Tree Genetics & Genomes (2008) 4:611–623
DOI 10.1007/s11295-008-0137-1

ORIGINAL PAPER

Phylogeography of *Larix sukaczewii* Dyl. and *Larix sibirica* L. inferred from nucleotide variation of nuclear genes

Neyton H. T. Araki · Ismael A. Khatab · Kariyawasam K. G. U. Hemamali · Nobuyuki Inomata · Xiao-Ru Wang · Alfred E. Szmidt

Received: 4 April 2007 / Revised: 18 November 2007 / Accepted: 28 January 2008 / Published online: 4 April 2008

Abstract We investigated phylogeography of *Larix sukaczewii* and *Larix sibirica* using nucleotide variation at three following nuclear gene regions: 5.8 S rDNA including two internal transcribed spacers (*ITS*), cinnamyl alcohol dehydrogenase (*CAD*), and phytochrome-O (*PHYO*). We also included sequences of the 4-coumarate: coenzyme A ligase (*4CL*) gene region obtained in our recent study. *CAD* and *PHYO* showed very low nucleotide variation, but *ITS* and *4CL* had levels of variation similar to those reported for other conifers. Pleistocene refugia have been hypothesized to exist in the Southern Urals and South Central Siberia, where four out of nine of the investigated populations occur. We found moderate to high levels of population differentiation (F_{ST}=0.115–0.531) in some pairwise comparisons suggesting limited gene flow and independent evolution of some refugial populations. In *L. sukaczewii*, low levels of differentiation were found among populations from areas glaciated during the Pleistocene, indicating their recent origin. Our results also suggest these populations were created by migrants from multiple, genetically distinct refugia. Furthermore, some haplotypes observed in populations from previously glaciated areas were not found in putative refugial populations, suggesting these populations might have contributed little to the extant populations created after the Last Glacial Maximum. Some authors regard *L. sukaczewii* and *L. sibirica* as a single species, while others consider them as separate species. The observed conspicuous differences in haplotype composition and distribution between *L. sukaczewii* and *L. sibirica*, together with high values of F_{ST} between populations of the two species, appear to support the latter classification.

Keywords Phylogeography · *Larix* · Conifer · Internal transcribed spacer (*ITS*) · 5.8 S nuclear ribosomal DNA (rDNA) · 4-coumarate: coenzyme A ligase (*4CL*) · Pleistocene refugia

Electronic supplementary material The online version of this article (doi:10.1007/s11295-008-0137-1) contains supplementary material, which is available to authorized users.

Ismael A. Khatab and Kariyawasam K.G.U. Hemamali contributed equally to this work.

Communicated by S. Aitken

N. H. T. Araki · I. A. Khatab · N. Inomata · A. E. Szmidt (✉)
Deparment of Biology, Graduate School of Science,
Kyushu University,
6-10-1, Hakozaki,
Higashi-ku, Fukuoka 812-8581, Japan
e-mail: aszmiscb@mbox.nc.kyushu-u.ac.jp

K. K. G. U. Hemamali
Deptarment of Applied Genetics and Pest Management,
Faculty of Agriculture, Kyushu University,
6-10-1, Hakozaki,
Higashi-ku, Fukuoka 812-8581, Japan

X.-R. Wang
Department of Ecology and Environmental Science,
Umeå University,
901 87 Umeå, Sweden

I. A. Khatab
Department of Genetics, Faculty of Agriculture,
Kafr El-Sheikh University,
Kafr El-Sheikh, Egypt

Introduction

The history of postglacial dispersal of many plant species has been clarified by phylogeographic studies. However, there is still little knowledge on the phylogeography of the genus *Larix* Mill. The biogeographic history of *Larix* and

other plants in Eurasia has been shaped by Pleistocene glaciations (Hewitt 2000). During the Last Glacial Maximum (LGM), most of Northwestern Eurasia was covered by glaciers or tundra, to approximately 57° N in European Russia, as well as in Western and Central Siberia (Svendsen et al. 1999; Tarasov et al. 2000). According to fossil data, forest refugia were present in the southern mid-latitudes of Eurasia, such as the north of the Sea of Azov and east of the Ural watershed (north of Caspian Sea, northwest of Aral Sea and Southern Urals; Hewitt 2004; Tarasov et al. 2000). Other refugia were in the Tien-Shan Mountains (Kazakhstan) and in northern Mongolia (Tarasov et al. 2000). Furthermore, in the Altai region, forests could only grow at altitudes lower than 1,000 m during the LGM and migration of trees into the Altai from nearby refugia occurred solely after deglaciation (Blyakharchuk et al. 2004).

In addition to such complex history of Eurasia, the reproductive biology of *Larix* species suggests that their populations may have high levels of differentiation because under normal conditions, their pollen and seeds usually disperse over less than 100 m (Brown et al. 1988; Duncan 1954; Hall 1986; Knowles et al. 1992). Therefore, extant populations of *Larix* species are likely to have complex origins and genetic structures. Yet, most previous studies on *Larix* suggested relatively simple pictures, such as low levels of genetic differentiation at both inter and intraspecific levels. Recent speciation on the geological time scale, lack of reproductive isolation, and recent divergence of extant populations were given as explanations for such low differentiation (Gros-Louis et al. 2005; Larionova et al. 2004; Lewandowski 1997; Semerikov and Lascoux 2003; Semerikov and Lascoux 1999; Timerjanov 1997; Wei et al. 2003; Wei and Wang 2004b).

The classification status of *Larix* populations occurring from Western Russia to Central Siberia is controversial. Some authors have considered these populations as a single species, the *Larix sibirica* L. (Kullman 1998; Milyutin and Vishnevetskaia 1995; Semerikov et al. 1999; Wei and Wang 2004a). However, because they display a slight geographic gradient of morphological traits along their distribution range, populations westward of the Irtysh and Ob rivers are considered by some authors as an independent species: *Larix sukaczewii* Dylis (Bashalkhanov et al. 2003; Dylis 1947). In this classification, *L. sibirica* refers to populations found mainly in Central Siberia, while *L. sukaczewii* refers to those found in Western Russia (Abaimov et al. 2002; Abaimov et al. 1998; Timerjanov 1997).

While there is some information about *Larix* species for DNA markers (microsatellites, amplified fragment length polymorphism (AFLP) etc.), there is still very little information about levels and patterns of nucleotide variation in coding regions of nuclear genome. As demonstrated in our recent study of the Eurasian *Larix* species, such information can give important insights into history and classification of this genus (Khatab et al. 2008).

In this study, we directly sequenced partial regions of the 5.8 S rDNA gene including two internal transcribed spacers *ITS*1 and *ITS*2 (hereafter referred to as *ITS*), the cinnamyl alcohol dehydrogenase (*CAD*) and the phytochrome-O (*PHYO*). We also included sequence data for the partial region of the 4-coumarate: coenzyme A ligase (*4CL*) gene obtained in our previous study (Khatab et al. 2008). We examined six populations of *L. sukaczewii* and three populations of *L. sibirica*. All investigated populations of *L. sibirica* came from locations that were not glaciated during the Pleistocene and some of them are regarded as glacial refugia (Tarasov et al. 2000). On the other hand, four of the investigated populations of *L. sukaczewii* (populations 1 through 4) are located in a previously glaciated area, while the two remaining populations (4 and 5) are located in areas regarded as glacial refugia (Tarasov et al. 2000).

The main objectives of the present study were (1) to determine whether, as suggested by previous studies, populations of *L. sukaczewii* and *L. sibirica* are weakly differentiated; (2) to verify taxonomic status of *L. sukaczewii* and *L. sibirica*; and (3) to provide new information about the demographic history of both species.

Materials and methods

Seed samples of *L. sukaczewii* and *L. sibirica* were collected from natural forests in Russia (Abaimov et al. 2002). Details on number of samples per population and locations of the nine populations used in this study are shown in Table 1. The number of samples was not uniform among the investigated DNA regions and populations due to either non-amplification of the DNA target or depletion of DNA stock during experiments.

DNA extraction, amplification, and sequencing of the target DNA regions

Seeds were kept for 2–3 days on moist sterilized paper to facilitate separation of the maternal haploid megagametophyte tissues from seed coats and embryos. Total genomic DNA was isolated from megagametophytes using the SDS method (Ish-Horowicz 1989) with some modifications.

