*Annals of Botany* **129**: 231–245, 2022
https://doi.org/10.1093/aob/mcab140, available online at www.academic.oup.com/aob

ANNALS OF BOTANY
Founded 1887

# Influence of Pliocene and Pleistocene climates on hybridization patterns between two closely related oak species in China

**Yao Li[1,2], Xingwang Zhang[3], Lu Wang[1], Victoria L. Sork[4,5], Lingfeng Mao[1,2,*] and Yanming Fang[1,*]**

*[1]Co-Innovation Center for Sustainable Forestry in Southern China, College of Biology and the Environment, Key Laboratory of State Forestry and Grassland Administration on Subtropical Forest Biodiversity Conservation, Nanjing Forestry University, Nanjing, Jiangsu 210037, China, [2]Laboratory of Biodiversity and Conservation, College of Biology and the Environment, Nanjing Forestry University, Nanjing, Jiangsu 210037, China, [3]School of Life Sciences, Huaibei Normal University, Huaibei, Anhui 235000, China, [4]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095-7239, USA and [5]Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095-1496, USA*
** For correspondence. E-mail maolingfeng2008@163.com; jwu4@njfu.edu.cn*

Received: 27 August 2021 Returned for revision: 31 October 21 Editorial decision: 12 November 2021 Accepted: 26 November 2021
Electronically published: 10 December 2021

• **Background and Aims** Contemporary patterns of genetic admixture reflect imprints of both ancient and recent gene flow, which can provide us with valuable information on hybridization history in response to palaeoclimate change. Here, we examine the relationships between present admixture patterns and past climatic niche suitability of two East Asian Cerris oaks (*Quercus acutissima* and *Q. chenii*) to test the hypothesis that the mid-Pliocene warm climate promoted while the Pleistocene cool climate limited hybridization among local closely related taxa.
• **Methods** We analyse genetic variation at seven nuclear microsatellites (1111 individuals) and three chloroplast intergenic spacers (576 individuals) to determine the present admixture pattern and ancient hybridization history. We apply an information-theoretic model selection approach to explore the associations of genetic admixture degree with past climatic niche suitability at multiple spatial scales.
• **Key Results** More than 70 % of the hybrids determined by Bayesian clustering analysis and more than 90 % of the individuals with locally shared chloroplast haplotypes are concentrated within a mid-Pliocene contact zone between ~30°N and 35°N. Climatic niche suitabilities for *Q. chenii* during the mid-Pliocene Warm Period [mPWP, ~3.264–3.025 million years ago (mya)] and during the Last Glacial Maximum (LGM, ~0.022 mya) best explain the admixture patterns across all *Q. acutissima* populations and across those within the ancient contact zone, respectively.
• **Conclusions** Our results highlight that palaeoclimate change shapes present admixture patterns by influencing the extent of historical range overlap. Specifically, the mid-Pliocene warm climate promoted ancient contact, allowing widespread hybridization throughout central China. In contrast, the Pleistocene cool climate caused the local extinction of *Q. chenii*, reducing the probability of interspecific gene flow in most areas except those sites having a high level of ecological stability.

**Key words:** China, genetic admixture pattern, hybridization, introgression, palaeoclimate, Pleistocene, Pliocene, *Quercus acutissima*, *Quercus chenii*.

## INTRODUCTION

Contemporary patterns of genetic admixture reflect imprints of both ancient and recent hybridization, which can provide us with valuable information on the temporal and spatial dynamics of a hybrid zone in response to climate change (Buggs, 2007; Hewitt, 2011; Taylor *et al*., 2014, 2015; Gompert *et al.*, 2017; Ryan *et al.*, 2018). At a geological time scale, sister species that have diverged for a long time may have experienced dramatic range shifts (Hewitt, 2004). They could produce a hybrid zone that potentially has moved hundreds to thousands of kilometres over its lifetime (Wang *et al.*, 2014; Zohren *et al.*, 2016; Wielstra, 2019). The historical movement of a hybrid zone would leave behind a molecular footprint because selectively neutral alleles are expected to introgress from a receding species into an advancing species (Currat *et al.*, 2008; Wielstra *et al.*, 2017). On the other hand, a particular hybrid zone may have formed more than once as climatic fluctuations could lead to repeated range contractions and expansions of sister lineages, causing their ranges to come into contact many times (Hewitt, 2004, 2011). The signatures of ancient introgression among species that have coexisted in the past can still be detected even if the ancestors or parental species have suffered local extinction (Kim *et al.*, 2018; Ortego *et al.*, 2018; Crowl *et al.*, 2020). Given that present admixture patterns contain memories of early times that were particularly affected by range dynamics, it is possible to use such information, together with other lines of evidence from fossil records and species distribution modelling, to reconstruct the hybridization history among closely related species within the context of palaeoclimate change (Hewitt, 2011; Gompert *et al.*, 2017; Wielstra, 2019).

Oaks (*Quercus*) are a classic system for studying natural hybridization and introgression in plants because they often exchange alleles without disrupting species integrity

(Cannon and Petit, 2020). Ongoing hybridization has been widely detected within most sections of the genus, such as sect. *Quercus* (Lepais *et al.*, 2009; Zeng *et al.*, 2011; Ortego *et al.*, 2014; Burge *et al.*, 2019), sect. *Lobatae* (Dodd and Afzal-Rafii, 2004; Peñaloza-Ramírez *et al.*, 2010; Moran *et al.*, 2012), sect. *Ilex* (Neophytou *et al.*, 2011) and sect. *Cyclobalanopsis* (An *et al.*, 2017). Genomic signals of historical hybridization can be preserved over long periods of time, which can be detected in the genomes of modern oak taxa (e.g. Crowl *et al.*, 2020). For oaks on different continents, the history of ancient gene exchange differs depending on both climatic and topographic contexts that influence range dynamics and demographic processes (Kremer and Hipp, 2020). In Europe and eastern North America, the temperate white oaks experienced extensive introgression when postglacial recolonization restored genetic contacts among populations from isolated refugia (Leroy *et al.*, 2017, 2020; Petit *et al.*, 2002*a*). This process created opportunities for cytoplasmic capture, allowing the range-wide spread of a small number of shared chloroplast haplotypes (Whittemore and Schaal, 1991; Petit *et al.*, 2002*b*). In contrast, the western North American oaks maintained a relatively stable distribution and only experienced local migration across glacial–interglacial cycles, thus leaving a patchy distribution of a diverse set of haplotypes (Grivet *et al.*, 2006; Gugger *et al.*, 2013). The stable climate also enabled the long-term persistence of populations and preservation of hybridization events. Numerous cases reported here come from allopatric species that have coexisted in the past, associated with range shifts occurring at a much early time (e.g. the Pliocene or Pleistocene; Maze, 1968; Manos *et al.*, 1999; McVay *et al.*, 2017; Kim *et al.*, 2018; Ortego *et al.*, 2018).

East Asia is one of the global hotspots of oak diversity, including more than 100 species placed into four sections (Kremer and Hipp, 2020). Previous studies have stressed the role of ancient and recent gene flow in shaping the present admixture pattern among native oaks (Zeng *et al.*, 2011; Tamaki and Okada, 2014; An *et al.*, 2017), but the impact of palaeoclimate on hybridization history remains unclear. Compared with Europe and North America, East Asia was less affected by the Pleistocene glaciations (Feng *et al.*, 2016). The relatively stable climate, interacting with complex landscapes, has contributed to the long-term *in-situ* survival of local populations across multiple refugia (Zeng *et al.*, 2015; Zhang *et al.*, 2015, 2018; Li *et al.*, 2019*a*). Considering this history, we expect that the current admixture patterns among East Asian oaks would reflect the genetic imprints of ancient hybridization affected by historical climate change, probably before the Quaternary.

In this study, we reconstruct the hybridization history between two East Asian Cerris oaks, *Quercus acutissima* and *Q. chenii*, to assess the influence of palaeoclimate on present admixture patterns. We chose to focus on these particular taxa because they are sister species estimated to have diverged ~30–10 million years ago (mya) (Simeone *et al.*, 2018; Hipp *et al.*, 2020). Such a long evolutionary history makes it feasible to examine the impact of pre-Quaternary climate change. Currently, *Q. acutissima* is among the dominant tree species of the East Asian warm temperate deciduous forests (Fujiwara and Harada, 2015), while *Q. chenii* exhibits a much narrower distribution in East China. Diagnostic morphological features include leaf and acorn size. Nevertheless, a variety of *Q. chenii* showing an intermediate phenotype was identified in their overlapping ranges, suggesting that putative hybrids may occur (Liu, 1992).

Fossil evidence indicates that *Q. chenii* was a common tree species during the mid-Pliocene Warm Period (mPWP, ~3.264–3.025 mya) at more northerly latitudes than its current range limit. However, it later suffered local extinction as a result of both Pleistocene cooling and strengthening of the East Asian winter monsoon (Momohara, 2010, 2016). The range dynamics of *Q. chenii* would affect the hybridization history by triggering contact zone shifts. Here, we predict that the mid-Pliocene warm climate promoted while the Pleistocene cool climate limited the hybridization by influencing the extent of historical range overlap. To test this hypothesis, we first analyse the genetic variation found in bi-parentally inherited nuclear microsatellites (nSSRs) and maternally inherited chloroplast (cp) DNA sequences to determine the putative hybrids and present admixture pattern. Second, we combine the evidence from fossil records, demographic inference and ecological niche modelling (ENM) to explore the role of ancient gene flow in shaping the hybridization history. Finally, we apply an information-theoretic model selection approach to examine the associations between genetic admixture degree and past climatic niche suitability. If ancient hybridization has left imprints on admixture patterns, we expect to detect a positive association between past climatic niche suitability and degree of genetic admixture because hybrids are more likely to occur in habitats that have been suitable for parental species (Ortego *et al.*, 2014).

## MATERIAL AND METHODS

### *Sampling*

Between May 2014 and September 2017, we sampled 30 populations of *Quercus acutissima* ($n = 696$ individuals) and 18 populations of *Quercus chenii* ($n = 415$ individuals), encompassing the majority of the distributions of both species in China (Fig. 1A and Supplementary Data Table S1). Leaf tissues were quickly dried with silica gel and stored at room temperature in the laboratory. Spatially explicit information was recorded for each population using a handheld GPS unit (Magellan, USA). Voucher specimens of all individuals were deposited in the Herbarium of Nanjing Forestry University (HNFU).

