Journal of Biogeography (*J. Biogeogr.*) (2014) **41**, 1710–1720

ORIGINAL ARTICLE

Phylogeography of a widespread Asian subtropical tree: genetic east–west differentiation and climate envelope modelling suggest multiple glacial refugia

Miao-Miao Shi[1]*, Stefan G. Michalski[1], Erik Welk[2], Xiao-Yong Chen[3] and Walter Durka[1]

[1]Department of Community Ecology (BZF), Helmholtz Centre for Environmental Research – UFZ, D-06120 Halle, Germany, [2]Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg, D-06108 Halle, Germany, [3]School of Resources and Environmental Sciences, Tiantong National Station for Forest Ecosystem, East China Normal University, Shanghai 200062, China

ABSTRACT

Aim Fossil-based biome reconstructions predict that during the Last Glacial Maximum (LGM), the subtropical zone of East Asia was reduced to a narrow southern belt. In contrast, previous phylogeographical studies of subtropical plant species, many of which are rare, indicated different glacial refugia north of this predicted area. Here, we aim to elucidate the phylogeographical structure and putative refugia of *Castanopsis eyrei*, a widely distributed tree of subtropical evergreen broad-leaved forests of China.

Location Subtropical China.

Methods We compiled distribution data and employed climate envelope model projections to predict potential areas at the LGM. Microsatellite data and chloroplast DNA (cpDNA) sequence data were obtained for 31 populations sampled throughout the species' range. Microsatellites were analysed with Bayesian clustering. Relationships among cpDNA haplotypes were depicted in a statistical parsimony network. We analysed patterns of variation within and among populations and clusters and along latitudinal clines.

Results Modelling revealed a potential LGM distribution of *C. eyrei* in a broad but interrupted belt overlapping the southern part of the present range. Nuclear microsatellites revealed two main clusters, suggesting a split between the western and eastern range, and a south-to-north decline in genetic variation. The eastern cluster harboured significantly higher nuclear genetic diversity. Sixteen closely related cpDNA haplotypes were identified. Populations were strongly differentiated at cpDNA markers, but lacked phylogeographical structure. Both data sets revealed higher genetic differentiation in the western cluster than in the eastern cluster.

Main conclusions Our results suggest at least two putative refugia during the LGM, further refugia-within-refugia substructure and a post-glacial northwards recolonization. Topographical differences between the mountainous western and the lowland eastern refugia may have affected the patterns of genetic differentiation between the extant populations. Incongruence between nuclear and chloroplast data might be attributed to ancestral polymorphism of cpDNA and chloroplast capture, but does not contradict the hypothesis of multiple refugia. Our results are likely to represent a template for evolutionary history and phylogeography in this region.

Keywords

***Castanopsis eyrei*, chloroplast capture, climate envelope modelling, genetic structure, glacial refugia, Last Glacial Maximum, microsatellites, phylogeography, subtropical China.**

*Correspondence and current address: Miao-Miao Shi, South China Botanical Garden, Chinese Academy of Sciences, Xingke Road 723, Guangzhou 510650, China
E-mail: mmshi@scbg.ac.cn

http://wileyonlinelibrary.com/journal/jbi
doi:10.1111/jbi.12322

INTRODUCTION

The present-day distribution of plant populations is mainly determined by environmental factors, but also by historical processes (Avise, 2000). The repeated dramatic climate changes in the Quaternary have resulted in multiple contraction–expansion processes that have profoundly shaped the geographical patterns and current genetic structure in many species (Hewitt, 2004). Phylogeographical studies can shed light on the effects of climate changes on species distributions and help to unravel such historical processes (Avise, 2000; Hickerson *et al.*, 2010). Most studies have been performed for species of the temperate zone, in particular in Europe and North America (Taberlet *et al.*, 1998; Avise, 2000; Weiss & Ferrand, 2007), as well as in Asia (Cheng *et al.*, 2005; Hiraoka & Tomaru, 2009; López-Pujol *et al.*, 2011). Despite their highly diverse flora, however, subtropical areas have not been adequately studied. The Asian subtropics are a global biodiversity hotspot (Barthlott *et al.*, 2005) and are considered to be one of the most important refugia for lineages that evolved prior to the late Pliocene and Pleistocene glaciations (Axelrod *et al.*, 1996). The evidence for effects of the Quaternary on the phylogeography of taxa in the Asian subtropics is, however, surprisingly limited (Qiu *et al.*, 2011).

In China, the subtropical zone ranges from 34° N to 22° N and is characterized by evergreen broad-leaved forests (EBLF) (Wu, 1980). This region was not covered by large ice sheets during the Last Glacial Maximum (LGM) (Hewitt, 2000; Shi, 2002), but nevertheless underwent complex changes of climate and vegetation throughout the last ice-age cycles (Qiu *et al.*, 2011). At the LGM, the climate was cooler than today, by 4–6 °C (Zheng *et al.*, 2003), and markedly drier, by *c.* 400–600 mm yr^{-1} (Qiu *et al.*, 2011; Lu *et al.*, 2013). According to global climate circulation models and derived LGM biome maps, EBLFs were forced to retreat southwards into the current tropical zone (Ni *et al.*, 2010; Prentice *et al.*, 2011; Qiu *et al.*, 2011). Thus, potential refugia for EBLF have been hypothesized in the southernmost mainland regions of China (Qiu *et al.*, 2011). After the LGM, species would have moved northwards again and should reveal signs of northward expansion. However, LGM biome maps and global climate circulation models cannot resolve conditions at smaller scales, such as those within mountain ranges.

A number of empirical studies have revealed multiple isolated refugia within subtropical China that are located outside of the predicted refugia of subtropical vegetation (Liu *et al.*, 2012), mainly in the mountain regions of the Yunnan–Guizhou (Yungui) Plateau (Shen *et al.*, 2005), the Nanling Mountains and the far east of subtropical China (e.g. the Tianmu Mountains; Yan *et al.*, 2007; Gong *et al.*, 2008; Wang *et al.*, 2009; Zhou *et al.*, 2010). Most of these studies have focused on endangered species with narrow distribution ranges or on coniferous species. In order to reveal a more complete and general pattern of the phylogeography of the subtropical flora, common and widespread species of the subtropical biomes need to be investigated using both molecular markers and independent distribution-modelling approaches (Espíndola *et al.*, 2012; Hampe *et al.*, 2013). Climate envelope models relate current presence data of species to spatial patterns of environmental variables in order to infer models of climatic tolerances. In combination with down-scaled palaeoclimate reconstructions, climatic envelope models enable the hindcasting of potential refugia of a particular species and can thus generate patterns that can be evaluated using molecular data (Waltari *et al.*, 2007; Werneck *et al.*, 2011).

