

SCIENTIFIC REPORTS



OPEN

Contributions of historical and contemporary geographic and environmental factors to phylogeographic structure in a Tertiary relict species, *Emmenopterys henryi* (Rubiaceae)

Yong-Hua Zhang¹, Ian J. Wang², Hans Peter Comes³, Hua Peng⁴ & Ying-Xiong Qiu¹

Examining how historical and contemporary geographic and environmental factors contribute to genetic divergence at different evolutionary scales is a central yet largely unexplored question in ecology and evolution. Here, we examine this key question by investigating how environmental and geographic factors across different epochs have driven genetic divergence at deeper (phylogeographic) and shallower (landscape genetic) evolutionary scales in the Chinese Tertiary relict tree *Emmenopterys henryi*. We found that geography played a predominant role at all levels – phylogeographic clades are broadly geographically structured, the deepest levels of divergence are associated with major geological or pre-Quaternary climatic events, and isolation by distance (IBD) primarily explained population genetic structure. However, environmental factors are clearly also important – climatic fluctuations since the Last Interglacial (LIG) have likely contributed to phylogeographic structure, and the population genetic structure (in our AFLP dataset) was partly explained by isolation by environment (IBE), which may have resulted from natural selection in environments with divergent climates. Thus, historical and contemporary geography and historical and contemporary environments have all shaped patterns of genetic structure in *E. henryi*, and, in fact, changes in the landscape through time have also been critical factors.

Understanding the contemporary and historical ecological (climatic, geographical) factors shaping population genetic diversity, structure, and divergence is of great interest to molecular ecology, evolutionary biology and conservation biology^{1–3}. Populations separated by physical geographical barriers (including geographic distance) may diverge under any combination of natural selection and random genetic drift resulting from reduced gene flow and population connectivity^{4,5}. In the absence of extrinsic (e.g. physical) barriers to gene flow, population divergence may still occur when reproductive isolation evolves between populations as a result of ecologically-based divergent selection in different environments^{5–10}. Thus, population genetic divergence can result from both geographical and environmental factors (including climate and soil, among others). Disentangling the roles of geographic and environmental forces in driving genetic structure during certain periods (usually contemporary) has seen a large body of research in both plants and animals over recent years^{4,11–15}. However, few studies have examined how these factors contribute to genetic structure through time, including both historical and contemporary

¹Key Laboratory of Conservation Biology for Endangered Wildlife of the Ministry of Education, and College of Life Sciences, Zhejiang University, Hangzhou 310058, China. ²Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA. ³Department of Ecology & Evolution, Salzburg University, A-5020 Salzburg, Austria. ⁴Laboratory of Biodiversity and Biogeography, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650204, China. Correspondence and requests for materials should be addressed to Y.-X.Q. (email: qyxhero@zju.edu.cn)

Received: 20 October 2015

Accepted: 21 March 2016

Published: 03 May 2016

geographic and environmental factors, to better understand how changing climates and geographic landscape features can influence patterns of genetic structure observed presently^{16,17}.

For example, climatic fluctuations during the Quaternary which resulted in population isolation in multiple refugia are considered major drivers of population divergence and broad phylogeographic patterns^{18–20}. Of course, major geographic barriers, like oceans, rivers, and mountains, are also recognized as key drivers of biogeographic structure¹⁸, and thus, both environmental and geographic factors can contribute to genetic divergence at deeper evolutionary scales. Likewise, geography and the environment have also been recognized as critical factors underlying genetic differentiation over shorter evolutionary scales, like the evolution of population genetic structure among contemporary populations on a landscape^{5,21,22}.

So, how do historical and contemporary geographic and environmental factors contribute to genetic divergence at different evolutionary scales? This is a central yet largely unexplored question in ecology and evolution^{16,17}. In general, the flora of subtropical (Central-Southeast) China presents some excellent systems for such studies, including several genera of Tertiary relict trees that have inhabited topographically and ecologically heterogeneous environments (in terms of climate, soil, etc.) for millions of years^{23–26}. These genera (e.g. *Cathaya*, *Ginkgo*, *Metasequoia*, *Davidia*, *Emmenopterys*) are thought to represent remnants of the so-called ‘boreotropical flora’ that likely formed a belt of vegetation around the Northern Hemisphere during the Early Tertiary/Eocene^{27–29}. Here, we examine this key question by investigating how environmental and geographic factors across different epochs have driven genetic divergence at deeper (phylogeographic) and shallower (landscape genetic) evolutionary scales in the Chinese flowering tree *Emmenopterys henryi* Oliv. (Rubiaceae).

Emmenopterys henryi is a particular suitable species for addressing these issues. This deciduous tree, which is the only extant species of its genus, is native to subtropical China, where it occurs in disjunct montane valleys of mainly warm-temperate deciduous (WTD) forests, at elevations ranging from c. 400–1600 (2000) m above sea level³⁰ (Fig. 1A). Landscape characteristics, climatic conditions, and soil types vary between regions within the distribution range of *E. henryi*³¹. Well-preserved infructescences of now-extinct *Emmenopterys* species are known from the Eocene of North America and Germany³², but there are no reliable fossils of *E. henryi*. However, there is dated molecular evidence to suggest that the origin of this species dates back to the Early Miocene³³. Previous results based on inter-simple sequence repeat (ISSR) markers indicate that *E. henryi* exhibits significant population genetic structure and divergence³⁴, however it remains unclear whether this is the result of long-term geographical barriers to gene flow, ecologically-based divergent selection, or recent habitat fragmentation.

In this study, we integrate genetic markers that capture signatures from historical (chloroplast DNA) and contemporary (amplified fragment length polymorphisms; AFLPs) divergence³⁵, ecological niche modelling (ENM), and spatial genetic modelling approaches to disentangle the relative roles of geography, climate, and ecology in shaping the population genetic structure of *E. henryi* across subtropical China. Our main objectives were to: (i) estimate the timing and pattern of divergence among populations of *E. henryi*; (ii) investigate how climatic and geographical variation over space and time explain patterns of phylogeographic and population genetic structure; and (iii) explore the specific environmental variables that may underlie local adaptation through natural selection in divergent environments.

Results

cpDNA and ITS phylogeography and diversity. The three cpDNA-IGS regions surveyed across the 443 individuals of *E. henryi* were aligned along a total length of 2163 bp with 26 single-site mutations (including two 1-bp indels), 18 length polymorphisms (2–78 bp) and one inversion (27 bp) observed (Table S5). A total of 40 haplotypes (‘chlorotypes’; H1–40) were detected in the 38 *E. henryi* populations across subtropical China (Table S1; Fig. 1A). Of the 40 haplotypes, 20 were shared by at least two populations while the other 20 haplotypes were only found in a single population (Table S1; Fig. 1A). The most common haplotypes were H12 (found in 10 populations with a frequency of 0.25), H2 (17.5% of all populations), H28 (15%), H14 (12.5%), and H6 (10%). Total haplotype diversity (h_T) was estimated to be 0.928 and within-population diversity (h_S) was 0.332 (Table S1). Regression analyses showed that h_S was not dependent on longitude or latitude ($P = 0.23, 0.66$, respectively), but π_S was weakly related to latitude ($R^2 = 0.116, P = 0.036$).

The Bayesian haplotype tree from BEAST supported the monophyly of *E. henryi* [posterior probability (PP) = 1], and two main (‘Northern’ vs. ‘Southern’) lineages were recognized with weak support (PP = 0.57, 0.70, respectively; Fig. 1B). The haplotypes of the Northern lineage were located in the northern part of the species’ range, except for H12 from populations S17 and S18, while those of the Southern lineage were present only in southern populations (Fig. 1A). In the haplotype network, the two major lineages are also recognized (Fig. 1B,C). Most haplotypes close to each other in the phylogeny and haplotype network tended to occur in nearby populations (Fig. 1). In the SAMOVA analysis, F_{CT} values increased progressively as the value of K increased from 2 to 26. However, with K ranging between 5 and 26, F_{CT} values did not increase significantly, and in most cases the newly defined groups comprised single populations (Fig. S1). Thus, we retained the configuration of $K = 4$ ($F_{CT} = 0.523$). The four cpDNA groups identified are populations S1–S6 in the east (Southeast group), S7–S23 in the middle (Central-Southwest group), N1–N6 in the northeast (Northeast group), and N7–N15 in the northwest (Northwest group). This grouping is mostly consistent with the phylogenetic analyses (Fig. 1B,C).

Non-hierarchical AMOVA (Table 1) revealed a strong population genetic structure for cpDNA sequence variation at the species level ($\Phi_{ST} = 0.779; P < 0.001$). Hierarchical AMOVA, however, revealed that of the total genetic variation, 36.48% was distributed between the Northern lineage and the Southern lineage ($\Phi_{CT} = 0.365$), 45.43% was explained by variation among populations within regions ($\Phi_{SC} = 0.715$), and only 18.10% was found within populations ($\Phi_{ST} = 0.819$; Table 1). Nevertheless, there was greater population subdivision in the Southern lineage ($\Phi_{ST} = 0.823$) when compared to the Northern lineage ($\Phi_{ST} = 0.551$; Table 1). Moreover, significant phylogeographic structure for cpDNA was observed at the range-wide scale ($G_{ST}/N_{ST} = 0.704/0.802, P < 0.01$) and in the Southern lineage ($G_{ST}/N_{ST} = 0.757/0.868, P < 0.05$), but no such structure was found in the Northern lineage

