Admixture and selection patterns across the European distribution of Scots pine, Pinus sylvestris (Pinaceae)

Witold Wachowiak, Annika Perry, Julia Zaborowska, Santiago C. Gonzalez-Martinez, Stephen Cavers

Botanical Journal of the Linnean Society, 2022, **XX**, 1–17. With 4 figures.

Admixture and selection patterns across the European distribution of Scots pine, *Pinus sylvestris* (Pinaceae)

WITOLD WACHOWIAK[1,2,*], ANNIKA PERRY[3], JULIA ZABOROWSKA[1], SANTIAGO C. GONZÁLEZ-MARTÍNEZ[4] and STEPHEN CAVERS[3]

[1]*Institute of Environmental Biology, Faculty of Biology, Adam Mickiewicz University in Poznań, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland*
[2]*Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland*
[3]*UK Centre for Ecology & Hydrology Edinburgh, Penicuik, Midlothian, EH26 0QB, UK*
[4]*INRAE, Univ. Bordeaux, BIOGECO, F-33610 Cestas, France*

Received 18 March 2021; revised 13 December 2021; accepted for publication 12 February 2022

We analysed data on nucleotide polymorphism at 78 nuclear genes to search for signatures of divergence and selection in populations from across the distribution of *Pinus sylvestris* in Europe. Data on variation at a large set of maternally inherited mitochondrial DNA markers were also analysed. Most nuclear loci showed homogeneous patterns of variation across populations from Spain, Central Europe, Nordic countries and Scotland. Mitochondrial markers did not reveal strong population structure but suggested admixture of samples carrying different mitochondrial lineages in certain parts of the distribution. At nuclear DNA loci, we found high nucleotide polymorphism, generally low population differentiation and weak signals of isolation by distance. Together with the mitochondrial DNA results, these patterns seem most likely to be associated with a process of colonization from multiple sources and efficient gene flow. The results are in line with demographic inference indicating establishment of Central European and Nordic populations before the Last Glacial Maximum and admixture of colonization fronts from mainland Europe in Scottish populations. Coalescent and outlier detection methods identified several genes showing molecular signatures of selection especially in response to temperature variation in ecologically and phenotypically divergent populations. The study provides a large set of markers for analysis of genetic diversity in populations of *P. sylvestris*, including candidate genes that are implicated in local adaptation and population divergence.

ADDITIONAL KEYWORDS: DNA sequencing – local adaptation – mitochondrial DNA – nuclear loci – nucleotide polymorphisms – population structure.

INTRODUCTION

In the Northern Hemisphere, tree populations have experienced recurring waves of extinction and migration driven by glacial cycles, and the consequences of this for contemporary genetic variation are still being resolved (Hewitt, 2000; Petit, Hu & Dick, 2008). The expansion and range shifts of the species were accompanied by the development of distinct phenotypes as they adapted to new local environments. Experiments under controlled conditions (glasshouse and provenance trials) have shown the genetic basis for this phenotypic variation, which is consistent with a process of natural selection due to local adaptation (Hurme *et al.*, 1997; Wachowiak *et al.*, 2017). For wind-pollinated species such as pines (*Pinus* L.), this differentiation reflects a balance between the homogenizing effect of gene flow and selection (Savolainen, Pyhäjärvi & Knürr, 2007). Although distinct phenotypes are well characterized for many ecologically and economically important species, the availability of genomes for studies of natural selection was until recently very limited in non-model forest tree species. Consequently, the patterns of molecular evolution, processes contributing to standing genetic variation and segregation of polymorphisms that may underlie the development of distinct ecotypes are not fully resolved.

Scots pine (*Pinus sylvestris* L.) is one of the most extensively studied tree species due its high ecological

*Corresponding author. E-mail: witwac@amu.edu.pl

and commercial importance. It has the broadest distribution of all pines, with a natural range covering large parts of Europe and Asia from western Scotland to eastern Siberia and southern Spain to the Arctic Circle (Critchfield & Little, 1966). Provenance trials and common garden experiments have found strong ecotypic variation in growth traits (Oleksyn, 1991; Giertych & Oleksyn, 1992; Zadworny *et al.*, 2016). Clear patterns of adaptation to local environmental conditions, established since post-glacial migration, have been documented in traits such as phenology and abiotic stress resistance (Hurme *et al.*, 1997; Notivol *et al.*, 2007; Salmela *et al.*, 2011; Wachowiak *et al.*, 2017). Other studies have demonstrated the genetic basis of adaptive responses of Scots pine ecotypes to light spectra across a northern latitudinal cline and the absence of maternal effects on the performance of the seedlings (Ranade & García-Gil, 2013). However, progress towards a connection between variation in the phenotypic and growth performance traits and molecular variation remains slow.

Genetic studies in *P. sylvestris* have found low between-population differentiation across large geographical areas with isozymes, microsatellites and nuclear sequence data (Prus-Glowacki & Stephan, 1994; Cheddadi *et al.*, 2006; Wachowiak *et al.*, 2014). In general, these studies demonstrate that populations from southern regions of Europe show some unique patterns of genetic variation compared to areas recolonized after the Last Glacial Maximum. Genetic approaches using polymorphisms in nuclear genes and coalescent simulations indicate an ancient bottleneck of *P. sylvestris* and Norway spruce [*Picea abies* (L.) H.Karst.] populations in Europe that predates the last glacial period (Heuertz *et al.*, 2006; Pyhäjärvi *et al.*, 2007). Population structure assessments using polymorphisms in maternally inherited, seed-dispersed mitochondrial DNA markers have been largely limited to just a few genomic regions (Sinclair, Morman & Ennos, 1999; Soranzo *et al.*, 2000). Consequently, mitochondrial DNA studies of the species have had low spatial resolution and primarily identify broad-scale putative recolonization routes (Soranzo *et al.*, 2000; Naydenov *et al.*, 2007; Pyhäjärvi, Salmela & Savolainen, 2008; Floran, Sestras & García-Gil, 2011; Dering *et al.*, 2017; Tóth *et al.*, 2017; Semerikov *et al.*, 2018). Studies across large geographical areas and of different ecotypes have allowed identification of some candidate genes related to cold tolerance and phenology in Scots pine (Wachowiak, Balk & Savolainen, 2009; Kujala & Savolainen, 2012; Tyrmi *et al.*, 2020) and other conifer species including, e.g., Norway spruce and maritime pine (*Pinus pinaster* Aiton) (Pot *et al.*, 2006; Eveno *et al.*, 2008; Grivet *et al.*, 2011; Ranade & García-Gil, 2020). More recently, novel genomic resources have been developed in *P. sylvestris*, including information about polymorphisms at the whole transcriptome level (Wachowiak *et al.*, 2015) and *c.* 50 000 single nucleotide polymorphisms (SNPs), mostly comprising variation at coding regions of the genome and placed in genotyping arrays (Perry *et al.*, 2020; Kastally *et al.*, 2021). A key step towards deploying these resources for breeding and population genomic studies is to develop better understanding of the processes that have influenced patterns of standing genetic variation in populations of *P. sylvestris*. In this study, we investigated patterns of neutral and non-neutral genetic variation within and among European populations of *P. sylvestris*, looking at the distribution of nuclear variation in a large set of resequenced nuclear genes and mitochondrial DNA markers that have different modes of dispersal and inheritance in pines. We applied the markers to search for (1) signatures of population genetic structure in European *P. sylvestris*, considering its history inferred by approximate Bayesian computation analyses and (2) molecular-level patterns of divergence and selection in populations sampled across environmental gradients and known to be phenotypically differentiated based on quantitative genetic studies (Wachowiak *et al.*, 2017). We compared results from the present study and earlier investigations of the distribution of genetic variation and phenotypic divergence in natural populations to advance selection of suitable parts of the species range for future population genomic studies of the species.

MATERIAL AND METHODS

Sampling and DNA extraction

Seeds were collected from 14 autochthonous European populations: 135 samples were analysed, comprising up to ten different trees from each location (Table 1). Those populations represent different geographical stands and environments. Subsequent analyses were conducted for individual populations but also regional groups labelled Central Europe (Poland and Austria), Scotland, Nordic (Finland and Sweden) and Spain. Samples from the same locations were previously characterized for a set of quantitative traits (Wachowiak *et al.*, 2017). This study indicated clear ecotypic patterns of between-population differentiation at phenology and growth traits. The most significant trait differentiation was observed among populations from continental European locations that were highly correlated with elevation and latitude of the site of origin (Wachowiak *et al.*, 2017). Furthermore, the population division was also justified by earlier demographic inference analysis using an approximate Bayesian computation (ABC) framework implemented in DIYABC v.2.1.0 (Cornuet *et al.*, 2014). The results indicated that Central European and Nordic

Table 1. Location of 14 populations of *Pinus sylvestris* investigated in Europe

Acr.	Location (country, population)	*N*	Longitude	Latitude	Elevation m a.s.l.	Herbarium collection
SCO_S	Scotland, Shieldaig	10	–5° 38’ 24”	57° 30’ 35”	81	555554
SCO_GT	Scotland, Glen Tanar	10	–2° 51’ 36”	57° 02’ 60”	160	555555
SCO_R	Scotland, Rothiemurchus	10	–3° 46’ 12”	57° 08’ 60”	318	555552
SCO_GA	Scotland, Glen Affric	10	–4° 55’ 12”	57° 16’ 12”	256	555553
SCO_B	Scotland, Black Wood of Rannoch	10	–4° 19’ 12”	56° 40’ 12”	275	555560
AT	Austria, Pernitz	10	16° 00’ 00”	47° 54’ 50”	500	555561
ES_T	Spain, Trevenque	10	3° 32’ 51”	37° 05’ 47”	1170	555562
ES_V	Spain, Valsain	10	4° 02’ 08”	40° 51’ 55”	1350	555563
FI_P	Finland, Punkaharju	10	29° 23’ 21”	61° 45’ 33”	80	555556
FI_K	Finland, Kolari	10	24° 03’ 00”	67° 11’ 00”	190	555559
PL	Poland, Jarocin	10	17° 28’ 40”	51° 58’ 20”	120	555551
SE_T	Sweden, Krp. Tjärnbergsheden	8	20° 48’ 0”	64° 37’ 12”	110	555550
SE_M	Sweden, Väster Mjőingenn	7	13° 34’ 48”	62° 45’ 00”	640	555557
IT	Italy, Cella di Palmia	10	10° 09’ 56”	44° 37’ 59”	180	555558

Acr. – acronym for the population; *N* – number of samples analysed; Herbarium collection – voucher number in Herbarium Dendrologiae Instituti, Academia Scientiarum Polona

populations were most probably established before the Last Glacial Maximum (LGM) and did not derive from Iberian or Apennine populations, whereas Scottish populations have probably arisen through admixture of different colonization fronts from mainland Europe (González-Díaz, 2018). As we found an Italian population to be consistently distinct in different analyses, it was excluded from group comparisons. Genomic DNA was extracted from a single haploid megagametophyte per mother tree from germinated seeds using a DNeasy Plant Mini Kit (Qiagen).

