ORIGINAL RESEARCH PAPER

Heterogeneous patterns of genetic variation at nuclear genes and quantitative traits in a Scots pine provenance trial

Anna Maria Hebda[1,2]*, Witold Wachowiak[3,4]

[1] Department of Plant Biotechnology, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University in Kraków, Gronostajowa 7, 30-387 Kraków, Poland
[2] Department of Genetics and Forest Tree Breeding, Institute of Forest Ecology and Silviculture, Faculty of Forestry, University of Agriculture in Krakow, 29 Listopada 46, 31-425 Kraków, Poland
[3] Institute of Environmental Biology, Faculty of Biology, Adam Mickiewicz University, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland
[4] Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland

* Corresponding author. Email: a.hebda@uj.edu.pl

Abstract

We studied the genetic variation in a set of nuclear genes analyzed from 16 populations of Scots pine derived from a 50-year-old provenance trial in Poland. At the same set of loci, the pattern of genetic variation was compared to several reference populations from a latitudinal gradient in Northern and Central Europe. Similar levels of nucleotide diversity were observed between the defined groups of Polish populations representing three climatic regions (π_{total} = 0.0040–0.0051) in comparison with the reference samples (π_{total} = 0.0054–0.0058). Polish populations showed minor but heterogeneous patterns of genetic variation between regional groups (F_{ST} up to 6%), which were caused by differentiation at specific loci. When outlier loci were excluded from between group comparisons, there were no differences between the Polish populations. Loci related to glycosyltransferase and laccase were identified as outliers, and were correlated with phenotypic differentiation using mixed-linear models. Moreover, these genes were also found as being potentially under selection across the Scots pine distribution range as the patterns of nucleotide variation correlated with latitude and altitude of the maternal stands. The provenance trial measurements have characterized a set of growth and developmental traits over 50 years and forms a suitable experimental system for detailed genetic studies.

Keywords

provenance trial; sequence variation; natural selection; candidate genes; genetic correlation

Introduction

Provenance trial experiments, in which samples from distinct geographical locations are grown under similar climatic conditions, provide useful experimental designs for studying patterns of phenotypic and genetic variation within forest trees. Genetic variation within species is influenced by population history and evolutionary processes that drive adaptation [1]. Natural selection changes allele frequencies through various mechanisms that can lead to deviations in genetic diversity or differentiation of populations at specific regions in comparison with background variation [2]. Identification of genomic regions and genes under selection is important for a better understanding of the variations in phenotypic and adaptive traits from populations in different environments. Provenance trial experiments provide useful experimental settings for

testing neutral and adaptive genetic variation of populations grown under uniform environmental conditions.

The great adaptive potential of Scots pine (*Pinus sylvestris* L.) enables the species to grow in a wide range of natural environments. Differentiation has been observed in morphological, physiological, and ecological traits across the species geographical distribution. At the same time, a large phenotypic differentiation has been associated with a generally low genetic diversity at neutral DNA markers across the Scots pine continuous European range [3,4]. Only some refugial populations, which have been reservoirs of genetic variation since the last Ice Age and are found in areas of the Balkans, Iberian Peninsula, and isolated stands of the Apennine Peninsula [5,6], show slightly higher levels of heterozygosity and different haplotype diversity patterns compared to more northern areas of the species range [7,8]. Scots pine colonization of Europe could have potentially generated clinal variation patterns of allele frequencies [9,10]. However, for wind-pollinated forest tree species, gene flow has high homogenizing effects on the allele frequency distribution over large geographical areas [11,12], with species showing generally low genetic structures [3,7,13]. Gene flow is another important factor that stabilizes the level of genetic diversity within populations [14]. Additionally, the patterns of genetic diversity in Scots pine could have potentially been affected by the trade of reforestation material from European forests in the nineteenth and twentieth centuries [6,15].

Due to the high ecological and economic importance of Scots pine, different populations were extensively studied for breeding value (e.g., [16–19]) and for determining the influence of forest management on genetic diversity [20–23]. Patterns of high phenotypic variation between populations, which were usually accompanied by low genetic differentiation at neutral markers, indicate the role of some genomic regions in the development of adaptive traits. Most characters evaluated in provenance trials and seed orchard experiments are complex quantitative traits, such as biomass production, wood quality, and biotic and abiotic stress responses [24]. Recently, genetic studies were used for the identification of the associations between genomic regions and the individual loci underlying those traits [25]. In conifers, studies have focused on the associations of quantitative traits loci (QTL) [26–33] and the identification of variation in genomic regions that were potentially related to the development of forest tree phenotypic differentiation [32,34–36]. However, even if some associations related to temperature gradients, photoperiod, or water availability were found [26–33], the variation patterns at individual genes could not always be validated in populations from different locations [37,38]. So far, studies of associations in forest trees have identified some SNPs in genes related to the timing of the bud set that showed latitudinal clines of allele frequencies (e.g., [39]). Some variations in candidate genes were associated with cold hardiness [40], cessation of growth [41], and tree height [42]. Some evidence for selection in Scots pine has also been found in several gene fragments (e.g., [1,2]).

In the current study, genetic variation was analyzed in a set of nuclear gene loci from Scots pine populations growing in different climatic zones in Poland, which were used in a provenance trial experiment. The trial characterized phenotypical traits for growth performance and wood quality and showed significant variation between populations [43]. The differentiation of the quantitative traits was accompanied by the presence of a relatively homogenous structure in these populations at a set of neutral microsatellite loci [44,45]. Using this experimental design, we assessed whether gene variability patterns are similar to patterns identified at neutral loci and if any genetic variation was a departure from neutrality. Furthermore, we used the data from nucleotide diversity analysis and phenotypic differentiation of the studied pine populations to look for correlations between genetic variation and in the differentiation of selected phenotypic traits. We compared the patterns of the genetic variation of Polish populations to other population across a broad latitudinal gradient in Europe to determine if some loci showed deviations from neutrality in the provenance trial and the reference populations.

Material and methods

Study locations, DNA extraction and amplification

Sixteen Scots pine populations derived from a species provenance trial from an experimental plot in the Carpathian Mountains were used in the study. The provenance trial was established in 1966 with 1-year-old seedlings derived from seeds collected from natural populations in three climatic zones within the Polish distribution range of Scots pine (Tab. 1, Fig. 1). The climatic zones of the parental populations were defined using the length of the vegetation period and meteorological observations from 1931 to 1960 [43,46]. The detailed characteristics of the experimental plot and seed sources were presented in an earlier study [43].

