Botanical Journal of the Linnean Society, 2024, **XX**, 1–13
https://doi.org/10.1093/botlinnean/boad077
Advance access publication 10 February 2024
Original Article

OXFORD

Original Article

Allopatric speciation and secondary sympatry of *Fagus longipetiolata* and *F. lucida* (Fagaceae) in subtropical China

Danqi Li[2,3,‡], Lu Jiang[2,‡], Wei He[2], Dengmei Fan[1,2], Shanmei Cheng[2], Yi Yang[2], Meixia Wang[2], Shaoqing Tang[1], Yixuan Kou[1,2,*], and Zhiyong Zhang[1,2,4,*]

[1]Key Laboratory of Ecology of Rare and Endangered Species and Environmental Protection (Guangxi Normal University), Ministry of Education, Guilin 541004, China
[2]Laboratory of Subtropical Biodiversity, Jiangxi Agricultural University, Nanchang, Jiangxi 330045, China
[3]Lushan Botanical Garden, Jiangxi Province and Chinese Academy of Sciences, Jiujiang 332900, China
[4]Guangxi Key Laboratory of Landscape Resources Conservation and Sustainable Utilization in Lijiang River Basin, Guilin 541006, China

‡These authors contributed equally.

*Corresponding author. Key Laboratory of Ecology of Rare and Endangered Species and Environmental Protection (Guangxi Normal University), Ministry of Education, Guilin 541004, China. E-mail: pinus-rubus@163.com or zhangzy@gxnu.edu.cn; yixuankou@163.com

ABSTRACT

Inferring the geographic mode of speciation is essential for explaining the generation and assembly of biodiversity, but has been rarely applied to the temperate flora of China. *Fagus longipetiolata* and *F. lucida* are a sister pair of beech species with largely overlapping ranges in subtropical China, however, little is known about the geographic mode of speciation and the formation of their sympatric distributions. In this study, we used IMa2 and fastsimcoal2 to simulate the speciation history of the two sisters. On the basis of 11 nuclear loci screened for 27 populations/species, their divergence time was estimated to 8.37 Mya by IMa2, and weak interspecific gene flow was detected. The simulation by fastsimcoal2 found that *F. longipetiolata* and *F. lucida* diverged ~8.53 Mya, then came into contact and hybridized at 0.8 Mya. The extended Bayesian skyline plots indicated that *F. lucida* began demographic expansion prior to the Pleistocene at 3 Mya, whereas *F. longipetiolata* increased its effective population size during the Quaternary, between 0.8 and 0.5 Mya. Ecological niche analysis and niche modelling showed that the two beeches had subtle niche differentiation and their projected distributions were largely overlapped across the late Quaternary. These results, coupled with well-studied fossil records in *Fagus*, demonstrate that these two beech species diverged allopatrically at the high latitudes of the Northern Hemisphere, and came into contact and hybridized with each other after successive migrations into subtropical (i.e. warm temperate) China recently. However, the two beeches may have developed strong intrinsic reproductive barriers due to long-term allopatry, allowing them to coexist in subtropical China. Our results illuminate the assembly history of the temperate woody flora in subtropical China, supporting that many relict temperate lineages in subtropical China may have established their populations recently.

Keywords: allopatric speciation; coalescent simulation; *Fagus lucida*; *Fagus longipetiolata*; secondary sympatry; subtropical China.

INTRODUCTION

Understanding the sympatry/coexistence of closely related species lies at the nexus of disentangling how historical and ecological factors govern patterns of biodiversity (Weber and Strauss 2016). Sympatry or coexistence of closely related species can either result from a sympatric speciation process or from secondary contact due to range expansion after speciation (Ricklefs 2010, Mittelbach and Schemske 2015). Thus, inferring the geographic mode of speciation could be instrumental in revealing the evolutionary and ecological mechanisms that underlie the generation and assembly of biodiversity (Mittelbach and Schemske 2015, Skeels and Cardillo 2019). Allopatric speciation is the most common mode of speciation, when incipient divergence between populations is fostered by geographic isolation that precludes interspecific gene flow (Mayr 1942, Coyne and Orr 2004). Although some evidence suggests that sympatric speciation exists in nature (Schliewen *et al.* 1994, Savolainen *et al.* 2006, Papadopulos *et al.* 2011), the requirements for sympatric speciation are very stringent (species largely sympatric, reproductive isolation, sister relationships, and allopatric phases are unlikely, Coyne and Orr 2004, Foote 2018). In addition, sympatric speciation necessitates that

Received 15 August 2023; revised 30 October 2023; accepted 6 December 2023

populations establish sufficient ecological differences to coexist before and after reproductive isolation developed except for the case of allopolyploidization and resultant genetic incompatibility (Coyne and Orr 2004, Schuler *et al.* 2016).

Traditionally, a common approach to examine geographic mode of speciation is to apply age range correlations, in which time since divergence is regressed against range overlap (e.g. Barraclough and Vogler 2000, Losos and Glor 2003, Fitzpatrick and Turelli 2006, Anacker and Strauss 2014, Hodge and Bellwood 2016). A similar framework examines the correlation of the age and niche divergence of sister pairs (e.g. Tobias *et al.* 2014). However, such approaches cannot distinguish sympatric speciation from secondary contact after allopatric speciation, providing mixed evidence for the geographic mode of speciation (Losos and Glor 2003, Dong *et al.* 2020 but see Skeels and Cardillo 2019). Ideally, studies that attempt to disentangle mechanisms driving the formation of sympatry should explicitly test for the modes of geographic speciation (Warren *et al.* 2014). One of the most rigorous approaches is to apply a full phylogeographic framework to reconstruct the speciation history of closely related species (e.g. Pettengill and Moeller 2012, Dong *et al.* 2020). In those studies, coalescent-based simulations such as the isolation with migration model (IM; Hey 2010a, b), Approximate Bayesian computation (Csilléry *et al.* 2010), and faster continuous-time sequential Markovian coalescent algorithm (fastsimcoal2; Excoffier *et al.* 2021) have been used to distinguish various geographic speciation modes by assessing the presence of gene flow. Such studies, when complemented with ecological niche modelling or species distribution modelling (SDM), which identifies potential historical geographic ranges without inferring genealogical relationships, can provide deep insights into the speciation geography and the formation of secondary sympatry (Carstens and Richards 2007, Chan *et al.* 2011, Pettengill and Moeller 2012).

China is one of the richest countries regarding temperate plant diversity and one of the world's 17 mega-diversity countries (Mittermeier *et al.* 1997, Lu *et al.* 2018, Ding *et al.* 2020). One plausible explanation for the high temperate species richness is that China (especially subtropical China or warm temperate China) serves as a museum for many relict temperate woody lineages derived from the ancient Arcto-Tertiary flora that once inhabited in high latitudes of the Northern Hemisphere (Wolfe 1980, Manchester and Tiffney 2001). However, recent studies suggest that east China (central and southeast China) may have served as both a museum and a cradle for woody genera (López-Pujol *et al.* 2011, Lu *et al.* 2018). Therefore, how and when the major components of the temperate woody flora of China assembled to form the present-day vegetation remains a hot issue debated among biogeographers and botanists (Chen *et al.* 2018). To our knowledge, however, the approach of speciation geography has been rarely applied to elucidating the evolutionary history of the temperate woody flora of China.

Fagus L. (Fagaceae), a small group containing at least 11 broad-leaved deciduous tree species, is one of the most typical genera in perhumid temperate forests of the Northern Hemisphere (Shen 1992, Fang and Lechowicz 2006, Jiang *et al.* 2022). China is one of the distribution centre of the genus, comprising four species (Li *et al.* 2023). They often grow in the moist warm temperate forests of mountainous areas in subtropical China (Liu *et al.* 2003, Guo and Werger 2010, Li *et al.* 2023). The high species richness of *Fagus* in China has been interpreted as a refuge accumulation (museum) of lineages that had once inhabited higher latitudes of the Northern Hemisphere before the Miocene global cooling (Denk and Grimm 2009, Jiang *et al.* 2022). This notion implies that the precursors of Chinese beech species might have speciated allopatrically and their sympatric distributions might have established through southward migration into subtropical China and subsequent range expansion during the late Cenozoic. However, mountains in China have always been considered to be cradles of biological species (López-Pujol *et al.* 2011, Ding *et al.* 2020). Sympatric speciation has been proved to be highly likely in mountainous areas for plant species (Osborne *et al.* 2020) because mountains are characterized by steep changes of the physical and biotic environment that can promote adaptive divergence and ecological speciation (Abbott and Brennan 2014). It remains plausible that the overlapping distributions of Chinese beeches might be accomplished through sympatric (or parapatric) speciation and subsequent range expansions.

To elucidate the speciation history of Chinese beech species and shed new light on the assembly of the temperate flora of China, we chose to study the speciation history of two co-distributed beeches in China: *Fagus longipetiolata* and *F. lucida*. Morphologically, *F. longipetiolata* and *F. lucida* differ mainly in the size of cupule (1.29–2.45 vs. 0.52–1.26 cm), the length of peduncle (3.61–8.29 vs. 0.32–1.02 cm), the size and shape of bract of cupule (0.34–0.74 cm long linear vs. 0.05–0.08 cm short tuberculate and closely appressed) (Li *et al.* 2023). Little morphological difference has been observed in their flowers and catkins and the two species bloom in April and May synchronously. While *F. longipetiolata* always grows at lower altitudes mixed with evergreen broad-leaved trees, *F. lucida* is the dominant species in deciduous broad-leaved forests at higher elevations. However, they occur occasionally in mixed stands (Zhang *et al.* 2013). Our previous phylogeographic study found that the two species share cpDNA haplotypes extensively (Zhang *et al.* 2013), suggestive of the possibility of interspecific hybridization and introgression between the two species (Zhou *et al.* 2022). In addition, our phylogenetic investigation found that *F. longipetiolata* and *F. lucida* are the only sister species with overlapping distributions within *Fagus* (Jiang *et al.* 2022). Given their close phylogenetic affinity, sympatric distribution, and a possible complex hybridization history, this system may represent an ideal system for studying the beech's speciation and the assembly of beech communities in China. In this study, we simulated the speciation history of *F. longipetiolata* and *F. lucida* across their entire ranges based on multilocus nuclear sequences. Combined with ecological niche analysis and niche modelling, we address two questions: (i) what is the geographic mode of speciation of these two sister beeches? (ii) How did their sympatric distribution form?

