ARTICLES

The Location of the Northern Glacial Refugium of Scots Pine Based on Mitochondrial DNA Markers

JURATA BUCHOVSKA[1*], DARIUS DANUSEVIČIUS[2], DANAS BANIULIS[3], VIDMANTAS STANYS[3], JURATĖ BRONĖ ŠIKŠNIANIENĖ[3] AND DARIUS KAVALIAUSKAS[2]

[1]- Institute of Forestry, Lithuanian Research Centre for Agriculture and Forestry, Liepu str. 1, Girionys LT-53101, Kaunas distr., Lithuania. Tel. +370 37 547221, fax: +370 37 54 74 46.

[2]- Aleksandras Stulginskis University, Institute of Forest Biology and Silviculture, Studentu str. 11, LT-53361 Akademija, Kaunas reg., Lithuania. Tel. +370 37 75 23 00, Fax: + 370 37 39 75 00.

[3]- Institute of Horticulture, Lithuanian Research Centre for Agriculture and Forestry, Kauno str. 30, Babtai, LT-5433, Kaunas reg. Lithuania. Tel. +370-37-555395, Fax +370 37 55 51 76.

[*]Corresponding author:Jurata Buchovska, Institute of Forestry Lithuanian Research Centre for Agriculture and Forestry. Postal address: Liepų str. 1 Girionys LT-53101, Kaunas distr., Lithuania. Tel:+370 37 547289, Fax: +370 37 547446, E-mail: bukauskaite.jurate@gmail.com

Buchovska, J., Danusevičius, D., Baniulis, D., Stanys, V., Šikšnianienė, J. B. and Kavaliauskas, D. 2013. The Location of the Northern Glacial Refugium of Scots Pine Based on Mitochondrial DNA Markers. *Baltic Forestry* 19(1): 2–12.

Abstract

Several recent studies based on mitochondrial DNA markers suggested a northern refugium for Scots pine somewhere westwards of the southern part of the Ural Mountains. The objective of our study was to assess the mtDNA polymorphism of Scots pine at the Nad7-1 and Nad1-B/C loci with the aim of detecting the location of this northern glacial refugium and the associated post-glacial migration routes. We studied 54 populations densely covering the European part of Russia westwards of the Ural Mountains, but also populations from the Czech Republic, Poland, Sweden, Finland, Scotland, Georgia and eastern Siberia were included. For the Nad1-B/C locus, all our material was monomorphic. Of the total of 474 individuals tested at the Nad7-1 locus, 348 individuals (73 %) possessed the universal haplotype A of 300 bp and 126 individuals (27 %) – the northern haplotype B of 295 bp. Geographical distribution of the Nad7-1 northern B haplotype was not random (SAMOVA, BAPS) forming a consistent cline directed towards north-west of the south-eastern part of European Russia up to the Scandinavia in the north. This provides a stronger support for the south-eastern rather than the central European location of the northern glacial refugium. A possible location of the northern refugium could be at about 300 km south-east of Moscow, where the northern B haplotype occurs in high frequency and Scots pine could possibly survive during the LGM. There also is a possibility for a more southern location of the northern refugium, assuming that such signature was lost during the northward migration or via genetic drift.

Key words: differentiation, glaciation, LGM, *Pinus sylvestris*, phylogeography, post-glacial colonisation, mtDNA, organellar DNA

Introduction

Based on mitochondrial DNA (mtDNA) markers, Naydenov et al. (2007) revealed two distinct glacial refugia for Scots pine in the temperate and boreal zones of Eurasia: (a) multiple refugia south of the ice cover longitudinally stretching over all of Eurasia from Scotland to south western Siberia (Chernova et al. 1991, Cheddadi et al. 2006) sharing the universal haplotype for the mtDNA marker; (b) a northern refugium, much smaller in geographical area, mainly found in the Baltic, Karelian (north-western part of Russia) and Finnish populations sharing the northern haplotype for the mtDNA marker. The above mentioned northern European areas were covered by the arctic ice sheet during the last glacial maximum (LGM, (37 000 - 16 000 years ago, Webb and Bertlain 1992, Taberlet et al. 1998, Abbott and Brochmann 2003) and so cannot represent the location of this northern refugium. Similar results at the same mtDNA locus were independently obtained

by Pyhäjärvi at al. (2008) but with fewer populations. Unfortunately, the studies of Naydenov et al (2007) and Pyhäjärvi et al. (2008) included a very few of the central Russian populations. Therefore, the location of this northern Scots pine refugium and the associated major post-glacial migration routes remain still largely unknown and constitute the main problem addressed in our study. It is also important to confirm the inferences made by Naydenov et al. (2007) and Pyhäjärvi et al. (2008) and the other similar studies with another material representing areas not covered by these studies.

Understanding the patterns of the temporal and spatial distribution of the genetic variation of such wide-spread coniferous species as Scots pine is of major fundamental interest in revealing the evolution and present genetic structure of the species. This problem was first tackled with fossil pine pollen and tree fossil data (e.g. Willis et al. 1998, Froyd 2005). It was established that Scots pine reached northern Scandinavia at about 7800 BP (Huntley and Birks 1983, Willis et al. 1998) but the exact colonization routs are largely unknown. Later studies employed DNA markers (e.g Cheddadi et al. 2006). Because of the non-recombinant nature and the uniparental inheritance mode, the organellar DNA markers provide powerful tools to investigate the ancestral lineages of coniferous species (Petit et al. 2003). In coniferous species, the mitochondrial DNA (mtDNA) genome is maternally inherited (Neale and Sederoff 1989, Jaramillo-Correa et al. 2003) and so being free of the paternal effects during the generations of mating, and is transmitted through seeds so preserving the distinct genetic lineages of the evolutionary past (Sperisen et al. 2001). The mitochondrial DNA genome of conifers exhibits a relatively high level of structural evolution combined with a low rate of the DNA sequence mutations, which are very suitable features for studying the intra-population structure (Gros-Louis et al. 2005). An undesirable feature for the pine population studies, however, is that the mtDNA genome of pines possesses a low level of polymorphism (Wolfe et al. 1987) and only a few informative mtDNA markers are available (Naydenov et al. 2007).

Sinclair et al. (1998; 1999) studied the mtDNA polymorphism of mainly Scottish Scots pine populations using the *cox I* gene as a RFLP hybridization probe and found at least three different postglacial sources for the Scots pine populations of Western Europe. However, this type of marker is too laborious to use for large number of individuals. Soranzo et al. (1999; 2000) developed a PCR based mtDNA marker; however, even though it could discriminate between the pine species based on the different number of SSRs, it was monomorphic in Scots pine. To date, only two polymorphic PCR-based mtDNA markers are known for Scots pine: (a) insertions/deletions in the intron 1 of the mtDNA gene Nad7 (Nad7-1 locus) allowing to discriminate between the mitotypes representing the universal (found over whole Eurasia), the northern and the Turkish refugial zones (Naydenov et al. 2007), (b) intron B/C of the mtDNA gene Nad1 (Nad1-B/C locus), separating the populations from the Iberian peninsula from the rest of the species range (Soranzo et al. 2000).