Samples for genetic analyses were collected from randomly selected individuals from each population. In total, 192 samples from the provenance trial were analyzed at the molecular level and included 12 different trees from each population (Tab. 1). Nucleotide and haplotype variations of the Polish pine populations were determined for phenotypically diverse groups of samples within the three climatic zones. This procedure was also used to compare similar number of individuals in each group.

The reference populations used for comparing genetic variation of candidate genes included stands from Northern and Central Europe (Tab. 1, Fig. 1). Genomic DNA of the provenance trial was extracted from foliage using DNeasy Plant Mini Kit (Qiagen). Fifty-five reference samples sequenced from megagametophytes, a haploid nutritive tissue that surrounds the embryo in a mature seed, were collected from 10 trees growing in most of the populations (Tab. 1). The selected nuclear genes, which were identified from Scots pine expression studies, are considered potentially important in the species adaptive variation and phenotypic differentiation. Eight gene fragments used for sequencing included loci related to cellular metabolism, transport, signal transduction, and transcription regulation (Tab. S1). PCR-amplification was performed with Thermo MBS thermal cyclers and used 15-μL samples containing about 15 ng of haploid template DNA, 10 μM of dNTP, 0.2 μM of forward primer, 0.2 μM of reverse primer, 0.15 U Taq DNA polymerase, 1× BSA, 1.5 μM of $MgCl_2$, and 1× PCR buffer (BioLabs, New England, USA). Standard amplification procedures were used with initial denaturation at 94°C for 3 min, followed by 35 cycles of 30-s denaturation at 94°C, 30-s annealing at 60°C, 90-s extension at 72°C, and a final 5-min extension at 72°C. PCR fragments were purified using ExoI–Sap (Exonuclease I, Shrimp Alkaline Phosphatase) enzymatic treatment. About 20 ng of the PCR product was used as a template in 10-μL sequencing reactions with the Big Dye Terminator DNA Sequencing Kit (Applied Biosystems, Carlsbad, CA, USA). Analyses were conducted at Genomed (Poland, Warsaw). CodonCode Aligner software ver. 3.7.1 (CodonCode, Dedham, MA, USA) was used for editing the chromatograms and for the visual inspection of all detected and aligned polymorphic sites. When nucleotide and haplotype variation were evaluated, the haplotypic sequence phase was determined for samples extracted from diploid tissue using PHASE haplotype reconstruction option as implemented in DnaSP ver. 5 (http://www.ub.edu/dnasp/) [47]. The haplotypes at each locus were compared to the haplotypes reported for the Scots pine reference samples that were collected within the European distribution of the species (Tab. S1). The reference sample haplotypic phases were known as those samples that were directly sequenced from megagametophytes that were genetically equivalent to the haploid progeny with a genotype identical to the maternal gamete.

Nucleotide and haplotype polymorphisms

The patterns of nucleotide polymorphism within the loci were analyzed in selected groups from Poland, defined on the basis of climatic zones and phenotypic differentiation of populations [43] and compared to reference geographic locations from Northern and Central Europe (Tab. 1). Nucleotide diversity was measured as the average number of nucleotide differences per site (π) between two sequences [48]. The number of haplotypes (*N*) and haplotype diversity (H_d) were computed for each gene using DnaSP ver. 5 [47]. Deviations from the frequency distribution spectrum expected under the

Tab. 1 Location and characteristics of maternal trees from the provenance trial that provided seeds, and the geographic locations of the reference populations.

Population	Location (stand)	Region	Longitude E	Latitude N	Altitude (m)	Maternal trees			
						No.	Average age (years)	Average high (m)	Average diameter (cm)
Provenance trial populations									
1.	Karsko	PL_C_1	15°15	52°54	75	18	130	28.0	48.0
2.	Tabórz	PL_N_NE_2	19°55	53°52	110	34	187	33.0	54.7
3.	Dłużek	PL_N_NE_2	20°38	53°31	145	44	166	30.6	51.4
4.	Ruciane	PL_N_NE_3	21°35	53°37	145	16	169	29.4	51.7
5.	Rozpuda	PL_N_NE_3	22°55	53°55	205	33	166	30.5	44.5
6.	Supraśl	PL_N_NE_2	23°20	53°15	165	30	147	31.9	42.8
7.	Starzyna	PL_N_NE_1	23°37	52°38	170	17	135	30.4	49.8
8.	Gubin	PL_C_1	14°50	51°55	70	17	149	27.0	45.0
9.	Rychtal	PL_C_2	17°50	51°12	190	30	108	30.0	40.0
10.	Bolewice	PL_C_2	16°10	52°23	90	30	127	25.9	40.0
11.	Lipowa	PL_N_NE_1	18°15	53°44	130	51	139	25.1	40.2
12.	Jegiel	PL_N_NE_1	21°40	52°40	95	27	135	29.0	45.6
13.	Spała	PL_C_2	20°15	51°35	150	29	127	28.8	44.3
14.	Janów Lubelski	PL_C_1	22°25	50°40	250	22	120	28.5	40.0
15.	Nowy Targ	PL_S	20°20	49°20	590	10	110	28.0	39.0
16.	Piwniczna	PL_S	20°17	49°20	500	*	*	*	*
Reference stands									
17.	Austria_Pernitz	EU_C	16°00	47°54	500				
18.	Finland_Punkaharju	EU_N	29°23	61°45	80				
19.	Finland_Kolari	EU_N	24°03	67°11	190				
20.	Poland_Jarocin	EU_C	17°28	51°58	120				
21.	Sweden_Krp. Tjärnbergsheden	EU_N	20°48	64°37	110				
22.	Sweden_Väster Mjöingenn	EU_N	13°34	62°45	640				