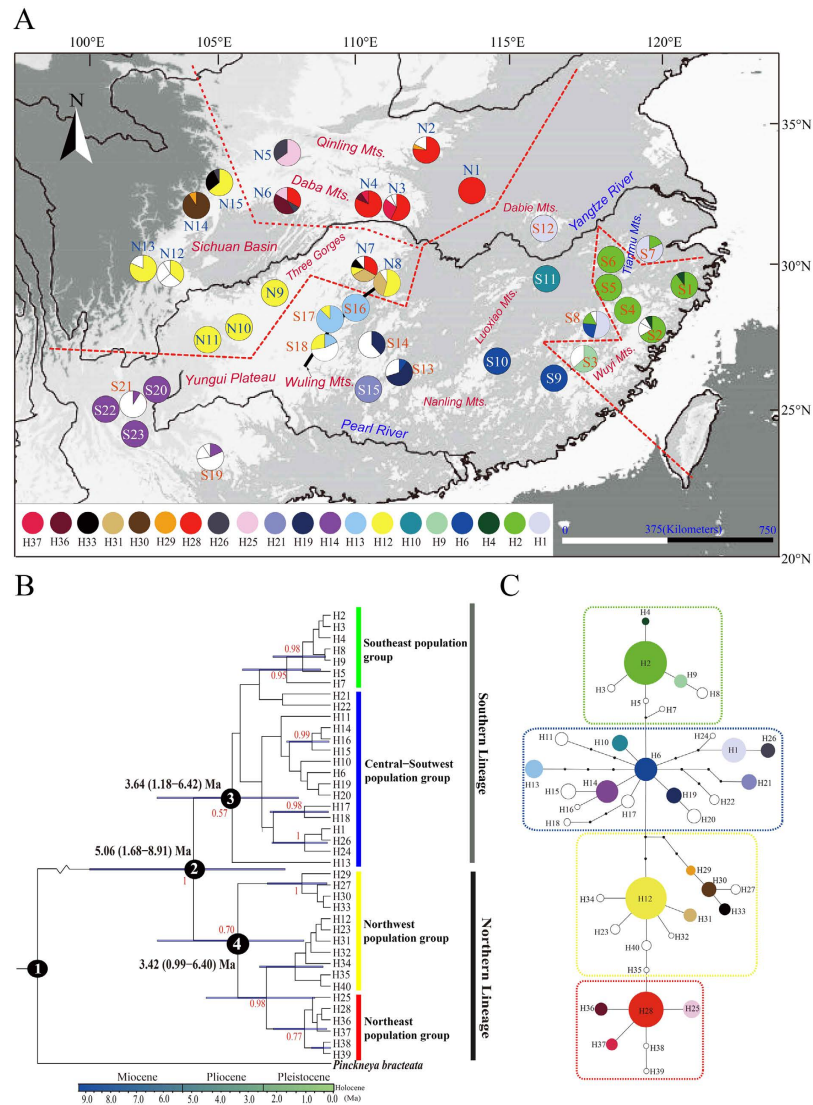


Figure 1. Analysis of cpDNA (*psbA-trnH*, *trnL-trnF*, *trnT-trnL*) haplotypes of *Emmenopterys henryi*. (A) Geographical distribution of haplotypes across 38 sampled populations. Pie charts represent haplotype proportions. Colored haplotypes are shared by two or more populations, and blank ones are private haplotypes. Population groups identified by SAMOVA for $K = 4$ are delimited by red dotted lines. (B) Fifty percent majority-rule consensus tree of the Bayesian phylogenetic analysis. Numbers on branches are Bayesian posterior probabilities. (C) Statistical parsimony network of cpDNA haplotypes. Lines represent single nucleotide substitutions, and the small black dots indicate missing haplotypes (extinct or not found). The sizes of circles are approximately proportional to sample size (n), with the smallest circles representing $n = 1$ and the largest representing $n = 64$. The map was drawn using ArcGIS v.9.3 (ESRI, Redlands, CA, USA)⁹⁴.

($G_{ST}/N_{ST} = 0.521/0.379$, $P > 0.05$; Table 1). Finally, Mantel tests uncovered a strong pattern of IBD for cpDNA both at the range-wide scale ($r = 0.237$, $P = 0.001$) and the region scale (Northern lineage: $r = 0.506$, $P = 0.001$; Southern lineage: $r = 0.293$, $P = 0.001$).

The ITS sequences of 212 individuals (38 populations) of *E. henryi* were aligned with a total of length of 767 bp, exhibiting 10 nucleotide substitutions (ITS-1: 4; ITS-2: 6; Table S6). Together, these 10 polymorphic sites identified nine ITS haplotypes ('ribotypes', R1–9; Table S6). Of those, five ribotypes were specific to the Southern cpDNA lineage (R2–6) and three to the Northern cpDNA lineage (R7–9; Fig. 2A). These lineage-specific ribotypes formed separate rcs clades, except for the shared ribotype R1 (Fig. 2B).

Molecular dating and historical demography based on cpDNA sequence variation. Based on the assumption that *E. henryi* and *P. bracteata* diverged at $c. 22$ Ma (Manns *et al.*³³; Fig. 1B; node 1), the time since divergence between the Northern and Southern lineages was estimated by BEAST analysis to have occurred 5.06 Ma (95% HPD: 1.68–8.91 Ma; node 2). The onset of lineage diversification was estimated as 3.42 Ma (Northern lineage, 95% HPD: 0.99–6.40 Ma; node 4) and 3.64 Ma (Southern lineage, 95% HPD: 1.18–6.42 Ma; node 3). For this cpDNA-IGS chronogram, we recovered an average substitution rate of 6.225×10^{-10} s/s/y by the BEAST

| Source of variation | cpDNA | | | | nrITS | | | AFLPs | | |
|----------------------------------|-------|----------------------------------|---------------------|----------------------|-------|----------------------------------|---------------------|-------|----------------------------------|---------------------|
| | d.f. | Percentage of total variance (%) | Fixation indices | G_{ST}/N_{ST} | d.f. | Percentage of total variance (%) | Fixation indices | d.f. | Percentage of total variance (%) | Fixation indices |
| all | | | | | | | | | | |
| Among populations | 37 | 77.88 | $\Phi_{ST} = 0.779$ | 0.704/0.802** | 37 | 55.78 | $\Phi_{ST} = 0.558$ | 36 | 34.41 | $\Phi_{ST} = 0.344$ |
| Within populations | 395 | 22.12 | | | 174 | 44.22 | | 357 | 65.59 | |
| South | | | | | | | | | | |
| Among populations | 22 | 82.27 | $\Phi_{ST} = 0.823$ | 0.757/0.868* | 22 | 55.77 | $\Phi_{ST} = 0.558$ | 21 | 41.63 | $\Phi_{ST} = 0.416$ |
| Within populations | 229 | 17.73 | | | 103 | 44.23 | | 212 | 58.37 | |
| North | | | | | | | | | | |
| Among populations | 14 | 55.07 | $\Phi_{ST} = 0.551$ | 0.521/0.379 | 14 | 18.41 | $\Phi_{ST} = 0.184$ | 14 | 17.54 | $\Phi_{ST} = 0.175$ |
| Within populations | 166 | 44.93 | | | 71 | 81.59 | | 145 | 82.46 | |
| South + North | | | | | | | | | | |
| Among regions | 1 | 36.48 | $\Phi_{CT} = 0.365$ | | 1 | 25.37 | $\Phi_{CT} = 0.254$ | 1 | 6.07 | $\Phi_{CT} = 0.061$ |
| Among populations within regions | 36 | 45.43 | $\Phi_{SC} = 0.715$ | | 36 | 36.06 | $\Phi_{SC} = 0.483$ | 35 | 30.35 | $\Phi_{SC} = 0.323$ |
| Within populations | 395 | 18.1 | $\Phi_{ST} = 0.819$ | | 174 | 38.56 | $\Phi_{ST} = 0.614$ | 357 | 63.58 | $\Phi_{ST} = 0.364$ |

Table 1. Analyses of molecular variance (AMOVAs) based on cpDNA chlorotype frequencies, nrITS ribotype frequencies, and amplified fragment length polymorphism (AFLP) allele frequencies for populations of *E. henryi*. For cpDNA, population differentiation for unordered (G_{ST}) and ordered (N_{ST}) haplotypes were calculated respectively at the different levels. d.f., degrees of freedom. N_{ST} significantly different from G_{ST} is shown in bold (** $P < 0.01$; * $P < 0.05$). Fixation index values were significant at all levels.

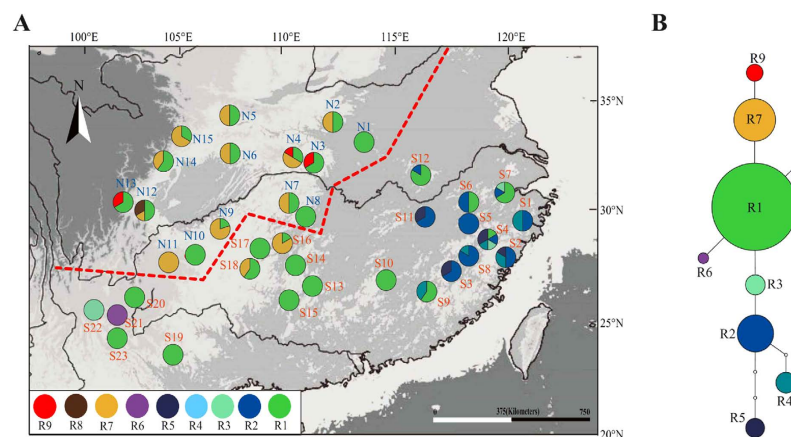


Figure 2. Analysis of internal transcribed spacer (ITS) ribotypes of *E. henryi*. (A) Geographical distribution of ITS ribotypes across sampled populations. Pie charts represent ribotype proportions. (B) Statistical parsimony network of ITS ribotypes. Lines represent single nucleotide substitutions, and the small white dots indicate missing ribotypes (extinct or not found). The sizes of circles are approximately proportional to sample size (n), with the smallest circles representing $n = 2$ and the largest representing $n = 110$. The map was drawn using ArcGIS v.9.3 (ESRI, Redlands, CA, USA)⁹⁴.

analysis, which is much lower than the average values generally reported for noncoding regions of the chloroplast genome (e.g. $1.2\text{--}1.7 \times 10^{-9}$ s/s/y)³⁶, but in accordance with that in woody taxa and/or phylogenetic relicts (viz. 'living fossils')^{37–40}.