Initially, the studies based on the fossil pollen and macrofossil remains of trees revealed that the glacial refugia for the tree species were located in discontinuous areas in the southern European peninsulas such as Iberia, Italy and Balkans (Huntley and Birks 1983, Birks 1989). However, later studies, in addition to employing DNA markers, also raised a possibility for pine refugia at more northern latitudes than believed earlier and suggested refugia in the Hungarian Plain, in the eastern Alps, the Danube basin the Czech Republic and norwards of the Black sea (Kremenetski et al. 1998, Sinclair et al. 1999, Stewart and Lister 2001, Willis and van Andel 2004, Cheddadi et al. 2006). The recent studies based on new mtDNA markers and covering more of the eastern distribution of Scots pine provided evidence for the existence of even more northern refugia of Scots pine somewhere westwards of the Ural Mountains up to the latitudes of 50°N (Magri et al. 2006, Naydenov et al. 2007, Pyhäjärvi et al. 2008). Our study is based on the findings of Naydenov et al. (2007) and Pyhäjärvi et al. (2008) of the northern haplotype B with a 5bp deletion at the mtDNA Nad7-1 locus common among in the north-eastern European populations of Scots pine. The later authors logically suggest than a recent mutation at the Nad7-1 locus giving rise to the northern B haplotype is unlikely, because of such a wide range of this haplotype. Most of the populations contained both the universal and the northern haplotypes that suggests co-migration from several refugial zones. However, due to the low representation of the area westwards of the Ural Mountains in the above mentioned studies, the original post-glacial migration sources are still unclear.

The objective of our study was to assess the mtDNA polymorphism of Scots pine at the Nad7-1 and Nad1-B/C loci mainly aiming to obtain more information on the location of the northern glacial refugium of Scots pine and the associated post-glacial migration routes. In contrast to the earlier studies, our material densely covers the suspected areas of the possible location of the northern refugium of Scots pine westwards of the Ural Mountains up to the Baltic States.

Material and methods

Population sampling

The needles for the DNA study were mainly sampled in the provenance field trial of the Prokazin series (Shutyaev and Giertych 1998) located in Kazlų Rūda, the central part of Lithuania. The provenance seed lot used for establishment of the field test was a commercial seed mixture from a number of natural stands within a forest district occupying an area of ca 50 ha. The seed collection in the former Soviet Union was organised so that the seed crops from the stands within a forest district were mixed and considered as one seed lot. Such a seed mixture represents the variation from several stands within a restricted area. For the aims of our study, this representation is more suitable than samples from one stand only.

The Prokazin series of Scots pine provenance tests were established over a number of locations in the former Soviet Union and served a similar purpose as the contemporary IUFRO series in Western Europe. Only natural stands were sampled for the Prokazin test series. In our study, the initial aim was to sample 15 trees per population. However, owing to the DNA quality and financial reasons, the population sample was approximately 10 trees per population (Table 1). In total, samples were obtained from 474 individuals from 54 populations. The populations mainly represent the European part of Russia, but also populations from Czech Republic, Poland, Sweden, Finland, Scotland, Georgia and Eastern Siberia were included (Table 1, Figure 1).

Country[1]	Region[2]	Population	Id	LAT	LONG	ALT	H295B	H300(A)	N	FreqB	H_e	SAMOVA group	BAPS group
BY	BALT	Gardin	3	53.25	25.15	105	2	8	10	0.20	0.36	2	1
ET	BALT	Estonia	1	58.10	26.28	103	6	5	11	0.55	0.55	1	2
LT	BALT	Kazlų Rūda	4	54.45	23.35	80	2	8	10	0.20	0.36	2	1
LT	BALT	Ignalina	5	55.16	25.47	168	1	4	5	0.20	0.40	2	1
LT	BALT	Mažeikiai	6	57.15	22.40	72	1	6	7	0.14	0.29	2	1
LT	BALT	Neringa	7	55.31	21.06	35	2	29	31	0.06	0.12	2	1
LV	BALT	Latvia	2	56.42	25.10	60	3	7	10	0.30	0.47	2	1
	BALT Average						**17**	**67**	**84**	**0.24**	**0.36**		
CZ	EE	V.Chvojno	10	50.03	16.99	408	2	7	9	0.22	0.39	2	1
PL	EE	Bialystok	11	52.02	21.03	120	4	6	10	0.40	0.53	2	1
UA	EE	Lviv	8	48.07	24.30	610	0	9	9	0.00	0.00	2	1
UA	EE	Cherkasy	9	49.27	39.03	294	0	5	5	0.00	0.00	2	1
	EE Average						**6**	**27**	**33**	**0.16**	**0.23**		
GE	KAUK	Boržomsk	12	42.00	44.00	1903	3	6	9	0.33	0.50	2	1
	KAUK Average						3	6	9	0.33	0.50	2	1
RU	RU-C	Smolensk	14	54.00	33.00	198	3	7	10	0.30	0.47	2	1
RU	RU-C	Kalinin	15	57.45	36.40	133	3	6	9	0.33	0.50	2	1
RU	RU-C	Moscow	16	55.40	37.10	203	5	3	8	0.63	0.54	1	2
RU	RU-C	Vladimir	17	56.21	41.15	105	4	6	10	0.40	0.53	2	1
RU	RU-C	Kostroma	18	58.00	41.00	135	0	8	8	0.00	0.00	2	1
RU	RU-C	Riazan	19	54.40	39.45	110	5	4	9	0.56	0.56	1	2
RU	RU-C	Briansk	20	53.20	34.15	158	6	4	10	0.60	0.53	1	2
	RU-C Average						**26**	**38**	**64**	**0.40**	**0.45**		
RU	RU-E	Uljanovsk	21	54.14	49.35	96	1	9	10	0.10	0.20	2	1
RU	RU-E	Tatarstan	22	55.50	48.09	91	0	6	6	0.00	0.00	2	1
RU	RU-E	Tver	23	56.50	35.56	152	1	6	7	0.14	0.29	2	1
RU	RU-E	Udmurtia	24	57.00	54.00	194	0	9	9	0.00	0.00	2	1
RU	RU-E	Bashkiria	25	55.30	54.40	106	2	14	16	0.13	0.23	2	1
RU	RU-E	Orienburg	26	52.47	52.15	149	0	8	8	0.00	0.00	2	1
	RU-E Average						**4**	**52**	**56**	**0.06**	**0.12**		
RU	RU-EE	Karpinsk	27	60.00	60.00	65	5	3	8	0.63	0.54	1	2
RU	RU-EE	Tiumensk	28	57.00	66.00	106	1	4	5	0.20	0.40	2	1
	RU-EE Average						**6**	**7**	**13**	**0.41**	**0.47**		
RU	RU-FE	Chabarovsk	29	48.33	135.28	56	0	10	10	0.00	0.00	2	1
RU	RU-FE	Tygdinsk	30	54.00	126.00	289	0	8	8	0.00	0.00	2	1
RU	RU-FE	Tuganski	31	57.00	86.00	197	0	6	6	0.00	0.00	2	1
RU	RU-FE	Ingodinsk	32	51.00	112.00	612	3	2	5	0.60	0.60	4	2
	RU-FE Average						**3**	**26**	**29**	**0.15**	**0.15**		

