

# Biogeography of larches in Eastern Siberia - using single nucleotide polymorphisms derived by genotyping by sequencing

Sarah Haupt<sup>1,2</sup>, Nadine Bernhardt<sup>3</sup>, Stefanie Killing<sup>2</sup>, Stefano Meucci<sup>1,2</sup>, Ulrike Herzschuh<sup>1,2,4</sup>, Evgenii Zakharov<sup>5,6</sup>, Dörte Harpke<sup>7</sup>, Lyudmila Pestryakova<sup>5</sup>, and Stefan Kruse<sup>1</sup>

<sup>1</sup>Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, 14473 Potsdam, Germany

<sup>2</sup>Institute of Biochemistry and Biology, University of Potsdam, 14476 Potsdam, Germany

<sup>3</sup>Julius Kühn-Institut, Federal Research Centre for Cultivated Plants, Institute for Resistance Research and Stress Tolerance, 06484 Quedlinburg, Germany

<sup>4</sup>Institute of Environmental Sciences and Geography, University of Potsdam, 14476 Potsdam, Germany

<sup>5</sup>Institute of Natural Sciences, North-Eastern Federal University of Yakutsk, 677000 Yakutsk, Russia

<sup>6</sup>Institute for Biological Problems of the Cryolithozone, Russian Academy of Sciences, Siberian branch, 677000 Yakutsk, Russia

<sup>7</sup>Leibniz Institute of Plant Genetics and Crop Plant Research Gatersleben, 06466 Stadt Seeland, Germany

September 6, 2023

## Abstract

The present distribution of Siberian boreal forests that are dominated by larches is influenced, to an unknown extent, by the glacial history. Knowing the past treeline response to climate shifts can improve our understanding of future treeline dynamics under changing climate. Here, we study patterns in the genetic variability of Siberian larches (*Larix* spp.) that can help us to unravel biogeographic migration routes since the Last Glacial Maximum (LGM). We sampled 148 larch individuals from Eastern Siberia. For each individual, genome-wide single nucleotide polymorphisms (SNPs) were derived through genotyping by sequencing (GBS). We inferred the spatial distribution from 14,003 SNPs with a cluster analysis. To infer the postglacial demographic history of *Larix*, we applied an Approximate Bayesian Computation. The Bayesian population assignment statistically supported three to four clusters from Western to Eastern Siberia that correspond well to the geographic ranges of the main Siberian larch species *Larix sibirica*, *L. gmelinii*, and *L. cajanderi*. Using four plausible clusters, the tested hypotheses in DIYABC show that the existing populations seem to have been initiated long before the LGM. We presume that the different populations originate from larch populations that survived the glacial periods. From our genetic studies, we deduce that *Larix* was more likely to have survived the cold LGM in northern refugia, from where a fast colonization of Siberia was possible, rather than *Larix* completely repopulating Siberia in the postglacial spreading out from southern areas with less harsh climatic conditions. The northernmost expansion during the Holocene seems to have benefitted from refugial populations ahead of the treeline at that time, which explains the existence of *Larix* in the far north. We expect from our results that the present migration will be slow at first as there are currently no refugial populations far north, as there probably were in the Holocene.

## 1 Introduction

The extensive forests in East 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 2009). The 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.

The speed and spatial pattern of a species' response to climate change are determined by several processes. The biogeographic history of larches (*Larix* spp.) 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 invaded in the postglacial (Bennett and Provan 2008). Relict forest stands of past warm phases which have survived in refugia ahead of the current treeline can speed up the migration rate of treelines estimated from observational studies (Stewart and Lister 2001; Holtmeier and Broll 2007; MacDonald et al. 2008; Välranta et al. 2011) and modelling (Kruse et al. 2019). In line with these general findings, radiocarbon-dated macrofossil findings suggest a fast range expansion of the 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 current responses of the Siberian larch forest reveals that tree stands respond with a pronounced time lag to the current 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, knowledge about the influence of relict trees from an earlier wider extent of the forest on migration dynamics is of great importance.

Fossil records of *Larix* show that the genus became a well-established forest constituent in northeastern Russia by the late Oligocene (26.5 to 24 mya) (LePage and Basinger 1995). Recent Siberian larch species, especially *L. gmelinii*, probably formed as an adaptation to the increasing climatic continentality in the Pleistocene (Abaimov 2010). *Larix* formed forests in Siberia throughout the Holocene (Cao et al. 2019). Several 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 BP), even when there was very likely a relatively low density of larch (Cao et al. 2020). Pollen and macrofossil records suggest that local *Larix* populations persisted 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). However, 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 the larches in Siberia (Herzschuh 2020). Araki et al. (2008) suggest that the investigated *Larix sukaczewii* and *L. sibirica* populations were founded by migrants 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* had a very limited contribution to the current populations of the central and northern parts of its range. In contrast, a recent study was able to distinguish between *L. sibirica* and *L. gmelinii* in glacial refugial populations by enriching sedimentary ancient DNA extracts for chloroplast genome sequences (Schulte et al. 2021; Schulte et al. 2022b). These studies reveal that northern refugial populations existed during the LGM and were almost exclusively composed of *L. gmelinii*. *Larix sibirica*, on the other hand, recolonized from southern refugia. However, for the vast distribution area of larches in Eastern Siberia and the other Siberian

larch species the knowledge is very limited.

The genetic markers previously used to study phylogeography were limited to a few informal sequences and mostly localized to plastids. This limitation can be overcome by the recently developed approach of genotyping by sequencing (GBS) that enables the analysis of hundreds to thousands of nuclear loci at a low cost and gives deep insights into their phylogeography as well as adaptation. GBS combines complexity reduction, multiplexing of samples, and the use of next-generation sequencing (NGS) methods for genotyping of whole mapping populations (Wendler et al. 2014) and is feasible for high diversity, large genome species such as larches (*Larix* spp.) (Elshire et al. 2011). In addition, the analysis of nuclear DNA has other advantages, as it contains complementary information to plastid DNA due to different modes of inheritance (Petit et al. 2004). Many gymnosperms inherit plastid DNA from male pollen which is spread in the wind and mitochondrial DNA is normally inherited exclusively from the mother (Freeland 2005). Additionally, for land plants, the plastid and nuclear genomes have a 3- to 10-fold greater mutation rate than the mitochondrial genome (Smith 2015). Thus, we could expect to detect higher levels of genetic structure among populations using biparentally inherited nuclear DNA from the GBS analyses. To our knowledge, GBS has not yet been used to infer population structures and hence unravel the biogeography of larches (*Larix* Mill.).

The overall aim of this study is 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 examine genealogical interrelations of representatively sampled populations across Eastern Siberia (*Larix sibirica* Ledeb., *L. gmelinii* (Rupr.) Rupr., *L. cajanderi* Mayr.) using single nucleotide polymorphisms (SNPs) derived by GBS. The specific objectives are (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 Last Glacial Maximum (LGM).

## 2 Material and Methods

### 2.1 Study genus

*Larix* Mill. is the most abundant genus in boreal coniferous forests of Eurasia and *Larix* species cover 70% of the forest biome globally (Abaimov 2010). Larches are well adapted to the harsh climate conditions, that is, cold winters to below -40 degC and short summers with only 60–90 days exceeding the freezing point (Franz 1973). According to different taxonomic systems, there are different opinions regarding the number of larch species and the boundaries of their habitats (Mayr 1906; Dylis 1961; Bobrov 1972). The distribution areas of some *Larix* species overlap and larches can hybridize easily in natural habitats (Semerikov et al. 2007; Abaimov 2010). This especially applies to the Siberian larch species. For this study, following Bobrov (1972), we recognize three species of *Larix* across their Siberian distribution from east to west, *L. sibirica* Ledeb., *L. gmelinii* (Rupr.) Rupr., and *L. cajanderi* Mayr, although we acknowledge there might be considerable hybridization between the several species and varieties. The natural distribution of Siberian larch species extends from west to east in large, relatively discrete ranges with hybridization occurring in the boundary zones (Abaimov 2010).

