

From glacial refugia to wide distribution range: demographic expansion of *Loropetalum chinense* (Hamamelidaceae) in Chinese subtropical evergreen broadleaved forest

Wei Gong^{1,2} · Wanzhen Liu¹ · Lei Gu^{1,3} · Shingo Kaneko⁴ · Marcus A. Koch⁵ · Dianxiang Zhang¹

Received: 1 July 2015 / Accepted: 19 November 2015 / Published online: 1 December 2015
© Gesellschaft für Biologische Systematik 2015

Abstract The subtropical evergreen broadleaved forest (STEBF) in China is globally one of the most diverse and biologically important forest systems. There has been a long-term debate whether this region was affected dramatically during Pleistocene glaciation and deglaciation cycles, e.g., in terms of range dimensions and changes in species richness. Here we report a large-scale phylogeographic study, focusing on a widespread and typical constituent species of the Chinese STEBF, *Loropetalum chinense* (R. Br.) Oliver (Hamamelidaceae). In total, 56 populations spanning the entire distribution range of *L. chinense* were analyzed. Chloroplast DNA sequence variation and AFLPs as molecular marker systems were used in

combination with ecological niche modeling (ENM). ENM indicated that the distribution ranges of *L. chinense* were contracted remarkably and most populations retreated southward. Thereby, based on ENM, geographical distribution pattern of cpDNA haplotypes, and AFLP genetic clusters, one glacial refuge was inferred in the Nanling Mountains in southern China, and a second glacial refuge was identified in Three Gorges Area and Dabashan Mountains in Chongqing Province, southwestern China. In addition to ENM, with mismatch distribution analysis and Bayesian skyline plots, demographic expansion was inferred to take place about 10.6 kya. The current geographic distribution pattern of genetic variation might be shaped by northward and eastward expansion along Nanling Mountains and Wuyishan Mountains, respectively. Additionally, the two mountain ranges were supposed to act as geographical barriers restricting gene flow between the southern and northern populations. Herewith, we aim to further contribute a case study of the phylogeographic history of this vegetation type, which will help to improve deeper understanding of past vegetation dynamics and floristic evolutionary pathways of the Chinese STEBF.

Electronic supplementary material The online version of this article (doi:10.1007/s13127-015-0252-4) contains supplementary material, which is available to authorized users.

✉ Marcus A. Koch
marcus.koch@cos.uni-heidelberg.de

✉ Dianxiang Zhang
dx-zhang@srbg.ac.cn

- ¹ Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China
- ² College of Life Sciences, South China Agricultural University, Guangzhou 510642, China
- ³ Peking University, Conservation Biology Building, Beijing 100871, China
- ⁴ Laboratory of Forest Biology, Division of Forest and Biomaterials Science, Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan
- ⁵ Centre for Organismal Studies (COS) Heidelberg, Department of Biodiversity and Plant Systematics, Heidelberg University, Im Neuenheimer Feld 345, 69120 Heidelberg, Germany

Keywords Widespread species · Phylogeography · Postglacial history · Evergreen broadleaved forest

Introduction

Climatic oscillations during the Quaternary glaciation and deglaciation cycles have long been considered to have had profound effects on the geographic distribution and spatial genetic pattern of extant species (Hewitt 2004). An overall “contract-expand” model and genetic diversity pattern of “southern richness versus northern purity” have often been observed for species occurring in European or northern American regions

(Hewitt 2000). Accumulating evidence suggested that the genetic structure of modern plant populations reflect aspects of vegetational dynamics since at least the Last Glacial Maximum (LGM: c. 21 kya), e.g., the locations of glacial refugia, trajectories of postglacial colonization and associated demographic processes, as well as the aggregation and disaggregation of forest communities (Hewitt 2000; Petit et al. 2008; Hu et al. 2009). Genetic evidence is now required to provide new insights into past vegetational history and address important paleoecological questions (Hu et al. 2009; McLachlan et al. 2005; Shepherd et al. 2007). Phylogeographic studies of widespread species or constituent residents in a specific forest system have been especially important for interpreting the evolutionary history of species and their associated vegetation.

The subtropical evergreen broadleaved forest (STEBF) in China is one of the most important forest systems. It has outstanding plant richness and is one of the main biodiversity hot spots in Southeast Asia due to its large geographic and climatic heterogeneity (Hou 1983; Wang 1992a, b; Myers et al. 2000; Ying 2001). Several localized centers of plant diversity with high levels of endemism have been identified in topographically diverse mountain ranges (Ying 2001). Simulated vegetation maps have indicated that climatic cooling during the LGM caused southward shifts in the geographical distribution of all temperate vegetation species, being confined to the most southern tropical part of China and forming a continuous vegetation band eastward to the Japanese archipelago via the East China Sea shelf (Harrison et al. 2001; Sun et al. 1999; Ni et al. 2010). The refugial populations of the STEBF expanded and advanced northward during the middle Holocene (c. 6 kya) (Sun et al. 1999; Ni et al. 2010). Meanwhile, evidence obtained from geographical data of the Quaternary glaciation and deglaciation cycles has suggested an alternative hypothesis. No glaciers developed at low elevations below 2500 m in the eastern China mountainous regions (defined here as the subtropical region to the east of Qinghai-Tibet Plateau) according to Shi et al. (2006) and Li et al. (2004), but a cold and dry climate formed due to a strengthening winter monsoon (Shi et al. 2006). The effects of lower temperature and precipitation on subtropical region might also cause the latitudinal and altitudinal migrations, population extinction, and range fragmentation and/or expansion.

The increasing large-scale phylogeographic studies of plant species in China supported multiple glacial refugia present in the widespread or component species in STEBF (Qiu et al. 2011) (Table 1). The Hengduan Mountains and the Yungui Plateau in western China were identified as glacial refugia for *Taxus wallichiana*, *Fagus longipetiolata*, and *Pinus kwangtungensis* (Gao et al. 2007; Liu 2008; Tian et al. 2010). The Nanling Mountains in southern China were revealed to be glacial refugia for *F. longipetiolata*, *P. kwangtungensis*, and *Eurycorymbus cavaleriei* (Liu 2008; Tian et al. 2010; Wang et al. 2009). The region of eastern

subtropical China was inferred to be a glacial refugium for *T. wallichiana* and *Ginkgo biloba* (Gao et al. 2007; Gong et al. 2008). Additionally, the southwestern Dabashan Mountains, Daloushan Mountains, and the Sichuan Basin were inferred to be glacial refugia for *Cathaya argyrophylla*, *Cercidiphyllum japonicum*, *T. wallichiana*, *G. biloba*, and *E. cavaleriei* (Wang and Ge 2006; Gao et al. 2007; Gong et al. 2008; Wang et al. 2009; Qi et al. 2012). Moreover, localized or fast range expansions from several and different refugia were detected in *T. wallichiana*, *P. kwangtungensis*, *Quercus variabilis*, *Platycarya strobilacea*, and *Rhododendron simsii* (Gao et al. 2007; Tian et al. 2008, 2015; Chen et al. 2012a, b; Li et al. 2012; Xu et al. 2014). Evidence for eastward migration is also obtained in widespread species such as *F. longipetiolata* (Liu 2008).

The research presented here focused on *Loropetalum chinense* (R. Br.) Oliver, an evergreen shrub or small tree, belonging to the Hamamelidaceae family. It is one of the dominant residents of the STEBF with widespread distribution across the major mountain ranges in subtropical China. The most widely found populations of *L. chinense* display white flowers with terminal inflorescence. Two morphologically variable populations were found with narrow distributions on the southern limit and exhibit either small white flowers with axillary inflorescence or red flowers with terminal inflorescence. Previous studies revealed that seeds of *L. chinense* show ballistic dispersal model and fall down by gravity (Du et al. 2009). Pollination studies indicated self-compatibility of *L. chinense*; however, the possibility of cross-pollination cannot be excluded as *Thrips* sp. has been detected to exist inside the flowers (Gu and Zhang 2008; Gu 2008). Its congener *Loropetalum subcordatum* (Benth) Oliver, an endemic endangered species with contrastingly narrow and scattered distribution pattern in China, shows similar ecological preferences and is suggested to be autogamy based on previous studies (Gu and Zhang 2008; Gong et al. 2010). The third species in the genus, *Loropetalum lanceum* Hand.-Mzt., has not been recorded in recent decades and therefore was not included in the present study. The variety *L. chinense* Oliver var. *rubrum* Yieh, which has red flowers and terminal inflorescence, was originally found in central China and is now cultivated widely.

Given the wide distribution of *L. chinense* in subtropical China and ballistic dispersal seeds by gravity, this species represents a good system to test if the STEBF component plant species conform the “contract-expand” model and assist in a better understanding of the vegetative history of STEBF. Despite an increasing number of phylogeographic studies in China, few of them focus on the widespread or component plant species in STEBF with extensive sampling. In the present study, a wide-ranging phylogeographic analysis was conducted using chloroplast DNA (cpDNA) sequence variation and amplified fragment length polymorphisms (AFLP)

Table 1 Life form, genetic diversity, glacial refugia, and postglacial history for selected constituent plant species in subtropical China

	Life form	Pollination model	Seed dispersal model	Haplotype diversity	Gene diversity	Glacial refugia	Postglacial history	References
<i>Sargentodoxa cuneata</i>	Woody deciduous climber	Wind pollinated	By birds	0.665	Not done	Nanling	Repeated expansion	Tian et al. (2015)
<i>Quercus glauca</i>	Evergreen tree	Wind pollinated	By gravity	0.920	Not done	Dabashan, Xuefengshan, Wuyishan, etc	Range expansion	Xu et al. (2014)
<i>Quercus variabilis</i>	Deciduous, monoecious, autochthonous	Wind and insect pollinated	/	0~0.821	Not done	Dabieshan, Dabashan, etc	Range expansion	Chen et al. (2012a, b)
<i>Rhododendron simsii</i>	Semi-evergreen shrub	Animal pollinated	/	0.71	Not done	Sichuan Basin	Range expansion	Li et al. (2012)
<i>Cercidiphyllum japonicum</i>	Tall canopy tree, dioecious	Wind pollinated	By wind	Not provided	0.901	Sichuan Basin, Middle Yangtze region	Postglacial northward recolonization	Qi et al. (2012)
<i>Eurycorymbus cavaleriei</i>	Deciduous, small tree	Insect pollinated	By gravity	0.834	Not done	Nanling, Dabashan, Daloushan, Sichuan Basin	/	Wang et al. (2009)
<i>Fagus longipetiolata</i>	Widespread tree in subtropical China	Wind pollinated	By gravity or animals	Not provided	0.793	Nanling, Yungui Plateau	Eastward migration	Liu (2008)
<i>Pinus kwangtungensis</i>	Five-needled pine tree in isolated montane areas	Wind pollinated	/	Not done	0.629	Nanling, Yungui Plateau	Range expansion	Tian et al. (2008)
<i>Taxus wallichiana</i>	Dioecious tree	Wind pollinated	By birds and mammal	0.884	Not done	Nanling, Yungui Plateau, Dabashan, Daloushan, Sichuan Basin	Range expansion	Gao et al. (2007)
<i>Cathaya argyrophylla</i>	Coniferous tree	Wind pollinated	/	/	/	Dabashan, Daloushan, Sichuan Basin	/	Wang and Ge (2006)
<i>Hibiscus tiliaceus</i>	Long-lived woody perennial	Insect pollinated	Water dispersal	Not done	0.826	/	/	Tang et al. (2003)

analysis. Maxent-based ecological niche modeling was performed to further make some predictions on past and potential distribution area and climatically suitable range.

