Phylogeography of *Pterocarya hupehensis* reveals the evolutionary patterns of a Cenozoic relict tree around the Sichuan Basin

Zi-Jia Lu[1,2#], Tian-Rui Wang[1#], Si-Si Zheng[1#], Hong-Hu Meng[3,4], Jian-Guo Cao[2*], Yi-Gang Song[1,5,6*] and Gregor Kozlowski[1,6,7]

[1] Eastern China Conservation Centre for Wild Endangered Plant Resources, Shanghai Chenshan Botanical Garden, Shanghai 201602, China
[2] College of Life Sciences, Shanghai Normal University, Shanghai 200234, China
[3] Plant Phylogenetics and Conservation Group, Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming 650223, China
[4] Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences, Naypyidaw 05282, Myanmar
[5] College of Forestry and Biotechnology, Zhejiang A&F University, Hangzhou 311300, China
[6] Department of Biology and Botanic Garden, University of Fribourg, Fribourg 1700, Switzerland
[7] Natural History Museum Fribourg, Fribourg 1700, Switzerland
# Authors contributed equally: Zi-Jia Lu, Tian-Rui Wang, Si-Si Zheng
* Corresponding authors, E-mail: cao101@shnu.edu.cn; cherish-faith@163.com

Abstract

Environmental factors such as mountain tectonic movements and monsoons can enhance genetic differentiation by hindering inter- and intra-specific gene flow. However, the phylogeographic breaks detected within species may differ depending on the different molecular markers used, and biological traits may be a major confounding factor. *Pterocarya hupehensis* is a vulnerable relict species distributed throughout the Sichuan Basin. Here, we investigated the phylogeographic patterns and evolutionary history of *P. hupehensis* using chloroplast DNA and restriction site-associated DNA sequencing data from 18 populations around the Sichuan Basin. The 24 chloroplast haplotypes separated into western and eastern lineages at approximately 16.7 Mya, largely coincident with a strengthening of the East Asian monsoon system during the early to middle Miocene. Both cpDNA and nuclear DNA datasets consistently identified distinct western and eastern lineages whose phylogeographic break conformed to the boundary of the Sino-Himalayan and Sino-Japanese forest sub-kingdoms. However, in contrast to the nuclear gene data, the cpDNA data revealed further divergence of the eastern lineage into northern and southern groups along the Yangtze River, a result that likely reflects differences in the extent of pollen vs seed dispersal. During the temperature decline in the penultimate (Riss) glacial period of the Pleistocene epoch, *P. hupehensis* experienced a genetic bottleneck event, and ecological niche modeling suggests that a subsequent population expansion occurred during the last interglacial period. Our findings not only establish a basis for conservation of this species, but also serve as a case study for the effects of geography and climate change on the evolutionary history of wind-pollinated relict plants.

Introduction

The distribution and genetic structure of many species have been influenced by environmental factors such as monsoons, mountain tectonics, and other historical/ecological processes[1–3]. In addition, the presence of geographic barriers, combined with species-specific characteristics such as dispersal mode of seeds or pollen, can potentially effect the genetic differentiation of species[4–6]. Thus, understanding phylogeographic patterns and their potential influences on biomes is a primary objective of conservation and evolutionary biology[1,7].

The Sino-Japanese Floristic Region (SJFR), known for its abundant diversity of temperate flora, has attracted significant attention from phylogeographers and paleo-ecologists[8–10]. Its high biodiversity is typically explained by the absence of continental glaciation and a smaller magnitude of Quaternary environmental change[11]. The Sichuan Basin, a unique geological structure in East Asia, acts as a mountain refuge for numerous relict species[12–14]. The phylogeographic break of relict plants around the Sichuan Basin is mainly related to uplift of the Qinghai-Tibet Plateau (QTP) during the late Pliocene, as well as intensification of the East Asian monsoon system (EAMS)[15–17]. Climate fluctuations since the Miocene may also have been a key determinant of the differentiation and colonization of ancient taxa[18,19]. However, the dynamics of species around the Sichuan Basin during the Miocene climate change—and how species characteristics and geographic barriers affected their genetic patterns—remain poorly understood.

The climate of Asia has changed dramatically since the Miocene[20–22]. Between the early and middle Miocene, the pattern of aridity began to change from 'planetary' subtropical to 'inland'[21], and this was followed by formation and dominance of the monsoon climate[23,24]. This phenomenon is generally explained by uplift of the QTP[25,26] and cooling of the global climate during the middle Miocene[27], which further enhanced East Asian summer and winter monsoons during this period (*c.* 15–10 Mya)[28–30].

Many phylogeographic studies have shown that several post-Miocene uplift and monsoon events are related to the genetic structure and genetic differentiation of plants[14,18,31].

Monsoons can also strengthen geographic barriers by giving rise to different climatic environments on either side of the barrier, promoting the formation of lineage discontinuities, such as the 'Tanaka–Kaiyong Line' (TKL)[32,33]. Abundant summer precipitation, brought about by strengthening of the East Asian monsoon, is critical to the development and maintenance of subtropical evergreen broad-leaved forests in China[34,35]. Aridity is also an important factor affecting plant distribution[36]. With uplift of the QTP and strengthening of the Asian monsoon in the Early Miocene, inland drought began to occur and intensify in Asia[31,37,38]. This forced plants in the interior of Asia to retreat southward to find suitable habitats in subtropical regions with better hydrothermal conditions.

The Sichuan Basin, which is located in the second step of China's terrain, is surrounded by mountains. These mountains exhibit complex geomorphological and climatic characteristics that have created diverse habitats for relict species along steep ecological gradients[39,40]. Many physical and ecological barriers, such as mountains, rivers, and climate, are believed to drive population diversification and speciation around the Sichuan Basin[6,16]. Among the main phylogeographic patterns around the Sichuan Basin, the best-known is the 105°E line[6,41]. This line runs longitudinally across the Sichuan Basin, coincident with the boundary of the Sino-Himalayan and Sino-Japanese forest sub-kingdoms, dividing taxa into eastern and western lineages[13,42]. The TKL is another major phytogeographic boundary that traverses the mountains in the southwestern Sichuan Basin[43]. In addition, the Yangtze River (Three Gorges region) traverses the mountains in the eastern Sichuan Basin, forming a phylogeographic break that divides taxa into northern and southern lineages[31,44]. Monsoons, QTP uplift, mountains, and river basins have thus isolated biodiversity into areas of endemism or created lineages by impeding gene flow in dispersal-limited organisms[16,40,45].

Most phylogeographic studies of plants in East Asia have used chloroplast molecular markers[14,17,18,46,47]. Chloroplast DNA (cpDNA) is passed down by uniparental maternal inheritance in most angiosperms and is transmitted by seeds alone, thus providing no pollen-related information[48]. However, pollen-mediated gene flow may dominate in wind-pollinated trees[5,49]. Asymmetrical gene flows mediated by pollen and seeds have been consistently demonstrated in phylogenetic studies of many temperate tree species[5,41,50]. Different transmission characteristics may result in different intraspecific genetic structures. For example, inconsistencies between cpDNA and nuclear DNA of *Juglans cathayensis* and *Populus lasiocarpa* have been attributed to biological traits (e.g. extensive pollen exchange and wind-dispersed seeds) that partly delay the genetic imprinting of long-term isolation[5,6]. When exploring the effects of past events on the evolutionary history of taxa, species-specific biological characteristics should therefore be taken into account.

Pterocarya hupehensis is a tree species endemic to China that is found in mountainous areas (between 700 and 2,000 m above sea level) around the Sichuan Basin[45,51,52]. It is a typical riparian relict species of subtropical evergreen broad-leaved forests[51]. *P. hupehensis* is monoecious, producing flowers in catkins (amenta), with female amenta terminals on new growth; it is thus a typical anemophilic, cross-pollinated species[53,54]. *P. hupehensis* has been assessed as vulnerable (VU) and is considered to be of conservation concern[52]; a better understanding of its population genetics and phylogeography

can therefore guide appropriate protective actions. It is also an excellent case study for investigating potential drivers of genetic patterns in wind-pollinated relict species around the Sichuan Basin.

Here, we used cpDNA and restriction site-associated DNA sequencing (RAD-seq) datasets to reconstruct the evolutionary history and phylogeography of *P. hupehensis*. We then performed ecological niche modeling (ENM) to explore its suitable habitats from the past to the future. In particular, we addressed three questions. First, what is the genetic structure of *P. hupehensis* populations around the Sichuan Basin? Second, how did this species respond to climatic fluctuations from the Neogene to the Quaternary periods? Third, what evolutionary processes contributed to the observed genetic patterns?

Materials and methods

Sampling and sequencing

Leaves were collected from 18 natural populations covering the distribution range of *P. hupehensis* around the Sichuan Basin (Table 1). Individuals in each population were at least 50 m away from each other, and 156 individuals were sampled. Fresh leaves were dried using silica gel, and DNA was extracted from the dried leaf tissue using a modified CTAB method[55]. After removal of samples with low-quality DNA, 141 individuals were selected for cpDNA sequencing and 122 individuals for RAD sequencing (Supplemental Table S1). Voucher specimens of each individual were stored at the Shanghai Chenshan Botanical Garden Herbarium (CSH). After screening previously published universal primers, six cpDNA loci (*psbD-trn*T[56], *trn*V(UAC)x2-*ndh*C[57], *trn*L(UAG)-*rpL*32-F[56], *trn*S-*trn*fM[58], *trn*G-*trn*S[59], and *trn*D-*trn*Y[60]) were selected for use in this study. PCR amplification was performed as described previously[61]. The amplified PCR products were sequenced by Sangon Bioengineering Co., Ltd. (Shanghai, China).

Chloroplast DNA sequence analysis

Chloroplast DNA sequences were assembled and checked using Sequencher v4.1.4 (Gene Codes Corp., Ann Arbor, MI, USA). The sequences were aligned and calibrated using ClustalW implemented in MEGA v11[62] and then manually calibrated and adjusted. The haplotypes of the chloroplast fragments were extracted using DnaSP v6.0 with default parameters[63], and a haplotype distribution map was constructed using ArcGIS v10.2. Haplotype diversity (H_d) and nucleotide diversity (π) were calculated with Arlequin v3.5[64]. Total gene diversity (H_T), within-population gene diversity (H_S), and population differentiation indices (G_{ST} and N_{ST}) were calculated using PERMUT v2.0[65]. The Median Joining model of NETWORK v10.2.0.0[66] was used to construct the haplotype network. BARRIER v2.2 was used to detect biogeographic boundaries evaluated by 100 replicates of population average pairwise difference matrices[67]. Analysis of molecular variance (AMOVA) with 1,000 permutations was performed to examine genetic variation among and within populations using Arlequin v3.5[64].

