Tree Genetics & Genomes (2023) 19:16
https://doi.org/10.1007/s11295-023-01592-z

ORIGINAL ARTICLE

Late Quaternary history of Siberian stone pine as revealed by genetic and paleoecological data

Denis N. Shuvaev[1] · Vladimir L. Semerikov[1] · Galina V. Kuznetsova[2] · Yuliya A. Putintseva[3]

Received: 4 September 2022 / Revised: 18 February 2023 / Accepted: 23 February 2023 / Published online: 7 March 2023

Abstract

Current climatic trends raise concerns about the fate of boreal forests and associated communities. To understand forest vegetation dynamics in the face of future changes, it is necessary to know how trees have responded to past climatic fluctuations. We investigated genetic data and a past distribution based on paleorecords and species distribution modeling data to determine refugia and migration routes of Siberian stone pine (*Pinus sibirica*) from the late Pleistocene to the present. Thus, we gained a deep insight into the past of Siberian stone pine. We revealed four genetic clusters of Siberian stone pine located in the following mountain systems of Siberia: the West Sayan and the Altai (WSA), the East Sayan (ES), the Kuznetsk Alatau (KA), and the Urals (U). Paleorecords and species distribution modeling indicated preservation of Siberian stone pine in these mountains and near big lakes (Lake Baikal and Lake Teletskoe) during the Last Glacial Maximum (LGM). Nonetheless, processes of migration from these refugia were heterogeneous: WSA has never expanded after the LGM beyond its current range; ES occupied the eastern geographic range of the species; KA and U had common history before the LGM and got isolated during LGM. Moreover, refugia and migration routes of Siberian forest species (*Larix sibirica* and *Abies sibirica*) were different from those of *P. sibirica* despite similar ecological preferences. *Pinus sibirica* occupied the West Siberian Plain from the Urals and Kuznetsk Alatau, whereas *A. sibirica* and *L. sibirica* expanded from Lake Baikal and the Sayan Mountains, respectively. Thus, our results substantially complement existing notions of South Siberia Mountains, Lake Baikal area, and Urals as glacial refugia for Siberian stone pine.

Keywords *Pinus sibirica* · Genetic markers · Paleorecord · Species distribution modeling · Late Quaternary · Refugia

Introduction

Siberian stone pine (*Pinus sibirica* Du Tour) is the main forest species in large parts of Siberia. The species' geographic range extends from the western Urals through the entire West Siberian Plain to Transbaikalia and the Aldan Highlands (Yakutia) (Fig. 1). *Pinus sibirica* has an important effect on environmental conditions of taiga forests because its seeds represent a major food resource for taiga mammals and birds. Nevertheless, climate changes in recent decades have led to shifts of species ranges and population sizes as well as genetic changes (Graham et al. 2019). Furthermore, current climate trends are a cause for serious concern about the fate of *P. sibirica*. For instance, climate-induced mortality of Siberian stone pine has been reported in some studies on the southern Siberian Mountains and on the region of Lake Baikal (Kharuk et al. 2017a, b). Therefore, we need a basis for principles for the placement of future *P. sibirica* nature reserves. Data about past favorable locations could give valuable information for the planning of the nature reserves (Rossetto and Kooyman 2021). At present, there is still no accurate information on locations of refugia of Siberian stone pine.

Communicated by J. Beaulieu

✉ Denis N. Shuvaev
denis.shuvaev@gmail.com

1 Institute of Plant and Animal Ecology, Ural Branch of the Russian Academy of Sciences, 8 Marta Str. 202, Ekaterinburg 620144, Russia

2 Sukachev Forest Institute of FRC KSC SB RAS, Akademgorodok 50/28, Krasnoyarsk 660036, Russia

3 Department of Biophysics, Institute of Fundamental Biology and Biotechnology, Siberian Federal University, 79 Svobodny Ave., Krasnoyarsk 660041, Russia

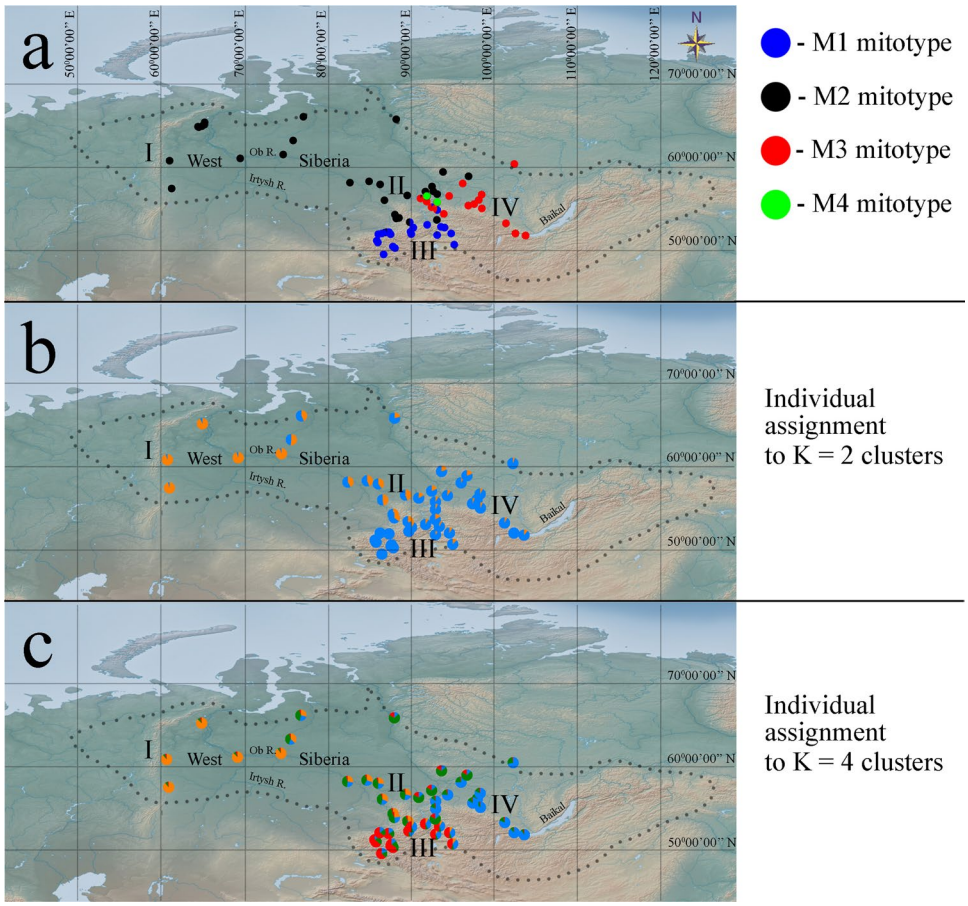

Fig. 1 (**a**) The mitotype distribution based on mtDNA data of *Pinus sibirica*. (b–c) The genetic structure inferred from STRUCTURE applied to nuclear microsatellite data. Population-level assignment to K = 2 clusters (**b**) and K = 4 clusters (**c**) for all 48 genotyped populations is visualized as pie charts. I: the Urals, II: the Kuznetsk Alatau, III: the Altai and the West Sayan, IV: the East Sayan. The dotted line shows the *P. sibirica* range

The late Quaternary was characterized by climate oscillations that induced the largest-scale transformations of flora and fauna complexes of Siberia. Temperate-zone and boreal trees were profoundly influenced by Pleistocene glacial cycles (Hewitt 2000). In these adverse periods, *P. sibirica* could survive in refugia where reduced population sizes and isolation affected its genetic structure. Afterwards, *P. sibirica* spread from these climate refugia carrying their genetic footprints. The harshest cold-dry climate fell on the Last Glacial Maximum (LGM) (Tarasov et al. 2000; Velichko et al. 2011) and may have caused a dramatic contraction of the *P. sibirica* range to limited areas in the south of Siberia and of the Urals. On the basis of the Holocene pollen data, Blyakharchuk (2010) hypothesized that the post-glacial expansion of Siberian stone pine proceeded from the mountains of South Siberia and the southern Urals. Nonetheless, it is still unclear how many refugia existed during the LGM, where they were, and what the possible migration routes were.

On the other hand, the current genetic data are also insufficient to draw any reliable phylogeographic conclusions. Previous allozyme studies on *P. sibirica* have shown several genetically distinct population groups: Altai-West Sayan, East Sayan, and northern populations of West Siberia (Goncharenko et al. 1991; Krutovsky et al. 1988, 1989; Politov et al. 1992). Some narrow-range studies have been performed on nuclear microsatellite loci (e.g., Oreshkova et al. 2014). In addition, genetic differences were found between Ural and Kuznetsk Alatau populations (Shuvaev and Ibe 2021). Nevertheless, biparentally inherited markers alone are not usually enough for tracking long-term migration history because these markers transmitted by seed and pollen are subject to strong gene flow. By contrast, mitochondrial genetic markers (mtDNA) in Pinaceae are transmitted by seeds, are subject to much more limited gene flow, and are associated with twofold smaller effective population size. Therefore, these markers are the most suitable for revealing a migration history and deeper interpopulation differences.

In general, the biogeographic history of Siberian conifers has shown differences in refugia locations. For example, genetic data have confirmed that Siberian larch (*Larix sibirica* Ledeb.) and common juniper (*Juniperus communis* L.) could have survived during the LGM at northern latitudes, but their northern secondary refugia

contributed differently to the recolonization (Hantemirova et al. 2017; Semerikov et al. 2013). Nonetheless, the last paleogenomic investigation of Siberian larch contradicts the hypothesis of the northern refugia and implies a Holocene expansion of larch to the north (Schulte et al. 2022). Scots pine (*Pinus sylvestris* L.) apparently colonized Siberia from European refugia undergoing a dramatic decline of gene diversity (Semerikov et al. 2018). Besides, two closely related boreal spruces—Siberian spruce (*Picea obovata* Ledeb.) and Norway spruce (*Picea abies* [L.] H. Karst.)—have persisted in their own refugia: the former has survived in southern Siberia, and the latter has inhabited the East European Plain (Tollefsrud et al. 2015). Siberian fir (*Abies sibirica* Ledeb.) has also remained in glacial refugia of several geographic regions, but the main recolonization wave went from the East Sayan and Baikal area (Semerikov et al. 2019). Siberian stone pine, just as Siberian fir, is a major forest-forming species of the dark taiga. Furthermore, these species have similar environmental requirements (Krylov 1961). It would be interesting to find out, based on these premises, whether refugia and recolonization patterns were similar or different between these species.

