

Nuclear and chloroplast DNA phylogeography reveal two refuge areas with asymmetrical gene flow in a temperate walnut tree from East Asia

Wei-Ning Bai^{1,2}, Wan-Jin Liao² and Da-Yong Zhang²

¹State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China; ²MOE Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing 100875, China

Summary

Author for correspondence:

Wei-Ning Bai

Tel: +86 10 62836514

Email: baiwn@ibcas.ac.cn

Received: 13 May 2010

Accepted: 25 June 2010

New Phytologist (2010) **188**: 892–901

doi: 10.1111/j.1469-8137.2010.03407.x

Key words: gene flow, glacial refugia, *Juglans mandshurica*, northern and northeastern China, phylogeography.

- Recently, there has been a debate about whether the temperate forests of East Asia merged or fragmented during glacial periods in the Pleistocene. Here, we tested these two opposing views through phylogeographical studies of the temperate-deciduous walnut tree, *Juglans mandshurica* (Juglandaceae) in northern and northeastern China, as well as Japan and Korea.
- We assessed the genetic structure of 33 natural populations using 10 nuclear microsatellite loci and seven chloroplast DNA (cpDNA) fragments.
- The cpDNA data showed the complete fixation of two different haplotype lineages in northeastern vs northern populations. This pronounced phylogeographic break was also indicated by nuclear microsatellite data, but there were disparities regarding individual populations. Among those populations fixed for haplotype A (the northeastern group), three were clustered in the northern group and four showed evidence of mixed ancestry based on microsatellite data.
- Our results support the hypothesis that two independent refugia were maintained across the range of *J. mandshurica* in the north of China during the last glacial maximum, contrary to the inference that all temperate forests migrated to the south (25–30°N). The discordance between the patterns revealed by cpDNA and microsatellite data indicate that asymmetrical gene flow has occurred between the two refugia.

Introduction

There are two opposing views on forest responses to the Quaternary climatic changes in East Asia (Qian & Ricklefs, 2000; Harrison *et al.*, 2001). Qian & Ricklefs (2000) suggested that multiple refugia for temperate forests might have existed in coastal areas in the north of China during the Last Glacial Maximum (LGM), which may have promoted the current species diversity through allopatric speciation.

Moreover, temperate forests would have extended across the continental shelf to link populations in China, Korea and Japan during glacial periods, whereas higher sea levels during interglacial periods isolated these regions. Conversely, palaeovegetation data from east Asia show that temperate forests in these regions were considerably more restricted than today and would have retreated southward to *c.* 30°N during the LGM, calling into question the

existence of northern refugia and the coalescence of tree populations required by the hypothesis of Qian & Ricklefs (Harrison *et al.*, 2001). It is not possible to determine whether temperate forests of eastern Asia coalesced or fragmented during the LGM without more detailed information (Qian & Ricklefs, 2001). Fortunately, molecular evidence has provided an effective approach, independent of fossil information, for testing the range dynamics of most organisms during the Quaternary (Avice, 2000), as climatic oscillations have left genetic signatures in current populations (Hewitt, 2000, 2004).

Tian *et al.* (2009) examined the phylogeographical pattern of a temperate deciduous shrub species (*Ostryopsis davidiana*) in northern China. They found multiple refugia were maintained across the range, contrasting with the conclusions of Harrison *et al.* (2001) that temperate forests would have retreated southward to 30°N during the LGM.

Chen *et al.* (2008) reached similar conclusions concerning a cold-resistant conifer species, *Pinus tabulaeformis*. Until now, there were no independent phylogeographical studies of temperate deciduous tree species in eastern Asia to test the two hypotheses, especially in northeastern and northern China as well as Japan and Korea. Northeastern China (NEC) is a megadiversity area with complex topography within 40–50°N, adjacent to the Korean Peninsula, including the NEC plain and the major mountain ranges, such as the Daxing'anling Mountain range and the Changbai Mountains (China EPA, 1998; Xu *et al.*, 1999) (Fig. 1). Northern China is an important part of the north–south vegetation transect, south to the Qinling Mountains and north to the Yanshan Mountain, including the northern China plain and the Taihang Mountains (Ren, 1985) (Fig. 1). Both regions are a mosaic of mountains and were characterized by a relatively mild Pleistocene climate (Weaver *et al.*, 1998; Ju *et al.*, 2007), potentially hosting microclimatic zones capable of supporting a variety of habitats in relative stability (Qian & Ricklefs, 2000). Thus, glacial refugia may have been available in these regions for East Asian species. However, another possibility that cannot be ruled out is that a refugium might have been present in the southern Korean peninsula and/or Japan, and the species colonized NEC from there during postglacial times. More phylogeographical studies in these regions would contribute to a resolution of this issue.

Most phylogeographic studies of plants have been based on chloroplast DNA (cpDNA) and have revealed genetic heterogeneity throughout the range of a species and allowed an inference of historical range shifts and recolonization routes (Taberlet *et al.*, 1998; Wares & Cunningham, 2001;

Petit *et al.*, 2003). Although the merits of the nonrecombinant and maternally inherited cpDNA for phylogeographic studies in most angiosperms have been demonstrated, the genetic structure only reflects the history of a single gene (Hey & Machado, 2003). In addition, cpDNA can only reveal seed gene flow, providing no information about pollen gene flow, a main component shaping the organization of genetic diversity within and among populations. Liepelt *et al.* (2002) showed, using cpDNA and mtDNA markers, that a high level of pollen-mediated gene flow between refugia in a conifer species (*Abies alba*) eliminated the genetic imprints of Pleistocene refugial isolation. Recently, it has been shown that very different phylogeographic structures can be resolved by DNAs that contrast in rates of gene flow (Currat *et al.*, 2008; Du *et al.*, 2009; Zhou *et al.*, 2010). In view of these findings, it is desirable that plant phylogeographic studies should employ both nuclear and cpDNA markers, as use of both types of marker could reveal more information than a single type of marker about the history of range shifts and postglacial gene flow among refugia during Pleistocene climatic oscillations (e.g. Schonswetter *et al.*, 2005; Alsos *et al.*, 2007; Edh *et al.*, 2007).

