


Tracing the Holocene hybrid origin of cultivated walnut in southwestern China

Jie Liu^{1*}, Hai-Ling Qi^{1,2}, Peng-Zhen Fan^{1,3}, Chu-Mei Liu^{1,3}, Zeng-Yuan Wu¹, Ya-Huang Luo⁴, Richard I. Milne⁵, Raees Khan¹, Lian-Ming Gao⁴, Yue-Hua Wang², Moses C. Wambulwa⁶, Robabeh Shahi Shavvon^{7*}  and De-Zhu Li^{1,4*}

¹ Germplasm Bank of Wild Species & Yunnan Key Laboratory of Crop Wild Relatives Omics, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

² School of Ecology and Environmental Sciences, Yunnan University, Kunming 650091, China

³ University of the Chinese Academy of Sciences, Beijing 100049, China

⁴ Center for Interdisciplinary Biodiversity Research & College of Forestry, Shandong Agricultural University, Tai'an 271018, China

⁵ Institute of Molecular Plant Sciences, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JH, UK

⁶ Department of Life Sciences, School of Science and Computing, South Eastern Kenya University, Kitui 170-90200, Kenya

⁷ Department of Biology, Faculty of Science, Yasouj University, Yasouj 74934-75918, Iran

* Correspondence: liujie@mail.kib.ac.cn (Liu J); r.shahi@yu.ac.ir (Shahi Shavvon R); dzl@mail.kib.ac.cn (Li DZ)

Abstract

Genetic exchanges underpin population adaptation and species evolution. However, knowledge of the genetic impacts of historical, anthropogenic crop-to-wild gene flow remains limited. In southwestern China, the long cultivation and parapatric distribution of *Juglans regia* and *J. sigillata* provide an ideal system to investigate this process, yet the evolutionary origin of the cultivated walnut in this region remains a mystery. Using 31 microsatellite loci to genotype 2,866 individuals, along with 716 chloroplast genomes, we evaluated genetic diversity, population structure, and maternal lineage patterns across the region. Population demographic histories were subsequently inferred using approximate Bayesian computation. Our analysis revealed that *J. sigillata* has higher genetic diversity than *J. regia*, attributed to its long-term persistence in heterogeneous environments. We observed extensive hybridization between the two species around the Sichuan Basin, forming two geographically distinct subgroups: Hybrid1 in the northwest and Hybrid2 in the south. Both subgroups exhibit variable parental contributions, but with the chloroplast genome entirely introgressed from *J. regia*. Furthermore, the Sichuan Basin and Yangtze River acted as major geographic barriers to north-south gene flow. Our divergence time estimates showed that the two species diverged during the Pleistocene, followed by an admixture event between *J. regia* and *J. sigillata*, forming Hybrid2 during the early Holocene. Subsequent admixture between Hybrid2 and *J. regia* produced Hybrid1 during the middle Holocene. We propose that cultivated walnuts in southwestern China originated through introgression between *J. regia* and wild *J. sigillata* during the Holocene, prior to the introduction of agriculture to the region, a process likely facilitated by historical southward human migrations. This research elucidates how long-term human influence has subtly reshaped the genetic landscape of walnuts in southwestern China through gene flow, providing valuable insights for the study and management of other tree crops.

Keywords: Cultivated walnut, Chloroplast introgression, Hybridization, *Juglans regia*, *Juglans sigillata*, Southwestern China

Citation: Liu J, Qi HL, Fan PZ, Liu CM, Wu ZY, et al. 2026. Tracing the Holocene hybrid origin of cultivated walnut in southwestern China. *Forestry Research* 6: e018 <https://doi.org/10.48130/forres-0026-0018>

Introduction

Analysis of the genetic structure, variation, and relatedness within populations can reveal the dynamic history of species and evolutionary relationships of populations^[1,2]. Insights from such analyses can also help address questions about species origins^[3,4] or centers of diversification^[5,6]. Interspecific hybridization is one of the most significant factors influencing the genetic variation of plant populations, closely connected with the gene flow patterns. In plant populations, gene flow is determined by mating systems, pollinators, population structure, and geographical barriers^[7,8], and it can occur between wild and cultivated plants^[9]. Many domesticated crops, including bananas^[10] and certain citrus fruit^[11], are believed to have evolved from hybridization between two wild relatives, often accompanied by allopolyploidization^[11–13]. Previous studies demonstrate the widespread occurrence of gene flow from crops to wild populations in many different species, such as date palms^[14], coffee^[15], almond^[16], apples^[17–20], and walnuts^[21–24]. Nevertheless, the contribution of gene transfer to the origin of cultivated walnut remains unclear.

Common walnut (*J. regia*) and its sister species, the iron walnut (*J. sigillata*)^[25,26] are deciduous trees, both diploid with $2n = 32$ ^[27,28].

The two species are ideal for examining genetic exchange and human effects due to their overlapping natural and cultivation ranges in southwestern China. Although *J. regia* is cultivated worldwide, wild *J. sigillata* is narrowly distributed in southwestern China and the Eastern Himalaya^[22,29–32]. However, it is extensively cultivated within this range, particularly in the Chinese provinces of Yunnan, Sichuan, and Guizhou^[24]. Hybridization has been observed in much of the *J. sigillata* distribution range (e.g., southwestern China and the Himalaya) where the two species coexist^[22,24,32,33]. Moreover, it has been demonstrated that human activities play a significant role in shaping the genetic diversity patterns of walnuts in central Asia and the Himalaya^[22,34]. Therefore, the origins and history of cultivated walnut in southwestern China might have involved repeated selection and adaptation, and/or interspecific gene flow, and hence represents an instructive case of how cultivated tree genotypes arise.

The arrangement of mountains and rivers in southwestern China likely limited agricultural expansion, leading to diverse planting practices and variation between locations during the Late Bronze Age^[35,36]. The earliest evidence of agriculture in southwestern China is dated around 5,000 years before present (BP), particularly in

northwestern Sichuan^[37,38]. The formation of alluvial terraces and floodplains around 5,500 years BP, possibly triggered by changes to the East Asian Summer Monsoon (EASM), might have contributed to subsequent human settlements in the Sichuan Basin^[39]. Consistent with this, Gunn et al.^[33] posited that *J. regia* and *J. sigillata* remained allopatric until humans introduced *J. regia* from Xinjiang to the Hengduan Mountains. They suggested that human-assisted dispersal could have occurred along key trading routes, such as the Tea-Horse Road along the steep Mekong River canyons in northwestern Yunnan. Long-distance contact networks, oriented along north-south mountain ranges, facilitated not only human migration, but also the exchange of agricultural practices and crops^[40,41]. The expansion of these networks during the Bronze Age^[42] provides a plausible, though not exclusive, framework for understanding how walnuts might have spread alongside human movements. Hence, from this point onward, human-mediated germplasm exchange between introduced species and their wild relatives became increasingly feasible, a mechanism well documented in plants^[43,44]. Therefore, understanding the genetic history of walnuts in this region can provide insights into broader patterns of agricultural development.

Analyzing molecular characteristics is a key method for deciphering genetic diversity and relationships among walnut species and their cultivars. One of the markers commonly used to assess genetic diversity is Simple Sequence Repeats (SSRs). This marker has proven extremely valuable for elucidating genetic diversity and the relationships between commercially important species such as *J. regia* and *J. sigillata*^[23,32,45]. More recently, genome sequencing has been employed to elucidate the genetic architecture and evolutionary history of *Juglans* species. Studies based on nuclear and chloroplast genomes have provided insights into the population structure and demographic history of *J. regia* and *J. sigillata*^[22,23,46]. Nevertheless, these studies remain insufficient to fully clarify the origin and cultivation history of walnut in southwestern China. Furthermore, the anthropogenic dimension of this history remains completely unexamined. In this study, by integrating biparentally inherited microsatellite markers and maternally inherited chloroplast genomes,

we utilized a comprehensive genetic dataset in southwestern China to explore: (1) the genetic diversity and population structure of *J. sigillata* and *J. regia* populations, and their genetic relationships; and (2) the mechanism underlying the genetic origin of cultivated walnut. By elucidating genetic patterns within walnut populations in southwestern China, our study provides valuable insights into their origins and guides future conservation and management strategies.

Materials and methods

Field sampling

Based on the walnut distribution data from digital specimens and literature, we conducted field sampling of both *J. regia* and *J. sigillata* in northern and southwestern China between 2013 and 2020 (Fig. 1, Supplementary Table S1). The initial species identification of sampled trees was based on key diagnostic morphological characters. This distinction was primarily based on leaflet number and nut morphology^[33]. *J. regia* possesses 5–7 leaflets, while *J. sigillata* exhibits 9–17. Furthermore, *J. sigillata* is characterized by pitted nut surfaces and dark kernels with tough septa, contrasting with the wrinkled nut surfaces and light-colored kernels typical of *J. regia*. We collected 5–30 individuals per population, with a minimum distance of > 100 km between populations and > 100 m between individuals (Supplementary Table S1). Additionally, we tried to sample ancient trees with a diameter at breast height (DBH) exceeding 100 cm and an estimated age of over 100 years, as inferred from local knowledge gathered during field surveys. Young leaf materials were collected and immediately dried using silica gel. In total, 2,866 samples from 123 populations (42 *J. regia* and 81 *J. sigillata*) were collected across 13 provinces in China. These included data previously published^[22,32,47] (detailed collection information is provided in Supplementary Table S1). In addition, a total of 716 chloroplast genomes were used in our study (Supplementary Table S2), including 204 newly generated in this work, and 512 assembled from previous studies^[23,48].

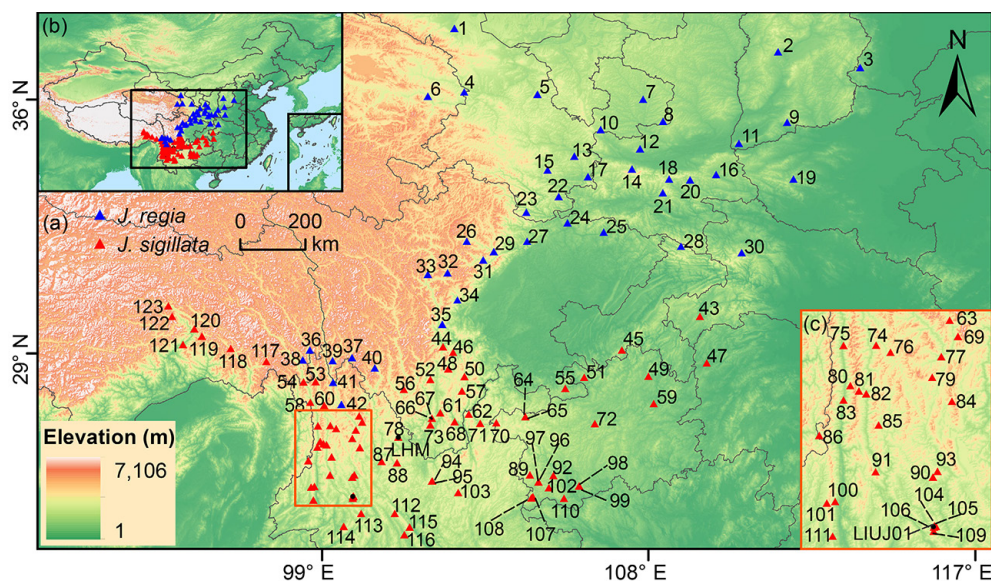


Fig. 1 Geographical distribution of 123 populations studied in this study. (a) Sampling sites of *Juglans regia* and *J. sigillata* populations in northern and southwestern China. (b) Geographical location of study region in Asia. (c) Sampling sites of *J. sigillata* in the Three Parallel Rivers region. The numbers in the map represent populations IDs (Supplementary Table S1). Blue triangles represent *J. regia* populations; red triangles indicate *J. sigillata* populations.

DNA extraction, PCR amplification, microsatellite genotyping, and sequencing

Approximately 0.20 g of silica gel-dried leaf material was used for DNA extraction following a modified CTAB method^[49,50]. Subsequently, DNA quality and concentration were assessed using 1% agarose gel electrophoresis and a Nano Drop® ND-2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). DNA samples meeting PCR quality standards were diluted to a concentration of 30–50 ng/μL and stored at –20 °C for further use.