We believe that the complicated past of Siberian stone pine can be better revealed using an integrated approach, which can give a deeper insight into the history of *P. sibirica*. In general terms, it may include paleontological and genetic data and species distribution modeling (SDM) methods to make joint inferences about ancient history and to reveal species biogeographic patterns (Gavin et al. 2014; Hao et al. 2018). As a result, disadvantages of one method are compensated by advantages of the others. In recent years, this approach has become more feasible due to the development of new genetic markers and models of past climate for SDM methods. In addition, now we have enough information about Siberian paleogeography and paleorecords.

This study was aimed at searching for main LGM refugia and reconstructing recolonization patterns for Siberian stone pine from the late Pleistocene to the present. Moreover, it would be interesting to determine whether Blyakharchuk's hypothesis is correct and *P. sibirica* could survive in the mountains of South Siberia and the Urals during the harsh LGM climate. For this purpose, we analyzed widely sampled populations of *P. sibirica* by means of new mtDNA markers that were developed based on de novo next-generation sequencing of the mitochondrial genome. We employed nuclear microsatellite markers, which are well suited for assaying genetic variation and for establishing genetic relationships among populations (Selkoe and Toonen 2006). Finally, we compared the genetic data with paleorecords and environmental-niche modeling results to create an additive model of Siberian stone pine history.

Materials and methods

Plant material

The geographic distribution of *P. sibirica* has a wide ecological amplitude (Fig. 1). Its northern habitat reaches the border of the forest-tundra ecotone, and at the southern edge, *P. sibirica* dominates in mountain forests of South Siberia, thus representing the main component of dark-coniferous forests of the Altai–Sayan Mountains (Krylov 1961). Although Siberian stone pine is a typical mountain species, it also occurs in flatland forests of West Siberia along river valleys at sufficient soil and air humidity (Krylov et al. 1983).

All 67 *P. sibirica* populations (1821 trees) were sampled in natural forest stands from the Urals to Lake Baikal (Fig. 1). We did not collect material from the eastern range of *P. sibirica* (from the east coast of Lake Baikal to the Aldan Highlands) because it contains a zone of introgressive hybridization with closely related Siberian dwarf pine (*Pinus pumila* [Pall.] Regel) (Belokon et al. 2022).

Genomic DNA was extracted from fresh needles by a standard CTAB-method (Devey et al. 1996). Samples from 66 populations (1470 trees) (including populations from Shuvaev and Ibe (2021)) representing 4–24 individuals per population (Table S1.1) were used for mtDNA investigation. Forty-two of the newly sampled populations (1130 trees) were additionally studied using a set of eight nuclear microsatellites (Table S4.7). This robust marker set was previously selected from a full marker set of Belokon et al. (2016) and was analyzed as described by Shuvaev and Ibe (2021). Furthermore, we added microsatellite data from 10 populations published in Shuvaev and Ibe (2021) to this study. In total, we analyzed data on 52 populations with eight microsatellites. The 15 remaining populations were at short geographic distances from the populations under study and therefore not genotyped at the microsatellite loci.

Development of mtDNA markers and genotyping

We sequenced two mitochondrial genomes of *P. sibirica* trees growing in the Urals and the West Sayan and discovered two single-nucleotide polymorphisms (SNPs) (Tables S2.2 and S2.3). SNP genotyping was carried out with the single-strain conformation polymorphism (SSCP) method (Fujita and Silver 1994) as described for *A. sibirica* by Semerikov et al. (2019) (Fig. S2.1). For additional verification, we analyzed mtDNA fragments by sequencing of at least one specimen for each detected mitotype from each studied population.

Data analyses

Analysis of genetic data

A cluster analysis of the mtDNA data was carried out in SAMOVA v.1.0 (Dupanloup et al. 2002). We chose the best value of K clusters (from 1 to 5) based on the maximal F_{CT} level.

Genotyping errors and null alleles of nuclear microsatellites were checked using MICRO-CHECKER 2.2.3 (Oosterhout et al. 2004) and FREENA (Chapuis and Estoup 2007), respectively. Hardy–Weinberg equilibrium and linkage disequilibrium tests were performed in GENEPOP ON THE WEB with default settings (Raymond and Rousset 1995; Rousset 2008).

The standard genetic statistics (e.g., H_O and H_E) were calculated in GENALEX 6.5 (Peakall and Smouse 2006). The microsatellite clustering was based on a matrix of genetic distances (D_A) (Nei et al. 1983) and principal coordinate analysis (Kendall and Stuart 1976) using GENALEX 6.5. We performed a hierarchical analysis of molecular variance (AMOVA) and the pairwise exact test of genetic differentiation (Raymond and Rousset 1995) among AMOVA population groups by means of ARLEQUIN 3.5.2 with 10,000 permutations (Excoffier and Lischer 2010). Analysis of population genetic structure was carried out by the Bayesian clustering method in STRUCTURE 2.3.4 (Pritchard et al. 2000). Parameters of the simulation experiment included Admixture and Locprior models (Hubisz et al. 2009). The number of Monte Carlo Markov Chain simulations was 500,000 with a burn-in of 100,000 simulations. Ninety iterations were carried out for each K from 1 to 12. CLUMPACK and DISTRUCT applications (Kopelman et al. 2015) were used for alignment of cluster assignments across replicate analyses and for plotting of a K graph by the ΔK method (Evanno et al. 2005) and the maximum likelihood for increasing values of K (Pritchard et al. 2000). Then, we assessed ΔK values and LnP(K) (natural logarithm of probability of data) plot to choose the value best explaining K clusters.

We applied a neighbor-joining tree representation (Saitou and Nei 1987) of relationships between populations to obtain preliminary inferences about their histories (Fig. S5.6). The neighbor-joining tree was built with POPTREE2 (Takezaki et al. 2010) using a matrix of genetic distances (D_A) based on the studied microsatellites. Additionally, we examined a mitotype distribution to discard unlikely population histories.

Paleorecord review

The Neotoma Paleoecology Database (http://www.neotomadb.org/) was utilized to obtain paleorecords. Other paleorecords were taken from the literature (Binney et al. 2009; Binney et al. 2017; see also references in Appendix S7 in Supplementary Material). We calculated pollen percentages for *P. sibirica* on the basis of the total pollen number of the TRSH (trees and shrubs) ecology group and applied a 10% pollen threshold as an indication of the species' presence within paleo-sections. Although the threshold values for *Pinus* (haploxylon-type) recommended by Lisitsyna et al. (2011) were assessed at 1−5%, we decided to employ the higher value. All calibrated pollen/macrofossil paleorecords were examined to survey spatial changes of *P. sibirica*'s range from the late Pleistocene to the Holocene. These dated paleorecords were put on maps for the following time periods: (a) marine isotope stage (MIS) 5—130−71 cal ka BP (calibrated kiloanni before 1950 Common Era, hereinafter referred to as ka; thousand years ago); (b) MIS 3—57−29 ka; (c) MIS 2—29−14 ka (Lisiecki and Raymo 2005); and (d) Holocene times—14−10 ka, 10−8 ka, 8−4 ka, 4−0 ka (Fig. 3). Unfortunately, we did not find any published paleorecords showing the presence of Siberian stone pine or associated species during MIS 4 (71−57 ka). We visualized genetic data (as pie charts) and paleontological finds using GENGIS 2.5.3 (Parks et al. 2013). Geographical maps were downloaded from the Natural Earth website (www.naturalearthdata.com). The maps are based on the geographic coordinate system (projection) and WGS84 datum.

Species distribution modeling

Occurrence data These data were retrieved from the GBIF digital biodiversity data portal (https://www.gbif.org/), herbarium collections of the Moscow Digital Herbarium (www.plant.depo.msu.ru) (last accessed: October 20, 2021), and our field work. In total, 4723 occurrence records were assessed to eliminate data from outside the natural range, erroneous data, and duplicated occurrence records, thus yielding 2571 occurrences. The occurrence data may contain a strong sampling bias of the survey effort; this bias can decrease the accuracy of SDM methods (Hortal et al. 2008; Veloz 2009). Consequently, we performed spatial filtering (Boria et al. 2014) and randomly removed occurrences that were within 10 km of one another. After that, we divided this species' range into geographic regions (Urals, West Siberian Plain, and others) to equalize the number of records by randomly selecting occurrences in order to obtain regions with the same density and got 305 unique localities (Fig. 4a). For pseudo-absence data, presence-only distribution models need to estimate extrapolating reliability of the species distribution across time (VanDerWal et al. 2009). Accordingly, we chose random selection of pseudo-absences, which has a better effect on model specificity thereby reducing false positive error (Barbet-Massin et al. 2012). To this end, we

generated absence data in MaxEnt v3.4.1 (Phillips et al. 2006) using 100,000 background points from which we randomly selected 306 points. This pseudo-absence dataset was employed instead of true absences. We included pseudo-absence observations that fall within grid cells along with presence observations to avoid preconceived assumptions about environments (un)favorable for Siberian stone pine (Stokland et al. 2011).