In this study we examined the phylogeographic patterns of *Juglans mandshurica*, a temperate deciduous tree distributed in northern and northeastern China, and locally scattered in the Russian Far East, Korea and Japan. This species is wind-pollinated, and its fruits (walnuts) typically fall in the vicinity of parental plants, so seed dispersal distance is generally much more limited than pollen. Flowering is heterodichogamous (Bai *et al.*, 2006), that is there are two temporal morphs (protandry and protogyny)

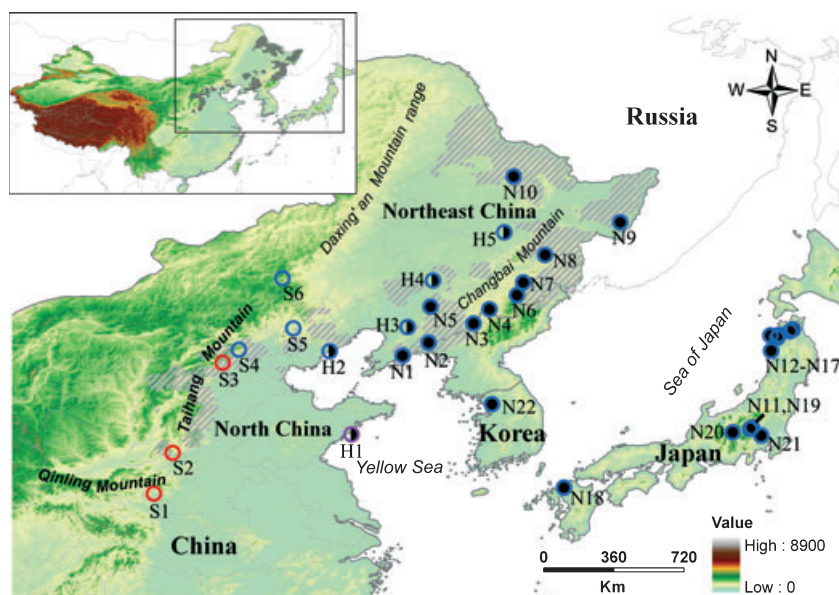


Fig. 1 Locations for 33 populations of *Juglans mandshurica* sampled. The present-day distribution of *J. mandshurica* is shaded gray. Circles inside colors correspond to two genetic clusters identified by the program STRUCTURE: blank, northern group; half-blank and half-black, admixed group; black, northeastern group. Red circles, haplotype B; blue circles, haplotype A; purple circles, haplotype A and B.

within a population, with the flowering periods of the two morphs reciprocal and synchronous. We used cpDNA, which is maternally inherited in *Juglans* species (Potter *et al.*, 2002), as well as 10 nuclear microsatellite loci, which were biparentally inherited. Data from the two marker systems were compared at the population level in the phylogeographic analysis of *J. mandshurica*. The main goals of the study were: to infer the existence and locations of past glacial refugia for *J. mandshurica*; to reconstruct its postglacial history and gene flow according to the two different molecular marker systems; and to ascertain whether temperate forest communities merged or fragmented during glacial periods.

Materials and Methods

Population sampling

We collected leaf samples from 670 adult individuals of 33 natural populations covering the whole range of *J. mandshurica* Maxim. (Juglandaceae) in northern China and NEC, Japan and Korea. Around 30 individuals at least 30 m apart from each other were sampled in most Chinese populations, whereas only 1–10 individuals were sampled from Japanese and Korean populations (see the Supporting Information Table S1).

DNA extraction and microsatellite genotyping

Total genomic DNA was extracted from dried leaf tissue by a modified cetyltrimethylammonium bromide (CTAB) method (Bai *et al.*, 2007). Genotypes of DNA samples were scored using 10 pairs of the microsatellite PCR primers that were developed for *Juglans nigra* (Woeste *et al.*, 2002; Dangl *et al.*, 2005; Table S2). To ensure that our PCR products were indeed microsatellite fragments, the products of one or two individuals for each of the 10 loci (Table S2) were sequenced using an ABI 3100 automated sequencer (Applied Biosystems, Foster City, CA, USA).

Polymerase chain reaction amplification of primer pairs was performed with a PTC-200 thermal cycler (MJ Research Inc., Waltham, MA, USA) using 20- μ l reactions. The PCR reaction mixture contained 50 mM Tris-HCl (pH 8.0), 500 mg ml⁻¹ KCl, 1.5 mM MgCl₂, 200 μ M dNTP and 0.4 μ M (each) primer; the upper primers were labeled with fluorescent dye, 6-FAM, TAMRA or HEX (Applied Biosystems), 20 ng of DNA template, and 0.6 U *Taq* polymerase (TaKaRa, Tokyo, Japan). The PCR amplifications were performed as follows: an initial denaturation step at 94°C for 5 min followed by 35 cycles of 30 s at 94°C, 1 min at an annealing temperature and 1 min at 72°C, and a final extension step at 72°C for 10 min. The annealing temperatures were: 54°C for WGA32, WGA72 and WGA79; 52°C for WGA4; and 50°C for WGA7,

WGA009, WGA089, WGA118, WGA202 and WGA276. The PCR products were separated on an ABI 3100 automated sequencer using a 50-cm capillary, polymerPOP-6 and ROX 500 (both Applied Biosystems) as an internal standard. Fragment sizes were assessed using GENEMAPPER software version 3.7 (Applied Biosystems). Allele size determination was performed twice manually to reduce scoring error.

Chloroplast DNA sequencing

Seven cpDNA intergenic spacer regions: *psaI-accD*, *rpl36-infA*, *trnQ-5' rps16*, *rpl32-trnL* (Shaw *et al.*, 2007), *trnQ-trnS* (Kanno *et al.*, 2004), *trnH-trnK* (Demesure *et al.*, 1995) and *psbB-psbT-psbN* (Hamilton, 1999) were sequenced. We sequenced > 5000 bp of seven regions for 161 individuals (mean = 5 per population, range = 1–10 individuals) in 33 populations (Table S1). A PCR was performed with a PTC-200 thermal cycler (MJ Research Inc.) using 25- μ l reactions containing the following reaction components: 1 μ l template DNA (c. 10–100 ng), 1 \times buffer, 200 mmol⁻¹ each dNTP, 3.0 mmol l⁻¹ MgCl₂, 0.1 mmol l⁻¹ each primer and 1.25 units *Taq* (TaKaRa). The cycling conditions were template denaturation at 80°C for 5 min followed by 30 cycles of denaturing for 1 min at 95°C, annealing at 50°C for 1 min with a ramp of 0.3°C s⁻¹ to 65°C, and extension for 4 min at 65°C. A final extension followed at 65°C for 4 min.

The PCR products were purified before sequencing with the quick PCR Purification Kit (Qiagen). All DNA sequencing was performed with the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit, v. 2.0 or 3.1 (Perkin-Elmer/Applied Biosystems, Foster City, CA, USA), using the thermal cycle parameters 80°C, 5 min; 30 \times (10 s at 96°C; 5 s at 50°C; and 4 min at 60°C). The products were electrophoresed and detected on an ABI Prism 3100 automated sequencer. All sequences were deposited in GenBank under accession numbers HM466687–HM466693.