The *ITS* is a multi-copy region that is believed to evolve in a concerted fashion (Wei and Wang 2003). Previous phylogeographic studies of trees and annual plants, such as *Fraxinus* sp. (Jeandroz et al. 1997), *Olea europaea* L. (Hess et al. 2000), *Helenium virginicum* (Simurda and Knox 2000), *Saxifraga oppositifolia* (Holderegger and Abbott 2003), *Pritzelago alpina* (Kropf et al. 2003), and *Clausia aprica* (Franzke et al. 2004), have successfully used the *ITS*

Table 1 Sample sizes and locations of *Larix* populations used in this study

	Sample Size		Location of populations			
Population	*ITS* region	*4CL* region	Region name in Russia	Latitude (N)	Longitude (E)	Altitude (m)
Larix sukaczewii						
1 (1D)	11	10	Nizhnij Novgorod	57°3''	45°1''	n.a.
2 (2A)	11	9	Plesetsk	63°0''	40°2''	100
3 (4A)	8	10	Petchora	66°0''	57°4''	n.a.
4 (5A)	10	10	Selechard	63°4''	66°4''	60
5 (6B)	10	9	Perm	56°0''	59°3''	460
6 (7A)	8	10	Ufa	55°4''	56°5''	370
Total	58	58				
Larix sibirica						
7 (9A)	10	10	Boguchany	58°3''	97°3''	158
8 (10A)	10	8	Novokuznetsk	53°4''	88°0''	n.a.
9 (11A)	10	10	Altai	50°1''	87°5''	1630
Total	30	28				

Original population designations used by Abaimov et al. (2002) are given in parentheses
n.a. Not available

region as DNA marker. The *4CL* is a low-copy gene that has been used in phylogenetic studies of *Pinaceae* (Wang et al. 2000). Two to three copies of the *4CL* gene exist in *Larix* species (Wei and Wang 2004a). The *CAD* gene has been reported to exist as a single copy in *Pinus taeda* (MacKay et al. 1995) and as a small gene family in *Picea abies* (Schubert et al. 1998). Both the *4CL* and the *CAD* genes play roles in the lignin biosynthetic pathway (Wei and Wang 2004a; Whetten and Sederoff 1995). The *PHYO* gene was used in a study of nucleotide diversity along a latitudinal cline in *Pinus sylvestris* (Garcia-Gil et al. 2003). Phytochrome acts as the photoreceptor that mediates red light effects on various physiological and molecular responses in plants (Sharrock and Quail 1989).

Primers for the *ITS*, *CAD*, and *PHYO* gene regions were designed based on conserved DNA sequence regions of *Larix* species from the GenBank using Primer3 (Rozen and Skaletsky 2000) and GeneFisher (Reeder et al. 2006), both of which are website-based primer designers. The primer sequences (5′–3′) used in this study were as follows: *ITS* Fwd: TGCGGTAGGATCATTGATAGCA, Rev: AGCCCA AACCTATCCATCCGA, *CAD* Fwd: CACTTACACTCTC AGGTACA, Rev: GAAGGGCCAGATAAGGTTCCA, *PHYO* Fwd: GAGGTAGTTGCAGAGATGAGA, Rev: ATATTGGGAGTCTGAGACACA. The polymerase chain reaction (PCR) mixture was prepared to the total volume of 50 µl containing 50–100 ng DNA template, 50 mM KCL, 10 mM Tris-HCl pH 8.3, 1.5 mM $MgCl_2$, 2.5 pmol of each primers and 2 mM each of dATP, dGTP, dCTP, and dTTP (Amersham Bioscience, USA) and 1 unit of Taq polymerase. The amplification of the *ITS* region was carried out after denaturing the DNA at 95°C for 5 min followed by 35 cycles of 30 s at 95°C, 45 s at 55°C for annealing, 60 s at 72°C, and ending with 7 min at 72°C for further extension. The amplification of the *PHYO* region was as follows: 95°C for 5 min, 35 cycles of 30 s at 95°C, 30 s at 55°C, 30 s at 72°C and then a further extension of 7 min at 72°C. The amplification of the *CAD* region was 95°C for 5 min, followed by 35–40 cycles of 30 s at 95°C, 45 s at 50–55°C, 60 s at 72°C, ending with 7 min at 72°C for further extension. The amplification of the *4CL* region was performed according to (Wang et al. 2000). All PCR products were purified using Wizard^{R} SV Gel and PCR Clean-Up System (Promega, USA) following manufacturer's instructions. Purified PCR products were directly sequenced on the ABI Prism 3100 Genetic Analyzer (Applied Biosystems), using the BigDye^{TM} Terminator (v 3.1) Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems) according to manufacturer's instructions, and sequences were determined for both strands. Additional internal primers were used during sequencing (data not shown). Sequences obtained in this study have been deposited in the GenBank with the following accession numbers: *ITS* EU441958–EU441983, *CAD* EU441953–EU441957, and *PHYO* EU441947–EU441952.

Data analyses

Sequences of both strands were checked using the Sequence Navigator 1.01 (Applied Biosystems, Foster City, CA, USA) and the ATGC program version 4 (GENETYX Corporation). Complete sequences of individuals were aligned using the ClustalX program ver. 1.83 (Thomson et al. 1997). The DnaSP program ver. 4.10.9 (Rozas et al. 2003) was used to perform the following sequence

analyses: (1) nucleotide diversity per site (π; Nei 1987), (2) nucleotide polymorphism (θ) (Watterson 1975), (3) the number of haplotypes (H), and (4) the following neutrality tests: Tajima's D (Tajima 1989), Fu and Li's D^* and F^* (Fu and Li 1993), and Hudson, Kreitman, and Aguadé's (HKA) test (Hudson et al. 1987). The HKA test was calculated by direct input mode because multilocus analyses could not be performed due to differences in number of samples between the investigated DNA regions (Table 1). Measures of population differentiation (F_{ST}; Hudson et al. 1992) were performed using the ProSeq ver. 2.9 program (Filatov 2002) with 1,000 permutations to obtain p values. When the obtained p values were close to the significance level of 0.05, calculations were repeated with 10,000 to 100,000 permutations to obtain more accurate significance estimates. Two types of treatments were used: one where gaps were considered as segregating sites and the other where they were excluded.

To determine whether the size of the investigated populations remained constant or experienced expansion events, we carried out coalescent simulations using the approach developed by Hudson (2002) and Hudson and Coyne (2002). Simulations were performed using the ms program developed by Hudson (2002). Populations 1, 2, 3, and 4 were considered as a single population in coalescent analyses because no significant population differentiation was found among them (Table 3). In simulations, we used the number of samples examined (n=40 for the *ITS* and 39 for the *4CL* region) and the observed number of segregating sites (S). The observed number of segregating sites was the same (S = 9) for both *ITS* and *4CL* regions.

In simulations for population expansion, we examined the range of the following parameters: three different values of possible today's effective population size (No; 1.0×10^4, 1.0×10^5, and 1.0×10^6), five different starting times of population expansion in generations before present (T=10, 50, 100, 500, 1000 generations), and the ratio of expansion (No/Nt), where Nt is the population size at a given time in generations in the past. The parameter T was assigned for a wide range of time of population expansion, considering that the minimum generation time of *Larix* species varies between 10 and 20 years (Young and Young 1992). Therefore, the time of population expansion examined ranged from as a few hundred years (≈10 generations) to as old as the time of the LGM (≈1,000 generations). The values of the ratio No/Nt were, 10, 100, 1,000, and 10,000. Ratios of population expansion larger than 10,000 are not realistic because most Pleistocene *Larix* refugia occupied relatively large areas (Tarasov et al. 2000). Additional values within the range shown above of T and No/Nt were also used to obtain closer matches between the simulation and empirical results.

Both T and No/Nt were entered as their input parameters t and α (http://home.uchicago.edu/~rhudsonl/source/niksamples/msdir). The parameter t is the time before present, measured in units of $4No$ ($t=T/4No$), and α is the growth parameter ($\alpha=-1/t$ log (Nt/No)). In each simulation for population expansion, the parameter α was set to zero at time t, when the expansion was supposed to have started. The values of nucleotide diversity per locus (πT) and Tajima's D were obtained for each combination of t and α based on 10,000 independent samples generated during simulations. The same number of independent samples was generated when no change in population size was assumed.

Haplotype networks were constructed using median-joining method as implemented in the NETWORK program ver. 4.2.0.1 (Bandelt et al. 1999) to visualize relationships and frequencies of individual haplotypes (indels were considered in the analysis).