Bioclimatic variables The distribution of Siberian stone pine was modeled for current time and projected into the mid-Holocene (8.326–4.2 ka), the LGM (21 ka), and the LIG (the last interglacial; 130 ka) climate scenarios. Bioclimatic layers of 19 predictors (the mid-Holocene and the LIG) were downloaded from the PaleoClim v1.0 database (www.paleoclim.org) (Brown et al. 2018; Fordham et al. 2017; Otto-Bliesner et al. 2006). Present-day and LGM layers were retrieved from the CHELSA v1.2b database (www.chelsa-climate.org) (Karger et al. 2017). All layers were used at a spatial resolution of 2.5 min and represented coupled atmosphere–ocean general circulation models from the Hadley Centre Coupled Model Version 3 (HadCM3) (Brown et al. 2018). We performed Pearson's correlation analysis on standardized predictors and retained the subset of predictors with $|r| < 0.8$ (Table S3.4). These calculations were performed in STATISTICA 8.0 (StatSoft, Inc., USA). Bio3, Bio8, and Bio15 show no correlation coefficients higher than |0.8| with any of the other predictors. Therefore, we included them in the model. The other five predictors (Bio7, 10, 11, 18, and 19) were added in the model from correlated pairs based on the criteria of the highest model contribution and training gain using the jackknife test in MaxEnt v3.4.1 (Phillips et al. 2006). In this way, we successively discarded Bio1, 2, 4, 5, 6, 9, 12, 13, 14, 16, and 17 to create the final set of eight variables (Table S3.4).

Development of models A series of candidate Maxent models at different combinations of parameter settings was developed by means of the KUENM R package (Cobos et al. 2019) in R 4.0.3 (R Core Development Team 2013). These combinations included five regularization multiplier values (0.5, 1, 2, 3, and 4), three feature class combinations (lq, lqp, and q, where "l" = linear, "q" = quadratic, and "p" = product), and 37 combinations of uncorrelated environmental variables. The set of variables was as follows: Bio3, Bio7, Bio8, Bio10, Bio11, Bio15, Bio18, and Bio19 (Table S3.4). Thus, we designed and evaluated a total of $5 \times 3 \times 37 = 555$ candidate models. The best model was selected based on pROC (the partial area under the receiver operating characteristic curve) analysis, a threshold of omission error ($\leq 5\%$), and the least Akaike information criterion (corrected for small sample size: AICc) (Table S3.5; Figs. S3.2, S3.4, and S3.5). To convert Maxent continuous logistic output into binary presence-absence maps, we constructed a confusion matrix and calculated the true skill statistic for pseudo-absence data (TSS′) (Liu et al. 2016). The threshold value that maximizes TSS′ (0.765) was chosen at the 30% level (Fig. S3.3) and was utilized for further binary transformation of all prediction maps by a custom C++ script.

All manipulations of the climate layers and maps were performed using GLOBAL MAPPER v20.0.0 (Blue Marble Geographics, USA), ArcGIS 10.8.2 (ESRI, USA), and SDMtoolbox 2.0 (Brown et al. 2017).

Results

Spatial pattern of mtDNA polymorphisms

We found SNPs in two regions of the mitochondrial genome, which were designated in accordance with mtDNA assembly contigs: 455 and 161 (Table S2.3). One nucleotide substitution (SNP) distinguished two variants of each marker. Thus, the mitochondrial markers produced four mitotypes (Table 1).

SAMOVA detected the highest differentiation among populations at $K = 3$ ($F_{CT} = 0.88$), and three SAMOVA clusters showed high among-group variation (Table 2) ($P < 0.001$). The first cluster (mitotype M2) covers the vast territory of Siberia from the Urals (I) to the Kuznetsk Alatau (II) (Ural-K.Alatau group). Two other clusters belong to mountain systems of South Siberia: (1) the West Sayan and Altai group (III) (mitotype M1, the WSA group) and (2) the East Sayan group (IV) (mitotype M3, the ES group) (Fig. 1(a)). Admixed populations (including mitotypes M1 and M2) were found between the southern part of the Kuznetsk Alatau and the WSA group. Other admixed populations were found between groups WSA and ES (including mitotypes M1 and M3) and between the eastern part of the Ural-K.Alatau group and ES group (including mitotypes M2 and M3). Curiously, we found these admixed populations in Central Siberia. Apparently, this area has become an intraspecies admixture zone for the discovered population groups. Probably, this is where the fourth recombinant

Table 1 MtDNA mutations and haplotypic combinations in *Pinus sibirica*. Colors of the mitotypes refer to those displayed in Fig. 1(a)

Names of mitotypes (color)	455 marker nucleotide	161 marker nucleotide
M1 (blue)	C	C
M2 (black)	A	C
M3 (red)	C	A
M4 (green)	A	A

Table 2 Results of spatial analysis of molecular variance (AMOVA) based on mtDNA data for 66 Siberian stone pine populations, with the values presented for joining populations into three groups according to the maximum among-group fixation index F_{CT}

Variation source	Degrees of freedom	Sum of squares	Variance components	Variation, %
Among 3 SAMOVA groups	2	809.31	0.827	88.15
Among 66 populations within SAMOVA groups	63	94.39	0.065	6.91
Within populations	1416	65.69	0.046	4.95
Total	1481	970.01	0.938	–

Fixation indices: F_{CT}=0.882*** among SAMOVA groups; F_{SC}=0.583*** among populations within SAMOVA groups; F_{ST}=0.951*** among all populations within the total population. ***$P<0.001$

mitotype (M4) originated, which occurred in this narrow zone, where three population groups (WSA, ES, and the south-east part of the Ural-K.Alatau group) could admix.

Genetic structure inferred from nuclear microsatellite markers

Null alleles were revealed at locus PS_364418 (14.47%). To assess an influence of the null alleles on the genetic differentiation, we calculated F_{ST} values by the *ENA* (excluding null alleles) correction algorithm (Chapuis and Estoup 2007). We did not detect any differences between populations' F_{ST} values ($P>0.05$) with and without the null alleles. We also did not observe significant differences in the analogous test of D_A values ($P>0.05$). Therefore, we chose to retain this null-allelic locus for further analyses, except for the analysis with STRUCTURE. The pairwise tests of the loci did not reveal linkage disequilibrium ($P>0.05$).

The STRUCTURE analysis produced a dual clustering result. The maximal ΔK and the LnP(K) values separated all populations into two *K* genetic clusters: Ural and South Siberia groups. In contrast, the lower ΔK and the LnP(K) peaks corresponded to four *K* clusters: groups Ural, KA (Kuznetsk Alatau), WSA, and ES (Fig. 1(b−c) and S6.7). We suspect that such a dual result is possible for continuous populations of *P. sibirica*, inside of which closer populations are located. The population structure of close South Siberia groups (KA, WSA, and ES) may have been homogenized by pollen flow over time. By contrast, Ural populations have not been exposed to pollen flow from South Siberia owing to the huge geographical distances. However, despite this, the result for four *K* clusters is biologically meaningful that is supported by the different mitotypes of the South Siberia groups.

The principal coordinate analysis based on the genetic distances (D_A) yielded four main population clusters (Fig. 2). Two microsatellite clusters matched the mitochondrial clusters completely. Firstly, the microsatellite WSA cluster corresponded to mitotype group M1. Similarly, the microsatellite ES cluster matched mitotype cluster M3. By contrast, mitotype group M2 (Ural-K.Alatau) was divided into Ural and Kuznetsk Alatau subgroups according to the microsatellite data. Lastly, recombinant mitotype M4 matched the central admixed zone of the main microsatellite clusters (Fig. 2).

The pairwise exact test uncovered genetic differences among four microsatellite groups (WSA, ES, KA, and Ural) ($P<0.01$). Pairwise F_{ST} among these groups varied from 0.011 to 0.062 ($P<0.01$). AMOVA revealed that 2.24% of these differences were attributable to group divergence,

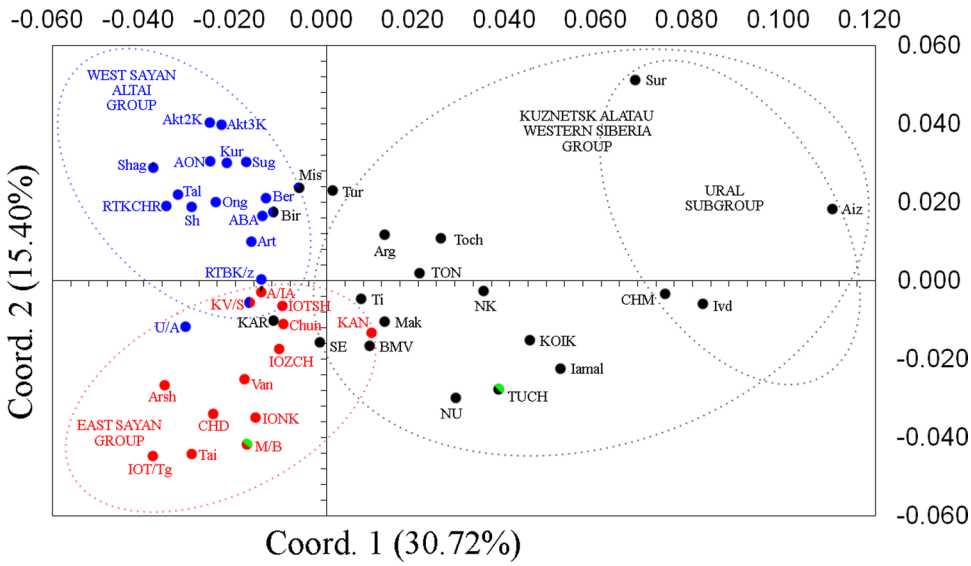

Fig. 2 Plot of two main principal coordinates and their proportion (%) of the total variance according to principal coordinate analysis (PCoA) based on nuclear microsatellite data. A matrix of pairwise Nei's genetic distances (D_A) between the 48 Siberian stone pine populations was used for PCoA. Sample codes of *P. sibirica* populations are explained in Table S1.1. The colored pie charts show the mitotype groups (Fig. 1(a)): blue, M1; black, M2; red, M3; and green, M4

1.78% to population differences within groups, and 87.19% to an individual level (Table 3).

