Mitochondrial DNA in Siberian conifers indicates multiple post-glacial colonization centers

Vladimir L. Semerikov[1*] e-mail: semerikov@ipae.uran.ru

Svetlana A. Semerikova[1] e-mail: s.a.semerikova@ipae.uran.ru

Yuliya A. Putintseva[2] e-mail: yaputintseva@mail.ru

Natalia V. Oreshkova[2,3] e-mail: oreshkova@ksc.krasn.ru

Konstantin V. Krutovsky[2,4,5,6] e-mail: konstantin.krutovsky@forst.uni-goettingen.de

[1]Institute of Plant and Animal Ecology, Ural Branch of the Russian Academy of Sciences, 620144 Ekaterinburg, Russia

[2]Laboratory of Forest Genomics, Genome Research and Education Center, Siberian Federal University, 660036 Krasnoyarsk, Russia

[3]Laboratory of Forest Genetics and Selection, V. N. Sukachev Institute of Forest, Federal Research Center "Krasnoyarsk Science Center" of the Siberian Branch of the Russian Academy of Sciences, 660036 Krasnoyarsk, Russia

[4]Laboratory of Population Genetics, N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, 119333 Moscow, Russia

[5]Department of Forest Genetics and Forest Tree Breeding, Georg-August University of Göttingen, 37077 Göttingen, Germany

[6]Department of Ecosystem Science and Management, Texas A&M University, College Station, TX 77843-2138, USA

*Corresponding author

Abstract

The geographic variation of the mitochondrial DNA was studied in Siberian fir using the newly developed markers and compared with the phylogeographic pattern of another previously studied Siberian coniferous - Siberian larch. Similar to Siberian larch the distribution of mtDNA haplotypes in Siberian fir revealed clear differentiation among distinct geographic regions of southern Siberia and the Urals, likely indicating post-glacial re-colonization from several sources. The northern part of the range of both species was genetically homogeneous, which is probably due to its recent colonization from one of the glacial refugia. This conclusion is in agreement with published pollen and macrofossil data in Siberian fir and with the reconstruction of environmental niches indicating a dramatic reduction of the range and a likely survival of fir in certain southern areas during the last glacial maximum (LGM) – 21 thousand years ago (kya). Although the modeling of Siberian larch ecological niche reconstructed a shift of the range to the south at that period, the paleontological data indicated the presence of this species in most areas of the current range during LGM, that corresponds to the results of previous historical demography study suggesting the population expansion preceding the LGM.

***Key words*:** mitochondrial DNA, NGS, phylogeography, *Abies sibirica, Larix sibirica,* refugia, pollen data, macrofossils, environmental niche modelling

1. Introduction

The response of the boreal forest species to the Pleistocene glacial cycles was different from that of the temperate forests. During the last glacial maximum (LGM), their ranges were not dramatically reduced, and they were able to survive near the glacial sheet (Huntley and Birks 1983; Willis and van Andel 2004). Unlike Europe, Siberia during the Pleistocene was less exposed to the influence of glacial cover, but the glacial sheet reached the latitude of 60º N by about 250-270 thousand years ago (kya) and the latitude of 62º N by about 130-190 kya (Volkova et al. 2002). Consequently, the recolonization of northern Siberia by woody species could occur no earlier than those glacial intervals, but likely earlier than the LGM. Although the paleontological and genetic data indicate a relatively recent settlement of the northern part of the range of larch *Larix gmelinii* (Rupr.) Rupr., *L. cajanderi* Mayr (Polezhaeva et al. 2010), *L. sibirica* Ledeb. (Semerikov et al. 2013), common juniper *Juniperus communis* L. (Hantemirova et al. 2017), wood lemming *Myopus schisticolor* Lilljeborg (Fedorov et al. 2008), and others (Goropashnaya et al. 2004; Oshida et al. 2005; Zink et al. 2002; Kohli, 2015) the age of those colonizations is older than the age of LGM, during which the species probably survived in numerous northern micro-refugia.

Siberian fir (*Abies sibirica* Ledeb.) is more demanding for temperature and humidity than other taiga trees (Krylov et al. 1986), which can lead to a specific reaction of this species to Quaternary climate fluctuations. Siberian fir fossils related to the Late Pleistocene are rare even in the southern part of the range, which makes it difficult to determine the location of glacial refugia. Previous range-wide studies of genetic diversity of Siberian fir were based on allozymes (Semerikova and Semerikov 2006), chloroplast microsatellite loci (Semerikova and Semerikov 2007), and AFLP (Semerikova and Semerikov 2011). They revealed several genetically distinct geographic groups, which were probably the result of post-glacial dispersion out of a few isolated refugia. Such refugia were hypothesized in South

Siberia (Altai Mountains, Sayan Mountains, the Baikal Lake area and the South Urals). Northern Siberia and the Northern Urals were suggested to have been colonized primarily from hypothetical refugia located in the Baikal Lake area (Semerikova and Semerikov, 2011).

A study of the phylogeography of Siberian larch, another representative of the Pinaceae family in the flora of Siberia and Eastern Europe, which has the range close to that of Siberian fir, using mitochondrial and chloroplast DNA markers, also revealed a few geographic groups of populations (Semerikov et al. 2013). Similarly, these groups can be regarded as the result of dispersion out of several refugia located in the Urals and in South Siberia. Possibly, two more refugia existed in more northern areas in the middle of West Siberia (Semerikov et al. 2013). Based on the similarity in the mitotype distribution in northern Siberia and the territory in the northern foothills of the Sayan Mountains, it was concluded that the latter area was the primary source of recolonization of northern Siberia.

Comparison of the phylogeography of Siberian fir and Siberian larch helps to identify common features and differences in the history of the modern populations, to reveal glacial refugia, and time and direction of re-colonization. For this purpose, this study of the Siberian fir phylogeography was conducted using the markers of the mitochondrial DNA (mtDNA), which is maternally inherited in the Pinaceae family and transmitted via seeds, unlike the paternally inherited chloroplast DNA transmitted via pollen and the biparentally inherited nuclear DNA transmitted by both seeds and pollen (Neale and Sederoff 1989). Due to this property, the mtDNA markers are especially informative for describing migrations associated with seed dispersion, including recolonization from glacial refugia. During the preliminary study, we did not detect any variation in the fragments of the mtDNA amplified using "universal primers" based on conservative annealing sites in mitochondrial genes of plants (Demesure et al. 1995; Dumolin-Lapegue at al. 1997), also there were no publicly available mtDNA markers specific for the Siberian fir, therefore we used NGS data to develop four new mtDNA markers.

In addition to the genetic data we analyzed distribution of Siberian fir and Siberian larch during and after the LGM using available paleontological data and also conducted environmental niche modelling to reconstruct the expected ranges of these species during the LGM.

The main aims of this study were to test the hypotheses regarding the location of glacial refugia, time and routes of post-glacial migrations of Siberian fir and to compare the observed biogeographic pattern with one found in other Siberian conifer – *Larix sibirica*.

2. Materials and methods

2.1. Development of mtDNA markers

To develop mtDNA markers, we searched for polymorphism in the mitochondrial genome (mitogenome) of Siberian fir. The approach included the following steps:

1) Relatively low coverage paired-end (PE) sequencing of the entire Siberian fir genome using Illumina HiSeq 2000. For this sequencing we used the PE DNA library with the insert size of 200 bp produced using total DNA isolated with the CTAB method from needles of a single Siberian fir tree growing in a natural population (56° 39' N 59° 16' E). The library preparation was performed following a standard Illumina protocol (www.bu.edu/iscf/files/2011/05/TruSeq_DNA_SamplePrep_Guide_15005180_A.pdf). For sequencing we used 2×100 cycles Illumina Kit. In total, 22,821,847 pairs of reads were generated. We used FastQC and Trimmomatic for quality control and adapter trimming.

