MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 4958–4971 doi: 10.1111/mec.12433

Southern montane populations did not contribute to the recolonization of West Siberian Plain by Siberian larch (*Larix sibirica*): a range-wide analysis of cytoplasmic markers

VLADIMIR L. SEMERIKOV,* SVETLANA A. SEMERIKOVA,* MARIA A. POLEZHAEVA,* PAVEL A. KOSINTSEV* and MARTIN LASCOUX†

*Institute of Plant and Animal Ecology, Ural Branch of Russian Academy of Science, 8 Marta Str., 202, 620144 Ekaterinburg, Russia, †Department of Ecology & Genetics, EBC, Uppsala University, Norbyvägen 18D, 75236 Uppsala, Sweden

Abstract

While many species were confined to southern latitudes during the last glaciations, there has lately been mounting evidence that some of the most cold-tolerant species were actually able to survive close to the ice sheets. The contribution of these higher latitude outposts to the main recolonization thrust remains, however, untested. In the present study, we use the first range-wide survey of genetic diversity at cytoplasmic markers in Siberian larch (*Larix sibirica*; four mitochondrial (mt) DNA loci and five chloroplast (cp) DNA SSR loci) to (i) assess the relative contributions of southern and central areas to the current *L. sibirica* distribution range; and (ii) date the last major population expansion in both *L. sibirica* and adjacent *Larix* species. The geographic distribution of cpDNA variation was uninformative, but that of mitotypes clearly indicates that the southernmost populations, located in Mongolia and the Tien-Shan and Sayan Mountain ranges, had a very limited contribution to the current populations of the central and northern parts of the range. It also suggests that the contribution of the high latitude cryptic refugia was geographically limited and that most of the current West Siberian Plain larch populations likely originated in the foothills of the Sayan Mountains. Interestingly, the main population expansion detected through Approximate Bayesian Computation (ABC) in all four larch species investigated here pre-dates the LGM, with a mode in a range of 220 000–1 340 000 years BP. Hence, *L. sibirica*, like other major conifer species of the boreal forest, was strongly affected by climatic events pre-dating the Last Glacial Maximum.

Keywords: biogeography, cpDNA, glacial refugia, *Larix*, mtDNA, West Siberia

Received 18 December 2012; revision received 26 June 2013; accepted 29 June 2013

Introduction

Both palaeoecological and genetic data show that glacial cycles of the Pleistocene (2.6 Myr–11 kyr) have profoundly affected the ranges of animal and tree species of the Northern Hemisphere and thereby the level and geographical distribution of genetic diversity (Hewitt 2000). However, the data accumulated over the last decade have also suggested a large diversity of Quaternary histories among species (Petit *et al.* 2003; Lascoux *et al.* 2004). While some species, such as oaks, were confined to southern refugia during cold periods, others were able to survive much closer to the ice sheets (Huntley & Birks 1983; Willis *et al.* 2000). For instance, palaeoecological data indicate that boreal tree species survived at high latitudes during the Last Glacial Maximum (LGM, 18–20 ky before present) both in Europe (Willis & van

Correspondence: Vladimir L. Semerikov, Fax: 7(343) 260 65 00; E-mail: semerikov@ipae.uran.ru

Andel 2004) and in Siberia (Brubaker *et al.* 2005; Lapteva & Korona 2008; Binney *et al.* 2009). However, it remains unclear whether the presence of macrofossils of forest trees at high latitudes implies the existence of refugia that contributed to postglacial recolonization and, if they did contribute what the extent of this contribution was. Many of the species that are associated with the taiga forest, for example, wood lemming (*Myopus schisticolor*; Fedorov *et al.* 2008), tundra shrew (*Sorex tundrensis*; Hope *et al.* 2011), flying squirrel (*Pteromys volans*; Oshida *et al.* 2005), Siberian fir (*Abies sibirica*; Semerikova & Semerikov 2006), Dahurian larch (*Larix gmelinii*; Polezhaeva *et al.* 2010) seem to have survived the adverse climatic phases of the Pleistocene in isolated refugia. When the locations of these refugia could be inferred (Semerikova & Semerikov 2006; Polezhaeva *et al.* 2010), they were not limited to southern regions, but could also be found at intermediate latitudes.

The Siberian larch (*Larix sibirica*) is extremely resistant to cold and drought and consequently has an unusually large range size, extending from the Polar tree line at a latitude of approximately 70ºN to the desert regions of northwestern China, that is, at a latitude of 43ºN and from the Ob River in the west to the Lake Baikal in the east. In the west, *L. sibirica* is replaced by *Larix sukaczewii* Dyl. (Dylis 1947; Semerikov *et al.* 1999, 2003), whereas in the east, it forms a vast zone of introgressive hybridization with *L. gmelinii* (Dylis 1947; Kruklis & Milyutin 1977). The area today inhabited by larch has been only intermittently glaciated, and the glaciated areas were circumscribed to the north. The Pleistocene palaeogeography of West Siberia is determined by the alternating periods of formation and disappearance of an ice sheet in the north. The West Siberian Plain is tilted to the north, and rivers flow north into the Arctic Ocean basin leading to the formation of large ice-dammed lakes when the northern ice sheet blocked the water flow. The depth and size of these lakes varied through time as the ice sheet increased or receded (Arkhipov *et al.* 1970; Volkov 1980). In contrast to continuous ice sheets, the ice-dammed lakes did not prevent the survival of trees because the most elevated areas were not flooded. The glacier reached its maximal size during the Middle Pleistocene, approximately between 270 and 250 kyr ago (MIS8, Volkova *et al.* 2002). At that time, its southern border was a little north of 60°N. Another major glaciation occurred in the Late Pleistocene, between about 190 and 130 kyr ago (MIS6, Volkova *et al.* 2002). Its southern border in Western Siberia corresponded approximately to latitude 62–63°N. After this period, the presence of glaciers in Western Siberia was limited to a short period around 90 to 70 kyr ago (MIS5b-MIS4, Svendsen *et al.* 2004; Astakhov 2012), and their southern border was close to the Arctic Circle (Svendsen *et al.* 2004; Astakhov 2009). In summary, Western Siberia was never completely glaciated, and larch could have persisted during the whole Pleistocene in Western and in Middle Siberia south of 60°N and settled areas close to the Arctic Circle as early as 70 000 years ago. Did larch seize upon this opportunity? A recent palaeoecological study suggests that during the Late Pleistocene, Western Siberia was mostly a cold desert with relicts of arboreal vegetation in large river valleys (Velichko *et al.* 2011). This result suggests that vegetation was poor. However, large animal bones (mammoth, rhinoceros, horse, reindeer, bison, etc.) are found across Western Siberia during all the Late Pleistocene without any marked interruptions indicating that, if Western Siberia had indeed been a cold desert, the vegetated areas, including forested ones, could have been rather extensive. Further evidence of the presence of trees and shrubs in Western Siberia during the Pleistocene is provided by the analysis of woolly mammoths and woolly rhinos diet. Macrofossils of *Larix, Salix, Betula* sect. *Nanae* and *Alnus* were found in all intestines of adult animals analysed so far (Ukraintseva 1993; Tomskaya 2000; Boeskorov *et al.* 2011; Kosintsev *et al.* 2012). Additionally, mammoths and rhinos lived across the entire Western Siberia during the Late Pleistocene (Kuzmin & Orlova 2004; Kosintsev 2008; Kosintsev & Vasiliev 2009). It, therefore, seems that trees and shrubs were also spread throughout Western Siberia during this period even though open landscapes might still have occupied a more important area than forests as typical forest animals like moose were very rare during the Pleistocene and only found in the south.

The present study is based on an extensive genetic survey encompassing the whole range of *L. sibirica* as well as parts of the ranges of *L. sukaczewii* and *L. gmelinii*. All individuals were genotyped at both chloroplastic and mitochondrial markers. In conifers, mitochondrial DNA is maternally inherited and dispersed only through seeds while chloroplast (cp) DNA is paternally inherited and dispersed through both seeds and pollen (Neale & Sederoff 1989). We used these data to address key questions on the quaternary history of *L. sibirica*: (i) Did southern populations contribute to the recolonization of the vast West Siberian Plain after the LGM or do genetic data provide support for a significant contribution from high latitude refugia?; (ii) Do the genetic patterns in Siberian *Larix* species provide evidence of population expansion and, if so, when did this expansion start? *L. sibirica* is a particularly good model species to address these questions as there is ample evidence of survival at high latitudes during the Late Glacial and the LGM (Binney *et al.* 2009).