### *Ecological niche modelling*

We used MAXENT 3.4.1 (Phillips *et al.*, 2018) to estimate the distributions of *Q. acutissima* and *Q. chenii* for four periods whose climate data are available: the present, Last Glacial Maximum (LGM, ~ 0.022 mya), Marine Isotope Stage 19 (MIS19, ~ 0.787 mya, an analogue for the Holocene) in the Pleistocene and mPWP. Species occurrence records were collected from field observations as well as from specimen databases (Chinese Virtual Herbarium, http://www.cvh.ac.cn/; Specimen Resources Sharing Platforms for Education, http://mnh.scu.edu.cn/; Plant Photo Bank of China, http://ppbc.iplant.cn/). After removing duplicates and retaining only one

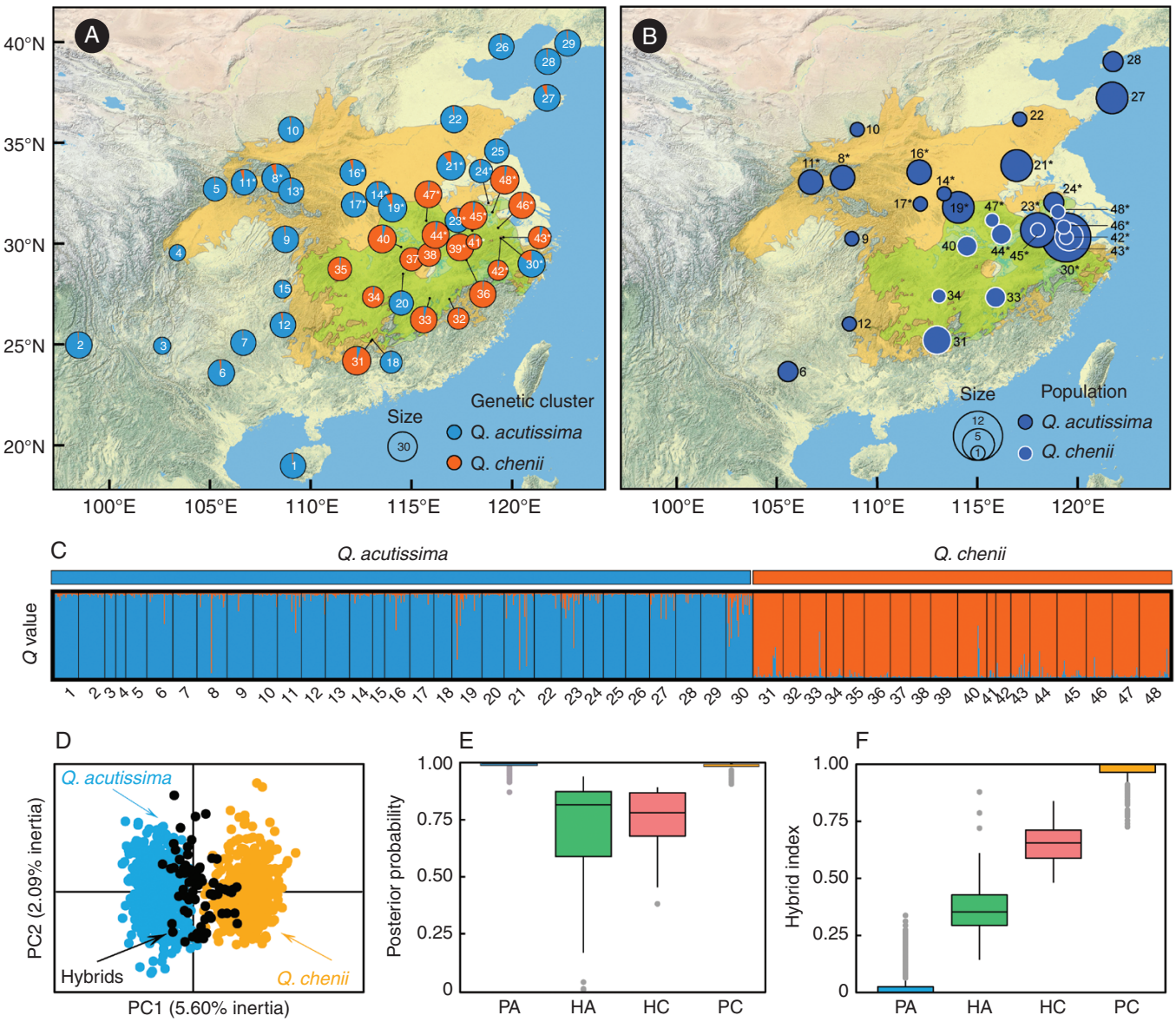


FIG. 1. Results of STRUCTURE, PCA, NewHybrids and INTROGRESSION analyses based on the genetic variation at seven nuclear microsatellite loci. (A) Geographical distributions of two genetic clusters corresponding to *Quercus acutissima* (blue) and *Q. chenii* (orange) as inferred by STRUCTURE analyses. Circle area is proportional to the sample size. (B) Number of putative hybrids at each site. In A and B, orange and green areas indicate the climatically suitable areas for *Q. chenii* during the mid-Pliocene Warm Period (mPWP) and during the present, respectively. They were created from the logistic outputs of MAXENT using the maximum sensitivity plus specificity threshold (Gugger *et al.*, 2013). Population codes (*Q. acutissima*: 1–30; *Q. chenii*: 31–48) are described in Supplementary Data Table S1. *Populations inside the ancient contact zone. (C) Histogram of an individual's probability of assignment to *Q. acutissima* (blue) and *Q. chenii* (orange) in STRUCTURE analyses. Each vertical bar represents one individual. (D) PCA result for *Q. acutissima*, *Q. chenii* and their hybrids. (E) Posterior probability (*PP*) assigned to *Q. acutissima* for parental individuals of *Q. acutissima* (group PA, $n = 642$) and hybrids morphologically assigned to *Q. acutissima* (group HA, $n = 54$), and to *Q. chenii* for parental individuals of *Q. chenii* (group PC, $n = 395$) and hybrids morphologically assigned to *Q. chenii* (group HC, $n = 20$). The *PP* values are averaged over ten independent runs of NewHybrids analyses. (F) Maximum-likelihood hybrid index (*H*) estimated by the R package introgression for the PA, HA, HC and PC groups.

record among all locations falling within the same 2.5′ × 2.5′ grid, a total of 457 and 54 presence points were obtained for *Q. acutissima* and *Q. chenii*, respectively. We selected four bioclimatic variables that show low to moderate correlations ($|r| < 0.7$) to prevent potential over-fitting: annual mean temperature (bio1), temperature seasonality (bio4), annual precipitation (bio12) and precipitation seasonality (bio15). Seasonal temperature extremes were not chosen because they are not available for the mPWP and MIS19. Environmental layers of the four variables were retrieved from the WorldClim 1.4 (Hijmans *et al.*, 2005) and PaleoClim (Brown *et al.*, 2018) databases at a grid resolution of 2.5 arc-minutes. Palaeoclimate data were generated using the downscaled outputs of general circulation models, including the Community Climate System Model version 4 (CCSM4) and the Hadley Centre Coupled Model version 3 (HadCM3; Hill, 2015). All the layers were clipped to the same spatial range (15°–45°N, 90°–145°E) using the package raster 2.8-19 (Hijmans, 2019) in R (R Core Team, 2018). ENM

analyses were performed with default settings. Model performance was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC) estimated by 10-fold cross-validation. Niche suitability maps were created based on the logistic outputs of MAXENT.

### *DNA extraction, sequencing and microsatellite genotyping*

Total genomic DNA was extracted from 30 mg leaf tissue of each individual using the Tiangen Plant Genomic DNA Kit (Tiangen, Beijing, China). Three chloroplast intergenic spacers (*atp*B-*rbc*L, *psb*A-*trn*H, *trn*S-*trn*G; Supplementary Data Table S2) were sequenced for 401 individuals of *Q. acutissima* and 175 individuals of *Q. chenii* following the protocols of Zhang *et al.* (2015). All the 1111 samples were genotyped using seven nuclear microsatellite markers (Supplementary Data Table S3) previously developed for other oak species. Polymerase chain reactions (PCRs) and genotyping were performed as described by Zhang *et al.* (2018). Deviation from Hardy–Weinberg equilibrium (HWE), linkage disequilibrium between each locus pair, and the presence of null alleles was tested for each population following Zhang *et al.* (2018).

### *Nuclear microsatellite diversity and differentiation*

We estimated allelic richness ($A_R$) and genetic diversity within populations ($H_S$) for each population using Fstat 2.9.3.2 (Goudet, 2001). We calculated $F_{ST}$ and the standardized measure $G'_{ST}$ to determine the levels of genetic differentiation across all populations with MSA 4.05 (Dieringer and Schlötterer, 2003). Genetic statistics $A_R$, $H_S$ and $F_{ST}$ were compared between *Q. acutissima* and *Q. chenii* using 10 000 permutations with Fstat. We also performed an analysis of molecular variance (AMOVA) to examine the patterns of genetic variance partitioned among species, among populations within species and within populations using Arlequin 3.5 (Excoffier and Lischer, 2010). The significance of fixation indices ($F_{CT}$, $F_{SC}$ and $F_{ST}$) was tested by 10 000 permutations.

### *Genetic structure and hybrid identification*

To detect the underlying genetic structure and potential hybrids between the two species, we employed Bayesian clustering analyses. The analyses were performed using an admixture model with correlated allele frequencies in the program STRUCTURE 2.3.4 (Pritchard *et al.*, 2000). Ten independent runs were conducted for each number of clusters ($K$) varying from 1 to 10. In each run, we used 1 000 000 Markov chain Monte Carlo (MCMC) iterations following a burn-in step of 100 000 replications. Both log probabilities of $X|K$ (Pritchard *et al.*, 2000) and the $\Delta K$ method (Evanno *et al.*, 2005) were implemented in Structure Harvester (Earl and vonHoldt, 2012) to determine the best-supported number of clusters (Janes *et al.*, 2017). Optimal alignments of inferred clusters across different replicates and different values of $K$ were identified by the CLUMPAK pipeline (Kopelman *et al.*, 2015). We expected that the most likely number of clusters was $K = 2$, and the two clusters corresponded to *Q. acutissima* and *Q. chenii*. In this case, individuals with a probability of membership ($Q$) to the cluster of the morphologically inferred parental species (i.e. $Q_{ACU}$ for *Q. acutissima*, $Q_{CHE}$ for *Q. chenii*) ≥ 0.9 were classified as parental individuals, while those with $Q < 0.9$ were considered to be hybrids (Burgarella *et al.*, 2009; Lyu *et al.*, 2018).

Complementary to Bayesian clustering analyses, we applied three different methods to validate the results of STRUCTURE. First, we performed principal component analysis (PCA) using the R package ade4 1.7-13 (Dray and Dufour, 2007). PC scores on the first axis were compared between parental individuals and hybrids using the pairwise Wilcoxon rank-sum test. $P$-values were adjusted through the Holm–Bonferroni method (Holm, 1979). Second, we employed a Bayesian statistical approach implemented in NewHybrids 1.1 beta (Anderson and Thompson, 2002) to distinguish parental individuals from hybrids. According to the 3rd criterion of Burgarella *et al.* (2009), only individuals with posterior probability ($PP$) assigned to the parental species class ≥0.9 were classified as parental individuals. This analysis was repeated ten times with a 50 000-iteration burn-in followed by 150 000 steps. The Jeffreys-like priors were selected for both the mixing proportions and the allele frequencies. Finally, we calculated the maximum-likelihood hybrid index ($H$) for each individual using the R package introgress 1.2.3 (Gompert and Buerkle, 2010). This index is an estimate of the proportion of alleles that were inherited from one of the two parental species (Buerkle, 2005). Parental groups identified by the Bayesian clustering analyses were specified as parental populations. Members of these groups should have a value of $H$ near 0 or 1.

### *Gene flow and demographic history*

We inferred the temporal pattern of interspecific gene flow under an approximate Bayesian computation (ABC) framework. Four distinct demographic models were statistically compared: strict isolation (SI), isolation with migration (IM), ancient migration (AM) and secondary contact (SC) (Fig. 2). The SI model assumed that no gene flow occurred after the initial divergence of the two species at $T_{DIV}$, whereas the IM model assumed that they continuously exchanged alleles after $T_{DIV}$. Under the AM model, gene flow occurred after $T_{DIV}$ but stopped at $T_{AM}$. Under the SC model, gene flow occurred after $T_{SC}$ following an early stage of strict isolation. For each model, we assumed that both the ancestral and the daughter populations had constant effective population sizes ($N_{ACU}$, $N_{CHE}$, $N_{ANC}$) sampled from a log-uniform distribution between 10 and 1 000 000. A previous oak phylogenomic study has indicated that *Q. acutissima* and *Q. chenii* diverged ~30–10 mya (Hipp *et al.*, 2020). Given an average generation time of 100 years for both species (Zeng *et al.*, 2015), we used a uniform distribution from 100 000 to 300 000 for $T_{DIV}$, and a log-uniform distribution from 1 to 300 000 for $T_{AM}$ ($<T_{DIV}$) and $T_{SC}$ ($<T_{DIV}$). The effective migration rate ($4N_em$) was drawn from a uniform distribution between 0 and 500.