Castanopsis eyrei (Champ. ex Benth.) Tutch. (Fagaceae) is one of the dominant tree species in late successional EBLF in subtropical China. It is monoecious and is pollinated by wind and insects. The acorn seeds are predominantly dispersed by gravity and small rodents (Li & Jin, 2006), presumably leading to spatially more restricted gene flow by seeds than by pollen. *Castanopsis eyrei* occurs from 300 to 1700 m a.s.l. (Huang *et al.*, 1999). It is frequent in south-eastern China and more scattered in the south-west (Fig. 1). Whereas the eastern part of the distribution range of *C. eyrei* is relatively flat, leading to a more coherent distribution, the western part is characterized by more complex topography, with numerous mountains and including unsuitable karst habitats, resulting in a fragmented distribution. Thus, gene flow among populations, especially by seed dispersal, is likely to be more obstructed in the west, and higher genetic differentiation is expected.

In this study, we used climatic envelope modelling combined with down-scaled high-resolution estimates of LGM climate parameters, maternally inherited cpDNA and biparentally inherited nuclear microsatellite loci to investigate the phylogeography of *C. eyrei*. In particular, we test whether southern refugia suggest retreat and recolonization or whether northern glacial refugia suggest *in situ* survival. We ask: (1) Which areas were climatically suitable for *C. eyrei* during the LGM? (2) Is there evidence for the differentiated gene pools indicative for multiple refugia? (3) Are there clines of genetic diversity indicating refugia and postglacial colonization?

MATERIALS AND METHODS

LGM distribution model

To identify potential glacial refugia of *C. eyrei*, we calibrated climatic envelope models using georeferenced native presence records for the species based on herbarium specimen data and the Maxent algorithm. A description of the compilation of occurrence information of *C. eyrei* is given in Appendix S1 in the Supporting Information. A cross-validated model was then projected onto scenarios of the Pleistocene LGM about 21,000 years ago using MAXENT 3.3.3k (Phillips *et al.*, 2006; Elith *et al.*, 2011) (see Appendix S1 for details of the climate envelope modelling). In short, we compiled occurrence information and WorldClim 1.4 bioclimatic variables (Hijmans *et al.*, 2005; available at http://www.worldclim.org/). We used

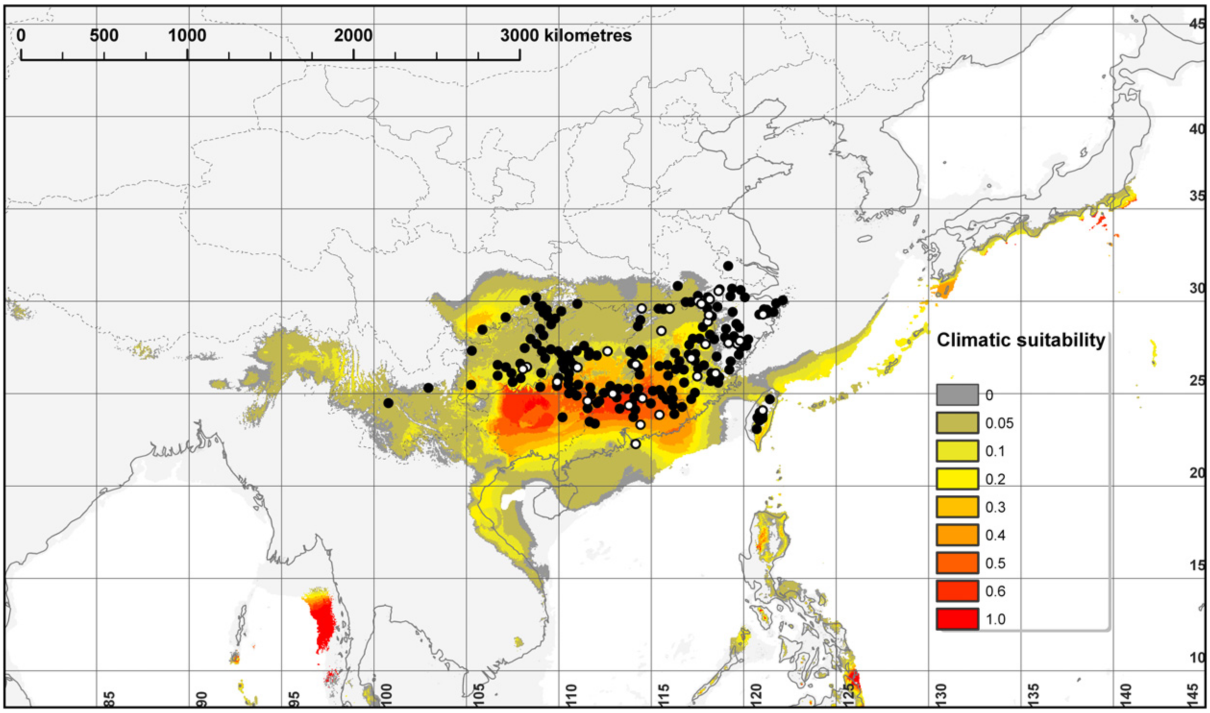

Figure 1 Current distribution (black dots) and mean predicted climatic suitability during the Last Glacial Maximum (LGM) for *Castanopsis eyrei* across Southeast Asia based on MAXENT modelling using the CCSM LGM climate scenario. Climatic suitability is indicated as a gradient from red (high suitability) to yellow (low suitability) and grey (no suitability). White dots with black outlines represent the sampling sites used in this study.

principal components analysis in SYSTAT 13 (Systat, 2009) to reduce collinearity among predictors. Potential refugia were mapped by projecting the resulting potential distribution onto the LGM using palaeoclimatic layers simulated under two general atmospheric circulation models: the Community Climate System Model (CCSM; Collins *et al.*, 2006) and the Model for Interdisciplinary Research on Climate (MIROC; Hasumi & Emori, 2004).

Population sampling and genotyping

We selected 31 natural populations of *C. eyrei* from across the species' range (Table 1, Fig. 2; see Table S1 in Appendix S1 for details). *Castanopsis eyrei* is a morphologically well-defined species and hybridization with sympatric congeners is unknown, except for hybrids with *Castanopsis lamontii* known from Shanghang, Fujian Province (Huang *et al.*, 1999). In each population, leaves from 2–31 (mean 24) individuals were collected and dried with silica gel. As putative outgroups for the cpDNA analysis, we sampled two individuals from each of the sympatric species *Castanopsis fargesii* Franch., *Castanopsis carlesii* (Hemsl.) Hayata, *Castanopsis tibetana* Hance and *Castanopsis sclerophylla* (Lindl. & Paxton) Schottky in Gutianshan Nature Reserve (29°0′19″ N, 118°03′50″ E).

All samples of *C. eyrei* were genotyped at eight nuclear microsatellite loci following Shi *et al.* (2011). A total of 271 individuals, including eight outgroup samples, were sequenced for two chloroplast intergenic spacer regions: *trn*T–*trn*L (Taberlet *et al.*, 1991) (redesigned reverse primer: 5′-TCGAAGATCCAGAGTTGATCC-3′) and *pet*G–*trn*P (Hwang *et al.*, 2000). The mean sample size per population was 8.5 (Table 1; see Appendix S2 for detailed laboratory protocol). DNA sequences were deposited in GenBank, under accession numbers JX215141–JX215239.