Materials and methods

Plant materials and laboratory methods

Needle samples were collected in 116 larch populations in Russia, Mongolia, Kazakhstan and NorthWest China (Table S1, Supporting information, Fig. 1). 1595 individuals in 116 populations were genotyped with mtDNA markers, and 1455 of them in 112 populations were also genotyped with cpDNA markers. A limited part of the material has been analysed earlier with mitochondrial DNA markers (Semerikov *et al.* 2007). Genomic DNA was extracted from dried needles with the CTAB method (Devey *et al.* 1996). PCR conditions, electrophoresis, silver staining and scoring of fragments for both types of markers are described in Polezhaeva *et al.* (2010).

To assess cpDNA variation, we used five cpSSR loci (Pt26081, Pt30204, trnLV, Pt9393 and Pt9833) that were initially developed based on the cp genome of *Pinus thunbergii* (Vendramin *et al.* 1996). They were subsequently optimized for *Larix* species from Eastern Siberia and the Far East where they were found to be informative (Polezhaeva *et al.* 2010).

Three mtDNA fragments [*nad4(3c-4r)* (Dumolin-Lapégue *et al.* 1997), *UBC460* and *R11* (Semerikov *et al.* 2006)] were previously shown to be variable in *L. sibirica* or in *L. sibirica* × *L. gmelinii* hybrid populations (Semerikov *et al.* 2007; Polezhaeva *et al.* 2010). The *Atp-A* gene and its flanking areas were polymorphic in larch populations of Eastern Siberia (Polezhaeva *et al.* 2010). The PCR primers for *Atp-A* fragment are as follows: 5′-TCTCGTGCTAGTTCGTGGAA-3′ and 5′-CAAACCGGGATGTACTGCTC-3′. In *L. sibirica*, the *Atp-A* gene has polymorphic nucleotides at positions 486 (T/G) and 694 (A/C) (GenBank Accession nos JQ 411195-JQ 411201). These polymorphisms correspond to *Mse*I restriction sites. One more polymorphism in the *Atp-A* gene corresponds to an indel variation. The polymorphism in *nad4(3c-4r)* corresponds to *Hinf*I site. These polymorphisms were therefore scored via PCR-RFLP following (Polezhaeva *et al.* 2010). The R11 locus carries a minisatellite with a 31-bp motif and was scored according to Semerikov *et al.* (2007).

Genetic diversity

PCR band patterns from mt and cp loci, respectively, were concatenated, resulting in two haplotypes per

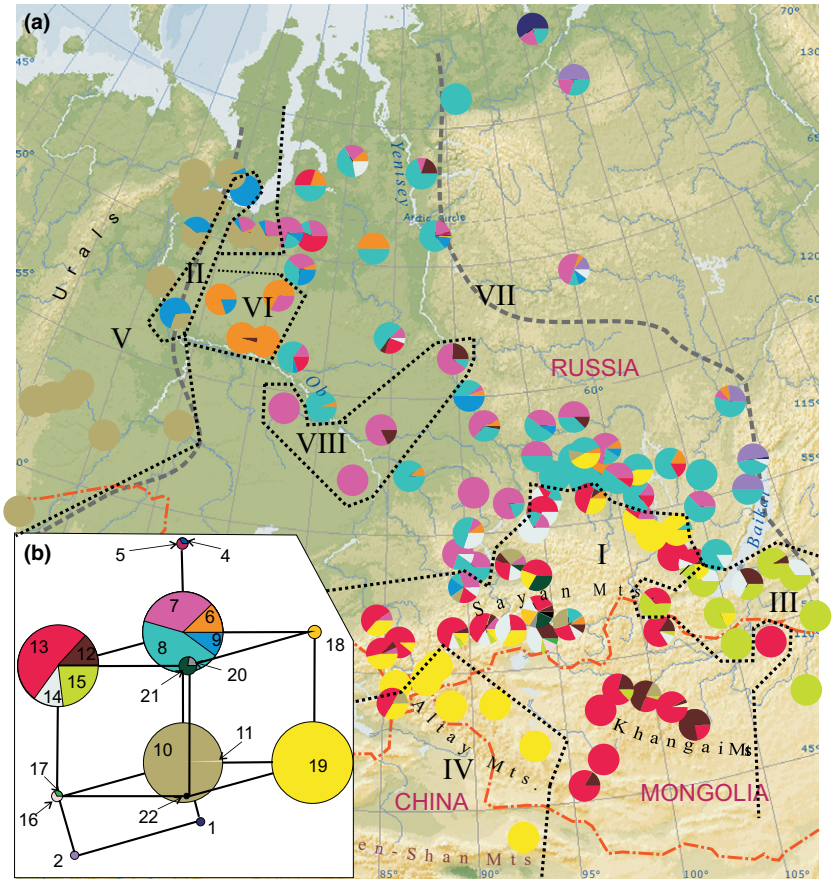

Fig. 1 (a) Geographic distribution of sampled populations of *L. sibirica*. The mitotype frequencies in each population are represented with pie diagrams. SAMOVA population groups (I–VIII) are delineated by dashed line. Approximate borders between *Larix* species are marked with thick dashed line. (b) Mitotype network. The network was constructed ignoring the information from the *R11* locus and pie diagrams correspond to the mitotype frequencies in the total sample.

individual. Diversity measures were calculated within each population for both types of markers: the total number of haplotypes (N_o), the effective number of haplotypes (N_e) and the unbiased haplotype diversity (H_e) (Nei 1987), using the following formulas: $N_e = \frac{1}{\sum_{i=1}^{n} p^2}$ and $H_e = \frac{n}{n-1}(1 - \sum_{i=1}^{k} p_i^2)$, where n is the size of the population, k is the number of haplotypes and p_i is the frequency of the ith haplotype in the population.

Population structure and the presence of southern refugia

The fixation index G_{ST} (Nei 1987) was computed for both kind of markers. We also calculated R_{ST} (Goldstein *et al.* 1995) and N_{ST} for cpSSRs and mtDNA markers, respectively. G_{ST} is based on the allele frequencies only while R_{ST} and N_{ST} take into account the genetic distances between haplotypes. N_{ST} counts the number of mutations and R_{ST} uses the genetic distance D^2_{sh} that was calculated as:

$$D^2_{sh}(i,j) = K^{-1}[\Sigma|a_{ik} - a_{jk}|]^2,$$

where a_{ik} and a_{jk} are the lengths the ith and jth alleles for of kth SSR locus, and K is the total number of loci (Goldstein *et al.* 1995). The comparison of G_{ST} and N_{ST} or R_{ST} was conducted with PermutCpSSR v.1.0 (www.Pierroton.intra.fr/genetics/labo/Software/) (Pons & Petit 1996). If N_{ST} (or R_{ST}) > G_{ST}, different but genetically close haplotypes tend to co-occur, indicating that haplotypes originate in the same populations where they were detected.

A median-joining network (Bandelt *et al.* 1999) relating the mitotypes was constructed using NETWORK 4.6.1.0. (available at http://www.fluxus-engineering.com; Fig. 1). The spatial genetic structure of all populations was analysed by spatial analysis of molecular variance (SAMOVA; Dupanloup *et al.* 2002). This method aims to identify groups of populations that are maximally differentiated from each other. A simulated annealing process is run until the configuration of K groups with the largest F_{CT} (the proportion of total genetic variance due to differences among groups of populations) is obtained, and single-population groups are absent. The analysis was done for K varying from 2 to 12. The hierarchical partitioning of genetic diversity among groups of populations identified with SAMOVA, populations and individuals was carried out with the AMOVA module of the program ARLEQUIN 3.01 (Excoffier *et al.* 2006). There were eight groups in the case of mtDNA and a single group in the case of cpDNA.

Isolation by distance for both types of markers was tested by Mantel test (Mantel 1967) using the NTSYS-pc v. 2.02h software (Rohlf 1988) with 1000 random permutations of the relationship between the matrix of pairwise statistics N_{ST} or R_{ST} (for mtDNA and cpDNA, respectively) and that of geographical distances.

Demographic inferences: population growth or constant population size?