We used a generalized stepwise mutation (GSM) model for the seven microsatellite loci (Estoup *et al.*, 2002). This model included two parameters, mutation rate ($\mu$) and geometric parameter ($P_{GSM}$). We set mean $\mu$ across loci to $5 \times 10^{-4}$ mutations per locus per generation (Aoki *et al.*, 2019), and sampled individual locus $\mu$ from a Gamma distribution with *shape* and

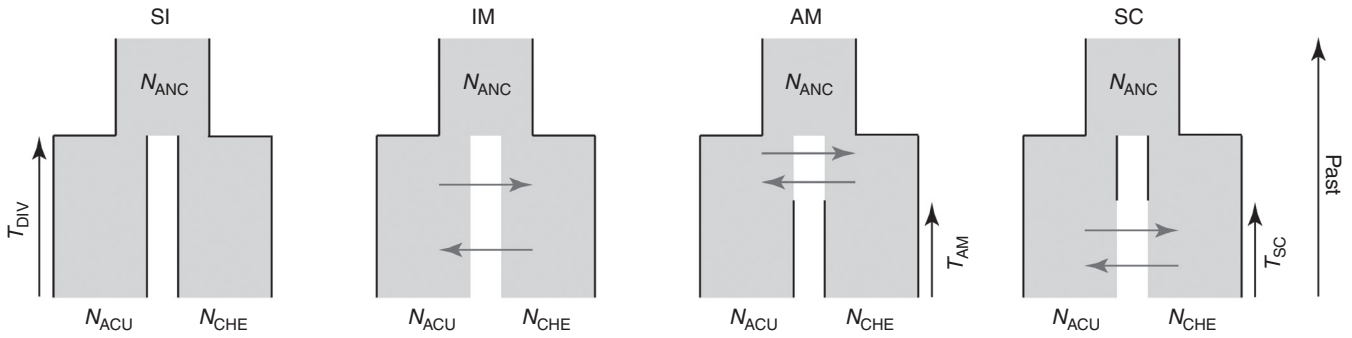


FIG. 2. Four demographic scenarios tested by approximate Bayesian computation (ABC): strict isolation (SI), isolation with migration (IM), ancient migration (AM) and secondary contact (SC). $N_{ACU}$, $N_{CHE}$ and $N_{ANC}$ are effective population sizes of *Quercus acutissima*, *Q. chenii* and their ancestor, respectively. $T_{DIV}$ is the number of generations since the initial divergence. $T_{AM}$ is the number of generations since the two species stopped exchanging alleles. $T_{SC}$ is the number of generations since the two species began exchanging alleles. Dark grey arrows denote the effective migration rate from *Q. acutissima* to *Q. chenii* ($4N_{CHE}m_{A\to C}$) and that in the opposite direction ($4N_{ACU}m_{C\to A}$).

*rate* parameters. The value of *shape* was drawn from a uniform distribution between 0.5 and 2, and the *rate* parameter was calculated by *shape*/mean $\mu$. We used a uniform distribution between 0 and 0.8 for mean $P_{GSM}$ and assumed individual locus $P_{GSM}$ followed a Beta distribution $\beta(a, b)$, where $a = 0.5 + 199 \times$ mean $P_{GSM}$ and $b = a \times (1 -$ mean $P_{GSM})$/mean $P_{GSM}$ (Excoffier *et al.*, 2005).

We simulated genotype data for 100 diploid individuals per population to reduce the computational costs. Model parameter values were sampled from the prior distributions using ABCsampler in ABCtoolbox (Wegmann *et al.*, 2010). One million replicates were run for each scenario with fastsimcoal2 2.6.0.3 (Excoffier and Foll, 2011; Excoffier *et al.*, 2013). Seventeen summary statistics were computed for both observed and simulated datasets in Arlsumstat 3.5.2 (Excoffier and Lischer, 2010). The observed dataset included 100 individuals randomly subsampled for each population, with sampling biases checked following Aoki *et al.* (2019). We employed both the multinomial logistic regression (MLR) approach in the R package abc 2.1 (Csilléry *et al.*, 2012) and the random forest (RF) approach in the R package abcrf 1.8.1 (Pudlo *et al.*, 2016) to estimate the posterior probability and classification error rate of each model. The goodness of fit (GoF) was assessed using the *gfit* function in the R package abc. The 1 % (i.e. 10 000) simulated datasets closest to the observed data were chosen to infer the posterior distributions of model parameters through a local linear regression approach implemented in the R package abc. Finally, posterior predictive checks were conducted by comparing the summary statistics of the observed dataset with those calculated for the 1000 new simulations generated from the model's posterior (Csilléry *et al.*, 2012).

### *Genetic admixture pattern*

We used an information-theoretic model selection approach to assess the relationships between genetic admixture and three sets of explanatory variables for all *Q. acutissima* populations and those within the ancient contact zone. The probability of membership to the genetic cluster of *Q. chenii* ($Q_{CHE}$) was used to measure the degree of admixture. The predictor matrices included: (1) presence of *Q. chenii* in a given locality (*Presence*; 0/1); (2) distance to the nearest occurrence record of *Q. chenii*; and (3) climatic niche suitability (*S*) for both parental species during the present, LGM, MIS19 and mPWP. We constructed a series of general linear models (GLMs) with a normal error structure and an identity link function. Each model included one or two predictors and excluded those variables showing strong multicollinearity problems (variance inflation factor > 2). We selected among competing models using Akaike's information criterion corrected for small sample size (AICc; Burnham and Anderson, 1998). Models with the lowest AICc and those with ΔAICc (AICc of a model – minimum AICc) ≤ 2 were considered equally plausible. We computed model-averaged estimates for each variable when multiple candidate models were supported. These analyses were performed at both population and individual levels using the R package AICcmodavg 2.2-2 (Mazerolle, 2019). To assess the modality of the frequency distributions of $Q_{ACU}$ values, we conducted Hartigans' dip test (Hartigan and Hartigan, 1985) using the R package diptest 0.75-7 (Martin, 2016).

### *Chloroplast DNA sequence analyses*

Sequences of the three cpDNA fragments were proofread, aligned and concatenated using BioEdit 7.2.5 (Hall, 1999) and FasParser 2.1.1 (Sun, 2017). Indels and inversions were coded as binary characters (Simmons and Ochoterena, 2000; Young and Healy, 2003). Chloroplast haplotypes were determined by DnaSP 5.10 (Librado and Rozas, 2009). A median-joining network was constructed with PopART 1.7 (Leigh and Bryant, 2015). Genetic diversity statistics were estimated for each species (average genetic diversity within populations, $h_S$; total genetic diversity, $h_T$) and each population (haplotype diversity, $H_d$) using Permut 2.0 (Pons and Petit, 1996) and DnaSP. The phylogeographical structure was assessed by testing the significance of the difference between two measures of population differentiation ($G_{ST}$ and $N_{ST}$) with 10 000 permutations in Permut. AMOVA was performed using Arlequin with 10 000 permutations.

## RESULTS

### *Ecological niche modelling*

The mean AUC values (± s.d.) on the test dataset indicated a good fit of ENMs to the observed occurrence data

of both species (*Q. acutissima*: 0.849 ± 0.021; *Q. chenii*: 0.968 ± 0.014; Supplementary Data Fig. S1). The models showed that *Q. acutissima* has had a relatively stable distribution since the mPWP, while *Q. chenii* has experienced more pronounced range shifts (Fig. 3). Furthermore, the two species have had a mid-Pliocene contact zone beyond the current northern range limit of *Q. chenii* (between ~30°N and 35°N; Fig. 3). Eleven *Q. acutissima* populations and nine *Q. chenii* populations were found to be within this region (Fig. 1A), most of which showed moderate to high mPWP climatic suitability for both species (*Q. acutissima*: $S_{mPWP}$ = 0.200 ± 0.089; *Q. chenii*: $S_{mPWP}$ = 0.806 ± 0.170). However, along with the following range contraction and southward retreat of *Q. chenii* (Fig. 3), this ancient contact zone may have disappeared during the cooler period of the Early Pleistocene.

### *Nuclear microsatellite diversity and differentiation*

The frequency of null alleles was estimated to be lower than the threshold of 0.1 at each of the seven loci across populations (Supplementary Data Table S3). No significant linkage disequilibrium was detected for all locus pairs in each population ($P < 0.05$ after Bonferroni correction), and no evidence of deviation from HWE was observed for each population of both species ($P < 0.05$ after Bonferroni correction; Supplementary Data Table S4). Higher within-population genetic diversity was detected in *Q. acutissima* compared to *Q. chenii* ($P = 0.001$ for both $A_R$ and $H_S$; Supplementary Data Table S5). Intraspecific genetic differentiation was significant but lower for both species (*Q. acutissima*: $F_{ST}$ = 0.057, $P < 0.001$; *Q. chenii*: $F_{ST}$ = 0.057, $P < 0.001$), and the levels did not differ substantially between the two taxa ($P$ = 0.983; Supplementary Data Table S5). Interspecific differentiation was moderate and significant ($F_{CT}$ = 0.146, $P < 0.00001$), with 14.6 % of the variation partitioned among species, and the most (80.5 %) partitioned within populations (Supplementary Data Table S6).

### *Genetic structure and hybrid identification*

Bayesian clustering analyses indicated an optimal value of $K$ = 2 (Supplementary Data Fig. S2), separating all populations into two distinct genetic clusters congruent with the observed phenotypes of *Q. acutissima* and *Q. chenii* (Fig. 1C). Considering a threshold value of $Q$ = 0.9, we identified 74 putative hybrids of which 54 were morphologically assigned to *Q. acutissima* (i.e. group HA) and 20 were classified as *Q. chenii* (i.e. group HC) (Supplementary Data Table S4). Although hybrids were widely detected in 28 populations, nearly 75.9 % of those in group HA (41/54) and 55.0 % of those in HC (11/20) were found to occur in the ancient contact zone recovered by the ENM analyses (Fig. 1B). A much higher proportion of hybrids was detected inside the contact zone than outside this zone for *Q. acutissima* (inside: 14.7 %, 41/279; outside: 3.1 %, 13/417), but not for *Q. chenii* (inside: 5.3 %, 11/209; outside: 4.4 %, 9/206). Groups HA and HC showed similar levels of genetic admixture (mean ± s.d., HA: $Q_{CHE}$ = 0.290 ± 0.214, HC: $Q_{ACU}$ = 0.250 ± 0.142; Wilcoxon rank-sum test: $W$ = 565, $P$ = 0.766). Only five individuals (four in HA and one in HC) exhibited ~50 % admixture ($0.4 < Q < 0.6$) for both parental species.

The results of hybrid identification were supported by three different analyses. First, PCA separated all the samples into two distinct groups corresponding to the two parental species (Fig. 1D). No overlap along PC1 was detected between parental individuals of *Q. acutissima* (i.e. group PA, $n$ = 642 individuals) and parental individuals of *Q. chenii* (i.e. group PC, $n$ = 395 individuals). Putative hybrids in groups HA and HC were characterized by intermediate scores along PC1 (pairwise Wilcoxon rank-sum test: all $P$ values < 0.001 after Holm–Bonferroni correction; Supplementary Data Table S7). Second, all the samples in group HC and 44 individuals (81.5 %) in group HA were classified as hybrids (*PP* assigned to the corresponding parental species class <0.9) in the NewHybrids analysis. Although ten of those in HA were identified as parental individuals of *Q. acutissima*, the *PP* values were found to be close to the threshold value of 0.9 (0.901–0.939). Almost all the samples in PA (99.8 %) and PC (98.2 %) were assigned to the parental groups successfully (Fig. 1E). Finally, the INTROGRESS analysis showed that compared with parental individuals of *Q. acutissima* (mean ± s.d., $H$ = 0.034 ± 0.068) and parental individuals of *Q. chenii* ($H$ = 0.971 ± 0.058), members in both hybrid groups exhibited intermediate values of the maximum-likelihood hybrid index (HA: $H$ = 0.383 ± 0.141, HC: $H$ = 0.654 ± 0.097) (Fig. 1F). All pairwise comparisons for $H$ were highly significant (Wilcoxon rank-sum test: all $P$ values < 0.001 after the Holm–Bonferroni correction; Supplementary Data Table S7), supporting the separation among the parental and hybrid individuals identified by Bayesian clustering analyses.