For the four cpDNA population groups of *E. henryi*, at least one of the estimates of Tajima's D and Fu's F_S was generally significant, except for the Central-Southwest group (Table 2). In contrast, almost all four groups failed to reject the spatial expansion model (SSD , H_{Rag} values $P > 0.05$), with the exception of the Northeast group ($P < 0.05$; Table 2). In addition, the mismatch distributions of the Southeast, Northeast and Northwest groups were unimodal, except for the Central-Southwest group (Fig. S3). Consequently, the Southeast, Northeast and Northwest groups were found to fit to the distributions expected under a spatial expansion model. Based on the corresponding τ values, and assuming a substitution rate of 6.225×10^{-10} s/s/y (see above), we dated the three spatial expansions to the last glacial cycle(s) (Southeast: c. 0.23 Ma, 95% CI: 0.000–0.811 Ma; Northwest: c. 0.19 Ma, 95% CI: 0.111–0.314 Ma; Northeast: c. 0.26 Ma, 95% CI: 0.133–0.376 Ma; Table 2).

AFLP analysis. The nine primer combinations employed with samples from 37 populations (394 individuals) of *E. henryi* generated a total of 457 fragments, of which 431 (94.31%) were polymorphic. Different

| Group | Parameter (τ) | Expansion time (t , Ma) | SSD | P | H_{Rag} | P | Fu's F_S | P | Tajima's D | P |
|------------------------------------|----------------------|----------------------------|--------|-------|-----------|-------|------------|--------------|--------------|--------------|
| Southeast population group | 0.605 (0.000–2.184) | 0.225 (0.000–0.811) | 0.0003 | 0.723 | 0.189 | 0.638 | −3.060 | 0.039 | −1.204 | 0.110 |
| Central-Southwest population group | 3.569 (1.888–5.673) | NC | 0.011 | 0.313 | 0.033 | 0.538 | −0.696 | 0.469 | −0.812 | 0.219 |
| Northwest population group | 0.511 (0.299–0.847) | 0.190 (0.111–0.314) | 0.002 | 0.186 | 0.152 | 0.446 | −4.469 | 0.005 | −1.474 | 0.045 |
| Northeast population group | 0.686 (0.359–1.011) | 0.255 (0.133–0.376) | 0.008 | 0.017 | 0.130 | 0.154 | −3.388 | 0.039 | −0.925 | 0.197 |

Table 2. Summary of mismatch distribution parameters and neutrality tests for four geographical groups of *E. henryi*. Estimates were obtained under models of spatial expansion using ARLEQUIN. NC, not calculated. There is a significant difference at the $\alpha = 0.05$ level⁸⁵ for Tajima's D and Fu's F_S .

primer pairs amplified variable numbers of fragments, from 38 to 69, with an average of 50.8 ± 10.1 fragments per primer combination (Table S3). The percentage of polymorphic fragments varied among primer pairs from 88.89 to 98.55% (Table S3). Because of the high degree of polymorphism, these primer combinations distinguished all 394 individuals as separate phenotypes. AFLP variation within populations (in terms of PPF , I , H_E , DW) varied widely (Table S1). The highest genetic diversity was observed in a northern population (N12), with $PPF = 74.18\%$, $I = 0.243$, and $H_E = 0.162$. By contrast, genetic diversity was lowest in a southern population (S22), with $PPF = 57.77\%$, $I = 0.035$, and $H_E = 0.023$. Regression analyses showed that H_E was weakly related to latitude ($R^2 = 0.110$, $P = 0.045$).

The partition with the highest log marginal likelihood (spatial: −51313.01, non-spatial: −51225.49) produced by BAPS identified nine different clusters (Fig. S2A). All populations from the Northern and two (S17, S18) from the Southern (cpDNA) lineage formed one cluster, while the remaining populations from the Southern lineage were further divided into eight clusters (Fig. 3A,B). In the PCoA-plot, the first axis (42.20% of the variation) and the third axis (12.84%) indicated that the main split in the dataset was between the central-southeastern and the northern samples, while the second axis (16.10%) separated the southwestern samples from the northern group (Fig. 3C). The NJ analysis based on genetic distances among populations also identified three regional groups, again comprising populations from the central-southeastern region (67% bootstrap support), southwestern region (92%), and northern region (<50%; Fig. S2B). With few exceptions, S16 and S17 clustered with the northern group. Populations from central-southeastern and southwestern China showed higher among-population divergence when compared with those from within the northern group (Figs 3C and S2B). Non-hierarchical AMOVA indicated high overall levels of population differentiation for AFLPs in *E. henryi* ($\Phi_{ST} = 0.344$; Table 1). When accounting for the species' significant hierarchical (regional) substructure ($\Phi_{CT} = 0.061$), levels of population subdivision still remained high ($\Phi_{SC} = 0.364$), but were markedly higher in the Southern lineage ($\Phi_{ST} = 0.416$) than in the Northern one ($\Phi_{ST} = 0.175$; Table 1).

Ecological niche modelling across temporal scales. The MAXENT model for *E. henryi* had high predictive power and did not overfit the presence data ($AUC = 0.804 \pm 0.003$). The current distributional predictions (Fig. 4A) were accurate representations of the species' extant distribution, except for some predicted areas where the species does not occur at present (e.g. southeastern Qinghai–Tibetan Plateau and Taiwan). Palaeodistribution modeling suggested more restricted ranges of the species during the Last Interglacial (LIG) compared with its current distribution, particularly in north-central China (e.g. northern Sichuan Basin and Daba/Qinling Mts.; Fig. 4B), with subsequent expansion at the Last Glacial Maximum (LGM) to cover a slightly greater area than that predicted under current climatic conditions (Fig. 4A,C). However, ENMs for the future (2080) predict that climate change will result in a reduction of the species' potential range in China. Most evident is a loss of suitable habitat in areas south of the Yangtze Delta, where only small and disjunct mountain areas are predicted as suitable (Fig. 4D).

IBD and IBE. The structural equation modelling (SEM) and multiple matrix regression with randomization (MMRR) analyses both revealed significant effects of IBD and IBE on neutral (non-outlier) AFLP divergence (F_{ST}) in this species. In both analyses, IBD explained more of the variation in genetic differentiation than IBE (SEM: ≈ 2 times as much; MMRR: ≈ 1.25 times as much; Table 3). SEM identified significant pairwise relationships among environmental distance, geographic distance, and genetic distance ($P < 0.01$) and estimated that IBD ($\beta_D = 0.360 \pm 0.037$) contributed about twice as much as IBE ($\beta_E = 0.181 \pm 0.151$) to genetic distance. Similarly, MMRR also estimated the contribution of IBD ($\beta_D = 0.307$, $P < 0.001$) to be larger than that of IBE ($\beta_E = 0.247$, $P = 0.002$). Each analysis detected low levels of covariation between geographical and environmental distance (0.133 ± 0.028 in SEM and 0.488 in MMRR), although the covariation based on the MMRR analysis was approaching the reliability limit of 0.5 (Table 3). The difference between these estimates from each analysis is due to the different environmental distance matrices used, estimated as Euclidian distances among all observed environmental variables in MMRR and as a latent variable constructed from the observed environmental variables in SEM.

For the SEM analysis, we also quantified the contributions of the individual environmental variables to the composition of the environmental dissimilarity latent variable (Table S7). The results revealed that BIO4 (temperature seasonality) and BIO7 (temperature annual range) were the primary contributors, followed by BIO1 (annual mean temperature), BIO5 (max temperature of warmest month), BIO9 (mean temperature of driest quarter), and BIO15 (precipitation seasonality). By contrast, slope, soil, and BIO12 (annual precipitation) had very minor contributions.

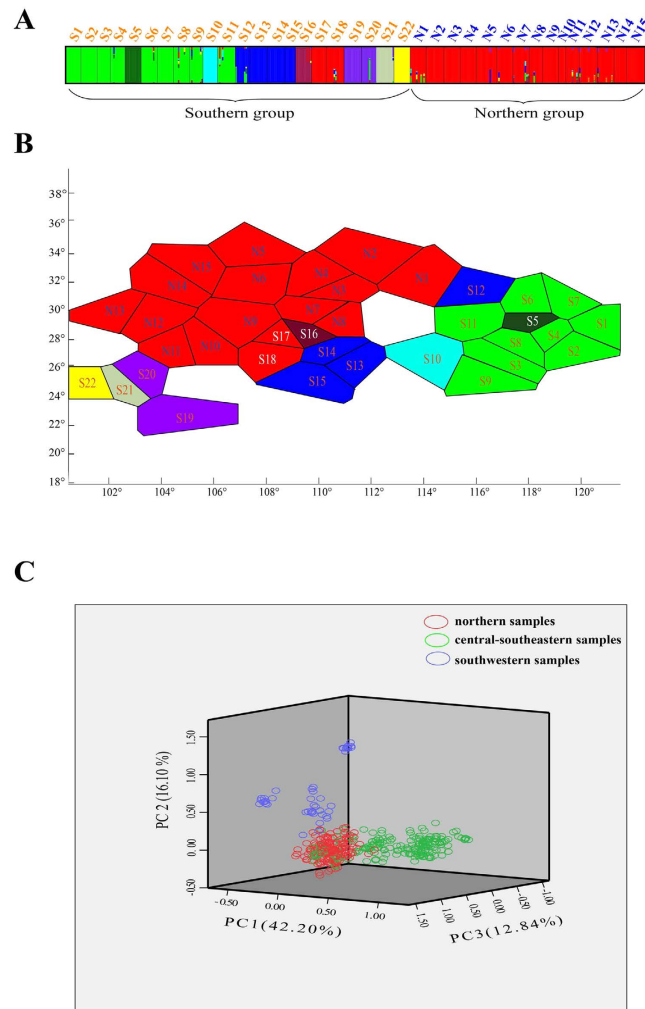


Figure 3. Results for 37 populations of *E. henryi* based on the analysis of 457 AFLP loci. (A) Histogram of the BAPS assignment test for 37 populations (394 individuals) ($K = 9$). Each vertical bar represents an individual. Southern and northern cpDNA lineages contain 9 and 1 AFLP clusters, respectively. Population codes are shown above the bar. **(B)** Spatial clustering of 37 populations in BAPS ($K = 9$). Each polygon represents one population corresponding to the histogram above, and colours of polygons indicate clusters. **(C)** PCoA for 37 populations (394 individuals). Percentages of total variance explained by the first three coordinates are shown on the respective axes.