2) Assembly of contigs using the CLC Assembly Cell software. The expected genome size for *Abies sibirica* is 15.452 Gbp (Ohri and Khoshoo 1986). Because of the low coverage, the genome assembly was very partial and included only 0.2% of the expected genome size. However, due to the fact that

109 there are multiple copies of the mitogenome per nuclear genome in one cell, we were able to identify
110 several mitochondrial contigs.
111 3) Search for mitochondrial contigs using BLASTn and all plant mitochondrial sequences available
112 in the NCBI Genbank and other public databases, such as ftp://plantgenie.org/ConGenIE and
113 https://treegenesdb.org/FTP/Genomes/Pita/mito for Norway spruce (*Picea abies* (L.) Karst.) and loblolly
114 pine (*Pinus taeda* L.), respectively. The matching Siberian fir contigs in the BLASTn hits with the
115 alignment length of more than 100 bp and similarity higher than 90% were selected for further analysis.
116 In total, 87 contigs with the total length of 958,226 bp were selected, which represents a significant part
117 of the mitochondrial genome, considering 5.9 Mb of the mitochondrial genome assembled in another
118 conifer *Picea glauca* (Moench) Voss (Jackman et al. 2016). The selected mtDNA contigs were then used
119 to design PCR primers and to search for polymorphism by partial amplicon-based resequencing of eight
120 individuals representing different parts of the Siberian fir range including Altai, Kuznetsk Alatau and the
121 Sayan Mountains, the Lake Baikal Region, the Southern and the Northern Urals (Table S1 in the
122 Supplementary material). The PCR primers (Table 1) were designed using the Primer3 software (Rozen
123 and Skaletsky 2000).
124 The PCR was performed in a volume of 25 ul, containing about 250 ng of genomic DNA, 1X PCR
125 buffer (75 mm Tris-HCl, 20 mM $(NH_4)_2SO_4$, 0.1% Tween-20), 2.5 mM $MgCl_2$, 200µM of each dNTP,
126 0.2µM forward and reverse primers, 0.32 units of *Taq* polymerase (SibEnzyme Ltd., Novosibirsk,
127 Russia). The PCR program consisted of initial denaturation at 94ºC for 5 min and 35 cycles of
128 amplification: 94ºC – 30 sec, 60ºC – 45 sec, 72ºC – 2 min. The final elongation was 7 min at 72ºC. The
129 PCR product was checked using electophoresis in 1% agarose gel, purified using ExoSAP-IT®
130 (Affimetrix Inc., Santa Clara, CA, USA) and then sequenced using the BigDye v.3.1. kit and
131 GeneAnalyser 3130 (Applied Biosystems, Thermo Fisher Scientific Inc., Waltham, MA, USA). The
132 obtained nucleotide sequences were edited and aligned using CodonCode v. 1.2.4 and BioEdit v. 7.2.5

133 (Hall 1999). Four single nucleotide polymorphisms (SNPs) were detected in four different contigs,
134 respectively (Table. 2). All four SNPs were biallelic. The identified SNPs were further used as genetic
135 markers.

136 **2.2. Genotyping**

137 For routine genotyping, the SSCP (single strain conformation polymorphism) method (Fujita and
138 Silver 1994) was applied with minor modifications. For better SSCP resolution, additional PCR primers
139 were developed to amplify shorter fragments (less than 250 bp) containing SNPs (Table 1). All four
140 fragments were amplified in a single 10 µl multiplex reaction. Its composition and PCR conditions were
141 identical to the described above, except for a 1 min shorter elongation time. The PCR product was further
142 subjected to digestion with the restriction enzyme *Rsa*I in order to obtain shorter fragments containing
143 polymorphism. As a result, the restriction fragment containing A37 was 95 bp in length, and the
144 restriction fragment containing marker A126 – 130 bp. The restriction product was 5X diluted with a
145 loading buffer containing 95% formamide and denatured at 95° C for 3 minutes before electrophoresis
146 in a 8% polyacrylamide gel and 1X TBE electrode buffer containing 10% glycerol. The gel and buffer
147 were pre-cooled in a refrigerator to 0-4ºC, and the electrophoresis was carried out at 4ºC. The electric
148 power was stabilized at 15 watts, and the electrophoresis was run at 4000 volts × hours in the Model S2
149 sequencing Gel Electrophoresis System (Applied Biosystems, Thermo Fisher Scientific Inc., Waltham,
150 MA, USA). The DNA in the gel was visualized after the electrophoresis by silver staining. An example
151 of an SSCP gel and observed polymorphism is demonstrated in Fig. S1 (Supplementary material).

152 **2.3. Plant material**

153 Samples from 45 populations representing 8 - 24 individuals per population (Fig. 1; Table S1 in the
154 Supplementary material) used previously for allozyme, chloroplast microsatellite, and AFLP studies

155 (Semerikova and Semerikov 2006, 2007, 2011) were genotyped using the SSCP method to study mtDNA
156 diversity. In addition, for verification the mtDNA fragments were sequenced in at least one individual
157 for each detected mitotype in each studied population.

158 **2.4. Population genetic differentiation analysis**

159 A hierarchical analysis of molecular variation (AMOVA) within and between populations, and
160 within and between groups was performed using the Arlequin program v.3.5 (Excoffier et al. 2006). The
161 statistical significance of the fixation indices was estimated using 1000 permutations. The populations
162 were grouped based on their clustering into geographic groups using the SAMOVA program (Dupanloup
163 et al. 2002). The algorithm of the program is aimed at clustering geographically adjacent populations in
164 K groups, where K is set a priori, by maximizing differentiation between groups (F_{CT}). The analysis was
165 performed at K = 2, 3, 4, 5, 6. The G_{ST} (Nei 1987) and N_{ST} fixation indices were also calculated based
166 only on haplotype frequencies or taking into account also the genetic distance between the haplotypes
167 (Pons and Petit 1996), respectively. The comparison of G_{ST} with N_{ST} was carried out using PermutCpSSR
168 v.1.0 (http://www6.bordeaux-aquitaine.inra.fr/biogeco/Production-scientifique/Logiciels/Contrib-
169 Permut/Permut) (Burban et al. 1999). If $N_{ST} > G_{ST}$, genetically similar haplotypes tend to coexist in the
170 same population.

171 **2.5. Phylogenetic analysis**

172 To investigate the phylogenetic relationships of the identified haplotypes (mitotypes) of *A. sibirica*,
173 we sequenced polymorphic fragments in one sample in each of the two related species (Semerikova et
174 al. 2018): *A. nephrolepis* (Trautv. ex Maxim.) Maxim. (Russian Far East) and *A. semenovii* B. Fedtsch.
175 (Western Tien Shan). The latter was classified (Farjon and Rushforth 1989) as a subspecies of Siberian
176 fir (*A. sibirica* subsp. *semenovii* (B. Fedtsch.) Farjon), but their species rank was confirmed by molecular

177 data (Semerikova et al. 2012; Semerikova and Semerikov 2016). To infer the relationships of mitotypes,
178 we used the method of Median-Joining Network, performed with the software NETWORK v. 5.0.0.1
179 (Bandelt et al., 1999).

180 **2.6. Environmental niche modelling**

181 We used environmental niche modelling to reconstruct putative ranges of Siberian fir and Siberian
182 larch during the LGM (21 kya). To do so we used data on the current distribution of fir (*A. sibirica, A.*
183 *nephrolepis, A. sachalinensis* and *A. semenovii*) and larch (*L. sibirica, L. gmelinii, L. olgensis, L.*
184 *kamtchatica* and *L. kajanderi*) species, as well as present and past climatic parameter distributions and
185 the machine learning method based on maximum entropy implemented in the program MAXENT 3.3.3
186 (Phillips et al. 2006). The fir and larch distribution data were retrieved from the Global Bioinformation
187 Facility database (https://www.gbif.org, accessed on June 17, 2018). The data set was expanded by
188 adding 170 occurrences of larch from our field records (Polezhaeva et al. 2010; Semerikov et al. 2013)
189 and fir occurrences from this study and Semerikova et al. (2011). The environmental data describing the
190 baseline climate (19 BioClim layers for the 1950–2000 period at a spatial resolution of 2.5 arc min), the
191 LGM climate (BioClim layers derived from the Coupled Model Intercomparison Project Phase 5) were
192 retrieved from the WorldClim database (Hijmans et al. 2005). To reduce the effect of association between
193 climate parameters, we computed correlation between all pairs of the 19 parameters for the geographic
194 points of the species occurrence. For the parameters with correlation 0.8 or more we presented only one
195 parameter. As a result, we used 9 layers: bio1 - Annual Mean Temperature, bio2 - Mean Diurnal Range,
196 bio3 - Isothermality, bio4 - Temperature Seasonality, bio5 - Max Temperature of Warmest Month, bio8 -
197 Mean Temperature of Wettest Quarter, bio12 - Annual Precipitation, bio15 - Precipitation Seasonality,
198 bio18 - Precipitation of Warmest Quarter. Default settings of MAXENT were used.

2.7. Paleodata

Published paleodata on pollen (Binney et al. 2017) and macrofossils (Binney et al. 2009; Kosintsev et al. 2012) were used to test the inferences from genetic data on the history of species distribution after the LGM. We selected data sites in the range of latitude 41°- 75° and longitude 29° - 177°. To reduce the number of erroneously interpreted cases of fir presence due to redeposition of pollen from older layers or long-distance pollen dispersion, we selected only cases with the proportion of pollen of the species in question of above 1% in the database. Since larch pollen has poor preservation and insufficient mobility, for larch we took into account all the samples where larch pollen was noted. We selected the pollen and macrofossil data younger than 21500 year old and combined them according to calibrated radiocarbon age into the following eight categories: 0 - 0.5, 0.5 - 3.5, 3.5 - 6.5, 6.5 - 9.5, 9.5 - 12.5, 12.5 - 15.5, 15.5 - 18.5, and 18.5 - 21.5 kya.