### *Gene flow and demographic history*

ABC analyses favoured the AM model as it received the highest posterior probability when using both the MLR (0.950) and RF (0.791) approaches to select the best model (Supplementary Data Tables S8 and S9). Cross-validations indicated that the overall prior error rate was moderate (MLR: 0.334, RF: 0.332), while the risk of selecting the AM model instead of others was much lower (average type II error, MLR: 0.052, RF: 0.054; Supplementary Data Table S10). A comparison of variable importance showed that $F_{ST}$ provided the most information in distinguishing the various models (Supplementary Data Fig. S3). This statistic was more easily underestimated under the SC and IM models, but overestimated under the SI model (Supplementary Data Fig. S4). Posterior predictive checks confirmed that the observed values of all the 17 summary statistics can be reproduced well under the AM scenario (Supplementary Data Figs S5–S8). The GoF test further demonstrated the ability of this model to generate simulations close to the observed data ($P$ = 0.990; Supplementary Data Fig. S9). Under the AM model, all the posterior distributions of $N_{ACU}$, $N_{CHE}$, $T_{AM}$ and mean $P_{GSM}$ contained a single peak different from the prior distributions (Supplementary Data Fig. S10). Posterior modes (95 % HPD) of $N_{ACU}$ and $N_{CHE}$ were 4688 (676–19 324) and 6898 (261–16 642), respectively. The posterior mode (95 % HPD) of $T_{AM}$ was 7831 (42–51 749) generations ago. The posterior mode (95 % HPD) of mean $P_{GSM}$ was 0.135 (0.021–0.354) (Table 1). Posterior distributions of other

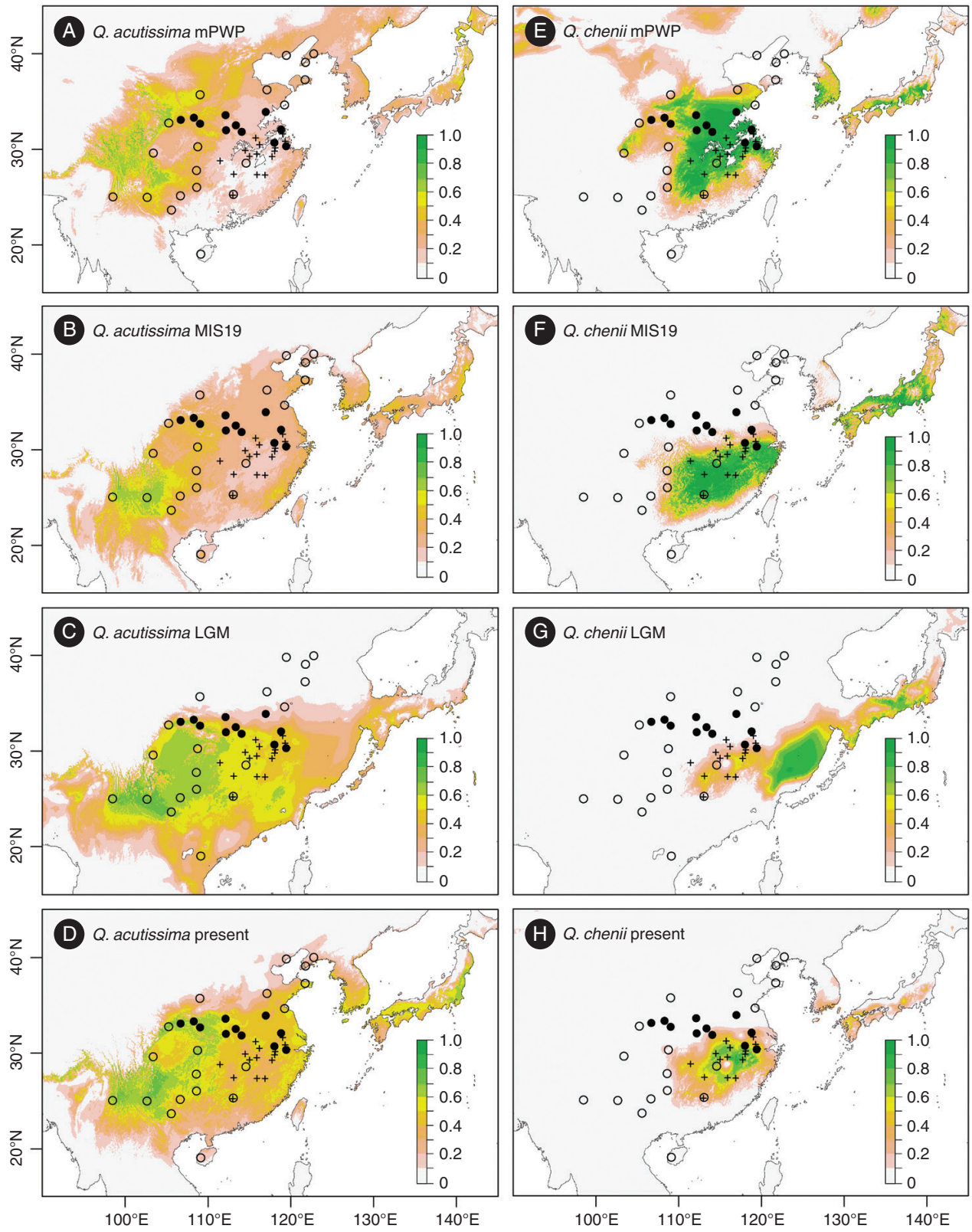


FIG. 3. MAXENT prediction maps showing the climatically suitable areas of *Quercus acutissima* (A–D) and *Q. chenii* (E–H) during the mid-Pliocene Warm Period (mPWP), Marine Isotope Stage 19 (MIS19) in the Pleistocene, Last Glacial Maximum (LGM) and the present. Plus signs represent the sampling sites of *Q. chenii*. Solid and open dots represent the sampling sites of *Q. acutissima* inside and outside the ancient contact zone, respectively.

TABLE 1. *Prior and posterior distributions of model parameters under the ancient migration (AM) scenario.*

| Parameter | Prior [type (min, max)] | Posterior | | | | |
|---|---|---|---|---|---|---|
| | | Median | Mean | Mode | 95 % HPD | |
| | | | | | Lower | Upper |
| $N_{ACU}$ | loguniform (10, 1000 000) | 4440.17 | 4223.77 | 4688.13 | 676.24 | 19 324.13 |
| $N_{CHE}$ | loguniform (10, 1000 000) | 5025.74 | 3921.93 | 6897.63 | 260.92 | 16 641.79 |
| $N_{ANC}$ | loguniform (10, 1000 000) | 2926.85 | 2966.18 | 16 084.22 | 13.41 | 738 074.16 |
| $T_{DIV}$ | uniform (100 000, 300 000) | 204 728.78 | 203 128.13 | 281 470.94 | 104 866.99 | 296 033.33 |
| $T_{AM}$ | loguniform (1, 300 000) | 3390.00 | 2414.35 | 7830.69 | 41.85 | 51 748.77 |
| $4N_{CHE}m_{A\to C}$ | uniform (0, 500) | 215.43 | 226.59 | 38.31 | 8.05 | 485.73 |
| $4N_{ACU}m_{C\to A}$ | uniform (0, 500) | 232.93 | 238.64 | 38.40 | 7.36 | 491.02 |
| mean $P_{GSM}$ | uniform (0, 0.8) | 0.17 | 0.17 | 0.13 | 0.02 | 0.35 |
| *shape* | uniform (0.5, 2) | 1.10 | 1.15 | 0.92 | 0.55 | 1.94 |

Under the AM model, the initial divergence of *Quercus acutissima* and *Q. chenii* occurred at $T_{DIV}$. The two newly formed populations continued to exchange alleles until time $T_{AM}$ ($T_{AM} < T_{DIV}$). $N_{ACU}$, $N_{CHE}$ and $N_{ANC}$ are effective population sizes of *Q. acutissima*, *Q. chenii* and the ancestral population, respectively. $4N_{CHE}m_{A\to C}$ and $4N_{ACU}m_{C\to A}$ are effective migration rates between the two species. Mean $P_{GSM}$ and *shape* are parameters of the generalized stepwise mutation (GSM) model. Note that the posterior estimate values were converted from log to linear scale.

parameters ($N_{ANC}$, $T_{DIV}$, $4N_{CHE}m_{A\to C}$, $4N_{ACU}m_{C\to A}$ and *shape*) were found to be poorly differentiated from the prior distributions (Supplementary Data Fig. S10), indicating that the current microsatellite dataset cannot provide enough information for these parameters.

### *Genetic admixture pattern*

The *Q. acutissima* populations inside the ancient contact zone showed a higher level of genetic admixture ($Q_{CHE}$) than those outside this zone, regardless of whether the two sympatric sites are included or not (*t*-test: all *P* values < 0.05; Fig. 4 and Supplementary Data Fig. S11). Both population- and individual-level GLMs indicated that the degree of admixture tended to be higher in localities where *Q. chenii* is present or those exhibiting more suitable climatic conditions for *Q. chenii* during the mPWP (Table 2; Supplementary Data Table S11 and Fig. S12). Within the ancient contact zone, we detected a positive effect of both the presence of *Q. chenii* in a given locality and the climatic niche suitability for *Q. chenii* during the Pleistocene (i.e. the MIS19 and LGM) (Table 2; Supplementary Data Table S11 and Fig. S12). Although the climatic niche suitability for *Q. acutissima* during the LGM was included in the individual-level model, it showed a non-significant effect with the unconditional 95 % confidence interval crossing zero (Table 2 and Supplementary Data Table S11). Hartigans' dip test indicated that the ancient contact zone formed a sharp bimodal distribution of ancestry estimates (Fig. 5). Similar patterns were also observed for the two sympatric populations of *Q. acutissima* and *Q. chenii*, i.e. populations 23 and 45 (AQY and CQY), and populations 30, 42 and 43 (AWTM, CWTM, and CZN) (all *P* values < 0.05; Fig. 5).

### *Chloroplast DNA sequence analyses*

The lengths of consensus sequences after alignment of *atp*B-*rbc*L, *psb*A-*trn*H and *trn*S-*trn*G were 718, 645, and 607 bp, respectively. We identified 29 haplotypes based on 18 nucleotide substitutions, seven indels and a 32-bp-long inversion (Supplementary Data Table S12). Fourteen haplotypes were specific to *Q. acutissima*, ten were private to *Q. chenii* and only five were shared by the two species (Fig. 6). The most widespread shared haplotype (H1) was found to occur in 23 populations of *Q. acutissima* (76.7 %) and seven populations of *Q. chenii* (38.9 %), while the most narrowly distributed one (H17) was only detected in a sympatric population in East China (i.e. populations 23 and 45; Fig. 6A). Nearly 93.8 % of the individuals with locally shared haplotypes (H7, H10, H17 and H18) were found to occur in the ancient contact zone detected by the ENM analyses (Fig. 6A and Supplementary Data Table S4).

The median-joining network grouped all the haplotypes into three non-species-specific clades (Fig. 6B). Neither species exhibited a significant phylogeographical structure (*Q. acutissima*: $N_{ST} > G_{ST}$, $P = 0.482$; *Q. chenii*: $N_{ST} < G_{ST}$, $P = 0.453$; Supplementary Data Table S5). Interspecific differentiation at cpDNA markers was weak ($F_{CT} = 0.025$, $P = 0.096$), with only 2.5 % of the variation partitioned among species (Supplementary Data Table S6).