MATERIALS AND METHODS

Sampling and DNA sequencing

We collected 157 individuals from 27 populations of *Fagus longipetiolata* and 154 individuals from 27 populations of *F. lucida* across their geographic ranges in subtropical China, each species containing 13 sympatric and 14 allopatric populations

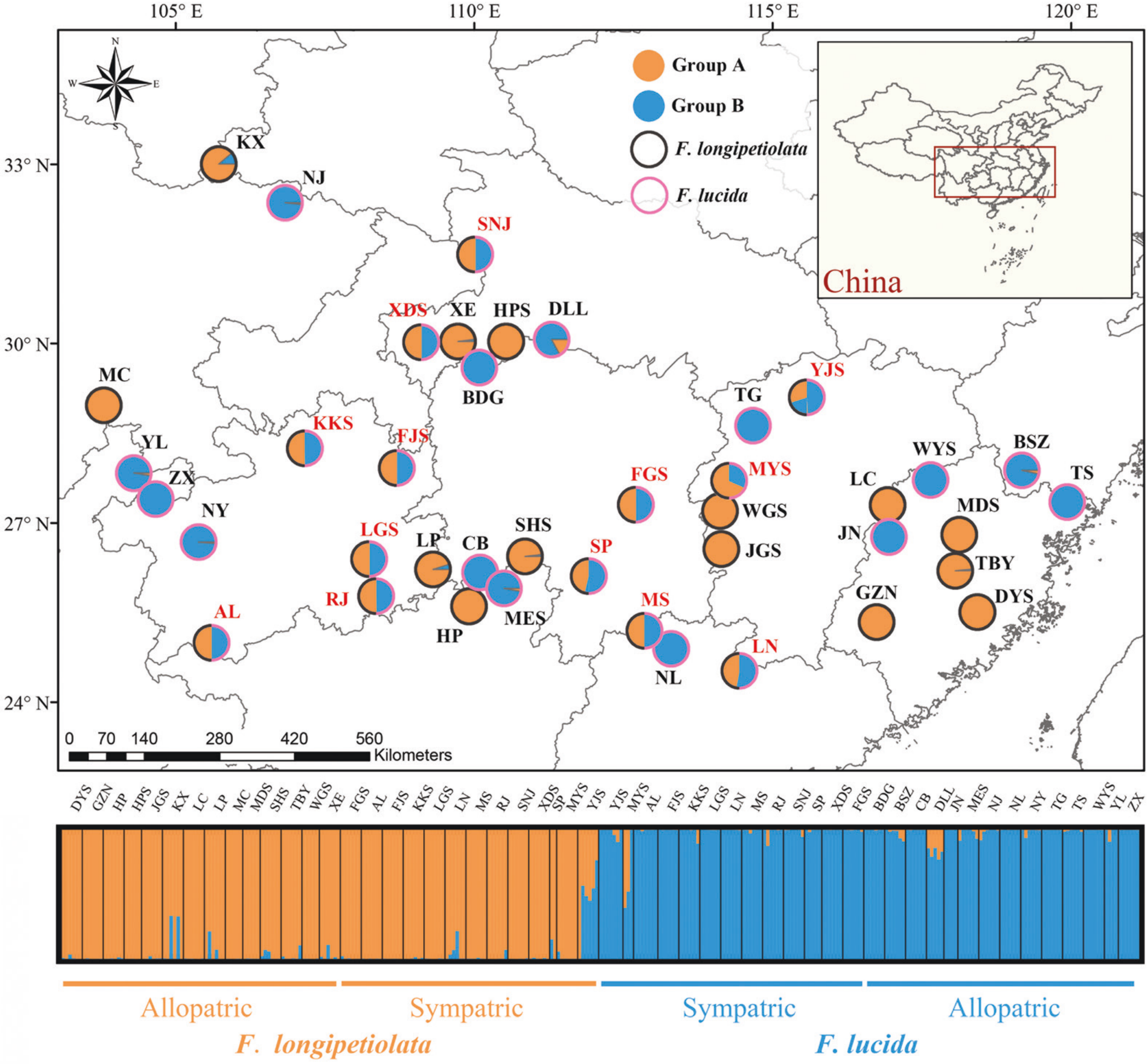

Figure 1. Distribution of genetic components (orange/light and blue/dark colours) of *Fagus longipetiolata* and *F. lucida* (top) revealed by STRUCTURE analysis when $K = 2$ (bottom), which was determined to be the optimal K value using the ΔK method. Black/dark cycle: *F. longipetiolata*; pink/light cycle: *F. lucida*; bi-coloured cycle (black and pink): sympatric populations of the two species. The red lettering on the map indicates the locations of sympatric populations.

(Fig. 1, Table S1, Supporting Information). Fresh leaves or bark tissues (only six individuals from SNJ) of six individuals were collected for each population except for those with very small population sizes. The samples were then dried in silica gel immediately. All sampled individuals from each population were spaced at least 50–100 m apart, and all voucher specimens were deposited in the Herbarium of Jiangxi Agricultural University (Table S1, Supporting Information).

Total DNA was extracted from leaves or bark tissues using a modified cetyltrimethylammonium bromide protocol (Doyle and Doyle 1987). Eleven low-copy nuclear genes (F138, F159, F253, F286, F289, P4, P49, P50, P52, P54, and P97, Table S2, Supporting Information) from a 28-gene set that had been developed in our previous study (Jiang *et al.* 2022) were screened and sequenced. The markers were selected to ensure the highest possible discrimination capacity between the two species. Protocols of DNA extraction, PCR amplification, and DNA sequencing are detailed in Jiang *et al.* (2022).

Raw chromatograms were checked and edited using Sequencher v.5.4.6 (GeneCodes Corporation, Ann Arbor, MI, USA), aligned using BioEdit v.7.2 (Hall 1999), and then refined manually in MEGA v.7.0 (Kumar *et al.* 2016). The sequences were phased in DNAsp v.5.10 (Librado and Rozas 2009) for performing downstream analyses. All sequences have been deposited in GenBank under the accession numbers OR106156–OR112581.

Sequence variation, genetic diversity, and neutrality test

After sequence phasing using the phase algorithm in DNAsp v.5.10 (Librado and Rozas 2009), we used this program to estimate the population genetic parameters of both species for each nuclear locus, including nucleotide diversity (π) (Tajima 1983), Watterson's parameter (θ_w) (Watterson 1975), the number of segregating sites (S), haplotype diversity (H_d), the number of haplotypes (N_h), and the minimum number of recombinant events (R_m).

Neutrality tests were performed for each nuclear locus using Tajima's D (Tajima 1989), Fu and Li's D^* and F^* (Fu and Li 1993), and Fay and Wu's H (Fay and Wu 2000) test in DNAsp v.5.10. We further performed multilocus Hudson–Kreitman–Aguadé (HKA) (Hudson *et al.* 1987) test to assess the fit of data to neutral equilibrium in DNAsp v.5.10. Finally, the maximum frequency of derived mutations (MFDM) (Li 2011) test was used to examined the likelihood of natural selection occurred at individual locus. Sequences of *F. engleriana* belong to subgen. *Engleriana* were downloaded from Jiang *et al.* (2022) and were used as an outgroup when performing Fay and Wu's H, HKA, and MFDM tests.

Genetic differentiation and population structure

Wright's fixation index (F_{ST}) and hierarchical analysis of molecular variance between *Fagus longipetiolata* and *F. lucida* were estimated for each locus in Arlequin v.3.5 (Excoffier and Lischer 2010). Genetic differentiation (F_{ST}) within species and among sympatric and allopatric populations of the two species was also estimated separately. The significance of F_{ST} was tested by 10 000 permutations of sequences among species and populations. Furthermore, the genealogical relationships of nuclear haplotypes at each locus were constructed using the median-joining network method in POPART v.1.7 (Leigh and Bryant 2015).

To examined the genetic structure of two beech species and to detect possible hybridization events, we implemented the Bayesian assignment algorithm in STRUCTURE v.2.3.4 (Hubisz *et al.* 2009) with the admixture model using segregating sites without significant linkage disequilibrium after Bonferroni correction. Twenty independent runs were performed for each number of clusters (K) from 1 to 27 with burn-ins of 20 000 and 200 000 iterations, respectively. The most likely number of clusters was estimated using LnP(D) (Pritchard *et al.* 2000) and ΔK statistics (Evanno *et al.* 2005). The population cluster graphics were visualized by DISTRUCT v.1.1 (Rosenberg 2004).

Inference of speciation and demographic history

Before these analyses, we omitted the admixed individuals detected in STRUCTURE analysis (those with the ancestry coefficient, q, below 0.05 or above 0.95). To transform the time estimates into years, we calculated the mutation rate according to $\mu = K_S/2T$, where K_S is the average divergence at silent sites between *Fagus longipetiolata*/*F. lucida* and outgroup *F. engleriana*, and T is the divergence time between subgen. *Fagus* and subgen. *Engleriana* (~32.7 Mya) based on fossil record (*F. pacifica* Chaney, which has affinity to modern taxa of subgen. *Fagus* Meyer and Manchester 1997, Denk and Grimm 2009, Jiang *et al.* 2022). Ultimately, the resulting geometric average mutation rate was 2.56×10^{-10} per site per year and the generation time of beeches was set to 50 (Merzeau *et al.* 1994).

The demographic history of *F. longipetiolata* and *F. lucida* was each estimated using the Extended Bayesian skyline plot (EBSP) analysis in BEAST v.2.6.2 (Bouckaert *et al.* 2014). We used JC substitution model determined by jModeltest v.2.1.7 (Darriba *et al.* 2012) and set the MCMC to 100 million steps with sampling every 100 000 steps, then set the burn-in at 20%. The program Tracer v.1.7.1 (Rambaut *et al.* 2018) was used to check and ensure the effective sample size not less than 200. The result was visualized in R v.4.1.0 (R Core Team 2021) using the 'plotEBSP.R' script download from the Beast2 website (http://www.beast2.org/tutorials/).