Results

Sequence variation

The obtained lengths of the aligned sequences (including indels) were: the *ITS* region=1777 bp, the *4CL* region=758 bp, the *CAD* region=1331 bp, and the *PHYO*=565 bp. In the *ITS* region, the total number of S was 31, including 17 singletons and one indel. Twenty nine segregating sites, including 16 singletons, were found in the *ITS*1 region (18 – 1,390 bp position). Two segregating sites including one singleton were observed in the *ITS*2 region (1,553–1,777 bp) in *L. sibirica* (Fig. 1). No variation was found in the 5.8 S rDNA region (1,391 – 1,552 bp). Eleven segregating sites were found in the *4CL* region, including five singletons and two indels. Eight segregating sites including one replacement were found in exon 1 (1 – 654 bp). Three segregating sites (one singleton and two indels) were found in the intron (655 – 736 bp) and no variation was observed in exon 2 (737 – 758 bp; Fig. 2). In the *CAD* region, only three haplotypes were found among 48 sequences of *L. sukaczewii* and five sequences of *L. sibirica*. Forty-seven sequences of *L. sukaczewii* and one of *L. sibirica* represented only one haplotype. Four sequences of *L. sibirica* represented another haplotype, which differed from the previous one by only one non-synonymous substitution at 607 bp position in exon 3. The third haplotype was found in only one individual of *L. sukaczewii* and differed from the most common haplotype by one indel of two bp at 248–49 bp positions in intron 1. The partial sequence of the *PHYO* region analyzed in our study is composed of only one exon. Only one synonymous segregating site (a 'T/C' nucleotide substitution) was found in this region at 367 bp position in

Nucleotide position (bp)	91	92	218	232	341	351	363	405	408	435	498	522	569	576	641	653	682	701	744	848	991	1012	1031	1091	1100	1121	1132	1230	1300	1356	1415
H01	T	G	C	C	C	C	C	T	G	C	G	G	A	C	G	T	C	C	C	C	C	G	C	G	—	C	G	G	A	A	C
H02	A	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	T	·	G	·	·	·	C	·	·	·	·	·	T
H03	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	T	·	G	·	·	·	C	·	·	·	·	·	T
H04	·	·	T	G	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	G	·	·	·	C	·	·	·	·	·	T
H05	·	·	·	G	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	C	·	·	·	·	·	T
H06	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	C	·	·	·	·	·	T
H07	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	T
H08	·	·	·	G	·	·	·	A	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	A	·	·	·
H09	·	·	·	G	·	·	·	·	·	·	C	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	A	·	·	·
H10	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	A	·	·	·	A	·	·	·
H11	·	·	·	G	·	·	·	·	·	·	C	·	·	·	·	·	·	·	·	·	G	·	·	A	·	T	·	A	·	·	T
H12	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	A	·	·	·	A	·	·	·
H13	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	T	A	·	T	G	·	·	A	·	·	·	A	·	·	·
H14	·	C	·	·	T	T	·	·	·	·	·	·	·	·	·	·	T	·	·	·	G	·	·	A	·	·	·	A	·	·	·
H15	·	C	·	·	T	T	·	·	·	·	·	·	·	·	·	·	T	·	·	·	G	·	·	A	·	·	·	A	·	·	T
H16	·	C	·	·	T	T	·	·	·	·	·	·	T	A	·	T	·	·	·	·	G	·	·	A	·	T	·	A	·	G	T
H17	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	A	·	·	·	A	·	·	T
H18	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	A	·	·	·	A	·	·	·
H19	·	·	·	G	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	A	·	·	·	A	·	·	T
H20	·	·	·	G	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	A	·	·	·	A	·	·	·
H21	·	·	·	G	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	A	·	·	T
H22	·	·	·	G	·	T	·	·	·	·	·	·	·	·	C	·	·	·	·	·	G	·	·	·	·	·	·	A	G	·	T
H23	·	·	·	G	·	T	·	·	·	·	·	·	·	·	C	·	·	·	·	·	G	·	·	·	·	·	·	A	·	·	T
H24	·	·	·	G	·	T	·	·	T	·	·	·	G	·	·	·	·	·	·	·	G	·	·	·	·	·	·	A	·	·	T
H25	·	·	·	·	·	T	·	·	·	·	·	T	·	·	·	C	·	·	·	·	G	C	T	·	·	·	T	A	·	·	T

Fig. 1 Summary of segregating sites in the *ITS* region. *H01~H25* represent haplotypes. The *dash* (–) represents an indel

both *L. sukaczewii* and *L. sibirica*. However, this site was ambiguous (showing both 'T' and 'C' nucleotides) in several sequences of both species. This result might have been caused by, e.g., recent duplication of the gene. Because both *CAD* and *PHYO* regions showed very low nucleotide variation, they were excluded from further analyses.

The nucleotide diversity ($\pi_{\text{all sites}}$) in the *ITS* region ranged from 0.0007 (population 3) to 0.0026 (population 9) and the nucleotide polymorphism ($\theta_{\text{all sites}}$) from 0.0007 (population 3) to 0.0026 (population 8). In the *4CL* region, $\pi_{\text{all sites}}$ ranged from 0.0013 (population 6) to 0.0036 (population 9) and $\theta_{\text{all sites}}$ from 0.0014 (population 1) to 0.0037 (population 4). Polymorphisms at non-synonymous, noncoding and synonymous, as well as silent sites (synonymous and noncoding) are shown in Table 2. Values of $\pi_{\text{all sites}}$ and $\theta_{\text{all sites}}$ were, in general, lower in *L. sukaczewii* (*ITS*/*4CL* over all populations: $\pi_{\text{all sites}}$=0.0010/0.0020; $\theta_{\text{all sites}}$=0.0013/0.0026) than in *L. sibirica* (*ITS*/*4CL* over all populations: $\pi_{\text{all sites}}$=0.0026/0.0033; $\theta_{\text{all sites}}$=0.0031/0.0027; Table 2). But, similar levels of $\pi_{\text{all sites}}$ and $\theta_{\text{all sites}}$ were found in comparisons between populations of *L. sukaczewii* from putative refugia (5 and 6) and populations created after deglaciation (1–4; Table 2).

Haplotypes

The constructed haplotype networks, including indels, are shown in Fig. 3a (*ITS*) and b (*4CL*). Twenty-five haplotypes (including indels) that relate to each other in a complex network were found in the *ITS* region (Fig. 3a). Ten haplotypes were found in *L. sukaczewii* and 16 in *L. sibirica* and only one haplotype (H21) was shared by both species (Figs. 3a and 4). Haplotypes H06, H21, and H23 were the most frequent in *L. sukaczewii* while in *L. sibirica*, H14 was the most frequent haplotype (Fig. 3a). Some haplotypes differed from each other by only one mutational step (e.g., H06 and H07 differed only by an indel at 1,200 bp position; H19 and H21 differed by one nucleotide substitution at 1,191 bp position), while others were several mutational steps apart (e.g., H16 and H25, the two most isolated haplotypes; Figs. 1 and 3a). Haplotypes H02 × H07 found in *L. sukaczewii* and haplotypes H14 and H15 found in *L. sibirica* differed by eight or more mutational steps and

Nucleotide position (bp)	145	246	341	348	423	454	505	604	676	752	792
H01	A	A	C	C	G	T	C	C	C	T	T
H02	·	·	·	·	·	C	·	G	·	·	·
H03	·	·	·	·	C	C	·	G	·	·	·
H04	·	·	·	G	·	C	·	G	·	·	·
H05	·	·	·	·	·	C	·	G	·	·	—
H06	·	·	·	·	·	C	T	G	·	·	·
H07	·	·	·	·	·	C	T	G	·	·	—
H08	·	·	·	·	·	C	T	G	T	·	·
H09	T	G	·	·	·	C	·	G	·	·	—
H10	T	G	·	·	·	C	·	G	·	·	·
H11	T	G	·	G	·	C	·	G	·	·	·
H12	T	G	T	·	·	C	·	G	·	·	·
H13	T	G	·	·	·	C	·	G	·	—	·

Fig. 2 Summary of segregating sites in the *4CL* region. The nucleotide 'T' at 244 bp position of the H12 haplotype (*third column*) is a replacement. *Dashes* (–) represent indels

Table 2 Summary of total number of segregating sites (S), nucleotide diversity (π) and nucleotide polymorphism (θ) in the *ITS* and *4CL* regions