Samples were divided into groups from central Poland (PL_C_1 and 2), north and northeastern Poland (PL_N_NE_1, 2, and 3), southeastern Poland (PL_S), Northern Europe (EU_N), and Central Europe (EU_C) (see "Material and methods" for more details). * Temporary seed production stand. There is no information about the detailed characteristics of trees. Due to the nature of the stand, it should be assumed that it was native, and the collection included trees about 100-year old, characterized by good quality and abundance of seeds.
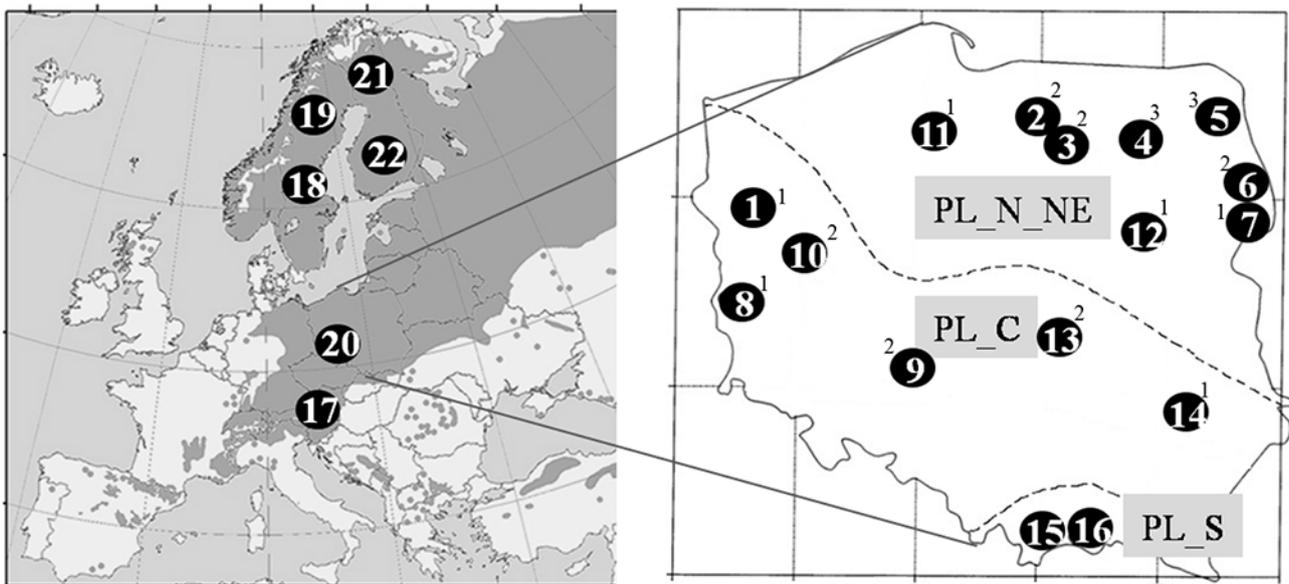
Fig. 1 Location of the Polish and reference Scots pine populations (geographic details are provided in Tab. 1) within the species European distribution range (grey). Numbers in circles are the specific populations and numbers adjacent to the circles are climatic subgroups at each population.

standard neutral model of evolution were assessed using a frequency spectrum test and coalescence-based approaches [49,50]. The distributions of Tajima's and the Fu and Li's D test statistics were investigated for each loci and selected groups of populations. Significance levels of these tests were determined by carrying out 10,000 coalescent simulations.

Outlier detection and tests for population differentiation

The significance of genetic differentiation at the loci measured as the Wright's fixation index, F_{ST} [51], was evaluated with 1,000 permutations of the samples between populations and regional groups using ARLEQUIN ver. 3.5 [52]. The full SNP dataset was used to test for loci under selection with the hierarchical analysis of Excoffier et al. [52] in ARLEQUIN ver. 3.5. Simulations estimated the null distribution and confidence intervals around the observed values, which allowed identification of outliers among locus-specific F_{ST} values. Each simulated group consisted of 100 subpopulations, and 20,000 replicates of the coalescent were used to identify the expected distribution of F_{ST}. The significance thresholds of the F_{ST} values were set at 95% and 99% of the F_{ST} values [52].

The hierarchical distribution of multiloci genetic variation among populations and tested groups was estimated using an analysis of molecular variance (AMOVA) in ARLEQUIN ver. 3.5. The standard AMOVA computations for haplotypic data was used and the significance of population genetic structure was tested using 10,000 permutations. The measurements were performed for SNPs identified in eight genes, and in genes limited to where no consistent signature of selection had been detected. Clustering analysis to examine the relationships between individual Polish populations and reference locations of the species was conducted using BAPS 6.0 software [53]. The genetic mixture analysis, which was based on all detected polymorphic sites, had ten independent runs conducted for each K (1–22) to estimate the number of clusters for the combined samples.

Correlations between SNP, phenotype, and geographic data

We analyzed and compared the patterns of nucleotide sequence variation to phenotypic and neutral variation within the same plant material. Single locus based tests were

performed for selected qualitative and quantity traits that significantly differentiated Polish Scots pine populations. Growth characteristics were available from previous research [43] and included: stand volume (Volume, $m^3 \times ha^{-1}$) that was calculated from mean tree volume in a population (m^3) divided by the area of each population at the experimental site (0.1275 ha); diameter at breast height (Diameter, cm); and diameter of approximately 50 trees per population which had the largest diameters after 47 years of growth in the experiment (Diameter_50 select). Stand volume reflects a relative measure of productivity that could be closely related to local adaptation or diameter of selected trees; possibly indicating how natural selection affects the growth of the pine populations used for genetic correlations. Furthermore, the quality traits stem straightness (SS) and crown width (CW) were included in the analysis. Those traits were scored with a 5-step scale, where: 1 was given for very crooked stems, loss of leader shoots (SS) and very wide crowns (CW), and 5 was given for very straight stems (SS) and very narrow crowns (CW) (for more details, see Tab. S3 in [43]). Differences in the traits between populations were determined using an analysis of covariance (ANCOVA). Detailed descriptions of measured and scored traits were presented in an earlier study [43].

Eight genes were used in a correlation analysis of phenotype and SNP variation and checked using a mixed linear model [54–57]. This model was fitted to represent the phenotypic traits and SNP variation that was determined as mean nucleotide and/or haplotype variation at the loci. Because phenotypic traits were measured for all trees growing in the experimental trial (about 2,300 individuals), estimated breeding values based upon a covariance analysis (ANCOVA) [43] were used as the observations for phenotypes, taking into account the genetic correlation analyses. This approach was also used to verify the possible dependence of SNP variation to the geographic characteristics latitude, longitude, and altitude of the original stands. The provenance was defined by minimal or lack of population structure, as revealed from the analysis of neutral markers [44,45]. Therefore, the signatures of selection could be effectively contrasted with background genetic variation. All correlations included Bonferroni corrections for multiple testing analyses and were conducted with the statistical software STATISTICA [58].

