Molecular Phylogenetics and Evolution 132 (2019) 36–45

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Land bridges in the Pleistocene contributed to flora assembly on the continental islands of South China: Insights from the evolutionary history of *Quercus championii*

Xiao-Long Jiang[a,b,e], Elliot M. Gardner[c], Hong-Hu Meng[d], Min Deng[b,e,*], Gang-Biao Xu[a,*]

[a] The Laboratory of Forestry Genetics, Central South University of Forestry and Technology, Changsha 410004, Hunan, China
[b] Shanghai Chenshan Botanical Garden/ Shanghai Chenshan Plant Science Research Center, Chinese Academy of Sciences, Shanghai 201602, China
[c] The Morton Arboretum, 4100 IL-53, Lisle, IL 60532, USA
[d] Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming 650223, China
[e] Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences, Yezin, Nay Pyi Taw 05282, Myanmar

ARTICLE INFO

Keywords:
Demographic dynamics
Approximate Bayesian computation
Species dispersal
Local adaptation
RAD-seq
Oaks

ABSTRACT

The South China Mainland (SCM) and its adjacent continental islands are a global biodiversity hotspot. However, how and when plants dispersed between SCM and Hainan/Taiwan Islands remains largely unknown. In this study, we used restriction site-associated DNA sequencing (RAD-seq) to identify the demographic dynamics and local adaptation of *Quercus championii*, a dominant forests tree distributed in SCM and Hainan/Taiwan Islands. Through phylogenetic reconstruction, principal components analysis (PCA) and structure analysis, we identified four distinct *Q. championii* lineages that correspond to its geographical distribution. The genetic structure of Hainan Island population was distinct, possibly reflecting an introgression. We conducted an approximate Bayesian computation analyses and found that *Q. championii* originated from Southwest China-Northern Vietnam, then dispersed to Southeast China as the climate warmed. During the Pleistocene glacial period, land bridges arose between SCM and Hainan/Taiwan Islands, and the land bridges likely facilitated species dispersal from SCM to these islands. We found a strong correlation between genetic variation and isothermality through a gradient forest analysis and identified precipitation seasonality as a key driver to the local adaptation of *Q. championii*. Finally, we analyzed putative adaptation loci and identified genes regulating vegetative and reproductive organ development as important for the adaptation of *Q. championii* to heterogeneous environments. We provide new insights into the evolutionary history and local adaptation of biotas in Southern China and adjacent islands.

1. Introduction

Islands have provided a framework for understanding evolution from the very earliest studies of evolutionary theory (Darwin, 1839; Wallace, 1869). Islands can form in different ways. Oceanic islands usually arise from volcanic activity, and continental islands form when rising sea levels separate a portion of a mainland into an island. Continental islands are typically separated from continents by narrow and shallow waters, and can be considered relicts of the adjacent mainland (De Queiroz, 2005; Whittaker and Fernández-Palacios, 2007). In contrast to oceanic islands that generally do not have connection to a continent, fluctuating sea levels may cause continental islands to connect and separate from the mainland repeatedly, a process that can promote the evolution of unique biotas in continental islands (Chen et al., 2015; Whittaker and Fernández-Palacios, 2007). Isolated on small landmasses, insular biotas may be more susceptible to global climate change compared to mainland organisms (Bellard et al., 2014; Crowl et al., 2016; Taylor and Kumar, 2016). Disentangling the biogeographic history and local adaptation of species with disjunct distributions on a mainland and adjacent islands can provide insights into the evolution of continental island biotas and can guide the conservation of islands biotas facing climate change.

Adjacent to the Pacific coast of Southern China are numerous continental islands, including the largest islands Taiwan and Hainan. Species diversity on the two islands is quite similar, with ca. 3600 species from 1201 genera in Taiwan and ca. 3800 species from 1283

* Corresponding authors.
E-mail addresses: xiaolongjiang1@gmail.com (X.-L. Jiang), egardner@mortonarb.org (E.M. Gardner), menghonghu@xtbg.ac.cn (H.-H. Meng), dengmin@sibs.ac.cn (M. Deng), gangbiaoxu@163.com (G.-B. Xu).

https://doi.org/10.1016/j.ympev.2018.11.021
Received 9 August 2018; Received in revised form 19 October 2018; Accepted 26 November 2018
Available online 01 December 2018

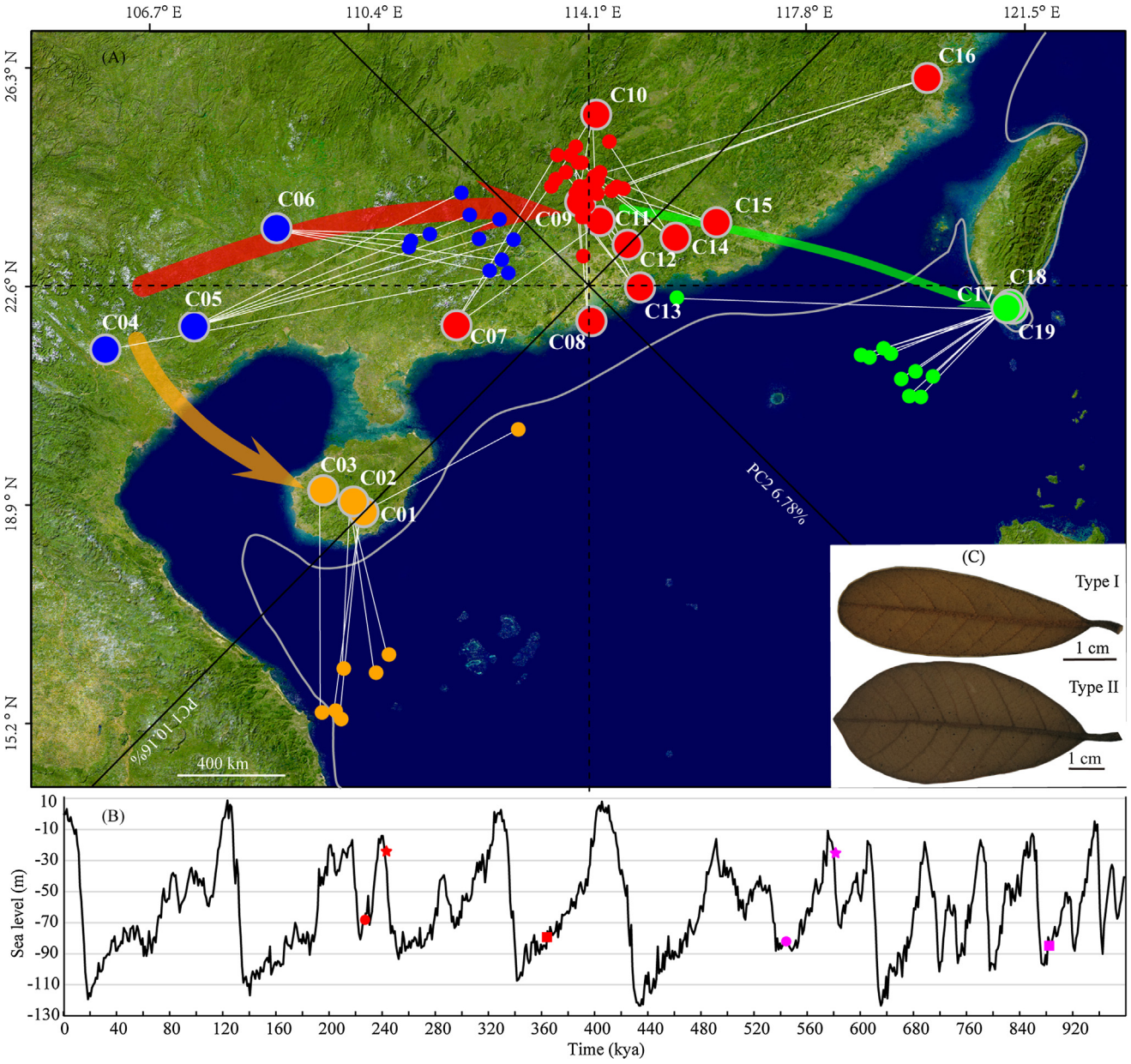

Fig. 1. (A) Procrustes-transformed principal component analysis (PCA) plot of genetic variation of 56 *Quercus championii* individuals mapped onto the first two principal components (the small circles) relative to the geographic location of populations (the larger circles). Colors distinguish different geographic locations, where yellow represents HN, blue represents SW, red represents SE and green represents TW. White straight lines show the orientations of PC1 and PC2 for the genetic data relative to the longitude and latitudinal axes. Grey curves represent the coastline when sea level was 80 m lower than today. Arrows point to the potential migration route of *Q. championii* inferred from DIYABC, and arrow thickness is proportional to the effective population size of each group. (B) Sea level change in the SCM region over time. Red and pink marks represent the migration times of *Q. championii* calculated using 50 and 120 years as generation times, respectively. Squares correspond to the time of species migration to Hainan Island from Southwest China and Northern Vietnam, stars correspond to the time of migration to Southeast China from Southwest China and circles correspond to the time of migration to Taiwan Island from southeast China. (C) Morphological characteristics of *Q. championii* leaves under different habitats. Species with type I leaves is distribution in mountainsides or mountaintop, and type II leaves is distribution in ravines or riverside. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