BEAST v2.5 was used to estimate chloroplast haplotype divergence times under a log-normal relaxed clock[68]. We chose *Juglans regia* as an outgroup[69]. On the basis of the Akaike information criterion (AIC) implemented in Modeltest v3.7[70], the HKY + I + G model was selected as the best alternative model. The age of the earliest conclusive *Juglans* L. fossil was used as the minimum age to constrain the stem of the

Table 1. Sample codes, sample locations, and genetic diversity indices for 18 populations of *Pterocarya hupehensis* based on RAD-seq data and cpDNA data.

Code	Site	Longitude (°E)	Latitude (°N)	RAD-SNPs					cpDNA		
				n	H_O	H_E	π	F_{is}	h	$\pi \times 10^3$	Haplotype (number of individuals)
HH	Xian, Shaanxi	107.93	33.87	6	0.199	0.173	0.203	0.007	0	0	H1 (10)
HXC	Ankang, Shaanxi	109.49	32.72	3	0.207	0.165	0.209	0.004	0.42	0.005	H2 (1), H3 (1), H4 (7)
FLC	Ankang, Shaanxi	109.4	31.95	5	0.192	0.143	0.174	−0.033	0	0	H5 (9)
DSP	Baoji, Shaanxi	107.49	33.84	4	0.183	0.144	0.177	−0.013	0	0	H6 (8)
QDZ	Nanyang, Henan	111.97	33.52	6	0.210	0.175	0.197	−0.027	0.39	0.004	H7 (7), H8 (2)
HKC	Nanyang, Henan	112.02	33.57	6	0.193	0.184	0.209	0.033	0.5	0	H7 (6), H9 (3)
TSG	Nanyang, Henan	111.72	33.63	6	0.182	0.171	0.198	0.031	0.33	0	H7 (5), H10 (1)
LJL	Nanyang, Henan	111.70	33.63	4	0.191	0.164	0.207	0.027	0	0	H7 (4)
LYG	Tianshui, Gansu	106.1	34.23	5	0.189	0.152	0.177	0.021	0	0	H1 (10)
YPC	Longnan, Gansu	106.23	33.67	5	0.206	0.149	0.170	0.068	0	0	H11 (7)
HJG	Kangxian, Gansu	105.51	33.39	5	0.184	0.141	0.162	0.041	0	0	H12 (7)
MYG	Yangba, Gansu	105.74	33.03	6	0.191	0.170	0.196	0.010	0	0	H13 (6)
SNJ	Shennongjia, Hubei	110.92	31.65	7	0.216	0.199	0.217	0.004	0	0	H14 (7)
XJZ	Shennongjia, Hubei	110.58	31.59	11	0.209	0.209	0.220	0.033	0.47	0.067	H15 (7), H16 (3)
SHJZ	Enshi, Hubei	109.8	30.16	11	0.206	0.179	0.194	−0.026	0.76	0.067	H17 (5), H18 (1), H19 (2), H20 (1), H21 (1)
JSZ	Nanchuan, Chongqing	107.14	29.02	7	0.213	0.212	0.223	0.029	—		—
DFX	Bijie, Guizhou	105.88	27.33	12	0.224	0.212	0.222	−0.001	0.36	0.004	H22 (8), H23 (2)
NYX	Bijie, Guizhou	105.47	26.7	13	0.227	0.195	0.204	−0.047	0	0	H24 (10)
AVERAGE					0.201	0.174	0.198	−0.006	0.19	0.009	

n, the number of samples used for RAD-seq analysis; H_O, observed heterozygosity; H_E, expected heterozygosity; π, nucleotide diversity; F_{is}, inbreeding coefficient; h, haplotype diversity.

haplotype tree (~50 Mya: 44.4–57.88 Mya)[71]. We also used *Pterocarya fraxinifolia* and *Pterocarya macroptera* as outgroups, and the age of the earliest *Pterocarya* fossil was used to constrain the crown group (~30 Mya: 28–34 Mya)[51]. Markov chain Monte Carlo runs were performed for 10 million iterations with parameter sampling every 1,000 generations. Convergence was assessed using Tracer v1.7[72], and the effective sample size (ESS) for all parameters was calculated. The first 20% of each run was discarded as burn-in using TreeAnnotator v1.8 (http://beast.bio.ed.ac.uk/TreeAnnotator). Mismatch distribution analysis and two neutrality tests, Tajima's D[73] and Fu's F_S[74], were performed to estimate historical demographic expansions using Arlequin v3.5[64].

RAD-seq data processing and analysis

SNP calling and filtering

The raw reads from all 122 individuals were cleaned to remove reads with uncalled bases and low quality scores using the process_radtags module in Stacks v2.62[75]. The process_radtags module was also used to truncate the final reads to 120 bp. The *Pterocarya stenoptera*[76] reference genome was indexed with BWA v0.7.17[77], and the *P. hupehensis* RAD sequences were aligned to the indexed reference genome. The 'flagstat' command in SAMtools v1.16 was used to calculate the mapping rates and read numbers[78]. The Stacks v2.62 pipeline was used to process the RAD-seq reads[75]. The gstacks module was used to identify single nucleotide polymorphisms (SNPs) at each locus in the population and genotype each individual for each identified SNP. The resulting BAM files were sorted with SAMtools[78]. The populations module in Stacks was used for data filtering and SNP calling with the following criteria: (1) greater than 80% of individuals in each population were processed for each locus using the parameter 'r = 0.8'; (2) the maximum observed heterozygosity was set to 0.7 with the parameter 'max-obs-het = 0.7'; (3) the minimum minor allele frequency (MAF) was set to 0.05; and (4) only the first SNP locus

of each read was retained to avoid physical linkage. The variant dataset was then filtered for missing data using VCFtools v0.1.16[79] with the parameter 'max-missing = 0.8'.

Population structure and genetic diversity

Bayesian clustering was performed using Admixture[80]. The most probable values of K for explaining population structure were determined using the lowest cross-validation (CV) error rate. R v4.1.0[81] was used to visualize the curve of the CV error rate from one to ten and the population structure histogram. Population structure was also investigated using principal component analysis (PCA) for 122 individuals with the R package 'adegenet'[82]. The optimal number of lineages was selected on the basis of the lowest associated Bayesian information criterion. An individual-based maximum likelihood (ML) tree was constructed using IQ-TREE v1.6.12[83] and contained three outgroups: *P. macroptera*, *Cyclocarya paliurus*, and *Juglans mandshurica*[45]. The nucleotide diversity (π), the expected and observed heterozygosities (H_E and H_O), and the fixation index (F_{is}) among populations were calculated using the populations module in Stacks. AMOVA was performed using Arlequin v3.5[64] to explore the degree of genetic differentiation among lineages, populations within lineages, and populations.

Population demographic histories

We used TreeMix v1.13[84] to infer possible hybridization events among populations by obtaining allele frequencies from multiple populations and generating a ML tree. Migration events were analyzed from one to ten and then calibrated on the ML tree. The parameter '-noss' was used to prevent overcorrection. We calculated the percent variance explained in order to judge the migration events using the script 'treemixVarianceExplained.R'. We set the program to use 10 migration events for generation of the ML tree. The standard errors of all entries in the covariance matrix estimated from the data were used to construct a heatmap. The migration tree and heat map were visualized using R v4.1.0.

Fluctuations in effective population size were inferred using Stairway Plot 2[85], which implements an unsupervised learning strategy for model selection and supports both folded and unfolded site frequency spectra (SFS). The *P. stenoptera* reference genome was indexed, and unfolded SFSs were generated for both genetic groups and total populations using ANGSD v0.939[86]. The effective population size was inferred for each lineage and for all populations using a mutation rate of 2.06×10^{-9} per locus per year (with reference to *Juglans*)[87] and a generation time of 30 years. We used the recommended percentage of training sites (67%) to run the Stairway Plot 2 program. By default, 200 input files were created for each estimation.

BEAST v2.5[68] was used to estimate the nuclear genome (RAD-seq) divergence time for *P. hupehensis*. According to the phylogeny of Juglandaceae, *Cyclocarya paliurus* and *Pterocarya macroptera* were used as outgroups[71]. We used the oldest fossil of *Cyclocarya* to calibrate the stem age of *Pterocarya* (65–55 Mya)[51] and the oldest fossil of *Pterocarya* to calibrate the crown age of *Pterocarya* (34–28 Mya)[51]. MCMC chains were run for 50,000,000 generations under the GTR + I + G model chosen by jModelTest[88]. Tracer v1.7 was used to evaluate convergence and calculate the ESS (http://tree.bio.ed.ac.uk/software/tracer/). The first 30% of samples were discarded as burn-in using TreeAnnotator v1.8, and the phylogeny was visualized using Figtree v1.4.4 (http://tree.bio.ed.ac.uk/software/Figtree/).

Ecological niche modeling

The potential distribution areas of *P. hupehensis* were generated on the basis of all known distribution points using M$_{AXENT}$ v3.4.4[89]. The distribution points included the 18 sampling points used here, as well as distribution records from the National Specimen Information Infrastructure (www.nsii.org.cn/) and the Chinese Virtual Herbarium (www.cvh.ac.cn). We deleted points corresponding to non-natural populations (parks, urban areas, etc.) using a coordinate backcheck. The R package 'dismo'[90] was used to delete missing, duplicate, and incorrect coordinates. To reduce the influence of spatial autocorrelation on climate variables, we performed grid screening to remove coordinate points less than 2.5 arc-min (~4.6 km) apart using the R package 'raster'[91]. We acquired the standard 19 bioclimatic variables from the WorldClim website (https://worldclim.org/) at a 2.5-arc-min spatial resolution[92] for four time periods: the present (1970–2000), the last glacial maximum (LGM, *c.* 21 kya)[93], the last interglacial period (LIG, *c.* 120 kya)[94], and the 2061–2080 period under the representative concentration pathway (RCP) 4.5 scenario. We used the function variance inflation factor (VIF) from the R package 'usdm'[95] to remove environmental factors with correlation coefficients > 0.8. The following environmental factors were retained as predictors: annual mean temperature, mean diurnal range, temperature seasonality, precipitation of wettest month, precipitation seasonality, and precipitation of driest quarter. In total, 25% of the data were used for model testing and validation. Ten independent replicates were analyzed using the bootstrap method. The average results were used to reclassify the suitable areas in ArcGIS v10.2. The parameter for reclassification had five levels, 0–0.08, 0.08–0.25, 0.25–0.5, 0.5–0.75, and 0.75–1, indicating the degree of habitat suitability from low to high.

