Genetic Diversity and Population Structure of Scots Pine (*Pinus sylvestris* L.) in Middle Siberia

Marina Sheller [1,2,*], Endre György Tóth [3], Elena Ciocîrlan [2], Pavel Mikhaylov [1], Sergey Kulakov [1,4], Nadezhda Kulakova [1], Natalia Melnichenko [1], Aleksey Ibe [5], Tatyana Sukhikh [5] and Alexandru Lucian Curtu [2]

[1] Scientific Laboratory of Forest Health, Reshetnev Siberian State University of Science and Technology, 660037 Krasnoyarsk, Russia

[2] Faculty of Silviculture and Forest Engineering, Transilvania University of Brasov, 500123 Brasov, Romania

[3] Forest Research Institute (FRI), University of Sopron (UOS), Várkerület 30/a, 9600 Sárvár, Hungary

[4] Sukachev Institute of Forest SB RAS, Federal Research Center «Krasnoyarsk Science Center SB RAS», 660036 Krasnoyarsk, Russia

[5] Branch of the Russian Centre of Forest Health—Centre of Forest Health of Krasnoyarsk Krai, 660036 Krasnoyarsk, Russia

* Correspondence: maralexsheller@mail.ru

Abstract: Scot pine (*Pinus sylvestris* L.) is one the main forest-forming tree species in Russia, and is of great ecological, economic and social value. In order to manage the genetic resources of this important forest tree species efficiently, knowledge of its genetic diversity and structure is needed. Here, we assessed the genetic variation of 17 populations representing different parts of Scots pine range in Russia by using nuclear microsatellite markers (nSSR). Specifically, 14 populations were chosen within the natural distribution range of the species in Middle Siberia and three distant populations were sampled from the European part of Russia, the West Siberia and the Russian Far East. All populations showed high values of genetic diversity (H_E = 0.514) and A_R = 4.150. However, the easternmost population has shown the lowest level of genetic diversity (H_E = 0.433) and allelic richness (A_R = 3.505). Five genetic groups could be detected that correspond to: the European part of Russia, the south of Middle Siberia, the northwest of Middle Siberia, West and Middle Siberia, and the Russian Far East. However, the European population was the most genetically distinct one. The variation among Scots pine populations accounted for only 5% of the total variance. The highest level of genetic differentiation was found only between westernmost and easternmost populations (F_{ST} = 0.097). Our genetic data may contribute to a better understanding of the pattern of genetic diversity of Scots pine populations in Middle Siberia and help the conservation efforts of these genetic resources.

Keywords: Scots pine; *Pinus sylvestris* L.; genetic diversity; genetic structure; nSSR

1. Introduction

Genetic diversity is important for the long-term survival of species and plays a key role in their conservation [1]. It can contribute to the adaptability of species and increase chances that at least some individuals in a population are capable to withstand changing environmental conditions [2]. Meta-analyses of hundreds of datasets showed that genetic diversity is decreasing because of habitat degradation and population loss, unsustainable harvest, invasive species and increasing extreme climatic events [3]. Low level of genetic diversity will increase extinction risk [1]. Thus, to efficiently conserve the genetic diversity of a species, the level of genetic diversity should be defined [4].

Scots pine (*Pinus sylvestris* L.) is one of the most widespread tree species in the boreal forest of Eurasia. It has great ecological, economic and social importance [5–8]. Within its vast distribution area, Scots pine grows in various soils and climatic conditions and differs significantly in morphological, ecological, and physiological characteristics [9]. More than

150 morphological varieties have been characterized along the species range [10]. In Russia, pine forests are concentrated in Siberia, where they account for up to 30% of all coniferous forests [11]. In Siberia, pine forests are distributed as a complex spatial mosaic and only in the Angara river basin do they form a continuous zone. Permafrost and pyrogenic factors play a major role in determining the northern and southern boundaries of the species' distribution in Siberia [12].

In Russia, Scots pine forests are in the first place in terms of the timber production and clear cuttings are widely used [13,14]. At the same time, reforestation processes are passive, often significantly delayed and occur with an undesirable succession of tree species [13]. Due to natural disturbances (mainly forest fires), overexploitation and mismanagement, the area of pine forests in Russia has been decreased by two million ha over the last ten years [15]. Therefore, the study of Scots pine genetic resources in Russia is highly relevant for a sustainable use in breeding and conservation programs [16,17].

Over the past decades, using different marker systems, extensive information has been obtained on the genetic diversity and structure of Scots pine [18,19]. Microsatellites or simple sequence repeats (SSRs) markers are widely used for population genetic studies of Scots pine [5,6,20,21]. Due to their properties, such as high polymorphism, amenability to multiplexing, as well as inter- and intralaboratory reproducibility, they can be useful to study genetic diversity and population structure analysis of coniferous tree species characterized by a low level of inter-population diversity.

In this study, we explored the genetic diversity and population structure of 17 Scots pine populations from different locations of the natural distribution range in Russia, using nuclear microsatellite markers. We addressed the following questions: (i) What is the current level of genetic diversity and differentiation within and among Scots pine populations across Russia? (ii) Is there a population structure among the studied Scots pine populations?

2. Materials and Methods

2.1. Plant Material

A total of 17 natural populations of Scots pine were collected from the Russian distribution of the species. Among them, 14 populations were chosen within the natural distribution range of the species in Middle Siberia. Three distant populations were included in the study: one from the European part of Russia (SHA), one from West Siberia (ARO) and one from the Russian Far East (SVO) (Figure 1, Table 1). Geographic distances between populations ranged from 14 km (VAN2 and VAN1) up to 9656 km (SHAT and SVO).

Table 1. Geographic location of 17 Scots pine populations across Russia.