### 2.2 Study area and sample collection

The study area includes arctic and subarctic areas located in the forest-tundra transition zone in Siberia. The sites are located in Siberia from the Ural Mountains in the West, followed by the Taymyr Peninsula, to the Lena Delta and the Verkhoyansk Range reaching to Chukotka in the Far East; as well as Kamchatka in the southeast, via Magadan and southern Yakutsk to the Altai Mountains (c. 50–80degN; 50–180degE).

For genetic analyses, fresh needles from natural populations of *Larix sibirica*, *L. gmelinii*, and *L. cajanderi* were taken during summer fieldwork (2011–2021) (Fig. 1). Samples were selected to cover a wide distribution area of Siberia. Needle samples were dried on silica gel during fieldwork and stored at 4degC. In total, 148 individuals were processed for which the identifier, geographic position, and height were recorded (Appendix S1).

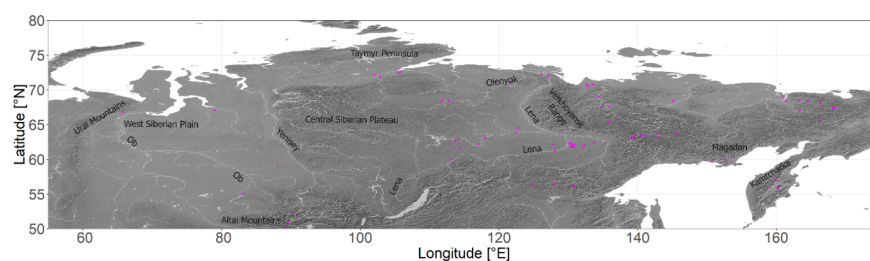


Fig. 1 Sampling locations

These sequence data have been submitted to the European Nucleotide Archive (ENA) under accession number PRJEB65323. (not yet published)

### 2.3 DNA extraction

The genomic DNA was extracted from 40 to 80 mg of dried needles after grinding them for 2 minutes in a FastPrep®-24 instrument (MP BIOMEDICALS) device using the Invisorb® Spin Plant Mini Kit (STRATEC MOLECULAR), DNeasy® Plant Mini Kit (QIAGEN) or DNeasy® Plant Pro Kit (QIAGEN), following the supplier's protocol, but the buffers were heated before use.

### 2.4 GBS, de novo assembly, and variant call

Genotyping by sequencing (GBS) was conducted following Wendler et al. (2014) to obtain genome-wide single nucleotide polymorphisms (SNPs). Genomic DNA (200 ng) was digested using restriction enzymes PstI-HF and MspI. PstIHF is a rare-cutting enzyme (recognition site: CTGCA'G) and the methylation-sensitive enzyme MspI (recognition site: C'CGG) is a frequent cutting enzyme. Libraries were sequenced on an Illumina HiSeq 2000 and NovaSeq6000, generating single-end reads of 100-120 bp.

Barcoded Illumina reads were de-multiplexed using the Casava pipeline 1.8 (Illumina), trimmed, filtered, and de novo assembled with ipyrad v. 0.7.30 (Eaton and Overcast 2020).

A maximum of 5 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 6. 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 60%. Heterozygous sites were allowed for a maximum of 50% of the samples.

For additional details, the ipyrad parameter file is provided in Appendix S2. The geno file was used to generate the input file for DIYABC (Appendix S3).

## 2.5 GBS data analysis

The genetic structure among populations and between individuals has the potential to unravel spatial distribution patterns. Individuals will have more similar genotypes when they originate from the same population. Thus, it is possible to evaluate how they cluster based on their genotypes (Cornuet et al. 1999). Here we use the Bayesian population assignment using the R package LEA to serve as an illustration of the distribution of ancestry coefficients. It represents the spatial predictions on a geographic map. The acquired SNP data are linked to geographical information and patterns are inferred through the analysis of ancestry coefficients and admixture rates (cluster analysis) (Frichot and François 2015). Subsequently, the evaluation of potential biogeographical dynamics of different larch populations in Russia will be performed via the model-based approach Approximate Bayesian Computation, implemented in the software DIYABC (Cornuet et al. 2014).

### 2.5.1 Spatial distribution of SNPs

We used optimized versions of principal component analysis (PCA) and non-negative matrix factorization algorithms (sNMF) (Frichot et al. 2014) as implemented in the R package LEA version 2.8.0 (Frichot and François 2015; François 2016). To define the optimal number of genetic clusters (K) we tested K values from 1 to 6 based on knowledge about the number of larch species and main hybrids in Siberia. The final number of clusters (K) was selected by choosing the K-value with the lowest cross-entropy. Only the K with the maximal local contribution to ancestry is represented at each point of the geographic map.

### 2.5.2 Biogeographic inference - Approximate Bayesian Computation (ABC)

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

We applied an Approximate Bayesian Computation (ABC) using the software DIYABC v. 2.1.0 (Cornuet et al. 2014). The potential demographic history was inferred via complex evolutionary scenarios. To keep the scenarios in the ABC simple, four populations were defined based on the results of the cluster analysis and the admixture plots: Pop1 (Western Siberia= WSib), Pop2 (Western Yakutia= WYak), Pop3 (Eastern Yakutia= EYak), and Pop4 (Chukotka= Chuk). Alternative scenarios including the predefined four distinct genetic groups for the estimation of the population demographic history were constructed. In these scenarios,  $t\#$  represents the time scale, measured by generation time, and  $N\#$  represents the effective population size of the corresponding populations. The 12 examined scenarios viewed backward in time, are as follows (Fig. 2):

Scenario 1: Pop4/ Chuk merged with Pop3/ EYak at t1, and then they merged with Pop2/ WYak at t2. They merged with Pop1/ WSib at t3. Scenario 2: Pop3/ EYak merged with Pop2/ WYak at t1, and then Pop4/ Chuk merged with them at t2. They merged with Pop1/ WSib at t3. Scenario 3: Pop3/ EYak was created by an admixture of Pop2/ Wyak and Pop4/ Chuk at t1, and then Pop4/ Chuk merged with Pop2/ WYak at t2. Pop2 merged with Pop1/ WSib at t3. Scenario 4: Three populations (Pop1/ WSib, Pop3/ EYak, Pop4/ Chuk) split at the same time, namely t1, and then Pop2/ WYak merged with them at t2. Scenario 5: All four populations (Pop1/ WSib, Pop2/ WYak, Pop3/ EYak, Pop4/ Chuk) split at the same time, namely t1. Scenario 6: The order of the hierarchical splits is the same as in scenario 1 but a recent bottleneck was modeled in all four populations at t1-db (db: time scale to bottleneck before t1): population size changed from N1 to N1b, from N2 to N2b, from N3 to N3b, and from N4 to N4b. N4b merged with N3b at t2, and then they merged with N2b at t3. They merged with N1b at t4. Scenarios 7–12: These are the same as scenarios 1–6 but an ancient and severe bottleneck was modeled: population size changed from N1 to N1a (and from N1b to N1a for scenario 12) at t4-db for scenarios 7–9, at t3-db for scenario 10, at t2-db for scenario 11, and at t5-db for scenario 12.