Materials and methods

Population sampling

Between 2006 and 2009, samples from 56 populations of *L. chinense* (numbered from 1 to 56) from across its entire geographic range within the STEBF were collected, including two unusual and morphologically very variable populations and two geographically distinct Japanese populations. A total of 462 and 470 individuals were subjected to cpDNA and AFLP analysis, respectively, with eight to nine individuals per population. The Chinese populations were classified into four regions based on geographically defined distributions and respective altitudes and climate: (1) The southwestern populations (SW; numbers 1–9) include those at the Three Gorges area in the Yangtze River valley and the ones in the southwestern mountains at high altitudes of more than 700 m, such as Dabashan Mountains and Daloushan Mountains. (2) The central populations (CEN; numbers 10–24) are composed of those populations at lower altitudes than 500 m in the mountain ranges of Xuefengshan Mountains and Luoxiaoshan Mountains central region. (3) The southern populations (STH; numbers 25–42) include populations at middle to high altitudes at the Nanling Mountains and those near the southern coastline. Morphologically variable populations (MPV; numbers 41–42) were located in the STH region. (4) The eastern populations (EST; numbers 43–54) comprise the ones with relatively lower altitudes than 500 m in the eastern Wuyishan Mountains or along the eastern coastline. Two Japanese populations (JP; numbers 55–56) were collected from Mei Prefecture and Arao, Kyushu. As for *L. chinense* var. *rubrum* (CUL; numbers 57–58), nine individuals of two cultivated populations were analyzed comparatively. Twelve individuals of two populations of *L. subcordatum* (LS; numbers 59–60) were used as an outgroup in the analysis. The locations of all the sampled populations and the number of individuals analyzed per population are presented in Table S1 and Fig. 1. Leaf material was collected at intervals of at least 10 m in each population. Total genomic DNA was extracted from a silica gel-dried leaf tissue using a modified CTAB method (Doyle 1991; Doyle and Doyle 1987).

Ecological niche modeling

Ecological niche modeling (ENM) is a means of characterizing the spatial distribution of suitable conditions for species, and in conjunction with traditional molecular phylogeographic inferences, it has been the main method for locating

potential distributional areas at the LGM (Soberón and Peterson 2005; Waltari et al. 2007). The modeling approach assumes equilibrium between a species and its environmental adaptation and relates the known occurrence of the species to the environmental parameters of landscape-level variation. Thus, it can be used to predict the suitable distribution range of a species (Peterson 2003). ENM was performed based on high-resolution paleoclimate data inferred for the Last Interglacial (LIG), Last Glacial Maximum (LGM), Middle Holocene (MH), and current. Bioclimatic variables were downloaded from the WorldClim database (<http://worldclim.org/download>) for the four different stages with its conditions at 2.5' spatial resolution. The LIG, LGM, and MH data were obtained from circulation model simulation of the Community Climate System Model (CCSM) (Collins et al. 2006), which provides downscaled high-resolution estimates of the climate parameters (Hijmans et al. 2005). A maximum entropy modeling technique (Maxent v3.3.2) (Phillips et al. 2006) was used in our study. Current *L. chinense* herbarium records were used for accurate presence data and projection into the past. Occurrence points were put into “samples,” current bioclimatic variables into “environmental layers,” and past 19 bioclimatic variables separately into “projection layers directory.” This setting gives a solid model or specific niche built by combining current bioclimatic variables with occurrence points and projected into past period layers. When running the model, the default convergence threshold (10^{-5}) and maximum number of iterations (500) were used together with a regularization multiplier of 1. Model accuracy was assessed by evaluating the area under the curve (AUC) of the receiver-operating characteristic (ROC) plot (Phillips et al. 2006), where scores higher than 0.7 were considered to show a good performance of the model (Fielding and Bell 1997). To test each model, 25 % of the data in each run was randomly selected by Maxent to generate a test file, which was then compared with the model output calculated with the remaining 75 % of the presence data. We adopted “minimum training presence logistic threshold (MTP)” as a stringent threshold to classify predicted map into “suitable habitat” and “unsuitable habitat.” Cells were coded “suitable” if the Maxent output suitability value was greater than or equal to the lowest output value for the training occurrence points (Richmond et al. 2010). This approach is thus conservative, identifying the minimum predicted area possible while maintaining zero omission error in the training dataset (Pearson 2007).

Molecular marker systems

An initial screen for DNA sequence variation of various chloroplast markers using universal primers was conducted for one individual from each population. Variation was observed in three intergenic spacer (IGS) regions: *psaA-ycf3* (Ickert-Bond and Wen 2006), *psbA-trnH* (Hamilton 1999), and

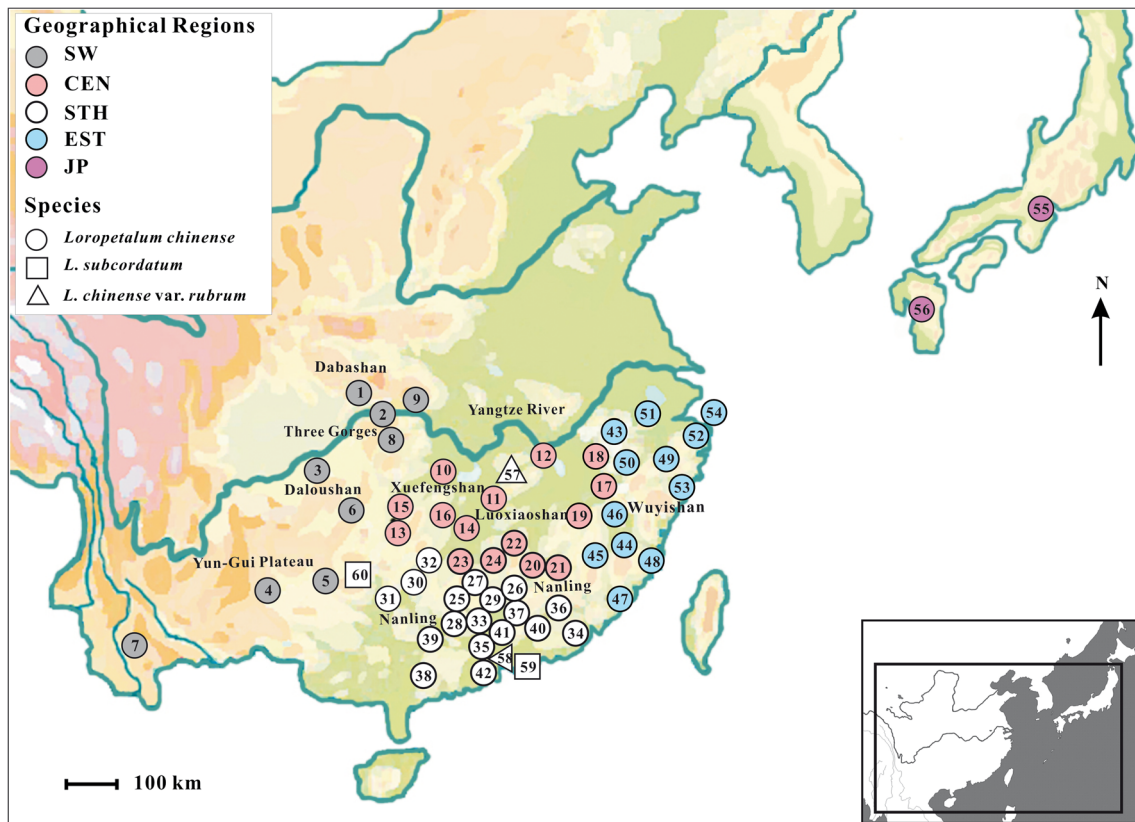


Fig. 1 Map of the sample locations. The 56 populations (1–56) of *L. chinense* were examined across its whole distribution range in China and Japan. Two populations of *L. chinense* var. *rubrum* (57–58) and two populations of *L. subcordatum* (59–60) were also sampled, respectively. Numbers 1–60 (ingroup and outgroup populations) denote the population

order and correspond to those used in Table S1. The four Chinese geographical regions (SW, CEN, STH, and EST) and one Japanese region (JP) of *L. chinense* are differentiated by circles filled with different colors. The major mountain ranges in China are also annotated on the map. Additional accession information is given in Table S1

atpB-rbcL (Chiang et al. 1998), and one region of the *trnL* intron (Taberlet et al. 1991). PCR was carried out using a reaction mixture with a total volume of 50 μ L, containing 50 ng of genomic DNA, 2 mmol/L $MgCl_2$, 0.2 mmol/L dNTPs, 0.2 μ mol/L of each primer, and 1.25 U DNA polymerase (Takaya Bio Inc.). The amplification program consisted of a 5-min denaturation step at 94 $^{\circ}C$, followed by 30 cycles of 1 min denaturation at 94 $^{\circ}C$, 1 min annealing at 52 $^{\circ}C$, and 90 s elongation at 72 $^{\circ}C$. Amplification ended with an elongation phase lasting 8 min at 72 $^{\circ}C$ and a final hold at 10 $^{\circ}C$. Amplification products were checked for length and concentration on 1.5 % agarose gels and sent to Shanghai Invitrogen Biotechnology Co., Ltd. for commercial sequencing. Haplotype sequences identified in *L. chinense* and *L. subcordatum* were deposited in GenBank under accession numbers JN542723–JN542828 and HM369421, HM369423, HM369425, HM369426, HM369428–HM369430, and HM369432–HM369438.