3. Results

3.1. New markers revealed a strong spatial structure of the mtDNA variation in *A. sibirica*

Sequencing of randomly selected 33 regions of 20 Siberian fir mtDNA contigs in eight trees from geographically distinct populations with the total length of 49,000 bp identified four biallelic SNPs (A167, A65, A126 and A37) in four contigs – 167, 65, 126, and 37, respectively (Table 1). The study of mtDNA variation in 45 populations of Siberian fir revealed three mitotypes different by 1-4 nucleotides: two of them were relatively frequent – mitotype M1 (GACC haplotype, according to the nucleotide alleles in A167, A65, A126, and A37 SNPs, respectively) and mitotype M2 (TCAC). M3 (GACA) was rare and different from M1 by one SNP (Fig. 1). M4 was found only in *A. semenovii* (GCCC) and M5 – only in *A. nephrolepis* (GCAC). Apart from four SNPs polymorphic in *A. sibirica,* one more mutation was found specific to mitotype M4 of *A. semenovii* and seven mutations specific to M5 of *A. nephrolepis*

(Fig. 1). A closer relationship between *A. sibirica* and *A. semenovii* compared with *A. nephrolepis* is expected, considering that, based on the previously used mitochondrial DNA fragments, *A. sibirica* and *A. nephrolepis* have different, albeit closely similar haplotypes, while no differences between *A. semenovii* and *A. sibirica* were found (Semerikova et al. 2018). Most of the studied populations of *A. sibirica* contained only one mitotype (Fig. 1). M2 was fixed in the Baikal region, north-east of the Eastern Sayan, the Middle and Lower Yenisei, most of West Siberia and the Northern and Subpolar Urals. M1 was fixed or dominant in the mountain ridges of the Western Sayan, Kuznetsk Alatau and the Altai. These two mitotypes M1 and M2 formed a mosaic structure co-occurring in some populations in the Middle and Southern Urals, the west of the West Siberia Plain and European Russia.

The mitotype M3 was fixed or present as an admixture to M1 in the southernmost populations of the Altai and Kuznetsk Alatau and completely absent in more northern populations. The phylogenetic structure was not pronounced, and N_{ST} was higher than G_{ST} (0.897 vs. 0.866), but the difference was not statistically significant (P = 0.25). At K = 3, SAMOVA divided the populations into three geographic groups according to the areas of the three identified mitotypes. At the same time, F_{ST} was very high (0.933, Table 2), which is typical for mtDNA markers in conifers. The highest differentiation, based on the SAMOVA grouping, was found between populations within the total sample and between groups (Table 2).

3.2. The niche modelling indicates a stronger reduction of the range of *A. sibirica* compared to the *Larix* species during the LGM

The computed *Abies* and *Larix* ranges in Northern Eurasia based on the environment niche modeling for the present day largely coincided with the current ranges of these species (Fig. 2: H and P). The computed area potentially favorable for fir during the LGM was at the mid-latitudes (south of Moscow's latitude – about 55 lat.) (Fig. 2, A). It intermittently stretches from west to east and partly overlaps with

244 the most southern parts of the present range of Siberian fir, such as the Southern Urals, Altai, the Sayans
245 and the south of the Baikal region. This range included also areas where fir species are completely absent
246 now: the plains of Eastern Europe and Central Kazakhstan.

247 The potential area reconstructed for larch during the LGM (Fig. 2, I) was much wider, and the northern
248 limit of larch distribution located much further to the north than for fir, reaching the latitude of Surgut
249 town (about 61 lat.). In the southern and western parts, the range of larch in the LGM period could
250 probably significantly expand into Eastern Europe, Central Kazakhstan and East Asia.

251 **3.3. Peleodata indicates a more limited distribution of fir in the LGM and a later expansion to the**
252 **north after climate improvement in comparison to larch**

253 Pollen and macrofossil records of *Abies* in the time intervals close to the LGM (21.5–18.5 and 18.5–
254 15.5 kya; Fig. 2, A and B) are extremely rare. The pollen records along the coasts of the Kara and Okhotsk
255 seas, as well as the macrofossil sample taken near the mouth of the Yenisei River, seem suspicious, as
256 they are located far beyond the reconstructed range of fir during the LGM and even beyond the modern
257 range. We believe that the wrong time placement of fir pollen in these records may be due to the re-
258 deposition of older pollen, and the wrong time placement of the macrofossil may be the result of an error
259 in determining the radiocarbon age. In the more recent time (12.5–15.5 kya; Fig. 2, C), fir records become
260 more abundant, however, relatively reliable ones are still limited to southern Siberia and the south of the
261 Russian Far East. Near the beginning of the Holocene (10–12 kya; Fig. 2, D), fir suddenly becomes
262 common both in the south of Western Siberia and in the north, where it is noted at several locations.
263 Then, until about the middle of the Holocene, fir remains abundant in northern Siberia, and 7–9 kya it
264 was noted in the center of Yakutia (Fig. 2, E). After 4 kya, fir reduces the area in the north (Fig. 2, G and
265 H). Some pollen deposition corresponding to this time period was found in the north close to the Lena
266 River delta, but it could be the result of re-deposition of older pollen from the early Pleistocene or an

occasional long-distance pollen dispersion. The same trend was observed in the south of the Yamal Peninsula and north of West Siberia where fir pollen was continuously present in the peat deposits beginning from 8 kya and disappeared after 5 kya (Panova et al. 2010). It is interesting to note that similar dynamics of fir was observed in the mountains of southwestern Mongolia, about 600 km beyond the southern limit of the present range, where, according to radiocarbon dating of fir macrofossils (wood) in the Holocene peat deposits, fir was present in the middle Holocene and disappeared approximately after 3.5 kya (Dorofeyuk and Tarasov 2000). Similar, fir appeared in peatlands in Northern Kazakhstan (far to the south of the modern range) in the Holocene optimum and later disappeared (Gorchakovsky 1987).

Unlike fir, during the LGM larch is noted in pollen records in northeastern Siberia, and macrofossils are found in the north of Western Siberia, i. e. much to the north of the area reconstructed by MAXENT (Fig. 2, I), and 18.3 kya, according to macrofossils – on the coast of the Kara Sea. However, a noticeable increase in the amount of pollen and macrofossil records throughout the present range is observed only after 15.5 kya (Fig. 2, K), and especially significantly with the beginning of the Holocene (Fig. 2, L). By the end of the Holocene, larch retreats to the limits of the modern range.

4. Discussion

4.1. NGS facilitates development of mtDNA markers in plants

The maternal inheritance of mtDNA markers makes them very valuable for population and phylogeographic studies, especially in the Pinaceae family, where chloroplast DNA has paternal inheritance. MtDNA markers could be useful also in other plants, because plant mitogenomes are much larger than animal ones ranging from 208 kb in *Brassica hirta* to 11.3 Mb in *Silene conica* (see Liao et al. 2018 for review). However, until recently, the use of mtDNA was limited, since available markers were restricted to a few known intron and intergenic spacer regions amplified by the "universal" PCR

289 primers. For many conifers polymorphism in these regions is either absent or not sufficiently informative.
290 Developing species-specific markers requires *de novo* sequencing of a sufficient part of the mitogenome,
291 which, prior to the appearance of NGS methods, was a time-consuming task. This is exacerbated by the
292 low nucleotide variation in plant mtDNA, much lower than the variation in nuclear and chloroplast DNA.
293 The latter circumstance makes it inefficient to search for variation in plant mitogenomes using the RAD-
294 seq or other methods that involve sequencing of small fractions of the genome. In this study, we used the
295 results of the whole-genome-shotgun (WGS) sequencing of the Siberian fir genome to develop species-
296 specific markers of mtDNA. Since the WGS was based only on one tree, to search for polymorphism we
297 re-sequenced randomly selected portions of the mitogenome in a few trees using the Sanger method. We
298 used the same approach recently to develop seven markers for mtDNA in Scots pine (Semerikov et al.
299 2015, 2018). Unfortunately, this approach implies a large volume of capillary sequencing. The use of
300 NGS instead of the Sanger method for the sequencing of several trees in the search for polymorphism in
301 mtDNA of Scots pine and related species proved to be more effective (Donnelly et al. 2017). Thus, at
302 present the progress of NGS methods makes it relatively inexpensive and very quick to develop mtDNA
303 markers suitable for population and phylogenetic studies of any plant species.

