

ORIGINAL ARTICLE

Population structure and historical demography of *Dipteronia dyeriana* (Sapindaceae), an extremely narrow palaeoendemic plant from China: implications for conservation in a biodiversity hot spot

C Chen^{1,2}, RS Lu¹, SS Zhu¹, I Tamaki³ and YX Qiu¹

Inferring past demography is a central question in evolutionary and conservation biology. It is, however, sometimes challenging to disentangle their roles of contemporary versus historical processes in shaping the current patterns of genetic variation in endangered species. In this study, we used both chloroplast microsatellite (cpSSR) loci and nuclear microsatellite (nSSR) loci to assess the levels of genetic differentiation, genetic effective population size, contemporary/historical levels of gene flow and demographic history for five populations sampled across the range of *Dipteronia dyeriana*, an endangered palaeoendemic from Southwestern China. We found that *D. dyeriana* had a mixed pattern of moderate genetic diversity and high inbreeding. Bayesian clustering divided *D. dyeriana* populations into two nSSR genetic clusters. Coalescent-based approximate Bayesian computation analyses suggest the western and eastern groups of *D. dyeriana* likely persisted in a long-term refuge in Southern China since the beginning of the last glacial period, whereas increasingly colder and arid climates at the onset of the last glacial maximum might have fostered the fragmentation of *D. dyeriana* within refugia. Following their divergence, the western group kept relatively stable effective population size, whereas the eastern group had experienced 500-fold population expansion during the Holocene. Although clear loss of genetic diversity by human activities was not suggested, recent habitat fragmentation has led to a reduction of population connectivity and increased genetic differentiation by ongoing genetic drift in isolated populations, possibly owing to decreased population size in recent dozen years. Finally, we discussed the implications of these results on conservation policies.

Heredity (2017) **119**, 95–106; doi:10.1038/hdy.2017.19; published online 5 April 2017

INTRODUCTION

The Southwestern China, with ~13 000 vascular plant species (López-Pujol *et al.*, 2011), constitutes one of the world's major biodiversity hot spots (Myers *et al.*, 2000). Approximately 29.2% of plant species are endemic to this region (López-Pujol *et al.*, 2011). Of these endemic plant species, some are defined as 'palaeoendemic lineages' whose historical distribution was much wider and today is restricted to a few, often disjunct refugia. Designing effective conservation actions for these narrow palaeoendemics often requires knowledge on the contemporary population structure of the species, along with an understanding of the main processes shaping it (Schwartz *et al.*, 2007). It is now well established that, even within the glacial sanctuary for plants in Southwestern China, Pleistocene climate cycling had an important influence on their distribution, genetic structure and phylogeographic patterns (for example, Qiu *et al.*, 2011; Liu *et al.*, 2012). Besides, all forest stands in Southwestern China suffered some kind of land-cover change since the mid-twentieth century because of logging and clearing for agriculture (Zhang, 2000), including the destruction of natural habitat of subtropical plant species (Long *et al.*, 2003), as possibly revealed by distinct genetic signatures of population

bottlenecks, reduced within-population genetic diversity and/or elevated levels of inbreeding (Dubreuil *et al.*, 2010). Regardless of the widespread recognition that both contemporary and historical processes may play important roles in shaping contemporary genetic patterns, it has proved challenging to disentangle their respective contributions (Zellmer and Knowles, 2009). Various approaches have been used to disentangle the relative role of contemporary and historical processes, including utilizing historical (museum) samples to compare the genetic variation before and after the decline (Gutiérrez-Rodríguez *et al.*, 2011), or using different types of analytical methods (Johnson *et al.*, 2009; Chiocchi and Gibbs, 2010) to test the historical and contemporary gene flow or effective population size (N_e) based on extant samples. Recently, a very powerful and flexible approach, approximate Bayesian computation (ABC), has been developed to estimate demographic and historical parameters (for example, effective population size, divergence time) and quantitatively compare alternative scenarios (Beaumont, 2010). In addition, it should be stressed that molecular markers with different rates of substitution can capture signatures of historical and contemporary processes, and could therefore help to determine the relative contribution of postglacial

¹Key Laboratory of Conservation Biology for Endangered Wildlife of the Ministry of Education, and College of Life Sciences, Zhejiang University, Hangzhou, People's Republic of China; ²College of Life Sciences, Nanjing Agricultural University, Nanjing, People's Republic of China and ³Gifu Academy of Forest Science and Culture, Mino, Japan
Correspondence: Professor YX Qiu, College of Life Sciences, Zhejiang University, 866 Yuhangtang Road, Hangzhou, Zhejiang 310058, People's Republic of China.
E-mail: qyxhero@zju.edu.cn

Received 7 December 2016; revised 2 February 2017; accepted 3 February 2017; published online 5 April 2017

migration patterns and more recent gene flow to the genetic structure. However, no studies have so far attempted to disentangle the relative contributions of long-term (Late Quaternary) range fragmentation vs more recent (human-induced) deforestation to the spatial distribution of genetic diversity of palaeoendemic plant species from Southwestern China. In this study, we investigated the influence of past climate changes and contemporary anthropogenic disturbance on population genetic structure and connectivity of *Dipteronia dyeriana*, a palaeoendemic species.

Dipteronia Oliv. (Angiosperm Phylogeny Group (APG), 1998; McClain and Manchester, 2001) is one of the numerous Cenozoic woody relict genera of the North Temperate floristic region that were once more species rich and widely distributed in both Asia and North America during the Paleocene and Paleocene–Oligocene, respectively (McClain and Manchester, 2001; Manchester *et al.*, 2009). Today, however, the genus just contains two extant species, *D. sinensis* Oliv. (Oliver, 1889) and *D. dyeriana* Henry (Henry, 1903), both of which are hypothesized to be palaeoendemisms in mainland China (Manchester *et al.*, 2009). Phylogenetic results are congruent with the palaeoendemic status of two extant species (Renner *et al.*, 2008). The extant distribution of *D. dyeriana*, the focal species of the present study, is restricted to the southeastern part of Yunnan Province (~6 km²) (Figure 1), although its occurrences are historically documented in both south-central Yunnan and the neighboring Province of Guizhou, but now extinct (Zhang, 2000; Su *et al.*, 2006). In

contrast, *D. sinensis* has a wider allopatric distribution in central and southern China (McClain and Manchester, 2001). *D. dyeriana* is a diploid ($2n = 18$), deciduous shrub or small tree, up to 10–15 m tall, with pinnately compound leaves (Ying *et al.*, 1993). Currently, there are only five natural populations of *D. dyeriana* in southeast Yunnan, that is, three in the county of Wenshan, and one each in the counties of Mengzi and Pingbian, respectively (Figure 1). However, there is no documented evidence that forest-clearing activities of humans since the mid-twentieth century have contributed to population extinction and/or disruption of population connectivity of this species. *D. dyeriana* is andromonoecious with flowering from April to May and fruiting between August and October. The species, at least in part, reproduces by insect-mediated outcrossing. These winged fruits are apparently dispersed by wind (Tian *et al.*, 2001). As to its overall conservation status, *D. dyeriana* is listed by both the Chinese Plant Red Book (Wang and Xie, 2004) and the IUCN (International Union for Conservation of Nature; Sun, 1998) as being 'endangered'. The geographic patterns of genetic variation in five and three natural populations of *D. dyeriana* have been surveyed previously for, respectively, inter-simple sequence repeats (ISSRs) (Qiu *et al.*, 2007) and random amplified polymorphic DNA (RAPD) markers (Li *et al.*, 2005). The results revealed high genetic differentiation (ISSRs: $G_{ST} = 0.375$; RAPD: $G_{ST} = 0.421$) and low within-population diversity (ISSRs: 0.091 for Nei's gene diversity; RAPD: 0.159 for Nei's gene diversity). This information has been critical in the conservation of