Data analysis

Identification of genetic clusters from microsatellites

To determine spatial genetic structure, a Bayesian clustering approach was employed using STRUCTURE 2.3.3 (Pritchard *et al.*, 2000). STRUCTURE probabilistically assigns individual genotypes to K gene pools and estimates the posterior probability of the data, given the value of K and assuming Hardy–Weinberg equilibrium. An admixture model was run with correlated allele frequencies allowing individuals to be jointly assigned to two or more gene pools if their genotypes indicate that they are admixed. Each run was pursued for 1,000,000 Markov chain Monte Carlo (MCMC) cycles with a burn-in of 100,000. The algorithm was run 10 times for each value of K from 1 to 10. The mean log-likelihood for each K, $[\ln \Pr(X|K)]$ and ΔK were used to estimate the most likely number of clusters, following Evanno *et al.* (2005). As with the whole dataset, two clusters were identified and we repeated the STRUCTURE analysis independently for these two clusters.

Table 1 Details of population locations, sample sizes and parameters of genetic diversity for each population of *Castanopsis eyrei* in China.

		SSR				cpDNA			
Sampling sites	Population ID	n	A	A_{R_13}	H_E	n	Haplotypes	h	$\pi \times 10^3$
Datang town	DT	22	9.9	7.1	0.78	9	H8 (1), H9 (1), H10 (7)	0.417	2.233
Fangxiang town	FX	15	8.8	7.4	0.81	8	H1 (8)	0	0
Huaping Nature Reserve	HP	30	10.6	6.5	0.71	8	H8 (2), H9 (2), H11 (4)	0.714	2.441
Shunhuangshan	SH	19	4.0	3.5	0.57	8	H1 (6), H3 (2)	0.429	0.862
Guposhan	GP	25	11.9	7.6	0.75	9	H1 (5), H8 (4)	0.556	0.559
Nanyue	NY*					3	H2 (3)	0	0
Mangshan Nature Reserve	MS	18	11.0	8.4	0.84	9	H1 (9)	0	0
Yingde	YD	24	11.3	7.4	0.79	9	H15 (9)	0	0
Hongkong	HK*					2	H8 (2)	0	0
Jinggangshan	JG	30	14.8	9.0	0.84	10	H12 (10)	0	0
Xiangtoushan	XT	23	13.9	9.3	0.86	10	H1 (8), H2 (2)	0.356	0.358
Tongshan County	TS*					3	H1 (1), H3 (1), H14 (1)	1.000	1.341
Quannan Maoshan Forestry	QN	21	5.8	4.8	0.70	10	H4 (1), H13 (9)	0.200	1.005
Qimuzhang Nature Reserve	QM	29	9.1	6.5	0.79	9	H1 (7), H2 (2)	0.389	0.391
Yuhuashan	YH	13	5.8	5.3	0.70	13	H16 (13)	0	0
Lushan	LS	31	11.6	7.8	0.80	13	H1 (13)	0	0
Taining	TN*					7	H2 (7)	0	0
Tianbaoyan Nature Reserve	TB	15	11.0	8.9	0.89	9	H2 (8), H5 (1)	0.222	0.224
Guniujiang Nature Reserve	GN	19	9.6	7.8	0.84	9	H1 (6), H3 (3)	0.500	1.006
Chawan Nature Reserve	CW	22	10.0	7.3	0.81	9	H3 (1), H4 (8)	0.222	1.118
Wuyishan	WY	29	11.5	7.5	0.80	10	H1 (10)	0	0
Sanqingshan	SQ	29	13.9	8.6	0.86	10	H1 (10)	0	0
Gutianshan Nature Reserve	GT	30	12.4	8.1	0.83	8	H3 (8)	0	0
Huangshan	HS	22	10.8	7.6	0.83	12	H1 (2), H2 (2), H3 (8)	0.546	0.793
Youxi County	YX	29	15.0	9.2	0.87	9	H2 (4), H6 (5)	0.556	1.118
Banqiao Nature Reserve	BQ	19	8.8	6.6	0.71	8	H2 (8)	0	0
Baishanzu Nature Reserve	BS	31	13.6	8.3	0.83	10	H1 (6), H2 (4)	0.533	0.538
Wuyanling Nature Reserve	WL	25	12.3	8.4	0.85	9	H2 (9)	0	0
Wencheng County	WC	20	11.9	8.5	0.84	9	H2 (1), H3 (5), H6 (1), H7 (2)	0.694	2.068
Taiwan	TW*					3	H8 (3)	0	0
Tiantai	TT	24	10.3	7.3	0.79	8	H2 (7), H5 (1)	0.250	0.252
mean		23.6	10.7	7.5	0.79	8.5		0.245	0.526
overall		614	28.1	10.1	0.80	263		0.814	1.750

n, sample size; A, number of alleles per locus; A_{R_13}, allelic richness based on 13 samples; H_E, expected heterozygosity; h, haplotype diversity; π, nucleotide diversity; *populations with small sample size ($n < 10$) were excluded when analysing microsatellite data. Values in parentheses indicate the frequency of each haplotype.

Nuclear genetic diversity and differentiation

Genetic variation was assessed at the species level, as gene diversity in the total population (H_T) and average gene diversity within populations (H_S) (Nei, 1987), and at the level of genetic clusters and populations, as the mean number of alleles per locus (A), allelic richness (A_R, correcting for sample size by rarefaction) and gene diversity (H_E). These calculations were performed using FSTAT 2.9.3.2 (Goudet, 1995), excluding populations with small sample sizes ($n < 10$) in order to minimize sample-size effects. We compared genetic diversity and differentiation (see below) between clusters by permuting populations 1000 times in FSTAT. We tested for the correlation of genetic variation with geographical gradients by performing a multiple regression with backward elimination of geographical factors (latitude, longitude and cluster factor: western or eastern) in R 2.14 (R Core Team, 2012). We determined levels of population differentiation (F_{ST}; Weir & Cockerham, 1984) in FSTAT and standardized genetic differentiation as $F'_{ST} = F_{ST}/F_{STmax}$ (Hedrick, 2005). F_{STmax} was calculated with FSTAT after recoding the data using RecodeData 0.1 (Meirmans, 2006).