Paleorecord data

Paleobotanical MIS 2 finds of Siberian stone pine are scarce. Nevertheless, they clearly indicate that Siberian stone pine could have survived the LGM in mountains of Urals and South Siberia as well as on the Lake Baikal coast (Fig. 3(e)). Perhaps from here, the Holocene expansion of *P. sibirica* began, which can be deduced from an increase in pollen records during the Holocene in Siberia (Fig. 3(a–d)).

The more ancient macrofossils and pollen were found for MIS 3 (57−29 ka) in the Ob–Irtysh interfluve and near Lake Baikal. The other part of pollen data has to do with MIS 5 (130−71 ka) and implies a favorable environment for *P. sibirica* at northern latitudes (Fig. 3(f–g)).

Species distribution modeling

The modern range of *P. sibirica* was predicted with good accuracy (the area under the curve [AUC] > 0.90; Fig. 4a and S3.4–S3.5) (Swets 1988). TSS′ was the highest at the 30% threshold level and corresponded to 76.5% of correct classifications (Fig. S3.3). We observed a seeming "shifting" of the predicted area to the west. This shift was due to two circumstances. Firstly, the retreat of the western edge of the modern range is associated with the disappearance of Siberian stone pine owing to massive deforestation since the beginning of economic development (Krylov et al. 1983).

Table 3 Analysis of molecular variance (AMOVA) based on the nuclear microsatellites for 48 Siberian stone pine populations

Variation source	Degree of freedom	Sum of squares	Variance components	Variation, %
Among 4 groups[a]	3	97.76	0.043	2.24
Among 48 populations within groups[a]	44	174.15	0.034	1.78
Among individuals within populations	1339	2678.76	0.168	8.79
Within individuals	1387	2309	1.665	87.19
Total	2773	5259.68	1.909	100

[a]Grouping of populations was based on the results of the principle coordinate analyses (see Fig. 2). Fixation indices: F_{CT}=0.022*** among four groups; F_{SC}=0.018*** among 48 populations within groups; F_{IS}=0.09*** among individuals within populations; F_{IT}=0.13*** among individuals within the total population; F_{ST}=0.042*** among all populations within the total population. ***P<0.001

In other words, Siberian stone pine existed westward in past centuries, and the SDM result showed this past, westward extended edge of the species' range. On the other hand, SDM did not show the presence of Siberian stone pine near the eastern edges of its range (except the Aldan Highlands). In some sense, this is not surprising because this is a less-favored zone for *P. sibirica* growth owing to the extreme continental climate of East Siberia. Here, competing with Siberian dwarf pine, Siberian stone pine selectively grows near rivers and lakes (for example, in the Leprindo lake system) and prefers moist mountain slopes at an altitude of 450–650 m a.s.l. (Krylov et al. 1983). Perhaps for this reason, the eastern range of *P. sibirica* needs a separate modeling with high-resolution bioclimatic layers.

The LGM extrapolation predicted suitable climatic conditions in the southern Siberian Mountains, the southern Urals, and the southern coast of Lake Baikal (Fig. 4c). In the mid-Holocene (8.326−4.2 ka), areas suitable for *P. sibirica* widened to the Yamal Peninsula and to a part of the Taymyr Peninsula (i.e., displacement to the north) (Fig. 4b). Additionally, the potential distribution of *P. sibirica* was broad on the Russian Plain. Similar results were obtained for the LIG time (130 ka) (Fig. 4d). Both periods are characterized by a generally warm and humid climate known as the Holocene climatic optimum and the beginning of the Eemian interglacial (MIS 5e) (Kalis et al. 2003; NEEM community members 2013). At the Holocene optimum, the average air temperatures in the temperate latitudes of the northern Hemisphere were 2–3 °C higher than the current ones, and annual precipitation exceeded the current values by 75–100 mm (Neishtadt 1983). These favorable conditions may have contributed to intensive expansion into the northern part of West Siberia.

Discussion

Discovery of glacial refugia in South Siberia and Urals

Combined analysis of mtDNA and nuclear microsatellites delineated clear-cut genetic structure of *P. sibirica* populations (Fig. 1). We suppose that this spatial-genetic pattern is tightly related to the late Quaternary history of Siberian stone pine in Siberia. We discovered four mitotypes and suggest several origins of *P. sibirica* postglacial dispersal. The fourth mitotype (M4) was recombinant and originated from a contact area of the other mitotypes. Therefore, we did not regard this zone as a historical cluster for further historical reconstructions. According to the other three mitotypes, these origins may be related to several LGM refugia within South Siberia: East Sayan (and/or Lake Baikal) (M3), Kuznetsk Alatau (M2), and Altai-West Sayan

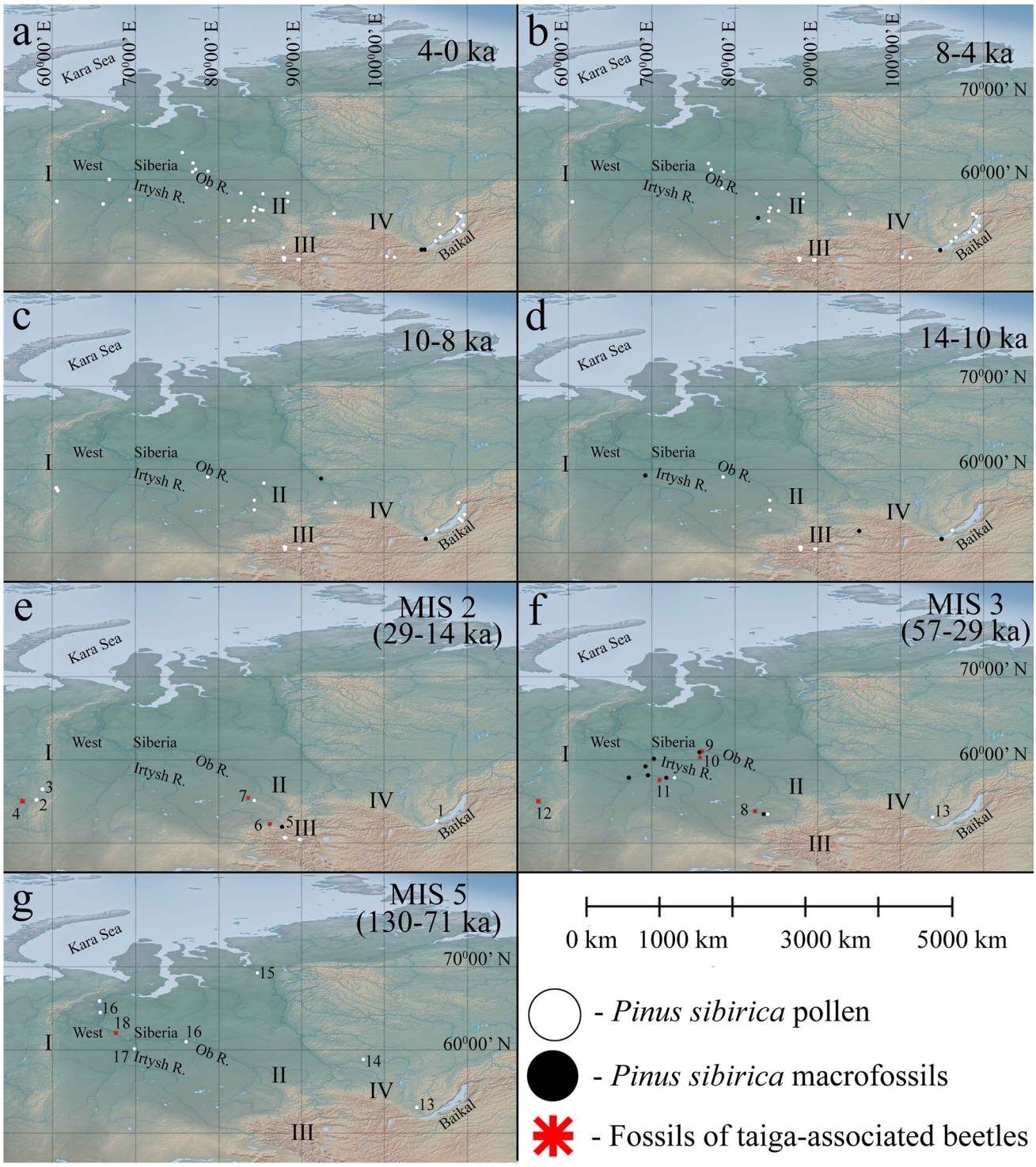

Fig. 3 Map representation of paleorecords. (**a**–**d**) Holocene time series; e marine isotope stage (MIS) 2; f MIS 3; g MIS 5. I: the Urals, II: the Kuznetsk Alatau, III: the Altai with the West Sayan, IV: the East Sayan. For reference numbers see Appendix S7 of Supplementary Materials

(M1). In addition, LGM pollen and SDM results indicate a contraction of the *P. sibirica* range toward South Siberia Mountains and Urals (Figs. 3(e) and 4c). In these glacial refugia, Siberian stone pine could persist, and further postglacial expansion gave rise to the modern populations, as is the case for other coniferous species (e.g., *A. sibirica* and *P. obovata*) (Semerikov et al. 2013, 2019; Tollefsrud et al. 2015).