Table 1. Origin of the populations studied, the frequencies the mtDNA Nad7-1 haplotypes, intra-population diversity and the SAMOVA grouping. Population Id corresponds with the number in Figure 1. N is total tree number. H295B and H300A – number of individulas with the northern B and universal A haplotypes. FrqB – frequency of the B haplotype. H_e – within population diversity index, see M&M section

Table 1 continued

Country[1]	Region[2]	Population	Id	LAT	LONG	ALT	H29 5B	H30 0(A)	N	FreqB	H_e	SAMOVA group	BAPS group
RU	RU-NE	Kozvinsk	34	65.00	57.00	151	0	3	3	0.00	0.00	1	1
	RU-NE Average						**0**	**3**	**3**	**0.00**	**0.00**		
FI	RU-NW	Tuusula	41	61.11	26.20	64	2	6	8	0.25	0.43	2	1
RU	RU-NW	Karelia	35	61.40	33.40	120	3	7	10	0.30	0.47	2	1
RU	RU-NW	Karelia	36	61.50	30.28	84	7	2	9	0.78	0.39	1	2
RU	RU-NW	Karelia	37	61.40	36.33	131	1	9	10	0.10	0.20	2	1
RU	RU-NW	Leningrad	38	60.00	30.25	14	3	7	10	0.30	0.47	2	1
RU	RU-NW	Archangesk	39	62.54	40.24	70	0	10	10	0.00	0.00	2	1
RU	RU-NW	Vologda	40	59.15	39.30	169	0	10	10	0.00	0.00	2	1
	RU-NW Average						**16**	**51**	**67**	**0.25**	**0.28**		
RU	RU-SE	Tambov	42	53.12	41.20	152	5	5	10	0.50	0.56	1	2
RU	RU-SE	Voronez	43	51.50	39.20	94	4	3	7	0.57	0.57	1	2
RU	RU-SE	Michailovsk	44	51.00	43.00	101	1	4	5	0.20	0.40	2	1
	RU-SE Average						**10**	**12**	**22**	**0.42**	**0.51**		
BY	RU-SW	Magiliov	45	53.18	28.40	159	9	2	11	0.82	0.33	5	2
UA	RU-SW	Kijev	47	50.50	31.20	122	0	7	7	0.00	0.00	2	1
UA	RU-SW	Sumsk	48	52.01	34.01	185	0	10	10	0.00	0.00	2	1
UA	RU-SW	Rovno	46	51.30	26.40	153	4	6	10	0.40	0.53	1	1
	RU-SW Average						**13**	**25**	**38**	**0.30**	**0.22**		
BY	RU-W	Vitebsk	51	55.25	30.20	161	4	6	10	0.40	0.53	2	1
RU	RU-W	Pskov	49	57.43	30.31	89	9	1	10	0.90	0.20	3	2
RU	RU-W	Novgorad	50	58.15	33.28	194	6	4	10	0.60	0.53	1	2
	RU-W Average						**19**	**11**	**30**	**0.63**	**0.42**		
SE	SE	Kalmar	52	57.32	15.45	183	3	7	10	0.30	0.47	2	1
	SE Average						3	7	10	0.30	0.47	2	1
DE	WE	Erfurt	53	50.52	11.17	305	0	9	9	0.00	0.00	2	1
UK	WE	Scotland	54	55.23	-3.44	107	0	7	7	0.00	0.00	2	1
	WE Average						**0**	**16**	**16**	**0.00**	**0.00**		
	Grand Average									**0.26**	**0.30**		

[1] Country codes: BY – Byelorussia, DE – Germany, ET – Estonia, FI – Finland, GE – Georgia, PL – Poland, LT – Lithuania, LV – Latvia, RU – Russia, SE – Sweden, UA – Ukraine, UK – United Kingdom.

[2] Regions in Russia: RU-W Russia west, RU-SW Russia south west, RU-SE Russia south east, RU-NW Russia north-west, RU-NE Russia north-east, RU-E Russia east (west of Ural), RU-EE Russia east (east of Ural), RU-FE Russia far east.

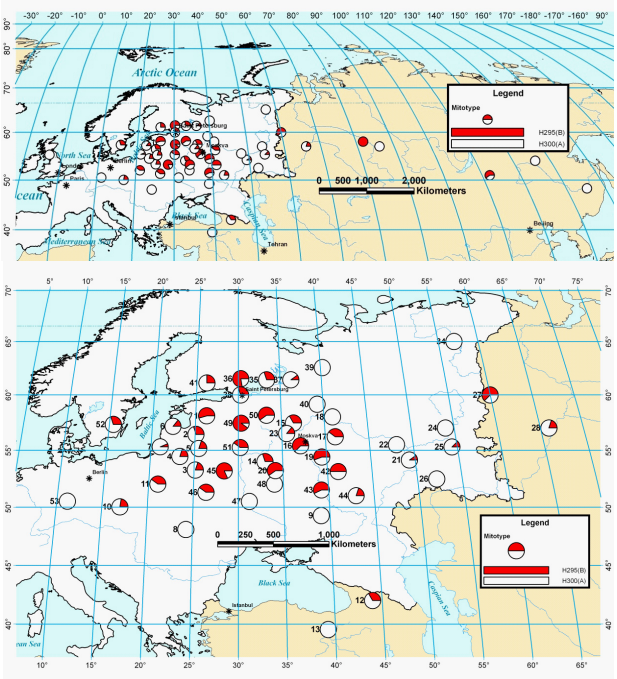

Figure 1. The geographical distribution of the Nad7-1 A and B haplotypes in Eurasia (upper map) and Europe (lower map). The population location is marked by the pie charts showing relative frequency of the mtDNA haplotypes. The solid line outlines the BAPS separated cluster of populations containing the type B haplotype at frequencies over 0.5. The dotted line delineates the region where woody vegetation could possibly survive during the LGM (Chaddadi et al. 2006)