PCR Amplification and sequencing

Each sample was genotyped at a set of nuclear and mitochondrial regions. For the nuclear DNA analysis, 78 gene fragments associated with a broad range of different functions were selected for sequencing (Supporting Information, Table S1). PCR primers for each nuclear gene fragment were designed based on unique cDNA originally sequenced in *Pinus taeda* L. (Mosca *et al.*, 2012). PCR amplification was performed on Thermo MBS thermal cyclers and carried out in a total volume of 15 µL containing about 15 ng haploid template DNA, 0.1 µM each dNTP, 0.2 µM each of forward and reverse primer, 0.15 µl of 5U Taq DNA polymerase (BioLabs), 1× bovine serum albumin (BSA), 1.5 µM $MgCl_2$ and 1× PCR buffer. Amplification conditions included initial denaturation at 94 °C for 3 min followed by 35 cycles with 30 s denaturation at 94 °C, 30 s annealing at 60 °C and 1 min 30 s extension at 72 °C and a final 5 min extension at 72 °C. PCR fragments (including mitochondrial DNA region Pr34) were purified using exonuclease I-shrimp alkaline phosphatase enzymatic treatment. About 20 ng of PCR product was used as a template in 10 µL sequencing reactions with the Big Dye Terminator DNA Sequencing Kit (Applied Biosystems) performed by Edinburgh Genomics (University of Edinburgh, UK) and Genomed (Poland). Multilocus haplotypes were determined by direct sequencing of haploid DNA. CodonCode Aligner (Codon Code Corporation) was used to edit and align sequences. Sequence data are deposited in the NCBI repository (Supporting Information, Table S1). For the mitochondrial DNA analysis, 15 regions were genotyped in total using PCR-RFLP technique or, in the case of locus Pr34, Sanger sequencing (Supporting Information, Table S2). PCR amplification of 13 regions were carried out in a total volume of 15 µL containing 15 ng template DNA, 0.016 µM each dNTP, 0.3 µM each of forward and reverse primer, 0.15 µl of 5U Taq DNA polymerase, 1.5 µM $MgCl_2$ and 1× PCR buffer (Novazym). Standard amplification procedures were used with initial denaturation at 94 °C for 3 min followed by 32 cycles with 30 s denaturation at 94 °C, 30 s annealing at 57 °C and 1 min 20 s extension at 72 °C, and a final 10 min extension at 72 °C. Digestions of RFLP markers were conducted in 10 µL reactions containing 3 µL of PCR product, 2× buffer and 2.5 U restriction enzyme (Supporting Information, Table S2), and samples were incubated overnight for at least 10 h (temperature specified by the manufacturer, Thermo Scientific). Products were subjected to electrophoresis on 1.5–2.0% agarose gel and visualized with GelRed stain in UV light. The remaining two mitochondrial DNA regions previously used in phylogeographic studies of *P. sylvestris* (*nad1* and *nad7*) were amplified according

to Soranzo *et al.* (2000) and Jaramillo-Correa, Beaulieu & Bousquet (2004), respectively, and genotyped by restriction enzyme digestion.

Nucleotide and haplotype variation

We compared the distribution of nucleotide polymorphism across populations and geographical regions. Pairwise nucleotide diversity at nuclear genes was measured as the average number of differences per site (π; Nei, 1987). Correlation between polymorphic sites (ZnS statistics), was measured as average pairwise linkage disequilibrium between all SNPs over a given single locus (Kelly, 1997). To assess skew in allelic frequency distribution, multilocus Tajima's *D* (Tajima, 1989) was computed using the difference between two estimates of the scaled mutation parameter theta for each locus and statistical significance was evaluated by comparison to a distribution generated by 1 000 coalescent simulations using DNASP v.5 software (Librado & Rozas, 2009). The same program was used to compute the number of haplotypes (H_n) and haplotype diversity (H_d) at individual populations (Hudson, Boos & Kaplan, 1992). The number of shared and exclusive haplotypes between regional groups of populations for each nuclear locus was determined using SITES v.1.1 (Hey & Wakeley, 1997). Deviations from neutrality at nuclear loci were estimated using two compound neutrality tests (HEW and DHEW), that are robust to demographic processes (Zeng, Shi & Wu, 2007). Significance levels were determined by 10 000 coalescent simulations on the basis of Watterson's estimator of theta as implemented in the DH software (Zeng *et al.*, 2007). The distribution of the test statistic was investigated for each locus in all samples. All samples were genotyped using all mitochondrial DNA markers, but 27 individuals with ≥ 10% missing data were excluded from further analysis (Table 2). Phylogenetically informative gaps (indels) in Pr34, *nad1* and *nad7* were coded as single mutation events for analyses. For haplotype (mitotype) analysis, single missing positions in individual genotypes were imputed according to the most similar haplotypes and/or to the variants dominating in the same population: the creation of *de novo* haplotypes was avoided. A median joining network of the mitotypes was constructed with POPART v.1.7 (Bandelt, Forster & Röhl, 1999).

Table 2. Nucleotide and haplotype diversity at 78 nDNA and 15 mtDNA loci in *Pinus sylvestris* populations and regional groups

	Nuclear DNA							Mitochondrial DNA				
Populations	N_s	S_a	π_{tot}	ZnS	*D*	H_n	H_d (SD)	*N*	*S*	*K*	H_n	H_d (SD)
SCO_S	9.6	4.7	0.0041	0.451	0.021	3.23	0.554 (0.017)	3	0	0.000	1	0.000 (0)
SCO_GT	9.4	4.7	0.0041	0.457	−0.004	3.40	0.586 (0.017)	10	0	0.000	1	0.000 (0)
SCO_R	9.7	4.7	0.0044	0.368	0.152	3.54	0.594 (0.013)	8	0	0.000	1	0.000 (0)
SCO_GA	9.5	4.3	0.0039	0.477	0.062	3.35	0.556 (0.016)	10	0	0.000	1	0.000 (0)
SCO_B	9.6	5.2	0.0047	0.396	−0.001	3.69	0.612 (0.017)	9	0	0.000	1	0.000 (0)
AT	9.8	4.9	0.0043	0.384	−0.097	3.65	0.591 (0.016)	10	11	5.489	5	0.800 (0.100)
ES_T	9.7	4.4	0.0038	0.435	0.008	3.26	0.562 (0.016)	9	0	0.000	1	0.000 (0)
ES_V	9.6	4.9	0.0044	0.453	−0.019	3.53	0.586 (0.017)	9	0	0.000	1	0.000(0)
FI_P	9.5	5.4	0.0045	0.389	−0.252	3.74	0.609 (0.018)	9	3	1.167	2	0.389 (0.164)
FI_K	8.8	4.5	0.0042	0.482	0.006	3.19	0.567 (0.018)	8	9	3.964	3	0.607 (0.164)
PL	9.0	5.2	0.0045	0.354	−0.185	3.74	0.642 (0.020)	5	1	0.400	2	0.400 (0.237)
SE_T	7.3	4.3	0.0043	0.469	−0.095	3.21	0.583 (0.021)	6	0	0.000	1	0.000 (0)
SE_M	6.6	4.6	0.0046	0.489	0.025	3.19	0.624 (0.023)	4	11	6.000	3	0.833 (0.222)
IT	9.3	2.3	0.0026	0.788	0.798*	1.95	0.380 (0.012)	8	0	0.000	1	0.000 (0)
Regional groups												
SCO	47.28	8.2	0.0042	0.189	−0.176	6.73	0.587 (0.003)	40	0	0.000	1	0.000 (0)
CE	18.76	6.7	0.0044	0.247	−0.205	5.26	0.623 (0.008)	15	12	4.914	7	0.857 (0.057)
ES	19.49	5.9	0.0043	0.330	−0.005	4.55	0.587 (0.008)	18	1	0.529	2	0.529 (0.040)
NORD	32.08	8.2	0.0045	0.221	−0.316	6.51	0.607 (0.005)	27	11	5.288	8	0.826 (0.042)
All samples	128.94	13.3	0.0045	0.094	−0.677*	12.2	0.607 (0.001)	108	13	3.731	14	0.795 (0.032)

N_s – average sample size; S_a – average number of segregating sites; π_{tot} – nucleotide diversity (Nei, 1987); ZnS – linkage disequilibrium statistics (Kelly, 1997); *D* – Tajima's *D* test (Tajima, 1989); *N* – number of individuals genotyped; *S* – number of segregating sites; *K* – average number of nucleotide differences between two randomly chosen sequences; H_n – number of haplotypes; H_d – haplotype diversity (standard deviation); $^*P < 0.01$
Regional groups: SCO – Scotland, CE – Central Europe (Poland and Austria); ES – Spain, NORD - Finland and Sweden

Population structure

The hierarchical distribution of genetic variation at nuclear loci among populations and regional groups was estimated using analysis of molecular variance (AMOVA) (ARLEQUIN v.3.5.22; Excoffier & Lischer, 2010). Estimates were based on all polymorphic sites and a set of putatively neutral SNPs, which were selected by excluding one locus showing F_{ST} outlier behaviour and all SNPs in genes that departed from neutrality. Population structure based on all and a set of 661 putatively neutral polymorphic sites was explored using STRUCTURE v.2.2 (Pritchard, Stephens & Donnelly, 2000) [*K* from one to ten, estimates averaged over ten independent runs, burn-in set to 10 000 and run length to 100 000, results processed using Structure Harvester (Earl & vonHoldt, 2012)] and discriminant analysis of principal components [DAPC, *adegenet* package in R v.4.02, with number of principal components to retain (50) selected using the cross-validation function *xvalDapc*; Jombart & Ahmed (2011)]. The relationship between genetic and geographic distance was evaluated in MEGA 7 (Kumar, Stecher & Tamura, 2016) using a matrix of pairwise net divergence between populations and F_{ST} values (ARLEQUIN), based on all polymorphic sites and a subset of neutral SNPs. To avoid bias due to overrepresentation of Scottish samples and considering observed genetic similarity, a single population from Scotland (Glen Affric) was included in this analysis. Mantel tests for isolation by distance were run with and without the outlier Italian population. Using the mitochondrial DNA data, we looked at the spatial distribution of mitotypes: genetic distance among populations based on the mean number of differences across all mitochondrial DNA sites was calculated in MEGA 7 and used in principal coordinate analysis (PCoA) in GENALEX v.6.503 software (Peakall & Smouse, 2012). Mantel tests were completed in GENALEX, with 999 random permutations of the genetic and geographic distance matrices. For genetic distance estimates, both between-population distance (d_{XY}) and pairwise net divergence (d) between populations (calculated in MEGA) were used. Spatial AMOVA was conducted with Spatial Analysis of Molecular Variance software SAMOVA 2.0 (Dupanloup, Schneider & Excoffier, 2002), using *K* values from two to 13. Haplotype differentiation among populations was estimated using an AMOVA (ARLEQUIN v.3.5.22). Finally, G_{ST} (Nei, 1973) and N_{ST} (Lynch & Crease, 1990) were calculated and compared using a permutation test with 1000 replicates [PERMUTCPSSR v.2.0 (Burban *et al.*, 1999; Pons & Petit, 1996)] to test for phylogeographic structure.