Results

cpDNA sequence variation

The alignments of the *psb*D-*trn*T, *trn*V-*ndh*C, *trn*L-*rpL*32, *trn*S-*trn*fM, *trn*G-*trn*S, and *trn*D-*trn*Y sequences were 1,578, 709, 849, 1305, 1,524, and 780 bp in length, respectively. After concatenation, the resulting 6,745 bp of chloroplast DNA sequence was found to contain 91 polymorphisms and 24 haplotypes in the 17 examined populations (Table 1, Fig. 1). The haplotype diversity (H_d) and nucleotide diversity (π) of *P. hupehensis* were 0.929 and 0.00379, respectively. Population SHJZ had the highest haplotype diversity ($h = 0.76$), and populations XJZ and SHJZ had the highest nucleotide diversity ($\pi = 0.0067$) (Table 1). Total diversity ($H_T = 0.963$) was much higher than the average within-population diversity ($H_S = 0.189$). Significant phylogeographic structure was detected among the populations ($G_{ST} = 0.804 < N_{ST} = 0.969$, $P < 0.05$). The BARRIER analysis revealed genetic barriers between the eastern and western lineages (near the Qinling Mountains and the southern 105°E line) and also between the northern and southern Yangtze River (Supplemental Fig. S1). AMOVA showed that 75.05% of the total genetic variation occurred between the two lineages ($F_{CT} = 0.75$), 22.64% occurred among populations within lineages ($F_{SC} = 0.91$), and only 2.31% occurred within populations ($F_{ST} = 0.98$; Table 2).

The phylogenetic network resolved two main haplotype lineages with 30 step mutations (western and eastern lineages) located in the eastern and western parts of the Sichuan Basin. The eastern lineage was further divided into two clades (clades II and III). Clade II included haplotypes from the eastern Sichuan Basin and the western Qinling population (YPC). Clade III was composed mainly of haplotypes from the southeastern Sichuan Basin (Fig. 1). The haplotype structures revealed by BEAST were similar to the groupings revealed by the phylogenetic network. The haplotype divergence between the western and eastern lineages dated to the middle Miocene (16.7 Mya), and clade II separated from clade III at approximately 8.5 Mya. The crown age of clade II was estimated at 5.3 Mya and that of clade III at 5.8 Mya (Fig. 2).

The values of Tajima's *D* and Fu's *F*s were positive for *P. hupehensis* ($D = 0.675$, $P = 0.693$; $F_S = 0.196$, $P = 0.626$), suggesting no expansion of its distribution. Mismatch distribution analysis showed multimodal distributions for all samples, also suggesting that this species has not undergone a recent demographic expansion (Supplemental Fig. S2).

Population structure and genetic diversity

We obtained 87 million RAD-seq reads from 122 samples. The average alignment rates of the samples to the reference genome was 86.72% (Supplemental Table S2). After restricting variants to those processed in > 80% of the individuals in each population, 695,262 variant sites remained. After filtering, 2,427 and 2,889 SNPs with and without outgroups, respectively, were obtained for subsequent analyses.

Admixture analysis revealed that the genetic structure of *P. hupehensis* consisted of two lineages (Fig. 3; Supplemental Fig. S3). Populations located in the northwestern Qinling Mountains, together with one population (NYX) from the southwest, were assigned to the western lineage. The eastern lineage included all populations from the eastern Sichuan Basin, as well as one population from the northwest (YPC) (Fig. 3a). Genetic introgression was detected in five populations (western
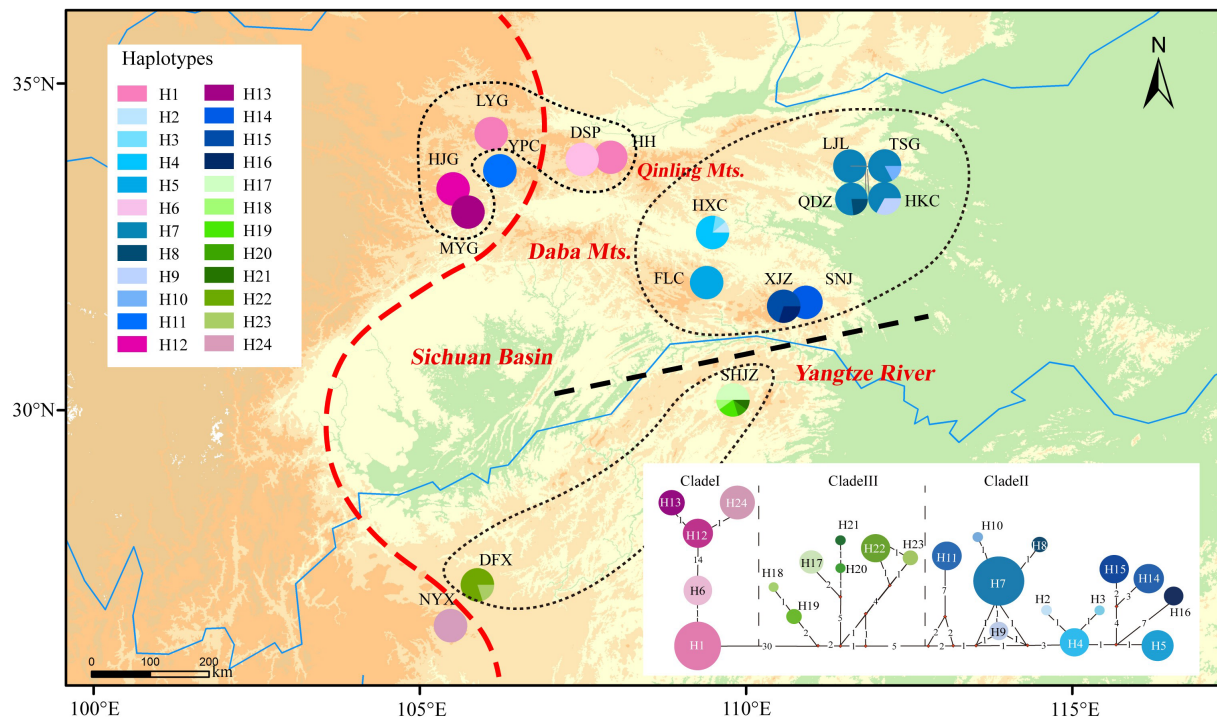
Fig. 1 Distribution of 24 chloroplast haplotypes of *Pterocarya hupehensis*. The lower right panel shows the minimum spanning network of 24 chlorotypes. Circle sizes are proportional to the number of samples per haplotype. The red dashed line on the map represents the Sino-Himalayan/Sino-Japanese forest boundary. The black dashed line represents the phylogeographic break along the Yangtze River in the eastern Sichuan Basin. The black dotted lines delineate three phylogroups with closely related chlorotypes.

Table 2. Analyses of molecular variance (AMOVA) based on cpDNA data and RAD-seq data for *Pterocarya hupehensis* from western and eastern lineages.

Source of variation	cpDNA			RAD-SNPs		
	d.f.	Percentage variation (%)	Fixation indices	d.f.	Percentage variation (%)	Fixation indices
Among lineages	1	75.05	$F_{CT} = 0.75$	1	22.22	$F_{CT} = 0.22$
Among populations within lineages	15	22.64	$F_{SC} = 0.91$	16	11.47	$F_{SC} = 0.15$
Within populations	124	2.31	$F_{ST} = 0.98$	226	66.31	$F_{ST} = 0.34$

lineages LYG, MYG, DSP, and HH and eastern lineage FLC) (Fig. 3a, c). The core populations of the eastern lineage always clustered together when $K = 2$ to 5 (Supplemental Fig. S4). PCA also divided all populations into two lineages. The percentages of variation explained by PC1 and PC2 were 9.2% and 3.9%, respectively (Fig. 3). A ML tree showed that the eastern lineage evolved earlier than the western lineage (Supplemental Fig. S5), and the hybrid populations were primarily located at the branch ends of the ML tree.

Genetic diversity of nuclear DNA varied among populations as assessed by H_O (0.182–0.227), H_E (0.141–0.212), π (0.162–0.223), and F_{is} (−0.047–0.068) (Table 1). Genetic diversity was higher in the southern populations, including NYX, DFX, and JSZ. AMOVA revealed that genetic differentiation occurred mainly within populations ($F_{ST} = 0.34$, 66.31%; Table 2).

Population demographic histories

The migration events in the ML tree showed two strong signals with a high migration weight, indicating unidirectional gene flow from the DFX population to the common ancestral populations of NYX and HJG and also from the HJG to MYG populations (Fig. 4a). Eight other gene flow signals were also detected. The longest horizontal branch was that of the MYG population, indicating that it had undergone the greatest genetic drift compared with the other populations (Fig. 4a, Supplemental Table S3). The topology of the genetic relationships in the ML tree was consistent with the results of Admixture and IQ-TREE analyses. Stronger introgression events among populations were shown by the residual heatmap than by the ML tree inferred with Treemix (especially in the SHJZ, YPC, DFX, and QDZ populations) (Fig. 4a, b). The BEAST result based on the phylogeny from the RAD-seq datasets showed that the western lineage separated from the eastern lineage at 16.78 Mya (95% HPD: 11.99–22.25 Mya) (Supplemental Fig. S6), very similar to the divergence time of haplotypes between the western and eastern lineages estimated from cpDNA data (16.7 Mya, 95% HPD: 11.8–21.9 Mya).