Microsatellite data analysis

Genetic diversity within populations For each microsatellite locus, genetic diversity was assessed by calculating the observed number of alleles (A_o), observed heterozygosity (H_o), gene diversity (H_S) within populations and total gene diversity (H_T). For each population, genetic diversity was estimated across all loci using A_o , H_o , H_S and allele richness (R_S), a sample-size independent measure of the number of alleles (Petit *et al.*, 1998). Allele richness was standardized for 10 individuals (by discarding all populations with < 10 individuals). These statistics were estimated using FSTAT 2.9.3 (Goudet, 2001). The significance of deviations from Hardy–Weinberg equilibrium, given by deviations of

fixation index (F_{is}) from zero, was tested by randomization using *FSTAT*. Genotypic disequilibrium was tested for all locus pairs in each population by randomization and the obtained *P*-values ($= 0.05$) were adjusted applying a sequential Bonferroni correction (Rice, 1989) to avoid false positives, using *FSTAT*. We calculated the genetic diversity of 22 populations with a sample size ≥ 10 to increase the detection power. To determine whether genetic variation within populations was correlated with geographical gradients, Pearson correlations between statistics of variation (R_S and H_S) and geographic ordinates (latitude) for each population were analysed.

Genetic differentiation and relationships among populations

We determined levels of genetic differentiation among populations using θ (F_{ST}) (Weir & Cockerham, 1984) and the standardized genetic differentiation G'_{ST} (Hedrick, 2005) across 10 loci with the web-based software *SMOGL* (Crawford, 2010). Genetic distances (D_A ; Nei *et al.*, 1983) were calculated from allele frequencies and the resulting distance matrix was used to create a neighbor-joining tree. We assessed the reliability of the tree with 1000 bootstraps using *DISPAN* (Ota, 1993).

STRUCTURE version 2.1 (Pritchard *et al.*, 2000) applies a Bayesian method to infer the number of clusters (K) without using prior information of individual sampling locations. This program distributes individuals among K clusters based on their allelic frequencies and estimates the posterior probability of the data given each particular K . *STRUCTURE* was run for $K = 1$ to $K = 23$ clusters. Each run was pursued for 1 000 000 MCMC interactions, with an initial burn-in of 100 000, and an ancestry model that allowed for admixture, with the same alpha for all populations. To assess stability, 10 independent simulations were run for each K . The final posterior probability of K , $\Pr(X|K)$, was computed, as suggested by Pritchard *et al.* (2000), using the runs with highest probability for each K . However, as indicated in the *STRUCTURE* documentation and Evanno *et al.* (2005), $\Pr(X|K)$ usually plateaus or increases slightly after the 'right' K is reached. Thus, following Evanno *et al.* (2005), ΔK , where the modal value of the distribution is located at the real K , was calculated. The sensitivity of the final result to specific previous assumptions of alpha and independence of allelic frequencies was also computed.

Regional genetic diversity and structure To assess whether genetic diversity and structure differ between regions, we first calculated average values of R_S , H_S and F_{ST} within each region and then tested differences in these parameters among regions using a permutation test with *FSTAT*. Bayesian clustering was also tested for each region with the same methods as described earlier. The patterns of spatial genetic structure described as isolation-by-distance (IBD) models (Wright, 1943) were evaluated according to Rousset

(1997) using a Mantel test with 9999 random permutations, which was performed between the matrix of pairwise population differentiation in terms of $F_{ST}/(1 - F_{ST})$ and the matrix of the natural logarithm of geographic distance. These IBD evaluations were tested for all populations (sample size > 10) and populations within each region using *GENALEX* (Peakall & Smouse, 2001).

Gene flow

To verify whether there was asymmetrical gene flow between groups (cpDNA division and microsatellite *STRUCTURE* division), we used the software package *MIGRATE* version 3.1.3 (Beerli, 2006) to estimate the effective number of migrants ($4Nm$, where N is the effective population size and m is the migration rate) entering and leaving each group per generation. The *MIGRATE* program calculates maximum-likelihood (ML) estimates for both migration rates and effective population size between pairs of populations using a coalescent approach (Beerli & Felsenstein, 1999). To avoid the confounding effects of differences in sample size on the estimate of gene flow, the software picked a random subset of individuals from the larger group, with the number of individuals in the subset the same as in the smaller group. We relied on maximum likelihood estimation and used 10 short chains (10 000 trees) and three long chains (1 000 000) with 10 000 trees discarded as initial 'burn-in', replicates = YES: 5, randomtree = YES, heating = ADAPTIVE: 1{1 1.2 1.5 3.0}. We ran *MIGRATE* five times to verify the consistency of our results. For each run, we changed the random number seed and the starting values for θ ($4 N\mu$, where μ is the mutation rate) and $4 Nm$. In the first run θ and $4 Nm$ were estimated from F_{ST} values and in the subsequent runs estimates of θ and $4 Nm$ from the previous run were used. The estimates from the final run are reported here.

Divergence time between lineages

Sequences were aligned using *CLUSTAL_X* version 1.81 (Thompson *et al.*, 1997). Net pairwise divergence per base pair (d_A), which is proportional to time since divergence (T) of two clades assuming homogeneity of mutation rates across lineages, was calculated using *MEGA 4* (Tamura *et al.*, 2007) under the Kimura-2 model. Divergence time was calculated as $T = d_A/2 \mu$, where μ is the rate of nucleotide substitution (Nei & Kumar, 2000). An average overall rate of cpDNA sequence divergence was taken as 0.772×10^{-10} substitutions per site per year based on fossil records of *Pterocarya* and *Juglans* (c. 54 million yr ago (Mya)) (Aradhya *et al.*, 2005). However, this average rate was not appropriate for all the species in *Juglans*, because those in section *Cardiocaryon* have evolved at significantly different rates from taxa in section *Rhysocaryon*. An appropriate rate has not been calibrated for the seven chloroplast intergenic

spacer (IGS) regions, so we used the average rate for cpDNA in seed plants (1.01×10^{-9} substitutions per site per year) (Graur & Li, 1999) and *Juglans* (0.772×10^{-10} substitutions per site per year) as an approximation.