304 **4.2. Colonization of the northern part of the range of Siberian fir and Siberian larch**

305 The histories of Siberian larch and Siberian fir in the LGM and after it apparently differed significantly
306 from each other. According to the reconstruction of ecological niches, pollen and macrofossil data,
307 Siberian larch was present in northern Siberia during the LGM, although treeless landscapes dominated
308 much of Northern Eurasia (Tarasov et al. 2000; Binney et al. 2017). Siberian larch reached the latitude
309 of 72° already by 18.37 kya (Kosintsev et al. 2012). Consequently, the colonization of northern Siberia
310 by larch began before the LGM that agrees with the results of the Bayesian analysis of the historical
311 demography of Siberian larch based on the chloroplast microsatellite data, which gave the estimate of

312 the age of population expansion that significantly exceeded the age of the LGM and probably
313 corresponded to the onset of migration from the refugia after one of the intensive Middle-Pleistocene
314 glaciations (Semerikov et al. 2013). Despite the fact that the structure of the mtDNA variation in Siberian
315 larch corresponds to the colonization of northern Siberia primarily from a single southern refugium,
316 heterogeneity of the mtDNA variation in the northern Siberia supported by the spatial analysis of
317 molecular variation (SAMOVA) suggests post-LGM dispersion out from several secondary refugia
318 located in the north of Western Siberia (Semerikov et al. 2013).

319 In contrast to larch, Siberian fir had a very limited distribution in Siberia and the Urals during the
320 LGM, according to both the paleodata and niche modelling. During the LGM, it apparently could exist
321 only in some mountain and foothill areas of the Southern Urals, the south of Siberia and the Baikal region.
322 A significant increase in the number of samples containing *Abies* pollen or macrofossils occurred after
323 15 kya and only in the southern part of the range. However, by the beginning of the Holocene, Siberian
324 fir had already reached the Kara Sea. The mtDNA data in Siberian fir correspond to the colonization of
325 the north range from a limited area, because only one of the two most common haplotypes is distributed
326 in the north of Western Siberia and the north of the Urals. Such a source could be the Baikal region,
327 where this haplotype is fixed. Data on the allozyme and AFLP variation (Semerikova and Semerikov
328 2011) are compatible with this hypothesis: based on the allele frequencies, the populations of northern
329 Siberia and the northern Urals are similar to Baikal populations (Semerikova and Semerikov 2011).
330 Moreover, higher diversity in populations of southern Siberia and the Baikal region compared to northern
331 Siberia and the Urals was revealed with chloroplast microsatellites (Semerikova and Semerikov 2007),
332 which is in agreement with the hypothesis of migrations from southern refugia. The contribution of only
333 one hypothetical glacial refugium in the recolonization of the North suggested by the mtDNA data
334 contrasts Siberian fir with Siberian larch, for which the existence of several secondary refugia can be
335 assumed in the north of the range during the LGM. This circumstance also distinguishes it from boreal

conifers in North America, where several refugia were reconstructed for each of the species studied: balsam fir (Cinget et al. 2015), black spruce (Gérardi et al. 2010; Jaramillo-Correa et al. 2004), tamarack (Warren et al. 2016), and jack pine (Godbout et al. 2010). This difference is probably related to the dry, extremely continental climate of Siberia, which could have been drier in the late glacial and probably restricted the fir to the most favorable areas of south Siberia and the Urals, unlike the larch, which survived close to the present Kara sea, and prevented the fir from crossing the southern, most arid belt of the West Siberian Plain after warming. As a result, only the populations of the refugium located in the mountains around the Baikal Lake could do this by spreading along the Angara and Yenisei Rivers.

The phylogeographic pattern of the presence of a common haplotype in the north of Western Siberia and in the Baikal region found in Siberian fir, is not quite common for the species of the taiga biota of Northern Eurasia. For instance, the populations of Siberian larch (Semerikov et al. 2013), wood lemming (Fedorov et al. 2008), flying squirrel (Oshida et al. 2005), and Siberian pine (Dr. D. Shuvaev, pers. comm.) in the north of Western Siberia had haplotypes common with haplotypes in the regions located west of Lake Baikal. In addition to the peculiarities of the ecological properties of Siberian fir, its unusual phylogeographic pattern may be a consequence of the random nature of the long-distance seed dispersion, which undoubtedly plays an important role in the process of colonization.

The mitotypes M1 and M3, common in the Altai, the Sayans, and Kuznetsk Alatau Mountains, were absent in northern Siberia, indicating the lack of a significant contribution of populations of these regions to the modern northern populations, which coincides with a similar conclusion about the role of southern mountain populations of Siberian larch (Semerikov et al. 2013).

4.3. MtDNA data indicate repeated migrations of fir to the Urals

Floristic surveys suggest the Siberian origin of most of the taiga forests species of the Urals (Krasheninnikov 1937; Hulten 1937; Gorchakovsky 1969). Accordingly, the mitotypes of Siberian fir

and Siberian larch in the Urals arose in Siberia or have originated from related mitotypes common in Siberia. Unlike Siberian larch, which has only one mitotype across the Urals, Siberian fir in the Urals has two mitotypes, M1 and M2, which are also common in Siberia. It is noteworthy that in the northern part of the Urals (starting from population #5 to the north) mitotype M2 is almost fixed (Fig. 1). M2 is also fixed in northern Siberia, which probably indicates that the colonization of the north of the Urals took place together with the settlement of the north of Siberia by a single migration wave. This assumption is supported by the results of studies of variation of allozyme and AFLP loci, which make however the story more complex: populations in the northern Urals genetically are similar both to the populations of the Baikal region, as well as to the populations of the Urals south (Semerikova and Semerikov 2006, 2011), suggesting the admixed origin of the populations in the northern Urals. It is interesting that the fir mitotypes in the south Ural and neighbor areas have a mosaic geographical distribution, and the majority of the investigated populations contain a single mitotype, which may be the result of dispersion from multiple local refugia, that suggests that M1 and M2 mitotypes were present here before LGM. Moreover, they could not come to the Urals together, during the post-LGM colonization of northern Siberia, since only one of them is present in the center of Western Siberia. Consequently, the fir migrations to the Urals most likely occurred several times or in different ways, for example, through more southerly regions, which are now outside the range of fir. For example, according to the findings of the *Abies* pollen in the peat bog near Lake Karasie, Kokshetau region of Kazakhstan (Gorchakovsky 1987, and refs. cited therein), the area of fir in the early Holocene could have expanded southward, capturing Northern Kazakhstan. In addition, the MAXENT reconstruction allows the presence of Siberian fir in Central Kazakhstan in the LGM (Fig. 2, A) and, accordingly, does not exclude the possibility of migrations from the Altai to the Urals across this area.

4.4. Markers with different inheritance mode reflect different aspects of the colonization

382 The populations in Western Siberia, despite their Baikalian origin suggested by the mtDNA, according
383 to the allozyme, AFLP and cpDNA data (Semerikova and Semerikov 2006, 2007, 2011) contain an
384 admixture of genes from the populations of Altai, the Urals and the Sayans. Probably, this discrepancy
385 of mitochondrial vs. nuclear and chloroplast markers can be explained by the difference in their
386 inheritance mode and, as a result, the second and third are transferred by both seeds and more mobile
387 pollen in contrast to the first transferred only by seeds. This feature determines faster homogenization of
388 the spatial structure for nuclear and chloroplast markers due to the pollen mediated gene flow that more
389 efficiently connects remote regions.

390 ## 5. Conclusions

391 The study of mtDNA variation, analysis of paleontological data and environmental niche modelling
392 shed light on the history of two Siberian conifers in the Late Pleistocene and Holocene and suggested the
393 most likely scenario of the dynamics of their ranges. The geographical heterogeneity of the mtDNA
394 variation of Siberian fir and Siberian larch in the southern part of their range is in agreement with their
395 distribution from several glacial refugia, while the homogeneity of populations in the northern part of the
396 range indicates that its colonization involved only one of the southern refugia, which, however, was not
397 the same in these species. In addition, the beginning of the colonization of the northern part of the range
398 of Siberian larch as indicated by the paleontological data and the results of modeling of ecological niches,
399 predated the last glacial maximum unlike in Siberian fir. According to the mtDNA data the migrations of
400 Siberian fir to the Urals probably occurred more than once.

401 ## Acknowledgments

402 This study was supported by the State Contract of the Institute of Plant and Animal Ecology, Ural
403 Branch, Russian Academy of Sciences and partly by the project “Genomics of the Key Boreal Forest

Conifer Species and Their Major Phytopathogens in the Russian Federation" funded by the Government of the Russian Federation (grant no. 14.Y26.31.0004). The laboratory experiments were supported by the Russian Foundation for Basic Research (grant no. 16-04-00607). We thank Vladimir Mikryukov for help with environmental niche modeling. Authors are also thankful to the Associate Editor and two anonymous reviewers for their suggestions that helped us improve the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

References

Bandelt, H. J., Forster, P., and Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. **16**: 37–48.

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., Valiranta, 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. doi:10.1016/j.quascirev.2009.04.016.

Binney, H., Edwards, M., 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., Ryabogina, N., Solovieva, N., Willis, K., and Zernitskaya, V. 2017. Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. Quat. Sci. Rev. **157**: 80-97. doi:10.1016/j.quascirev.2016.11.022.

