Research Article

Combined genotype and phenotype analyses reveal patterns of genomic adaptation to local environments in the subtropical oak *Quercus acutissima*

Jie Gao[1,2†]* 🄳, Zhi-Long Liu[3†], Wei Zhao[4†], Kyle W. Tomlinson[2,5], Shang-Wen Xia[1], Qing-Yin Zeng[6], Xiao-Ru Wang[4], and Jin Chen[1]

[1]CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun 666303, Yunnan, China
[2]Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Menglun 666303, Yunnan, China
[3]Experimental Center of Tropical Forestry, Chinese Academy of Forestry, Pingxiang 532600, Guangxi Zhuang Autonomous Region, China
[4]Department of Ecology and Environmental Science, UPSC, Umeå University, Umeå 90187, Sweden
[5]Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun 666303, Yunnan, China
[6]State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing 100091, China
[†]These authors contributed equally to this work.
*Author for correspondence. E-mail: gaojie@xtbg.org.cn

Abstract Understanding the effects of the demographic dynamics and environmental heterogeneity on the genomic variation of forest species is important, not only for uncovering the evolutionary history of the species, but also for predicting their ability to adapt to climate change. In this study, we combined a common garden experiment with range-wide population genomics analyses to infer the demographic history and characterize patterns of local adaptation in a subtropical oak species, *Quercus acutissima* (Carruthers). We scanned approximately 8% of the oak genome using a balanced representation of both genic and non-genic regions and identified a total of 55 361 single nucleotide polymorphisms (SNPs) in 167 trees. Genomic diversity analyses revealed an east–west split in the species distribution range. Coalescent-based model simulations inferred a late Pleistocene divergence in *Q. acutissima* between the east and west groups as well as subsequent preglaciation population expansion events. Consistent with observed genetic differentiation, morphological traits also showed east–west differentiation and the biomass allocation in seedlings was significantly associated with precipitation. Environment was found to have a significant and stronger impact on the non-neutral than the neutral SNPs, and also significantly associated with the phenotypic differentiation, suggesting that, apart from the geography, environment had played a role in determining non-neutral and phenotypic variation. Our approach, which combined a common garden experiment with landscape genomics data, validated the hypothesis of local adaptation of this long-lived oak tree of subtropical China. Our study joins the small number of studies that have combined genotypic and phenotypic data to detect patterns of local adaptation.

Key words: demographic history, isolation by distance, isolation by environment, local adaptation, phenotype variation, RAD sequencing.

1 Introduction

Detecting signatures of adaptation across conspecific populations is important for understanding both the genetic basis of adaptation to the local environment and the potential effects of climate change on species (Savolainen et al., 2013; Prunier et al., 2016; Yeaman et al., 2016). Species occupying spatially heterogeneous environments are usually subject to heterogeneous selective pressures. Natural selection acts initially on phenotypic traits, over time resulting in changes of allele frequencies at the specific genetic loci underlying those traits and leading finally to locally adapted populations (Orsini et al., 2013). There is evidence of widespread local adaptation both in plants (Leimu & Fischer, 2008) and animals (Fraser et al., 2011). Tree species are of particular interest for research on local

adaptation because of their economic and ecological importance (Hoffmann & Sgro, 2011). Dominant tree species are keystone species for many complex ecological communities and many occupy highly heterogeneous habitats. In addition, trees typically have high outcrossing rates and large effective population sizes, resulting in low levels of linkage disequilibrium (Moritsuka et al., 2012; Parchman et al., 2012). Their genomes could therefore contain genetic material of unrealized adaptive potential (Sork et al., 2013; Iida et al., 2014).

The architecture of adaptive traits and evolutionary history are important for studying the adaptation of tree species (Alberto et al., 2013; Kremer et al., 2014). Historical demographic processes such as range expansion and serial colonization could determine population structure and mimic the patterns produced by natural selection (Excoffier et al., 2009; Siol et al., 2010; Hoban et al., 2016). Although many common garden and reciprocal-transplant experiments have been carried out to demonstrate the adaptation of tree populations to local environments (reviewed by Savolainen et al., 2007, 2013; Lind et al., 2017), it is still somewhat difficult to reveal the genetic basis of local adaptation. The lack of genome-wide investigation limits the ability to identify the genomic regions that are under selection as, in most cases, the traits that confer local adaptation in trees are suggested as polygenic quantitative traits (Le Corre & Kremer, 2012). Thus, many markers dispersed throughout the genome are needed to characterize both adaptive and random changes (Bragg et al., 2015; Chen et al., 2016). Given that few tree species have substantially developed genomic resources, restriction site-associated DNA sequencing (RAD-seq) (Miller et al., 2007; Baird et al., 2008) was a cost-effective and efficient method for generating single-nucleotide polymorphisms (SNPs) markers distributed across the whole genome. It could provide both genome-wide amounts of neutral and non-neutral loci for genomic studies of demographic history and for adaptation in trees (Parchman et al., 2018).

Oak (*Quercus* L. and related genera in the Fagaceae) is a powerful model for studying the adaptation of forest trees to global environmental changes in Eurasia and North America (Petit et al., 2013; Cannon et al., 2018). The combined effects of geographic distance and adaptation on the genetic and morphological differentiation of oak populations has been widely reported both in Europe (Alberto et al., 2013; Rellstab et al., 2016; Pina-Martins et al., 2019) and in California (Riordan et al., 2016; Sork et al., 2016b; Gharehaghaji et al., 2017). In addition, evidence of spatially divergent selection on climate-associated candidate genes in natural populations of valley oak (*Quercus lobata* Née) has provided preliminary indications of local adaptation (Sork et al., 2016b). Moreover, a study of three oak species (*Q. petraea* [Matt.] Liebl., *Q. pubescens* Willd., and *Q. robur* L.) in Switzerland provided an example of adaptive genetic variation associated with local topography, historical climate, and soil characteristics (Rellstab et al., 2016). However, in tropical and subtropical Asia, most genetic studies of *Quercus* and related genera have primarily focused on phylogeography and genetic diversity (Zhang et al., 2013; Shi et al., 2014; Xu et al., 2015; Zeng et al., 2015; Du et al., 2017; Meng et al., 2017; Deng et al., 2018; Jiang et al., 2018). Only two studies, one focusing on *Castanopsis carlesii* (Hemsley) Hayata (Sun

et al., 2016) and another on *Q. championii* Benth. (Jiang et al., 2019), have used genome-wide markers to address whether climatic factors affect genetic variation in natural populations. Both studies suggested that geography and climate jointly shape the genetic structure of these species. Furthermore, specific climate variables such as seasonal precipitation were identified as key drivers of the local adaptation of *Q. championii* (Jiang et al., 2019). Further research, and in particular common garden experiments undertaken in concert with landscape genomics assessments, could simultaneously quantify the roles of phenotypic plasticity and local adaptation in oak species' response to environmental heterogeneity.

Quercus acutissima (Carruthers) is a dominant species occupying a wide range of environmental conditions in subtropical and warm temperate zones in China. Owing to its high ecological and economic value, it has been listed as a precious timber species in China. Populations of this species can persist across large temperature and precipitation ranges, which are associated with continuous gradients in latitude and longitude (and distance from the ocean), respectively. An investigation of the seeds of *Q. acutissima* found throughout its natural distribution showed significant differences between populations, especially in seed mass and free sugar content (Liu et al., 2011). Based on observed patterns of phenotypic differentiation, we hypothesized that environmental heterogeneity might exert strong selective pressures on these populations, and could thereby drive phenotypic and genotypic divergence in populations of *Q. acutissima.* The aim of the present study was to test the hypothesis of genetic-based climate adaptation in *Q. acutissima* by combining two complementary approaches, that is, an analysis of genome-wide genetic markers obtained using RAD-seq, and direct phenotypic evidence from a common garden experiment. As some of the phenotypic differences observed between populations in a common garden experiment are heritable (and could therefore be adaptive) rather than plastic (Gonda et al., 2011; Brachi et al., 2013), this experiment also enabled us to detect the local adaption on the phenotypic variation among populations from different environments. Our specific goals were to: (i) characterize the genetic structure of *Q. acutissima* populations; (ii) quantify phenotypic differentiation and the main environmental variables correlated with measured traits; (iii) examine the roles played by geographic isolation and adaptation in determining phenotypic and genotypic variation; and (iv) identify candidate outlier loci that might be involved in local adaptation.

2 Material and Methods

2.1 Sample collection and common garden setup
We collected acorns or leaves from 27 natural populations of *Quercus acutissima* in 2006. The locations of the sampled populations covered most of the species' natural distribution in mainland subtropical China, over a latitudinal range from 21.936°N to 35.8°N and a longitudinal range from 100.159°E to 119.767°E (Fig. 1; Table S1).

We set up a common garden at Chuzhou, Anhui province, China (118.06°E, 32.16°N) using acorns from 15 to 20 adult
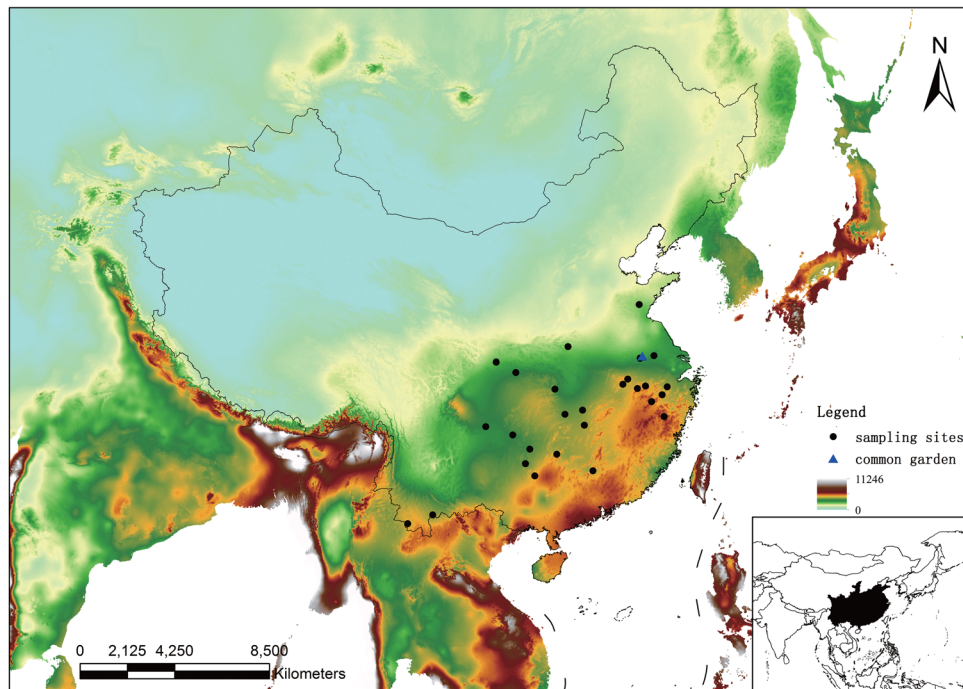
Fig. 1. Locations of the sampled *Quercus acutissima* populations and the common garden site in China, with the map showed the gradient of the mean annual total precipitation in the sampled area. The maps are taken from the http://www.tianditu.gov.cn/ and the mean annual total precipitation data from the http://www.worldclim.org/.

trees of each population for long-term observation (Fig. 1). The common garden is located in a transitional area between the humid zone in central China and the semi-arid temperate zone in north China, with an annual mean temperature of 15.2℃ and annual precipitation of 1041.6 mm. We established the common garden using a random block design, with 15 seeds (mixed seeds from all individuals per population) in each population plot and three blocks. To prevent edge effects, we planted the same species tree around the trial. The morphological traits of seedlings were measured at 1 year old.