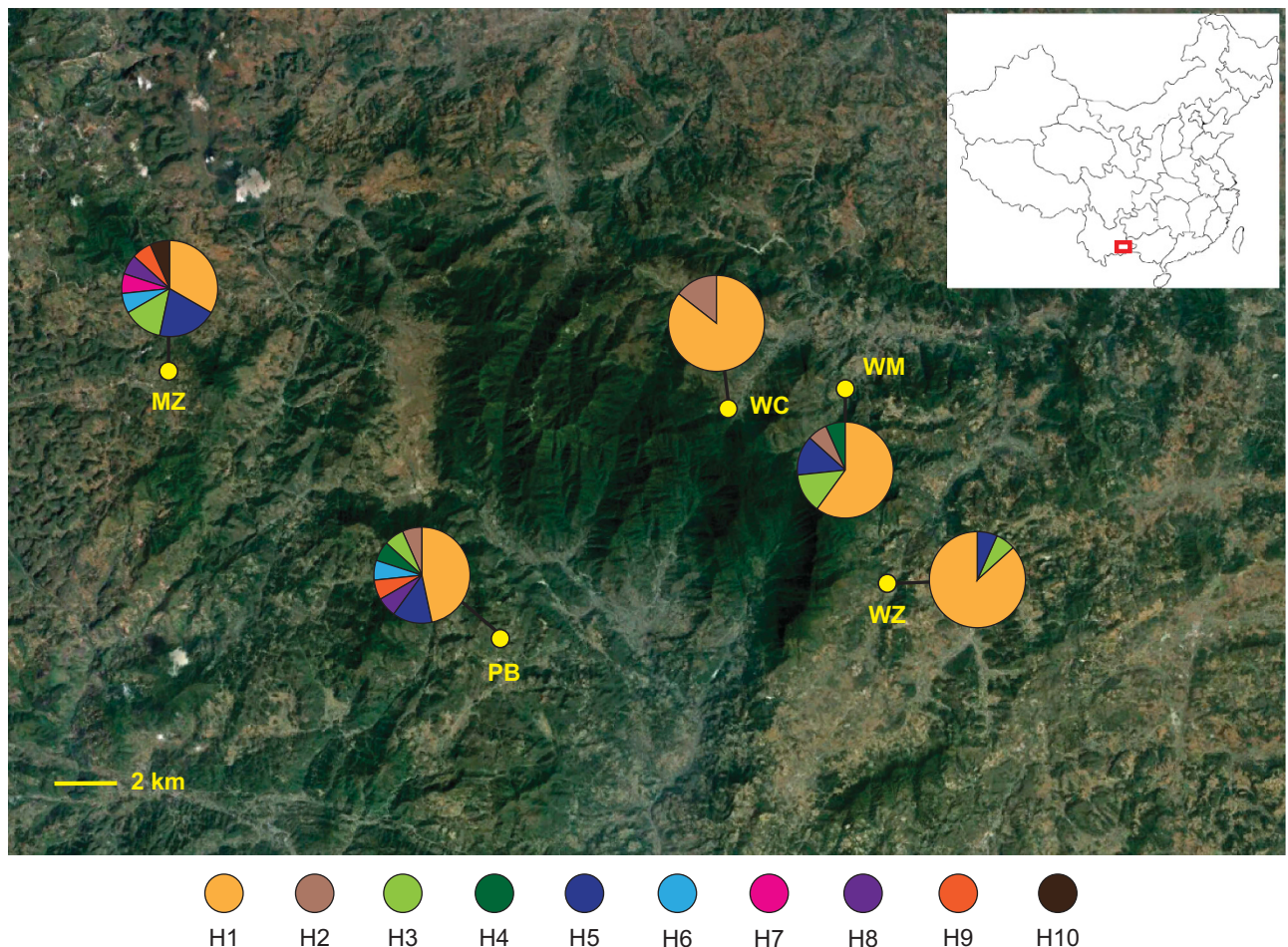


Figure 1 Location of the five extant natural populations of *Dipteronia dyeriana* in Yunnan Province, China. Also shown is the geographic distribution of 10 cpSSR haplotypes (H1–H10) found in 74 individuals from these 5 populations.

this species. However, because of their dominant nature, previous ISSR and RAPD data failed to provide detailed insights into the timescales of population divergence and the roles of contemporary vs historical processes in determining the current patterns of disjunct population remnants and genetic variation.

Here, we gathered both nuclear (codominant) and chloroplast microsatellite (SSR) data from the five natural populations of *D. dyeriana* to gain further insights into the historical and contemporary evolutionary processes shaping the genetic architecture of this paleoendemic species. This will allow us to assess the effects of anthropogenic habitat modification on this taxon, thereby informing conservation efforts towards this species. Our specific goals were to (1) characterize the overall patterns of genetic diversity as well as range-wide genetic structure of *D. dyeriana*, (2) investigate the levels of recent and historical gene flow among populations and (3) test several alternative hypotheses regarding the demographic history of the species (including changes in population size and diversification among populations over evolutionary time) using coalescent-based ABC methods.

MATERIALS AND METHODS

Sample collection

Silica-dried leaf material of *D. dyeriana* was obtained from all presently known natural populations (WZ, WM, WC, PB, MZ; Figure 1; Table 1). Approximately 14 to 15 plants were sampled from each site, with 74 individuals in total (Table 1), and these samples were previously surveyed for ISSRs (Qiu *et al.*, 2007). Voucher specimens representative of all populations sampled are stored at the Herbarium of Yunnan Laboratory for Conservation of Rare, Endangered and Endemic Forest Plants (YCP) (Qiu *et al.*, 2007). All of these 74 individuals were employed for both nuclear microsatellite (nSSR) and chloroplast microsatellite (cpSSR) analysis.

Chloroplast and nuclear microsatellite genotyping

Total genomic DNA was extracted from the dried leaf tissue using DNA Plantzol Reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. We PCR amplified 10 cpSSR regions (*Ycf5*, *Ycf5-2*, *PsbC-TrnS*, *Rm5-TrnR*, *RpoB*, *Rps3-Rp122*, *TrnR-Rm5*, *Rp12-Rp123*, *TrnL-16S*rRNA and *TrnL*) and 13 nSSR loci using primers developed specifically for *Dipteronia* (Yang *et al.*, 2008) and *D. dyeriana* (Chen *et al.*, 2011), respectively. Primers were labeled with either 6-FAM or HEX fluorescent dyes (Applied Biosystems, Foster City, CA, USA). PCR amplifications were performed on a GeneAmp9700 DNA Thermal Cycler (Perkin-Elmer, Waltham, MA, USA) using protocols described in He *et al.* (2012) and Chen *et al.* (2011), and PCR products were separated on a MegaBACE 1000 (GE Healthcare Biosciences, Pittsburgh, PA, USA). Alleles were scored manually with the aid of GENETIC PROFILER (version 2.2; GE Healthcare Biosciences).

Genetic variation and population differentiation at chloroplast SSR loci

Because of the non-recombining nature of the chloroplast genome, different alleles at each of the cpSSR loci in each individual were combined into a haplotype. For each population, we computed haplotypic diversity (H_{cp}) (Pons and Petit, 1995), the number of chlorotypes (na_{cp}) and rarefied haplotypic richness (H_R) (standardized for 14 individuals) using the program CONTRIB (version 1.02; Petit *et al.*, 1998). Population structure was analyzed by comparing two coefficients of population divergence, G_{ST} and R_{ST} , using the program PERMUT and CPSSR (version 2.0; Pons and Petit, 1996). G_{ST} is only based on haplotype frequencies, whereas R_{ST} takes into account both haplotype frequencies and haplotype similarities. Significantly higher R_{ST} versus G_{ST} values (1000 permutations; $P < 0.05$) indicate that genealogically closely related haplotypes tend to occur together within populations, implying the presence of phylogeographical structure (Pons and Petit, 1996).

Table 1 Geographic and genetic characteristics of the six *Dipteronia dyeriana* populations studied

Population code	Location	Altitude (m)	Latitude (°N)	Longitude (°E)	nSSR					cpSSR				
					N	A_R	H_O	H_E	F_{IS}	Modal F	N_e	H_R	H_{cp}	Haplotype
WC	Chenjiashai, Wenshan County, Yunnan Province	1975–2008	23°21', 103°59'		14	2.192	0.384	0.563	0.318	0.107	1.054	1.670	0.484	H1(12), H2(2)
WM	Majiazhai, Wenshan County, Yunnan Province	1967–1999	23°21', 103°60'		15	2.007	0.375	0.426	0.119	0.225	1.187	2.667	0.638	H1(9), H2(1), H3(2), H4(1), H5(2)
WZ	Zhongcaozhi, Wenshan County, Yunnan Province	2011	23°17', 103°60'		15	1.598	0.308	0.359	0.140	0.371	1.047	1.067	0.257	H1(13), H3(1), H5(1)
MZ	Laozhai, Mengzi County, Yunnan Province	1902	23°24', 103°47'		15	2.522	0.708	0.735	0.036	0.026	1.336	4.383	0.867	H1(5), H3(2), H5(3), H6(1), H7(1), H8(1), H9(1), H10(1)
PB	Pingbian County, Yunnan Province	2155–2283	23°18', 103°60'		15	2.411	0.475	0.617	0.230	0.186	1.272	4.000	0.790	H1(7), H2(1), H3(1), H4(1), H5(2), H6(1), H7(1), H8(1)
Mean/Overall					74	2.140	0.450	0.540	0.160	0.183	1.179	2.757	0.607	H1(46), H2(4), H3(6), H4(2), H5(8), H6(2), H7(1), H8(2), H9(2), H10(1)

Abbreviations: A_R , allelic richness for nSSR; cpSSR, chloroplast microsatellite; H_{cp} , haplotypic diversity; H_E , expected heterozygosity for nSSR; H_O , observed heterozygosity for nSSR; H_R , haplotypic richness; F_{IS} , the within population inbreeding coefficient for nSSR; Modal F , probability of allelic coancestry for nSSR; N_e , sample size for nSSR/cpSSR analysis; N_e , effective number of alleles for cpSSR; nSSR, nuclear microsatellite.

Genetic variation and population differentiation at nuclear SSR loci

CERVUS (version 2.0; Kalinowski *et al.*, 2007) and FSTAT (version 2.9.3.2; Goudet, 2012) were used to calculate, for each nSSR locus, the number of alleles (A), observed heterozygosity (H_O), expected heterozygosity (H_E) and the within-population inbreeding coefficient (F_{IS}). The significance of departures from Hardy–Weinberg equilibrium (HWE) was tested using a Markov chain (dememorization 1000, 100 batches, 1000 iterations per batch) in GENETOP (version 4.2; Rousset, 2008). Null allele frequencies were obtained using the program FRENA (Chapuis and Estoup, 2007) following the Expectation Maximization method described by Dempster *et al.* (1977). For each population, we estimated two measures of genetic variability, that is, allelic richness, AR (standardized for 14 individuals), H_O and H_E , as well as F_{IS} , across all loci by the software FSTAT (version 2.9.3.2; Goudet, 2012). Using the same program, differentiation among populations was calculated as F_{ST} (Weir and Cockerham, 1984).

Genetic subgroups in the entire nSSR sample were identified using the Bayesian clustering model implemented in STRUCTURE (version 2.3.4; Pritchard *et al.*, 2000). This model infers population structure by clustering individual multilocus genotypes into a given number of populations (K) by minimizing Hardy–Weinberg and linkage disequilibrium (Gao *et al.*, 2007). This Bayesian analysis was performed under the admixture model, assuming independent allele frequencies among populations. The number of K was set to vary from 1 to 5. For each value of K , we performed 10 runs with a burn-in of 10^4 and a run length of 10^5 iterations. The posterior probability of the data ($\ln P(D)$) for a given K was computed using the runs with the highest probability for each K (see Pritchard *et al.*, 2000). In addition, we used the method suggested by Evanno *et al.* (2005) to estimate a parameter (ΔK) that evaluates the second-order rate of change of the likelihood function with respect to K . To further explore the genetic relationships between populations, genetic distances (D_A , Nei *et al.*, 1983) among all pairs of populations were calculated from allele frequencies using POPTREE2 (Takezaki *et al.*, 2010). The resulting D_A matrix was subjected to a cluster analysis using the unweighted pair-group method with arithmetic means (UPGMA), and bootstrap values from 5000 replicates was performed (Sneath and Sokal, 1973). Isolation-by-distance (IBD) effects were tested based on pairwise geographical and genetic distances (F_{ST}) with the Isolation by Distance Web Service (Jensen *et al.*, 2005) (<http://ibdws.sdsu.edu/~ibdws/>), and running 10 000 permutations.