Detecting potential loci under selection: outlier loci test and MLR analysis. Using *FDIST*, 67 of 457 loci (14.66%) showed high probability (99.5%) for divergent selection among the nine BAPS groups, including 21 loci under directional selection and 46 loci under balancing selection (Fig. 5A). Additionally, we identified 16 outlier loci in *BAYESCAN* (3.50% of all 457 loci; Fig. 5B). However, only six loci were identified by both *FDIST* and *BAYESCAN* (Fig. 5). Eventually, four potential loci under selection (L128, L144, L294, and L305) were confirmed by the MLR analysis with $R_{adj}^2 > 0.5$ (Table S8). When we ran linear regressions using each environmental variable individually, all six loci were significantly ($P < 0.05$) associated with at least two of the eight selected environmental variables (Table S7). Among the environmental variables, BIO2 (mean diurnal temperature range), BIO4 (temperature seasonality), BIO5 (maximum temperature of the warmest month), BIO12 (annual precipitation) and BIO15 (precipitation seasonality) were most highly associated with the potential loci under selection.

Discussion

The results of our phylogeographic and landscape genetic analyses reveal how historical and contemporary environmental and geographic factors have all contributed to presently observed patterns of genetic divergence in *E. henryi*. At broad scales, the results of our cpDNA and ITS phylogeography show that genetic variation is, largely, geographically structured. In the cpDNA phylogeography, we found two major lineages corresponding to the Northern and Southern populations (with minor exceptions). The boundary between these lineages is located in the region of the Yangtze River and next to the Three Gorges Mountain Region (TGMR; Figs 1–3). We found evidence of admixture of Northern lineage cpDNA haplotypes into the Southern lineage (S16, S17, S18), which is most likely explained by both introgression following secondary contact during glacial periods and incomplete

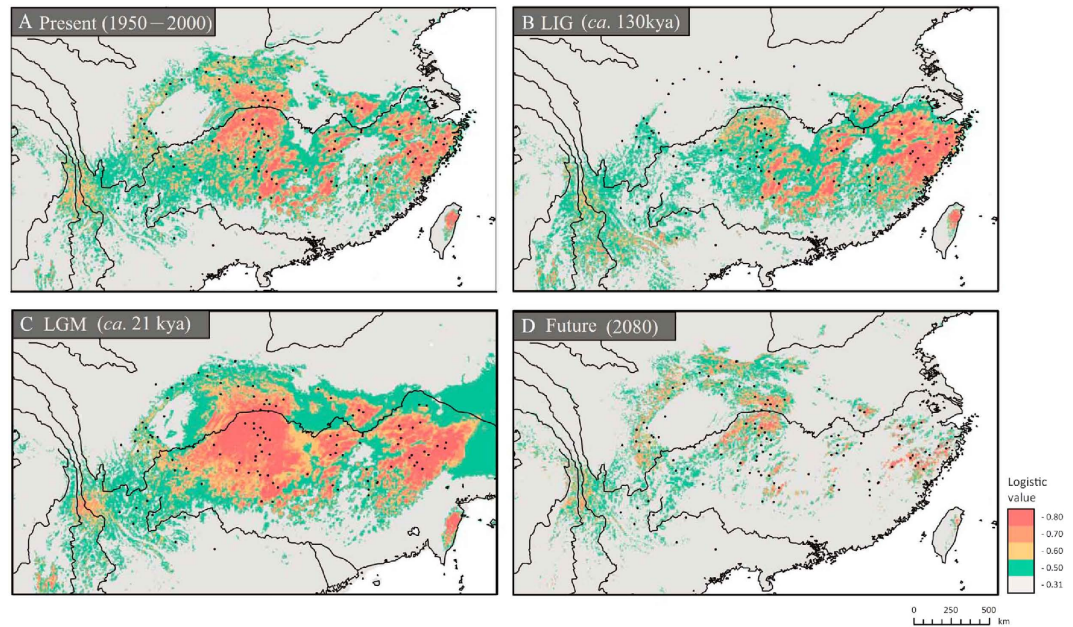


Figure 4. Potential distributions as probability of occurrence for *E. henryi* in China (A) at present (1950–2000), (B) at the Last Interglacial (LIG; c. 130kya), (C) at the Last Glacial Maximum (LGM; c. 21kya), and (D) in the future (2080, A2 scenario). Ecological niche models were established with current bioclimatic variables on the basis of extant occurrence points (black dots) of the species using MAXENT v.3.2.1. Predicted distribution probabilities (in logistic values) are shown in each 2.5 arc-min pixel. Maps were generated using ArcGIS v.9.3 (ESRI, Redlands, CA, USA)⁹⁴.

| Model | IBD | IBE | TOTAL | COVAR. | CONTRIB. VARS. |
|-------|------------------|------------------|-------|---------------|----------------|
| SEM | 0.360 ± 0.037*** | 0.181 ± 0.151*** | 0.541 | 0.133 ± 0.028 | Temp. |
| MMRR | 0.307*** | 0.247** | 0.554 | 0.488* | Temp. |

Table 3. Proportions of spatial genetic divergence explained by isolation by distance (IBD) and isolation by environment (IBE) based on SEM analysis and MMRR analysis using the neutral (non-outlier) AFLP dataset. The sums of IBD and IBE (TOTAL), the covariation between these variables (COVAR.) and the primary contributors (CONTRIB. VARS.) to the environmental dissimilarity latent variable are listed. *** $P < 0.001$; ** $P < 0.005$; * $P < 0.05$.

lineage sorting due to recent, postglacial divergence. Nuclear ITS ribotype data did not show such a distinct division between the Northern and Southern populations but did, nevertheless, indicate geographically structured ribotype frequency differences between these sets of populations and the presence of ribotypes found only in the Northern and Southern lineages. In fact, of the nine ribotypes we detected, only one was shared between lineages (R1), which was found in many populations, including those around the Yungui Plateau (Fig. 2). Several ribotypes and chlorotypes are also unique to Southeastern populations, where the climate is warmest and wettest, or to the Northernmost populations, where the climate is coolest and driest. Whether these haplotypes and populations sharing similar haplotype frequencies are clustered in these areas because they are environmentally divergent from the intermediate climate of central China, where the most haplotype diversity is found, or because of geographical restrictions to the mountain chains in these regions is still unclear. Nevertheless, it appears that major geographic regions harbor phylogeographic structure, but that geographic distances and major geographic barriers (like the Yangtze River) are not always major barriers to genetic introgression. For example, there is cpDNA evidence to suggest that two populations from the Northern lineage (N6 and N15 in the Qinling Mts.) were shaped by dispersal events from a southern population (S12 in central China), as a chlorotype (H1) unique to this latter population differs from its derived haplotype (H26) in the Qinling Mts. by only one step (Fig. 1A,C; Table S1). These populations are separated by several hundred kilometers, suggesting that very long distance dispersal may be possible in this species, that these populations have some shared ancestry, or that they may have been more widespread and come in contact in the past, possibly during the LGM when there was much more suitable habitat available to *E. henryi* (Fig. 4). Hence, the observed phylogeographic patterns may also reflect historical factors, particularly relating to the expansion of suitable habitat during the LGM from heavily fragmented habitat during the LIG, followed by the restriction, again, of suitable habitat leading to the present disjunct montane distribution of *E. henryi* populations (Fig. 4).

Our ENM analysis through time suggested that *E. henryi* populations likely experienced cycles of expansion and retraction into and out of local refugia, and Quaternary climatic fluctuations may have played a role in