2.2 Restriction site-associated DNA sequencing and SNP calling

We isolated DNA from leaves of eight randomly selected and unrelated individuals from each population in the common garden and generated RAD sequencing data (Baird et al., 2008). In total, we obtained high-quality data for 167 individuals. The RAD-seq library was prepared using 200 ng DNA from each individual tree; after extraction, DNA samples were digested with the restriction enzyme *Eco*RI. Illumina sequencing adaptor and a unique barcode were ligated to *Eco*RI-digested fragments. All the individual libraries were pooled and randomly sheared, and amplified by polymerase chain reaction using Illumina sequencing primers (Doc. S1). DNA fragments with insert fragment sizes from 200 to 400 bp were isolated from a 2.0% agarose gel using QiaQuick gel extraction kits (Qiagen, Hilden, Germany). These fragments were then double-end sequenced on an Illumina HiSeq2500 (Illumina, San Diego, CA, USA).

The quality of the raw sequence reads was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/); adaptor sequences and low-quality bases (i.e., with base quality scores < 20) at the start and/or end of each read were removed using Trimmomatic (Bolger et al., 2014). Reads shorter than 36 bp were discarded. We used BWA-MEM (Li, 2013) to map the remaining reads to the *Q. lobata* reference genome (Sork et al., 2016a) with default parameters. *Quercus lobata* is the closest relative to *Q. acutissima* with a relatively good quality of genome sequence available. We carried out SNP calling using GATK version 3.2.2 with HaplotypeCaller and GenotypeGVCFs (DePristo et al., 2011). We also used Vcftools (Danecek et al., 2011) to further filter low-quality SNPs according to the following criteria: (i) only biallelic SNPs that were at least 5 bp away from any indel and with mapping quality > 40 were retained; (ii) genotypes with a genotype quality score < 20 and sequencing depth < 10 were treated as missing; and (iii) SNP with a genotype missing call rate ≥ 30% and a minor allele frequency < 0.05 across the whole sample set were removed. After these filtering steps, 55 361 SNPs were retained for genetic analyses. For genetic structure and coalescent simulation analysis, we applied different filtering strategies as described below.

2.3 Genetic diversity, genetic structure, and population demography

We estimated gene diversity (π) for each population using Arlequin version 3.5 (Excoffier & Lischer, 2010), and the per-individual based heterozygosity of each population using VCFtools version 1.012 (Danecek et al., 2011).
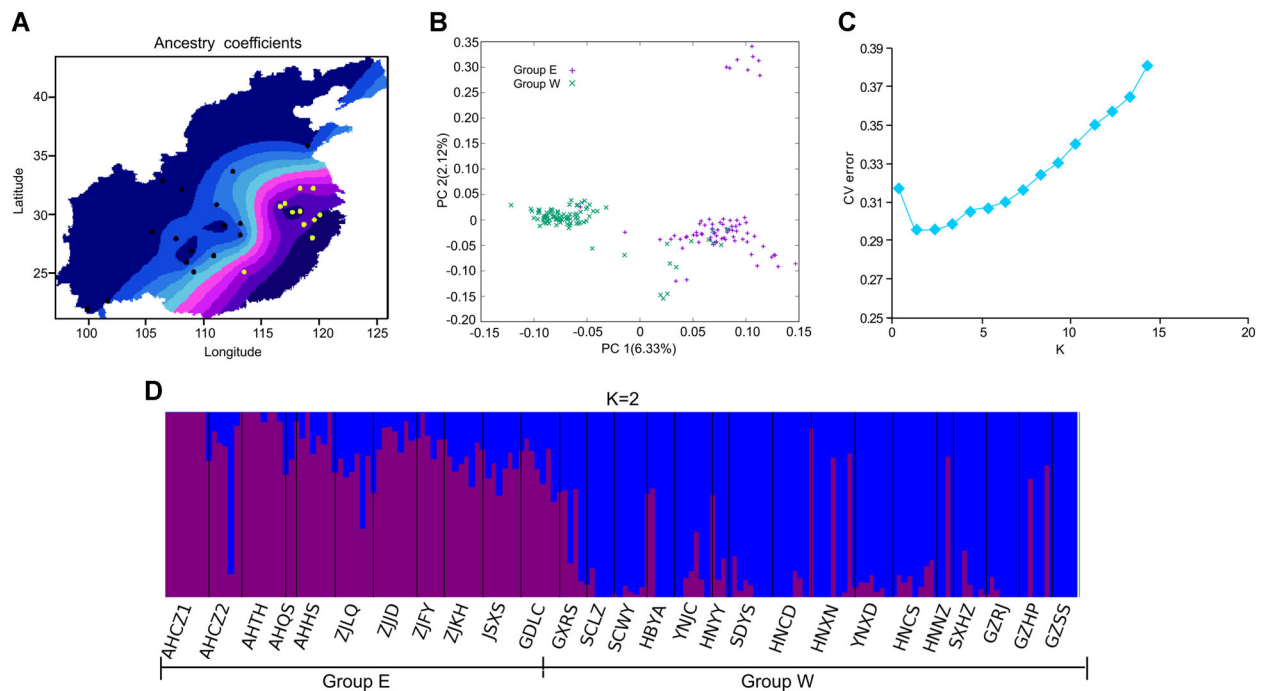
Fig. 2. Genetic structure, principal component analysis (PCA) plots of the restriction site-associated DNA sequencing variation in 27 populations and the admixture proportions of each population from $K = 2$. **A,** Genetic structure of the two groups from the method of Admixture (black dots represent group W and yellow dots represented group E). **B,** Distribution of genetic variance by PCA. **C,** Cross-validation value of each K from the admixture analysis. **D,** Admixture proportions of each population from $K = 2$.

We investigated the genetic structure of *Q. acutissima* populations using the model-based clustering algorithm implemented in Admixture version 1.23 (Alexander et al., 2009). As linkage disequilibrium could violate the statistical assumptions involved in this test, we pruned the dataset (kept 19 834 SNPs) using the indep-pairwise function (50 5 0.05) in PLINK (https://www.cog-genomics.org/plink/1.9/). We ran Admixture with $K = 1$-15 and repeated the process 20 times with random seeds. A 10-fold cross-validation procedure was used for evaluating the runs with different K values in Admixture. We also assessed the distribution of genetic variance by principal component analysis (PCA) as implemented in EIGENSOFT version 6.0 (Patterson et al., 2006).

In cases of uneven sampling across species' range, the model-based methods for population structure may tend to partition continuous variation into spurious clusters and overestimate the number of potential clusters. The conStruct software (https://github.com/gbradburd/conStruct) infers genetic structure by taking into account the continuous pattern of differentiation in geographically separated populations (Bradburd et al., 2018). It examines two models, a non-spatial model, which is similar to the model used in Admixture, and a spatial model that allows for testing for the presence of isolation-by-distance. We tested both the spatial and non-spatial models for a range of $K = 1$-10, with 10 repetitions per K value and 50 000 iterations per repetition. We used 10-fold cross-validation and calculated the relative contribution of each cluster to the total covariance to determine the most likely number of clusters (K value).

To gain insight into the historical processes that generated the observed population structure, we tested eight demographic models using fastsimcoal2 (Excoffier & Foll, 2011; Excoffier et al., 2013) based on the identified two clusters (East and West) in the sampled populations (see "Results"). The models tested describe different demographic scenarios for the two groups. These models include: M1-2, divergence with or without gene flow, and descendant populations have constant population sizes; M3-4, after divergence, the descendant populations show exponential population size changes in both groups (with or without gene flow); M5-6, after divergence, the descendant populations show stepwise population size changes followed by exponential growth (with or without gene flow); and M7-8, after divergence, the descendant populations show constant exponential growth from different start times (see Fig. 3). Since fastsimcoal2 models with missing data can result in biased estimates of the site frequency spectrum (SFS), we performed a down-sampling procedure following Thomé & Carstens (2016) to maximize the number of segregating sites retained for SFS estimation. Finally, we generated a joint folded SFS using Arlequin v3.5 (Excoffier & Lischer, 2010) for Group W ($n = 36$) and Group E ($n = 30$) using unpruned SNPs with > 10x coverage, including low-frequency SNPs (without filtering out SNPs with minor allele frequency < 0.05) and no missing data (4247 SNPs left). We used a mutation rate per site per year value of 1.8×10^{-9}, as previously estimated for the genus *Quercus* (Cavender-Bares et al., 2011) as well as a generation time of 50 years (Xu et al., 2016) for all coalescent
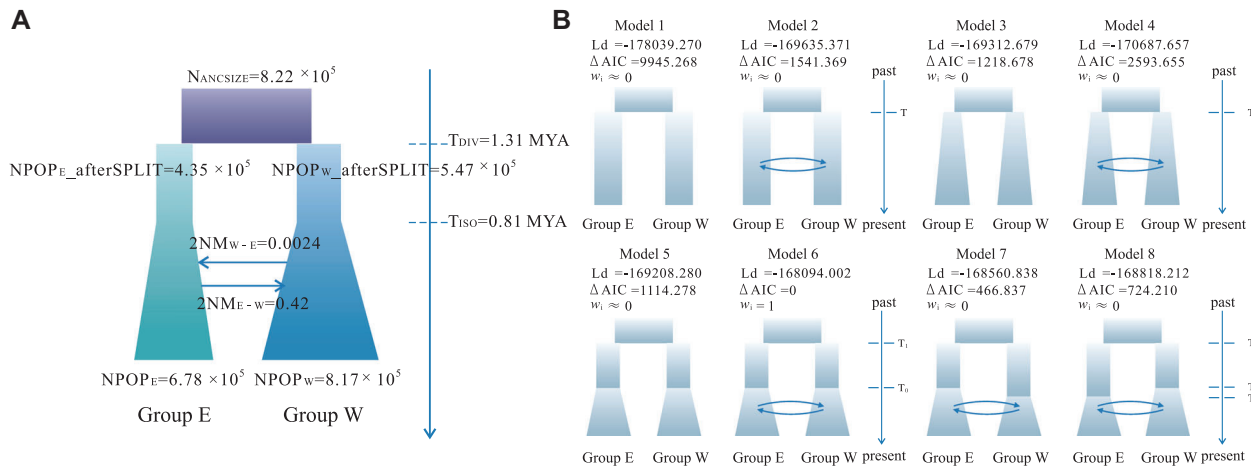
Fig. 3. A, Summary of the best demography model for the two *Quercus acutissima* groups. Eight demographic parameters estimated by fastsimcoal2 are shown. Each block represents a current or ancestral population with its estimated effective population size (Ne). Arrows denote the direction of gene flow with the estimated migration rate labeled above or below the arrow. The timings of the splitting events and expansion events are indicated in Mya. **B,** Eight demographic scenarios of the two groups: M1-2, divergence with or without gene flow, and descendant populations have constant population sizes; M3-4, after divergence, the descendant populations show exponential population size changes in both groups (with or without gene flow); M5-6, after divergence, the descendant populations show stepwise population size changes followed by exponential growth (with or without gene flow); M7-8, after divergence, the descendant populations show constant exponential growth from different start times.

simulations. Each model proceeded for 100 replications of 200 000 simulations to calculate composite likelihoods (Papadopoulou & Knowles, 2015). Model selection was based on maximum likelihood over 100 independent runs using the Akaike information criterion (Excoffier et al., 2013), and the relative likelihood w_i. We calculated the 95% confidence intervals (CI) for all parameter estimates from 100 parametric bootstrap replicates by simulating the SFS from the maximum composite likelihood estimates; parameters were reestimated each time.