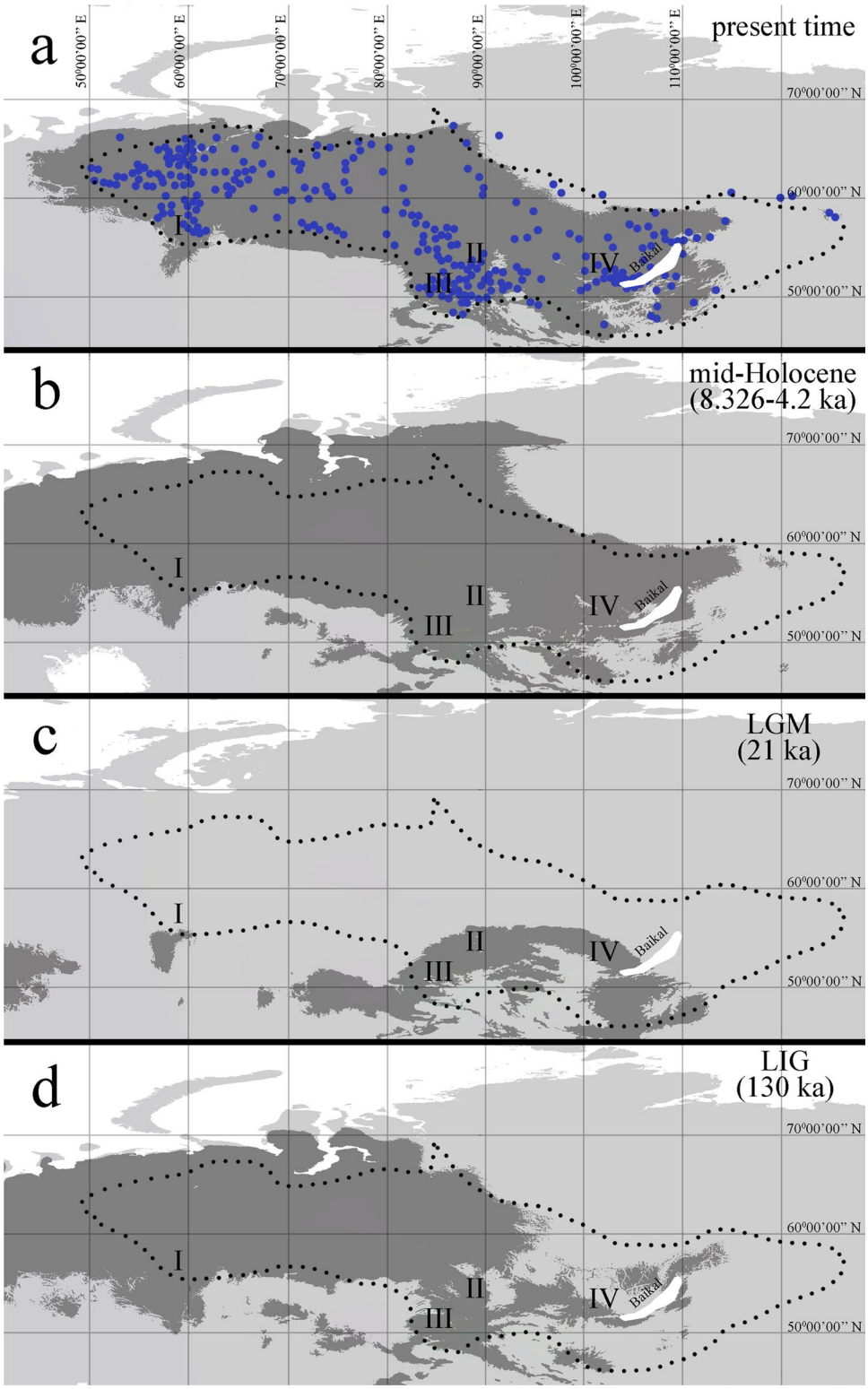

Fig. 4 Areas of *Pinus sibirica* identified as suitable for four time periods are highlighted in dark gray. The suitable conditions for **a** the present time; **b** the middle Holocene (8.326–4.2 ka); **c** the last glacial maximum (LGM, 21 ka); **d** the last interglacial (LIG, 130 ka). I: the Urals, II: the Kuznetsk Alatau, III: the Altai with the West Sayan, IV: the East Sayan. The dotted line denotes the natural range of *Pinus sibirica*. Blue dots on the upper map (**a**) show true occurrences after spatial filtering

Pollen data suggested that the broad expansion of Siberian stone pine occurred in the Holocene in whole Siberia (Fig. 3(a–d)). Nonetheless, Altai and West Sayan populations (the WSA group) remained within the modern area after the LGM because mitotype M1 did not occur anywhere else (Fig. 1(a)). A similar pattern was observed for common juniper from Altai populations, which did not make a genetic contribution to the main geographic range (Hantemirova et al. 2017). Survival of Siberian stone pine in Altai refugia was also confirmed by fossil seeds near Lake Teletskoe

(16−15 ka) (Baryshnikov 1995, 1996) and by pollen records along the Anui River valley (Bolikhovskaya and Shunkov 2014). Moreover, LGM *Larix-Pinus sibirica* forests near Lake Teletskoe were predicted by species distribution modeling (Hais et al. 2015).

Another mitotype (M3) was detected across the East Sayan and the eastern part of Siberia (Lake Baikal). On the Lake Baikal coast, pollen of Siberian stone pine was found that is related to the LGM (Kobe et al. 2022), thereby indicating a refugium near Lake Baikal. Furthermore, the predicted LGM geographic range revealed favorable habitats on the southern coast of Lake Baikal (Fig. 4c). As for the other forest species, there is evidence that the origin of *A. sibirica* populations in West Siberia may be located in a Lake Baikal refugium (Semerikov et al. 2019).

Thus, our observations are concordant with the hypothesis of the southern mountain refugia (Blyakharchuk 2010). The mountain relief was able to divide the common harsh climatic pattern of the LGM into more protected locations characterized by favorable microclimates, for instance, along the northern Altai foothills, allowing survival of boreal trees (Chlachula 2001). An influence of large water bodies on the local climate cannot be ruled out either. As stated above, Siberian stone pine is a mountain species preferring ecotopes with humid air and soil. Apparently, the LGM climate was favorable for *P. sibirica* forests along mountain river valleys and lakes. Big fault lakes (i.e., Lake Teletskoe and Lake Baikal) and ancient ice-dammed lakes (Adamenko et al. 2015; Astakhov 2006; Carrivick and Tweed 2013; Reuther et al. 2006) could positively affect local climate, for example, on the slopes surrounding these lakes during the LGM. By contrast, unsuitable conditions for *P. sibirica* on the West Siberian Plain correspond to an extremely harsh dry climate during the LGM (Velichko et al. 2011).

The M2 mitotype was common for the Kuznetsk Alatau and Ural groups, but we cannot determine the distribution routes of Siberian stone pine on the West Siberian Plain from mitochondrial data alone. On the other hand, we proposed refugia status of the Urals owing to paleorecords of the Urals, SDM results, and the microsatellite separation of the Ural group from the other groups (Fig. S6.7). Moreover, the survival of *P. sibirica* forests during the LGM in the Urals may be indirectly confirmed by recent phylogeographic studies on some tree species such as Scots pine, Siberian fir, small-leaved lime (*Tilia cordata* Mill.), and pedunculate oak (*Quercus robur* L.) (Semerikov et al. 2018, 2019; Semerikova et al. 2020, 2021). The lack of mitotype differentiation between the Ural and Kuznetsk Alatau groups and clear microsatellite differentiation imply their later separation. Nevertheless, the possibility of several events of Ural colonization cannot be ruled out. For example, similar associations have been reported for *A. sibirica* populations, indicating several ancient migrations into the Urals (Semerikov et al. 2019). On the other hand, this observation is self-evident because these Ural populations possessed Altai's and Baikal's mitotypes. Obviously, the colonization routes of Siberian stone pine differ from those of *A. sibirica*, despite their similar ecological requirements. And furthermore, *P. sibirica* routes clearly differ from *L. sibirica*, which was possibly spread from the Sayan Mountains (Semerikov et al. 2013). Additional difficulties derive from nonanalogous climates and communities of the past (Williams et al. 2004). Presumably, the ecological similarities are unreliable evidence because we cannot determine with certainty a tolerance limit of a species in a past adverse environment.

Siberian stone pine on the West Siberian Plain

We can advance some hypotheses about migration routes of Siberian stone pine on the West Siberian Plain on the basis of mtDNA and of the microsatellite patterns, paleoecological premises, and SDM results. The common M2 mitotype is a sign of earlier colonization events. Besides, this mitotype may imply an ancient common range between the Ural and Kuznetsk Alatau populations. In turn, the microsatellite differentiation may be a consequence of an isolation between these populations during glaciations. It is possible that MIS 4 (71−57 ka) (Early Zyryanka glaciation in the Siberian regional stratigraphic scheme (Volkova and Babushkin 2000)) and/or the LGM were quite serious drivers behind such isolation. In particular, this assumption is deduced from the absence of paleorecords and favorable areas on the West Siberian Plain during the LGM (Figs. 3(e) and 4c). Unfortunately, we have no bioclimatic layers or *P. sibirica* paleorecords for MIS 4. Therefore, we can only speculate about the absence of *P. sibirica* on the West Siberian Plain during MIS 4.

We propose that the modern population structure of *P. sibirica* forests on the West Siberian Plain could have formed at the following main stages. Firstly, the absence of pollen and suitable habitats on the West Siberian Plain during the LGM suggests that the divergence between the Ural and Kuznetsk Alatau populations occurred during the LGM. Secondly, their common mitotype (M2) may be evidence of the existence of a wide ancestral range before the LGM, as confirmed by paleoecological data on MIS 3 and MIS 5 (Fig. 3(f–g)).