Based on our previous research^[32,51,52], a total of 31 microsatellite loci (Supplementary Table S3) were used to genotype the samples. PCR amplification was performed using a multiplex PCR system on a Veriti® 96-well Thermo-Cycler (Applied Biosystems, Foster City, California, USA), following the protocol developed by Xiahou et al.^[52]. PCR products were verified by 1% agarose gel electrophoresis, and fragment sizes for each individual were determined using an ABI3730xl DNA analyzer (Applied Biosystems, CA, USA). Genotyping data were processed using GeneMarker v2.2.0 (SoftGenetics, State College, PA, USA) and formatted for downstream analyses using GenAlEx v6.51b2^[53].

For newly generated chloroplast genomes, genomic DNA was fragmented into ~500 bp inserts for library preparation using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Massachusetts, USA), following the manufacturer's instructions. Sequencing was performed on the DNBSEQ-T7 platform (BGI, Shenzhen, China), producing 150 bp paired-end reads. For each individual, approximately 20× coverage was obtained, yielding over 12 Gb of high-quality, clean data.

Microsatellite data analysis

Clonal individuals and Hardy–Weinberg equilibrium

Genetic homogeneity among clones can bias diversity estimates by artificially lowering metrics such as expected heterozygosity while simultaneously inflating others, like observed heterozygosity^[54–56], thereby confounding inferences of population structure^[57]. Consequently, we first used GenAlEx 6.51b2 to define clonal individuals as those with identical genotypes across all 31 loci examined. Additionally, clustering analysis (see the 'Population inference structure' section below) was used to check for unusual groupings among the 2,866 individuals. In the case of anomalies, the individuals were visually checked and re-genotyped whenever necessary. Furthermore, Arlequin v3.5.2^[58] was used to estimate whether populations were in Hardy–Weinberg equilibrium.

Genetic diversity estimation

Genetic diversity parameters for loci, populations, and different groups were calculated using GenAlEx. These parameters comprised the number of alleles (N_A), the effective number of alleles (N_E), observed heterozygosity (H_O), expected heterozygosity (H_E), unbiased expected heterozygosity (uH_E), and the fixation index (F). Additionally, polymorphism information content (PIC) was calculated to

characterize the genetic polymorphism of loci using the PIC_CALC software (https://github.com/luansheng/PIC_CALC). For genetic diversity at population and group levels, the total number of alleles (N_T) and the number of private alleles (N_P) were also calculated using GenAlEx, while the allelic richness (A_R) was calculated using the *hierfstat* package in R (<https://cran.r-project.org/web/packages/hierfstat>). Inbreeding coefficients (F_{IS}) for populations and groups were calculated using FSTAT v2.9.4 (www2.unil.ch/popgen/softwares/fstat.htm) (Table 1).

Population structure inference

Genetic differentiation coefficients (F_{ST}) between populations and groups were calculated using Arlequin. Standard genetic distances D_A between populations and groups were calculated using MSA v4.05^[59]. Analysis of molecular variance (AMOVA) was conducted using GenAlEx for all populations and groups.

Bayesian clustering analysis was performed using STRUCTURE v2.3.4^[60]. The number of groups (K) was set from 1 to 10, with 100,000 burn-in iterations and 1,200,000 MCMC replications after burn-in, repeated 20 times for each K . The optimal K value was determined using STRUCTURE HARVESTER v0.6.1^[61], and repeated sampling analysis was performed using CLUMPP v1.1.2^[62]. The results were visualized using DISTRUCT v1.1^[63].

Based on the admixture coefficient (Q) calculated from STRUCTURE analysis at $K = 2$, individuals were classified into pure groups if they had a Q value ≥ 0.8 , while individuals with a Q value $0.2 \leq Q < 0.8$ were assigned as hybrids^[22]. This genetics-based classification scheme was employed to provide an objective baseline assessment of admixture, independent of morphological traits that may be influenced by environmental factors and phenotypic plasticity. It therefore serves as an unbiased reference framework for comparison with the phenotypic identification. Ultimately, this classification resulted in three groups: pure *J. sigillata*, pure *J. regia*, and hybrids. To infer relatedness among all individuals, principal coordinates analysis (PCoA) based on genetic distance D_A was performed using GenAlEx, and visualized with R package *ggplot2* in R^[64]. A Neighbor-joining (NJ) analysis was conducted using Populations v1.2.31^[65], and the resulting clustering tree was graphically represented using the R package *ggtree*^[66]. A neighbor-net diagram was constructed using SplitsTree v4.18.3^[67] to visualize the genetic relationships among all individuals.

Spatial genetic patterns, Mantel test, and barrier analyses

To illustrate the spatial distribution of genetic structure among populations, STRUCTURE results at $K = 2$ were mapped using ArcMap v10.7 (ESRI, Redlands, CA, USA). Additionally, interpolation analysis was performed using inverse distance weighting (IDW) in ArcMap v10.7 to visualize the geographic distribution of genetic diversity parameters, such as allelic richness (A_R) and expected heterozygosity (H_E).

We performed Mantel tests using the mantel function in the R package *vegan*^[68] to assess the patterns of Isolation by Distance

Table 1. Genetic diversity within three groups based on microsatellite data. JR: *Juglans regia*; JS: *J. sigillata*.

Group	N	N_T	N_P	N_A	N_E	H_O	H_E	uH_E	A_R	F_{IS}
JR	871	164	3	5.29	1.94	0.37	0.42	0.42	6.00	0.12
JS	821	237	36	7.65	2.62	0.51	0.58	0.58	8.16	0.13
Hybrid	1,011	207	5	6.68	2.67	0.51	0.58	0.58	7.37	0.11
Total	2,703	251	–	8.10	2.88	0.46	0.61	0.61	8.05	0.24

N , sample size; N_T , total number of alleles; N_P , private alleles; N_A , number of alleles; N_E , effective number of alleles; H_O , observed heterozygosity; H_E , expected heterozygosity; uH_E , unbiased expected heterozygosity; A_R , allelic richness; F_{IS} , inbreeding coefficient.

(IBD). To distinguish broad-scale, landscape-level patterns from those specific to major genetic groups, we employed a two-level approach. First, we conducted a global analysis of all 123 populations, correlating pairwise genetic differentiation (F_{ST}) with geographic distance (km) and testing the relationships between genetic diversity (H_E) and altitude, longitude, and latitude using the Pearson correlation. Subsequently, group-level analyses were performed separately for *J. regia* (JR), *J. sigillata* (JS), and hybrid populations. This stratified approach is essential because a significant global signal can be driven by a single group; separate tests allow us to identify whether IBD or environmental correlations operate consistently across the genetic landscape or are group-specific.

To analyze isolation barriers, we employed Monmonier's maximum-difference algorithm in Barrier v2.2^[69], a method effective for identifying sharp genetic discontinuities that may correspond to barriers to gene flow. For this analysis, we used the Pairwise genetic distance matrices (D_A) generated from 100 bootstrap replicates in Microsatellite Analyzer (MSA), which were then integrated with the geographical coordinates (Supplementary Table S1). The algorithm initially connects the geographic coordinates through Delaunay triangulation, followed by generation of the corresponding Voronoi tessellations. Barriers were outlined based on these tessellations, and the robustness of the identified boundaries was assessed using Monmonier's maximum difference algorithm. While absolute genetic distance values depend on the specific metric used, the relative spatial positions of identified barriers and the overall pattern are robust to the choice of distance measure. Potential geographical barriers from the Barrier analysis were subsequently displayed in space using ArcMap v10.7.

Inference of demographic history

We employed the approximate Bayesian computation (ABC) approach using supervised machine learning in DIYABC-RF v1.2.1^[70] to explore the demographic history of the genetic groups identified by STRUCTURE at $K = 2$. For this analysis, the hybrid group was further divided into Hybrid1 and Hybrid2, which were treated separately. We developed five plausible demographic scenarios, as illustrated in Fig. 2. All parameters, except admixture and divergence time, were set to default values (Supplementary Table S4), with the generation time assumed to be 15 years^[48]. For model calibration, we used a mean microsatellite mutation rate of 5×10^{-4} mutations per locus per generation (a standard value widely used in population genetic studies of eukaryotes^[71–73]). This allowed us to convert the scaled time parameters estimated by DIYABC-RF into absolute times in generations, which were then converted into years using the generation time. While the absolute timing of events relies on the chosen mutation rate and generation time, the sequence of those events, for instance, whether divergence came before admixture, is derived from the model's fit to the genetic data. This relative ordering remains stable under reasonable adjustments to these calibration values. We generated approximately 2,000 simulations per scenario for scenario choice, and 10,000 simulations for parameter estimation. The choice of scenario and estimation of the posterior probability of the best-supported scenario was performed using the Random Forest algorithm module within DIYABC-RF, which produced 2,000 Random Forest trees for each analysis^[70]. Additionally, we performed Bayesian analysis of historical migration rates with Migrate-n 5.0.4^[74]. One long chain was run, retaining 50,000 sampled generations at intervals of 100 generations after a burn-in period of 1,000,000 generations.

Chloroplast genome data analysis

A total of 716 individuals of *J. regia* and *J. sigillata* were newly assembled and analyzed for their chloroplast genomes, including 204 that were sequenced in this study. The remaining 512 individuals, comprising 460 from Ji et al.^[48] and 52 from Ding et al.^[23], with project numbers PRJNA721107 and PRJNA356989, respectively, were downloaded from the National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov) using the SRA Toolkit v2.10.8^[75]. The chloroplast genomes were assembled and annotated using GetOrganelle v1.7.7.0^[76], with the *J. regia* chloroplast genome^[22] serving as the reference sequence. The chloroplast genomes were initially aligned using Geneious v9.0.2^[77]. The resulting alignment files were imported into R, where haplotypes were identified using the *haplotypes* package v1.1.2^[78]. The identified haplotype sequences were then aligned using Geneious v9.0.2. A haplotype network was constructed using PopART v1.7^[79], and a geographical distribution map of the chloroplast genome haplotypes was created using QGIS 3.44.0^[80].

Phylogenetic relationships among haplotypes were inferred using Maximum Likelihood (ML) in IQ-TREE (v2.3.5)^[81] and Bayesian Inference (BI) in MrBayes v3.2.7a^[82]. The following seven chloroplast genomes were downloaded from NCBI and used as outgroups owing to their close phylogenetic relationship with the *Juglans*: *Pterocarya stenoptera* (MN262640), *Cyclocarya paliurus* (MW118603), *Platycarya strobilacea* (MH189595), *Annamocarya sinensis* (MN911165), *Carya cathayensis* (MW410227), *Engelhardia fenzelii* (MT991009), and *Rhoiptelea chiliantha* (NC_053773). The resulting trees were visualized using iTOL (<https://itol.embl.de>). To infer relatedness among haplotypes, principal components analysis (PCA) was carried out using the R package *adeigenet* v2.1.8^[83], and visualized with the R package *ggplot2*.

Maternal lineage bias assessment

Given that the separation of pure *J. regia* and *J. sigillata* individuals was consistently supported by both nuclear and chloroplast data, the maternal parent of each hybrid individual was inferred through phylogenetic analysis of chloroplast genome haplotypes. Consequently, hybrid individuals were unambiguously classified as originating from either *J. regia* or *J. sigillata*. Among the 1,011 individuals identified as hybrids based on microsatellite data, we counted the number possessing the *J. regia* and *J. sigillata* chloroplast genome haplotype. The proportion of hybrids with maternal *J. regia* lineage was calculated along with its 95% confidence interval (using the Wilson score method). To statistically evaluate whether the maternal contribution was symmetric, we tested the observed proportion against a null expectation of 0.5 using a two-sided exact binomial test.

Results

Microsatellite data

Clonal individuals and Hardy–Weinberg equilibrium

Out of the 2,866 individuals screened, 163 were identified as clones. After removing these clonal individuals, the final microsatellite dataset comprised 2,703 individuals from 123 populations. Genetic diversity results for 31 microsatellite loci are provided in Supplementary Table S5. The number of alleles (N_A) across marker loci ranged from 4 (in JS28 and JR07) to 13 (in JS22). The mean effective number of alleles (N_E), expected heterozygosity (H_E), and observed heterozygosity (H_O) were 2.88, 0.61, and 0.46, respectively.