2.4 Outlier detection and genetic-environmental association analysis

To detect putative outlier loci that may be associated with local adaptation, we used the allele frequency-based *pcadapt* (Luu et al., 2017) and a genetic-environmental association-based method known as Bayenv2 (Coop et al., 2010; Günther & Coop, 2013). We used the dataset including 55 361 SNPs (without pruning) for these analyses, considering that significant associations are likely to be near the true target loci of selection. The *pcadapt* takes account of the population structure with PCA and using the Mahalanobis distance to detected outliers with respect to how they are related to population structure (Luu et al., 2017). We ran *pcadapt* iteratively for $K = 1$-10 latent factors for 2000 iterations with scaled allele frequencies (i.e., the default parameter); and used Cattell's graphical rule to choose the number of principal components (K). We used the R package qvalue method (Storey et al., 2015) to select SNPs at false discovery rate 0.05.

Bayenv2 tests for correlation between allele frequencies and environmental variables while accounting for sampling

and covariance due to shared population history. It uses a set of putatively neutral loci as a null model to estimate an arbitrary covariance matrix of allele frequency for calibration (Günther & Coop, 2013). We generated a set of putatively neutral and independent SNPs by excluding the outliers identified by *pcadapt* and pruned the dataset using the indep-pairwise (50 5 0.05) implemented in PLINK. This set of SNPs was then used to estimate the covariance matrix over 100 000 iterations generated by three independent runs with different seed numbers to ensure convergence. The environmental correlations were determined by averaging five independent runs of Bayenv2 with 100 000 Markov chain Monte Carlo iterations with different random seeds and using the covariance matrix estimated by the null model as a calibration. We only considered SNPs that were in the top 1% of the observed Bayes factor values (i.e., > 3) and in the top 5% of the absolute Spearman's rank correlation coefficient (ρ) results to be robust candidate adaptive loci. The climate data included in our genetic-environmental association analysis were obtained from Worldclim (available at http://www.worldclim.org/) (Hijimans et al., 2005); soil organic carbon, soil pH, and growing degree day values were from the Center for Sustainability and the Global Environment (https://nelson.wisc.edu/sage/data-and-models/); ground-frost frequency, wet day frequency, and vapor pressure values were taken from the IPCC database (http://www.ipcc-data.org/observ/index.html); the global UV-B seasonality values were taken from the global UV-B radiation dataset (https://www.ufz.de/gluv/). To avoid multicollinearity, we selected 12 environmental variables that had pairwise Spearman correlation coefficients < 0.8 in our sampling range (Table S1).

www.jse.ac.cn

J. Syst. Evol. 59(3): 541–556, 2021

2.5 Isolation by distance and isolation by environment

To understand the relative contributions of geography and environment to population differentiation, we undertook a series of redundancy analyses (RDA) to divide the among-population genetic variation into IBD (isolation by distance) and IBE (isolation by environment) fractions. We included two independent matrices, the 12 environmental variables (representing IBE) and geographic variables (IBD). The geographical matrix was based on the geographic coordinates (in decimal degrees) transformed by the *pcnm* function in the *vegan* R package (Borcard & Legendre, 2002). To prevent over-fitting, we used a forward selection procedure on geographic and environment matrix variables that retained only the relevant polynomial terms. After forward selection, the geographic variables of PCNM1, PCNM2, PCNM3, PCNM6, PCNM7, and PCNM9 and the environment variables of bio7, bio12, bio14, and bio19 were left. We used the Hellinger-transformed allele frequencies prior to RDA (Legendre & Gallagher, 2001), and we calculated the adjusted R^2 (variance explained by the model) to correct for the number of explanatory variables using the *varpart* and *rda* functions of the vegan package (Oksanen et al., 2016). The statistical significance of each partition was tested using the *anova.cca* function of the *vegan* package with a step size of 1000, resulting in at least 999 permutations. We undertook RDAs on both neutral (19 029 SNPs: pruned SNPs excluding the outlier identified by both *pcadapt* and *bayenv2*) and outliers identified by both *pcadapt* and *bayenv2* (119 SNPs).

2.6 Phenotypic trait analysis

We scored the phenotypic traits of three 1-year-old seedlings that were randomly selected from each population grown in the common garden. We took smaller numbers (three individuals per population) for phenotype analysis as the biomass measurement of the seedlings will completely destroy the plants. We only collected the leaves for genotyping in two populations (YNXD and YNJC), and in another population (GZHP) we did not obtain phenotype data, thus we undertook the phenotypic trait analysis based on the remaining 24 populations. The traits that we measured included total mass, leaf area, leaf mass fraction (LMF, g leaf mass g^{-1} total mass), stem mass fraction (SMF, g stem mass g^{-1} total mass), root mass fraction (RMF, g root mass g^{-1} total mass), and specific taproot length (STRL, mm root length g^{-1} root mass), which measures the root extension achieved per unit root mass. The LMF, SMF, and RMF represent biomass partitioning among the major organs, and STRL represents the efficiency of soil depth penetration to access deeper water (Tomlinson et al., 2012).

We ran a PCA on these morphological traits for all populations to examine whether morphological and genetic variables showed the same structure. Then we tested whether any of the traits differed between the two groups using Welch's two-sample *t*-tests, allowing for unequal variances. Next, we ran multiple regression models to find the best environmental predictors for each trait. We rescaled all predictors to a mean of zero and a variance of one so that coefficients could be compared directly. Because we had few replication samples, we used forward selection to maximize the degrees of freedom of the error. Parameters were added into the model only if they decreased the Akaike information criterion. All statistical analyses were undertaken using R (R Development Core Team, 2016). In addition, we examined the association of morphological variation with environmental and geographical factors using RDA.

3 Results

3.1 Restriction site-associated DNA sequencing data, genetic diversity, and structure

Sequencing of the RAD library generated 2 016 014 750 reads in total and with an average of 12 071 945 reads per individual. After trimming and quality filtering, we retained an average of 11 359 235 reads per individual (Table S2). Relative to the *Quercus lobata* genome size of 0.76 Gbp, our RAD sequences covered approximately 8% of the genome (average 58.88 Mbp, with a range of 27.69 to 93.81 Mbp among individuals) with a mean depth of 15.5 × (Table S2; Fig. S1). The GATK pipeline identified 856 474 RAD loci in total. We applied a stringent reference genome-guided SNP calling strategy, and obtained a total of 55 361 SNPs from 167 samples, of which 23 190 (42.7%) were located in genic regions. The distributions of the missing data over all individuals were between 0.03 and 0.29 (Fig. S1; Table S2).

All methods used to infer the genetic structures (i.e., admixture, conStruct, and PCA) have concordance results that revealed two genetic clusters in our samples of *Q. acutissima* which split the sampled populations into the East (E) and West (W) groups (Fig. 2). Group E included 11 populations (AHCZ1, AHCZ2, AHTH, AHQS, AHHS, ZJLQ, ZJJD, ZJFY, ZJKH, JSXS, and GDLC) ranging from Anhui to Zhejiang provinces, and group W included 16 populations (GXRS, SCLZ, SCWY, HBYA, HNYY, SDYS, HNCD, HNXN, YNXD, YNJC, HNCS, HNNZ, SXHZ, GZRJ, GZHP, and GZSS) ranging from Hubei to Yunnan in SW China (Figs. 2A, 2C). For the conStruct analysis, the cross-validation plot showed that the predictive accuracy improved very little after $K = 2$, although the highest predictive accuracy continued to $K = 5$ in the non-spatial model, which is indicative of overestimation of the potential cluster number (Fig. S2). For the spatial model, the highest predictive accuracy was $K = 3$; however, the first two layers contributed the most to the total covariance (Fig. S3), indicating that $K = 2$ sufficiently described the population structure. The spatial model displayed a better fit than the non-spatial model at $K = 2$ (Fig. S2), and the decay of genetic relatedness against geographic distance ($\alpha_D = 0$ for the non-spatial model, $\alpha_{D(layer\ 1)} = 0.005$ and $\alpha_{D(layer\ 2)} = 0.001$ for the spatial model (Fig. S3; Table S3) suggested weak IBD within each ancestral layer. Together, the conStruct results support the East and West groups and moderately indicate the presence of IBD. One population (AHTH) from the East group differed from other populations in the structure analysis when $K = 3$ and 4, which might be due to the genetic introgression from other sympatric-related species (*Q. variabilis* Blume and *Q. chenii* Nakai). Thus, this population was excluded in the following population history analysis. Partitioning the genetic variance from the PCA analysis also clearly separated the populations into these two groups (Figs. 2B, S4). Measurements of genetic diversity and the average observed heterozygosity in the sampled populations are summarized in Table S1. The estimates of gene diversity (π) and observed heterozygosity (H_o) were slightly higher in group E

(0.2319 and 0.2772, respectively) than in group W (0.2222 and 0.2593).