DNA analysis

For the mtDNA Nad7-1 locus, we made an adjustment to avoid the restriction step in the laboratory analyses described by Naydenov et al. (2007). We have used the GeneBank deposited sequences of the Naydenov et al. (2007) mtDNA Nad7-1 haplotypes (DQ665913 to DQ665915) to design the primers flanking the 5 bp deletion and 32 bp deletion so that the PCR would result in the following three fragments (Naydenov et al. 2007), 300 bp (the universal A haplotype) and 295 bp (the northern B haplotype) and 268 bp (the C haplotype found in Turkish populations only) to be resolved on a DNA sequencer. The primers were designed with the BLAST primer designing tool: the forward primer sequence is 5' 6FAM-ATACCGTCTGGCGAAAACGCCG-3' (6FAM dye to be resolved in the Applied Biosystems sequencer); the reverse primer sequence was 5'-GGCCTCTCCATTTCCAATGACCCG-3' (no dye). The DNA was extracted from 100 mg of fresh needles by the CTAB protocol (Doyle and Doyle 1990). The PCR was carried out in 10 μl volumes, containing 10 × reaction buffer, 2.5 mM $MgCl_2$, 0.2 mM each of dNTP, 30 ng of DNA and 0.5 units of *Taq* DNA polymerase with "Mastercycler" thermocycler (Eppendorf). The PCR started with 2 min. of denaturation at 95°C, followed by 25 cycles of denaturation for 30 s at 94°C, annealing for 45 s at 60°C and extension at 72°C for 10 min. The fluorescence-labelled PCR products were separated by capillary electrophoresis with the GeneScan 500 LIZ standard on the ABI 3130 genetic analyser (Applied Biosystems). The fragments were analysed with the GeneMapper 4.0 software (Applied Biosystems).

For the mtDNA Nad1-B/C locus the PCR amplification results in two alleles: the universal haplotype found in whole Eurasia (217 bp, abbreviated here as A2) and the Iberian (248 bp, B2) (Soranzo et al. 2000). The primers used were 5'-TTAATC AAAAGGTCCGGAG-3' (forward); 5'-GTTGTACCGTAAACCTGCTC-3' (reverse) (Soranzo et al. 2000). The PCR mix is described above. The PCR steps were: 4 min. of denaturation at 94°C, followed by 30 cycles of denaturation for 30 s at 92°C, annealing for 30 s at 55°C and extension at 72°C for 6 min. The amplified fragments were hybridised with ethidium bromide and resolved on an agarose gel (Top Vision™ LE GQ Agarose) under the UV light in a PCR box.

Statistical analyses

The Nei's index of genetic diversity estimated without bias (H_e, equivalent to the expected heterozygosity for diploid data, Nei 1973) were used an estimate of intra-population diversity (calculated with the Haplotype Analysis ver. 1.05 software (Eliades and Eliades 2009).

The hierarchal AMOVA was run to assess the regional and population effects by using the F_{ST} distance type based on the number of different alleles (2000 permutations; Arlequin ver. 3.1., Excoffier and Lischer 2010). The AMOVA was carried out for both the original regional structure (based on the Scots pine regions of provenance in the former Soviet Union, Table 1) and for the regional structure defined by SAMOVA (see below). The AMOVA significance tests were run with 1023 permutations.

Presence of phylogeographic structure in the molecular data was tested by comparing the G_{ST} and N_{ST} fixation indexes calculated with the PERMUT software (Pons and Petit 1996). G_{ST} is an F_{ST} estimate that is based on the haplotype frequencies alone (not considering the population structure of haplotypes), while the N_{ST} considers the genetic relatedness among the haplotypes (Pons and Petit 1996). If the N_{ST} value is greater than G_{ST} value for the molecular data, then the closely related haplotypes tend to be located within a similar area, indicating a geographic structure of the molecular data.

The spatial analysis of molecular variance (SAMOVA) was used to assess geographical population structure (SAMOVA 1.0 software Dupanloup et al. 2002). The SAMOVA is based on the simulated annealing procedure, which aims to maximise the proportion of the total genetic variance due to the difference between the groups of populations (the F_{CT} index). The number of predefined groups (K) resulting in the maximum F_{CT} value was chosen.

In addition, the optimum number of population groups was investigated by the Bayesian analysis of population structure implemented the BAPS software version 5.3 (Corander et al. 2008a). The spatial group genetic mixture analysis was applied to populations, where the geographical location of the populations is considered in the clustering by assigning the biologically relevant non-uniform prior distribution over the space of clustering solutions under expectation that underlying clusters are spatially smooth to a certain degree (Corander et al. 2008b). The methods use a Markov chain Monte Carlo simulation method to group populations into optimum number of groups K, where multiple maximum K values were tested from 1 to 5 (each value repeated 3 times for verification). What BAPS does for each K value (even the replicates of the same value) is to find the optimal partitions with K less than the maximum K set by the user and after all the K values have been processed, it returns the clustering solution with the greatest marginal log-likelihood ratio value.

Results

For the Nad1-B/C locus, all our material was fixed at the A2 haplotype (217 bp, common in whole Eurasia, Soranzo et al. 2000, Naydenov et al. 2007) and further on the polymorphism at the Nad7-1 locus with the three A/B/C haplotypes is reported only.

Of the total of 474 individuals tested at the Nad7-1 locus, 348 individuals (73%) possessed the universal haplotype A of 300 bp (abbreviated as H300A) and 126 individuals (27%) possessed northern haplotype B of 295 bp (H295B) (Table 1). No haplotype C (Naydenov et al. 2007) with the 32bp indel was detected at the Nad7-1 locus in our material.

The within population diversity index (H_e) varied from 0 (100 % possession of either A or B haplotype) to 0.57 (A and B haplotypes in equal proportions) (Table 1). Thus, the greatest within population diversity at the Nad-7-1 locus was in the central part of European Russia where most of the populations possessed A and B haplotypes in near to equal proportions (Figure 2).

The population differentiation was moderate (G_{ST}= 0.239, s.e.= 0.0523). The AMOVA on the allele frequencies revealed significant population effect within region (F_{SC}=0.20; P<0.001) but not significant region effect (F_{CT}=0.07; P=0.07; Table 2). Probably, the regional structure based on small geographic zones was too detailed for detection of the large-scale spreading pattern of the mtDNA haplotypes. The N_{ST} was not greater but equal to G_{ST} indicating no phylogeny structure. However, this may be caused by a lack of statistical power because most populations contained only two Nad7-1 haplotypes differing in presence/absence of the 5 bp indel.

