

ECOGRAPHY

Research article

Biogeography of larches in eastern Siberia – using single nucleotide polymorphisms derived by genotyping by sequencing

Sarah Haupt^{1,2}, Nadine Bernhardt³, Stefanie Killing^{1,2}, Stefano Meucci^{1,2}, Ulrike Herzschuh^{1,2,4}, Evgenii S. Zakharov⁵, Dörte Harpke⁶, Luidmila A. Pestryakova⁵ and Stefan Kruse¹

¹Polar Terrestrial Environmental Systems, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany

²Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany

³Julius Kühn-Institut, Federal Research Centre for Cultivated Plants, Institute for Resistance Research and Stress Tolerance, Quedlinburg, Germany

⁴Institute of Environmental Sciences and Geography, University of Potsdam, Potsdam, Germany

⁵Institute of Natural Sciences, North-Eastern Federal University of Yakutsk, Yakutsk, Russia

⁶Leibniz Institute of Plant Genetics and Crop Plant Research Gatersleben, Stadt Seeland, Germany

Correspondence: Sarah Haupt (sarah.haupt@awi.de), Stefan Kruse (stefan.kruse@awi.de)

Ecography

2024: e07092

doi: [10.1111/ecog.07092](https://doi.org/10.1111/ecog.07092)

Subject Editor: Wolf L. Eiserhardt

Editor-in-Chief:

Jens-Christian Svenning

Accepted 3 April 2024



The present distribution of Siberian boreal forests that are dominated by larches (*Larix* spp.) is influenced, to an unknown extent, by glacial history. Knowing the past treeline dynamics can improve our understanding of future treeline shifts under changing climate. Here, we study patterns in the genetic variability of Siberian *Larix* to help unravel biogeographic migration routes since the Last Glacial Maximum (LGM).

We infer the spatial distribution and the postglacial demographic history of *Larix* using genome-wide single nucleotide polymorphisms (SNPs) derived through genotyping by sequencing (GBS) from 130 individuals sampled across eastern Siberia.

Our analysis gives statistical support for two or three clusters, spanning from western to eastern Siberia. These clusters reveal a genetic structure influenced by isolation resulting from geographical distance, barriers imposed by geographic features, and distinct glacial histories. Assuming three clusters, our demographic inference indicates that the common ancestor of the current *Larix* populations existed in northeast Siberia well before the LGM. This suggests that *Larix* persisted in the northern region throughout previous glacials.

Our genetic studies suggest that *Larix* likely survived the cold LGM in northern refugia, enabling a fast colonization of Siberia. Instead of complete repopulation from southern areas postglacially, the northernmost *Larix* expansion during the Holocene seems to have benefitted from refugial populations ahead of the treeline. Present-day migration is expected to be slow initially, due to the absence of current refugial populations in the far north, in contrast to the early-Holocene situation.

Keywords: boreal forests, demographic history, genotyping by sequencing (GBS), glacial refugia, deciduous larch, Last Glacial Maximum, phylogeography, Siberia, single nucleotide polymorphisms (SNPs)



www.ecography.org

© 2024 The Authors. Ecography published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Introduction

The extensive forests in eastern Siberia are dominated by deciduous larch species (*Larix* spp.). These unique forests mostly grow on continuous permafrost, to which the larches with their shallow roots are well adapted (Kajimoto 2010). Larches also stabilize the permafrost (Zhang et al. 2011). Changing climate forces species worldwide to migrate (Pearson et al. 2013) and exceptional attention is paid to the high northern latitudes since climate change here is faster and more severe than in other regions (IPCC 2022). Siberian boreal forests are expected to expand northwards in the course of modern global warming (MacDonald et al. 2008, Andreev et al. 2011). This tundra–taiga transition will be accompanied by an albedo decrease, which in turn will raise regional temperatures (Bonan 2008). This additional warming can offset the negative forcing that is expected from carbon sequestration (Zhang et al. 2013). However, the processes of the treeline ecotone transition that would lead to different climate feedbacks, as well as timing, are still poorly understood (Harsch et al. 2009).

The speed and spatial pattern of a species' response to climate change are determined by several processes. The biogeographic history of *Larix* in Eurasia has been predominantly shaped by the alternating glacial and interglacial periods over the last three million years (Hewitt 2000). The trees either persisted in northern refugia or they survived in southern areas under less harsh climatic conditions from where they spread postglacially (Bennett and Provan 2008). Relict forest stands from previous warm phases, surviving in refugia ahead of the current treeline, can accelerate treeline migration rates as estimated from observational studies (Stewart and Lister 2001, Holtmeier and Broll 2007, MacDonald et al. 2008, Väiliranta et al. 2011) and modeling (Kruse et al. 2019). Radiocarbon-dated macrofossil findings also indicate rapid expansion of boreal *Larix* forests at the end of the Late Glacial and the beginning of the Holocene (MacDonald et al. 2000, Andreev et al. 2022). However, modeling the current response of Siberian larch forests reveals a pronounced time lag in their response to ongoing climate change (Kruse et al. 2016). To predict the future climate response under the current global warming, we need more reliable estimates of species range-expansion rates to implement in simulation models. Therefore, understanding the influence of relict trees from earlier forest extents on migration dynamics is of great importance.

Fossil records show that *Larix* became a well-established forest constituent in northeastern Russia by the late Oligocene (26.5–24 million years ago (Ma)) (LePage and Basinger 1995). Recent Siberian larch species, especially *L. gmelinii*, probably formed as an adaptation to increasing continentality in the Pleistocene (Abaimov 2010). *Larix* formed forests in Siberia throughout the Holocene (Cao et al. 2019). Paleoecological studies indicate the existence of refugia for *Larix* in northern areas during the coldest phases of the latest Pleistocene, particularly the Last Glacial Maximum (LGM; approximately 21 000 years before present (ka BP)), even when *Larix* density was likely relatively

low (Cao et al. 2020). Pollen and macrofossil records suggest the persistence of local *Larix* populations in northern Asia in general (Schulte et al. 2022a) and within western Beringia (Brubaker et al. 2005, Lozhkin et al. 2018), in the western foreland of the Verkhoyansk Mountains (Tarasov et al. 2009, Müller et al. 2010), and even on the Taymyr Peninsula (Binney et al. 2009). Although these previous studies reveal examples of Siberian refugial populations that persisted during the LGM, it remains unclear as to what extent and how these Siberian glacial refugial populations genetically contributed to postglacial recolonization and the modern genetic pool.

Genetic analyses have been used to reveal local dispersal patterns and population genetics and to infer the historical biogeography of *Larix* in Siberia (Herzschuh 2020). Araki et al. (2008) suggest that *Larix sukaczewii* and *L. sibirica* populations originated from migrants originating from multiple, genetically distinct refugia. Semerikov et al. (1999, 2013) conducted one of the few species-wide population genetic studies of *L. sibirica* and reveal, using cytoplasmic markers, that the southernmost populations of *L. sibirica* made a very limited contribution to the current populations of the central and northern parts of its range. In contrast, recent research has successfully differentiated between *L. sibirica* and *L. gmelinii* in glacial refugial populations by enriching sedimentary ancient DNA extracts for chloroplast genome sequences (Schulte et al. 2021, 2022b). These investigations highlight the existence of northern refugial populations during the LGM, primarily composed of *L. gmelinii*. *Larix sibirica*, however, recolonized from southern refugia. Yet, knowledge regarding the vast distribution of larches in eastern Siberia and other Siberian larch species remains limited.

Genetic markers previously used to study phylogeography were limited to a few informal sequences and mostly localized to plastids. However, recent advancements in genotyping by sequencing (GBS) have revolutionized this approach, allowing for the analysis of hundreds to thousands of nuclear loci at a low cost, providing deep insights into phylogeography as well as adaptation. GBS combines complexity reduction, sample multiplexing, and next-generation sequencing (NGS) methods for genotyping entire mapping populations (Wendler et al. 2014), making it applicable to high diversity, large-genome species like *Larix* spp. (Elshire et al. 2011). Moreover, the analysis of nuclear DNA offers other advantages, as it contains complementary information to plastid DNA due to different modes of inheritance (Petit et al. 2004). In many gymnosperms, plastid DNA is inherited from male pollen dispersed by wind, while mitochondrial DNA is normally inherited exclusively from the mother (Freeland 2005). Additionally, for land plants, plastid and nuclear genomes exhibit a mutation rate approximately 3- to 10-fold greater than the mitochondrial genome (Smith 2015). Therefore, biparentally inherited nuclear DNA from GBS analyses is expected to reveal higher levels of genetic structure among populations. To date, GBS has not yet been utilized to infer population structures and hence unravel the biogeography of *Larix*.

The overall aim of this study was to provide a better understanding of the importance of glacial relict trees ahead of the treeline on the postglacial migration rate of Siberian larches across the continent. Therefore, we examined geneecological interrelations of representatively sampled populations across eastern Siberia using single nucleotide polymorphisms (SNPs) derived by GBS. The specific objectives were 1) to reveal patterns in the genetic composition of Siberian larches (*Larix* spp.) by assessing the spatial distribution of SNPs using cluster analysis, and 2) to unravel the potential demographic history to test whether the initiation of the current populations can be dated before the LGM.