Our AFLP procedure was performed following the protocol established by Vos et al. (1995) with some modifications. One individual was selected from each population and they were tested using 64 primer combinations (ABI plant

protocol) to screen for those which produced the most readable and informative profiles. Three primer combinations yielding the most polymorphic and readable band patterns were selected for the final AFLP analysis: *EcoRI*-ACA/*MseI*-CTC (FAM), *EcoRI*-AAC/*MseI*-CTC (HEX), and *EcoRI*-AGG/*MseI*-CAG (TET). Three standard samples were used in each run to serve as a control for scoring the reproducibility of the banding patterns of each primer combination. Three differentially labeled fluorescent primer pairs were multiplexed prior to electrophoresis. Multiplexed selective PCR reactions were diluted 50 times with ddH₂O, and 6 μ L of the resulting mixture was extracted and mixed with 1.5 μ L ROX500. After denaturation at 95 $^{\circ}C$ for 5 min, samples were run on a MegaBase 1000 automated sequencer (Amersham Biosciences). Raw data were imported into GeneMarker v1.85 (Softgenetics LLC, PA, USA), and amplified fragments between 50 and 500 bp were scored by visual inspection for their presence (1) or absence (0) in the output traces. Only distinct peaks were counted and the manual scoring procedure was repeated three times to reduce inconsistencies in scoring. The binary matrix from each primer combination was combined into one dataset. To allow further analysis, the data

matrix was converted into appropriate input matrices for different genetic software using the program AFLPDAT (Ehrich 2006).

Data analysis

To analyze cpDNA data, sequences were edited and assembled with Sequencher v4.1.4 (Gene Codes Corporation). Variable sites in the data matrix were double-checked against the original electrophorogram. Multiple alignments of the cpDNA sequences were made using MUSCLE (Edgar 2004a, b) with subsequent manual adjustment in BioEdit 7.0.4.1 (Hall 1999). Haplotype diversity (h) and nucleotide diversity (π) (Nei 1987) as well as haplotype frequency and distribution were calculated using Arlequin 3.0 (Excoffier et al. 2005). For AFLP data, we assessed the total number of fragments (NT), the percent polymorphic fragments (PPF), and gene diversity for all populations and geographical regions using AFLPDAT (Ehrich 2006). As the species are self-compatible, estimations of inbreeding coefficients were evaluated using FAFLPcalc (Dasmahapatral et al. 2008) with FAFLP values calculated for each individual. For both datasets, analysis of molecular variance (AMOVA) was implemented in Arlequin 3.0, and a hierarchical analysis of population subdivision was performed (Excoffier et al. 2005). Pairwise F_{st} values were also calculated using Arlequin 3.0 (Excoffier et al. 2005) in order to estimate the genetic differentiation among localities. The correlation between geographical and genetic distance was investigated by performing a Mantel test using GenAlEx6 (Peakall and Smouse 2006) and computing the geographical and genetic distance matrices after 999 permutations.

Phylogenetic analysis of cpDNA sequence data was conducted using Bayesian inference (BI) in BEAST v1.5.3 (Drummond and Rambaut 2007). *L. subcordatum* was used as outgroup for the phylogenetic analysis. The GTR+G model was selected as the most appropriate substitution model based on Akaike information criterion (AIC) implemented in Modeltest v3.7 (Posada and Crandall 1998). Statistical support for branching patterns was estimated by bootstrap replication (NJ, ML: 1000 replicates). BI was conducted by using samples drawn every 1000 generations over a total of 1×10^7 iterations to obtain the posterior estimates of the age of the most recent common ancestor (TMRCA) by a Markov chain Monte Carlo (MCMC) analysis. The first 25 % of trees were discarded as burn-in. The genealogical degree of relatedness among cpDNA haplotypes of *L. chinense* was estimated with a 95 % connection limit using TCS version 3.11 (Clement et al. 2000). In this analysis, indels were treated as single mutation events and coded as substitutions (A or T). In both parsimonious and network analyses, length variations in mononucleotide repeats (poly A or T) were excluded because of their tendency for homoplasy.

Loropetalum and *Hamamelis-Corylopsis* are two highly supported sister clades (Li et al. 1999a, b; Ickert-Bond and Wen 2006). To estimate a reliable divergence time between *L. chinensis* and *L. subcordatum*, we employed the four chloroplast noncoding regions (the *psaA-ycf3*, *psbA-trnH*, *atpB-rbcL* IGSs, and the *trnL* intron) from *Corylopsis* (DQ352363, AB237065, GU576577, and DQ352352), *Hamamelis mollis* (GU576726, GU576760, GU576588, and AF147475) and *Hamamelis virginiana* (GU576735, GU576770, AM183419, and DQ352196). Two fossils were used as calibration points to infer both absolute ages of clades and maximum or minimum age constraints. Based on fossil flowers of *Androdecidua endressii* (Magallón et al. 2001) and fossil leaves of *Corylopsis reedae* (Radtko et al. 2005), the difference in time between Hamamelidae and Corylopsideae was constrained to be a minimum of 85 my [F1], and the age of *Corylopsis* at least 50 my [F2]. The divergence time of *L. chinense* and *L. subcordatum* was estimated with the aid of BEAST v1.5.3 (Drummond and Rambaut 2007) using the nucleotide substitution model GTR+G, which gave the best fit according to Modeltest v3.7 (Posada and Crandall 1998). Posterior estimates of the age of the TMRCA were obtained by a MCMC analysis with samples drawn every 1000 generations over a total of 2×10^8 iterations. The molecular clock model was applied for estimations of the cpDNA lineage divergence time. The rate constancy of cpDNA haplotype evolution in *Loropetalum* was evaluated by relative rate tests (Tajima 1993) in MEGA5.0 (Tamura et al. 2007) using *Corylopsis* as an outgroup. Following the discovery of rate constancy, the time since divergence (T) of haplotype lineages was inferred from their net pairwise sequence divergence per base pair (d_A) using the Kimura-2-p model. Divergence time was calculated as $T = d_A / 2\mu$, where μ is the rate of nucleotide substitution (Nei and Kumar 2000). The value of μ was derived from the pairwise genetic distance (d_A), and the divergence time (T) between *L. chinense* and *L. subcordatum* estimated based on fossil calibrations with the aid of the BEAST program.

In order to test the demographic expansion scenario, mismatch distribution analysis of cpDNA data was performed using a sudden (stepwise) expansion model (Rogers and Harpending 1992) for three haplotype groups C, D, and E as well as at species levels. Goodness-of-fit was tested using the sum of squared deviations (SSD) between observed and expected mismatch distributions and Harpending's (Harpending 1994) raggedness index (HRag). The number of sites where two DNA sequences differ is proportional to the expansion time, making it possible to determine the date of population expansions directly from mismatch distributions (Rogers and Harpending 1992; Schneider and Excoffier 1999). For groups with expanding populations, the expansion parameter (τ) was converted to an estimate of time (T , in number of generations) since the start of expansion began using $T = \tau / 2u$ (Rogers and Harpending 1992; Rogers 1995). In this formula, the neutral mutation rate for the

entire sequence (haplotype) per generation of u is calculated as $u = \mu kg$, where μ is the substitution rate, k is the average sequence length of the DNA region under study, and g is the generation time in years. The value of g was approximated as 10 years in shrubs. A parametric bootstrap approach (Schneider and Excoffier 1999) with 1000 replicates was used to assess the goodness-of-fit of the observed mismatch distribution to the sudden expansion model, to test the significance of HRag, and to obtain 95 % confidence intervals (CIs) around τ . Significantly negative D and F_s values often indicate recent population expansion following a severe reduction in population size, i.e., causing a severe bottleneck (Fu 1997). For the total dataset and each geographical region, Tajima's D (Tajima 1989) and Fu's F_s (Fu 1997) tests of selective neutrality were conducted. All the above analyses were carried out with the aid of Arlequin 3.0 (Excoffier et al. 2005). Analysis of Bayesian skyline plot (BSP) was also conducted to test the change of effective population size. The BEAST v1.5.3 software (Drummond and Rambaut 2007) was used to create the BSP with 10 steps. The analysis was run for 3×10^7 iterations with a burn-in of 3×10^7 using the GTR+G model obtained from Modeltest v3.7 (Posada and Crandall 1998). Convergence of parameters and mixing of chains were followed by visual inspection of parameter trend lines and checking of ESS values by performing three preruns. Adequate sampling and convergence to a stationary distribution were checked using TRACER v1.4.1 (Rambaut and Drummond 2009). Posterior estimates of parameters were all distinctly unimodal with all parameters identifiable.

For AFLP data, population structure was examined at an individual level by genetic admixture analysis using the program STRUCTURE 2.1 (Pritchard et al. 2000), which implements a model-based clustering method to infer population structure and is widely used for inferring gene-pool structure in genetic data. The assumed K populations ranged from 1 to 10, with 10 replicate runs for each K , a burn-in period of 2×10^5 , and 5×10^4 MCMC replicates after burn-in. The admixture model and uncorrelated allele frequencies were chosen for the analysis. The estimated mean logarithmic likelihood of K values and delta K values were both calculated and ranged from 1 to 10, and the optimal K was determined using the online version of Structure Harvester. For further details of the abovementioned parameters and methods, refer to Ehrich (2006) and Evanno et al. (2005). The assignment of individuals from each population to the clusters with optimal K was displayed as a proportional bar, with each cluster represented by a different color. The clustering result was displayed geographically by the program distruct 1.1 (Rosenberg 2004). Additionally, the population structure was also examined using the program BAPS v3.2 (Corander et al. 2003, 2004; Corander and Marttinen 2006) in order to compare with the result of STRUCTURE 2.1. BAPS was run with the most likely number of groups (K) set to 2 to 15. Each run was replicated three times.

Results

Ecological niche modeling

The predictions of the bioclimatically suitable areas for *L. chinense* during the LIG (120–140 kya), LGM (~21 kya), MH (6 kya), and current are shown in Fig. 2a, d. Evaluation of model performance based on both training and test sample data indicated that the models had high predictive abilities (AUC=0.995 and 0.994, respectively). A vegetation map from the LGM (http://intarch.ac.uk/journal/issue11/2/map/download_page_js.htm) was compared in parallel (Fig. 2b). The geographical distribution range of *L. chinense* at the LGM was predicted to be remarkably contracted and retreated southward occupying the previous tropical regions, reaching the land bridge and extending to the Japanese archipelago. The potentially suitable distribution range of *L. chinense* was mostly confined to the southern subtropical mountainous regions, i.e., Nanling Mountains. The suitable distribution range was also found to be located in the southwestern mountainous ranges, i.e., the Three Gorges Area and Dabashan Mountains. Based on LGM vegetation map, the potentially suitable distribution range of *L. chinense* was covered by tropical grassland (number 6) in South Sea Land Bridge, forest steppe (number 18) in southern China, semiarid temperate woodland or shrub (number 12) in southern China and East Sea Land Bridge, and open boreal woodlands (number 11) in southwestern China during the LGM (Ray and Adams 2001). Comparably, the distribution range of *L. chinense* during the MH demonstrated a large expansion, occupying most of the Chinese subtropical region. No obvious changes of the geographical distribution were detected between MH and current.