The SAMOVA indicates that the geographical distribution of the Nad7-1 alleles is not random and separates population groups with prevalence of either A or B haplotype by also considering population location (within region F_{CT} = 0.4; P<0.0001; Tables 1 and 2). Note, that after the SAMOVA delineation of the regions, the region effect in AMOVA became highly significant and the effect of population within region turned to not significant (Table 2). The most reasonable was the 5 region structure providing maximum F_{CT} value, where the largest group contained most of the populations with dominance of the universal A haplotype (43 populations), followed by a smaller group of 10 populations with more or less equal proportions of the A and B haplotypes (mainly the centre of European part of Rus-

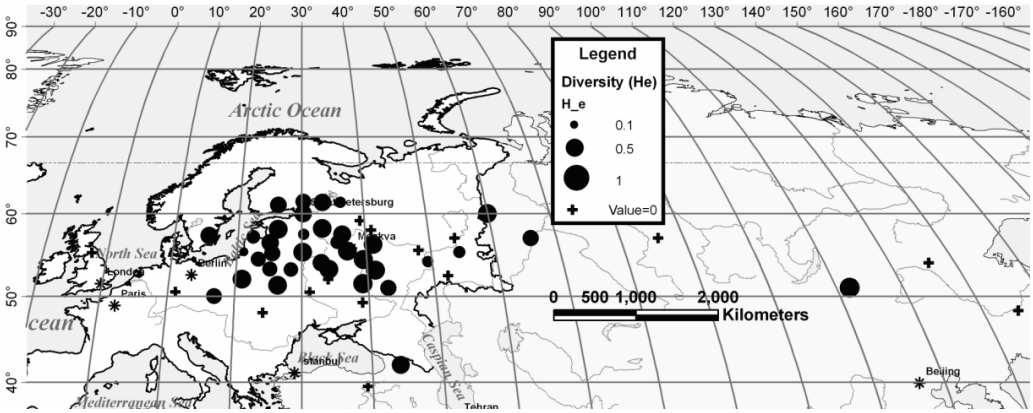

Figure 2. Geographical distribution of the within population diversity (H_e) at the mtDNA Nad7-1 locus. The population location is marked by filled circles sized by the within population diversity. Large circles mean high diversity. The small crosses indicate populations with H_e = 0

Table 2. Results of AMOVA and SAMOVA on the genetic effects on the variation at the mtDNA Nad7-1 locus. Distance method is number of different alleles (F_{ST})

Source of variation		d.f.	% of variation	Fixation Indices	P value
Among regions	AMOVA	13	7	F_{CT}=0.07	0.0655+-0.0073
	SAMOVA	4	44	F_{CT}=0.44	0.0000+-0.0000
Among populations within groups	AMOVA	38	18	F_{SC}= 0.20	0.0000+-0.0000
	SAMOVA	47	0.3	F_{SC}= 0.01	0.0000+-0.0000
Within populations	AMOVA	422	75	F_{ST}= 0.25	0.0000+-0.0000
	SAMOVA	422	55	F_{ST}=0.45	0.0000+-0.0000

sia) and 3 single population groups containing high frequency of the B haplotype but geographically separated from the rest of the populations with the dominance of the B haplotype (Table 1).

The BAPS clustering with the spatial option returned two cluster structure as optimum possessing the greatest value of the marginal log-likelihood ratio with a very low probability for presence of a third group (*p* equal to 0.002). The BAPS groups can simply be described based on the prevalence of either of the two haplotypes (Table 1) and so delineating the occurrence of the B haplotype group of the populations with the B haplotype frequency greater than 0.5 (Figure 1).

The A haplotype group is spread throughout Eurasia from Scotland to the eastern Asia and was found in all populations (Figure 1). The northern B haplotype group geographically was much smaller and formed a cline directed north-westwards of the south-eastern part of European Russia corresponding to the direction from Moscow to St. Petersburg (populations outlined by the solid line in Figure 1). The highest concentration of the northern B haplotype was found few hundred kilometres north-westwards of Moscow (Western Russia (Mogilov, Pskov regions), Southern Karelia, Figure 1). The frequency of the B haplotype decreased gradually when moving north-east and south-west of this north-westwards situated B haplotype cline. The Baltic populations contained the B haplotype but at low frequency. The southern Swedish population also contained the northern B haplotype (Figure 1). Populations eastwards of the Ural Mountains also possessed the B haplotype (Figure 1). The outliers from the above described trend for distribution of the B haplotype were populations No. 27 (Karpinsk at the foot of the Ural Mountains), No. 12 (Borzomi in Georgia), No. 32 (far-east population of Ingodisk at longitude of 122 degrees).

The data on the vegetation modelling during the LGM indicate that woody plants including Scots pine may have survived southeast of Moscow as shown by dotted line in Fig 1 based on Cheddadi et al. (2006). Consequently, the area where the prevalence of the northern mtDNA B haplotype overlaps with the possible zone of the remaining vegetation during LGM may indicate a possible location the northern refugium of Scots pine. Our data indicate a possible centre of this type B northern refugium of Scots pine at about 300 km south-east of Moscow at the intersection of latitudes 50°-55° N and longitudes 35°-45° E (Figure 1.). Occurrence of the B haplotype at a very low frequency immediately east and south of the suggested B haplotype refugial zone strongly supports our observation (Figure 1). Geographical distribution of the B haplotype found in the LGM glaciated areas may reflect possible post-glacial migration routes of Scots pine (Figure 1). According to our data, the major direction of the post-glacial migration for the type B haplotype is north-westwards of the area at the latitudes of 50°-55° N and longitudes 35°-45° towards St. Petersburg , Karelia and Finland (Figure 3).