	ITS region					*4CL* region										
	Non-coding regions		All sites			Silent		Non-coding regions		Only synonymous		Only non-synonymous		All sites	
Population	S	π	θ	π	θ	S	π	θ	π	θ	π	θ	π	θ	π	θ
L. sukaczewii																
1	5	0.0012	0.0011	0.0011	0.0010	3	0.0047	0.0043	0.0000	0.0000	0.0070	0.0063	0.0000	0.0000	0.0016	0.0014
2	6	0.0014	0.0013	0.0013	0.0012	4	0.0049	0.0059	0.0000	0.0000	0.0073	0.0088	0.0000	0.0000	0.0016	0.0019
3	3	0.0008	0.0007	0.0007	0.0007	5	0.0056	0.0057	0.0000	0.0000	0.0084	0.0084	0.0004	0.0007	0.0021	0.0023
4	5	0.0011	0.0011	0.0010	0.0010	8	0.0086	0.0114	0.0058	0.0044	0.0100	0.0147	0.0000	0.0000	0.0028	0.0037
5	4	0.0010	0.0009	0.0009	0.0008	4	0.0052	0.0044	0.0000	0.0000	0.0076	0.0066	0.0004	0.0007	0.0020	0.0019
6	5	0.0012	0.0012	0.0011	0.0011	3	0.0038	0.0043	0.0000	0.0000	0.0057	0.0063	0.0000	0.0000	0.0013	0.0015
Overall pop.	11	0.0011	0.0015	0.0010	0.0013	9	0.0057	0.0070	0.0013	0.0027	0.0079	0.0090	0.0001	0.0004	0.0020	0.0026
L. sibirica																
7	12	0.0025	0.0026	0.0022	0.0024	7	0.0099	0.0099	0.0025	0.0044	0.0135	0.0126	0.0000	0.0000	0.0032	0.0033
8	13	0.0021	0.0029	0.0019	0.0026	5	0.0076	0.0077	0.0000	0.0000	0.0113	0.0115	0.0000	0.0000	0.0025	0.0025
9	11	0.0029	0.0024	0.0026	0.0022	7	0.0108	0.0099	0.0044	0.0043	0.0140	0.0126	0.0000	0.0000	0.0036	0.0033
Overall pop.	22	0.0028	0.0034	0.0026	0.0031	8	0.0102	0.0083	0.0025	0.0032	0.0140	0.0107	0.0000	0.0000	0.0033	0.0027

Fig. 3 Haplotype networks (unrooted minimum spanning trees): **a** the *ITS* region; and **b** the *4CL* region. *Small gray circles* in (**a**) represent nodes. *All other circles* represent haplotypes. The sizes of circles are proportional to the haplotype frequency. Branch lengths longer than one mutational step are marked with *numbers*

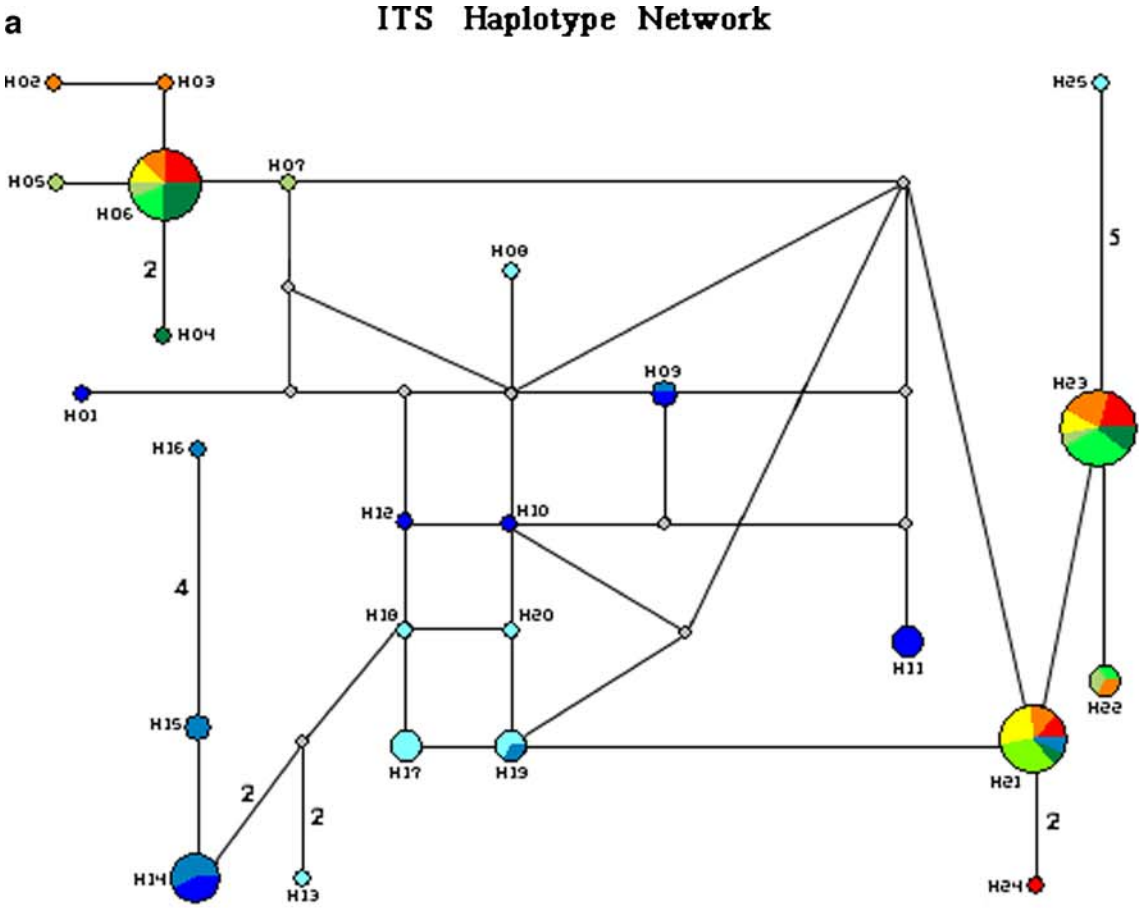
formed the two most distinct groups in the network (Figs. 1 and 3a).

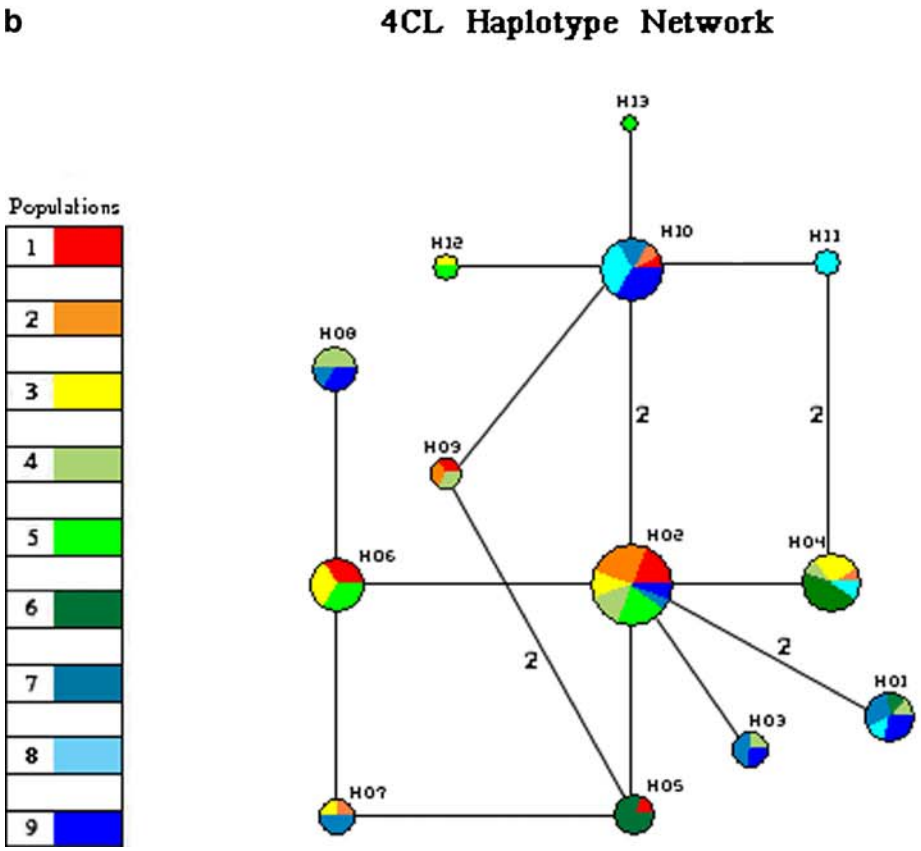
Haplotype network of the *4CL* region (Fig. 3b) was simpler than that of the *ITS* region (Fig. 3a). Thirteen haplotypes (indels included) were found in this region. Five of them were found only in *L. sukaczewii*, and one was found only in *L. sibirica*; seven haplotypes were shared by both species. As in the *ITS* region, some haplotypes differed from each other by only one mutational step (e.g., H06 and H08), others were up to five mutational steps apart from each other (e.g., H01 and H12; Figs. 2 and 3b). Haplotypes H02–H08 and H09–H13 appeared to form two separate groups and the haplotype H01 appeared to be isolated from these two groups. Haplotypes H02, H04, H05, and H06 of the first group were more frequent in *L. sukaczewii* than in *L. sibirica,* while the haplotype H10 of the second group as well as the haplotype H01 were more frequent in *L. sibirica* (Figs. 2 and 3b).