genera in Hainan (Ying and Xu, 2002; Zhu, 2016). The flora of Hainan is most similar to Guangxi province in Southern China and Northern Vietnam (Zhu, 2016), and the flora of Taiwan is most similar to the Southeast China mainland (Ying and Xu, 2002). Most mountain regions of East Asian mainland and its nearby continental islands are biodiversity hotspot (Myers et al., 2000; Tang et al., 2006), and numerous studies have focused on the evolution of the flora of this region recently. Phylogeographic studies of species distributions in mainland China, Japan, and the Korean Peninsula (CJK) have illuminated the evolution of disjunct species in these temperate regions (Bai et al., 2010; Qiu et al., 2009a; Qiu et al., 2009b; Sakaguchi et al., 2012; Ye et al., 2017). However, little is known about the evolution and biogeographic history of the flora in the South China mainland (SCM) and its adjacent islands. One hypothesis posited that the formation of Hainan Island and its flora were tightly related to its southeastern drifting from the south China plate during the Cenozoic (Zhu, 2016, 2017), and this hypothesis is based on the similarity of regional floras from South China and Hainan; however, it does not take into consideration the phylogeography of representative disjunct species. Sea level fluctuations in the Pleistocene resulted in recurring land bridges

connecting SCM and the continental islands (Voris, 2000), which may have provided suitable habitats for species to migrate (Heaney, 1985). Few biogeographic studies have explored the evolution of species with disjunct distributions in SCM and Hainan/Taiwan Islands, and the existing studies have found divergent biogeographic patterns. For example, climate change during the Pleistocene contributed to divergence and population expansion of the bat *Hipposideros armiger*, and *H. armiger* populations from Taiwan and Hainan islands are closely related to those from Southeast China (Lin et al., 2014); gibbons appeared to have migrated from eastern Indochina to Hainan Island (Chatterjee, 2006). Therefore, the assemblage of biotas on these continental islands along SCM needs further investigation.

The evergreen oak *Quercus championii* Benth. (Fagaceae) belongs to section *Cyclobalanopsis* (ring-cupped oaks). Its sister species is *Q. gilva* Blume (Deng et al., 2018), which is widely distributed in East China and Japan, but the morphological differences of the two species are prominent. For example, *Q. gilva* has leaves with serrations, an acuminate apex, and 11–18 secondary veins on each side of the midvein, while *Q. championii* has leaves with a recurved and entire margin, a short, blunt, rarely retuse apex, and 6–10 secondary veins on each side of the midvein. The distribution of *Q. championii* is southeast China, southwest China, northern Vietnam, and islands adjacent to SCM. While *Q. championii* has been a widespread and dominant tree historically, its populations have declined recently from a rapid habitat loss caused by anthropogenic activities. Although *Q. championii* on SCM is found in subtropical evergreen forests, it is restricted to tropical mountains on Hainan and Taiwan islands (Yu et al., 2000). Oak seeds are typically dispersed only short distances by gravity and rodents. Therefore, the distribution patterns and seed dispersal mode of *Q. championii* provide a unique proxy to gain insights into the evolution and biogeographic history of SCM and its adjacent islands, as well as the local adaptation to varying environments.

Here, we examined the genetic structure, demographic history and environmental adaptation of *Q. championii* using single-nucleotide polymorphisms (SNPs) from restriction site associated DNA sequencing (RAD-seq). RAD-seq can generate genome-wide sequence data from thousands of orthologous loci without a reference genome (Andrews et al., 2016), and has been an effective approach to investigate the evolutionary history of long-lived and frequently-hybridizing species, such as oaks (Cavender-Bares et al., 2015; Deng et al., 2018; Hipp et al., 2014; Hipp et al., 2017; McVay et al., 2017). The aims of the current study were: (1) to reveal the spatial genetic structure of *Q. championii* at the genomic level; (2) to illuminate the biogeographic history of *Q. championii* on SCM and its continental islands; and (3) to explore local adaptation of *Q. championii* to environmental heterogeneity.

2. Materials and methods

2.1. Sampling and RAD-seq library preparation

Samples from 262 *Q. championii* individuals from 19 natural populations in Southern China and Northern Vietnam were collected during 2010–2015 (Fig. 1A). RAD-seq provides finer resolution for evolutionary studies even with comparatively few samples compared to other methods such as *cp*DNA or SSRs (Cao et al., 2018). To reduce the sequencing costs, we selected 1–5 individuals from each of 19 population, for a total of 60 individuals (Table 1). Voucher specimens for all trees were deposited at the Herbarium of the Shanghai Chenshan Botanical Garden (CSH).

Total genomic DNA of *Q. championii* was isolated from silica-dried leaf tissue using a modified CTAB method (Doyle, 1987) or the DNeasy plant tissue kit (DNeasy, Qiagen, Valencia, CA, USA). RAD-seq library preparation using restriction enzyme *Pst*I (5′- CTGCA|G-3′) and sequenced by Personalbio Company (Shanghai). All 60 samples were sequenced on Illumina Hiseq 2500 with 2 × 150. During library constructions and sequencing, background error rates can affect the quality of SNP calling, including quality of DNA, PCR error and bias, sequencing error and restriction site variation (Mastretta-Yanes et al., 2015). To account for this background error rate and to optimize assembly parameters, five random replicate samples from different populations were sequenced ****(Tables 1 and S1). RADseq data of *Q. championii* were submitted to NCBI (SRA accession: SAMN09694015-SAMN09694075).

2.2. SNP calling

Analyses were carried out using Stacks v1.46 (Catchen et al., 2013) and SNP calling was conducted only on R1 reads. Raw reads were filtered for quality check and trimmed to 130 bp using the *process_radtags* module, as the quality of the reads were low after position 130 for some individuals. To test whether different *de novo* assembly parameters in Stacks affected the number and error rates of SNPs, five replicate individuals were used to run Stacks repeatedly with varying parameters. The four parameters with the greatest impact on the number and error rate of SNPs were tested using different values (Mastretta-Yanes et al., 2015): the minimum depth of coverage required for creating a stack (-m: 2–6, default: 2), the maximum distance (in nucleotides) allowed between stacks (-M: 2–12, default: 2), the number of mismatches allowed between sample loci when building the catalog (-n: 1–6, default: 1), and the maximum number of stacks at a single *de novo* locus (–max_locus_stacks: 2–7, default: 3). Only one parameter was changed at a time while other parameters were kept at their default values. Calculations of numbers of loci, numbers of missing loci, and SNPs error rates were performed in R v3.2.3 (Core Team, 2012). An optimal parameter set was developed based on the above calculations and used for SNP calling of all individuals in the study.

2.3. Genetic diversity and structure

Genetic diversity indices such as nucleotide diversity (π), expected heterozygosity (*He*), observed heterozygosity (H_o), and genetic differentiation (F_{st}) were estimated using the *populations* module on Stacks v1.46 (Catchen et al., 2013). Population genetic structure was estimated using two methods, Bayesian clustering and principal component analysis (PCA). Bayesian clustering was performed in STRUCTURE 2.3.4 (Pritchard et al., 2000) with the admixture model, and this analysis assumes that loci are unlinked. We exported SNPs from the *populations* module in Stacks with the *-write_single_snp* option. To determine the optimal number of groups (K), we ran STRUCTURE 10 times for each K value from 1 through 10. Each run was performed for 200,000 Markov Chain Monte Carlo (MCMC) generations with a burn-in period of 100,000 generations. The optimal *K* was determined using STRUCTURE HARVESTER (Earl and vonHoldt, 2012) using the delta-K method (Evanno et al., 2005). The PCA was calculated using *glPCA* function using the “adegenet” package (Jombart, 2008) in R v3.2.3 (Core Team, 2012).

Phylogenetic relationships of *Q. championii* were reconstructed using maximum likelihood (ML) and Bayesian inference (BI) methods. The closely related species *Quercus sichourensis* and *Q. kiukiangensis* (Deng et al., 2018) were chosen as outgroups to root the tree. The parallel version of MrBayes v3.2.6 (MPI) (Ronquist et al., 2012) was used to reconstruct a BI tree with the GTR + GAMMA model. Markov chains were run for 1,000,000 generations and sampled at each 100 generations. The first 40% of the samples were discarded as burn-in. The ML tree was reconstructed using the RAxML v8.2.4 with the GTR + GAMMA model (Stamatakis, 2014). 1000 non-parametric bootstrap replicates was used to estimate the reliability of the tree topology.

The relationship between geographic coordinates and genetic structure, which was defined as the first two components of the PCA, was calculated using the Procrustes analysis from the “vegan” package (Oksanen et al., 2007) in R. Procrustes rotation rotates a matrix and minimizes the sum of squared Euclidean distances to achieve a

Table 1
Sample information and genetic diversity of *Quercus championii*.