Polymorphic information content (PIC) ranged from 0.17 (in JM5446) to 0.76 (in BFU-Jr38), with a mean of 0.55. (Supplementary Table S5). The Hardy–Weinberg equilibrium test indicated that 93.97% of the population–locus combinations conformed to Hardy–Weinberg expectations. Accordingly, the complete dataset comprising 2,703 individuals (Supplementary Table S6) was retained for downstream analyses.

Genetic diversity of populations and groups

The number of total alleles (N_T) per population ranged from 55 to 127, and private alleles per population ranged from 0 to 4. The mean number of alleles (N_A) and effective number of alleles (N_E) across the 123 populations were 3.21 and 2.12, respectively. The mean expected heterozygosity (H_E) and allelic richness (A_R) were 0.46 and

1.99, respectively. Additionally, 39 populations had F_{IS} values below 0, which may indicate heterozygote excess relative to Hardy–Weinberg expectations (Supplementary Table S1).

Based on the results of genetic structure (Fig. 2a), all individuals were categorized into three groups: JR (*J. regia*), JS (*J. sigillata*), and a Hybrid group (all admixture individuals). For genetic diversity analyses, hybrids were treated as a single group, as these metrics are not designed to distinguish between potentially different hybrid origins. Among the three groups, JS had higher diversity than JR for all variables calculated (Table 1). For example, JS had the highest number of alleles (237) and private alleles (36), compared to 164 and three in the JR group. Additionally, the JS and Hybrid groups showed higher values of observed heterozygosity (H_O), expected heterozygosity (H_E), and allelic richness (A_R) compared to the JR group.

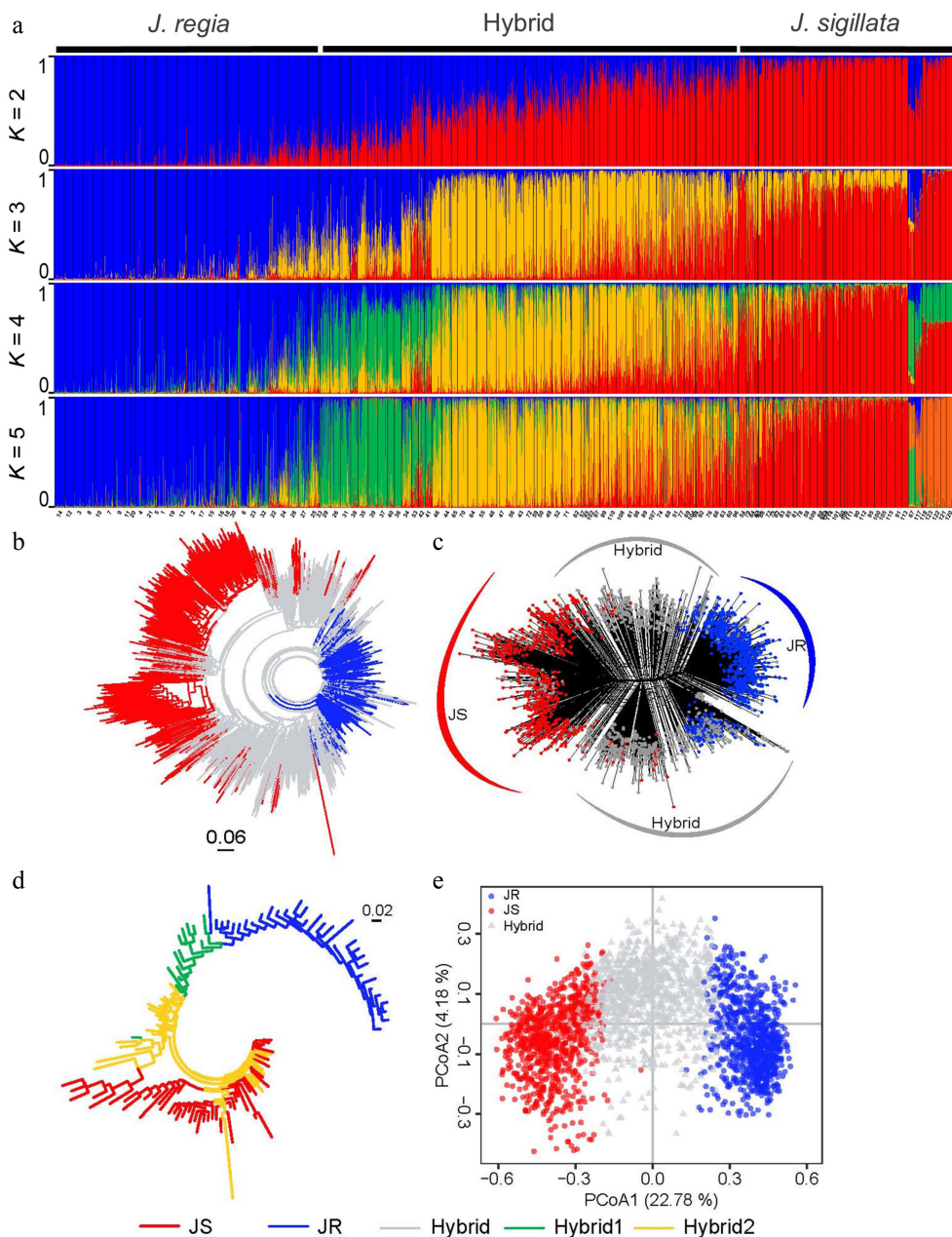


Fig. 2 Population structure of 2,703 walnut individuals (*Juglans regia* and *J. sigillata*) based on 31 microsatellite loci. In all five diagrams, various colors indicate distinct genetic groups: blue for JR (*J. regia*), red for JS (*J. sigillata*), grey for Hybrid, green for Hybrid1, and yellow for Hybrid2. (a) STRUCTURE results for $K = 2$ to $K = 5$. (b) Neighbor-joining tree, and (c) neighbor-net tree of all individuals based on Nei’s genetic distance (D_A). (d) Neighbor-joining tree for 123 populations based on Nei’s genetic distance. (e) Principal coordinates analysis (PCoA) of all individuals based on D_A genetic distance.

Genetic structure and its spatial patterns

The analysis of 2,703 individuals across 123 populations revealed a significant drop in Delta K values after $K = 2$, indicating that $K = 2$ was the optimal value. Based on the set Q value threshold, the 2,703 individuals were divided into three groups: JS (*J. sigillata*, 821 individuals), JR (*J. regia*, 871 individuals), and the Hybrid groups (1,011 individuals). Specifically, 871 individuals were assigned to the JR group, 821 individuals to the JS group, and 1,011 individuals to the Hybrid group (Fig. 2a, Supplementary Table S7). This finding was further supported by the STRUCTURE results at $K3-K5$ (Fig. 2a), which not only revealed extensive admixture between the JR and JS groups, but also identified subgroups within the Hybrid group, designated as Hybrid1 (green) and Hybrid2 (yellow).

Phylogenetic analyses based on the Neighbor-joining (NJ) and Neighbor-net trees were generally consistent with STRUCTURE results, clearly delineating the three genetic groups (Fig. 2b, c). Additionally, the Neighbor-joining tree revealed that the Hybrid group was split into Hybrid1 and Hybrid2 at the population level (Fig. 2d). The PCoA analysis also yielded similar results, further supporting the clustering of individuals and populations into three different groups (JR, JS, and Hybrid groups), with the first two axes together explaining 26.96% of the total genetic variation (PCoA1: 22.78%, PCoA2: 4.18%). Significant differences among groups were observed only along the PCoA1 axis, with the JS and JR groups clustering on opposite sides, while the Hybrid group clustered between them (Fig. 2e). Therefore, the PCoA, NJ, and Neighbor-net analyses all corroborated the STRUCTURE results, confirming the clustering of individuals into two main groups (JR and JS) and one Hybrid group, comprising two subgroups (Hybrid1 and Hybrid2).

Overall, the pairwise genetic differentiation (F_{ST}) results for the 123 populations indicated moderate genetic differentiation (Supplementary Table S8, Supplementary Fig. S1), ranging from -0.04 between CYM and LYM (both *J. sigillata*) to 0.63 between FGT (*J. sigillata*) and XJR (*J. regia*). Additionally, pairwise comparative analysis of genetic distance (D_A) showed a minimum of 0.02 between GQR and HSR, SGR and TBR, XXR and TBR, JCR and WQR, SGR and JCR (all *J. regia*), and a maximum of 0.64 between TBR (*J. regia*) and MMR (*J. sigillata*) (Supplementary Table S8, Supplementary Fig. S1).

Analysis of Molecular Variance (AMOVA) at different hierarchical scales revealed distinct patterns of genetic variation partitioning and genetic differentiation (Table 2). When all populations were analyzed without prior grouping, the majority of genetic variation (64%) was found within populations, while variation among populations accounted for 36% ($F_{ST} = 0.379$, $p = 0.001$). When AMOVA was done separately for the three groups, 71% of the genetic variation was within groups and 29% among groups ($F_{ST} = 0.290$, $p = 0.001$). However, when the two hybrid subgroups (Hybrid1 and Hybrid2) were further distinguished, the proportion of variation among groups decreased to 18%, accompanied by a reduction in genetic differentiation ($F_{ST} = 0.183$, $p = 0.001$). Additionally, genetic differentiation between the JR and JS groups is high ($F_{ST} = 0.464$, $p = 0.001$), with nearly half of the total genetic variation partitioned 46% between them, compared to within groups. Furthermore, the JS-Hybrid comparison shows 16% genetic variation between groups (84% within groups; $F_{ST} = 0.163$, $p = 0.001$), and the JR-Hybrid comparison shows 21% variation between groups (79% within groups; $F_{ST} = 0.215$, $p = 0.001$).

Spatial mapping of the three genetic groups indicated distinct geographic ranges for the JR and JS groups, with the JR group primarily distributed north of the Sichuan Basin and the JS group mainly distributed in southeastern Tibet and western Yunnan (Fig. 3c). The Hybrid group occupied regions adjacent to the JR and

Table 2. Analysis of molecular variance (AMOVA) of 123 populations and three groups of *Juglans regia* and *J. sigillata*.

Scale	Source	d.f.	Sum of squares	Mean squares	Percentage of variation (%)
Total	Among pops	122	24,768.26	203.02	36
	Within pops	2,580	38,593.24	14.96	64
	Total	2,702	63,361.50		100
Groups (JS/JR/Hybrid)	F_{ST}	0.37*			
	Among groups	2	13,572.47	6,786.24	29
	Within groups	2,700	49,789.03	18.44	71
JS/JR/Hybrid1/Hybrid2	Total	2,702	63,361.50		100
	F_{ST}	0.29*			
	Among groups	3	7,182.079	2,394.026	18
JS/JR	Within groups	2,699	43,865.67	8.12	82
	Total	2,702	51,047.749		100
	F_{ST}	0.18*			
JS/Hybrid	Among groups	1	12,759.23	12,759.23	46
	Within groups	1,690	29,482.96	17.45	54
	Total	1,691	42,242.19		100
JR/Hybrid	F_{ST}	0.46*			
	Among groups	1	3,599.82	3,599.82	16
	Within groups	1,830	37,035.60	20.24	84
JS/Hybrid	Total	1,831	40,635.42		100
	F_{ST}	0.16*			
	Among groups	1	4,518.02	4,518.02	21
JR/Hybrid	Within groups	1,880	33,059.49	17.58	79
	Total	1,881	37,577.51		100
	F_{ST}	0.21*			

d.f., degree of freedom. * $p = 0.001$.

JS groups, such as the Sichuan Basin, eastern Yunnan, and Guizhou. Regarding the distribution of the Hybrid group, hybrids in the north-western Sichuan Basin and the upper Three Parallel Rivers region were dominated by the JR genetic component. Conversely, hybrids in other parts of Yunnan, the southeastern Sichuan Basin, and Guizhou were primarily composed of the JS genetic component. Consequently, these subgroups were distinguished as Hybrid1 and Hybrid2, respectively, which exhibited distinct geographic distributions. Hybrid1 populations occurred primarily on the northwestern periphery of the Sichuan Basin, whereas Hybrid2 was predominantly found in cultivated stands to the south. This spatial separation suggests divergent origins. Genetically identified pure *J. sigillata* (JS) accessions corresponded closely to wild populations. In contrast, the Hybrid2 group consistently consisted of cultivated populations known locally as soft-shell walnut (Chinese: 绵核桃, Mian Hetao). These results indicate that much of the material cultivated in this region under the name *J. sigillata* is in fact of hybrid origin.