Diversity and differentiation at chloroplast DNA sequences

The sequences of the two cpDNA markers were combined and aligned in BioEdit 7.0.5.0 (Hall, 1999). Three variable mononucleotide repeats were removed from the alignment. Insertions and deletions (indels) were treated as single mutations. The relationships among haplotypes were visualized as a statistical parsimony network computed with TCS (Clement *et al.*, 2000). We calculated population-level haplotype diversity (h), nucleotide diversity (π) and pairwise differentiation (F_{ST}) using ARLEQUIN 3.5 (Excoffier & Lischer, 2010). We

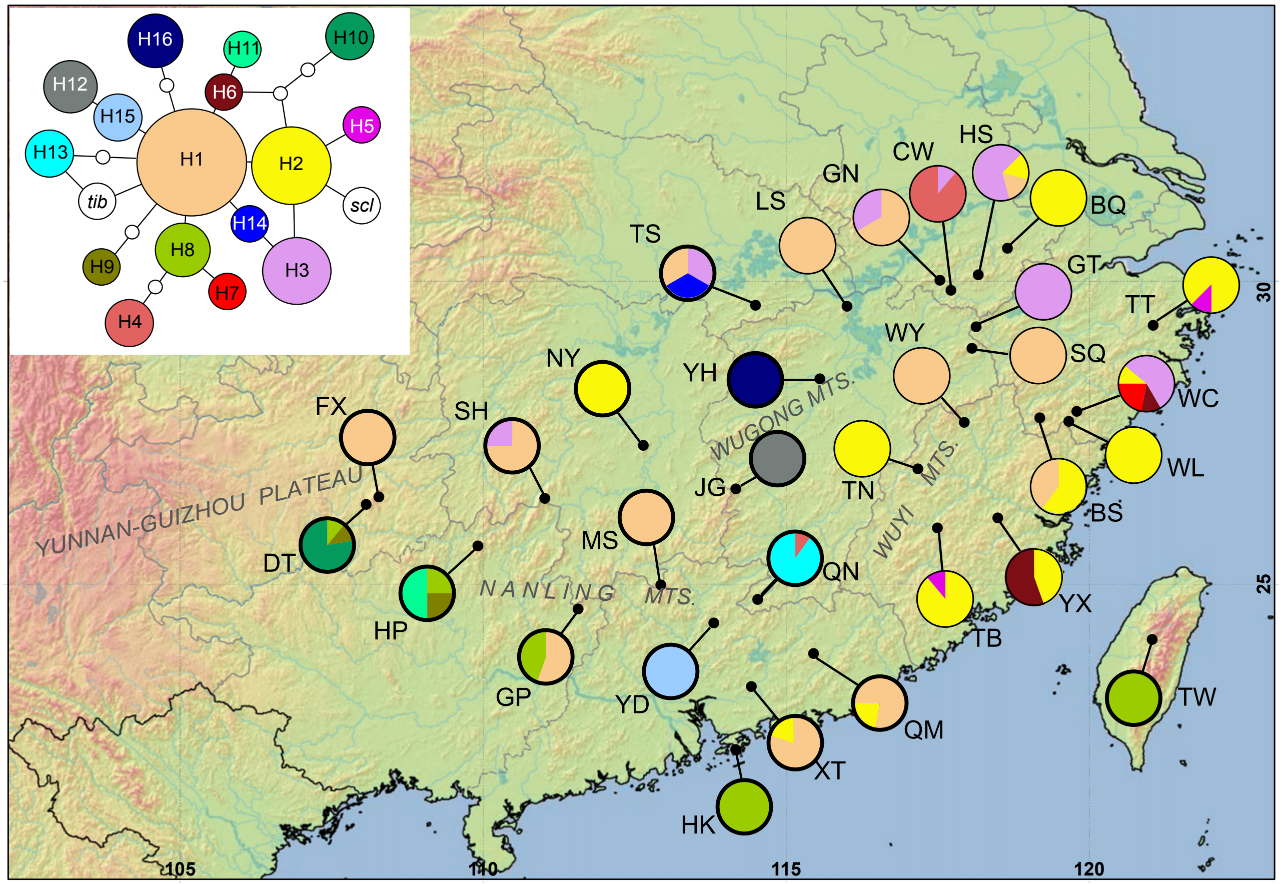

Figure 2 Locations of *Castanopsis eyrei* populations sampled in this study and the geographical distribution and frequency of chloroplast DNA (cpDNA) haplotypes H1 to H16 in subtropical China. Bold and thin outlines indicate the two genetic clusters, western and eastern, respectively, resulting from STRUCTURE analysis of microsatellites (see Fig. 3). Colours of haplotypes correspond to those of the small figure in the left corner, where the statistical parsimony network of cpDNA haplotypes is shown. Size of circles indicates haplotype frequency. Open circles indicate missing haplotypes. Note that *C. fargesii* and *C. carlesii* shared H2; *tib* and *scl* correspond to haplotypes of *C. tibetana* and *C. sclerophylla*, respectively.

computed mean within-population (h_S) and total gene diversity (h_T) based on unordered alleles, and also the equivalent parameters (v_S and v_T) based on ordered alleles, as well as differentiation G_{ST} for unordered and N_{ST} for ordered alleles with PERMUT 1.0 (Pons & Petit, 1996). A higher N_{ST} than G_{ST} usually indicates the presence of phylogeographical structure. The comparison of G_{ST} and N_{ST} was conducted based on 2000 random permutations. Pollen-to-seed migration ratio (r) was estimated using Ennos' (1994) method. We tested for isolation by distance by evaluating the significance of the correlation between pairwise genetic differentiation and log geographical distances with a Mantel test in R.

RESULTS

LGM projection of current climate distribution models

The cross-validation of the climate envelope models revealed a high mean model fit with AUC = 0.91 (SD 0.028). Estimation of the relative contributions of the environmental variables to the Maxent model suggests that winter coldness, seasonality and summer drought (monsoon activity) are the most important. The resulting potential distribution, based on climate, indicates a number of lowland and hill areas in China (especially in western Guizhou, north-eastern Hunan, northern Jiangxi and north-western Guangxi) that are climatically suitable, but where the species is currently not known to occur (Fig. S1 in Appendix S1).

The two global circulation model simulations of the LGM climate (CCSM and MIROC) revealed strongly dissimilar inferences regarding the potential palaeodistribution of *C. eyrei*. The CCSM model (Fig. 1, Fig. S2 in Appendix S1) inferred a broad southern belt and two geographically separated regions with high suitability expanding northwards into the current distribution, one in the east, in the borderlands of southern Jiangxi, south-eastern Zhejiang and northern Fujian, and one in the west, stretching from the northern Nanling Mountains in north-eastern Guangxi to southern Hunan. In contrast, the MIROC model (Fig. S2b) inferred a single large area along the coast of eastern China including the East China Sea shelf and a second area mainly in Hunan

Province. Further areas of putative high suitability, but without contact to the current distribution are located in northern Vietnam, north-eastern India and Myanmar. A considerable difference between the CCSM and the MIROC projection is the portion of the current occurrences (4% vs. 55%) that are located within the potential LGM range with medium to high (> 0.5) suitability values. In CCSM, 92% of the current occurrences had low (< 0.4) suitability values and are thus likely to be located outside LGM refugia, indicating considerable range shifts. In contrast, only 16% of the current occurrences are assigned low suitability in MIROC, suggesting overall range stability. A common feature of both models, however, is the inference of two suitable regions in the west and east, respectively.

Genetic population structure at nuclear microsatellites

The most likely number of clusters using Bayesian cluster analysis was two (Fig. S3 in Appendix S3). These largely corresponded to a split between western and eastern parts of the distribution range (Fig. 3). Some populations in the contact area of the two clusters showed mixture between gene pools (JG and QN; see Table 1 for population identifiers). Both main clusters were further subdivided into two subclusters (Fig. 3). Three populations in the Nanling Mountains (HP, SH and GP) formed a coherent subcluster within the western cluster and, in the eastern cluster, populations TB, SQ, YX and BS formed a subcluster. While this further subdivision was very pronounced in the western cluster, the two subclusters within the eastern cluster showed considerable mixture in many populations (Fig. 3).