In the *ITS* region, population 6 differed from other populations of *L. sukaczewii* mainly in haplotype frequencies rather than in haplotype composition. Each population of *L. sibirica* appeared to be unique in haplotype composition with only few shared haplotypes observed among populations 7, 8, and 9 (Fig. 4). In the *4CL* region, population 6 of *L. sukaczewii* and population 8 of *L. sibirica* were most distinct. Population 6 shared haplotype H04 with populations 2, 3, and 4 and haplotype H05 with population 1, but frequencies of these haplotypes differed. Populations 5 and 6 did not share any haplotype. Among the four haplotypes observed in population 8, two haplotypes (H04 and H11) were absent in populations 7 and 9. The haplotypes H01 and H10 were shared among all three populations of *L. sibirica*; but, in population 8, their frequencies differed from populations 7 and 9 (Fig. 4). The most marked characteristic of the haplotype pattern of both DNA regions, though, was the apparent distinction between *L. sukaczewii* and *L. sibirica* in both composition and frequencies of haplotypes (Fig. 4).

Genetic differentiation among populations

The F_{ST} values obtained with and without indels were very similar to each other; therefore, only F_{ST} values with indels are presented. In the *ITS* region, the highest values of F_{ST} were found in comparisons between populations of *L. sukaczewii* and *L. sibirica* (Table 3). The range of pairwise F_{ST} values in this region varied from negative (e.g., populations 1–4 of *L. sukaczewii*) to as high as 0.531


a
ITS Haplotype Network
H02
H03
H05
H06
H07
H08
H04
H09
H01
H25
5
H23
H16
H12
H10
4
H18
H20
H11
H22
H15
H17
H19
H21
2
2
2
H14
H13
H24


b
4CL Haplotype Network
Populations
1
2
3
4
5
6
7
8
9
H13
H12
H10
H11
H08
H09
2
2
H06
H02
H04
2
2
H01
H03
H07
H05

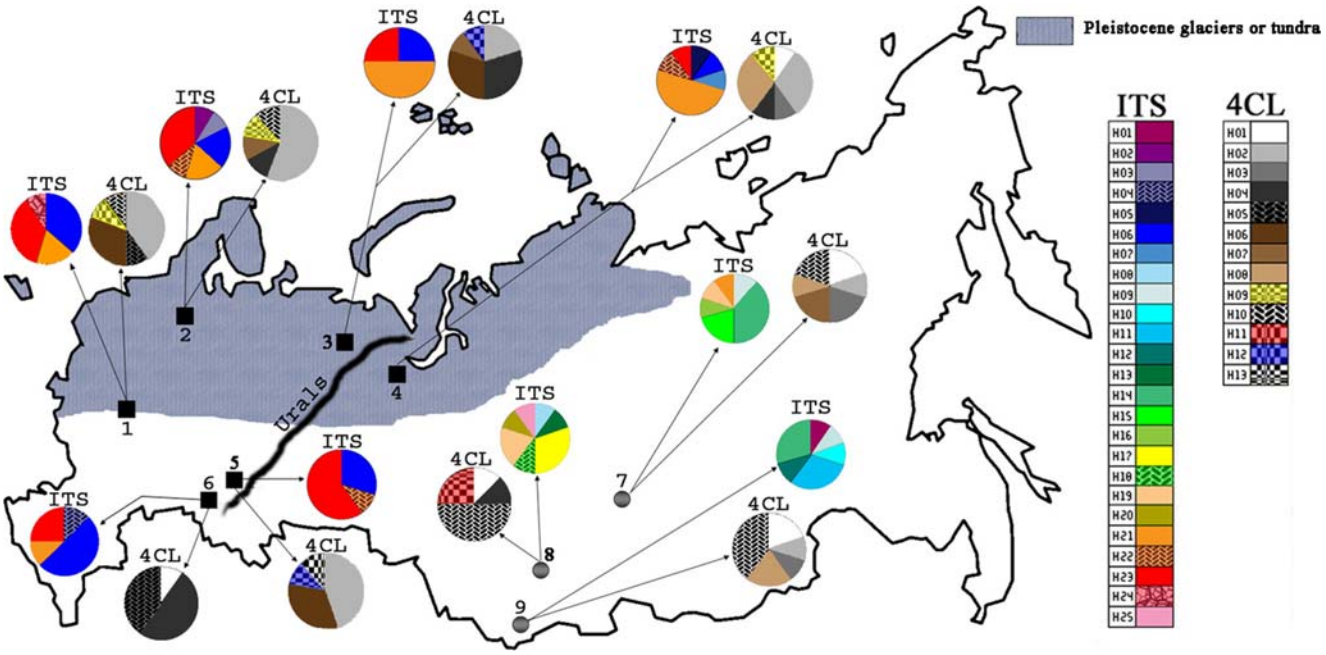


Fig. 4 Distribution of *ITS* and *4CL* haplotypes among populations. *Larix sukaczewii* populations are represented by *black squares* and *L. sibirica* with by *gray circles*

between populations 6 (*L. sukaczewii*) and 7 (*L. sibirica*; Table 3). With few exceptions, moderate to high F_{ST} values were found when population 6 was compared to other populations and all pairwise comparisons among *L. sibirica* populations were moderate to high. All F_{ST} values for pairwise comparisons between *L. sukaczewii* and *L. sibirica* populations were high and statistically significant ($p<0.05$; Table 3).

In the *4CL* region, populations 6 and 8 were the most differentiated. Moderate to high values of F_{ST} were found in pairwise comparisons involving one of these two populations, and the F_{ST} value for the comparison between populations 6 and 8 was the highest (0.407; Table 3). However, low levels of differentiation were found among most *L. sukaczewii* populations (1–5) and similarly low levels were also observed between some *L. sukaczewii* and *L. sibirica* populations when populations 6 and 8 were not involved in comparisons (Table 3).

Tests of neutrality and population size changes

No statistically significant result was obtained in any of the neutrality tests (Tajima's *D*, Fu and Li's *D** and *F**, and HKA), and there was no tendency toward negative or positive values in Tajima's *D* and Fu and Li's *D** and *F** (data not shown). Therefore, no deviations from neutrality were detected.

Using the constant population size and expansion models, we carried out coalescent simulations to find the range of parameters that fit our results. The observed nucleotide diversity per locus (π_T) calculated by multiplying the nucleotide diversity per site (π) by the length of the DNA region without indels and Tajima's *D* for the combined population (1 through 4) were: *ITS* (πT= 1.7415; D=−0.5124) and *4CL* (πT=1.5367; D=−0.8257). Ranges of the results from coalescent simulations for population expansion are shown in S1. Several different combinations among *No*, *T*, and *No*/*Nt* have generated π_T and *D* values that were consistent with empirical results (S1). Therefore, we could not reject any model used in simulations and do not discuss them further.

Discussion

DNA sequences and polymorphism

The *ITS* region is, perhaps, the most commonly used sequence in population genetic and phylogenetic studies (Alvarez and Wendel 2003). However, some authors have argued that for various reasons such as, e.g., the presence of multiple copies, compensatory base changes, and difficulties in alignment, the use of *ITS* for such studies is problematic (Alvarez and Wendel 2003; Bailey et al. 2003; Campbell et al. 2005; Gernandt and Liston 1999). Indeed, the presence of multiple copies of the *ITS* region was reported for some *Larix* species (Gernandt and Liston 1999; Gernandt et al. 2001; Wei et al. 2003; Wei and Wang 2004b). Yet, there is also evidence suggesting that different multiple copies of the *ITS* region were not amplified in our study. If such multiple copies were present in our material, we would expect to observe multiple peaks during

Table 3 Pairwise F_{ST} (with indels) for the *ITS* region (below diagonal) and the *4CL* region (above diagonal)

Pop.	1	2	3	4	5	6	7	8	9
1	–	−0.069	−0.006	−0.010	−0.106	0.238**	−0.029	0.310*	0.014
2	−0.066	–	0.008	0.007	−0.060	0.133	−0.006	0.237	0.006
3	−0.078	−0.048	–	−0.004	−0.026	0.087	0.028	0.328*	0.099
4	−0.054	−0.033	−0.106	–	−0.009	0.123	−0.060	0.306*	−0.005
5	−0.028	−0.050	0.005	0.030	–	0.227*	−0.032	0.277*	0.003
6	−0.001	0.016	0.081	0.075	0.115	–	0.158*	0.407*	0.215**
7	0.491***	0.484***	0.496***	0.484*	0.523**	0.531*	–	0.232	−0.061
8	0.322**	0.320***	0.306*	0.299***	0.362***	0.404**	0.232*	–	0.093
9	0.366***	0.364**	0.377*	0.359***	0.413***	0.389***	0.134	0.135*	–

*p<0.05, **p<0.02, ***p<0.01

sequencing such as those reported by Gernandt et al. (2001). Yet, our *ITS* chromatograms obtained using ABI 3100 automatic sequencer had no ambiguous nucleotide sites. We, therefore, believe that direct sequencing method used in our study detected only one copy of the *ITS* region or multiple copies, which had identical sequence. Based on our data alone, we cannot determine the reason why additional copies of the *ITS* region were not detected in our study. Nevertheless, such selective amplification has been often reported in other studies, and its possible causes have been reviewed by, e.g., Wagner et al. (1994).