Genetic diversity (A_R , H_E) interpolation for the 123 populations suggested that *J. sigillata* populations have higher genetic diversity compared to the *J. regia* and Hybrid populations. Notably, high genetic diversity of *J. sigillata* was observed in Yunnan, Guizhou, and adjacent regions in southwestern China, which correspond to lower latitudes, while populations at the northern edge (higher latitudes) of the species range exhibited low genetic diversity (Fig. 3a, b, Table 1, Supplementary Table S7).

The Mantel test revealed a significant positive correlation between the geographical distance and genetic distance across populations ($r = 0.698$, $p = 0.000$) (Supplementary Fig. S4a). At the group level, the JR ($r = 0.595$, $p = 0.000$) and JS ($r = 0.611$, $p = 0.000$) groups exhibited significantly higher values than the Hybrid group ($r = 0.191$, $p = 0.014$) (Supplementary Fig. S4b–S4d). These results indicate that isolation by distance is stronger in the parental species

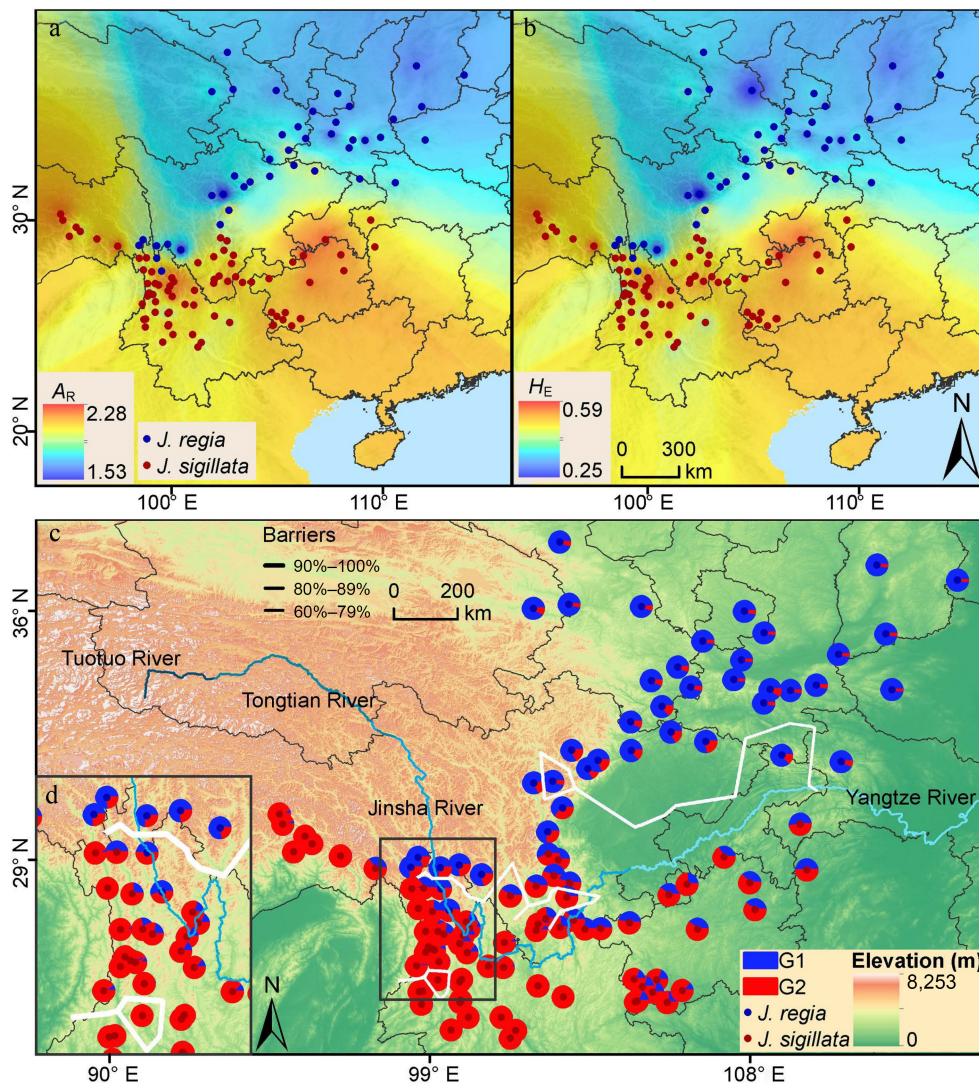


Fig. 3 Spatial patterns of genetic diversity and population structure in 123 walnut populations. Genetic diversity was mapped using IDW (Inverse distance weighted) interpolation in ArcMap: (a) allelic richness (A_R); (b) expected heterozygosity (H_E); (c) geographical distribution of the population structure. The map is based on the results of STRUCTURE at $K = 2$. (d) The inset shows the population from the Three Parallel Rivers region. The solid, thick white lines indicate the locations of the two most likely barriers identified through the Barrier analysis.

than in hybrids, consistent with the expectation that admixture can weaken spatial genetic structure. Moreover, the analysis of the relationship between genetic diversity (H_E) and latitude demonstrated a moderately negative correlation ($r = -0.481$, $p = 1e-04$), suggesting that genetic diversity decreased with increasing latitude. In contrast, correlations between genetic diversity (H_E) and both altitude and longitude were weakly negative but statistically significant ($r = -0.105$, $p = 0.004$; $r = -0.155$, $p = 3e-04$, respectively) (Supplementary Fig. S2).

The Barriers analysis identified two distinct geographical barriers, both with a strength of 99% (Fig. 3c). One barrier was located in the Sichuan Basin and Yangtze River, separating the *J. regia* and *J. sigillata* populations, and dividing the Hybrid group into two subgroups (i.e., Hybrid1 and Hybrid2). The second geographic barrier was located in the Three Parallel Rivers region, effectively separating the northern *J. regia* and southern *J. sigillata* populations. Although different distance metrics or analytical parameters could slightly alter the precise positions of the identified barriers, the overall spatial patterns remain robust. Consequently, the identified barriers provide strong geographical corroboration for the genetic

structure of the three groups (JR, JS, and Hybrid) inferred across the study region (Fig. 3c).

Inference of demographic history

While previous analyses treated hybrids as a single contemporary group, our demographic model explicitly distinguishes between two hybrid subgroups (Hybrid1 and Hybrid2) to model distinct temporal origins and differing parental contributions. DIYABC-RF analyses showed that the best fit model was scenario 5 (976/2,000 votes), with a posterior probability of 0.61 (Fig. 4, Supplementary Table S4). This model indicated that JS split from JR around 2.08 Ma BP (95% CI: 0.84–5.29 Ma BP), followed by an admixture event between JR and JS forming Hybrid2 around 9.75 ka BP (95% CI: 2.75–14.61 ka BP), with subsequent admixture between Hybrid2 and JR producing Hybrid1 around 6.51 ka BP (95% CI: 1.49–14.45 ka BP). The effective population sizes of JR, Hybrid1, Hybrid2, and JS were 2,413 (95% CI: 826–4,080), 4,182 (95% CI: 752–9,180), 4,004 (95% CI: 930–8,629), and 3,959 (95% CI: 1,438–6,636), respectively. Historical migration rates inferred from Migrate revealed relatively symmetrical historical migration rates between groups (Supplementary

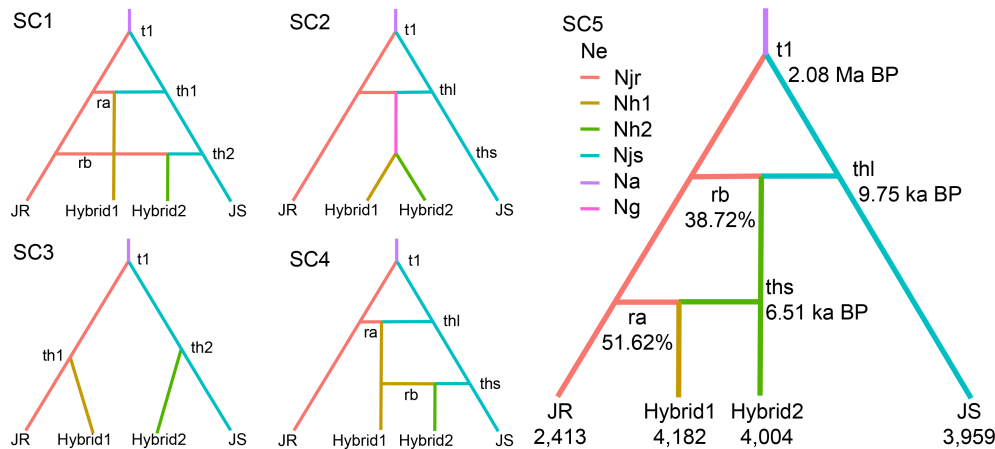


Fig. 4 Population demographic history of walnuts inferred from DIYABC-RF analysis. Demographic and historical parameters include six effective population sizes (N_{jr} , N_{h1} , N_{h2} , N_{js} , N_a , and N_g) and five events representing population divergence or admixture (t_1 , th_1 , th_2 , th_l , and th_s). For the scenarios with admixture, the parameters ra and rb represent the genetic contribution of each of the source groups. Branch colors indicate discrete effective population sizes in the model, with time measured in generations, and is not calibrated to scale. The descriptions of model parameters are provided in [Supplementary Table S4](#).

[Table S9](#)). The highest migration rate was observed from Hybrid2 to Hybrid1 (0.83; 95% CI: 0.61–1.04), whereas the lowest rate was detected from Hybrid1 to JS (0.72; 95% CI: 0.50–0.92).

Chloroplast genome

Genome data acquisition and assembly

A total of 204 individuals were newly sequenced, and 512 additional individuals were successfully downloaded from the NCBI database, resulting in 716 whole genome sequences in total. The sequencing depth of the downloaded whole-genome sequences ranged from $11\times$ to $91\times$, averaging $25\times$. After assembly and annotation, 716 chloroplast genomes were successfully obtained ([Supplementary Table S2](#)).

Haplotype clustering and phylogenetic analyses

Eighteen haplotypes (H1–H18) were identified across the 716 chloroplast genomes ([Supplementary Table S2](#)). Network analysis revealed two major clades: one dominated by 11 *J. regia* haplotypes (H1–H11) and another by seven *J. sigillata* haplotypes (H12–H18) ([Fig. 5c](#)). Among these, H1 and H5 were widespread, while H3, H4, and H6–H11 were private in *J. regia* ([Fig. 5c](#), [Supplementary Table S2](#)). Despite 92 mutation steps between the two clades, the presence of haplotypes from both species among individuals identified as hybrids is consistent with past hybridization. The star-shaped pattern observed in the *J. regia* clade is compatible with a recent population expansion. Moreover, multiple haplotypes (H1, H2, H5, H12, and H14) were shared across *J. regia* and *J. sigillata*, patterns that may reflect historical introgression and/or incomplete lineage sorting. Specifically, the widespread occurrence of haplotype H5 not only across its own species but also in *J. sigillata* individuals and hybrids, together with the population structure revealed by microsatellite data ([Fig. 3](#)), supports the possibility of chloroplast introgression from *J. regia* into *J. sigillata*.

Phylogenetic analysis revealed three highly supported clades (Maximum Likelihood Bootstrap Proportion (MBP) = 100, Posterior Probability [PP] = 1) in both the maximum likelihood (ML) and Bayesian inference (BI) phylogenetic trees ([Fig. 5a, b](#)). These clades comprised the outgroups, the *J. sigillata* clade and the *J. regia* clade, with the latter comprising two sub-clades (H1–H4 and H5–H11). No relationships within sub-clade H1–H11 had any statistical support,

mirroring the star-shaped pattern in the network analysis ([Fig. 5](#)). The most common shared haplotypes, H1 and H5, were located within subclades H1–H4 and H5–H11, respectively, with both subclades also containing private haplotypes.

Principal component analysis (PCA) of the chloroplast genomes revealed a clear separation between *J. regia* and *J. sigillata* along the first principal component (PCA1, explaining 86.61% of the variance) ([Fig. 5b](#)). *J. regia* was predominantly represented by a single high-frequency haplotype (H5; 521 samples), whereas *J. sigillata* was primarily associated with haplotype H12 (42 samples). The PCA2 further subdivided the *J. regia* clade into two subgroups, corresponding to haplotypes H1–H4 and H5–H11. Overall, the PCA clustering was consistent with the network and phylogenetic structure of haplotypes ([Fig. 5b](#)).