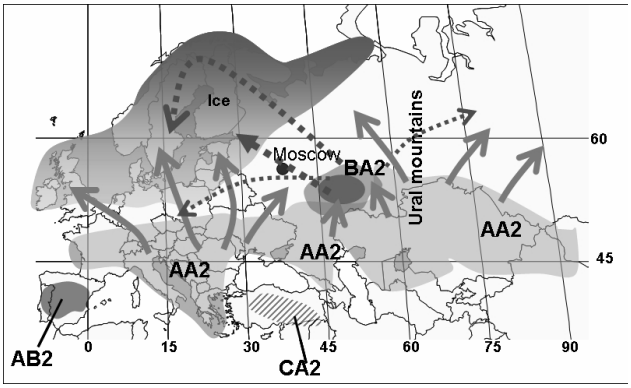

Figure 3. A simplified scenario of postglacial migration and location of the type B northern refugium of Scots pine based on our study and Naydenov et al. (2007). The haplotypes are coded as follows: the first letter indicates the allele of the Nad7-1 locus; the second letter the haplotype of Nad1-B/C locus

Discussion

Our study strongly supports the conclusions of Naydenov et al. (2007) and Pyhäjärvi et al. (2008) on the presence of a distinct northern refugium of Scots pine eastwards of Moscow. The geographical distribution patterns of the northern type B Nad7-1 haplotype in our study correspond well with that found by Naydenov et al. (2007) and Pyhäjärvi et al. (2008): (a) the increasing concentration of the type B haplotype towards northwest of Moscow; greatest in southern Karelia; (b) low frequency of the B haplotype immediately westwards of the Ural mountains; (c) presence of the B haplotype at low frequency in Poland, Czech Republic and southern Sweden. One difference of our study is that the B haplotype dominated in less of the north-eastern populations than in Naydenov et al. (2007) and Pyhäjärvi et al. (2008). This may be a consequence of smaller sample size and fewer populations from northern Karelia and Finland in our study.

In pines, the population differentiation G_{ST} value based on mtDNA markers is usually higher than obtained in our study (G_{ST}=0.239 compare with e.g. G_{ST} = 0.657 and 0.82 for Scots pine in Naydenov et al. 2007 and Pyhäjärvi et al. (2008), respectively or G_{ST}=0.569 for Jack pine in Godbout et al. 2005). However, most

of these studies used a relatively higher number of mtDNA alleles detected either by covering evolutionary distinct parts of the species range (e.g. Naydenov et al. 2007) or a greater number of polymorphic mtDNA loci (Godbout et al. 2005). High number of allelic variants increases the variation in the molecular data so increasing the chances to detect greater population differentiation values than in case of two alleles highly shared among populations as obtained in our study.

Our results should be interpreted bearing in mind the comparably low number of trees per population (an average of 10 instead of common 15 or 20 trees per population). However, in both Pyhäjärvi et al. (2008) and Naydenov et al. (2007) there were 8 – 10 populations with less than 11 trees. Given our budget, we chose to analyse more populations with less trees than vice versa.

Geographic distribution of the Nad7-1 haplotypes in our study and in Naydenov et al. (2007) as well as Pyhäjärvi et al. (2008) complement each other providing an evidence of the glacial refugium in the area where Scots pine could possibly survive during the LGM and where our study revealed the highest frequency of the type B northern haplotype. This area is at the intersection of latitudes 50°-55° N and longitudes 35°-45° E, (ca 300 km south-east of Moscow, Fig. 1). Scots pine is known as a very plastic pioneer tree species capable to withstand very low sub-zero temperatures (Rehfeldt et al. 2002). The general circulation modelling indicates possible Scots pine refugia during the LGM in the south-eastern part of European Russia and northern Ukraine, at the intersection of latitudes 45°-50°N and longitudes 30°-40°E, which is similar to our findings (Cheddadi et al. 2006). A more southern or eastern location is less likely because of the low frequency of haplotype B in the populations immediately east and south of the suggested refugium (Fig. 1, Naydenov et al. 2007, Pyhäjärvi et al. 2008). Nevertheless, the occurrence of the northern B haplotype at a comparably high frequency as far west as southern and northern Poland as well as the Czech Republic (Fig. 1, Naydenov et al. 2007, Pyhäjärvi et al. 2008) may indicate an additional source of the northern B haplotype, especially considering the remarks on the patchiness of the refugia by Cheddadi et al. (2006) where the Czech Republic was suggested to contain possible refugia. However, our study indicates that this seems to be less likely because of the low frequency or even absence of the type B haplotype in the Ukrainian and southern Russian populations, which should otherwise contain more of the type B haplotype if the source of migration would be located in the present-day Czech Republic (Fig. 1; Naydenov et al. 2007). Alternatively, it may be a post-glacial migration branch from the northern refugium south westwards as in Norway spruce (Lagercrantz and Ryman 1990).

Except for the southern-most parts of Eurasia, the Nad7-1 universal haplotype A dominates both numerically and geographically and is present to a high degree even in the populations from the suggested northern type B refugium (Figure 1; Naydenov et al. 2007, Pyhäjärvi et al. 2008). This observation supports the suggestions that the northern B refugium was much smaller in size and co-migration may have occurred from both from the A and B refugia. If the glacial refugia were unevenly distributed and patchy as suggested by Cheddadi et al. (2006), the A and B refugia could even be intermixed or, alternatively, the refugium B spread first and was later invaded by A from more southern sources. The comparably uneven distribution of the B haplotype in north-eastern Europe found in our study may be a result of successive founder events during colonization (Ibrahim et al. 1996, Naydenov et al. 2007) and/or the relatively smaller sampling sizes in our study (Table 1). A mixed co-migration for several refugia was discussed as a possible scenario for many forest trees (Petit et al. 2003). There also is a possibility that the B haplotype occurred in more southern locations during the last glaciation assuming that such signature was lost during the northward migration or via genetic drift.

The trans-Eurasian location of the A mtDNA haplotype found in our study and in Naydenov et al. (2007) indicates its origin from multiple refugia that could primarily be the southern parts of central and eastern Europe (Cheddadi et al. 2006) as well as south-western Siberia (Chernova et al. 1991). For *Picea abies,* another cold tolerant coniferous species, studies based on fossil pollen data and genetic inferences suggest a northern refugium near Moscow (Huntley and Birks 1983, Lagercrantz and Ryman 1990). A simplified picture of the post-glacial refugia based on our results as well as Naydenov et al. (2007) and Pyhäjärvi et al. (2008) studies on the polymorphism at the mtDNA Nad7-1 and Nad1-B/C loci is given in Fig. 3. The type A refugium is given in a simplified continuous manner, because the location of the refugial patches is largely unknown for the universal A haplotype.

Our study, together with Naydenov et al. (2007), indicates that the major post-glacial migration route of the northern type B haplotype was towards the north-west via St. Petersburg and further north to Karelia and northern Finland. The low frequency of the Nad7-1 B haplotype eastwards of the suggested northern refugium implies that the migration of the B haplotype towards north-east and east was much weaker

(Figure 1; Naydenov et al. 2007). A feasible interpretation of such observation is (a) as regards migration north-westwards, the type B northern refugium populations were superior in frost hardness and possessed an adaptive advantage over the more southerly type A populations in colonising the northern territories; (b) type B refugium could have been geographically closer to spread north-westwards than type A refugia; (c) the migration of type B haplotype north-eastwards and eastwards could have been relatively weaker, because type A populations were already there or it could have been much shorter distance for the type A populations to reach the eastern territories, where the existence of the western Siberian refugia was suggested by Chernova et al. (1991).