We first used the IM model in the IMa2 software (Hey 2010a, b) to infer the speciation history between the two species. After extracting the non-recombining region of each locus using DNAsp v.5.10 (Librado and Rozas 2009, Strasburg and Rieseberg 2010), the simulations ran in different random seeds with 2 000 000 steps and 100 000 burn-in steps under the HKY mutation model. Second, speciation history was further explored using the fastsimcoal2 (Excofffier *et al.* 2021) with the two-dimensional joint site frequency spectrum (2D-SFS) as input file generated by Arlequin v.3.5. We inferred the best speciation model among the four classical models proposed by Nielsen and Wakeley (2001): the ancestral population of size N_A split in the past to form two populations of size N_{lo} (*F. longipetiolata*) and N_{lu} (*F. lucida*) at time t_1, then (i) there is no gene flow between species (Model 1, divergence without gene flow); (ii) there is continuous bi-direction gene flow (Model 2, isolation with gene flow); (iii) incipient bi-directional gene flow ceases at time t_2 (Model 3, incipient gene flow followed by isolation); and (iv) secondary contact and bi-directional gene flow starts at time t_3 (Model 4, secondary contact following isolation; Fig. S1, Supporting Information). Then, 100 000 coalescent simulations and 20 loops of the likelihood for a set of models were performed in 100 independent runs. The best model was identified by the Akaike's information criterion (AIC) and the maximum Akaike's weight (w_i) (Excoffier *et al.* 2013), and 95% confidence intervals (CIs) for the parameters were estimated by bootstrapping 100 2D-SFS.

Ecological divergence and niche modelling

A total of 107 occurrence records of *Fagus longipetiolata* and 62 of *F. lucida* retrieved from Chinese Virtual Herbarium (https://www.cvh.ac.cn/) and our previous research (Zhang *et al.* 2013) were used for niche divergence analysis and niche modelling (Table S3, Supporting Information). We performed a principal component analysis (PCA) on 19 climate variables of the occurrence sites obtained from WorldClim database (http://www.worldclim.org/) (Table S4, Supporting Information). Then, a non-parametric Kruskal–Wallis test was used to assess the differences of each ecological variable of the two species and displayed in kernel density plots by R v.4.1.0 (RCoreTeam 2021).

Nineteen bioclimatic variables were downloaded from the WorldClim database, with a 2.5 arc minute resolution in four periods: the present, the Last Glacial Maximum (LGM), the Mid-Holocene (MH), and the Last Interglacial (LIG). For MH and LGM, we used paleoclimate data based on the MIROC-ESM model. Only environmental variables with a pairwise Pearson correlation coefficient $r \leq 0.7$ were retained in niche modelling. We used the default parameters in MAXENT v.3.4.1

(Phillips *et al.* 2006) and included 80% of the species records for training and 20% for testing the model. The model with an area under the ROC curve (AUC) value above 0.9 was considered better discrimination. The tool 'Distribution changes between binary SDMs' in SDM toolbox v.2.5 (Brown *et al.* 2017) was used to calculate the co-distribution and private distribution areas between these two species for different periods. The statistics of Schoener's *D* and standardized Hellinger distance (*I*) (Schoener 1968) were calculated by niche overlap and identity tests in ENMTools v.1.4.4 (Warren *et al.* 2010).

RESULTS

Genetic diversity and neutrality test

Eleven low-copy nuclear genes were sequenced and aligned for 311 individuals from 54 populations of *Fagus longipetiolata* and *F. lucida.* The total length of the aligned genes was 4858 bp with 325 SNPs. A 4 bp long indel was detected in P97 and excluded in subsequent analyses. The average nucleotide diversity (π), Watterson's parameter (θ_w), number of segregating sites (S), haplotype diversity (H_d), the number of haplotypes (N_h), and the minimum number of recombinant events (R_m) across 11 nuclear genes were similar between these two species ($\pi = 0.00416$, $\theta_w = 0.00732$, $S = 20$, $H_d = 0.604$, $N_h = 23$, $R_m = 3$ in *F. longipetiolata*, and $\pi = 0.00454$, $\theta_w = 0.00693$, $S = 18$, $H_d = 0.659$, $N_h = 25$, $R_m = 4$ in *F. lucida*) (Table S5, Supporting Information).

In each species, negative values of Tajima's *D*, Fu and Li's *D** and *F** were detected at most of the 11 nuclear loci and several showed a significant level (Table S6, Supporting Information). MFDM tests detected the likelihood of natural selection acting on four loci (F159, P4, P50, and P54), but the multilocus HKA test found no loci deviated from neutrality in both species. Taken together, no loci deviated from neutral model in all neutrality tests, thus were not considered to be under selection pressure.

Interspecific differentiation and genetic structure

Overall interspecific genetic differentiation across the 11 nuclear loci was highly variable, with the lowest of $F_{ST} = 0.116$ for F286 and the highest of $F_{ST} = 0.832$ for P50 (Table S7, Supporting Information). The intraspecific genetic differentiation was relatively moderate, with $F_{ST} = 0.061$–0.213 in *F. longipetiolata* and $F_{ST} = 0.049$–0.385 in *F. lucida* (Table S8, Supporting Information). Furthermore, interspecific genetic differentiation was slightly higher in sympatric populations ($F_{ST} = 0.410$) than in allopatric populations ($F_{ST} = 0.393$; Table S8, Supporting Information). Haplotype genealogies constructed by median-joining network showed that haplotype sharing was ubiquitous between *Fagus longipetiolata* and *F. lucida* in all 11 loci (Fig. S2, Supporting Information). However, most shared haplotypes positioned centrally, indicating they are retained ancestral polymorphisms. Occasional sharing of derived haplotypes occurred in some loci (e.g. F138, F159, F286, F289, P4, P49, P54; Fig. S2, Supporting Information), which could derive from introgressive hybridization between the two species.

STRUCTURE analyses showed that Δ*K* was the highest when $K = 2$ but Ln*P*(*D*) increased slowly after $K > 2$ (Fig. S3A, Supporting Information). Therefore, $K = 2$ may be the most likely number of genetic clusters, which clearly corresponds to the two species. Some individuals displayed admixed ancestry and this pattern was a bit more frequent in sympatric populations (6.3% for *F. longipetiolata* and 3.5% for *F. lucida*) than in allopatric populations (1.8% for *F. longipetiolata* and 2.2% for *F. lucida*) (Fig. 1). No meaningful substructure was detected within each species (Fig. S3B, Supporting Information), possibly due to the small number of SNPs or prevalent pollen-mediated gene flow in wind-pollinated tree species (e.g. Bai *et al.* 2014, Kou *et al.* 2020).

Demographic history and speciation history

The EBSP analyses showed that the two species experienced different demographic histories. The effective population size of *F. longipetiolata* began to increase abruptly at 0.8 Mya and followed by a sharp decrease from 0.5 Mya to the present day. However, the effective population size of *F. lucida* increased steadily since *c.* 3 Mya (Fig. 2).

The initial divergence time between *F. longipetiolata* and *F. lucida* was estimated to the Late Miocene (8.37 Mya, 95% HPD: 5.89–11.20 Mya) by IMa2 (Table S9, Supporting Information). The effective population size (N_e) of *F. lucida* (2.47×10^5, 95% HPD: 2.08–2.87 × 10^5) was estimated as 25% larger than that of *F. longipetiolata* (1.95×10^5, 95% HPD: 1.63–2.28 × 10^5) and both were about twice as large as their ancestral populations (0.74×10^5, 95% HPD: 0.38–1.13 × 10^5). Migration rate (2*Nm*) was asymmetrical, with gene flow higher from *F. lucida* to *F. longipetiolata* (0.85, 95% HPD: 0.55–1.15) than that in the reverse direction (0.52, 95% HPD: 0.24–0.82) (Fig. 3D, Table S9, Supporting Information).

Among the four speciation models, the fastsimcoal2 simulation found that model 4 (secondary contact following isolation) had the lowest AIC values (AIC = 2224.08 and $w_i = 1$; Table S10, Supporting Information), The simulation gave the same Late Miocene estimate for the initial divergence (8.53 Mya, 95% CI: 8.46–8.60), then secondary contact and hybridization occurred in mid-Pleistocene at 0.8 Mya (95% CI: 0.77–0.83) following a long period of isolation (Fig. 3D, Table 1). Migration rate (*m*) from *F. lucida* to *F. longipetiolata* (4.15×10^{-6}, 95% CI: 4.03–4.27 × 10^{-6}) were higher than vice versa (3.22×10^{-6}, 95% CI: 3.12–3.32 × 10^{-6}). The estimated effective population size was smaller for *F. lucida* (1.91×10^5, 95% CI: 1.89–1.93 × 10^5) and only slightly larger than that of *F. longipetiolata* (1.69×10^5, 95% CI: 1.67–1.71 × 10^5). Both species had greater N_e than their ancestral populations (1.38×10^5, 95% CI: 1.34–1.41 × 10^5).