ID	Location	Longitude	Latitude	Sample	*He*	*Ho*	π
HN group				9/7	0.088	0.090	0.096
C1	Lingshui, Hainan	109.87	18.72	**3/3**	0.072	0.087	0.093
C2	Wuzhishan, Hainan	109.70	18.90	3/3	0.078	0.092	0.095
C3	Changjiang, Hainan	109.19	19.09	3/1	–	–	–
SW group				11/11	0.084	0.076	0.089
C4	Vĩnh Phúc, Vietnam	105.51	21.47	1/1	–	–	–
C5	Chongzuo, Guangxi	107.01	21.86	**5/5**	0.077	0.083	0.087
C6	Nanning, Guangxi	108.40	23.52	5/5	0.075	0.074	0.086
SE group				29/28	0.077	0.070	0.079
C7	Yangchun, Guangdong	111.44	21.88	3/3	0.060	0.071	0.075
C8	Zhuhai, Guangdong	113.72	21.95	**2/2**	0.058	0.077	0.078
C9	Conghua, Guangdong	113.53	23.96	3/2	0.044	0.070	0.076
C10	Chenzhou, Hunan	113.80	25.44	3/3	0.062	0.072	0.076
C11	Huizhou, Guangdong	113.86	23.64	3/3	0.056	0.062	0.070
C12	Huizhou, Guangdong	114.33	23.24	3/3	0.061	0.075	0.074
C13	Shenzheng, Guangdong	114.54	22.52	3/3	0.057	0.069	0.072
C14	Heyuan, Guangdong	115.14	23.36	**3/3**	0.063	0.074	0.077
C15	Jieyang, Guangdong	115.83	23.62	3/3	0.062	0.068	0.075
C16	Fuzhou, Fujian	119.39	26.06	3/3	0.055	0.067	0.068
TW group				11/10	0.066	0.067	0.070
C17	Pingdong, Taiwan	120.84	22.16	3/3	0.048	0.066	0.065
C18	Pingdong, Taiwan	120.79	22.23	**4/4**	0.058	0.066	0.067
C19	Pingdong, Taiwan	120.73	22.17	4/3	0.056	0.072	0.068

Note: In the column of sample, left and right of slash represents the sequencing and successful sequencing samples, respectively; bold populations represent one of samples in this population as repeat sequencing.

maximum similarity to a target matrix (Wang et al., 2012). The similarity of the two matrixes was quantified using the formula $t_0 = sqrt(1 - D)$, where D is the minimum sum of the squared Euclidean distances between the two matrixes. The "protest" function in vegan with 100,000 permutations was used to test the probability of observing a similarity statistic higher than the observed t_0 when no geographic pattern is assumed (Wang et al., 2012).

2.4. Divergence and demographic history

To estimate the demographic history of *Q. championii*, the Approximate Bayesian Computation (ABC) algorithm was implemented in DIYABC 2.1 (Cornuet et al., 2014). To reduce the analysis time and computational resources, a subset of the dataset containing 1105 SNPs with ≤3 missing individuals per SNP was used. The 19 populations of *Q. championii* were divided into four groups for the DIYABC analysis. The groups were based on species distribution, genetic structure, and phylogenetic relationships of *Q. championii*: the HN group consisted of three populations from Hainan Island; the SW group consisted of three populations from Southwest China-North Vietnam; the SE group consisted of 10 populations from Southeast China; and the TW group consisted of three populations from Taiwan. Six alternative demographic models were tested ***(Fig. S1). In models 1–3, group HN was set as the ancestral population, based on this group phylogenetic position in the ML and BI trees. In models 4–5, group SW was set as the ancestral population, as SW China is the center of diversity for *Quercus* section *Cyclobalanopsis* (Luo and Zhou, 2001). In model 6, group SE set as the ancestral population because the core distribution of *Q. championii* is currently in SE China. All models assumed that group TW was derived from group SE based on their similar genetic structure, geographic proximity, and phylogenetic relationship.

For all population size and divergent time parameters, the analysis employed a uniform prior distribution ranging from 10 to 100,000. A total of 10,000,000 simulated datasets were tested with an average 1,660,000 per scenario. The posterior probabilities and parameters for each scenario were estimated using logistic regression with the 100,000 simulated datasets (1% of the total data) that most closely matched the observed data. Type I and type II model error rates were calculated to evaluate the reliability of each scenario.

2.5. Landscape genomic patterns

To estimate the contributions of environmental variables to population genetic structure and to understand the turnover of allele frequencies along a climate gradient, a gradient forest (GF) analysis was performed using the R package "gradientForest" (Ellis et al., 2012). GF is an effective method for analyzing nonlinear associations between spatial, environmental, and allelic variables (Bay et al., 2018; Fitzpatrick and Keller, 2015; Gugger et al., 2018) and GF defines the amount of variation explained as 'split importance' values. Moving along the environmental gradient, the split importance values are added cumulatively to produce a step-like curve. Therefore, areas with many large steps in a row indicate significant allelic change. Present environmental data on temperature and precipitation (Fick and Hijmans, 2017) were downloaded from Worldclim (http://www.worldclim.org/). Paired Pearson correlation coefficients of 19 climate factors were calculated in R. Subsets of climate variables with higher correlations ($r \geq 0.9$) were reduced to single variables, and a total of 12 variables were used in the gradient forest analysis (GF).

The six environmental variables that the gradient forest analysis revealed as high contributors to genetic variation were used to investigate adaptation to local environments. The analyses were performed by latent factor mixed modeling (LFMM) in the R package "LEA" (Frichot and François, 2015). LFMM can efficiently estimate the random effects from population history and isolation by distance patterns, and it has a relatively low false detection rate compared to other methods (e.g. regression models) (Frichot et al., 2013). The optimal latent factor number of $K = 3$ was identical to the STRUCTURE result. The iterations and burn-in numbers were 100,000 and 30,000, respectively. The loci associated with adaptation were aligned to the *Quercus suber* reference genome using Blastn (Zhang et al., 2000).

3. Results

3.1. SNPS from RAD-seq analysis of Q. championii

We conducted RAD-seq analysis of *Q. championii* from 19 populations, and of the samples that were processed, four samples with less than 800,000 raw reads were excluded from further analysis (Table 1).

The average numbers of raw reads per sample was 5,019,878 (range: 933,001–29,148,566, median: 2,927,005; ***Table S1). After filtering out low quality reads, the average number of raw reads was 4,939,595 (range: 808,223–28,995,062, median: 2,913,129; ***Table S1).

We optimized the parameters for Stack analysis using five replicate samples. SNP error rates, loci error rates and number of missing loci were reduced when parameter *m* modified from 1 to 6 ***(Fig. S2), but changes in parameters *M* and max_locus_stacks (*MLS*) had limited influence ***(Fig. S2). SNP error rates increased when the parameter *n* was modified from 1 to 6, but this parameter change had limited impacts on loci error rates and number of missing loci ***(Fig. S2). Increasing *M* and *MLS* reduced and increased the number of assembled loci, respectively ***(Fig. S2). Moreover, when parameters *m* and *n* were increased, the number of loci increased at first but then decreased after the maximum number of loci was achieved at $n = 2$ or $m = 3$ ***(Fig. S2).

The optimal parameters to obtain the most SNPs at the lowest error rate were $m = 6$, $M = 2$, $MLS = 3$, and $N = 1$, and this setting was used to analyze all samples. The final data set contained 5207 SNPs from 2513 loci. The average SNP error rate was 0.00098 (range: 0–0.00123), the average loci error rate was 0.01461 (range: 0.00120–0.02275), and the average number of mismatches loci per sample was 12.2 (range: 1–19).

3.2. Genetic diversity and structure

At the population level, nucleotide diversity (π), expected heterozygosity (*He*) and observed heterozygosity (*Ho*) were 0.065–0.095, 0.044–0.078 and 0.062–0.092, respectively (Table 1). Populations from Wuzhishan, Hainan province, China (C02) had the highest genetic diversity indices (Table 1). Populations C09 and C17 had the highest genetic differentiation ($F_{st} = 0.258$), and populations C05 and C06 had the lowest genetic differentiation ($F_{st} = 0.085$, ***Table S2). At the group level, genetic diversity indices from high to low were: HN, SW, SE and TW (Table 1). Groups HN and TW had the highest genetic differentiation ($F_{st} = 0.116$), while groups SE and SW had lowest genetic differentiation ($F_{st} = 0.034$, ***Table S3).

We next conducted a Bayesian clustering analysis to analyze the genetic structure of *Q. championii*, and found that the best-fit number for clustering (highest Δ*K* value) was 3 ***(Fig. S3). Individuals from the HN group contained nearly equal portions of cluster A and cluster B ***(Fig. S4). Individuals from group SW were composed of clusters A, B and C, with cluster A being the dominant cluster at between 0.76 and 0.84 ***(Fig. S4). Individuals from group SE primarily had cluster A (probability > 0.93; ***Fig. S4). Individuals from group TW were composed of clusters A and C, but cluster A was dominant (between 0.84 and 0.87; ***Fig. S4). The first three axes of the PCA indicated that the species genetic structure was divided into four clusters corresponding to the geographical groups (Figs. 1A and ***Fig. S5). The first three principal components explained 10.16%, 6.78% and 4.78% of total genetic variation. The Procrustes analysis using the first two axes identified a significant similarity score ($t_0 = 0.765$, $p < 0.001$) between genetic structure and geographic distribution (Fig. 1A).

The topologies of the ML and BI trees of *Q. championii* were similar. Groups HN, SW and TW formed clades, but group TW was nested within group SE (Fig. 2). The phylogenetic tree revealed that group HN was sister to all other groups, while group SW was sister to all groups except HN (Fig. 2).