Demographic inferences were based on cpSSR data. The decision to use cpSSR rather than mtDNA PCR-RFLP polymorphism was based on two grounds. First, several recent publications (e.g. Du *et al.* 2011) showed that cpDNA is more efficient than mtDNA to delineate species. Second, SSR markers were easier to use in the Approximate Bayesian Computation (ABC) analysis. We used two types of analyses to identify the signal of past demographic events. First, Fu's (1997) neutrality test was used to detect population growth. Following Navascués *et al.* (2006), the cpSSR data were coded in a binary way: each locus consists of n_{max} characters: n '1' followed by $n_{max}-n$ '0', where n is the number of repeats for a given haplotype at a given locus, and n_{max} is the maximum number of repeats at this locus. ARLEQUIN 3.01 (Excoffier *et al.* 2006) was used to test the significance of Fu's F_S. Owing to the particular behaviour of the F_S-statistic, a test with $P < 0.02$ was considered as evidence for population expansion at the significance level of $\alpha = 0.05$ (Fu 1997). Second, we applied the ABC method implemented in DIYABC v. 1.0.4.36 (Cornuet *et al.* 2010) to the cpSSR markers. Chloroplast SSR loci were regarded as equivalent to Y-linked markers, and the GSM mutation model was used. Four larch taxa were considered: *L. sibirica* itself (848 individuals) and *L. sukaczewii* (103 individuals) in the west and *L. gmelinii* (145 individuals) and *L. cajanderi* (386 individuals) in the east (Polezhaeva *et al.* 2010). Admixed populations located near the border between any two taxa were not included in the combined data set used in the ABC analysis. Hybrid populations were determined based on morphology (Dylis 1947; Dylis 1961; Kruklis & Milyutin 1977) and on presence of mitotypes and/or chlorotypes specific of the sister species of *L. sibirica*. In ABC, a large number of coalescent genealogies are generated conditional on a demographic model (Fig. 2a) that is defined by a set of parameters. For instance, in a standard neutral model, the only parameters are the effective population size N_e and the mutation rate, μ, while model with a discrete change of effective population size, will have four parameters: the current effective population size N_0; the time at which population size changed, t; the population size before this time, N_A; and the mutation rate, μ. The values of the population sizes, the mutation rates and the divergence time are randomly drawn from prior distributions.

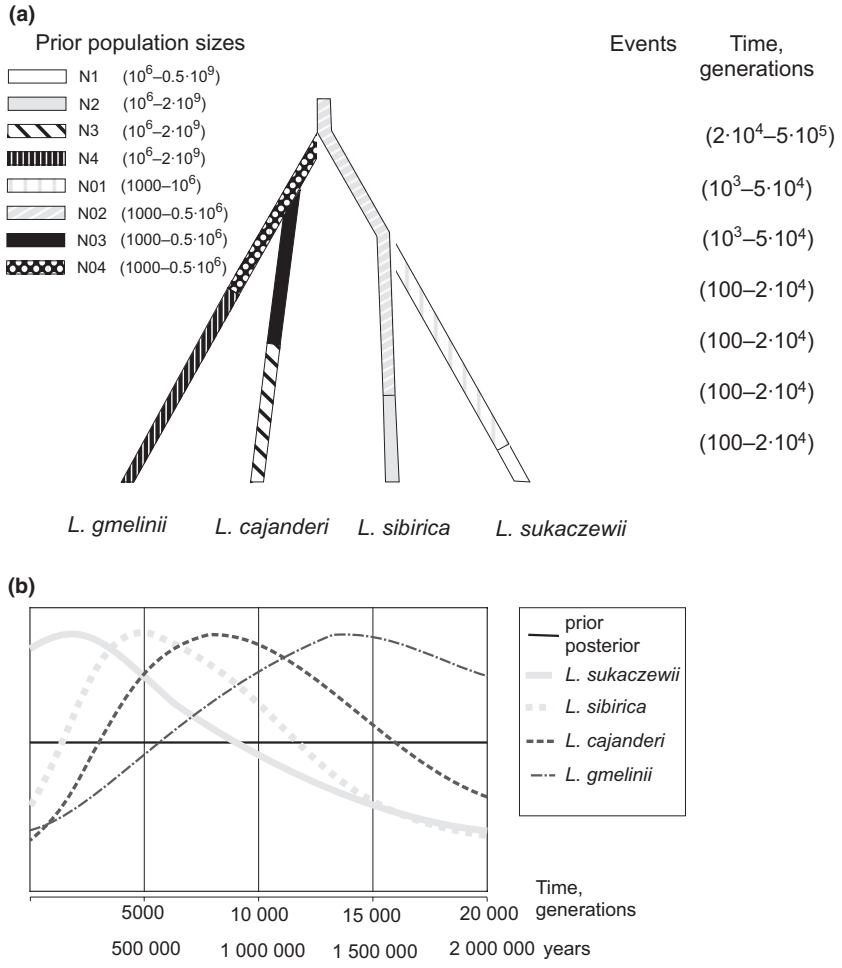

Fig. 2 (a) Graphical representation of the scenario used for ABC simulations (see text for details). The common ancestor of the four species split into two populations (*L. sibirica* and *L. gmelinii*) t7 generations ago and after that they split again at t6 and t5 into four populations (*L. sukaczewii*, *L. sibirica*, *L. cajanderi* and *L. gmelinii*). Each of these population changes size at times t1, t2, t3 and t4, respectively. Priors are uniformly distributed within the indicated ranges. (b) Posterior distributions of the time of population expansion in *L. sukaczewii*, *L. sibirica*, *L. cajanderi* and *L. gmelinii* estimated using Approximate Bayesian Computation.

After that a set of summary statistics is computed for each simulation and compared with observed values. We used all the summary statistics available in the DIYABC for four 'populations' (taxa in our case). The parameter values of the 1% of the simulations yielding the summary statistics closest to the summary statistics of the observed data are then retained for the computation of the posterior distributions of the model parameters using a local linear regression. We used the following model parameters: the present effective population sizes, N1, N2, N3 and N4, of the four taxa *L. sukaczewii*, *L. sibirica*, *L. cajanderi* and *L. gmelinii*, respectively, each with priors uniformly distributed in an interval: 10^6–0.5×10^9 N1 and 10^6–2×10^9 the rest. At time *t*1, *t*2, *t*3 and *t*4 (100–2×10^4 generations going back in time), the population size changed abruptly, taking the values N01, N02, N03 and N04 in a range of 10^3–10^6 individuals. The lower limit of *t*1-*t*4 was set at 100 generations, given that during the Holocene (the last 10 000 years), no significant reduction in population size occurred, and assuming a generation time of approximately 100 years (Provan *et al.* 1999). Estimating generation time remains difficult in species with long reproductive spans like forest trees (Petit & Hampe 2006). Larch trees can start reproducing when they are 20–30 years old, but the trees acquire their full reproductive potential later on and can still contribute to reproduction after up to 200 years of age. We chose 100 years as an average age, but this estimate should be taken with a grain of salt. The upper limit was set at the beginning of the Pleistocene (2 Mya), as the Pleistocene is characterized by climatic instability and, consequently, boreal species could be expected to have experienced large population changes during this period. *L. sukaczewii* merged with *L. sibirica* at time *t*5 (10^3–5×10^4 generations). The lower bound was set to 10^3

generations (10^5 years) because *L. sibirica* was observed in the Urals that is within the geographic range of *L. sukaczewii* at least since the last interglacial, that is, about 100 000 years ago (Stefanovskii 2006). The upper bound was set at the beginning of the Pliocene because the oldest larch macrofossils similar to *L. sibirica* were dated to the Pliocene (Bobrov 1972). *L. gmelinii* merged with *L. cajanderi* at time *t*6 (10^3–5×10^4 generations). We do not have a solid justification for this interval except for the analogy with *L. sukaczewii*-*L. sibirica* that are also closely related sister taxa. At time *t*7 ($2–50 \times 10^4$), the ancestors of *L. sibirica* and *L. gmelinii* merged. These limits were taken because phylogenies indicate that *L. sibirica* cp DNA is the most basal among the larch species (Semerikov *et al.* 2003). So, given that the *Larix* origin was dated to the Eocene (33–55 My, Wang *et al.* 2000), we put the upper limit to 50×10^4 generations and the lower limit to the Pleistocene-Pliocene border (2 My) because a large number of *Larix* macrofossils discovered in Europe and Asia were dated to the Pliocene, precluding post-Pliocene divergence of *L. sibirica* cpDNA from other larches. The mutation rate was in the range of 10^{-7}– 2×10^{-5} per generation. These limits were chosen so that in their vicinity, the posterior value was close to zero. For the chosen scenario, parameter values drawn from the posterior parameter distributions were used to generate one thousand simulated data sets. Summary statistics were computed for each of these predicted data sets, and their distributions were used to evaluate the goodness of fit of the model. That was done in the following way. We first calculated the PCA of summary statistics obtained from (i) 1000 simulated data sets based on parameter values drawn from the prior distributions; (ii) 100 simulated data sets based on parameters estimates from the posterior distributions; and (iii) the observed data. We then checked the position of the summary statistics of the observed data set relative to distribution of the summary statistics of the data sets simulated from the posteriors.