Fluctuations in the effective population size of *P. hupehensis* were estimated to have occurred from approximately 1.0 Mya in the Late Pleistocene (Fig. 4c). Stairway Plot 2 analysis revealed a genetic bottleneck event at about 200–400 kya in which the effective population size dropped to about 1.4×10^4 to 2×10^3 individuals. The effective population size then climbed to an upper limit of $\sim 3.2 \times 10^4$ individuals at about 200 kya and remained steady between 120–140 kya during the LIG. From the LGM period to the Holocene, effective population size gradually decreased to its lowest level of 0.1×10^3 individuals (Fig. 4c). The western and eastern lineages exhibited similar trends in population size fluctuation. However, the bottleneck event occurred slightly earlier in the western lineage than in the eastern lineage (Supplemental Fig. S7a, b).
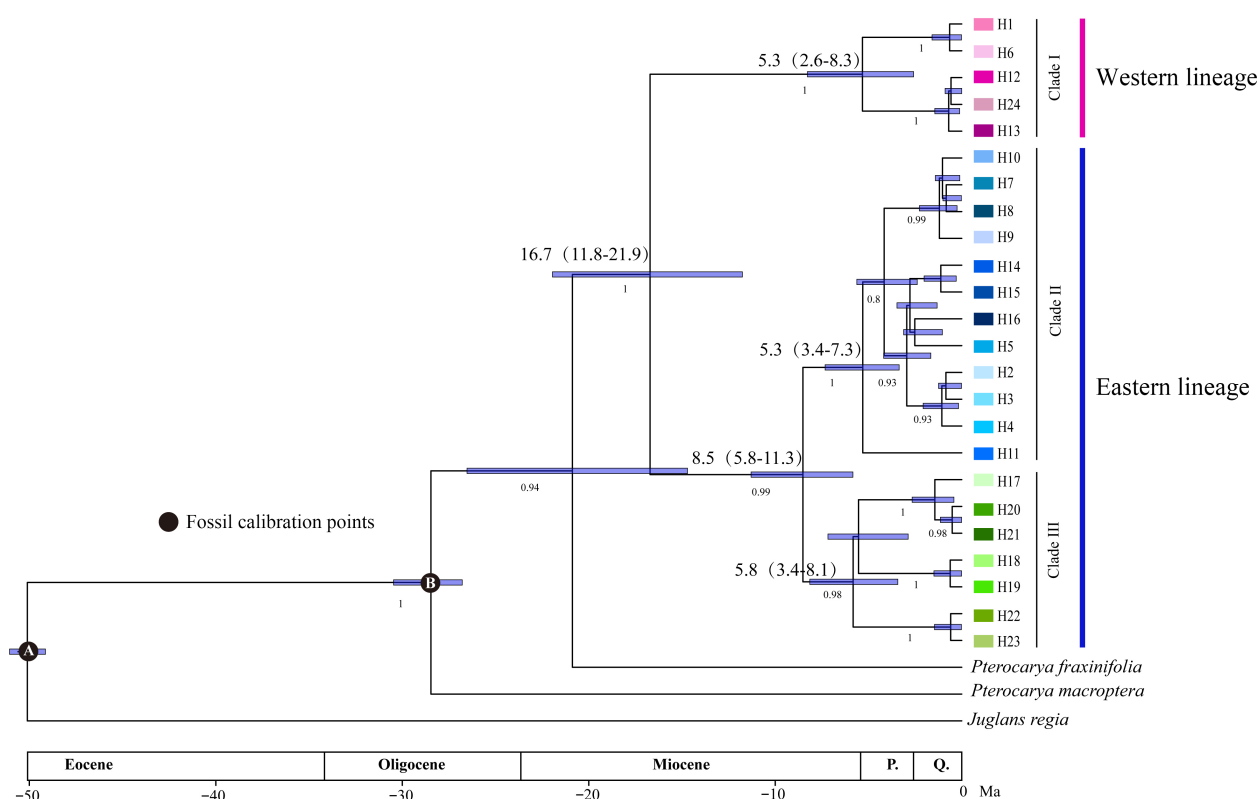
Fig. 2 BEAST-derived chronogram of 24 *Pterocarya hupehensis* haplotypes based on six chloroplast DNA (cpDNA) fragments. The divergence time are shown above branches, with blue bars indicating the 95% highest posterior densities (HPDs). Posterior probabilities are labeled at each node.

Ecological niche modeling

The Area Under Curve (AUC) value of the receiver operating characteristic (ROC) curve was high (> 0.977) in the four periods (Supplemental Fig. S8a–d). Annual mean temperature (35%) and mean diurnal range (26%) made the greatest contributions to the model under the current climate (Supplemental Fig. S9, Supplemental Table S4). The potential distribution of *P. hupehensis* under the present climate matched its current distribution; suitable area for its growth was limited to the mountains around the Sichuan Basin, especially in the northern and eastern regions (Fig. 5a). During the LGM period, the suitable area was significantly smaller in the southern and western regions of the Sichuan Basin but greater in the Qinling Mountains (Fig. 5b). The suitable area for *P. hupehensis* around the Sichuan Basin was greatest during the LIG period (Fig. 5c). Notably, during this period, an area of high suitability appeared in the Hengduan and Daliang Mountains. Under the RCP 4.5 scenario for 2061–2080, the suitable area for *P. hupehensis* was predicted to be just slightly smaller than at present, and the suitable areas in the western and southern Sichuan Basin were predicted to shrink under global warming (Fig. 5d).

Discussion

Phylogeographic conflicts between chloroplast and nuclear DNA

Significant chloroplast genetic structure

Pterocarya hupehensis, like other relict tree species in subtropical China, shows strong phylogeographic structure (N_{ST} > G_{ST}) based on cpDNA markers[5,41,42]. Most of the haplotypes are

private, except for H1 and H7, which are shared among populations. As reported for other angiosperm species, genetic differentiation among lineages of *P. hupehensis* was higher when assessed with cpDNA (F_{CT} = 0.75) than with nuclear DNA (F_{CT} = 0.22). Such significant chloroplast genetic structure is typically attributed to maternal inheritance, as chloroplasts are only transmitted by seeds, which have a limited dispersal distance. Seed-mediated gene flow can be influenced by seed dispersal ability, germination and dormancy, seedling establishment ability, and so forth. The wingnuts of *P. hupehensis* can be carried over short distances by wind and dispersed over long distances along rivers. *P. hupehensis*, as a relict species, is present in restricted microhabitats and shows strong niche conservatism (moist riparian forests along rivers with an elevational range of 700–2,000 m in the mountains)[45,96–99]. It produces relatively few seedlings, even when abundant seeds are present, owing to low seed quality and difficult seedling establishment[51,52]. Taken together, these factors have impeded the exchange of chloroplast genes among populations and shaped current genetic patterns.

Different nuclear genetic structure

Optimal clustering based on nuclear genetic structure (from RAD-seq data) divided *P. hupehensis* into two genetic clades in both Admixture analysis and PCA. More subdivided genetic structures were recognized in the Admixture analyses when K values were larger. Most populations in the eastern lineage were maintained in a single group from K = 2 to K = 5, indicating extensive gene flow *via* pollen dispersal. However, the numerous and private chloroplast haplotypes in the eastern lineage suggested their independent evolution (Fig. 1,
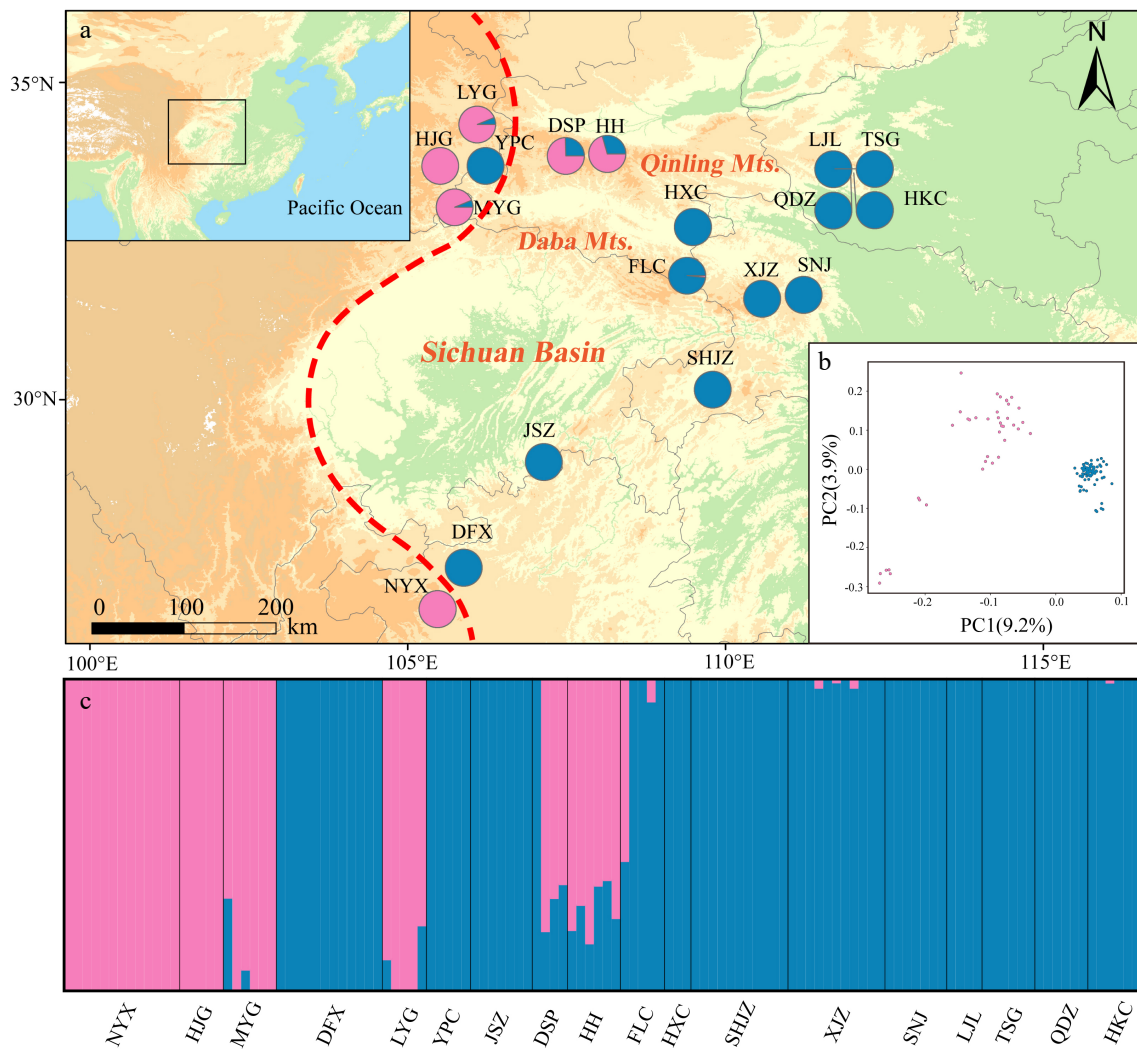
Fig. 3 Genetic structure of *Pterocarya hupehensis* based on 2,889 SNPs dataset. (a) Geographic origins of 18 *P. hupehensis* populations and their color-coded grouping at close to $K = 2$. The red dashed line represents the Sino-Himalayan/Sino-Japanese forest boundary. (b) Principal component analysis (PCA), with pink and blue colors representing two clusters. (c) Histogram of the Admixture analysis for *P. hupehensis* with $K = 2$.