Outlier detection tests

We tested for selection at nuclear loci among populations and regional groups that in our previous studies showed clear patterns of divergence in quantitative traits such as phenology and growth performance (Wachowiak *et al.*, 2017). Our sample sizes and inclusion of different ecotypes allowed testing for signatures of selection as shown in previous studies (Kujala & Savolainen, 2012). However, low sample sizes may affect estimates of allele frequencies across populations and particularly the application of F_{ST} based genome scans from reduced representation data that have known limitations in determining regions under selection (Lowry *et al.*, 2017). Differentiation between samples was estimated locus by locus at both haplotype and SNP/indel level. Significance was estimated by 1 000 permutations of samples between populations and regional groups (ARLEQUIN). False discovery rate (FDR) adjustment was conducted (QVALUE, lambda = 0.15, FDR level = 0.01). Estimates of the overall proportion of true null hypotheses (π_0) and *q* values (which reflect the expected proportion of false positives among significant results) were calculated on the basis of the distribution of *P* values. Additionally, the full SNP dataset was used to test for outlier loci, which may indicate directional or balancing selection. First, coalescent simulations were used to estimate a null distribution and confidence intervals to allow identification of outliers by locus-specific F_{ST} conditioned on the multilocus distribution of F_{ST} values. Simulated groups consisted of 100 subpopulations, and 20 000 replicates of the coalescent were used to identify the expected distribution of F_{ST}. Significance thresholds were set at 95 and 99% (ARLEQUIN). Second, we used BAYESCAN (Foll & Gaggiotti, 2008) for outlier detection, contrasting models with and without selection based on differences in allele frequencies between populations. We tested SNPs with minor allele frequency (MAF) > 0.05. Sample size of 5 000 was used, with thinning intervals of 10, and 20 pilot runs each of 5 000 and burn-in of 50 000. We considered genes as outliers if they showed significant deviation from neutrality and a signature of selection for both haplotypes and SNPs.

Outlier-environment association tests

Allele frequencies of the outlier SNPs were tested for correlation with environmental variables in BAYENV2 software (Coop *et al.*, 2010). For this purpose, a dataset of 19 basic bioclimatic variables was obtained from the World Climate database v.2 with a spatial resolution of 30 s (*c.* 1 km^2) represented as geographic information system continuous raster surfaces (Fick & Hijmans, 2017). These data are derived from the monthly temperature and rainfall values and represent annual trends, seasonality and extreme environmental factors (Supporting Information, Table S3). All climatic information, together with data on latitude, longitude and elevation, were standardized by the *Z* score normalization. We retained

all factors throughout the analysis. The matrix prepared for association searches, which serves as the null model of allele frequency covariance across populations and helps to control for the effects of population structure, was estimated using the full set of polymorphic sites identified in the 78 genes, and three independent repetitions of 100 000 MCMC algorithm iterations. Subsequently, each outlier SNP was tested against each variable in three runs of 10 000 000 iterations.

RESULTS

Nucleotide and haplotype variation at nuclear and mitochondrial DNA loci

The 78 loci produced an alignment of > 32 kbp, with 1 046 polymorphic sites. Mean population diversity levels (the range of π_{tot} = 0.0038 – 0.0047) were similar to those in regional groups (π_{tot} = 0.0042 – 0.0045), although it was low in the Italian population (π_{tot} = 0.0026, Table 2). Linkage disequilibrium levels ranged from 0.354 to 0.489 in populations (0.788 for Italy) and from 0.189 to 0.330 for regional groups. An excess of singleton mutations (significantly negative multilocus Tajima's *D*) was found across all samples (*D* = –0.677; *P* < 0.01, Table 2). A significant excess of common variants was present in the Italian population (*D* = 0.798; *P* < 0.01). The average haplotype diversity was similar for each population (H_d = 0.554–0.642) and regional group (H_d = 0.587–0.607) except Italy (0.380) (Table 2). Most polymorphic sites (70–76%) and haplotypes (50–58%) were shared among regional groups (Supporting Information, Table S4A), and the proportion of unique haplotypes was similar across regional groups (18–23%, Supporting Information, Table S4B). Deviations from neutrality were found at 13 loci for the compound neutrality tests applied across all genes (Supporting Information, Table S5). All 15 mitochondrial DNA markers were polymorphic and three mutations were singletons [at Pr31 in Austria, at Pr13 and *nad1* in an individual from Spain (ES_V)]. The average difference between two sequences in the whole sample was K = 3.731 (Table 2). The 15 loci characterized 15 haplotypes, six occurred only once in the sample. One individual from Spain (ES_V) was distinct: as well as two singleton mutations, it also had a unique combination of polymorphisms at other loci, similar to haplotypes found in closely related *P. uncinata* Ramond ex DC. (Zaborowska, Łabiszak & Wachowiak, 2020); it was excluded from further analysis. The remaining 14 haplotypes were used to look at population diversity and structure. Haplotype diversity for the whole collection was H_d = 0.795 (SD = 0.032) (Table 2), although there was a lot of variation: one haplotype was found across Scotland (number of segregating sites, *S* = 0), whereas five haplotypes (*S* = 11) occurred in the Austrian population. More than half of the populations were fixed for one haplotype (all Scottish, SE_T, ES_T and IT; Supporting Information, Table S6, Fig. 1). Three populations (AT, SE_M and FI_K) had haplotype diversities > 0.5 (Table 2).

Population structure and differentiation at nuclear loci

STRUCTURE analysis at nuclear DNA loci suggested the presence of three genetic groups. The Italian population was consistently an outlier, and two additional genetic groups were evident but without clear geographic structure (Fig. 2, Supporting Information, Figs S1 and S2). A distinction between populations from Italy and Spain and all others was evident in the DAPC analysis based on polymorphic sites detected at nuclear genes (Fig. 3). The population from Italy was also a clear outlier at a subset of neutral SNPs (Supporting Information, Fig. S3). Significant differentiation was found in pairwise comparisons between some populations at all SNPs combined and a subset of neutral SNPs (Supporting Information, Table S7). At neutral loci the most distinctive populations were from Italy and southern Spain (Trevenque). Two Scottish populations from Glen Affric and Glen Tanar also showed significant differentiation from both Finnish populations. No significant differentiation was found between Scottish populations at either the full or neutral set of polymorphisms (Supporting Information, Table S7). Regional groups showed significant differentiation across all polymorphic sites (F_{ST} = 0.024–0.066) except between Central Europe and Nordic region (F_{ST} = 0.007) (Supporting Information, Table S8). Low but significant differentiation was also found at neutral SNPs in pairwise comparisons between sampled locations, except Central Europe and Nordic regions, and Central Europe and Spanish populations (Supporting Information, Table S8). In the AMOVA using all polymorphic sites, *c.* 2.6 and 2.9% of the variation was found among groups and among populations within groups, respectively (Table 3). Most of the variation (*c.* 95 - 97% depending on the set of polymorphic sites used) was found within populations (Table 3). Excluding the Italian population, significant isolation by distance was found based on both net divergence (*d*) at all polymorphic sites and a subset of neutral SNPs (r^2 = 0.236, *P* = 0.01) and F_{ST} matrix (all: r^2 = 0.259, *P* = 0.01; neutral: r^2 = 0.223, *P* = 0.01) (Supporting Information, Table S9).

Population differentiation at mitochondrial DNA markers

The median joining haplotype network at mitochondrial DNA regions (Fig. 4) showed two main groups (Gr1 and