Results

Genetic variation in Polish populations

Over 3,000 nucleotide sites were sequenced from 247 individuals (Tab. 2), providing a set of 200 single nucleotide polymorphic sites. Similar levels of nucleotide polymorphisms were observed between defined groups of Polish Scots pine populations (π_{total} = 0.0040–0.0051) and were compared to the reference samples from Europe (π_{total} = 0.0054–0.0058) (Tab. 2). There were large nucleotide variation differences between the loci with the highest polymorphisms at Pr1_26 (π_{total} = ~0.02; Tab. S2) and the lowest polymorphism at Pr4_12 (π_{total} = ~0.0003; Tab. S2). Relative to neutral expectations, there were excesses of low frequency variants in almost all regional defined population groups as measured by the negative values of Tajima's, and the Fu and Li's, D statistics (Tab. 2). A significant positive Tajima's D statistic, indicating an intermediate alleles frequency, was found at gene Pr1_19 (populations PL_N_NE_3; Tab. S2). Simultaneously, numerous low frequency variants were found at three genes including Pr4_4 (populations EU_N), Pr4_12 (PL_C_2, PL_N_NE_3, EU_N, and EU_C), and Pr_4_19 (PL_C_1 and PL_C_2) (Tab. S2). Haplotype diversity was similar across groups of populations (0.545–0.613; Tab. 2).

Variation at individual loci and populations

In pairwise comparisons, the six groups of Polish pine populations differed significantly at five loci, including Pr1_26, Pr4_4, Pr4_12, Pr4_41, and Pr4_19 (Tab. S3). Pines from central Poland (PL_C_1 and PL_C_2) were the most differentiated and differed by at least one locus from all other groups of populations. Outlier SNPs that had high allele

Tab. 2 Summary statistics of nucleotide and haplotype variation and frequency distribution across eight nuclear genes in tested groups of Polish Scots pine populations and reference stands in Europe.

Region	n	L	Nucleotide diversity			Tajima's D	Fu and Li D	Haplotype diversity		
			SNPs	Sing	π_{total}			N	H_d	SD
PL_C_1	57.8	3,158	75	19	0.0040	−0.936	−0.384	82	0.566	0.057
PL_C_2	51.0	3,158	75	21	0.0048	−0.697	−0.591	83	0.597	0.134
PL_N_NE_1	57.0	3,158	69	24	0.0044	−0.404	−0.589	73	0.588	0.052
PL_N_NE_2	56.0	3,158	72	18	0.0042	−0.577	0.342	73	0.545	0.054
PL_N_NE_3	37.0	3,158	67	19	0.0051	−0.248	−0.181	65	0.613	0.062
PL_S	33.5	3,158	58	14	0.0045	−0.371	0.006	57	0.587	0.068
EU_N	32.5	3,158	81	28	0.0058	−0.288	−0.617	57	0.561	0.059
EU_C	19.5	3,158	66	28	0.0054	−0.299	−0.416	44	0.609	0.080
Mean	43.0	31,580	70.4	21.4	0.0048	−0.478	−0.304	66.8	0.583	0.071

n – average number of samples analyzed per locus; L – average length of the sequences in base pairs excluding indels; SNPs – number of polymorphic sites detected; Sing – number of singleton mutations; π – nucleotide diversity [48]; Tajima's D – multiloci Tajima's D statistics [49]; Fu and Li D – multiloci Fu and Li D statistics [50]; N – number of haplotypes; H_d – haplotype diversity; SD – standard deviation of haplotype diversity.

Tab. 3 F_{ST} for all polymorphic sites combined across eight genes (below the diagonal), and across groups of nuclear genes where no consistent signature of selection was detected (Pr1_12, Pr4_4, Pr4_5, and Pr4_12; above diagonal, grey distinction) between the tested groups of Polish and reference Scots pine populations.

	1	2	3	4	5	6	7	8
1. PL_C_1	0	−0.044	−0.012	0.033	−0.010	−0.043	0.051	0.042
2. PL_C_2	0.020	0	−0.017	0.008	−0.016	−0.036	0.018	0.009
3. PL_N_NE_1	0.025	0.023	0	0.026	0.014	−0.009	0.028	0.020
4. PL_N_NE_2	0.010	0.028	0.020	0	0.066	0.030	−0.044	−0.051
5. PL_N_NE_3	0.043	0.044*	0.012	0.001	0	−0.038	0.084	0.073
6. PL_S	0.061*	0.032	0.027	0.061*	0.044	0	0.046	0.037
7. EU_N	0.152*	0.194*	0.224*	0.154*	0.222*	0.329*	0	0.000
8. EU_C	0.104*	0.134*	0.174*	0.110*	0.172*	0.266*	−0.020	0

* $p < 0.01$.

frequency differences between groups of Polish Scots pine populations were found at two loci (Pr4_19 and Pr4_41) (Tab. S4). Polish populations differed at six loci from the reference populations in Europe (Pr1_19, Pr1_26, Pr4_4, Pr4_12, Pr4_19, and Pr4_41) (Tab. S3). At these loci (except Pr4_4 and Pr4_12), there were 13 SNPs that showed significant differences in frequency between the Polish and reference populations (Tab. S4).

For all polymorphic sites, genetic differentiation between selected groups of Polish populations was up to 6% (Tab. 3). Higher genetic variation was found between Polish Scots pine populations and reference groups from Northern Europe (15–33%) (Tab. 3). However, when loci considered as outliers due to patterns of genetic variation (locus Pr1_19, Pr1_26, Pr4_19, and Pr4_41) were excluded from between group comparisons, there was no differentiation between Polish populations in comparison with the reference samples (Tab. 3). Similarly, in a hierarchical AMOVA, nearly 10% of the variation was found at the between population level (Tab. S5). However, when loci showing signs of selection were excluded from the analyses, most of the genetic variation (nearly 100%) was identified within populations (data not shown). No evidence of population structure was found in the BAPS analyses (data not shown).