In spite of the fact that three copies of the *4CL* region exist in the genus *Larix*, direct sequencing method used by Khatab et al. (2008) detected only the *4CL*-B copy (as determined by comparisons with *4CL* sequences of *Larix* from the GenBank). Therefore, similarly to *ITS*, the *4CL* region was treated as a single locus in the present study.

It is often assumed that long noncoding regions of the DNA harbor more nucleotide variation than shorter coding regions. Although this may be true in most cases, in this study, most segregating sites (eight out of 11) in the *4CL* region were found in the exon 1 (size=654 bp). On the other hand, the *CAD* region was almost monomorphic despite its total size of 1,331 bp, including more than 600 bp of introns. The reasons for such low nucleotide variation in the *CAD* region remain a question for further investigation. The low nucleotide variation in the *CAD* and *PHYO* regions and the ambiguity observed at the only segregating site in the *PHYO* region prevented their utilization in this study.

The levels of π (Table 2) revealed in our study were similar to nucleotide variation reported in other studies on conifers using nuclear gene regions. For instance, values of π were in approximately the same order of magnitude as those reported for *Pinus taeda* (ranges of 19 loci: $\pi_{\text{all sites}}$= 0.00027–0.01728, π_{silent}=0.00042–0.01975; Brown et al. 2004); *P. sylvestris* (PHYP: $\pi_{\text{all sites}}$=0.0010, π_{syn}=0.0020; *PHYO*: $\pi_{\text{all sites}}$=0.0004, π_{syn}=0.0013; Garcia-Gil et al. 2003); *P. sylvestris* (*pal1*: $\pi_{\text{all sites}}$=0.0014, π_{syn}=0.0049; Dvornyk et al. 2002); *P. tabuliformis*, *P. yunnanensis*, *P. densata* (ranges of seven loci: $\pi_{\text{all sites}}$=0.0064–0.0092; π_{silent}=0.0087–0.0128; Ma et al. 2006) and *Cryptomeria japonica* (ranges of seven loci: $\pi_{\text{all sites}}$=0.00004 ×0.00519; π_{silent}=0.00017–0.00813; Kado et al. 2003). Similar levels of nucleotide diversity were observed in the *C3H* nuclear gene region of *L. sukaczewii* ($\pi_{\text{all sites}}$=0.0016) and *L. sibirica* ($\pi_{\text{all sites}}$=0.0020; Khatab et al. 2008).

In overall population comparisons, the values of π were lower in *L. sukaczewii* (*ITS*: $\pi_{\text{all sites}}$=0.0010, $\pi_{\text{noncoding}}$= 0.0011; *4CL*: $\pi_{\text{all sites}}$=0.0020, π_{silent}=0.0057) than in *L. sibirica* (*ITS*: $\pi_{\text{all sites}}$=0.0026, $\pi_{\text{noncoding}}$=0.0028; *4CL*: $\pi_{\text{all sites}}$=0.0033, π_{silent}=0.0102; Table 2). This result was concordant with that reported for the *C3H* region, where the levels of π were also slightly lower in *L. sukaczewii* (Khatab et al. 2008). However, similar levels of variation in nuclear AFLP between these two species were reported (Semerikov and Lascoux 2003). This difference could be due to the different ways AFLP markers and sequencing nuclear gene loci sample the existing genetic variation. Hence, it seems that the available data is still insufficient to make general inferences about the levels of polymorphism in these two species.

Population differentiation

Usually, low levels of genetic differentiation among local populations of conifers are expected because of their outbreeding and wind-pollination behavior (Loveless and Hamrick 1984). In the genus *Larix*, however, high levels of population differentiation could be expected because its pollen does not have air-sacs (Owens et al. 1998) and, thus, cannot disperse for long distances. For example, it has been reported that, under normal conditions, most of *L. laricina* pollen falls less than 50 m away from the parent tree (Hall 1986; Knowles et al. 1992). Seeds are not easily disseminated either, being generally dispersed over distances equivalent to less than two-tree heights (Brown et al. 1988; Duncan 1954; Knowles et al. 1992). Therefore, geographic isolation has been considered as a barrier to

gene flow among *Larix* populations (Lewandowski et al. 1994; Young and Young 1992). Yet, most previous studies on *Larix* revealed low population differentiation (Larionova et al. 2004; Lewandowski 1997; Semerikov and Lascoux 2003; Semerikov and Lascoux 1999; Timerjanov 1997; Wei et al. 2003; Wei and Wang 2004b). Recent divergence of extant populations was suggested as the cause of the low genetic differentiation within and among Eurasian species from the genus *Larix* (Semerikov and Lascoux 2003; Wei et al. 2003; Wei and Wang 2004b). In this study, we found both low and high levels of differentiation among populations. The lack of differentiation among populations 1 through 4 (F_{ST} values were close to zero; Table 3) that occupy previously glaciated areas on the plains of Northwestern Russia is concordant with results from previous studies and is consistent with their recent divergence on geological time scale. No population differentiation was observed either in the *C3H* region for the same populations (Khatab et al. 2008). Moderate to high levels of differentiation among populations (F_{ST}>0.075) were found in many pairwise comparisons involving populations 6, 7, 8, and 9, which (except population 7) occur in, or near, areas of putative refugia (Table 3). These results are consistent with a history of long time isolation of these populations or their respective sources during the Pleistocene. Only few other similar results of moderate to high levels of population differentiation in *Larix* species have been reported. In randomly amplified polymorphic DNA, analyses of *Larix* species (Kozyrenko et al. 2004) found an overall G_{ST}= 0.1864. In a study of *L. sukaczewii* using allozymes, one highly differentiated population from southern Urals (near the location of population 6) was reported, in spite of a low overall population differentiation (F_{ST}=0.028; Timerjanov 1997). The author concluded that this result may be due to isolation of this area from other parts of *L. sukaczewii* distribution during the LGM. An additional reason for the high levels of differentiation among some populations revealed in our study could be the fact that four out of nine of the investigated populations are located in, or near, areas of different and isolated putative Pleistocene refugia, which have rarely been investigated before. Populations from these areas might have evolved independently for a long time with little, if any, gene flow among individual refugia.

Two recent studies on populations of *Larix* species have also revealed moderate to high levels of population differentiation. For instance, in the *C3H* gene region, populations 6 and 8 showed significant levels of population differentiation when compared to other populations of the corresponding species (Khatab et al. 2008). In a study of mtDNA variation (Semerikov et al. 2007), the observed overall F_{ST} of 45.7% was very similar to the levels of population differentiation revealed in our present study (Table 3). The divergent haplotype distribution between *L. sukaczewii* and *L. sibirica*, as well as among *L. sibirica* populations reported by Semerikov et al. (2007) were also very similar to our results, especially those for the *ITS* region (Fig. 4).

Demography

Populations of *L. sukaczewii* were sampled both from areas of recent colonization (1, 2, 3, and 4) and from putative Pleistocene refugia. Thus, they were probably created by migrants coming from southern refugia (likely from areas near the Sea of Azov, etc.) and might have started occupying extant locations around 7,500–8,700 years before present (Kullman 1998). Some haplotypes found in populations 1 through 4 were not found in populations 5 and 6 in both gene regions (*ITS*: H02, H03, H05, H07, and H24; *4CL*: H03, H07, H08, H09, and H10; Figs. 3a,b and 4). The differences in haplotype composition observed between populations of *L. sukaczewii* from refugial areas in the Southwestern Urals (5 and 6) and populations, from Northwestern Russia, which was glaciated during the LGM (1 through 4; Tarasov et al. 2000) suggest that populations 5 and 6 are not the likely sources of postglacial expansion into that region. It is possible that the extant populations in Northwestern Russia have originated from several sources located in other refugial areas that existed during the LGM, such as the surroundings of the Sea of Azov and other locations within the Urals watershed (Tarasov et al. 2000). However, to our knowledge, the part of southern Urals, where populations 5 and 6 are located, is currently the southernmost limit of extant populations of *L. sukaczewii*, and *Larix* species no longer grow in areas farther south and near the Sea of Azov because those areas are now dominated by steppe vegetation or desert. It, thus, appears that some refugial populations, which gave rise to the extant populations in Northwestern Russia, went extinct. The similar levels of π and θ observed in comparisons between populations from putative refugia (5 and 6) and populations from newly colonized areas (1 × 4; Table 2) confirms the findings reported by Khatab et al. (2008) and are also concordant with our suggestion that populations 1, 2, 3, and 4 were created by migrants from different refugia, the admixture effect as proposed by Widmer (2001). That is because populations occurring in refugial areas or created by migrants from different and genetically distinct refugia are expected to harbor higher levels of genetic diversity than those occurring in newly colonized deglaciated areas.