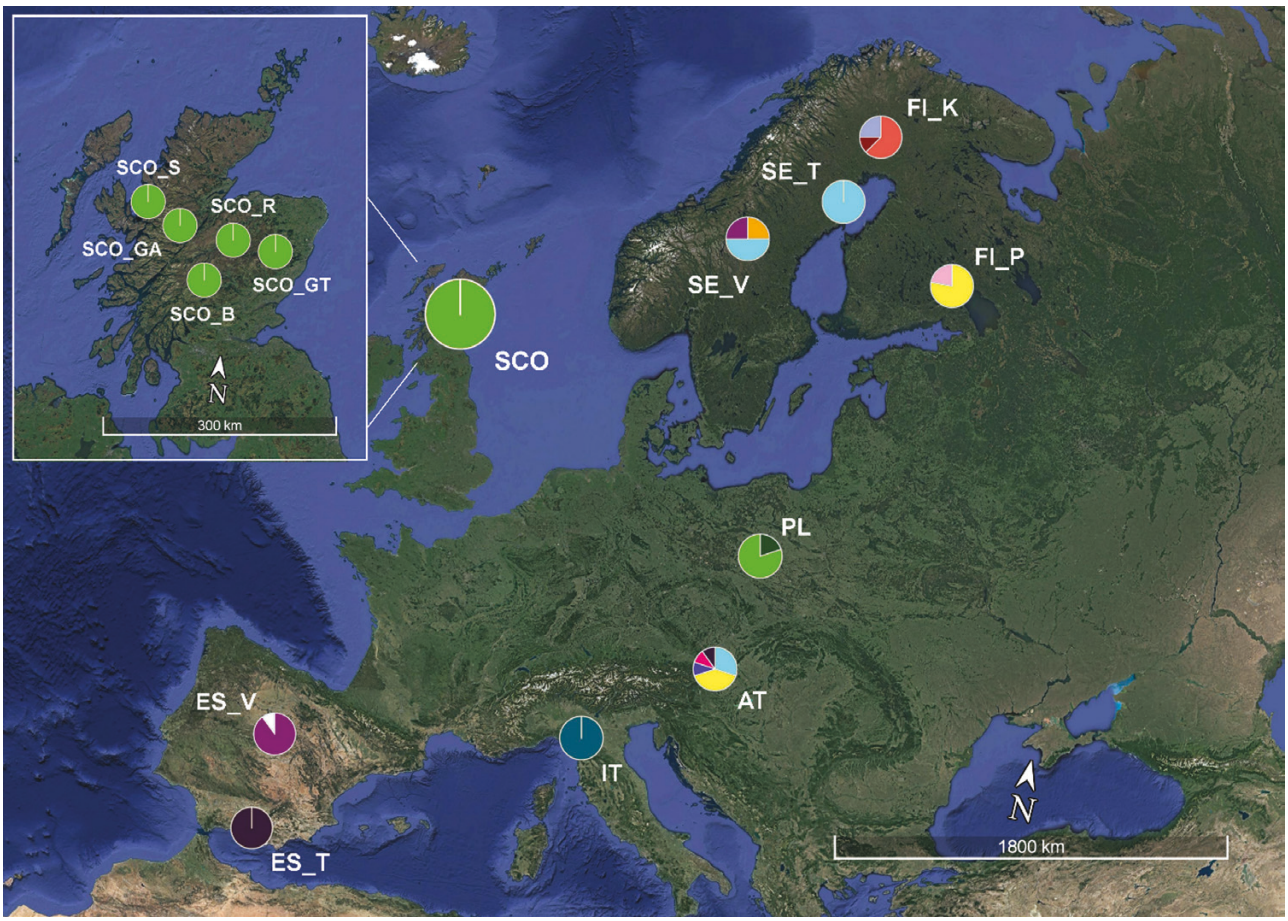

Figure 1. The geographical distribution of 14 mitochondrial DNA haplotypes in 14 populations of *Pinus sylvestris* across Europe. Colours correspond to those used in haplotype network in Figure 4.

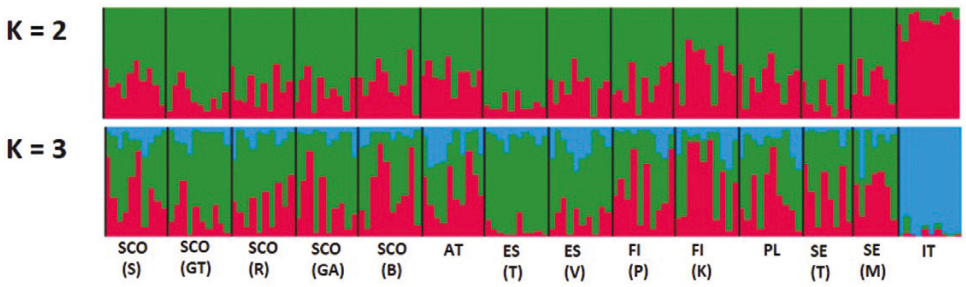

Figure 2. Outputs of clustering analysis of 14 European populations of *Pinus sylvestris* performed using all polymorphic nuclear sites for $K = 2$ and $K = 3$ in STRUCTURE software; likelihood and delta K values are shown in Figure S1 (in the Supporting Information).

Gr2), separated by four substitutions. Although the two groups were not geographically distinct, Gr1 was predominantly located in Finland (Fig. 1), whereas Gr2 was spread across most of the rest of populations. This reveals a novel but substantial genetic subdivision within Nordic populations. Looking separately at *nad1* and *nad7*, previously used in phylogeographic studies (Naydenov *et al.*, 2007; Pyhäjärvi *et al.*, 2008), we confirmed the prevalence in Europe of their variants *a* (217 bp long) and *A* (1175 bp), respectively. Variant *b* (248 bp) of *nad1* was found in Spanish population (ES_V) in the individual that was excluded, and variant *B* (1170 bp) of *nad7* was present in Austria and some Nordic populations (SE_M, FI_K and FI_P). Mantel

tests [using both between-population distance (d_{XY}) and net between-population divergence (d) and a single Scottish population (SCO_GA)], were not significant (d_{XY}: $r^2 = 0.0001$ and $P > 0.05$, d: $r^2 = 0.019$, $P > 0.05$). Comparison of N_{ST} and G_{ST} provided no indication of phylogeographic structure ($N_{ST} = 0.682 < G_{ST} = 0.746$), due to strong intra-regional differences, particularly among Nordic populations, and haplotype sharing among geographically distant populations (Hap_2 fixed in all Scottish and present in Polish population; Hap_5 occurred in central Spanish and Swedish populations). Population structure was evident in PCoA based on the genetic distance d_{XY}. PC1 (73.36% of variation) primarily separated two Finnish populations from the rest, and PC2 (12.15%) differentiated the remaining populations including Scottish and Polish vs. Austrian and Swedish populations (Supporting Information, Fig. S3). SAMOVA emphasized the distinctiveness among populations (Supporting Information, Figs S4 and S5), suggesting $K = 8$, which grouped Scottish and Polish populations, and one Swedish (SE_M) site with the Austrian samples. In AMOVA, 71.23% ($P < 0.001$) of the variation was among populations, and 77.34% ($P < 0.001$) was among SAMOVA-defined groups.

Outlier detection

Significant differentiation among populations was found at individual genes (Supporting Information, Table S10). The FDR suggested a low probability that statistically significant tests at $P < 0.01$ were false positives, with q values < 0.05 in most (87%) cases (Supporting Information, Table S10). Correspondingly, the expected proportion of significant tests ($1 - \pi_0$) was generally high, ranging from 0.640 to 0.843. Nine of the genes were significantly differentiated in at least two different regional comparisons, mostly involving Central Europe and Spain (seven genes) and Scotland versus Central Europe and the Nordic region (five genes each) (Supporting Information, Table S10). Using BAYESCAN, evidence for selection was found in two genes including a transcription factor (Pr1-9) and putative phytocyanin (*phy*). The outlier detection approach allowed identification of 74 outlier SNPs (32 excluding closely linked SNPs) between different regional groups (Supporting Information, Table S11). In total, we identified 12 loci under selection based on the compound neutrality tests and the outlier patterns of polymorphisms including significant F_{ST} ($P < 0.01$) and/or outlier SNPs among populations. They represent different gene categories

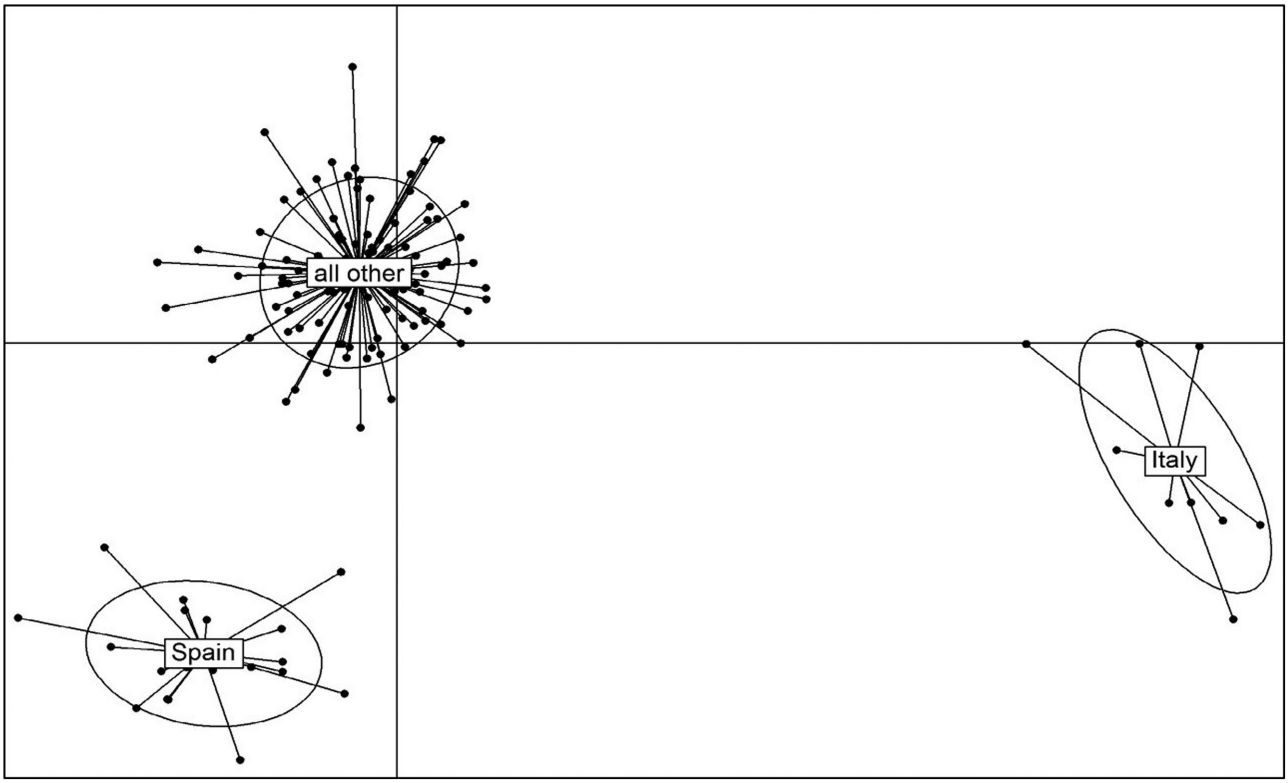

Figure 3. The outputs of the DAPC analysis for all 14 analysed populations of *Pinus sylvestris* in Europe (above) and excluding the most distinct samples from southern Europe (Italy and Spain) (next page). The plot represents the PC eigenvalues of (above) all individuals on principal component (PC) 1 (x-axis) and PC2 (y-axis) identified by DAPC.