3.3. Divergence and demographic history

In DIYABC, *Q. championii* originated from Southwest China-Northern Vietnam, then dispersed to Hainan Island and Southeast China in turn (scenario 5) had the highest posterior probability of the logistic regression (0.993, 95% CI: 0.982–1.000), which was much higher than scenario 1 (0.001, 95% CI: 0.000–1.000), scenario 2 (0.000, 95% CI: 0.000–1.000), scenario 3 (0.000, 95% CI: 0.000–1.000), scenario 4 (0.000, 95% CI: 0.000–1.000), and scenario 6 (0.005, 95% CI: 0.000–1.000). Scenario 5 showed that the median effective population sizes were 72,000 (95% CI: 35,700–96,800) for SW, 9230 (95% CI: 2270–52,200) for TW, 65,400 (95% CI: 45,400–75,800) for HN and 93,500 (95% CI: 69,300–99,500) for SE. The median values for the generation of divergence between TW and SE (*t*1) was 4550 (95% CI: 2170–7700), between groups SE and SW (*t*2) was 4840 (95% CI: 2630–9190), and between groups HN and SW (*t*3) was 7260 (95% CI: 2840–34,100). The most commonly used generation time for oaks is 50 years (Bagnoli et al., 2016; Chen et al., 2012; Xu et al., 2015), although some scholars have used 120 years as the generation time for evergreen oaks in the USA and Mexico (Cavender-Bares et al., 2011). We adopted 50 years as the generation time, the estimated divergence time for *t*1, *t*2 and *t*3 was 227,500, 242,000 and 363,000 years ago, respectively. When generation time was set to 120 years, the divergence time for *t*1, *t*2 and *t*3 was 546,000, 580,800 and 871,200 years ago, respectively. Fourteen of 24 summary statistics (F_{ST} values in four group population pairs) showed significant differences between the observed and simulated data based on the posterior distributions ***(Table S4). Type I and type II error rates from the logistic regression were 0.290 and 0.047, respectively.

3.4. Landscape genomics patterns

GF analyses indicated that isothermality was the most important predictor of species allele frequencies (R^2 weighted importance: 0.037) among the 12 environmental variables that were tested. Annual precipitation (R^2 weighted importance: 0.030), precipitation of the coldest quarter (R^2 weighted importance: 0.028) and temperature seasonality (R^2 weighted importance: 0.027) were also of high importance. Maximum temperature of the warmest month (R^2 weighted importance: 0.023) and precipitation seasonality (R^2 weighted importance: 0.022) showed moderate importance on species allele frequencies. The other six variables had little contribution (Fig. 3A). Of the parameters that were analyzed, allelic composition changes were sharp when isothermality was between 40 and 42, annual precipitation between 2000 and 2500 mm, and precipitation of the coldest quarter 80–120 mm (Fig. 3B).

LFMM analysis found significant associations between 48 SNPs and precipitation seasonality (bio15), 30 SNPs and annual precipitation (bio12), and 17 SNPs with temperature seasonality (bio4) ***(Fig. S6). The 48, 30 and 17 SNPs were from 33, 18 and 9 loci, respectively. Five loci were associated with all three variables; six loci were associated with variables bio15 and bio12; and three loci were associated with variables bio15 and bio4 (Fig. 4). In a total, 39 loci were associated with environment variables. Twenty-five of 39 loci were successfully aligned to the *Q. suber* genome using blastn with cut-off of e-value < 1E−40. Among these were genes related to the regulation of leaf functions (Table 2), such as serine—glyoxylate aminotransferase, with converts glyoxylic acid produced by oxidation in peroxisomes to glycine (Somerville and Ogren, 1980). These loci also included genes related to the development of reproductive organs with adaption to precipitation seasonality; for example, MADS-box protein FBP24 (De Folter et al., 2006) and tubulin-folding cofactor *D* (Kirik et al., 2002).

4. Discussion

4.1. Genetic structure and introgression of Q. championii

We identified four distinct regional groups of *Q. championii* based on phylogenetic and PCA analyses (Figs. 1A, 2, ***S4 and S5): SE China, Taiwan Island, Hainan Island, and SW China-North Vietnam. The Bayesian clustering analysis also showed different genetic structure among these four groups, although the best cluster number was three ***(Fig. S4). These genetic groups are consistent with the geographical

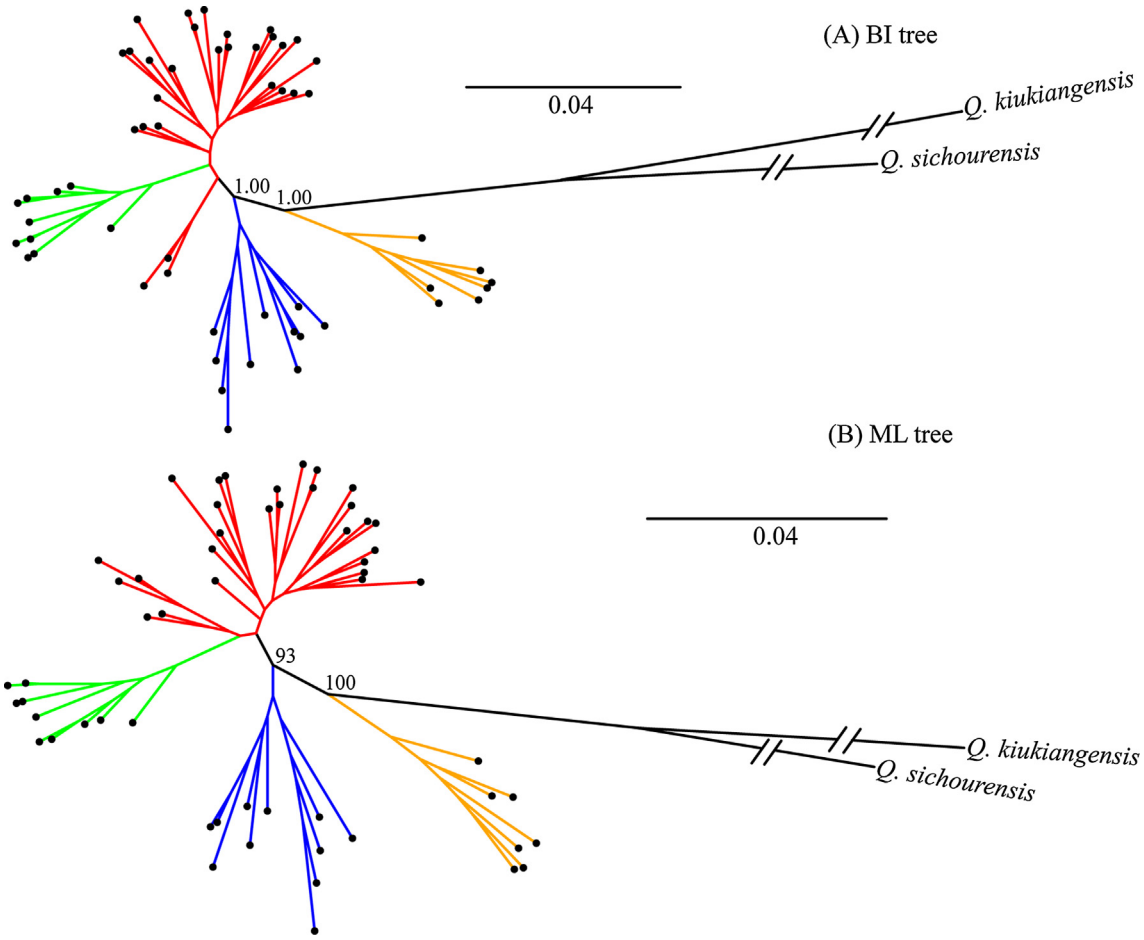

Fig. 2. Bayesian inference (BI; A) and Maximum likelihood (ML; B) phylogenetic trees for 56 *Quercus championii* individuals and two outgroup species (*Q. kiukiangensis* and *Q. sichourensis*). Branches are color-coded to indicate different groups of *Q. championii* with yellow (HN), blue (SW), green (TW) and red (SE). Bootstrap values or posterior probability of the derived of main clades were mark in the node. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

distributions of the species, which indicates that *Q. championii* populations have strong geographic structure, which is limited by the lack of long-distance dispersal and gene flow between the groups. Acorns are dispersed by gravity and rodents, which result in a limited dispersed range (Pesendorfer et al., 2018). Although dispersal of acorns by corvids has been reported (Pesendorfer et al., 2017), they are not efficient at long distance dispersal and can only move up to 750 m (Pesendorfer et al., 2016; Sork, 1984). While wind-dispersed oak pollen may facilitate gene flow through long distance dispersal between highly fragmented habitats. The complex topography and high environment heterogeneity of Southern China combined with long distance isolation among the groups may block gene flow among these groups even if there was pollen dispersal among these populations. Furthermore, a high similarity score ($t_0 = 0.765$, $p < 0.001$; Fig. 1A) between the rotated genetic space and geographic distributions that we found in our analysis also demonstrates that *Q. championii* has strong geographic structure.