Materials and methods

Study genus

Larix is the most abundant genus in Eurasian boreal coniferous forests, covering 70% of the global forest biome (Abaimov 2010). Larches are well adapted to a harsh climate with winters below -40°C and short summers (Franz 1973). The natural distribution of Siberian *Larix* species extends from west to east, with overlapping and hybridization occurring in the boundary zones (Semerikov et al. 2007, Abaimov 2010). This especially applies to the Siberian *Larix* species. Following Bobrov (1972), this study recognizes three Siberian *Larix* species, *L. sibirica*, *L. gmelinii*, and *L. cajanderi*, acknowledging that some authors consider *L. cajanderi* a synonym and part of *L. gmelinii* (Borsch et al. 2020).

Study area and sample collection

The study covers arctic and subarctic regions in the Siberian forest–tundra transition zone, spanning from the Taymyr

Peninsula in the west to Chukotka in the Far East, and including the Lena Delta, Verkhoyansk Range, Kamchatka, Magadan, and southern Yakutsk (ca. $50\text{--}80^{\circ}\text{N}$; $100\text{--}170^{\circ}\text{E}$, Fig. 1).

For genetic analyses, fresh needles from trees in natural populations were taken during summer fieldwork (2011–2021). Needle samples were collected to cover a wide distribution area of Siberia, restricted by accessibility. Needle samples were dried on silica gel during fieldwork and stored at 4°C . In total, 148 individuals were processed (Supporting information). Samples from the same sampling location were usually spaced ~ 25 m apart from each other to mitigate biases in genetic analyses. Additionally, closely related samples were excluded during subsequent analyses (section ‘Genetic relatedness’).

DNA extraction

Genomic DNA was extracted from 40 to 80 mg of dried needles after grinding them for 2 min in a FastPrep[®]-24 instrument (MP Biomedicals, USA) using the Invisorb[®] Spin Plant Mini Kit (Stratec Molecular, Germany), DNeasy[®] Plant Mini or Pro Kit (Qiagen, Germany), following the supplier’s protocol with preheating of buffers. DNA extractions of low quality or low concentration were repeated and those with a distinct band, high molecular weight of at least $20\text{ ng }\mu\text{l}^{-1}$, and minimal smearing in 1% agarose gel electrophoresis were selected for sequencing. All DNA extractions were standardized to a concentration of $20\text{ ng }\mu\text{l}^{-1}$.

Genotyping by sequencing (GBS)

GBS followed Wendler et al. (2014) to obtain genome-wide SNPs. Genomic DNA (200 ng) was digested with rare-cutting enzyme PstI-HF (recognition site: CTGCA’G) and

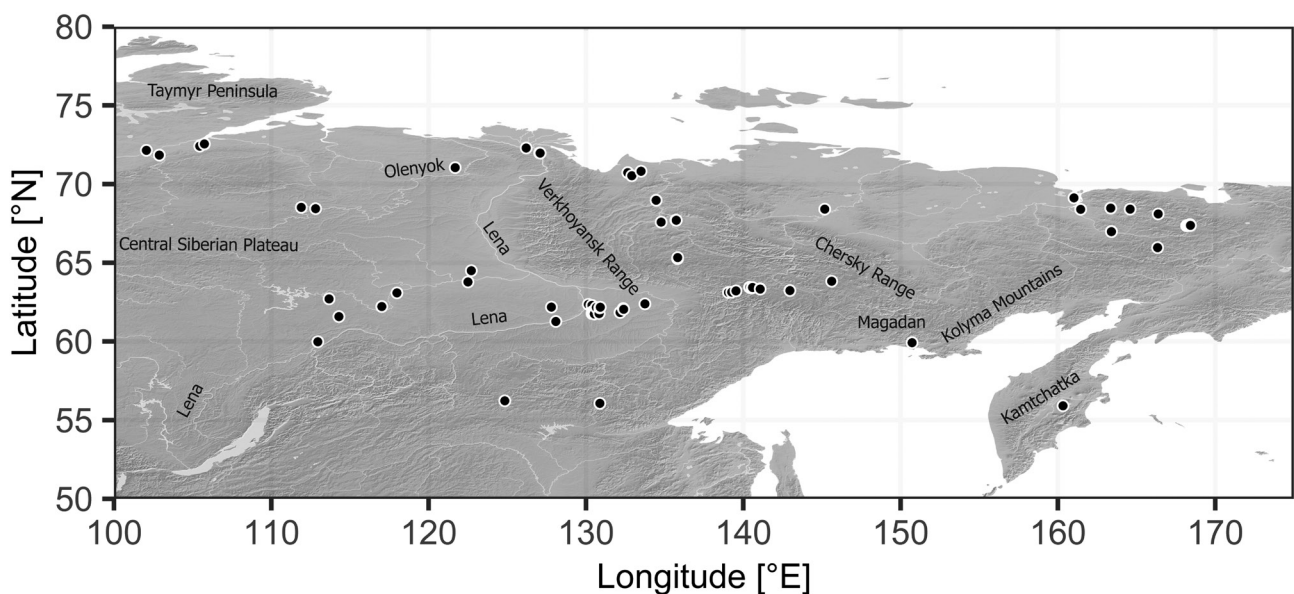


Figure 1. Map of the study area. Black dots represent 130 individuals at the sampling locations.

frequent cutting enzyme MspI (recognition site: C'CGG). Sequencing was performed on an Illumina HiSeq 2000 and NovaSeq 6000, generating 100–120 bp single-end reads.

Demultiplexing, de novo assembly, and variant calling

Barcoded Illumina reads were demultiplexed using the Casava pipeline 1.8 (Illumina).

The sequence data have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB71740 (<https://www.ebi.ac.uk/ena/data/view/PRJEB71740>), using the data brokerage service of the German Federation for Biological Data (GFBio, Diepenbroek et al. 2014).

Subsequently, de novo assembly and variant calling were performed using ipyrad (ver. 0.9.84) (Eaton and Overcast 2020). A maximum of five low-quality ($Q < 20$) bases were allowed in a read. Consensus base calling was also part of the ipyrad pipeline with a minimum depth set at six. A threshold of at least 85% sequence similarity was set to identify homologous sequences, and thus cluster together. The minimum number of samples that must have data at a given locus was set to 80%. Heterozygous sites were allowed for a maximum of 50% of the samples. See Supporting information for details of the ipyrad parameters.

Genetic quality assessment

SNPs were linkage disequilibrium (LD) pruned with a window of 50 SNPs, a step size of 20 markers, and r^2 threshold 0.05. To minimize sequencing and assembly errors, the SNPs showing severe deviation from Hardy Weinberg equilibrium (HWE) ($p < 0.05$) and minor allele frequency (MAF) lower than 5%, as well as SNP markers with missing data (MISS) above 20%, were discarded from further analysis. LD, HWE, MAF, and MISS of the SNP dataset were inferred using the software PLINK (Purcell et al. 2007).

Genetic relatedness

Due to the influence of genetic relatedness on SNP association analysis (Gross et al. 2017), samples exhibiting close genetic relationships were systematically excluded. Kinship coefficients were calculated using the *snpGdsIBDML*E function from the 'SNPRelate' package (ver. 1.32.2) (Zheng et al. 2012) with parameter settings 'kinship=TRUE' and 'auto-some.only=FALSE', implemented in R (ver. 4.2.1) (www.r-project.org). Samples were excluded pairwise if their kinship exceeded 0.125 (second-degree relatives), retaining samples with fewer missing loci from each of these pairs. Out of the initial 148 samples, 18 were excluded due to close genetic relatedness (Supporting information).

The remaining 130 samples experienced demultiplexing, de novo assembly, variant calling (section 'Demultiplexing, de novo assembly, and variant calling'), and genetic quality assessment (section 'Genetic quality assessment') with the same parameters as in the initial analysis.

Individual-based genetic structure and admixture

Genetic structure reveals spatial distribution patterns, with individuals from the same population exhibiting more similar genotypes (Cornuet et al. 1999). Here we used the Bayesian cluster assignment to serve as an illustration of the distribution of ancestry coefficients. We used optimized versions of principal component analysis (PCA) and sparse non-negative matrix factorization (sNMF) algorithms (Frichot et al. 2014) as implemented in the 'LEA' R package (ver. 2.8.0) (Frichot and François 2015, Francois 2016). The acquired SNP data are linked to geographical information, and patterns are inferred through the analysis of ancestry coefficients and proportions of admixture (cluster analysis) (Frichot and François 2015).

Ancestry coefficients were estimated for 1–10 ancestral clusters (K) using 100 repetitions for each K. The final number of clusters (K) was selected by choosing the K value with the lowest cross-entropy.

The relationship of the clusters of assigned individuals was visualized through discriminant analysis of principal components (DAPC) (*dapc* function from 'adegenet' R package (ver. 2.1.10) (Jombart 2008) with parameter settings 'pca.select='percVar', 'perc.pca=90', and 'n.da=5').

Genetic diversity and differentiation within/among clusters

To explore differences between inferred clusters, standard population genetic statistics were calculated in R using default settings. For each cluster, the intrapopulation genetic diversity was evaluated by calculating the observed (H_O) and expected heterozygosity (H_E) (Nei 1987) (*summary* function from 'adegenet' R package (ver. 2.1.10) (Jombart (2008))) and the fixation index (F_{IS}) (Wright 1965) (*wc* function from 'hierfstat' R package (ver. 0.5.11) (Goudet and Jombart 2022)). Allele set diversity within each cluster was assessed by allelic richness and private alleles.