Correlations between genetic diversity and quantitative traits and geographical location of pine populations from Poland

A statistically significant positive correlation was found between Diameter_50 select. and nucleotide diversity at gene Pr1_19 ($r = 0.565$, $p \leq 0.05$; Tab. S6). Crown width was negatively correlated with haplotype diversity at gene Pr4_19 ($r = -0.515$, $p \leq 0.05$; Tab. S6). Across the tested Polish pine populations, significant correlations were found between the geographic location and the diversity of the three genes Pr1_19, Pr4_19, and Pr4_4 (Tab. S6). Positive correlations were identified with population latitude and the nucleotide and haplotype diversity at Pr1_19 gene ($r = 0.526$, $r = 0.425$, respectively, $p \leq 0.05$; Tab. S6), which remained significant after a Bonferroni correction for multiple testing ($p = 0.007$; Fig. 2). Some positive correlations were observed between stand altitude and the nucleotide and haplotype diversity of gene Pr4_19 ($r = 0.626$ and $r = 0.490$, respectively, $p \leq 0.05$; Tab. S6). A negative correlation was found between altitude and the haplotype diversity of gene Pr4_4 ($r = -0.576$, $p \leq 0.05$; Tab. S6). However, these positive and negative correlations were not significant after Bonferroni adjustments.
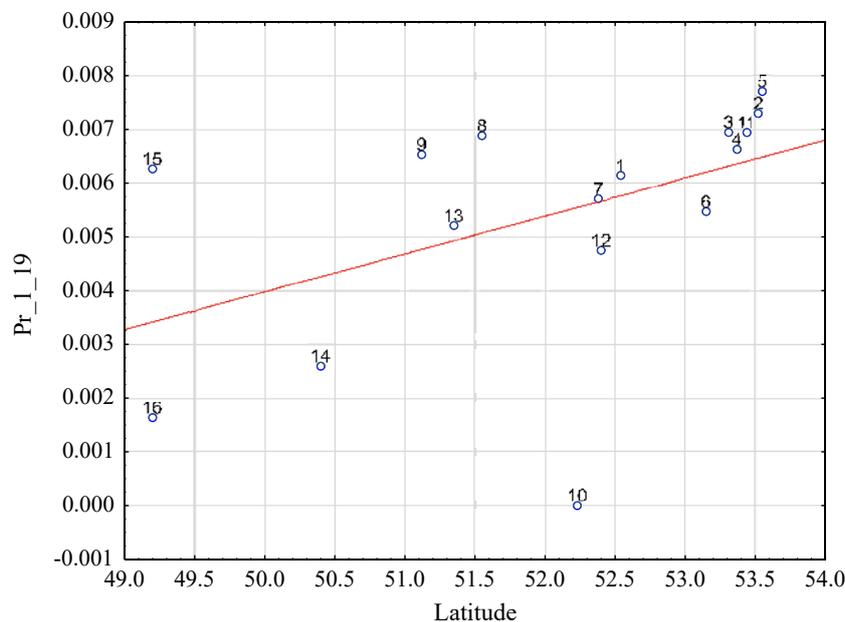

Fig. 2 Correlation between population latitude and nucleotide diversity at the Pr1_19 gene ($r = 0.526$, $p = 0.007$). 1–16 – population number.

Discussion

Nucleotide polymorphisms

In this study, we analyzed genetic differentiation within and among 16 Scots pine populations that were representative of the species distribution range and three climatic zones in Poland. We studied nucleotide polymorphisms at nuclear loci to investigate patterns of genetic variation between the Polish populations and reference samples from Northern and Central Europe. Our data indicated no differences in the nucleotide diversity at the studied loci between tested groups of Polish populations. Overall, we observed very similar haplotype and SNP frequencies among Polish populations from different climatic zones. The Polish populations also showed very similar levels of nucleotide polymorphism to the reference Scots pine populations from Northern and Central Europe, which were determined from 16 nuclear loci ($\pi = 0.0052$ [59]) and 13 cold-related genes ($\pi = 0.006$ [1]). Similar low levels of nucleotide diversity and average levels of haplotype diversity were reported for other conifers, including *Picea abies* [60] and *Pseudotsuga menziesii* [40].

Population structure

The structural homogeneity of the Polish populations, which were based on reference loci, follows patterns expected for large populations exposed to efficient gene flow [61]. Furthermore, the populations from the provenance trial showed an overall excess of low frequency mutations, as shown by the negative multiloci Tajima's *D* statistics. The pattern of genetic variation observed in the Polish populations and in the reference locations from Northern and Central Europe, is expected for recently expanded populations [61]. There were no clear signatures of selection for population structure in the European Scot pine distribution at the set of analyzed loci. Low genetic differentiation between populations (~2%) in this part of the Scots pine distribution was previously found as microsatellite [44,62] and nuclear sequence variations [61]. Gene flow by pollen carried over long geographical distances [21] has a general homogenizing effect on genetic variation in outcrossing forest tree species, which we observed as a set of neutrally evolving loci. Furthermore, forest trees species including Scots pine, have a stable genetic structure that preserves an appropriate level of population genetic variability and adaptability [14].

Patterns of genetic variation from the analyzed populations were influenced by past demographic factors [59] and also by anthropogenic factors related to historical seed transfer and forest management activities. However, the seeds that were used to establish the experimental trial in 1966 in the Carpathian Mountains were collected from old trees (aged from 195 to 108 years; 130-year old on average). Forest management in Poland would recognize such trees as of native origin, because they would predate the period of an intensification of the seed trade, which occurred between 1860 and 1910 [63]. As the trial represents half-sib progeny from natural stands, cross-pollination by maladapted foreign breeding populations cannot be completely excluded. However, even though foreign alleles possibly exist at some loci, it should not significantly bias the average estimates obtained at the population level. Thus, the trial is considered representative of the natural distribution of Scots pine in Poland, even though anthropogenic influences on the genetic variability patterns of some populations cannot be completely excluded due to the complex history of management in Europe [6,15].

Signatures of selection and genetic correlations

Considering the similar genetic backgrounds of the populations reported in earlier studies, the outlier patterns of differentiation at some loci maybe due to non-neutral processes. We identified some genes where the between-population differentiation was correlated with phenotypic variation of the populations. In our dataset, the most consistent patterns of significant population differentiation based on haplotype structure and allele frequencies were found between selected Polish groups from central, eastern and northern locations (Tab. S2 and Tab. S4). Departures from neutrality were observed at the Pr1_19 locus in populations 4-Ruciane and 5-Rozpuda from northeast Poland. The excellent breeding performance of pines from northeast Poland have been reported based on provenance trials [64,65]. Strong outlier patterns of variation for both haplotype structure and allele frequencies were found at four loci (Pr1_19, Pr1_26, Pr4_19, and Pr4_41) in pairwise comparisons between the Polish populations and reference Northern European stands. Despite the similar genetic background in this part of the species distribution at neutral markers [44], we identified several loci and SNPs, where allele frequencies significantly differed between tested regions (Tab. S6). When those loci were excluded from the between group comparisons, there were no differentiation between Polish and North European samples (Tab. 3). These results suggest that natural selection may have shaped the patterns of genetic variation in these genes, resulting in departures from neutral expectations.