Estimating contemporary and historical gene flow

We estimated ‘contemporary’ interpopulation migration rates (over the past few generations; m_c , fraction of immigrant individuals) using a Bayesian approach in BAYESASS (version 3.03; Wilson and Rannala, 2003). We took contemporary timescale to be <5 generation for *D. dyeriana*. This method allows for deviation from HWE but assumes linkage disequilibrium and constant migration rates for two generations before sampling. To maximize the log likelihood values, Δ values were adjusted to optimize terminal proposed changes between chains (40–60% of the total iterations) to ensure sufficient parameter space was searched (Wilson and Rannala, 2003). Each Markov chain Monte Carlo run was performed with 3×10^6 iterations after a burn-in of the first 10^6 iterations. Model convergence was assessed by comparison of posterior probability densities of inbreeding coefficients and allele frequencies across 10 replicate runs, each with a different initial seed. We compared the log-probability of the each run with the program TRACER (version 1.4.8; Rambaut and Drummond, 2007). The Bayesian deviance measure was used to determine the run that displayed the best model fit (Spiegelhalter *et al.*, 2002). Historical gene flow (much longer period of time, $\sim 4 N_e$ generations in the past; Beerli and Felsenstein, 2001) was estimated using Bayesian inference in MIGRATE (version 6.0; Beerli, 2008). We estimated mutation-scaled rates of migration (M ; $M = m_h/\mu$, where m_h is the historical migration rate and μ is the mutation rate per generation (estimated for nSSRs as 10^{-3} ; Udupa and Baum 2001)). We ran MIGRATE using a Brownian motion mutation model with constant mutation rates for all loci and starting parameters based on F_{ST} calculations. We used slice sampling and uniform prior distribution to estimate M (range = 0–1000, mean = 500, and delta = 100). Following a burn-in of 50 000 iterations, each run visited a total of 1 000 000 parameter values and recorded 20 000 genealogies at a sampling increment of 50. We used a static heating scheme

at four temperatures (1, 1.5, 3 and 6) to efficiently search the genealogy space. In the results, we report the mean and 95% credible intervals for m_h .

ABC analysis of population demography

To infer the population demography of *D. dyeriana* we used ABC approach. ABC enables us to compare the different population demographic models and to estimate their parameters without calculating complex likelihood (Beaumont, 2010). Five populations were divided into two genetic groups (WC, WM and WZ for east group; MZ and PB for west group) according to the result of the STRUCTURE analysis (see details in the Results section). First, four different population size change models (Figure 2), which were nearly the same as used in Tamaki *et al.* (2016), were built and were applied to the each group (for detailed description of the four models see Supplementary Appendix S1). Generalized stepwise mutation model was used for the mutation model of nSSR and cpSSR (Estoup *et al.*, 2002). Generalized stepwise mutation model has two parameters, mutation rate per generation (μ) and geometric parameter for GSM (P). P ranges from 0 to 1 and represents the proportion of mutations that will change the allele size by more than one step. A value of zero means a strict stepwise mutation model (SMM). For nSSR, we did not use six loci that contained high frequencies of null alleles (DAG14, DAG38, DAG88, DAG327 and DAG368, see Results) and large indel (DAG307) and only used remaining seven loci in this analysis. We simulated seven independent loci whose upper limit in repeat number was set to up to 30. Mean μ among seven loci was fixed into 10^{-3} (Udupa and Baum, 2001) and each locus value of μ was randomly drawn from Gamma (shape, rate) in every simulation. Prior distribution of the shape parameter was drawn from Uniform (0.5, 5) and the rate parameter was then calculated by shape/mean μ . Prior distribution of mean value of P was drawn from Uniform (0, 1) and each locus value was randomly drawn from β (a, b). The value of a and b were calculated by $0.5 + 199 \times \text{mean } P$ and $a \times (1 - \text{mean } P) / \text{mean } P$, respectively, according to Excoffier *et al.* (2005). For cpSSR, we simulated five linked regions whose upper limit in repeat number was set to up to 5. Prior distributions of μ was drawn from log-uniform (10^{-5} , 10^{-3}), because reported values of μ in cpSSR was much lower than those in nSSR (Provan *et al.*, 1999). P was set to 0 and SMM was assumed. The same values of μ and P were used for all five linked regions. Therefore, all four models have three additional free parameters related to the mutation model, shape for nSSR, mean P for nSSR and μ for cpSSR.

All priors were generated using R (version 3.3.1; R Core Team, 2016) and simulations were conducted using FASTSIMCOAL2 (version 2.5.2.21; Excoffier and Foll, 2011). When simulating cpSSR, effective population size in cpSSR was set to half of that in nSSR because *D. dyeriana* was andromonoecious and all individuals could become a maternal tree. Therefore, the values of $2 \times N_{CUR}$ and N_{CUR} , that is, numbers of gene copies, were set to the coalescent simulator for nSSR and cpSSR, respectively. Simulations were repeated 10^6 times and summary statistics were calculated using ARLEUMSTAT (version 3.5.2; Excoffier and Lischer, 2010) and our own R script for nSSR and cpSSR, respectively, in each model and group. Average and s.d. of number of alleles, expected

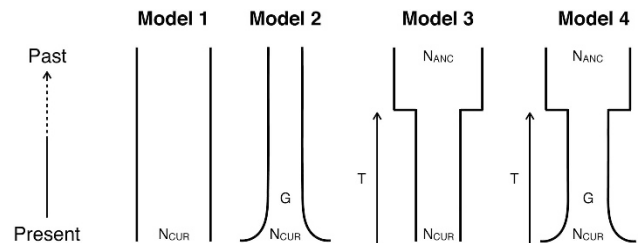


Figure 2 Compared four population size change models that were applied for each east and west group. Model 1, standard neutral model; model 2, exponential growth model; model 3, instantaneous size change model; and model 4, exponential growth after instantaneous size change model. N_{CUR} , current effective population size in the number of diploid individuals; G , growth rate ($N_T = N_{CUR} \times \exp(G \times T)$); T , time when the population size changed and is scaled in generations; N_{ANC} , ancestral effective population size in the number of diploid individuals.

heterozygosity and allele size range were used for summary statistics in nSSR. Number of haplotypes and gene diversity were used for summary statistics in cpSSR. Therefore, a total of eight summary statistics was used for the following analysis. Tolerance rate was set to 0.001 and 1000 data sets nearest to the observed data set were used for model comparison and parameter estimation. Neural network regression method implemented in ABC package (version 2.1; Wegmann *et al.*, 2010) was used for calculating posterior probability of the models and posterior distribution of the parameters (Csilléry *et al.*, 2012). Logistic transformation of parameters were applied when the estimation of the posterior distributions of them to keep the adjusted values within a prior range. Posterior mode and 95% highest posterior density were calculated using CODA package (version 0.18; Plummer *et al.*, 2006). To convert the time parameters from generation to year, we used 20 years per generation.

Second, we built the four population divergence models (Figure 3; for detailed description of the four models see Supplementary Appendix S1). For the mutation model, although there were three free parameters, shape, mean P and μ for cpSSR, in the population size change analysis, we fixed the mean P and μ for cpSSR into 0.6 and 2.2×10^{-4} , respectively, according to the results of the population size change analysis (see details in the Results section). Prior distribution of shape parameter was also drawn from Uniform (0.5, 5). Therefore, all population divergence models have one additional free parameter related to the mutation model, shape. Simulations were conducted as like a population size change analysis. We also calculated an average F_{ST} among 7 nSSR loci and F_{ST} of cpSSR haplotypes for summary statistics, and a total of 18 summary statistics were used for model comparison and parameter estimation. Model comparison and parameter estimation were also conducted as like a population size change analysis. Finally, to verify the model fitting for the observed data, posterior predictive simulations using 1000 samples randomly drawn from the posterior distribution were conducted in both population size change and divergence analyses. Summary statistics were calculated and were compared with the observed data.

Analysis of recent population bottlenecks

We used Wilcoxon's signed rank test vs the mode-shift test implemented in BOTTLENECK (version 1.2.02; Piry *et al.*, 1999) to detect population declines over extended vs more recent time scales ($\sim 2N_e - 4N_e$ generations in the past vs a few dozen generations ago; Luikart and Cornuet, 1998). In Wilcoxon's test, recently bottlenecked populations are assumed to exhibit higher observed heterozygosity than expected under mutation-drift equilibrium. In the mode-shift test, non-bottlenecked populations at mutation-drift equilibrium are expected to have a larger proportion of low-frequency alleles ($>10\%$) compared with those at intermediate frequency ('L-shaped' distribution), whereas in bottlenecked populations the situation should be reversed ('shifted mode' distribution; Luikart and Cornuet, 1998; Spencer *et al.*, 2000). We performed 10 000 simulations under the SMM and the two-phase model, with a variance of 12 as recommended by Piry *et al.* (1999). Significant P -values from Wilcoxon's test and/or 'shifted mode' distributions were taken as evidence of bottlenecks.

RESULTS

Genetic diversity and differentiation for cpSSRs

Ten cpSSR markers were used to genotype 74 individuals from 5 populations of *D. dyeriana*, yet only 5 markers (*Ycf5*, *Ycf5-2*, *Rrn5-TrnR*, *RpoB* and *Rps3-Rp122*) showed polymorphism, and the other 5 loci (*PsbC-TrnS*, *Rp12-Rp123*, *TrnL-16SsrRNA*, *TrnL* and *TrnR-Rrn5*) showed monomorphism. Here, we selected the five polymorphic loci for further analysis of variation.