Genetic diversity and phylogenetic analysis of cpDNA haplotypes

The four sequenced cpDNA noncoding regions originating from the 462 individuals of *L. chinense* ranged from 2232 to 2259 bp and were aligned with a consensus length of 2281 bp. Twenty-seven haplotypes (H1–H27) were identified based on 17 polymorphic loci in *L. chinense* (Table S2). Two species-specific haplotypes (H28 and H29) were found for the outgroup sequences from *L. subcordatum* (Table S2). Phylogenetic reconstructions of BI analyses generated weakly supported clades. Despite low Bayesian posterior values, the BI tree was broadly congruent with the TCS network (Fig. 3). An overall star-like shape was displayed based on the TCS network. A total of 12 haplotype groups (A to L) were identified based on TCS network analysis (Fig. 3). Genetic diversity indices at both the population and regional levels were estimated and shown in Table S1.

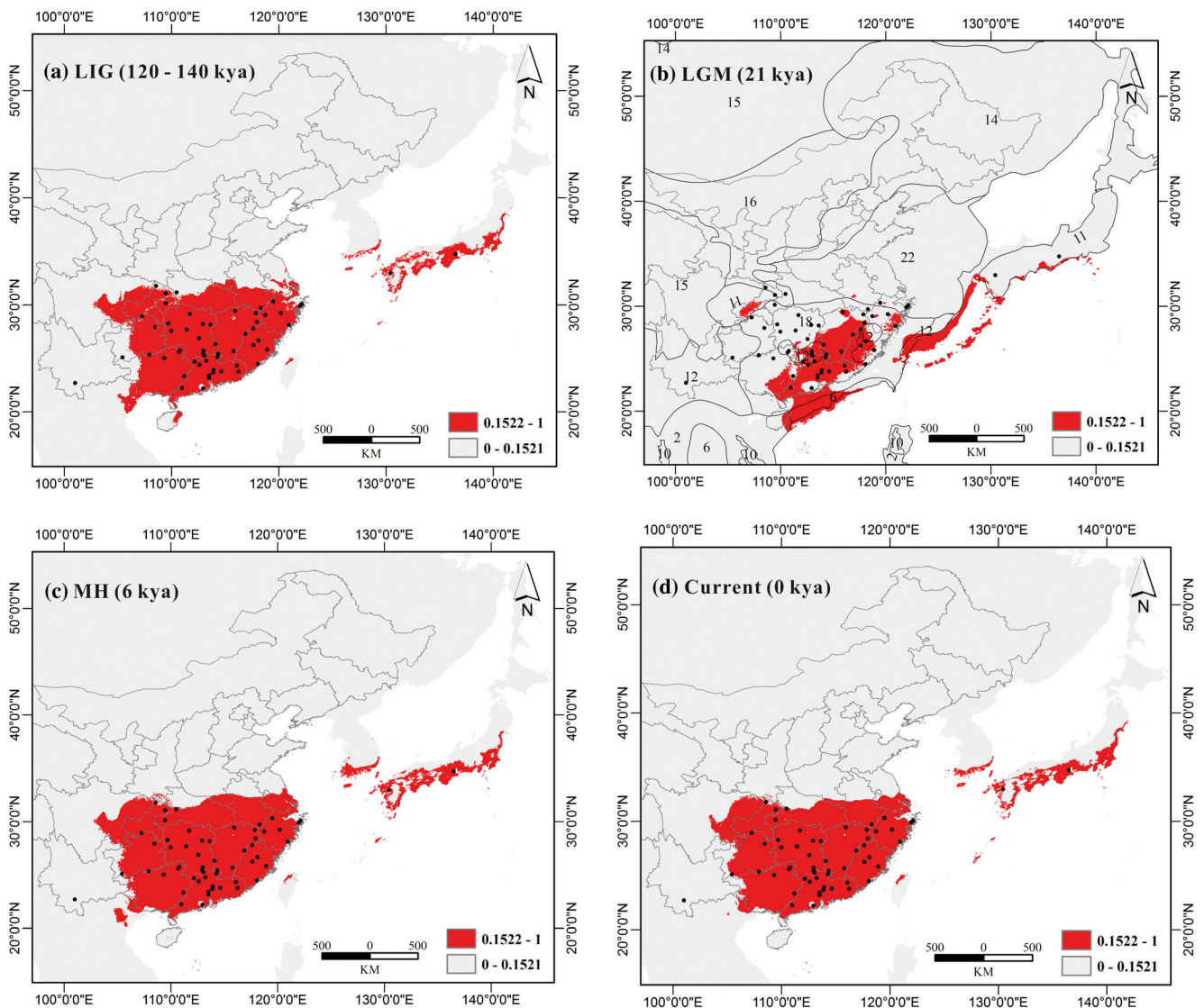


Fig. 2 Potentially suitable areas for *L. chinense* predicted by ecological niche modeling (ENM) using climatic variables at four different periods. Suitable and unsuitable habitats are displayed as colors of *red* and *gray*, respectively, where *red* represents the habitat suitability (occurrence probability) higher than 15%. *Black dots* represent the points of sampled populations in our study. **a** The simulated distribution range at the Last Interglacial (LIG). **b** Potentially suitable areas projected in comparison with a layer of GIS-based vegetation map at the Last

Glacial Maximum (LGM). Numbers from 1 to 26 represent different vegetation types at the LGM. These suitable areas were proven to be covered by tropical grassland (number 6) in South Sea Land Bridge, forest steppe (number 18) in southern China, semiarid temperate woodland or shrub (number 12) in southern China and East Sea Land Bridge, and open boreal woodlands (number 11) in southwestern China. **c** The simulated distribution range at the Middle Holocene (MH). **d** The current potential distribution range

Genetic diversity indices are given in Table S1. Based on cpDNA data, the total haplotype diversity (h) was 0.7666 and the nucleotide diversity (π) was 1.421×10^{-3} . At the population level, the haplotype diversity (h) varied from 0.0000 to 0.8333 and nucleotide diversity (π) from 0.0000 to 1.845×10^{-3} among all populations. The highest genetic diversity was found in CQFJ (number 2; $h=0.8333$, $\pi=1.805 \times 10^{-3}$) in the SW region, GXJWS (number 16; $h=0.8333$, $\pi=1.845 \times 10^{-3}$) in the STH region, and JXHC (number 19; $h=0.8333$, $\pi=1.648 \times 10^{-3}$) in the CEN region. At the regional level, the highest haplotype diversity and nucleotide diversity were

revealed in the CEN region ($h=0.8360$) and the SW region ($\pi=1.871 \times 10^{-3}$), respectively. The AFLP data revealed that amplification from the three primer combinations for all the accessions produced a total of 214 fragments, ranging from 50 to 500 bp. Bands not reproducible in the active control were excluded as they were likely to be artifacts. The mean scoring error, calculated in the three replicates, was $<5\%$ for the four primer combinations. A total of 213 bands displayed polymorphism (PPF=99.53%) and the total gene diversity was 0.2634. The highest gene diversity was detected in the STH region (0.2595), followed by the CEN region (0.2569). Based

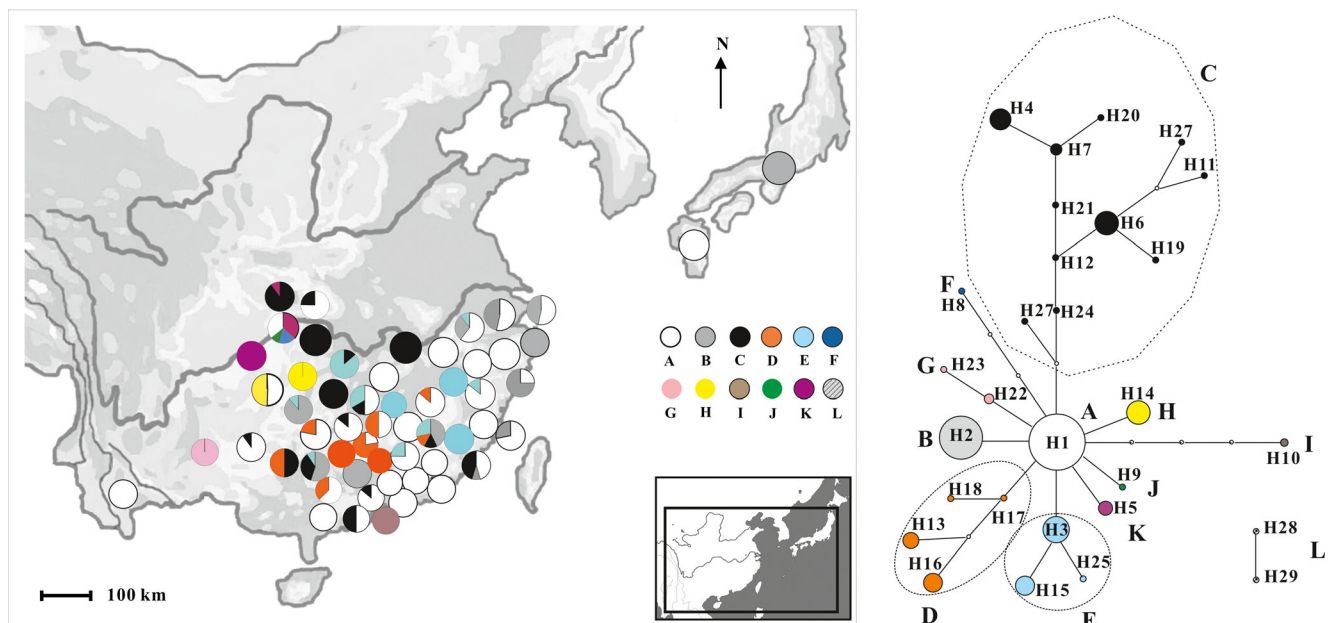


Fig. 3 The geographic distribution and respective frequency of the 12 haplotypes (A–L) among the 56 populations and the 95 % plausible network of these cpDNA haplotypes. The size of each circle corresponds to the frequency of each haplotype. *Small solid circles*

without haplotype codes indicate unsampled or extinct haplotypes. Haplotypes H1–H27 belong to *L. chinense*, and haplotypes H28–H29 occur in *L. subcordatum* only

on both cpDNA and AFLP dataset, low genetic diversity ($h=0.0000$) and gene diversity (0.0000) were detected in MPV (numbers 41–42). The JP populations (numbers 55–56) displayed lower genetic diversity ($h=0.4583$) and gene diversity (0.1596) based on both datasets. The cultivated *L. chinense* var. *rubrum* exhibited extremely low genetic diversity ($h=0.0000$) and gene diversity (0.2143). Concerning the inbreeding level, the average FAFLP values at the species level was 0.2237. At the population level, the highest inbreeding levels were revealed in Japanese populations and morphologically variable populations (average FAFLP values of 0.98 and 1.86, respectively).

