Low 95CI	3.02	0.00188983	0.00225759	0.00412203	0.00490509	0.00074009	0.00091097
High 95CI	3.70	0.00238664	0.00281423	0.00548911	0.00630463	0.00112082	0.00133772
1,000 bootstraps (364 loci)							
Low 95CI	3.38	0.00307339	0.00342669	0.00466651	0.00524553	0.00103738	0.00122625
High 95CI	4.48	0.00439749	0.00481324	0.00800163	0.00884918	0.00232990	0.00244981

S total: number of segregating sites per locus; π (total, syn, nsyn): Tajima's nucleotide diversity (Tajima 1989) for total sites, synonymous sites, and non-synonymous sites; θ_w (total, syn, nsyn): Watterson's nucleotide diversity (Watterson 1975) for total sites, synonymous sites, and non-synonymous sites; 95Cl: 95% Confidence Intervals.

Table S8. Summary statistics for all sites, synonymous sites, non-synonymous sites and silent sites, for the two conifer species for all loci (target and reference genes) for the CRSP and for the "extended CRSP" datasets. Divergence estimates are given using loblolly pine as reference. Statistics were normalized for varying sample size across loci with the SDMTools package in R.

	Scots pine (CRSP)			
	all	syn	nsyn	silent
θ_w (stdev)	0.003302 (0.003664)	0.006871 (0.008962)	0.001470 (0.002685)	0.005588 (0.006476)
π (stdev)	0.002926 (0.003886)	0.006346 (0.009822)	0.001254 (0.002494)	0.004978 (0.006712)
Divergence (JC) (stdev)	0.016598 (0.010255)	0.032276 (0.025316)	0.006973 (0.008064)	0.028613 (0.018187)

	maritime pine (CRSP)			
	all	syn	nsyn	silent
θ_w (stdev)	0.002141 (0.002162)	0.004189 (0.005440)	0.000823 (0.001393)	0.003749 (0.003783)
π (stdev)	0.002236 (0.002507)	0.004426 (0.006942)	0.000905 (0.001803)	0.003785 (0.004145)
Divergence (JC) (stdev)	0.019092 (0.014916)	0.037465 (0.031322)	0.007029 (0.007974)	0.033709 (0.025498)

	maritime pine ("extended CRSP")			
	all	syn	nsyn	silent
θ_w (stdev)	0.002588 (0.001915)	0.00463 (0.005147)	0.001016 (0.00138)	0.004193 (0.003292)
π (stdev)	0.002912 (0.00252)	0.00543 (0.006992)	0.001083 (0.001996)	0.004768 (0.00428)
Divergence (JC) (stdev)	0.021141 (0.015417)	0.041243 (0.036966)	0.008142 (0.009089)	0.037145 (0.030381)

Table S9. Nucleotide diversity for maritime pine with CRSP (364 loci + 8 previously studied loci = 372 loci) and "extended CRSP" (128 + 8 previously studied loci = 136 loci) datasets using bootstrap resampling with replacement for different number of loci (indicated in brackets).

	CRSP		"extended CRSP"	
	π silent	θ_w silent	π silent	θ_w silent
Full dataset	0.003785	0.003749	0.004768	0.004193
Lower 95% CI	0.003282	0.003174	0.004076	0.003580
Higher 95% CI	0.004337	0.004358	0.005530	0.004770
Mean 1,000 bootstrats (10)				
Lower 95% CI	0.001220	0.001436	0.002448	0.002319
Higher 95% CI	0.007082	0.007389	0.007804	0.006756
Mean 1,000 bootstrats (50)				
Lower 95% CI	0.002380	0.002398	0.003612	0.003248
Higher 95% CI	0.005065	0.005451	0.005902	0.005192
Mean 1,000 bootstrats (100)				
Lower 95% CI	0.002783	0.002826	0.003939	0.003579
Higher 95% CI	0.004773	0.004946	0.005631	0.004904

Statistics were normalized for varying sample size across loci with the SDMTools package in R.