Results

Diversity and population structure

MtDNA. Six polymorphisms were found at the four mtDNA loci: one restriction site in *nad4/3-4*, one indel in *UBC460*, two restriction sites and one indel in *Atp-A* and one minisatellite in *R11*. The latter had 28 alleles. In total, 64 mitotypes were observed. The relationships between them are detailed in Table 1. MtDNA diversity varies strongly among populations and was highly structured (Table 2). The most diverse populations are located in the central part of the *L. sibirica* range, and peripheral populations are generally less diverse (Table S1, Supporting information, Fig. 1). The westernmost populations (2–12) corresponding to *L. sukaczewii* and the populations from the Altai and Tien-Shan (95–101) were all monomorphic. With an estimated value of 0.617 (Table 3), N_{ST} was significantly higher than G_{ST} (0.564) suggesting the presence of a phylogeographic structure.

Substitution rates in plant mitochondrial DNA are remarkably low (Wolfe *et al.* 1987), but the mutation rate in minisatellites could be exceptionally high (Rogstad *et al.* 2003). Because of this strong difference among markers, it does not seem correct to combine in the same analysis of mtDNA molecular diversity the point mutations and indel variation with the minisatellite variation of the R11 fragment. On the other hand, the R11 fragment behaves as a minisatellite only if the repeat number exceeds 1. If there is only one repeat, then its mutation rate should be similar to that of the other mtDNA loci. The SAMOVA of mtDNA variation was therefore conducted in two different ways: in the first case, we discarded the information of the minisatellite repeat number, and in the second, we used it. In the former, we combined the alleles with repeat number higher than one into one allele giving four possible alleles: (i) no amplification; (ii) no repeat motif within the PCR fragment; (iii) the motif is present in single copy; and (iv) the motif is repeated at least twice. This reduced the mitotype number to 21 (Table 1 and Fig. 1). Based on the geographical distribution of the resulting mitotypes and their network relationships, four major groups can be distinguished: *L. sukaczewii* west of the Ob River (V), populations from the Central West Siberian Plains (VII), populations around and west of the Baikal Lake (III, I), and finally populations from the Altai and Tien-Shan mountains (IV) (Fig. 1). In addition to these four major groups, the SAMOVA analysis also identified minor group located near the border between *L. sukaczewii* and *L. sibirica* (II) and two smaller subgroups within the West Siberian Plain (VI, VIII).

The analysis using all minisatellite alleles gave similar results to the 4-allele analysis if the number of groups identified by SAMOVA is below six. When the number of population groups was six or higher, the algorithm produced additional single-population groups and did not recognize groups I, II, VII and VIII (data not shown). This large number of single-population clusters is likely due to the presence of numerous rare haplotypes and is not very informative on ancient demographic events. Therefore, we chose to only present results based on the restricted number of alleles. The mitotype spatial distributions are presented separately for each mtDNA polymorphic nucleotide, indels and for the minisatellite in Fig. S1a–e (Supporting information).

Finally, to evaluate the phylogenetic network of the mitotypes, we omitted the minisatellite locus and used

Table 1 Description of the mitotypes based on the polymorphism detected at four mtDNA loci

	Loci†					
Mitotype	*nad4(3c-4r)*, pos. 1433	*Atp-A*, pos. 486	*Atp-A*, pos. 694	*Atp-A*, 5 bp. deletion	*UBC460*, 266 bp. deletion	*R11* minisatellite, repeats number
1	C	G	A	–	–	–*
2	C	G	C	–	–	–
3	A	G	A	–	–	–
4	A	G	A	+	+	>1
5	A	G	A	+	+	1
6	A	G	A	–	+	–
7	A	G	A	–	+	>1
8	A	G	A	–	+	1
9	A	G	A	–	+	0
10	A	G	A	–	–	–
11	A	G	A	–	–	>1
12	A	G	C	–	+	–
13	A	G	C	–	+	>1
14	A	G	C	–	+	1
15	A	G	C	–	+	0
16	A	G	C	–	–	–
17	A	G	C	–	–	1
18	A	T	A	–	+	1
20	A	T	C	–	+	–
21	A	T	C	–	+	1
22	A	T	C	–	–	–

†C/A, G/T, A/C-nucleotide at the variable site, *–/+ absence/presence of deletion in fragments *UBC460* or *Atp-A* or amplification in *R11,* number-number of minisatellite repeats in *R11* fragment.

Table 2 Analysis of molecular variance (AMOVA) in mtDNA and cpDNA in studied populations of *L. sibirica, L. sukaczewii* and *L. sibirica* × *L. gmelinii* hybrids

Marker	Source of variation	d.f.	SS	Variance components	Percentage of variation	*F*-statistics
mtDNA	Among 8 groups defined by SAMOVA	7	329.144	0.25834	51.85	F_{CT} = 0.51852**
	Among populations within groups	108	111.834	0.06389	12.82	F_{SC} = 0.26635**
	Within populations	1479	260.296	0.17599	35.32	F_{ST} = 0.64676**
	Total	1594	701.274	0.49823		
cpDNA	Among 112 populations	111	1172.251	0.71735	36.14	
	Within populations	1343	1702.315	1.26755	63.86	F_{ST} = 0.36141**
	Total	1454	2874.565	1.98490		

**$P < 0.01$.
d.f., degrees of freedom; SS, sum of squares.

Table 3 Average genetic diversity within populations (H_S), total genetic diversity (H_T) and fixation indices for mtDNA (G_{ST} and N_{ST}) and cpDNA (G_{ST} and R_{ST})

Marker	No. of populations	H_S (SE)	H_T (SE)	G_{ST} (SE)	N_{ST} or R_{ST} (SE)
mtDNA	116	0.386 (0.0283)	0.886 (0.0071)	0.564 (0.0318)	0.617 (0.0350)*
cpDNA	112	0.918 (0.0076)	0.960 (0.0038)	0.044 (0.0077)	0.232 (0.0614)***
Excluding *L. gmelinii* hybrids	106	0.917 (0.0079)	0.957 (0.0039)	0.041 (0.0081)	0.047 (0.0226)

*N_{ST} or R_{ST} is significantly different from G_{ST} ($P < 0.01$), ***$P < 0.001$.
SE, standard error; NC, not calculated because of low variation among populations.

only the polymorphic nucleotides and indels (Fig. 1). Plant mtDNA can recombine with low frequency due to crossing of individuals bearing different mitotypes and paternal leakage of mtDNA (Jaramillo-Correa & Bousquet 2005; Wang *et al.* 2011). The four gametes test of Hudson & Kaplan (1985) supports the presence of recombination in *L. sibirica* mtDNA. For instance, all four possible allele combinations between the two polymorphic nucleotides within *Atp-A* fragment are observed in larch populations: GA, GC, TA and TC.

cpDNA. Combination of the five cpSSR loci gives 205 cpDNA haplotypes (chlorotypes). We did not sequence the cpSSR fragments, and therefore, we scored alleles according to their relative sizes (Table S1, Supporting information). The haplotype diversity, *H*, was high exceeding 0.90 in most populations. None of the cpSSR diversity parameters were correlated with latitude except $\overline{D^2_{sh}}$ that was positively correlated (Spearman rank correlation $P = 0.001$). This correlation is partly explained by the high value of this parameter in the *L. sukaczewii* populations (Fig. S2, Supporting information). However, the correlation was still significant ($P = 0.0033$) when *L. sukaczewii* and the putative hybrid *L. sukaczewii* x *L. sibirica* populations were excluded. The correlation is explained by lower $\overline{D^2_{sh}}$ values in the southern mountain populations. The R_{ST} estimate greatly exceeded G_{ST} (0.232 vs. 0.044). However, R_{ST} was indistinguishable from G_{ST} once the six populations located in the *L. sibirica – L. gmelinii* contact zone (populations 28, 29, 30, 31, 82, 83, Table S1, Supporting information), and containing *L. sibirica* and *L. gmelinii* chlorotypes, were removed (0.047 vs. 0.041, Table 3). The lack of meaningful population structure was confirmed by the SAMOVA analysis. For any K > 1, a large group was formed that held the majority of the populations, together with several groups, each containing a single population. These outliers were generally the populations located within the zone of contact with *L. gmelinii*, or they corresponded to populations at the periphery of the range, for example, Polar tree line and Tien-Shan (data not shown). Nonetheless, between-population pairwise R_{ST} and geographic distances were highly correlated (Mantel test Z statistics was 0.14396, *P*[random Z > = observed Z] = 0.0080), and variation at cpDNA was useful to distinguish species, with haplotypes specific to the contact zones with *L. sukaczewii* and *L. gmelinii* (Figs S3 and S4, Supporting information).