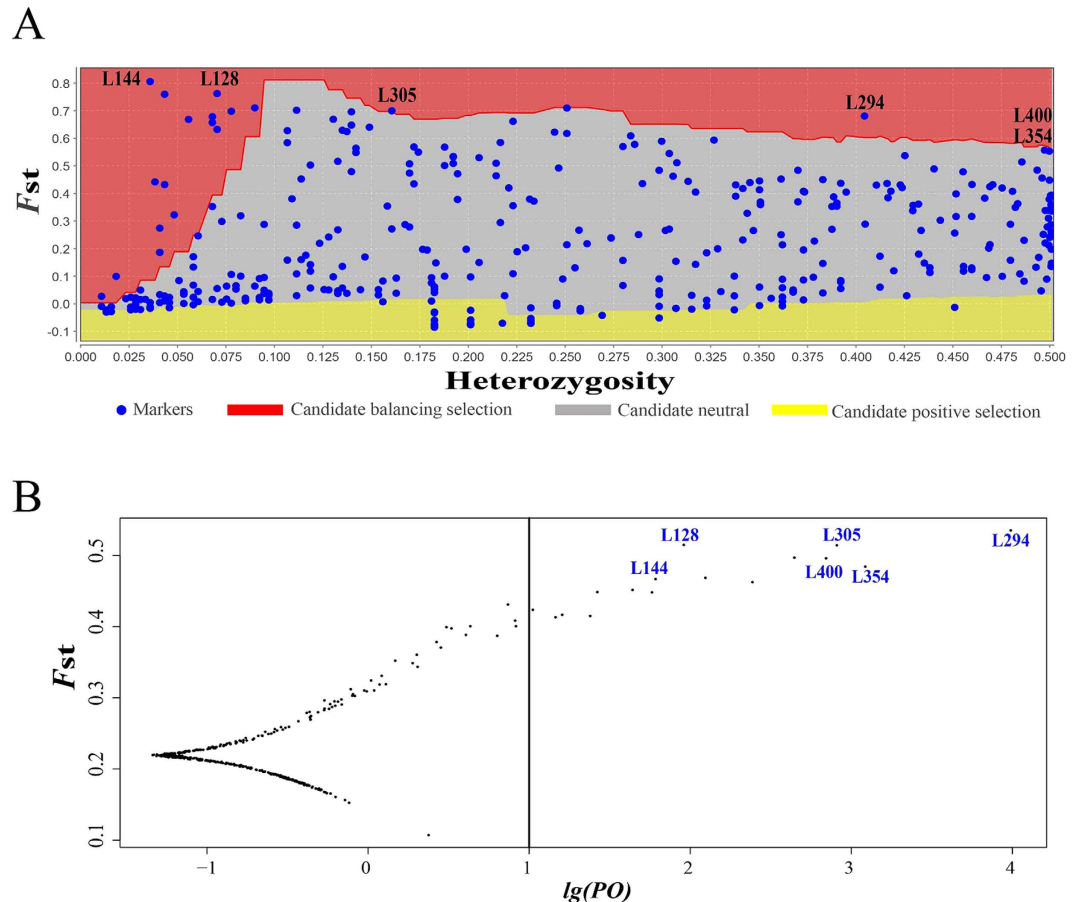


Figure 5. The outlier tests for *E. henryi* using multi-methods based on 457 AFLP loci. The outlier loci were detected by *FDIST* in *MCHEZA* (A) and by *BAYESCAN* (B), respectively, among the nine genetic clusters identified from the *BAPS* analysis. Six outlier loci were identified by both methods together.

generating phylogeographic structure as well^{26,41,42}. In fact, the earliest diversification events in *E. henryi* are associated with major geological and climatic events. The estimated divergence time between the two major lineages (Northern and Southern) of *E. henryi*, at approximately the Mio-/Pliocene boundary [c. 5.06 Ma (1.68–8.91) Ma; see node 2 in Fig. 1B], coincides with global climate cooling during the Late Miocene/Early Pliocene^{43,44}. This cooling is hypothesized to have been a key trigger of aridification in East Asia along with stronger winter and/or weaker summer monsoon circulations^{45,46}. Concurrently, the mid-Pliocene abrupt uplift of the eastern Tibetan Plateau (c. 3.4 Ma)^{46,47} induced dramatic geomorphological changes in Southwest China. These climatic and geological changes may have triggered early lineage diversification in *E. henryi* through habitat fragmentation and the formation of physical barriers to gene flow⁴⁶, which we see reflected in early branching events within each of the major Northern and Southern lineages beginning around 3.64 to 3.42 Ma (see nodes 3 and 4 in Fig. 1B). Such tectonic/climate-induced vicariance has also been invoked to explain similar patterns of north-south differentiation in other forest tree species from subtropical China (e.g. *Taxus wallichiana*⁴⁸, *Cercidiphyllum japonicum*³⁹, *Kalopanax septemlobus*⁴⁹). Expansion of potential suitable habitat of *E. henryi* during the last glaciations, indeed, is supported by our ENM analysis, which indicates larger distribution ranges of this lineage at the LGM compared to both the LIG and present (Fig. 4A–C).

In addition to the interplay between geographic and climatic factors, population demography and colonization history also likely contributed to the observed pattern of phylogeographic structure. For instance, our phylogeographic and ENM analyses are consistent with the relatively recent expansion of *E. henryi* into the northernmost regions (Northwest and Northeast groups) of its range. These groups (particular the Northeast group; N1–6) harbor lower genetic diversity (Fig. S3) and the most derived haplotypes found in our *E. henryi* phylogeny (Fig. 1B) and *tcs* networks (Fig. 1C), which both suggest populations founded following range expansion. Additionally, the MDA indicated a spatial (and demographic) expansion within the Northwest and Northeast groups at c. 0.19 Ma (95% CI: 0.111–0.314 Ma) and c. 0.26 Ma (95% CI: 0.133–0.376 Ma; Table 2), respectively, possibly coinciding with the penultimate (Riss) glacial (c. 0.12–0.35 Ma), although, due to the broad confidence intervals, this interpretation requires some caution. In fact, the species is not predicted to have occurred in the northern Sichuan Basin and the Daba/Qinling Mts. during the LIG (Fig. 4B), i.e. when temperatures were at least 5°C higher than at present⁴⁷. Accordingly, we hypothesize that the climate might have warmed enough during the LIG to extirpate *E. henryi* from the above regions, followed by recolonization via northward expansion during the

LGM (Fig. 4C). Notably, such a significant impact of the LIG has also been detected for other plants and animals in subtropical China (e.g. *Pinus kwangtungensis*⁵⁰, *Aegithalos concinnus*⁵¹, *Parus monticolus*⁵²).

The Southeast group may also carry signatures of expansion – these populations also have reduced haplotype diversity (Fig. S3A) and significantly negative Tajima's D and Fu's F_s statistics (Table 2). This provides strong evidence for a relatively recent spatial (and demographic) expansion in the Tianmu and Wuyi Mts. located southwest of the Yangtze Delta Region (YDR). Although the range of our time estimates for this expansion, again, is broad (approx. 0–0.81 Ma; Table 2), the respective point estimate (c. 0.23 Ma) once more coincides with the penultimate (Riss) glacial (see above). In addition to the ENM and MDA analyses (Figs 4B,C and S3A), this expansion scenario is further supported by vegetation reconstructions based on fossil pollen data, which indicate that the YDR still sustained patches of northern peripheral WTD forest in the lowland areas of East China during the LGM⁵³. Thus, we again find evidence that Quaternary climatic oscillations across geographic regions of China likely contributed to phylogeographic patterns in this species.

Clearly, at broad phylogeographic scales, our results demonstrate distinct geographic genetic structure across the range of *E. henryi*. At finer spatial and evolutionary scales, our landscape genetic analysis of AFLP genotype data also revealed that geographical isolation played the predominant role in driving genetic divergence of *E. henryi* populations. However, we also found that environmental/climatic variation also contributed significantly to explaining population genetic structure (SEM: IBD \approx 36.0 vs. IBE \approx 18.1%; MMRR: IBD \approx 30.72 vs. IBE \approx 24.67%). Thus, we found evidence of both isolation by distance (IBD)²¹ and isolation by environment (IBE)⁵ in our population genetic dataset. IBD could be explained by geographical distance, extensive restriction and fragmentation of WTD forests⁵⁴, physical barriers to gene flow (like the Yangtze River), or increasing fragmentation from other land uses in subtropical China (such as forest, agriculture and residential), all of which are commonly barriers to gene flow in plants³. Patterns of IBE often result from divergent selection between different environments; however, environmental factors can also shape gene flow through other processes (e.g. environmental differences affecting phenological differences among populations)⁵⁵. Thus, IBE may not be due solely to selection but could also be explained by other diverse mechanisms associated with environmental factors⁵. Hence, multiple processes, associated with both geographic and environmental variables, all could have effectively disrupted gene flow between *E. henryi* populations, and broad-scale geographic and climatic factors appear to have shaped the spatial distribution of genetic variation in this species at both deeper evolutionary and more recent ecological time scales.

To examine the potential mechanism driving the pattern of IBE, we tested for loci under selection, in our AFLP dataset, and their associations with environmental/climatic variables. In identifying outlier AFLP loci, we sought to determine how selection might play a role in shaping genetic differentiation of the nine BAPS clusters of *E. henryi* along environmental clines. All four loci identified by both F_{DIST} and $BAYESCAN$ as undergoing putative diversifying selection (Fig. 5) were associated with environmental predictors across environmental gradients (Table S7), suggesting these regions of the genome are diverging and that climate may play a role. As expected, temperature and precipitation were estimated as the major driving factors influencing allele frequencies at outlier loci, consistent with other studies examining drivers of adaptive genetic divergence in plants⁵⁶. These variables were also identified as important drivers of neutral genetic divergence by our SEM analysis (specifically BIO5, maximum temperature of the warmest month, and BIO15, precipitation seasonality). Altogether, these results are consistent with population genetic divergence under natural selection in divergent environments, suggesting this may have been the process generating IBE in this system.

In conclusion, by using phylogeographic, landscape genetic, and ecological niche modeling analysis together, we were able to identify the many factors, both historical and contemporary, that have shaped spatial genetic structure in this Tertiary relict species. At various spatial and evolutionary scales, we found that both geographical and environmental/climatic factors contribute to patterns of genetic structure. Geography played a predominant role at all levels – phylogeographic clades are broadly geographically structured, the deepest levels of divergence are associated with major geological or pre-Quaternary climatic events, and IBD primarily explained population genetic structure. However, environmental factors are clearly also important – climatic fluctuations since the LIG have likely contributed to phylogeographic structure, and the population genetic structure (in our AFLP dataset) was partly explained by IBE, which may have resulted from natural selection in environments with divergent climates. Thus, historical and contemporary geography and historical and contemporary environments have all shaped patterns of genetic structure in *E. henryi*, and, in fact, changes in the landscape through time have also been critical factors.

Methods

Plant material and sampling design. We obtained silica-dried leaf material from 38 populations of *Emmenopterys henryi* throughout its range (Table S1, Fig. 1A; see Supplementary Method S1 for more details of the study species). In each population, representative samples of 10–20 plants were taken, resulting in a total of 433 individuals. Total genomic DNA was extracted from the dried leaf tissue using a DNeasy plant tissue kit (Qiagen). All samples were sequenced at three intergenic spacer (IGS) regions of chloroplast DNA (cpDNA), while a subset of individuals ($n = 212$) was also sequenced at the entire internal transcribed spacer (ITS) region of nuclear ribosomal DNA (nrDNA). Of these 433 individuals, 394 (representing 37 populations; population S23 with 1 individual was excluded) were surveyed for AFLPs (Table S1; Fig. 1A). *Pinckneya bracteata* (Bartram) Raf., collected from the JC Raulston Arboretum (Raleigh, NC, USA), was selected as an outgroup for the phylogenetic analyses based on a previous molecular phylogenetic study of Rubiaceae³³. Voucher specimens of this species and all sampled populations of *E. henryi* are stored at the Herbarium of Zhejiang University (HZU; Hangzhou, Zhejiang, China).