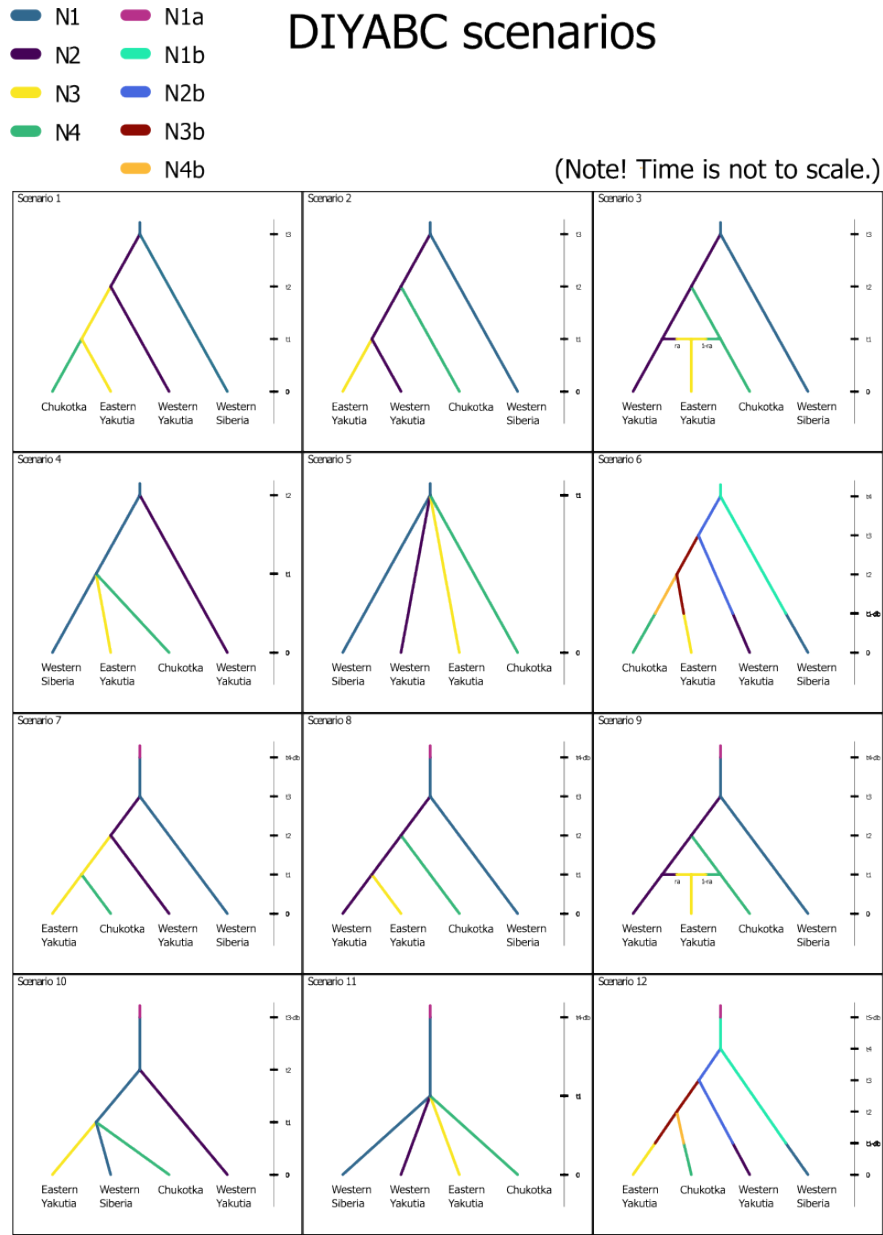


Fig. 2: The twelve constructed scenarios tested with Approximate Bayesian Computation. In these scenarios,  $t\#$  represents the time scale in number of generations and  $N\#$  represents the effective population size during the time period (e.g.,  $0-t_1$ ,  $t_1-t_2$ ).

These 12 scenarios constructed by us are the ones among a multitude of pre-tested scenarios that most closely correspond to the possible migration dynamics.

For each scenario, 100,000 simulations were run. The models were compared using the logistic regression method, and the scenario with the highest posterior probability was determined to be the most realistic one.

To obtain the calibrated years before present (BP), the number of generations has to be multiplied by a given generation time. In the present study, the generation time of *Larix* is set to 25 years, since larch trees can start reproducing when they reach an age of 20–30 years (Semerikov et al. 2013).

### 3 Results

#### 3.1 GBS data assessment

The DNA extracted from the needles of the 148 individuals was genotyped by GBS and returned a maximum of 4,822,890 and a minimum of 417,255 reads with an average of 1,630,465 raw reads per sample.

The raw reads were assembled into 14,565 filtered loci among the 148 individuals with an average of 11,285, a maximum of 14,003, and a minimum of 2,209 loci per sample. A total of 91,203 SNPs with 28.6% missing sites was detected by consensus base calling. An average of 0.02050 and 0.004297 estimated heterozygosity and error rate was detected, respectively. Full statistical results are provided in Appendix S4.

#### 3.2 Spatial distribution of SNPs – Cluster analysis

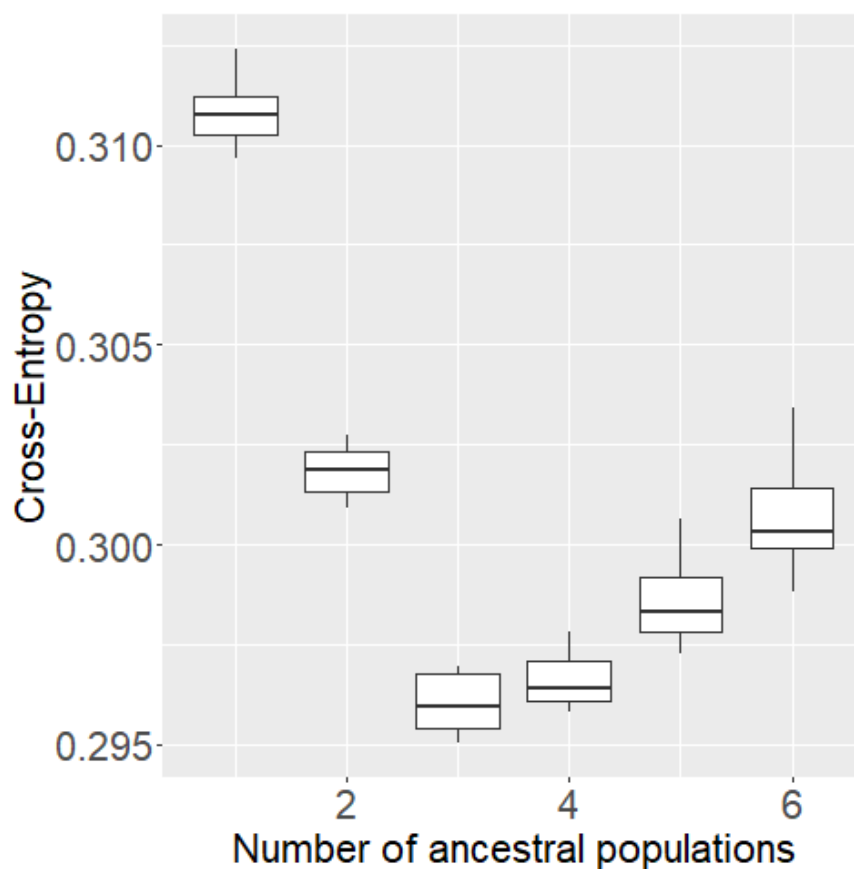


Fig. 4: Cluster size evaluation: Cross-entropy plot for the number of clusters  $K=1-6$  with 10 repetitions for randomization. Value of the cross-entropy criterion as a function of the number of populations in sNMF. The retained value of  $K$  is  $K=3$ .

We assessed the spatial distribution of SNPs to ascertain whether they can be detected in aggregated cluster formations. The results of the cross-validation revealed the presence of three statistically verified clusters, but the difference in cross-entropy from  $K=3$  to  $K=4$  is negligible (Fig. 4). In addition, four clusters remain stable even if the cluster analysis is run several times. If more clusters are set ( $K=5$  or  $K=6$ ), the resulting cluster 5 and cluster 6 look different for each run.