Demographic inferences based on cpSSR data

F_s was significantly negative in most populations suggesting population expansion (Table S1, Supporting information). The ABC analysis also supports a population expansion in all four species (Table 4). Assessment of the goodness of fit of the demographic model via comparison of the summary statistics of the real cpDNA data set with the simulated data sets indicates that the real data set lies fairly close to the data sets simulated from the posterior distributions of model parameters (Fig. S5, Table S2, Supporting information). The modes of the posterior of the population expansion time lie between 2000 and 13 000 generations (Fig. 2b). Broad confidence intervals cover several glacials and interglacials. The upper 95% confidence limit of this parameter in *L. sukaczewii* is 440 generations, that is, within the early stage of the late glacial if one assumes a generation time of 100 years (Zyryanka glaciation, 10 000–120 000 years ago), but in *L. sibirica*, *L. cajanderi* and *L. gmelinii*, it corresponds to 1370, 1900 and 1690 generations, respectively, well predating the Zyryanka glaciation. The divergence time estimates between *L. sibirica* and *L. sukaczewii* and between *L. cajanderi* and *L. gmelinii* were 31 800 and 28 100 generations (3.2 and 2.8 Ma), respectively.

Discussion

The aim of the present study was to address two main questions: (i) Did southern populations contribute to the recolonization of the vast West Siberian Plains after the LGM or do genetic data provide support for a significant contribution of high latitude refugia?; and (ii) Does the distribution of polymorphism in *L. sibirica* species offer evidence of population expansions and if so, when did those expansions start? Below we will discuss these two questions in turn, as well as the phylogeography of populations from the southern Mountain ranges and Mongolia and the presence of introgression at the margins of the species.

Recolonization: not from the mountains but primarily from the south with a limited contribution from high altitude refugia?

The distribution of mitochondrial variation across the whole range of *L. sibirica* range unambiguously shows that populations from the southern mountain ranges (Tien-Shan and Sayan ranges), and Mongolia had a very limited contribution, if any, to the recolonization of the West Siberian Plain after the LGM. Indeed, the set of haplotypes found in these geographical regions are virtually absent elsewhere. While this allows us to rule out these mountain ranges as major glacial refugia for the current larch population of Western Siberia, it does not preclude the presence of refugial populations in the northern foothills of the Sayan mountains as the haplotypes that are today frequent there are also present across the whole range of *L. sibirica*. This would be in

Table 4 Priors and posteriors for the parameters of the demographic inference performed with the Approximate Bayesian Computation (ABC) approach of DIYABC for *L. sukaczewii, L. sibirica, L. cajanderi* and *L. gmelinii* (see Materials and methods for details)

		Posterior distribution		
Parameter	Priors†	Mode	95% confidence limits	
N1	(10^6–0.5×10^9)	6.71×10^7	1.32×10^7	4.88×10^8
N2	(10^6–2×10^9)	1.66×10^9	5.91×10^7	1.95×10^9
N3	(10^6–2×10^9)	1.68×10^9	5.37×10^7	1.95×10^9
N4	(10^6–2×10^9)	1.81×10^9	5.91×10^7	1.95×10^9
t1	(100–20 000)	2.22×10^3	4.45×10^2	1.87×10^4
N01	(10^3–10^6)	3.38×10^5	5.44×10^4	9.72×10^5
t2	(100–20 000)	4.17×10^3	1.37×10^3	1.85×10^4
N02	(10^3–0.5×10^6)	2.99×10^5	8.83×10^4	4.83×10^5
t3	(100–20 000)	7.66×10^3	1.90×10^3	1.90×10^4
N03	(10^3–0.5×10^6)	1.49×10^5	4.03×10^4	4.53×10^5
t4	(100–20 000)	1.34×10^4	1.69×10^3	1.95×10^4
N04	(10^3–0.5×10^6)	9.53×10^4	2.07×10^4	4.76×10^5
t5	(10^3–5×10^4)	3.58×10^4	7.32×10^3	4.86×10^4
t6	(10^3–5×10^4)	2.94×10^4	6.34×10^3	4.80×10^4
t7	(2×10^4–5×10^5)	3.36×10^5	1.12×10^5	4.90×10^5
μ	(1×10^{-7}–2×10^{-5})	2.00×10^{-6}	5.00×10^{-7}	1.30×10^{-5}

†N1 - N04 presented in units of individuals number, *t*1–*t*7 in number of generations and μ in mutations number per generation.

line with the main paradigm that dominated much of phylogeography until recently, whereby southern refugia were the main contributors to the recolonization of formerly glaciated or inhospitable areas (Hewitt 2000). While this paradigm seems to hold true for a large array of temperate species, it may be more questionable for cold-adapted ones. In the case of forest trees, it has been questioned on the basis of (i) the absence of a phylogeographical pattern suggestive of southern refugia in species such as willows or birches (Lascoux *et al.* 2004); and (ii) the presence of macrofossils at higher latitudes (Willis *et al.* 2000; Binney *et al.* 2009; Välirantä *et al.* 2011; Parducci *et al.* 2012). Even if some authors suggested that scattered late-glacial tree populations acted as dispersal nuclei for forest development (e.g. Välirantä *et al.* 2011; Parducci *et al.* 2012), the existence of some of those high-latitude cryptic populations remain controversial (Birks *et al.* 2012), and, more importantly, the balance between their contribution and the contribution of 'classical' southern refugia to modern populations is still an open and difficult question.

In the case of *L. sibirica*, one can distinguish a couple of limited, genetically distinct, geographic groups of populations within the West Siberian Plain, for instance groups II, VI and VIII of the SAMOVA analysis (Fig. 1). This pattern could indicate the survival of *L. sibirica* in isolated refugia that would have acted as dispersal nuclei subsequently, as suggested by Välirantä *et al.* (2011). If so, our data would indicate a rather local contribution of cryptic refugia located along the Ob River to modern populations, the largest part of the latter still originating from more southern refugia in the foothills of the Sayan Mountains. The strong homogeneity of the West Siberian Plains would also suggest that recolonization was rapid, an observation that is in agreement with data in other tree species that also show marked homogeneity over large geographic regions [*Picea obovata* (J. Chen, Y. Tsuda, M. Stocks, T. Källman, G. G. Vendramin, V. L. Semerikov & M. Lascoux, unpublished), *A. sibirica* (Semerikova & Semerikov 2006), *L. gmelinii* (Polezhaeva *et al.* 2010)]. For example, *Picea obovata* population located from 55°N to 60°N along the Yenisei river were genotyped with 15 SSR and 349 SNP, and in both cases showed no population structure at all (J. Chen, Y. Tsuda, M. Stocks, T. Källman, G. G. Vendramin, V. L. Semerikov & M. Lascoux, in prep.). Interestingly, this is in strong contrast to the situation in Scandinavia (Tollefsrud *et al.* 2009; Chen *et al.* 2012), where population genetic structure is important and reflects a complex immigration history after the LGM.

What does the fossil record suggest? In comparison with Western Europe, where detailed fossil pollen maps are available, our knowledge of the palaeoecology of Western Siberia remains incomplete. A recent palaeoecological study suggested that the West Siberian Plain was a cold desert with relicts of arboreal vegetation in large river valleys (Velichko *et al.* 2011). As pointed out in the introduction, findings of large animal macrofossils across West Siberia during the Late Pleistocene suggest that, if West Siberia was a cold desert, the areas of arboreal vegetation could still have been rather extensive. The large diversity found in both cpDNA and

mtDNA, as well as the presence of clusters within the area would tend to support a presence more important than a few scattered cryptic populations. In any case, even if the extent of the forested area is still uncertain, the trees were present as indicated by findings of macrofossils (Binney *et al.* 2009). Furthermore, we recently extracted ancient DNA from (i) larch macrofossils collected in the south of the Yamal peninsula (latitude 67–68ºN) and dated between 2000 and 7000 years (Hantemirov & Shiyatov 2002); and (ii) a fragment of larch branch with a radiocarbon years age of 16 000 extracted from the intestine of a mammoth excavated in the north of the Gydan peninsula (72ºN, Lapteva & Korona 2008; V. L. Semerikov & M. A. Polezhaeva, unpublished). The sample from the mammoth intestine had the mitochondrial DNA haplotype characteristic of Siberian larches now living in most of West Siberia, while the samples from the southern Yamal peninsula carried both this haplotype and the haplotype characteristic of *L. sukaczewii*. The latter indicates that admixed *L. sibirica x L. sukacewii* populations already existed in southern Yamal (Fig. 1) in the early Holocene while the former further confirms the presence of larch at high latitudes. These data do not prove a major role of northern refugia in the recolonization, but suggest that these putative high-latitude refugia were genetically very similar to the modern larch populations of northern Siberia.