Our analyses of Polish Scots pine populations showed significant correlations between diameter of selected trees (Diameter_50 select.) and crown width and polymorphisms at the Pr1_19 and Pr4_19 genes. Additionally, some trends of correlations were identified between geographic coordinates of populations and diversity of SNPs. The correlation between latitude and nucleotide diversity of Pr1_19 proved to be statistically significant after a Bonferroni correction. However, it is important to note that the Bonferroni

correction is very conservative and therefore, a less powerful test for identifying correlations between SNP variations and quantitative traits that are shaped by small effects from many genes. Gene Pr1_19 is involved in glycosyltransferase functions related to biosynthesis of polysaccharides and glycoproteins in the plant cell wall [66], which have crucial roles in plant growth and responses to biotic and abiotic stresses [67–70]. Alternatively, gene Pr4_19 encodes enzymatic proteins from the laccases group involved in lignification in conifers [67].

Considering the weak divergence between geographical regions, which was related to neutral loci with high differentiation of quantitative traits of adaptive importance, it is possible that differences in the frequency and distribution of polymorphisms at some loci may be due to diversifying selection across the Scots pine range. The capability to detect selection in genomic regions depends on the time since selection and the number of loci involved [36]. The large conifer genome size and complex interactions of many traits are the primary reasons for the identification of only a few candidate loci associated with phenotypic traits. A few candidate genes, including dehydrin [1] and the ft/tfl1-like and pseudo response regulator 1 genes [2], were identified in Scots pine as potentially being under selection. In loblolly pine (*Pinus taeda*), four cell wall genes explained about 3% of the variation in wood traits [55]. In Monterey pine (*Pinus radiata* D. Don), nine genes were associated with wood quality [68]. In maritime pine (*Pinus pinaster*), González-Martínez et al. [55] reported genetic variation and linked mutations underlying phenotypic variability. Our study provides a set of new genes distinct in their variations in comparison with neutral markers that are especially interesting for further investigations.

Conclusions

Our study showed close genetic relationships between populations from a provenance trial. Across a similar genetic background at neutral loci, we identified some outlier patterns of nucleotide diversity in the European species distribution. Additionally, nucleotide polymorphisms at some of the genes studied were significantly correlated with phenotypic variation among populations from different environments. Considering that the provenance trial examined here has been well characterized for growth and developmental traits during nearly 50 years of measurements [43], it forms a suitable experimental system for detailed comparative-association genetic studies using available genomic resources [71].

Supplementary material

The following supplementary material for this article is available at http://pbsociety.org.pl/journals/index.php/asbp/rt/suppFiles/asbp.3623/0:

Tab. S1   Analyzed loci.

Tab. S2   Summary statistics of nucleotide and haplotype variation and frequency distribution spectra at eight nuclear genes in the tested groups of Polish populations of Scots pine and reference stands in Europe.

Tab. S3   Significant values of F_{ST} statistics for corresponding loci in pairwise comparisons between groups of Polish and reference Scots pine populations.

Tab. S4   Outlier SNPs detected across loci in pairwise comparisons between Polish and reference populations.

Tab. S5   AMOVA based on all polymorphic sites detected.

Tab. S6   Correlation between total nucleotide diversity and haplotype diversity at individual loci and phenotypic traits variation as well as geographic coordinates of maternal stands of Polish pine populations.

References

1. Wachowiak W, Balk P, Savolainen O. Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in Scots pine (*Pinus sylvestris* L.). Tree Genet. Genomes. 2009;5:117–132. https://doi.org/10.1007/s11295-008-0188-3

2. Kujala ST, Savolainen O. Sequence variation patterns along a latitudinal cline in Scots pine (*Pinus sylvestris*): signs of clinal adaptation? Tree Genet Genomes. 2012;8:1451–1467. https://doi.org/10.1007/s11295-012-0532-5

3. Wójkiewicz B, Wachowiak W. Substructuring of Scots pine in Europe based on polymorphism at chloroplast microsatellite loci. Flora. 2016;220:142–149. https://doi.org/10.1016/j.flora.2016.03.005

4. Sebastiani F, Pinzauti F, Kujala ST, González-Martínez SC, Vendramin GG. Novel polymorphic nuclear microsatellite markers for *Pinus sylvestris* L. Conserv Genet Resour. 2012;4:231–234. https://doi.org/10.1007/s12686-011-9513-5

5. Petit R, Aguinagalde I, Beaulieu J, Bittkau C, Brewer S, Cheddadi R, et al. Glacial refugia: hotspots but not melting pots of genetic diversity. Science. 2003;300:1563–2035. https://doi.org/10.1126/science.1083264

6. Krakau UK, Liesebach M, Aronen T, Lelu-Walter MA, Schneck V. Scots pine (*Pinus sylvestris* L.). In: Pâques LE, editor. Forest tree breeding in Europe. Dordrecht: Springer; 2013. p. 267–323. (Managing Forest Ecosystems; vol 25). https://doi.org/10.1007/978-94-007-6146-9

7. Pyhäjärvi T, Salmela MJ, Savolainen O. Colonization routes of *Pinus sylvestris* inferred from distribution of mitochondrial DNA variation. Tree Genet Genomes. 2008;4(2):247–254. https://doi.org/10.1007/s11295-007-0105-1

8. Belletti P, Ferrazzini D, Piotti A, Monteleone I, Ducci F. Genetic variation and divergence in Scots pine (*Pinus sylvestris* L.) within its natural range in Italy. Eur J For Res. 2012;131(4):1127–1138. https://doi.org/10.1007/s10342-011-0584-3

9. Excoffier L, Ray N. Surfing during population expansions promotes genetic revolutions and structuration. Trends Ecol Evol. 2008;23:347–351. https://doi.org/10.1016/j.tree.2008.04.004

10. Frichot E, Schoville SD, de Villemereuil P, Gaggiotti OE, François O. Detecting adaptive evolution based on association with ecological gradients: orientation matters! Heredity. 2015;115:22–28. https://doi.org/10.1038/hdy.2015.7

11. Burczyk J, DiFazio SP, Adams WT. Gene flow in forest trees: how far do genes really travel? For Genet. 2004;11(2–3):1–14.

12. Varis S, Pakkanen A, Galofre A, Pulkkinen P. The extent of south-north pollen transfer in Finnish Scots pine. Silva Fennica. 2009;43:717–726. https://doi.org/10.14214/sf.168

13. Wachowiak W. Relacje genetyczne pomiędzy polskimi i referencyjnymi populacjami sosny zwyczajnej (*Pinus sylvestris* L.) z Europy w analizie polimorfizmu sekwencji nukleotydowych loci nDNA. Sylwan. 2015;159:53–61.