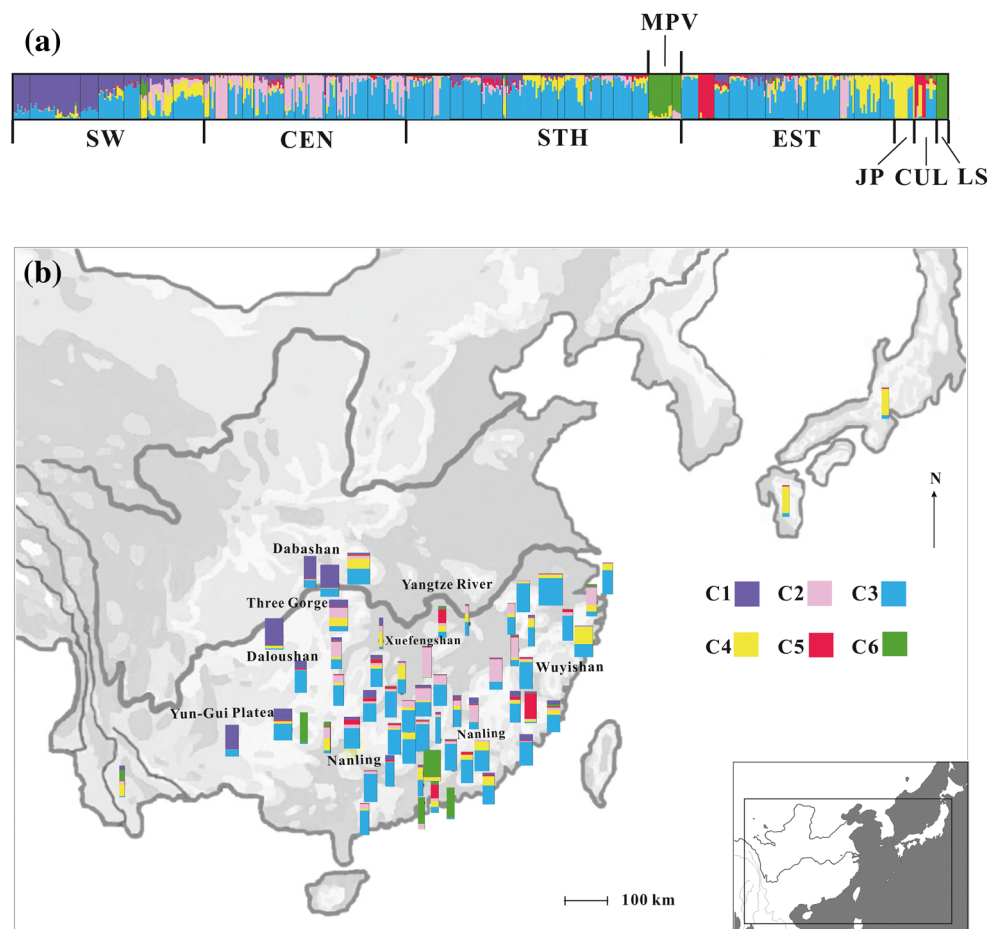
Spatial geographical distribution of genetic variation and population genetic structure

The geographic distribution and respective frequencies of the 12 haplotype groups (A–L) among the 56 populations are shown in Fig. 3. The most widely distributed haplotype (H1) was found in 34 populations and 211 individuals. It was inferred to be the ancestral haplotype with more frequency in the STH and EST regions. The second most frequent haplotype was H2 (14.3 % of all samples) with higher frequency in the EST region. Haplotype group C was shared among the four regions, but with comparably lower frequency in the EST region. Haplotype group D occurred mainly in Nanling Mountains in the STH region. Haplotype group E occurred mostly in the CEN and EST regions. Haplotype groups of F, G, H, J, and K were

geographically restricted to the SW region. The highest number of haplotypes was identified in STH with 16 haplotypes, among which seven were specific (H10, H11, H12, H17, H18, H20, and H21). The second highest number of haplotypes was found in CEN with 13 haplotypes, including three private ones (H25, H26, and H27). The SW region possessed less haplotype numbers (11 haplotypes) than the CEN region, but owned a higher number of seven unique ones (H4, H5, H8, H9, H22, H23, and H24). Only four haplotypes were detected in the EST region without any specific ones. The richness of the haplotypes decreased from west to east. Haplotype H10 was specific to MPV populations and haplotypes H28 and H29 were unique to *L. subcordatum*. No interspecific haplotypes were detected between *L. chinense* and *L. subcordatum*.

Genetic admixture analysis was conducted on the AFLP data for all populations of *L. chinense* using STRUCTURE 2.1 (Pritchard et al. 2000). Based on the estimated mean logarithmic likelihood of K values, $\text{Ln}''(K)=320.6800$ with $K=6$ was higher than $\text{Ln}''(K)=292.9000$ with $K=4$. Therefore, the optimal K was determined to be 6 (Fig. 4a; Fig. S1). Though the $\text{Ln}''(K)$ value when $K=2$ was extremely high, we did not consider it as the optimal one. When $K=2$, the corresponding geographically structured results provided insufficient information for further interpretation, because only individuals of MPV and *L. subcordatum* were successfully separated (Fig. S2). Additionally, the $\text{Ln}''(K)$ values at $K=2$ often and consistently tend to be very high and significant for population data even for other species

Fig. 4 **a** Genetic admixture analysis conducted on AFLP data for *L. chinense*, *L. chinense* var. *rubrum* (*CUL*), and *L. subcordatum* (*LS*). Each vertical bar represents an individual and its assignment proportion into one of six population clusters. **b** Genetic structuring of populations based on AFLP data. The map of the individual assignment in each population to $K = 6$ clusters (C1–C6) is based on STRUCTURE analysis of the AFLP data. Each cluster is represented by a different color



based on the method of Evanno et al. (2005). This method was shown to favor $K=2$, causing artificially maximal $\text{Ln}''(K)$ values (Vigouroux et al. 2008; Iorizzo et al. 2013). Therefore, the optimal K was often adopted with the second largest $\text{Ln}''(K)$ values (Vigouroux et al. 2008; Iorizzo et al. 2013). Finally, we conducted BAPS analysis to further test for congruent results using higher K values. The optimal K was given to be 11 by BAPS program, which rejected $K=2$ as the optimal one. Though BAPS gave incongruent optimal K values compared to STRUCTURE, BAPS at least did not suggest that $K=2$ was the optimal one. Additionally, we found converging and very similar results between BAPS ($K=8$) and STRUCTURE using the optimal $K=6$ (Fig. S2).

The geographical assignment pattern of the individuals of *L. chinense* in the six clusters based on Bayesian analysis was generally consistent to the geographical regions of SW, STH, CEN, EST, and JP. It also showed similar geographical distribution pattern with the cpDNA haplotypes. Cluster C1 (purple) was specific to SW. The populations in STH and EST mainly comprised individuals from cluster C3 (blue) with a minor part from C2 (pink). The populations in the CEN region were assigned to clusters C2 (pink) and C3 (blue). The Japanese populations contain individuals assigned to cluster C4 (yellow).

Cultivated populations belonged to clusters C3 (blue) and C5 (red). The MPV showed the same assignment of individuals with *L. subcordatum*, belonging to cluster C6 (green).

Analysis of the molecular variance of the cpDNA data (Table 2) indicated significant genetic differentiation among the 56 populations ($F_{st}=0.7053$, $P<0.001$), of which the variation among the geographic regions accounted for 4.40 %. Analysis of the molecular variance of the AFLP data indicated relatively high genetic differentiation within populations (76.39 %), but lower among populations (15.22 %) or among the geographic regions (8.39 %), with $F_{st}=0.2255$ ($P<0.001$). The results of the Mantel test implied that there was no significant correlation between genetic distances and geographic distances ($P=0.423$) for the cpDNA data. However, isolation by distance was detected for the AFLP data ($R_{xy}=0.3270$; $P=0.001$).

Estimation of divergence time and inference of demographic history

To estimate the Bayesian divergence time, we constrained the stem lineage to a minimum of 85 my and the age of *Corylopsis*

Table 2 Analysis of molecular variance (AMOVA) for all populations of *L. chinense* based on cpDNA and AFLP data at both population and regional levels

Source of variation	df	Sum of squares	Variance components	Percentage of variance	F_{st}
cpDNA					
Among groups	4	67.670	0.0727	4.40 ns	
Among populations within groups	51	483.477	1.0938	66.13***	
Within populations	409	199.352	0.4874	29.47***	
Total	464	750.499	1.6540		0.7053***
AFLP					
Among groups	4	719.395	1.2134	8.39***	
Among populations within groups	51	3243.491	5.1227	15.22***	
Within populations	415	9032.198	21.7643	76.39***	
Total	469	13,118.509	28.3683		0.2255***

ns not significant

$p > 0.05$; ** $p < 0.05$; *** $p < 0.001$

to a minimum of 50 my. With fossil calibration, the estimated interspecific divergence time between *L. chinense* and *L. subcordatum* was 45.1 (18.5–72.5) my during the middle Eocene (Tertiary) (Fig. 5). This result was comparable with the result of Ickert-Bond and Wen (2006), which estimated the two species divergence time to be 46.07 (27.42–63.10) my. Relative rate constancy tests detected no significant rate heterogeneity among the cpDNA haplotypes of *Loropetalum* compared with *Corylopsis* (all $P > 0.05$), indicating that a clock-like evolution model was applicable for *Loropetalum*. The corrected cpDNA sequence divergence (d_A) between *L. chinense* and *L. subcordatum* was 0.0113. Using the divergence time (45.1 my) estimated from fossil calibrations, the substitution rate was estimated to be 1.2525×10^{-10} s/s/y for the four combined noncoding cpDNA regions of *L. chinense* and *L. subcordatum*. Neutral test showed the result of negative

D (−1.0313) and F_s (−5.8698) values at the species level. Mismatch distribution analysis detected unimodal distributions with nonsignificant SSD (0.0125; $P = 0.6710$) and HRag (0.0256; $P = 0.8780$) statistics at the species level (Fig. 6). Additionally, Bayesian skyline plots displayed slightly ascending curves (Fig. S3). All of the above analyses provide evidence for demographic expansion and population growth of *L. chinense*. Accordingly, the expansion time was estimated to be 10.6 kya between the LGM and MH. For cpDNA haplotype groups C, D, and E, either D or F_s showed a negative value, but not both of them together (neutrality test, Fig. S4). Unimodal distributions were detected with significant SSD and HRag statistics for cpDNA haplotype groups D and E, but not for group C based on mismatch distribution analysis (Fig. S4). Therefore, the evidence for demographic expansion for each cpDNA haplotype group seemed to be relatively weak.