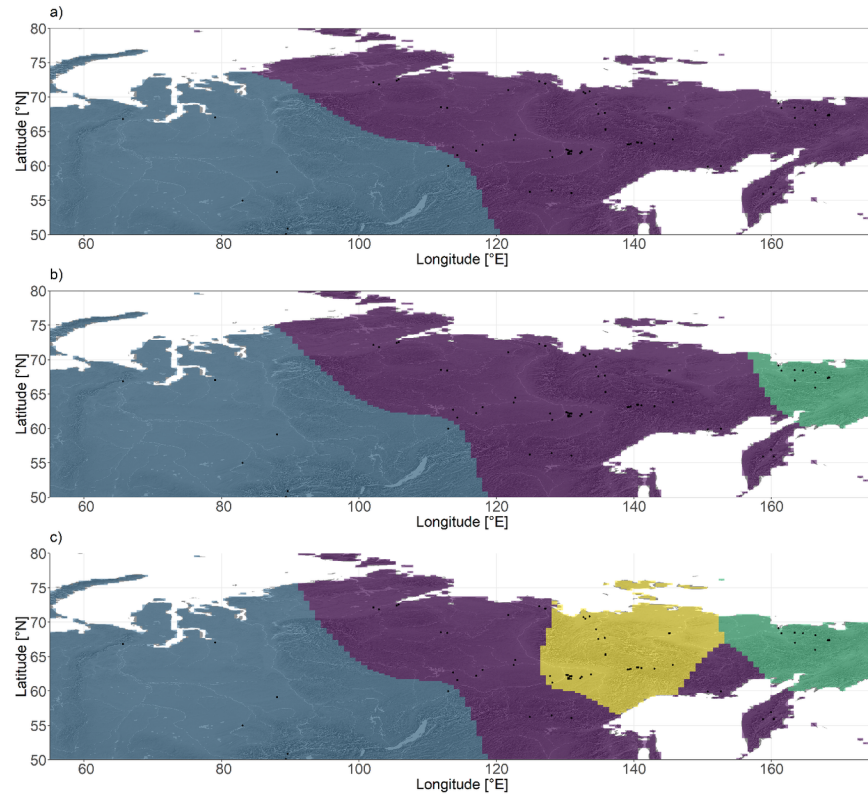


Fig. 3 a) Ancestry coefficients for a) two clusters ( $K=2$ ), b) three clusters ( $K=3$ ), and c) four clusters ( $K=4$ ).

Four geographically distinct groupings with sharp species boundaries can be distinguished and assigned to the given regions from Western to Eastern Eurasia. When two clusters were set ( $K=2$ ), cluster 1 (Western Siberia) was genetically distinct from cluster 2 (Eastern Siberia) (Fig. 3 a). Cluster 1 spreads out from Northern to Southern Siberia over the entire zone of the West Siberian Plain. Cluster 2 runs from across the Taymyr Peninsula in the northern part of Siberia through Yakutia to Bolshoi Toko in Southern Yakutia. Between cluster 1 and cluster 2, the Yenisey River and the Central Siberian Plateau form a clear boundary. Under three clusters ( $K=3$ ), the former Eastern Siberian cluster divides into two distinct clusters (Fig. 3 b): cluster 2 (Yakutia) and further east in the region of Chukotka cluster 3 (Chukotka). Under four clusters ( $K=4$ ), the Yakutian cluster is divided into a western and an eastern cluster (Fig. 3 c). Between cluster 3 (Western Yakutia) and cluster 4 (Eastern Yakutia), the Lena River is visible, which is located parallel to the Verkhoyansk Range.

Many individuals (63%) can be assigned with 75–100% confidence to a definite cluster (Fig. 5). When comparing the admixture plots with the spatial distribution of the clusters (Fig. 3 c), it becomes apparent that the individuals in the core area of the clusters have a clear affiliation to exclusively one cluster and the individuals at the edge of the clusters have a fragmented affiliation to different clusters. This is also apparent in the pie charts on a map provided in Appendix S5.

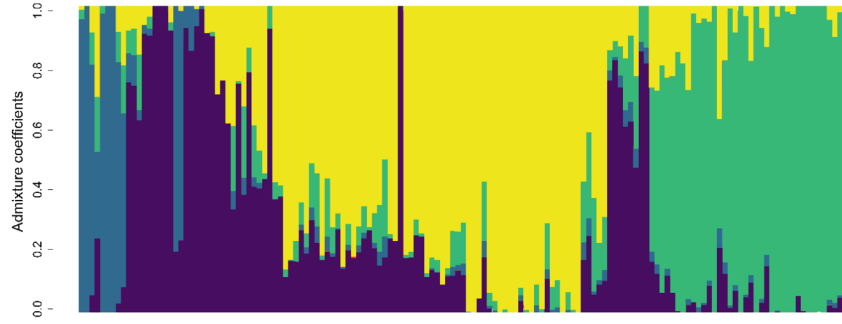


Fig. 5: Bayesian population assignment for four clusters ( $K=4$ ): Admixture plots are histograms of admixture coefficients and show the relative assignment of the individuals to the different relevant clusters. On the y-axis, the proportion of admixture per individual is shown. The x-axis spans the entire dataset with the current percentage of allocation to a cluster. Each bar represents one individual. A one-colored bar belongs exclusively to one cluster, whereas a multi-colored bar indicates an affiliation to different clusters. The individuals on the x-axis are sorted by longitude from west to east: blue=Western Siberia, purple=Western Yakutia, yellow=Eastern Yakutia, green=Chukotka.

### 3.3 Biogeographic inference - Approximate Bayesian Computation

Since linkage disequilibrium (LD) can affect biogeographic inference, SNPs with strong LD were pruned. In order to minimize sequencing and assembly errors the SNPs showing severe distortion of the HWE ( $p < 0.05$ ), MAF lower than 5%, and MISS above 20% were filtered out. Thereby, the number of SNPs was reduced to 2733.

Table 1: Posterior probability and confidence interval (CI) for every DIYABC scenario

| Scenario | Posterior probability (95%CI) |
|----------|-------------------------------|
| 1        | 0.0016 (0.0000-0.6450)        |
| 2        | 0.0001 (0.0000-0.6451)        |
| 3        | 0.0068 (0.0000-0.6448)        |
| 4        | 0.0000 (0.0000-0.6451)        |
| 5        | 0.0000 (0.0000-0.6451)        |
| 6        | 0.0001 (0.0000-0.6451)        |
| 7        | 0.2347 (0.0000-0.6457)        |
| 8        | 0.0225 (0.0000-0.6446)        |
| 9        | 0.7344 (0.5771-0.8916)        |
| 10       | 0.0000 (0.0000-0.6451)        |
| 11       | 0.0000 (0.0000-0.7910)        |
| 12       | 0.0000 (0.0000-0.6451)        |

## 4 Discussion

4.1 The genetically distinct clusters from Western to Eastern Siberia correspond well to the main larch species. The GBS resulted in a total of 14,565 variable SNPs among the 148 individuals, which is a high number in comparison to other studies with Pinaceae (e.g. Johnson et al. 2017; Eckert et al. 2010; Chen et al. 2013).

With GBS the complexity of the large genome of *Larix* was efficiently reduced for downstream analyses (Dong et al. 2019; Deschamps et al. 2012).

The cluster analysis based on all SNPs identified four distinct genetic clusters. The boundaries of the genetically inferred main clusters correspond well to the boundaries of the main Siberian larch species described by Abaimov (2010). As expected, the Western Siberian cluster (cluster 1) can be assigned to *L. sibirica* and corresponds to its distribution across the territory between the Ob and Yenisey Rivers (Abaimov 2010). The easternmost site belonging to this cluster is located at Lake Khamra (59.97° N, 112.96° E). The Western Yakutian cluster (cluster 2) shows the distribution area of *L. gmelinii*, which extends from the basin of the Pyasina River in the west to the Lena River delta in the east. The five tree individuals from the Taymyr Peninsula region are the westernmost trees in this *L. gmelinii* cluster in our study. According to Abaimov (2010), the northernmost *L. gmelinii* forests are located on the Taymyr Peninsula. The Eastern Yakutian cluster (cluster 3) is attributable to *L. cajanderi*, whose western border habitat goes along the Lena River. However, in the discussion about the taxonomic status of *L. cajanderi*, there are different views on where to place the boundary between *L. gmelinii* and *L. cajanderi*. Some authors classify *L. cajanderi* Mayr as a synonym of *L. gmelinii* (Rupr.) Kuzen (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 also coincides with the transition zone between our separate Western Yakutian and Eastern Yakutian clusters and supports the differentiation of 4 rather than 3 clusters in our study. 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). In summary, we argue that the genetic differences between *L. gmelinii* and *L. cajanderi* are very small but existent.