14. Savolainen O. Guidelines for gene conservation based on population genetics. In: Krishnapillay B, Soepadmo E, Arshad NL, Wong A, Appanah S, Chik SW, et al., editors. Forest and society: the role of reaserch. XXI IUFRO World Congress; 2000 Oct 7–12; Kuala Lumpur, Malaysia. Vienna: IUFRO; 2000. p. 100–109.

15. Mason WL, Alía R. Current and future status of Scots pine (*Pinus sylvestris*) forests in Europe. Investigación Agraria, Sistemas y Recursos Forestales. 2000;1:317–335.

16. Write J, Bull W. Geographic variation in Scots pine. Silvae Genet. 1963;12:1–40.

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

18. Eriksson G. *Pinus sylvestris* recent genetic research. Uppsala: Swedish University of Agricultural Sciences; 2008.

19. Barzdajn W, Kowalkowski W, Chmura DJ. Variation in growth and survival among European provenances of *Pinus sylvestris* in a 30-year-old experiment. Dendrobiology. 2016;75:67–77. https://doi.org/10.12657/denbio.075.007

20. Savolainen O, Kärkkäinen K. Effect of forest management on gene pools. New For (Dordr). 1992;6:329–345. https://doi.org/10.1007/978-94-011-2815-5_17

21. Burczyk J. Systemy krzyżowania drzew leśnych. Bydgoszcz: Wydawnictwo Uczelniane WSP; 1998.

22. Kosińska J, Lewandowski A, Chałupka W. Genetic variability of Scots pine maternal populations and their progenies. Silva Fennica. 2007;41(1):5–12. https://doi.org/10.14214/sf.304

23. Cunningham SC, Mac Nally R, Baker PJ, Cavagnaro TR, Beringer J, Thomson JR, et al. Balancing the environmental benefits of reforestation in agricultural regions. Perspect Plant Ecol Syst. 2015;17(4):301–317. https://doi.org/10.1016/j.ppees.2015.06.001

24. Plomion C, Cooke J, Richardson T, Mackay J, Tuskan G. Report on the forest trees workshop at the plant and animal genome conference. Comp Funct Genomics. 2003;4:229–238. https://doi.org/10.1002/cfg.262

25. Kujala ST, Knürr T, Kärkkäinen K, Neale DB, Sillanpää MJ, Savolainen O. Genetic heterogeneity underlying variation in a locally adaptive clinal trait in *Pinus sylvestris* revealed by a Bayesian multipopulation analysis. Heredity. 2016;118(5):413–423. https://doi.org/10.1038/hdy.2016.115

26. Groover A, Devey M, Fiddler T, Lee J, Megraw R, Mitchel-Olds T, et al. Identification of quantitative trait loci influencing wood specific gravity i n an outbred pedigree of Loblolly pine. Genetics. 1994;138:1293–1300.

27. Hurme P, Repo T, Savolainen O, Pääkkönen T. Climatic adaptation of bud set and frost hardiness in Scots pine (*Pinus sylvestris*). Can J For Res. 1997;27:716–723. https://doi.org/10.1139/x97-052

28. Jermstad KD, Bassoni DL, Wheeler NC, Anekonda TS, Aitken SN, Adams WT, et al. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas fir. II. Spring and fall cold-hardiness. Theor Appl Genet. 2001;102:1152–1158. https://doi.org/10.1007/s001220000506

29. Jermstad KD, Bassoni DL, Jech KS, Ritchie GA, Wheeler NC, Neale DB. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas fir. III. Quantitative trait loci-by-environment interactions. Genetics. 2003;165:1489–506.

30. Sewell MM, Davis MF, Tuskan GA, Wheeler NC, Elam CC, Bassoni DL, et al. Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). II. Chemical wood properties. Theor Appl Genet. 2002;104:214–222. https://doi.org/10.1007/s001220100697

31. Sewell MM, Neale DB. Mapping quantitative traits in forest trees. In: Jain SM, Minocha SC, editors. Molecular biology of woody plants. Dordrecht: Springer; 2000. p. 407–423. (Forestry Sciences; vol 64). https://doi.org/10.1007/978-94-017-2311-4_17

32. Wheeler NC, Jermstad KD, Krutovsky K, Aitken SN, Howe GT, Krakowski J, et al. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas fir. IV. Cold-hardiness QTL verification and candidate gene mapping. Mol Breed. 2005;15:145–156. https://doi.org/10.1007/s11032-004-3978-9

33. Salmela MJ, Cavers S, Cottrell JE, Iason GR, Ennos RA. Spring phenology shows genetic variation among and within populations in seedlings of Scots pine (*Pinus sylvestris* L.) in the Scottish Highlands. Plant Ecol Divers. 2013;6:523–536. https://doi.org/10.1080/17550874.2013.795627

34. Gion JM, Rech P, Grima-Pettenati J, Verhaegen D, Plomion C. Mapping candidate genes in *Eucalyptus* with emphasis on lignification genes. Mol Breed. 2000;6:441–449. https://doi.org/10.1023/A:1026552515218

35. Neale DB, Sewell MM, Brown GR. Molecular dissection of the quantitative inheritance of wood property traits in loblolly pine. Ann For Sci. 2002;59:595–605. https://doi.org/10.1051/forest:2002045

36. Wright SI, Gaut BS. Molecular population genetics and the search for adaptive evolution in plants. Mol Biol Evol. 2005;22:506–519. https://doi.org/10.1093/molbev/msi035

37. Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen THH. From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. Can J Bot. 2003;81:1247–1266. https://doi.org/10.1139/B03-141

38. Savolainen O, Bokma F, Knürr T, Kärkkäinen K, Pyhäjärvi T, et al. Adaptation of forest trees to climate change. In: Koskela J, Buck A, Teissier du Cros E, editors. Climate change and forest genetic diversity: implications for sustainable forest management in Europe. Rome: Bioversity International; 2007. p. 19–30.