3.2 Population history

We tested several demographic scenarios between the two groups in *Q. acutissima* using fastsimcoal2. The best model was the one that included recent exponential growth in population size of both groups after splitting (i.e., Model 6) (Fig. 3; Table S4). Assuming a generation time of 50 years and a mutation rate per year of 1.8×10^{-9} (as previously estimated for the genus *Quercus*, see Cavender-Bares et al., 2011), the divergence time between groups E and W was estimated at 1.31 Mya (95% CI, 1.27–1.34) and the effective ancestral population size was 8.2×10^5 (95% CI, 7.94×10^5–8.5×10^5) (Table S5). This model suggested that the two groups experienced stepwise population size changes until 0.81 Mya (95% CI, 0.77–0.84 Mya), followed by exponential population growth. The estimated current population sizes were 6.77×10^5 (95% CI, 6.52×10^5–7.03×10^5) and 8.17×10^5 (95% CI, 7.86×10^5–8.48×10^5) for groups E and W, respectively. The estimated current gene flow was 0.42 migrants per generation (95% CI, 0.32–0.52) from group E to group W and 0.0024 (95% CI, 0.0018–0.003) from group W to group E (Fig. 3; Table S5).

3.3 Outlier detection and genetic–environmental association analysis

Analyses using *pcadapt* detected 2628 SNPs (4.75% of 55 361 SNPs) as significant outliers, while Bayenv2 identified 431 outlier SNPs (0.78% of 55 361 SNPs). The number of SNPs associated with environmental variables is shown in Table S6. To diminish the frequency of false positives, we selected 119 SNPs identified by both methods as robust outliers. The potential functions of the genes carrying these 119 SNPs were deduced from annotations of their homologs in the *Q. lobata* genome, which showed that 55 were functional genes. Thirty-seven genes are associated with temperature and 12 genes are likely related to adaptation to precipitation (Table S7). There are also seven genes that correlate to soil pH and four genes related to global UV-B radiation.

To understand the impacts of environment and geography on genetic and morphological differentiation, we undertook a series of RDAs on both neutral (pruned SNPs excluding the outliers from both the *pcadapt* and bayenv2 analysis) and putatively adaptive SNPs (outliers detected by both the *pcadapt* and bayenv2 analysis) as well as phenotypic traits (Table 1). We found that, when environment was controlled for, geography alone explained a significant proportion (10.11%) of the variation on neutral loci, but when geography was controlled for, the effect of environment was not significant. Regarding outlier loci, both geography (22.36%, $P = 0.001$, constrained by environment) and environment (5.16%, $P = 0.029$, constrained by geography) showed significant contributions. The combined effect of geography and environment was 44.06% on outlier loci and 20.80% on neutral loci (Table 1).

3.4 Phenotypic and climate variation between groups

The two population groups differed with respect to stem mass fraction (SMF; t-value = 2.78, $P = 0.012$) and root mass fraction (RMF; t-value = 2.46, $P = 0.023$), but not with respect to leaf mass fraction (LMF; t-value = 1.08, $P = 0.293$), leaf area (t-value = 0.33, $P = 0.744$), STRL (t-value = 0.39, $P = 0.704$), or total mass (t-value = 0.93, $P = 0.36$) (Fig. 4).

The two population groups appear to inhabit different precipitation zones, with Group E being found in areas with moister conditions than group W. Group E areas were found to be associated with greater mean annual precipitation (t-value = 2.92, $P < 0.01$), greater precipitation in the driest month (t-value = 3.08, $P < 0.01$), and greater precipitation of coldest quarter (t-value = 1.73, $P < 0.01$) (Table 2).

Mean annual rainfall affected biomass traits, with a positive effect on leaf and stem mass fractions and a negative effect on root mass fraction (Fig. 5; Table 3). Soil organic carbon also affected several traits, notably total mass (negative), leaf area (positive), and specific taproot length (positive; see Table 3). The diurnal temperature range had a positive effect on leaf area, whereas soil pH and

Table 1 Redundancy analysis (RDA) was used to partition population genetic variation (allele frequencies, F) and phenotypic variation (phenotypical variables, Morp) into two components: (i) environmental (IBE); and (ii) geography (IBD) in *Quercus acutissima*

	Outlier (119 SNPs)	Neutral loci (19 029 SNPs)	Phenotype	
	R^2	R^2		R^2
Combined fractions				
F~clim.	0.2169***	0.0729***	Morp~clim.	0.2360*
F~geog.	0.3890***	0.1655***	Morp~geog.	0.0546^{ns}
Individual fractions			Individual fractions	
F~geog.|clim.	0.2236**	0.1011***	Morp~geog.|clim.	0.1081^{ns}
F~clim.|geog.	0.0516*	0.0085^{ns}	Morp~clim.|geog.	0.2895*
F~(geog. + clim.)	0.4406	0.1740	Morp~(geog. + clim.)	0.3441
Total explained	0.716	0.284	Total explained	0.7417
Total unexplained	0.284	0.716	Total unexplained	0.2583
Total	1.0000	1.0000	Total	1.0000

Redundancy analysis test is of the form: F~independent matrices|covariate matrices. Significance of confounded fractions between environment and geography was not tested.
*$P < 0.1$; **$P < 0.01$; ***$P < 0.001$. ns, not significant; Total explained, total adjusted R^2 of individual fractions.
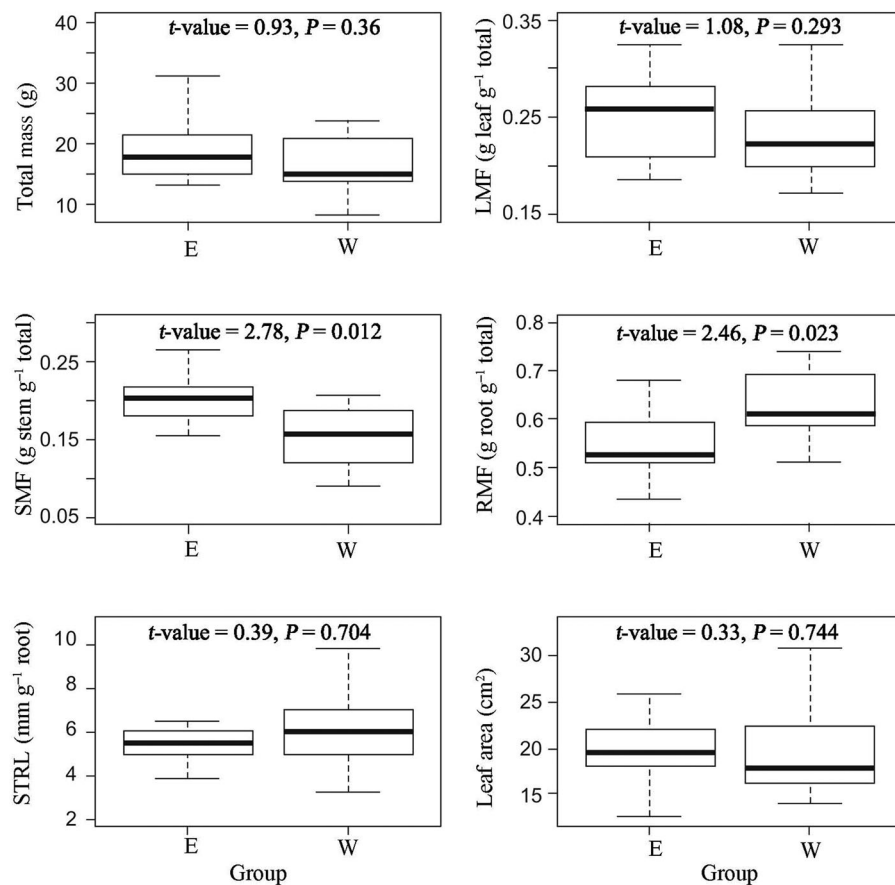
Fig. 4. Spread of trait values for each population group of *Quercus acutissima*. LMF, leaf mass fraction; RMF, root mass fraction; SMF, stem mass fraction; STRL, specific taproot length.

Uvb both had negative effects on leaf area (Table 3). Two components, which together accounted for 72.4% of the variation, were identified by the morphological PCA test and were consistent with the pattern of genetic structure. The root, stem, and leaf mass fractions were all correlated with the first PCA axis, whereas specific taproot length, total mass, and leaf area were correlated with the second axis (Fig. S4). With respect to morphological variation, only the environment was a significant predictor (28.95%, $P = 0.035$), once geography was controlled for. The combined effect of geography and environment explained 34.41% of the morphological variation (Table 1).

Table 2 Comparison of mean environmental conditions for each population group according to the environmental predictors included in this analysis, using Welch's two-sample *t*-tests, allowing for unequal variances

Environmental parameter	Units	Group E	Group W	t-value	d.f.	P-value
Bio 1: Mean annual temperature	(°C × 10)	163.80	156.90	0.80	21.60	0.431
Bio 2: Mean diurnal temperature range	(°C × 10)	81.30	83.50	−0.62	15.60	0.543
Bio 4: Temperature seasonality	(SD×100)	8348.80	7944.10	1.30	20.30	0.208
Bio 5: Max temperature of warmest month	(°C × 10)	321.70	309.50	1.60	22.00	0.124
Bio 7: Temperature annual range	(°C × 10)	317.90	308.30	1.01	16.90	0.329
Bio 12: Mean annual precipitation	(mm)	1476.50	1183.10	2.92	20.40	0.008*
Bio 14: Precipitation of driest month	(mm)	42.40	27.50	3.08	17.54	0.007*
Bio 15: Precipitation seasonality	(CV)	52.80	63.20	−2.10	12.99	0.055
Bio 19: Precipitation of coldest quarter	(mm)	181.40	103.60	3.60	21.20	0.002*
Soil organic carbon content	(g/cm³)	4.60	4.61	−0.02	21.10	0.983
Soil pH	(pH)	6.04	6.14	−0.63	21.10	0.535
UVb2: Global UV-B seasonality	(J/m²/day)	128 830.80	131 309.60	−1.25	21.20	0.226

*Significant differences ($P < 0.01$), including climate parameters (temperature, rainfall) and soil fertility parameters.
CV, coefficient of variation; *d.f.*, degrees of freedom; SD, standard deviation.
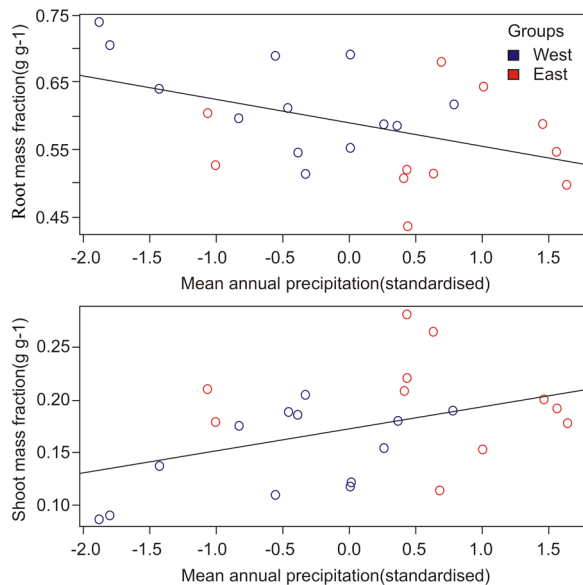
Fig. 5. Relationships of *Quercus acutissima* population-level shoot mass fraction (SMF) and root mass fraction (RMF) with mean annual precipitation. Each population is colored by its assigned population group (blue = West; red = East).