The highest genetic admixture and genetic diversity were detected in group HN, which was derived from the SW group. Introgression or contact between genetically differentiated populations can increase genetic diversity and genetic admixture (Ortego et al., 2015; Rieseberg and Wendel, 1993). If genetic diversity increases by admixture of two genetically differentiated lineages, the mixed population will have a similar genetic structure to the source lineage with lower genetic divergence (Petit et al., 2003). Group HN showed significant differentiation from the others three groups (Fig. 1A and ***Table S3), and individuals in the Hainan group showed admixture of clusters A and B in nearly equal portions ***(Fig. S4). Cluster A was the dominant cluster in all other groups, and cluster B mainly existed in group HN. These results suggest that Hainan Island population may have experienced introgression, which led to the increased genetic diversity and unique gene pool of the HN group. Hybridization and introgression between close related species and sympatric oaks is a common phenomenon (Burgarella et al., 2009; Eaton et al., 2015; Ortego and Bonal, 2010). Moreover, interspecific hybridization can occur easily on islands, where natural and anthropogenic disturbances can bring congeneric species into contact in a dynamic habitat confined to a small physical space (Crawford and Archibald, 2017; Kerbs et al., 2017). There are about 14 sympatric species in section *Cyclobalanopsis* on Hainan Island, including *Q. austrocochinchinensis*, *Q. kerrii*, *Q. edithae*, *Q. patelliformis*, *Q. neglecta*, *Q. hui*, *Q. litoralis*, *Q. blakei*, and *Q. tiaoloushanica* (Chun, 1965). Introgression alleles from one or more of these species may be the source of the HN group's genetic differentiation from the other *Q. championii* populations. Further population genetic studies on sympatric oaks on Hainan Island can provide insights into the contribution of interspecies gene flow to the genetic diversity of the HN group of *Q. championii*.

4.2. Origin and colonization route

Approximate Bayesian Computation inferences in DIYABC indicated that the ancestral range of *Q. championii* was Southwest China-Northern Vietnam, which is in accordance to the recent biogeographic analyses on section *Cyclobalanopsis* that found that these regions were involved

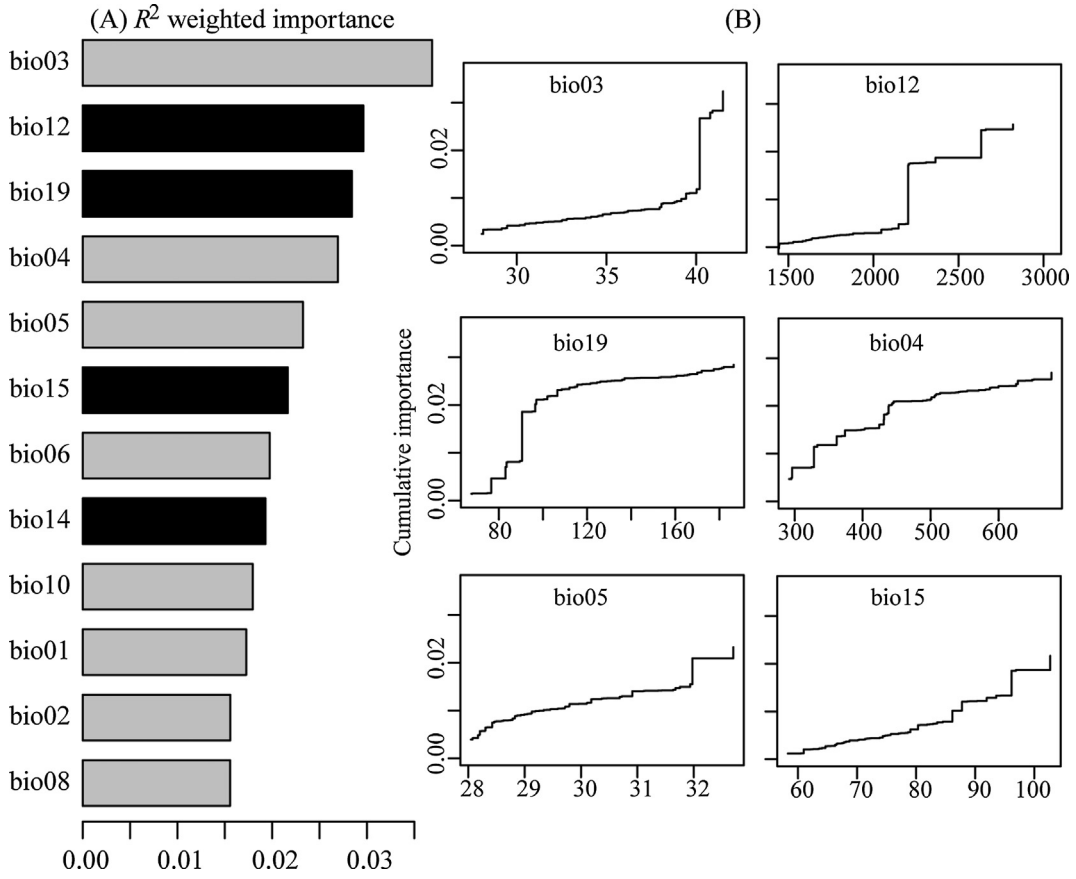

Fig. 3. (A) R^2-weighted importance of environmental variables that explain genetic gradients from gradient forest analysis. (B) Cumulative importance of allelic change along the first six environmental gradients.

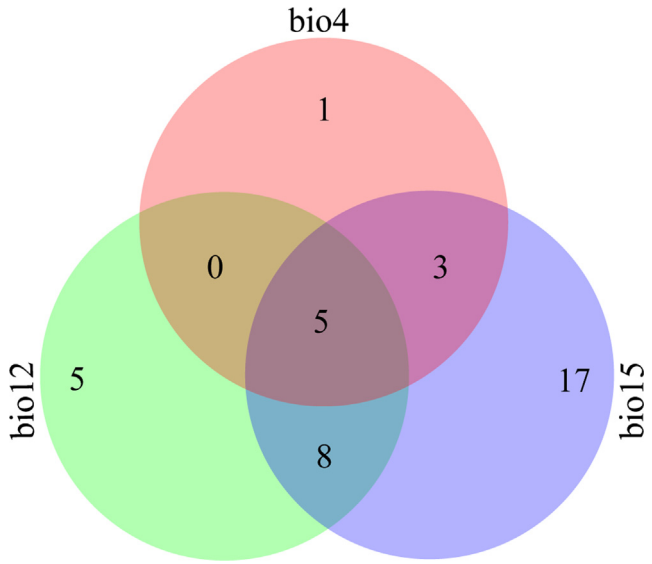

Fig. 4. Venn diagram showing the distribution of the number of adaptive loci among the three climate variables.

in the origin and early divergence of the section (Deng et al., 2018), especially the CTB lineage. Most species in the CTB lineages are found in Indo-China and Southwest China (e.g. *Q. sichourensis, Q. delavayi, Q. kerrii, Q. rex* and *Q. austrocochinchiensis*), and older fossils of this lineage were found in these regions as well (Writing Group of Cenozoic Plants of China, 1978; Xu et al., 2016). Regardless, the main contemporary distribution area of *Q. championii* is Southeast China, with populations extending to Southwest China and North Vietnam. Moreover, *Q. prechampionii*, a fossil taxa with close affinities to *Q. championii*, was found in the Oligocene from Jinggu, West Yunnan (Writing Group of Cenozoic Plants of China, 1978) and in the Pliocene from Shuifu, Guizhou (Zhang, 1978). *Quercus championii* therefore has a long evolutionary history in Southwest China and was likely widespread in Southwest China by at least the late Neogene. Likewise, a recent study on flora assembly of East Asia inferred similar evolutionary dynamics for the regional flora as a whole, finding that Southwest China was the "cradle" of East Asian flora (Lu et al., 2018). Since the late Miocene, Southwest China has undergone complex tectonic activities and climate changes, including the intensified Asian winter monsoons, the rapid uplift of mountains, and glacial/interglacial cycles (An et al., 2001; An et al., 2011; Miao et al., 2012), all of which impacted species dispersal, extinction, speciation and local adaptation (Du et al., 2017; Jiang et al., 2016; Meng et al., 2017; Xing and Ree, 2017). These geological and climatic events may have wiped out habitats of *Q. championii* in Southwest China, resulting in the patchy distribution we observe today.

Hainan Island and the adjacent mainland share a similar flora, and the distribution of species on Hainan Island likely arose through vicariance at the time the island formed (Zhu, 2016, 2017). Hainan Island has a continental origin and a long geological history. It once belonged to the South China block, and became an island by southeastern drift from the continent shelf during the early Cenozoic (Li et al., 1995; Liu and Morinaga, 1999). Eocene fossil sites from the Changchang formation in Hainan Island contains high numbers of Lauraceae-Fagaceae elements, indicating that the island was a cool environment at this time (Spicer et al., 2014). Although long-term geographical isolation is a major driver of speciation and local adaptation of organisms (Dionne et al., 2008; Endler, 1977), the dispersal times of *Q. championii* to Hainan and Taiwan islands inferred in our study were much later than

Table 2
List of 25 candidate loci under selection that the E-value of blastn is less than 1E−40.