Niche similarity between *F. longipetiolata* and *F. lucida* and distribution changes over time

The non-parametric Kruskal–Wallis tests indicated that only seven temperature-related bioclimatic variables (Bio01, annual mean temperature; Bio05, max temperature of warmest month; Bio06, min temperature of coldest month; Bio08, mean temperature of wettest quarter; Bio09, mean temperature of driest quarter; Bio10, mean temperature of warmest quarter; and Bio11, mean temperature of coldest quarter) showed significant differences between *F. longipetiolata* and *F. lucida* (Fig. S4, Supporting Information). The first two components explained 68.7% of the total bioclimatic variance (PC1 37.9% and PC2

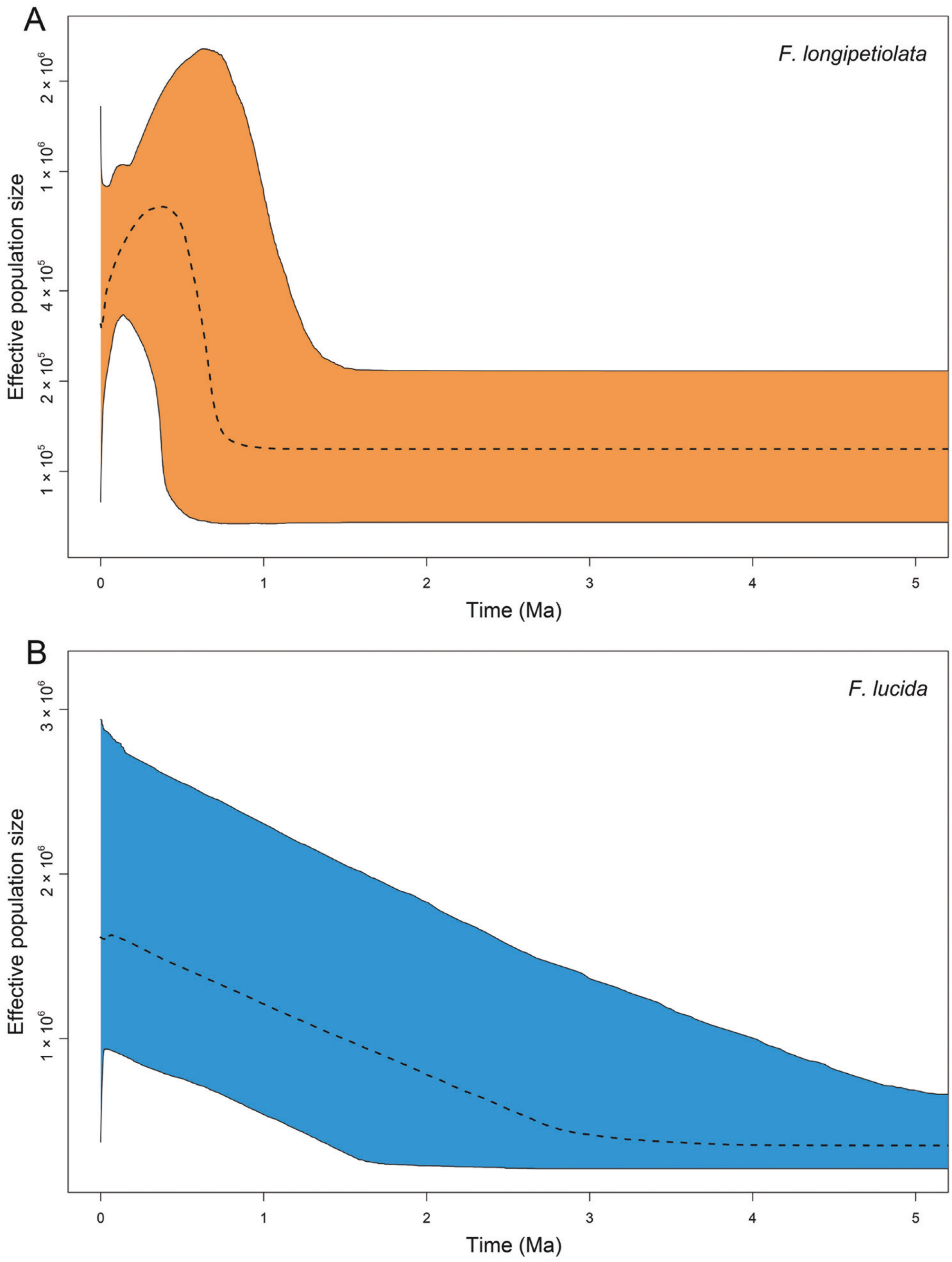

Figure 2. Extended Bayesian skyline plot (EBSP) representing historical changes in effective population size (N_e) of A, *Fagus longipetiolata* and B, *F. lucida*. The *x*-axis represents time in units of million years; the *y*-axis represents the effective population size. Mean estimates of N_e are indicated by a dashed line; the thin solid lines delineate the 95% highest posterior density.

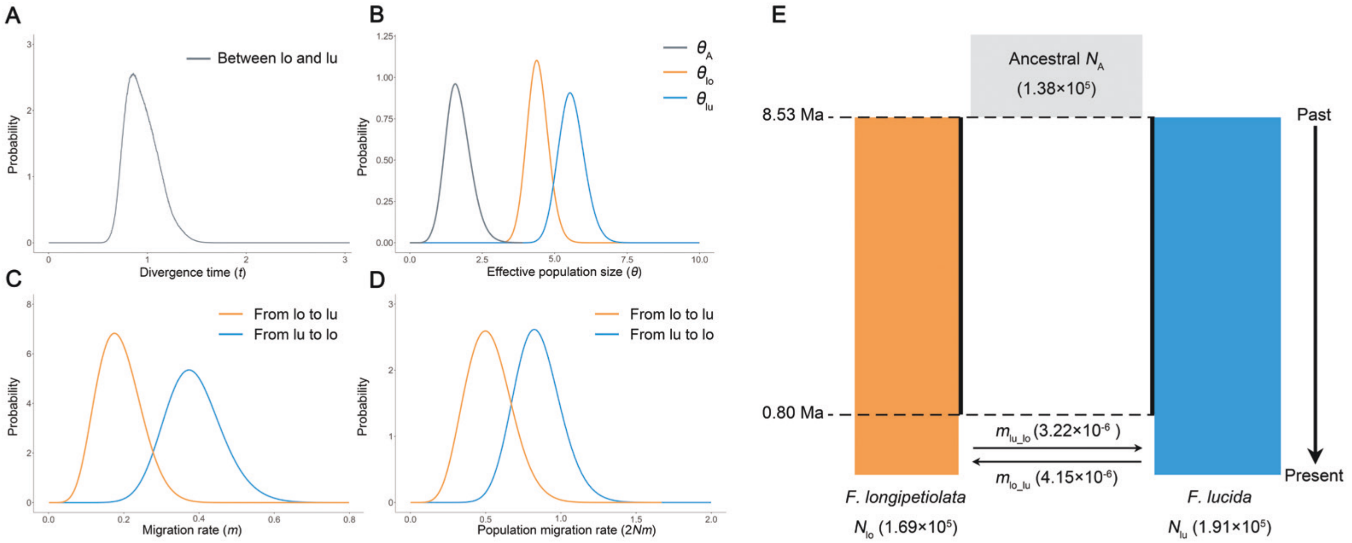

Figure 3. Posterior probability distribution of divergence time, effective population size, and migration rate of *Fagus longipetiolata* and *F. lucida* estimated in IMa2 (A–D) and schematic representation of the best demographic model simulated by fastsimcoal2 (E). A, Divergence time between two species. B, Effective population size for *F. longipetiolata* (θ_{lo}), *F. lucida* (θ_{lu}), and their common ancestor (θ_A). C, Migration rate (*m*). D, Population migration rate (2*Nm*) from *F. longipetiolata* to *F. lucida* (from lo to lu), and the reverse migration rate (from lu to lo). E, Dashed lines are time points at which populations split (8.53 Mya) and secondary contact and gene flow started (0.80 Mya). Arrows represent migration rate per generation from *F. longipetiolata* to *F. lucida* (m_{lu_lo}) and the revise direction (m_{lo_lu}). The vertical solid black lines represent strict isolation. The current population sizes of *Fagus longipetiolata* and *F. lucida* are represented by N_{lo} and N_{lu}, respectively, and population size of their ancestor is denoted by N_A. Divergence and secondary contact times, effective population sizes, and migration rates corresponding to 95% CIs obtained from this model are shown in Table 1.

Table 1. Demographic parameters and their 95% confidence intervals (CIs) from the best model (model 4) for the two *Fagus* species by fastsimcoal2.

Parameter	N_{lo}	N_{lu}	N_A	m_{lo_lu}	m_{lu_lo}	T_A (Ma)	T_{recent} (Ma)
Mode	1.69×10^5	1.91×10^5	1.38×10^5	4.15×10^{-6}	3.22×10^{-6}	8.53	0.80
95% CI lower	1.67×10^5	1.89×10^5	1.34×10^5	4.03×10^{-6}	3.12×10^{-6}	8.46	0.77
95% CI upper	1.71×10^5	1.93×10^5	1.41×10^5	4.27×10^{-6}	3.32×10^{-6}	8.60	0.83

N_{lo}, effective population size of *F. longipetiolata*.
N_{lu}, effective population size of *F. lucida*.
N_A, effective population size of ancestral population.
m_{lo_lu}, migration rate from *F. lucida* to *F. longipetiolata*.
m_{lu_lo}, migration rate from *F. longipetiolata* to *F. lucida*.
T_A, time since species divergence.
T_{recent}, time of secondary contact starts.

30.8%), however, PCA did not find obvious climatic niche differentiation between *F. longipetiolata* and *F. lucida* in the first two components (Fig. S5, Supporting Information).

In the MAXENT, seven bioclimatic variables with $r \leq 0.7$ (Bio02, mean diurnal range; Bio04, temperature seasonality; Bio06, min temperature of coldest month; Bio08, mean temperature of wettest quarter; Bio10, mean temperature of warmest quarter; Bio14, precipitation of driest month; and Bio16, precipitation of wettest quarter) were retained in each species and all niche models with high AUC values (> 0.97). The models predicted that the distributions of *F. longipetiolata* and *F. lucida* were relatively stable across all periods (present day, MH, LGM, and LIG) (Fig. S6, Supporting Information). In each period, their distributions were overlapped to some extent (Fig. 4). The results of ENMTOOLS showed that the observed measures of niche similarity (*D* and *I*) were consistent with null distributions for *F. longipetiolata* and *F. lucida*, suggesting niche differentiation between the two species is weak (Fig. S7, Supporting Information).