Spatial distribution of haplotypes

The geographic distribution of haplotypes provides further insight into the history of cultivation. Geographically, the seven haplotypes of the *J. sigillata* clade (H12–H18) were restricted to western Yunnan and southeastern Tibet, comprising 61 individuals ([Fig. 5d](#)). In contrast, the 11 haplotypes of the *J. regia* clade (H1–H11) spanned southwestern to northern China. Of these haplotypes, nine were private, among which seven were distributed in northern China, whereas H3 and H4 were unique to southeastern Tibet ([Fig. 5d](#)). Haplotypes from the *J. regia* clade occurred within all *J. regia* individuals, some hybrids, and some *J. sigillata* individuals. The widespread shared haplotype H5, distributed throughout the sampling range, was present in 181 individuals of *J. regia*, 29 individuals of *J. sigillata*, and 311 individuals of hybrids, as defined by microsatellite data ([Supplementary Table S2](#)). The other common haplotype, H1, was found among 64 *J. regia* individuals and 61 individuals of hybrids. Haplotype H1, which likely evolved from H5, occurs more frequently in northern China than in southwestern China ([Fig. 5d](#)). These patterns reflect complex maternal lineage sharing, which may arise from historical movement, introgression, incomplete lineage sorting, or a combination of all.

Maternal lineage bias in hybrids

Maternal lineage analysis based on cpDNA revealed a striking directional bias in introgression. Of the 1,011 hybrid individuals, only two individuals in southeastern Tibet possessed *J. sigillata* cpDNA,

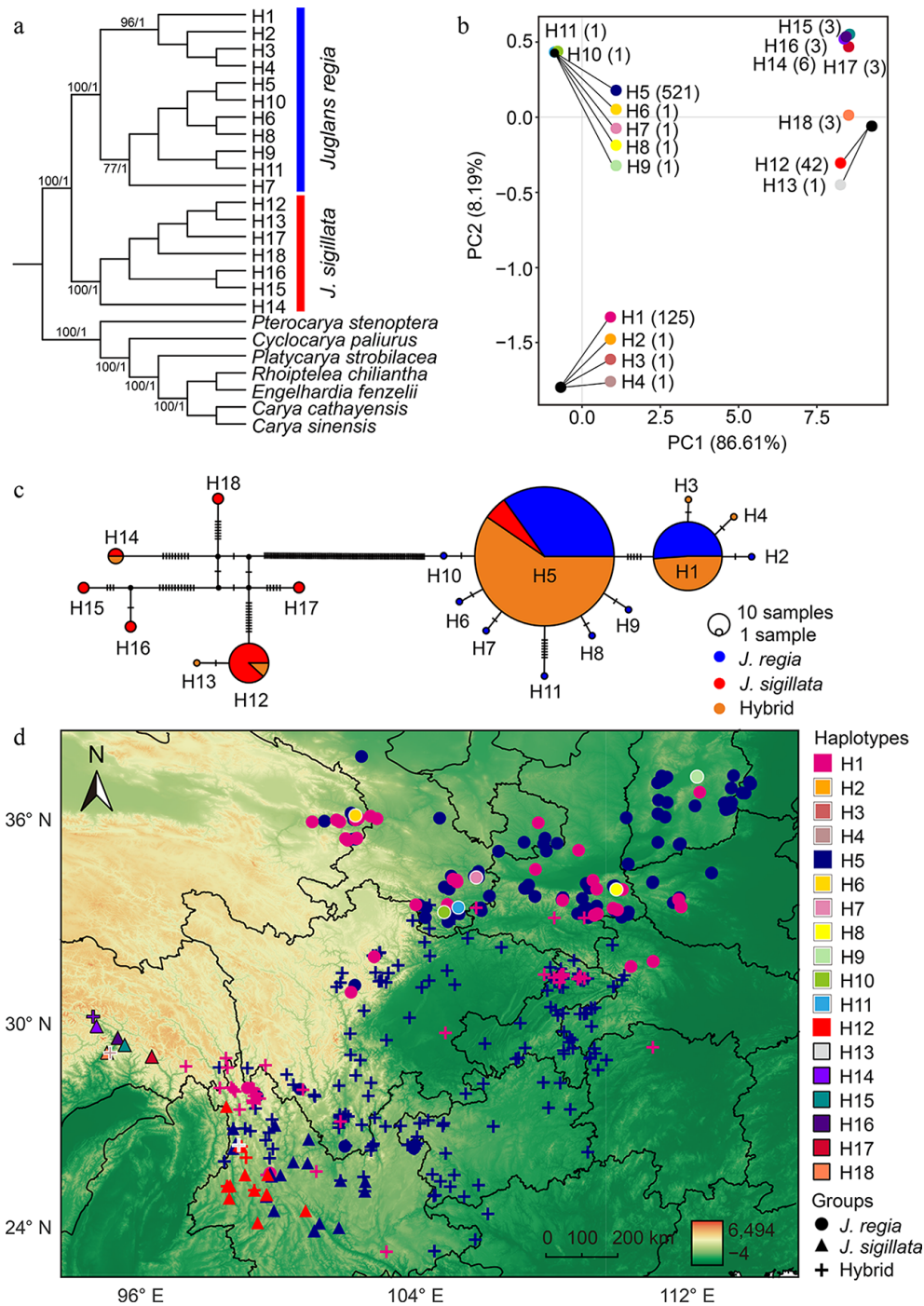


Fig. 5 Clustering relationships and geographical distribution of chloroplast genome haplotypes. (a) Phylogenetic tree inferred using maximum likelihood and Bayesian inference. Numbers on the branches indicate maximum likelihood bootstrap supports (BS) and Bayesian posterior probabilities (PP), respectively. Only major clades and values greater than 70 are shown. (b) Principal component analysis (PCA) of 18 haplotypes identified from 716 chloroplast genomes. The numbers next to the haplotypes indicate the number of individuals carrying each specific haplotype. Haplotypes that overlap due to small genetic distances are indicated by external lines. (c) Network of 18 haplotypes based on 716 chloroplast genomes. The size of each circle corresponds to the proportion of the specific haplotype. Vertical lines indicate mutational steps between haplotypes. (d) Geographical distribution of 18 haplotypes in 716 walnut individuals. Private haplotypes were indicated with white and black outlines for *J. regia* and *J. sigillata*, respectively. Different colors correspond to distinct haplotypes, while different shapes represent *J. regia* (circle), *J. sigillata* (triangle), and hybrids (plus sign).

while the remaining 1,009 (99.8%, 95% CI: 99.2%–99.97%) across the hybrid zone inherited their chloroplasts from *J. regia*. This proportion is significantly different from a 50:50 expectation (binomial exact test, $p < 0.0001$), indicating that *J. regia* acts almost exclusively as the maternal parent in interspecific hybridization events.

Integrating nuclear and chloroplast evidence

The nuclear and chloroplast data present a clear but complex history of walnut hybridization (Fig. 6). The biparental nuclear microsatellites show that hybridization was a two-way process, creating the Hybrid1 and Hybrid2 groups with mixed ancestry from both

species (Figs. 2–4). In contrast, the chloroplast genome tells a different part of the story. The maternally-inherited chloroplast genome is one of strong asymmetry; the vast majority of hybrids carry chloroplasts from *J. regia* (Fig. 5, Supplementary Table S2). The widespread *J. regia* H5 haplotype (Fig. 5d), found even in some pure *J. sigillata* underscores this pattern. So, while genes from both parents mixed freely in the hybrids, the maternal line was consistently dominated by *J. regia*.

Discussion

Drivers of walnut genetic diversity

Much of the cultivated walnuts in the study area of southwestern China comprises hybrids between *J. regia* and *J. sigillata*, including all examined populations that were cultivated as *J. sigillata*. According to various parameters, genetic diversity was higher in wild *J. sigillata* (JS) than the hybrids or *J. regia* (JR) (Table 1), consistent with Fan et al.^[32]. Additionally, the greatest genetic differentiation (F_{ST}) and genetic distance were observed between the JR and JS groups, while the hybrid group was positioned between them (Supplementary Fig. S3, Table 2). The unique geographical conditions and historical factors within the native range of *J. sigillata*, such as limited influence from Pleistocene glacial periods in

southwestern China^[84–87], may have contributed to preserving population-level genetic variation. However, *J. sigillata* has experienced long-term human disturbance, which may have led to the elimination of some alleles, resulting in low genetic variability in the cultivated populations. Furthermore, low genetic variation in cultivated populations could be a consequence of artificial selection and/or a bottleneck due to breeding from a small number of genotypes^[88,89]. A particular issue in this genus is that the extensive grafting of cultivars onto natural populations of *J. sigillata*, has led to rapid genetic erosion over the past three decades^[24]. Conversely, founder effects following their ancient introductions of *J. regia* from the Western Himalaya via Central Asia^[23,32,46,47] might be responsible for the lower genetic diversity identified in *J. regia* populations in northern China. Notably, hybrids are discussed as a single genetic group in this context. The finer subdivision into Hybrid1 and Hybrid2 is applied specifically in the demographic analyses to interpret distinct hybridization histories.

Latitude was negatively correlated with genetic diversity based on microsatellite markers and latitude (Supplementary Fig. S2). Populations of *J. sigillata* (JS) and the hybrids groups had higher genetic diversity at lower latitudes. This genetic diversity gradient in walnuts likely reflects the impacts of cultivation, selection, and long-term agricultural practices such as the domestication and spread of walnuts across southwestern China. Therefore, although it meets

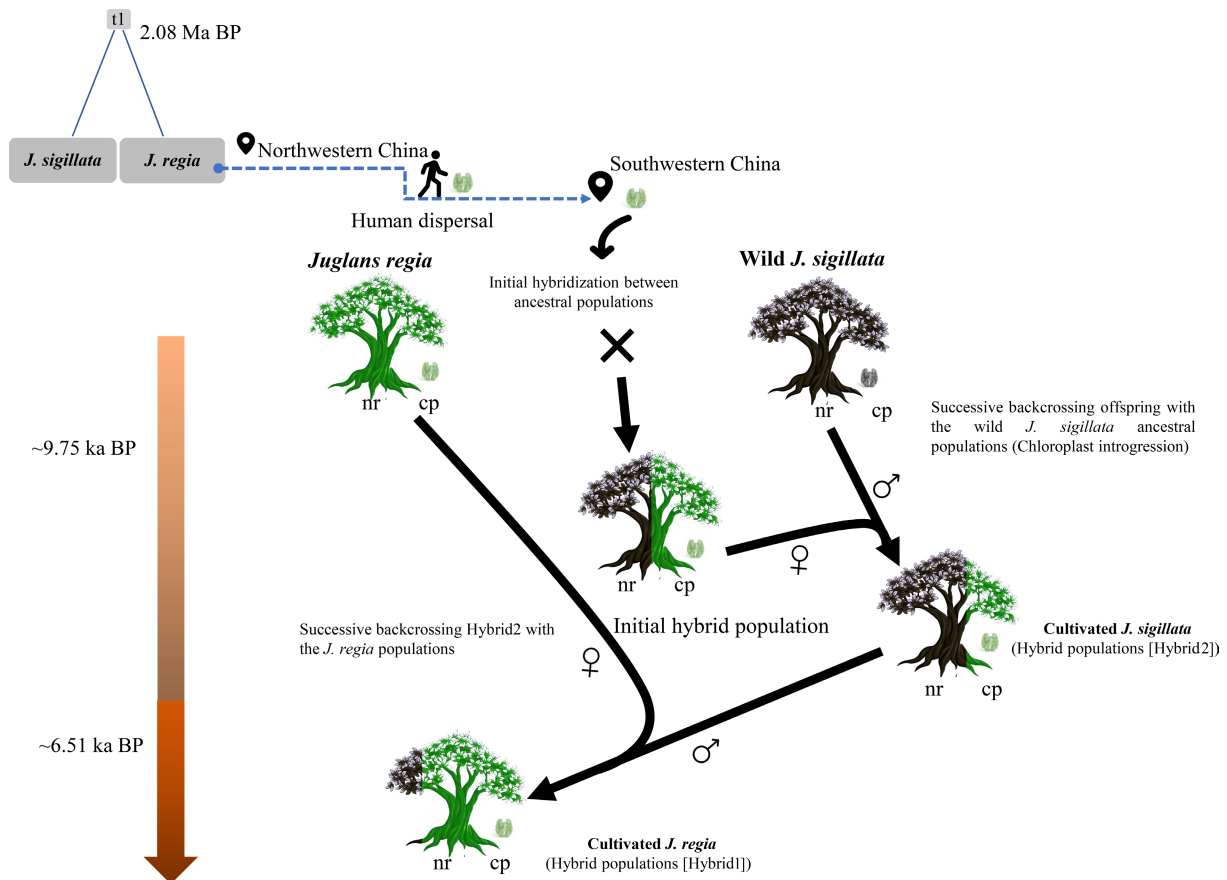


Fig. 6 Schematic model for the hybrid origin of cultivated walnuts in southwestern China. *Juglans sigillata* and *J. regia* diverged approximately 2.08 million years ago. Briefly, during the Holocene, *J. regia* was spread to northern China and was later introduced into southwestern China following the emergence of cultivated hybrid walnuts there. The conceptual diagram depicts successive hybridization events and nuclear genome (nr) exchange, as shown by the color-coded tree stems representing nuclear genome composition. The chloroplast genome (cpDNA) is maternally inherited and is thus denoted by the female symbol (♀) to indicate its maternal origin, whereas the paternal parent is represented by the male symbol (♂). This model proposes the gradual chloroplast introgression in *J. sigillata*, driven by repeated genetic exchanges through hybridization. Notably, introgression is an ongoing process, not a discrete event.