When combined, the 5 polymorphic cpSSR markers identified 10 haplotypes (H1–H10; Supplementary Appendix S2; Figure 1), resulting in high total haplotypic diversity with $H_{cp} = 0.627$ at species level. The haplotypic diversity per population (H_{cp}) ranged from 0.257 (WZ) to 0.867 (MZ), with an average of 0.607 (Table 1). Similarly, we observed major variations in haplotypic richness H_R among populations, from 1.067 (WZ) to 4.383 (MZ), with an average of 2.757. Populations from western group (MZ, PB) had on average higher value of haplotypic diversity ($H_{cp} = 0.676$) than those from eastern group (WZ, WM, PC with the mean H_{cp} was 0.449). H1 was present in all populations, and most other haplotypes were found in two (that is, H4; H5, H8, H9) or more populations (that is, H2, H3, H5), except two (H7, H10) restricted to (the most diverse) population MZ. Overall levels of population differentiation were very low ($G_{ST} = 0.062$, $R_{ST} = 0.023$), whereby R_{ST} was not significantly greater than G_{ST} ($P = 0.732 > 0.05$) that thus rejected the presence of a phylogeographic structure at cpDNA in *D. dyeriana*.

Genetic diversity and differentiation for nuclear microsatellites

Screening all 74 individuals of *D. dyeriana* at the 13 nSSR loci revealed a total of 208 alleles, with 9 to 38 alleles per locus (Supplementary Appendix S3). The observed (H_O) and expected (H_E) heterozygosity per locus over all populations ranged from 0.068 to 0.649 and from 0.391 to 0.952, respectively. Of the 13 loci, 5 (DAG88, DAG14, DAG368, DAG327 and DAG38) significantly deviated from HWE after Bonferroni correction (Supplementary Appendix S3), and exceeded the threshold null allele frequency ($v = 0.15$) across all five populations. As the presence of null alleles is known to cause bias in estimators of genetic diversity and differentiation (Chapuis and Estoup, 2007), the five non-HWE loci with high frequencies of null alleles were excluded from subsequent analyses.

At the population level, average estimates of genetic diversity were generally high ($AR = 2.140$, $H_E = 0.540$, $H_O = 0.450$), being lowest in population WZ (1.598, 0.308 and 0.359) and highest in population MZ (2.522, 0.708 and 0.735; Table 1). Populations from western group had on average higher levels of diversity (for example, $H_E = 0.829$) than those from eastern group ($H_E = 0.460$). The values

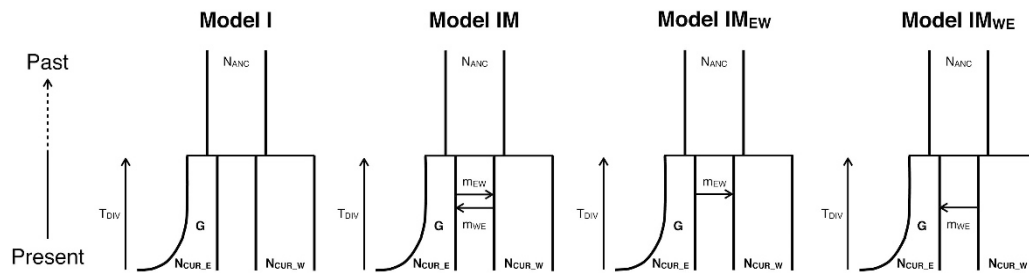


Figure 3 Compared four population divergence models. $N_{CUR,E}$ and $N_{CUR,W}$, current effective population sizes in the number of diploid individuals for east and west groups, respectively; G , growth rate ($N_t = N_{CUR} \times \exp(G \times T)$); T_{DIV} , divergence time that is scaled in generations; m_{EW} and m_{WE} , migration rate per generation from east to west and from west to east for backward in time, respectively; N_{ANCE} , ancestral effective population size in the number of diploid individuals. Fixed values were used for parameters shown in bold ($N_{CUR,E}$ and $N_{CUR,W} = 1100$, and $G = -0.0084$).

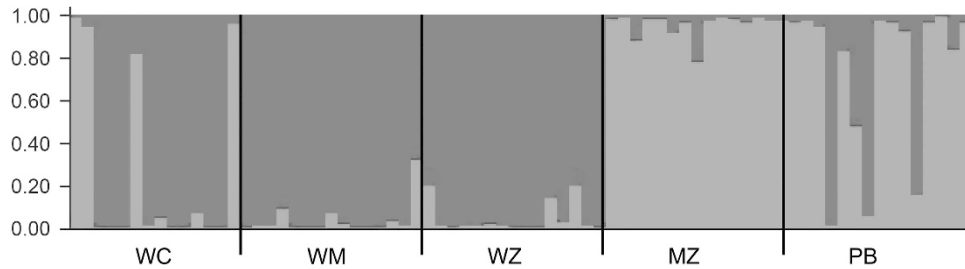


Figure 4 Histogram of the STRUCTURE analysis for the model with $K=2$ (showing the highest ΔK). Each color corresponds to a suggested cluster, and a vertical bar represents a single individual. The x axis corresponds to population codes. The y axis presents the estimated membership coefficient (Q) for each individual in the different clusters.

of inbreeding (F_{IS}) ranged from 0.036 to 0.318, with an average of 0.160 (Table 1). The overall F_{ST} value (0.140) was significantly different from zero at the species level, indicating moderate genetic differentiation among populations (Balloux and Lugon-Moulin, 2002).

In the STRUCTURE analysis, the true number of gene pools (K) in the data were not straightforward to determine from the posterior probability of the data, as values of $\ln P(D)$ increased progressively from $K=1$ to $K=9$ (Supplementary Appendix S4). In contrast, the ΔK statistic suggested a rate change in $\ln P(D)$ corresponding to $K=2$ (Supplementary Appendix S4). At $K=2$, all populations were separated into two clusters. Cluster I (red) was present at high frequency (87%) in three eastern populations (WC, WM, WZ), whereas the majority of individuals (84%) from two western populations MZ and PB were assigned to cluster II (green; Figure 4). In the UPGMA phenogram (Supplementary Appendix S5), populations tended to cluster according to geography: the three adjacent populations from the east (WC, WM, WZ) formed a strongly supported cluster (93% bootstrap support), whereas those in the west (PB, MZ), being further apart from each other (see Figure 1), were also genetically more distinct. Mantel tests of IBD revealed no significant correlation between geographical and genetic distances ($r=0.796$, $P=0.968$).

Contemporary and historical migration rates

Multiple runs of BAYESSASS yielded low levels of contemporary gene flow (m_c , fraction of individuals that are immigrants) among most populations (Table 2). Of the 20 pairwise estimates, only three showed some evidence of moderate contemporary gene flow/migration (m_c) between populations, with WZ serving as a main source of migrants (that is, from WZ to WC, WM and PB at $m_c=0.240$, 0.268 and 0.247, respectively), whereas 17 had m_c values less than 0.035 and their 95% confidence intervals overlapped zero (Table 2). Thus, when taken together, contemporary gene flow in *D. dyeriana* appears to be low to moderate and occur in an east to (north) west direction. In contrast, historical rates of migration ($m_h=M_{ij}$), as calculated in MIGRATE, were mostly significantly different from zero and generally relatively high, ranging from 0.018 to 0.071 (Table 2).

Population size change history

In the east group, model 2 showed the highest posterior probability (0.566; Table 3). Structural parameters of model 2, N_{CUR} and G , showed a clear single peak (Supplementary Appendix S6). Although model 4 also showed relatively high value of posterior probability (0.345), common parameters of these two models, N_{CUR} and G , showed nearly equal values ~ 1100 and 8.4×10^{-3} , respectively and trajectories of these two models showed nearly the same shape for the past 68 000 years (Supplementary Appendix S7). These two models

Table 2 Mean ($\pm 95\%$ CI) contemporary migration rate (m_c , fraction of immigrant individuals) estimated from BAYESSASS and historical migration rate ($m_h=M_{ij}$) estimated from MIGRATE across the five natural populations of *Dipteronia dyeriana*

	Contemporary migration rate (m_c)			Historical migration rate (m_h)		
	Mean	Lower 95% CI	Upper 95% CI	Mean	Lower 95% CI	Upper 95% CI
<i>Migration into WC</i>						
From WM	0.020	0.000	0.056	0.043	0.011	0.082
From WZ	0.240	0.170	0.311	0.062	0.033	0.100
From MZ	0.021	0.000	0.059	0.050	0.021	0.083
From PB	0.022	0.000	0.063	0.031	0.004	0.054
<i>Migration into WM</i>						
From WC	0.016	0.000	0.047	0.062	0.072	0.100
From WZ	0.268	0.211	0.324	0.051	0.018	0.085
From MZ	0.016	0.000	0.046	0.039	0.016	0.065
From PB	0.016	0.000	0.046	0.071	0.066	0.099
<i>Migration into WZ</i>						
From WC	0.017	0.000	0.047	0.056	0.029	0.085
From WM	0.017	0.000	0.048	0.053	0.024	0.097
From MZ	0.017	0.000	0.049	0.018	0.000	0.044
From PB	0.016	0.000	0.047	0.055	0.020	0.082
<i>Migration into MZ</i>						
From WC	0.021	0.000	0.062	0.044	0.022	0.080
From WM	0.019	0.000	0.055	0.030	0.008	0.061
From WZ	0.026	0.000	0.072	0.028	0.006	0.035
From PB	0.034	0.000	0.078	0.060	0.027	0.097
<i>Migration into PB</i>						
From WC	0.017	0.000	0.048	0.046	0.012	0.086
From WM	0.016	0.000	0.046	0.054	0.024	0.093
From WZ	0.247	0.183	0.311	0.046	0.027	0.075
From MZ	0.017	0.000	0.049	0.036	0.013	0.069

Abbreviation: CI, confidence interval.

indicated that east group had experienced rapid population growth from 15 000 years ago after quite low level of effective population size for more than 50 000 years ago. On the other hand, in the western group, models 1 and 3 showed almost the same levels of high posterior probabilities (0.394 and 0.373, respectively; Table 3). Structural

Table 3 Posterior model probability (P) and posterior mode and 95% HPD of parameters for the compared four population size change models for the east and west populations