Results

Microsatellite diversity within populations

Estimates of diversity varied among microsatellite loci (Table 1) and among populations (Table S3). The range in total number of alleles per locus in our sample of 670 individuals was 15–51 with a total of 275 alleles across the 10 loci. H_S among loci was 0.77, with a range of 0.64–0.92. H_T estimates among all populations ranged from 0.74 to 0.96 and averaged 0.85. Within populations, the range was 0.67–0.81 for H_S , 0.64–0.81 for H_o , and 4.8–8.5 for R_S . For each population, F_{is} showed no significant deviation from zero at any locus (Table S3), suggesting that Hardy–Weinberg equilibrium assumptions applied to the populations. No evidence was found for significant genotypic disequilibrium among the 45 locus pairs in each population, after correcting for multiple testing using the sequential Bonferroni procedure. From this we concluded that the 10 loci were sufficiently independent for use in Bayesian clustering methods.

Genetic differentiation between populations

Population differentiation was significant at each locus ($P < 0.05$; Table 1), with the average F_{ST} value for multilocus estimates equal to 0.091, (range 0.045–0.191). Pairwise- F_{ST} values between populations were all significant ($P < 0.05$), except between N5 and N6. The greatest value was 0.113, between H1 and N11, and the lowest was

0.015, between N5 and N6. Note that the F_{ST} for microsatellite variation may considerably underestimate the extent of population differentiation owing to high heterozygosity within populations (Hedrick, 2005). The standardized genetic differentiation, G'_{ST} , given by Hedrick (2005) was higher than F_{ST} across all loci ($G'_{ST} = 0.429$, range 0.316–0.616, Table 1).

When using the Bayesian clustering approach of Pritchard *et al.* (2000), the inference of the number of gene pools K was not straightforward as log-likelihood values for the data conditional on K , $\ln P(X|K)$, increased progressively as K increased (Fig. 2). In such a case it may not be possible to determine the true value of K and Falush *et al.* (2003) have suggested choosing the smallest value that captures the major structure in the data (software documentation available at http://pritch.bsd.uchicago.edu/software/structure_v.2.3.1/documentation.pdf). However, ΔK values (Evanno *et al.*, 2005) computed for all K classes indicated a strong signal for $K = 2$ (Fig. 2). Changing the assumptions of 'equal alpha for each population' and 'correlated allele frequencies' did not change this final result. The proportions of each individual in each population assigned into two clusters (clusters I and II) are shown in Fig. 2. The Bayesian clustering approach revealed that cluster I was mainly distributed in NEC, Japan and Korea, across 22 populations (N1–N22), while cluster II was mainly distributed in northern China across six populations (S1–S6). Five populations (H1–H5) comprised individuals representing both clusters (Figs 1 and 2). Although the bootstrap value of the neighbor-joining tree based on Nei *et al.*'s D_A (Fig. S1) was very low, two clusters of populations could be recognized.

Regional genetic diversity and structure

The STRUCTURE analyses indicated three 'regions' (northeast, admixed and north) that corresponded to the locations of the grouped populations. The values of the genetic diversity and differentiation parameters described as R_S , H_S and F_{ST} , obtained in the global and hierarchical analyses are shown in Table 2. Although measures of genetic diversity (H_o , H_S , F_{is} and F_{ST}) were not significantly different among the three groups, the difference in R_S was marginally significant ($P = 0.06$; northeast > admixed > north). Furthermore, the respective F_{ST} and G'_{ST} values among populations within northeastern (0.065 and 0.349) and northern regions (0.077 and 0.319), were lower than those among all 22 populations (0.092 and 0.429).

In the spatial genetic structure evaluations, there was significant IBD among all 22 populations (11 northeastern, six northern and five admix populations) ($r = 0.53$, $P < 0.01$) and among populations within northeastern ($r = 0.46$, $P = 0.03$) and northern regions ($r = 0.46$, $P = 0.05$), but not within admixed regions ($r = 0.36$, $P = 0.24$). However,

Table 1 A comparison of genetic diversity at 10 microsatellite loci in 22 *Juglans mandshurica* populations

Locus	A_o	H_o	H_S	H_T	F_{ST}	G'_{ST}
WGA4	19	0.578	0.718	0.802	0.113	0.384
WGA7	14	0.588	0.635	0.754	0.173	0.442
WGA32	51	0.916	0.924	0.964	0.043	0.622
WGA72	15	0.726	0.687	0.848	0.191	0.415
WGA79	27	0.777	0.808	0.879	0.078	0.576
WGA009	30	0.620	0.669	0.744	0.103	0.307
WGA089	26	0.732	0.803	0.862	0.063	0.388
WGA118	24	0.797	0.846	0.890	0.044	0.352
WGA202	18	0.724	0.728	0.796	0.077	0.358
WGA276	49	0.865	0.890	0.936	0.048	0.479
Mean	27.5	0.734	0.771	0.847	0.091	0.432

A_o , Observed allele number; H_o , observed heterozygosity; H_S , gene diversity; H_T , overall gene diversity; F_{ST} , among population differentiation; G'_{ST} , standardized genetic differentiation.