Pop ID	Population	Sample Size	Latitude/ Longitude	Altitude (m)
BOR	Bor	29	61.58/89.99	65–67
VAN2	Vanavara_2	29	60.36/102.37	260–270
VAN1	Vanavara_1	23	60.24/102.43	350–360
CHU	Chunoyar	22	57.44/97.37	165–170
ARO	Aromashevo	30	57.21/69.14	70–72
BORZ	Borzovo	27	57.17/97.27	275–280
SUK2	Sukhobuzimskoye_2	18	56.50/93.22	155–160
SUK1	Sukhobuzimskoye_1	14	56.30/92.97	280–290
ZEL	Zeledeevo	17	56.22/92.24	325–330
SHA	Shalo	26	55.76/93.76	340–350
NAR	Narva	16	55.45/93.73	360–380
KOS	Kostino	26	55.00/89.40	550–560
SAR	Sarala	29	54.87/89.22	500–520
SHAT	Shatsk	30	53.53/41.23	148–150
YRB	Yrban	18	52.72/95.74	860–980
SVO	Svobodny	30	51.16/127.97	167–170
BAL	Balgazyn	22	51.08/95.09	970–1060
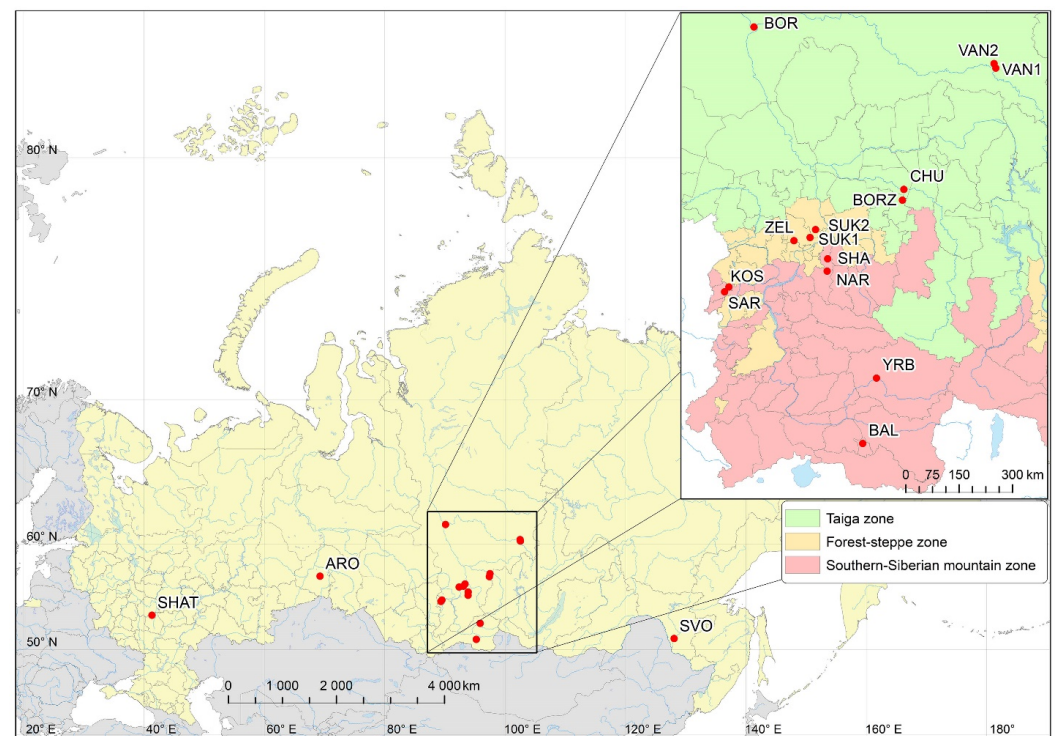
Figure 1. Map of Russia showing the geographic locations of the studied 17 Scots pine populations (acronyms are as in Table 1).

Initially, 510 individuals were sampled across Russia. Due to PCR failure, the number of studied trees was reduced to 406.

2.2. DNA Isolation and Microsatellite Genotyping

DNA isolation was performed according to the CTAB method [22]. The quality and quantity of the isolated DNA was measured with a Nanodrop 8000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA), then adjusted to a concentration of 10–20 ng/μL. Initially, ten SSR primer pairs were used in the genetic analysis: Psyl16, Psyl17, Psyl42, Psyl44, Psyl57 [23]; PtTX2146 [24]; PtTX4001 [25]; lw_isotig04195, lw_isotig04306, lw_isotig07383 [26]. All primers were combined into three multiplex sets: set 1 consisted of Psyl44, Psyl57 and lw_isotig04306; set 2 comprised of Psyl16, PtTX2146 and lw_isotig07383; and set 3 contained Psyl17, Psyl42, lw_isotig04195 and PtTX4001. Reverse primers were labelled for use in GenomeLab GeXP genetic analyzer (Beckman Coulter, Fullerton, CA, USA). The chosen loci were amplified using Qiagen Multiplex PCR Kits (Qiagen, Hamburg, Germany) under conditions recommended by the manufacturer. PCR reactions were performed under the following conditions: an initial denaturation of 5 min at 95 °C, then 32 cycles of 30 s at 95 °C, 90 s at 58 °C, 50 s at 72 °C and a final extension of 7 min at 72 °C. Genotyping of individuals was performed using the GenomeLab GeXP software (Version 10.2, Beckman Coulter, Fullerton, CA, USA).

2.3. Data Analysis

Micro-Checker software [27] was used to check all SSR loci for null alleles and possible misprints. The presence of null alleles at high frequencies was indicated for two microsatellite markers (Psyl17 and PtTX4001), which were excluded from further analysis. Then, standard genetic diversity parameters (number of alleles (N_A), number of effective alleles (N_E), inbreeding coefficient of an individual relative to the subpopulation (F_{IS}), inbreeding coefficient of an individual relative to the total population (F_{IT}), genetic differentiation coefficient (F_{ST}), observed heterozygosity (H_O) and expected heterozygosity (H_E) were

calculated for each population, using GenAlEx v. 6.5 software [28]. The allelic richness (A_R) was calculated in R [29] using the "hierfstat" package [30].

Analysis of molecular variance (AMOVA), implemented in GenAlEx v. 6.5 software, was used to determine the partitioning of the genetic variation among populations. The significance of differences was estimated using a permutation approach with 999 replications. Besides, the unweighted pair-group method with arithmetic mean (UPGMA) was used to perform cluster analysis on the Nei's genetic distances data [31] and a Principal Coordinates Analysis (PCoA) was conducted using the "ade4" package in R [32] to compare genetic differentiation among populations. In addition, population pairwise F_{ST} values were calculated based on Nei [33], plotted in the form of a heat map, and paired with UPGMA and the cluster assignment detailed below.

STRUCTURE software v.2.3.4 was used to analyze population structure with a Bayesian clustering approach [34]. We tested twenty independent runs with *K* from one to 10, each run had a burn-in period of 100,000 iterations and 500,000 Monte Carlo Markov iterations, assuming admixture model (with LocPrior) with correlated allele frequencies. The studied populations were separated into groups by the web-based Structure HARVESTER [35] based on ΔK and mean L(*K*) values [36]. The average matrices of individual membership proportions for each population were estimated using CLUMPP v.1.1.2. [37].

BARRIER software v.2.2 [38] was applied to detect any potential barriers to gene flow among the studied populations using Monmonier's maximum-difference algorithm [39]. One thousand distance matrices (*D*) (Nei's standard genetic distance corrected for sample size) were generated in MSA software [40] by bootstrapping over the eight nSSR loci. The generated matrices were used to assess possible boundaries between populations.

A Mantel test, the correlation between geographical distances and genetic distances, was conducted to test for the Isolation-By-Distance (IBD) hypothesis. The significance of Mantel test was checked by comparing the absence of spatial structure against the correlation between the distance matrices (the observed value). The absence of spatial structure was simulated with a Monte-Carlo permutation approach applying 9999 repetitions using the "adegenet" package in R [41]. Furthermore, to test whether IBD present in clines of genetic differentiation or in distant patches, a 2-D kernel density estimation was applied using the "mass" and "ggplot" R packages [42].