Paleontological evidence (Krivonogov 1988) points to a hypothetical discontinuous "forest belt," which probably connected the Ural and Kuznetsk Alatau populations across a central part of the West Siberian Plain (Fig. 3(f)) during MIS 3 (57−29 ka) (also called the "Karginsky interstadial" in the Siberian regional stratigraphic scheme) (Astakhov and Mangerud 2005). The Karginsky interstadial is the most problematic Pleistocene period in West Siberia because many ^{14}C ages obtained in the 1960–1970s were

incorrect and have been stratotypes of the pre-Karginsky age (Laukhin et al. 2012). Therefore, we took into consideration the calibrated paleo-sections that corresponded to the Karginsky interstadial after Laukhin's revision (Laukhin 2011). Karginsky macrofossils indicate that Siberian stone pine grew on the Ob–Irtysh interfluve 46−31 ka, and several sites with fossils of wood-associated beetles (e.g., *Polygraphus* sp.) were found near middle reaches of the Ob River (Fig. 3(f)). The fossil beetles from the southern Urals are also assigned to forest fauna of the boreal humid type and are related to MIS 3 (Dudko et al. 2022). Nonetheless, the climate of the Karginsky interstadial was not especially favorable for the widespread advance of taiga forests (Andreev et al. 2006; Laukhin et al. 2015; Sher et al. 2005). In our opinion, the Karginsky paleorecords reflect a weak expansion of Siberian stone pine after MIS 4 (Early Zyryanka glaciation: 71−57 ka). Therefore, we can assume that the Karginsky *P. sibirica* forests might have only partially recovered in limited areas in the Ob–Irtysh interfluve. Probably, more ancient *P. sibirica* forests from the Kuznetsk Alatau and Urals were connected via the northern taiga forests during MIS 5 (130–71 ka) ("Kazantsevo" in the Siberian regional stratigraphic scheme). Furthermore, in the warmest Eemian interval (131–120 ka) (Kukla et al. 2002; Lozhkin et al. 2007), dark coniferous forests spread to the north and might have created a northern continuous taiga-forest belt across the northern Siberia (Arkhipov and Volkova 1994; Borisova 2005; Sheinkman et al. 2021; Volkova et al. 2002). For instance, abundant pollen of *Picea* and *P. sibirica* with some *Abies* was found near lower reaches of the Yenisei River (121–98 ka), thereby pointing to favorable environments for Siberian stone pine at high latitudes (Fig. 3(g)) (Astakhov and Mangerud 2005; Kind 1974).

According to all available information, we propose several assumptions about the recolonization routes of Siberian stone pine during the late Quaternary: (1) a northern connection of *P. sibirica* forests during MIS 5 (130−71 ka); (2) the cold interval of MIS 4 (71−57 ka) might have disrupted the northern forest range; (3) this range recovered partially during MIS 3 (57−29 ka) from Ural and Kuznetsk Alatau refugia with the formation of a discontinuous "forest belt" across the West Siberian Plain. Nonetheless, additional genetic data—and possibly a paleogenomic investigation—are necessary to confirm or refute these suppositions. Finally, the last disjunction of the MIS 3 "forest belt" occurred during MIS 2 (29−14 ka), because for MIS 2, pollen finds of *P. sibirica* on the West Siberian Plain have not been reported (Fig. 3(e)). There is evidence that during the culmination of MIS 2 (LGM), the West Siberian Plain was dominated by cryoarid steppe vegetation with rare patches of larch forests (Krivonogov 1988; Volkova 2002). A large part of the West Siberian Plain represented a (sub)arctic climatic zone similar to cold deserts (Velichko et al. 2011). Nevertheless, Siberian spruce might have persisted in the southern part of the West Siberian Plain during the LGM (Tarasov et al. 2000; Tollefsrud et al. 2015). However, the last supposition for *P. sibirica* has never been confirmed by paleontological finds. Therefore, after the LGM, Siberian stone pine probably colonized the West Siberian Plain from Ural and Kuznetsk Alatau refugia. Apparently, along springs of the Ob River, these refugial populations could spread to form the separated populations on the West Siberian Plain.

Conclusions

The combined data presented in this study allowed us to draw the following major conclusions: (1) Siberian stone pine populations from the West Sayan and the Altai (WSA), the East Sayan (ES), the Kuznetsk Alatau (KA), and the Urals are four genetically distinct populations that occupied mountainous refugia during the LGM; (2) the mountains and ancient lakes could play a key role in the survival of the forests by establishing a locally favorable climate; (3) the West Sayan and Altai populations are a single population that has not expanded after the LGM beyond its current borders; (4) *Pinus sibirica* began to occupy the whole West Siberian Plain from the Urals and the Kuznetsk Alatau during the Holocene; (5) At the same time, the eastern populations of *P. sibirica* originated from the East Sayan and/or the Lake Baikal coast and spread into Pre-Baikalia and parts of Middle Siberia. All these findings clearly indicate that recolonization routes of Siberian stone pine differ from those of other Siberian forest species.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11295-023-01592-z.

Acknowledgements We would like to thank the following people, without whom we would not have been able to conduct this research. Firstly, we express our gratitude to the team of the Russian Centre of Forest Health for the help with the collection of the materials. We wish to thank Dr. E.V. Zinovyev for guiding us to important paleoecological publications and for a productive discussion. D.N. Shuvaev is especially grateful to his supervisor, Dr. V.V. Tarakanov, whose support allowed us to complete the study. The English language was corrected and certified by shevchuk-editing.com. This research was financially supported by a grant from the Russian Science Foundation, No. 22-24-00665, https://rscf.ru/en/project/22-24-00665/.

Author contribution Conceptualization, D.N.S.; methodology, D.N.S. and V.L.S.; software, D.N.S.; validation, D.N.S. and V.L.S.; formal analysis, D.N.S.; investigation, D.N.S.; resources, D.N.S., V.L.S., Y.A.P., and G.V.K.; data curation, D.N.S.; writing−original draft preparation, D.N.S.; review and editing, D.N.S. and V.L.S.; visualization, D.N.S.; supervision, D.N.S. and V.L.S.; project administration, D.N.S.; funding acquisition, D.N.S. and V.L.S. All authors have read and agreed to the final version of the manuscript.

Data Availability The data that support the findings of this study are openly available in www.figshare.com at https://doi.org/10.6084/m9.figshare.19291856 and https://doi.org/10.6084/m9.figshare.20892247.

Declarations

Conflict of interest The authors declare no competing interests.

Data archiving The microsatellite dataset and the pseudo-absence occurrences related to this article can be found at https://doi.org/10.6084/m9.figshare.19291856 and https://doi.org/10.6084/m9.figshare.20892247, respectively.

References

Adamenko MM, YaM G, Solomina ON (2015) Glacial history of the Kuznetsky Alatay Mountains. Environ Earth Sci 74:2065–2082. https://doi.org/10.1007/s12665-015-4621-2

Andreev AA, Forman SL, Ingólfsson Ó, Manley WF (2006) Middle Weichselian environments on western Yamal Peninsula, Kara Sea based on pollen records. Quatern Res 65(02):275–281. https://doi.org/10.1016/j.yqres.2005.11.005

Arkhipov SA, Volkova VS (1994) Geological history, pleistocene landscapes and climate in West Siberia. OIG GM SO RAN, Novosibirsk. Retrieved January 19, 2022, from https://www.geokniga.org/bookfiles/geokniga-geologicheskaya-istoriya-landshafty-i-klimaty-pleystocena-zapadnoy-sibiri.pdf

Astakhov VI (2006) Evidence of Late Pleistocene ice-dammed lakes in West Siberia. Boreas 35(4):607–621. https://doi.org/10.1111/j.1502-3885.2006.tb01167.x

Astakhov VI, Mangerud J (2005) The age of the Karginsky Interglacial strata on the Lower Yenisei. Doklady Earth Sciences 403(5):673–676. https://people.uib.no/ngljm/PDF_files/AstakhovandMangerud2005

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol Evol 3(2):327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Baryshnikov GYA (1995) Fossil floras of the transitional zones of the Altai. Altai Republic Flora's Journal 1:111–115

Baryshnikov GYA (1996) Fossil vegetation in terraced complexes of the Altai. Proceedings of the South Siberian Botanical Garden 129–135

Belokon MM, Politov DV, Mudrik EA, Polyakova TA, Shatokhina AV, Belokon YuS, Oreshkova NV, Putintseva YuA, Sharov VV, Kuzmin DA, Krutovsky KV (2016) Development of microsatellite genetic markers in Siberian stone pine (*Pinus sibirica* Du Tour) based on the *de novo* whole genome sequencing. Russ J Genet 52:1263–1271. https://doi.org/10.1134/s1022795416120036

Belokon MM, Belokon YS, Petrova EA, Vasilyeva GV, Efimova AP, Zakharov ES, Goroshkevich SN, Politov DV (2022) Ancient and modern hybridization of Siberian stone pine and Dwarf siberian pine in the south of Yakutia. Russ J Genet 58:1287–1296. https://doi.org/10.1134/S1022795422110023

Binney HA, Willis KJ, Edwards ME, Bhagwat SA, Anderson PM, Andreev AA, Blaauw M, Damblon F, Haesaerts P, Kienast F, Kremenetski KV, Krivonogov SK, Lozhkin AV, MacDonald GM, Novenko EY, Oksanen P, Sapelko TV, Valiranta M, Vazhenina L (2009) The distribution of late-Quaternary woody taxa in northern Eurasia: Evidence from a new macrofossil database. Quatern Sci Rev 28(23–24):2445–2464. https://doi.org/10.1016/j.quascirev.2009.04.016

Binney H, Edwards M, Macias-Fauria M, Lozhkin A, Anderson P, Kaplan JO, Andreev A, Bezrukova E, Blyakharchuk T, Jankovska V, Khazina I, Krivonogov S, Kremenetski K, Nield J, Novenko E, Ryabogina N, Solovieva N, Willis K, Zernitskaya V (2017) Vegetation of Eurasia from the last glacial maximum to present: key biogeographic patterns. Quatern Sci Rev 157:80–97. https://doi.org/10.1016/j.quascirev.2016.11.022

Blyakharchuk TA (2010) Postglacial dynamics of the vegetation cover of West Siberian Plain and the Altai-Sayan mountain region (according to the spore-pollen analysis of marsh and lake deposits). Dissertation, Institute of Monitoring of Climatic and Ecological Systems of the Siberian Branch of the Russian Academy of Sciences