Supplemental Fig. S3). Genetic differentiation among all populations was higher for cpDNA ($F_{ST} = 0.98$) than for nuclear DNA ($F_{ST} = 0.34$). Compared with maternal inheritance of cpDNA, parental inheritance of nuclear DNA typically exhibits more imprints from pollen flow. Thus, the inconsistency in genetic structure between cpDNA and nuclear DNA (especially for wind-pollinated species) can be understood as arising from differences between seed-mediated and pollen-mediated gene flows[5,100,101]. The present study provides another example of a temperate tree species with stronger genetic structure in the chloroplast genome (seed-mediated gene flow) than in the nuclear genome (pollen-mediated)[42,102,103]. Gene flows mediated by long-distance pollen dispersal were detected here and have been demonstrated in many other anemophilous tree species, such as *P. fraxinifolia*, *Quercus robur*, *Zelkova carpinifolia*, and others[104–106]. The strong East Asian monsoon that began in the early Miocene may have enabled the spread of pollen over long distances and thus promoted gene flow among populations[21].

Extensive pollen flows of anemophilous tree species have been reported to facilitate genetic exchange and delay genetic differentiation in species with restricted distributions[5]. Such genetic exchange can improve population adaptation, particularly for tree species with slow evolutionary rates, high pollen dispersal capacity, and weak reproductive ability[107,108]. Compared with other wind-pollinated species, *P. hupehensis* has a relatively high level of genetic differentiation[109,110], which may reflect the influence of slower pollen-mediated gene flow, higher levels of genetic drift, and local adaptation due to selection pressure associated with long-term environmental heterogeneity[109–111]. The bottleneck event and small population sizes of *P. hupehensis* may have led to high levels of genetic drift. In addition, long-established small and isolated populations of *P. hupehensis* are likely to have experienced more environmental selection pressure, which may also have contributed to a high level of genetic differentiation[112,113].

Species differentiation in the early to middle Miocene and later diversification

Colonization occurs during the wet and rainy monsoon

Reconstructions of divergence times based on cpDNA and nuclear DNA revealed that the eastern and western lineages of
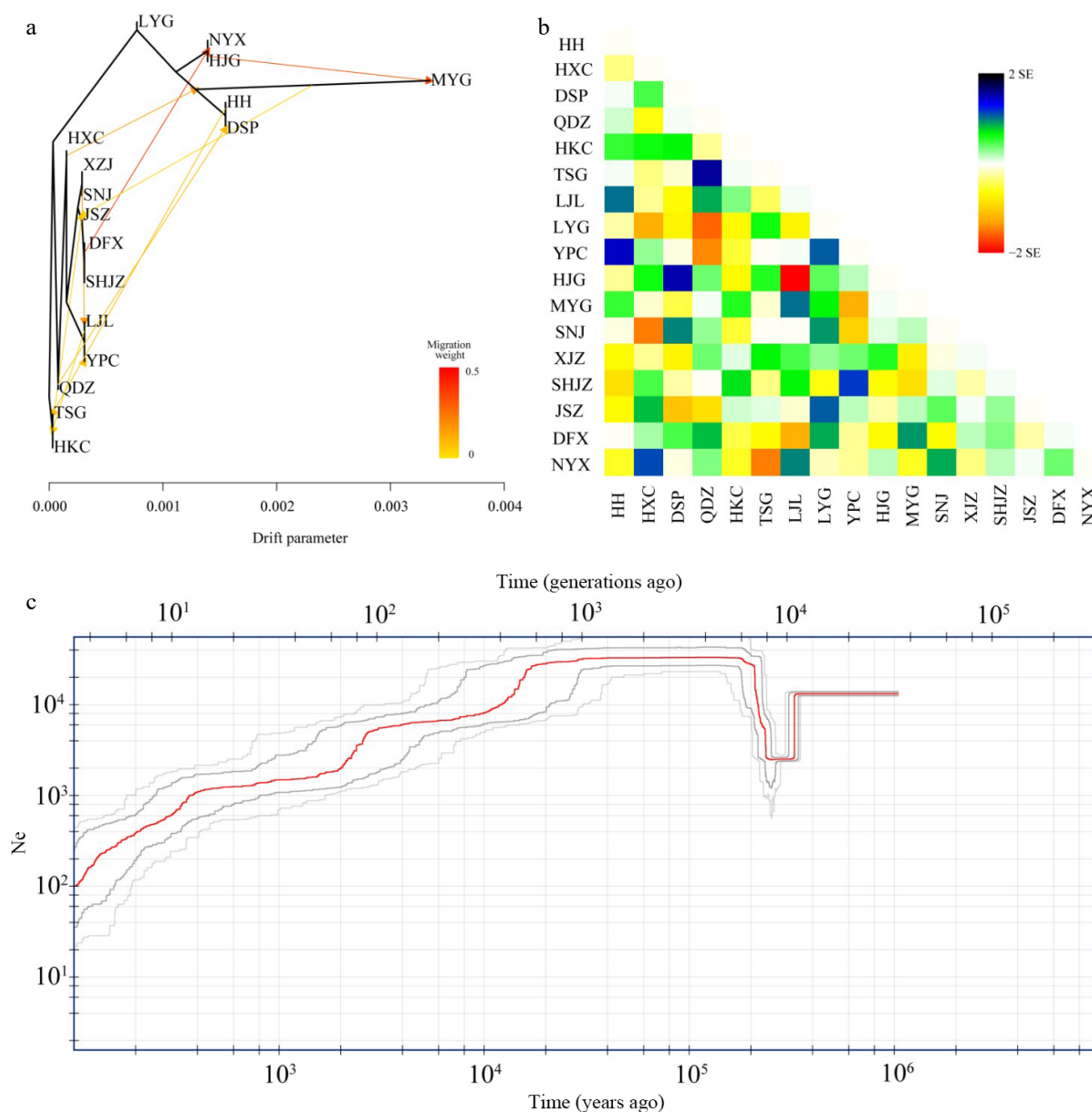
Fig. 4 Hybridization among populations of *Pterocarya hupehensis*. (a) Maximum likelihood (ML) tree inferred using Treemix; gene-flow events are depicted with arrows, and ten migration events were allowed. Migration arrows are colored according to their weight. Horizontal branch lengths are proportional to the amount of genetic drift that occurred on each branch. The scale bar shows 10× the average standard error of the entries in the sample covariance matrix. (b) Heatmaps of residual fit from the ML tree. Residuals white through blue indicate that the corresponding populations are more closely related to each other than on the ML tree, suggesting confounding events between these populations. (c) Demographic history of *P. hupehensis* inferred with Stairway Plot 2 using unfolded site frequency spectra. The 95% confidence interval for estimated effective population size is shown with gray lines.

P. hupehensis diverged during the early to middle Miocene. A similar pattern was reported for *Cyclocarya paliurus*, which also belongs to the Juglandaceae[18]. By contrast, most relict tree species in this area diverged during the Pliocene, including *Davidia involucrata* (4.81 Mya), *Euptelea pleiosperma* (3.64 Mya), and *Populus lasiocarpa* (3.66 Mya)[6,13,42]. Initial intensifications of the East Asian summer and winter monsoons began in the early Miocene owing to the rapid uplift of the Tibetan Plateau[20,28]. Abundant precipitation associated with the monsoons, together with subsequent cooling during the mid-to-late Miocene and early Pliocene, promoted speciation and lineage differentiation of plants in East Asia[16,114,115]. The SJFR