3. Results

Six to 19 alleles were detected per locus, with a total of 70 alleles across all loci and populations (Table 2). Inbreeding coefficient of an individual relative to the subpopulation (F_{IS}) ranged from −0.115 (Psyl42) to 0.061 (lw_isotig04195), with an average value of −0.033 per locus. Genetic differentiation (F_{ST}) varied in the range between 0.024 (Psyl42) to 0.071 (lw_isotig07383) with an average value of 0.041 per locus.

Table 2. Diversity statistics of the eight nuclear SSR loci across 406 Scots pine individuals.

Locus	N_A	F_{IS}	F_{IT}	F_{ST}
Psyl44	6	−0.059	−0.034	0.024
Psyl42	7	−0.115	−0.068	0.042
Psyl57	7	−0.101	−0.065	0.033
Psyl16	7	0.014	0.055	0.042
lw_isotig04306	10	−0.005	0.034	0.039
lw_isotig07383	8	−0.014	0.058	0.071
lw_isotig04195	6	0.061	0.091	0.033
PtTX2146	19	−0.042	0.006	0.046
Mean	8.750	−0.033	0.010	0.041
SD	±1.638	±0.021	±0.021	±0.005

Note: number of alleles (N_A); inbreeding coefficient of an individual relative to the subpopulation (F_{IS}); inbreeding coefficient of an individual relative to the total population (F_{IT}); genetic differentiation coefficient (F_{ST}); ± standard deviation (SD).

The mean number of alleles present per population ranged from 4.0 (SUK1) to 5.5 (BOR) with an overall mean of 4.618 (Table 3). Effective number of alleles (N_E) varied between 2.045 in SVO population and 2.869 in SHAT population, with an average of 2.494 per population. Shannon Information Index (I) ranged from 0.822 (SVO) to 1.133 (SHAT) population. The SVO population had the lowest values for allelic richness ($A_R = 3.505$) and the VAN2 population had the highest value ($A_R = 4.764$). The expected heterozygosity (H_E) ranged from 0.433 (SVO) to 0.580 (SHAT). The inbreeding coefficient (F_{IS}) indicated an excess of homozygotes in BOR, VAN2, SUK1, NAR and KOS populations.

The hierarchical analysis of molecular variance (AMOVA) showed high molecular variance within populations (95%) and a low molecular variance among populations (5%) (Table 4). The number of migrants per generation (Nm) was estimated at 4.437. The Mantel test of isolation by distance yielded a significant correlation between genetic differentiation and geographical distance among populations ($R^2 = 0.43$, $p = 0.01$). A higher level of differentiation was found between two most distant populations (SHAT and SVO) ($F_{ST} = 0.097$, $p = 0.001$) and the number of migrants per generation (Nm) was estimated at 2.331.

The highest ΔK value of the Structure analysis was observed at $K = 2$; however, there was a second, lower ΔK peak at $K = 5$, indicating the presence of sub-clusters within the dataset (Figure 2A,B). The populations were first grouped according to $K = 2$, and then according to $K = 5$ (Figure 2C). For $K = 2$, one cluster clearly corresponds to the SHAT population (orange color) and one to the 15 Siberian populations (blue color). Individuals from the BOR population contained two genetic groups that were highly admixed. The dataset contained sub-clusters that showed the most distant populations of SHAT, SVO, BAL, BOR, and ARO formed distinct clusters, while the remaining populations were highly admixed. For $K = 5$, SHAT, BOR, BAL and SVO populations were clearly separated from the remaining Scots pine populations (Figure 2D).

Two barriers with bootstrap support between 50% and 90% were detected, using Barrier software (Figure 2D). The first barrier with the highest bootstrap value (85%) separates the easternmost population (SVO) from the remaining ones. The second barrier, with bootstrap support between 52% and 79%, delineates SHAT, ARO, BOR, KOS and SAR populations. All the other barriers between the populations were weak and showed a non-significant separation with <50% bootstrap support.

UPGMA analysis confirmed the STRUCTURE results for $K = 2$, separating the SHAT population from all the others (Figure 3A). It also showed similarities to $K = 5$ in which SHAT, BAL, SVO, BOR differentiated from the rest of the populations. Congruencies were also apparent on the heat map where the highest F_{ST} values were detected for those populations that were separated in the UPGMA (Figure 3B).

The first three axes of the Principal Coordinates Analysis (PCoA) accounted for 76.93% of the accumulated variability (Figure 4). The clusters distinguished by the PCoA analysis were found to be similar to the results of the Structure analysis at $K = 5$. At the first two axes, which explained 63.12% of the total variation, the population from the European part of Russia (SHAT), as well as the easternmost (SVO) and southernmost (BAL) populations were separated. This pattern was confirmed by the second and third axes. The remaining populations from the central part of the Siberian sampling range in the Asian part of Russia were not separated.

The Mantel test detected significant correlation between F_{ST} and geographical distances between the populations ($R^2 = 0.429$, $p = 0.008$), indicating that genetic differentiation among the populations significantly increases with geographic distance (Figure 5). However, the 2-D kernel density estimate indicated patches instead of a continuous cloud, suggesting genetic discontinuity. The SHAT population, as the most geographically distant population, showed discontinuity in genetic differentiation and spatial segregation.

Table 3. Genetic diversity indices averaged across eight microsatellite loci for each Scots pine population.