DISCUSSION

Allopatric speciation and secondary sympatry are the most likely scenario for *Fagus longipetiolata* and *F. lucida*

Estimating the extent and timing of gene flow during divergence is crucial to understand the geographic model of speciation (Dong *et al.* 2020, Muto and Kai 2023, Qin *et al* 2023). The presence of gene flow at the initial stage suggests that speciation was initially driven by divergent selection without a geographic barrier (sympatric or parapatric speciation), whereas the complete absence of gene flow suggests allopatric speciation with or without the effect of selection (Nosil 2012, Sousa and Hey 2013). Allopatric divergence can be confounded by secondary sympatry due to range shift, however, the two scenarios can be distinguished from

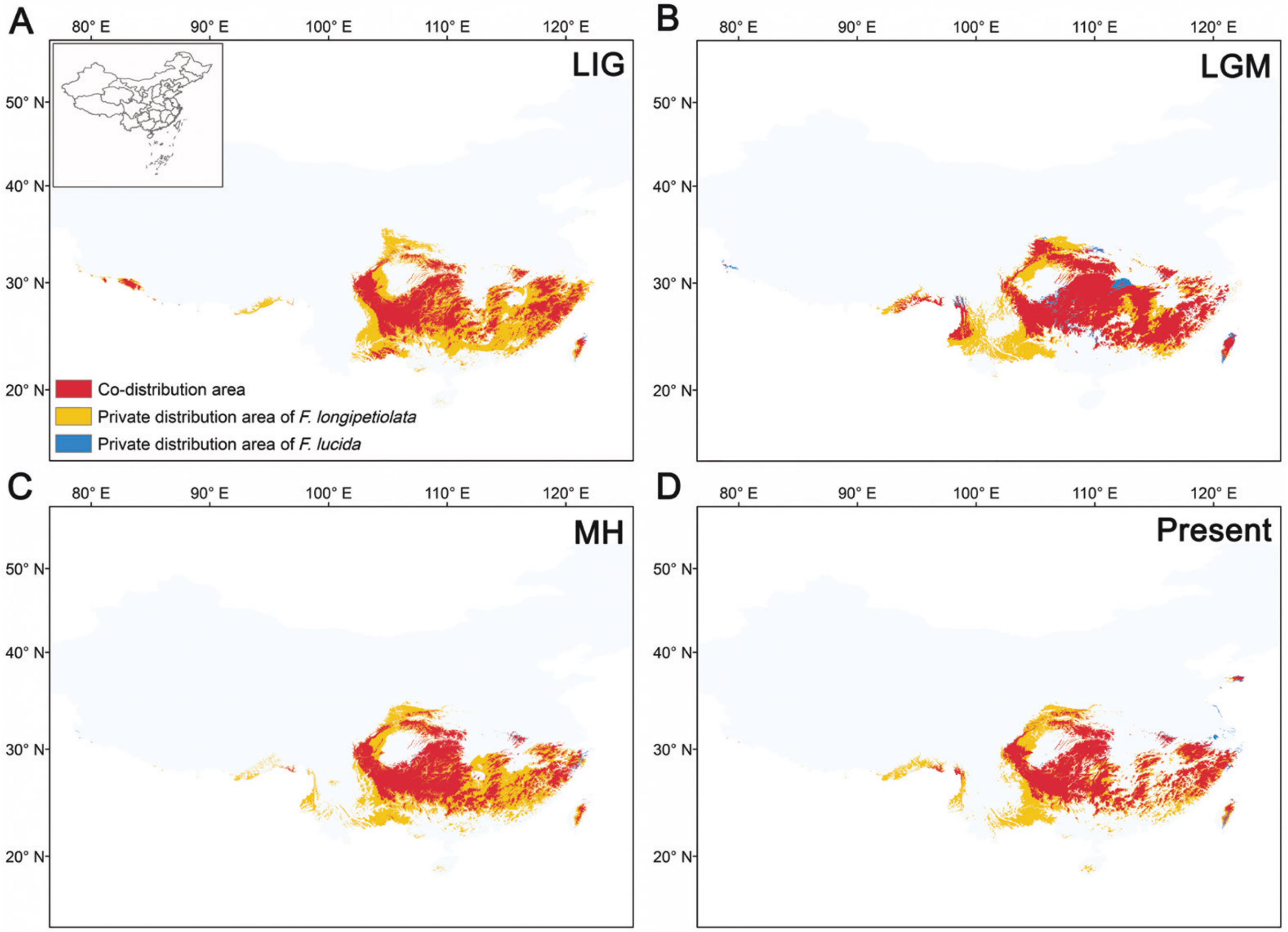

Figure 4. The distribution overlapping of *Fagus longipetiolata* and *F. lucida* across different periods inferred from niche modelling in MAXENT: A, the Last Interglacial; B, the LGM; C, the MH; and D, present day.

each other by the late timing of gene flow (Sousa and Hey 2013). In this study, we found that *Fagus longipetiolata* and *F. lucida* initially diverged in the Late Miocene (~8.5 Mya by fastsimcoal2 and IMa2) (Fig. 3; Tables 1, S9, Supporting Information), however, gene flow between the two sympatric beeches only happened during the mid-Pleistocene (0.8 Mya by fastsimcoal2) (Fig. 3D; Table 1). These results suggest that the two species could have speciated allopatrically and came into contact with each other recently. This finding is consistent with the estimation of Weir and Price (2011) that secondary sympatry generally requires millions of years to achieve before complete reproductive isolation evolves in allopatry. Note that the divergence time of *F. longipetiolata* and *F. lucida* estimated in this study is older than that (5.8 Mya) estimated in our phylogenetic study (Jiang *et al.* 2022) but much younger than the late Eocene/early Oligocene estimates of Renner *et al.* (2016) using a fossilized birth-death dating approach and a set of 53 fossils. The younger age in Jiang *et al.* (2022) can be explained by secondary gene flow during the secondary contact and lack of discriminative signals in some nuclear loci (Cardoni *et al.* 2022). The much higher estimates in Renner *et al.* (2016) may be caused by the two moderately to high-divergent noncoding regions, one of which shows a deep, profound dimorphism (cf. Denk *et al.* 2005). In spite of this incongruence, our phylogenetic and population genetic studies reconcile with each other that *F. longipetiolata* and *F. lucida* represent relatively deep diverged lineages.

Although a scenario of allopatric speciation and secondary sympatry is indicated in this study, it is interesting to determine where the two species originated because of the ubiquitous range shift during the late Cenozoic climate changes (Davis and Shaw 2001). It is equally possible that present sympatric distribution of the two species is achieved through range expansion after *in situ* allopatric speciation by vicariance within the mountain ranges of subtropical China, or through immigrations after allopatric speciation outside this region. Fortunately, beeches have an exceptionally well-studied fossil record extending back to the early Cenozoic (Denk and Grimm 2009, Renner *et al.* 2016), which offers an exceptional opportunity to examine the historical biogeography (Denk and Grimm 2009, Jiang *et al.* 2022) as well as the speciation geography of the genus. By integrating fossil records into historical biogeography, Denk and Grimm (2009) and Jiang *et al.* (2022) concluded that *Fagus* could have originated in the high latitudes of the Northern Hemisphere (North Pacific region) in late early Eocene. In the Miocene, *Fagus* is found throughout the Northern Hemisphere up to very high latitudes. However, modern lineages could have originated during the global cooling since mid-Miocene at high latitudes where *Fagus* underwent range reduction and extinction in its original areas. Later, the Late Miocene accelerated global cooling and the Pleistocene glaciations could have driven beeches into lower latitudes of East Asia and North America. The present high beech species richness in China can be explained as young relics resulting from several migrations from central Asia and north-eastern Asia (see fig. 10 in Denk and Grimm 2009). In support of these conclusions, there are no fossil records of *Fagus* recovered in subtropical China prior to the mid-Miocene (Table 1 and Fig. 3 in Denk and Grimm 2009), except for late Eocene to early Oligocene fossil (*Fagus* sp. 1 in Southcentral China, Liu *et al.* 1996) that cannot unambiguously be placed within *Fagus*. Therefore, *F. longipetiolata* and *F. lucida* could have undergone divergence allopatrically at higher latitudes due to range fragmentation of their common ancestor during the Late Miocene global cooling and Quaternary glacial periods, and then acquired sympatry in subtropical China through immigrations from central Asia (*F. lucida*) and north-eastern Asia (*F. longipetiolata*) during the late Cenozoic.

The EBSP analyses indicated that *F. lucida* began demographic expansion since 3 Mya whereas *F. longipetiolata* increased its effective population size only between 0.8 and 0.5 Mya (Fig. 2), notably, coinciding with the estimated secondary gene flow. The distinct demographic histories of the two beech species have several implications for their speciation geography. First, it is most likely that *F. lucida* colonized subtropical China and enlarged its population size earlier than *F. longipetiolata*, although the beginning of N_e increase cannot be translated into the arrival in subtropical China directly. It is also plausible that *F. longipetiolata* colonized subtropical China prior to *F. lucida* but expanded its range recently because post-dispersal establishment and expansion are a process with large uncertainties and their success is influenced by many factors (Wu *et al.* 2023). Second, the sudden increase of the effective population size coinciding with the time of gene flow between the two species (0.8 Mya; Fig. 3E, Table 1) suggests that the immigration of *F. longipetiolata* and subsequent range expansion during the mid-Pleistocene resulted in secondary contact and gene flow with *F. lucida* in subtropical China. If we are reminiscent of the historical biogeography of *Fagus*, our inference of secondary sympatry history sounds more reasonable. As discussed before, *F. lucida* could have immigrated from central Asia, however, no beech grows there right now because central Asia is covered by an extensive dryland (Gobi and other deserts, cold steppes in the lowlands). Therefore, the ancestors of *F. lucida* must have been driven to subtropical China before the climate of central Asia became inhospitable for beeches. In fact, the level of aridity in central Asia remained at moderate levels throughout the Miocene and major increases in aridity occurred only after ~3.5 Mya (Guo *et al.* 2002). This major increases in aridity corresponds well to the beginning time of N_e increase of *F. lucida* (3 Mya), indicating that *F. lucida* could have been driven to subtropical China during the Pliocene by the intensified aridity in central Asia. On the contrary, north-eastern Asia has not undergone any long-term aridification, beech forests can still be found as high as 42° N of Japan now (Okaura and Harada 2002). The entering into subtropical China of *F. longipetiolata* might be associated with the mid-Pleistocene transition (MPT, 0.8–1.2 Mya), a period characterized by an increase in the severity of glaciations and the emergence of the ~100-kyr glacial cycles (Liu *et al.* 2003, Clark *et al.* 2006).