The genetic differentiation among the clusters was assessed by calculating the pairwise population differentiation measures: Weir and Cockerham's F_{ST} (Weir and Cockerham 1984) (*wc* function from 'hierfstat' R package), Jost's D (Jost 2008) (*pairwise_D* function from 'mmod' R package (ver. 1.3.3) (Winter 2012)), Hedrick's G_{ST} (Hedrick 2005) (*pairwise_Gst_Hedrick* function from 'mmod' R package), and Nei's G_{ST} (Nei 1973) (*pairwise_Gst_Nei* function from 'mmod' R package), and the significance of these values was tested by comparison with the 95% confidence interval (CI) under the null hypothesis of no genetic differentiation, acquired from 100 randomizations of the assigned cluster. Subsequently, we assessed whether differentiation among *Larix* clusters across geographic regions resulted from isolation by distance by conducting a Mantel test using the *mantel.randtest* function from the 'ade4' R package (ver. 1.7.22) (Dray and Dufour 2007). This analysis utilized the pairwise Euclidean genetic distance and compared it with a geographical distance – both for the complete dataset and separately for each cluster. The significance was assessed through 1000 permutations.

Biogeographic inference using approximate Bayesian computation (ABC)

The potential demographic history of the three different larch populations in Siberia derived from the Bayesian cluster assignment was inferred via complex evolutionary scenarios using the model-based approach approximate Bayesian computation (ABC), implemented in DIYABC Random Forest ver. 1.0 (Collin et al. 2021), an extended iteration of DIYABC ver. 2.1.0 (Cornuet et al. 2014). This method integrates supervised machine learning to select the most likely demographic scenario and determine parameter robustness through backward simulations. Following the recommendations in the manual, we conducted 1000 simulations per scenario for scenario choice and 100 000 for parameter estimations for the scenario based on the highest posterior probability. The models were evaluated through logistic regression analysis, wherein the relative posterior probabilities were compared against 1% of the simulated datasets closest to the observed data using logistic regression. Subsequently, the scenario with the highest posterior probability was identified as the most realistic one. Additionally, an examination was conducted to ensure that the observed data were situated centrally within the posterior distribution in the principal component analysis (PCA).

To keep the scenarios in the ABC simple, three populations were defined based on the results of the cluster analysis and the admixture plots: Pop1 (western Yakutia), Pop2 (eastern Yakutia), and Pop3 (Chukotka). Alternative scenarios including the predefined three distinct genetic groups for the estimation of the population demographic history were constructed. We estimated the parameters for divergence time ($t\#$), measured by generation time, and the effective population size ($N\#$) of the corresponding populations. The three examined scenarios are viewed backward in time (Fig. 2).

Scenario 1: Hierarchical split model; Pop1 merged with Pop2 at t_1 , and then they merged with Pop3 at t_2 .

Scenario 2: Isolation with admixture model; Pop2 was created by an admixture of Pop1 and Pop3 at t_1 , and then Pop1 merged with Pop3 at t_2 .

Scenario 3: Simple split model; all three populations Pop1, Pop2, and Pop3 split at the same time, namely t_2 .

Based on the Bayesian cluster assignment, Pop2 is identified as an admixed population, supporting the likelihood of scenario 2. We hypothesized a decrease in effective population size at t_3 from N_b (the ancestral population) to N_a (population before the divergence during t_3-t_2), attributed to a bottleneck event, followed by another decline at t_2 when, in all scenarios, the ancestral population underwent a split.

To obtain the calibrated years before present (BP), the number of generations has to be multiplied by a given generation time. Although larches can initiate reproduction at 20–30 years of age, they do not attain their maximum reproductive capacity until later, with the potential for continued reproduction until 200 years of age. Consequently, an average age of 100 years was selected as the generation time (Semerikov et al. 2013).

Results

Genetic quality assessment

The DNA from our processed dataset of 130 samples returned a maximum of 4 822 890 and a minimum of 479 617 reads with an average of 1 631 356 raw reads per sample.

The raw reads were assembled into 8698 filtered loci among the 130 individuals with an average of 7572, a maximum of 8605, and a minimum of 1887 loci per sample. A total of 45 004 SNPs with 18.3% missing sites was detected by consensus base calling. An average of 0.01447 and 0.003394 estimated heterozygosity and error rate was detected, respectively. Further statistical results are provided in the Supporting information.

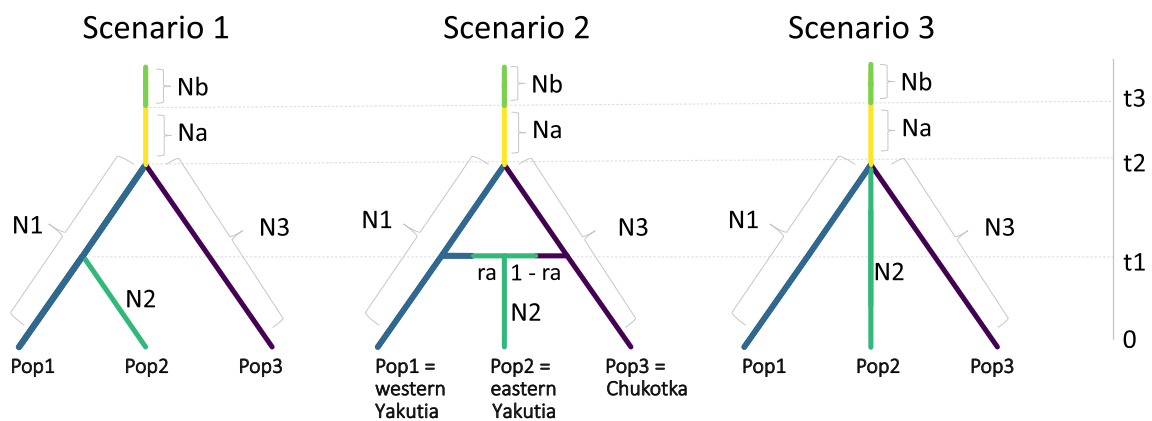


Figure 2. The three constructed scenarios tested with approximate Bayesian computation. In these scenarios, $t\#$ represents the time scale in number of generations and $N\#$ represents the effective population size of the corresponding clusters during the relevant time period (e.g. $0-t_1$, t_1-t_2). Pop1 (western Yakutia), Pop2 (eastern Yakutia), and Pop3 (Chukotka); 'a' (before divergence), 'b' (ancestral population), and 'ra' (relative assignment probability).

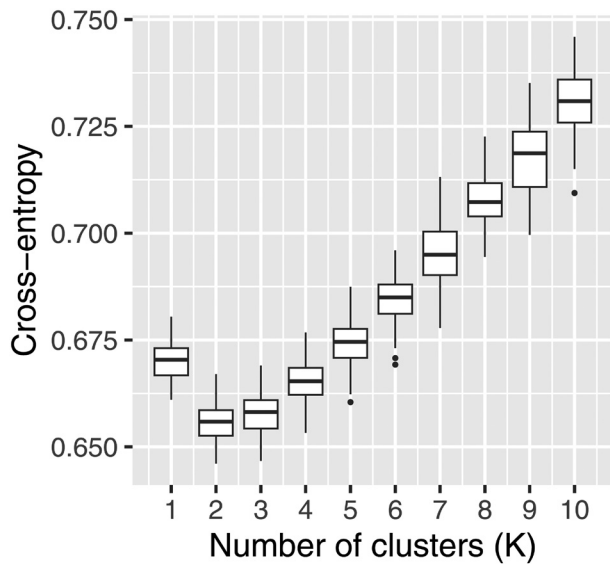


Figure 3. Cluster size evaluation: cross-entropy plot for the number of clusters $K=1-10$ with 100 repetitions for randomization. Value of the cross-entropy criterion as a function of the number of clusters in sparse non-negative matrix factorization (sNMF).

Individual-based genetic structure and admixture

The cross-validation with sNMF showed that the number of clusters (K) with the lowest cross-entropy was $K=2$ in 89 out of 100 repetitions. But the difference in cross-entropy between $K=2$ and $K=3$ is negligible (Fig. 3). Assignment of individuals to the three clusters remains stable even in repeated cluster analyses. Considering higher number of clusters ($K=4$ or $K=5$), the assignment of individuals to clusters 4 and 5 exhibits variability across different runs.

The two or three distinct clusters revealed by the analysis of ancestry coefficients and proportions of admixture characterize a genetic structure extending across the geographical area from western to eastern Eurasia. When two clusters are considered ($K=2$), Cluster 1 (Yakutia) is genetically distinct from Cluster 2 (Chukotka) (Fig. 4a). Cluster 1 runs from across the Taymyr Peninsula in the northern part of Siberia

through Yakutia to Bolshoi Toko in southern Yakutia. Cluster 2 spreads out further east in the region of Chukotka. Under three clusters ($K=3$), the Yakutian cluster is divided into a western and an eastern cluster (Fig. 4b). Between Cluster 1 (western Yakutia) and Cluster 2 (eastern Yakutia), the Lena River is located parallel to the Verkhoyansk Range.

Individuals in the core areas exhibit higher ancestry coefficients, indicating a clear affiliation with a single cluster (Fig. 4–6). In contrast, those at the peripheries demonstrate a more fragmented affiliation with multiple clusters, suggesting potential admixture or hybridization. Specifically, for $K=3$, the eastern Yakutian cluster in the center appears to represent an admixed region, displaying genetic characteristics indicative of contributions from both adjacent clusters.