ID		N_A	N_E	I	A_R	H_O	H_E	F_{IS}
BOR	Mean	5.500	2.612	1.092	4.486	0.522	0.542	0.004
		(±0.707)	(±0.352)	(±0.172)	(±0.621)	(±0.080)	(±0.083)	(±0.076)
VAN2	Mean	5.375	2.722	1.130	4.764	0.478	0.548	0.116
		(±0.653)	(±0.465)	(±0.162)	(±0.406)	(±0.074)	(±0.078)	(±0.052)
VAN1	Mean	4.500	2.426	0.962	4.031	0.511	0.495	−0.021
		(±0.655)	(±0.458)	(±0.166)	(±0.593)	(±0.097)	(±0.078)	(±0.068)
CHU	Mean	4.375	2.413	0.993	4.022	0.608	0.517	−0.144
		(±0.375)	(±0.371)	(±0.132)	(±0.339)	(±0.098)	(±0.069)	(±0.069)
ARO	Mean	4.750	2.558	1.053	4.172	0.554	0.538	−0.037
		(±0.453)	(±0.348)	(±0.157)	(±0.459)	(±0.081)	(±0.080)	(±0.026)
BORZ	Mean	4.750	2.435	0.971	4.104	0.500	0.483	−0.006
		(±0.675)	(±0.413)	(±0.191)	(±0.618)	(±0.112)	(±0.094)	(±0.077)
SUK2	Mean	4.875	2.418	1.060	4.596	0.569	0.529	−0.062
		(±0.295)	(±0.317)	(±0.127)	(±0.307)	(±0.084)	(±0.067)	(±0.065)
SUK1	Mean	4.000	2.361	0.906	4.000	0.455	0.467	0.013
		(±0.627)	(±0.474)	(±0.182)	(±0.670)	(±0.095)	(±0.088)	(±0.076)
ZEL	Mean	4.375	2.481	1.005	4.182	0.544	0.521	−0.042
		(±0.532)	(±0.397)	(±0.155)	(±0.534)	(±0.086)	(±0.075)	(±0.065)
SHA	Mean	4.500	2.553	1.030	4.053	0.587	0.532	−0.106
		(±0.567)	(±0.385)	(±0.163)	(±0.528)	(±0.087)	(±0.078)	(±0.048)
NAR	Mean	4.125	2.706	1.035	4.042	0.523	0.542	0.021
		(±0.611)	(±0.522)	(±0.166)	(±0.627)	(±0.079)	(±0.076)	(±0.059)
KOS	Mean	4.625	2.499	1.010	4.194	0.486	0.504	0.020
		(±0.680)	(±0.449)	(±0.185)	(±0.615)	(±0.085)	(±0.089)	(±0.052)
SAR	Mean	4.875	2.619	1.032	4.097	0.543	0.522	−0.013
		(±0.666)	(±0.415)	(±0.179)	(±0.513)	(±0.107)	(±0.092)	(±0.050)
SHAT	Mean	5.125	2.869	1.133	4.373	0.600	0.580	−0.038
		(±0.953)	(±0.454)	(±0.170)	(±0.677)	(±0.082)	(±0.077)	(±0.037)
YRB	Mean	4.375	2.514	1.020	4.183	0.549	0.521	−0.043
		(±0.498)	(±0.365)	(±0.168)	(±0.512)	(±0.094)	(±0.085)	(±0.035)
SVO	Mean	4.125	2.045	0.822	3.505	0.463	0.433	−0.031
		(±0.515)	(±0.300)	(±0.147)	(±0.435)	(±0.105)	(±0.082)	(±0.096)
BAL	Mean	4.250	2.172	0.888	3.746	0.534	0.473	−0.123
		(±0.412)	(±0.316)	(±0.142)	(±0.419)	(±0.085)	(±0.072)	(±0.055)
Total mean		4.618	2.494	1.008	4.150	0.531	0.514	−0.029
		(±0.141)	(±0.094)	(±0.038)	(±0.522)	(±0.021)	(±0.019)	(±0.015)

Note: number of alleles (N_A); number of effective alleles (N_E); Shannon Information Index (I); allelic richness (A_R); observed heterozygosity (Ho); expected heterozygosity (H_E); fixation index (F_{IS}); ± standard deviation.

Table 4. Hierarchical Analysis of Molecular Variance (AMOVA).

| Source	df	SS	MS	Est. Var.	%	p
Among populations	16	156.456	9.779	0.235	5	
Within populations	389	1624.588	4.176	4.176	95	0.001
Total	405	1781.044		4.412	100	