Sympatric speciation is unlikely for *F. longipetiolata* and *F. lucida*

Recently, the recognition that speciation can also occur in the absence of geographic barriers (under sympatric conditions) has increased. Despite extensive searches for examples of sympatric speciation in the wild, only a few cases have been convinced in a small space or isolated islands, where ruling out historical allopatric scenarios is relatively easy (Richards *et al.* 2019, Wang *et al.* 2022). For species thriving across large open areas such as *F. longipetiolata* and *F. lucida,* it is extremely difficult to fulfil the requirement of the nonexistence of an allopatric phase because they could have experienced extensive range shifts over the thousands or millions of years required to attain reproductive isolation, especially during the Quaternary glacial-interglacial periods (Fig. 4 and Fig. S6, Supporting Information).

More importantly, sympatric speciation necessitates that populations establish reproductive isolation through divergent natural selection and strong assortative mating to counteract the eroding force of gene flow and recombination (Dieckmann and Doebeli 1999, Coyne and Orr 2004, Schuler *et al.* 2016). However, this study found that most bioclimatic variables (12 out of 19) of *F. longipetiolata* and *F. lucida* overlapped in kernel density plots (Fig. S4, Supporting Information). PCA analysis and ENMTOOLS also detected weak niche differentiation between them (Figs S5 and S7, Supporting Information). Meanwhile, the MAXENT model predicted that the distributions of *F. longipetiolata* and *F. lucida* are widely overlapped across different periods of the late Quaternary (Fig. 4), suggesting they tend to track similar ecological environments, thus could have experienced weak disruptive selection (Cai *et al.* 2021).

Moreover, sympatric speciation is always associated with pollination syndromes in plants (such as flower time, Savolainen *et al.* 2006; pollinator specificity) that is subject to divergent selection also controls non-random mating (Gavrilets 2004, Hernández-Hernández *et al.* 2021). However, beeches are outcrossing and anemophilous plants with little variation in their pollination syndrome (Shen 1992). Our field observations indicate that the flowering time of *F. longipetiolata* and *F. lucida* is largely synchronous. Still, speciation in sympatry or parapatry may be enabled by the incidence of ploidy shifts in plants (Weber and Strauss 2016). However, beech species are all diploids with 24 chromosomes as far as studied, sympatric speciation through polyploidization and subsequent mitotic-genetic incompatibility can be ruled out. Taken together, different lines of evidence suggest that it is unlikely for *F. longipetiolata* and *F. lucida* to initiate sympatric and even parapatric divergence by disruptive selection or through polyploidization. Rather, allopatric speciation caused by long-term geographic separation may be more likely for the two beeches owing to the high phylogenetic niche conservatism in *Fagus* (Wiens 2004).

The implications for the assembly of the relict temperate woody flora of China

Subtropical China is the core area of so called 'Metasequoia flora' (i.e. the Sino-Japanese forest subkingdom), where many relict temperate woody lineages (living fossils) such as *Tetracentron*

Oliv., *Cercidiphyllum* Sieb. & Zucc., *Davidia* Baill., *Trochodendron* Sieb. & Zucc., *Euptelea* Sieb. & Zucc., *Ginkgo* L., *Cathaya* Chun & Kuang, and *Metasequoia* Hu & W. C. Cheng locate (Wu and Wu 1996, Chen *et al.* 2018). The accumulation history of this museum flora has been addressed by ancestral area reconstruction of specific taxa based on dated phylogeny (see a meta-analysis by Chen *et al.* 2018 and references therein), or by examining the dated phylogeny of the entire flora (Lu *et al.* 2018, Ding *et al.* 2020). There is a growing consensus that the flora may be much younger than the stem ages of the living fossils suggest, with most of its clades occurring in subtropical China since the Miocene (13.6 Mya for 'Metasequoia flora', Chen *et al.* 2018; 22.04–25.4 Mya for the flora of eastern China, Lu *et al.* 2018). However, most of these studies have focused on the stem ages or crown ages of a specific lineage (always a genus or a group of species within a genus, Chen *et al.* 2018) rather than the time of population divergence and hybridization within a species or closely related species, which may overestimate the origin time of a specific species or the whole flora. For example, *Pseudotaxus* W. C. Cheng, a relict genus with only one species in subtropical China, diverged from *Taxus* L. 54–65 Mya, however, population genetic investigations suggested that the last common ancestor of its extant populations appeared in subtropical China only after 3.7 Mya (Kou *et al.* 2020). Similar situations have been also observed in other living fossils such as *Ginkgo*, *Taiwania* Hayata, and among others (Chou *et al.* 2011, Hohmann *et al.* 2018). This incongruence calls for more population genetic investigations into relict woody species to complement the evolutionary history of the temperate flora of China.

In this study, we found that *F. longipetiolata* and *F. lucida*, a species pair belonging to one of the most representative woody genera in temperate forests of China, were diverged allopatrically by the Late Miocene and hybridized during the mid-Pleistocene (Fig. 3E). In addition, *F. lucida* began demographic expansion since 3 Mya whereas *F. longipetiolata* increased its effective population size between 0.8 and 0.5 Mya (Fig. 2). These findings, coupled with extensive fossil records of *Fagus*, strengthen the point of view that the relict lineages that originated in high latitudes of the Northern Hemisphere may be very ancient, whereas their extant populations could have become established in subtropical China more recently ('old lineage(s) young populations', Kou *et al.* 2020). In addition, these findings suggest that the geological and climatic events during the Pliocene (5.33–2.58 Mya) such as the abrupt global cooling (Zachos *et al.* 2001), the sharp increase in aridification in central Asia (~3.5 Ma, Guo *et al.* 2002), the major uplift of the eastern Qinghai-Tibetan Plateau (~3.4 Ma, Sun *et al.* 2011, Favre *et al.* 2015; but see Su *et al.* 2019 and Ding *et al.* 2022), and the intensification of Asian monsoon (3.6–2.6 Mya, An *et al.* 2001), as well as the glacial climate changes during the Quaternary, may be the major driving forces for the colonization of relict woody lineages into subtropical China from higher latitudes. This time frame is largely consistent with the idea that 'Metasequoia flora' are relatively young (Chen *et al.* 2018), however, it further implies that 'Metasequoia flora' might be younger than the estimated age based on the dated phylogenies. Interestingly, the evolutionary scenario of 'old lineages young populations' provides a reasonable explanation for the coexistence of genetically discrete sibling that belong to the morphologically and ecologically conservative woody genera (Wen 1999, Milne and Abbott 2002), because these lineages may have developed strong intrinsic reproductive barriers due to long-term isolation in allopatry as suggested by ancient divergence and weak gene flow in this study and other species pairs (e.g. *Pinus massoniana* Lamb. and *P. hwangshanensis* W. Y. Hsia, Zhou *et al.* 2017).

CONCLUSIONS

By investigating the geographic mode of speciation using population genetic analyses, coupled with well-studied beech's fossil records, this study clearly suggests that *F. longipetiolata* and *F. lucida* diverged allopatrically at the high latitudes of the Northern Hemisphere, and then came into contact and hybridized with each other after successive migrations into subtropical China (by Pliocene for *F. lucida* and by mid-Pleistocene for *F. longipetiolata*). Although more studies are needed, this study exemplifies that the study of speciation geography may provide an insightful perspective into the assembly and coexistence of the rich temperate woody plant species of China. However, this study did not detect meaningful substructure within each beech species possibly due to the small amount of SNPs and/or strong pollen-mediated gene flow, preventing from detailing the secondary sympatry history by reconstructing the ancestral area within species (e.g. Schneeweiss *et al.* 2017) and by analysing the spatial and temporal extent of gene flow (e.g. Grummer *et al.* 2015). Such studies using high-resolution genome-scale SNPs (such as whole-genome resequencing data) are critically needed to offer more historical details about the speciation geography of the two beech species.

SUPPLEMENTARY DATA

Supplementary data is available at *Botanical Journal of the Linnean Society* online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (grant numbers 32360064, 31770236 and 30760016) and the Training Program for Academic and Technical Leaders (leading talents) in Major Disciplines of Jiangxi Province (grant to Zhiyong Zhang, 20213BCJ22006). Appreciations go to Dr Guido W. Grimm for his insightful comments and linguistic help.

DATA AVAILABILITY

The data underlying this article are available in the GenBank Nucleotide Database at National Center for Biotechnology Information (nih.gov), and can be accessed with accession numbers from OR106156 to OR112581.

REFERENCES

Abbott RJ, Brennan AC. Altitudinal gradients, plant hybrid zones and evolutionary novelty. *Philosophical Transactions of the Royal Society of*

London, Series B: Biological Sciences 2014;**369**:20130346. https://doi.org/10.1098/rstb.2013.0346

An Z, Kutzbach JE, Prell WL *et al*. Evolution of Asian monsoons and phased uplift of the Himalaya–Tibetan plateau since Late Miocene times. *Nature* 2001;**411**:62–6. https://doi.org/10.1038/35075035

Anacker BL, Strauss SY. The geography and ecology of plant speciation: range overlap and niche divergence in sister species. *Proceedings Biological Sciences* 2014;**281**:20132980. https://doi.org/10.1098/rspb.2013.2980

Bai WN, Wang WT, Zhang DY. Contrasts between the phylogeographic patterns of chloroplast and nuclear DNA highlight a role for pollen-mediated gene flow in preventing population divergence in an East Asian temperate tree. *Molecular Phylogenetics and Evolution* 2014;**81**:37–48. https://doi.org/10.1016/j.ympev.2014.08.024

Barraclough TG, Vogler AP. Detecting the geographical pattern of speciation from species-level phylogenies. *The American Naturalist* 2000;**155**:419–34. https://doi.org/10.1086/303332

Bouckaert R, Heled J, Kühnert D *et al*. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 2014;**10**:e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Brown JL, Bennett JR, French CM. SDMtoolbox 20: the next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ* 2017;**5**:e4095. https://doi.org/10.7717/peerj.4095