39. Ingvarsson PK, Garcia MV, Luquez V, Hall D, Jansson S. Nucleotide polymorphism and phenotypic associations within and around the phytochrome B2 locus in European aspen (*Populus tremula*, Salicaceae). Genetics. 2008;178:2217–2226. https://doi.org/10.1534/genetics.107.082354

40. Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky KV, et al. Association genetics of coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-hardiness related traits. Genetics. 2009;182:1289–1302. https://doi.org/10.1534/genetics.109.102350

41. Ma XF, Hall D, Onge KRS, Jansson S, Ingvarsson PK. Genetic differentiation, clinal variation and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. Genetics. 2010;186:1033–1044. https://doi.org/10.1534/genetics.110.120873

42. Prunier J, Pelgas B, Gagnon F, Desponts M, Isabel N, Beaulieu J, et al. The genomic architecture and association genetics of adaptive characters using a candidate SNP approach in boreal black spruce. BMC Genomics. 2013;14:368. https://doi.org/10.1186/1471-2164-14-368

43. Hebda A, Wachowiak W, Skrzyszewski J. Long-term growth performance and productivity of Scots pine (*Pinus sylvestris* L.) populations. Acta Soc Bot Pol. 2017;86:3521. https://doi.org/10.5586/asbp.3521

44. Hebda A, Wójkiewicz B, Wachowiak W. Genetic characteristics of Scots pine in Poland and reference populations based on nuclear and chloroplast microsatellite markers. Silva Fennica. 2017;51(2):1721. https://doi.org/10.14214/sf.1721

45. Hebda A, Skrzyszewski J, Wachowiak W. Zróżnicowanie fenotypowe i zmienność tła genetycznego polskich proweniencji sosny zwyczajnej. Sylwan. 2017;161(4):277–286.

46. Sabor J. Zmienność wewnątrzgatunkowa drzew leśnych. Sosna zwyczajna i świerk pospolity. In: Saobr J, editor. Elementy genetyki i hodowli selekcyjnej drzew leśnych. Warszawa: CILP; 2006. p. 127–142.

47. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–1452. https://doi.org/10.1093/bioinformatics/btp187

48. Nei M. Molecular evolutionary genetics. New York, NY: Columbia University Press; 1987. https://doi.org/10.7312/nei-92038

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

50. Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics. 1993;133:693–709.

51. Weir BS, Cockerham CC. Estimating *F*-statistics for the analysis of population structure. Evolution. 1984;38:1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

52. Excoffier L, Lischer HEL. ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10:564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

53. Corander J, Tang J. Bayesian analysis of population structure based on linked molecular information. Math Biosci. 2007;205:19–31. https://doi.org/10.1016/j.mbs.2006.09.015

54. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet. 2006;38:203–208. https://doi.org/10.1038/ng1702

55. González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB. Association genetics in *Pinus taeda* L. I. Wood property traits. Genetics. 2007;175:399–409. https://doi.org/10.1534/genetics.106.061127

56. Beaulieu J, Doerksen T, Boyle B, Clément S, Deslauriers M, Beauseigle S, et al. Association genetics of wood physical traits in the conifer white spruce and relationships with gene expression. Genetics. 2011;188:197–214. https://doi.org/10.1534/genetics.110.125781

57. Guerra FP, Wegrzyn JL, Sykes R, Davis MF, Stanton BJ, Neale DB. Association genetics of chemical wood properties in black poplar (*Populus nigra*). New Phytol. 2013;197:162–176. https://doi.org/10.1111/nph.12003

58. StatSoft. STATISTICA (data analysis software system). Version 10.0 [Software]. Tulusa, OK: StatSoft; 2011.

59. Pyhäjärvi T, García-Gil MR, Knürr T, Mikkonen M, Wachowiak W, Savolainen O. Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. Genetics. 2007;177:1713–1172. https://doi.org/10.1534/genetics.107.077099

60. Heuertz M, de Paoli E, Källman T, Larsson H, Jurman I, Morgante M, et al. Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst)]. Genetics. 2006;174:2095–2105.

https://doi.org/10.1534/genetics.106.065102

61. Wachowiak W, Wójkiewicz B, Cavers S, Lewandowski A. High genetic similarity between Polish and North European Scots pine (*Pinus sylvestris* L.) populations at nuclear gene loci. Tree Genet Genomes. 2014;10:1015–1025. https://doi.org/10.1007/s11295-014-0739-8

62. González-Díaz P, Jump AS, Perry A, Wachowiak W, Lapshina E, Cavers S. Ecology and management history drive spatial genetic structure in Scots pine. For Ecol Manage. 2017;400:68–76. https://doi.org/10.1016/j.foreco.2017.05.035

63. Fonder W, Załęski A, Matras J. Leśna baza nasienna w Polsce. Warszawa: Dyrekcja Generalna Lasów Państwowych; 2007.

64. Białobok S. Ochrona. In: Białobok S, Boratyński A, Bugała W, editors. Biologia sosny zwyczajnej. Poznań-Kórnik: Sorus; 1993. p. 459–496.

65. Jaworski A. Sosna zwyczajna. *Pinus sylvestris* L. In: Jaworski A, editor. Charakterystyka hodowli drzew i krzewów leśnych. Tom III. Hodowla lasu. Warszawa: Powszechne Wydawnictwo Rolnicze i Leśne; 2011. p. 19–85.

66. Hansen SF, Harholt J, Oikawa A, Scheller HV. Plant glycosyltransferases beyond CAZy: a perspective on DUF families. Front Plant Sci. 2012;3:59. https://doi.org/10.3389/fpls.2012.00059

67. Richardson A, Duncan J, McDougall G. Oxidase activity in lignifying xylem of a taxonomically diverse range of trees: identification of a conifer laccase. Tree Physiol. 2000;20:1039–1047. https://doi.org/10.1093/treephys/20.15.1039

68. Dillon SK, Nolan M, Li W, Bell C, Wu HX, Southerton SG. Allelic variation in cell wall candidate genes affecting solid wood properties in natural populations and land races of *Pinus radiata*. Genetics. 2010;185:1477–1487. https://doi.org/10.1534/genetics.110.116582

69. Huang J, Pang C, Fan S, Song M, Yu J, Wei H, et al. Genome-wide analysis of the family 1 glycosyltransferases in cotton. Mol Genet Genomics. 2015;290(5):1805–1818. https://doi.org/10.1007/s00438-015-1040-8

70. Li P, Li YJ, Zhang FJ, Zhang GZ, Jiang XY, Yu HM, et al. The *Arabidopsis* UDP-glycosyltransferases UGT79B2 and UGT79B3, contribute to cold, salt and drought stress tolerance via modulating anthocyanin accumulation. Plant J. 2017;89(1):85–103. https://doi.org/10.1111/tpj.13324

71. Wachowiak W, Trivedi U, Perry A, Cavers S, Perry A. Comparative transcriptomics of a complex of four European pine species. BMC Genomics. 2015;16:234. https://doi.org/10.1186/s12864-015-1401-z