Note: df = degrees of freedom; SS = sum of squares; MS = mean of the squares; Est. Var. = estimated variance of components; % = percentage of total variance contributed by each component.
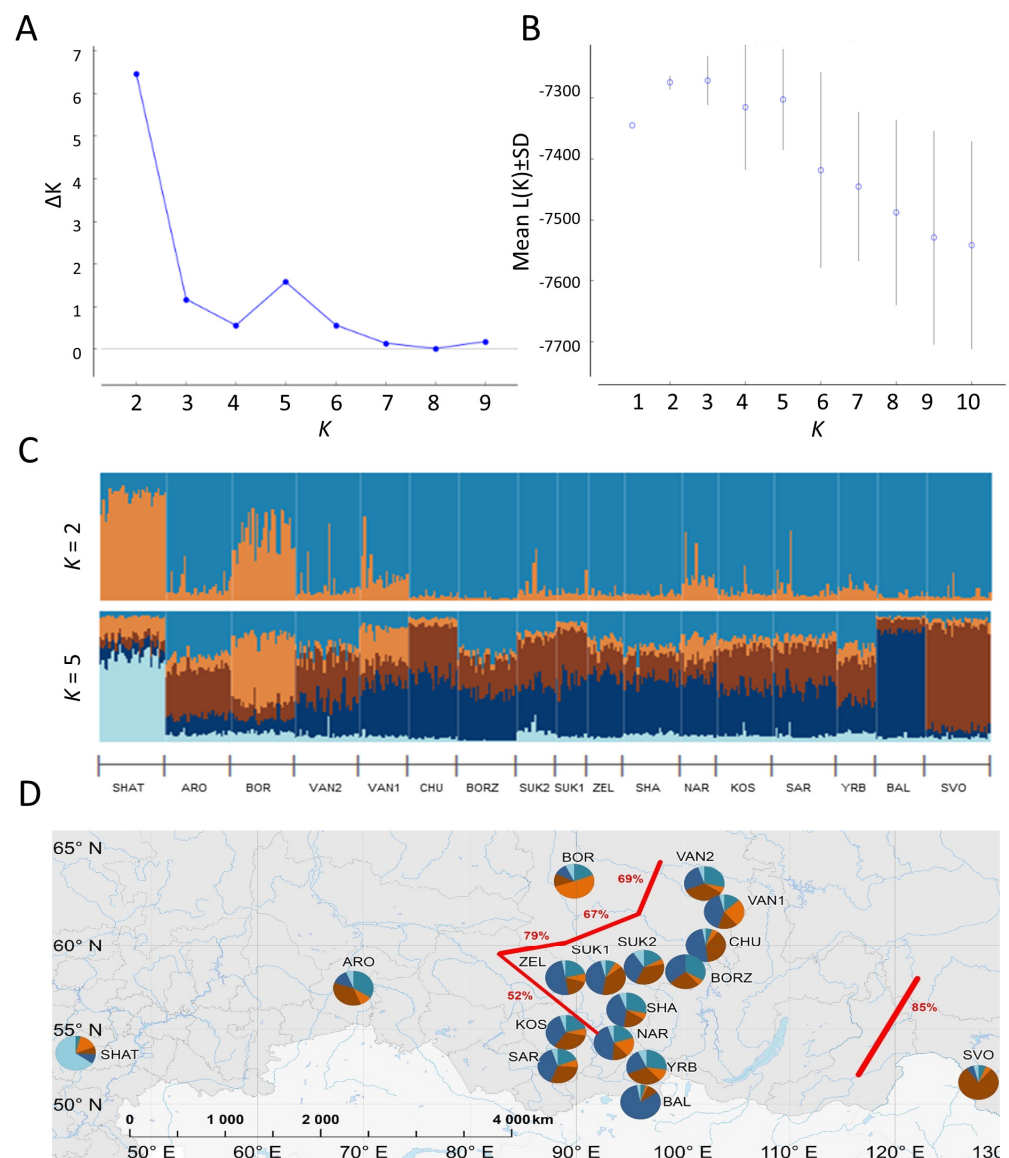
Figure 2. Results of population genetic structure analysis of 17 Scots pine populations (acronyms are as in Table 1). (**A**) Estimated population structure (*K* = 2 and *K* = 5). (**B**) Estimation of the best subpopulation numbers based on Δ*K* and Mean L(*K*) ±SD values. (**C**) Barplot of genetic structure for *K* = 2 and *K* = 5. Light blue, blue, dark blue, orange and brown colors correspond to different clusters. (**D**) Geographic distribution of the five genetic clusters and the two barriers revealed by Barrier analysis (barriers are shown in red bold lines with bootstrap value).
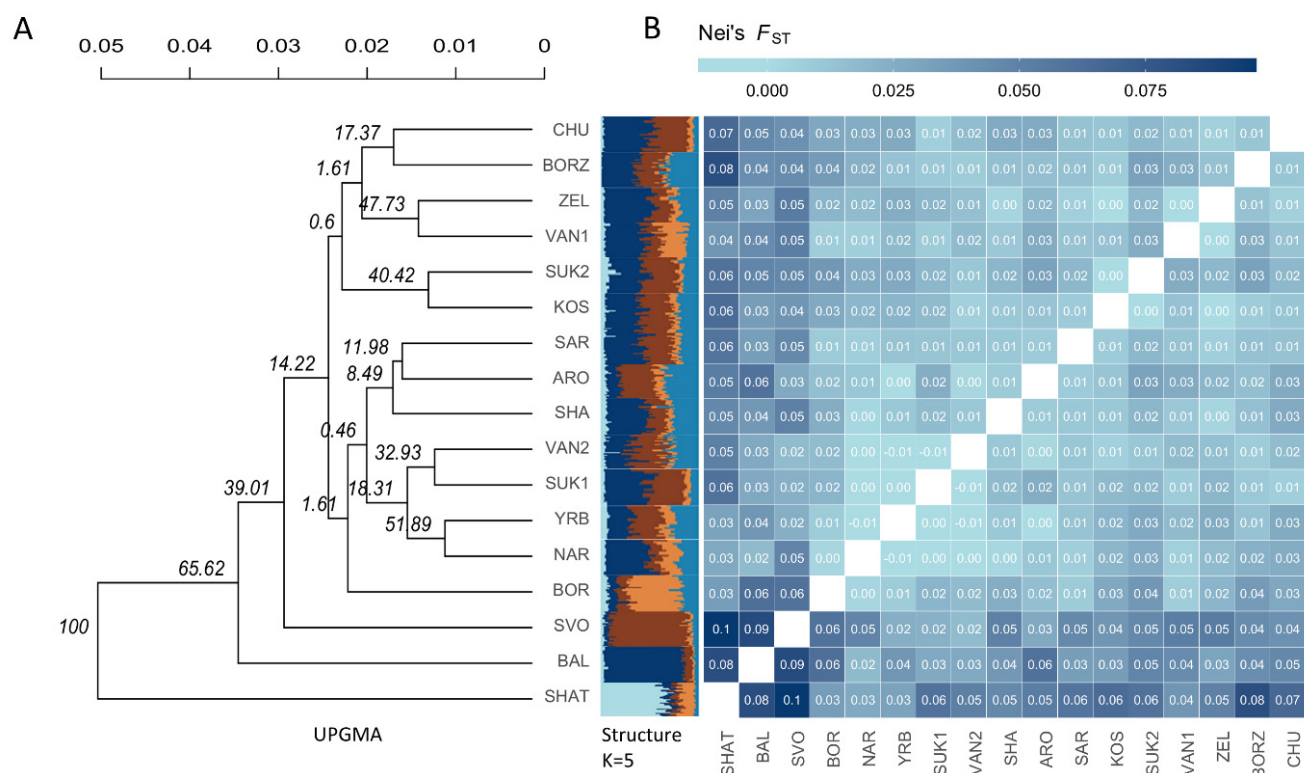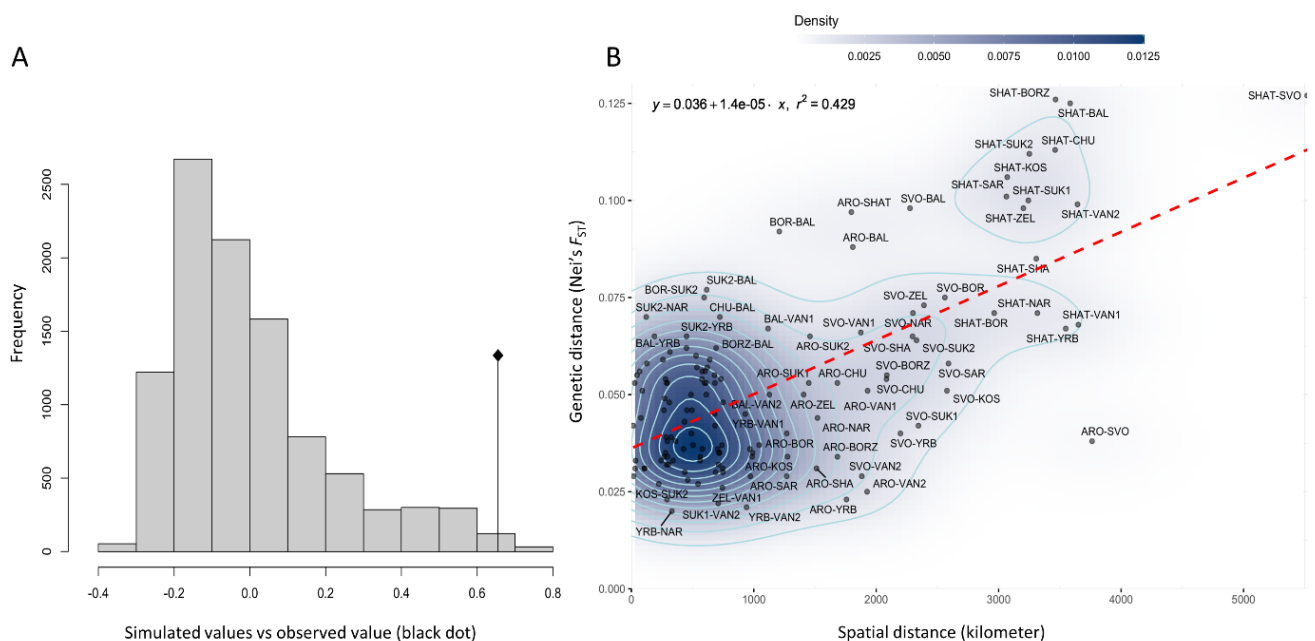
Figure 3. Bootstrapped UPGMA dendrogram (**A**), cluster assignments at *K* = 5, and the heatmap of pairwise of F_{ST} between the 17 Scots pine populations investigated (**B**). Acronyms are as in Table 1.
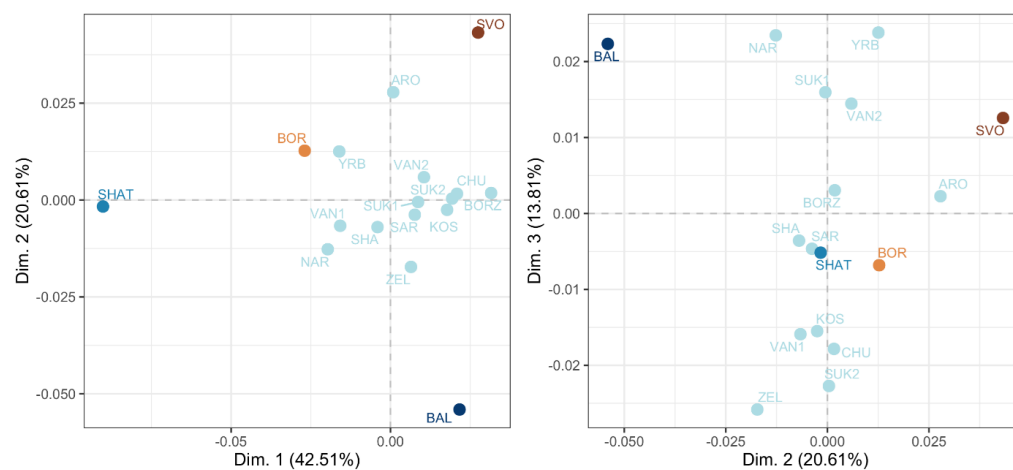
Figure 4. Two-dimensional plot of the three main principal components and their part of the total variance in % using Principal Coordinates Analysis (PCoA). Acronyms are as in Table 1.