contains a rich diversity of temperate flora, which benefited from the changes in precipitation pattern and incomplete glacial coverage of the Quaternary glaciation[10, 116–118]. Our species distribution modeling provides further evidence that temperature and precipitation are the most important climatic predictors of suitable habitat for *P. hupehensis*. Both *P. hupehensis* and *C. paliurus* inhabit wet habitats near riverbanks or streams with high-humidity microclimates. Previous research has suggested that the characteristic of naked buds on temperate trees, as exhibited by *P. hupehensis*, may be associated with colder temperatures and summer precipitation[18,119].
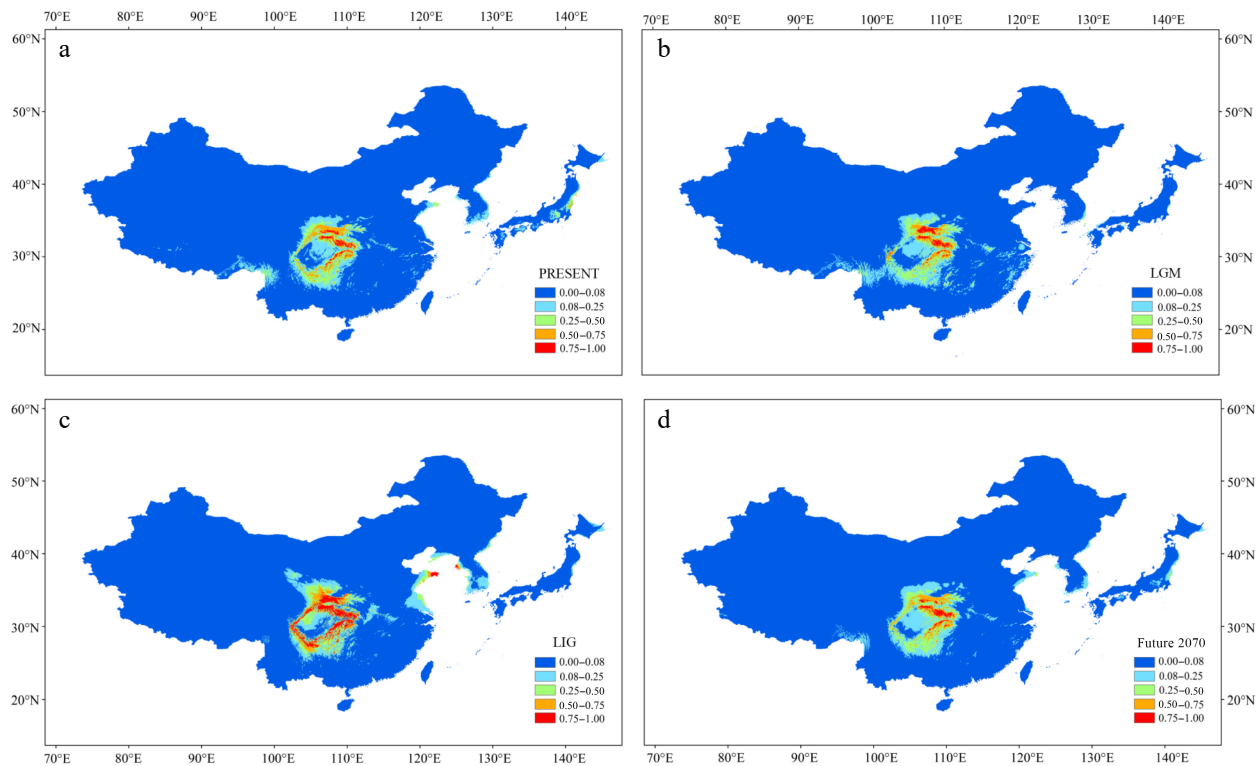
Fig. 5 Results of ecological niche modeling of *P. hupehensis* in four time periods from past to future. (a) Average projection of the model to present climatic conditions. (b) Average projection of the model to the last glacial maximum (LGM: *c.* 21 kya BP (before present)). (c) Average projection of the model for the last interglacial (LIG: *c.* 120–140kya BP). (d) Average projection of the model to the year 2070 (2061–2080) under an intermediate climate warming scenario (RCP 4.5). Colors from blue to red represent the degree of habitat suitability for *P. hupehensis* survival, from unsuitable to suitable.

Multiple driving forces for lineage differentiation

The Sichuan Basin acts as a geographic barrier between the Sino-Himalayan and Sino-Japanese Forest subkingdoms (more or less along the 105°E line)[120,121], affecting patterns of genetic diversity and structure for many relict species (e.g., *Davidia involucrata*[13], *Dysosma versipellis*[17], and *Primula ovalifolia*[122]). Both chloroplast and nuclear evidence demonstrate that *P. hupehensis* has also been influenced by this geographic barrier. The NYX and DFX populations provide a clear illustration of this barrier, as they are close geographically but contain chloroplast haplotypes and nuclear genes from different lineages. The western and eastern lineages of *P. hupehensis* appear to have diverged at the end of the early Miocene, which was followed by intensification of the EAMS. A previous study also detected a monsoon-driven phylogeographic break between western and eastern lineages of relict species around the Sichuan Basin[16].

North–south lineage divergence in the Three Gorges region of the east Sichuan Basin has been documented in plants[14,31,44] and animals[121]. Here, chloroplast haplotypes in the eastern lineage of *P. hupehensis* were further divided into northern and southern clades by the Yangtze River in the Three Gorges region. The barrier of the Three Gorges region blocks gene flow by limiting seed dispersal and animal migration. The absence of this phylogeographic break in nuclear gene analyses can be attributed to long-distance pollen dispersal by wind, and the results presented here for *P. hupehensis* are a good example of this phylogeographic inconformity.

The origin and formation of the Yangtze River and the Sichuan Basin were synchronized with the uplift of the Tibetan Plateau[123], although details of the age and developmental history of the Yangtze River have been vigorously debated for more than 100 years[124]. Our results suggest that the Three Gorges barrier to plant gene flow may be traced back to the late Miocene, 5 Mya before the speculated formation time of the Yangtze River[125]. Thus, the phylogeographic break for some relict trees in the Three Gorges region may have occurred in the late Miocene. Overall, our data suggest that significant changes in climate and geography around the Sichuan Basin promoted phylogeographic breaks between the western and eastern lineages, whereafter in the eastern lineage of *P. hupehensis*.

Population demographic history after the Pleistocene

Climatic fluctuations in the Pleistocene glacial and interglacial periods had substantial effects on many relict plants. *P. hupehensis* experienced a population expansion during the warm LIG period and a slight shrinkage during the LGM period. However, we estimated relatively little fluctuation in its distribution around the Sichuan Basin from the Pleistocene to the future, consistent with findings for some other species[6,121]. Because the micro-environment of the mountains around the Sichuan Basin may have been more stable than that of other regions, East Asia acted as a biodiversity sanctuary during the LGM period. There was a brief cooling period at 120–350 kya (penultimate [Riss] glacial period) before the longest period of Pleistocene warmth[126], and we detected a bottleneck event during this glacial period (200–400 kya). We speculate that temperature may have had an important effect on the distribution of *P. hupehensis*. The bottleneck effect can cause a series of

negative chain reactions, especially for small populations, resulting in a loss of genetic diversity[127,128], high levels of population isolation[129], and altered fitness because of genetic drift and inbreeding[130]. The occurrence of bottlenecks or adversity is likely to lead to high levels of genetic drift[127,131].

Our field investigations revealed that the current population of *P. hupehensis* is small and fragmented, with only a few dozen individuals in some populations[52]. *P. hupehensis* is currently listed as vulnerable on the IUCN Red List of Threatened Species. Our findings highlight the risk of a gradual decline in effective population size in the event of renewed adversity (Fig. 4c). Thus, future work should aim to assess the genomic vulnerability of each population, and both *ex situ* and *in situ* conservation of these small populations should be improved[111]. We should focus not only on the effect of global warming and greenhouse gas emissions on this species but also on the interference of human activities with its natural habitat.

Conclusions

We used cpDNA and nuclear DNA data to reconstruct the phylogeographic history of *P. hupehensis*. Both cpDNA and nuclear genetic data revealed two distinct lineages corresponding to two phylogeographic regions. However, the cpDNA data suggest a relatively isolated and stronger phylogeographic structure than the nuclear data. This result suggests that pollen flow plays a more important role than seed flow in shaping genetic structure. External geologic and climatic changes have also influenced current genetic distribution patterns. Strengthening of the EAMS during the early to middle Miocene appears to have been the main driver of colonization and differentiation in *P. hupehensis*. The Three Gorges region, which acts as a seed dispersal barrier, promoted further north–south differentiation among the eastern lineages. In addition to population genetics studies and modeling, more efforts should be directed toward searching for *P. hupehensis* populations in the western Sichuan Basin. Genetic resources of *P. hupehensis* from the western and southwestern Sichuan Basin should be given priority.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Song YG, Wang TR, Cao JG; data collection: Song YG; analysis and interpretation of results: Lu ZJ, Zheng SS; draft manuscript preparation: Lu ZJ, Song YG, Wang TR, Meng HH, Cao JG, Kozlowski G. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The haplotype sequences of chloroplast DNA and the RAD-seq data have been deposited at the National Center for Biotechnology Information (NCBI) with GenBank accession numbers QQ884193–QQ884252 and PRJNA967132.

Acknowledgments

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

References

1. Hewitt GM. 2004. The structure of biodiversity – insights from molecular phylogeography. *Frontiers in Zoology* 1:4

2. Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B* 359:183−95

3. Qiu YX, Fu CX, Comes HP. 2011. Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. *Molecular Phylogenetics and Evolution* 59:225−44

4. Liepelt S, Bialozyt R, Ziegenhagen B. 2002. Wind-dispersed pollen mediates postglacial gene flow among refugia. *Proceedings of the National Academy of Sciences of the United States of America* 99:14590−94

5. Bai WN, Wang WT, Zhang DY. 2014. Contrasts between the phylogeographic patterns of chloroplast and nuclear DNA highlight a role for pollen-mediated gene flow in preventing population divergence in an East Asian temperate tree. *Molecular Phylogenetics and Evolution* 81:37−48

6. Li X, Ruhsam M, Wang Y, Zhang HY, Fan XY, et al. 2023. Wind-dispersed seeds blur phylogeographic breaks: the complex evolutionary history of *Populus lasiocarpa* around the Sichuan Basin. *Plant Diversity* 45:156−68

7. Usinowicz J, Chang-Yang CH, Chen YY, Clark JS, Fletcher C, et al. 2017. Temporal coexistence mechanisms contribute to the latitudinal gradient in forest diversity. *Nature* 550:105−08

8. Wu Z, Wu S. 1998. A proposal for a new floristic Kingdom (realm)- the E. Asiatic kingdom, its delineation and characteristic. In *Floristic Characteristic and Diversity of East Asian Plants*, eds. Zhang A, Wu S. China: China Higher Education Press. pp. 3–42.