The results of the discriminant analysis of principal components (DAPC) (Supporting information) match with the sNMF clustering. Clusters 1 and 2 are closely positioned and only separated on the second linear discriminant (LD2), indicating substantial spatial overlap, possibly suggestive of admixture regions. In contrast, Cluster 3 is situated at a greater distance to the other clusters with a pronounced gap on LD1, signifying distinct genetic differentiation.

Genetic diversity and differentiation within/among clusters

Chukotka exhibits the highest number of assignments, although the sizes of the clusters are comparable. The allelic richness shows minimal variation, approximately 0.4%, among the clusters, with similar numbers of alleles observed across clusters. Private alleles are identified exclusively within Chukotka. Observed heterozygosity (H_o) levels are lower than expected (H_e), and the highest inbreeding coefficients (F_{is}) are observed in the eastern Yakutian cluster (Table 1).

All four pairwise genetic difference measures indicate a very low level of genetic differentiation between western and eastern Yakutia (i.e. a high degree of genetic similarity); a moderate level of genetic differentiation between eastern Yakutia and Chukotka; and a relatively high genetic differentiation between western Yakutia and Chukotka (Table 2).

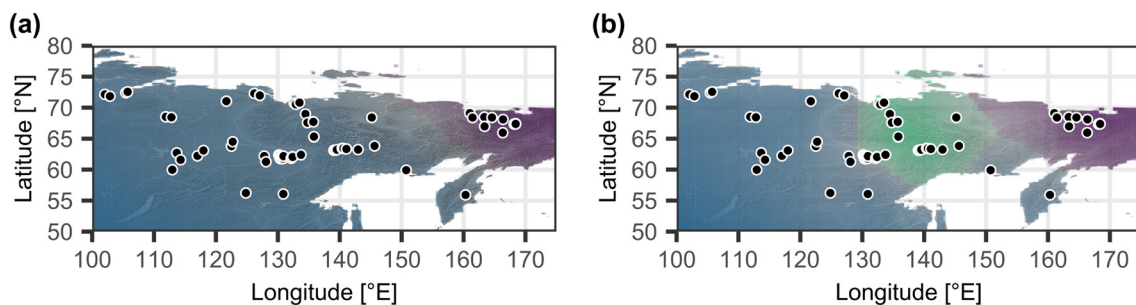


Figure 4. Interpolated values of ancestry coefficients for (a) two clusters ($K=2$) and (b) three clusters ($K=3$). Only the cluster with the maximal local contribution to ancestry is represented at each geographic point of the map. The results of the run with the lowest cross-entropy are shown (run64 for $K=2$ and run3 for $K=3$). The values were interpolated at the 130 sampling locations represented as black dots. The distinct colors signify individual clusters and the transparency of the representation corresponds to the magnitude of the ancestry coefficient values (blue = western Yakutia, green = eastern Yakutia, purple = Chukotka).

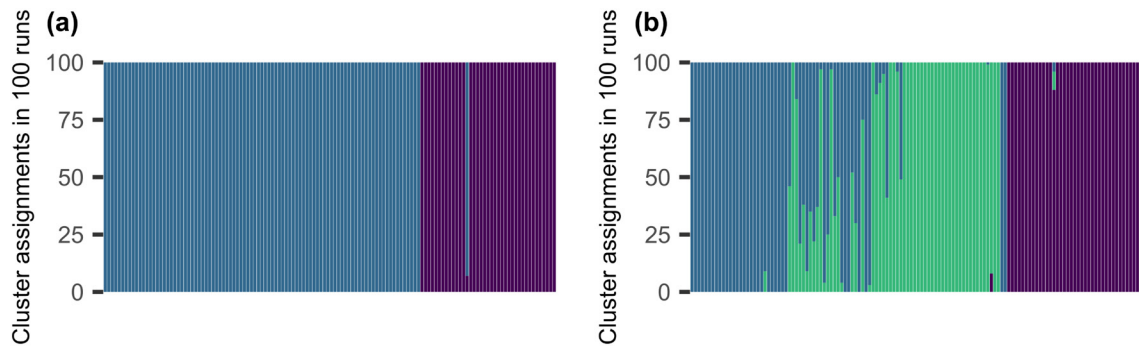


Figure 5. Bayesian cluster assignment derived from individual admixture coefficients using 100 repetitions of sparse non-negative matrix factorization (sNMF) computed in 'LEA', an R package for landscape and ecological association studies for (a) two clusters ($K=2$) and (b) three clusters ($K=3$). On the y-axis, the proportion of admixture per individual is shown and the different groups are seen in different colors (blue=western Yakutia, green=eastern Yakutia, purple=Chukotka). The individuals on the x-axis are sorted geographically by longitude from west to east.

The Mantel test demonstrates significant positive correlations between Euclidean genetic distance and geographical distance (complete dataset, $r=0.64$, $p=0.001$) in Cluster 1 ($r=0.56$, $p=0.001$) and Cluster 2 ($r=0.59$, $p=0.001$), and an even stronger relationship in Cluster 3 ($r=0.86$, $p=0.001$) (Supporting information).

Biogeographic inference using approximate Bayesian computation (ABC)

The ABC analysis returned the highest posterior probability for scenario 2 (posterior probability, 66.3%), the model with a relatively recent admixture event (Fig. 7). For this scenario, the median values of effective population size are given in Table 3.

The median value of the relative assignment probability (r_a) from Pop1 (western Yakutia) to Pop2 (eastern Yakutia) (hybrid) is 0.83 (90% CI, 0.41–0.97) and thus from Pop3 (Chukotka) to Pop2 (eastern Yakutia) is 0.17 ($1 - r_a$), suggesting that Pop1 makes a higher contribution to the hybrids than Pop3 (Table 3).

The median time of population decline (t_3) is 3530 (90% CI, 396–9271) generations ago. The median divergence time (t_2 : split of Pop1 and Pop3) is 204 (90% CI, 28–1400), and the admixture event that initiated Pop2 (hybrid) from parental

populations of Pop1 and Pop3, t_1 , is 46 (90% CI, 13–116) generations ago (Table 4). The calibrated time using 100 years as the generation time for the first observable larch colonization (t_3 , population Nb) is approximately 353 000 years BP (90% CI, 39 600–927 100 BP). The following divergence time, t_2 , can be dated to 20 400 years BP (90% CI, 2800–140 000 BP). The admixture event (t_1) can be dated to approximately 4600 years BP (90% CI, 1300–11 600 BP).

Discussion

Genetic structure of Eurasian larches is likely influenced by geographical factors

The GBS resulted in a total of 8698 variable SNPs among the 130 individuals, which is an appropriate genomic dataset for the research question and an acceptable number in comparison to other studies with Pinaceae (Eckert et al. 2010, Chen et al. 2013, Johnson et al. 2017). With GBS the complexity of the large genome of *Larix* was efficiently reduced for downstream analyses (Deschamps et al. 2012, Dong et al. 2019).

The cluster analysis based on SNPs revealed two or three geographically distinct genetic clusters in Eurasia.

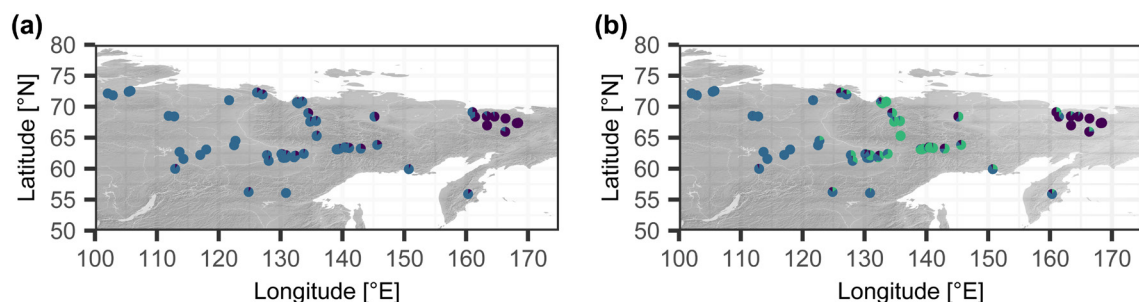


Figure 6. Pie charts for (a) two clusters ($K=2$) and (b) three clusters ($K=3$) showing the proportion of admixture for each individual. The results of the run with the lowest cross-entropy are shown (run64 for $K=2$ and run3 for $K=3$). The distinct colors signify individual clusters (blue = western Yakutia, green = eastern Yakutia, purple = Chukotka).

Table 1. Summary of standard population genetic statistics for three clusters in Siberia.

	Number of individuals	Number of alleles	Allelic richness	Number of private alleles	H _O	H _E	F _{IS}
Cluster 1/ western Yakutia	39	9230	1.99094	0	0.2336	0.2453	0.063
Cluster 2/ eastern Yakutia	41	9260	1.997412	0	0.2344	0.2482	0.07
Cluster 3/ Chukotka	50	9239	1.992882	2	0.2369	0.2475	0.0542

Considering two clusters ($K=2$), according to the taxonomic system of Bobrov (1972) of two *Larix* species within the modern recognized *L. gmelinii* distribution area, the Yakutian cluster in the west would be designated as *L. gmelinii*, while the Chukotka cluster in the Far East would be designated as *L. cajanderi* (Abaimov 2010). Further, the individuals in the core area of these clusters have an affiliation to exclusively one cluster, but at the periphery, individuals are admixed. In the contact zone of their geographic ranges, hybrids from *L. gmelinii* and *L. cajanderi* form transitional populations (Abaimov 2010, Isaev et al. 2010). This inference aligns with expectations, as cross-pollination occurs in these boundary areas (Baltunis et al. 1998). Considering three different clusters, these hybrid individuals form the eastern Yakutian cluster (Cluster 2), representing an admixed region in the center with its western border along the Lena River, characterized by genetic contributions from both Cluster 1 and Cluster 3.