Figure 5. Isolation-by-distance analyses (IBD), including the output of a Mantel test (**A**), were performed to indicate the empirical genetic distance vs. geographic distance correlation (black dot), which was plotted over a histogram of Monte Carlo simulations in the absence of spatial structure. A kernel density estimate of pairwise correlations revealing multiple clusters (**B**). Acronyms are as in Table 1.

4. Discussion

The genetic diversity and population structure of Scots pine populations from different parts of Russia were assessed based on the polymorphism of eight nuclear microsatellite markers. Our results showed that Scots pine in Russia harbor high genetic diversity (H_E = 0.514), however our estimates are lower than those obtained in other regional studies for Italian (H_E = 0.81) [43], Romanian and Hungarian (H_E = 0.55) [20], Turkish (H_E = 0.772) [44], Georgian (H_E = 0.577) [45] and Lithuanian populations (H_E = 0.59) [21]. The highest genetic diversity was observed in SHAT population (H_E = 0.580) located in the European part of Russia while the lowest genetic diversity was detected in SVO population (H_E = 0.433) from the Russian Far East. Previous studies conducted on Scots pine based on mitochondrial and chloroplast DNA markers also indicate a decrease in genetic diversity in Asian Scots pine populations compared to European ones [46,47]. Semerikov et al. [46] suggested that this fact might be the result of the loss of genetic variability due to multiple so-called "bottleneck" processes during the eastward distribution of the species. Besides heterozygosity, which is important for the long-term adaptation of populations to novel environmental conditions, allelic richness is also an important estimator of genetic diversity in populations [48–51]. Our results showed that the highest value of allelic richness was observed in one of the northernmost populations (VAN2) (A_R = 4.764) while the lowest value was detected in the easternmost one (SVO) (A_R = 3.505).

In most populations, there were no signs of inbreeding, F_{IS} values were overall negative (−0.029), except in the case of BOR, VAN2, SUK1, NAR and KOS populations, for which F_{IS} values were found to be positive (0.004, 0.116, 0.013, 0.021 and 0.02 respectively). The highest F_{IS} value (0.116) was detected in the VAN2 population. Heterozygote deficiency is a well-known characteristic of conifers and it can be the result of selection against heterozygotes, assortative mating, or the presence of null alleles [52].

AMOVA showed that only 5% of the total genetic variation occurred among populations, even if they are distributed at distances of up to 9656 km. The low genetic differentiation among Scots pine populations revealed in our study is in agreement with previous reports on Scots pine using nuclear simple sequence repeat (nSSR) markers [20,21,53]. Prob-

ably, free gene flow over large areas without any significant geographic barriers may have a homogenizing effect on the gene pool of Scots pine populations. Due to wind pollination and high outcrossing rates, pine species often exhibit high gene flow among populations and, as a result, a low level of genetic differentiation [54]. The high rate of gene flow among the studied populations, was confirmed by the high number of migrants per generation (Nm = 4.437). As expected, a significantly higher level of genetic differentiation was found only between the two most distant populations (SHAT and SVO) (F_{ST} = 0.097, p = 0.001).

Using the software STRUCTURE, we found support for two and five genetic clusters. According to K = 2, all Siberian populations except for the BOR population were grouped into a single cluster while one European population (SHAT) was distinct. The presence of two genetic clusters in the studied Scots pine populations was also confirmed by UPGMA analysis. The clustering of studied Siberian populations into one group might indicate their common ancestry. However, based on the STRUCTURE results for K = 5, the studied Scots pine populations were separated into five distinct genetic groups, as follows: the European part of Russia (SHAT), the south of Middle Siberia (BAL), the northwest of Middle Siberia (BOR), West and Middle Siberia (13 populations) and the Russian Far East (SVO). Two barriers against gene flow with more than 50% bootstrap support were detected. The barrier showing the highest bootstrap values (85%) separate the easternmost population (SVO) from the remaining ones. Despite the identified barriers, they cannot be considered as impervious, because neither the high level of gene flow nor the low inbreeding values support this.

Our results indicate that despite the balanced diversity and substantial gene flow among Scots pine populations in Russia, there is evidence of genetic differentiation in the nuclear genomes. The main factor contributing to this is most likely the spatial segregation of populations due to the large geographic distances, which was confirmed by our significant Isolation-By-Distance hypothesis, as well as the inflated F_{ST} values of SHAT, BOR, BAL, SVO. This might be a possible mechanism shaping the present distribution of genetic diversity. It should also be considered that distant populations could have originated from different glacial refugia and, thus, correspond to different genetic lineages. For example, the westernmost population could have originated from the Balkans or the Carpathians [19,55], the Caucasus region [45], or even from the Ural mountains [47,56], which are all known refugial regions and could have been the source of northward colonization. However, this later hypothesis and the geographic regions of the segregated populations require further detailed investigations.

5. Conclusions

The investigation of the genetic background of natural Siberian Scots pine populations is crucial for the sustainable management of forest genetic resources in light of forest biodiversity and climate change adaptation in Russia. Our study showed that the Siberian populations of Scots pine harbor a large amount of genetic diversity, despite having a low level of genetic differentiation among its vast distribution area in the Eurasian part of its natural range. Genetic diversity found to decrease from west to east, and the easternmost population has shown the lowest level of genetic diversity including allelic richness. For these reasons, this population should be in the focus of conservation efforts and further studies on population dynamics should be facilitated. Patterns of differentiation indicated separate genetic clusters of Scots pine in the European part of Russia, the south of Middle Siberia, the northwest of Middle Siberia, West and Middle Siberia and the Russian Far East, respectively. In these regions, besides the spatial segregation that indicated restrictions to gene flow, past demographic events also potentially affected neutral genetic variation. Our findings can be used in long-term monitoring of the state of Scots pine genetic resources in Russia and can provide guidance for future studies of population genetics. In addition, further studies encompassing more populations from these aforementioned regions could reveal the primary source of genetic variation and can provide insights into the adaptive genetic variation of Siberian Scots pine populations.