Bolikhovskaya NS, Shunkov MV (2014) Pleistocene environments of northwestern Altai: vegetation and climate. Archaeol Ethnol Anthropol Eurasia 42(2):2–17. https://doi.org/10.1016/j.aeae.2015.01.001

Boria RA, Olson LE, Goodman SM, Anderson RP (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecol Model 275:73–77. https://doi.org/10.1016/j.ecolmodel.2013.12.012

Borisova OK (2005) Vegetation and climate changes at the Eemian/Weichselian transition: new palynological data from central Russian plain. Polish Geological Institute Special Papers 16:9–17. https://bibliotekanauki.pl/articles/1186124

Brown JL, Bennett JR, French CM (2017) SDMtoolbox 2.0: the next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. PeerJ 5:e4095. https://doi.org/10.7717/peerj.4095

Brown JL, Hill DJ, Dolan AM, Carnaval AC, Haywood AM (2018) PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. Scientific Data 5(180254). https://doi.org/10.1038/sdata.2018.254

Carrivick JL, Tweed FS (2013) Proglacial lakes: character, behaviour and geological importance. Quatern Sci Rev 78:34–52. https://doi.org/10.1016/j.quascirev.2013.07.028

Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol 24(3):621–631. https://doi.org/10.1093/molbev/msl191

Chlachula J (2001) Pleistocene climate change, natural environments and palaeolithic occupation of the Altai area, west-central Siberia. Quatern Int 80–81:131–167. https://doi.org/10.1016/S1040-6182(01)00023-4

Cobos ME, Peterson AT, Barve N, Osorio-Olvera L (2019) Kuenm an R package for detailed development of ecological niche models using Maxent. PeerJ 7:e6281. https://doi.org/10.7717/peerj.6281

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Roy Stat Soc: Ser B (methodol) 39(1):1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Devey ME, Bell JC, Smith DN, Neale DB, Moran GF (1996) A genetic linkage map for *Pinus radiata* based on RFLP, RAPD, and microsatellite markers. Theor Appl Genet 92:673–679. https://doi.org/10.1007/BF00226088

Dudko RYu, Danukalova GA, Gurina AA, Ivanov AV, Mikhailov YE, Osipova EM, Prosvirov AS, Solodovnikov AYu, Legalov AA, Zinovyev EV (2022) Insects and molluscs of the Late Pleistocene at the Gornovo site (Southern Ural foreland, Russia): new data on palaeoenvironment reconstructions. Quatern Int 632:154–177. https://doi.org/10.1016/j.quaint.2021.10.003

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. Mol Ecol 11(12):2571–2581. https://doi.org/10.1046/j.1365-294X.2002.01650.x

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14(8):2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux

and Windows. Molecular Ecol Resour 10(3):564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fordham DA, Saltre F, Haythorne S, Wigley TML, Otto-Bliesner BL, Chan KC, Brook BW (2017) PaleoView: a tool for generating continuous climate projections spanning the last 21 000 years at regional and global scales. Ecography 40(11):1348–1358. https://doi.org/10.1111/ecog.03031

Fujita K, Silver J (1994) Single-strand conformational polymorphism. Genome Research 4:S137−S140. https://genome.cshlp.org/content/4/3/S137.short

Gavin DG, Fitzpatrick MC, Gugger PF, Heath KD, Rodríguez-Sánchez F, Dobrowski SZ, Hampe A, Hu FS, Ashcroft MB, Bartlein PJ, Blois JL, Carstens BC, Davis EB, de Lafontaine G, Edwards ME, Fernandez M, Henne PD, Herring EM, Holden ZA, Kong W, Liu J, Magri D, Matzke NJ, McGlone MS, Saltré F, Stigall AL, Tsai Y-HE, Williams JW (2014) Climate refugia: joint inference from fossil records, species distribution models and phylogeography. New Phytol 204(1):37–54. https://doi.org/10.1111/nph.12929

Goncharenko GG, Padutov VE, Silin AE (1991) Degree of genetic subdivision and differentiation in the natural Siberian stone pine populations of the USSR. Reports USSR Acad Sci 317:1477–1483

Graham V, Baumgartner JB, Beaumont LJ, Esperón-Rodríguez M, Grech A (2019) Prioritizing the protection of climate refugia: designing a climate-ready protected area network. J Environ Planning Manage 62(14):2588–2606. https://doi.org/10.1080/09640568.2019.1573722

Hais M, Komprdova K, Ermakov N, Chytry M (2015) Modelling the last glacial maximum environments for a refugium of Pleistocene biota in the Russian Altai Mountains, Siberia. Palaeogeogr Palaeoclimatol Palaeoecol 438:135–145. https://doi.org/10.1016/j.palaeo.2015.07.037

Hantemirova EV, Heinze B, Knyazeva SG, Musaev AM, Lascoux M, Semerikov VL (2017) A new Eurasian phylogeographical paradigm? Limited contribution of southern populations to the recolonization of high latitude populations in *Juniperus communis* L. (Cupressaceae). Journal of Biogeography 44(2):271–282. https://doi.org/10.1111/jbi.12867

Hao Q, de Lafontaine G, Guo D, Gu H, Hu FS, Han Y, Song Z, Liu H (2018) The critical role of local refugia in postglacial colonization of Chinese pine: joint inferences from DNA analyses, pollen records, and species distribution modeling. Ecography 41(4):592–606. https://doi.org/10.1111/ecog.03096

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature 405:907–913. https://doi.org/10.1038/35016000

Hortal J, Jimenez-Valverde A, Gomez JF, Lobo JM, Baselga A (2008) Historical bias in biodiversity inventories affects the observed environmental niche of the species. Oikos 117(6):847–858. https://doi.org/10.1111/j.0030-1299.2008.16434.x

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9(5):1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x

Kalis AJ, Merkt J, Wunderlich J (2003) Environmental changes during the Holocene climatic optimum in central Europe - human impact and natural causes. Quatern Sci Rev 22(1):33–79. https://doi.org/10.1016/S0277-3791(02)00181-6

Karger DN, Conrad O, Bohner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, Kessler M (2017) Climatologies at high resolution for the earth's land surface areas. Scientific Data 4(170122). https://doi.org/10.1038/sdata.2017.122

Kendall MG, Stuart A (1976) The advanced theory of statistics. Volume 3: Design and analysis, and time-series. Charles Griffin, New York

Kharuk VI, Im ST, Petrov IA, Golyukov AS, Ranson KJ, Yagunov MN (2017a) Climate-induced mortality of Siberian pine and fir in the Lake Baikal Watershed, Siberia. For Ecol Manage 384:191–199. https://doi.org/10.1016/j.foreco.2016.10.050

Kharuk VI, Im ST, Petrov IA, Dvinskaya ML, Fedotova EV, Ranson KJ (2017b) Fir decline and mortality in the southern Siberian Mountains. Reg Environ Change 17:803–812. https://doi.org/10.1007/s10113-016-1073-5

Kind NV (1974) Late quaternary geochronology according to isotopes data. Publishing Office «Nauka», Moscow. Retrieved February 5, 2022, from https://www.geokniga.org/bookfiles/geokniga-2571974kindquaternary.pdf

Kobe F, Leipe C, Shchetnikov AA, Hoelzmann P, Gliwa J, Olschewski P, Goslar T, Wagner M, Bezrukova EV, Tarasov PE (2022) Not herbs and forbs alone: pollen-based evidence for the presence of boreal trees and shrubs in Cis-Baikal (Eastern Siberia) derived from the Last Glacial Maximum sediment of Lake Ochaul. J Quat Sci 37(5):868–883. https://doi.org/10.1002/jqs.3290

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15(5):1179–1191. https://doi.org/10.1111/1755-0998.12387

Krivonogov SK (1988) Stratigraphy and paleogeography of the Lower Irtysh river in the last glaciation (by carpological data). Nauka Press, Novosibirsk. Retrieved February 3, 2022, from https://www.geokniga.org/bookfiles/geokniga-stratigrafiya-i-paleogeografiya-nizhnego-priirtyshya-v-epohu-poslednego-oledene.pdf

Krutovsky KV, Politov DV, Altukhov YuP (1988) Genetic variability of Siberian stone pine *Pinus sibirica* Du Tour. Rep. II. Allozyme variability levels in the natural population of Western Sayan. Russ J Genet 24:118–124

Krutovsky KV, Politov DV, Altukhov YuP, Milutin LI, Kuznetsova GV, Iroshnikov AI, Vorobyov VN, Vorobyova NA (1989) Genetic variability of Siberian stone pine *Pinus sibirica* Du Tour. Rep. IV. Genetic diversity and the degree of genetic differentiation between populations. Russ J Genet 25:2009–2032

Krylov GV (1961) Forests of Western Siberia History of study, types of forests, zoning, ways of using and improving. Academy of Sciences of the USSR, Moscow

Krylov GV, Talantsev NK, Kozakova NF (1983) Kedr [Cedar Pine]. Lesnaya promyshlennost, Moscow

Kukla GJ, Bender ML, de Beaulieu J-L, Bond G, Broecker WS, Cleveringa P, Gavin JE, Herbert TD, Imbrie J, Jouzel J, Keigwin LD, Knudsen K-L, McManus JF, Merkt J, Muhs DR, Müller H, Poore RZ, Porter SC, Seret G, Winograd IJ (2002) Last interglacial climates. Quaternary Res 58(1):2–13. https://doi.org/10.1006/qres.2001.2316

Laukhin SA (2011) "Warm" stages in the West Siberian Late Pleistocene. Quatern Int 241(1–2):51–67. https://doi.org/10.1016/j.quaint.2011.05.022

Laukhin SA, Pushkar BC, Cherepanova MV (2012) Current state of reconstruction of the environment in the north of Siberia in Karginsk time (late Pleistocene). Bulletin of the Moscow Society Of Naturalists 87(6):37–48. https://www.researchgate.net/publication/306251018