Nuclear genetic diversity and differentiation

Across all microsatellite loci, high levels of gene diversity (H_T = 0.88; mean H_S = 0.80) were observed (Table 1), with a significant overall population differentiation of F_{ST} = 0.097 (range: 0.057–0.149 across loci). The standardized genetic differentiation (F'_{ST}) was 0.443 and was thus much higher than F_{ST}. Genetic diversity at the population level (A_R and H_E) was higher in the eastern cluster than in the western cluster (Table 2), although total genetic diversity was similar ($H_{T\ west}$ = 0.866; $H_{T\ east}$ = 0.867). Accordingly, populations in the west were more strongly differentiated than those in the east (Table 2). Whereas the western subclusters did not differ in genetic diversity and differentiation (P > 0.2), in the east, the subcluster consisting of TB, SQ, YX and BS harboured higher genetic diversity (A_R: P = 0.003), but differentiation did not differ between the subclusters (F_{ST}: P = 0.262). In the multiple regression model of genetic variation, latitude and the cluster factor were retained, indicating that the genetic variation was significantly different in the two clusters (P = 0.001) and affected by latitude (P = 0.018). In the eastern cluster, allelic richness decreased with latitude, whereas in the western cluster, the decrease was not significant (Fig. 4). A significant pattern of isolation by distance was found for all populations (r = 0.407, P = 0.001) and in both the eastern and the western cluster (Fig. S4 in Appendix S3).

Table 2 Comparison of microsatellite genetic diversity within and differentiation among populations of *Castanopsis eyrei* in the western and eastern cluster within China.

Parameters	Western	Eastern	P
A_R	6.890	8.000	0.036
H_E	0.764	0.823	0.019
F_{ST}	0.122	0.052	< 0.001
F'_{ST}	0.542	0.309	< 0.001

A_R, allelic richness based on 13 samples; H_E, expected heterozygosity; F_{ST}, Weir & Cockerham's genetic differentiation; F'_{ST}, standardized genetic differentiation; P-value indicates the significance of the difference between the western and eastern group.

Genetic variation at chloroplast DNA sequences

Alignment lengths in *C. eyrei* were 568 bp for *trn*T–*trn*L and 474 bp for *pet*G–*trn*P. In total, 19 polymorphic informative sites, comprising 12 point mutations and seven indels, were detected, defining 16 haplotypes (Table S3 in Appendix S2). The geographical distribution of haplotypes is shown in Fig. 2. The most common haplotypes, H1 and H2, were found in 13 (42%) and 12 (39%) populations, respectively. When populations were grouped according to the microsatellite clusters, nine haplotypes (56%; H8–H16) were restricted to the western and three (19%; H5–H7) to the eastern cluster. Eight haplotypes occurred in a single population each, seven of them in the western cluster, particularly

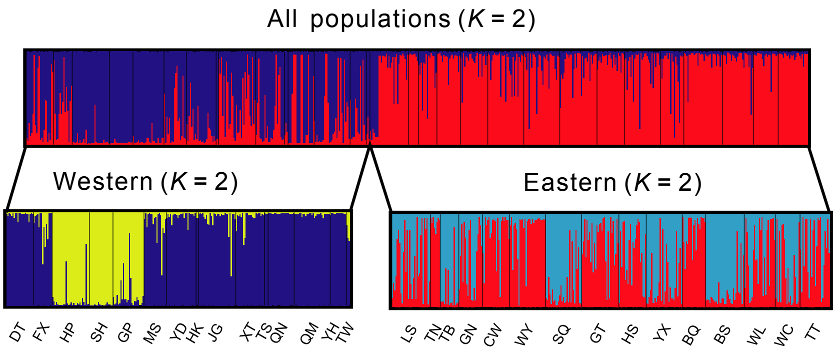

Figure 3 Results of Bayesian cluster analysis with STRUCTURE based on microsatellite data of *Castanopsis eyrei* in subtropical China. Thin bars represent the cluster membership of 629 individuals in 30 populations at K = 2 and further separate analyses of the western and eastern cluster each with K = 2 (see Appendix S3).

in the Nanling Mountains and the Wugong Mountains (Fig. 2).

The parsimony network revealed many closely related haplotypes (Fig. 2), with the most abundant haplotype, H1, being likely to represent the direct ancestor of numerous tip haplotypes. The outgroups *C. fargesii* and *C. carlesii* shared haplotype H2 with *C. eyrei*, whereas *C. tibetana* and *C. sclerophylla* exhibited new haplotypes differing from H1 or H2 by one mutation step (Fig. 2).

Overall, *C. eyrei* showed high haplotype and nucleotide diversity ($h_T = 0.842$, $\pi_T = 1.750 \times 10^{-3}$), with diversity within populations varying strongly from 0 (both h and π) to $h = 1$ and $\pi = 2.441 \times 10^{-3}$ (Table 1). Diversity of cpDNA was slightly higher in western populations than in eastern populations (Table 3).

Population differentiation in cpDNA was substantial as revealed by high values of both G_{ST} (0.709) and N_{ST} (0.729), which did not differ significantly from each other either overall or for the western and eastern cluster ($P > 0.594$), indicating that related haplotypes were not clustered. Western populations exhibited slightly stronger differentiation than eastern populations (Table 3). Although most population pairs were significantly differentiated ($P < 0.05$), no pattern of isolation by distance was found either overall or in the separate clusters ($P > 0.64$). Relative gene flow by pollen was high in *C. eyrei* as indicated by the pollen-to-seed migration ratio of $r = 25$.

DISCUSSION

Contrasting patterns at plastid and nuclear genomes

A clear geographical split between western and eastern populations was revealed by nuclear markers, although no such pattern was found in cpDNA. Similar inconsistencies between gene trees from cytoplasmic and nuclear genes can result from a variety of factors, most prominently incomplete lineage sorting of ancestral polymorphisms (Comes & Abbott, 2001) and chloroplast capture, i.e. introgression of chloroplasts from related species (Rieseberg & Soltis, 1991). Both ancestral cpDNA polymorphism (Premoli *et al.*, 2012) and chloroplast capture through recent hybridization between closely related species (Petit *et al.*, 2003) are quite common in Fagaceae. We do not, however, consider current hybridization as a likely explanation, because only one locally restricted hybridization is known for *C. eyrei* (with *C. lamontii*, in Shanghang, Fujian Province; Huang *et al.*, 1999). Rather, ancient chloro-plast capture from other *Castanopsis* species and ancestral polymorphism are likely causes of haplotype sharing that pre-dates glacial range changes. Thus, for *C. eyrei*, cpDNA markers are of limited phylogeographical use. Petit & Excoffier (2009) have pointed out that markers experiencing high levels of gene flow should be more informative regarding intraspecific phylogeography in the face of interspecific gene flow. *Castanopsis eyrei* is a species with strong pollen flow, showing a higher pollen-to-seed migration ratio ($r = 25$) than the reported median value for seed plants of 17 (Petit *et al.*, 2005). Thus, nuclear markers should be more effective in delimiting potential phylogeographical units.