425 Burban, C., Petit, R.J., Carcreff, E., and Jactel, H. 1999. Rangewide variation of the maritime pine bast
426 scale *Matsucoccus feytaudi* Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure
427 of its host. Mol. Ecol. **8**(10): 1593-1602. doi:10.1046/j.1365-294x.1999.00739.x.
428 Cinget, B., Gérardi, S., Beaulieu, J., and Bousquet, J. 2015. Less pollen-mediated gene flow for more
429 signatures of glacial lineages: congruent evidence from balsam fir cpDNA and mtDNA for multiple
430 refugia in Eastern and Central North America. PLoS ONE, **10**(4): e0122815. doi:10.1371/journal.
431 pone.0122815.
432 Demesure, B., Sodzi, N., and Petit, R.J. 1995. A set of universal primers for amplification of polymorphic
433 noncoding regions of mitochondrial and chloroplast DNA in plants. Mol. Ecol. **4**(1): 129-131.
434 doi:10.1111/j.1365-294X.1995.tb00201.x
435 Donnelly, K., Cottrell, J., Ennos, R.A., Vendramin, G.G., A'Hara, S., King, S., Perry, A., Wachowiak, W.,
436 and Cavers, S. 2017. Reconstructing the plant mitochondrial genome for marker discovery: a case
437 study using Pinus. Mol. Ecol. Res. **17**(5): 943-954. doi:10.1111/1755-0998.12646.
438 Dorofeyuk, N.I., and Tarasov, P.E. 2000. Vegetation of western and southern Mongolia in Late
439 Pleistocene and Holocene. Botanicheskii Zhurnal **85**(2): 1–17 (in Russian with Abstract in
440 English).
441 Dumolin-Lapegue, S., Pemonge, M.H., and Petit, R.J. 1997. An enlarged set of consensus primers for
442 the study of organelle DNA in plants. Mol. Ecol. **6**(4): 393-397. doi:10.1046/j.1365-
443 294X.1997.00193.x
444 Dupanloup, I., Schneider, S., and Excoffier, L. 2002. A simulated annealing approach to define the
445 genetic structure of populations. Mol. Ecol. **11**: 2571–2581. doi:10.1046/j.1365-
446 294X.2002.01650.x.

447 Excoffier, L., Laval. G., and Schneider, S. 2006. ARLEQUIN ver. 3.1: An integrated software package
448 for population genetics data analysis. Computational and Molecular Population Genetics Lab
449 (CMPG). Institute of Zoology, University of Bern, Bern, Switzerland.
450 Farjon, A., and Rushforth, K.D. 1989. A classification of *Abies* Miller. (Pinaceae). Notes Roy Bot. Gard.
451 Edinburg, **46**: 59–79.
452 Fedorov, V.B., Goropashnaya, A.V., Boeskorov, G.G., and Cook, J.A. 2008. Comparative
453 phylogeography and demographic history of the wood lemming (*Myopus schisticolor*):
454 implications for late Quaternary history of the taiga species in Eurasia. Mol. Ecol. **17**: 598–610.
455 doi:10.1111/j.1365-294X.2007.03595.x.
456 Fujita, K., and Silver, J. 1994. Single-strand conformational polymorphism. Genome Res. **4**: S137-S140.
457 Gérardi, S., Jaramillo-Correa, J.P., Beaulieu, J., and Bousquet, J. 2010. From glacial refugia to modern
458 populations: new assemblages of organelle genomes generated by differential cytoplasmic gene
459 flow in transcontinental black spruce. Mol. Ecol. **19**: 5265–5280. doi:10.1111/j.1365-
460 294X.2010.04881.x PMID: 21044193.
461 Godbout, J., Beaulieu, J., and Bousquet, J. 2010. Phylogeographic structure of jack pine (*Pinus*
462 *banksiana*; Pinaceae) supports the existence of a coastal glacial refugium in Northeastern North
463 America. Am. J. Bot. **97**(11): 1903-1912. doi:10.3732/ajb.1000148.
464 Gorchakovsky, P.L. 1969. Major Problems in the Historical Phyto-geography of the Urals. Ural Div. of
465 Academy of Sciences of USSR, Sverdlovsk, USSR. 286 p. (in Russian).
466 Gorchakovsky, P.L. 1987. Forest oases of the Kazakh Upland. Nauka, Moscow, USSR. 159 p. (in
467 Russian).
468 Goropashnaya, A.V., Fedorov, V.B., Seifert, B., Pamilo, P. 2004. Limited phylogeographic structure
469 across Eurasia in two red wood ant species Formica pratensis and F. lugubris (Hymenoptera,
470 Formicidae). Mol. Ecol. **13**: 1849–1858.

Oshida, T., Abramov, A., Yanagawa, H., Masuda, R. 2005. Phylogeography of the Russian flying squirrel (*Pteromys volans*): implication of refugia theory in arboreal small mammal of Eurasia. Mol. Ecol. **14**: 1191–1196.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. **41**: 95-98.

Hantemirova, E.V., Heinze, B., Knyazeva, S.G., Musaev, A.M., Lascoux, M., and Semerikov, V.L. 2017. A new Eurasian phylogeographical paradigm? Limited contribution of southern populations to the recolonization of high latitude populations in *Juniperus communis* L. (Cupressaceae). J. Biogeogr. **44**: 271-282 doi:10.1111/jbi.12867.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. **25**: 965–978. doi:10.1002/joc.1276.

Hulten, E. 1937. Outline of the History of Arctic and Boreal biota during the Quaternary Period. Acturlbolaget Thule, Stockholm, Sweden. 168 p.

Huntley, B., and Birks, H.J.B. 1983. An Atlas of Past and Present Pollen Maps for Europe: 0-13000 Years Ago. Cambridge University Press, Cambridge, UK

Jackman, S. D., Warren, R. L., Gibb, E. A., Vandervalk, B. P., Mohamadi, H., Chu, J., Raymond, A., Pleasance, S., Coope, R., Wildung, M. R., Ritland, C. E., Bousquet, J., Jones, S. J. M., Bohlmann, J., and Birol, I. 2016. Organellar Genomes of White Spruce (*Picea glauca*): Assembly and Annotation. Genome Biol. Evol. **8**(1): 29-41. doi:10.1093/gbe/evv244

Jaramillo-Correa, J.P., Beaulieu, J., and Bousquet J. 2004. Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. Mol. Ecol. **13**: 2735–2747. PMID: 15315685.

494 Kohli, B. A., Fedorov, V. B., Waltari, E., & Cook, J. A. 2015. Phylogeography of a Holarctic rodent
495 (*Myodes rutilus*): testing high-latitude biogeographical hypotheses and the dynamics of range
496 shifts. J. Biogeogr. **42**(2): 377-389. doi:10.1111/jbi.12433
497 Kosintsev, P.A., Lapteva, E.G., Korona, O.M., and Zanina, O.G. 2012. Living environments and diet of
498 the Mongochen mammoth, Gydan Peninsula, Russia. Quat. Int. **276**: 253-268.
499 doi:10.1016/j.quaint.2011.11.004.
500 Krasheninnikov, L.M. 1937. Analysis of the relict flora of the Southern Urals in connection with the
501 history of vegetation and the palaeogeography of Pleistocene. Soviet Botany **4**: 16-45. (in Russian)
502 Krylov, G.V., Maradudin, I.I., Mikheev, N.I., and Kozakova, N.F. 1986. Pikhta (Fir). Agropromizdat,
503 Moscow, (in Russian)
504 Liao, X., Zhao, Y., Kong, X., Khan, A., Zhou, B., Liu, D., Kashif, M.H., Chen, P., Wang, H., and Zhou,
505 R. 2018. Complete sequence of kenaf (*Hibiscus cannabinus*) mitochondrial genome and
506 comparative analysis with the mitochondrial genomes of other plants. Sci. Rep. **8**: 12714.
507 doi:10.1038/s41598-018-30297-w.
508 Neale, D.B., and Sederoff, R.R. 1989. Paternal inheritance of chloroplast DNA and maternal inheritance
509 of mitochondrial DNA in loblolly pine. Theor. Appl. Genet. **77**: 212–216.
510 Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press. New York. NY, USA
511 Ohri, D., and Khoshoo, T.N. 1986. Genome size in Gymnosperms. Plant Syst. Evol. **153**(1-2): 119-132.
512 doi:10.1007/bf00989421.
513 Panova, N.K., Trofimova, S.S., Antipina, T.G., Zinoviev, E.V., Gilev, A.V., and Erokhin, N.G. 2010.
514 Holocene dynamics of vegetation and ecological conditions in the southern Yamal Peninsula
515 according to the results of comprehensive analysis of a relict peat bog deposit. Russ. J. Ecol. **41**(1):
516 20-27. doi:10.1134/s1067413610010042.