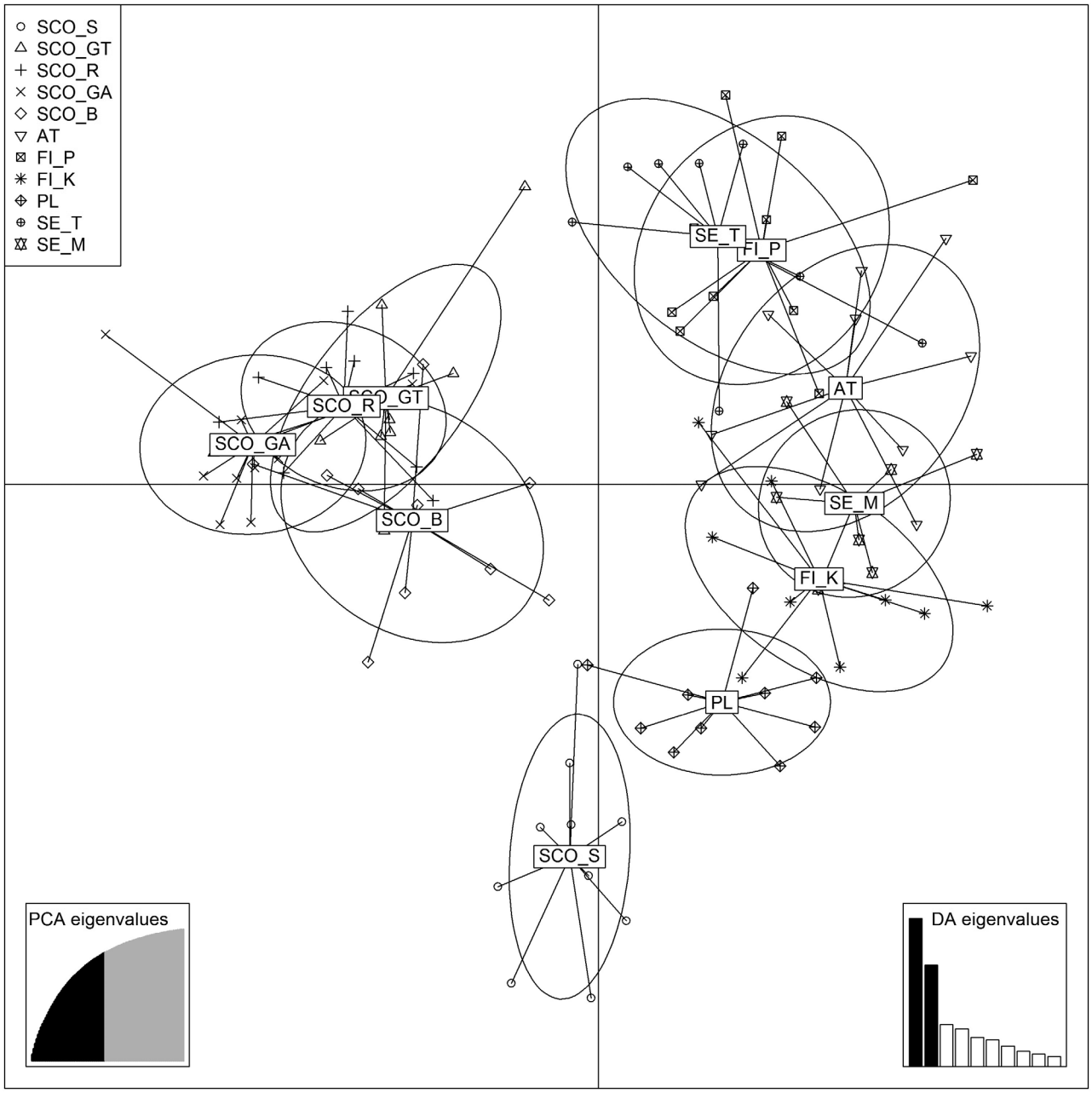

Figure 3. Continued.

and, among others, included ligase, oxidoreductase, kinase, helicase, peroxidase, hydrolase, caffeoyl CoA O-methyltransferase, phytocyanin, laccase and genes involved in transport and gene expression regulation (Supporting Information, Table S12).

SNP ASSOCIATIONS WITH ENVIRONMENTAL FACTORS

Correlations between bioclimatic and geographical variables and the set of 74 outlier SNPs were tested for by BAYENV2 using both a Bayesian approach and non-parametric Spearman's comparison of ranks. Runs of 100 000 MCMC algorithm iterations conducted to construct a background matrix for association searches showed both quick stabilization of the results within and a high convergence between the three repetitions. However, this did not ensure repeatability of the test results for the focal SNPs after 100 000 iterations, and therefore we increased that number to 10 000 000 and ran each association test three times. Subsequently,

Table 3. Analysis of molecular variance (AMOVA) for *Pinus sylvestris* populations and regional groups at a full set of polymorphic sites and a subset of neutral SNPs (estimates including population from Italy are marked in parenthesis)

Source of variation	Sum of squares	Variance components	Percentage variation
all SNPs – 14 populations (with IT)			
Among populations	1248.8 (1577.2)	3.65 (5.78)	4.91 (7.78)
Within populations	7404. 9 (7743.1)	70.73 (68.59)	95.09 (92.22)
Total	8653.7 (9320.3)	74.38 (74.37)	
all SNPs – four regional groups, no IT			
Among groups	436.0	1.94	2.59
Among populations, within groups	812.8	2.16	2.88
Within populations	7404.9	70.73	94.52
Total	8653.7	74.83	
neutral SNPs – 14 populations (with IT)			
Among populations	668.07 (773.3)	1.39 (2.01)	3.13 (4.65)
Within populations	4493.3 (4635.4)	43.01 (41.19)	96.88 (95.35)
Total	5161.4 (5408.7)	44.40 (43.20)	
neutral SNPs – four regional groups, no IT			
Among groups	189.1	0.35	0.78
Among populations, within groups	479.0	1.12	2.53
Within populations	4493.3	43.01	96.70
Total	5161.4	44.48	

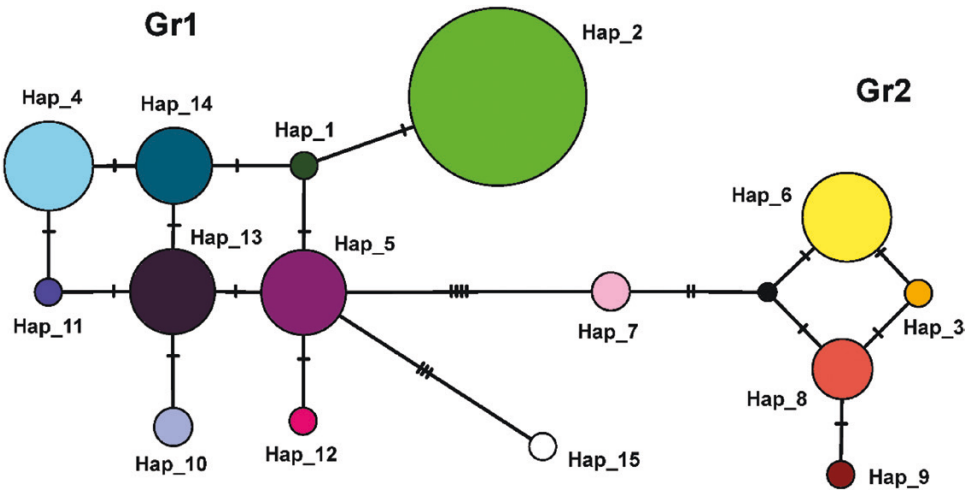

Figure 4. The median joining network of haplotypes detected at 15 mitochondrial DNA regions analysed for 14 European populations of *Pinus sylvestris*. Areas of the circles are proportional to haplotype frequencies and hatch marks represent numbers of nucleotide differences between them.

only SNPs with Bayes factor (BF) > 3 and modulus of Spearman's rank correlation coefficient $\rho > 0.3$, in all three runs, were considered as potentially associated with corresponding environmental factors. These two thresholds were chosen based on standard interpretation (e.g. Kass & Raftery, 1995) of both measures (BF values > 3.2 are regarded as substantial; values of $\rho > 0.25$–0.30 are commonly considered to indicate at least weak correlation) and slightly adjusted according to the observed results. Although the threshold values may seem soft, the combined approach should be robust (Günther & Coop, 2013). Finally, we identified 54 associations between 16 outlier SNPs from seven genes and 19 environmental

factors examined (Supporting Information, Table S13). Correlations with mean annual temperature, isothermality or cold-related variables were most common but none of the markers appeared linked to the variation in elevation, precipitation seasonality or precipitation of the warmest quarter (Supporting Information, Table S13).

DISCUSSION

Our results demonstrate that adaptive divergence in *P. sylvestris*, described in earlier quantitative genetic studies including for instance latitudinal patterns of variation in bud set, bud burst and growth rate (Wachowiak *et al.*, 2017), is maintained in populations that show the genetic signatures of range shifts and admixture. Overall, we do not see reduced genetic variation at nuclear and mitochondrial DNA markers compatible with a simple post-glacial recolonization model of populations. Instead, the level of genetic variation, haplotype diversity, linkage disequilibrium, correlation between polymorphic sites and number of unique polymorphisms in populations at higher latitudes (Nordic and Scottish stands) are similar to those found in stands at lower latitudes. Patterns of nucleotide polymorphism (excess of low frequency variants) indicated signs of population expansion, except in the Italian population. Despite their complex demographic history these distinct geographical locations showed relatively low population structure at nuclear genes and weak signs of isolation by distance due to many shared polymorphisms segregating in the populations. However, some molecular signatures indicating different origins were evident. The mitochondrial markers characterized haplotypes arranged in two genetic groups. To some extent, these were geographically distributed: southern Spanish, Scottish and Italian populations were restricted to a single genetic group, whereas Scottish and Central Spanish mitochondrial DNA haplotypes (mitotypes) were present at low frequency in Central and Northern Europe. The data provided evidence of a persistent and ancient divergence of populations from Italy and southern Spain (ES_T): each was fixed for a single unique mitochondrial haplotype, shared fewest SNPs and haplotypes and had a large number of exclusive high frequency polymorphisms at nuclear loci. The Italian population, with substantially reduced nucleotide diversity and significant excess of polymorphic sites with intermediate frequency variants, was a clear outlier in this study. The patterns of genetic variation suggest significant population size reduction of this stand. Some previous studies have also demonstrated genetic differentiation between Italian populations of *P. sylvestris* from the Apennine region and Italian Alps as the likely result of their geographical isolation and limited subsequent gene flow (Belletti *et al.*, 2012).