Genetic group	Model	Posterior probability	Parameter (mode/95% HPD)						
			N _{CUR}	G (×10 ⁻³)	T (kya)	N _{ANC}	Shape	P	μ for cpSSR (×10 ⁻⁴)
East	Model 1	0.029	322 (134 to 937)				4.47 (1.11 to 4.99)	0.571 (0.357 to 0.748)	1.75 (0.33 to 7.39)
	Model 2	0.566	1154 (480 to 2837)	-8.31 (-26.31 to -2.53)			3.14 (1.10 to 4.98)	0.581 (0.419 to 0.730)	2.15 (0.48 to 6.23)
	Model 3	0.060	933 (438 to 2913)		1.7 (0.1 to 33.5)	217 (68 to 825)	3.55 (1.00 to 4.99)	0.591 (0.371 to 0.706)	1.79 (0.20 to 8.76)
	Model 4	0.345	1000 (529 to 2999)	-8.48 (-41.84 to -0.12)	68.7 (0.1 to 1999.3)	609 (96 to 2691)	1.49 (0.68 to 4.78)	0.570 (0.260 to 0.760)	2.00 (0.30 to 8.48)
West	Model 1	0.394	1007 (521 to 2018)				4.61 (1.51 to 5.00)	0.582 (0.407 to 0.728)	2.20 (0.74 to 7.99)
	Model 2	0.043	1643 (752 to 2994)	-2.15 (-6.39 to -0.07)			4.59 (1.74 to 4.99)	0.647 (0.511 to 0.743)	3.08 (0.89 to 6.94)
	Model 3	0.373	1194 (408 to 2890)		45.6 (0.0 to 1959.8)	754 (116 to 2706)	4.10 (1.41 to 5.00)	0.599 (0.393 to 0.731)	2.64 (0.81 to 7.89)
	Model 4	0.191	2344 (554 to 2998)	-5.87 (-42.86 to -0.10)	3.6 (0.0 to 15.3)	951 (493 to 2864)	4.50 (1.54 to 4.99)	0.613 (0.348 to 0.738)	2.70 (0.84 to 7.56)

Abbreviations: cpSSR, chloroplast microsatellite; HPD, highest posterior density; N_{ANC} , ancestral effective population size in the number of diploid individuals; N_{CUR} , current effective population size in the number of diploid individuals. Best-supported model is shown in bold.

parameters of models 1 and 3 showed a clear single peak except for T in the model 3 (Supplementary Appendix S8). N_{CUR} for these two models showed similar value, ~ 1100 . Although model 3 was instantaneous size change model, the ratio between N_{CUR} and N_{ANC} was only 1.58 ($=1194/754$), and hence they did not change (Supplementary Appendix S7). Therefore, the west group kept relatively stable effective population size for a long time. We confirmed that predictive simulations of the best-supported model in each group showed good agreement between observed and simulated data sets (Supplementary Appendices S9 and S10). In addition, regardless of the mutation model used (two-phase model or SMM), Wilcoxon's test revealed no significant heterozygote excess in any population (all $P > 0.05$; Supplementary Appendix S11), as expected in the absence of a historical bottleneck. Similarly, in the mode-shift test, all populations showed an L-shaped distribution of alleles, providing no evidence of a recent bottleneck.

Population divergence history

Model IM_{WE}, isolation with one directional migration model from the west to the east, showed the highest posterior probability of all compared models (0.737; Table 4). All posterior distributions of structural parameters of the model showed different shape from their priors (Supplementary Appendix S12). Posterior modes (95% highest posterior density) of N_{ANC} , T_{DIV} and m_{WE} were 1004 (33–2736) and 17.3 (6.7–1689.3) in kya and 0.0017 (0.0006–0.0029), respectively. We also calculated number of migrants per generation ($N_e m_{WE}$) by multiplying N_{CUR} ($=1100$) and m_{WE} , and the value was 1.9 (0.7–3.2). We confirmed that predictive simulations of the best-supported model in each groups showed good agreement between observed and simulated data sets (Supplementary Appendix S13).

DISCUSSION

Southwestern China is one of centers of temperate plant species diversity and endemism (López-Pujol *et al.*, 2011) and well-known glacial refugia for a number of Tertiary relict trees (Zou *et al.*, 2002; Qiu *et al.*, 2011), where logging and agricultural expansion has fragmented lowland forest habitats. Our study of *D. dyeriana*, a critically endangered paleoendemic, provides insights into opportunities for ecological and genetic restoration efforts, and explores how past climate changes and contemporary anthropogenic disturbance influence the genetic structure and connectivity of *D. dyeriana*, in particular and other paleoendemic species more generally.

Patterns of genetic diversity

Plant species that are endemic, restricted to islands or have relatively small population sizes often have reduced within-population diversity compared with more common and widespread species because of drift, founder events and other stochastic processes (Frankham, 1997; Cole, 2003). In the present study, our nSSR-measured H_E value ($H_E = 0.540$) is higher than average nSSR-based H_S (the equivalent to H_E in the present study) in endemic species ($H_S = 0.42$), as reviewed by Nybom (2004), and only slightly lower than the average nSSR-based H_S in regional or widespread species ($H_S = 0.65$ and 0.62, respectively). However, our nSSR-measured H_E value is comparable to other endemic, outcrossing plant species in East Asia, for example, *Michelia coriacea* (0.505; Zhao *et al.*, 2012) and *Kirengeshoma palmata* (0.60; Yuan *et al.*, 2014). As expected for codominant microsatellite markers with high mutation rates, *D. dyeriana* was found to exhibit much higher average-within population diversity at nSSRs than previously reported for anonymous dominant markers (RAPD: $H_S = 0.159$, Li *et al.*, 2005; ISSRs: $H_S = 0.093$, Qiu *et al.*, 2007). In

Table 4 Posterior model probability and posterior mode and 95% HPD of parameters for the compared four population divergence models

Model	Posterior probability	Parameter (mode/95% HPD)							
		N_{ANC}	$T_{DIV}(kya)$	$m_{EW}(\times 10^{-4})$	$m_{WE}(\times 10^{-4})$	Nm_{EW}	Nm_{WE}	β	Shape
Model I	0.006	189 (9–756)	4.0 (2.9–10.5)						3.92 (2.05–4.99)
Model IM _{EW}	0.004	240 (17–1169)	4.9 (3.6–16.2)	0.9 (0.1–9.1)		0.1 (0.0–1.0)		0.444 (0.024–0.961)	4.31 (1.29–5.00)
Model IM _{WE}	0.737	1004 (33–2736)	17.3 (6.7–1689.3)			16.9 (6.50–29.3)	1.9 (0.7–3.2)	0.875 (0.177–0.999)	4.42 (1.66–5.00)
Model IM	0.254	657 (63–2795)	35.1 (17.7–1918.9)	0.3 (0.1–7.3)	19.5 (5.5–36.7)	0.3 (0.0–0.8)	2.1 (0.6–4.0)	0.912 (0.241–0.999)	3.45 (1.29–5.00)

Abbreviations: HPD, highest posterior density; m_{EW} and m_{WE} , migration rate per generation from east to west and from west to east for backward in time, respectively; T_{DIV} , divergence time that is scaled in generations. Best-supported model is shown in bold.

Abbreviations: HPD, highest posterior density; m_{EW} and m_{WE} , migration rate per generation from east to west and from west to east for backward in time, respectively; T_{DIV} , divergence time that is scaled in generations. Best-supported model is shown in bold.

terms of cpDNA diversity, *D. dyeriana* also retains a high level of genetic diversity within population as revealed by cpSSR analysis ($H_{cp} = 0.607$; $H_R = 2.757$). The scarcity of population genetic studies based on chloroplast microsatellite data for trees or shrubs precludes a direct comparison of diversity statistics with those of the present study for this marker. However, the number of cpSSR haplotypes found in *D. dyeriana* falls between those observed in some more widely distributed angiosperms, such as *Vitellaria paradoxa* (7 haplotypes; Fontaine *et al.*, 2004), *Castanopsis hystrix* (14 haplotypes; Li *et al.*, 2007) or *Saruma henryi* (11 haplotypes; Zhou *et al.*, 2010), although comparisons among different genera must be taken with caution.

High genetic diversity maintained in rare plants is attributable to a number of factors (Zawko *et al.*, 2001), such as demographic stability, recent reduction of population size plus insufficient time for isolation or extensive, recurrent gene flow (Maguire and Sedgley, 1997; Chiang *et al.*, 2006). A better comprehension of these factors is important to understand not only the evolutionary history of species, but also their conservation perspectives (Frankham, 1995). For *D. dyeriana*, the moderate genetic diversity within populations and overall is unlikely to be attributable to ongoing or recent contemporary (pollen and seed) (refer to the results of respective gene flow analyses in BAYESASS, Table 2), but this rather points at relative demographic stability and/or expansion over historical and even more ancient timescales (Figures 2 and 3), as then discussed in the next section. In addition, none of the five extant populations seems to have suffered from recent genetic bottlenecks (see below). These results support our idea that *D. dyeriana* populations have persisted in Southwestern China for a long time, including the last glacial maximum (LGM, ~18 000 years before present). This may not be such a surprising conclusion because the latitude band in subtropical China between 23° N and 25° N, which encompasses the extant range of *D. dyeriana* (Table 1; Figure 1), has long been regarded as crucial refugia for warm-temperate ('broad-leaved') evergreen (WTE) forest species that resulted from southward range shifts during the LGM (Yu *et al.*, 2000; Qiu *et al.*, 2011). This role of refugia as long-term stores of genetic variation in plant species and, consequently, high levels of genetic diversity is widely recognized (Hewitt, 2000; Lee *et al.*, 2006; Qiu *et al.*, 2011; Sakaguchi *et al.*, 2011). Some factors that may have contributed to maintaining these high levels of genetic diversity are the outcrossing mating system of the species, the long lifespan of adult plants and their high resprouting ability (Qiu *et al.*, 2007). Considering groups of populations, western populations (MZ and PB) were found to harbor considerably more genetic diversity than eastern cluster (WC, WM, WZ) at both cpSSRs and nSSRs loci (Table 1). The larger long-term evolutionary effective population sizes in the former (see results of ABC) have likely contributed to their relatively higher diversity.