517 Phillips, S.J., Anderson, R.P., and Schapire, R.E. 2006. Maximum entropy modeling of species
518 geographic distributions. Ecol. Modelling, **190:** 231–259. doi:10.1016/j.ecolmodel.2005.03.026
519 Polezhaeva, M.A., Lascoux, M., and Semerikov, V.L. 2010. Cytoplasmic DNA variation and
520 biogeography of *Larix* Mill. in Northeast Asia. Mol. Ecol. **19**(6): 1239-1252. doi:10.1111/j.1365-
521 294X.2010.04552.x.
522 Pons, O., and Petit, R.J. 1996. Measuring and testing genetic differentiation with ordered versus
523 unordered alleles. Genetics, **144**: 1237-1245.
524 Rozen, S., and Skaletsky, H.J. 2000. Primer3 on the WWW for general users and for biologist
525 programmers. *In* Bioinformatics Methods and Protocols: Methods in Molecular Biology. *Edited by*
526 S. Krawetz and S. Misener. Humana Press, Totowa, NJ, p. 365–386.
527 Semerikov, V.L., Semerikova, S.A., Polezhaeva, M.A., Kosintsev, P.A. and Lascoux, M. 2013. Southern
528 montane populations did not contribute to the recolonization of West Siberian Plain by Siberian
529 larch (*Larix sibirica*): a range-wide analysis of cytoplasmic markers. Mol. Ecol. **22**(19): 4958-
530 4971. doi:10.1111/mec.12433.
531 Semerikov, V. L., Putintseva, Y. A., Oreshkova, N. V., Semerikova, S. A., and Krutovsky, K. V. 2015.
532 Development of new mitochondrial DNA markers in Scots pine (*Pinus sylvestris* L.) for
533 population and phylogeographic studies. Russ. J. Genet. **51**(12): 1199-1203.
534 doi:10.1134/s1022795415120108.
535 Semerikov, V. L., Semerikova, S. A., Putintseva, Y. A., Tarakanov, V. V., Tikhonova, I. V., Vidyakin, A.
536 I., Oreshkova, N. V., and Krutovsky, K. V. 2018. Colonization history of Scots pine in Eastern
537 Europe and North Asia based on mitochondrial DNA variation. Tree Genet. Genomes, **14**(1): 8.
538 doi:10.1007/s11295-017-1222-0.

539 Semerikova, S.A., and Semerikov, V.L. 2006. Genetic variation and population differentiation in Siberian
540 fir *Abies sibirica* Ledeb. inferred from allozyme markers. Russ. J. Genet. **42**(6): 636-644.
541 doi:10.1134/s102279540606008.
542 Semerikova, S.A., and Semerikov, V.L. 2007. The diversity of chloroplast microsatellite loci in Siberian
543 fir (*Abies sibirica* Ledeb.) and two Far East fir species *A. nephrolepis* (Trautv.) Maxim. and *A.*
544 *sachalinensis* Fr. Schmidt. Russ. J. Genet. **43**(12): 1373-1381. doi:10.1134/s102279540712006x.
545 Semerikova, S.A., and Semerikov, V.L. 2011. Genetic variability of Siberian fir *Abies sibirica* Ledeb.
546 inferred from AFLP markers. Russ. J. Genet. **47**(2): 241-246. doi:10.1134/s1022795411020153.
547 Semerikova, S. A., Semerikov, V. L., & Lascoux, M. 2011. Post-glacial history and introgression in *Abies*
548 (Pinaceae) species of the Russian Far East inferred from both nuclear and cytoplasmic markers. J.
549 Biogeog. **38**: 326-340. doi:10.1111/j.1365-2699.2010.02394.x
550 Semerikova, S.A., Lascoux, M., and Semerikov, V.L. 2012. Nuclear and cytoplasmic genetic diversity
551 reveals long-term population decline in *Abies semenovii*, an endemic fir of central Asia. Can. J.
552 Forest Res. **42**: 2142-2152. doi:10.1139/cjfr-2012-0158.
553 Semerikova, S.A., and Semerikov, V.L. 2016. Phylogeny of Firs (Genus *Abies*, Pinaceae) Based on
554 Multilocus Nuclear Markers (AFLP). Russ. J. Genet. **52**: 1164-1175. doi:10.1134/s102.
555 Semerikova, S.A., Khrunyk, Y.Y., Lascoux, M., and Semerikov, V.L. 2018. From America to Eurasia: a
556 multigenomes history of the genus *Abies*. Mol. Phylogenet. Evol. **125**: 14-28.
557 doi:https://doi.org/10.1016/j.ympev.2018.03.009.
558 Tarasov, P. E., Volkova, V. S., Webb, T., Guiot, J., Andreev, A. A., Bezusko, L. G., Bezusko, T. V., Bykova,
559 G. V., Dorofeyuk, N. I., Kvavadze, E. V., Osipova, I. M., Panova, N. K., and Sevastyanov, D. V.
560 2000. Last glacial maximum biomes reconstructed from pollen and plant macrofossil data from
561 northern Eurasia. J. Biogeography, **27**(3): 609-620. doi:10.1046/j.1365-2699.2000.00429.x.

562 Volkova, V.S., Arkhipov, S.A., Babushkin, A.E., Kulkova, I.A., Guskov, S.A., Kuzmina, O.B., Levchuk,
563 L.K., Mikhailova, I.V., and Sukhorukova, S.S. 2002. Stratigraphy of oil and gas basins of Siberia.
564 Cenozoic of West Siberia. Novosibirsk: Publ. House of SB RAS, Department "GEO". 246 p.
565 Warren, E., de Lafontaine, G., Gerardi, S., Senneville, S., Beaulieu, J., Perron, M., Pablo Jaramillo-
566 Correa, J., and Bousquet, J. 2016. Joint inferences from cytoplasmic DNA and fossil data provide
567 evidence for glacial vicariance and contrasted post-glacial dynamics in tamarack, a transcontinental
568 conifer. J. Biogeography, **43**(6): 1227-1241. doi:10.1111/jbi.12675.
569 Willis, K.J., and van Andel, T.H. 2004. Trees or no trees? The environments of central and eastern Europe
570 during the Last Glaciation. Quat. Sci. Rev. **23**: 2369-2387. doi:10.1016/j.quascirev.2004.06.002
571 Zink, R.M., Drovetski, S.V., and Rohwer, S. 2002. Phylogeographic patterns in the great spotted
572 woodpecker *Dendrocopos major* across Eurasia. J Avian Biol. **33**(2): 175-178. doi:10.1034/j.1600-
573 048X.2002.330208.x
574
575

576 **Table 1.** Mitochondrial contigs, mitotypes, GenBank accession numbers, SNPs detected, nucleotide
577 sequences of primers used for sequencing and SSCP genotyping, and length of the PCR product amplified
578 in *Abies sibirica* (mitotypes M1, M2 and M3), *A. nephrolepis* (M5) and *A. semenovii* (M4).

Contig number	Mitotype	GenBank accession number	SNP (nucleotide position)	Primer nucleotide sequence	Length of the PCR fragment in *A. sibirica*, bp
167	M2	MH070276	T (992)	A167F: AGCTGATCCGCTGAATGACT	1339
	M1, M3	MH070277	G (992)	A167R: ACTTCGTCCCTGAAGCAAGA	1339
	M5	MH070278	G (992)		1339
	M4	MH070279	G (992)		1339
				A167NL[a]: AACAATGGGATTTGGAATGC	221
				A167NR[a]: CTCGTCCAATTGATCAAGCA	
65	M2	MH070284	C (46)	A65F: CCGGAGTGGTTTTGTTGAGT	1113
	M1, M3	MH070285	A (46)	A65R: TATGCCTTCTCGGAAACACC	1113
	M5	MH070286	C (46)		1096
	M4	MH070287	C (46)		1113
				A65NR[a,b]: TGGCCCCTAATGGTGTTAGA	165
126	M2	MH070272	A (1254)	A126-2F: TGTGGGGATGGATCTCTAGC	1452
	M1,M3	MH070273	C(1254)	A126-2R: AGGGGTGGTGTGGTCAATAA	1452
	M5	MH070274	A(1250)		1448
	M4	MH070275	C(1254)		1452
				A126NL[a]: CTCCTCACCCTTCGACTCAC	182
				A126NR[a]: CCAGAACGGGTGAGTCACTT	
37	M2, M1	MH070280	C (1191)	A37-1F: GGCGACGAATAAATCAGGAA	1431
	M3	MH070281	A (1191)	A37-1R: TCTTGCTTGTTTTGGTGCTG	1431
	M5	MH070282	C (1191)		1431
	M4	MH070283	C (1191)		1431
				A37NL[a]: CTACAGCGGCACATAGATCG	250
				A37NR[a]: GTGGAGAGCTCTGCGCTAAT	

579 [a]Used for the SSCP genotyping

580 [b]Used in a pair with A65F

581

582 **Table 2.** Hierarchical analysis of the mtDNA genetic variation (AMOVA) based on the SAMOVA