Table 3 Environmental variables that predict variation in measured traits of *Quercus acutissima* across all sampled populations from the common garden experiment

Trait	Intercept	bio 2	bio 12	SC	SpH	Uvb2
Total mass (ln)	2.851			−0.156		
LMF	0.237		0.013			
SMF	0.173		0.021			
RMF	0.589		−0.035			
STRL (ln)	1.744			0.106		
Leaf area (ln)	20.545	0.103		0.068	−0.174	−0.109

Model predictor variables were selected using forward selection using the criterion that variables were included if the model Akaike information criterion decreased.

bio 2, mean diurnal temperature range; bio 12, annual precipitation; LMF, leaf mass fraction; RMF, root mass fraction; SC, soil organic carbon; SMF, stem mass fraction; SpH, soil pH; STRL, specific taproot length; Uvb2, global UV-B radiation seasonality.

4 Discussion

4.1 Spatial genetic structure and divergence history of *Quercus acutissima*

This study surveyed approximately 8% of the oak genome and recovered 55 361 SNPs from sampled populations of *Q. acutissima,* and these sites included a balanced proportion of both genic and non-coding regions. Genetic structure and PCA analyses of these populations indicated an east-west split. Concerning the two geographic discrete groups separated by nearly 300 km, other published reports suggest there is an unsampled natural population that exists in-between the two regions we covered (Zhang et al., 2015, 2018). The Admixture method might over-estimate potential clustering by partitioning a continuous variation into false separated clusters. To further assess the spatial pattern of diversity, we used the conStruct analysis to infer the genetic structure, which accounts for the continuous pattern of differentiation in geographically separated populations (Bradburd et al., 2018). The conStruct spatial model supports a weak IBD and revealed that two ancestry layers contributed the most to the total covariance between populations, indicating that two potential clusters sufficiently described the population structure. A typical IBD refers to a pattern of geographically restricted dispersal, allowing genetic drift to establish differentiation between distant locations (Wright, 1943). However, the weak IBD observed in the conStruct analysis of the *Q. actuissima* populations suggests that other historical processes, like geographic/environmental isolation or past refugia, resulted in the isolation of the eastern and western groups. A previous study of this species using three chloroplast DNA fragments did not detect any obvious geographical structure (Zhang et al., 2015), likely due to incomplete lineage sorting and shared chlorotypes among populations (Simeone et al., 2016; Vitelli et al., 2017). However, a study using 10 nuclear microsatellites (Simple sequence repeats) (Zhang et al., 2018) of this species uncovered a clear east-west divergence, supporting the pattern derived from the whole genome analysis in this study. A similar phylogeographic difference between plant populations found in subtropical west and temperate east China has been found in other Fagaceae species, including *Castanopsis fargesii* Franch. (Li et al., 2014; Sun et al., 2014), *C. eyrei* (Champ. ex Benth.) Tutch. (Shi et al., 2014), and *C. carlesii* (Sun et al., 2016), as well as many other plant species (reviewed by Qiu et al., 2011; see also Liu et al., 2012). Taken together, these results suggest that an east-west split could be common for tree species in this area.

The estimated divergence time between the two groups of *Q. acutissima* was approximately 1.31 Mya (95% CI, 1.27–1.34 Mya). When assuming an average generation time of 50 years, the divergence time estimated by simple sequence repeat markers for *Q. acutissima* (Zhang et al., 2018) was 1.775–1.185 Mya; this value also supports our results, and this estimate is similar to that of another sympatric oak species, *Q. variabilis* (1.45 Mya, Chen et al., 2012). The margins of the east and west populations occur along the eastern margin of the Yunnan–Guizhou Plateau and is the boundary between the second and third Chinese steppe regions (Committee of Chinese Academy of Sciences for Physical Geography of China, 1985). Therefore, the genetic divergence between the east and west populations might be the result of regional responses to the uplift of the Yunnan-Guizhou Plateau and associated mountain ranges in adjacent areas. The uplift of the Yunnan-Guizhou Plateau initiated approximately 2.8 Mya ago (Tang et al., 1994) and several mountain ranges (i.e., the Wuyi and Nanling ranges) gradually uplifted in the late

Pleistocene (Chen & Zhou, 1993; Luo et al., 2010). These physical features could have restricted inter-regional gene flow. Furthermore, the intensified East Asia monsoon during the Pleistocene (Clift et al., 2008; An et al., 2015) increased the climatic differences between the second and third steppe regions, that is, a moist maritime climate in the east and seasonal inland climate in west China. An obvious climatic gradient can be observed from west to east, especially with respect to precipitation in the driest month, the driest quarter, and the coldest quarter; each of these variables is generally low in the west and higher in the east (Li et al., 2014). Taken together, ancient regional geological events along with changes in climate likely drove west-east lineage isolation and increased the habitat heterogeneity of plant populations distributed in East Asia (Yan et al., 2012, 2013; Meng et al., 2017; Jiang et al., 2018), including the habitat of *Q. acutissima.*

The estimated gene diversity in *Q. acutissima* is similar to that of *C. carlesii*, another sympatric Fagaceae species (the mean gene diversity for the western and eastern group was 0.250 and 0.32, respectively, in Sun et al., 2016) with a higher diversity in the eastern group than in the western group; however, these differences are not significant. The two groups each occupied a large area with constant population size over a long period of time following divergence, therefore indicating a relatively consistent genetic diversity. However, the slightly higher diversity in the eastern group might be because genetic diversity was increased by introgression between genetically differentiated populations, as hybridization and introgression are common in sympatric and closely related oak species (Cannon et al., 2018), as documented in *C. carlesii* (Ortego et al., 2015; Sun et al., 2016) as well as in *Q. acutissima*. We found that the rate of gene flow from the east group to the west group was much higher than that from the west group to the east group. Asymmetric gene flow between populations or species is usually caused by a combination of several factors. First, asymmetric gene flow can result from intrinsic factors, such as hybrid incompatibility (Ortego et al., 2016), and extrinsic factors, such as wind direction (Friedman & Barrett, 2009). *Quercus actuissima* is located in the East Asian monsoon region; it is wind pollinated and flowers in early spring (March-April). The subtropical regions of China are mainly affected by western and southern winds from March to May, which facilitate gene flow from the eastern to the southwestern group through wind-borne pollen (Xu & Gao, 1962; Bai et al., 2010). Moreover, a high level of environmental heterogeneity often leads to directional gene flow (Paz-Vinas et al., 2013). Second, population abundance also affects gene flow. Although the estimated effective population size of the east group was slightly lower than that of the west group, potentially leading to a greater gene flow rate from the west group to the east group, this effect might be counteracted by other factors. Third, it is possible that gene flow estimated with the fastsimcoal2 model is biased by genetic introgression from a third, unsampled species (Strasburg & Rieseberg, 2010; Gao et al., 2012). Introgression from the sympatric closely related oak species like *Q. variabilis* (Chen et al., 2012) and *Q. chenii* (Li

et al., 2016) in the east area might result in an overestimation of the gene flow from the east group to the west group and an underestimation of that in the opposite direction. Further analysis should include sister species in the adjacent regions to better understand the patterns of introgression.

4.2 Signature of environmental adaptation

Partitioning the relative effects of geographic distance and environmental differences in *Q. acutissima* showed that geography and environment combined could explain 17.40% of the genetic differentiation of neutral loci and 44.06% of the variation of outlier SNPs. Geography alone explained 10.11% and 22.36% of the variation at neutral and outlier SNPs, respectively, indicating a strong effect of geographic isolation. This suggests that gene flow in this species is geographically limited, which is consistent with the demographic model. Environment had a significant and stronger impact on outlier loci and morphology. Environment alone explained 5.16% of the variation of outlier loci, and 28.95% of phenotype differentiation, but geography had no significant effect on phenotype. Taken together, the genetic and phenotypic analyses revealed a joint effect of geography and environment on *Q. acutissima*. The proportion of the variation of outlier SNPs explained by environment was much smaller in *Q. acutissima* than that reported in two previous studies in *Q. lobata* (Gugger et al., 2016; Sork et al., 2016b). This could be due to the differences in genetic data; the two *Q. lobata* studies used genetic markers either putatively involved in local adaptation or obtained from the transcriptome data. These markers are more likely linked to genes that underwent adaptive selection. However, the proportion of variation explained by environment in our study was larger than those estimated in three California oaks (*Q. engelmannii* Greene, *Q. berberidifolia* Liebm., and *Q. cornelis-mulleri* Nixon & K.P. Steele) using microsatellite markers (Riordan et al., 2016), suggesting the effectiveness of our genome-wide scan method.

Distinguishing isolation by colonization (IBC) from IBD and IBE is difficult as their contributions are often confounded. The serial colonization events could lead to a similar pattern as IBD/IBE (Ramachandran et al., 2005; DeGiorgio et al., 2009; Henn et al., 2012) when colonization routes correlate with the environmental/geographic gradients (Orsini et al., 2013; Dewoody et al., 2015). However, the genetic signatures of IBC could be eliminated if the contemporary level of gene flow and the time since colonization is sufficient (Xia et al., 2018). Tree species in subtropical China were less affected by Pleistocene glaciations, and they have been preserved in multiple localized glacial refugia in central and southwestern China during the glaciations (Hewitt, 2000; Shi, 2002). The previous study of the phylogeographic pattern revealed by chloroplast DNA variation in *Q. acutissima* shows no signs of long-distance colonization (Zhang et al., 2015). Our historical demography inferred a much earlier divergence time (1.31 Mya) of the two groups of *Q. acutissima*, as well as the pre-last glacial maximum expansion (0.81 Mya) of the two groups, support the previous results that it did not experience the postglaciations recolonizaiton. Thus, a long time since the population divergence enabled this species to produce a higher number

of generations with contemporary levels of gene flow, which could erode the influence of historical factors. Moreover, morphological differentiation, in contrast to genetic variation, is widely considered adaptive and is rarely considered as historic phenotypes of the ancestral gene pools from colonization (Kremer et al., 2002; Gailing et al., 2012). The environment, rather than geography, was significantly associated with phenotypic differentiation in our study supported the effect of environmental isolation (Orsini et al., 2013).