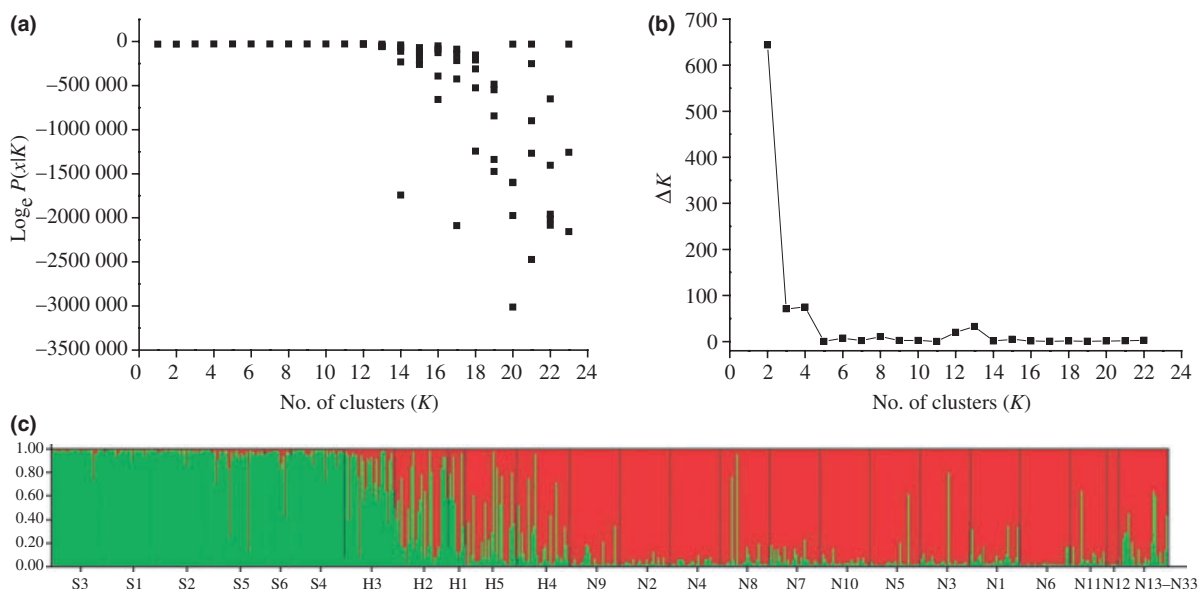


Fig. 2 Bayesian inference of the number of clusters (K) of *Juglans mandshurica*. K was estimated using (a) the posterior probability of the data given each K (10 replicates), (b) the distribution of ΔK , and (c) the two clusters were detected from STRUCTURE analysis.

Table 2 Differences between groups of populations of *Juglans mandshurica* in the mean values of R_s , H_o , H_s , F_{ST} and F_{is} investigated with 1000 permutation tests

Groups	R_s	H_o	H_s	F_{is}	F_{ST}
Northeastern	7.0	0.724	0.756	0.042	0.065
Northern	6.3	0.736	0.780	0.057	0.077
Admixed	6.6	0.756	0.784	0.041	0.065

R_s , Allele richness; H_o , observed heterozygosity; H_s , gene diversity; F_{ST} , among population differentiation; F_{is} , fixation index.

there was no significant IBD in the northeastern group when population N11, located in Japan, was excluded ($r = 0.19$, $P = 0.26$). Pearson correlation analysis showed that intrapopulation genetic diversity statistics (R_s and H_s) were not significantly correlated with latitude either in the northern ($r = -0.55$, $P = 0.26$) or the northeastern group ($r = -0.19$, $P = 0.59$).

Gene flow

Estimates of gene flow calculated with MIGRATE software and based on the pooled data indicated high levels of gene flow between populations. Unidirectional estimates of $4 Nm$ from lineage B to lineage A based on cpDNA haplotype division (see the next section) was 45.5623, much $> 4 Nm_{A \rightarrow B}$ of 32.5327. Of the six pairwise comparisons among three groups based on microsatellite STRUCTURE division (northeast, admixed and north), $4 Nm_{\text{north} \rightarrow \text{northeast}}$ was the highest value, 38.3212, and $4 Nm_{\text{northeast} \rightarrow \text{north}}$ was the lowest value of 28.0890 (Table 3).

Chloroplast DNA diversity and population structure

The aligned sequences of *psaI-accD*, *rpl36-infA*, *trnQ-trnS*, *trnH-trnK*, *trnQ-5'rps16*, *rpl32-trnL*, and *psbB-psbT-psbN* spacer were 944, 1155, 974, 552, 605, 857 and 763 bp in length, respectively. Nucleotide substitutions occurred at only one site in each spacer, and a total of seven polymorphic sites resulted in the resolution of only two haplotypes (Table S4). Haplotype frequencies in each population and geographical distribution are presented in Table S1. The geographical distribution of haplotypes was highly structured: 10 NEC populations, four admixed populations (H2–H5), three northern Chinese populations (S4–S6), and the Japanese and Korean populations were fixed for haplotype A (group A), while the other three northern populations (S1–S3) were fixed for haplotype B (group B). Only population H1 was polymorphic for both haplotype A and B. As $d_A = 0.001$, this indicated that haplotypes A and B diverged from each other between 0.5 Mya and 6.5 Mya.

Discussion

Population structure and phylogeographical history

Our survey of cpDNA variation in *J. mandshurica* resolved two different haplotype lineages, A and B, within the species. Geographically, lineage A covered the entire sampling region of NEC, Japan and Korea, while lineage B was restricted to northern China. This pronounced phylogeographic break in the species was also indicated by the analysis of the nuclear microsatellite variation, and this leads

Table 3 Estimates of gene flow ($4Nm$) among three groups of *Juglans mandshurica* based on microsatellite STRUCTURE division

Group, <i>i</i>	θ_i^*	Northeastern	Admixed	Northern
		NEC $\rightarrow i$	Admix $\rightarrow i$	North $\rightarrow i$
Northeastern	4.4156		37.0654	38.3212
	4.2408–4.5998		35.9513–38.2090	37.1821–39.4860
Admixed	3.5129	28.3751		30.0437
	3.3842–3.6479	27.5175–29.2498		29.1568–30.9492
Northern	3.4838	28.0899	29.8028	
	3.3560–3.6188	27.2421–28.9555	28.9245–30.6974	

*The range estimates given below each value are the 95% confidence limits. Numbers in bold identify comparisons where asymmetrical gene flow was detected.

us to propose that in the past the species distribution was fragmented into two independent refugia.

There are two hypotheses concerning the occurrence of refugia and postglacial expansion in NEC, Japan and Korea. One is that a refugium existed in the southern Korean peninsula and/or south Japan, and the species subsequently colonized NEC. This hypothesis is supported by an analysis of the phylogeography and the demographic history of the Chinese black-spotted frog (*Pelophylax nigromaculata*) (H. Zhang *et al.*, 2008). H. Zhang *et al.* (2008) obtained evidence that the frog lineage of the Korean peninsula colonized NEC via the Changbai Cordillera and remained *in situ* rather than expanding southward after recolonizing NEC. The second hypothesis is that a refugium was mainly in NEC and refugial populations colonized Japan and Korea from NEC via the Korean peninsula, as suggested by a study of the widespread, cold-temperate spruce species, *Picea jezoensis*, in northeast Asia (Aizawa *et al.*, 2007). Our cpDNA data cannot distinguish which of these two hypotheses is more likely, but our nuclear microsatellite data provide some clues in this regard. Genetic diversity statistics (R_s and H_s) were not significantly correlated with latitude in this region, which contrasts with what might be expected if a large-scale, post-glacial range expansion had occurred from refugia in Japan and Korea with this scenario reduced levels of genetic variation would have been expected throughout the recolonized NEC region. In the northeast, the species comprises populations mainly located in the Changbai Mountain (41°58'–42°06'N), which has a variety of forest types (Zheng *et al.*, 2001). At altitudes < 1100 m there is a typical mixed coniferous and broad-leaved forest zone, dominated by *Pinus koraiensis*, *Quercus mongolica* and *J. mandshurica*, etc.; at 1100–2000 m, a cool temperate conifer and subalpine forest zone, dominated by *Pinus jezoensis* var. *komarovii*, *P. jezoensis* var. *komarovii* and *Betula ermani*; and at > 2000 m, there is an alpine tundra zone (Huang *et al.*, 1959). Based on a recent study of the late Pleistocene glaciations in the Changbai Mountain region (W. Zhang *et al.*, 2008), glacial advances took place only at elevations above 2000 m. Thus the lower elevations may have been characterized by a relatively mild Pleistocene climate,

potentially including microclimatic zones capable of supporting various habitats with relative stability. This conclusion is supported by the results of a phylogeographic study conducted on *Fraxinus mandshurica*, a widely distributed temperate tree species in NEC (Hu *et al.*, 2008). Bearing this in mind and given the present-day very small and fragmented populations of *J. mandshurica* in Korea and Japan, we favor the idea that a major refugium for *J. mandshurica* existed in the past in NEC rather than in southern Korea or Japan.