In the Nordic part of the distribution, our data indicate genetic differentiation at relatively short geographical distances between some Finnish and Swedish populations. Similarly, greater and more significant differentiation was found at nuclear loci between two geographically close populations in Spain and within the Nordic group than among populations distributed across large geographical distances, such as Sweden and Austria. Differentiation between geographically close populations may indicate admixture of refugial populations originating from different parts of the species range (Savolainen *et al.*, 2011; Zimmer & Sønstebø, 2018). In addition to several unique mitotypes, the Nordic stands included mitochondrial variants found in populations from Central Europe and central Spain. The highest haplotype diversity in a single population was observed in populations located in the central part of sampled distribution (Austria). The observed mixture of unique and shared mitotypes in distant populations in Central Europe and the Nordic region suggests a historical signal of colonization from multiple sources. Furthermore, ABC analysis conducted using the same nuclear SNPs analysed in this study indicated that Central European and Nordic populations were established before the LGM and did not derive from Iberian or Apennine populations (González-Díaz, 2018). Earlier genetic approaches using polymorphism at nuclear genes and coalescent simulation approaches indicated an ancient bottleneck of populations of *P. sylvestris* in Europe that predates the last glacial period (Pyhäjärvi *et al.*, 2007). As reviewed by Wójkiewicz, Cavers & Wachowiak (2016), in addition to southern locations several other areas have been proposed, based on macrofossil, biochemical and molecular studies, to had harboured populations of the species during the LGM in Scandinavia and mid-European latitudes (Kullman, 2008; Parducci, 2012), eastern Alps, Hungarian plain and Danube region (Willis, Rudner & Sümegi, 2000; Willis & van Andel, 2004; Cheddadi *et al.*, 2006) and regions to the south-east of Moscow (Buchovska *et al.*, 2013). The admixture of genetically distinct Nordic populations is also supported by the patterns of population differentiation in this area found in other forest tree species such as in the European aspen (*Populus tremula* L.) (De Carvalho *et al.*, 2010). Furthermore, the spatial distribution of unique mitochondrial haplotypes and equally high nucleotide polymorphisms at nuclear loci across the analysed stands (except the Italian population) does not reflect a pattern of southern richness and northern poverty, expected under a simple post-glacial population expansion and recolonization scenario. Similarly, a simple recolonization model does not fully account for

the high or comparable population differentiation and genetic diversity in Scottish populations in the most north-western parts of the distribution and in mainland populations (Kinloch, Westfall & Forrest, 1986; Provan *et al.*, 1998; Robledo-Amuncio *et al.*, 2004; Wachowiak *et al.*, 2011; González-Díaz *et al.*, 2017). Demographic inference indicated that the Scottish populations have probably arisen through admixture of colonization fronts from mainland Europe (González-Díaz, 2018). The presence of a single mitotype in Scottish samples suggests that a single recolonization wave played a predominant role in *P. sylvestris* and that, if multiple waves contributed to colonization, the remnants of the preceding fronts are now rare.

In our study, we also found some indications for the influence of local adaptation on molecular variation. Overall, we found low between-population variation at nuclear loci, with the majority of variation distributed within populations. Genetic relationships between populations supported those inferred from previous studies of nucleotide sequence and microsatellite data, indicating the closest relationship between Central European and Finnish populations as compared to the most southern locations of the species (Wachowiak *et al.*, 2014; Wójkiewicz & Wachowiak, 2016). High genetic similarity at pollen mediated markers may result from gene exchange after secondary contact among populations of different origins, especially considering observed patterns of seed mediated mitochondrial DNA variation. A strong balance between migration and selection in populations of *P. sylvestris* has been postulated in numerous quantitative genetic studies showing high phenotypic differentiation between populations at evolutionarily important traits such as phenology or abiotic stress resistance (Salmela *et al.*, 2011; Wachowiak *et al.*, 2017). For highly outcrossing and wind-pollinated species, this variation is thought to be maintained as a balance between gene flow and natural selection across an environmentally variable landscape. As the samples analysed in the present study were derived from geographically but also environmentally different populations, it seems reasonable to conclude that departures from neutrality might be driven by adaptive divergence. Furthermore, exactly the same populations showed differentiation at phenology and growth traits, with Nordic stands most diverged as compared to other locations (Wachowiak *et al.*, 2017).

Overall, we identified 12 loci that showed evidence for selection in neutrality tests and outlier analysis (Supporting Information, Table S12). In the group of more functionally diverse enzymes, engaged in a variety of biological processes, only general molecular functions could be recognized. Among these we found electron transport proteins including NAD(P)-linked oxidoreductase and phytocyanin, involved in numerous processes, including abiotic stress responses (Cao *et al.*, 2015; Gogoi, 2020); the peroxidases, related to scavenging of reactive oxygen species and defence against oxidative damage in response to heat, cold stress, high salinity and high metal ions concentrations (Pandey *et al.*, 2017); mitogen-activated protein-kinases (MAPKs), participating in signalling MAPK cascades directing cellular responses to mitogens, hormones or stressors (Jagodzik *et al.*, 2018) that together with another candidate gene (laccase) are linked to xylem development, regulation of gene expression and lignification (Sato, Wuli, Sederoff & Whetten, 2001; Morse *et al.*, 2009; Janusz *et al.*, 2020). Of these loci only one, phytocyanin (*phy*) SNP position 264, exhibited correlation with environmental variables, with a BF score > 10 in association tests with latitude, maximum temperature of warmest month (BIO5) and mean temperature of warmest quarter (BIO10). This result again points at an important role of this particular phytocyanin gene in adaptation of species of *Pinus* L. as shown in earlier gene expression and association genetics studies (Zhang, Sederoff & Allona, 2000; Palmé *et al.*, 2009; Wachowiak *et al.*, 2018). Phytocyanins (blue copper proteins specific for plants, taking part in redox reactions) were found to be linked to xylem formation, drought responses, mechanical stress response and primary defence response (Nersissian *et al.*, 1998; Zhang *et al.*, 2000; Trupiano *et al.*, 2012; Cao *et al.*, 2015). Another potential target of selection was a putative caffeoyl CoA O-methyltransferase, that represents cytosolic proteins primarily ascribed to processes of lignin biosynthesis and reinforcement of cell walls, responding to wounding, pathogen challenge, drought stress and to provision of feruloyl-CoA for other biosynthetic pathways (Li *et al.*, 1999). Polymorphisms at this gene showed also signatures of non-neutral evolution in *Pinus taeda* (González-Martínez *et al.*, 2006) and were significantly associated with carbon isotope discrimination in that species (González-Martínez *et al.*, 2008).

These loci complement a growing list of genes identified so far as potentially involved in local adaptation in *P. sylvestris* (Wachowiak *et al.*, 2009; Kujala & Savolainen, 2012) and would be excellent candidates for more detailed investigations of adaptive variation using much larger sampling and association genetic approaches. Suitable genomic resources for such analysis in *P. sylvestris*, including genotyping arrays with c. 50 000 SNPs, have been recently developed (Perry *et al.*, 2020; Kastally *et al.*, 2021). Considering its complex demographic history, analysis of phenotypically diverged stands at short geographical separation would minimize the confounding effects of population structure in association genetic studies. Quantitative genetic analyses provide numerous examples of such ecotype divergence at small spatial scales in different parts of the species distribution, e.g. Central Europe (Hebda

& Wachowiak, 2019), Scotland (Salmela *et al.*, 2013) and Scandinavia (Hurme *et al.*, 1997). At larger spatial scales, transects across latitudinal gradients grouping Swedish stands with Western European locations, and Finish populations with Central-Eastern European and potentially Balkan locations would pair sites that show phenotypic divergence but share much of their background genetic variation (Wójkiewicz & Wachowiak, 2016; Dering *et al.*, 2017).

CONCLUSIONS

At mitochondrial markers, we found divergence at a relatively fine spatial scale but no loss of diversity among northern populations of *P. sylvestris* at nuclear loci suggesting colonization from multiple sources and admixture of populations from different origins. Substantial genetic divergence at mitochondrial DNA loci was identified in Nordic populations. Colonization of Scotland from multiple routes was suggested by polymorphisms at nuclear loci but not mitochondrial markers, for which populations were fixed for a single mitotype. Our study describes new genetic markers for inference of population history in *P. sylvestris* and a set of candidate genes showing molecular signatures of natural selection. It provides evidence for the underlying genetic differentiation for local adaptation, in particular to temperature variation.

FUNDING

The research was financially supported by the Natural Environment Research Council, UK (grant no. NE/H003959/1). WW and JZ acknowledge financial support from Polish National Science Centre (grant no. UMO-2015/19/B/NZ9/00024).

DATA AVAILBILITY

The Scots pine data underlying this article are available in the GenBank Nucleotide Database (accession numbers KC979165-KC980899).

REFERENCES

Bandelt HJ, **Forster P**, **Röhl A. 1999.** Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16:** 37–48.

Belletti P, **Ferrazzini D**, **Piotti A**, **Monteleone I**, **Ducci F. 2012.** Genetic variation and divergence in Scots pine (*Pinus sylvestris* L.) within its natural range in Italy. *European Journal of Forest Research* **131:** 1127–1138.

Buchovska J, **Danusevicius D**, **Baniulis D**, **Stanys V**, **Siksnianiene JB**, **Kavaliauskas D. 2013.** The location of the northern glacial refugium of Scots pine based on mitochondrial DNA markers. *Baltic Forestry* **19:** 2–12.

Burban C, **Petit RJ**, **Carcreff E**, **Jactel H. 1999.** Rangewide variation of the maritime pine bast scale *Matsucoccus feytaudi* Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host. *Molecular Ecology* **8:** 1593–1602.

Cao J, **Li X**, **Lv Y**, **Ding L. 2015.** Comparative analysis of the phytocyanin gene family in 10 plant species: a focus on *Zea mays*. *Frontiers in Plant Science* **6:** 515.

Cheddadi R, **Vendramin G**, **Litt T**, **François L**, **Kageyama M**, **Lorentz S**, **Laurent J**, **Beaulieu J**, **Sadori L**, **Jost A**, **Lunt D. 2006.** Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*. *Global Ecology and Biogeography* **15:** 271–282.