Taxonomic discussions vary on the boundary between *L. gmelinii* and *L. cajanderi*, with some authors considering *L. cajanderi* a synonym or part of *L. gmelinii* (Borsch et al. 2020). Dylis (1961), who divided *L. dahurica* into a western and an eastern race, drew a dividing line between them at 120–123° E longitude, which Bobrov (1972) later described as two species: *L. gmelinii* and *L. cajanderi*. This line coincides with the boundary between our western Yakutian and eastern Yakutian clusters. The western Yakutian cluster (Cluster 1) corresponds to the distribution of *L. gmelinii*, limited to the east by the Lena River delta (Abaimov 2010). Trees from the Taymyr Peninsula are westernmost in our *L. gmelinii* cluster, aligning with Abaimov (2010). Notably, the hybrid individuals of Cluster 2 classified by Bobrov (1972) as *L. cajanderi* are genetically closer to *L. gmelinii* individuals from western Yakutia (Cluster 1) than to *L. cajanderi* individuals from Chukotka (Cluster 3). Further evidence for a separation has been found by molecular studies using both mitochondrial and chloroplast markers, which also reveal weak but visible

genetic differentiation between *L. gmelinii* and *L. cajanderi* (Polezhaeva et al. 2010).

Particularly striking are the topographical conditions like mountain ranges or river valleys along the cluster boundaries, especially as the positions of the samples were not considered in the cluster inference. The border between the western Yakutian cluster (Cluster 1) and the eastern Yakutian cluster (Cluster 2) closely follows the Lena River. Nevertheless, it can be assumed that at this point the Verkhoyansk Range serves as a stronger geographical barrier, which is responsible for the genetic differentiation between Cluster 1 and Cluster 2. For a lowland species like larch, mountains serve as a persistent geographical barrier to expansion. In addition to long-term geographical isolation, mountains can increase climatic gradients and thus facilitate population differentiation through environmental isolation (Li et al. 2019). Gene flow may occur between adjacent catchments, mountain ranges, or valleys, but less than between populations without landscape separation (Schnell and Priyadarshan 2012). Such separation may maintain genetic isolation, fostering genetic divergence (Rowe et al. 2017, Tsumura et al. 2020). Genetic deviations between Cluster 2 (eastern Yakutia) and Cluster 3 (Chukotka) may result from separation by the Chersky Range and the Kolyma Mountains, with varied microclimates at different elevations fostering regional climatic divergence. Genetic differentiation can also increase with geographical distance, even without apparent gene flow barriers. Limited gene flow between the population in Chukotka and the main distribution area within the large boundary area of central Yakutia may have spurred genetic drift during the glacial and inter-glacial cycles (Rousset 1999). A significant positive correlation between genetic and geographical distance provides strong support for isolation by distance, indicating increased genetic divergence as geographical distance grows among clusters. The genetic variability of *L. cajanderi* is also supported by observations showing that *L. cajanderi* is involved in hybridization with

Table 2. Pairwise genetic difference measures for *Larix* clusters in Siberia for Weir and Cockerham's F_{ST} , Jost's D , Hedrick's G_{ST} , and Nei's G_{ST} .

Pairwise Weir and Cockerham's F_{ST} values for clusters			Hedrick's G_{ST} values for clusters		
	Cluster 2/ eastern Yakutia	Cluster 3/ Chukotka		Cluster 2/ eastern Yakutia	Cluster 3/ Chukotka
Cluster 1/ western Yakutia	0.0013	0.0249*	Cluster 1/ western Yakutia	0.0031	0.0342*
Cluster 2/ eastern Yakutia		0.0155*	Cluster 2/ eastern Yakutia		0.0218*
Jost's D values for clusters			Nei's G_{ST} values for clusters		
	Cluster 2/ eastern Yakutia	Cluster 3/ Chukotka		Cluster 2/ eastern Yakutia	Cluster 3/ Chukotka
Cluster 1/ western Yakutia	0.0008	0.0088*	Cluster 1/ western Yakutia	0.013	0.0082*
Cluster 2/ eastern Yakutia		0.0056*	Cluster 2/ eastern Yakutia		0.0011*

Values marked with an asterisk (*) fall outside the 95th percentile of the bootstrapped distribution ($n=100$).

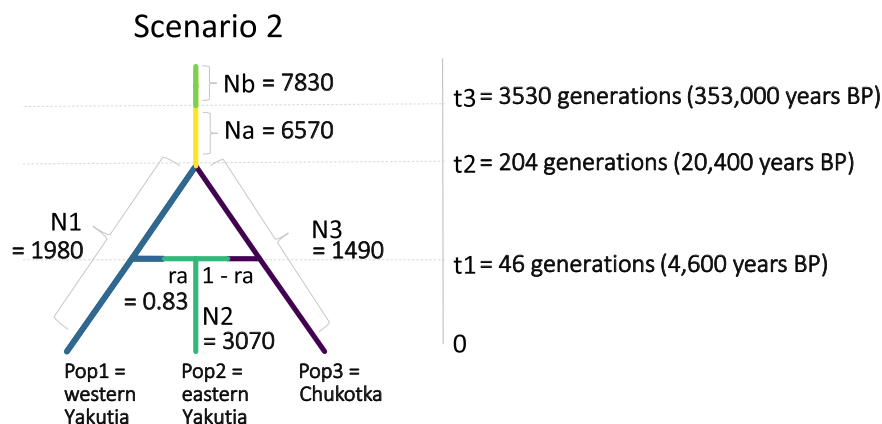


Figure 7. DIYABC-RF Scenario 2: events, time scale in number of generations, and effective population sizes of the corresponding clusters.

L. kamschatica and other *Larix* hybrids in the far northeast (Abaimov 2010).

The observed slight differences in allelic richness suggest subtle variations in the genetic diversity of the clusters. Private alleles in Chukotka may indicate unique genetic variations influenced by local factors like historical isolation, adaptive processes, or genetic drift. All clusters exhibit comparable mean observed heterozygosity values, suggesting similar proportions of heterozygous loci within clusters. Mean expected heterozygosity values are also close among clusters, indicating similar genetic diversity levels. Each cluster exhibits a positive F_{IS} , signifying slight excess homozygosity and some inbreeding, though levels are mild, reflected by relatively small F_{IS} values. Genetic differentiation analysis among clusters reveals consistent patterns. Pairwise comparisons consistently show genetic similarity between western Yakutia and eastern Yakutia, but significant differentiation between eastern Yakutia and Chukotka, and between western Yakutia and Chukotka. These patterns may reflect historical and demographic factors shaping the genetic structure of these clusters.

In summary, subtle spatial genetic structure and small but existent genetic differences between clusters (Table 1–2), mirroring the distribution of described *Larix* (sub)species *L. gmelinii* and *L. cajanderi*, result from a combination of isolation by distance, reproductive barriers imposed by geographical features, and independent glacial histories affecting these (sub)species.

Table 3. Approximate Bayesian computation – random forest (DIYABC-RF) analysis results. Parameter estimation for scenario 2. Effective population sizes of the corresponding clusters and relative assignment probability. Median and 90% confidence interval (CI).

Type	Parameter	Median (90% CI)
Effective population size	Nb	7830 (4190–9900)
Effective population size	Na	6570 (3380–9600)
Effective population size	N1	1980 (530–6640)
Effective population size	N2	3070 (340–7720)
Effective population size	N3	1490 (310–6090)
Relative assignment probability	ra	0.83 (0.41–0.97)

Refugia likely assisted the colonization of northern areas of the forest–tundra ecotone

The calibrated demographic event times from the biogeographic inference using ABC seem to fall within marine isotope stages (MIS; Railsback 2006) (Table 4, Fig. 7). The first larch colonization, according to our genetic data, traces back to MIS 11 (before t_3 , population Nb), an interglacial period during the Pleistocene, conducive to larch establishment. The subsequent bottleneck (t_3 , population Nb to population Na) can be dated to 353 000 years BP, within the glacial stage MIS 10 (approximately 365 000–335 000 years ago), likely causing diversity loss and population size reduction. The subsequent split event (t_2) at 20 400 years BP, within MIS 2 (containing the LGM), likely decreased gene flow on both sides of the Verkhoyansk Mountain Range and initiated population separation. The following admixture event (t_1) can be dated to 4600 years BP in MIS 1, thus in the Holocene with better conditions such as greater fitness of pollen or faster growth.

The ABC analysis suggests that the common ancestors of today's larch populations (Na and Nb) likely existed in northeast Siberia long before the last glacial period. However, precise timing of recent and ancient historical events remains challenging due to possible biases, particularly in estimating the average generation time of long-lived trees. Given the generation time is likely greater than 100 years during colder climate stages (Semerikov et al. 2013), the inferred divergence time is probably underestimated. Thus, we speculate that larch populations could have survived in isolated refugia during the last glacial, a hypothesis supported by similar findings in genetic studies indicating the presence of several refugia during Pleistocene glacial intervals (Polezhaeva et al. 2010, Semerikov et al. 2013). Pollen and macrofossil evidence also suggests *Larix* survival in northern regions throughout the LGM in multiple, often isolated refugia (Khatrab et al. 2008, Binney et al. 2009, Müller et al. 2010). Furthermore, we find genetic differentiation indicating long-term isolation of recent populations within geographically disconnected refugia (Tóth et al. 2019). Alaska also shows indications of possible in situ larch persistence during the LGM (Napier et al.