Particularly striking are the topographical conditions such as mountain ranges or river valleys along the cluster and species boundaries, especially as the positions of the samples were not considered in the cluster inference. The border between the West Siberian cluster (cluster 1) and the Western Yakutian cluster (cluster 2) corresponds almost exactly with the course of the Yenisey River. Nevertheless, it can be assumed that at this point the Central Siberian Plateau serves more as a geographical barrier which is responsible for the genetic differentiation between *L. sibirica* and *L. gmelinii*. Similarly, the Verkhoyansk Range likely serves as a geographical barrier between the Western Yakutian cluster (cluster 2) and the Eastern Yakutian cluster (cluster 3) rather than the Lena River and is thus responsible for the separation between *L. gmelinii* and *L. cajanderi*. For a lowland species like larch, the mountains present a persistent geographic barrier to expansion. In addition to long-term geographical isolation, mountains can increase climatic gradients and thus further facilitate population differentiation through environmental isolation (Li et al. 2019). There may be gene flow between adjacent watersheds, mountain ranges, or valleys, but less than between populations that are not separated by landscape components (Priyadarshan and Schnell 2012). The separation could be responsible for keeping two populations genetically isolated so that they slowly diverge genetically from each other (Rowe et al. 2017). A study on the effects of the last glacial period on genetic diversity and genetic differentiation in *Cryptomeria japonica* in East Asia (Tsumura et al. 2020) concludes that geographical barriers between populations are one of the strongest drivers for genetic variety.

The individuals in the core area of the clusters have an affiliation to exclusively one cluster, but at the periphery, individuals have a fragmented affiliation to different clusters. In the contact zone of their geographic ranges, we found hybrids from the *L. gmelinii* and *L. cajanderi* clusters forming a stripe of transitional populations on the left bank of the Lena River (Abaimov 2010; Isaev et al. 2010). These findings meet expectations, since cross-pollination of two different plant varieties is taking place in these transitional areas (Baltunis et al. 1998). Unfortunately, no tree genetic samples from the mountainous region of Krasnoyarsk Krai between cluster 1 and cluster 2 were available in this study. However, it can be assumed that the species in the contact zone between the eastern limit of *L. sibirica* and the western limit of *L. gmelinii* are involved in natural hybridization (Abaimov 2010; Isaev et al. 2010). Szafer (1913) described this stripe of hybrid populations as *L. czekanowskii*. From the observation of genetic clusters and hybrid zones, the modern and historical genetic influences can be deduced. Populations that have evolved independently due to isolation by

geographical barriers or because they retreated into refugia during glacial periods met again when physical barriers disappeared or when the populations expanded again during warm periods (Abaimov 2010; Isaev et al. 2010).

Noticeable is the significant geographical distinction of the Chukotka cluster (cluster 4), which can be considered as a variant of *L. cajanderi*. The intraspecific genetic deviations between cluster 3 (Western Yakutia) and cluster 4 (Chukotka) could stem from separation by the Chersky Range and the Kolyma Mountains, where varied microclimates could be present at different elevations, which could have led to differentiation due to regional climatic differences. Genetic differentiation can also increase with geographical distance even when there are no apparent barriers to gene flow. Since very little gene flow occurs between the *L. cajanderi* population in Chukotka and the main distribution area of the species, then during the glacial and interglacial cycles, these Chukotka populations could have diverged genetically due to gene drift, a process called isolation by distance (Rousset 1999). The genetic variability of *L. cajanderi* is also supported by observations showing that *L. cajanderi* is involved in hybridization with *L. kamschatica* and other hybrid forms of *Larix* species in the far northeast (Abaimov 2010).

4.2 Refugia likely assisted the invasion of northern areas of the forest-tundra ecotone

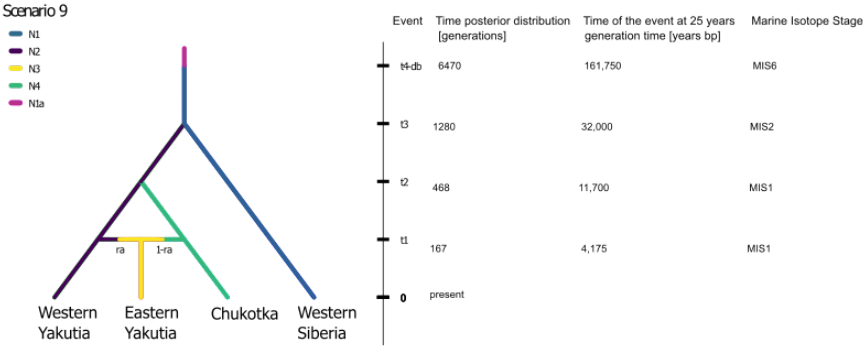


Fig. 6: DIYABC Scenario 9 - Events and marine isotope stage (MIS) assignment

According to our results from the ABC, the calibrated times of demographic events can be assigned to marine isotope stages (MIS; Railsback 2006) (Fig. 6). The first larch colonization (t4, population N1a) has been recorded at around 224,000 years BP, which is equivalent to MIS7 and specified as a moderate interglacial. This suggests that the climate was favorable for larch establishment. The subsequent ancient and severe bottleneck (t4-db, population N1a to population N1) can be dated to 161,750 years BP, corresponding to MIS6, when the penultimate glacial period occurred, which was a more severe glaciation than the LGM. These harsh climate conditions may have caused loss of diversity and a decrease in population size. The population N1a may have been threatened with extinction, but after the bottleneck event, the population size increased again. The subsequent split event (t3) is registered at 32,000 years BP within MIS2, containing the LGM. This period of moderate glaciation did not cause a significant decrease in population size. The

following split event (t2) and the admixture event (t1) can be dated at 11,700 years BP and 4,175 years BP, respectively, both in MIS1, thus in the Holocene. The split events and the higher population size indicate better conditions such as greater fitness of pollen or faster growth.

The ABC leads to the assumption that the common ancestors of today's larch populations (N1 and N2) must have been present in northeast Siberia long before the last glacial. However, the exact timing of both recent and ancient historical events should be considered a rough estimate because of possible biases, as it is difficult to estimate the average generation time of long-lived trees. Because the generation time is likely to be greater than 25 years during colder climate stages, the inferred divergence time is most probably underestimated. Hence, we can deduce that larch populations must have survived in isolated refugia during the last glacial. This hypothesis is also supported by similar findings in other genetic studies that suggest the presence of several refugia during Pleistocene glacial intervals (Polezhaeva et al. 2010; Semerikov et al. 2013). Both pollen and macrofossil evidence indicates the survival of *Larix* in northern regions throughout the LGM in multiple and often isolated refugia (Khatab et al. 2008; Binney et al. 2009; Müller et al. 2010). Furthermore, we detected a genetic differentiation that serves as an indicator of the mentioned long-term isolation of the recent populations within geographically disconnected refugia (Tóth et al. 2019). For Alaska, there are also indications of a possible in situ persistence of larches during the LGM (Napier et al. 2020). Binney et al. (2017) mention that northern Eurasia is topographically complex so it is likely that the wide range of local climates provided conditions for refugial populations to persist. In the region of the Eastern Yakutian cluster, the 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 provided shelter for the persistence of *Larix* during the LGM.