Populations 8 and 9 of *L. sibirica* are in the areas of putative refugia in the South Central Siberia and Altai (Blyakharchuk et al. 2004; Tarasov et al. 2000). On the other hand, there is no information about the presence of *Larix* refugia in the Upper Tunguska region where population 7 is located, near the banks of the Angara River. It is possible that this population was created by

migrants from refugia other than those where populations 8 and 9 occur because in the *ITS* region, population 7 showed moderate to high levels of differentiation in relation to the other two populations (Table 3; Fig. 4). The Angara River, which flows out of Lake Baikal, could have been the main route of colonization of that area, most likely from Northern Mongolia through the surroundings of Lake Baikal. If this scenario is correct, our results for the *ITS* region give support to the results reported by Semerikov et al. (2007), where haplotype frequencies observed in populations of *L. sibirica* from the southern coast of Lake Baikal, also suggested their independent origin from western populations. Population 9 (Altai region) appears to have been created by migrants from nearby refugial areas located at lower altitudes because no forest was present at its current altitude (1,630 m) during Pleistocene glaciation (Blyakharchuk et al. 2004; Tarasov et al. 2000). Finally, unique haplotype composition of population 8 and high F_{ST} values in pairwise comparisons with the other two *L. sibirica* populations (7 and 9; Table 3 and Fig. 4) suggest that despite relative geographic proximity, it has evolved in isolation from populations occurring in other parts of the Siberian Central Plateau.

L. sukaczewii and *L. sibirica*

Following an extensive study, Dylis (1947) found that populations from Western Russia differ from those occurring in Central and Eastern Siberia with respect to a considerable number of characters such as, e.g., cone variability, seeds, shoots, crown shape, stem, and physical and mechanical properties of wood. Based on these results, he proposed to regard populations from Western Siberia as a separate taxon: *L. sukaczewii*. Results from karyotypic analyses (Muratova 1991) gave further support to such classification, and analysis of phylogenetic relationships between *L. sibirica* and *L. sukaczewii* using the chloroplast DNA *trnK* intron sequences (Bashalkhanov et al. 2003) revealed interspecific levels of genetic distances between *L. sibirica* and *L. sukaczewii*. In this study, haplotype composition of the investigated populations showed a conspicuous separation between *L. sibirica* and *L. sukaczewii* in the *ITS* region. Among the 25 *ITS* haplotypes, only one (H21) was shared by both taxa (Figs. 3a and 4) and this haplotype was frequent in *L. sukaczewii*, but it was found in only one individual of *L. sibirica*. Further evidence of the divergence between these two taxa was given by the haplotype network, which showed two groups of haplotypes (H02–H07, *L. sukaczewii* and H13–H16, *L. sibirica*) separated by several mutational steps (Figs. 1 and 3a). In the *4CL* region, haplotypes of the two species were more similar to each other than those observed in the *ITS* region. Seven out of 13 haplotypes were shared, but noticeable differences in haplotype frequencies were also observed when populations of *L. sibirica* were compared to those of *L. sukaczewii*. Some *4CL* haplotypes that were frequent in *L. sibirica* (e.g., H01 and H10) were rarely found in *L. sukaczewii* and vice versa (e.g., H02; Figs. 3b and 4). The high F_{ST} values obtained in most pairwise comparisons in the *ITS* region when populations of *L. sibirica* were compared to populations of *L. sukaczewii* also suggest a considerable divergence between these two taxa (e.g., pop. 6 vs. 7, F_{ST}=0.531; Table 3). Therefore, our results provide partial support for the classification of *L. sibirica* and *L. sukaczewii* as two distinct taxa. However, as phylogeography of *L. sibirica* and *L. sukaczewii* seems to be much more complex than previously suggested, further studies that include *L. sibirica* populations from areas colonized after the LGM are necessary for a better comprehension of the postglacial history of these species.

Acknowledgements We wish to thank Drs. Ove Martinsson JiLU, Bispgården, Sweden and Katsuhiko Takata, Institute of Wood Technology Akita Prefectural University, Japan for providing seed samples. We also wish to thank Dr. Vladimir L. Semerikov, Institute of Plant and Animal Ecology, Ural Division, Russian Academy of Sciences, Yekaterinburg, Russia for help in obtaining literature related to morphological studies on *L. sukaczewii* and *L. sibirica*. This work was partly supported by Grant No. 16-260 from the Sasakawa Scientific Research Grant, The Japan Science Society to NHTA and by the Grants No. 13575002 and 17405032 from the Ministry of Education, Culture, Sports, Science, and Technology of Japan to AES.

References

Abaimov AP, Barzut VM, Berkutenko AN, Buitnk J, Martinsson O, Milyutin LI, Polezhaev A, Putenikhin VP, Katsuhiko T (2002) Seed collection and seed quality of *Larix* spp. from Russia—initial phase of the Russian–Scandinavian Larch project. Eur J For Research 4:39–49

Abaimov AP, Lesinski JA, Martinsson O, Milyutin LI (1998) Variability and ecology of Siberia Larch species. Department of Silviculture, Report No. 43. Swedish University of Agricultural Sciences, Umeå, p 123

Alvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. Mol Phyl Evol 29:417–434

Bailey CD, Carr TG, Harris SA, Hughes CE (2003) Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Mol Phyl Evol 29:435–455

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Bashalkhanov SI, Konstantinov YM, Verbitskii DS, Kobzev VF (2003) Reconstruction of phylogenetic relationships of larch *Larix sukaczewii* Dyl. based on chloroplast DNA trnK intron sequences. Rus J Genet 39:1322–1327

Blyakharchuk TA, Wright HE, Borodavko PS, van der Knaap WO, Ammann B (2004) Late Glacial Holocene vegetational changes on the Ulagan high-mountain plateau, Altai Mountains, southern Siberia. Palaeogeogr Palaeoclimatol Palaeoecol 209:259–279

Brown KR, Zobel D, Zasada J (1988) Seed dispersal, seedling emergence and early survival of *Larix laricina* (DuRoi) K. Koch