Laukhin SA, Pushkar BC, Cherepanova MV (2015) Experience of the correlation of natural events in the Pleistocene Karginian time (analogs of MIS−3) from the Ob region to the Okhotsk Sea. Bulletin of the Moscow Society Of Naturalists 90(2):23−34. https://elibrary.ru/item.asp?id=23865548

Lisiecki LE, Raymo ME (2005) A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. Paleoceanography 20(2):PA1003. https://doi.org/10.1029/2004PA001071

Lisitsyna OV, Giesecke T, Hicks S (2011) Exploring pollen percentage threshold values as an indication for the regional presence of

major European trees. Rev Palaeobot Palynol 166(3–4):311–324. https://doi.org/10.1016/j.revpalbo.2011.06.004

Liu C, Newell G, White M (2016) On the selection of thresholds for predicting species occurrence with presence-only data. Ecol Evol 6(1):337–348. https://doi.org/10.1002/ece3.1878

Lozhkin A, Anderson P, Matrosova T, Minyuk P (2007) The pollen record from El'gygytgyn Lake: implications for vegetation and climate histories of northern Chukotka since the late middle Pleistocene. J Paleolimnol 37:135–153. https://doi.org/10.1007/s10933-006-9018-5

NEEM community members (2013) Eemian interglacial reconstructed from a Greenland folded ice core. Nature 493:489–494. https://doi.org/10.1038/nature11789

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J Mol Evol 19:153–170. https://doi.org/10.1007/BF02300753

Neishtadt MI (1983) To the issue of some concepts of the Holocene stratigraphy. Izvestiia AN SSSR 2:103–108

Oosterhout C, Hutchinson WF, Wills Derek PM, Shipley P (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4(3):535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x

Oreshkova NV, Sedelnikova TS, Pimenov AV, Efremov SP (2014) Analysis of genetic structure and differentiation of the bog and dry land populations of *Pinus sibirica* Du Tour based on nuclear microsatellite loci. Russ J Genet 50:934–941. https://doi.org/10.1134/S1022795414090105

Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A (2006) Simulating arctic climate warmth and icefield retreat in the Last Interglaciation. Science 311(5768):1751–1753. https://doi.org/10.1126/science.11208

Parks DH, Mankowski T, Zangooei S, Porter MS, Armanini DG, Baird DJ, Langille MGI, Beiko RG (2013) GenGIS 2: geospatial analysis of traditional and genetic biodiversity, with new gradient algorithms and an extensible plugin framework. PloS ONE 8:e69885. https://doi.org/10.1371/journal.pone.0069885

Peakall R, Smouse PE (2006) GenAlEx V6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6(1):288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190(3–4):231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Politov DV, Krutovsky KV, Altukhov YuP (1992) Characteristics of the gene pools of subsect. Cembrae populations by a set of isozyme loci. Russ J Genet 28:93–114

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2,1):945–959. https://doi.org/10.1093/genetics/155.2.945

R Core Development Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org/

Raymond M, Rousset F (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. J Heredity 86(3):248–249. https://doi.org/10.1093/oxfordjournals.jhered.a111573

Reuther AU, Herget J, Ivy-Ochs S, Borodavko P, Kubik PW, Heine K (2006) Constraining the timing of the most recent cataclysmic flood event from ice-dammed lakes in the Russian Altai Mountains, Siberia, using cosmogenic in situ 10Be. Geology 34(11):913–916. https://doi.org/10.1130/G22755A.1

Rossetto M, Kooyman R (2021) Conserving refugia: what are we protecting and why? Diversity 13(2):1–10. https://doi.org/10.3390/d13020067

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol Ecol Resour 8(1):103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Schulte L, Meucci S, Stoof-Leichsenring KR, Heitkam T, Schmidt N, von Hippel B, Andreev AA, Diekmann B, Biskaborn BK, Wagner B, Melles M, Pestryakova LA, Alsos IG, Clarke C, Krutovsky KV, Herzschuh U (2022) *Larix* species range dynamics in Siberia since the Last Glacial captured from sedimentary ancient DNA. Communications Biology 5(570). https://doi.org/10.1038/s42003-022-03455-0

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol Lett 9(5):615–629. https://doi.org/10.1111/j.1461-0248.2006.00889.x

Semerikov VL, Semerikova SA, Polezhaeva MA, Kosintsev PA, Lascoux M (2013) Southern montane populations did not contribute to the recolonization of West Siberian Plain by Siberian larch (*Larix sibirica*): a range-wide analysis of cytoplasmic markers. Mol Ecol 22(19):4958–4971. https://doi.org/10.1111/mec.12433

Semerikov VL, Semerikova SA, Putintseva YA, Oreshkova NV, Krutovsky KV (2019) Mitochondrial DNA in Siberian conifers indicates multiple postglacial colonization centers. Can J for Res 49(8):875–883. https://doi.org/10.1139/cjfr-2018-0498

Semerikov VL, Semerikova SA, Putintseva YA, Tarakanov VV, Tikhonova IV, Vidyakin AI, Oreshkova NV, Krutovsky KV (2018) Colonization history of Scots pine in Eastern Europe and North Asia based on mitochondrial DNA variation. Tree Genetics & Genomes 14(8). https://doi.org/10.1007/s11295-017-1222-0

Semerikova SA, Isakov IY, Semerikov VL (2020) Chloroplast DNA variation shed light on the history of lime tree (*Tilia cordata* s.l.) in the eastern part of the range. Russ J Genet 56:192–203. https://doi.org/10.1134/S1022795420020118

Semerikova SA, Isakov IY, Semerikov VL (2021) Chloroplast DNA variation and phylogeography of pedunculate oak *Quercus robur* L. in the eastern part of the range. Russ J Genet 57:47–60. https://doi.org/10.1134/S1022795421010130

Sheinkman V, Sedov S, Shumilovskikh LS, Bezrukova E, Dobrynin D, Timireva S, Rusakov A, Maksimov F (2021) A multiproxy record of sedimentation, pedogenesis, and environmental history in the north of West Siberia during the late Pleistocene based on the Belaya Gora section. Quatern Res 99:204–222. https://doi.org/10.1017/qua.2020.74

Sher AV, Kuzmina SA, Kuznetsova TV, Sulerzhitsky LD (2005) New insights into the Weichselian environment and climate of the East Siberian Arctic, derived from fossil insects, plants, and mammals. Quatern Sci Rev 24(5–6):533–569. https://doi.org/10.1016/j.quascirev.2004.09.007

Shuvaev DN, Ibe AA (2021) Genetic structure and postglacial recolonization of *Pinus sibirica* Du Tour in the West Siberian Plain, inferred from nuclear microsatellite markers. Silvae Genetica 70:99–107. https://doi.org/10.2478/sg-2021-0008

Stokland JN, Halvorsen R, Stoa B (2011) Species distribution modelling—effect of design and sample size of pseudo-absence observations. Ecol Model 222(11):1800–1809. https://doi.org/10.1016/j.ecolmodel.2011.02.025

Swets JA (1988) Measuring the accuracy of diagnostic systems. Science 240(4857):1285–1293. https://doi.org/10.1126/science.3287615

Takezaki N, Nei M, Tamura K (2010) POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. Mol Biol Evol 27(4):747–752. https://doi.org/10.1093/molbev/msp312

Tarasov PE, Volkova VS, Webb T, Guiot J, Andreev AA, Bezusko LG, Bezusko TV, Bykova GV, Dorofeyuk NI, Kvavadze BV, Osipova IM, Panova NK, Sevastyanov DV (2000) Last glacial maximum biomes reconstructed from pollen and plant macrofossil data from

northern Eurasia. J Biogeogr 27(3):609–620. https://doi.org/10.1046/j.1365-2699.2000.00429.x

Tollefsrud MM, Latalowa M, van der Knaap WO, Brochmann C, Sperisen C (2015) Late Quaternary history of North Eurasian Norway spruce (*Picea abies*) and Siberian spruce (*Picea obovata*) inferred from macrofossils, pollen and cytoplasmic DNA variation. J Biogeogr 42(8):1431–1442. https://doi.org/10.1111/jbi.12484

VanDerWal J, Shoo LP, Graham C, Williams SE (2009) Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? Ecol Model 220(4):589–594. https://doi.org/10.1016/j.ecolmodel.2008.11.010

Velichko AA, Timireva SN, Kremenetski KV, MacDonald GM, Smith LC (2011) West Siberian Plain as a late glacial desert. Quatern Int 237(1–2):45–53. https://doi.org/10.1016/j.quaint.2011.01.013

Veloz SD (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. J Biogeogr 36(12):2290–2299. https://doi.org/10.1111/j.1365-2699.2009.02174.x

Volkova VS (2002) Palynological and radiocarbon data from Late Quaternary deposits of the Western Siberian Plain. In: Anderson PM and Lozhkin AV (Eds.) Late Quaternary vegetation and climate of Siberia and the Russian Far East (palynological and radiocarbon database). North East Science Center, Far East Branch, Russian Academy of Sciences, Magadan, pp 196–257. Retrieved February 12, 2022, from https://search.rsl.ru/ru/record/01001850460

Volkova VS, Babushkin AE (2000) Unified regional stratigraphic scheme of quaternary deposits of West Siberian Plain. SNIIGGIMS, Novosibirsk

Volkova VS, Arkhipov SA, Babushkin AE, Kulkova IA, Guskov SA, Kuzmina OB, Levchuk LK, Mikhailova IV, Sukhorukova SS (2002) Stratigraphy of oil and gas basins of Siberia. Cenozoic of Western Siberia. Publishing house of SB RAS GEO, Novosibirsk. Retrieved February 10, 2022, from https://www.geokniga.org/bookfiles/geokniga-cenozoicstratigraphywestsiberia2002.pdf

Williams JW, Shuman BN, Webb TIII, Bartlein PJ, Leduc PL (2004) Late-Quaternary vegetation dynamics in North America: scaling from taxa to biomes. Ecol Monogr 74(2):309–334. https://doi.org/10.1890/02-4045