Values of F_{IS} per population ranged from 0.036 to 0.318, with an average of 0.160 (Table 1), indicating an overall excess of homozygotes in most populations (except MZ) compared with HWE expectations. In fact, a highly selfing breeding system might exist in *D. dyeriana* that is known to be self-compatible (Qiu *et al.*, 2007). Although *D. dyeriana* has samara borne in pairs and the seeds dispersed by winds, it is mainly distributed on the edges of WTE forest. Their fragmented occurrence in a forest margin habitat may act as a barrier to effective seed dispersal and reduce the potential for long-distance seed dispersal. Besides, insect pollination can also be restricted and may largely occur within populations (Zhang *et al.*, 2006). Corroborating this, our BAYESASS analyses show that contemporary levels of gene flow between populations were low (Table 2). It is worthwhile to note that our field observations have revealed that seed production in natural populations of *D. dyeriana* is generally high, whereas seedling

establishment is often surprisingly low (Z-Q Ouyang, unpublished data). This may indicate that local populations suffer from a deficit in outcross pollen, and inbreeding depression due to selfing or mating between close relatives (Qiu *et al.*, 2007).

Population structure and contemporary vs historical gene flow

The inter-population differentiation assessed with microsatellite markers ($F_{ST}=0.140$) was lower than previous estimates using RAPD ($G_{ST}=0.421$; Li *et al.*, 2005) and ISSRs ($G_{ST}=0.375$; Qiu *et al.*, 2007). For nSSR marker, homoplasy because of the high mutation rate may increase the underestimation of differentiation (Estoup *et al.*, 2002). On the other hand, the dominant property of ISSRs and RAPD markers is also responsible for a bias in the estimation of differentiation. Although 'absolute values' of differentiation may differ between markers, they lead to similar conclusions: (1) a high differentiation among populations was found both with codominant microsatellites and dominant RAPD/ISSRs markers at the small spatial scale; and (2) a predominant part of the total differentiation is due to the subdivision between the eastern and western part of the species' range (see below). The congruent results obtained with different markers most likely indicate that marker differences (mutation rates, homoplasy, estimation biases) had only a minor impact on differentiation (Mariette *et al.*, 2001).

Because of their uniparental inheritance, cytoplasmic genomes typically have a lower effective population size than nuclear genomes and are therefore expected to show a higher level of population differentiation (Birky *et al.*, 1983). Moreover, as cytoplasmic genomes are maternally inherited in most plant species, the more limited dispersal ability of seeds compared with pollen may further enhance population differentiation of cytoplasmic loci relative to nuclear loci (Ennos, 1994). As a result, F_{ST} estimates for cytoplasmic markers are often higher than for nuclear markers (see, for example, Petit *et al.*, 2005). However, in this study, *D. dyeriana* displays the opposite results. Population divergence based on the cpSSR data was unexpectedly low ($G_{ST}=0.062$). Relatively low levels of genetic differentiation in *D. dyeriana* at maternally inherited cpDNA markers can be attributed to either recent gene exchange between long-term isolated populations or recent fragmentation of a large population with recurrent gene flow. In the situation of long-term isolation, each separate population would have independently experienced genetic drift, and resulted in high frequencies of rare haplotypes in each population (Kimura *et al.*, 2002). Alternatively, if the populations were recently separated, there would be insufficient time for genetic drift to have occurred in each small fragmented population, and most populations would share common haplotypes (Printzen *et al.*, 2003). In line with the latter interpretation, except for two haplotypes (H7, H10) specific to population MZ, the remaining haplotypes were found to be shared among populations (Table 1; Figure 1). Therefore, the hypothesis of recent fragmentation of a large population with recurrent gene flow is highly probable. This hypothesis is also corroborated by the ABC analysis (see below).

In the STRUCTURE analysis (Figure 4), individuals from the eastern vs western part of the species' range were largely clustered into two groups ($K=2$). In addition, the UPGMA analysis showed much closer genetic relationships among the eastern populations (WZ, WM, WC) compared with the more disjunct populations from the west (MZ and PB) (Supplementary Appendix S5). According to our ABC analysis, the divergence time of the two groups was estimated to 865 generations ago, or 17 299 years ago when we assume a generation time of 20 years for *D. dyeriana*. Estimating generation time remains difficult in forest trees with long reproductive spans. Although

D. dyeriana bears first fruits at 10 years of age, it only reaches the canopy after ~20 years at which stage the trees acquire their full reproductive potential (Ouyang *et al.*, 2005). If we assume that the generation time is ~25 years, then the divergence time is consistent with a climate-induced vicariant event at the LGM (Clark *et al.*, 2009). Both STRUCTURE and BAYESASS analyses indicated that these populations from two groups are not completely isolated. STRUCTURE analyses showed admixture of two clusters in PB (Figure 4), whereas contemporary gene flow appears to occur in an east to west direction (Table 2). Similarly, our ABC analyses also revealed a unidirectional contemporary gene flow from eastern to western groups ($N_{em}=1.9$ (0.7–3.2)). This unidirectional contemporary gene flow is likely associated with recent postglacial range expansion of the eastern group (see below). Despite restricted contemporary gene flow between most pairs of populations, the historical rates of migration were mostly significantly different from zero and generally relatively high (Table 2). The historical estimates summarize gene flow over many generations that represents a period much longer than the time over which contemporary gene flow occurred. The observed inconsistency between contemporary and historical gene flows in this study suggests that habitat destruction mainly due to anthropogenic deforestation and land use, particularly from mid-twentieth century (Zhang, 2000), may have contributed to a reduction of population connectivity and increased genetic differentiation by ongoing genetic drift in isolated populations, possibly owing to decreased population size as a result of more recent bottlenecks (Zhang, 2000; Su *et al.*, 2006). Yet, our BOTTLENECK analyses showed no evidence of a more recent bottleneck for all studied populations. Nonetheless, considering this species has a long generation time of ~20 years and low historical effective population size (Table 3), it is possible that populations suffering a recent reduction in census size may not suffer a severe reduction in effective population size (a genetic bottleneck) (Pimm *et al.*, 1989; Yuan *et al.*, 2014). In fact, according to our ABC analyses, the effective population size of *D. dyeriana* is historically low because severe climatic oscillations during the last glacial–interglacial cycle have forced populations into small refugial areas (see below). We also found some evidence of inbreeding within *D. dyeriana* populations that may be because of small effective population size and limited contemporary gene flow.

Population history

Fossil pollen analyses have previously indicated that during the glacials, WTE forest was forced to retreat southward as far as 24° N on the Asiatic mainland (Yu *et al.*, 2000; Harrison *et al.*, 2001). The mountainous areas of subtropical China may have harbored refugial populations of WTE species through periods of adverse climatic conditions (Provan and Bennett, 2008; Qiu *et al.*, 2011). Nevertheless, the increasingly colder and arid climates at the onset of the LGM (Zhang *et al.*, 2005) would have caused further regression of subtropical forest, with grassland in the lowlands and open woodland in the upland regions (Adams and Faure, 1997; see also Malhotra and Thorpe, 2004). In line with this palaeovegetation-modeling predictions, our ABC analyses suggest the western and eastern groups of *D. dyeriana* likely persisted in a long-term refuge in Southern China since the beginning of the last glacial period ~100 000 years ago, whereas increasingly colder and arid climates at the onset of the LGM might have fostered the fragmentation of *D. dyeriana* within refugia (Ouyang *et al.*, 2001; Sun, 2002). Although both groups have similar current effective population size, western and eastern population groups experienced strikingly different histories following their initial divergence. For the western group, our ABC analyses suggest that this

group kept relatively stable effective population size in a long-term refuge. Genetic evidence of persistence of *D. dyeriana* at the western margin of the Laojunshan Nature Reserve of Yunnan Province is noteworthy, where the geographical landscape features likely restricted WTE habitats during postglacial period. In contrast, the eastern group had experienced 500-fold population expansion from before ~15 000 years ago. It could well indicate that this species actually benefited from WTE forest expansion as a result of the warmer and more humid climates promoted by a stronger Asian summer monsoon during the Holocene in China (Zheng, 2000; Wu *et al.*, 2010). Recent population expansions have previously been inferred among several plant species from China, and generally are interpreted as postglacial expansions from southern LGM refugia (Huang *et al.*, 2002, 2003; Su *et al.*, 2005).

Conclusions and conservation implications

The patterns of genetic diversity and gene flow of *D. dyeriana* appears to reflect both the species' natural population history and recent human impact. In line with this palaeovegetation-modeling predictions, coalescent-based ABC analyses suggest the western and eastern groups of *D. dyeriana* likely persisted in a long-term refuge in Southwestern China since the beginning of the last glacial period ~100 000 years ago, whereas increasingly colder and arid climates at the onset of the LGM might have fostered the fragmentation of *D. dyeriana* within refugia. Following their divergence, the western group kept relatively stable effective population size, whereas the eastern group had experienced 500-fold population expansion as a result of the warmer and more humid climates promoted by a stronger Asian summer monsoon during the Holocene. Although clear loss of genetic diversity by human activities was not suggested by this study, recent habitat fragmentation has led to a reduction of population connectivity and increased genetic differentiation by ongoing genetic drift in isolated populations, possibly owing to decreased population size in recent dozen years.