Differences in environment between the two spatially and genetically distinct groups reflect differences in precipitation. Relative to the western group (Group W), the eastern group (Group E) is found in areas with greater mean annual precipitation, greater precipitation in the driest month, and lower precipitation seasonality. Consistent with this environmental differentiation, these two populations showed signs of phenotypical divergence that reflect adaptation to the major climate variable (i.e., precipitation). Seedlings from Group W had more below-ground biomass, a characteristic of plants adapted to drier areas, whereas seedlings from Group E had greater allocation to above-ground biomass, likely reflecting competition for growth. Biomass allocation traits were found to be significantly associated with precipitation, thus further corroborating the role of environment in driving the differentiation of these populations. In general, precipitation critically shapes many plant adaptive traits, including leaf size, branching architecture, and growth rate (Nicotra & Davidson, 2010; Moles et al., 2014). Several studies have found that species adapted to low or seasonal precipitation tend to show increased allocation to root mass (Mokany et al., 2006; Poorter et al., 2012). Furthermore, correlations between biomass and/or ecophysiological traits and climatic variables related to precipitation or temperature have also been reported (De Kort et al., 2014; Lind et al., 2017). For example, *Solidago virgaurea* L., a species distributed in dry serpentine habitats, shows higher resource allocation to below-ground biomass, a trait involved in local adaptation to dry sites (Sakaguchi et al., 2017). We detected 236 outliers that correlated with precipitation-related variables using bayenv2. Fifty-five annotated genes were identified based on the 119 most robust outlier SNPs, of which 12 genes were related to precipitation response. Therefore, precipitation was associated with both the morphological traits and the outlier SNPs, further supporting the notion that precipitation had driven the local adaptation of this species. The role of precipitation variables driving genetic adaptation in subtropical China has also been confirmed in the sympatric Fagaceae species pair *C. fargesii* (Li et al., 2014) and *Q. championii* (Jiang et al., 2019). In another study, the precipitation of the warmest quarter of the growing season explained a larger proportion of the variation of the geographic distribution of the valley oak *Q. lobata* in North America and was also found to lead to genetic differentiation (Gugger et al., 2013). Individual climatic variables varied with genetic and morphological variations in the three sympatric Californian oaks, suggesting a species-specific response. However, precipitation seasonality was an exception, which is only associated with genetic variations in all the three species (Riordan et al., 2016). Together, these studies indicate that precipitation is a potential common selective pressure for oak species. However, it is worth noting that many of the SNPs we identified were associated with temperature-related variables and several SNPs were related to soil organic carbon, soil pH, and global UV-B radiation. Manel et al. (2012) investigated 13 European alpine species and found that temperature and precipitation both strongly influenced the adaptive genetic variation in those species. Thus, temperature might also have a minor impact on the adaptive genetic variation in *Q. acutissima*.

In this study, many of the genes identified by our functional annotations of the outliers play roles in growth, development, and disease resistance. One of the detected outlier SNPs (SNP_ID: scaffold2960: 20721) was functionally assigned to a sugar transport protein (i.e., a sucrose transporter). This locus might therefore be involved in the biomass allocation patterns in plants. In plants, sucrose is important as an energy vector, a signaling molecule, and as a source material for the building of carbon skeletons (Lemoine, 2000). Sugar transport proteins mediate cell-to-cell and long-distance transport of sugars throughout the plant (Williams et al., 2000). Some functional genes have been identified in other plant species confirming the robustness of our outlier identification; for instance, the loci coded for glutathione S-transferase (*GST*) was also indicated to play a role in environmental adaptation and is significantly associated with bud phenology traits in oaks (Alberto et al., 2013; Sork et al., 2016b); the SNPs annotated as marking a *GTP*-binding protein, a *NAC* domain-containing protein, and *UDP*-glucosyltransferase in *Q. acutissima* have homologous genes in *Populus* L. (Geraldes et al., 2014). *NAC* genes are involved in drought and salt resistance as well as the regulation of drought-response genes (Hu et al., 2006; Lu et al., 2007), and have also been identified as linked to precipitation in genetic association studies of black spruce (Prunier et al., 2011).

5 Conclusions and Implications

Adaptive phenotypic traits are usually controlled by multiple genes, each with small effects spread across the genome (Le Corre & Kremer, 2012; Hendry, 2013). Detecting the actual loci linked to phenotypic variability from a genome-wide association study strongly depends on the extent of linkage disequilibrium and the density of markers along the genome. The RAD-seq technique is capable of uncovering a large number of markers throughout the genome. In this study, we sampled approximately 8% of the oak genome, a higher degree of coverage than other genetic studies on oak species undertaken to date (e.g., Eation et al., 2015; Deng et al., 2018; Ortego et al., 2018; Pina-Martins et al., 2019). Most importantly, 42.7% of the 55 361 SNPs identified in this study were located in genic regions, a finding rarely reported in other studies. This high-density dataset should provide a sufficient amount of information to assess the population structure and demographic history of populations of *Quercus acutissima*, and might also be sufficient to detect outlier loci in genic regions, thereby increasing the probability of inferring potentially adaptive variation. Another important advantage of this study is that we used a common garden experiment to detect adaptive

phenotypic divergence, enabling us to distinguish such divergence from phenotypic plasticity. Moreover, thus far, few studies (De Kort et al., 2014; Sork et al., 2016b; Vangestel et al., 2018) have combined genome-wide scans with common garden experiments, especially in forest trees with large genomes and longer generation times. To our knowledge, this study is the first to attempt to combine a genome scan and phenotype evaluation in a common garden in an Asian subtropical oak species; it will provide a valuable study system to compare with temperate oaks in North America and Europe. We also note the existence of a few caveats in the interpretation of our study, including the small sample size used to estimate phenotypic variation in each population, and the fact that measurements were taken for only 1 year. Although young seedling traits cannot fully reflect adult phenotypes, they still provide useful information on adaptation, and were used in the neo-tropical oak *Q. oleoides* to infer the evolutionary potential to climate change (Ramírez-Valiente et al., 2018) as well as in other studies (Horacio & Miguel, 2003; Savolainen et al., 2004). In the future, long-term observations of a large number of individuals will be carried out in the common garden experiment, which should provide us with additional information regarding adaptive variation in *Q. acutissima*. Thus far, the adaptive morphological traits and candidate genes discovered in *Q. acutissima* could help in effective artificial selection and screening of stock material for transplanting, thereby assisting both breeding programs and management decisions for predicting tree performance in response to climate change.

Acknowledgements

Conflict of Interest

The authors declared that there is no conflict of interest.

References

Alberto FJ, Derory J, Boury C, Frigerio JM, Zimmermann NE, Kremer A. 2013. Imprints of natural election along environmental gradients in phenology-related genes of *Quercus petraea*. *Genetics* 195: 495–512.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655–1664.

An ZS, Wu GX, Li JP, Sun YB, Liu YM, Zhou WJ, Cai YJ, Duan AM, Li L, Mao JY, Cheng H, Shi ZG, Tan LC, Yan H, Ao H, Chang H, Feng J. 2015. Global monsoon dynamics and climate change. *Annual Review of Earth and Planetary Sciences* 43: 29–77.

Bai WN, Liao WJ, Zhang DY. 2010. Nuclear and chloroplast DNA phylogeography reveal two refuge areas with asymmetrical gene flow in a temperate walnut tree from East Asia. *New Phytologist* 188: 892–901.

Baird NA, Etter PD, Atwood TS, Gurrey MG, Shiver AL, Lewis ZA, Sellker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3: e3376.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Borcard D, Legendre P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbor matrices. *Ecological Modelling* 153: 51–68.

Brachi B, Villoutreix R, Faure N, Hautekèete N, Piquot Y, Pauwels M, Roby D, Cuguen J, Bergelson J, Roux F. 2013. Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in *Arabidopsis thaliana*. *Molecular Ecology* 22: 4222–4240.

Bradburd GS, Coop GM, Ralph PL. 2018. Inferring continuous and discrete population genetic structure across space. *Genetics* 210: 33–52.

Bragg JG, Supple MA, Andrew RL, Borevitz JO. 2015. Genomic variation across landscapes: Insights and applications. *New Phytologist* 207: 953–967.

Cannon CH, Brendel O, Deng M, Hipp AL, Kremer A, Kua CS, Plomion C, Romero-Severson J, Sork VL. 2018. Gaining a global perspective on Fagaceae genomic diversification and adaptation. *New Phytologist* 218: 894–897.

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. *Journal of Biogeography* 38: 962–981.

Chen C, Liu ZG, Pan Q, Chen X, Wang HH, Guo HK, Liu SD, Lu HF, Tian SL, Li RQ, Shi W. 2016. Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinisxinyuan* n. ssp. *Molecular Biology and Evolution* 33: 1337–1348.

Chen DM, Zhang XX, Kang HZ, Sun X, Yin S, Du HM, Yamanaka N, Gapare W, Wu HX, Liu CJ. 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 JY, Zhou XD. 1993. Relationship between palaeoenvironmental change and soil development in Wuyi Mountains. *Geology of Fujian* 12: 228–231.

Clift PD, Hodges KV, Heslop D, Hannigan R, Long HV, Calves G. 2008. Correlation of Himalayan exhumation rates and Asian monsoon intensity. *Nature Geoscience* 1: 875–880.

Committee of Chinese Academy of Sciences for Physical Geography of China. 1985. Physical geography of China: Pandect. Beijing: Science Press.

Coop G, Witonsky D, Di RA, Pritchard JK. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185: 1411–1423.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

De Kort H, Vandepitte K, Bruun HH, Closset-Kopp D, Honnay O, Mergeay J. 2014. Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. *Molecular Ecology* 23: 4709–4721.

DeGiorgio M, Jakobsson M, Rosenberg NA. 2009. Out of Africa: Modern human origins special feature: Explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. *Proceedings of the National Academy of Sciences USA* 106: 16057–16062.

Deng M, Jiang XL, Hipp AL, Manos PS, 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. *Molecular Phylogenetics and Evolution* 119: 170–181.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43: 491–498.

Dewoody J, Trewin H, Taylor G. 2015. Genetic and morphological differentiation in *Populus nigra* L.: Isolation by colonization or isolation by adaptation? *Molecular Ecology* 24: 2641–2655.

Du FK, Hou M, Wang WT, Mao KS, 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. *Journal of Biogeography* 44: 294–307.

Eation DA, Hipp AL, 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.

Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genetics* 9: e1003905.

Excoffier L, Foll M. 2011. Fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27: 1332–1334.

Excoffier L, Hofer T, Foll M. 2009. Detecting loci under selection in a hierarchically structured population. *Heredity* 103: 285–298.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.

Fraser DJ, Weir LK, Bernatchez L, Hansen MM, Taylor EB. 2011. Extent and scale of local adaptation in salmonid fishes: Review and meta-analysis. *Heredity* 106: 404–420.

Friedman J, Barrett SC. 2009. Wind of change: New insights on the ecology and evolution of pollination and mating in wind-pollinated plants. *Annals of Botany* 103: 1515–1527.

Gailing O, Lind J, Lilleskov E. 2012. Leaf morphological and genetic differentiation between *Quercus rubra* L. and *Q. ellipsoidalis* E. J. Hill populations in contrasting environments. *Plant Systematics and Evolution* 298: 1533–1545.

Gao J, Wang BS, Mao JF, Ingvarsson P, Zeng QY, Wang XR. 2012. Demography and speciation history of the homoploid hybrid pine *Pinus densata* on the Tibetan Plateau. *Molecular Ecology* 21: 4811–4827.