Table 4. Approximate Bayesian computation – random forest (DIYABC-RF) analysis results. Parameter estimation for scenario 2. Events, time scales in number of generations, and marine isotope stage (MIS) assignment. Median and 90% confidence interval (CI). LGM, Last Glacial Maximum.

Events	Time scale median [number of generations] (90% CI)	Time of the event assuming 100-year generation time [years BP] (90% CI)	Marine isotope stage
t1	46 (13–116)	4600 (1300–11 600)	MIS 1 Holocene
t2	204 (28–1400)	20 400 (2800–140 000)	MIS 2 LGM
t3	3530 (396–9271)	353 000 (39 600–927 100)	MIS 10 Glacial

2020). Northern Eurasia's complex topography likely provided a wide range of local climates conducive to refugial populations' persistence (Binney et al. 2017). In the eastern Yakutian cluster region, sheltered valleys of the Verkhoyansk Mountains (Tarasov et al. 2009) or the Tschuch'ye Lake area in eastern Yakutia with its deep, protected valleys (Lozhkin et al. 2018) could have served as refugia for *Larix* during the LGM.

Our ABC supports the hypothesis that the populations persisting in northern refugia during the LGM genetically contributed to post-LGM recolonization. This is consistent with studies indicating these populations were established well before the LGM, likely originating from a single source population (western Siberia) with a small effective size and low recent gene flow (Semerikov et al. 2013, Ma et al. 2020). The population in Chukotka likely originated earlier, with the population in eastern Yakutia emerging subsequently through an admixture event between populations from western Yakutia and Chukotka. Schulte et al. (2022b) state that *L. sibirica* had to recolonize northern areas from southern refugia postglacially. While *L. sibirica* may have locally retreated during the LGM from the lake region investigated in their study, survival in other northern areas is plausible. Refugia likely facilitated colonization of northerly forest–tundra ecotone areas by providing seed sources and shelter for larch regeneration (Kharuk et al. 2013). During range expansion and refugia reconnection, hybridization occurs, a common process in many forest tree species, including Siberian larch species (Semerikov et al. 2007).

Absence of northern refugia ahead of the current treeline could possibly explain the current treeline migration lag

The refugial populations likely served as a starting point for the rapid colonization of areas north of the treeline in the early Holocene (Tarasov et al. 2009, Epp et al. 2018), explaining the presence of *Larix* in the far north at that period (Bigelow 2003). Migration rates at the treelines were slow during this time. However, the current situation differs. Climate cooling during the little ice age (LIA), spanning from the 16th to the 19th centuries, negatively impacted tree population densities and led to range contraction. Despite enhanced recruitment in the 20th century, it has not sufficiently compensated for this range contraction (MacDonald et al. 2008). Consequently, there are currently no refugial populations in northern Siberia, as was most likely the case in the early Holocene. This circumstance may explain the lag in treeline advance behind climate warming, supported by simulation

studies (e.g. with LAVESI, Kruse et al. 2016, 2019). However, if individual trees establish themselves in the tundra area ahead of the treeline in the future, they could serve as the initial spark for rapid boreal coniferous forest dispersal (Wieczorek et al. 2017). Should progressive forest expansion align with climate change in the future, as suggested by various studies (MacDonald et al. 2008, Pearson et al. 2013, Kruse and Herzsuh 2022), tundra habitats are at risk of receding or disappearing completely.

Conclusions and outlook

We inferred the spatial distribution patterns of genetic variability in Siberian larches by GBS. Our data suggest the presence of two or three genetic groups. However, ecological considerations suggest that differentiation into three clusters could unveil admixture regions, highlighting not only main areas but also potential emergence of distinct clusters within the same species. The genetic structure arises from a confluence of factors including isolation by distance, reproductive barriers dictated by geographical features, and distinct glacial histories, which have independently influenced the genetic differentiation of the Eurasian larch species.

Our aim was to answer the question of whether refugia existed in northern areas during the LGM and to understand the temporal classification of possible demographic events. Overall, the ABC analysis supports a scenario in which present Siberian larch populations survived the LGM in northern refugia, rather than migrating postglacially from the south. The existence of northern LGM refugia may explain the early presence of larches in the far north during the Holocene and their continued dominance. The current lack of northern refugia will potentially hinder treeline advance to the north despite climate warming. The results of this study provide a better understanding of how refugial populations contribute to the treeline migration of Siberian larches.

Acknowledgements – Many people of the logistic and scientific staff at the Alfred Wegener Institute (AWI) and North-Eastern Federal University Yakutsk (NEFU) contributed to the success of the expeditions (2011–2021), where all the tree genetic samples used in this study were taken. We would like to thank two anonymous reviewers for their valuable feedback which greatly improved the manuscript and Cathy Jenks for English language proofreading. We also thank Sisi Liu for proofreading.

Funding – The PhD position of Sarah Haupt was funded by the Russian–German DFG-project 'Climate-induced treeline dynamics in the Ural Mountains: drivers, constraints, and the role

of genetic adaptation' (UralTreelines, <https://gepris.dfg.de/gepris/projekt/448651799>). We acknowledge support by the Open Access Publication Funds of Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung.

Author contributions

Sarah Haupt: Data curation (lead); Formal analysis (lead); Project administration (lead); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review and editing (lead). **Nadine Bernhardt:** Conceptualization (supporting); Investigation (supporting); Methodology (supporting); Writing – review and editing (supporting). **Stefanie Killing:** Formal analysis (supporting); Methodology (supporting); Software (supporting); Validation (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Stefano Meucci:** Data curation (supporting); Methodology (supporting); Software (supporting). **Ulrike Herzschuh:** Conceptualization (supporting); Funding acquisition (supporting); Resources (supporting); Supervision (supporting); Writing – review and editing (supporting). **Evgenii S. Zakharov:** Funding acquisition (supporting); Investigation (supporting); Writing – review and editing (supporting). **Dörte Harpke:** Data curation (supporting); Resources (supporting); Writing – review and editing (supporting). **Luidmila A. Pestryakova:** Funding acquisition (supporting); Resources (supporting); Writing – review and editing (supporting). **Stefan Kruse:** Conceptualization (lead); Formal analysis (lead); Funding acquisition (lead); Investigation (lead); Methodology (lead); Project administration (lead); Software (lead); Supervision (lead); Validation (lead); Visualization (lead); Writing – review and editing (lead).

Transparent peer review

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ecog.07092>.

Data availability statement

Data are available from the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB71740: <https://www.ebi.ac.uk/ena/data/view/PRJEB71740> (Haupt et al. 2024), using the data brokerage service of the German Federation for Biological Data (GFBio, Diepenbroek et al. 2014).

Supporting information

The Supporting information associated with this article is available with the online version.

References

Abaimov, A. P. 2010. Geographical distribution and genetics of Siberian larch species. – In: Osawa, A., Zyryanova, O. A., Matsuura, Y., Kajimoto, T. and Wein, R. W. (eds), Permafrost eco-