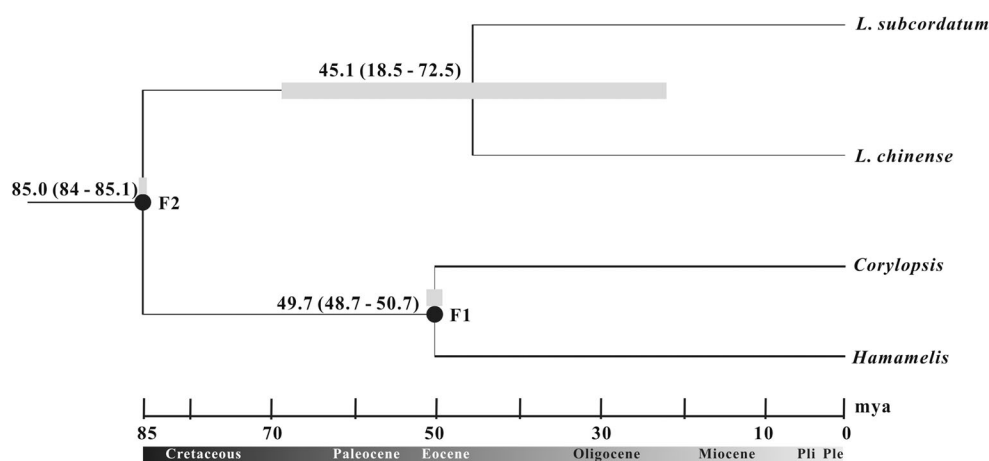
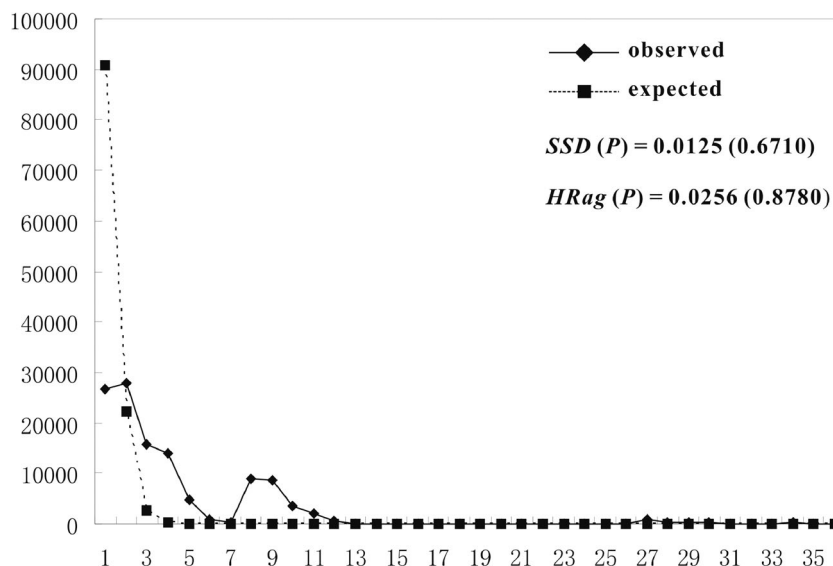


Fig. 5 Chronogram of the Bayesian tree for divergence time estimations. Branch lengths were transformed via Markov chain Monte Carlo (MCMC) simulations in the Bayesian time estimation. Gray boxes indicate 95 % confidence intervals. Fossils were used to calibrate divergence time estimates (F1 = 49.7 (48.7–50.7) my: *Corylopsis*

reedae, and F2 = 85.0 (84.0–85.1) my: *Androdecidua*). The divergence time between *L. chinense* and *L. subcordatum* is estimated to be 45.1 (18.5–72.5) my based on the combination of the four cpDNA sequences of *L. chinense*, *L. subcordatum*, *Hamamelis*, and *Corylopsis*

Fig. 6 Mismatch distribution analysis detected unimodal distributions with SSD and HRag statistics at the species level



Discussion

Genetic diversity and population differentiation

Our results revealed a high level of genetic diversity (cpDNA: $h=0.7666$; AFLP: $h=0.2634$) with 27 cpDNA haplotypes and 99.53 % AFLP polymorphic fragments in *L. chinense*. Other species that have similar geographic distributions compared to *L. chinense* in subtropical China, such as *T. wallichiana* ($h=0.8840$) (Gao et al. 2007) and *E. cavaleriei* ($h=0.8340$) (Wang et al. 2009), also exhibit high levels of cpDNA genetic diversity (Table 1). Based on AFLP data, relatively high genetic diversity has also been reported in other subtropical plants such as *Hibiscus tiliaceus* (Malvaceae) with PPB=88.5 % and $h=0.2430$ (Tang et al. 2003). The high genetic diversity of *L. chinense* might be due to its wide and continuous distribution range, large population size, and putatively long evolutionary history, which would have provided ample opportunity for the accumulation and maintenance of high levels of genetic variation (Hamrick and Godt 1996; Chiang and Schaal 1999; Huang et al. 2001; Chiang et al. 2006). The MPV that are located in the southern marginal region displayed low genetic diversity (cpDNA: $h=0.5556$; AFLP: gene diversity=0.1918, PPF=56.07 %). Japanese populations also possessed low genetic diversity (cpDNA: $h=0.4583$; AFLP: gene diversity=0.1596, PPF=38.79 %). The low level of genetic diversity in MPV and JP could be interpreted as a result of long-term geographical isolation and genetic drift. As for the cultivated red-flowered variety *L. chinense* var. *rubrum*, low genetic diversity was detected (cpDNA: $h=0.0000$; AFLP: gene diversity=0.2134, PPF=44.39 %). The apparent loss of genetic variation in the cultivated populations might be attributed to the simple germplasm resources.

Significant genetic divergence among the populations based on the analyses of cpDNA data ($F_{st}=0.7053$, $P<0.001$) was found. This may be because *L. chinense* produces gravity-dispersal seeds, which discourages long-distance seed flow among populations, allowing genetic drift to have a significant impact on subpopulations, and ultimately leads to distinct differentiation among populations (Hamrick and Godt 1996; Abbott et al. 2000; Petit et al. 2003). Based on the AFLP data, genetic divergence was again evident among the populations ($F_{st}=0.2255$, $P<0.001$), but was lower than that obtained from cpDNA data. This might suggest a low level of inbreeding of *L. chinense*, which was detected in our result (average FAFLP values=0.2232). Though a previous pollination study revealed self-compatibility in *L. chinense*, the possibility of cross-pollination cannot be excluded, as thrips have been observed inside the flowers (Gu 2008). Thrips have been proven to be effective pollinators in many plants (Sakai 2001; Moog et al. 2002; Peñalver et al. 2012). Consequently, the possibility of outbreeding either mediated by insects or wind is suggested to exist in *L. chinense*.

Glacial refugia, demographic expansion, and geographical barriers

Our findings suggested the existence of two glacial refugia within the STEBF in China, constituted during the last Pleistocene glaciation. One was situated in Nanling Mountains in southern China. The second one was located in Three Gorges Area and Dabashan Mountains, Chongqing Province, in southwestern China. ENM predicted the potentially suitable distribution range of *L. chinense* at the LGM to be isolated in southern and southwestern China, which was consistent with the simulated vegetation map and geographical data (Harrison

et al. 2001; Sun et al. 1999; Ni et al. 2010; Shi et al. 2006; Li et al. 2004). Molecular evidence revealed high genetic diversity, haplotype diversity, and unique haplotype numbers in the two potential glacial refugia. Climate oscillations during the Pleistocene ice age resulted in glacial-interglacial cycles with associated expansion and contraction of species' ranges (Hewitt 2000, 2004; Comes and Kaderit 1998; Taberlet et al. 1998; Abbott and Brochmann 2003). There were indications that mountainous regions in southern and southwestern China with high genetic diversity and ancient persistent lineages were dominated by evergreen broadleaved trees during the LGM (Fig. 2b; Xiao et al. 2007). To serve as refugia, these regions would need to exhibit comparably stable ecological conditions during periods of environmental fluctuations, allowing the accumulation of genetic diversity (Hewitt 1996). Mountainous regions are usually regarded as critical glacial refugia because their high relief offers scope for shifts in altitudinal range in response to major climatic changes (Bennett et al. 1991). Nanling Mountains and Dabashan Mountains possess high topographic heterogeneity, with especially high species richness and endemism (Ying 2001; Wang and Wang 1994). With the complex and varied physical topography, these regions were thought to have been only slightly affected by the Pleistocene glaciations, becoming important refugia for many Tertiary plant species (Ying 2001; Zheng 2000). Meanwhile, the Nanling Mountains originated during the late Triassic as a result of the Indo-China movement, also allowing it to be potential glacial refugia for many relic plant species (Ying 2001). Additionally, the Yangtze River winds through southwestern China with formation of numerous steep cliffs and deep valleys, which consequently serve as geographic barriers, restricting gene flow and seed dispersal, and ultimately, resulting in the unique genetic makeup of the resident plant species (Gao et al. 2007; Wang et al. 2009).