Table 3 Population diversity and differentiation in the chloroplast DNA of *Castanopsis eyrei*.

Parameters	Total	Western	Eastern
h_T	0.842 (0.038)	0.876 (0.053)	0.758 (0.050)
v_T	0.554 (0.085)	0.636 (0.121)	0.487 (0.121)
h_S	0.245 (0.051)	0.254 (0.078)	0.235 (0.068)
v_S	0.150 (0.037)	0.158 (0.056)	0.171 (0.058)
G_{ST}	0.709 (0.061)	0.710 (0.093)	0.690 (0.085)
N_{ST}	0.729 (0.061)	0.751 (0.079)	0.649 (0.098)
$N_{ST} - G_{ST}$	0.020	0.041	−0.041

h_T and v_T, total gene diversity based on unordered and ordered alleles, respectively; h_S and v_S, mean within-population gene diversity based on unordered and ordered alleles, respectively; G_{ST} and N_{ST}, genetic differentiation based on unordered and ordered alleles, respectively. Standard errors are shown in parentheses.

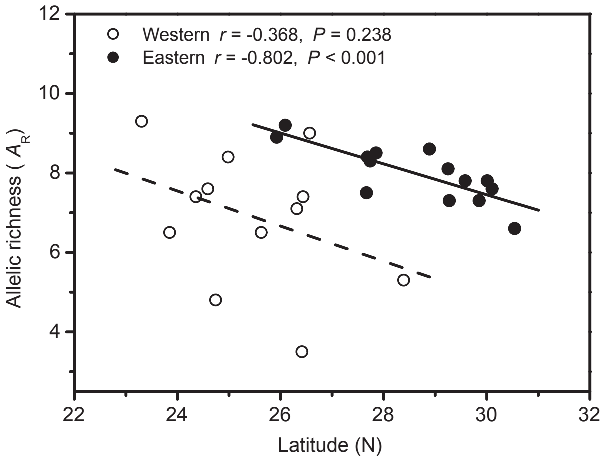

Figure 4 The relationship between allelic richness (A_R) and latitude in the western and eastern clusters of *Castanopsis eyrei* in subtropical China.

Phylogeographical history

Quaternary glacial cycles are considered to have strongly affected the distribution of plant species and the structure of genetic variation within species. Although south-eastern China was never covered by large ice sheets, most subtropical species are thought to have retreated to lower latitudes or warmer lowland areas during the LGM (Qiu *et al.*, 2011). The climate-based envelope model of the current distribution characterizes *C. eyrei* as a species with a distinctly oceanic distribution strictly confined to humid, mild and seasonally balanced climates. The projection to LGM climatic conditions inferred a potential range belt of the palaeodistribution in the southern mainland of China with very limited overlap with the current distribution for the CCSM model. In contrast, the projection to the MIROC model obtained a potential refugium in the East China Sea shelf. Recent research results on palaeoclimate and palaeovegetation of East China (Yue *et al.*, 2012; Lu *et al.*, 2013) contradict the MIROC-based projection. Shu & Wang (2012) and Xu *et al.* (2010) delivered fossil-based evidence that subtropical forest types with *Lithocarpus* and *Castanopsis* spp. were widely distributed only from 11 to 5 cal. kyr BP in higher-elevated shelf areas and the lower reaches of the Yangtze. As discussed below, the molecular analyses also indicated the existence of northern and southern glacial refugia and supported two routes of post-glacial recolonization. These results are consistent with the LGM distribution as inferred under the CCSM, but not under the MIROC palaeoclimatic model.

The longitudinal differentiation, with higher nuclear diversity in the east and higher genetic differentiation in the west, and with significant isolation by distance only in the east but not in the west, seems to be well supported by the contrasting topographical heterogeneity. After the LGM, the colonization of the eastern lowland and shelf regions might have supported gene flow and higher effective population sizes, whereas the more fragmented mountainous areas of the west increasingly suffered from genetic drift, stronger differentiation and more extinctions.

The development of climate envelope models assumes a species' current large-scale geographical distribution to be largely in equilibrium with the environment. A further assumption in the application of climate envelope models for predicting past distribution is niche stability over time (Peterson, 2003). Because large-scale distribution modelling cannot integrate local habitat climatic variation (Daly *et al.*, 2010), the existence of additional small refugia within the current range is conceivable. Although the current distribution of *C. eyrei* is well captured by the climate-based envelope model (Fig. S1), both underestimation and overestimation of the potential climatic niche are possible. Warmer lowland habitat conditions are likely to be underrepresented in the current distribution range because of its land-use history and, as a consequence, the LGM projection might be more prone to underestimate the southern refugial belt. The molecular data support the prediction of multiple refugia for *C. eyrei*. Two nuclear gene pools and a cline of decreasing genetic variation towards the north were evident, suggesting at least two main independent southern refugia. Additionally, significant substructures, notably in the western cluster (yellow cluster in Fig. 3) may imply the presence of a series of minor 'refugia within refugia' (Gómez & Lunt, 2007). Further, although no east–west pattern was detected at the cpDNA level, a high level of population differentiation for cpDNA, especially in the western cluster, with many cases of fixation of different haplotypes is consistent with multiple northern, potentially *in situ*, refugia during the LGM (Zhou *et al.*, 2010). Multiple refugia, with little admixture among populations from the different refugia, have been suggested for the Chinese endemic *Fagus engleriana* (Lei *et al.*, 2012), as well as for other species with a similar distribution, such as *Platycarya strobilacea* (Chen *et al.*, 2012) and *Rhododendron simsii* (Li *et al.*, 2012b). Similarly, in the temperate tree *Pteroceltis tatarinowii* (Li *et al.*, 2012a), most regions possessed a unique set of haplotypes, suggesting multiple refugia in mountain areas in southern China. All these findings suggest the existence of multiple suitable areas in subtropical China during LGM.

Strong population differentiation in the western range

Over its whole distribution range, *C. eyrei* presented a high level of nuclear differentiation ($F'_{ST} = 0.443$) compared with results of a previous smaller-scale study located in the eastern part of the range ($F'_{ST} = 0.15$) (Shi *et al.*, 2011). Indeed, western populations exhibited significantly higher differentiation than eastern populations in both microsatellites (Fig. S4, Table 2) and cpDNA (Fig. 2), suggesting that severe genetic drift occurred in the west. A number of factors may have contributed to the observed pattern. First, topographical isolation may be the most important factor. In the western distribution range of *C. eyrei*, large mountain areas prevail and the Nanling Mountains, which extend from east to west, may have acted as a geographical barrier, obstructing migration between southern and northern populations (Wang *et al.*, 2004). As a result of restricted northward gene flow, some haplotypes may be confined to the south, a pattern shown by haplotype H8, which was found in four populations, all south of the Nanling Mountains. Second, fragmentation always results in small and isolated populations, where low genetic diversity and high differentiation arise (Young *et al.*, 1996). The distribution range of *C. eyrei* is more fragmented in the west, especially in the south-west, where the predominating calcareous karst habitats are not suitable for *C. eyrei* (Guo *et al.*, 2011). Third, the western cluster included a number of populations which represent the current southern distribution margin of *C. eyrei* (e.g. DT, HP and GP). Small size and prolonged isolation in marginal populations have presumably resulted in reduced within-population genetic diversity and pronounced genetic differentiation between populations (Eckert *et al.*, 2008).