A rather common presence across the range is also in line with the fact that the ABC analysis based on cpDNA does not indicate any population expansion corresponding to the LGM in *L. sibirica* and closely related species. The posterior distributions of the age of population expansion in all larch species investigated here had a mode in a range of 220 000 – 1 340 000 years BP (95%CI = 44 000 – 1 900 000 years BP) suggesting a postbottleneck population recovery following some mid-Pleistocene glaciation well before the LGM. Of course, one has to consider the possibility that the signal of the LGM was not detected by ABC as it was obscured by the stronger signal of a more ancient population expansion.

A similar pattern was observed in many taiga species. For example, wood lemmings (*M. schisticolor*), a small rodent associated with the taiga forest, a population expansion occurred about 125 000 years ago (Fedorov *et al.* 2008), that is, soon after the Saalian (glacial period 130 000–190 000 years ago, Svendsen *et al.* 2004), which corresponds to a period where the extent of the ice sheet in Siberia was larger than in most recent glacial periods. Similarly, the time of the population expansion in Western Siberia also preceded the LGM in tundra shrew (Hope *et al.* 2011). In contrast, though, brown bear experienced a population bottleneck during the LGM (Korsten *et al.* 2009). However, it should be mentioned that an exceptionally high mtDNA mutation rate was estimated within the clade of brown bears by Korsten *et al.* (2009) that was not supported by more extensive studies on bear mtDNA genomes (see, for example Lindqvist *et al.* 2010). Therefore, it is quite possible that the bottleneck in brown bear also pre-dated the LGM. It should also be noted that the algorithm used in DIYABC, does not include any migration among species, and therefore that all time estimates are likely to be underestimates. The estimations of the divergence time between *L. sibirica* and *L. sukaczewii* (31 800 generations ago) or the divergence between *L gmelinii* and *L. cajanderi* (28 100 generations) would correspond to the Pliocene (Table 4). It is difficult to verify the age of the separation of Siberian larch species because of the dearth of larch macrofossils from the Pliocene, and the rare finds of pollen fossil (Nikitin 1957) do not allow species identification. It is, however, important to keep in mind that many of these studies, might lack power to detect recent population size changes. In humans, for instance, recent exponential population growth is only captured with large genomic data and very large sample size (Keinan & Clark 2012; Nelson *et al.* 2012). The signal used to detect recent population growth is an excess of rare variants compared with the standard neutral expectations, which is only captured with such large samples. In the same vein, the ancient bottlenecks generally observed in tree species in western Europe, on a timescale ranging from a few hundred of thousand years to a few million years (Heuertz *et al.* 2006; Pyhäjärvi *et al.* 2007; Ingvarsson 2008) might as well reflect the limited sample sizes used in these studies rather than the absence of population changes after those events.

Phylogeography of larches from Southern Mountains and Mongolia

The marked differences among the three SAMOVA clusters observed in the southern mountain ranges, with an almost complete mitotype fixation in the Altay Mountains, contrasting with a high mitotype diversity in the Sayan and Khangai Mountains of Mongolia are suggestive of different demographic histories that would warrant further study, in particular demographic inferences based on multilocus nuclear data. In the absence of those, inferences on past history can only be very speculative, but the available genetic data, together with palaeoecological data can still help to suggest plausible scenarios. Haplotypes found in Mongolia belong to cluster I and are well separated from haplotypes in the Altay Mountains (cluster IV) and those in the East (cluster III). Mongolia and northwestern China are a series of mountain ranges separated by dry

steppes and deserts. In contrast to Siberia, in this part of the range, *L. sibirica* usually forms pure, small stands, located on the northern slopes of high mountains that are isolated from each other by large expanses of alpine and steppe landscapes absolutely insurmountable for tree species (Dylis 1947). Most likely the present distribution of larch in this area is the result of past expansions, which took place during more humid periods favourable for forest growth. In Central Asia, the global warming and cooling cycles that determined the alternation of glacial and interglacial phases at higher latitudes corresponded to a succession of arid and relatively pluvial periods (Dodonov 2002). Existing palaeoecological data from Central Asia indicate a shift from a cold and arid Late Pleistocene towards a warmer and pluvial Holocene optimum 10–4 ky BP ago (Felauer *et al.* 2012). Findings of fir, spruce, Siberian stone pine and larch macrofossils, as well as palynological data indicate that the expansion of taiga vegetation in today's treeless Govi Altai (eastern edge of the Altai mountains) occurred in the middle of the Holocene (Dorofeyuk & Tarasov 2000). However, an earlier migration cannot be excluded as there is evidence that the Late Pleistocene period (13000–10 000 years ago) in the Semipalatinsk region (southeast Kazakhstan) was characterized by temperatures and precipitations close to the levels currently favourable to tree species (Kremenetski *et al.* 1997).

Introgression

Our data confirmed the asymmetric pattern of the introgression of mitochondrial DNA and cp DNA in the contact zone of *L. sibirica* and *L. gmelinii* that was previously observed (Semerikov & Lascoux 2003; Semerikov *et al.* 2007; Katyshev *et al.* 2009): the populations neighbouring pure *L. sibirica* stands, carried only *L. sibirica* mitotypes together with a mixture of chlorotypes from both species. Easternmost populations that are far from the pure *L. sibirica* stands, exclusively carry chlorotypes specific to *L. gmelinii* in combination with a mixture of *L. sibirica* and *L. gmelinii* mitotypes. The reason of such an asymmetry is a difference in geneflow level between cpDNA and mtDNA that are transmitted via pollen and seeds, respectively. The asymmetric introgression is also associated with the natural invasion of *L. gmelinii* into the range of *L. sibirica*. Indeed, many foresters and botanists noted that *L. gmelinii* is better adapted to the present climate of East Siberia than *L. sibirica* (Dylis 1947). *L. gmelinii* can survive across a broader array of habitats and *L. gmelinii* grows faster and can displace *L. sibirica* over large areas. Traces of a wider past distribution of *L. sibirica* in East Siberia are found in the form of morphological features specific to *L. sibirica* in populations of *L. gmelinii* located many hundreds of kilometres away from the nearest populations of *L. sibirica* (Dylis 1947). Currat *et al.* (2008) also showed through simulations that when two species hybridize, the introgression goes primarily from the local species to the invading one. In line with this explanation, *L. gmelinii*, being the invading species displacing *L. sibirica*, captures the *L. sibirica* mtDNA, but does not capture the cpDNA.

Besides *L. gmelinii*, *L. sibirica* also has a common border with *L. sukaczewii* in the west. The populations bearing the mixture of mitotypes of the two taxa are spread along the Ob river from the mouth of Irtish River in the south up to the Yamal peninsula in the north. In the northern part of this area, the distance between the westernmost and easternmost admixed populations (pop. 15 and 16, Table S1, Supporting information) is around 380 km. Unfortunately, our cpSSR data cannot distinguish *L. sukaczewii* and *L. sibirica*, and therefore are not informative on the width and limits of the *L. sibirica - L. sukaczewii* contact zone for plastid genome.

Conclusion

To the best of our knowledge, our study is the most extensive survey of genetic variation for a Siberian tree species. The *L. sibirica* natural range can be subdivided into three main domains: the southern mountains and Mongolia, the central Western Siberian Plains and the margins. The southern mountains and Mongolia are themselves composed of subdomains corresponding to the different mountain ranges. The history of *Larix* in these areas is still poorly understood and a large-scale multilocus survey with nuclear markers would certainly unravel further subdivisions and help reconstruct the history of this area and provide a link to the larch species from the Qinghai Tibetan Plateau and the Himalayas. One of the main findings of the present study is that the contribution of this first domain to the recolonization of the West Siberian Plains was limited. The West Siberian Plain was therefore recolonized by populations in the northern foothills of the Sajan mountain range or from cryptic refugia along the major rivers that have played such an important part in shaping the history of the region. Similarly, multilocus surveys would be required to apportion the relative contributions of the different areas. The fact that population genetic structure is limited over very large swaths of lands make this area uniquely interesting for association studies or studies of clinal variation at candidate genes and ideal to replicate studies done in western European counterparts whose history, and resulting population genetic structure, are vastly more complex. Such a strategy has been used successfully in spruce species (J. Chen, Y. Tsuda, M. Stocks,

T. Källman, G. G. Vendramin, V. L. Semerikov & M. Lascoux, in prep.), but it would be interesting to implement it in larches too. Finally, we have large hybrid zones at the Eastern and Western borders of *L. sibirica*. These hybrid zones could also be useful for association studies.