References

1. Spielman, D.; Brook, B.W.; Frankham, R. Most species are not driven to extinction before genetic factors impact them. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15261–15264. [CrossRef]
2. Radu, R.G.; Curtu, L.A.; Spârchez, G.; Șofletea, N. Genetic diversity of Norway spruce [*Picea abies* (L.) Karst.] in Romanian Carpathians. *Ann. For. Res.* **2014**, *57*, 19–29. [CrossRef]
3. Hoban, S.; Bruford, M.; D'Urban Jackson, J.; Lopes-Fernandes, M.; Heuertz, M.; Hohenlohe, P.A.; Paz-Vinas, I.; Sjögren-Gulve, P.; Segelbacher, G.; Vernesi, C.; et al. Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biol. Conserv.* **2020**, *248*, 108654. [CrossRef]
4. Graudal, L.; Loo, J.; Fady, B.; Vendramin, G.; Aravanopoulos, F.A.; Baldinelli, G.; Bennadji, Z.; Ramamonjisoa, L.; Changtragoon, S.; Kjær, E.D. Indicators of the genetic diversity of trees—State, Pressure, benefit and response. In *State of the World's Forest Genetic Resources—Thematic Study*; FAO: Rome, Italy, 2020; p. 92.
5. Șofletea, N.; Mihai, G.; Ciocîrlan, E.; Curtu, A.L. Genetic Diversity and Spatial Genetic Structure in Isolated Scots Pine (*Pinus sylvestris* L.) Populations Native to Eastern and Southern Carpathians. *Forests* **2020**, *11*, 1047. [CrossRef]
6. Sheller, M.; Ciocîrlan, E.; Mikhaylov, P.; Kulakov, S.; Kulakova, N.; Ibe, A.; Sukhikh, T.; Curtu, A.L. Chloroplast DNA Diversity in Populations of *P. sylvestris* L. from Middle Siberia and the Romanian Carpathians. *Forests* **2021**, *12*, 1757. [CrossRef]
7. Vasilyeva, Y.; Chertov, N.; Nechaeva, Y.; Sboeva, Y.; Pystogova, N.; Boronnikova, S.; Kalendar, R. Genetic Structure, Differentiation and Originality of *Pinus sylvestris* L. Populations in the East of the East European Plain. *Forests* **2021**, *12*, 999. [CrossRef]
8. Przybylski, P.; Tereba, A.; Meger, J.; Szyp-Borowska, I.; Tyburski, Ł. Conservation of Genetic Diversity of Scots Pine (*Pinus sylvestris* L.) in a Central European National Park Based on cpDNA Studies. *Diversity* **2022**, *14*, 93. [CrossRef]
9. Ekart, A.K.; Larionova, A.Y.; Kravchenko, A.N.; Tikhonova, I.V.; Zatsepina, K.G.; Tarakanov, V.V.; Zhamyansuren, S. Genetic diversity and differentiation of populations of Scots pine in Southern Siberia and Mongolia. *Contemp. Probl. Ecol.* **2014**, *7*, 52–59. [CrossRef]
10. Tóth, E.G.; Vendramin, G.G.; Bagnoli, F.; Cseke, K.; Höhn, M. High genetic diversity and distinct origin of recently fragmented Scots pine (*Pinus sylvestris* L.) populations along the Carpathians and the Pannonian basin. *Tree Genet. Genomes* **2017**, *13*, 47. [CrossRef]
11. Ivanova, G.A.; Kukavskaya, E.A.; Zhila, S.V. Fire impact on carbon balance parameters and ecosystem components of light-coniferous forests of central Siberia. In *GEO-Siberia-2010: International Exhibition and Scientific Congress*; SSGA: Novosibirsk, Russia, 2010.
12. Pimenov, A.V. Biodiversity of the Scots pine (*Pinus sylvestris* L.) in contrasting ecotopes of the south of Siberia. Ph.D. Dissertation, Sukachev Forest Institute of the Siberian Branch of the Russian Academy of Sciences, Krasnoyarsk, Russia, 2015; p. 406.
13. Markatyuk, A.A.; Runova, E.M.; Gavrilin, I.I.; Vedernikov, I.B. Current state of East Siberian boreal forests in the aspect of natural *Pinus sylvestris* L. regeneration Systems. *Methods. Technol.* **2013**, *1*, 163–169.
14. Forest Online. Available online: https://www.lesonline.ru (accessed on 14 December 2022).
15. Parliamentary Newspaper. Available online: https://www.pnp.ru/ (accessed on 14 December 2022).
16. Ilyinov, A.A.; Raevsky, B.V. The current state of *Pinus sylvestris* L. gene pool in Karelia. *Sib. J. For. Sci.* **2016**, *5*, 45–54.
17. Torbik, D.N.; Bedrickaya, T.V.; Vlasova, M.M.; Sinelnikov, I.G. Genetic diversity of natural populations of *Pinus sylvestris* L. In Proceedings of the Federal Budget Institution «Northern Research Institute of Forestry»; Demidova, N.A., Ed.; Northern Research Institute of Forestry: Arkhangelsk, Russia, 2019; pp. 91–99.
18. Floran, V.; Sestraş, R.E.; Gil, M.R.G. Organelle genetic diversity and phylogeography of Scots pine (*Pinus sylvestris* L.). *Not. Bot. Horti Agrobot. Cluj-Napoca* **2011**, *39*, 317–322. [CrossRef]
19. Tóth, E.G.; Köbölkuti, Z.A.; Pedryc, A.; Höhn, M. Evolutionary history and phylogeography of Scots pine (*Pinus sylvestris* L.) in Europe based on molecular markers. *J. For. Res.* **2017**, *28*, 637–651. [CrossRef]
20. Bernhardsson, C.; Floran, V.; Ganea, S.L.; García-Gil, M.R. Present genetic structure is congruent with the common origin of distant Scots pine populations in its Romanian distribution. *For. Ecol. Manag.* **2016**, *361*, 131–143. [CrossRef]
21. Kavaliauskas, D.; Danusevičius, D.; Baliuckas, V. New Insight into Genetic Structure and Diversity of Scots Pine (*Pinus sylvestris* L.) Populations in Lithuania Based on Nuclear, Chloroplast and Mitochondrial DNA Markers. *Forests* **2022**, *13*, 1179. [CrossRef]