583 grouping of the *Abies sibirica* populations into three groups.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	2	422.082	1.244	92.23
Among populations within groups	42	12.793	0.014	1.03
Within populations	648	58.903	0.091	6.74
Total	692	493.778	1.349	
	Fixation indices		*P*-value	
F_{SC}	0.133		<0.00001	
F_{ST}	0.933		<0.00001	
F_{CT}	0.922		<0.00001	

584

585

586 ## Figure captions

587

588 **Fig. 1. (a)** Geographic distribution of the studied fir populations. The mitotype frequencies in each

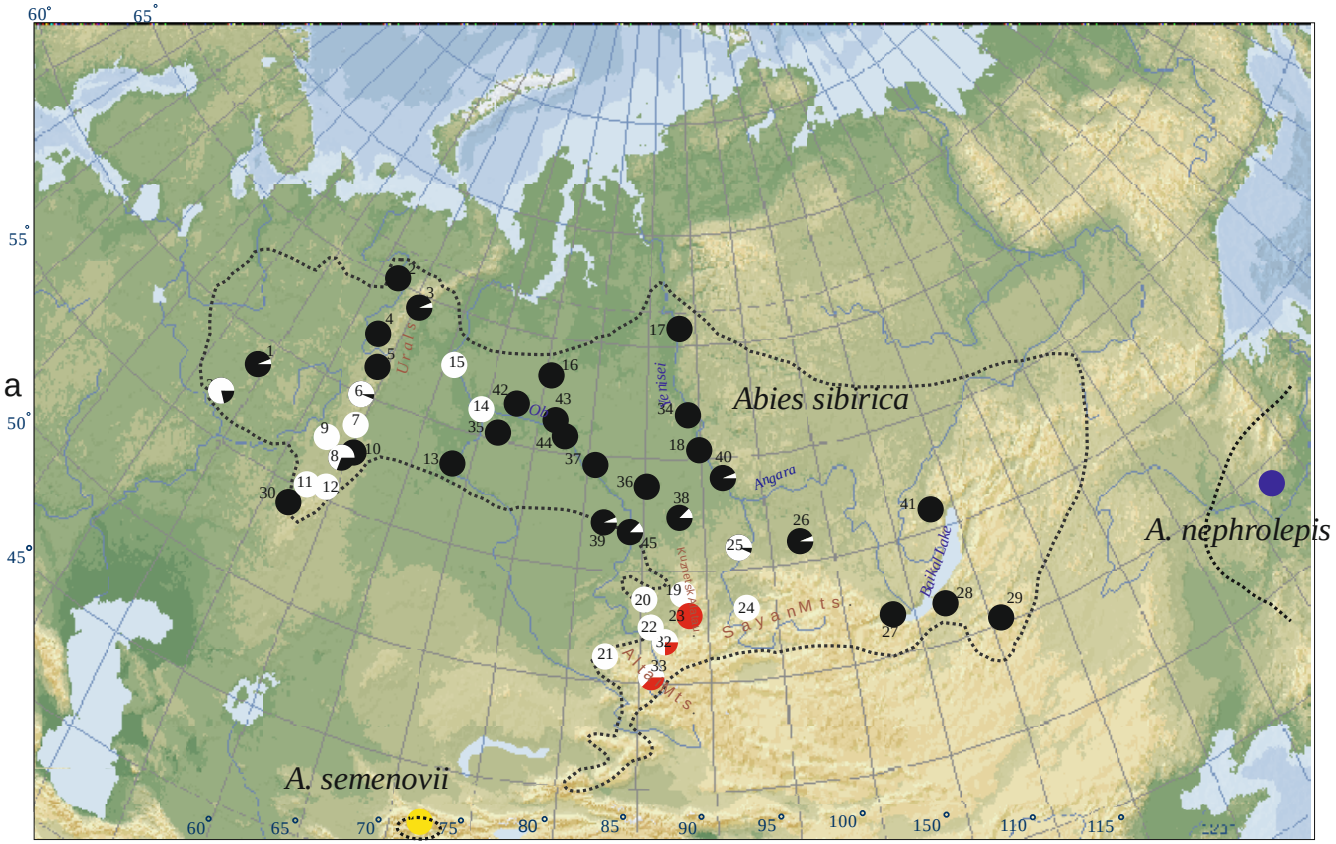
589 population are represented as a pie diagram. See population names in Table S1 (Supplementary

590 material). (**b)** A network of three mitotypes (M1-M3) found in *Abies sibirica* in Eastern Europe and

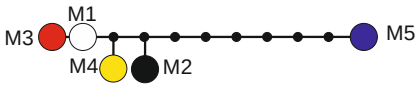
591 Northern Asia and two mitotypes found in *A. semenovii* (M4) and *A. nephrolepis* (M5).

Fig. 2. Mapped pollen and macrofossil paleorecords of *Abies* sp. and *Larix* sp. Modern ranges are outlined by the thick purple line. Existed pollen records for particular time interval are depicted as white circles, pollen records of *Abies* – as red circles, *Abies* macrofossils – as red diamonds, pollen record of *Larix* – as plume circles, *Larix* macrofossils – as plume diamond. Distributions of species predicted by MAXENT for LGM (A and I) and present day (H and P) are highlighted by the tone of red color. More saturated red color reflects conditions more appropriate for the species, less saturated - less suitable conditions. Time span (kya) for *Abies*: A – 18.5–21.5, B – 15.5-18.5, C – 12.5–15.5, D – 9.5–12.5, E – 6.5–9.5, F – 3.5–6.5, G – 0.5–3.5, H - 0–0.5, and *Larix*: I – 18.5–21.5, J – 15.5-18.5, K – 12.5–15.5, L – 9.5–12.5, M – 6.5–9.5, N – 3.5–6.5, O – 0.5–3.5, P - 0–0.5.

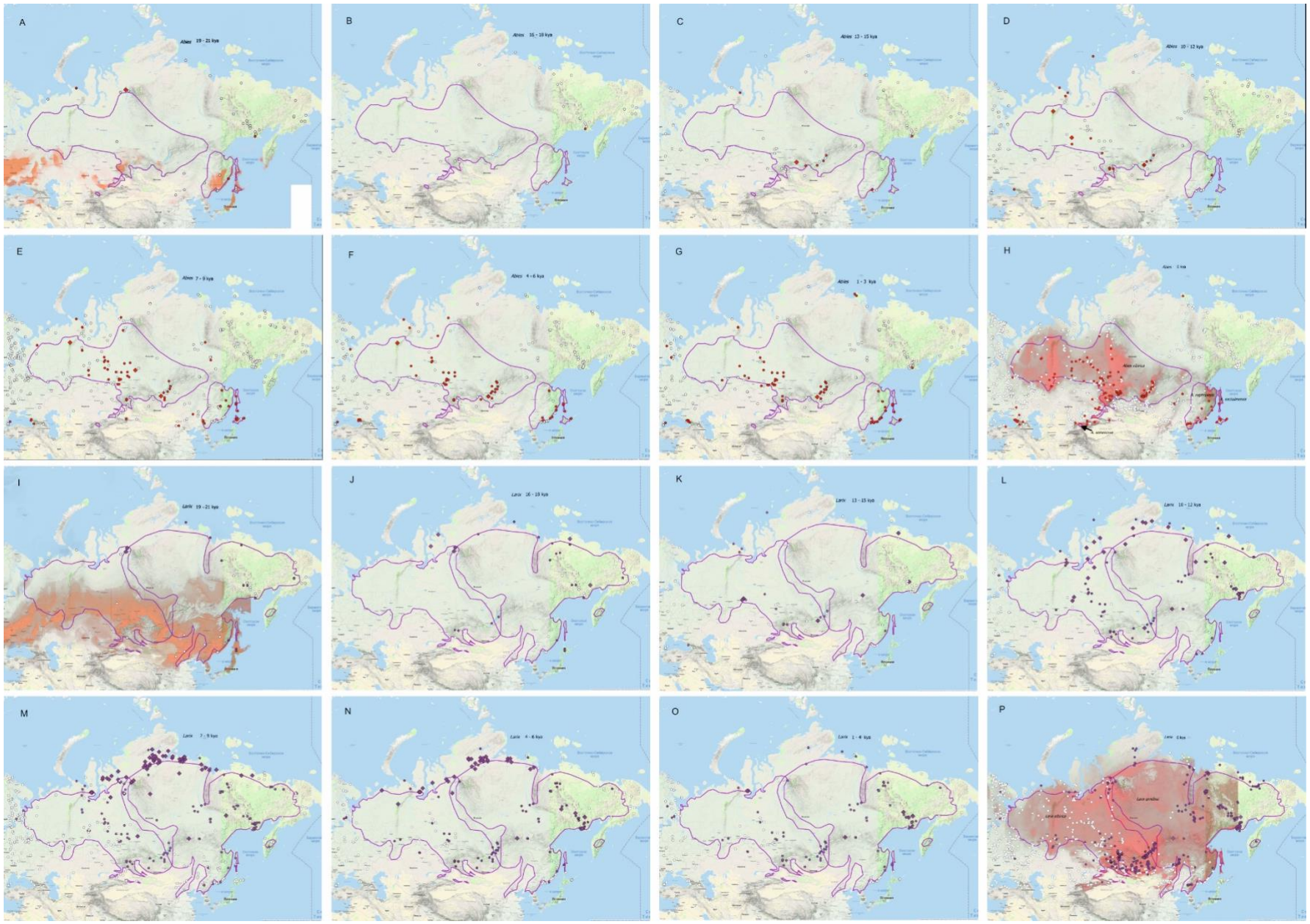
a


Abies sibirica
A. nephrolepis
A. semenovii
Angara
Baikal Lake
Yenisei
Kuznetsk Alatau
Sayan Mts.
Altai Mts.
Ural Mts.
60°
65°
70°
75°
80°
85°
90°
95°
100°
150°
110°
115°
55°
50°
45°

b


M1
M3
M5
M4
M2

601



602

603 Supplementary material

604 **A range-wide study of the mitochondrial DNA diversity of the Siberian fir**
605 **indicates multiple post-glacial colonization centers**