Our ABC shows that the populations that persisted during the LGM in northern refugia have genetically contributed to post-LGM recolonization. This is also corroborated by other studies that suggest that these populations were established well before the LGM from a single source population (Western Siberia) with probably a small effective size and low recent gene flow (Ma et al. 2020; Semerikov et al. 2013). The population in Chukotka probably originated earlier and the population in Eastern Yakutia subsequently emerged in the course of an admixture event of the populations from Western Yakutia and Chukotka. However, our results contradict the conclusions of Schulte et al. (2022b), who state that *L. sibirica* had to recolonize northern areas from refugia in the south in the postglacial. It is possible that *L. sibirica* retreated locally from the region of the lake investigated in their study during the LGM but survived the LGM in other northern areas. In general, the existing refugia are likely to have strongly assisted the colonization of more northerly areas of the forest-tundra ecotone since they provide a seed source and shelter for recruitment of larch regeneration (Kharuk et al. 2013). During range expansion and reconnection of refugia or at contact areas, hybrids form and the tendency for this process to occur is common in many forest tree species and is known for Siberian larch species (Semerikov et al. 2007).

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

The refugial populations may have served as a starting point for rapid colonization of the areas north of the treeline in the early Holocene (Tarasov et al. 2009; Epp et al. 2018). This explains the existence of *Larix* in the far north at that time (Bigelow 2003), although the migration rates at the treelines were slow. Today the initial situation is different. The climate cooling during the Little Ice Age (LIA; extending from the 16<sup>th</sup> to the 19<sup>th</sup> centuries) negatively impacted tree population densities and caused range contraction, while the enhanced recruitment in the twentieth century has not been of sufficient magnitude to compensate for this range contraction (MacDonald et al. 2008). As a result, there are currently no refugial populations in northern Siberia, as was most likely the case in the early Holocene. This fact could possibly explain why the current treeline advance is lagging behind climate warming. However, if individual trees establish themselves in the tundra area ahead of the treeline in the future, they could be the initial spark for rapid dispersal of the boreal coniferous forest. If the progressive forest expansion keeps pace with climate change in the future, as various studies assume (MacDonald et al. 2008; Pearson et al. 2013; Kruse and Herzsuh 2022), the habitats of tundra are threatened and could recede or disappear completely. This knowledge can be implemented in

simulation models such as LAVESI (*Larix* vegetation simulator) (Kruse et al. 2016; 2019).

## 5 Conclusions and outlook

We inferred spatial distribution patterns of the genetic variability of Siberian larches by GBS. The data are best explained by three and four genetic groups. However, from an ecological point of view, a differentiation of five to six clusters has the potential to reveal admixture regions and not just the main areas in the given region but even, for example, the emergence of two clusters belonging to the same species but differing from one another. According to Bobrov's taxonomic system (1972), the four statistically verified main clusters match well with the expected distinction into the three Siberian larch species *L. sibirica*, *L. gmelinii*, and *L. cajanderi* from Western to Eastern Eurasia. The most eastern cluster is in Chukotka and seems to be another aggregation of *L. cajanderi*. Furthermore, the geographical barriers correspond to the habitat zones of the different species.

Our aim was to answer the question of whether refugia existed in northern areas during the LGM and to get an idea of the temporal classification concerning possible demographic events. Altogether the ABC supports a scenario whereby the present Siberian larch populations have survived the LGM in refugia in the north, rather than migrating in the postglacial from the south. The presence of northern LGM refugia may explain the early existence of larches in the far north in the Holocene and their dominance until today. In contrast to the past situation, there are no northern refugia today, which could delay the treeline advance to the north despite climate warming.

The results of this study provide a better understanding on how refugial populations contribute to the treeline migration of Siberian larches. Furthermore, cluster analysis could be used to search for possible refugial populations on a small scale for conservation purposes. Additionally, more complex scenarios, complementary to the most probable scenario detected in the present study, could be analyzed using the ABC method.

## References

- Abaimov, A. P. (2010). Geographical distribution and genetics of Siberian larch species. In A. Osawa, O. A. Zyryanova, Y. Matsuura, T. Kajimoto, & R. W. Wein (Eds.), *Permafrost ecosystems: Siberian larch forests* (Ecological Studies Vol. 209, pp. 41–58). Springer, Dordrecht. [https://doi.org/10.1007/978-1-4020-9693-8\\_3](https://doi.org/10.1007/978-1-4020-9693-8_3)
- Andreev, A. A., Nazarova, L. B., Lenz, M. M., Böhmer, T., Syrykh, L., Wagner, B., Melles, M., Pestryakova, L. A., & Herzschuh, U. (2022). Late Quaternary paleoenvironmental reconstructions from sediments of Lake Emanda (Verkhoyansk Mountains, East Siberia). *Journal of Quaternary Science*, 37(5), 884–899. <https://doi.org/10.1002/jqs.3419>
- Andreev, A. A., Schirrmeister, L., Tarasov, P. E., Ganopolski, A., Brovkin, V., Siebert, C., Wetterich, S., & Hubberten, H.-W. (2011). Vegetation and climate history in the Laptev Sea region (Arctic Siberia) during Late Quaternary inferred from pollen records. *Quaternary Science Reviews*, 30(17–18), 2182–2199. <https://doi.org/10.1016/j.quascirev.2010.12.026>
- Araki, N. H. T., Khatab, I. A., Hemamali, K. K. G. U., Inomata, N., Wang, X.-R., & Szmidt, A. E. (2008). Phylogeography of *Larix sukaczewii* Dyl. and *Larix sibirica* L. inferred from nucleotide variation of nuclear genes. *Tree Genetics & Genomes*, 4(4), 611–623. <https://doi.org/10.1007/s11295-008-0137-1>
- Baltunis, B. S., Greenwood, M. S., & Eysteinsson, T. (1998). Hybrid vigor in *Larix*: Growth of intra- and interspecific hybrids of *Larix decidua*, *L. laricina*, and *L. kaempferi* after 5 years. *Silvae Genetica*, 47(5–6), 288–293.
- Bennett, K., & Provan, J. (2008). What do we mean by ‘refugia’? *Quaternary Science Reviews*, 27(27–28), 2449–2455. <https://doi.org/10.1016/j.quascirev.2008.08.019>
- 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. *Journal of Geophysical Research*, 108(D19), 8170. <https://doi.org/10.1029/2002JD002558>

- 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., Väliänta, M., & Vazhenina, L. (2009). The distribution of late-Quaternary woody taxa in northern Eurasia: Evidence from a new macrofossil database. *Quaternary Science Reviews*, 28(23–24), 2445–2464. <https://doi.org/10.1016/j.quascirev.2009.04.016>
- 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., & Zernitskaya, V. (2017). Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. *Quaternary Science Reviews*, 157, 80–97. <https://doi.org/10.1016/j.quascirev.2016.11.022>
- Bobrov, E. G. (1972). History and systematics of *Larix* . *Journal of Botany*, 60(6), 797–805.
- Bonan, G. B. (2008). Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science*, 320(5882), 1444–1449. <https://doi.org/10.1126/science.1155121>
- Borsch, T., Berendsohn, W., Dalcin, E., Delmas, M., Demissew, S., Elliott, A., Fritsch, P., Fuchs, A., Geltman, D., Güner, A., Haevermans, T., Knapp, S., Roux, M. M., Loizeau, P., Miller, C., Miller, J., Miller, J. T., Palese, R., Paton, A., ... Zamora, N. (2020). World Flora Online: Placing taxonomists at the heart of a definitive and comprehensive global resource on the world’s plants. *TAXON*, 69(6), 1311–1341. <https://doi.org/10.1002/tax.12373>
- Brubaker, L. B., Anderson, P. M., Edwards, M. E., & Lozhkin, A. V. (2005). Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapped pollen data. *Journal of Biogeography*, 32(5), 833–848. <https://doi.org/10.1111/j.1365-2699.2004.01203.x>
- Cao, X., Tian, F., Andreev, A., Anderson, P. M., Lozhkin, A. V., Bezrukova, E., Ni, J., Rudaya, N., Stobbe, A., Wicczorek, M., & Herzschuh, U. (2020). A taxonomically harmonized and temporally standardized fossil pollen dataset from Siberia covering the last 40 kyr. *Earth System Science Data*, 12(1), 119–135. <https://doi.org/10.5194/essd-12-119-2020>
- Cao, X., Tian, F., Li, F., Gaillard, M.-J., Rudaya, N., Xu, Q., & Herzschuh, U. (2019). Pollen-based quantitative land-cover reconstruction for northern Asia covering the last 40 ka cal BP. *Climate of the Past*, 15(4), 1503–1536. <https://doi.org/10.5194/cp-15-1503-2019>
- Chen, C., Mitchell, S. E., Elshire, R. J., Buckler, E. S., & El-Kassaby, Y. A. (2013). Mining conifers’ megagenome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform. *Tree Genetics & Genomes*, 9(6), 1537–1544. <https://doi.org/10.1007/s11295-013-0657-1>
- Cornuet, J.-M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J.-M., & Estoup, A. (2014). DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, 30(8), 1187–1189. <https://doi.org/10.1093/bioinformatics/btt763>
- Deschamps, S., Llacá, V., & May, G. D. (2012). Genotyping-by-Sequencing in Plants. *Biology*, 1(3), 460–483. <https://doi.org/10.3390/biology1030460>
- Dong, M., He, Q., Zhao, J., Zhang, Y., Yuan, D., & Zhang, J. (2019). Genetic Mapping of Prince Rupprecht’s Larch (*Larix principis-rupprechtii* Mayr) by Specific-Locus Amplified Fragment Sequencing. *Genes*, 10(8), 583. <https://doi.org/10.3390/genes10080583>
- Dylis, N. V. (1961). Larch species of East Siberia and the Far East. AN SSSR, Moscow, 209.
- Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, 36(8), 2592–2594. <https://doi.org/10.1093/bioinformatics/btz966>