Based on ENM, as well as mismatch distribution analysis and the Bayesian skyline plots, demographic expansion was observed and estimated to be 10.6 kya between the LGM and MH. We speculated that the refugial populations expanded northward and eastward along Nanling Mountains and Wuyishan Mountains that are going in the southwest-northeast direction. (i) Given the AFLP clustering results and the eastward decreasing cpDNA haplotype diversity, one migration route was supposed to start from Nanling Mountains eastward along Wuyishan Mountains. This migration route corresponds well with Wang (1992a, b) based on flora investigation in the eastern Asiatic region. Long-distance dispersal is considered to contribute to the losing of genetic diversity of cpDNA haplotypes. However, the relatively low altitudes and flatter landscape in the east allows more pollen flow and long-distance dispersal, which to some extent decreases the inbreeding level in the EST region based on AFLP data. Evidence was shown that cpDNA haplotype groups B and E, as well as AFLP cluster C2, were mostly

present in the northern part of the two mountain ranges, but AFLP cluster C3 majorly occurred in the southern part. Therefore, we supposed that Nanling Mountains and Wuyishan Mountains have acted as geographic barriers during the migration process. (ii) Another migration route was inferred to be from Nanling Mountains northward along Xuefengshan Mountains and Luoxiaoshan Mountains in central China. This migration model corresponded with the hypothesis that the STEBF refugial populations advanced northward during the middle Holocene (Sun et al. 1999; Ni et al. 2010). Short-distance dispersal was suggested among these mountain ranges due to the complex geographical topography, ultimately resulting in the preservation of genetic diversity and the unique genetic makeup in the populations in CEN (Fig. 3). Evidence was shown that cpDNA haplotypes groups D and E were restricted to the northern part of Nanling Mountains and never occurred to the southern part. Similarly, AFLP clusters C2 and C3 were also isolated between the northern and southern regions. Therefore, geographical barriers were again supposed to be formed by Nanling Mountains during the migration process, which restricted gene flow and seed dispersal between its northern and southern parts. Similar geographical barrier models have been proposed for other common residents of the STEBF, such as *E. cavaleriei* (Wang et al. 2009) and *F. longipetiolata* (Liu 2008). (iii) As for the potential glacial refugium of Three Gorges Area and Dabashan Mountains, no obvious range expansion was suggested due to the high number of geographic unique haplotypes and the specific AFLP cluster C1 in the SW region. By considering the shared cpDNA haplotype group D between SW and CEN regions, the only possibility was inferred that seeds or individuals were dispersed or transported by human beings from west to east along the Yangtze River. As high levels of genetic diversity and gene diversity as well as high number of cpDNA haplotypes were present in CEN, a contact zone might be suggested as a result of genetic mixture between SW and STH regions.

Insights into the past vegetation history of the STEBF in China

The genetic structure of current plant populations is thought to reflect vegetation dynamics since at least the LGM (Hewitt 2000). The present study not only provides a comprehensive elucidation of the phylogeographic history of the common residents in the STEBF of China, but also improves our understanding of the past vegetation dynamics and the floristic evolution within it. The results revealed two glacial refugia and two major migration routes of habitant plant species within the STEBF. The occurrence of the two glacial refugia supported both the vegetation dynamics simulations based on paleo-biome reconstructions (Harrison et al. 2001) and the Quaternary theory of glaciations and deglaciations in subtropical China proposed by Shi et al. (2006). The refugial

populations of the STEBF not only are restricted to the modern southern tropical region but could also be identified in the southwestern mountainous region. The subtropical Chinese mountain ranges, such as Nanling Mountains and Wuyishan Mountains with varying geographic conditions running southwest-northeast, have acted as geographical barriers limiting gene flow between the south and north. The STEBF in China is regarded as one of the most important forest systems in the world, serving as a critical glacial refugium and having particularly high biodiversity. The phylogeographic studies, particularly of a comparative nature on the wide range of constituent species, are further required to provide a deeper understanding of the past vegetation changes in the Chinese STEBF.

Acknowledgments This project was supported by the National Science Foundation of China (31300174; 31470312) and the Doctoral Program of Higher Education Research Fund New Teacher Class Jointly Funded Project. We are indebted to Dr. Atsushi Kawakita (Kyoto University, Japan) and Dr. Hanghui Kong, Bo Li, Mr. Xiangxu Huang, Mr. Binghui Chen, and Mr. Lianxuan Zhou (South China Botanical Garden, CAS) for collecting samples and/or assistance in the field and Dr. Tiejiao Tu (South China Botanical Garden, CAS) for data analysis. Special thanks go to the various members of the Department of Biodiversity and Plant Systematics at COS Heidelberg, in particular Dr. Markus Kiefer, Dr. Juraj Paule, and Dr. Roswitha Schmickl for providing substantial support while establishing and running the AFLP experiments. Prof. Peter Del Tredici (The Arnold Arboretum of Harvard University) helped in linguistic corrections, which is greatly acknowledged here.

References

- Abbott, R. J., & Brochmann, C. (2003). History and evolution of the arctic flora: in the footsteps of Eric Hultén. *Molecular Ecology*, *12*, 299–313.
- Abbott, R. J., Smith, L., Milne, R. I., Crawford, R. M., Wolf, K., & Balfour, J. (2000). Molecular analysis of plant migration and refugia in the Arctic. *Science*, *289*, 1343–1346.
- Bennett, K. D., Tzedakis, P. C., & Willis, K. J. (1991). Quaternary refugia of north European trees. *Journal of Biogeography*, *18*, 103–115.
- Chen, D. M., Zhang, X. X., Kang, H. Z., Sun, X., Yin, S., Du, H. M., Yamanaka, N., Gapare, W., Wu, H. X., & Liu, C. J. (2012a). Phylogeography of *Quercus variabilis* based on chloroplast DNA sequence in East Asia: multiple glacial refugia and mainland-migrated island populations. *Plos One*, *7*, e47168.
- Chen, S. C., Zhang, L., Zeng, J., Shi, F., Yang, H., Mao, Y. R., et al. (2012b). Geographic variation of chloroplast DNA in *Platycarya strobilacea* (Juglandaceae). *Journal of Systematics and Evolution*, *50*, 374–385.
- Chiang, T. Y., & Schaal, B. A. (1999). Phylogeography of North American populations of the moss species *Hylocomium splendens* based on the nucleotide sequence of internal transcribed spacer 2 of nuclear ribosomal DNA. *Molecular Ecology*, *8*, 1037–1042.
- Chiang, T. Y., Schaal, B. A., & Peng, C. I. (1998). Universal primers for amplification and sequencing a noncoding spacer between *atpB* and *rbcL* genes of chloroplast DNA. *Botanical Bulletin of Academia Sinica*, *39*, 245–250.
- Chiang, Y. C., Huang, K. H., Schaal, B. A., Ge, X. J., Hsu, T. W., & Chiang, T. Y. (2006). Contrasting phylogeographical patterns between mainland and island taxa of the *Pinus luchuensis* complex. *Molecular Ecology*, *15*, 765–779.
- Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, *9*, 1657–1659.
- Collins, W. D., Bitz, C. M., Blackmon, M. L., Bonan, G. B., Bretherton, C. S., et al. (2006). The community climate system model version 3 (CCSM3). *American Meteorological Society*, *19*, 2122–2143.
- Comes, H. P., & Kaderit, J. W. (1998). The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, *3*, 432–438.
- Corander, J., & Marttinen, P. (2006). Bayesian identification of admixture events using multi-locus molecular markers. *Molecular Ecology*, *15*, 2833–2843.
- Corander, J., Waldmann, P., & Sillanpää, M. J. (2003). Bayesian analysis of genetic differentiation between populations. *Genetics*, *163*, 367–374.
- Corander, J., Waldmann, P., Marttinen, P., & Sillanpää, M. J. (2004). Baps 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics*, *20*, 2363–2369.
- Dasmahapatra, K. K., Lacy, R. C., & Amos, W. (2008). Estimating levels of inbreeding using AFLP markers. *Heredity*, *100*, 286–295.
- Doyle, J. J. (1991). DNA protocols for plants: CTAB total DNA isolation. In G. M. Hewitt & A. W. B. Johnston (Eds.), *A molecular techniques in taxonomy* (pp. 283–293). Berlin: Springer.
- Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf material. *Phytochemistry Bulletin*, *19*, 11–15.
- Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, *7*, 214–221.
- Du, Y. J., Mi, X. C., Liu, X. J., Chen, L., & Ma, K. P. (2009). Seed dispersal phenology and dispersal syndromes in a subtropical broad-leaved forest of China. *Forest Ecology and Management*, *258*, 1147–1152.
- Edgar, R. C. (2004a). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797.
- Edgar, R. C. (2004b). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*, 113–131.
- Ehrich, D. (2006). AFLPdat: a collection of R functions for convenient handling of AFLP data. *Molecular Ecology Notes*, *6*, 603–604.
- 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.
- Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, *1*, 47–50.
- Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, *24*, 38–49.
- Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, *147*, 915–925.
- Gao, L. M., Möller, M., Zhang, X. M., Hollingsworth, M. L., Liu, J., Mill, R. R., et al. (2007). High variation and strong phylogeographic pattern among cpDNA haplotypes in *Taxus wallichiana* (Taxaceae) in China and North Vietnam. *Molecular Ecology*, *16*, 4684–4698.
- Gong, W., Chuan, C., Dobeš, C., Fu, C. X., & Koch, M. A. (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*, 1094–1105.
- Gong, W., Gu, L., & Zhang, D. X. (2010). Low genetic diversity and high genetic divergence caused by inbreeding and geographical isolation in the populations of endangered species *Loropetalum subcordatum*