Loci ID	Adapted climate	Matching accession	Sequence description	Blastn E-value
453	bio4, bio12, bio15	XM_024059323.1	DNA polymerase alpha catalytic subunit	2.00E−58
2781	bio4, bio12, bio15	XM_024036455.1	sodium channel modifier 1-like	9.18E−52
12003	bio15	XM_024059860.1	TATA-binding protein-associated factor 2N-like	9.11E−57
14505	bio4, bio15	XM_024038282.1	topless-related protein 2	3.28E−56
15758	bio15	XM_024047427.1	probable receptor-like protein kinase At1g49730	1.97E−53
22366	bio12	XM_024044898.1	pentatricopeptide repeat-containing protein At3g61360	4.21E−60
23803	bio12, bio15	XM_024031182.1	uncharacterized	3.28E−56
29523	bio12, bio15	XM_024028758.1	TORTIFOLIA1-like protein 2	1.18E−55
34800	bio12	XM_024067422.1	soluble starch synthase 1, chloroplastic/amyloplastic	1.53E−54
47868	bio15	XM_024056073.1	chitotriosidase-1-like	1.96E−58
52906	bio15	XM_024058235.1	tubulin-folding cofactor D	1.18E−55
53245	bio15	XM_024050581.1	protein NETWORKED 2D	9.11E−57
55524	bio15	XM_024043332.1	zinc transporter 2-like	3.30E−51
56222	bio12, bio15	XM_024019693.1	protein NRT1/ PTR FAMILY 3.1-like	4.21E−60
56903	bio15	XM_024016873.1	uncharacterized	1.96E−58
57968	bio12, bio15	XM_024016426.1	protein SMAX1-LIKE 3-like	1.97E−53
63939	bio15	XM_024064131.1	MADS-box protein FBP24	1.96E−58
69231	bio4, bio12, bio15	XM_024019490.1	uncharacterized	3.28E−56
71080	bio12, bio15	XM_024066236.1	increased DNA methylation 1-like	1.97E−53
71961	bio15	XR_002884923.1	26S proteasome regulatory subunit 6A homolog	5.45E−59
72916	bio15	XM_024040642.1	uncharacterized WD repeat-containing protein sll0163-like	1.96E−58
74122	bio12, bio15	XM_024040564.1	AP2/ERF and B3 domain-containing transcription factor At1g50680-like	4.21E−60
91468	bio12	XM_024030576.1	cytochrome P450 71A1-like	1.97E−53
92414	bio15	XM_024029951.1	serine-glyoxylate aminotransferase	7.05E−58
93283	bio15	XM_024042790.1	serine carboxypeptidase-like 18	2.55E−52

the geographic split between the islands and the mainland. This result was consistent when we used 50 and 120 years as the generation time of *Q. championii*. In addition, the East Asian flora is quite young; 66% of the angiosperm genera in this region originated around the Miocene (Chen et al., 2017; Lu et al., 2018). Likewise, the divergence of the main lineages in *Quercus* section *Cyclobalanopsis* occurred during the later Neogene (Deng et al., 2018). Furthermore, Hainan's flora has a strong affinity to tropical Asian species and very low endemism at the genus and species levels, implying a continental origin (Zhu, 2016). Multiple lines of evidences indicate that the flora of Hainan is "young" and was mainly established by colonization/dispersal from mainland Asia after the Miocene rather than by ancient vicariance.

Glaciation and interglaciation events were accompanied by rising and falling sea levels during the late Pleistocene. These events had a profound impact on land-sea changes in Southeast Asia. The estimated dispersal time of *Q. champioii* from Soutwest China-Northern Vietnam to Taiwan and Hainan Islands coincided with the glacial periods with low sea levels (Miller et al., 2011). Nevertheless, colonization from Southwest China to Southeast China occurred during a post-glacial period with higher sea levels (interglacial periods; Fig. 1B). Whether the generation time was set to 50 or 120 years, sea level was about 85 m lower than today when *Q. championii* dispersed from Southwest China-Northern Vietnam to Hainan Island. Likewise, when *Q. championii* dispersed to Taiwan Island from Southeast China, sea level was between 64 m (generation time: 50 years 50) and 85 m (generation time: 120 years) lower than today. A land bridge could have connected Hainan and Taiwan Islands with SCM at a time when the sea level was at least 80 m lower than today (Fig. 1A). Because we could not confidently determine the optimal generation time for *Q. championii*, the precise dispersal times estimated here should be treated with caution. However, the inferred dispersal time of *Q. championii* from the mainland to the islands coincides well to the period of sea level was low and a land bridge may have existed, regardless of whether 50 or 120 years as the generation time. Our results therefore suggest that land bridges during glacial periods may have allowed *Q. championii* to colonize the islands. Several recent biogeographical studies have focused on species with disjunct distributions in SCM and the adjacent islands (Chatterjee, 2006; Chen et al., 2015; Lin et al., 2014; Qu et al., 2015), but only very few have estimated dispersal time between the islands and mainland (Chatterjee, 2006). Migration of gibbons that were adapted to (sub) tropical forests, from eastern Indo-China to Hainan island was dated to *ca.* 1.8–0.3 Ma (Chatterjee, 2006), and this timescale partly coincides with the estimated dispersal of *Q. championii* to Hainan Island. Migration corridors over land bridges with canopy trees such *Q. championii* may have supported the (sub)tropical forest stretching between SCM and the continental islands, enabling biotic exchange between the mainland and islands.

4.3. Adaptation to local environments

In addition to geographic isolation, environmental variables also likely contributed to the genetic patterns observed in *Q. championii*. In our analysis, we found that isothermality was most strongly associated with genetic variation in the *Q. championii* populations. Sharp changes in genetic composition occurred when isothermality was ca. 40, which is close to the maximum range of the species distribution. Only populations in Hainan and Taiwan islands had the isothermality value close to the threshold. *Quercus championii* on Hainan and Taiwan islands are found in montane tropical rainforests. Stable temperatures in tropical montane regions could have favored the long-term survival of *Q. championii*, maintaining higher allele frequencies on the islands.

LFMM analysis indicated that precipitation seasonality along an environment gradient contributed greatly to the local adaptation of *Q. championii*. Leaf traits and functions, such as leaf shape, size and specific leaf area, are strongly influenced by genetic and environmental factors (Dudley, 1996; Ramírez-Valiente et al., 2010; Tsukaya, 2004). Our field survey suggested that leaf traits might vary with differences in water supply. Individuals bearing thick leathery leaves with obvious recurved margins (Fig. 1C) are usually found in open windy sites with seasonal dry periods of low humidity, such as mountainsides and mountaintop. By contrast, populations growing in habitats with high humidity all year around, like ravines or streams, have thin leathery leaves with flat or only slightly recurved margins (Fig. 1C). We detected a series of putative adaptation loci related to physiological trait regulation of leaves that might be under selection and involved in adaptation to precipitation seasonality. These genes include Serine-glyoxylate aminotransferase, a crucial enzyme in the photorespiratory pathway (Somerville and Ogren, 1980) that converts glyoxylic acid

produced by oxidation in peroxisomes to glycine. We also identified adaptive loci with genes involved in the development of reproductive organs, with may be related to adaptation to precipitation seasonality. This included MADS-box protein FBP24, a gene involved in ovule and seed development (De Folter et al., 2006), and tubulin-folding cofactor *D*, which is required for development of embryo and endosperm tissue (Kirik et al., 2002). These results suggest that *Q. championii* responded to climate change by altering vegetative and reproductive organs. Further work, in particular a common garden experiment combined with transcriptomic and DNA methylation data could further illuminate these issues by quantifying the relative roles of phenotypic plasticity and local adaptation in the response of *Q. championii* to environmental heterogeneity.

In summary, our analyses demonstrated that sea-level changes caused by the Pleistocene glacial-interglacial climate oscillation impacted the (sub)tropical biotic exchange between SCM and adjacent islands. The dispersal of *Q. championii* provides new insight into the mechanism of flora assembly on the continental islands of SCM. Geographical isolation and environmental heterogeneity contributed to spatial genetic patterns observed in *Q. championii* populations, and the landscape genomic analyses identified potential adaptive mechanisms in evergreen broad-leaf oaks to environmental heterogeneity in (sub) tropical Asia.

Acknowledgments

We would like to thank Guan-Peng Ren from University of Lausanne and Miguel Navascués from University of Montpellier for their great help on analyzing the data. We are grateful to Yi-Gang Song and Yan-Shi Xiong from Shanghai Chenshan Botanical Garden for their help on sampling. This work was supported by grants from the National Natural Science Foundation of China (31700174), the Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences (Y4ZK111B01) and Shanghai Municipal Administration of Forestation and City Appearances (G182427, G172406, G162404 and G162405). H.-H. MENG was supported by Youth Innovation Promotion Association, CAS (2018432).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2018.11.021.

References

An, Z.S., Kutzbach, J.E., Prell, W.L., Porter, S.C., 2001. Evolution of Asian monsoons and phased uplift of the Himalaya-Tibetan plateau since Late Miocene times. Nature 411, 62.

An, Z.S., Clemens, S.C., Shen, J., Qiang, X., Jin, Z., Sun, Y., Prell, W.L., Luo, J., Wang, S., Xu, H., Cai, Y., Zhou, W., Liu, X., Liu, W., Shi, Z., Yan, L., Xiao, X., Chang, H., Wu, F., Ai, L., Lu, F., 2011. Glacial-interglacial Indian summer monsoon dynamics. Science 333, 719–723.

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Natl. Sci. Rev. 17, 81–92.