Within the northern China region, populations S1–S3 were fixed for a different cpDNA haplotype (B) to all other populations and were clustered into a northern group according to a STRUCTURE analysis of microsatellite variation. Population S1 was located on the Funiu Mountain (32°14'–34°14'N), which is characterized by a stable, temperate mountain climate that presumably is favorable to temperate trees (Zhang & Zeng, 2000). Funiu Mountain is an eastern extension of the Qinling Mountains, an important dividing line between the warm temperate zone and the northern subtropic zone in China. The specific geographic location and complex ecological conditions, combined with less human interference, is likely to have preserved a rich biodiversity in this region (Shang *et al.*, 1998). Indeed, Tian *et al.* (2009) have suggested that multiple refugia for *Ostryopsis davidiana*, a species with reproductive biology similar to *J. mandshurica*, that is, wind-pollinated with fruits dispersed mainly by gravity and hoarding rodents (Chen, 1994), were maintained both north and south of the Qinling Mountains. Thus, we suggest that the Qinling Mountains could have been a refugium for temperate deciduous species in the past. The S2 and S3 populations, are located in the Taihang Mountains, a large mountain range in northern China that has abundant temperate tree species (Zhang *et al.*, 2006). Although the Taihang Mountains are north of the Qinling Mountains, fossil pollen data show that subtropical conifer, hardwood and deciduous broad-leaved trees, that is, *Quercus*, *Carpinus*, *Ulmus* and *Juglans* spp. occurred in this region during the LGM (Li, 1998). Therefore, the Taihang Mountains, together with the north of the Qinling Mountains, may represent another refugium for

J. mandshurica during the LGM. It should be noted that population S4, located quite near to S3, was fixed with haplotype A (characteristic of the northeastern group) instead of the haplotype B. It is feasible that the S4 and S3 populations represent the borderline between lineages A and B, respectively, but more sampling in the area is clearly needed for an unambiguous conclusion to be reached on this point. Interestingly, both populations were clustered in the northern group by microsatellite data, presumably owing to asymmetrical gene flow between the two refugia (see later).

Overall, our data indicate that two separate refugia were maintained across the range of *J. mandshurica*. One was in NEC (40–45°N), mainly in the Changbai Mountains, and the other in the Qinling and Taihang Mountains (32–38°N). The conclusion is contrary to the inference by Harrison *et al.* (2001) that temperate forests migrated to the south (25–30°N) without leaving forest stands in the north of China during the LGM. We found only one population, H1, located in Shandong Peninsula, containing both haplotypes A and B (Fig. 1). This implies that the distribution of *J. mandshurica* in northern China, Japan and Korea may have extended across the continental shelf through the Yellow Sea during the LGM, consistent with the hypothesis of Qian & Ricklefs (2000) concerning population coalescence. However, considering that H1 is only a remnant population containing only 10 individuals, and is located quite near to the Taoist temple of the Laoshan Mountain, the possibility of an introduction from other locations by visitors and pilgrims cannot be ruled out. Collectively, our results support the hypothesis of Qian & Ricklefs (2000) that multiple LGM refugia may have allowed species to persist across northern China, and the presently disjunct temperate forests in this region coalesced through the Yellow Sea during the LGM.

Contrasting patterns of molecular diversity

Although cpDNA and microsatellite markers were broadly consistent in distinguishing the northeastern and northern lineages, there were disparities regarding individual populations. Thus, populations S1–S6 were clustered in the northern group by microsatellite data, whereas three of them, populations S4–S6, were fixed for the northeastern haplotype A. Moreover, STRUCTURE analysis showed evidence of mixed ancestry in five populations (H1–H5), while all of these populations were fixed for haplotype A except H1 which was polymorphic for both haplotypes. Unidirectional gene flow $4Nm_A \rightarrow B$ was much $< 4Nm_B \rightarrow A$, and there was apparently asymmetrical gene flow from northern populations to northeastern populations via admixed populations (Table 3). So, we may infer that asymmetrical gene flow occurred between the two population groups, with the northern group nuclear genome being introduced into the northeastern group via pollen. The result that allelic

richness of the northeastern group was marginally higher than that of the northern group (Table 2) also suggests that new alleles might have migrated to the northeastern from the northern group by pollen flow.

Juglans mandshurica is wind pollinated and commences flowering in early spring (April–May). On average, its pollen likely travels far greater distances than its seed, which travels relatively limited distances because it is dispersed by gravity and rodents (Ma *et al.*, 2005). So, seed flow would be very limited compared with pollen flow. *Juglans mandshurica* is located in the east Asian monsoon region, which is the driving engine of the northwest and southeast surface winds in winter and summer, respectively (Nakagawa *et al.*, 2006; He *et al.*, 2007). The transition stage of the winter to summer monsoons is in March–May (Xu & Gao, 1962). In this period the summer monsoon displays a distinct stepwise northward and northeastward advance, while the winter monsoon displays a stepwise western retreat. Thus, in March–May, northern China is mainly affected by western and southern winds (Xu & Gao, 1962; Ren, 1985), which would facilitate pollen gene flow from the northern to the northeastern group.

In addition to wind direction, the heterodichogamous flowering of *J. mandshurica* (Bai *et al.*, 2006) may strengthen asymmetrical gene flow between the two groups. The flowering periods of the two morphs (protandry and protogyny) are reciprocal and synchronous, that is, the female flowering phase of protogynous individuals is synchronized with the male flowering phase of protandrous individuals for the first flowering period, and vice versa for the second flowering period. Protogynous individuals in the early-flowering, northern populations have no chance of receiving pollen from the late-flowering, northeastern populations, whereas the reverse is possible. Indeed, the pollen flow rate for the first flowering period was found to be higher than the second period in a northern population (Bai *et al.*, 2006, 2007). Thus, the northeastern populations in the first flowering period could more easily accept external pollen from northern populations in the second flowering period.