Acknowledgements

We are grateful to A. Iroshnikov, Yu. Isakov, R. Hantemirov, L. Agafonov and S. Shijatov for help with material collections. This work was financially supported by the Russian Basic Research Foundation, grant nos. 10-0496000 and 11-0491396. Martin Lascoux would like to acknowledge support from the EU project Noveltree and the BioDiversa projects Linktree and TipTree.

References

Arkhipov SA, Vdovin VV, Mizerov BV, Nikolaev VA (1970) *Western Siberian Plain*. Nauka, Moscow. 280 p. (in Russian)

Astakhov VI (2009) The middle and late Neopleistocene of The Glacial Zone of Western Siberia: problems of stratigraphy and paleogeography. *Bulletin of the commission for the study of the Quaternary Period*, **69**, 8–24. (in Russian).

Astakhov VI (2012) Pleistocene glaciations of northern Russia – a modern view. *Boreas*, doi:10.1111/j.1502-3885.2012.00269.x.

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Binney AH, Willis KJ, Edwards ME *et al.* (2009) The distribution of late-Quaternary woody taxa in northern Eurasia: evidence from a new macrofossil database. *Quaternary Science Reviews*, **28**, 2445–2464.

Birks HH, Giesecke T, Hewitt GM, Tzedakis PC, Bakke J, Birks HJB (2012) Comment on "Glacial Survival of Boreal Trees in Northern Scandinavia". *Science*, **338**, 742.

Bobrov EG (1972) *History and Systematics of Larix*. Nauka, Leningrad (in Russian).

Boeskorov GG, Lazarev PA, Sher AV *et al.* (2011) Woolly rhino discovery in the lower Kolyma River. *Quaternary Science Reviews*, **30**, 2262–2272.

Brubaker LB, Anderson PM, Edwards ME, Lozhkin AV (2005) Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapped pollen data. *Journal of Biogeography*, **32**, 833–848.

Chen J, Källman T, Ma XF *et al.* (2012) Disentangling the roles of history and local selection in shaping clinal variation in allele frequency and gene expression in Norway spruce (*Picea abies*). *Genetics*, **191**, 865–881.

Cornuet JM, Ravigne V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, **11**, 401.

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution*, **62**, 1908–1920.

Devey ME, Bell JS, Smith DN, Neale DB, Moran GF (1996) A genetic linkage map for *Pinus radiate* based on RFLP, RAPD and microsatellite markers. *Theoretical and Applied Genetics*, **92**, 673–679.

Dodonov AE (2002) *Quaternary of Middle Asia: Stratigraphy, Correlation, Paleogeography*. Transactions of GIN RAS, V. 546. Geos, Moscow, Russia. 250 p. (in Russian)

Dorofeyuk NI, Tarasov PE (2000) Vegetation of western and southern Mongolia in Late Pleistocene and Holocene. *Botanicheskii Zhurnal*, **85**, 1–17. (in Russian).

Du FK, Peng XL, Liu JQ, Lascoux M, Hu FS, Petit RJ (2011) Direction and extent of organelle DNA introgression between two spruce species in the Qinghai-Tibetan Plateau. *New Phytologist*, **192**, 1024–1033.

Dumolin-Lapégue S, Pemonge MH, Petit RJ (1997) An enlarged set of consensus primers for the study of organelle DNA in plants. *Molecular Ecology*, **6**, 393–397.

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

Dylis NW (1947) *Siberian Larch*. Published by the Moscow Society of Naturalists, Moscow (in Russian).

Dylis NW (1961) *Larch of Eastern Siberia and Far-East*. Academy of Sciences of USSR, Moscow, Russia.

Excoffier L, Laval G, Schneider S (2006) ARLEQUIN ver. 3.1: An Integrated Software Package for Population Genetics Data Analysis. Computational and Molecular Population Genetics Lab (CMPG). Institute of Zoology, University of Bern, Bern, Switzerland.

Fedorov VB, Goropashnaya AV, Boeskorov GG, Cook JA (2008) Comparative phylogeography and demographic history of the wood lemming (*Myopus schisticolor*): implications for late Quaternary history of the taiga species in Eurasia. *Molecular Ecology*, **17**, 598–610.

Felauer T, Schlütz F, Murad W, Mischke S, Lehmkuhl F (2012) Late Quaternary climate and landscape evolution in arid Central Asia: a multiproxy study of lake archive Bayan Tohomin Nuur, Gobi desert, southern Mongolia. *Journal of Asian Earth Sciences*, **48**, 125–135.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139**, 463–471.

Hantemirov RM, Shiyatov SG (2002) A continuous multimillennial ring-width chronology in Yamal, northwestern Siberia. *Holocene*, **12**, 717–726.

Heuertz M, De Paoli E, Källman T *et al.* (2006) Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics*, **174**, 2095–2105.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hope AG, Waltari E, Fedorov VB, Goropashnaya AV, Talbot SL, Cook JA (2011) Persistence and diversification of the Holarctic shrew, *Sorex tundrensis* (Family Soricidae), in response to climate change. *Molecular Ecology*, **20**, 4346–4370.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequence. *Genetics*, **111**, 147–164.

Huntley B, Birks HJB (1983) *An Atlas of Past and Present Pollen Maps for Europe: 0–13000 Years Ago*. Cambridge University Press, Cambridge, UK.

Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics*, **180**, 329–340.

Jaramillo-Correa JP, Bousquet J (2005) Mitochondrial genome recombination in the zone of contact between two hybridizing conifers. *Genetics*, **171**, 1951–1962.

Katyshev AI, Semerikov VL, Konstantinov YM (2009) *L. sibirica - L. gmelinii* hybridization in the middle taiga and sub-taiga zone revealed by genetic markers. In: *International Conference ''Conservation of Forest Genetic Resources of Siberia''*, 3–9 August, Novosibirsk, Russia (in Russian).

Keinan A, Clark AG (2012) Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science*, **336**, 740–743.

Korsten M, Ho SYW, Davison J *et al.* (2009) Sudden expansion of a single brown bear maternal lineage across northern continental Eurasia after the last ice age: a general demographic model for mammals? *Molecular Ecology*, **18**, 1963–1979.

Kosintsev PA (2008) *Mammoth Fauna of the Yuribey river Basin (Yamal Peninsula)*. Faunae and florae of Northern Eurasia in the Late Cenozoic, Ekaterinburg-Chelyabinsk. pp. 147–157. (in Russian).

Kosintsev PA, Vasiliev SK (2009) The large mammal fauna in the Neopleistocene of Western Siberia. *Bulletin of the commission for the study of the Quaternary Period*, **69**, 94–105. (in Russian).

Kosintsev PA, Lapteva EG, Korona OM, Zanina OG (2012) Living environments and diet of the Mongochen mammoth, Gydan Peninsula, Russia. *Quaternary International*, **276–277**, 253–268.

Kremenetski CV, Tarasov PE, Cherkinsky AE (1997) The Latest Pleistocene in southwestern Siberia and Kazakhstan. *Quaternary International*, **41**, 125–134.

Kruklis MV, Milyutin LI (1977) *Larix Czekanovskii*. Nauka, Moscow, Soviet Union (in Russian).

Kuzmin YV, Orlova LA (2004) Radiocarbon chronology and environment of woolly mammoth (*Mammuthus primigenius* Blum.) in Northern Asia: result and perspectives. *Earth-Science Reviews*, **68**, 133–169.

Lapteva EG, Korona OM (2008) Paleobotanical investigation of the recovery site of the Mongochenskii mammoth (Gydan peninsula). In: *Palynology: Stratigraphy and Geology*. Collection of the scientific works of the XII all-Russian palynological conference (September 29 – October 4 2008, S. – Petersburg). Vol. II. P. 149–154. (in Russian).