- (Hamamelidaceae) endemic to China. *Conservation Genetics*, 11, 2281–2288.
- Gu, L. (2008). *Pollination biology of representative species in Hamamelidaceae*. Guangzhou: Southern China Botanical Garden, the Chinese Academy of Sciences, PhD thesis.
- Gu, L., & Zhang, D. X. (2008). Autogamy of an endangered species: *Loropetalum subcordatum* (Hamamelidaceae). *Journal of Systematics and Evolution*, 46, 651–657.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95–98.
- Hamilton, M. B. (1999). Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology Notes*, 8, 521–523.
- Hamrick, J. T., & Godt, M. J. (1996). Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 351, 1291–1298.
- Harpending, R. C. (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology*, 66, 591–600.
- Harrison, S. P., Yu, G., Takahar, H., & Prentice, I. C. (2001). Diversity of temperate plants in East Asia. *Nature*, 413, 129–130.
- Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58, 247–276.
- Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice age. *Nature*, 405, 907–913.
- Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 359, 183–195.
- Hijmans, R. J., Cameron, S., Parra, J., Jones, P., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978.
- Hou, H. Y. (1983). Vegetation of China with reference to its geographical distribution. *Annals of the Missouri Botanical Garden*, 70, 509–548.
- Hu, S. H., Hampe, A., & Petit, R. J. (2009). Paleoeecology meets genetics: deciphering past vegetational dynamics. *Frontiers in Ecology and the Environment*, 7, 371–379.
- Huang, S., Chiang, Y. C., Schaal, B. A., Chou, C. H., & Chiang, T. Y. (2001). Organelle DNA phylogeography of *Cycas taitungensis*, a relict species in Taiwan. *Molecular Ecology*, 10, 2669–2681.
- Ickert-Bond, A., & Wen, J. (2006). Phylogeny and biogeography of Altingiaceae: evidence from combined analysis of five non-coding chloroplast regions. *Molecular Phylogenetics and Evolution*, 39, 512–528.
- Iorizzo, M., Senalik, D. A., Ellison, S. L., Grzebelus, D., Cavagnaro, P. F., Allender, C., Brunet, J., Spooner, D. M., Van Deynze, A., & Simon, P. W. (2013). Genetic structure and domestication of carrot (*Daucus carota* subsp. *sativus*) (Apiaceae). *American Journal of Botany*, 100, 930–938.
- Li, J. H., Bogle, A. L., & Klein, A. S. (1999a). Phylogenetic relationships of the Hamamelidaceae inferred from sequences of internal transcribed spacers (ITS) of nuclear ribosomal DNA. *American Journal of Botany*, 86, 1027–1037.
- Li, J. H., Bogle, A. L., & Klein, A. S. (1999b). Phylogenetic relationships in the Hamamelidaceae: evidence from the nucleotide sequences of the plastid gene matK. *Plant Systematics and Evolution*, 218, 205–219.
- Li, J. J., Shu, Q., Zhou, S. Z., Zhao, Z. J., & Zhang, J. M. (2004). Review and prospects of Quaternary glaciation research in China. *Journal of Glaciology and Geocryology*, 3, 235–243.
- Li, Y., Yan, H. F., & Ge, X. J. (2012). Phylogeographic analysis and environmental niche modeling of widespread shrub *Rhododendron simsii* in China reveals multiple glacial refugia during the last glacial maximum. *Journal of Systematics and Evolution*, 50, 362–373.
- Liu, M. H. (2008). *Phylogeography of Fagus longipetiolata: insights from nuclear DNA microsatellites and chloroplast DNA variation*. Shanghai: East China Normal University, PhD thesis.
- Magallón, S., Crane, P. R., & Herendeen, P. S. (2001). *Androecidia endressii*, gen. et sp. nov., from the Late Cretaceous of Georgia (United States): further floral diversity in Hamamelidoideae (Hamamelidaceae). *International Journal of Plant Sciences*, 162, 963–983.
- McLachlan, J. S., Clark, J. S., & Manos, P. S. (2005). Molecular indicators of tree migration capacity under rapid climate change. *Evolution*, 86, 2088–2098.
- Moog, U., Fiala, B., Federle, W., & Maschwitz, U. (2002). Thrips pollination of the dioecious ant plant *Macaranga hullettii* (Euphorbiaceae) in Southeast Asia. *American Journal of Botany*, 89, 50–59.
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., de Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspot for conservation priorities. *Nature*, 403, 853–858.
- Nei, M. (1987). *Molecular evolutionary genetics*. New York: Columbia University Press.
- Nei, M., & Kumar, S. (2000). *Molecular evolution and phylogenetics*. New York: Oxford University Press.
- Ni, J., Yu, G., Harrison, S. P., & Prentice, I. C. (2010). Palaeovegetation in China during the late Quaternary: biome reconstructions based on a global scheme of plant functional types. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 289, 44–61.
- Peakall, R., & Smouse, P. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6, 288–295.
- Pearson, R. G. (2007). Species' distribution modeling for conservation educators and practitioners. American Museum of Natural History, Synthesis. Available at <http://ncep.amnh.org>
- Peñalver, E., Labandeira, C. C., Barrón, E., Delclòs, X., Nel, P., Nel, A., Tafforeau, P., & Soriano, C. (2012). Thrips pollination of Mesozoic gymnosperms. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 8623–8628.
- Peterson, A. T. (2003). Predicting the geography of species' invasions via ecological niche modelling. *The Quarterly Review of Biology*, 78, 419–433.
- Petit, R. J., Auinagalde, I., de Beaulieu, J. L., Bittkau, C., & Brewer, S. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, 300, 1563–1565.
- Petit, R. J., Hu, F. S., & Dick, C. W. (2008). Forests of the past: a window to future changes. *Science*, 320(5882), 1450–1452.
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231–259.
- Posada, D., & Crandall, K. A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics*, 14, 817–818.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Qi, X. S., Chen, C., Comes, H. P., Sakaguchi, S., Liu, Y. H., Tanaka, N., Sakio, H., & Qiu, Y. X. (2012). Molecular data and ecological niche modelling reveal a highly dynamic evolutionary history of the East Asian Tertiary relict *Cercidiphyllum* (Percidiphyllaceae). *New Phytologist*, 196, 617–630.
- Qiu, Y. X., Fu, C. X., & Comes, H. P. (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, 225–244.
- Radtke, M. G., Pigg, K. B., & Wehr, W. C. (2005). Fossil *Corylopsis* and *Fothergilla* leaves (Hamamelidaceae) from the Lower Eocene Xora of Republic, Washington, USA, and their evolutionary and

- biogeographic significance. *International Journal of Plant Sciences*, 166, 347–356.
- Rambaut, A., Drummond, A. J. (2009). *Tracer: MCMC trace analysis tool*. 1.4.1 edition. (<http://tree.bio.ed.ac.uk/software/tracer/>). Institute of Evolutionary Biology, University of Edinburgh.
- Ray, N., & Adams, J. M. (2001). A GIS-based vegetation map of the world at the last glacial maximum (25,000–15,000 BP). *Internet Archaeology*, 11.
- Richmond, O. M. W., McEntee, J. P., Hijmans, R. J., & Brashares, J. S. (2010). Is the climate right for Pleistocene rewilding? Using species distribution models to extrapolate climatic suitability for mammals across continents. *PLoS ONE*, 5, e12899.
- Rogers, A. (1995). Genetic evidence for a Pleistocene population explosion. *Evolution*, 49, 608–618.
- Rogers, A. R., & Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, 9, 552–569.
- Rosenberg, N. A. (2004). Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4, 137–138.
- Sakai, S. (2001). Thrips pollination of androdioecious *Castilla elastica* (Moraceae) in a seasonal tropical forest. *American Journal of Botany*, 88, 1527–1534.
- Schneider, S., & Excoffier, L. (1999). Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, 152, 1079–1089.
- Shepherd, L. D., Perrie, L. R., & Brownsey, P. J. (2007). Fire and ice: volcanic and glacial impacts on the phylogeography of the New Zealand forest fern *Asplenium hookerianum*. *Molecular Ecology*, 16, 4536–4549.
- Shi, Y. F., Cui, Z. J., & Su, Z. (2006). *The Quaternary glaciation and environmental variations in China*. Hebei: Hebei Science and Technology Publishing House.
- Soberón, J., & Peterson, A. T. (2005). Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, 2, 1–10.
- Sun, X. J., Song, C. Q., & Chen, X. D. (1999). China Quaternary pollen database (CPD) and Biome 6000 project. *Advance in Earth Sciences*, 8, 407–411.
- Taberlet, P., Gielly, L., Pautou, G., & Bouvet, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, 17, 1105–1109.
- Taberlet, P., Fumagalli, L., Wust-Saucy, A. G., & Cosson, J. F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, 7, 453–464.
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585–595.
- Tajima, F. (1993). Measurement of DNA polymorphism. In N. Takahata & A. G. Clark (Eds.), *Mechanisms of molecular evolution* (pp. 37–59). Sunderland: Sinauer Associates Inc.
- 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.
- Tang, T., Zhong, Y., Jian, S. G., & Shi, S. H. (2003). Genetic diversity of *Hibiscus tiliaceus* (Malvaceae) in China assessed using AFLP markers. *Annals of Botany*, 92, 409–414.
- Tian, S., Luo, L. C., Ge, S., & Zhang, Z. Y. (2008). Clear genetic structure of *Pinus kwangtungensis* (Pinaceae) revealed by a plastid DNA fragment with a novel minisatellite. *Annals of Botany*, 102, 69–78.
- Tian, S., López-Pujol, J., Wang, H., Ge, S., & Zhang, Z. Y. (2010). Molecular evidence for glacial expansion and interglacial retreat during Quaternary climatic changes in a montane temperate pine (*Pinus kwangtungensis* Chun ex Tsiang) in southern China. *Plant Systematics and Evolution*, 284, 219–229.
- Tian, S., Lei, S. Q., Hua, W., Deng, L. L., Li, B., Meng, Q. L., et al. (2015). Repeated range expansions and inter/postglacial recolonization routes of *Sargentodoxa cuneata* (Oliv.) Rehd. et Wils. (Lardizabalaceae) in subtropical China revealed by chloroplast phylogeography. *Molecular Phylogenetics and Evolution*, 85, 238–246.
- Vigouroux, Y., Glaubitz, J. C., Matsuoka, Y., Goodman, M. M., Sánchez, J. G., & Doebley, J. (2008). Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *American Journal of Botany*, 95, 1240–1253.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, 23, 4407–4414.
- Waltari, E., Hijmans, R. J., Peterson, A. T., Nyári, Á. S., Perkins, S. L., & Guralnick, R. P. (2007). Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PLoS One*, 2, e563.
- Wang, W. C. (1992a). On some distribution patterns and some migration routes found in the eastern Asiatic region. *Acta Phytotaxonomica Sinica*, 30, 1–24.
- Wang, W. C. (1992b). On some distribution patterns and some migration routes found in the eastern Asiatic region. *Acta Phytotaxonomica Sinica*, 30, 97–117.
- Wang, H. W., & Ge, S. (2006). Phylogeography of the endangered *Cathaya argyrophylla* (Pinaceae) inferred from sequence variation of mitochondrial and nuclear DNA. *Molecular Ecology*, 15, 4109–4122.
- Wang, Z. Y., & Wang, Y. L. (1994). The biodiversity and characters of spermatophytic genera endemic to China. *Acta Botanica Yunnanica*, 16, 209–220.
- Wang, J., Gao, P. X., Kang, M., Andrew, J. L., & Huang, H. W. (2009). Refugia within refugia: the case study of a canopy tree (*Eurycorymbus cavaleriei*) in subtropical China. *Journal of Biogeography*, 36, 2156–2164.
- Xiao, J. Y., Lu, H. B., Zhou, W. J., Zhao, Z. J., & Hao, R. H. (2007). Evolution of vegetation and climate since the last glacial maximum recorded at Dahu peat site, south China. *Science in China Series D: Earth Sciences*, 50, 1209–1217.
- Xu, J., Deng, M., Jiang, X. L., Westwood, M., Song, Y. G., & Turkington, R. (2014). 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, 1–17.
- Ying, T. S. (2001). Species diversity and distribution pattern of seed plants in China. *Biodiversity Science*, 9, 393–398.
- Zheng, Z. (2000). Late Quaternary vegetational and climatic changes in the tropical and subtropical areas of China. *Acta Micropalaeontologica Sinica*, 17, 125–146.