- systems: Siberian larch forests (Ecological Studies), vol. 209. Springer, pp. 41–58.
- Andreev, A. A., Schirrmeister, L., Tarasov, P. E., Ganopolski, A., Brovkin, V., Siebert, C., Wetterich, S. and Hubberten, H.-W. 2011. Vegetation and climate history in the Laptev Sea region (Arctic Siberia) during Late Quaternary inferred from pollen records. – *Quat. Sci. Rev.* 30: 2182–2199.
- Andreev, A. A., Nazarova, L. B., Lenz, M. M., Böhmer, T., Syrykh, L., Wagner, B., Melles, M., Pestryakova, L. A. and Herzschuh, U. 2022. Late Quaternary paleoenvironmental reconstructions from sediments of Lake Emada (Verkhoyansk Mountains, East Siberia). – *J. Quat. Sci.* 37: 884–899.
- Araki, N. H. T., Khatib, I. A., Hemamali, K. K. G. U., Inomata, N., Wang, X.-R. and Szmidi, A. E. 2008. Phylogeography of *Larix sukaczewii* Dyl. and *Larix sibirica* L. inferred from nucleotide variation of nuclear genes. – *Tree Genet. Genomes* 4: 611–623.
- Baltunis, B. S., Greenwood, M. S. and Eysteinnsson, T. 1998. Hybrid vigor in *Larix*: growth of intra- and interspecific hybrids of *Larix decidua*, *L. laricina*, and *L. kaempferi* after 5 years. – *Silvae Genet.* 47: 288–293.
- Bennett, K. and Provan, J. 2008. What do we mean by 'refugia'? – *Quat. Sci. Rev.* 27: 2449–2455.
- Bigelow, N. H. 2003. Climate change and Arctic ecosystems: 1. Vegetation changes north of 55°N between the Last Glacial Maximum, mid-Holocene, and present. – *J. Geophys. Res.* 108: 8170.
- Binney, H. A., Willis, K. J., Edwards, M. E., Bhagwat, S. A., Anderson, P. M., Andreev, A. A., Blaauw, M., Damblon, F., Haesaerts, P., Kienast, F., Kremenetski, K. V., Krivonogov, S. K., Lozhkin, A. V., MacDonald, G. M., Novenko, E. Y., Oksanen, P., Sapelko, T. V., Välranta, M. and Vazhenina, L. 2009. The distribution of late-Quaternary woody taxa in northern Eurasia: evidence from a new macrofossil database. – *Quat. Sci. Rev.* 28: 2445–2464.
- Binney, H. A., Edwards, M. E., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J. O., Andreev, A., Bezrukova, E., Blyakharchuk, T., Jankovska, V., Khazina, I., Krivonogov, S., Kremenetski, K., Nield, J., Novenko, E. Y., Ryabogina, N., Solovieva, N., Willis, K. J. and Zernitskaya, V. 2017. Vegetation of Eurasia from the last glacial maximum to present: key biogeographic patterns. – *Quat. Sci. Rev.* 157: 80–97.
- Bobrov, E. G. 1972. History and systematics of *Larix*. – *J. Bot.* 60: 797–805.
- Bonan, G. B. 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. – *Science* 320: 1444–1449.
- Borsch, T. et al. 2020. World flora online: placing taxonomists at the heart of a definitive and comprehensive global resource on the world's plants. – *Taxon* 69: 1311–1341.
- Brubaker, L. B., Anderson, P. M., Edwards, M. E. and Lozhkin, A. V. 2005. Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapped pollen data. – *J. Biogeogr.* 32: 833–848.
- Cao, X., Tian, F., Li, F., Gaillard, M.-J., Rudaya, N., Xu, Q. and Herzschuh, U. 2019. Pollen-based quantitative land-cover reconstruction for northern Asia covering the last 40 ka Cal BP. – *Clim. Past* 15: 1503–1536.
- Cao, X., Tian, F., Andreev, A., Anderson, P. M., Lozhkin, A. V., Bezrukova, E., Ni, J., Rudaya, N., Stobbe, A., Wiczorek, M. and Herzschuh, U. 2020. A taxonomically harmonized and temporally standardized fossil pollen dataset from Siberia covering the last 40 kyr. – *Earth Syst. Sci. Data* 12: 119–135.

- Chen, C., Mitchell, S. E., Elshire, R. J., Buckler, E. S. and El-Kassaby, Y. A. 2013. Mining conifers' mega-genome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform. – *Tree Genet. Genomes* 9: 1537–1544.
- Collin, F. D., Durif, G., Raynal, L., Lombaert, E., Gautier, M., Vitalis, R., Marin, J. M. and Estoup, A. 2021. Extending approximate Bayesian computation with supervised machine learning to infer demographic history from genetic polymorphisms using DIYABC Random Forest. – *Mol. Ecol. Resour.* 21: 2598–2613.
- Cornuet, J. M., Piry, S., Luikart, G., Estoup, A. and Solignac, M. 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. – *Genetics* 153: 1989–2000.
- Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J. M. and Estoup, A. 2014. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. – *Bioinformatics* 30: 1187–1189.
- Deschamps, S., Llaca, V. and May, G. D. 2012. Genotyping-by-sequencing in plants. – *Biology* 1: 460–483.
- Diepenbroek, M., Glöckner, F., Grobe, P., Güntsch, A., Huber, R., König-Ries, B., Kostadinov, I., Nieschulze, J., Seeger, B., Tolksdorf, R. and Triebel, D. 2014. Towards an integrated biodiversity and ecological research data management and archiving platform: the German federation for the curation of biological data (GFBio). – In: Plödereder, E., Grunske, L., Schneider, E. and Ull, D. (eds), *Informatik 2014 – big data Komplexität meistern*. GI-Edition: lecture notes in informatics (LNI) – Proceedings. GI edn., vol. 232. Köllen Verlag, pp. 1711–1724.
- Dong, M., He, Q., Zhao, J., Zhang, Y., Yuan, D. and Zhang, A. J. 2019. Genetic mapping of Prince Rupprecht's larch (*Larix principis-rupprechtii* Mayr) by specific-locus amplified fragment sequencing. – *Genes* 10: 583.
- Dray, S. and Dufour, A.-B. 2007. The ade4 package: implementing the duality diagram for ecologists. – *J. Stat. Softw.* 22: 1–20.
- Dylis, N. V. 1961. Larch species of East Siberia and the Far East. – *AN SSSR*.
- Eaton, D. A. R. and Overcast, I. 2020. ipyrad: interactive assembly and analysis of RADseq datasets. – *Bioinformatics* 36: 2592–2594.
- Eckert, A. J., Bower, A. D., González-Martínez, S. C., Wegrzyn, J. L., Coop, G. and Neale, D. B. 2010. Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). – *Mol. Ecol.* 19: 3789–3805.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. and Mitchell, S. E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. – *PLoS ONE* 6: e19379.
- Epp, L. S., Kruse, S., Kath, N. J., Stoeff-Leichsenring, K. R., Tiedemann, R., Pestryakova, L. A. and Herzschuh, U. 2018. Temporal and spatial patterns of mitochondrial haplotype and species distributions in Siberian larches inferred from ancient environmental DNA and modeling. – *Sci. Rep.* 8: 17436.
- François, O. 2016. Running structure-like population genetic analyses with R. – *Univ. of Grenoble-Alpes*, pp. 1–8.
- Franz, H. J. 1973. *Physische Geographie der Sowjetunion*. – Haack.
- Freeland, J. R. 2005. *Molecular ecology*. – Wiley.
- Frichot, E. and François, O. 2015. LEA: an R package for landscape and ecological association studies. – *Methods Ecol. Evol.* 6: 925–929.
- Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. and François, O. 2014. Fast and efficient estimation of individual ancestry coefficients. – *Genetics* 196: 973–983.
- Goudet, J. and Jombart, T. 2022. hierfstat: estimation and tests of hierarchical F-statistics. – R package ver. 0.5-11, <https://CRAN.R-project.org/package=hierfstat>.
- Gross, A., Tönjes, A. and Scholz, M. 2017. On the impact of relatedness on SNP association analysis. – *BMC Genet.* 18: 104.
- Harsch, M. A., Hulme, P. E., McGlone, M. S. and Duncan, R. P. 2009. Are treelines advancing? A global meta-analysis of treeline response to climate warming. – *Ecol. Lett.* 12: 1040–1049.
- Haupt, S., Bernhardt, N., Killing, S., Meucci, S., Herzschuh, U., Zakharov, E. S., Harpke, D., Pestryakova, L. A. and Kruse, S. 2024. Data from: Biogeography of larches in eastern Siberia – using single nucleotide polymorphisms derived by genotyping by sequencing. – European Nucleotide Archive, <https://www.ebi.ac.uk/ena/data/view/PRJEB71740>.
- Hedrick, P. W. 2005. A standardized genetic differentiation measure. – *Evolution* 59: 1633–1638.
- Herzschuh, U. 2020. Legacy of the Last Glacial on the present-day distribution of deciduous versus evergreen boreal forests. – *Global Ecol. Biogeogr.* 29: 198–206.
- Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. – *Nature* 405: 907–913.
- Holtmeier, F. K. and Broll, G. E. 2007. Treeline advance – driving processes and adverse factors. – *Landscape Online* 1: 1–33.
- IPCC. 2022. *Climate change 2022: impacts, adaptation and vulnerability. Contribution of working group II to the sixth assessment report of the intergovernmental panel on climate change* (Pörtner, H.-O., Roberts, D. C., Tignor, M., Poloczanska, E. S., Mintenbeck, K., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., Okem, A. and Rama, B. (eds)). – Cambridge Univ. Press.
- Isaev, A. P. et al. 2010. Vegetation of Yakutia: elements of ecology and plant sociology. – In: Troeva, E. I., Isaev, A. P., Cherosov, M. M. and Karpov, N. S. (eds), *The far north: plant biodiversity and ecology of Yakutia*. Springer, pp. 143–260.
- Johnson, J., Chhetri, P., Krutovsky, K. and Cairns, D. 2017. Growth and its relationship to individual genetic diversity of mountain hemlock (*Tsuga mertensiana*) at alpine treeline in Alaska: combining dendrochronology and genomics. – *Forests* 8: 418.
- Jombart, T. 2008. adegenet: an R package for the multivariate analysis of genetic markers. – *Bioinformatics* 24: 1403–1405.
- Jost, L. 2008. G_{ST} and its relatives do not measure differentiation. – *Mol. Ecol.* 17: 4015–4026.
- Kajimoto, T. 2010. Root system development of larch trees growing on Siberian permafrost. – In: Osawa, A., Zyryanova, O. A., Matsuura, Y., Kajimoto, T. and Wein, R. W. (eds), *Permafrost ecosystems: Siberian larch forests (ecological studies)*, vol. 209. Springer, pp. 303–330.
- Kharuk, V. I., Ranson, K. J., Im, S. T., Oskorbin, P. A., Dvinskaya, M. L. and Ovchinnikov, D. V. 2013. Tree-line structure and dynamics at the northern limit of the larch forest: Anabar Plateau, Siberia, Russia. – *Arct. Antarct. Alp. Res.* 45: 526–537.
- Khatib, I. A., Ishiyama, H., Inomata, N., Wang, X. R. and Szmidt, A. E. 2008. Phylogeography of Eurasian *Larix* species inferred from nucleotide variation in two nuclear genes. – *Genes Genet. Syst.* 83: 55–66.
- Kruse, S. and Herzschuh, U. 2022. Regional opportunities for tundra conservation in the next 1000 years. – *eLife* 11: e75163.