DNA extraction, DNA sequencing and AFLP fingerprinting. For the phylogeographic DNA analyses, we sequenced three cpDNA-IGS regions (*psbA-trnH*, *trnL-trnF*, *trnT-trnL*) and the ITS region (i.e. ITS1 + 5.8S + ITS2). The primers and methodology for amplification of these four DNA regions were described in Shaw *et al.*⁵⁷ and Gielly *et al.*⁵⁸. Sequences were generated on an ABI 377XL DNA sequencer and were edited, assembled, and aligned in GENEIOUS v.4.8.5⁵⁹. All sequences were deposited in GenBank (see Table S2 for accession numbers).

The AFLP protocol followed the procedure described by Vos *et al.*⁶⁰, with minor modifications that included the use of fluorescent-dye-labeled primers (Applied Biosystems, Foster City, California, USA) for selective amplification in multiplex analysis. Selective primer pairs were initially screened on 50 individuals from 25 populations. Of the 64 primer pair combinations tested, nine pairs that gave the best results with respect to polymorphism and clarity of AFLP profiles (Table S3) were chosen for the full survey (further details in Supplementary Method S2). We also performed four separate test runs with 15 individuals and each chosen primer combination. These test runs confirmed that AFLP fragments were reproducible, resulting in an average error rate of 0.21% (± 0.07) per fragment (see also Knowles & Richards⁶¹). The AFLP data matrix is available in Supplementary Dataset.

Phylogeographical and population genetic data analyses. For both cpDNA and ITS, haplotype (h) and nucleotide (π) diversities were calculated for each population and the species overall using DNASP v.5.10⁶². Genealogical relationships of the haplotypes identified were inferred from a 95% statistical parsimony network constructed in TCS v.1.21⁶³, with gaps (indels) coded as substitutions (A or T). For cpDNA, the permutation test implemented in PERMUT was employed to compare parameters of population differentiation with unordered and ordered alleles (G_{ST} and N_{ST} , respectively) based on 1000 random permutations⁶⁴. Spatial analysis of molecular variance (SAMOVA), as implemented in SAMOVA v.1.0⁶⁵, was used to identify regional groups of populations that are geographically homogeneous and maximally differentiated from each other. For cpDNA data, isolation by distance (IBD) effects²¹ were assessed by regressions of $F_{ST}/(1 - F_{ST})$ against the logarithm (\log_{10}) of geographic distance for all pairs of populations, or subsets thereof, following Rousset⁶⁶.

For the AFLP dataset, we calculated the total number of AFLP fragments per population (F_{+}), the percentage of fragments that are polymorphic within each population (PPF), Nei's (1973) gene diversity (H_E), and Shannon's information index I following the method of Lynch & Milligan⁶⁷ using AFLPSURV v.1.0⁶⁸. In addition, 'frequency-down-weighted-marker values' (DW ; according to Schönswetter & Tribsch⁶⁹), were calculated for each population. To infer the most likely number of population genetic clusters (K) in the AFLP dataset, we used three approaches. First, we used BAPS v.6.0⁷⁰ to detect clusters of genetically similar populations and to estimate individual coefficients of ancestry (q) with regard to the detected clusters. Second, we utilized GENALEX v.6.4⁷¹ to run a principal coordinates analysis (PCoA), which non-hierarchically grouped the samples without prior knowledge of their source location. Third, we constructed an unrooted neighbour-joining (NJ) tree using the PHYLIP package 3.6 with 1000 bootstraps (see Supplementary Method S3 for more details). To quantify variation in cpDNA sequences and AFLPs among populations and genetic clusters (as identified by SAMOVA), we performed analyses of molecular variance (AMOVAs) in ARLEQUIN v.3.5⁷² using Φ -statistics, respectively. The significance of fixation indices was tested using 10,000 permutations⁷³.

Divergence time estimation and demographic analyses based on cpDNA sequences. To relate differentiation among cpDNA haplotypes of *E. henryi* to pre-Quaternary and Quaternary events, we estimated divergence time under a Bayesian approach as implemented in BEAST v.1.8.0⁷⁴. We estimated the divergence time using an unlinked substitution model (*psbA-trnH*: HKY; *trnL-trnF*: HKY; *trnT-trnL*: GTR+I) selected by JMODELTEST v.2.0⁷⁵. Given the intraspecific nature of our data, an uncorrelated lognormal distributed relaxed clock (UCLD) model, and a coalescent model assuming constant population size were used to model the tree prior. Since no reliable fossils of *E. henryi* are currently known (see Introduction), we used the divergence time between *E. henryi* and *P. bracteata*, inferred from a fossil-calibrated cpDNA phylogeny of Rubiaceae, as a secondary calibration point to calibrate the stem node of *E. henryi*³³ (node 1 in Fig. 1B: mean = 21.77 Ma; SD = 5.0 Ma; 95% CI = 7.92–33.64 Ma). For divergence time estimations, the Markov chain Monte Carlo (MCMC) was run for three independent 50 million generation chains, sampling from the chain every 5000 generations. The output was visualized using TRACER v.1.5 (<http://tree.bio.ed.ac.uk/software/tracer/>) to ensure that parameter values were fluctuating at stable levels. Based on these results, the first 5000 trees were discarded as burn-in, and the remaining samples were summarized as a maximum clade credibility (MCC) tree with mean divergence times and 95% highest posterior density (HPD) intervals of age estimates in TREEANNOTATOR v.1.8.0⁷⁶. Finally, these results were summarized in a single tree visualized in FIGTREE v.1.3.1 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Tajima's D ⁷⁷ and Fu's F_s ⁷⁸ neutrality tests were used to assess population demographic history. We also used ARLEQUIN to infer the demographic history of two main lineages identified ('Northern' vs. 'Southern'; see the Results Section) from their combined cpDNA-IGS sequences by mismatch distribution analysis (MDA) (further details in Supplementary Method S4).

Present, past and future distribution modelling. Assuming *E. henryi* has not changed (and will not change) its climatic preference over at least the last glacial/interglacial cycle and into the future⁷⁹, we reconstructed ENMs to determine the potential distribution of *E. henryi* across its range under present (1950–2000), past [the Last Glacial Maximum (LGM, c. 21 kya)^{80,81} and the Last Interglacial (LIG, c. 130–114 kya)⁸²] and future (2080) climatic conditions using the maximum entropy method implemented in MAXENT v.3.2.1. In addition to the 38 distribution records included in this study, 76 herbarium-recorded presences were sourced from the Chinese Virtual Herbarium (www.cvh.org.cn) and the National Specimen Information Infrastructure of China (www.nsi.org.cn). Based on a total of 114 records, a current distribution model was developed using six bioclimatic data layers (annual mean temperature, annual precipitation, precipitation of wettest, driest, warmest and coldest quarter) available from the WorldClim database (www.worldclim.org)⁸³ at 2.5 arc-min resolution

for the present (1950–2000), assumed to be important for temperate species in East Asia^{39,49}. This restricted dataset was used to avoid including highly correlated variables (data not shown), and thus to prevent potential over-fitting⁸⁴. The established model was then projected onto the set of climatic variables simulated by the Model for Interdisciplinary Research on Climate (MIROC)⁸⁵ to infer the extent of suitable habitat during the LGM and the LIG (see above). Future projections for 2080 were performed based on the Canadian Centre for Climate Modelling and Analysis model (CCCMA-CGCM31) under the A2 scenario that were provided by the CIAT downscaled GCM Data Portal (<http://gisweb.ciat.cgiar.org/GCMPPage/>) with 30 arc-second resolution.

Model performance was evaluated using receiver operating characteristic (ROC) analyses in MAXENT. Values of the area under the ROC curve (AUC) between 0.7 and 0.9 indicate good fit⁸⁶. We modelled the modern distribution 10 times, using different subsets of 70% of the localities to train the model and 30% to test the model, and visually compared AUC scores and jackknife tests of variable importance to assess consistency between runs.

IBD and IBE analyses. To quantify the roles of geography and ecology in spatial genetic divergence, we quantified isolation by distance (IBD)²¹ and isolation by environment (IBE)⁸⁷ using our AFLP dataset, because multi-locus datasets with highly variable markers are ideal for this purpose^{5,88,89}. We computed population pairwise genetic distances (F_{ST}) with ARLEQUIN using all putatively neutral (non-outlier) AFLP loci (i.e. 451 loci) described above. We obtained a total of 21 environmental and geographical variables for our study area, including 19 bioclimatic variables with 30 arc-second resolution (<http://www.bioclim.org>), a slope layer based on a digital elevation model with 1-km resolution from the USGS EROS database (<http://eros.usgs.gov>), and soil type from the Chinese soil taxonomy record⁹⁰. We used resistance-based distances, to reflect biological connectivity between populations, instead of direct geographical distances⁸⁹ (see Supplementary Method S5 for more details).