Cai Q, Welk E, Ji C *et al*. The relationship between niche breadth and range size of beech (*Fagus*) species worldwide. *Journal of Biogeography* 2021;**48**:1240–53. https://doi.org/10.1111/jbi.14074

Cardoni S, Piredda R, Denk T *et al*. 5S-IGS rDNA in wind-pollinated trees (*Fagus* L.) encapsulates 55 million years of reticulate evolution and hybrid origins of modern species. *The Plant Journal: for Cell and Molecular Biology* 2022;**109**:909–26. https://doi.org/10.1111/tpj.15601

Carstens BC, Richards CL. Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution* 2007;**61**:1439–54. https://doi.org/10.1111/j.1558-5646.2007.00117.x

Chan LM, Brown JL, Yoder AD. Integrating statistical genetic and geospatial methods brings new power to phylogeography. *Molecular Phylogenetics and Evolution* 2011;**59**:523–37. https://doi.org/10.1016/j.ympev.2011.01.020

Chen YS, Deng T, Zhou Z *et al*. Is the East Asian flora ancient or not? *National Science Review* 2018;**5**:920–32. https://doi.org/10.1093/nsr/nwx156

Chou YW, Thomas PI, Ge XJ *et al*. Refugia and phylogeography of *Taiwania* in East Asia. *Journal of Biogeography* 2011;**38**:1992–2005. https://doi.org/10.1111/j.1365-2699.2011.02537.x

Clark PU, Archer D, Pollard D *et al*. The middle Pleistocene transition: characteristics, mechanisms, and implications for long-term changes in atmospheric pCO_2. *Quaternary Science Reviews* 2006;**25**:3150–84. https://doi.org/10.1016/j.quascirev.2006.07.008

Coyne JA, Orr HA. *Speciation*. Sunderland, MA: Sinauer Associates, 2004.

Csilléry K, Blum MGB, Gaggiotti OE *et al*. Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* 2010;**25**:410–8. https://doi.org/10.1016/j.tree.2010.04.001

Darriba D, Taboada GL, Doallo R *et al*. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 2012;**9**:772–772. https://doi.org/10.1038/nmeth.2109

Davis MB, Shaw RG. Range shifts and adaptive responses to Quaternary climate change. *Science* 2001;**292**:673–9. https://doi.org/10.1126/science.292.5517.673

Denk T, Grimm GW. The biogeographic history of beech trees. *Review of Palaeobotany and Palynology* 2009;**158**:83–100. https://doi.org/10.1016/j.revpalbo.2009.08.007

Denk T, Grimm GW, Hemleben V. Patterns of molecular and morphological differentiation in *Fagus*: implications for phylogeny. *American Journal of Botany* 2005;**92**:1006–16. https://doi.org/10.3732/ajb.92.6.1006

Dieckmann U, Doebeli M. On the origin of species by sympatric speciation. *Nature* 1999;**400**:354–7. https://doi.org/10.1038/22521

Ding L, Kapp P, Cai F *et al*. Timing and mechanisms of Tibetan Plateau uplift. *Nature Reviews Earth & Environment* 2022;**3**:652–67. https://doi.org/10.1038/s43017-022-00318-4

Ding WN, Ree RH, Spicer RA *et al*. Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. *Science* 2020;**369**:578–81. https://doi.org/10.1126/science.abb4484

Dong F, Hung CM, Yang XJ. Secondary contact after allopatric divergence explains avian speciation and high species diversity in the Himalayan-Hengduan Mountains. *Molecular Phylogenetics and Evolution* 2020;**143**:106671. https://doi.org/10.1016/j.ympev.2019.106671

Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 1987;**19**:11–5.

Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 2005;**14**:2611–20. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier L, Dupanloup I, Huerta-Sanchez E *et al*. Robust demographic inference from genomic and SNP data. *PLoS Genetics* 2013;**9**:e1003905. https://doi.org/10.1371/journal.pgen.1003905

Excoffier L, Lischer HEL. Arlequin suite ver 35: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 2010;**10**:564–7. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Excofffier L, Marchi N, Marques DA *et al*. fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics* 2021;**37**:4882–5. https://doi.org/10.1093/bioinformatics/btab468

Fang J, Lechowicz MJ. Climatic limits for the present distribution of beech (*Fagus* L) species in the world. *Journal of Biogeography* 2006;**33**:1804–19. https://doi.org/10.1111/j.1365-2699.2006.01533.x

Favre A, Päckert M, Pauls SU *et al*. The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biological Reviews* 2015;**90**:236–53. https://doi.org/10.1111/brv.12107

Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. *Genetics* 2000;**155**:1405–13. https://doi.org/10.1093/genetics/155.3.1405

Fitzpatrick BM, Turelli M. The geography of mammalian speciation: mixed signals from phylogenies and range maps. *Evolution* 2006;**60**:601–15. https://doi.org/10.1111/j.0014-3820.2006.tb01140.x

Foote AD. Sympatric speciation in the genomic era. *Trends in Ecology & Evolution* 2018;**33**:85–95. https://doi.org/10.1016/j.tree.2017.11.003

Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics* 1993;**133**:693–709. https://doi.org/10.1093/genetics/133.3.693

Gavrilets S. *Fitness Landscapes and the Origin of Species*. Princeton: Princeton University Press, 2004.

Grummer JA, Calderón-Espinosa ML, Nieto-Montes de Oca A *et al*. Estimating the temporal and spatial extent of gene flow among sympatric lizard populations (genus *Sceloporus*) in the southern Mexican highlands. *Molecular Ecology* 2015;**24**:1523–42. https://doi.org/10.1111/mec.13122

Guo K, Werger MJA. Effect of prevailing monsoons on the distribution of beeches in continental East Asia. *Forest Ecology and Management* 2010;**259**:2197–203. https://doi.org/10.1016/j.foreco.2009.11.034

Guo ZT, Ruddiman WF, Hao QZ *et al*. Onset of Asian desertification by 22 Myr ago inferred from loess deposits in China. *Nature* 2002;**416**:159–63. https://doi.org/10.1038/416159a

Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. *Nucleic Acids Symposium Series* 1999;**41**:95–8.

Hernández-Hernández T, Miller EC, Román-Palacios C *et al*. Speciation across the tree of life. *Biological Reviews of the Cambridge Philosophical Society* 2021;**96**:1205–42. https://doi.org/10.1111/brv.12698

Hey J. The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Molecular Biology and Evolution* 2010a;**27**:921–33. https://doi.org/10.1093/molbev/msp298

Hey J. Isolation with migration models for more than two populations. *Molecular Biology and Evolution* 2010b;**27**:905–20. https://doi.org/10.1093/molbev/msp296

Hodge JR, Bellwood DR. The geography of speciation in coral reef fishes: the relative importance of biogeographical barriers in separating sister-species. *Journal of Biogeography* 2016;**43**:1324–35. https://doi.org/10.1111/jbi.12729

Hohmann N, Wolf EM, Rigault P *et al.* *Ginkgo biloba*'s footprint of dynamic Pleistocene history dates back only 390,000 years ago. *BMC Genomics* 2018;**19**:299. https://doi.org/10.1186/s12864-018-4673-2

Hubisz MJ, Falush D, Stephens M *et al.* Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 2009;**9**:1322–32. https://doi.org/10.1111/j.1755-0998.2009.02591.x

Hudson RR, Kreitman M, Aguadé M. A test of neutral molecular evolution based on nucleotide data. *Genetics* 1987;**116**:153–9. https://doi.org/10.1093/genetics/116.1.153

Jiang L, Bao Q, He W *et al.* Phylogeny and biogeography of *Fagus* (Fagaceae) based on 28 nuclear single/low-copy loci. *Journal of Systematics and Evolution* 2022;**60**:759–72. https://doi.org/10.1111/jse.12695

Kou Y, Zhang L, Fan D *et al.* Evolutionary history of a relict conifer, *Pseudotaxus chienii* (Taxaceae), in south-east China during the late Neogene: old lineage, young populations. *Annals of Botany* 2020;**125**:105–17. https://doi.org/10.1093/aob/mcz153

Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 70 for bigger datasets. *Molecular Biology and Evolution* 2016;**33**:1870–4. https://doi.org/10.1093/molbev/msw054

Leigh JW, Bryant DP. full-feature software for haplotype network construction. *Methods in Ecology and Evolution* 2015;**6**:1110–6. https://doi.org/10.1111/2041-210X.12410

Li DQ, Jiang L, Liang H *et al.* Resolving a nearly 90-year-old enigma: the rare *Fagus chienii* is conspecific with *F. hayatae* based on molecular and morphological evidence. *Plant Diversity* 2023;**45**:544–51. https://doi.org/10.1016/j.pld.2023.01.003

Li H. A new test for detecting recent positive selection that is free from the confounding impacts of demography. *Molecular Biology and Evolution* 2011;**28**:365–75. https://doi.org/10.1093/molbev/msq211

Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009;**25**:1451–2. https://doi.org/10.1093/bioinformatics/btp187

Liu H, Xing Q, Ji Z *et al.* An outline of Quaternary development of *Fagus* forest in China: palynological and ecological perspectives. *Flora - Morphology, Distribution. Flora - Morphology, Distribution, Functional Ecology of Plants* 2003;**198**:249–59. https://doi.org/10.1078/0367-2530-00098

Liu YS, Momohara A, Mei SW. A revision on the Chinese megafossils of *Fagus* (Fagaceae). *Journal of Japanese Botany* 1996;**71**:168–77.

López-Pujol J, Zhang FM, Sun HQ *et al.* Centres of plant endemism in China: places for survival or for speciation? *Journal of Biogeography* 2011;**38**:1267–80. https://doi.org/10.1111/j.1365-2699.2011.02504.x

Losos JB, Glor RE. Phylogenetic comparative methods and the geography of speciation. *Trends in Ecology & Evolution* 2003;**18**:220–7. https://doi.org/10.1016/s0169-5347(03)00037-5

Lu LM, Mao LF, Yang T *et al.* Evolutionary history of the angiosperm flora of China. *Nature* 2018;**554**:234–8. https://doi.org/10.1038/nature25485

Manchester SR, Tiffney BH. Integration of paleobotanical and neobotanical data in the assessment of phylogeographic history of Holarctic angiosperm clades. *International Journal of Plant Sciences* 2001;**162**:S19–27. https://doi.org/10.1086/323657

Mayr E. *Systematics and the Origin of Species*. New York: Columbia University Press, 1942.