## DISCUSSION

### *Genetic legacy of ancient hybridization in a mid-Pliocene contact zone*

Our current study provides evidence from both nSSR and cpDNA markers in support of the occurrence of hybridization between *Q. acutissima* and *Q. chenii*. We identified a total of 74 putative hybrids (~6.7 % of all the samples) based on the results of STRUCTURE. These findings were confirmed by a combination of PCA, NewHybrids analysis and maximum-likelihood hybrid index calculations. We also detected 112 individuals (~19.4 % of the tested samples) harbouring locally shared chloroplast haplotypes between the two species. The geographically associated pattern of shared haplotypes is more likely to be a result of chloroplast capture, i.e. chloroplast of one species being transferred to another through hybridization

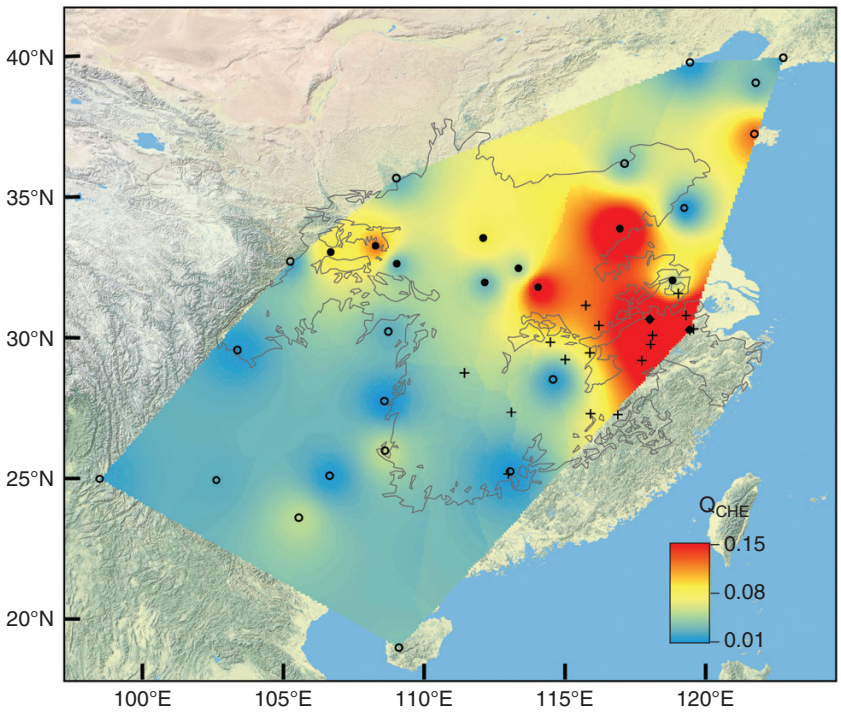


FIG. 4. Spatial interpolation for the probability of membership of each *Quercus acutissima* population to the *Q. chenii* genetic cluster ($Q_{CHE}$) as inferred by STRUCTURE analyses. The interpolated surface was obtained using the inverse distance-weighted method in ArcGIS 10.3 (ESRI). Plus signs represent the sampling sites of *Q. chenii*. Solid and open dots represent the sampling sites of *Q. acutissima* inside and outside the ancient contact zone, respectively.

(Acosta and Premoli, 2010), rather than incomplete lineage sorting, which would be evidenced by a random distribution of shared polymorphisms between species (Zhou *et al.*, 2017). Notably, we found that more than 70 % of the hybrids determined by STRUCTURE analysis and more than 90 % of the individuals with locally shared haplotypes are concentrated within a region between ~30°N and 35°N. These trees occur not only in the sympatric populations of eastern China but also in the western populations of *Q. acutissima* separated by ~400–600 km from the closest presence points of *Q. chenii*. Two hypotheses may explain our findings: (1) long-distance gene flow (Dodd and Afzal-Rafii, 2004; Ortego *et al.*, 2017) and (2) historical hybridization in an ancient contact zone followed by the local extinction of *Q. chenii* (Ortego *et al.*, 2014, 2018). In this case, we combine multiple lines of evidence to argue for the latter scenario.

First, ENMs show that *Q. acutissima* and *Q. chenii* may have had an ancient contact zone between ~30°N and 35°N during the mPWP (Figs 1A and 3). This interpretation is supported by the common presence of *Q. chenii* macrofossils in Late Pliocene sediments (~3.6–2.6 mya) at more northerly latitudes (~34–35°N, the Osaka and Tokai Groups of central Japan) than the species' current range limit (Momohara, 2010, 2016). Following this warm period, vegetation in northern China experienced a transition from open deciduous woodland with steppe to spruce–pine conifer forest (Liu *et al.*, 2002; Salzmann *et al.*, 2008). Together with other thermophilous subtropical tree species (Liu *et al.*, 2002; Qin *et al.*, 2011; Li *et al.*, 2019*b*), *Q. chenii* has suffered local extinction at the northern margin of its distribution (last occurrence at ~2.5 mya in the Tokai Group, Japan; Momohara, 2016) as a result of both climate cooling and strengthening of the winter monsoon in East Asia (Li *et al.*, 2004; Ge *et al.*, 2013). Such a change coincides with the ENM predictions that *Q. chenii* is likely to have retreated southward to its current distribution at least since the Early Pleistocene (Fig. 3), implying that the ancient contact zone may have disappeared along with the increased dryness and coldness during the glacial periods.

Second, our ABC analyses strongly support the scenario that there had been a long period of ancient contact followed by a recent period of strict isolation. This model is exactly in agreement with the expectation of historical hybridization followed by the local extinction of one parental species. Using an average generation time of 100–150 years for *Q. acutissima* and *Q. chenii* (Zeng *et al.*, 2015; Zhang *et al.*, 2018), we inferred that the mode of the time when the two species stopped exchanging alleles was 1.175–0.783 mya (Table 1). Although the estimated date must be interpreted with extreme caution, given the uncertainty in mutation models and generation time (Lande *et al.*, 2003), it is consistent with the ENM results that the degree of range overlap decreased significantly during the Early Pleistocene (Fig. 3). It should also be noted that we did not consider the variation of population size or recurrent cycles of contacts and isolation driven by the Pleistocene climate oscillations (Sun *et al.*, 2015). Thus, our demographic inference reflects a simplified scenario that allopatric divergence between the two species has been reinforced through isolation since the onset of the glacial periods.

Finally, we detect hybrids exhibiting a low level of admixture in almost all the allopatric populations of *Q. acutissima* inside the ancient contact zone (Figs 1 and 4). One possibility is that the two species have hybridized widely, followed by recurrent

TABLE 2. *General linear models (GLMs) showing relationships between genetic admixture and three sets of explanatory variables for all* Quercus acutissima *populations and those inside the ancient contact zone at both population and individual levels.*

| Parameter | Estimate | USE | Lower 95 % CI | Upper 95 % CI | $\Sigma\omega_i$ |
|---|---|---|---|---|---|
| **All *Q. acutissima* populations ($n$ = 28†)** | | | | | |
| Intercept | 0.024* | 0.008 | 0.009 | 0.038 | – |
| *Presence* | 0.086* | 0.019 | 0.049 | 0.123 | 0.61 |
| $S_{mPWP\_CHE}$ | 0.027* | 0.013 | 0.001 | 0.053 | 0.42 |
| **All *Q. acutissima* individuals ($n$ = 650†)** | | | | | |
| Intercept | 0.021* | 0.006 | 0.010 | 0.033 | – |
| *Presence* | 0.080* | 0.015 | 0.050 | 0.109 | 0.77 |
| $S_{mPWP\_CHE}$ | 0.032* | 0.012 | 0.009 | 0.055 | 0.77 |
| **All *Q. acutissima* populations with the two sympatric ones removed ($n$ = 26†)** | | | | | |
| Intercept | 0.022* | 0.008 | 0.007 | 0.037 | – |
| $S_{mPWP\_CHE}$ | 0.033* | 0.013 | 0.007 | 0.058 | 0.63 |
| $S_{MIS19\_CHE}$ | –0.037 | 0.020 | –0.077 | 0.002 | 0.20 |
| $S_{LGM\_CHE}$ | –0.155 | 0.097 | –0.346 | 0.036 | 0.12 |
| $S_{LGM\_ACU}$ | –0.024 | 0.017 | –0.057 | 0.008 | 0.10 |
| $S_{PRES\_CHE}$ | –0.055 | 0.042 | –0.137 | 0.027 | 0.08 |
| **All *Q. acutissima* individuals with the two sympatric populations removed ($n$ = 601†)** | | | | | |
| Intercept | 0.023* | 0.007 | 0.010 | 0.037 | – |
| $S_{mPWP\_CHE}$ | 0.038* | 0.011 | 0.016 | 0.059 | 0.65 |
| $S_{MIS19\_CHE}$ | –0.043* | 0.020 | –0.082 | –0.005 | 0.34 |
| $S_{LGM\_CHE}$ | –0.171 | 0.090 | –0.348 | 0.007 | 0.18 |
| $S_{LGM\_ACU}$ | –0.024 | 0.014 | –0.052 | 0.004 | 0.13 |
| ***Q. acutissima* populations inside the ancient contact zone ($n$ = 11)** | | | | | |
| Intercept | 0.046* | 0.010 | 0.026 | 0.066 | – |
| $S_{LGM\_CHE}$ | 0.425* | 0.125 | 0.180 | 0.671 | 0.23 |
| *Presence* | 0.078* | 0.023 | 0.033 | 0.123 | 0.22 |
| $S_{MIS19\_CHE}$ | 0.102* | 0.035 | 0.034 | 0.170 | 0.10 |
| ***Q. acutissima* individuals inside the ancient contact zone ($n$ = 279)** | | | | | |
| Intercept | 0.070* | 0.026 | 0.018 | 0.121 | – |
| $S_{LGM\_CHE}$ | 0.429* | 0.116 | 0.201 | 0.657 | 0.24 |
| *Presence* | 0.078* | 0.021 | 0.036 | 0.120 | 0.20 |
| $S_{LGM\_ACU}$ | –0.082 | 0.052 | –0.183 | 0.018 | 0.25 |

The model-averaged parameter estimates were computed among a set of candidate models with $\Delta AIC_C \leq 2$ (see Supplementary Data Table S11). USE, unconditional standard error; 95 % CI, unconditional 95 % confidence interval; $\Sigma\omega_i$, relative importance of a variable based on the sum of AICc weights of the models with $\Delta AIC_C \leq 2$; *Presence*, presence of *Q. chenii* in a given locality; *S*, climatic niche suitability for *Q. acutissima* (ACU) or *Q. chenii* (CHE) during the present (PRES), Last Glacial Maximum (LGM), Marine Isotope Stage 19 (MIS19) in the Pleistocene, and the mid-Pliocene Warm Period (mPWP).

*Significant estimate whose unconditional 95 % CI did not cross zero.

†Two *Q. acutissima* populations (ALYG and AQHD) were excluded from our analysis because they were below sea level during the mPWP (i.e. $S_{mPWP}$ values are missing).