In summary, the strong differentiation of populations of the species into two highly distinct geographical groups based on cpDNA and microsatellite nuclear variation indicates that each of these groups is derived from separate refugia. Our study provides no evidence for the hypothesis of Harrison *et al.* (2001) that temperate forests were restricted in distribution and migrated to the south during the LGM, but supports the hypothesis of Qian & Ricklefs (2000) that multiple LGM refugia may have allowed species to persist in the north of China. Disparities between cpDNA and nuclear variation in *J. mandshurica* are likely to result from the greater mobility of pollen relative to seed and asymmetrical pollen gene flow from the northern to the northeastern group.

Acknowledgements

We thank Dr Peter Beerli for his help with MIGRATE version 3.1.3, Dr Yukawa Tomohisa, Yukawa Junichi and N-S. Lee for sample collection, and Dr Hong-Fang Wang for making the distribution map. We are grateful to the editor and three anonymous reviewers, Prof. Shou-Hsien Li, Song Ge, Xiao-Quan Wang and Yi-Bo Luo, for their helpful suggestions and valuable comments on the manuscript. This work was supported by the National Basic Research Program of China (2007CB411600) and the National Natural Science Foundation of China (30430160).

References

- Aizawa M, Yoshimaru H, Saito H, Katsuki T, Kawahara T, Kitamura K, Shi F, Kaji M. 2007. Phylogeography of a northeast Asian spruce, *Picea jezoensis*, inferred from genetic variation observed in organelle DNA markers. *Molecular Ecology* 16: 3393–3405.
- Alsos IG, Eidesen PB, Ehrich D, Skrede I, Westergaard K, Jacobsen GH, Landvik JY, Taberlet P, Brochmann C. 2007. Frequent long-distance plant colonization in the changing Arctic. *Science* 316: 1606–1609.
- Aradhya KM, Potter D, Gao FY, Simon CJ. 2005. Cladistic biogeography of *Juglans* (Juglandaceae) based on chloroplast DNA inter-genic spacer sequences. In: Motley TJ, Zerega N, Cross H, eds. *Darwin's harvest – new approaches to the origin, evolution, and conservation of crops*. New York, NY, USA: Columbia University Press, 143–170.
- Avice JC. 2000. *Phylogeography: the history and formation of species*. Cambridge, MA, USA: Harvard University Press.
- Bai WN, Zeng YF, Liao WJ, Zhang DY. 2006. Flowering phenology and wind-pollination efficacy of heterodichogamous *Juglans mandshurica* (Juglandaceae). *Annals of Botany* 98: 397–402.
- Bai WN, Zeng YF, Zhang DY. 2007. Mating patterns and pollen dispersal in a heterodichogamous tree, *Juglans mandshurica* (Juglandaceae). *New Phytologist* 176: 699–707.
- Beerli P. 2006. Comparison of Bayesian and Maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22: 341–345.
- Beerli P, Felsenstein J. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152: 763–773.
- Chen ZD. 1994. Phylogeny and phylogeography of the Betulaceae. *Acta Phytotaxonomica Sinica* 32: 1–153.
- Chen KM, Abbott RJ, Milne RI, Tian XM, Liu JQ. 2008. Phylogeography of *Pinus tabulaeformis* Carr. (Pinaceae), a dominant species of coniferous forest in northern China. *Molecular Ecology* 17: 4276–4288.
- China EPA. 1998. *Chinese country study on biological diversity*. Beijing, China: China Environmental Science Press.
- Crawford NG. 2010. SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources* 10: 556–557.
- Currat M, Ruedi M, Petit RJ, Excoffier L. 2008. The hidden side of invasions: massive introgression by local genes. *Evolution* 62: 1908–1920.
- Dangl GS, Woeste K, Aradhya MK, Koehmstedt A, Simon C, Potter D, Leslie CA, McGranahan G. 2005. Characterization of 14 microsatellite markers for genetic analysis and cultivar identification of walnut. *Journal of the American Society for Horticultural Science* 130: 348–354.
- Demesure B, Sodzi N, Petit RJ. 1995. A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* 4: 129–131.
- Du FK, Petit RJ, Liu JQ. 2009. More introgression with less gene flow: chloroplast vs. mitochondrial DNA in the *Picea asperata* complex in China, and comparison with other Conifers. *Molecular Ecology* 18: 1396–1407.
- Edh K, Widen B, Ceplitis A. 2007. Nuclear and chloroplast microsatellites reveal extreme population differentiation and limited gene flow in the Aegean endemic *Brassica cretica* (Brassicaceae). *Molecular Ecology* 16: 4972–4983.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620.
- Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.
- Goudet J. 2001. *FSTAT, a program to estimate and test gene diversities and fixation indices, version 2.9.3*. <http://www2.unil.ch/popgen/softwares/fstat.htm>
- Graur D, Li WH. 1999. *Fundamentals of molecular evolution*. Sunderland, MA, USA: Sinauer & Associates.
- Hamilton MB. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* 8: 521–523.
- Harrison SP, Yu G, Takahara H, Prentice IC. 2001. Palaeovegetation – diversity of temperate plants in east Asia. *Nature* 413: 129–130.
- He JJ, Ju JH, Wen ZP, Lu JM, Jin QH. 2007. A review of recent advances in research on Asian monsoon in China. *Advances in Atmospheric Sciences* 24: 972–992.
- Hedrick PW. 2005. A standardized genetic differentiation measure. *Evolution* 59: 1633–1638.
- Hewitt GM. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405: 907–913.
- Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359: 183–195.
- Hey J, Machado CA. 2003. The study of structured populations – new hope for a difficult and divided science. *Nature Reviews Genetics* 4: 535–543.
- Hu LJ, Uchiyama K, Shen HL, Saito Y, Tsuda Y, Ide Y. 2008. Nuclear DNA microsatellites reveal genetic variation but a lack of phylogeographical structure in an endangered species, *Fraxinus mandshurica*, across north-east China. *Annals of Botany* 102: 195–205.
- Huang X, Liu DS, Li Z. 1959. Natural scenery zone in the north of Changbai Mountains. *Acta Geographica Sinica* 25: 435–446.
- Ju LX, Wang HK, Jiang DB. 2007. Simulation of the Last Glacial Maximum climate over East Asia with a regional climate model nested in a general circulation model. *Palaeogeography Palaeoclimatology Palaeoecology* 248: 376–390.
- Kanno M, Yokoyama J, Suyama Y, Ohshima M, Itoh T, Suzuki M. 2004. Geographical distribution of two haplotypes of chloroplast DNA in four oak species (*Quercus*) in Japan. *Journal of Plant Research* 117: 311–317.
- Li WY. 1998. *Chinese Quaternary vegetation and environment*. Beijing, China: Science Press.
- Liepelt S, Bialozyt R, Ziegenhagen B. 2002. Wind-dispersed pollen mediates postglacial gene flow among refugia. *Proceedings of the National Academy of Sciences, USA* 99: 14590–14594.
- Ma WL, Luo JC, Jing T, Kujansuu J. 2005. Ecological studies and prospect of cultivation on *Juglans mandshurica* population. *Journal of Inner Mongolia Normal University* 34: 489–492.
- Nakagawa T, Tarasov PE, Kitagawa H, Yasuda Y, Gotanda K. 2006. Seasonally specific responses of the East Asian monsoon to deglacial climate changes. *Geology* 34: 521–524.
- Nei M, Kumar S. 2000. *Molecular evolution and phylogenetics*. New York, NY, USA: Oxford University Press.
- Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution* 19: 153–170.