Geraldes A, Farzaneh N, Grassa CJ, McKown A, Guy RD, Mansfield SD, Douglas CJ, Cronk QCB. 2014. Landscape genomics of *Populus trichocarpa*: The role of hybridization, limited gene flow,

and natural selection in shaping patterns of population structure. *Evolution* 68: 3260–3280.

Gharehaghaji M, Minor ES, Ashley MV, Abraham ST, Koenig WD. 2017. Effects of landscape features on gene flow of valley oaks (*Quercus labata*). *Plant Ecology* 218: 487–499.

Gonda A, Herczeg G, Merila J. 2011. Population variation in brain size of nine-spined sticklebacks (*Pungitius pungitius*) local adaptation or environmentally induced variation? *BMC Evolutionary Biology* 11: 1–11.

Gugger PE, Cokus SJ, Sork V. 2016. Association of transcriptome-wide sequence variation with climate gradients in valley oak (*Quercus lobata*). *Tree Genetics & Genomes* 12: 15.

Gugger PF, Ikegami M, Sork V. 2013. Influence of late Quaternary climate change on present patterns of genetic variation in valley oak, *Quercus lobata* Née. *Molecular Ecology* 22: 3598–3612.

Günther T, Coop G. 2013. Robust identification of local adaptation from allele frequencies. *Genetics* 195: 205–220.

Hendry AP. 2013. Key questions in the genetics and genomics of eco-evolutionary dynamics. *Heredity* 111: 456–466.

Henn BM, Cavalli-Sforza LL, Feldman MW. 2012. The great human expansion. *Proceedings of the National Academy of Sciences USA* 109: 17758–17764.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405: 907–913.

Hijimans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.

Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd GS, Lowry DB, Poss ML, Reed LK, Storfer A, Whitlock MC. 2016. Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *The American Naturalist* 188: 379–397.

Hoffmann AA, Sgro CM. 2011. Climate change and evolutionary adaptation. *Nature* 470: 479–485.

Horacio P, Miguel MR. 2003. Seed mass and seedling performance within eight species of *Psychotria* (Rubiaceae). *Ecology* 84: 439–450.

Hu HH, Dai MQ, Yao JL, Xiao BZ, Li XH, Zhang QF, Xiong LZ. 2006. Overexpressing a *NAM*, *ATAF*, and *CUC* (*NAC*) transcription factor enhances drought resistance and salt tolerance in rice. *Proceedings of the National Academy of Sciences USA* 103: 12987–12992.

Iida Y, Kohyama TS, Swenson NG, Su SH, Chen CT, Chiang JM, Sun IF. 2014. Linking functional traits and demographic rates in a subtropical tree community: The importance of size dependency. *Journal of Ecology* 102: 641–650.

Jiang XL, An M, Zheng SS, Deng M, Su ZH. 2018. Geographical isolation and environmental heterogeneity conribute to the spatial genetic patterns of *Quercus kerrii* (Fagaceae). *Heredity* 120: 219–233.

Jiang XL, Gardner EM, Meng HH, Deng M, Xu GB. 2019. Land bridges in the Plestocene contributed to flora assembly on the continental islands of South China: Insights from the evolutionary history of *Quercus championii*. *Molecular Phylogenetics and Evolution* 132: 36–45.

Kremer A, Kleinschmit J, Cottrell J, Cundall EP, Deans JD, Ducousso A, König AO, Lowe AJ, Munro RC, Petit RJ, Stephan BR. 2002. Is there a correlation between chloroplastic and nuclear divergence, or what are the roles of history and selection on genetic diversity in European oaks? *Forest Ecology and Management* 156: 75–87.

www.jse.ac.cn

J. Syst. Evol. 59(3): 541–556, 2021

Kremer A, Potts BM, Delzon S. 2014. Genetic divergence in forest trees: Understanding the consequence of climate change. *Functional Ecology* 28: 22–36.

Le Corre V, Kremer A. 2012. The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology* 21: 1548–1566.

Legendre P, Gallagher E. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129: 271–280.

Leimu R, Fischer M. 2008. A meta-analysis of local adaptation in plants. *PLoS One* 3: e4010.

Lemoine R. 2000. Sucrose transporters in plants: Update on function and structure. *Biochimica et Biophysica Acta – Biomembranes* 1465: 246–262.

Li C, Sun Y, Huang HW, Cannon CH. 2014. Footprints of divergent selection in natural populations of *Castanopsis fargesii* (Fagaceae). *Heredity* 113: 533–541.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: 1303.3997.

Li Y, Zhang XW, Fang YM. 2016. Responses of the distribution pattern of *Quercus chenii* to climate change following the Last Glacial Maximum. *Chinese Journal of Plant Ecology* 40: 1164–1178.

Lind BM, Friedline CJ, Wegrzyn JL, Maloney PE, Vogler DR, Neale DB, Eckert AJ. 2017. Water availability drives signatures of local adaptation in whitebark pine (*Pinus albicaulis* Engelm.) across fine spatial scales of the Lake Tahoe Basin, USA. *Molecular Ecology* 26: 3168–3185.

Liu JQ, Sun YS, Ge XJ, Gao LM, Qiu YX. 2012. Phylogeographic studies of plants in China: Advances in the past and directions in the future. *Journal of Systematics and Evolution* 50: 267–275.

Liu ZL, Yu MK, Ma Y, Tang LZ, Fang SZ. 2011. A trend surface analysis of geographic variation in the triats of seeds and seedlings from different *Quercus acutissima* provenances. *Acta Ecologica Sinica* 31: 6796–6804. (in Chinese with English abstract)

Lu PL, Chen NZ, An R, Su Z, Qi BS, Ren F, Chen J, Wang XC. 2007. A novel drought inducible gene, *ATAF1*, encodes a NAC family protein that negatively regulates the expression of stress-responsive genes in *Arabidopsis*. *Plant Molecular Biology* 63: 289–305.

Luo D, Xiao YF, Zhan L, Zhang L. 2010. The tectogenesis control of Danxia landform in southeast China. *Journal of East China Institute of Technology* 33: 147–153.

Luu K, Bazin E, Blum MG. 2017. *pcadapt*: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources* 17: 67–77.

Manel S, Gugerli F, Thuiller W, Alvarez N, Legendre P, Holderegger R, Gielly L, Taberlet P, Consortium I. 2012. Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular Ecology* 21: 3729–3738.

Meng HH, Su T, Gao XY, Li J, Jiang XL, Sun H, Zhou ZK. 2017. Warm-cold colonization: Response of oaks to uplift of the Himalaya-Hengduan Mountains. *Molecular Ecology* 26: 3276–3294.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* 17: 240–248.

Mokany K, Raison RJ, Prokushkin A. 2006. Critical analysis of root: Shoot ratios in terrestrial biomes. *Gobal Change Biology* 12: 84–96.

Moles AT, Perkins SE, Laffan SW, Flores-Moreno H, Awasthy M, Tindall ML, Sack L, Pitman A, Kattge J, Aarssen LW, Anand M, Bahn M, Blonder B, Cavender-Bares J, Cornelissen JHC, Cornwell WK, Díaz S, Dickie JB, Freschet GT, Griffiths JG, Gutierrez AG, Hemmings FA, Hickler T, Hitchcock TD, Keighery M, Kleyer M, Kurokawa H, Leishman MR, Liu K, Niinemets Ü, Onipchenko V, Onoda Y, Penuelas J, Pillar VD, Reich PB, Shiodera S, Siefert A, Sosinski EE, Soudzilovskaia NA, Swaine EK, Swenson NG, van Bodegom PM, Warman L, Weiher E, Wright IJ, Zhang H, Zobel M, Bonser SP. 2014. Which is a better predictor of plant traits: Temperature or precipitation? *Journal of Vegetation Science* 25: 1167–1180.

Moritsuka E, Hisataka Y, Tamura M, Uchiyama K, Watanabe A, Tsumura Y, Tachida H. 2012. Extended linkage diseuilibrium in noncoding regions in a conifer, *Cryptomeria japonica*. *Genetics* 190: 1145–1148.

Nicotra AB, Davidson A. 2010. Adaptive phenotypic plasticity and plant water use. *Functional Plant Biology* 37: 117–127.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O' Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2016. Vegan: Community ecology package. R package version 2.4-1.

Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L. 2013. Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology* 22: 5983–5999.

Ortego J, Gugger PF, Sork VL. 2015. Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak. *Journal of Biogeography* 42: 328–338.

Ortego J, Gugger PF, Sork VL. 2016. Impacts of human-induced environmental disturbances on hybridization between two ecologically differentiated Californian oak species. *New Phytologist* 213: 942–955.

Ortego J, Gugger PF, Sork VL. 2018. Genomic data reveal cryptic lineage diversification and introgression in Californian golden cup oaks (section *Protobalanus*). *New Phytologist* 218: 804–818.

Papadopoulou A, Knowles L. 2015. Species-specific responses to island connectivity cycles: Refined models for testing phylogeographic concordance across a Mediterranean Pleistocene Aggregate Island Complex. *Molecular Ecology* 24: 4252–4268.

Parchman TL, Gompert Z, Mudge J, Schilkey FD, Benkman CW, Buerkle CA. 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology* 21: 2991–3005.

Parchman TL, Jahner JP, Uckele KA, Galland LM, Eckert AJ. 2018. RADseq approaches and application for forest tree genetics. *Tree Genetics & Genomes* 14: 39.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genetics* 2: e190.

Paz-Vinas I, Qumr E, Chikhi L, Loot G, Blanchet S. 2013. The demographic history of populations experiencing asymmetric gene flow: Combining simulated and empirical data. *Molecular Ecology* 22: 3279–3291.

Petit RJ, Carlson J, Curtu AL, Loustau ML, Plomion C, González-Rodríguez A, Sork V, Ducousso A. 2013. Fagaceae trees as models to integrate ecology, evolution and genomics. *New Phytologist* 197: 369–371.

Pina-Martins F, Baptista J, Pappas G, Jr, Paulo OS. 2019. New insights on adaptation and population structure of cork oak using genotyping by sequencing. *Gobal Change Biology* 25: 337–350.

Poorter H, Niklas KJ, Reich PB, Oleksyn J, Poot P, Mommer L. 2012. Biomass allocation to leaves, stems and roots: Meta-analyses of interspecific variation and environmental control. *New Phytologist* 193: 30–50.

Prunier J, Laroche J, Beaulieu J, Bousquet J. 2011. Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Molecular Ecology* 20: 1702–1716.

Prunier J, Verta JP, MacKay JJ. 2016. Conifer genomics and adaptation: At the crossroads of genetic diversity and genome function. *New Phytologist* 209: 44–62.

Qiu YX, Fu CX, Comes HP. 2011. Plant molecular phylogeography in China and adjacent regions: Tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. *Molecular Phylogenetics and Evolution* 59: 225–244.