Merzeau D, Comps B, Thiébaut B *et al.* Genetic structure of natural stands of *Fagus sylvatica* L (beech). *Heredity* 1994;**72**:269–77. https://doi.org/10.1038/hdy.1994.37

Meyer HW, Manchester SR. The Oligocene Bridge Creek Flora of the John Day Formation, Oregon. Berkeley: University of California Publications in Geological Sciences, 1997.

Milne RI, Abbott RJ. The origin and evolution of tertiary relict floras. *Advances in Botanical Research* 2002;**38**:282–314. https://doi.org/10.1016/S0065-2296(02)38033-9

Mittelbach GG, Schemske DW. Ecological and evolutionary perspectives on community assembly. *Trends in Ecology and Evolution* 2015;**30**:241–7. https://doi.org/10.1016/j.tree.2015.02.008

Mittermeier RA, Robles-Gil P, Mittermeier GC. *Megadiversity. Earth's Biologically Wealthiest Nations*. Mexico City: CEMEX/Agrupaciaon Sierra Madre, 1997.

Muto N, Kai Y. Allopatric origin, secondary contact and subsequent isolation of sympatric rockfishes (Sebastidae: *Sebastes*) in the north-western Pacific. *Biological Journal of the Linnean Society* 2023;**138**:37–50. https://doi.org/10.1093/biolinnean/blac135

Nielsen R, Wakeley J. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 2001;**158**:885–96. https://doi.org/10.1093/genetics/158.2.885

Nosil P. *Ecological Speciation*. Oxford: Oxford University Press, 2012.

Okaura T, Harada K. Phylogeographical structure revealed by chloroplast DNA variation in Japanese Beech (*Fagus crenata* Blume). *Heredity* 2002;**88**:322–9. https://doi.org/10.1038/sj.hdy.6800048

Osborne OG, Kafle T, Brewer T *et al.* Sympatric speciation in mountain roses (*Metrosideros*) on an oceanic island. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 2020;**375**:20190542. https://doi.org/10.1098/rstb.2019.0542

Papadopulos AS, Baker WJ, Crayn D *et al.* Speciation with gene flow on Lord Howe Island. *Proceedings of the National Academy of Sciences of the United States of America* 2011;**108**:13188–93. https://doi.org/10.1073/pnas.1106085108

Pettengill JB, Moeller DA. Phylogeography of speciation: allopatric divergence and secondary contact between outcrossing and selfing Clarkia. *Molecular Ecology* 2012;**21**:4578–92. https://doi.org/10.1111/j.1365-294X.2012.05715.x

Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 2006;**190**:231–59. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;**155**:945–59. https://doi.org/10.1093/genetics/155.2.945

Qin HT, Möller M, Milne R *et al.* Multiple paternally inherited chloroplast capture events associated with *Taxus* speciation in the Hengduan Mountains. *Molecular Phylogenetics and Evolution* 2023;**189**:107915. https://doi.org/10.1016/j.ympev.2023.107915

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 2021. https://www.R-project.org/ (30 July 2023, date last accessed).

Rambaut A, Drummond AJ, Xie D *et al.* Posterior summarization in Bayesian phylogenetics using Tracer 17. *Systematic Biology* 2018;**67**:901–4. https://doi.org/10.1093/sysbio/syy032

Renner SS, Grimm GW, Kapli P *et al.* Species relationships and divergence times in beeches: new insights from the inclusion of 53 young and old fossils in a birth–death clock model. *Philosophical Transactions of the Royal Society London Series B. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 2016;**371**:20150135. https://doi.org/10.1098/rstb.2015.0135

Richards EJ, Servedio MR, Martin CH. Searching for sympatric speciation in the genomic era. *BioEssays* 2019;**41**:e1900047. https://doi.org/10.1002/bies.201900047

Ricklefs RE. Host-pathogen coevolution, secondary sympatry and species diversification. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 2010;**365**:1139–47. https://doi.org/10.1098/rstb.2009.0279

Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology* 2004;**4**:137–8. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Savolainen V, Anstett MC, Lexer C *et al.* Sympatric speciation in palms on an oceanic island. *Nature* 2006;**441**:210–3. https://doi.org/10.1038/nature04566

Schliewen UK, Tautz D, Pääbo S. Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* 1994;**368**:629–32. https://doi.org/10.1038/368629a0

Schneeweiss GM, Winkler M, Schönswetter P. Secondary contact after divergence in allopatry explain current lack of ecogeographic isolation in two hybridizing alpine plant species. *Journal of Biogeography* 2017;**44**:2575–84. https://doi.org/10.1111/jbi.13071

Schoener TW. The Anolis lizards of Bimini: resource partitioning in a complex fauna. *Ecology* 1968;**49**:704–26. https://doi.org/10.2307/1935534

Schuler H, Hood GR, Egan SP *et al*. Modes and mechanisms of speciation. *Reviews in Cell Biology and Molecular Medicine* 2016;**2**:60–93. https://doi.org/10.1002/3527600906.mcb.201600015

Shen CF. A monograph of the genus *Fagus* Tourn. ex L. (Fagaceae). Ph.D. Thesis, The City University of New York, New York, 1992.

Skeels A, Cardillo M. Reconstructing the geography of speciation from contemporary biodiversity data. *The American Naturalist* 2019;**193**:240–55. https://doi.org/10.1086/701125

Sousa V, Hey J. Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics* 2013;**14**:404–14. https://doi.org/10.1038/nrg3446

Strasburg JL, Rieseberg LH. How robust are 'isolation with migration' analyses to violations of the IM model? A simulation study. *Molecular Biology and Evolution* 2010;**27**:297–310. https://doi.org/10.1093/molbev/msp233

Su T, Farnsworth A, Spicer RA *et al*. No high Tibetan Plateau until the Neogene. *Science Advances* 2019;**5**:eaav2189. https://doi.org/10.1126/sciadv.aav2189

Sun BN, Wu JY, Liu YS *et al*. Reconstructing Neogene vegetation and climates to infer tectonic uplift in western Yunnan, China. *Palaeogeography, Palaeoclimatology, Palaeoecology* 2011;**304**:328–36. https://doi.org/10.1016/j.palaeo.2010.09.023

Tajima F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 1983;**105**:437–60. https://doi.org/10.1093/genetics/105.2.437

Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989;**123**:585–95. https://doi.org/10.1093/genetics/123.3.585

Tobias JA, Cornwallis CK, Derryberry EP *et al*. Species coexistence and the dynamics of phenotypic evolution in adaptive radiation. *Nature* 2014;**506**:359–63. https://doi.org/10.1038/nature12874

Wang Y, Qiao Z, Mao L *et al*. Sympatric speciation of the spiny mouse from Evolution Canyon in Israel substantiated genomically and methylomically. *Proceedings of the National Academy of Sciences of the United States of America* 2022;**119**:e2121822119. https://doi.org/10.1073/pnas.2121822119

Warren DL, Cardillo M, Rosauer DF *et al*. Mistaking geography for biology: inferring processes from species distributions. *Trends in Ecology & Evolution* 2014;**29**:572–80. https://doi.org/10.1016/j.tree.2014.08.003

Warren DL, Glor RE, Turelli M. ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* 2010;**33**:607–11. https://doi.org/10.1111/j.1600-0587.2009.06142.x

Watterson GA. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 1975;**7**:256–76. https://doi.org/10.1016/0040-5809(75)90020-9

Weber MG, Strauss SY. Coexistence in close relatives: beyond competition and reproductive isolation is sister taxa. *Annual Review of Ecology, Evolution, and Systematics* 2016;**47**:359–81. https://doi.org/10.1146/annurev-ecolsys-112414-054048

Weir JT, Price TD. Limits to speciation inferred from times to secondary sympatry and ages of hybridizing species along a latitudinal gradient. *The American Naturalist* 2011;**177**:462–9. https://doi.org/10.1086/658910

Wen J. Evolution of eastern Asian and eastern North American disjunct distributions in flowering plants. *Annual Review of Ecology and Systematics* 1999;**30**:421–55. https://doi.org/10.1146/annurev.ecolsys.30.1.421

Wiens JJ. Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. *Evolution* 2004;**58**:193–7. https://doi.org/10.1111/j.0014-3820.2004.tb01586.x

Wolfe JA. Tertiary climates and floristic relationships at high latitudes in the Northern Hemisphere. *Palaeogeography, Palaeoclimatology, Palaeoecology* 1980;**30**:313–23. https://doi.org/10.1016/0031-0182(80)90063-2

Wu ZY, Milne RI, Liu J *et al*. The establishment of plants following long-distance dispersal. *Trends in Ecology & Evolution* 2023;**38**:289–300. https://doi.org/10.1016/j.tree.2022.11.003

Wu ZY, Wu SG. A proposal for a new Floristic Kingdom (realm): the E. Asiatic Kingdom, its delineation and characteristics. In: Zhang AL, Wu SG (eds), *Proceedings of the First International Symposium on Floristic Characteristics and Diversity of East Asian Plants*. Beijing: Chinese Higher Education Press, 1996.

Zachos J, Pagani M, Sloan L *et al*. Trends, rhythms, and aberrations in global climate 65 Mya to present. *Science* 2001;**292**:686–93. https://doi.org/10.1126/science.1059412

Zhang ZY, Wu R, Wang Q *et al*. Comparative phylogeography of two sympatric beeches in subtropical China: species-specific geographic mosaic of lineages. *Ecology and Evolution* 2013;**3**:4461–72. https://doi.org/10.1002/ece3.829

Zhou Y, Duvaux L, Ren G *et al*. Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions. *Heredity* 2017;**118**:211–20. https://doi.org/10.1038/hdy.2016.72

Zhou BF, Yuan S, Crowl AA *et al*. Phylogenomic analyses highlight innovation and introgression in the continental radiations of Fagaceae across the Northern Hemisphere. *Nature Communications* 2022;**13**:1320. https://doi: 10.1038/s41467-022-28917-1s