606 Vladimir L. Semerikov, Svetlana A. Semerikova, Yuliya A. Putintseva, Natalia V. Oreshkova,
607 Konstantin V. Krutovsky

608

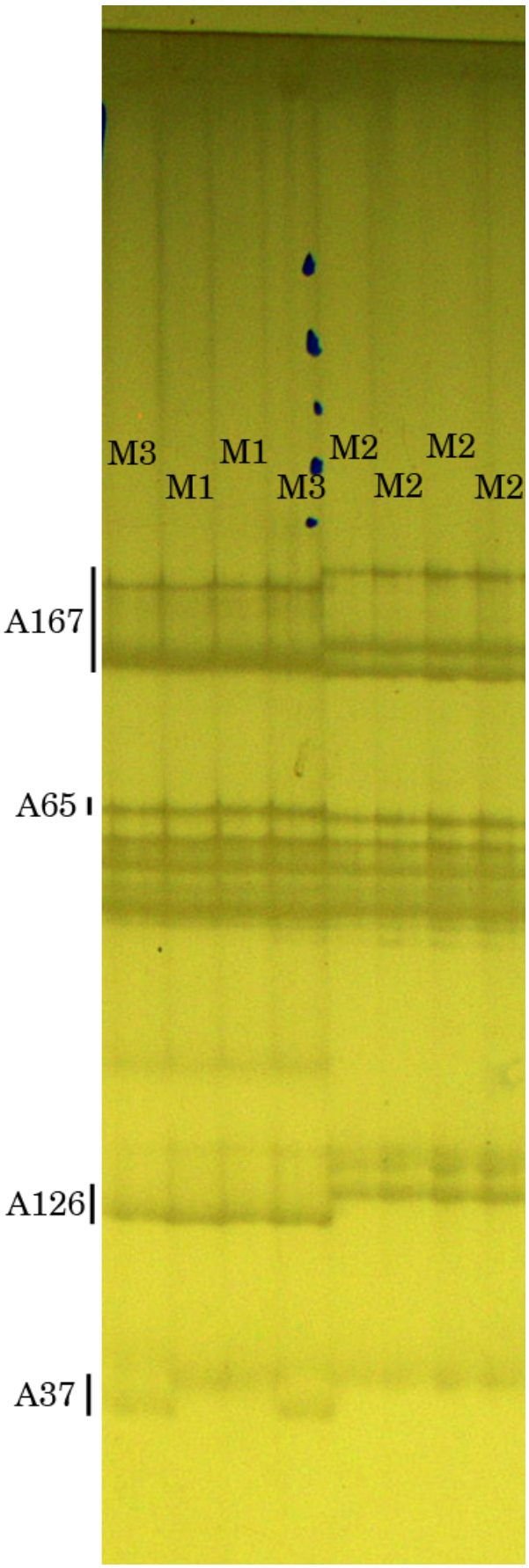
609 **Table S1.** Studied fir populations and their mitotypes.

No.	Population	Location (Northern latitude / Eastern longitude)	Sample size	Mitotype frequencies	Diversity, H_e
		A. sibirica			
1	Viyatka	58°40’/ 49°30’	16	M1:1,M2:15	0.094
2	Pechera	65º00’/ 57º30’	12	M2:12	0
3	Manya	64º30’/ 60º50’	16	M1:1,M2:15	0.094
4	Pechero-Ilychsky	61º50’/ 57º00’	24	M2:24	0
5	Denegkin	60º09’/ 59º57’	16	M2:16	0
6	Kongakovskii	59º40’/ 59º10’	16	M1:15,M2:1	0.094
7	Kushva	58º18’/ 59º41’	16	M1:16	0
8	Homutovka	56º48’/ 59º57’	13	M1:9,M2:4	0.346
9	Chusovoy	57º17’/ 57º49’	16	M1:16	0
10	Tavatuy	56º50’/ 60º20’	16	M2:16	0
11	Sim	54º59’/ 57º41’	12	M1:12	0
12	Taganay	55º10’/ 59º40’	16	M1:16	0
13	Tobolsk	58º12’/ 68º16’	16	M2:16	0
14	Khanty-Mansiysk	61º00’/ 69º10’	16	M1:16	0
15	Oktyabrskoye	62º35’/ 66º05’	16	M1:16	0
16	Noyabrsk	63º12’/ 75º29’	16	M2:16	0
17	Turuhansk	65º48’/87º59’	16	M2:16	0
18	Yartsevo	60º14’/90º15’	16	M2:16	0
19	Kemerovo	55º20’/ 86º05’	24	M1:24	0
20	Salair	53º35’/ 85º70’	24	M1:24	0
21	Kolyvan	51º10’/ 82º50’	18	M1:18	0
22	Karasuk	51º58’/ 85º57’	16	M1:16	0
23	Tashtagol	52º40’/ 88º00’	16	M3:16	0
24	Tanzybey	52º50’/ 93º00’	16	M1:16	0
25	Divnogorsk	55º55’/ 92º30’	15	M1:14,M2:1	0.100
26	Taishet	55º57’/ 98º00’	16	M1:1,M2:15	0.094

No.	Population	Location (Northern latitude / Eastern longitude)	Sample size	Mitotype frequencies	Diversity, H_e
27	Sludyanka	51º38’/103º42’	24	M2:24	0
28	Ulan-Ude	51º50’/106º42’	16	M2:16	0
29	Sohondinskii	49º30’/111º00’	16	M2:16	0
30	Inzer	54º18’/57º23’	16	M2:16	0
31	Yoshkar-Ola	56°42’/47°55’	11	M1:8,M2:3	0.327
32	Artybash[1]	51°48’/87°15’	16	M1:9,M3:3	0.409
33	Multa	50° 00’ /85° 49’	8	M1:5,M3:3	0.534
34	Bor	61°30’/90°10’	16	M2:16	0
35	Salym	60°03’/71°27’	16	M2:16	0
36	BelYar	58°26’/85°06’	16	M2:16	0
37	Kargasok	59°00’/80°51’	16	M2:16	0
38	Teguldet	57°18’/88°14’	16	M1:2,M2:14	0.175
39	Bakchar	57°02’/82°03’	16	M1:1,M2:15	0.094
40	Eniseysk	58°25’/92°09’	16	M1:1,M2:15	0.094
41	Severobaikalsk	55°42’/109°03’	16	M2:16	0
42	Pochekuika	61°22’/73°46’	16	M2:16	0
43	Visokii	61°06’/76°00’	16	M2:16	0
44	Aleksandrov	60°25’/77°50’	8	M2:8	0
45	Tomsk	56°34’/84°00’	8	M1:1,M2:7	0.188
	Total		697		
		A. nephrolepis			
	Obluchye	49°01’/131°05’	1	M5:1	-
		A. semenovii			
	Sary-Chelek	41°54’/71°56’	1	M4:1	-

610
611

Fig.S1. Genotyping of Siberian fir mitochondrial DNA markers using the SSCP method (Fujita et al., 1994) in a non-denaturing polyacrylamide gel. Eight lines (from left to right) correspond to eight individuals with mitotypes M3, M1, M1, M3, M2, M2, M2, M2. Four variable electrophoretic zones of single-stranded fragments correspond to the markers A167, A65, A126 and A37 (from top to bottom), each of which has two alleles differing in one nucleotide and different mobility, due to conformational polymorphism. Marker A167 in these eight trees has a nucleotide: G, G, G, G, T, T, T, T. Marker A65 - A, A, A, A, C, C, C, C, respectively. Marker A126 - C, C, C, C. A, A, A, A. A37 - A, C, C, A, C, C, C, C.



621