Bagnoli, F., Tsuda, Y., Fineschi, S., Bruschi, P., Magri, D., Zhelev, P., Paule, L., Simeone, M.C., González-Martínez, S.C., Vendramin, G.G., 2016. Combining molecular and fossil data to infer demographic history of *Quercus cerris*: insights on European eastern glacial refugia. J. Biogeogr. 43, 679–690.

Bai, W.N., Liao, W.J., Zhang, D.Y., 2010. Nuclear and chloroplast DNA phylogeography reveal two refuge areas with asymmetrical gene flow in a temperate walnut tree from East Asia. New Phytol. 188, 892–901.

Bay, R.A., Harrigan, R.J., Le Underwood, V., Gibbs, H.L., Smith, T.B., Ruegg, K., 2018. Genomic signals of selection predict climate-driven population declines in a migratory bird. Science 359, 83–86.

Bellard, C., Leclerc, C., Courchamp, F., 2014. Impact of sea level rise on the 10 insular biodiversity hotspots. Global Ecol. Biogeogr. 23, 203–212.

Burgarella, C., Lorenzo, Z., Jabbour-Zahab, R., Lumaret, R., Guichoux, E., Petit, R.J., Soto, A., Gil, L., 2009. Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*). Heredity 102, 442–452.

Cao, Y.N., Wang, I.J., Chen, L.Y., Ding, Y.Q., Liu, L.X., Qiu, Y.X., 2018. Inferring spatial patterns and drivers of population divergence of *Neolitsea sericea* (Lauraceae), based on molecular phylogeography and landscape genomics. Mol. Phylogenet. Evol. 126, 162–172.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22, 3124–3140.

Cavender-Bares, J., González-Rodríguez, A., Eaton, D.A., Hipp, A., Beulke, A., Manos, P.S., 2015. Phylogeny and biogeography of the American live oaks (*Quercus* subsection *Virentes*): a genomic and population genetics approach. Mol. Ecol. 24, 3668–3687.

Cavender-Bares, J., Gonzalez-Rodriguez, A., Pahlich, A., Koehler, K., Deacon, N., 2011. Phylogeography and climatic niche evolution in live oaks (*Quercus* series *Virentes*) from the tropics to the temperate zone. J. Biogeogr. 38, 962–981.

Chatterjee, H.J., 2006. Phylogeny and biogeography of gibbons: a dispersal-vicariance analysis. Int. J. Primatol. 27, 699–712.

Chen, D., Chang, J., Li, S.H., Liu, Y., Liang, W., Zhou, F., Yao, C.T., Zhang, Z.W., 2015. Was the exposed continental shelf a long-distance colonization route in the ice age? The Southeast Asia origin of Hainan and Taiwan partridges. Mol. Phylogenet. Evol. 83, 167–173.

Chen, D., Zhang, X., Kang, H., Sun, X., Yin, S., Du, H., Yamanaka, N., Gapare, W., Wu, H.X., Liu, C., 2012. Phylogeography of *Quercus variabilis* based on chloroplast DNA sequence in East Asia: multiple glacial refugia and mainland-migrated island populations. PloS one 7, e47268.

Chen, Y.S., Deng, T., Zhou, Z., Sun, H., 2017. Is the East Asian flora ancient or not? Natl. Sci. Rev. https://doi.org/10.1093/nsr/nwx156.

Chun, W., 1965. Flora Hainanica (Vol. II). Science Press, Beijing.

R Core Team, 2012. R: A Language and Environment for Statistical Computing. Austria, Vienna.

Cornuet, J.M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J., Estoup, A., 2014. DIYABC v2. 0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism. DNA sequence and microsatellite data. Bioinformatics 30, 1187–1189.

Crawford, D.J., Archibald, J.K., 2017. Island floras as model systems for studies of plant speciation: prospects and challenges. J. Syst. Evol. 55, 1–15.

Crowl, A.A., Visger, C.J., Mansion, G., Hand, R., Wu, H., Kamari, G., Phitos, D., Cellinese, N., 2016. Evolution and biogeography of the endemic roucela complex (campanulaceae: *campanula*) in the eastern mediterranean. Ecol. Evol. 22, 5329–5343.

Darwin, C., 1839. Journal and Remarks: 1832-1836 (Vol. 3). Henry Colburn, London.

De Folter, S., Shchennikova, A.V., Franken, J., Busscher, M., Baskar, R., Crossniklaus, U., Angenet, G., Immink, R., 2006. A Bsister MADS-box gene involved in ovule and seed development in *petunia* and *Arabidopsis*. Plant J. 47, 934–946.

De Queiroz, A., 2005. The resurrection of oceanic dispersal in historical biogeography. Trends Ecol. Evol. 20, 68–73.

Deng, M., Jiang, X.L., Hipp, A.L., Manos, P.S., Hahn, M., 2018. Phylogeny and biogeography of East Asian evergreen oaks (*Quercus* section *Cyclobalanopsis*; Fagaceae): insights into the Cenozoic history of evergreen broad-leaved forests in subtropical Asia. Mol. Phylogenet. Evol. 119, 170–181.

Dionne, M., Caron, F., Dodson, J.J., Bernatchez, L., 2008. Landscape genetics and hierarchical genetic structure in Atlantic salmon: the interaction of gene flow and local adaptation. Mol. Ecol. 17, 2382–2396.

Doyle, J.J., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19, 11–15.

Du, F.K., Hou, M., Wang, W., Mao, K., Hampe, A., 2017. Phylogeography of *Quercus aquifolioides* provides novel insights into the Neogene history of a major global hotspot of plant diversity in south-west China. J. Biogeogr. 44, 294–307.

Dudley, S.A., 1996. The response to differing selection on plant physiological traits: evidence for local adaptation. Evolution 50, 103–110.

Earl, D.A., vonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4, 359–361.

Eaton, D.A., Hipp, A.L., González-Rodríguez, A., Cavender-Bares, J., 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. Evolution 69, 2587–2601.

Ellis, N., Smith, S.J., Pitcher, C.R., 2012. Gradient forests: calculating importance gradients on physical predictors. Ecology 93, 156–168.

Endler, J.A., 1977. Geographic Variation, Speciation, and Clines. Princeton University Press, Princeton.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 14, 2611–2620.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302–4315.

Fitzpatrick, M.C., Keller, S.R., 2015. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. Ecol. Lett. 18, 1–16.

Frichot, E., François, O., 2015. LEA: an R package for landscape and ecological association studies. Methods Ecol. Evol. 6, 925–929.

Frichot, E., Schoville, S.D., Bouchard, G., François, O., 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. Mol. Biol. Evol. 30, 1687–1699.

Gugger, P.F., Liang, C.T., Sork, V.L., Hodgskiss, P., Wright, J.W., 2018. Applying landscape genomic tools to forest management and restoration of Hawaiian koa *(Acacia koa*) in a changing environment. Evol. Appl. 11, 231–242.

Heaney, L.R., 1985. Zoogeographic evidence for middle and late Pleistocene land bridges to the Philippine Islands. Modern Quatern. Res. Southeast Asia 9, 127–144.

Hipp, A.L., Eaton, D.A., Cavender-Bares, J., Fitzek, E., Nipper, R., Manos, P.S., 2014. A framework phylogeny of the American oak clade based on sequenced RAD data. PloS one 9, e93975.

Hipp, A.L., Manos, P.S., González-Rodríguez, A., Hahn, M., Kaproth, M., McVay, J.,