the expectations of the Latitudinal Diversity Gradient (LDG) hypothesis, whereby biodiversity increases from the polar regions towards the equator^[90], the pattern in walnuts might be partially anthropogenic in its origin. Nonetheless, a correlation analysis suggests that environmental factors (Supplementary Fig. S2), which intensify with increasing latitude, may have contributed to the observed genetic variation gradient^[91,92] through natural selection. Therefore, the genetic differentiation observed in the latitudinal gradients of these populations likely results from a synergy of natural factors and human activities.

The highest genetic diversity was detected between 27° N and 30° N latitude, with prominent peaks at the western and eastern edges of southwestern China. Increased genetic differentiation at lower latitudes might result from limited gene flow in these regions^[93]. In contrast, high genetic diversity in the western part of this region may be due to topographic and climatic factors^[94]. Moreover, our findings indicated that the Three Parallel Rivers region harbors the highest genetic diversity (Fig. 3), which is consistent with previous meta-analyses of plant genetic diversity^[94,95]. The region is known to have served as a Pleistocene glacial refugia for many taxa^[95–97]. Because of the prolonged isolation and long-term persistence, populations in refugia are expected to retain higher genetic diversity^[98,99]. Furthermore, the introduction of non-native *J. regia* germplasm and subsequent introgression between species may have enhanced regional genetic diversity. This further underscores the role of human activity in shaping genetic landscapes.

Genetic landscape and origin of cultivated walnut

The population structure inferred from microsatellite data (Figs. 2, 3b) clearly distinguished the three groups (JR, JS, and Hybrid), with hybrids occupying an intermediate position while remaining clearly differentiated from both JR and JS. In contrast, chloroplast genome data (Fig. 5), reflecting maternal inheritance, revealed only two clades, corresponding to JR and JS. Within the JS group, seven haplotypes characterized by multiple mutational steps were identified, whereas the JR group contained 11 haplotypes separated by only a few mutational steps. Two haplotypes (H1 and H5) were dominant and widely distributed in JR, with the remaining haplotypes derived from H5. This topology forms a classic star-like phylogeny centered on H5, which strongly suggests a recent population expansion. These maternal lineage patterns are consistent with long-term persistence and isolation in *J. sigillata*, as opposed to a recent population expansion of *J. regia* in southwestern China, possibly reflecting anthropogenic dispersal.

Neighbor-joining analysis could not definitively determine whether the hybrid group was closer to the JR and JS groups (Fig. 2b, d). However, further analysis grouped populations into JR, JS, and two hybrid subgroups, Hybrid1 and Hybrid2, indicating their closer genetic affinity to JR and JS groups, respectively (Fig. 2d). Hybrid1 is morphologically similar to the JR group from northern China, while Hybrid2 consists of cultivated material morphologically identified as *J. sigillata* and commonly grown as soft-shell walnut (Chinese: 绵核桃, Mian Hetao). The JS group possessed far more private alleles (36) than the Hybrid (5) or JR groups (3). Likewise, JS scored highest on all other genetic diversity measures (Table 1). The low genetic diversity in both hybrid groups relative to JS (Supplementary Table S10) supports a scenario in which Hybrid1 and Hybrid2 originated from discrete hybridization events involving small effective population sizes (Fig. 6). This pattern is likely due to major geographic features, specifically, the mountains around Sichuan Basin and Yangtze River,

which acted as significant barriers to north-south gene flow. Subsequent human-mediated dispersal likely facilitated the spread of these hybrids while reducing genetic diversity along dispersal routes. This further supports the view that JS represents long-standing wild populations, while JR and the hybrids appear to have been introduced more recently.

According to ABC analysis, *J. regia* and *J. sigillata* diverged around 2.08 Ma, during the early Pleistocene (Fig. 4). This estimate is broadly consistent with the Pleistocene divergence reported by Fan et al.^[32] (~1.41 Ma) and Ding et al.^[23] (~0.85 Ma). Although the estimates differ in timing, all studies confirm a critical period of diversification in the early to middle Pleistocene. The older divergence date in our analysis may be due to incomplete lineage sorting (ILS)^[100,101], undetected ancient gene flow^[102], or differences in sampling range^[103]. A plausible explanation is that habitat fragmentation caused by Pleistocene climate oscillations promoted the divergence of *J. regia* and *J. sigillata*, likely in the Himalaya^[32,104]. In the early Holocene, *J. regia* was likely first domesticated in the western Himalaya^[32], after which it spread to Europe, Central Asia, and northern China^[32,46,47,105].

Our results (Figs. 4, 5), consistent with previous studies^[23,48,106–108], indicate gene flow between *J. regia* and *J. sigillata* in southwestern China. Notably, we identified two hybrid groups that arose at different times (Fig. 4). The introduction of domesticated *J. regia* from northern China and its cultivation in southwestern China likely facilitated their hybridization, leading to the emergence of Hybrid2 around the early Holocene (~9.75 ka BP) (Fig. 6). Chloroplast genome data, reflecting strictly maternal inheritance, revealed that most hybrid individuals carried haplotypes common in *J. regia*, indicating that maternally inherited lineages from JR are common among hybrids. Our formal assessment of maternal lineage bias statistically confirms this predominance, demonstrating a significant asymmetry. The dominance of a limited number of these haplotypes, particularly H5, indicates that these specific maternal lines were preferentially selected and propagated, potentially reflecting their adaptive benefits. This systematic bias demonstrates directed human action, the repeated use of introduced *J. regia* as the seed parent, rather than a natural demographic process. Consequently, this pattern aligns with a domestication scenario in which superior *J. regia* varieties were served as maternal stock, and their hybrid progeny were subsequently spread across the region. Further intensification of human activity may have promoted further northward admixture between Hybrid2 and *J. regia*, yielding Hybrid1 by the middle Holocene (~6.51 ka BP) (Fig. 6). This suggests that the origin of cultivated walnut in this region predates the introduction of agriculture^[46,109], highlighting the influence of early human activities, such as hunting and gathering.

Several lines of evidence point to human-mediated dispersal as the primary driver for the walnut spread. First, a significant correlation was found between genetic variation and geographical distance for all populations and each group (Supplementary Fig. S2), suggesting that human movement, climate, and geophysical factors likely played a key role in shaping the genetic patterns^[103]. Second, two haplotypes derived from *J. regia*, H5 and H1, were widely distributed the study region (Fig. 5). Most other haplotypes, many of which were private, were derived from H5 and differed by only one or a few mutational steps, often observed in single individuals, indicating a recent and rapid population expansion with insufficient time for substantial mutation accumulation. Third, although rodents contribute to walnut seed dispersal^[110], they are unlikely to account for the extensive and rapid dissemination of hybrids across heterogeneous terrain characterized by numerous mountains and rivers^[111].

Finally, this genetic pattern is consistent with historical evidence: multiple waves of early human migration from the Yellow River basin into southwest China have been documented through both genetic and archaeological evidence^[112,113], with particularly intensified population movements occurring during the Qin (221–206 BC) and Han (206 BC–AD 220) dynasties^[114], including regions where Hybrid1 and Hybrid2 are found. Furthermore, historical trade routes in ancient China, such as the 'Tibetan-Yi Corridor' and the 'Southwest Silk Road', enabled the long-distance movement of humans and crops^[115–117]. Collectively, our results indicate that large-scale human population movement may play a key role in the spread across southwestern China. The introgression of a few widespread *J. regia* haplotypes into diverse local *J. sigillata* populations provides a parsimonious explanation for the origin of cultivated walnut via human-mediated hybridization.

Haplotype H5 of the chloroplast genome was also present in many populations morphologically identified as *J. sigillata* (Fig. 5, Supplementary Table S2), consistent with additional contact and introgression between the two species, even where nuclear data suggest the samples as relatively pure JS (Fig. 2). The seven haplotypes originating from *J. sigillata* were confined to the westernmost part of its range. All private haplotypes were restricted to the eastern Himalaya. These populations occur in remote natural ecosystems, as confirmed by our field observations. Therefore, *J. regia* was likely brought south by humans, coming into contact with wild and/or cultivated *J. sigillata* and leading to the origin of multiple hybrid genotypes with varying parental contributions. Therefore, our data suggest a domestication history where human-introduced *J. regia* hybridized with local wild *J. sigillata*. This was not a random process. The frequent use of *J. regia* as the maternal parent, shown by the dominant H5 haplotype indicates a deliberate practice. We propose that people selectively cultivated and spread the hybrid offspring from these preferred *J. regia* trees, which ultimately formed the basis of the cultivated walnut germplasm known today as *J. sigillata* (Fig. 6). In essence, human-mediated introgression systematically transformed the local walnut population.

It is also important to consider the limits of our study. While our models point to specific times for divergence and hybridization, these are ultimately estimates based on the best-fitting scenarios, not precise dates. Furthermore, factors like incomplete lineage sorting or historical gene flow from populations we didn't sample could make the actual history more complex. Similarly, the genetic patterns we see fit very well with a story of human-mediated dispersal, especially the spread of a few walnut types across difficult terrain; it remains an inference based on patterns that are difficult to explain by natural dispersal alone. Finally, because we focused specifically on southwestern China, our findings might not fully unravel the entire history of these walnuts across their whole range. Despite these points, our study provides a powerful genetic framework for the origin of cultivated walnut, which can be tested more directly by future studies integrating ancient DNA, archaeology, and genomics.

Conclusions

Using microsatellite markers and chloroplast genome data, this study examined the genetic diversity and population structure of 123 walnut populations to elucidate the evolutionary history of cultivated walnuts in southwestern China. Overall, our findings point to a history where natural landscape and human activity together shaped the genetics of walnuts in this region. The deep genetic split between *J. regia* and *J. sigillata* appears to have been

maintained by the Sichuan Basin and the Yangtze River acting as long-term physical barriers. Their differing levels of diversity tell separate stories: *J. sigillata* likely persisted in glacial refugia, while *J. regia* shows signs of human-facilitated introduction. However, discordance between nuclear and chloroplast signals indicates that, despite the strong genetic division, gene flow has occurred since the Holocene, giving rise to two cultivated hybrid groups around the Sichuan Basin. Notably, *J. regia* appears to have contributed chloroplast genomes to the hybrids, as well as to some *J. sigillata* individuals. In the end, the patterns we see do not originate from just one process, but from the combined and sometimes contrasting effects of geography, climate history, and human activities. Future work will need more extensive geographic and genomic sampling to fully unravel the species' complex evolutionary history as well as the adaptive benefits and functional significance of gene exchange. Moving forward, efforts should focus on harnessing the superior germplasm hidden in natural mixed genetic pools, using an integrated genomic and ecological approach to bolster local adaptation and climate resilience.