- Ota T. 1993. *DISPAN: genetic distance and phylogenetic analysis, Version 1.1*. Pennsylvania State University, University Park. <http://homes.bio.psu.edu/people/faculty/nei/software.htm>
- Peakall R, Smouse PE. 2001. *GENALEX (version 5.1): genetic analysis in Excel. Population genetic software for teaching and research*. Canberra, Australia: Australian National University. <http://www.anu.edu.au/BoZo/GenAlEx/>.
- Petit RJ, Aguinalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M *et al.* 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**: 1563–1565.
- Petit R, El Mousadik A, Pons O. 1998. Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* **12**: 844–855.
- Potter D, Gao FY, Baggett S, McKenna JR, McGranahan GH. 2002. Defining the sources of Paradox: DNA sequence markers for North American walnut (*Juglans L.*) species and hybrids. *Scientia Horticulturae* **94**: 157–170.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Qian H, Ricklefs RE. 2000. Large-scale processes and the Asian bias in species diversity of temperate plants. *Nature* **407**: 180–182.
- Qian H, Ricklefs RE. 2001. Palaeovegetation – diversity of temperate plants in east Asia – Reply. *Nature* **413**: 130.
- Ren ME. 1985. *Introduction to physical geography of China*. Beijing, China: The Commercial Press.
- Rice WR. 1989. Analyzing tables of statistical tests. *Evolution* **43**: 223–225.
- Rousset F. 1997. Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics* **145**: 1219–1228.
- Schonswetter P, Stehlik I, Holderegger R, Tribsche A. 2005. Molecular evidence for glacial refugia of mountain plants in the European Alps. *Molecular Ecology* **14**: 3547–3555.
- Shang FD, Wang PJ, Feng GP. 1998. Study on the characteristics and causes of formation of plant diversity in the Funiushan transition region. *Journal of Henan University (Natural Science)* **28**: 54–60.
- Shaw J, Lickey EB, Schilling EE, Small RL. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* **94**: 275–288.
- Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**: 453–464.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**: 1596–1599.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**: 4876–4882.
- Tian B, Liu RR, Wang LY, Qiu Q, Chen KM, Liu JQ. 2009. Phylogeographic analyses suggest that a deciduous species (*Ostryopsis davidiana* Decne., Betulaceae) survived in northern China during the Last Glacial Maximum. *Journal of Biogeography* **36**: 2148–2155.
- Wares JP, Cunningham CW. 2001. Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution* **55**: 2455–2469.
- Weaver AJ, Eby M, Augustus FF, Wiebe EC. 1998. Simulated influence of carbon dioxide, orbital forcing and ice sheets on the climate of the Last Glacial Maximum. *Nature* **394**: 847–853.
- Weir BS, Cockerham CC. 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
- Woeste K, Burns R, Rhodes O, Michler C. 2002. Thirty polymorphic nuclear microsatellite loci from black walnut. *Journal of Heredity* **93**: 58–60.
- Wright S. 1943. Isolation by distance. *Genetics* **28**: 114–138.
- Xu SY, Gao YX. 1962. The East Asia monsoon and season. In: Gao YX, ed. *Several issues of the East Asia monsoon*. Beijing, China: Science Press, 88–103.
- Xu HG, Wang SQ, Xue DY. 1999. Biodiversity conservation in China: legislation, plans and measures. *Biodiversity and Conservation* **8**: 819–837.
- Zhang H, Yan J, Zhang GQ, Zhou KY. 2008. Phylogeography and demographic history of Chinese black-spotted frog populations (*Pelophylax nigromaculata*): evidence for independent refugia expansion and secondary contact. *BMC Evolutionary Biology* **8**: 16.
- Zhang JT, Xi Y, Li J. 2006. The relationships between environment and plant communities in the middle part of Taihang Mountain Range, North China. *Community Ecology* **7**: 155–163.
- Zhang JW, Zeng Y. 2000. Plant diversity and protection of Funiu Mountain. *Journal of Henan University (Natural Science)* **30**: 76–81.
- Zhang W, Niu YB, Yan L, Cui ZJ, Li CC, Mu KH. 2008. Late Pleistocene glaciation of the Changbai Mountains in northeastern China. *Chinese Science Bulletin* **53**: 2672–2684.
- Zheng Y, Xiao X, Guo Z, Howard T. 2001. A county-level analysis of the spatial distribution of forest resources in China. *Journal of Forest Planning* **7**: 69–78.
- Zhou FY, Abbott RJ, Jiang ZY, Du FK, Milne RI, Liu JQ. 2010. Gene flow and species delimitation: a case study of two pine species with overlapping distributions in southeast China. *Evolution*. doi: 10.1111/j.1558-5646.2010.00988.x.

Supporting Information

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

Fig. S1 Neighbor-joining tree of the 22 populations based on D_A distances (Nei *et al.*, 1983).

Table S1 Details of sample locations and sample sizes in 33 populations of *Juglans mandshurica*

Table S2 Microsatellite loci analysed, PCR conditions, motif and allele size ranges in *Juglans mandshurica*

Table S3 Descriptive statistics of genetic variability for 22 populations of *Juglans mandshurica*

Table S4 Designation of the chloroplast haplotypes detected at seven polymorphic cpDNA loci examined in 33 *Juglans mandshurica* populations

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.