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

Cornuet J-M, **Pudlo P**, **Veyssier J**, **Dehne-Garcia A**, **Gautier M**, **Leblois R**, **Marin J-M**, **Estoup A. 2014.** DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* **30:** 1187–1189.

Critchfield WB, **Little E. 1966.** *Geographic distribution of the pines of the World*. Washington: U.S. Department of Agriculture.

De Carvalho D, **Ingvarsson PK**, **Joseph J**, **Suter L**, **Sedivy C**, **Macaya-Sanz D**, **Cottrell J**, **Heinze B**, **Schanzer I**, **Lexer C. 2010.** Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Molecular Ecology* **19:** 1638–1650.

Dering M, **Kosiński P**, **Wyka TP**, **Pers-Kamczyc E**, **Boratyński A**, **Boratyńska K**, **Reich PB**, **Romo A**, **Zadworny M**, **Żytkowiak R**, **Oleksyn J. 2017.** Tertiary remnants and Holocene colonizers: genetic structure and phylogeography of Scots pine reveal higher genetic diversity in young boreal than in relict Mediterranean populations and a dual colonization of Fennoscandia. *Diversity and Distributions* **23:** 540–555.

Dupanloup I, **Schneider S**, **Excoffier L. 2002.** A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* **11:** 2571–2581.

Earl DA, **vonHoldt BM. 2012.** STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4:** 359–361.

Eveno E, **Collada C**, **Guevara M**, **Leger V**, **Soto A**, **Diaz L**, **Leger P**, **Gonzalez-Martinez S**, **Cervera M**, **Plomion C**, **Garnier-Gere P. 2008.** Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Molecular Biology and Evolution* **25:** 417–437.

Excoffier L, **Lischer HEL. 2010.** Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10:** 564–567.

Fick SE, **Hijmans RJ. 2017.** WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37:** 4302–4315.

Floran V, **Sestras RE**, **García Gil MR. 2011.** Organelle genetic diversity and phylogeography of Scots pine (*Pinus sylvestris* L.). *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* **39:** 317–322.

Foll M, **Gaggiotti O. 2008.** A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180:** 977–993.

Giertych M, **Oleksyn J. 1992.** Studies on genetic variation in Scots pine (*Pinus sylvestris L.*) coordinated by IUFRO. *Silvae Genetica* **41:** 133–143.

Gogoi SR. 2020. Applications of oxidoreductases. In: Mansour MA, ed. *IntechOpen book series, biochemistry, Vol. 19. Oxidoreductase.* London: IntechOpen, 17–30. Available at: 10.5772/intechopen.94409

González Díaz P. 2018. *Development and maintenance of genetic diversity in Scots pine,* Pinus sylvestris *(L.). PhD Thesis*, University of Stirling, 1–216.

González-Díaz P, **Jump AS**, **Perry A**, **Wachowiak W**, **Lapshina E**, **Cavers S. 2017.** Ecology and management history drive spatial genetic structure in Scots pine. *Forest Ecology and Management* **400:** 68–76.

González-Martínez SC, **Ersoz E**, **Brown GR**, **Wheeler NC**, **Neale DB. 2006.** DNA sequence variation and selection of tag SNPs at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* **172:** 1915–1926.

González-Martínez SC, **Huber D**, **Ersoz E**, **Davis JM**, **Neale DB. 2008.** Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. *Heredity* **101:** 19–26.

Grivet D, **Sebastiani F**, **Alía R**, **Bataillon T**, **Torre S**, **Zabal-Aguirre M**, **Vendramin GG**, **González-Martínez SC. 2011.** Molecular footprints of local adaptation in two Mediterranean conifers. *Molecular Biology and Evolution* **28:** 101–116.

Günther T, **Coop G. 2013.** Robust identification of local adaptation from allele frequencies. *Genetics* **195:** 205–220.

Hebda AM, **Wachowiak W. 2019.** Heterogeneous patterns of genetic variation at nuclear genes and quantitative traits in a Scots pine provenance trial. *Acta Societatis Botanicorum Poloniae* **88:** 3623.

Heuertz M, **De Paoli E**, **Kallman T**, **Larsson H**, **Jurman I**, **Morgante M**, **Lascoux M**, **Gyllenstrand N. 2006.** Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics* **174:** 2095–2105.

Hewitt GM. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405:** 907–913.

Hey J, **Wakeley J. 1997.** A coalescent estimator of the population recombination rate. *Genetics* **145:** 833–846.

Hudson RR, **Boos DD**, **Kaplan NL. 1992.** A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution* **9:** 138–151.

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 Forestière* **27:** 716–723.

Jagodzik P, **Tajdel-Zielinska M**, **Ciesla A**, **Marczak M**, **Ludwikow A. 2018.** Mitogen-activated protein kinase cascades in plant hormone signaling. *Frontiers in Plant Science* **9:** 1387.

Janusz G, **Pawlik A**, **Świderska-Burek U**, **Polak J**, **Sulej J**, **Jarosz-Wilkołazka A**, **Paszczyński A. 2020.** Laccase properties, physiological functions, and evolution. *International Journal of Molecular Sciences* **21:** 966.

Jaramillo-Correa JP, **Beaulieu J**, **Bousquet J. 2004.** Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. *Molecular Ecology* **13:** 2735–2747.

Jombart T, **Ahmed I. 2011.** adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27:** 3070–3071.

Kass RE, **Raftery AE. 1995.** Bayes factors. *Journal of the American Statistical Association* **90:** 773–795.

Kastally C, **Niskanen AK**, **Perry A**, **Kujala ST**, **Avia K**, **Cervantes S**, **Haapanen M**, **Kesälahti R**, **Kumpula TA**, **Mattila TM**, **Ojeda DI**, **Tyrmi JS**, **Wachowiak W**, **Cavers S**, **Kärkkäinen K**, **Savolainen O**, **Pyhäjärvi T. 2021.** Taming the massive genome of Scots pine with PiSy50k, a new genotyping array for conifer research. *Plant Journal* https://doi.org/10.1111/tpj.15628

Kelly JK. 1997. A test of neutrality based on interlocus associations. *Genetics* **146:** 1197–1206.

Kinloch BB, **Westfall RD**, **Forrest GI. 1986.** Caledonian Scots pine - origins and genetic structure. *New Phytologist* **104:** 703–729.

Kujala S, **Savolainen O. 2012.** Sequence variation patterns along a latitudinal cline in Scots pine (*Pinus sylvestris*): signs of clinal adaptation? *Tree Genetics & Genomes* **8:** 1451–1467.

Kullman L. 2008. Early postglacial appearance of tree species in northern Scandinavia: review and perspective. *Quaternary Science Reviews* **27:** 2467–2472.

Kumar S, **Stecher G**, **Tamura K. 2016.** MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33:** 1870–1874.

Li L, **Osakabe Y**, **Joshi CP**, **Chiang VL. 1999.** Secondary xylem-specific expression of caffeoyl-coenzyme A 3-O-methyltransferase plays an important role in the methylation pathway associated with lignin biosynthesis in loblolly pine. *Plant Molecular Biology* **40:** 555–565.

Librado P, **Rozas J. 2009.** DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451–1452.

Lowry DB, **Hoban S**, **Kelley JL**, **Lotterhos KE**, **Reed LK**, **Antolin MF**, **Storfer A. 2017.** Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources* **17:** 142–152.

Lynch M, **Crease TJ. 1990.** The analysis of population survey data on DNA sequence variation. *Molecular Biology and Evolution* **7:** 377–394.

Morse AM, **Whetten RW**, **Dubos C**, **Campbell MM. 2009.** Post-translational modification of an R2R3-MYB transcription factor by a MAP kinase during xylem development. *New Phytologist* **183:** 1001–1013.

Mosca E, **Eckert AJ**, **Liechty JD**, **Wegrzyn JL**, **La Porta N**, **Vendramin GG**, **Neale DB. 2012.** Contrasting patterns of nucleotide diversity for four conifers of Alpine European forests. *Evolutionary Applications* **5:** 762–775.

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.

Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, USA* **70:** 3321–3323.

Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.

Nersissian AM, **Valentine JS**, **Immoos C**, **Hill MG**, **Hart PJ**, **Williams G**, **Herrmann RG. 1998.** Uclacyanins, stellacyanins, and plantacyanins are distinct subfamilies of phytocyanins: plant-specific mononuclear blue copper proteins. *Protein Science* **7:** 1915–1929.

Notivol E, **Garcia-Gil MR**, **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-Revue Canadienne de Recherche Forestière* **37:** 540–551.

Oleksyn J. 1991. Inheritance of resistance to abiotic factors. In: Giertych M, Matyas C, eds. *Genetics of Scots pine*. Amsterdam: Elsevier, 219–229.

Palmé AE, **Pyhajarvi T**, **Wachowiak W**, **Savolainen O. 2009.** Selection on nuclear genes in a *Pinus* phylogeny. *Molecular Biology and Evolution* **26:** 893–905.

Pandey VP, **Awasthi M**, **Singh S**, **Tiwari S**, **Dwivedi UN. 2017.** A comprehensive review on function and application of plant peroxidases. *Biochemistry and Analytical Biochemistry* **6:** 1–16.

Parducci L, **Jørgensen T**, **Tollefsrud MM**, **Elverland E**, **Alm T**, **Fontana SL**, **Bennett KD**, **Haile J**, **Matetovici I**, **Suyama Y**, **Edwards ME**, **Andersen K**, **Rasmussen M**, **Boessenkool S**, **Coissac E**, **Brochmann C**, **Taberlet P**, **Houmark-Nielsen M**, **Larsen NK**, **Orlando L**, **Gilbert MTP**, **Kjær KH**, **Alsos IG**, **Willerslev E. 2012.** Glacial survival of boreal trees in northern Scandinavia. *Science* **335:** 1083–1086.

Peakall R, **Smouse PE. 2012.** GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28:** 2537–2539.

Perry A, **Wachowiak W**, **Downing A**, **Talbot R**, **Cavers S. 2020.** Development of a single nucleotide polymorphism array for population genomic studies in four European pine species. *Molecular Ecology Resources* **20:** 1697–1705.

Petit RJ, **Hu FS**, **Dick CW. 2008.** Forests of the past: a window to future changes. *Science* **320:** 1450–1452.

Pons O, **Petit RJ. 1996.** Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* **144:** 1237–1245.