Examining wide-scale genetic structure and population demography of the endangered tree species will provide useful information on conservation units and guidelines for restoration or plantation efforts in forest conservation and breeding programs of many tree species, especially economically important ones (Tsuda and Ide, 2005; Sutherland *et al.*, 2010; Bodare *et al.*, 2013). This information provided by the present study could also be relevant to the conservation of *D. dyeriana*. A moderate genetic diversity is detected at both the species and population levels that is evenly distributed within and among populations (Table 1). This seems to rule out genetic impoverishment and inbreeding depression as immediate causes of threat. According to the clusters found with nSSRs, we suggest that *D. dyeriana* should be managed as two separate units, corresponding to the WC, WZ, WM vs MZ, PB populations, respectively. The MZ-PB unit from outside the reserve preserves the highest genetic diversity (considering both nSSR and cpSSR markers; Table 1) and the largest numbers of private alleles (see also Qiu *et al.*, 2007). Clearly, as these latter populations are facing a high risk of losing genetic variation because of ongoing habitat destruction with likely detrimental consequences for the long-term evolutionary viability of the entire species, conservation action is urgently required. Therefore, a broad genetic sample from the two units should be preserved by *ex situ* conservation programs (seed banks and botanic gardens). This will potentially allow future reintroductions or population reinforcements, whose success will heavily depend upon the genetic quality of the available *ex situ* sample (see, for example, Rita and Cursach, 2013; Fernández-Mazuecos *et al.*, 2014). In addition, although currently large-scale transplantation of individuals may not be needed in order to mitigate inbreeding over

most of the species' range, the proposed conservation units would be informative for adaptive management of these natural resources. Moreover, it is required to evaluate the ecological dynamics of not only the target species but also the whole ecosystem associated to it and to design a more practical approach of ecosystem management in this biodiversity hot spot.

DATA ARCHIVING

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.t8q1g>.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We gratefully acknowledge helpful comments made by two anonymous reviewers and the editor on a previous version of this article. This research was supported by the National Natural Science Foundation of China (Grant Nos. 31370241, 31570214) and Science & Technology Basic Resources Investigation Program of China (2017FY100100). We are grateful to Dr Zhi-Qin Ouyang for collecting plant materials.

-
- Adams JM, Faure H (1997). Preliminary vegetation maps of the world since the last glacial maximum: an aid to archaeological understanding. *J Archaeol Sci* **24**: 623–647.
- Angiosperm Phylogeny Group [APG] (1998). An ordinal classification for the families of flowering plants. *Ann Mo Bot Gard* **85**: 531–553.
- Balloux F, Lugon-Moulin N (2002). The estimation of population differentiation with microsatellite markers. *Mol Ecol* **11**: 155–165.
- Beaumont MA (2010). Approximate Bayesian computation in evolution and ecology. *Annu Rev Ecol Evol S* **41**: 379–406.
- Beerli P (2008). Migrate version 3.0: a maximum likelihood and Bayesian estimator of gene flow using the coalescent. Distributed over the Internet at <http://popgen.sc.fsu.edu/Migrate/Migrate-n.html>.
- Beerli P, Felsenstein J (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc Natl Acad Sci USA* **98**: 4563–4568.
- Birky CW, Maruyama T, Fuerst P (1983). An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* **103**: 513–527.
- Bodare S, Tsuda Y, Ravikanth G, Uma Shaanker R, Lascoux M (2013). Genetic structure and demographic history of the endangered tree species *Diospyros malabaricum* (Meliaceae) in Western Ghats, India: implications for conservation in a biodiversity hotspot. *Ecol Evol* **3**: 3233–3248.
- Chapuis MP, Estoup A (2007). Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol* **24**: 621–631.
- Chen C, Ren BB, Xu XH, Fu CX, Qiu YX (2011). Isolation and characterization of microsatellite markers for *Dipterocarpus dyeriana* (Sapindaceae), an endangered endemic species in China. *Am J Bot* **98**: e271–e273.
- Chiang YC, Hung KH, Schaal BA, GE XJ, Hsu TW, Chiang TY (2006). Contrasting phylogeographical patterns between mainland and island taxa of the *Pinus luchuensis* complex. *Mol Ecol* **15**: 765–779.
- Chiucchi JE, Gibbs HL (2010). Similarity of contemporary and historical gene flow among highly fragmented populations of an endangered rattlesnake. *Mol Ecol* **19**: 5345–5358.
- Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B *et al.* (2009). The last glacial maximum. *Science* **325**: 710–714.
- Cole CT (2003). Genetic variation in rare and common plants. *Annu Rev Ecol Evol S* **2003**: 213–237.
- Csilléry K, François O, Blum MGB (2012). abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol Evol* **3**: 475–479.
- Dempster AP, Laird NM, Rubin DB (1977). Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc B* **39**: 1–38.
- Dubreuil M, Riba M, González-Martínez SC, Vendramin GG, Sebastiani F, Mayol M (2010). Genetic effects of chronic habitat fragmentation revisited: strong genetic structure in a temperate tree, *Taxus baccata* (Taxaceae), with great dispersal capability. *Am J Bot* **97**: 303–310.
- Ennos R (1994). Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* **72**: 250–259.
- Estoup A, Jarne P, Cornuet JM (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol* **11**: 1591–1604.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620.