22. Doyle, J.J.; Doyle, J.L. Isolation of plant DNA from fresh tissue. *Focus* **1990**, *12*, 13–15.
23. Sebastiani, F.; Pinzauti., F.; Kujala, S.T.; González-Martínez, S.C.; Vendramin, G.G. Novel polymorphic nuclear microsatellite markers for *Pinus sylvestris* L. *Conserv. Genet. Resour.* **2012**, *4*, 231–234. [CrossRef]
24. Elsik, C.G.; Minihan, V.T.; Hall, S.E.; Scarpa, A.M.; Williams, C.G. Low-copy microsatellite markers for *Pinus taeda* L. *Genome* **2000**, *43*, 550–555. [CrossRef] [PubMed]
25. Zhou, Y.; Bui, T.; Auckland, L.D.; Williams, C.G. Undermethylated DNA as a source of microsatellites from a conifer genome. *Genome* **2002**, *45*, 91–99. [CrossRef]
26. Fang, P.; Niu, S.; Yuan, H.; Li, Z.; Zhang, Y.; Yuan, L.; Li, W. Development and characterization of 25 454 EST-SSR markers in *Pinus sylvestris* var. mongolica (*Pinaceae*). *Appl. Plant Sci.* **2014**, *2*, 455. [CrossRef]
27. Van Oosterhout, C.; Weetman, D.; Hutchinson, W.F. Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Mol. Ecol. Notes* **2006**, *6*, 255–256. [CrossRef]
28. Peakall, R.; Smouse, P.E. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **2006**, *6*, 288–295. [CrossRef]
29. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013.
30. Goudet, J. Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **2005**, *5*, 184–186. [CrossRef]
31. Nei, M. Genetic distance between populations. *Am. Nat.* **1972**, *106*, 283–292. [CrossRef]
32. Thioulouse, J.; Dray, S.; Dufour, A.B.; Siberchicot, A.; Jombart, T.; Pavoine, S. Description of Species Structures. In *Multivariate Analysis of Ecological Data with ade4*; Springer: New York, NY, USA, 2018; pp. 97–117.
33. Nei, M. *Molecular Evolutionary Genetics*; Columbia University Press: New York, NY, USA, 1987; pp. 1–512.
34. Pritchard, J.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959. [CrossRef]
35. Earl, D.A.; Vonholdt, B.M. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **2012**, *4*, 359–361. [CrossRef]
36. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef]
37. Jakobsson, M.; Rosenberg, N.A. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **2007**, *23*, 1801–1806. [CrossRef]
38. Manni, F.; Guerard, E.; Heyer, E. Geographic Patterns of (Genetic, Morphologic, Linguistic) Variation: How Barriers Can Be Detected by Using Monmonier's Algorithm. *Hum. Biol.* **2004**, *76*, 173–190. [CrossRef]
39. Monmonier, M.S. Maximum-Difference Barriers: An Alternative Numerical Regionalization Method. *Geogr. Anal.* **2010**, *5*, 245–261. [CrossRef]
40. Dieringer, D.; Schlötterer, C. MICROSATELLITE ANALYSER (MSA): A platform independent analysis tool for large microsatellite data sets. *Mol. Ecol. Notes* **2003**, *3*, 167–169. [CrossRef]
41. Jombart, T.; Ahmed, I. Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* **2011**, *27*, 3070–3071. [CrossRef] [PubMed]
42. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S-PLUS*; Springer Science & Business Media: New York, NY, USA, 2013; Volume 5, pp. 1–6.
43. Scalfi, M.; Piotti, A.; Rossi, M.; Piovani, P. Genetic variability of Italian southern Scots pine (*Pinus sylvestris* L.) populations: The rear edge of the range. *Eur. J. Forest Res.* **2009**, *128*, 377–386. [CrossRef]
44. Bilgen, B.B.; Nuray, K. Genetic diversity among *Pinus sylvestris* L. populations and its implications for genetic conservation: Comparison of nuclear and chloroplast microsatellite markers. *Fresenius Environ. Bull.* **2017**, *26*, 6873–6881.
45. Dering, M.; Baranowska, M.; Beridze, B.; Chybicki, I.J.; Danelia, I.; Iszkuło, G.; Kvartskhava, G.; Kosiński, P.; Rączka, G.; Thomas, P.A.; et al. The evolutionary heritage and ecological uniqueness of Scots pine in the Caucasus ecoregion is at risk of climate changes. *Sci. Rep.* **2021**, *11*, 22845. [CrossRef]
46. Semerikov, V.L.; Semerikova, S.A.; Dymshakova, O.S.; Zatsepina, K.G.; Tarakanov, V.V.; Tikhonova, I.V.; Ekart, A.K.; Vidyakin, A.I.; Jamiyansuren, S.; Rogovtsev, R.V.; et al. Microsatellite loci polymorphism of chloroplast DNA of Scots pine (*Pinus sylvestris* L.) in Asia and Eastern Europe. *Russ. J. Genet.* **2014**, *50*, 577–585. [CrossRef]
47. Semerikov, V.; Semerikova, S.; Putintseva, J.; Tarakanov, V.; Tikhonova, I.; Vidyakin, A.; Oreshkova, N.; Krutovsky, K. Colonization history of Scots pine in Eastern Europe and North Asia based on mitochondrial DNA variation. *Tree Genet. Genomes* **2018**, *14*, 8. [CrossRef]
48. Reed, D.H.; Frankham, R. Correlation between fitness and genetic diversity. *Conserv. Biol.* **2003**, *17*, 230–237. [CrossRef]
49. Bashalkhanov, S.; Pandey, M.; Rajora, O.P. A simple method for estimating genetic diversity in large populations from finite sample sizes. *BMC Genet.* **2009**, *10*, 84. [CrossRef]
50. Aravanopoulos, F. Genetic monitoring in natural perennial plant populations. *Botany* **2011**, *89*, 75–81. [CrossRef]
51. Porth, I.; El-Kassaby, Y.A. Assessment of the genetic diversity in forest tree populations using molecular markers. *Diversity* **2014**, *6*, 283–295. [CrossRef]
52. Gil MR, G.; Floran, V.; Östlund, L.; Gull, B.A. Genetic diversity and inbreeding in natural and managed populations of Scots pine. *Tree Genet. Genomes* **2015**, *11*, 28.

53.　Hebda, A.; Wójkiewicz, B.; Wachowiak, W. Genetic characteristics of Scots pine in Poland and reference populations based on nuclear and chloroplast microsatellite markers. *Silva Fenn.* **2017**, *2*, 17. [CrossRef]

54.　Petit, R.J.; Duminil, J.; Fineschi, S.; Hampe, A.; Salvini, D.; Vendramin, G.G. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Mol. Ecol.* **2005**, *14*, 689–701. [CrossRef] [PubMed]

55.　Tóth, E.G.; Bede-Fazekas, Á.; Vendramin, G.G.; Bagnoli, F.; Höhn, M. Mid-Pleistocene and Holocene demographic fluctuation of Scots pine (*Pinus sylvestris* L.) in the Carpathian Mountains and the Pannonian Basin: Signs of historical expansions and contractions. *Quat. Int.* **2019**, *504*, 202–213. [CrossRef]

56.　Chertov, N.; Nechaeva, Y.; Zhulanov, A.; Pystogova, N.; Danilova, M.; Boronnikova, S.; Kalendar, R. Genetic structure of *Pinus* populations in the Urals. *Forests* **2022**, *13*, 1278. [CrossRef]