CONCLUSIONS

The molecular data presented here for *Castanopsis eyrei*, a major tree species in subtropical China, supports the CCSM model of LGM climate better than the MIROC model, suggesting that the inference of phylogeographical histories based on climate envelope modelling alone is insufficient and emphasizes that predictions of refugia should be undertaken using complementary molecular and modelling approaches. The projection of climate envelope models onto LGM climate reconstructions partly corroborated the hypothesized existence of refugia south of the current distribution for *C. eyrei*. This is consistent with a south–north cline of genetic variation that indicates post-glacial recolonization from the south. However, a western and an eastern phylogeographical lineage were identified and the presence of further genetic differentiation, at least in the west, suggests possible refugia within refugia. We therefore suggest: first, south–north range shifts with the existence of larger southern refugia during the LGM; and second, smaller, more northern refugia may have remained *in situ* as a result of range contractions, facilitated by complex geography and microclimatic suitability. Such a scenario best explains the diversity structure of *C. eyrei* and conforms with simulation studies that show that rapid range shifts lead to lower levels of diversity, whereas rapid range contractions preserve diversity (Arenas *et al.*, 2012). Thus, the complementary events of range shifts and *in situ* survival may represent a template for evolutionary history and phylogeography in this region.

ACKNOWLEDGEMENTS

We thank Chiang-Her Yen, Min Liu, Ming Jin, Nana Xu, Shipin Chen, Shuo Yu, Wenhui Guo, Xiaoyan Wang, Xinghua Hu, Xiushan Li and Yingliang Liu for helping with sample collections, local forestry stations in Anhui, Fujian, Guangdong, Guangxi, Hubei, Jiangxi and Zhejiang provinces for assistance in sampling, and Ina Geier and Martina Hermann for their kind help in the lab. This study was financially supported by the German Research Foundation (DFG FOR 891/1 and 891/2) and the China Scholarship Council (2008 3039). Various travel grants for project preparation financed by DFG, NSFC and the Sino-German Centre for Research Promotion in Beijing (GZ 524, 592, 698 and 699) are gratefully acknowledged.

REFERENCES

Arenas, M., Ray, N., Currat, M. & Excoffier, L. (2012) Consequences of range contractions and range shifts on molecular diversity. *Molecular Biology and Evolution*, **29**, 207–218.

Avise, J.C. (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, MA.

Axelrod, D.I., Al-Shehbaz, I. & Raven, P.H. (1996) Floristic characteristics and diversity of East Asian plants. *History of the modern flora of China* (ed. by A. Zhang and S. Wu), pp. 43–55. Springer, New York.

Barthlott, W., Mutke, J., Rafiqpoor, D., Kier, G. & Kreft, H. (2005) Global centers of vascular plant diversity. *Nova Acta Leopoldina*, **92**, 61–83.

Chen, S.-C., Zhang, L., Zeng, J., Shi, F., Yang, H., Mao, Y.-R. & Fu, C.-X. (2012) Geographic variation of chloroplast DNA in *Platycarya strobilacea* (Juglandaceae). *Journal of Systematics and Evolution*, **50**, 374–385.

Cheng, Y.-P., Hwang, S.-Y. & Lin, T.-P. (2005) Potential refugia in Taiwan revealed by the phylogeographical study of *Castanopsis carlesii* Hayata (Fagaceae). *Molecular Ecology*, **14**, 2075–2085.

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., Carton, J.A., Chang, P., Doney, S.C., Hack, J.J. & Henderson, T.B. (2006) The Community Climate System Model version 3 (CCSM3). *Journal of Climate*, **19**, 2122–2143.

Comes, H.P. & Abbott, R.J. (2001) Molecular phylogeography, reticulation, and lineage sorting in Mediterranean *Senecio* sect. *Senecio* (Asteraceae). *Evolution*, **55**, 1943–1962.

Daly, C., Conklin, D.R. & Unsworth, M.H. (2010) Local atmospheric decoupling in complex topography alters climate change impacts. *International Journal of Climatology*, **30**, 1857–1864.

Eckert, C.G., Samis, K.E. & Lougheed, S.C. (2008) Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology*, **17**, 1170–1188.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.

Ennos, R.A. (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **72**, 250–259.

Espíndola, A., Pellissier, L., Maiorano, L., Hordijk, W., Guisan, A. & Alvarez, N. (2012) Predicting present and future intra-specific genetic structure through niche hindcasting across 24 millennia. *Ecology Letters*, **15**, 649–657.

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. & Lischer, H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Gómez, A. & Lunt, D.H. (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *Phylogeography of southern European refugia* (ed. by S. Weiss and N. Ferrand), pp. 155–188. Springer, New York.

Gong, W., Chen, 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.

Goudet, J. (1995) FSTAT (version 1.2): a computer program to calculate *F*-statistics. *Journal of Heredity*, **86**, 485–486.

Guo, K., Liu, C.-C. & Dong, M. (2011) Ecological adaptation of plants and control of rocky-desertification on karst region of Southwest China. *Chinese Journal of Plant Ecology*, **35**, 991–999.

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.

Hampe, A., Rodríguez-Sánchez, F., Dobrowski, S., Hu, F.S. & Gavin, D.G. (2013) Climate refugia: from the Last Glacial

Maximum to the twenty-first century. *New Phytologist*, **197**, 16–18.

Hasumi, H. & Emori, S. (2004) *K-1 coupled model (MIROC) description*. K-1 Technical Report 1. Center for Climate System Research, University of Tokyo, Tokyo, Japan.

Hedrick, P.W. (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *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.

Hickerson, M.J., Carstens, B.C., Cavender-Bares, J., Crandall, K.A., Graham, C.H., Johnson, J.B., Rissler, L., Victoriano, P.F. & Yoder, A.D. (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution*, **54**, 291–301.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hiraoka, K. & Tomaru, N. (2009) Genetic divergence in nuclear genomes between populations of *Fagus crenata* along the Japan Sea and Pacific sides of Japan. *Journal of Plant Research*, **122**, 269–282.

Huang, C.J., Zhang, Y.T. & Bartholomew, B. (1999) Fagaceae. *Flora of China* (ed. by Z.Y. Wu and P.H. Raven), pp. 314–400. Sciences Press, Beijing.

Hwang, L.H., Hwang, S.Y. & Lin, T.P. (2000) Low chloroplast DNA variation and population differentiation of *Chamaecyparis formosensis* and *Chamaecyparis taiwanensis*. *Taiwan Journal of Forest Science*, **15**, 229–236.

Lei, M., Wang, Q., Wu, Z.-J., López-Pujol, J., Li, D.-Z. & Zhang, Z.-Y. (2012) Molecular phylogeography of *Fagus engleriana* (Fagaceae) in subtropical China: limited admixture among multiple refugia. *Tree Genetics and Genomes*, **8**, 1203–1212.