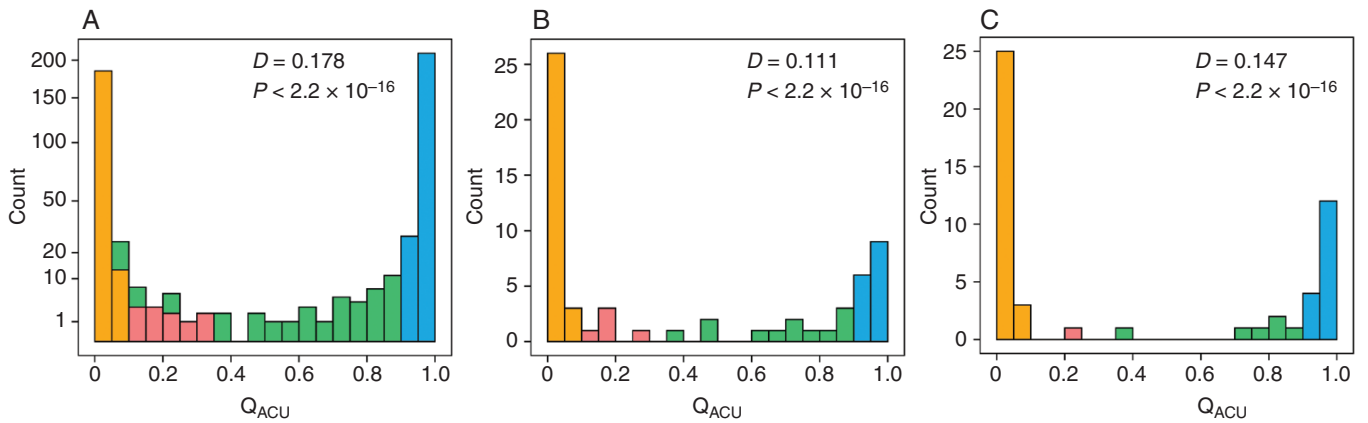


FIG. 5. Distribution of ancestry estimates in the ancient contact zone (A) and two sympatric populations of *Quercus acutissima* and *Q. chenii*, i.e. populations 23 and 45 (AQY and CQY) (B), and populations 30, 42 and 43 (AWTM, CWTM and CZN; see Fig. 1) (C). The ancestry is based on the probability of membership to the genetic cluster of *Q. acutissima* ($Q_{ACU}$). Hartigans' dip statistic *D* and its *P* value were computed using the R package diptest. Colours of bars represent four groups: parental individuals of *Q. acutissima* (group PA, blue), parental individuals of *Q. chenii* (group PC, orange), hybrids morphologically assigned to *Q. acutissima* (group HA, green) and hybrids morphologically assigned to *Q. chenii* (group HC, red).

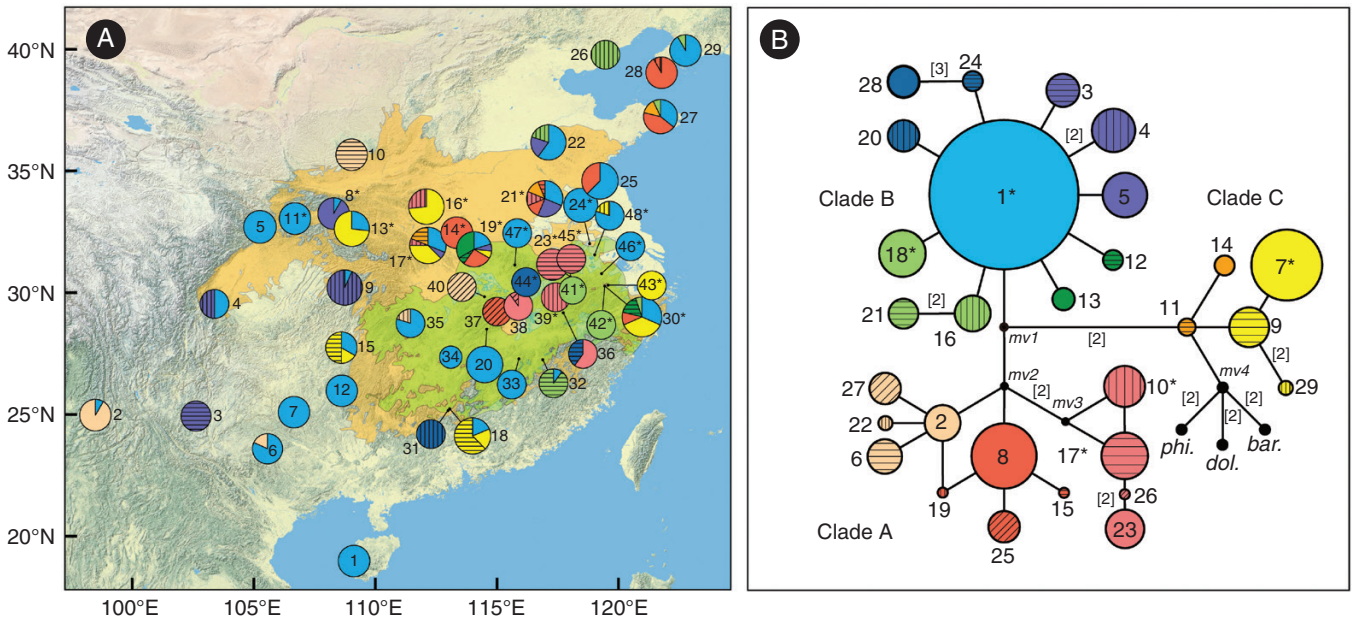


FIG. 6. Geographical distribution (A) and median-joining network (B) of 29 chloroplast DNA haplotypes of *Quercus acutissima* and *Q. chenii*. (A) Circle size is proportional to the sample size. Orange and green areas indicate the climatically suitable areas for *Q. chenii* during the mid-Pliocene Warm Period (mPWP) and during the present, respectively. They were created from the logistic outputs of MAXENT using the maximum sensitivity plus specificity threshold (Gugger *et al.*, 2013). Population codes (*Q. acutissima*: 1–30; *Q. chenii*: 31–48) are described in Supplementary Data Table S1. *Populations inside the ancient contact zone. (B) Circle size is proportional to the frequency of a haplotype across all populations. Small black dots indicate inferred intermediate haplotypes not detected in this investigation. Numbers in brackets on branches indicate the number of mutations between haplotypes when branches represent more than one mutation. Three outgroups include *Q. phillyraeoides* (*phi.*; GenBank accession number MK105462; Pang *et al.*, 2019), *Q. dolicholepis* (*dol.*; KU240010; Yang *et al.*, 2016) and *Q. baronii*. (*bar.*; KT963087; Yang *et al.*, 2017). *The haplotypes shared by the two species.

backcrosses with *Q. acutissima* in the absence of *Q. chenii* (Lepais *et al.*, 2009). Moreover, the observed hybrids present a scattered distribution across the mountainous areas in both western and eastern China. Previous phylogeographical studies have shown that these montane regions (e.g. the Qin Mountains, the Daba Mountains, the Dabie Mountains and the Tianmu Mountains) may have acted as glacial refugia for *Q. acutissima* (Zhang *et al.*, 2015, 2018). Thus, although *Q. chenii* suffered local extinction, the genetically admixed individuals can still survive *in situ* among isolated habitats with high ecological stability. However, the complex landscape in the ancient contact zone could have reduced long-distance pollen dispersal. Using nuclear genetic markers, previous work has revealed a strong phylogeographical break associated with the topographic, climatic and vegetational differences between western and eastern China for many woody plants (Sun *et al.*, 2016; Luo *et al.*, 2020), including wind-pollinated trees (Zhang *et al.*, 2018; Gao *et al.*, 2021). Thus, even considering that oaks have a high potential for long-distance pollen dispersal (Buschbom *et al.*, 2011), it seems unlikely that the reduced contemporary gene flow among western and eastern populations can fully explain the observed hybrid structure in the ancient contact zone.

Overall, we combine evidence from fossil records, demographic inference and ENM to support the occurrence of a mid-Pliocene contact zone between *Q. acutissima* and *Q. chenii*. These two species may have experienced extensive pre-glacial gene flow followed by reinforced post-glacial isolation. At present, we detect a clear-cut bimodal hybrid structure in the ancient contact zone (Fig. 5), suggesting that the accumulation of reproductive barriers may have significantly reduced the ongoing hybridization rate (Zeng *et al.*, 2011; Ortego *et al.*, 2014). Such a finding is consistent with the expanded biological species concept (Wang *et al.*, 2020), which expects substantial gene flow at the onset of speciation until reproductive isolation evolves after the gradual development of geographical isolation. However, the demographic history estimated here is different from that of European white oaks. Using genome-wide nuclear markers, recent studies have shown that long-term strict isolation followed by recent postglacial contacts has occurred between four white oak species in Europe (Leroy *et al.*, 2017, 2020). This difference in gene flow history may be attributed to the fact that the two East Asian Cerris oaks diverged ~ 20 mya earlier than their European congeners (Hipp *et al.*, 2020). Such a long period would provide sufficient time for *Q. acutissima* and *Q. chenii* to reinforce the reproductive barrier, which may have been strong enough to completely suppress the recent interspecific gene flow (Roux *et al.*, 2016).

### *Influence of Pliocene and Pleistocene climates on genetic admixture patterns*

Our investigation demonstrates that the *Q. acutissima* populations inside the ancient contact zone exhibit a higher level of genetic admixture than those outside this zone (Supplementary Data Fig. S11). One variable, the climatic niche suitability index for *Q. chenii* ($S_{CHE}$) during the mPWP, shows the best performance in explaining the overall patterns of admixture regardless of whether the two sympatric sites are included or not (Table 2; Supplementary Data Table S11 and Fig. S12). Such a result

is partially consistent with the findings of Ortego *et al.* (2014), who concluded that climatic niche suitability is associated with the observed admixture patterns between *Q. engelmannii* and the widespread scrub oaks in California. However, in contrast to previous work, our most parsimonious models only include $S_{CHE}$ for the mPWP rather than that for the present as an explanatory variable, suggesting a much stronger effect of mid-Pliocene climate in constraining the admixture patterns that may have originated from the ancient contact events during the mPWP.

Within the ancient contact zone, our analyses indicate that $S_{CHE}$ for the LGM plays a more important role in determining the local patterns of admixture than that for other periods (Table 2; Supplementary Data Table S11 and Fig. S12). The observed positive trend between $S_{LGM\_CHE}$ and $Q_{CHE}$ is mainly driven by the two sympatric sites that present the most suitable LGM climatic conditions for *Q. chenii* (Supplementary Data Fig. S12). One explanation is that the extremely cool climate during the glacial period has exerted a strong 'filter' effect on the occurrence of interspecific gene flow through demographic forces (Ortego *et al.*, 2018). This interpretation is supported by the ENM predictions showing that the two sympatric sites have maintained a moderate to high climatic suitability for both species throughout the Pleistocene ($S_{MIS19}$ and $S_{LGM} > 0.18$), whereas the other sites were unsuitable for *Q. chenii* at least in the LGM ($S_{LGM} < 0.04$; Fig. 3), thus reducing the probability of historical gene exchange. Furthermore, we detect a high level of nuclear diversity as well as the fixation of locally shared chloroplast haplotypes in the two sympatric populations (Fig. 6), suggesting that the two species may have persisted *in situ* within an ecologically stable area (Zhang *et al.*, 2018; Li *et al.*, 2019*a*). Combining all these lines of evidence, we infer that the mountainous areas around the two sympatric sites have acted as a shared glacial refugium for *Q. acutissima* and *Q. chenii*, allowing the long-term coexistence and historical hybridization between them. In contrast, the remaining sites are characterized by low ecological stability, which would reduce the survival rate of offspring and further restrict interspecific gene exchange.

Another explanation for the correlation between $S_{LGM\_CHE}$ and $Q_{CHE}$ is that sympatry (i.e. *Presence*) creates the opportunity for contemporary hybridization, but the current local distribution of *Q. chenii* is still constrained by LGM climate. This situation may occur when range shifts of oak trees lag behind climate change, while interspecific gene flow continues in previous overlapping distributions (Nathan *et al.*, 2011; Browne *et al.*, 2019). However, our present study provides weak support for this hypothesis because all the scenarios considering contemporary gene flow (i.e. the IM and SC models) are rejected by demographic inference (Supplementary Data Tables S8 and S9). Furthermore, both the sharp bimodal hybrid structure and the lack of $F_1$ hybrids suggest that the ongoing hybridization rate is rather low within the sympatric populations (Zeng *et al.*, 2011; Ortego *et al.*, 2014; Kim *et al.*, 2018). Although a previous study has described a putative hybrid variety (*Q. chenii* var. *linanensis*) from one of the sympatric sites (i.e. Lin'an County; Liu, 1992), the asymmetry in parental contributions to the ancestry of the hybrids sampled here suggested that this variety is more likely to have arisen from ancient introgression rather than contemporary hybridization between *Q. acutissima* and *Q. chenii* (Zhang *et al.*, 2019).