Author contributions

The authors confirm their contributions to the paper as follows: planned and designed the research: Liu J, Shahi Shavvon R, Li DZ; conducted the fieldwork: Liu J, Fan PZ, Wu ZY, Luo YH, Gao LM; performed the experiments: Qi HL; analyzed the data: Liu J, Qi HL, Fan PZ, Liu CM; wrote the first draft: Liu J, Shahi Shavvon R; input for the first draft: Milne RI, Wu ZY, Luo YH, Khan R, Gao LM, Wang YH, Wambulwa MC, Li DZ. All authors reviewed the results and approved the final version of the manuscript.

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (32170398, 31770367), the CAS 'Light of West China' Program, the Top-notch Young Talents Project of Yunnan Provincial 'Ten Thousand Talents Program' (YNWR-QNBJ-2018-146), the Natural Science Foundation of Yunnan (202201AT070222), the Key Research Program of Frontier Sciences, CAS (ZDBS-LY-7001), and the Fund of Yunnan Key Laboratory of Crop Wild Relatives Omics (CWR-2024-04). Jie Liu and Zeng-Yuan Wu were also supported by the China Scholarship Council (202304910135 and 202304910138) for a one-year study at the University of Toronto, Canada. Raees Khan was supported by the Postdoctoral International Exchange Program of the Office of China Postdoctoral Council, and the Caiyun Postdoctoral Program of Yunnan Province, including the Postdoctoral Targeted Funding and Postdoctoral Research Fund of Yunnan Province. Moses Wambulwa was supported by the CAS-ANSO Fellowship Program (CAS-ANSO-FS-2025-14).

Dr. Jie Liu extends his gratitude to his father, Mr. Xue-Wen Liu, his brother Mr. Tao Liu, and the students and volunteers for their efforts in field sampling. We also acknowledge Miss Ying Chen and Xing Kong, Dr. Lin-Jiang Ye, Mrs. Zuo-Ying Xiahou, Zu-Chang Xu, and Ye-Chuan Jin for their assistance with laboratory work. We sincerely thank the editor and the three anonymous reviewers for their constructive comments and valuable suggestions. Molecular experiments were conducted at the Laboratory of Molecular Biology, and

chloroplast genome data analysis was facilitated by the iFlora HPC Center (iFlora High-Performance Computing Center), Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences.

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary information accompanies this paper online at: <https://doi.org/10.48130/forres-0026-0018>.

Dates

Received 19 September 2025; Revised 9 February 2026; Accepted 24 March 2026; Published online 15 May 2026

References

- [1] Xia C, Zuo Y, Xue T, Kang M, Zhang H, et al. 2023. The genetic structure and demographic history revealed by whole-genome resequencing provide insights into conservation of critically endangered *Artocarpus nanchuanensis*. *Frontiers in Plant Science* 14:1224308
- [2] Choi IS, Han E, Wojciechowski MF, Heo TK, Park JS, et al. 2023. The genetic structure and demographic history of *Zabelia tyaihyonii*, endemic to Korean limestone karst forests, based on genome-wide SNP markers. *Ecology and Evolution* 13:e10252
- [3] Yakimowski SB, Southcott L, Barrett SCH. 2022. Contrasting patterns of genetic diversity and differentiation across the continental disjunct range of a sexually polymorphic aquatic plant. *Annals of Botany* 130:27–40
- [4] Besnard G, Terral JF, Cornille A. 2018. On the origins and domestication of the olive: a review and perspectives. *Annals of Botany* 121:385–403
- [5] Burle ML, Fonseca JR, Kami JA, Gepts P. 2010. Microsatellite diversity and genetic structure among common bean (*Phaseolus vulgaris* L.) landraces in Brazil, a secondary center of diversity. *Theoretical and Applied Genetics* 121:801–813
- [6] Yang MY, Yang EC, Kim MS. 2020. Genetic diversity hotspot of the amphi-Pacific macroalga *Gloiopeltis furcata sensu lato* (Gigartinales, Florideophyceae). *Journal of Applied Phycology* 32:2515–2522
- [7] Gamba D, Muchhala N. 2023. Pollinator type strongly impacts gene flow within and among plant populations for six Neotropical species. *Ecology* 104:e3845
- [8] Cheng J, Kao H, Dong S. 2020. Population genetic structure and gene flow of rare and endangered *Tetraena mongolica* Maxim. revealed by reduced representation sequencing. *BMC Plant Biology* 20:391
- [9] Hufford MB, Lubinsky P, Pyhäjärvi T, Devengeno MT, Ellstrand NC, et al. 2013. The genomic signature of crop-wild introgression in maize. *PLoS Genetics* 9:e1003477
- [10] Heslop-Harrison JS, Schwarzacher T. 2007. Domestication, genomics and the future for banana. *Annals of Botany* 100:1073–1084
- [11] Wu GA, Prochnik S, Jenkins J, Salse J, Hellsten U, Murat F, et al. 2014. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature Biotechnology* 32:656–662
- [12] Warschefsky EJ, von Wettberg EJB. 2019. Population genomic analysis of mango (*Mangifera indica*) suggests a complex history of domestication. *New Phytologist* 222:2023–2037
- [13] Olsen KM, Wendel JF. 2013. A bountiful harvest: genomic insights into crop domestication phenotypes. *Annual Review of Plant Biology* 64:47–70
- [14] Flowers JM, Hazzouri KM, Gros-Balthazard M, Mo Z, Koutroumpa K, et al. 2019. Cross-species hybridization and the origin of North African date palms. *Proceedings of the National Academy of Sciences of the United States of America* 116:1651–1658
- [15] Aerts R, Berecha G, Gijbels P, Hundera K, Van Glabeke S, et al. 2013. Genetic variation and risks of introgression in the wild *Coffea arabica* gene pool in south-western Ethiopian montane rainforests. *Evolutionary Applications* 6:243–252
- [16] Delplancke M, Alvarez N, Espindola A, Joly H, Benoit L, et al. 2012. Gene flow among wild and domesticated almond species: insights from chloroplast and nuclear markers. *Evolutionary Applications* 5:317–329
- [17] Cornille A, Giraud T, Smulders MJM, Roldán-Ruiz I, Gladieux P. 2014. The domestication and evolutionary ecology of apples. *Trends in Genetics* 30:57–65
- [18] Wedger MJ, Schumann AC, Gross BL. 2021. Candidate genes and signatures of directional selection on fruit quality traits during apple domestication. *American Journal of Botany* 108:616–627
- [19] Davies T, Watts S, McClure K, Migicovsky Z, Myles S. 2022. Phenotypic divergence between the cultivated apple (*Malus domestica*) and its primary wild progenitor (*Malus sieversii*). *PLoS One* 17:e0250751
- [20] Cornille A, Antolín F, Garcia E, Vernesi C, Fietta A, Brinkkemper O, et al. 2019. A multifaceted overview of apple tree domestication. *Trends in Plant Science* 24:770–782
- [21] Zhou H, Zhao P, Woeste K, Zhang S. 2021. Gene flow among wild and cultivated common walnut (*Juglans regia*) trees in the Qinling Mountains revealed by microsatellite markers. *Journal of Forestry Research* 32:2189–2201
- [22] Yan LJ, Fan PZ, Wambulwa MC, Qi HL, Chen Y, et al. 2024. Human-associated genetic landscape of walnuts in the Himalaya: implications for conservation and utilization. *Diversity and Distributions* 30:e13809
- [23] Ding YM, Cao Y, Zhang WP, Chen J, Liu J, et al. 2022. Population-genomic analyses reveal bottlenecks and asymmetric introgression from Persian into iron walnut during domestication. *Genome Biology* 23:145
- [24] Liu J, Magige EA, Fan PZ, Wambulwa MC, Luo YH, et al. 2023. Genetic imprints of grafting in wild iron walnut populations in southwestern China. *BMC Plant Biology* 23:423
- [25] Zhang Q, Ree RH, Salamin N, Xing Y, Silvestro D. 2021. Fossil-informed models reveal a boreotropical origin and divergent evolutionary trajectories in the walnut family (Juglandaceae). *Systematic Biology* 71:242–258
- [26] Aradhya MK, Potter D, Gao F, Simon CJ. 2007. Molecular phylogeny of *Juglans* (Juglandaceae): a biogeographic perspective. *Tree Genetics & Genomes* 3:363–378
- [27] Woodworth RH. 1930. Meiosis of microsporogenesis in the Juglandaceae. *American Journal of Botany* 17:863–869
- [28] Mu Y, Xi R, Lu Z. 1990. 核桃属部分种的小孢子发生观察及核型分析 [Microsporogenesis observation and karyotype analysis of some species in genus *Juglans* L.]. *植物科学学报* [Plant Science Journal] 8:301–310 (in Chinese)
- [29] Lu AM, Stone DE, Grauke LJ. 1999. Juglandaceae. In *Flora of China*, eds. Wu ZY, Raven PH. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press. pp. 277–285 www.researchgate.net/publication/285744209_Juglandaceae
- [30] Xi RT, Zhang YP. 1996. 核桃属植物志 [Walnut flora]. In *中国果树志* [China Fruit-Plant Monograph], eds. Zhao HC, Tian BF. Beijing: China Forestry Publishing House. 264 pp. (in Chinese) www.nhbs.com/china-fruit-plant-monograph-volume-3-walnut-flora-chinese-book?srsltid=AfmBOoqXBiu_oXNOuJ33chlgb51QUAYnFWUef4p5vyyC-cYUNuHEyTnY
- [31] Wang H, Pan G, Ma Q, Zhang J, Pei D. 2015. The genetic diversity and introgression of *Juglans regia* and *Juglans sigillata* in Tibet as revealed by SSR markers. *Tree Genetics & Genomes* 11:1
- [32] Fan PZ, Zhu GF, Wambulwa MC, Milne RI, Wu ZY, et al. 2026. Genetic origins and climate-induced erosion in economically important Asian walnuts. *Conservation Biology* 40:e70125
- [33] Gunn BF, Aradhya M, Salick JM, Miller AJ, Yang Y, et al. 2010. Genetic variation in walnuts (*Juglans regia* and *J. sigillata*; Juglandaceae): species distinctions, human impacts, and the conservation of agrobiodiversity in Yunnan, China. *American Journal of Botany* 97:660–671
- [34] Li X, Wang X, Zhang D, Huang J, Shi W, et al. 2024. Historical spread routes of wild walnuts in Central Asia shaped by man-made and nature. *Frontiers in Plant Science* 15:1394409