Li, J.-M. & Jin, Z.-X. (2006) Genetic diversity of *Castanopsis eyrei* populations in three forest communities with different succession stage. *Journal of Zhejiang University (Agricultural & Life Science)*, **32**, 232–236.

Li, X.-H., Shao, J.-W., Lu, C., Zhang, X.-P. & Qiu, Y.-X. (2012a) Chloroplast phylogeography of a temperate tree *Pteroceltis tatarinowii* (Ulmaceae) in China. *Journal of Systematics and Evolution*, **50**, 328–333.

Li, Y., Yan, H.-F. & Ge, X.-J. (2012b) 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, J.-Q., Sun, Y.-S., Ge, X.-J., Gao, L.-M. & Qiu, Y.-X. (2012) Phylogeographic studies of plants in China: advances in the past and directions in the future. *Journal of Systematics and Evolution*, **50**, 267–275.

López-Pujol, J., Zhang, F.-M., Sun, H.-Q., Ying, T.-S. & Ge, S. (2011) Centres of plant endemism in China: places for survival or for speciation? *Journal of Biogeography*, **38**, 1267–1280.

Lu, H., Yi, S., Liu, Z., Mason, J.A., Jiang, D., Cheng, J., Stevens, T., Xu, Z., Zhang, E., Jin, L., Zhang, Z., Guo, Z., Wang, Y. & Otto-Bliesner, B. (2013) Variation of East Asian monsoon precipitation during the past 21 k.y. and potential CO_2 forcing. *Geology*, **41**, 1023–1026.

Meirmans, P.G. (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution*, **60**, 2399–2402.

Nei, M. (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.

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.

Peterson, A.T. (2003) Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology*, **78**, 419–433.

Petit, R.J. & Excoffier, L. (2009) Gene flow and species delimitation. *Trends in Ecology and Evolution*, **24**, 386–393.

Petit, R.J., Bodénès, C., Ducousso, A., Roussel, G. & Kremer, A. (2003) Hybridization as a mechanism of invasion in oaks. *New Phytologist*, **161**, 151–164.

Petit, R.J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D. & Vendramin, G.G. (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14**, 689–701.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

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

Premoli, A.C., Mathiasen, P., Acosta, M.C. & Ramos, V.A. (2012) Phylogeographically concordant chloroplast DNA divergence in sympatric *Nothofagus* s.s. How deep can it be? *New Phytologist*, **193**, 261–275.

Prentice, I.C., Harrison, S.P. & Bartlein, P.J. (2011) Global vegetation and terrestrial carbon cycle changes after the last ice age. *New Phytologist*, **189**, 988–998.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

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.

R Core Team (2012) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rieseberg, L.H. & Soltis, D.E. (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants*, **5**, 65–84.

Shen, L., Chen, X.-Y., Zhang, X., Li, Y.-Y., Fu, C.-X. & Qiu, Y.-X. (2005) Genetic variation of *Ginkgo biloba* L. (Ginkgoaceae) based on cpDNA PCR-RFLPs inference of glacial refugia. *Heredity*, **94**, 396–401.

Shi, Y. (2002) Characteristics of late Quaternary monsoonal glaciation on the Tibetan Plateau and in East Asia. *Quaternary International*, **97–98**, 79–91.

Shi, M.-M., Michalski, S.G., Chen, X.-Y. & Durka, W. (2011) Isolation by elevation: genetic structure at neutral and putatively non-neutral loci in a dominant tree of subtropical forests, *Castanopsis eyrei*. *PLoS ONE*, **6**, e21302.

Shu, J.-W. & Wang, W.-M. (2012) A unique Middle Pleistocene beech (*Fagus*)-rich deciduous broad-leaved forest in the Yangtze Delta Plain, East China: its climatic and stratigraphic implication. *Journal of Asian Earth Sciences*, **56**, 180–190.

Systat (2009) *SYSTAT 13*. Systat Software, San Jose, CA.

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.

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, X.-P., Wang, Z.-H. & Fang, J.-Y. (2004) Mountain ranges and peaks in China. *Biodiversity Science*, **12**, 206–212.

Wang, J., Gao, P.X., Kang, M., Lowe, A.J. & 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.

Weir, B.S. & Cockerham, C.C. (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Weiss, S. & Ferrand, N. (2007) *Phylogeography of southern European refugia*. Springer, Berlin.

Werneck, F.P., Costa, G.C., Colli, G.R., Prado, D.E. & Sites, J.W., Jr (2011) Revisiting the historical distribution of seasonally dry tropical forests: new insights based on palaeodistribution modelling and palynological evidence. *Global Ecology and Biogeography*, **20**, 272–288.

Wu, Z.-Y. (1980) *Vegetation of China*. Science Press, Beijing.

Xu, D., Lu, H., Wu, N. & Liu, Z. (2010) 30 000-Year vegetation and climate change around the East China Sea shelf inferred from a high-resolution pollen record. *Quaternary International*, **227**, 53–60.

Yan, H.-F., Peng, C.I., Hu, C.M. & Hao, G. (2007) Phylogeographic structure of *Primula obconica* (Primulaceae) inferred from chloroplast microsatellites (cpSSRs) markers. *Acta Phytotaxonomica Sinica*, **45**, 488–496.

Young, A., Boyle, T. & Brown, T. (1996) The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology and Evolution*, **11**, 413–418.

Yue, Y., Zheng, Z., Huang, K., Chevalier, M., Chase, B.M., Carré, M. & Cheddadi, R. (2012) A continuous record of vegetation and climate change over the past 50,000 years in the Fujian Province of eastern subtropical China. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **365–366**, 115–123.

Zheng, Y., Yu, G., Wang, S., Xue, B., Liu, H. & Zeng, X. (2003) Simulations of LGM climate of East Asia by regional climate model. *Science in China Series D: Earth Sciences*, **46**, 753–764.

Zhou, Y.F., Abbott, R.J., Jiang, Z.Y., Du, F.K., Milne, R.I. & Liu, J.Q. (2010) Gene flow and species delimitation: a case study of two pine species with overlapping distributions in southeast China. *Evolution*, **64**, 2342–2352.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Climatic envelope model and LGM projection, and detailed information of populations sampled in this study.

Appendix S2 Laboratory protocol for cpDNA sequencing and haplotypes defined by two chloroplast intergenic spacers.

Appendix S3 Determination of most likely number of clusters from the STRUCTURE analysis and patterns of isolation by distance.

BIOSKETCH

Miao-Miao Shi is a researcher in Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences. This study was part of her PhD work. Her research interests are population genetics and phylogeography of plant species in East Asia.

Author contributions: M-M.S., X-Y.C. and W.D. conceived the ideas; M-M.S., W.D. collected the molecular data; E.W. collected the distribution data; M-M.S., S.G.M., E.W. and W.D. analysed the data; and M-M.S., S.G.M., E.W., X-Y.C. and W.D. contributed to the writing.

Editor: Mark Carine