- Excoffier L, Foll M (2011). Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**: 1332–1334.
- Excoffier L, Laval G, Schneider S (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* **1**: 47.
- Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* **10**: 564–567.
- Fernández-Mazuecos M, Jiménez-Mejías P, Rotllán-Puig X, Vargas P (2014). Narrow endemics to Mediterranean islands: moderate genetic diversity but narrow climatic niche of the ancient, critically endangered *Naufraga* (Apiaceae). *Perspect Plant Ecol* **16**: 190–202.
- Fontaine C, Lovett PN, Sanou H, Maley J, Bouvet JM (2004). Genetic diversity of the shea tree (*Vitellaria paradoxa* CF Gaertn), detected by RAPD and chloroplast microsatellite markers. *Heredity* **93**: 639–648.
- Frankham R (1995). Conservation genetics. *Annu Rev Genet* **29**: 305–327.
- Frankham R (1997). Do island populations have less genetic variation than mainland populations? *Heredity* **78**: 311–327.
- Gao H, Williamson S, Bustamante CD (2007). A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* **176**: 1635–1651.
- Goudet J (2002). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from <https://www2.unil.ch/popgen/softwares/fstat.htm>.
- Gutiérrez-Rodríguez C, Ornelas JF, Rodríguez-Gómez F (2011). Chloroplast DNA phylogeography of a distylous shrub (*Palicourea padifolia*, Rubiaceae) reveals past fragmentation and demographic expansion in Mexican cloud forests. *Mol Phylogenet Evol* **61**: 603–615.
- Harrison SP, Yu G, Takahara H, Prentice IC (2001). Palaeovegetation: diversity of temperate plants in East Asia. *Nature* **413**: 129–130.
- He S, Wang Y, Volis S, Li D, Yi T (2012). Genetic diversity and population structure: implications for conservation of wild soybean (*Glycine soja* Sieb. et Zucc) based on nuclear and chloroplast microsatellite variation. *Int J Mol Sci* **13**: 12608–12628.
- Henry A (1903). Chinese maples. *Gard Chron* **3**: 21–22.
- Hewitt G (2000). The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907–913.
- Huang SSF, Hwang SY, Lin TP (2002). Spatial pattern of chloroplast DNA variation of *Cyclobalanopsis glauca* in Taiwan and East Asia. *Mol Ecol* **11**: 2349–2358.
- Hwang SY, Lin TP, Ma CS, Lin CL, Chung JD, Yang JC (2003). Postglacial population growth of *Cunninghamia konishii* (Cupressaceae) inferred from phylogeographical and mismatch analysis of chloroplast DNA variation. *Mol Ecol* **12**: 2689–2695.
- Jensen J, Bohonak A, Kelley S (2005). Isolation by distance, web service. *BMC Genet* **6**: 13.
- Johnson JA, Tingay RE, Culver M, Hailer F, Clarke ML, Mindell DP (2009). Long-term survival despite low genetic diversity in the critically endangered Madagascar fish-eagle. *Mol Ecol* **18**: 54–63.
- Kalinowski ST, Taper ML, Marshall TC (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol* **16**: 1099–1106.
- Kimura R, Sato K, Fujimoto R, Nishio T (2002). Recognition specificity of self-incompatibility maintained after the divergence of *Brassica oleracea* and *Brassica rapa*. *Plant J* **29**: 215–223.
- Lee YJ, Hwang SY, Ho KC, Lin TP (2006). Source populations of *Quercus glauca* in the last glacial age in Taiwan revealed by nuclear microsatellite markers. *J Hered* **97**: 261–269.
- Li J, Ge XJ, Cao HL, Ye WH (2007). Chloroplast DNA diversity in *Castanopsis hystrix* populations in south China. *Forest Ecol Manag* **243**: 94–101.
- Li S, Qian Z, Cai Y, Zhao G (2005). A comparative study on the genetic diversity of *Dipteronia sinensis* Oliv. and *Dipteronia dyeriana* Henry. *Acta Phytocological Sinica* **29**: 785–792.
- Liu Q, Chen P, He K, Kilpatrick CW, Liu SY, Yu FH et al. (2012). Phylogeographic study of *Apodemus ilex* (Rodentia: Muridae) in Southwest China. *PLoS ONE* **7**: e31453.
- Long CL, Li H, Ouyang Z, Yang X, Li Q, Trangmar B (2003). Strategies for agrobiodiversity conservation and promotion: a case from Yunnan, China. *Biodivers Conserv* **12**: 1145–1156.
- López-Pujol J, Zhang FM, Sun HQ, Ying TS, Ge S (2011). Mountains of southern China as ‘plant museums’ and ‘plant cradles’: evolutionary and conservation insights. *Mt Res Dev* **31**: 261–269.
- Luikart G, Cornuet JM (1998). Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conserv Biol* **12**: 228–237.
- Maguire TL, Sedgley M (1997). Genetic diversity in *Banksia* and *Dryandra* (Proteaceae) with emphasis on *Banksia cuneata*, a rare and endangered species. *Heredity* **79**: 394–401.
- Malhotra A, Thorpe RS (2004). A phylogeny of four mitochondrial gene regions suggests a revised taxonomy for Asian pitvipers (*Trimeresurus* and *Ovophis*). *Mol Phylogenet Evol* **32**: 83–100.
- Manchester SR, Chen ZD, Lu AM, Uemura K (2009). Eastern Asian endemic seed plant genera and their paleogeographic history throughout the Northern Hemisphere. *J Syst Evol* **47**: 1–42.
- Mariette S, Chagné D, Lézier C, Pastuszka P, Raffin A, Plomion C, Kremer A (2001). Genetic diversity within and among *Pinus pinaster* populations: comparison between AFLP and microsatellite markers. *Heredity* **86**: 469–479.
- McClain AM, Manchester SR (2001). *Dipteronia* (Sapindaceae) from the Tertiary of North America and implications for the phylogeographic history of the *Aceroidae*. *Am J Bot* **88**: 1316–1325.
- Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J (2000). Biodiversity hotspots for conservation priorities. *Nature* **403**: 853–858.
- Nei M, Tajima F, Tateno Y (1983). Accuracy of estimated phylogenetic trees from molecular data. *J Mol Evol* **19**: 153–170.
- Nybom H (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol Ecol* **13**: 1143–1155.
- Oliver D (1889). *Dipteronia sinensis*. Hooker's *Icones Plantarum*, Third Series, vol. 9. Plate 1898.
- Ouyang Z, Su W, Zhang G (2005). Studies on character of seed germination of rare plant. *Acta Bot Yunnan* **28**: 509–514. [In Chinese with English summary].
- Ouyang ZQ, Zhou JZ, Huang QY, Li DX (2001). Ex situ conservation of *Dipteronia dyeriana*, a rare and endangered plant. *Yunnan Forest Sci Tech* **96**: 3–27.
- Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005). Invited review: comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Mol Ecol* **14**: 689–701.
- Petit RJ, El Mousadik A, Pons O (1998). Identifying populations for conservation on the basis of genetic markers. *Conserv Biol* **12**: 844–855.
- Pimm SL, Gittleman JL, Mccracken GF, Gilpin M (1989). Plausible alternatives to bottlenecks to explain reduced genetic diversity. *Trends Ecol Evol* **4**: 176–178.
- Piry S, Luikart G, Cornuet JM (1999). BOTTLENECK: a program for detecting recent effective population size reductions from allele data frequencies. *J Hered* **90**: 502–503.
- Plummer M, Best N, Cowles K, Vines K (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News* **6**: 7–11.
- Pons O, Petit RJ (1995). Estimation, variance and optimal sampling of gene diversity. *Theor Appl Genet* **90**: 462–470.
- Pons O, Petit RJ (1996). Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* **144**: 1237–1245.
- Printzen C, Ekman S, Tønsberg T (2003). Phylogeography of *Cavernularia hultenii*: evidence of slow genetic drift in a widely disjunct lichen. *Mol Ecol* **12**: 1473–1486.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Provan J, Bennett KD (2008). Phylogeographic insights into cryptic glacial refugia. *Trends Ecol Evol* **23**: 564–571.
- Provan J, Soranzo N, Wilson NJ, Goldstein DB, Powell W (1999). A low mutation rate for chloroplast microsatellites. *Genetics* **153**: 943–947.
- Qiu YX, Luo YP, Comes HP, Ouyang ZQ, Fu CX (2007). Population genetic diversity and structure of *Dipteronia dyeriana* (Sapindaceae), a rare endemic from Yunnan Province, China, with implications for conservation. *Taxon* **56**: 427–437.
- 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. *Mol Phylogenet Evol* **59**: 225–244.
- R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical computing: Vienna, Austria. Available from <http://www.R-project.org/>.
- Rambaut A, Drummond AJ (2007). Tracer v1. 4: MCMC trace analyses tool. Available from <http://tree.bio.ed.ac.uk/software/tracer/>.
- Renner SS, Grimm GW, Schneeweiss GM, Stuessy TF, Ricklefs RE (2008). Rooting and dating maples (*Acer*) with an uncorrelated-rates molecular clock: implications for North American/Asian disjunctions. *Syst Biol* **57**: 795–808.
- Rita J, Cursach J (2013). Creating new populations of *Apium bermejoi* (Apiaceae), a critically endangered endemic plant on Menorca (Balearic Islands). *Anales del Jardín Botánico de Madrid* **70**: 27–38.
- Rousset F (2008). GenePop'007: a complete re-implementation of the GenePop software for Windows and Linux. *Mol Ecol Resour* **8**: 103–106.
- Sakaguchi S, Takeuchi Y, Yamasaki M, Sakurai S, Isagi Y (2011). Lineage admixture during postglacial range expansion is responsible for the increased gene diversity of *Kalopanax septemlobus* in a recently colonised territory. *Heredity* **107**: 338–348.
- Schwartz MK, Luikart G, Waples RS (2007). Genetic monitoring as a promising tool for conservation and management. *Trends Ecol Evol* **22**: 25–33.
- Sneath PH, Sokal RR (1973). *Numerical Taxonomy. The Principles and Practices of Numerical Classification*. WH Freeman: San Francisco.
- Spencer CC, Neigel JE, Leberg PL (2000). Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. *Mol Ecol* **9**: 1517–1528.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002). Bayesian measures of model complexity and fit. *J Roy Stat Soc B* **64**: 583–639.
- Su WH, Zhang GF, Ouyang ZQ (2006). Community characteristics and conservation strategies of a rare species, *Dipteronia dyeriana*. *Acta Bot Yunnan* **28**: 54–58. [In Chinese with English summary].
- Su YJ, Wang T, Zheng B, Jiang Y, Chen GP (2005). Genetic differentiation of relic populations of *Alsephila spinulosa* in southern China inferred from cpDNA *trnL-F* noncoding sequences. *Mol Phylogenet Evol* **34**: 323–333.
- Sun W (1998). *Dipteronia dyeriana*. In: IUCN 2015. The IUCN Red List of Threatened Species. Available from <http://www.iucnredlist.org/>.
- Sun H (2002). Evolution of Arctic-Tertiary flora in Himalayan-Hengduan mountains. *Acta Bot Yunnan* **24**: 671–688. [In Chinese with English summary].
- Sutherland BG, Belaj A, Nier S, Cottrell JE, P Vaughan S, Hubert J et al. (2010). Molecular biodiversity and population structure in common ash (*Fraxinus excelsior* L.) in Britain: implications for conservation. *Mol Ecol* **19**: 2196–2211.
- Takezaki N, Nei M, Tamura K (2010). POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Mol Biol Evol* **27**: 747–752.

- Tamaki I, Kuze T, Hirota K, Mizuno M (2016). Genetic variation and population demography of the landrace population of *Camellia sinensis* in Kasuga, Gifu Prefecture, Japan. *Genet Resour Crop Evol* **63**: 823–831.
- Tian X, Xu TZ, Li DZ (2001). New data of some Aceraceae plants in Yunnan, China. *Acta Bot Yunnan* **23**: 291–292. [In Chinese with English summary].
- Tsuda Y, Ide Y (2005). Wide-range analysis of genetic structure of *Betula maximowicziana*, a long-lived pioneer tree species and noble hardwood in the cool temperate zone of Japan. *Mol Ecol* **14**: 3929–3941.
- Udupa S, Baum M (2001). High mutation rate and mutational bias at (TAA)_n microsatellite loci in chickpea (*Cicer arietinum* L.). *Mol Genet Genomics* **265**: 1097–1103.
- Wang S, Xie Y (2004). *China Species Red List*. Science Press: Beijing, pp 357.
- Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010). ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* **11**: 116.
- Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
- Wilson GA, Rannala B (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**: 1177–1191.
- Wu FL, Fang XM, Miao YF, Dong M (2010). Environmental indicators from comparison of sporopollen in early Pleistocene lacustrine sediments from different climatic zones. *Chin Sci Bull* **55**: 2981–2988.
- Yang J, Li S, Sun G, Yuan Y, Zhao G (2008). Population structure and genetic variation in the genus *Dipteronia* Oliv. (Aceraceae) endemic to China as revealed by cpSSR analysis. *Plant Syst Evol* **272**: 97–106.
- Ying TS, Zhang YL, Boufford DE (1993). *The Endemic Genera of Seed Plants of China*. Science Press: Beijing.
- Yu G, Chen X, Ni J, Cheddadi R, Guiot J, Han H *et al.* (2000). Palaeovegetation of China: a pollen data-based synthesis for the mid-Holocene and last glacial maximum. *J Biogeogr* **27**: 635–664.
- Yuan N, Sun Y, Comes HP, Fu CX, Qiu YX (2014). Understanding population structure and historical demography in a conservation context: population genetics of the endangered *Kirengeshoma palmata* (Hydrangeaceae). *Am J Bot* **101**: 521–529.
- Zawko G, Krauss SL, Dixon KW, Sivasithamparan K (2001). Conservation genetics of the rare and endangered *Leucopogon obtectus* (Ericaceae). *Mol Ecol* **10**: 2389–2396.
- Zellmer AJ, Knowles LL (2009). Disentangling the effects of historic vs contemporary landscape structure on population genetic divergence. *Mol Ecol* **18**: 3593–3602.
- Zhang HH (2000). *The Threatened Wild Plants in Guizhou Province*. China Forest Press: Beijing, pp 73–74.
- Zhang Q, Chiang TY, George M, Liu JQ, Abbott RJ (2005). Phylogeography of the Qinghai–Tibetan Plateau endemic *Juniperus przewalskii* (Cupressaceae) inferred from chloroplast DNA sequence variation. *Mol Ecol* **14**: 3513–3524.
- Zhang XP, Li XH, Qiu YX (2006). Genetic diversity of the endangered species *Kirengeshoma palmata* (Saxifragaceae) in China. *Biochem Syst Ecol* **34**: 38–47.
- Zhao X, Ma Y, Sun W, Wen X, Milne R (2012). High genetic diversity and low differentiation of *Michelia coriacea* (Magnoliaceae), a critically endangered endemic in southeast Yunnan, China. *Int J Mol Sci* **13**: 4396–4411.
- Zheng Z (2000). Vegetation and climate since the late Pleistocene in southern China. *J Geosci China* **2**: 7–20.
- Zhou TH, Qian ZQ, Li S, Guo Z G, Huang ZH, Liu ZL *et al.* (2010). Genetic diversity of the endangered Chinese endemic herb *Saruma henryi* Oliv. (Aristolochiaceae) and its implications for conservation. *Popul Ecol* **52**: 223–231.
- Zou XH, He P, Chen JM, Ma HJ (2002). Studies on floristic elements of rare and endangered plants and national conservative plants in Yunnan. *Southwest China Normal Univ (Nat Sci)* **27**: 939–944. [In Chinese with English summary].

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)