Pot D, **Rodrigues JC**, **Rozenberg P**, **Chantre G**, **Tibbits J**, **Cahalan C**, **Pichavant Fand Plomion C. 2006.** QTLs and candidate genes for wood properties in maritime pine (*Pinus pinaster* Ait.). *Tree Genetics & Genomes* **2:** 10–24.

Pritchard JK, **Stephens M**, **Donnelly P. 2000.** Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

Provan J, **Soranzo N**, **Wilson NJ**, **McNicol JW**, **Forrest GI**, **Cottrell J**, **Powell W. 1998.** Gene-pool variation in Caledonian and European Scots pine (*Pinus sylvestris* L.) revealed by chloroplast simple-sequence repeats. *Proceedings of the Royal Society of London Series B:Biological Sciences* **265:** 1697–1705.

Prus-Glowacki W, **Stephan BR. 1994.** Genetic-variation of *Pinus sylvestris* from Spain in relation to other European populations. *Silvae Genetica* **43:** 7–14.

Pyhäjärvi T, **Garcia-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.

Ranade SS, **García-Gil MR. 2013.** Ecotypic variation in response to light spectra in Scots pine (Pinus sylvestris L.). *Tree Physiology* **33:** 195–201.

Ranade SS, **García-Gil MR. 2020.** Molecular signatures of local adaptation to light in Norway spruce. *Planta* **253**(2), 53. https://doi.org/10.1007/s00425-020-03517-9

Robledo-Amuncio JJ, **Smouse PE**, **Gil L**, **Alia R. 2004.** Pollen movement under alternative silvicultural practices in native populations of Scots pine (*Pinus sylvestris* L.) in central Spain. *Forest Ecology and Management* **197:** 245–255.

Salmela MJ, **Cavers S**, **Cottrell JE**, **Iason GR**, **Ennos RA. 2011.** Seasonal patterns of photochemical capacity and spring phenology reveal genetic differentiation among native Scots pine (*Pinus sylvestris* L.) populations in Scotland. *Forest Ecology and Management* **262:** 1020–1029.

Salmela MJ, **Cavers S**, **Cottrell JE**, **Iason GR**, **Ennos RA. 2013.** Spring phenology shows genetic variation among and within populations in seedlings of Scots pine (*Pinus sylvestris* L.) in the Scottish Highlands. *Plant Ecology & Diversity* **6:** 523–536.

Sato Y, **Wuli B**, **Sederoff R**, **Whetten R. 2001.** Molecular cloning and expression of eight laccase cDNAs in loblolly pine (*Pinus taeda*). *Journal of Plant Research* **114:** 147–155.

Savolainen O, **Kujala ST**, **Sokol C**, **Pyhäjärvi T**, **Avia K**, **Knürr T**, **Kärkkäinen K**, **Hicks S. 2011.** Adaptive potential of northernmost tree populations to climate change, with emphasis on Scots Pine (*Pinus sylvestris* L.). *Journal of Heredity* **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.

Semerikov VL, **Semerikova SA**, **Putintseva YA**, **Tarakanov VV**, **Tikhonova IV**, **Vidyakin AI**, **Oreshkova NV**, **Krutovsky KV. 2018.** Colonization history of Scots pine in Eastern Europe and North Asia based on mitochondrial DNA variation. *Tree Genetics & Genomes* **14:** 8.

Sinclair WT, **Morman JD**, **Ennos RA. 1999.** The postglacial history of Scots pine (*Pinus sylvestris* L.) in western Europe: evidence from mitochondrial DNA variation. *Molecular Ecology* **8:** 83–88.

Soranzo N, **Alia R**, **Provan J**, **Powell W. 2000.** Patterns of variation at a mitochondrial sequence-tagged-site locus provides new insights into the postglacial history of European *Pinus sylvestris* populations. *Molecular Ecology* **9:** 1205–1211.

Tajima F. 1989. Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123:** 585–595.

Tóth EG, **Köbölkuti ZA**, **Pedryc A**, **Höhn M. 2017.** Evolutionary history and phylogeography of Scots pine (*Pinus sylvestris* L.) in Europe based on molecular markers. *Journal of Forestry Research* **28:** 637–651.

Trupiano D, **Rocco M**, **Renzone G**, **Scaloni A**, **Viscosi V**, **Chiatante D**, **Scippa GS. 2012.** The proteome of *Populus nigra* woody root: response to bending. *Annals of Botany* **110:** 415–432.

Tyrmi JS, **Vuosku J**, **Acosta JJ**, **Li Z**, **Sterck L**, **Cervera MT**, **Savolainen O**, **Pyhäjärvi T. 2020.** Genomics of clinal local adaptation in *Pinus sylvestris* under continuous environmental and spatial genetic setting. *G3 Genes|Genomes|Genetics* **10:** 2683–2696.

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:** 1174132.

Wachowiak W, **Perry A**, **Donnelly K**, **Cavers S. 2017.** Early phenology and growth trait variation in closely related European pine species. *Ecology and Evolution* **8:** 6554666.

Wachowiak W, **Salmela MJ**, **Ennos RA**, **Iason G**, **Cavers S. 2011.** High genetic diversity at the extreme range edge: nucleotide variation at nuclear loci in Scots pine (*Pinus sylvestris* L.) in Scotland. *Heredity* **106:** 775–787.

Wachowiak W, **Trivedi U**, **Perry A**, **Cavers S. 2015.** Comparative transcriptomics of a complex of four European pine species. *BMC Genomics* **16:** 234.

Wachowiak W, **Wójkiewicz B**, **Cavers S**, **Lewandowski A. 2014.** High genetic similarity between Polish and North European Scots pine (*Pinus sylvestris* L.) populations at nuclear gene loci. *Tree Genetics & Genomes* **10:** 1015–1025.

Wachowiak W, **Zaborowska J**, **Łabiszak B**, **Perry A**, **Zucca G**, **González-Martínez S**, **Cavers S. 2018.** Molecular signatures of divergence and selection in closely related pine taxa. *Tree Genetics & Genomes* **14:** 83.

Willis KJ, **Rudner E**, **Sümegi P. 2000.** The full-glacial forests of central and southeastern Europe. *Quaternary Research* **53:** 203–213.

Willis KJ, **van Andel TH. 2004.** Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. *Quaternary Science Reviews* **23:** 2369–2387.

Wójkiewicz B, **Cavers S**, **Wachowiak W. 2016.** Current approaches and perspectives in population genetics of Scots pine (*Pinus sylvestris* L.). *Forest Science* **62:** 343–354.

Wójkiewicz B, **Wachowiak W. 2016.** Substructuring of Scots pine in Europe based on polymorphism at chloroplast microsatellite loci. *Flora - Morphology, Distribution, Functional Ecology of Plants* **220:** 142–149.

Zaborowska J, **Łabiszak B**, **Wachowiak W. 2020.** Population history of European mountain pines *Pinus mugo* and *Pinus uncinata* revealed by mitochondrial DNA markers. *Journal of Systematics and Evolution* **58:** 474–486.

Zadworny M, **McCormack ML**, **Mucha J**, **Reich PB**, **Oleksyn J. 2016.** Scots pine fine roots adjust along a 2000-km latitudinal climatic gradient. *New Phytologist* **212:** 389–399.

Zeng K, **Shi S**, **Wu C-I. 2007.** Compound tests for the detection of hitchhiking under positive selection. *Molecular Biology and Evolution* **24:** 1898–1908.

Zhang Y, **Sederoff RR**, **Allona I. 2000.** Differential expression of genes encoding cell wall proteins in vascular tissues from vertical and bent loblolly pine trees. *Tree Physiology* **20:** 457–466.

Zimmer K, **Sønstebø JH. 2018.** A preliminary study on the genetic structure of northern European *Pinus sylvestris* L. by means of neutral nuclear microsatellite markers. *Scandinavian Journal of Forest Research* **33:** 6–13.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table S1. Analysed nuclear loci
Table S2. Mitochondrial markers analysed including specifications and frequency of the variants in the probe
Table S3. Description of 19 bioclimatic variables acquired from the World Climate database and tested for association with outlier SNP positions
Table S4. Percentage (%) of shared SNPs and haplotypes (in parentheses) in pairwise comparisons between regional groups of *Pinus sylvestris* at 78 nuclear loci
Table S5. Loci that showed deviations from neutrality ($P < 0.01$) based on both compound neutrality tests (HEW, DHEW)
Table S6. Polymorphic sites detected at the mitochondrial DNA regions and corresponding haplotypes
Table S7. Pairwise population differentiation (F_{ST}) at a set of all polymorphic sites detected (above diagonal) and a set of neutral SNPs (below diagonal)

Table S8. Pairwise F_{ST} at all polymorphic sites combined (above diagonal) and a subset of 661 SNPs considered as neutral among regional groups defined
Table S9. Results of Mantel test between geographical and genetic distance measured as net divergence (d) and F_{ST} matrix between populations based on a set of 1046 and 661 SNPs detected
Table S10. Pairwise F_{ST} values between analysed groups of populations defined at 78 genes
Table S11. Outlier SNPs detected across loci in pairwise comparisons between geographical groups defined
Table S12. Loci under selection based on the compound neutrality tests and the outlier patterns of polymorphisms including significant F_{ST} ($P < 0.01$) and/or outlier SNPs among populations
Table S13. Outlier SNPs expressing allele frequency correlation with environmental variables as identified by BAYENV2
Figure S1. The corresponding likelihood and delta K values in STRUCTURE clustering analysis of 14 European populations of *Pinus sylvestris*.
Figure S2. Clustering of 14 European populations of *Pinus sylvestris* at a set of 661 neutral SNPs using STRUCTURE. The highest support was obtained for $K = 3$.
Figure S3. Results of the DAPC analysis based on a set of 661 neutral SNPs of A, *Pinus sylvestris* for all 14 analysed populations and B, excluding most distinct samples from southern Europe (Italy and Spain). The plot represents the PC eigenvalues of all individuals on principal component (PC) 1 (xI-axis) and PC2 (y-axis) identified by DAPC.
Figure S4. PCoA of populations of *Pinus sylvestris*, based on genetic distance d_{XY} estimated from mitochondrial haplotypes. SAMOVA groups for $K = 8$ resulted in two clusters (circled in the figure) and six separate populations.
Figure S5. The results of SAMOVA analysis for different numbers of clusters K. Among group (genetic) variation reaches the highest value at $K = 8$ with $F_{CT} = 0.773$.