Avalos, S.v., Cavender-Bares, J., 2017. Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. New Phytol. 217, 439–452.
Jiang, X.L., Deng, M., Li, Y., 2016. Evolutionary history of subtropical evergreen broad-leaved forest in Yunnan Plateau and adjacent areas: an insight from *Quercus schottkyana* (Fagaceae). Tree Genet. Genomes 12, 104.
Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405.
Kerbs, B., Ressler, J., Kelly, J.K., Mort, M.E., Santos-Guerra, A., Gibson, M.J.S., Caujapé-Castells, J., Crawford, D.J., 2017. The potential role of hybridization in diversification and speciation in an insular plant lineage: insights from synthetic interspecific hybrids. AoB Plants 9, plx043.
Kirik, V., Grini, P.E., Mathur, J., Klinkhammer, I., Adler, K., Bechtold, N., Herzog, M., Bonneville, J., Hülskamp, M., 2002. The Arabidopsis TUBULIN-FOLDING COFACTOR A gene is involved in the control of the α/β-tubulin monomer balance. Plant Cell 14, 2265–2276.
Li, Z.X., Metcalfe, I., Wang, X., 1995. Vertical-axis block rotations in southwestern China since the Cretaceous: new Paleomagnetic results from Hainan Island. Geophys. Res. Lett. 22, 3071–3074.
Lin, A.Q., Csorba, G., Li, L.F., Jiang, T.L., Lu, G.J., Thong, V.D., Soisook, P., Sun, K.P., Feng, J., 2014. Phylogeography of *Hipposideros armiger* (Chiroptera: Hipposideridae) in the Oriental Region: the contribution of multiple Pleistocene glacial refugia and intrinsic factors to contemporary population genetic structure. J. Biogeogr. 41, 317–327.
Liu, Y.Y., Morinaga, H., 1999. Cretaceous palaeomagnetic results from Hainan Island in south China supporting the extrusion model of Southeast Asia. Tectonophysics 301, 133–144.
Lu, L.M., Mao, L.F., Yang, T., Ye, J.F., Liu, B., Li, H.L., Sun, M., Miller, J.T., Mathews, S., Hu, H.H., Niu, Y.T., Peng, D.X., Chen, Y.H., Smith, S.A., Chen, M., Xiang, K.L., Le, C.T., Dang, V.C., Lu, A.M., Soltis, P.S., Soltis, D.E., Li, J.H., Chen, Z.D., 2018. Evolutionary history of the angiosperm flora of China. Nature 554, 234.
Luo, Y., Zhou, Z., 2001. Phytogeography of *Quercus* subg *Cyclobalanopsis*. Acta Botanica Yunnanica 23, 1–16.
Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T.H., Piñero, D., Emerson, B.C., 2015. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Mol. Ecol. Resour. 15, 28–41.
McVay, J.D., Hipp, A.L., Manos, P.S., 2017. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. P. Roy. Soc. Lond. B. Bio. 284, 20170300.
Meng, H.H., Su, T., Gao, X.Y., Li, J., Jiang, X.L., Sun, H., Zhou, Z.K., 2017. Warm–cold colonization: response of oaks to uplift of the Himalaya-Hengduan Mountains. Mol. Ecol. 26, 3276–3294.
Miao, Y., Herrmann, M., Wu, F., Yan, X., Yang, S., 2012. What controlled Mid-Late Miocene long-term aridification in Central Asia?—Global cooling or Tibetan Plateau uplift: a review. Earth-Sci. Rev. 112, 155–172.
Miller, K.G., MouNtaiN, G.S., WriGht, J.D., BroWNiNG, J.V., 2011. A 180-million-year record of sea level and ice volume variations from continental margin and deep-sea isotopic records. Oceanography 24, 40–53.
Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853–858.
Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G.L., Solymos, P., Stevens, H.H. H., Wagner, H., 2007. The vegan package. R package version 2.5-2, URL < http://CRAN.R-project.org/ > .
Ortego, J., Bonal, R., 2010. Natural hybridisation between kermes (*Quercus coccifera* L.) and holm oaks (*Q. ilex* L.) revealed by microsatellite markers. Plant Biology 12, 234–238.
Ortego, J., Gugger, P.F., Sork, V.L., 2015. Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak. J. Biogeogr. 42, 328–338.
Pesendorfer, M.B., Baker, C.M., Stringer, M., McDonald-Madden, E., Bode, M., McEachern, A.K., Morrison, S.A., Sillett, T.S., 2018. Oak habitat recovery on California's largest islands: Scenarios for the role of corvid seed dispersal. J. Appl. Ecol. 55, 1185–1194.
Pesendorfer, M.B., Sillett, T.S., Koenig, W.D., Morrison, S.A., 2016. Scatter-hoarding corvids as seed dispersers for oaks and pines: a review of a widely distributed mutualism and its utility to habitat restoration. The Condor 118, 215–237.
Pesendorfer, M.B., Sillett, T.S., Morrison, S.A., 2017. Spatially biased dispersal of acorns by a scatter-hoarding corvid may accelerate passive restoration of oak habitat on California's largest island. Curr. Zool. 63, 363–367.
Petit, R.J., Aguinagalde, I., de Beaulieu, J.L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Müller-Starck, G., Demesure-Musch, B., Palmé, A., Martín, J.P., Rendell, S., Vendramin, G.G., 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300, 1563–1565.
Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.
Qiu, Y.X., Qi, X.S., Jin, X.F., Tao, X.Y., Fu, C.X., Naiki, A., Comes, H.P., 2009a. Population genetic structure, phylogeography, and demographic history of *Platycrater arguta* (Hydrangeaceae) endemic to East China and South Japan, inferred from chloroplast DNA sequence variation. Taxon 1226–1241.
Qiu, Y.X., Sun, Y., Zhang, X.P., Lee, J., Fu, C.X., Comes, H.P., 2009b. Molecular phylogeography of East Asian *Kirengeshoma* (Hydrangeaceae) in relation to Quaternary climate change and landbridge configurations. New Phytol. 183, 480–495.
Qu, Y., Song, G., Gao, B., Quan, Q., Ericson, P.G., Lei, F., 2015. The influence of geological events on the endemism of East Asian birds studied through comparative phylogeography. J. Biogeogr. 42, 179–192.
Ramírez-Valiente, J.A., Sánchez-Gómez, D., Aranda, I., Valladares, F., 2010. Phenotypic plasticity and local adaptation in leaf ecophysiological traits of 13 contrasting cork oak populations under different water availabilities. Tree Physiol. 30, 618–627.
Rieseberg, L.H., Wendel, J.F., 1993. Introgression and its consequences in plants. In: Harrison, R. (Ed.), Hybrid Zones and the Evolutionary Process. Oxford University Press, New York.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biol. 61, 539–542.
Sakaguchi, S., Qiu, Y.X., Liu, Y.H., Qi, X.S., Kim, S.H., Han, J., Takeuchi, Y., Worth, J., Yamasaki, M., Sakurai, Shogo, Isagi, Y., 2012. 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.
Somerville, C., Ogren, W., 1980. Photorespiration mutants of *Arabidopsis thaliana* deficient in serine-glyoxylate aminotransferase activity. Proc. Natl. Acad. Sci. USA 77, 2684–2687.
Sork, V.L., 1984. Examination of seed dispersal and survival in red oak, *Quercus rubra* (Fagaceae), using metal-tagged acorns. Ecology 1020–1022.
Spicer, R.A., Herman, A.B., Liao, W., Spicer, T.E., Kodrul, T.M., Yang, J., Jin, J., 2014. Cool tropics in the Middle Eocene: evidence from the Changchang Flora, Hainan Island, China. Palaeogeogr. Palaeocl. 412, 1–16.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Tang, Z., Wang, Z., Zheng, C., Fang, J., 2006. Biodiversity in China's mountains. Front. Ecol. Environ. 4, 347–352.
Taylor, S., Kumar, L., 2016. Global climate change impacts on pacific islands terrestrial biodiversity: a review. Trop. Conserv. Sci. 9, 203–223.
Tsukaya, H., 2004. Leaf shape: genetic controls and environmental factors. Int. J. Dev. Biol. 49, 547–555.
Voris, H.K., 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. J. Biogeogr. 27, 1153–1167.
Wallace, A.R., 1869. Sir Charles Lyell on geological climates and the origin of species. Quart. Rev. 126, 359–394.
Wang, C., Zöllner, S., Rosenberg, N.A., 2012. A quantitative comparison of the similarity between genes and geography in worldwide human populations. PLoS Genet. 8, e1002886.
Whittaker, R.J., Fernández-Palacios, J.M., 2007. Island Biogeography: Ecology, Evolution, and Conservation. Oxford University Press, Oxford.
Writing Group of Cenozoic Plants of China, 1978. Cenozoic Plants from China. Science Press, Beijing, Fossil Plants of China.
Xing, Y.W., Ree, R.H., 2017. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. Proc. Natl. Acad. Sci. USA 114, E3444–E3451.
Xu, H., Su, T., Zhang, S.T., Deng, M., Zhou, Z.K., 2016. The first fossil record of ring-cupped oak (*Quercus* L. subgenus *Cyclobalanopsis* (Oersted) Schneider) in Tibet and its paleoenvironmental implications. Palaeogeogr. Palaeocl. 442, 61–71.
Xu, J., Deng, M., Jiang, X.L., Westwood, M., Song, Y.G., Turkington, R., 2015. Phylogeography of *Quercus glauca* (Fagaceae), a dominant tree of East Asian subtropical evergreen forests, based on three chloroplast DNA interspace sequences. Tree Genet. Genomes 11, 1–17.
Ye, J., Zhang, Y., Wang, X., 2017. Phylogeographic breaks and the mechanisms of their formation in the Sino-Japanese floristic region. Chinese J. Plant Ecol. 41, 1003–1019.
Ying, J.S., Xu, G.S., 2002. An analysis of the flora of seed plants of Taiwan, China: its nature, characteristics, and relations with the flora of the mainland. Acta Phytotaxonomica Sinica 40, 1–51.
Yu, G., Chen, X.D., Ni, J., Cheddadi, R., Guiot, J., Han, H., Harrison, S.P., Huang, C., Ke, M., Kong, Z., Li, S., Li, W., Liew, P., Liu, G., Liu, J., Liu, Q., Liu, K.B., Prentice, T.C., Qui, W., Ren, G., Song, C., Sugita, S., Sun, X., Tang, L., Van Campo, E., Xia, Y., Xu, Q., Yan, S., Yang, X., Zhao, J., Zheng, Z., 2000. Palaeovegetation of China: a pollen data-based synthesis for the mid-Holocene and last glacial maximum. J. Biogeogr. 27, 635–664.
Zhang, J.H., 1978. Guizhou Branch Altas (II) Atlas of palaeobiology of SW China. Geological Publishing House, Beijing.
Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. 7, 203–214.
Zhu, H., 2016. Biogeographical evidences help revealing the origin of Hainan Island. PloS one 11, e0151941.
Zhu, H., 2017. Families and genera of seed plants in relation to biogeographical origin on Hainan Island. Biodivers. Sci. 25, 816–822.