9. Manchester SR, Chen ZD, Lu AM, Uemura K. 2009. Eastern Asian endemic seed plant genera and their paleogeographic history throughout the Northern Hemisphere. *Journal of Systematics and Evolution* 47:1−42

10. Tiffney BH. 1985. Perspectives on the origin of the floristic similarity between Eastern Asia and Eastern North America. *Journal of the Arnold Arboretum* 66:73−94

11. Liu KB. 1988. Quaternary history of the temperate forests of China. *Quaternary Science Reviews* 7:1−20

12. Wei X, Sork VL, Meng H, Jiang M. 2016. Genetic evidence for central-marginal hypothesis in a Cenozoic relict tree species across its distribution in China. *Journal of Biogeography* 43:2173−85

13. Ma Q, Du Y, Chen N, Zhang L, Li J, et al. 2015. Phylogeography of *Davidia involucrata* (Davidiaceae) inferred from cpDNA haplotypes and nSSR data. *Systematic Botany* 40:796–810

14. Sun Y, Moore MJ, Yue L, Feng T, Chu H, et al. 2014. Chloroplast phylogeography of the East Asian Arcto-Tertiary relict *Tetracentron sinense* (Trochodendraceae). *Journal of Biogeography* 41:1721–32

15. Gong W, Chen C, Dobeš C, Fu C, Koch MA. 2008. Phylogeography of a living fossil: Pleistocene glaciations forced *Ginkgo biloba* L. (Ginkgoaceae) into two refuge areas in China with limited subsequent postglacial expansion. *Molecular Phylogenetics and Evolution* 48:1094–105

16. Qiu Y, Lu Q, Zhang Y, Cao Y. 2017. Phylogeography of East Asia's Tertiary relict plants: current progress and future prospects. *Biodiversity Science* 25:136–46

17. Qiu Y, Guan B, Fu C, Comes HP. 2009. Did glacials and/or interglacials promote allopatric incipient speciation in East Asian temperate plants? Phylogeographic and coalescent analyses on refugial isolation and divergence in *Dysosma versipellis Molecular Phylogenetics and Evolution* 51:281–93

18. Kou Y, Cheng S, Tian S, Li B, Fan D, et al. 2016. The antiquity of *Cyclocarya paliurus* (Juglandaceae) provides new insights into the evolution of relict plants in subtropical China since the late Early Miocene. *Journal of Biogeography* 43:351–60

19. Wang L, Schneider H, Zhang X, Xiang Q. 2012. The rise of the Himalaya enforced the diversification of SE Asian ferns by altering the monsoon regimes. *BMC Plant Biology* 12:210

20. An Z, Kutzbach JE, Prell WL, Porter SC. 2001. Evolution of Asian monsoons and phased uplift of the Himalaya–Tibetan plateau since Late Miocene times. *Nature* 411:62–66

21. Guo ZT, Sun B, Zhang ZS, Peng SZ, Xiao GQ, et al. 2008. A major reorganization of Asian climate by the early Miocene. *Climate of the Past* 4:153–74

22. Holbourn AE, Kuhnt W, Clemens SC, Kochhann KGD, Jöhnck J, et al. 2018. Late Miocene climate cooling and intensification of southeast Asian winter monsoon. *Nature Communications* 9:1584

23. Clift P, Lee JI, Clark MK, Blusztajn J. 2002. Erosional response of South China to arc rifting and monsoonal strengthening; a record from the South China Sea. *Marine Geology* 184:207–26

24. Sun X, Wang P. 2005. How old is the Asian monsoon system?—Palaeobotanical records from China *Palaeogeography, Palaeoclimatology, Palaeoecology* 222:181–222

25. Harris N. 2006. The elevation history of the Tibetan Plateau and its implications for the Asian monsoon. *Palaeogeography, Palaeoclimatology, Palaeoecology* 241:4–15

26. Zhou Z, Huang J, Ding W. 2017. The impact of major geological events on Chinese flora. *Biodiversity Science* 25:123–35

27. Zachos J, Pagani M, Sloan L, Thomas E, Billups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292:686–93

28. Wan S, Li A, Clift PD, Stuut JBW. 2007. Development of the East Asian monsoon: mineralogical and sedimentologic records in the northern South China Sea since 20 Ma. *Palaeogeography, Palaeoclimatology, Palaeoecology* 254:561–82

29. Royden LH, Burchfiel BC, van der Hilst RD. 2008. The geological evolution of the Tibetan Plateau. *Science* 321:1054–58

30. Zhou S, Wang X, Wang J, Xu L. 2006. A preliminary study on timing of the oldest Pleistocene glaciation in Qinghai–Tibetan Plateau. *Quaternary International* 154–155:44–51

31. Zhang YH, Wang IJ, Comes HP, Peng H, Qiu YX. 2016. Contributions of historical and contemporary geographic and environmental factors to phylogeographic structure in a Tertiary relict species, *Emmenopterys henryi* (Rubiaceae). *Scientific Reports* 6:24041

32. Wang WM. 1994. Paleofloristic and paleoclimatic implications of Neogene palynofloras in China. *Review of Palaeobotany and Palynology* 82:239–50

33. Fan DM, Yue JP, Nie ZL, Li ZM, Comes HP, et al. 2013. Phylogeography of *Sophora davidii* (Leguminosae) across the 'Tanaka-Kaiyong Line', an important phytogeographic boundary in Southwest China. *Molecular Ecology* 22:4270–88

34. Ding Y. 1992. Summer monsoon rainfalls in China. *Journal of the Meteorological Society of Japan* 70:373–96

35. Yu G, Chen Z, Piao S, Peng C, Ciais P, et al. 2014. High carbon dioxide uptake by subtropical forest ecosystems in the East Asian monsoon region. *Proceedings of the National Academy of Sciences of the United States of America* 111:4910–15

36. Silva AC, Souza AF. 2018. Aridity drives plant biogeographical sub regions in the Caatinga, the largest tropical dry forest and woodland block in South America. *PLoS One* 13:e0196130

37. Sun J, Ye J, Wu W, Ni X, Bi S, et al. 2010. Late Oligocene–Miocene mid-latitude aridification and wind patterns in the Asian interior. *Geology* 38:515–18

38. Liu L, Eronen JT, Fortelius M. 2009. Significant mid-latitude aridity in the middle Miocene of East Asia. *Palaeogeography, Palaeoclimatology, Palaeoecology* 279:201–06

39. Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403:853–58

40. Luo D, Xu B, Li ZM, Sun H. 2021. Biogeographical divides delineated by the three-step landforms of China and the East China Sea: insights from the phylogeography of *Kerria japonica*. *Journal of Biogeography* 48:372–85

41. Ye JW, Bai WN, Bao L, Wang TM, Wang HF, et al. 2017. Sharp genetic discontinuity in the aridity-sensitive *Lindera obtusiloba* (Lauraceae): solid evidence supporting the Tertiary floral subdivision in East Asia. *Journal of Biogeography* 44:2082–95

42. Cao YN, Comes HP, Sakaguchi S, Chen LY, Qiu YX. 2016. Evolution of East Asia's Arcto-Tertiary relict *Euptelea* (Eupteleaceae) shaped by Late Neogene vicariance and Quaternary climate change. *BMC Evolutionary Biology* 16:66

43. Li XW, Li J. 1997. The Tanaka-Kaiyong Line-an important floristic line for the study of the flora of East Asia. *Annals of the Missouri Botanical Garden* 84:888–92

44. Qi XS, Chen C, Comes HP, Sakaguchi S, Liu YH, et al. 2012. Molecular data and ecological niche modelling reveal a highly dynamic evolutionary history of the East Asian Tertiary relict *Cercidiphyllum* (Cercidiphyllaceae). *New Phytologist* 196:617–30

45. Song YG, Li Y, Meng HH, Fragnière Y, Ge BJ, et al. 2020. Phylogeny, taxonomy, and biogeography of *Pterocarya* (Juglandaceae). *Plants* 9:1524

46. Zhang TC, Comes HP, Sun H. 2011. Chloroplast phylogeography of *Terminalia franchetii* (Combretaceae) from the eastern Sino-Himalayan region and its correlation with historical river capture events. *Molecular Phylogenetics and Evolution* 60:1–12

47. Chen SC, Zhang L, Zeng J, Shi F, Yang H, et al. 2012. Geographic variation of chloroplast DNA in *Platycarya strobilacea* (Juglandaceae). *Journal of Systematics and Evolution* 50:374–85

48. Dumolin S, Demesure B, Petit RJ. 1995. Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theoretical and Applied Genetics* 91:1253–56

49. Bai WN, Zeng YF, Zhang DY. 2007. Mating patterns and pollen dispersal in a heterodichogamous tree, *Juglans mandshurica* (Juglandaceae). *New Phytologist* 176:699–707

50. Lei M, Wang Q, Wu ZJ, López-Pujol J, Li DZ, et al. 2012. Molecular phylogeography of *Fagus engleriana* (Fagaceae) in subtropical China: limited admixture among multiple refugia. *Tree Genetics & Genomes* 8:1203–12

51. Kozlowski G, Sébastien B, Song YG. 2018. *Wingnuts (Pterocarya) & walnut family. Relict trees: linking the past, present and future.* Switzerland: Natural History Museum Fribourg. pp. 36–39.

52. Kozlowski G, Song Y, Bétrisey S. 2019. *Pterocarya hupehensis.* The IUCN Red List of Threatened Species 2019: e. T66816108A

152835141. http://dx.doi.org/10.2305/IUCN.UK.2019-3.RLTS.T66 816108A152835141.en. Accessed on 27 February 2024.

53. Fang J, Wang Z, Tang Z. 2011. *Atlas of woody plants in China*. Heidelberg: Springer Berlin. 2000 pp. https://doi.org/10.1007/ 978-3-642-15017-3

54. Wu ZY, Raven PH, Hong DY. 2003. *Flora of China. Volume 5: Ulmaceae through Basellaceae*. Beijing: Science Press, St. Louis: Missouri Botanical Garden Press. http://flora.huh.harvard.edu/ china/mss/volume05/index.htm

55. Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19:11−15

56. 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* 94:275−88

57. Qian ZH, Li Y, Li MW, He YX, Li JX, et al. 2019. Molecular phylogeography analysis reveals population dynamics and genetic divergence of a widespread tree *Pterocarya stenoptera* in China. *Frontiers in Genetics* 10:1089

58. Demesure B, Sodzi N, Petit RJ. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* 4:129−34

59. Sugahara K, Kaneko Y, Sakaguchi S, Ito S, Yamanaka K, et al. 2017. Quaternary range-shift history of Japanese wingnut (*Pterocarya rhoifolia*) in the Japanese Archipelago evidenced from chloroplast DNA and ecological niche modeling. *Journal of Forest Research* 22:282−93

60. Borkowski DS, McCleary T, McAllister M, Romero-Severson J. 2014. Primers for 52 polymorphic regions in the *Quercus rubra* chloroplast, 47 of which amplify across 11 tracheophyte clades. *Tree Genetics & Genomes* 10:885−93

61. Xu J, Deng M, Jiang XL, Westwood M, Song YG, et al. 2015. Phylogeography of *Quercus glauca* (Fagaceae), a dominant tree of East Asian subtropical evergreen forests, based on three chloroplast DNA interspace sequences. *Tree Genetics & Genomes* 11:805

62. Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Molecular Biology and Evolution* 38:3022−27

63. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, et al. 2017. DnaSP 6: DNA Sequence Polymorphism analysis of large data sets. *Molecular Biology and Evolution* 34:3299−302

64. Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10:564−67

65. Pons O, Petit RJ. 1996. Measwring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* 144:1237−45

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

67. Manni F, Guérard E, Heyer E. 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Human Biology* 76:173−90

68. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, et al. 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 15:e1006650

69. Hu Y, Woeste KE, Zhao P. 2016. Completion of the chloroplast genomes of five Chinese *Juglans* and their contribution to chloroplast phylogeny. *Frontiers in Plant Science* 7:1955

70. Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817−18

71. Zhang Q, Ree RH, Salamin N, Xing Y, Silvestro D. 2022. Fossil-informed models reveal a boreotropical origin and divergent evolutionary trajectories in the walnut family (Juglandaceae). *Systematic Biology* 71:242−58

72. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 67:901−04

73. Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585−95

74. Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915−25

75. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* 22:3124−40

76. Zhang WP, Cao L, Lin XR, Ding YM, Liang Y, et al. 2021. Dead-end hybridization in walnut trees revealed by large-scale genomic sequence data. *Molecular Biology and Evolution* 39:msab308

77. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows−Wheeler transform. *Bioinformatics* 25:1754−60

78. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078−79

79. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156−58

80. Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19:1655−64

81. R Core Team. 2021. *R: A language and environment for statistical computing*. https://www.R-project.org/

82. Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11:94

83. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32:268−74

84. Pickrell J, Pritchard J. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *Nature Precedings*

85. Liu X, Fu YX. 2020. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biology* 21:280

86. Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* 15:356

87. Bai WN, Yan PC, Zhang BW, Woeste KE, Lin K, et al. 2018. Demographically idiosyncratic responses to climate change and rapid Pleistocene diversification of the walnut genus *Juglans* (Juglandaceae) revealed by whole-genome sequences. *New Phytologist* 217:1726−36

88. Posada D. 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 25:1253−56

89. Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231−59

90. Hijmans RJ, Phillips S, Leathwick J, Elith J, Hijmans MRJ. 2017. Package 'dismo'. *Circles* 9:1−68

91. Hijmans RJ, Van Etten J, Cheng J, Mattiuzzi M, Sumner M, et al. 2015. Package 'raster'. *R package* 734:473

92. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25:1965−78

93. Braconnot P, Otto-Bliesner B, Harrison S, Joussaume S, Peterchmitt JY, et al. 2007. Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum − Part 1: experiments and large-scale features. *Climate of the Past* 3:261−77

94. Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A, et al. 2006. Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science* 311:1751−53

95. Naimi B, Hamm NA, Groen TA, Skidmore AK, Toxopeus AG. 2014. Where is positional uncertainty a problem for species distribution modelling? *Ecography* 37:191−203

96. Song YG, Walas Ł, Pietras M, Sâm HV, Yousefzadeh H, et al. 2021. Past, present and future suitable areas for the relict tree *Pterocarya fraxinifolia* (Juglandaceae): integrating fossil records, niche modeling, and phylogeography for conservation. *European Journal of Forest Research* 140:1323–39

97. Jump AS, Mátyás C, Peñuelas J. 2009. The altitude-for-latitude disparity in the range retractions of woody species. *Trends in Ecology & Evolution* 24:694–701

98. Reinmann AB, Hutyra LR. 2016. Edge effects enhance carbon uptake and its vulnerability to climate change in temperate broadleaf forests. *Proceedings of the National Academy of Sciences of the United States of America* 114:107–12

99. Czúcz B, Gálhidy L, Mátyás C. 2011. Present and forecasted xeric climatic limits of beech and sessile oak distribution at low altitudes in Central Europe. *Annals of Forest Science* 68:99–108

100. Zhou TH, Li S, Qian ZQ, Su HL, Huang ZH, et al. 2010. Strong phylogeographic pattern of cpDNA variation reveals multiple glacial refugia for *Saruma henryi* Oliv. (Aristolochiaceae), an endangered herb endemic to China. *Molecular Phylogenetics and Evolution* 57:176–88

101. Meng HH, Zhang CY, Song YG, Yu XQ, Cao GL, et al. 2022. Opening a door to the spatiotemporal history of plants from the tropical Indochina Peninsula to subtropical China. *Molecular Phylogenetics and Evolution* 171:107458

102. Chen D, Zhang X, Kang H, Sun X, Yin S, et al. 2012. Phylogeography of *Quercus variabilis* based on chloroplast DNA sequence in East Asia: multiple glacial refugia and mainland-migrated island populations. *PLoS One* 7:e47268

103. 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* 18:1396–407

104. Buschbom J, Yanbaev Y, Degen B. 2011. Efficient long-distance gene flow into an isolated relict oak stand. *Journal of Heredity* 102:464–72

105. Maharramova E, Huseynova I, Kolbaia S, Gruenstaeudl M, Borsch T, et al. 2018. Phylogeography and population genetics of the riparian relict tree *Pterocarya fraxinifolia* (Juglandaceae) in the South Caucasus. *Systematics and Biodiversity* 16:14–27

106. Maharramova E. 2016. *Genetic diversity and population structure of the relict forest trees Zelkova carpinifolia (Ulmaceae) and Pterocarya fraxinifolia (Juglandaceae) in the South Caucasus*. Dissertation. Berlin: Freie Universität Berlin. http://dx.doi.org/10.17169/refubium-9267

107. Hopkins R. 2013. Reinforcement in plants. *New Phytologist* 197:1095–103

108. Balao F, Lorenzo MT, Sánchez-Robles JM, Paun O, García-Castaño JL, et al. 2020. Early diversification and permeable species boundaries in the Mediterranean firs. *Annals of Botany* 125:495–507

109. de Lafontaine G, Ducousso A, Lefèvre S, Magnanou E, Petit RJ. 2013. Stronger spatial genetic structure in recolonized areas than in refugia in the European beech. *Molecular Ecology* 22:4397–412

110. Lind JF, Gailing O. 2013. Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. *Tree Genetics & Genomes* 9:707–22

111. Wang TR, Meng HH, Wang N, Zheng SS, Jiang Y, et al. 2023. Adaptive divergence and genetic vulnerability of relict species under climate change: a case study of *Pterocarya macroptera*. *Annals of Botany* 132:241–54

112. Gates K, Sandoval-Castillo J, Brauer CJ, Unmack PJ, Laporte M, et al. 2023. Environmental selection, rather than neutral processes, best explain regional patterns of diversity in a tropical rainforest fish. *Heredity* 130:368–80

113. Cortázar-Chinarro M, Lattenkam EZ, Meyer-Luch Y, Luquet E, Laurila A, et al. 2017. Drift, selection, or migration? Processes affecting genetic differentiation and variation along a latitudinal gradient in an amphibian *BMC Evolutionary Biology* 17:189

114. Winkler MG, Wang PK. 1993. The late-quaternary vegetation and climate of China. In *Global Climates since the Last Glacial Maximum*, eds. Wright HE, Kutzbach JE, WebbIII T, Ruddiman WF, Street-Perrott FA, et al. Minneapolis, MN: University of Minnesota Press. pp. 221–64

115. Willis KJ, Niklas KJ. 2004. The role of Quaternary environmental change in plant macroevolution: the exception or the rule? *Philosophical Transactions of the Royal Society B* 359:159–72

116. Sun Y, Ding Y. 2010. A projection of future changes in summer precipitation and monsoon in East Asia. *Science China Earth Sciences* 53:284–300

117. Wang B, Liu J, Kim HJ, Webster PJ, Yim SY. 2012. Recent change of the global monsoon precipitation (1979–2008). *Climate Dynamics* 39:1123–35

118. Li Y, Ding Y, Li W. 2017. Interdecadal variability of the Afro-Asian summer monsoon system. *Advances in Atmospheric Sciences* 34:833–46

119. Schoonderwoerd KM, Friedman WE. 2021. Naked resting bud morphologies and their taxonomic and geographic distributions in temperate, woody floras. *New Phytologist* 232:523–36

120. Qiao L, Wen G, Qi Y, Lu B, Hu J, et al. 2018. Evolutionary melting pots and reproductive isolation: a ring-shaped diversification of an odorous frog (*Odorrana margaratea*) around the Sichuan Basin. *Molecular Ecology* 27:4888–900

121. Wang Y, Feijó A, Cheng J, Xia L, Wen Z, et al. 2021. Ring distribution patterns—diversification or speciation? Comparative phylogeography of two small mammals in the mountains surrounding the Sichuan Basin *Molecular Ecology* 30:2641–58

122. Xie XF, Yan HF, Wang FY, Ge XJ, Hu CM, et al. 2012. Chloroplast DNA phylogeography of *Primula ovalifolia* in central and adjacent southwestern China: past gradual expansion and geographical isolation. *Journal of Systematics and Evolution* 50:284–94

123. Liu Y, Chang X. 2003. Modeling for the burial and subsidence history of the Sichuan Basin. *Chinese Journal of Geophysics* 46:283–90

124. Zheng H. 2015. Birth of the Yangtze River: age and tectonic-geomorphic implications. *National Science Review* 2:438–53

125. Li J, Xie S, Kuang M. 2001. Geomorphic evolution of the Yangtze Gorges and the time of their formation. *Geomorphology* 41:125–35

126. Gillespie A, Molnar P. 1995. Asynchronous maximum advances of mountain and continental glaciers. *Reviews of Geophysics* 33:311–64

127. Nei M, Maruyama T, Chakraborty R. 1975. The bottleneck effect and genetic variability in populations. *Evolution* 29:1–10

128. Borthakur D, Busov V, Cao XH, Du Q, Gailing O, et al. 2022. Current status and trends in forest genomics. *Forestry Research* 2:11

129. Xu LL, Yu RM, Lin XR, Zhang BW, Li N, et al. 2021. Different rates of pollen and seed gene flow cause branch-length and geographic cytonuclear discordance within Asian butternuts. *New Phytologist* 232:388–403

130. Ellstrand NC, Elam DR. 1993. Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics* 24:217–42

131. Allendorf FW. 1986. Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology* 5:181–90