Our results highlight that palaeoclimate change shapes present admixture patterns between the two East Asian Cerris oaks by influencing the extent of historical range overlap (Maze, 1968). Specifically, the mid-Pliocene warm climate promoted ancient contact, allowing widespread hybridization throughout central China. In contrast, the Pleistocene cool climate caused the local extinction of *Q. chenii*, reducing the probability of interspecific gene flow in most areas except those sites having a high level of ecological stability.

### *Limitations and future directions*

Our study has several limitations. First, the microsatellite dataset only includes a limited number of nuclear markers and represents a rather low proportion of the oak genome. These markers show good performance in identifying hybrids, but they do not provide sufficient information to reliably estimate the direction and extent of interspecific gene flow (Supplementary Data Fig. S10). Second, considering that most microsatellites are neutral loci reflecting the patterns of mutation, genetic drift and gene flow (Scotti-Saintagne *et al.*, 2004), it is difficult to assess whether selective forces (e.g. adaptive introgression) also contribute to the admixture pattern. Finally, compared with other types of markers, microsatellites are more likely to be affected by homoplasy because of their high mutation rates and allele size constraints, which also hampers the estimation of long evolutionary history (Estoup *et al.*, 2002; Camacho-Sanchez *et al.*, 2020). To overcome these shortcomings, future studies should utilize genome-wide informative markers from reduced library sequencing (e.g. Ortego *et al.*, 2018; Kim *et al.*, 2018; Burge *et al.*, 2019) or when possible whole genome sequences, to robustly evaluate the influence of past climate change on the extent, direction and timing of interspecific gene flow. In combination with the identification of genomic introgression regions, such work would provide novel insight into how ancient introgression has interacted with both non-adaptive and adaptive processes to shape the complex evolutionary history of East Asian Cerris oaks. Nonetheless, the use of multiple types of corroborative evidence supports the conclusions reported here regarding the associations between palaeoclimate change, historical range shifts and ancient hybridization history.

## CONCLUSIONS

Our study demonstrates that the current admixture patterns among East Asian Cerris oaks reflect the genetic imprints of ancient hybridization affected by both Pliocene and Pleistocene climates. This finding differs from that reported for European white oaks, whose hybridization history is mainly constrained by the last glacial–interglacial cycle (Petit *et al.*, 2002*a*; Leroy *et al.*, 2017, 2020), but is generally consistent with the patterns found in western North America, where genetically distant or allopatric oak species exchanged alleles in much older times and left signals of historical introgression in contemporary populations (e.g. Maze, 1968; McVay *et al.*, 2017; Kim *et al.*, 2018; Ortego *et al.*, 2018). Our results indicate that the hybridization history between sister species occurring in areas less affected by glaciations is more complex than that for species

seriously threatened by Pleistocene climate change. The impact of pre-Quaternary climate should be taken into account especially when exploring ancient gene flow among species native to East Asia.

## SUPPLEMENTARY DATA

Supplementary data are available online at https://academic.oup.com/aob and consist of the following. Table S1: Locations of 30 *Quercus acutissima* populations and 18 *Q. chenii* populations. Table S2: Primer sequences, annealing temperatures and lengths of consensus sequences for the three chloroplast intergenic spacers. Table S3: Primer sequences, annealing temperature and genetic statistics for the seven nuclear microsatellite loci. Table S4: Genetic statistics for each population of *Q. acutissima* and *Q. chenii*. Table S5: Genetic statistics of *Q. acutissima* and *Q. chenii*. Table S6: Analyses of molecular variance for *Q. acutissima* and *Q. chenii*. Table S7: Descriptive statistics and multiple comparisons for PC1 score and maximum-likelihood hybrid index among parental individuals of *Q. acutissima* and *Q. chenii* and their hybrids. Table S8: Model choice using the multinomial logistic regression approach. Table S9: Model choice using the random forest approach. Table S10: Fraction of models correctly predicted and classification error rate estimated by both multinomial logistic regression and random forest approaches. Table S11: Model selection to assess the relationships between genetic admixture and three sets of explanatory variables for all *Q. acutissima* populations and those inside the ancient contact zone at both population and individual levels. Table S12: Chloroplast haplotypes identified in this study based on 18 nucleotide substitutions, seven indels and one 32-bp inversion. Fig. S1: The receiver operating characteristic curves averaged over ten replicate runs. Fig. S2: $\Delta K$ and Ln Pr($X|K$) obtained in Bayesian cluster analysis for each value of $K$. Fig. S3: Variable importance estimated by the random forest approach. Fig. S4: Genetic differentiation index between *Q. acutissima* and *Q. chenii* under the four demographic models. Fig. S5: Posterior predictive checks for the ancient migration model. Fig. S6: Posterior predictive checks for the isolation with migration model. Fig. S7: Posterior predictive checks for the secondary contact model. Fig. S8: Posterior predictive checks for the strict isolation model. Fig. S9: Goodness of fit for the four demographic models. Fig. S10: Prior and posterior distributions of parameters in the ancient migration model. Fig. S11: Probability of membership to the genetic cluster of *Q. chenii* for *Q. acutissima* populations inside and outside the ancient contact zone. Fig. S12: Linear correlations between genetic admixture and climatic niche suitability for *Q. chenii* during the Last Glacial Maximum or mid-Pliocene Warm Period.

## ACKNOWLEDGEMENTS

We thank Qingliang Liu, Baokun Xu, Xuan Li, Zhongren Xiong, Xiaochen Zhang, Kaiwen Zhang, Sujing Fu, Heng Jia, Wenbin Xu, Huachen Wang, Gang Yao, Ye Tian, Xulan Shang, Shengzuo Fang and Qianru Liu for their help with fieldwork.

## FUNDING

This work was supported by the National Natural Science Foundation of China (31770699, 31870506, 31370666), the China Postdoctoral Science Foundation (2020M681629), the Jiangsu Postdoctoral Research Funding Program (2021K038A), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB31000000), the Natural Science Foundation of Jiangsu Province, China (BK20181398), the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYLX15_0922), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

## DATA ACCESSIBILITY

Sequence data are available on GenBank (http://www.ncbi.nlm.nih.gov/genbank/) under accession numbers KT152178–KT152200 and MH924168–MH92419. The microsatellite genotype dataset is available from Zenodo: https://doi.org/10.5281/zenodo.1442911. The input files and R scripts for the ABC analyses are available from Zenodo: https://doi.org/10.5281/zenodo.5631037.

## LITERATURE CITED

**Acosta MC, Premoli AC. 2010.** Evidence of chloroplast capture in South American *Nothofagus* (subgenus *Nothofagus*, Nothofagaceae). *Molecular Phylogenetics and Evolution* **54**: 235–242.

**An M, Deng M, Zheng SS, Jiang XL, Song YG. 2017.** Introgression threatens the genetic diversity of *Quercus austrocochinchinensis* (Fagaceae), an endangered oak: a case inferred by molecular markers. *Frontiers in Plant Science* **8**: 229.

**Anderson EC, Thompson EA. 2002.** A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**: 1217–1229.

**Aoki K, Tamaki I, Nakao K, *et al.* 2019.** Approximate Bayesian computation analysis of EST-associated microsatellites indicates that the broadleaved evergreen tree *Castanopsis sieboldii* survived the Last Glacial Maximum in multiple refugia in Japan. *Heredity* **122**: 326–340.

**Brown JL, Hill DJ, Dolan AM, Carnaval AC, Haywood AM. 2018.** PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Scientific Data* **5**: 180254.

**Browne L, Wright JW, Fitz-Gibbon S, Gugger PF, Sork VL. 2019.** Adaptational lag to temperature in valley oak (*Quercus lobata*) can be mitigated by genome-informed assisted gene flow. *Proceedings of the National Academy of Sciences, USA* **116**: 25179–25185.

**Buerkle CA. 2005.** Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes* **5**: 684–687.

**Buggs RJA. 2007.** Empirical study of hybrid zone movement. *Heredity* **99**: 301–312.

**Burgarella C, Lorenzo Z, Jabbour-Zahab R, *et al.* 2009.** Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*). *Heredity* **102**: 442–452.

**Burge DO, Parker VT, Mulligan M, Sork VL. 2019.** Influence of a climatic gradient on genetic exchange between two oak species. *American Journal of Botany* **106**: 864–878.

**Burnham KP, Anderson DR. 1998.** *Model selection and inference: a practical information-theoretic approach*. New York: Springer.

**Buschbom J, Yanbaev Y, Degen B. 2011.** Efficient long-distance gene flow into an isolated relict oak stand. *Journal of Heredity* **102**: 464–472.

**Camacho-Sanchez M, Velo-Antón G, Hanson JO, *et al.* 2020.** Comparative assessment of range-wide patterns of genetic diversity and structure with SNPs and microsatellites: a case study with Iberian amphibians. *Ecology and Evolution* **10**: 10353–10363.

**Cannon CH, Petit RJ. 2020.** The oak syngameon: more than the sum of its parts. *New Phytologist* **226**: 978–983.

**Crowl AA, Manos PS, McVay JD, Lemmon AR, Lemmon EM, Hipp AL. 2020.** Uncovering the genomic signature of ancient introgression between white oak lineages (*Quercus*). *New Phytologist* **226**: 1158–1170.

**Csilléry K, François O, Blum MGB. 2012.** abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution* **3**: 475–479.

**Currat M, Ruedi M, Petit RJ, Excoffier L. 2008.** The hidden side of invasions: massive introgression by local genes. *Evolution* **62**: 1908–1920.

**Dieringer D, Schlötterer C. 2003.** Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* **3**: 167–169.

**Dodd RS, Afzal-Rafii Z. 2004.** Selection and dispersal in a multispecies oak hybrid zone. *Evolution* **58**: 261–269.

**Dray S, Dufour AB. 2007.** The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* **22**: 1–20.

**Earl DA, vonHoldt BM. 2012.** STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**: 359–361.

**Estoup A, Jarne P, Cornuet JM. 2002.** Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* **11**: 1591–1604.

**Evanno G, Regnaut S, Goudet J. 2005.** Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–2620.

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

**Excoffier L, Estoup A, Cornuet JM. 2005.** Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* **169**: 1727–1738.

**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, 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.

**Feng G, Mao L, Sandel B, Swenson NG, Svenning JC. 2016.** High plant endemism in China is partially linked to reduced glacial-interglacial climate change. *Journal of Biogeography* **43**: 145–154.

**Fujiwara K, Harada A. 2015.** Character of warm-temperate *Quercus* Forests in Asia. In: Box E, Fujiwara K. eds. *Warm-temperate deciduous forests around the Northern Hemisphere*. Cham: Springer, 27–80.

**Gao J, Liu ZL, Zhao W, *et al.* 2021.** Combined genotype and phenotype analyses reveal patterns of genomic adaptation to local environments in the subtropical oak *Quercus acutissima*. *Journal of Systematics and Evolution* **59**: 541–556.

**Ge JY, Dai Y, Zhang ZS, *et al.* 2013.** Major changes in East Asian climate in the mid-Pliocene: triggered by the uplift of the Tibetan Plateau or global cooling? *Journal of Asian Earth Sciences* **69**: 48–59.

**Gompert Z, Buerkle CA. 2010.** Introgress: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources* **10**: 378–384.

**Gompert Z, Mandeville EG, Buerkle CA. 2017.** Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics* **48**: 207–229.

**Goudet J. 2001.** *FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3)*. http://www2.unil.ch/popgen/softwares/fstat.htm.

**Grivet D, Deguilloux MF, Petit RJ, Sork VL. 2006.** Contrasting patterns of historical colonization in white oaks (*Quercus* spp.) in California and Europe. *Molecular Ecology* **15**: 4085–4093.