- Eckert, A. J., Bower, A. D., González-Martínez, S. C., Wegrzyn, J. L., Coop, G., & Neale, D. B. (2010). Back to nature: Ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology*, 19(17), 3789–3805. <https://doi.org/10.1111/j.1365-294X.2010.04698.x>
- Epp, L. S., Kruse, S., Kath, N. J., Stoof-Leichsenring, K. R., Tiedemann, R., Pestryakova, L. A., & Herzsuh, U. (2018). Temporal and spatial patterns of mitochondrial haplotype and species distributions in Siberian larches inferred from ancient environmental DNA and modeling. *Scientific Reports*, 8(1), 17436. <https://doi.org/10.1038/s41598-018-35550-w>
- Francois, O. (2016). Running Structure-like Population Genetic Analyses with R. University of Grenoble-Alpes, 1–8.
- Franz, H. J. (1973). *Physische Geographie der Sowjetunion*. Haack.
- Freeland, J. R. (2020). *Molecular ecology*. John Wiley & Sons.
- Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8), 925–929. <https://doi.org/10.1111/2041-210X.12382>
- Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196(4), 973–983. <https://doi.org/10.1534/genetics.113.160572>
- Harsch, M. A., Hulme, P. E., McGlone, M. S., & Duncan, R. P. (2009). Are treelines advancing? A global meta-analysis of treeline response to climate warming. *Ecology Letters*, 12(10), 1040–1049. <https://doi.org/10.1111/j.1461-0248.2009.01355.x>
- Herzschuh, U. (2020). Legacy of the Last Glacial on the present-day distribution of deciduous versus evergreen boreal forests. *Global Ecology and Biogeography*, 29(2), 198–206. <https://doi.org/10.1111/geb.13018>
- Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), 907–913. <https://doi.org/10.1038/35016000>
- Holtmeier, F. K., & Broll, G. E. (2007). Treeline advance—Driving processes and adverse factors. *Landscape Online*, 1, 1–33. <https://doi.org/10.3097/LO.200701>
- 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 [H.-O. Portner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegria, M. Craig, S. Langsdorf, S. Loschke, V. Moller, A. Okem, B. Rama (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., <https://doi.org/10.1017/9781009325844>.
- Isaev, A. P., Protopopov, A. V., Protopopova, V. V., Egorova, A. A., Timofeyev, P. A., Nikolaev, A. N., Shurduk, I.F., Lytkina, L.P., Ermakov, N.B., Nikitina, N.V., Efimova, A.P., Zakharova, V.I., Cherosov, M.M., Nikolin, E.G., Sosina, N.K., Troeva, E.I., Gogoleva, P.A., Kuznetsova, L.V., Pestryakov, B.N., Mironova, S.I., and Sleptsova, N. P. (2010). Vegetation of Yakutia: elements of ecology and plant sociology. In E. I. Troeva, A. P. Isaev, M. M. Cherosov, & N. S. Karpov (Eds) *The Far North: plant biodiversity and ecology of Yakutia* (pp. 143–260), Springer. [https://doi.org/10.1007/978-90-481-3774-9\\_3](https://doi.org/10.1007/978-90-481-3774-9_3)
- Johnson, J., Chhetri, P., Krutovsky, K., & 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(11), 418. <https://doi.org/10.3390/f8110418>
- Kajimoto, T. (2010). Root system development of larch trees growing on Siberian permafrost. In A. Osawa, O. A. Zyryanova, Y. Matsuura, T. Kajimoto, & R. W. Wein (Eds.), *Permafrost ecosystems: Siberian larch forests* (Ecological Studies Vol. 209, pp. 303–330), Springer, Dordrecht. [https://doi.org/10.1007/978-1-4020-9693-8\\_16](https://doi.org/10.1007/978-1-4020-9693-8_16)
- Kharuk, V. I., Ranson, K. J., Im, S. T., Oskorbin, P. A., Dvinskaya, M. L., & Ovchinnikov, D. V. (2013). Tree-line structure and dynamics at the northern limit of the larch forest: Anabar Plateau, Siberia, Russia.