- [35] Li H, Zuo X, Kang L, Ren L, Liu F, Liu H, et al. 2016. Prehistoric agriculture development in the Yunnan-Guizhou Plateau, southwest China: archaeobotanical evidence. *Science China Earth Sciences* 59:1562–1573
- [36] Dal Martello R, Li X, Fuller DQ. 2021. Two-season agriculture and irrigated rice during the Dian: radiocarbon dates and archaeobotanical remains from Dayingzhuang, Yunnan, Southwest China. *Archaeological and Anthropological Sciences* 13:62
- [37] Chen J. 2007. 波西、营盘山及沙乌都——浅析岷江上游新石器文化演变的阶段性 [Boxi, Yingpanshan and Shawudu – the developmental stages of Neolithic cultural change in the upper Min River]. *考古与文物* [Archaeology and Cultural Relics] 5:65–70 (in Chinese)
- [38] Dal Martello R. 2022. The origins of multi-cropping agriculture in Southwestern China: archaeobotanical insights from third to first millennium B.C. Yunnan. *Asian Archaeology* 6:65–85
- [39] Zhang J, Storzum MJ, Chen W, Rao Z, Hamilton R, Zheng Z, et al. 2023. Climatic shifts, geomorphic change, ancient routes of migration and adaptation in southwestern China: Site formation processes at Luojiaba, Sichuan Province. *Geoarchaeology* 38:351–370
- [40] Hein A. 2014. Introduction: diffusionism, migration, and the archaeology of the Chinese border regions. In *The "Crescent-Shaped Cultural-Communication Belt": Tong Enzheng's Model in Retrospect*, ed. Hein A. Oxford: Archaeopress. pp. 1–17 www.academia.edu/75667182/Diffusionism_Migration_and_the_Archaeology_of_the_Chinese_Border_Region
- [41] Hein AM. 2014. Interregional contacts and geographic preconditions in the prehistoric Liangshan region, Southwest China. *Quaternary International* 348:194–213
- [42] Allard F. 2005. Frontiers and boundaries: the Han Empire from its southern periphery. In *Archaeology of Asia*, ed. Stark MT. Malden: Blackwell. pp. 233–245 doi: [10.1002/9780470774670.ch11](https://doi.org/10.1002/9780470774670.ch11)
- [43] Guo Q. 2014. Plant hybridization: the role of human disturbance and biological invasion. *Diversity and Distributions* 20:1345–1354
- [44] Burgarella C, Barnaud A, Kane NA, Jankowski F, Scarcelli N, et al. 2019. Adaptive introgression: an untapped evolutionary mechanism for crop adaptation. *Frontiers in Plant Science* 10:4
- [45] Zhang S, Li Y, Li Y, Zhang Y, Hao Y, et al. 2024. Identification and genetic diversity analysis of specific walnut F1 progeny based on SSR molecular markers: taking heart-shaped walnuts and Jinghong 1 as examples. *Scientific Reports* 14:27869
- [46] Chen CJ, Pang XX, Ding YM, Zhang WP, Yang Y, et al. 2026. Resolving sampling and population-size biases in domestication genomics supports a South Asian origin of walnuts. *Genome Biology* 27:61
- [47] Qi H, Fan P, Wang Y, Liu J. 2023. Genetic diversity and population structure of *Juglans regia* from six provinces in northern China. *Biodiversity Science* 31:23120
- [48] Ji F, Ma Q, Zhang W, Liu J, Feng Y, et al. 2021. A genome variation map provides insights into the genetics of walnut adaptation and agronomic traits. *Genome Biology* 22:300
- [49] Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19:11–15
- [50] Liu J, Gao LM. 2011. 三种红豆杉属植物总 DNA 提取方法的分析与比较 [Analysis and comparison of three different total DNA extraction methods for *Taxus* species]. *广西植物* [Guihaia] 31:244–249, 159 (in Chinese)
- [51] Xu ZC, Jin YC, Milne RI, Xiahou ZY, Qin HT, et al. 2020. Development of 32 novel microsatellite loci in *Juglans sigillata* using genomic data. *Applications in Plant Sciences* 8:e11328
- [52] Xiahou ZY, Wambulwa MC, Xu ZC, Ye LJ, Fan PZ, et al. 2023. A multiplex PCR system of novel microsatellite loci for population genetic application in walnuts. *Plants* 12:4101
- [53] Peakall R, Smouse PE. 2012. GenA1Ex 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539
- [54] Ellstrand NC, Roose ML. 1987. Patterns of genotypic diversity in clonal plant species. *American Journal of Botany* 74:123–131
- [55] Halkett F, Simon JC, Balloux F. 2005. Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution* 20:194–201
- [56] Balloux F, Lehmann L, de Meeùs T. 2003. The population genetics of clonal and partially clonal diploids. *Genetics* 164:1635–1644
- [57] Eckert CG. 2001. The loss of sex in clonal plants. *Evolutionary Ecology* 15:501–520
- [58] 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
- [59] Dieringer D, Schlotterer C. 2003. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* 3:167–169
- [60] Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- [61] 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
- [62] Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806
- [63] Rosenberg NA. 2004. Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes* 4:137–138
- [64] Wickham H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. 1st Edition. New York: Springer. 213 pp. doi: [10.1007/978-0-387-98141-3](https://doi.org/10.1007/978-0-387-98141-3)
- [65] Langella O. 2018. *POPULATIONS 1.2.31. Population genetic software*. France: CNRS UPR9034. www.bioinformatics.org/~tryphon/populations
- [66] Yu G, Smith DK, Zhu H, Guan Y, Lam TT. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 8:28–36
- [67] Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23:254–267
- [68] Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, et al. 2018. *Vegan: Community Ecology Package*. R package version 2.5-3. <https://cran.r-project.org/web/packages/vegan/index.html>
- [69] Manni F, Guérard E, Heyer E. 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Human Biology* 76:173–190
- [70] Collin FD, Durif G, Raynal L, Lombaert E, Gautier M, et al. 2021. Extending approximate Bayesian computation with supervised machine learning to infer demographic history from genetic polymorphisms using DIYABC Random Forest. *Molecular Ecology Resources* 21:2598–2613
- [71] Waples RS, Do C. 2010. Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications* 3:244–262
- [72] Blair C, Jiménez Arcos VH, de la Cruz FRM, Murphy RW. 2015. Historical and contemporary demography of leaf-toed geckos (Phyllodactylidae: *Phyllodactylus tuberculatus saxatilis*) in the Mexican dry forest. *Conservation Genetics* 16:419–429
- [73] Waples RS, Hindar K, Hard JJ. 2012. *Genetic risks associated with marine aquaculture*. U.S. Dept. Commer., NOAA Tech. Memo. NMFS-NWFSC-119. 149 pp. <https://repository.library.noaa.gov/view/noaa/4247>
- [74] Beerli P, Mashayekhi S, Sadeghi M, Khodaei M, Shaw K. 2019. Population genetic inference with MIGRATE. *Current Protocols in Bioinformatics* 68:e87
- [75] SRA Toolkit Development Team. 2020. *SRA Toolkit*. <https://github.com/ncbi/sra-tools>
- [76] Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, et al. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* 21:241
- [77] Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649
- [78] Aktaş C. 2025. *haplotypes: Manipulating DNA sequences and estimating unambiguous haplotype network with statistical parsimony*. <https://cran.r-project.org/web/packages/haplotypes/index.html>
- [79] Leigh JW, Bryant D. 2015. Popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* 6:1110–1116

- [80] Dawson N, Fischer J, Kuhn M, Pasotti A, mhugent, et al. 2025. QGIS: 3.44.0. Geneva, Switzerland: Zenodo. doi: [10.5281/zenodo.15705458](https://doi.org/10.5281/zenodo.15705458)
- [81] Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37:1530–1534
- [82] Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61:539–542
- [83] Jombart T. 2008. Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405
- [84] 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
- [85] Fu J, Wen L. 2023. Impacts of Quaternary glaciation, geological history and geography on animal species history in continental East Asia: a phylogeographic review. *Molecular Ecology* 32:4497–4514
- [86] Tang CQ, Matsui T, Ohashi H, Dong YF, Momohara A, et al. 2018. Identifying long-term stable refugia for relict plant species in East Asia. *Nature Communications* 9:4488
- [87] Zhao J, Wang J, Yang X. 2019. 中国东部 (东经 105° 以东) 第四纪冰川研究回顾、进展及展望 [Review, progress and prospect of the Quaternary glaciations in eastern China (east to 105° E)]. *冰川冻土 [Journal of Glaciology and Geocryology]* 41:75–92 (in Chinese)
- [88] Liu C, Wang J, Ko YZ, Shiao MS, Wang Y, et al. 2024. Genetic diversities in wild and cultivated populations of the two closely-related medical plants species, *Tripterygium wilfordii* and *T. hypoglaucom* (Celastraceae). *BMC Plant Biology* 24:195
- [89] Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, et al. 2006. Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences of the United States of America* 103:16666–16671
- [90] Willig MR, Kaufman DM, Stevens RD. 2003. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annual Review of Ecology, Evolution, and Systematics* 34:273–309
- [91] Dynesius M, Jansson R. 2000. Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences of the United States of America* 97:9115–9120
- [92] Sandel B, Arge L, Dalsgaard B, Davies RG, Gaston KJ, et al. 2011. The influence of late quaternary climate-change velocity on species endemism. *Science* 334:660–664
- [93] Gamba D, Muchhala N. 2020. Global patterns of population genetic differentiation in seed plants. *Molecular Ecology* 29:3413–3428
- [94] Wambulwa MC, Luo YH, Zhu GF, Milne R, Wachira FN, et al. 2022. Determinants of genetic structure in a highly heterogeneous landscape in southwest China. *Frontiers in Plant Science* 13:779989
- [95] Wambulwa MC, Zhu GF, Luo YH, Wu ZY, Provan J, et al. 2025. Incorporating genetic diversity to optimize the plant conservation network in the third pole. *Global Change Biology* 31:e70122
- [96] Qiu YX, Fu CX, Comes HP. 2011. Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. *Molecular Phylogenetics and Evolution* 59:225–244
- [97] Liu J, Möller M, Provan J, Gao LM, Poudel RC, et al. 2013. Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. *New Phytologist* 199:1093–1108
- [98] Abbott RJ, Smith LC, Milne RI, Crawford RMM, Wolff K, et al. 2000. Molecular analysis of plant migration and refugia in the Arctic. *Science* 289:1343–1346
- [99] Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405:907–913
- [100] Feng S, Bai M, Rivas-González I, Li C, Liu S, et al. 2022. Incomplete lineage sorting and phenotypic evolution in marsupials. *Cell* 185:1646–1660.e18
- [101] Pezzi PH, Wheeler LC, Freitas LB, Smith SD. 2024. Incomplete lineage sorting and hybridization underlie tree discordance in *Petunia* and related genera (Petunieae, Solanaceae). *Molecular Phylogenetics and Evolution* 198:108136
- [102] Tiley GP, Flouri T, Jiao X, Poelstra JW, Xu B, et al. 2023. Estimation of species divergence times in presence of cross-species gene flow. *Systematic Biology* 72:820–836
- [103] Soares AER, Schrago CG. 2012. The influence of taxon sampling and tree shape on molecular dating: an empirical example from mammalian mitochondrial genomes. *Bioinformatics and Biology Insights* 6:BBI.S9677
- [104] Yan P, Zhang L, Hao J, Sun G, Hu Z, et al. 2024. Construction of a core collection of Korean pine (*Pinus koraiensis*) clones based on morphological and physiological traits and genetic analysis. *Forests* 15:534
- [105] Ye L, Shavvon RS, Qi H, Wu H, Fan P, et al. 2024. Population genetic insights into the conservation of common walnut (*Juglans regia*) in Central Asia. *Plant Diversity* 46:600–610
- [106] Sun YW, Hou N, Woeste K, Zhang C, Yue M, et al. 2019. Population genetic structure and adaptive differentiation of iron walnut *Juglans regia* subsp. *sigillata* in southwestern China. *Ecology and Evolution* 9:14154–14166
- [107] Yuan XY, Sun YW, Bai XR, Dang M, Feng XJ, et al. 2018. Population structure, genetic diversity, and gene introgression of two closely related walnuts (*Juglans regia* and *J. sigillata*) in southwestern China revealed by EST-SSR markers. *Forests* 9:646
- [108] Wambulwa MC, Fan PZ, Milne R, Wu ZY, Luo YH, et al. 2022. Genetic analysis of walnut cultivars from southwest China: implications for germplasm improvement. *Plant Diversity* 44:530–541
- [109] Lenda M, Knops JH, Skórka P, Morón D, Woyciechowski M. 2018. Cascading effects of changes in land use on the invasion of the walnut *Juglans regia* in forest ecosystems. *Journal of Ecology* 106:671–686
- [110] Liu J, Milne RI, Zhu GF, Spicer RA, Wambulwa MC, et al. 2022. Name and scale matter: clarifying the geography of Tibetan Plateau and adjacent mountain regions. *Global and Planetary Change* 215:103893
- [111] Qi X, Cui C, Peng Y, Zhang X, Yang Z, et al. 2013. Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the Tibetan Plateau. *Molecular Biology and Evolution* 30:1761–1778
- [112] Liu L, Chen J, Wang J, Zhao Y, Chen X. 2022. Archaeological evidence for initial migration of Neolithic Proto Sino-Tibetan speakers from Yellow River valley to Tibetan Plateau. *Proceedings of the National Academy of Sciences of the United States of America* 119:e2212006119
- [113] Hsu CY, Baker TD, Duke MS. 2012. *China: A New Cultural History*. New York: Columbia University Press. 632 pp. www.jstor.org/stable/10.7312/hsu-15920
- [114] Shi S. 2018. Ethnic flows in the Tibetan-Yi corridor throughout history. *International Journal of Anthropology and Ethnology* 2:2
- [115] Fei X. 1982. 关于深入开展民族田野调查问题的探讨 [Debating the problem of carrying out in-depth field ethnographic surveys]. *中南民族大学学报 [Journal of South-Central University for Nationalities]* 3:2–6 (in Chinese)
- [116] Elias H. 2024. The Southwest Silk Road: artistic exchange and transmission in early China. *Bulletin of the School of Oriental and African Studies* 87:319–344
- [117] Anderson JA. 2009. China's southwestern silk road in world history. *World History Connected* 6:1–7



Copyright: © 2026 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.