We then employed two related methods to quantify IBD and IBE: multiple matrix regression with randomization (MMRR)⁸⁸ and structural equation modelling (SEM)⁸⁹. Both methods use a series of regression-based analyses to quantify the effects of multiple explanatory variables on a single response variable, in this case genetic distance (F_{ST}). MMRR provides a straightforward method for estimating linear regressions among distance matrices, while SEM allows for flexibility in defining the relationships among variables and utilizes latent variable modelling. For the MMRR analysis, we constructed an environmental distance matrix by taking the population pairwise Euclidian distances across all 21 environmental variables. We then analyzed the effects of environmental distance and geographic distance, as explanatory variables, on genetic distance, as the response variable, using the ‘MMRR’ function in R⁸⁸ with 10,000 permutations. For the SEM analysis, we used all 21 environmental variables to construct an environmental distance latent variable, and then quantified the relationships between environmental distance, geographic distance, and genetic distance using the ‘lavaan’ package in R (<http://cran.r-project.org/web/packages/lavaan/>). Prior to analysis, all environmental, geographic, and genetic distance matrices were standardized. In each case, the analysis was used to test three alternative hypotheses: (1) only geographical distance contributes significantly to genetic distance, (2) only environmental distance contributes significantly to genetic distance, and (3) geographical distance and environmental distance each contribute significantly to genetic distance. We performed both analyses on AFLP pairwise genetic distances (F_{ST}) for all populations.

Potential AFLP loci under selection: detection of outlier loci and associations between allele frequencies and environmental variables. Two different basic approaches allow screening for AFLP loci that are putatively under selection: (i) a genome scan test procedure based on F_{ST} comparisons to identify the loci that are significantly different than expected under neutrality and a given demographic model; and (ii) correlative methods that look for associations between genetic variation and environmental variables. The key to using these approaches together is that they identify links between putative loci under selection and individual environmental variables. Hence, our primary purpose for using these tests was to identify the specific environmental variables that are potential drivers of genetic differentiation via divergent natural selection (e.g. IBE resulting from selection in divergent environments).

For the genome scan test, both the F_{DIST} ⁹¹ and BAYESCAN⁹² approaches were implemented on 457 polymorphic AFLP loci using the nine genetic clusters defined in the BAPS analysis ($K=9$). Then, to detect associations between allele frequencies and environmental variables, we used Multiple Linear Regression (MLR)⁹³ in R v.3.1.1 (R Development Core Team 2011) to identify potential adaptive loci that are under selection from current environmental factors. For outlier loci confirmed by both F_{DIST} and BAYESCAN, we computed their population pairwise frequencies of AFLP alleles at the 37 sampling sites. We extracted values for the geographical coordinates of each population from 19 bioclimatic GIS data layers with 30 arc-second resolution (1950–2000; <http://www.bioclim.org>). We then regressed the allele frequencies of the retained outlier loci (dependent variables) on the selected environmental variables (explanatory variables; standardized) using the MLR model (see Table S4 for the retained allele frequencies and the selected environmental variables). Potential adaptive loci were identified as $R_{adj}^2 > 0.5$ and significantly correlated to at least one explanatory variable¹³. Univariate regressions were then conducted for each variable individually to estimate its significance (see Supplementary Method S6 for more details).

References

1. Loveless, M. D. & Hamrick, J. L. Ecological determinants of genetic structure in plant populations. *Annu. Rev. Ecol. Syst.* **15**, 65–95 (1984).
2. Lester, S. E., Ruttenberg, B. I., Gaines, S. D. & Kinlan, B. P. The relationship between dispersal ability and geographic range size. *Ecol. Lett.* **10**, 745–758 (2007).
3. Holderegger, R., Buehler, D., Gugerli, F. & Manel, S. Landscape genetics of plants. *Trends Plant Sci.* **15**, 675–683 (2010).
4. Nosil, P. Ecological speciation and its alternatives in *Ecological speciation*. (ed. Nosil, P.) 3–21 (Oxford University Press Inc., New York, 2012).
5. Wang, I. J. & Bradburd, G. S. Isolation by environment. *Mol. Ecol.* **23**, 5649–5662 (2014).

6. Garant, D., Kruuk, L. E. B., Wilkin, T. A., McCleery, R. H. & Sheldon, B. C. Evolution driven by differential dispersal within a wild bird population. *Nature* **433**, 60–65 (2005).
7. Postma, E. & van Noordwijk, A. J. Gene flow maintains a large genetic difference in clutch size at a small spatial scale. *Nature* **433**, 65–68 (2005).
8. Rundle, H. D. & Nosil, P. Ecological speciation. *Ecol. Lett.* **8**, 336–352 (2005).
9. Dyer, R. J., Nason, J. D. & Garrick, R. C. Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Mol. Ecol.* **19**, 3746–3759 (2010).
10. Freeland, J. R., Biss, P., Conrad, K. F. & Silvertown, J. Selection pressures have caused genome-wide population differentiation of *Anthoxanthum odoratum* despite the potential for high gene flow. *J. Evolution Biol.* **23**, 776–782 (2010).
11. Pease, K. M. *et al.* Landscape genetics of California mule deer (*Odocoileus hemionus*): the roles of ecological and historical factors in generating differentiation. *Mol. Ecol.* **18**, 1848–1862 (2009).
12. Freedman, A. H., Thomassen, H. A., Buermann, W. & Smith, T. B. Genomic signals of diversification along ecological gradients in a tropical lizard. *Mol. Ecol.* **19**, 3773–3788 (2010).
13. Manel, S., Poncet, B. N., Legendre, P., Gugerli, F. & Holderegger, R. Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Mol. Ecol.* **19**, 3824–3835 (2010).
14. Sork, V. L. *et al.* Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Nees) in the face of climate change. *Mol. Ecol.* **19**, 3806–3823 (2010).
15. Yan, F. *et al.* Geological events play a larger role than Pleistocene climatic fluctuations in driving the genetic structure of *Quasipaa boulengeri* (Anura: Dicroglossidae). *Mol. Ecol.* **22**, 1120–1133 (2013).
16. Epps, C. W., Wasser, S. K., Keim, J. L., Mutayoba, B. M. & Brashares, J. S. Quantifying past and present connectivity illuminates a rapidly changing landscape for the African elephant. *Mol. Ecol.* **22**, 1574–1588 (2013).
17. He, Q. X., Edwards, D. L. & Knowles, L. L. Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution* **67**, 3386–3402 (2013).
18. Lomolino, M. V., Riddle, B. R., Whittaker, R. J. & Brown, J. A. Glaciation and biogeographic dynamics of the Pleistocene in *Biogeography* 4th edn (eds Lomolino, M. V. *et al.*) Ch. 9, 313–357 (Sinauer Associates, Inc., Sunderland, Massachusetts, 2010).
19. Hewitt, G. The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913 (2000).
20. Kropf, M., Comes, H. P. & Kadereit, J. W. Past, present and future of mountain species of the French Massif Central – the case of *Soldanella alpina* L. subsp. *alpina* (Primulaceae) and a review of other plant and animal studies. *J. Biogeogr.* **39**, 799–812 (2012).
21. Wright, S. Isolation by distance. *Genetics* **28**, 114–138 (1943).
22. Slatkin, M. Gene flow and the geographic structure of natural populations. *Science* **236**, 787–792 (1987).
23. Ying, T. S. & Zhang, Y. L. *The endemic genera of seed plants of China*. 516–518 (Science Press, Beijing, 1994) (In Chinese).
24. Qian, H. A comparison of the taxonomic richness of temperate plants in East Asia and North America. *Am. J. Bot.* **89**, 1818–1825 (2002).
25. Wu, Z. Y., Sun, H., Zhou, Z. K., Peng, H. & Li, D. Z. Origin and differentiation of endemism in the flora of China. *Front. Biol. China* **2**, 125–143 (2007).
26. Qiu, Y. X., Fu, C. X. & Comes, H. P. 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 (2011).
27. Tiffney, B. H. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. *J. Arnold Arboretum* **66**, 243–273 (1985).
28. Latham, R. E. & Ricklefs, R. E. Continental comparisons of temperate-zone tree species diversity in *Species diversity in ecological communities* (eds Ricklefs, R. E. & Schluter, D.) 294–314 (University of Chicago Press, Chicago, IL, USA, 1993).
29. Kubitzki, K. & Krutzsch, W. Origins of east and south east Asian plant diversity in *Floristic characteristics and diversity of East Asian plants – Proceedings of the first International Symposium on Floristic Characteristics and Diversity of East Asian Plants* (eds Zhang, A. L. *et al.*) 25–27 (Higher Education Press, China, 1996).
30. Chen, T. & Taylor, C. M. Rubiaceae – *Emmenopterys* in *Flora of China*, Vol. 19 (eds Wu, Z. Y. & Raven, P. H.) 102 (Science Press & Missouri Botanical Garden Press, 1989–2013).
31. Huang, H. W., Oldfield, S. & Qian, H. Global significance of plant diversity in China in *Plants of China* (eds Hong, D. Y. & Blackmore, S.) Ch. 2, 7–31 (Science Press, Beijing, China, 2013).
32. Manchester, S. R., Chen, Z. D., Lu, A. M. & Uemura, K. Eastern Asian endemic seed plant genera and their paleogeographic history throughout the Northern Hemisphere. *J. Syst. Evol.* **47**, 1–42 (2009).
33. Manns, U., Wikström, N., Taylor, C. M. & Bremer, B. Historical biogeography of the predominantly neotropical subfamily Cinchonoideae (Rubiaceae): into or out of America? *Int. J. Plant Sci.* **173**, 261–286 (2012).
34. Li, J. M. & Jin, Z. X. Genetic structure of endangered *Emmenopterys henryi* Oliv. based on ISSR polymorphism and implications for its conservation. *Genetica* **133**, 227–234 (2008).
35. Wang, I. J. Recognizing the temporal distinctions between landscape genetics and phylogeography. *Mol. Ecol.* **19**, 2605–2608 (2010).
36. Graur, D. & Li, W. H. Rates and patterns of nucleotide substitution in *Fundamentals of molecular evolution*, 2nd edn. (eds Graur, D. & Li, W. H.) Ch. 4, 99–164 (Sinauer Associates, Inc., Sunderland, MA, USA, 2000).
37. Kay, K. M., Whittall, J. B. & Hodges, S. A. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. *BMC Evol. Biol.* **6**, 36 (2006).
38. Avise, J. C., Nelson, W. S. & Sugita, H. A speciation history of living fossils molecular evolutionary patterns in horseshoe crabs. *Evolution* **48**, 1986–2001 (1994).
39. Qi, X. S. *et al.* Molecular data and ecological niche modelling reveal a highly dynamic evolutionary history of the East Asian Tertiary relict *Cercidiphyllum* (Cercidiphyllaceae). *New Phytol.* **196**, 617–630 (2012).
40. Liu, J. *et al.* Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. *New Phytol.* **199**, 1093–1108 (2013).
41. Liu, H. Y., Xing, Q. R., Ji, Z. K., Xu, L. H. & Tian, Y. H. An outline of Quaternary development of *Fagus* forest in China: palynological and ecological perspectives. *Flora* **198**, 249–259 (2003).
42. Qiu, Y. X., Guan, B. C., Fu, C. X. & Comes, H. P. Did glacials and/or interglacials promote allopatric incipient speciation in East Asian temperate plants? Phylogeographic and coalescent analyses on refugial isolation and divergence in *Dyosma versipellis*. *Mol. Phylogenet. Evol.* **51**, 281–293 (2009).
43. Shackleton, N. J. New data on the evolution of Pliocene climatic variability in *Paleoclimate and evolution, with emphasis on human origins* (ed. Vrba, E. S.) 242–248 (Yale University Press, New Haven, USA, 1995).
44. Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).
45. Heermance, R. V., Chen, J., Burbank, D. W. & Wang, C. Chronology and tectonic controls of Late Tertiary deposition in the southwestern Tian Shan foreland, NW China. *Basin Res.* **19**, 599–632 (2007).
46. Ge, J. Y. *et al.* Major changes in East Asian climate in the mid-Pliocene: triggered by the uplift of the Tibetan Plateau or global cooling? *J. Asian Earth Sci.* **69**, 48–59 (2013).
47. Shi, Y. F. *et al.* Uplift of the Qinghai-Xizang (Tibetan) Plateau and east Asia environmental changes in the Late Cenozoic. *Acta. Geogr. Sin.* **54**, 10–20 (1999).