- Arctic, Antarctic, and Alpine Research, 45(4), 526–537. <https://doi.org/10.1657/1938-4246-45.4.526>
- Khatab, I. A., Ishiyama, H., Inomata, N., Wang, X.-R., & Szmidt, A. E. (2008). Phylogeography of Eurasian *Larix* species inferred from nucleotide variation in two nuclear genes. *Genes & Genetic Systems*, 83(1), 55–66. <https://doi.org/10.1266/ggs.83.55>
- Kruse, S., Gerdes, A., Kath, N. J., Epp, L. S., Stoof-Leichsenring, K. R., Pestryakova, L. A., & Herzsuh, U. (2019). Dispersal distances and migration rates at the arctic treeline in Siberia – a genetic and simulation-based study. *Biogeosciences*, 16(6), 1211–1224. <https://doi.org/10.5194/bg-16-1211-2019>
- Kruse, S., Gerdes, A., Kath, N. J., & Herzsuh, U. (2018). Implementing spatially explicit wind-driven seed and pollen dispersal in the individual-based larch simulation model: LAVESI-WIND 1.0. *Geoscientific Model Development*, 11(11), 4451–4467. <https://doi.org/10.5194/gmd-11-4451-2018>
- Kruse, S., & Herzsuh, U. (2022). Regional opportunities for tundra conservation in the next 1000 years. *ELife*, 11, e75163. <https://doi.org/10.7554/eLife.75163>
- Kruse, S., Wieczorek, M., Jeltsch, F., & Herzsuh, U. (2016). Treeline dynamics in Siberia under changing climates as inferred from an individual-based model for *Larix*. *Ecological Modelling*, 338, 101–121. <https://doi.org/10.1016/j.ecolmodel.2016.08.003>
- LePage, B. A., & Basinger, J. F. (1995). The evolutionary history of the genus *Larix* (Pinaceae). U.S. Department of Agriculture, Forest Service, Intermountain Research Station, GTR-INT-31 (pp. 19–219).
- Li, Y., Zhang, X., & Fang, Y. (2019). Landscape features and climatic forces shape the genetic structure and evolutionary history of an oak species (*Quercus chenii*) in East China. *Frontiers in Plant Science*, 10, 1060. <https://doi.org/10.3389/fpls.2019.01060>
- Lozhkin, A., Anderson, P., Minyuk, P., Korzun, J., Brown, T., Pakhomov, A., Tsygankova, V., Burnatny, S., & 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(3), 938–953. <https://doi.org/10.1111/bor.12316>
- Ma, L., Cao, L., Hoffmann, A. A., Gong, Y., Chen, J., Chen, H., Wang, X., Zeng, A., Wei, S., & Zhou, Z. (2020). Rapid and strong population genetic differentiation and genomic signatures of climatic adaptation in an invasive mealybug. *Diversity and Distributions*, 26(5), 610–622. <https://doi.org/10.1111/ddi.13053>
- MacDonald, G. M., Kremenetski, K. V., & Beilman, D. W. (2008). Climate change and the northern Russian treeline zone. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1501), 2283–2299. <https://doi.org/10.1098/rstb.2007.2200>
- 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., & Gattaulin, V. N. (2000). Holocene treeline history and climate change across northern Eurasia. *Quaternary Research*, 53(3), 302–311. <https://doi.org/10.1006/qres.1999.2123>
- Muller, S., Tarasov, P. E., Andreev, A. A., Tutken, T., Gartz, S., & 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. *Quaternary Science Reviews*, 29(17–18), 2071–2086. <https://doi.org/10.1016/j.quascirev.2010.04.024>
- Mayr, H. (1906). *Fremdlandische Wald-und Parkbaume fur Europa*. P. Parey.
- Napier, J. D., Fernandez, M. C., de Lafontaine, G., & Hu, F. S. (2020). Ice-age persistence and genetic isolation of the disjunct distribution of larch in Alaska. *Ecology and Evolution*, 10(3), 1692–1702.
- Pearson, R. G., Phillips, S. J., Loranty, M. M., Beck, P. S. A., Damoulas, T., Knight, S. J., & Goetz, S. J. (2013). Shifts in Arctic vegetation and associated feedbacks under climate change. *Nature Climate Change*, 3(7), 673–677. <https://doi.org/10.1038/nclimate1858>

- Petit, R. J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D., & Vendramin, G. G. (2004). Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, 14(3), 689–701. <https://doi.org/10.1111/j.1365-294X.2004.02410.x>
- Polezhaeva, M. A., Lascoux, M., & Semerikov, V. L. (2010). Cytoplasmic DNA variation and biogeography of *Larix* Mill. in northeast Asia. *Molecular Ecology*, 19(6), 1239–1252. <https://doi.org/10.1111/j.1365-294X.2010.04552.x>
- 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., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3), 559–575. <https://doi.org/10.1086/519795>
- Railsback, L. B. (2006). Some fundamentals of mineralogy and geochemistry. On-line book, quoted from: <http://railsback.org/FundamentalsIndex.html>.
- Rousset, F. (1999). Genetic differentiation within and between two habitats. *Genetics*, 151(1), 397–407. <https://doi.org/10.1093/genetics/151.1.397>
- Rowe, G., Sweet, M., & Beebe, T. J. C. (2017). *An Introduction to Molecular Ecology*. Oxford University Press.
- Schnell, R. J., & 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., & Herzschuh, U. (2021). Hybridization capture of larch (*Larix* Mill.) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest. *Molecular Ecology Resources*, 21(3), 801–815. <https://doi.org/10.1111/1755-0998.13311>
- Schulte, L., Li, C., Lisovski, S., & Herzschuh, U. (2022a). Forest-permafrost feedbacks and glacial refugia help explain the unequal distribution of larch across continents. *Journal of Biogeography*, 49(10), 1825–1838. <https://doi.org/10.1111/jbi.14456>
- 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., Krukovsky, K. V., & Herzschuh, U. (2022b). *Larix* species range dynamics in Siberia since the Last Glacial captured from sedimentary ancient DNA. *Communications Biology*, 5(1), 570. <https://doi.org/10.1038/s42003-022-03455-0>
- Semerikov, V. L., Iroshnikov, A. I., & Lascoux, M. (2007). Mitochondrial DNA variation pattern and postglacial history of the Siberian larch (*Larix sibirica* Ledeb.). *Russian Journal of Ecology*, 38(3), 147–154. <https://doi.org/10.1134/S1067413607030010>
- Semerikov, V. L., Semerikov, L. F., & Lascoux, M. (1999). Intra- and interspecific allozyme variability in Eurasian *Larix* Mill. species. *Heredity*, 82(2), 193–204. <https://doi.org/10.1038/sj.hdy.6884710>
- Semerikov, V. L., Semerikova, S. A., Polezhaeva, M. A., Kosintsev, P. A., & 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. *Molecular Ecology*, 22(19), 4958–4971. <https://doi.org/10.1111/mec.12433>
- Smith, D. R. (2015). Mutation rates in plastid genomes: they are lower than you might think. *Genome Biology and Evolution*, 7(5), 1227–1234. <https://doi.org/10.1093/gbe/evv069>
- Stewart, J. R., & Lister, A. M. (2001). Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution*, 16(11), 608–613. [https://doi.org/10.1016/S0169-5347\(01\)02338-2](https://doi.org/10.1016/S0169-5347(01)02338-2)

Szafer, W. (1913). Contribution to the knowledge of Eurasian larch species with particular attention paid to larch species occurring in Poland. *Kosmos* 38: 1021, 1281–1315 (in Polish).

Tarasov, P., Muller, S., Andreev, A., Werner, K., & Diekmann, B. (2009). Younger Dryas *Larix* in eastern Siberia: A migrant or survivor? *PAGES News*, 17(3), 122–123. <https://doi.org/10.22498/pages.17.3.122>

Toth, E. G., Tremblay, F., Housset, J. M., Bergeron, Y., & 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 Evolutionary Biology*, 19(1), 190. <https://doi.org/10.1186/s12862-019-1510-4>

Tsumura, Y., Kimura, M., Nakao, K., Uchiyama, K., Ujino-Ihara, T., Wen, Y., Tong, Z., & Han, W. (2020). Effects of the last glacial period on genetic diversity and genetic differentiation in *Cryptomeria japonica* in East Asia. *Tree Genetics & Genomes*, 16(1), 19. <https://doi.org/10.1007/s11295-019-1411-0>

Valiranta, M., Kaakinen, A., Kuhry, P., Kultti, S., Salonen, J. S., & Seppa, 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. *Journal of Biogeography*, 38(5), 922–932. <https://doi.org/10.1111/j.1365-2699.2010.02448.x>

Wendler, N., Mascher, M., Noh, C., Himmelbach, A., Scholz, U., Ruge-Wehling, B., & Stein, N. (2014). Unlocking the secondary gene-pool of barley with next-generation sequencing. *Plant Biotechnology Journal*, 12(8), 1122–1131.

Zhang, N., Yasunari, T., & Ohta, T. (2011). Dynamics of the larch taiga–permafrost coupled system in Siberia under climate change. *Environmental Research Letters*, 6(2), 024003. <https://doi.org/10.1088/1748-9326/6/2/024003>

Zhang, W., Miller, P. A., Smith, B., Wania, R., Koenigk, T., & Doscher, R. (2013). Tundra shrubification and tree-line advance amplify arctic climate warming: Results from an individual-based dynamic vegetation model. *Environmental Research Letters*, 8(3), 034023. <https://doi.org/10.1088/1748-9326/8/3/034023>

## Appendices

Appendix S1: Sample list (submit Excel table)

Appendix S2: Ipyrad parameter file (submit txt file)

Appendix S3: geno file (submit geno file)

Appendix S4: Final sample stats summary (submit txt file)

Appendix S5: Pie charts: Map of proportion of genetic types underpinning the four clusters (K=4) identified. (submit as PDF)

Appendix S6: DIYABC Analysis results, estimate posterior distribution of parameters for scenario 9 (submit as PDF)