Presence of type A haplotype in southern Sweden (Figure 1; Neydenov et al. 2007, Pyhäjärvi et al. 2008) may indicate migration via seeds over the Baltic Sea. Given the high proportions of the A haplotype in the Swedish population in our study and in Pyhäjärvi et al. (2008), this migration should have been of a significant amount. Another indication of the migration over the Baltic Sea could be a relatively higher haplotype sharing among the Baltic and south-eastern Swedish populations than among Baltic and the more northern Swedish populations as found for central Sweden by Neydenov et al. (2007). The co-migration of the A haplotype from northern Scandinavia together with the B haplotype is less likely because of the low frequency of the A haplotype in northern Finland (Naydenov et al. 2007). However, Pyhäjärvi et al. (2008) found high frequency of the A haplotype in northern Sweden, which is difficult to interpret because the adjacent populations from northern Finland, on the contrary, were almost fixed for the B haplotype. Another similar example of such inconsistency is the relatively lower frequency of the B haplotype in the southern Finnish population in our study (Figure 1) compared with Neydenov et al. (2007). A possible interoperation could be that in the regions exhibiting inconstant results after repeated sampling, the A and B haplotypes are highly intermixed and such intermixture may indicate a relatively higher genetic diversity enhanced by the presence of the descendents from the two different glacial refugial lines. If true, high genetic diversity can be interpreted as a genetic signature of such regions with heterogeneous refugial material.

In support to the above, our study revealed the highest within population diversity in the regions where the Nad7-1 haplotypes A and B occur in equal proportions indicating possible convergence of northern type B and universal type A post-glacial refugia (Figure 2). In this respect, it would be interesting to compare the genetic diversity at the cpDNA and nuclear DNA levels of the region containing similar frequencies of both A and B haplotypes with the other areas within the species range.

The two Asian populations (No. 33 and No. 32) containing unexpectedly high proportions of the B haplotype and so deviating from the above explained general patterns, could be interpreted as outliers owing to the low sampling size (No. 32 with 5 trees and 33 with 1 tree is given on the map in Fig. 1 only and not included in the further analysis). Sampling errors during the establishment of the Prokazin provenance test cannot completely be ruled out. However, Naydenov et al. (2007) also found a 20 % frequency of the Nad7-1 B haplotype in one population from this region.

In conclusion, our study based on the geographical distribution patters of the mtDNA nad7-1 haplotypes supports the hypothesis of the northern Scots pine glacial refugium and provides information on migration route as well as a possible location of this northern refugium. The group of populations with high frequency of the northern haplotype B constitutes a geographically consistent cline directed north-westwards of the south-eastern part of European Russia and so indicating the north-westward direction of the post-glacial migration route. This provides a stronger support for the eastern rather than the central European location of the northern glacial refugium. A possible location of this northern refugium could be at about 300 km south-eastwards of Moscow at the intersection of latitude 50°- 55° N and longitude 35°- 45° E, where the northern type B haplotype occurs in high frequency and Scots pine could possible survive during the LGM. However, there also is a possibility for a more southern location of the northern refugium during the last glaciation assuming that such signature was lost during the northward migration or via genetic drift.

Acknowledgments

We express sincere thanks to Assoc. Prof. Julius Danusevičius for the guidance in the field tests. Thanks are also due to reviewers for helpful comments.

References

Abbott, R. J. and Brochmann, C. 2003. History and evolution of the arctic flora: in the footsteps of Eric Hultén. *Molecular Ecology* 12: 299-313.

Birks, H.J.B. 1989. Holocene isochrone maps and patterns of tree-spreading in the British Isles. *Journal of Biogeography* 16: 503–540.

Cheddadi, R., Vendramin, G. G., Litt, T., François, L., Kageyama, M., Lorentz, S., Laurent, J. M., de Beaulieu, J. L., Sadori, L., Jost, A. and Lunt, D. 2006. Imprints of glacial refugia in the modern genetic diversi-

ty of *Pinus sylvestris*. *Global Ecology and Biogeography* 15: 271-282.

Chernova, G. M., Mikhailov, N. N., Denisenko, V. P., and Kozyreva, M. G. 1991. Some questions of paleogeography of Holocene of South Eastern Altai. *Izvestia of All-Union Geographical Society* 2 140-146.

Corander, J., Marttinen, P., Sirén, J. and Tang, J. 2008a. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinformatics 9: 539.

Corander, J., Sirén, J. and Arjas, E. 2008b. Bayesian spatial modelling of geneticpopulation structure. *Computational Statistics* 23: 111-129.

Doyle, J. J., and Doyle, J. L. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12 (1): 13-15.

Dupanloup, I., Schneider, S. and Excoffier, L. 2002. A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* 11: 2571-2581.

Eliades, N-G., and Eliades, D. G. 2009. HAPLOTYPE ANALYSIS: software for analysis of haplotypes data. Distributed by the authors. Forest Genetics and Forest Tree Breeding, Georg-Augst University Goettingen, Germany.

Excoffier, L. and Lischer, H. 2010. ARLEQUIN, VER 3.5.1.2 USER, MANUAL. Computational and Molecular Population Genetics Lab (CMPG), Institute of Ecology and Evolution University of Berne, Baltzerstrasse 6, 3012 Bern, Switzerland.

Froyd, C. A. 2005. Fossil stomata reveal early pine presence in Scotland: implications for post-glacial colonization analyses. *Ecology* 86: 579–586.

Godbout, J., Jaramillo-Correa, J.P., Beaulieu, J. and Bousquet, J. 2005.A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (Pinus banksiana), a broad-range North American conifer. *Molecular Ecology* 14: 3497–3512.

Gros-Louis, M-C., Bousquet, J., Pāques, L. E. and Isabel, N. 2005. Species-diagnostic markers in Larix spp. based on RAPDs and nuclear, cpDNA, and mtDNA gene sequences, and their hylogenetic implications. *Tree Genetics and Genomes* 1: 50-63.

Huntley, B., and Birks, H. J. B. 1983. An Atlas of Past and Present Pollen Maps for Europe: 0–13 000 years Ago Cambridge (UK): Cambridge University Press.

Ibrahim, K. M., Nichols, R. A. and Hewitt, G. M. 1996. Spatial patterns of genetic variation generated by different forma of dispersal during range expansion. *Heredity* 77: 282-291.

Jaramillo-Correa, J. P., Bousquet, J., Beaulieu, J., Isabel, N., Perron, M. and Bouillé, M. 2003. Cross-species amplification of mitochondrial DNA sequence-tagged-site markers in conifers: the nature of polymorphism and variation within and among species in *Picea*. *Theoretical and Applied Genetics* 106: 1353-1367.