R Development Core Team. 2016. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza L. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences USA* 102: 15942–15947.

Ramírez-Valiente J, Etterson JR, Deacon NJ, Cavender-Bares J. 2018. Evolutionary potential varies across populations and traits in the neotropical oak *Qercus oleoides*. *Tree Physiology* 39: 427–439.

Rellstab C, Zoller S, Walthert L, Lesur I, Pluess AR, Graf R, Bodenes C, Sperisen C, Kremer A, Gugerli F. 2016. Signature of local adaptation in candidate genes of oaks (*Quercus* spp.) with respect to present and future climatic conditions. *Molecular Ecology* 25: 5907–5924.

Riordan EC, Gugger PF, Ortego J, Smith C, Gaddis K, Thompson P, Sork V. 2016. Association of genetic and phenotypic variability with geography and climate in three southern California oaks. *American Journal of Botany* 103: 73–85.

Sakaguchi S, Horie K, Ishikawa N, Nagano AJ, Yasugi M, Kudoh H, Ito M. 2017. Simultaneous evaluation of the effects of geographic environmental and temporal isolation in ecotypic populations of *Solidago vigaurea*. *New Phytologist* 216: 1268–1280.

Savolainen O, Bokma F, Garcia-Gil R, Komulainen P, Repo T. 2004. Genetic variation in cessation of growth and frost hardiness and consequences for adaptation of *Pinus sylvestris* to climatic changes. *Forest Ecology and Management* 197: 79–89.

Savolainen O, Lascoux M, Merila J. 2013. Ecological genomics of local adaptation. *Nature Reviews Genetics* 14: 807–820.

Savolainen O, Pyhäjärvi T, Knurr T. 2007. Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution & Systematics* 38: 595–619.

Shi MM, Michalski SG, Welk E, Chen XY, Durka W. 2014. Phylogeography of a widespread Asian subtropical tree: Genetic east-west differentiation and climate envelope modeling suggest multiple glacial refugia. *Journal of Biogeography* 41: 1710–1720.

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

Simeone MC, Grimm GW, Papini A, Vessella F, Cardoni S, Tordoni E, Piredda R, Franc A, Denk T. 2016. Plastome data reveal multiple geographic origins of *Quercus* group Ilex. *PeerJ* 4: e1897.

Siol M, Wright SI, Barrett SCH. 2010. The population genomics of plant adaptation. *New Phytologist* 188: 313–332.

Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB. 2013. Putting the landscape into the genomics of trees: Approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes* 9: 901–911.

Sork VL, Fitz-Gibbon ST, Puiu D, Crepeau M, Gugger PF, Sherman R, Stevens K, Langley CH, Pellegrini M, Salzberg SL. 2016a. First draft assembly and annotation of the genome of a California endemic oak *Quercus lobata* Née (Fagaceae). *G3: Genes, Genomes, Genetics* 11: 3485–3495.

Sork VL, Squire K, Gugger PF, Steele SE, Levy ED, Eckert AJ. 2016b. Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, *Quercus lobata*. *American Journal of Botany* 103: 33–46.

Storey JD, Bass AJ, Dabney A, Robinson D. 2015. Qvalue: Q-value estimation for false discovery rate control. R package, version 2.2.2.

Strasburg JL, Rieseberg LH. 2010. How robust are "isolation with migration" analyses to violations of the IM model? A simulation study. *Molecular Biology & Evolution* 27: 297–310.

Sun Y, Hu HQ, Huang HW, Vargas-Mendoza CF. 2014. Chloroplast diversity and population differentiation of *Castanopsis fargesii* (Fagaceae): A dominant tree species in evergreen broad-leaved forest of subtropical China. *Tree Genetics & Genomes* 10: 1531–1539.

Sun Y, Surget-Groba Y, Gao SX. 2016. Divergence maintained by climatic selection despite recurrent gene flow: A case study of *Castanopsis carlesii* (Fagaceae). *Molecular Ecology* 25: 4580–4592.

Tang GB, Zhang JP, Yang XD, Luo BX, Wang YZ. 1994. Late Cenozoic palynoflora and environment changes in Yunnan-Guizhou Plateau. *Marine Geology & Quaternary Geology* 14: 91–104.

Thomé MTC, Carstens BC. 2016. Phylogeographic model selection leads to insight into the evolutionary history of four-eyed frogs. *Proceedings of the National Academy of Sciences USA* 113: 8010–8017.

Tomlinson KW, Sterck FJ, Bongers F, da Silva DA, Barbosa ERM, Ward D, Bakker FT, Kaauwen MV, Prins HHT, de Bie S, Langevelde FV. 2012. Biomass partitioning and root morphology of savanna trees across a water gradient. *Journal of Ecology* 100: 1113–1121.

Vangestel C, Eckert AJ, Wegrzyn JL, Clair JBS, Neale DB. 2018. Linking phenotype, genotype and environment to unravel genetic components underlying cold hardiness in coastal Douglas-fir (*Pseudotsuga menziesii* var. *menziesii*). *Tree Genetics & Genomes* 14: 10.

Vitelli M, Vessella F, Cardoni S, Pollegioni P, Denk T, Grimm GW, Simeone MC. 2017. Phylogeographic structuring of plastome diversity in Mediterranean oaks (*Quercus* group Ilex, Fagaceae). *Tree Genetics & Genomes* 13: 3.

Williams LE, Lemoine R, Sauer N. 2000. Sugar transporters in higher plants- a diversity of roles and complex regulation. *Trends in Plant Science* 5: 283–290.

Wright S. 1943. Isolation by distance. *Genetics* 38: 114–138.

Xia H, Wang B, Zhao W, Pan J, Mao JF, Wang XR. 2018. Combining mitochondrial and nuclear genome analyses to dissect the effects of colonization, environments and geography on population structure in *Pinus tabuliformis*. *Evolutionary Application* 11: 1931–1945.

Xu J, Deng M, Jiang XL, Westwood M, Song YG, 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 Genetics & Genomes* 11: 805.

Xu J, Jiang XL, Deng M, Westwood M, Song YG, Zheng SS. 2016. Conservation genetics of rare trees restricted to subtropical montane cloud forests in southern China: A case study from *Quercus arbutifolia* (Fagaceae). *Tree Genetics & Genomes* 12: 90.

www.jse.ac.cn

J. Syst. Evol. 59(3): 541–556, 2021

Xu SY, Gao YX. 1962. The East Asia monsoon and season. In: Gao YX ed. *Several issues of the East Asia monsoon*. Beijing: Science Press. 88–103.

Yan F, Zhou WW, Zhao HT, Yuan ZY, Wang YY, Jiang K, Jin JQ, Murphy RW, Che J, Zhang YP. 2013. Geological events play a larger role than Pleistocene climatic fluctuation in driving the genetic structure of *Quasipaa boulengeri* (Anura: Dicroglossidae). *Molecular Ecology* 22: 1120–1133.

Yan HF, Zhang CY, Wang FY, Hu CM, Ge XJ, Hao G. 2012. Population expanding with the phalanx model and lineage split by environmental heterogeneity: A case study of *Primula obconica* in subtropical China. *PLoS One* 7: e41315.

Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, Nurkowski KA, Smets P, Wang T, Gray LK, Liepe KJ, Hamann A, Holliday JA, Whitlock MC, Rieseberg LH, Aitken SN. 2016. Convergent local adaptation to climate in distantly related conifers. *Science* 353: 1431–1433.

Zeng YF, Wang WT, Liao WJ, Wang HF, Zhang DY. 2015. Multiple glacial refugia for cool-temperate deciduous trees in northern East Asia: The Mongolian oak as a case study. *Molecular Ecology* 24: 5676–5691.

Zhang XW, Li Y, Liu CY, Xia T, Zhang Q, Fang YM. 2015. Phylogeography of the temperate tree species *Quercus acutissima* in China: Inferences from chloroplast DNA variations. *Biochemical Systematics & Ecology* 63: 190–197.

Zhang XW, Li Y, Zhang Q, Fang YM. 2018. Ancient east-west divergence, recent admixture, and multiple marginal refugia shape genetic structure of a widespread oak species (*Quercus acutissima*) in China. *Tree Genetics & Genomes* 14: 88.

Zhang YY, Fang YM, Yu MK, Li XX, Xia T. 2013. Molecular characterization and genetic structure of *Quercus acutissima* germplasm in China using microsatellites. *Molecular Biology Reports* 40: 4083–4090.

Supplementary Material

The following supplementary material is available online for this article at http://onlinelibrary.wiley.com/doi/10.1111/jse.12568/suppinfo:

Fig. S1. Histogram of restriction site-associated DNA sequencing (RAD-seq) data quality. **A,** the percentage of the loci achieved the Q30 (%); **B,** the distribution of reserved reads ratio (%); **C,** the genome coverage (Mbp); **D, E,** the distribution of mean sequencing depth and missing rate (%). Red and blue vertical lines indicate the mean and median value of the distribution, respectively.

Fig. S2. Cross-validation results from K = 2 to K = 10 of the conStruct analysis. Green dots represent the non-spatial model; blue dots represent the spatial model.

Fig. S3. Genetic structure of the restriction site-associated DNA sequencing (RAD-seq) variation in 27 populations from conStruct. **A–C,** Results of the spatial model. **D–F,** Results of the non-spatial model. **A, D,** Maps of admixture proportions for K = 2. **B, E,** Corresponding layer-specific covariance curves for K = 2. **C, F,** Each cluster's relative contribution to the total covariance to each K value.

Fig. S4. Distribution of genetic variance by principal component analysis (PCA) for PC1 and PC3.

Fig. S5. Principal component analysis (PCA) of the morphological traits from the common garden. LMF, leaf mass fraction; RMF, root mass fraction; SMF, stem mass fraction; STRL, specific taproot length.

Table S1. Geographic locations, sample size (*N*), gene diversity (π), average heterozygosity per locus (*Het*), and environmental variables of the 27 *Quercus acutissima* populations.

Table S2. Sequencing quality information of each individual.

Table S3. Alpha D values based on the spatial model and the non-spatial model.

Table S4. Model composite likelihoods and Akaike information criterion model selection for eight alternative demographic models of *Quercus acutissima* from fastsimcoal2.

Table S5. Demography parameters inferred from the best model. Migration rates scaled according to population effective sizes (2Nm) are given forward in time. The 95% confidence intervals were generated from 100 bootstrapped datasets. Estimates of divergence and expansion times are given in years. Estimates of sizes (Ne) are given in the number of haploids.

Table S6. Number of putatively adapted single nucleotide polymorphisms detected using Bayenv2.

Table S7. Gene annotation identified using the outlier approach and by testing for association with environmental variables in *pcadapt* and Bayenv2.

Doc. S1. The protocol of the restriction site-associated DNA sequencing (RAD-seq) library preparation.

Data S1. Raw sequence reads and phenotype data, available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.q2bvq83fv