48. Gao, L. M. *et al.* High variation and strong phylogeographic pattern among cpDNA haplotypes in *Taxus wallichiana* (Taxaceae) in China and North Vietnam. *Mol. Ecol.* **16**, 4684–4698 (2007).
49. Sakaguchi, S. *et al.* Climate oscillation during the Quaternary associated with landscape heterogeneity promoted allopatric lineage divergence of a temperate tree *Kalopanax septemlobus* (Araliaceae) in East Asia. *Mol. Ecol.* **21**, 3823–3838 (2012).
50. Tian, S., Lopez-Pujol, J., Wang, H. W., Ge, S. & Zhang, Z. Y. Molecular evidence for glacial expansion and interglacial retreat during Quaternary climatic changes in a montane temperate pine (*Pinus kwangtungensis* Chun ex Tsiang) in southern China. *Plant Syst. Evol.* **284**, 219–229 (2010).
51. Dai, C. *et al.* Profound climatic effects on two East Asian black-throated tits (Ave: Aegithalidae), revealed by ecological niche models and phylogeographic analysis. *Plos One* **6**, e29329 (2011).
52. Wang, W. *et al.* Glacial expansion and diversification of an East Asian montane bird, the green-backed tit (*Parus monticolus*). *J. Biogeogr.* **40**, 1156–1169 (2013).
53. Harrison, S. P., Yu, G., Takahara, H. & Prentice, I. C. Palaeovegetation. Diversity of temperate plants in east Asia. *Nature* **413**, 129–130 (2001).
54. Li, M. S. *et al.* Assessment of forest geospatial patterns over the three giant forest areas of China. *J. Forestry Res.* **19**, 25–31 (2008).
55. Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J. & De Meester, L. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Mol. Ecol.* **22**, 5983–5999 (2013).
56. Yoder, J. B. *et al.* Genomic signature of adaptation to climate in *Medicago truncatula*. *Genetics* **196**, 1263–1275 (2014).
57. Shaw, J., Lickey, E. B., Schilling, E. E. & Small, R. L. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am. J. Bot.* **94**, 275–288 (2007).
58. Gielly, L., Yuan, Y. M., Küpfer, P. & Taberlet, P. Phylogenetic use of noncoding regions in the genus *Gentiana* L.: chloroplast *trnL* (UAA) intron versus nuclear ribosomal internal transcribed spacer sequences. *Mol. Phylogenet. Evol.* **5**, 460–466 (1996).
59. Drummond, A. J. *et al.* Geneious 4.8.5 (Biomatters Ltd., Auckland, New Zealand, 2009). Available at <http://www.geneious.com/previous-versions>.
60. Vos, P. *et al.* AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**, 4407–4414 (1995).
61. Knowles, L. L. & Richards, C. L. Importance of genetic drift during Pleistocene divergence as revealed by analyses of genomic variation. *Mol. Ecol.* **14**, 4023–4032 (2005).
62. Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497 (2003).
63. Clement, M., Posada, D. & Crandall, K. A. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**, 1657–1659 (2000).
64. Pons, O. & Petit, R. J. Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* **144**, 1237–1245 (1996).
65. Dupanloup, I., Schneider, S. & Excoffier, L. A simulated annealing approach to define the genetic structure of populations. *Mol. Ecol.* **11**, 2571–2581 (2002).
66. Rousset, F. Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics* **145**, 1219–1228 (1997).
67. Lynch, M. & Milligan, B. G. Analysis of population genetic structure with RAPD markers. *Mol. Ecol.* **3**, 91–99 (1994).
68. Vekemans, X., Beauwens, T., Lemaire, M. & Roldan-Ruiz, I. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Mol. Ecol.* **11**, 139–151 (2002).
69. Schönswetter, P. & Tribsch, A. Vicariance and dispersal in the alpine perennial *Bupleurum stellatum* L. (Apiaceae). *Taxon* **54**, 725–732 (2005).
70. Corander, J., Marttinen, P., Siren, J. & Tang, J. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9**, 539 (2008).
71. Peakall, R. & Smouse, P. E. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **6**, 288–295 (2006).
72. Excoffier, L. & Lischer, H. E. L. 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 (2010).
73. Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
74. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
75. Durrin, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772–772 (2012).
76. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
77. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
78. Fu, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925 (1997).
79. Nogues-Bravo, D. Predicting the past distribution of species climatic niches. *Global. Ecol. Biogeogr.* **18**, 521–531 (2009).
80. Braconnot, P. *et al.* Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 1: experiments and large-scale features. *Clim. Past* **3**, 261–277 (2007).
81. Braconnot, P. *et al.* Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 2: feedbacks with emphasis on the location of the ITCZ and mid- and high latitudes heat budget. *Clim. Past* **3**, 279–296 (2007).
82. Otto-Bliesner, B. L. *et al.* Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science* **311**, 1751–1753 (2006).
83. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
84. Peterson, A. T. & Nakazawa, Y. Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. *Global. Ecol. Biogeogr.* **17**, 135–144 (2008).
85. K-1 Model Developers. K-1 coupled GCM (MIROC) description in *K-1 Technical Report 1 of the Kyousei project* (eds Hasumi, H. & Emori, S.) 1–28 (Center for Climate System Research, University of Tokyo, Tokyo, 2004). URL http://ccsr.aori.u-tokyo.ac.jp/~hasumi/miroc_description.pdf.
86. Swets, J. A. Measuring the accuracy of diagnostic systems. *Science* **240**, 1285–1293 (1988).
87. Wang, I. J. & Summers, K. Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Mol. Ecol.* **19**, 447–458 (2010).
88. Wang, I. J. Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution* **67**, 3403–3411 (2013).
89. Wang, I. J., Glor, R. E. & Losos, J. B. Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecol. Lett.* **16**, 175–182 (2013).
90. Chen, H. Z. & Gong, Z. T. Regularity of soil distribution and soil division in Chinese soil taxonomic classification (ed. Gong, Z. T.) *Ch. 22*, 860–863 (China Science Press, Beijing, China, 1999).

91. Beaumont, M. A. & Nichols, R. A. Evaluating loci for use in the genetic analysis of population structure. *P. Roy. Soc. B.: Biol. Sci.* **263**, 1619–1626 (1996).
92. Foll, M. & Gaggiotti, O. A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977–993 (2008).
93. Zulliger, D., Schnyder, E. & Gugerli, F. Are adaptive loci transferable across genomes of related species? Outlier and environmental association analyses in Alpine Brassicaceae species. *Mol. Ecol.* **22**, 1626–1639 (2013).
94. Environmental Systems Research Institute. ArcGIS Desktop 9, ArcView 9.3, Single Use. (Environmental Systems Research Institute, Redlands, CA, USA, 2008). Available at <http://www.esri.com/software/arcgis/arcgis-for-desktop>.

Acknowledgements

We thank Shi-Liang Zhou for help for help with the AFLP analysis. We are grateful to Sheng-Nan Zhai, Yun-Rui Mao, Na Yuan, Yi Ren, Zhang-Fu Wang, and Deng-Long Ha for help for help with the field survey and leaf collection, and to Dr. Shota Sakaguchi from the University of Tokyo for help with the ENM analysis. This research was supported by the National Science Foundation of China (Grant Nos 31370241 and 31570214) and the International Cooperation and Exchange of the National Natural Science Foundation of China (Grant Nos 31511140095, 31561143015).

Author Contributions

Y.X.Q. conceived and designed the study, Y.H.Z. and H.P. collected samples, Y.H.Z. performed experiments, Y.H.Z. and I.J.W. analyzed the data, Y.X.Q., Y.H.Z., H.P.C. and I.J.W. wrote the manuscript. All authors revised the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zhang, Y.-H. *et al.* Contributions of historical and contemporary geographic and environmental factors to phylogeographic structure in a Tertiary relict species, *Emmenopterys henryi* (Rubiaceae). *Sci. Rep.* **6**, 24041; doi: 10.1038/srep24041 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>