Kremenetski, C.V., Liu, K. and MacDonald, G.M. 1998. The late Quaternary dynamics of pines in northern Asia. In: Richardson DM (ed.). Ecology and Biogeography of *Pinus*. Cambridge: Cambridge University Press, p. 95-106.

Lagercrantz, U., and Ryman, N. 1990. Genetic structure of Norway spruce (*Picea abies*) – concordance of morphological and allozymic variation. *Evolution* 44: 38-53.

Magri, D., Vendramin, G. G., Comps, B., Dupanloup, I., Geburek, T., Gömöry, D., Latalowa, M., Litt, T., Paule, L., Roure, J. M., Tantau, I., van der Knaap, W. O., Petit, R.J., and de Beaulieu, J. L. 2006. A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Physiologist* 171: 199-221.

Naydenov, K., Senneville, S., Beaulieu, J., Tremblay, F., and Bousquet, J. 2007. Glacial vicariance in Eurasia: mitochondrial DNA evidence from Scots pine for a complex heritage involving genetically distinct refugia at mid-northern latitudes and in Asia Minor. *BMC Evolutionary Biology* 7: 233, doi:10.1186/1471-2148-7-233.

Neale, D. B., and Sederoff, R. R. 1989. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. *Theoretical and Applied Genetics* 77: 212-216.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. In: Proceedings of National Academy of Sciences of USA 70: 3321-3323.

Petit, R. J., Aguinagalde, I., Beaulieu, J. L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Muller-Starck, G., Demesure-Musch, B., Palmé, A., Pedro Marti, J., Rendell, S., and Vendramin G. G. 2003 Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300: 1563–1565.

Pons, O., and Petit, R. J. 1996. Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* 144: 1237-1245.

Pyhäjärvi, T., Salmela, M. J., and Savolainen, O. 2008. Colonization routes of *Pinus sylvestris* inferred from distribution of mitochondrial DNA variation. *Tree Genetics and Genomes* 4: 247–254.

Rehfeldt, G. E., Tchebakova, N. M., Parfenova, Y. I., Wykoff, W. R., Kuzmina, N. A., and Milyutin, L. I. 2002 Intraspecific responses to climate in *Pinus sylvestris*. *Global Change Biology* 8: 912–929.

Shutyaev, A. M., and Giertych, M. 1998 Height growth variation in a comprehensive Eurasian provenance experiment of (*Pinus sylvestris* L.). *Silvae Genetica* 46: 332–349.

Sinclair, W. T, Morman, J. D., and Ennos, R. A. 1999. The postglacial history of Scots pine *(Pinus sylvestris* L.) in Western Europe: evidence from mitochondrial DNA variation. *Molecular Ecology* 8: 83-88.

Sinclair, W. T., Morman, J. D., and Ennos, R. N. 1998. Multiple origins of Scots pine in Scotland: evidence from mitochondrial DNA variation. *Heredity* 80**:** 233-240.

Soranzo, N., Provan, J., and Powell, W. 1999. An example of microsatellite length variation in the mitochondrial genome of conifers. *Genome* 42: 158-161.

Soranzo, N., Alía, R., Provan, J., and Powell, W. 2000. Patterns of variation at a mitochondrial sequence-tagged-site locus provides new insights into the postglacial history of European *Pinus sylvestris* populations. *Molecular Ecology* 9**:** 1205-1211.

Sperisen, C., Büchler, U., Gugerli, F., Mátyás, G., Geburek, T., and Vendramin, G. G. 2001. Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. *Molecular Ecology* 10: 257-263.

Stewart, J. R., and Lister, A. M, 2001. Cryptic northern refugia and the prigins of the modern biota. *Trends in Ecology and Evolution* 16**:** 608-613.

Taberlet, P., Fumagalli, L., Wust-Saucy, A. G., and Cosson, J. F. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7: 453-464.

Webb, T., and Bartlein, P. J. 1992. Global changes during the last 3 million years: climatic controls and biotic responses. *Annual Review of Ecology and Systematics* 23: 141-173.

Willis, K. J., Bennett, K. D., and Birks, H. J. B. 1998. The late Quaternary dynamics of pines in Europe. In:

Ecology and Biogeography of Pinus Edited by: Richardson DM. Cambridge (UK): Cambridge University Press 1998, p. 107-121.

Willis, K., and van Andel, V. H. 2004. Trees or no trees? The environments of central and eastern Europe during the last glaciation. *Quaternary Science Reviews* 23: 2369–2387.

Wolfe, K.H., Li, W. H., and Sharp, P. M. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of National Academy of Sciences of USA 84: 9054-9058.

Received 22 November 2012
Accepted 07 June 2013

МЕСТОПОЛОЖЕНИЕ СЕВЕРНОГО ПОСЛЕЛЕДНИКОВОГО РЕФУГИУМА СОСНЫ ОБЫКНОВЕННОЙ НА ОСНОВЕ МАРКЕРОВ МИТОХОНДРИАЛЬНОЙ ДНК

Ю. Буховска, Д. Данусявичюс, Д. Банюлис, В. Станис, Ю. Б. Шикшнянене и Д. Каваляускас

Резюме

По данным нескольких недавних исследований, основанных на маркерах митохондриальной ДНК, северный рефугиум сосны обыкновенной находился приблизительно на западе от южной части Уральских гор. Целью нашего исследования явилось изучение полиморфизма мтДНК сосны обыкновенной по локусам Nad7-1 и Nad1-B / C, в основном направленного на выявление местоположения этого рефугиума и связанных с ним послеледниковых путей миграции. Мы исследовали 54 популяции, плотно охватывающие европейскую часть России на запад от Уральских гор, и включили также популяции из Чехии, Польши, Швеции, Финляндии, Шотландии, Грузии и Восточной Сибири. По локусу Nad1-B / C весь наш материал получился мономорфным. Из общего количества 474 особей, испытанных по локусу Nad7-1, 348 особи (73%) обладали универсальным гаплотипом A в 300 пар нуклеотидов и 126 (27%) - северным гаплотипом в 295 п.н. Географическое распределение северного гаплотипа по локусу Nad7-1 не было случайным (SAMOVA, BAPS) и указывало на центр местоположения его послеледникового рефугиума примерно в 300 км к юго-востоку от Москвы на пересечении 50-55 градусов северной широты и 35-45 градусов восточной долготы. Основной послеледниковый путь миграции из этого северного рефугиума расположен на северо-запад в направлении Карелии и Северной Финляндии.

Ключевые слова: дифференциация, оледенение, ЛГМ, сосна обыкновенная, филогеография, послеледниковая колонизация древесных видов, митохондриальная ДНК, ДНК органелла