in the Tanana Valley, Alaska. Can J Forest Res—Revue Can Rech For 18:306–314
Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. Proc Natl Acad Sci U S A 101:15255–15260
Campbell CS, Wright WA, Cox M, Vining TF, Smoot Major C, Arsenault MP (2005) Nuclear ribosomal DNA internal transcribed spacer 1 (ITS1) in *Picea* (Pinaceae): sequence divergence and structure. Mol Phyl Evol 35:165–185
Duncan D (1954) A study of some of the factors affecting the natural regeneration of tamarack (*Larix laricina*) in Minnesota. Ecology 35:498–521
Dvornyk V, Sirvio A, Mikkonen M, Savolainen O (2002) Low nucleotide diversity at the pal1 locus in the widely distributed *Pinus sylvestris*. Mol Biol Evol 19:179–188
Dylis NV (1947) In: Sibirskaya listvennitsa. Moskovskoye Obshchestvo Ispytatelnej Prirody, Novaya seria. Otdel' botanicheskij, Moscow
Filatov DA (2002) PROSEQ: A software for preparation and evolutionary analysis of DNA sequence data sets. Mol Ecol Notes 2:621–624
Franzke A, Hurka H, Janssen D, Neuffer B, Friesen N, Markov M, Mummenhoff K (2004) Molecular signals for Late Tertiary Early Quaternary range splits of an Eurasian steppe plant: *Clausia aprica* (Brassicaceae). Mol Ecol 13:2789–2795
Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709
Garcia-Gil MR, Mikkonen M, Savolainen O (2003) Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. Mol Ecol 12:1195–1206
Gernandt DS, Liston A (1999) Internal transcribed spacer region evolution in *Larix* and *Pseudotsuga* (Pinaceae). Am J Bot 86:711–723
Gernandt DS, Liston A, Pinero D (2001) Variation in the nrDNA ITS of Pinus Subsection Cembroides: implications for molecular systematic studies of pine species complexes. Mol Phyl Evol 21:449–467
Gros-Louis MC, Bousquet J, Pâques LE, Isabel N (2005) Species-diagnostic markers in *Larix* spp. based on RAM and nuclear, cpDNA, and mtDNA gene sequences, and their phylogenetic implications. Tree Genet Genom 1:50–63
Hall J (1986) Growth and development of larch in Newfoundland. In: Murray MB (ed) Sixth International Workshop on Forest Regeneration: The Yield Advantages of Artificial Regeneration at High Altitudes. USDA Forest Service General Technical Report PNW-194
Hess J, Kadereit JW, Vargas P (2000) The colonization history of Olea europaea L. in Macaronesia based on internal transcribed spacer 1 (ITS-1) sequences, randomly amplified polymorphic DNAs (RAPD), and intersimple sequence repeats (ISSR). Mol Ecol 9:857–868
Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature 405:907–913
Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. Phil Transact Royal Soc London: Biol Sci 359:183–195
Holderegger R, Abbott RJ (2003) Phylogeography of the Arctic-Alpine Saxifraga oppositifolia (Saxifragaceae) and some related taxa based on cpDNA and its sequence variation. Am J Bot 90:931–936
Hudson R (2002) Generating samples under Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337–338
Hudson R, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159
Hudson RR, Coyne JA (2002) Mathematical consequences of the genealogical species concept. Evolution 56:1557–1565
Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. Genetics 132:583–589
Ish-Horowicz D (1989) Isolation of DNA from adult flies. In: Ashburner, M. (ed) *Drosophila*. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbour
Jeandroz S, Roy A, Bousquet J (1997) Phylogeny and phylogeography of the circumpolar genus *Fraxinus* (Oleaceae) based on internal transcribed spacer sequences of nuclear ribosomal DNA. Mol Phyl Evol 7:241–251
Kado T, Yoshimaru H, Tsumura Y, Tachida H (2003) DNA Variation in a Conifer, *Cryptomeria japonica* (Cupressaceae sensu lato). Genetics 164:1547–1559
Khatab IA, Ishiyama H, Inomata N, Wang X-R, Szmidt AE (2008) Phylogeography of Eurasian *Larix* species inferred from nucleotide variation in two nuclear genes. Genes Genet Syst 83: in press
Knowles P, Perry D, Foster H (1992) Spatial genetic structure in two tamarack (*Larix laricina* (Du Roi) K. Koch) populations with differing establishment histories. Evolution 46:572–576
Kozyrenko MM, Artyukova EV, Reunova GD, Levina EA, Zhuravlev YN (2004) Genetic Diversity and Relationships Among Siberian and Far Eastern larches Inferred from RAPD Analysis. Rus J Genet 40:401–409
Kropf M, Kandereit J, Comes H (2003) Differential cycles of range contraction and expansion in European high mountain plants during the late Quaternary: insights from *Pritzelago alpine* (L.) Kuntze (Brassicaceae). Mol Ecol 12:931–949
Kullman L (1998) Paleoecological, biogeographical and paleoclimatological implications of early Holocene immigration of *Larix sibirica* into the Scandes mountains, Sweden. Global Ecol Biogeo Lett 1:5
Larionova AY, Yakhneva NV, Abaimov AP (2004) Genetic diversity and differentiation of Gmelin larch *Larix gmelinii* populations from Evkenia (Central Siberia). Rus J Genet 40:1127–1133
Lewandowski A (1997) Genetic relationships between European and Siberian larch, *Larix* spp. (Pinaceae), studied by allozymes. Is the Polish larch a hybrid between these two species. Plant Syst Evol 204:65–73
Lewandowski A, Nikkanen T, Burczyk J (1994) Production of hybrid seed in a seed orchard of two Larch Species, *Larix sibirica* and *Larix decidua*. Scand J For Res 9:214–217
Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. Annu Rev Ecol Syst 15:65–95
Ma XF, Szmidt AE, Wang XR (2006) Genetic structure and evolutionary history of a diploid hybrid pine *Pinus densata* inferred from the nucleotide variation at seven gene loci. Mol Biol Evol 23:807–816
MacKay JJ, Liu W, Whetten R, Sederoff RR, O'Malley DM (1995) Genetic analysis of cinnamyl alcohol dehydrogenase in loblolly pine: single gene inheritance, molecular characterization and evolution. Mol Genet Genom 247:537–545
Milyutin L, Vishnevetskaia K (1995) Larch and Larch Forest in Siberia. In: Schmidt WC, McDonald KJ (eds) Ecology and Management of *Larix* forests: a look ahead. U.S.D.A. Forest Service Intermountain Research Station GTR-INT-319, Whitefish, pp 19–29
Muratova E (1991) Karyotypic analysis of *Larix sibirica* (Pinaceae) from various parts of the species area. Bot Zh 76:1586–1595
Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
Owens J, Takaso T, Runions C (1998) Pollination in conifers. Trends Plant Sci 3:479–485

Reeder J, Höchsmann M, Rehmsmeier M, Voss B, Giegerich R (2006) Beyond Mfold: recent advances in RNA bioinformatics. J Biotec 124:41–55

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols. Meth Mol Biol 132:365–386

Schubert R, Sperisen C, Müller-Starck G, Lascala S, Ernst D, Sandermann Jr H, Hager KP (1998) The cinnamyl alcohol dehydrogenase gene structure in *Picea abies* (L.) Karst.: genomic sequences, Southern hybridization, genetic analysis and phylogenetic relationships. Trees - Struct Funct 12:453–463

Semerikov V, Lascoux M (2003) Nuclear and cytoplasmic variation within and between Eurasian *Larix* (Pinaceae) species. Am J Bot 90:1113–1123

Semerikov VL, Iroshnikov AI, Lascoux M (2007) Mitochondrial DNA variation pattern and postglacial history of the Siberian Larch (*Larix sibirica* Ledeb.). Rus J Ecol 38:163–171

Semerikov VL, Lascoux M (1999) Genetic relationship among Eurasian and American *Larix* species based on allozymes. Heredity 83:62–70

Semerikov VL, Semerikov LF, Lascoux M (1999) Intra- and interspecific allozyme variability in Eurasian *Larix* Mill. species. Heredity 82:193–204

Sharrock RA, Quail PH (1989) Novel phytochrome sequences in *Arabidopsis thaliana*: structure, evolution, and differential expression of a plant regulatory photoreceptor family. Genes Develop 3: 1745–1757

Simurda M, Knox J (2000) ITS sequence evidence for the disjunct distribution between Virginia and Missouri of the narrow endemic *Helenium virginicum* (Asteraceae). J Torrey Bot Soc 127:316–323

Svendsen JI, Astakhov VI, Bolshiyanov DY, Demidov I, Downdeswell JA, Gataullin V, Hjort C, Hubberten HW, Larsen E, Mangerud J, Melles M, Möller P, Saarnisto M, Siegert MJ (1999) Maximum extent of the Eurasian ice sheets in the Barents and Kara Sea region during the Weichselian. Boreas 28:134–242

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Tarasov PE, Volkova VS, Webb TW III, Guiot J, Andreev AA, Bezusko LG, Bezusko TV, Bykova GV, Dorofeyuk NI, Kvavadze EV, Osipova IM, Panova NK, Sevastyanov DV (2000) Last glacial maximum biomes reconstructed from pollen and plant macrofossil data from northern Eurasia. J Biogeo 27:609–620

Thomson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequences alignment aided by quality analysis tools. Nucl Acids Res 25:4876–4882

Timerjanov AS (1997) Lack of allozyme variation in *Larix sukaczewii* Dyl. from the southern Urals. Silvae Genet 46:61–64

Wagner A, Blackstone N, Cartwright P, Dick M, Misof B, Snow P, Wagner GP, Bartels J, Murtha M, Pendleton J (1994) Surveys of gene families using polymerase chain reaction - PCR selection and PCR drift. Syst Biol 43:250–261

Wang XQ, Tank DC, Sang T (2000) Phylogeny and divergence times in Pinaceae: Evidence from three genomes. Mol Biol Evol 17:773–781

Watterson G (1975) On the number of segregating sites in genetical models without recombination. Theor Pop Biol 7:256–276

Wei H-X, Wang X-Q, Hong D-Y (2003) Marked intergenomic heterogeneity and geographical differentiation of nrDNA ITS in *Larix potaninii* (Pinaceae). J Mol Evol 57:623–635

Wei XX, Wang XQ (2003) Phylogenetic split of *Larix*: evidence from paternally inherited cpDNA trnT-trnF region. Plant Syst Evol 239:67–77

Wei XX, Wang XQ (2004a) Evolution of 4-coumarate: coenzyme A ligase (4CL) gene and divergence of *Larix* (Pinaceae). Mol Phyl Evol 31:542–553

Wei XX, Wang XQ (2004b) Recolonization and radiation in *Larix* (Pinaceae): evidence from nuclear ribosomal DNA paralogues. Mol Ecol 13:3115–3123

Whetten R, Sederoff R (1995) Lignin biosynthesis. Plant Cell 7:1001–1013

Widmer A (2001) Glacial refugia: sanctuaries for allelic richness, but not for gene diversity. Trends Ecol Evol 16:267–269

Young J, Young C (1992) *Larix* Mill., larch. In: Dudky T (ed) Seeds of woody plants in North America. Dioscorides, Portland