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

**Hall TA. 1999.** BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95–98.

**Hartigan JA, Hartigan PM. 1985.** The dip test of unimodality. *Annals of Statistics* **13**: 70–84.

**Hewitt GM. 2004.** Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* **359**: 183–195.

**Hewitt GM. 2011.** Quaternary phylogeography: the roots of hybrid zones. *Genetica* **139**: 617–638.

**Hijmans RJ. 2019.** *raster: Geographic Data Analysis and Modeling. R package version 2.8-19*. https://CRAN.R-project.org/package=raster.

**Hijmans 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.

**Hill DJ. 2015.** The non-analogue nature of Pliocene temperature gradients. *Earth and Planetary Science Letters* **425**: 232–241.

**Hipp AL, Manos PS, Hahn M, *et al.* 2020.** Genomic landscape of the global oak phylogeny. *New Phytologist* **226**: 1198–1212.

**Holm S. 1979.** A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**: 65–70.

**Janes JK, Miller JM, Dupuis JR, *et al.* 2017.** The $K = 2$ conundrum. *Molecular Ecology* **26**: 3594–3602.

**Kim BY, Wei XZ, Fitz-Gibbon S, Lohmueller KE, Ortego J, Gugger PF, Sork VL. 2018.** RADseq data reveal ancient, but not pervasive, introgression between Californian tree and scrub oak species (*Quercus* sect. *Quercus*: Fagaceae). *Molecular Ecology* **27**: 4556–4571.

**Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015.** Clumpak: a program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources* **15**: 1179–1191.

**Kremer A, Hipp AL. 2020.** Oaks: an evolutionary success story. *New Phytologist* **226**: 987–1011.

**Lande R, Engen S, Saether BE. 2003.** *Stochastic population dynamics in ecology and conservation*. Oxford: Oxford University Press.

**Leigh JW, Bryant D. 2015.** PopART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6**: 1110–1116.

**Lepais O, Petit RJ, Guichoux E, *et al.* 2009.** Species relative abundance and direction of introgression in oaks. *Molecular Ecology* **18**: 2228–2242.

**Leroy T, Rougemont Q, Dupouey JL, *et al.* 2020.** Massive postglacial gene flow between European white oaks uncovered genes underlying species barriers. *New Phytologist* **226**: 1183–1197.

**Leroy T, Roux C, Villate L, *et al.* 2017.** Extensive recent secondary contacts between four European white oak species. *New Phytologist* **214**: 865–878.

**Li XQ, Li CS, Lu HY, Dodson JR, Wang YF. 2004.** Paleovegetation and paleoclimate in middle-late Pliocene, Shanxi, central China. *Palaeogeography, Palaeoclimatology, Palaeoecology* **210**: 57–66.

**Li Y, Zhang Z, Ding G, *et al.* 2019*b*.** Late Pliocene and early Pleistocene vegetation and climate change revealed by a pollen record from Nihewan Basin, North China. *Quaternary Science Reviews* **222**: 105905.

**Li Y, Zhang XW, Fang YM. 2019*a*.** Landscape features and climatic forces shape the genetic structure and evolutionary history of an oak species (*Quercus chenii*) in East China. *Frontiers in Plant Science* **10**: 1060.

**Librado P, Rozas J. 2009.** DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.

**Liu GW, Leopold EB, Liu Y, Wang WM, Yu ZY, Tong GB. 2002.** Palynological record of Pliocene climate events in North China. *Review of Palaeobotany and Palynology* **119**: 335–340.

**Liu MC. 1992.** A new variety of *Quercus chenii* Nakai. *Bulletin of Botanical Research* **12**: 275–276.

**Luo D, Xu B, Li ZM, Sun H. 2020.** Biogeographical divides delineated by the three-step landforms of China and the East China Sea: Insights from the phylogeography of *Kerria japonica*. *Journal of Biogeography* **48**: 372–385.

**Lyu J, Song J, Liu Y, Wang YY, Li JQ, Du FK. 2018.** Species boundaries between three sympatric oak species: *Quercus aliena*, *Q. dentata*, and *Q. variabilis* at the northern edge of their distribution in China. *Frontiers in Plant Science* **9**: 414.

**Manos PS, Doyle JJ, Nixon KC. 1999.** Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Molecular Phylogenetics and Evolution* **12**: 333–349.

**Martin M. 2016.** *diptest: Hartigan's dip test statistic for unimodality - Corrected. R package version 0.75-7*. https://CRAN.R-project.org/package=diptest.

**Maze J. 1968.** Past hybridization between *Quercus macrocarpa* and *Quercus gambelii*. *Brittonia* **20**: 321–333.

**Mazerolle MJ. 2019.** *AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). R package version 2.2-2*. https://cran.r-project.org/package=AICcmodavg.

**McVay JD, Hauser D, Hipp AL, Manos PS. 2017.** Phylogenomics reveals a complex evolutionary history of lobed-leaf white oaks in western North America. *Genome* **60**: 733–742.

**Momohara A. 2010.** Stages of major floral and vegetational changes since the latest Neogene in central Europe and central Japan in connection with climatic changes. *The Quaternary Research* **49**: 299–308.

**Momohara A. 2016.** Stages of major floral change in Japan based on macrofossil evidence and their connection to climate and geomorphological changes since the Pliocene. *Quaternary International* **397**: 93–105.

**Moran EV, Willis J, Clark JS. 2012.** Genetic evidence for hybridization in red oaks (*Quercus* sect. *Lobatae*, Fagaceae). *American Journal of Botany* **99**: 92–100.

**Nathan R, Horvitz N, He Y, Kuparinen A, Schurr FM, Katul GG. 2011.** Spread of North American wind-dispersed trees in future environments. *Ecology Letters* **14**: 211–219.

**Neophytou C, Aravanopoulos FA, Fink S, Dounavi A. 2011.** Interfertile oaks in an island environment. II. Limited hybridization between *Quercus alnifolia* Poech and *Q. coccifera* L. in a mixed stand. *European Journal of Forest Research* **130**: 623–635.

**Ortego J, Gugger PF, Riordan EC, Sork VL. 2014.** Influence of climatic niche suitability and geographical overlap on hybridization patterns among southern Californian oaks. *Journal of Biogeography* **41**: 1895–1908.

**Ortego J, Gugger PF, Sork VL. 2017.** 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.

**Pang X, Liu H, Wu S, *et al.* 2019.** Species identification of oaks (*Quercus* L., Fagaceae) from gene to genome. *International Journal of Molecular Sciences* **20**: 5940.

**Peñaloza-Ramírez JM, González-Rodríguez A, Mendoza-Cuenca L, Caron H, Kremer A, Oyama K. 2010.** Interspecific gene flow in a multispecies oak hybrid zone in the Sierra Tarahumara of Mexico. *Annals of Botany* **105**: 389–399.

**Petit RJ, Brewer S, Bordács S, *et al.* 2002*a*.** Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management* **156**: 49–74.

**Petit RJ, Csaikl UM, Bordács S, *et al.* 2002*b*.** Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management* **156**: 5–26.

**Phillips SJ, Dudík M, Schapire RE. 2018.** *Maxent software for modeling species niches and distributions (Version 3.4.1)*. http://biodiversityinformatics.amnh.org/open_source/maxent/.

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

**Pritchard JK, Stephens M, Donnelly P. 2000.** Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

**Pudlo P, Marin J, Estoup A, Cornuet JM, Gautier M, Robert CP. 2016.** Reliable ABC model choice via random forests. *Bioinformatics* **32**: 859–866.

**Qin F, Ferguson DK, Zetter R, *et al.* 2011.** Late Pliocene vegetation and climate of Zhangcun region, Shanxi, North China. *Global Change Biology* **17**: 1850–1870.

**R Core Team. 2018.** *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

**Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. 2016.** Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology* **14**: e2000234.

**Ryan SF, Deines JM, Scriber JM, *et al.* 2018.** Climate-mediated hybrid zone movement revealed with genomics, museum collection, and simulation modeling. *Proceedings of the National Academy of Sciences, USA* **115**: 2284–2291.

**Salzmann U, Haywood AM, Lunt DJ, Valdes PJ, Hill DJ. 2008.** A new global biome reconstruction and data-model comparison for the middle Pliocene. *Global Ecology and Biogeography* **17**: 432–447.

**Scotti-Saintagne C, Mariette S, Porth I, *et al.* 2004.** Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* **168**: 1615–1626.

**Simeone MC, Cardoni S, Piredda R, *et al.* 2018.** Comparative systematics and phylogeography of *Quercus* Section *Cerris* in western Eurasia: inferences from plastid and nuclear DNA variation. *PeerJ* **6**: 5793.

**Simmons MP, Ochoterena H. 2000.** Gaps as characters in sequence based phylogenetic analyses. *Systematic Biology* **49**: 369–381.

**Sun YB. 2017.** FasParser: a package for manipulating sequence data. *Zoological Research* **38**: 110–112.

**Sun YS, Li LL, Li L, Zou JB, Liu JQ. 2015.** Distributional dynamics and interspecific gene flow in *Picea likiangensis* and *P. wilsonii* triggered by climate change on the Qinghai-Tibet Plateau. *Journal of Biogeography* **42**: 475–484.

**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.

**Tamaki I, Okada M. 2014.** Genetic admixing of two evergreen oaks, *Quercus acuta* and *Q. sessilifolia* (subgenus *Cyclobalanopsis*), is the result of interspecific introgressive hybridization. *Tree Genetics and Genomes* **10**: 989–999.

**Taylor SA, Larson EL, Harrison RG. 2015.** Hybrid zones: windows on climate change. *Trends in Ecology & Evolution* **30**: 398–406.

**Taylor SA, White TA, Hochachka WM, Ferretti V, Curry RL, Lovette I. 2014.** Climate-mediated movement of an avian hybrid zone. *Current Biology* **24**: 671–676.

**Wang N, Borrell JS, Bodles WJ, Kuttapitiya A, Nichols RA, Buggs RJ. 2014.** Molecular footprints of the Holocene retreat of dwarf birch in Britain. *Molecular Ecology* **23**: 2771–2782.

**Wang XF, He ZW, Shi SH, Wu CI. 2020.** Genes and speciation: is it time to abandon the biological species concept? *National Science Review* **7**: 1387–1397.

**Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L. 2010.** ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* **11**: 116.

**Whittemore AT, Schaal BA. 1991.** Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences, USA* **88**: 2540–2544.

**Wielstra B. 2019.** Historical hybrid zone movement: more pervasive than appreciated. *Journal of Biogeography* **46**: 1300–1305.

**Wielstra B, Burke T, Butlin RK, *et al.* 2017.** A genomic footprint of hybrid zone movement in crested newts. *Evolution Letters* **1**: 93–101.

**Yang Y, Zhou T, Duan D, Yang J, Feng L, Zhao G. 2016.** Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Frontiers in Plant Science* **7**: 959.

**Yang Y, Zhou T, Yang J, Meng X, Zhu J, Zhao G. 2017.** The complete chloroplast genome of *Quercus baronii* (*Quercus*, L.). *Mitochondrial DNA Part B* **28**: 290–291.

**Young ND, Healy J. 2003.** GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinformatics* **4**: 6.

**Zeng YF, Liao WJ, Petit RJ, Zhang DY. 2011.** Geographic variation in the structure of oak hybrid zones provides insights into the dynamics of speciation. *Molecular Ecology* **20**: 4995–5011.

**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 BW, Xu LL, Li N, *et al.* 2019.** Phylogenomics reveals an ancient hybrid origin of the Persian walnut. *Molecular Biology and Evolution* **36**: 2451–2461.

**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 and 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 and Genomes* **14**: 88.

**Zhou Y, Duvaux L, Ren G, Zhang L, Savolainen O, Liu J. 2017.** Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions. *Heredity* **118**: 211–220.

**Zohren J, Wang N, Kardailsky I, *et al.* 2016.** Unidirectional diploid-tetraploid introgression among British birch trees with shifting ranges shown by restriction site-associated markers. *Molecular Ecology* **25**: 2413–2426.