- Kruse, S., Wiczorek, M., Jeltsch, F. and Herzschuh, U. 2016. Treeline dynamics in Siberia under changing climates as inferred from an individual-based model for *Larix*. – *Ecol. Modell.* 338: 101–121.
- Kruse, S., Gerdes, A., Kath, N. J., Epp, L. S., Stoof-Leichsenring, K. R., Pestryakova, L. A. and Herzschuh, U. 2019. Dispersal distances and migration rates at the arctic treeline in Siberia – a genetic and simulation-based study. – *Biogeosciences* 16: 1211–1224.
- LePage, B. A. and Basinger, J. F. 1995. The evolutionary history of the genus *Larix* (Pinaceae). – US Department of Agriculture, Forest Service, Intermountain Research Station, GTR-INT-31, pp. 19–219.
- Li, Y., Zhang, X. and Fang, Y. 2019. Landscape features and climatic forces shape the genetic structure and evolutionary history of an oak species (*Quercus chenii*) in east China. – *Front. Plant Sci.* 10: 1060.
- Lozhkin, A., Anderson, P., Minyuk, P., Korzun, J., Brown, T., Pakhomov, A., Tsygankova, V., Burnatny, S. and Naumov, A. 2018. Implications for conifer glacial refugia and postglacial climatic variation in western Beringia from lake sediments of the Upper Indigirka Basin. – *Boreas* 47: 938–953.
- Ma, L., Cao, L., Hoffmann, A. A., Gong, Y., Chen, J., Chen, H., Wang, X., Zeng, A., Wei, S. and Zhou, Z. 2020. Rapid and strong population genetic differentiation and genomic signatures of climatic adaptation in an invasive mealybug. – *Divers. Distrib.* 26: 610–622.
- MacDonald, G. M., Velichko, A. A., Kremenetski, C. V., Borisova, O. K., Goleva, A. A., Andreev, A. A., Cwynar, L. C., Riding, R. T., Forman, S. L., Edwards, T. W. D., Aravena, R., Hammarlund, D., Szeicz, J. M. and Gattaulin, V. N. 2000. Holocene treeline history and climate change across northern Eurasia. – *Quat. Res.* 53: 302–311.
- MacDonald, G. M., Kremenetski, K. V. and Beilman, D. W. 2008. Climate change and the northern Russian treeline zone. – *Phil. Trans. R. Soc. B* 363: 2283–2299.
- Müller, S., Tarasov, P. E., Andreev, A. A., Tütken, T., Gartz, S. and Diekmann, B. 2010. Late Quaternary vegetation and environments in the Verkhoyansk Mountains region (NE Asia) reconstructed from a 50-kyr fossil pollen record from Lake Billyakh. – *Quat. Sci. Rev.* 29: 2071–2086.
- Napier, J. D., Fernandez, M. C., de Lafontaine, G. and Hu, F. S. 2020. Ice-age persistence and genetic isolation of the disjunct distribution of larch in Alaska. – *Ecol. Evol.* 10: 1692–1702.
- Nei, M. 1973. Analysis of gene diversity in subdivided populations. – *Proc. Natl Acad. Sci. USA* 70: 3321–3323.
- Nei, M. 1987. *Molecular evolutionary genetics*. – Columbia Univ. Press.
- Pearson, R. G., Phillips, S. J., Lorant, M. M., Beck, P. S. A., Damoulas, T., Knight, S. J. and Goetz, S. J. 2013. Shifts in Arctic vegetation and associated feedbacks under climate change. – *Nat. Clim. Change* 3: 673–677.
- Petit, R. J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D. and Vendramin, G. G. 2004. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. – *Mol. Ecol.* 14: 689–701.
- Polezhaeva, M. A., Lascoux, M. and Semerikov, V. L. 2010. Cytoplasmic DNA variation and biogeography of *Larix* Mill. in northeast Asia. – *Mol. Ecol.* 19: 1239–1252.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J. and Sham, P. C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. – *Am. J. Hum. Genet.* 81: 559–575.
- Railsback, L. B. 2006. Some fundamentals of mineralogy and geochemistry. – <http://railsback.org/FundamentalsIndex.html>.
- Rousset, F. 1999. Genetic differentiation within and between two habitats. – *Genetics* 151: 397–407.
- Rowe, G., Sweet, M. and Beebe, T. J. C. 2017. *An introduction to molecular ecology*. – Oxford Univ. Press.
- Schnell, R. J. and Priyadarshan, P. M. (eds). 2012. *Genomics of tree crops*. – Springer Science & Business Media.
- Schulte, L., Bernhardt, N., Stoof-Leichsenring, K. R., Zimmermann, H. H., Pestryakova, L. A., Epp, L. S. and Herzschuh, U. 2021. Hybridization capture of larch (*Larix* Mill.) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest. – *Mol. Ecol. Resour.* 21: 801–815.
- Schulte, L., Li, C., Lisovski, S. and Herzschuh, U. 2022a. Forest-permafrost feedbacks and glacial refugia help explain the unequal distribution of larch across continents. – *J. Biogeogr.* 49: 1825–1838.
- Schulte, L., Meucci, S., Stoof-Leichsenring, K. R., Heitkam, T., Schmidt, N., Von Hippel, B., Andreev, A. A., Diekmann, B., Biskaborn, B. K., Wagner, B., Melles, M., Pestryakova, L. A., Alsos, I. G., Clarke, C., Krutovsky, K. V. and Herzschuh, U. 2022b. *Larix* species range dynamics in Siberia since the Last Glacial captured from sedimentary ancient DNA. – *Commun. Biol.* 5: 570.
- Semerikov, V. L., Semerikov, L. F. and Lascoux, M. 1999. Intra- and interspecific allozyme variability in Eurasian *Larix* Mill. species. – *Heredity* 82: 193–204.
- Semerikov, V. L., Iroshnikov, A. I. and Lascoux, M. 2007. Mitochondrial DNA variation pattern and postglacial history of the Siberian larch (*Larix sibirica* Ledeb.). – *Russ. J. Ecol.* 38: 147–154.
- Semerikov, V. L., Semerikova, S. A., Polezhaeva, M. A., Kosintsev, P. A. and 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: 4958–4971.
- Smith, D. R. 2015. Mutation rates in plastid genomes: they are lower than you might think. – *Genome Biol. Evol.* 7: 1227–1234.
- Stewart, J. R. and Lister, A. M. 2001. Cryptic northern refugia and the origins of the modern biota. – *Trends Ecol. Evol.* 16: 608–613.
- Tarasov, P., Müller, S., Andreev, A., Werner, K. and Diekmann, B. 2009. Younger Dryas *Larix* in eastern Siberia: a migrant or survivor? – *Pages News* 17: 122–123.
- Tóth, E. G., Tremblay, F., Housset, J. M., Bergeron, Y. and Carcaillet, C. 2019. Geographic isolation and climatic variability contribute to genetic differentiation in fragmented populations of the long-lived subalpine conifer *Pinus cembra* L. in the western Alps. – *BMC Evol. Biol.* 19: 190.
- Tsumura, Y., Kimura, M., Nakao, K., Uchiyama, K., Ujino-Ihara, T., Wen, Y., Tong, Z. and Han, W. 2020. Effects of the last glacial period on genetic diversity and genetic differentiation in *Cryptomeria japonica* in East Asia. – *Tree Genet. Genomes* 16: 19.
- Väliranta, M., Kaakinen, A., Kuhry, P., Kultti, S., Salonen, J. S. and Seppä, H. 2011. Scattered late-glacial and Early Holocene tree populations as dispersal nuclei for forest development in north-eastern European Russia: holocene forest development in north-eastern European Russia. – *J. Biogeogr.* 38: 922–932.
- Weir, B. S. and Cockerham, C. C. 1984. Estimating F-statistics for the analysis of population structure. – *Evolution* 38: 1358–1370.
- Wendler, N., Mascher, M., Nöh, C., Himmelbach, A., Scholz, U., Ruge-Wehling, B. and Stein, N. 2014. Unlocking the secondary gene-pool of barley with next-generation sequencing. – *Plant Biotechnol. J.* 12: 1122–1131.

- Wieczorek, M., Kruse, S., Epp, L. S., Kolmogorov, A., Nikolaev, A. N., Heinrich, I., Jeltsch, F., Pestryakova, L. A., Zibulski, R. and Herzsuh, U. 2017. Dissimilar responses of larch stands in northern Siberia to increasing temperatures – a field and simulation based study. – *Ecology* 98: 2343–2355.
- Winter, D. J. 2012. MMOD: an R library for the calculation of population differentiation statistics. – *Mol. Ecol. Resour.* 12: 1158–1160.
- Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. – *Evolution* 19: 395–420.
- Zhang, N., Yasunari, T. and Ohta, T. 2011. Dynamics of the larch taiga–permafrost coupled system in Siberia under climate change. – *Environ. Res. Lett.* 6: 024003.
- Zhang, W., Miller, P. A., Smith, B., Wania, R., Koenigk, T. and Döscher, R. 2013. Tundra shrubification and tree-line advance amplify arctic climate warming: results from an individual-based dynamic vegetation model. – *Environ. Res. Lett.* 8: 034023.
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C. and Weir, B. S. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. – *Bioinformatics* 28: 3326–3328.