Lascoux M, Palme AE, Cheddadi R, Latta RG (2004) Impact of Ice Ages on the genetic structure of trees and shrubs. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**, 197–207.

Lindqvist C, Schuster SC, Sun YZ *et al.* (2010) Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. *Proceedings of the National Academy of Sciences of the USA*, **107**, 5053–5057.

Mantel NA (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Navascués MZ, Vaxevanidou S, González-Martínez J, Climent L, Gil L, Emerson BC (2006) Chloroplast microsatellites reveal colonization and metapopulation dynamics in the Canary Island pine. *Molecular Ecology*, **15**, 2691–2698.

Neale DB, Sederoff RR (1989) Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. *Theoretical and Applied Genetics*, **77**, 212–216.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, New York, USA.

Nelson MR, Wegmann D, Ehm MG *et al.* (2012) An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science*, **337**, 100–104.

Nikitin PI (1957) *Pliocene and Quaternary Floras of Voroneg Region*. Academy of Sciences of USSR, Moscow, Leningrad. pp. 206. (in Russian)

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

Parducci L, Jørgensen T, Tollefsrud MM *et al.* (2012) Glacial survival of boreal trees in Northern Scandinavia. *Science*, **335**, 1083–1086.

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution and Systematics*, **37**, 187–214.

Petit RJ, Aguinagalde I, de Beaulieu JL *et al.* (2003) Glacial refugia: hotspot but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

Polezhaeva MA, Lascoux M, Semerikov VL (2010) Cytoplasmic DNA variation and biogeography of *Larix* Mill. in Northeast Asia. *Molecular Ecology*, **19**, 1239–1252.

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered *versus* unordered alleles. *Genetics*, **144**, 1237–1245.

Provan J, Soranzo N, Wilson NJ, Goldstein DB, Powell W (1999) A low mutation rate for chloroplast microsatellites. *Genetics*, **153**, 943–947.

Pyhäjärvi T, García-Gil MR, Knürr T, Mikkonen M, Wachowiak W, Savolainen O (2007) Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. *Genetics*, **177**, 1713–1724.

Rogstad SH, Keane B, Collier MH (2003) Minisatellite DNA mutation rate in dandelions increases with leaf-tissue concentrations of Cr, Fe, Mn, and Ni. *Environmental Toxicology and Chemistry*, **22**, 2093–2099.

Rohlf EJ (1988) *Numerical Taxonomy and Multivariate Analysis System*. Exeter Publishing, New York.

Semerikov V, Lascoux M (2003) Nuclear and cytoplasmic variation within and between Eurasian *Larix* (Pinaceae) species. *American Journal of Botany*, **90**, 1113–1123.

Semerikov VL, Semerikov LF, Lascoux M (1999) Intra and Interspecific Allozyme Variability in Eurasian *Larix* Mil. *L. s*pecies. *Heredity*, **82**, 193–204.

Semerikov VL, Zhang H, Sun M, Lascoux M (2003) Conflicting phylogenies of *Larix* (Pinaceae) based on cytoplasmic and nuclear DNA. *Molecular Phylogenetics and Evolution*, **27**, 173–184.

Semerikov VL, Vendramin GG, Sebastiniani F, Lascoux M (2006) RAPD-derived, PCR-based mitochondrial markers for *Larix* species and their usefulness in phylogeny. *Conservation Genetics*, **7**, 621–625.

Semerikov VL, Iroshnikov AI, Lascoux M (2007) Mitochondrial DNA variation pettern and postglacial history of Siberian Larch (*Larix sibirica* Ledeb.). *Russian Journal of Ecology*, **38**, 147–154.

Semerikova SA, Semerikov VL (2006) Genetic variation and population differentiation in Siberian fir *Abies sibirica* Ledeb. inferred from allozyme markers. *Russian Journal of Genetics*, **42**, 636–644.

Stefanovskii VV (2006) *Pliocene and Quaternary of the Eastern Slope of the Urals and Trans-Urals*. Institute of Geology and Geochemistry, Ural Branch of Russian Academy of Sciences, Ekaterinburg. 224 p. (in Russian).

Svendsen JI, Alexanderson H, Astakhov VI *et al*. (2004) Late Quaternary ice sheet history of northern Eurasia. *Quaternary Science Reviews*, **23**, 1229–1271.

Tollefsrud MM, Sonstebo JH, Brochmann C, Johnsen O, Skroppa T, Vendramin GG (2009) Combined analysis of nuclear and mitochondrial markers provide new insight into the genetic structure of North European *Picea abies*. *Heredity*, **102**, 549–562.

Tomskaya AI (2000). *The Mammoth Food Supply in the Late Pleistocene of Yakutia*. Yakutsk City Press, Yakutsk. 59 p. (In Russian).

Ukraintseva VV (1993) *Vegetation Cover and Environment of the "Mammoth Epoch" in Siberia*. The Mammoth Site of Hot Springs South Dakota, Hot Springs, South Dakoda.

Välirantä M, Kaakinen A, Kuhry P, Kultti S, Salonen JS, Seppä H (2011) Scattered late-glacial and early Holocene tree populations as dispersal nuclei for forest development in north-eastern European Russia. *Journal of Biogeography*, **38**, 922–932.

Velichko AA, Timireva SN, Kremenetski KV, MacDonald GM, Smith LC (2011) West Siberian Plain as a late glacial desert. *Quaternary International*, **237**, 45–53.

Vendramin GG, Lelli L, Rossi P, Morgante M (1996) A set of primers for the amplification of 20 chloroplast microsatellites in Panaceae. *Molecular Ecology*, **5**, 111–114.

Volkov IA (1980) *Ice-Dammed Lacustrine Basin. Paleogeography of West-Siberian Plain at the Zyryanka Glaciation Maximum*, pp. 32–42. Nauka, Novosibirsk.

Volkova VS, Arkhipov SA, Babushkin AE *et al*. (2002) *Stratigraphy of Oil and Gas Basins of Siberia. Cenozoic of West Siberia*. Publ. Hous of SB RAS, Department "GEO", Novosibirsk. pp. 246.

Wang X-Q, Tank DC, Sang T (2000) Phylogeny and divergence times in Pinaceae: evidence from three genomes. *Molecular Biology and Evolution*, **17**, 773–781.

Wang B, Mao JF, Gao G, Zhao W, Wang XR (2011) Colonization of the Tibetan Plateau by the homoploid hybrid pine *Pinus densata*. *Molecular Ecology*, **20**, 3796–3811.

Willis KJ, van Andel TH (2004) Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. *Quaternary Science Reviews*, **23**, 2369–2387.

Willis KJ, Rudner E, Sumegi P (2000) The full-glacial forests of central and southeastern Europe. *Quaternary Research*, **53**, 203–213.

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant. *Proceedings of the National Academy of Sciences of the USA*, **84**, 9054–9058.

Vladimir Semerikov's team (V.S., M.P. and S.S.) is interested in the phylogeography and phylogeny of forest species in connection with the biogeography of Northern Eurasia. P.K. is a leader of a group of paleontologists and is interested in paleogeography and paleozoology of West Siberia. ML is interested in evolutionary genomics of adaptation and speciation using *Picea* species, *Capcella* sp. and other plants as study systems. V.S., M.P., S.S., P.K. and ML. contributed to the writing of the manuscript, V.S. and M.L. contributed to the data analysis, V.S., M.P. and S.S. – to laboratory setup.

Data accessibility

Cp- and mt-DNA haplotype frequencies and description of the studied populations are indicated in Table S1 (Supporting information). It is submitted to Dryad database (doi: 10.5061/dryad.jq712)

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Distribution of mitotypes in studied populations.

Table S2 Location of summary statistics of observed data set relatively to data sets simulated from the posteriors.

Fig. S1 Distribution of mtDNA haplotype at individual loci.

Fig. S2 Distribution of average pairwise difference between individuals $\overline{D^2_{sh}}$ based on cpSSR data. SAMOVA population groups based on mtDNA variation are indicated.

Fig. S3 Frequency of chlorotypes 41233, 46243, 35243, 46242, 36243, 42244, 35233, 36143, 35235, specific to western race of *L. sibirica*, chlorotypes 44223 and 36143 specific to Altai populations and Low Ob River area, respectively.

Fig. S4 Frequency of chlorotypes, 21422, 22423, 22432, 12422, 32422, 22322, specific to *L. gmelinii* × *L. sibirica* hybrids.

Fig. S5 Principal component analysis of one thousand data sets simulated from priors, one hundred data sets simulated from posteriors and observed data set, axes 1 and 2.