Earth History and Biodiversity 10 (2026) 100048

Contents lists available at ScienceDirect

Earth History and Biodiversity

journal homepage: www.sciencedirect.com/journal/earth-history-and-biodiversity/

DNA barcode assessment in *Pterocarya* with historical biogeographic insights into the relict species *Pterocarya fraxinifolia*

Seyedeh Alemeh Sabbagh [a], Hamed Yousefzadeh [b,*,1], Mohammad Vatanparast [c], Mohammad Reza Bakhtiaryzadeh [d], Gregor Kozlowski [e,f,g], Yi- Gang Song [g,*,2]

[a] Tarbiat Modares University, Faculty of Natural Sciences, Department of Environmental Science, Tehran, Iran
[b] Tarbiat Modares University, Faculty of Natural Sciences, Department of Forestry, Tehran, Iran
[c] Royal Botanic Gardens Kew, Richmond, Surrey, UK
[d] Department of Animal and Poultry Science, College of Aburaihan, University of Tehran, Tehran, Iran
[e] Department of Biology and Botanic Garden, University of Fribourg, Chemin du Musée 10, Fribourg, CH-1700, Switzerland
[f] Natural History Museum Fribourg, Chemin du Musée 6, Fribourg, CH-1700, Switzerland
[g] Eastern China Conservation Centre for Wild Endangered Plant Resources, Shanghai Chenshan Botanical Garden, Shanghai, 201602, China

ARTICLE INFO

Keywords:
DNA barcoding
Pterocarya fraxinifolia
MatK
RbcL
Historical biogeography
Fossil record
Juglandaceae

ABSTRACT

The genus *Pterocarya* (Juglandaceae) is a relict tree lineage with a disjunct distribution across East Asia and the Hyrcanian forests of West Asia. We compared the two core chloroplast barcodes, *matK* and *rbcL*, across six reference taxa and then tested candidate diagnostic states for *P. fraxinifolia* in 100 individuals from 20 Iranian populations. The same sequence data were also used for complementary phylogenetic, divergence-time, and biogeographic analyses. *matK* recovered four reference haplotypes and unique sequences for two of the six taxa; *rbcL* recovered three haplotypes and a unique sequence for one taxon, and combining the two loci did not improve on *matK*. Within *P. fraxinifolia*, we detected six *matK* haplotypes and two *rbcL* haplotypes. Four of the five states that appeared diagnostic in the reference comparison were polymorphic in the population sample, whereas one *matK* state remained fixed across all 100 individuals. The marker-specific trees and the concatenated maximum-likelihood analysis resolved several terminal East Asian relationships, but the deeper placement of *P. fraxinifolia* remained weakly supported. Fossil-constrained BEAST analysis placed crown *Pterocarya* at 30.81 Ma (95% HPD: 28.00–33.64 Ma). The two biogeographic models differed at the deeper nodes: DEC reconstructed a widespread West-East Asian crown range (AB), whereas the BBM graphic retained a mixed A/B crown state; B nevertheless had the highest numerical probability (96.70%) in the BBM output. Both models inferred one dispersal and one vicariance overall, although they placed these events at different nodes, and neither inferred extinction under the extant-only two-area framework. Overall, the barcode results show limited species-level resolution and demonstrate why putatively diagnostic states identified from single reference sequences require population-level validation. The biogeographic results likewise represent a simplified history of the surviving lineages and are best interpreted alongside the much richer fossil record of *Pterocarya*.

Introduction

The genus *Pterocarya* (Juglandaceae) comprises approximately eight recognised species of deciduous trees characterised by winged fruits and a disjunct distribution across East Asia and the Caucasus region (Lu, 1982; Song et al., 2020). *Pterocarya fraxinifolia* occurs in the Hyrcanian forests of northern Iran and the Caucasus Mountains, whereas the remaining species are concentrated mainly in East Asia, particularly central and southwestern China. Based on terminal-bud morphology, the genus has traditionally been divided into two sections: *Pterocarya*, with naked buds, and *Platyptera*, with scaled buds (Lu, 1982; Manning, 1978).

Despite the taxonomic importance of the genus, species delimitation within *Pterocarya* remains difficult because diagnostic morphological

* Corresponding authors.
E-mail addresses: h.yousefzadeh@modares.ac.ir (H. Yousefzadeh), cherish-faith@163.com (Y.G. Song).
[1] https://orcid.org/0000-0003-4326-4762
[2] https://orcid.org/0000-0003-2584-2338

https://doi.org/10.1016/j.hisbio.2026.100048
Received 27 June 2026; Received in revised form 22 August 2026; Accepted 15 September 2026
Available online 19 September 2026
2950-4759/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

traits overlap and its evolutionary history is complex (Song et al., 2020; Zhou et al., 2021). *The genus* is also biogeographically notable because fossils document a much broader Cenozoic distribution across the Northern Hemisphere than is represented by the modern East Asia-Caucasus disjunction (Manchester, 1987; Song et al., 2020; Yan et al., 2024). At the family level, fossil-informed analyses that integrate molecular, morphological, palaeoenvironmental, and extinct-lineage evidence support a Boreotropical origin of Juglandaceae and point to major roles for Europe and intercontinental land connections in historical dispersal and extinction (Zhang et al., 2021). Within *Pterocarya*, Yan et al. (2024) compiled 83 fossil records and showed that including extinct taxa greatly increased the inferred numbers of dispersal, vicariance, and local-extinction events compared with reconstructions based only on living species. The present disjunction therefore cannot be understood from extant distributions alone; fossil evidence is essential for evaluating historical range expansion and contraction. Cenozoic cooling and progressive aridification in Central Asia probably contributed to regional extinction and fragmentation of formerly continuous temperate forests, while intercontinental exchange was facilitated at different times by northern land connections and changing Eurasian seaways (Tiffney and Manchester, 2001; Zhang et al., 2021; Yan et al., 2024).

Earlier studies provide the broader phylogenetic context for our marker-level analysis. Xiang et al. (2014), using five chloroplast markers, recovered three main lineages, including a distinct *P. fraxinifolia* lineage. Song et al. (2020) subsequently obtained a much more resolved genus-wide phylogeny from RAD-seq data, supporting the two traditional sections and providing a fossil-calibrated biogeographic framework. More recently, Wang et al. (2026) showed that chloroplast topologies in *Pterocarya* can vary with the plastome partition analysed and cautioned against relying on chloroplast evidence alone for species identification or biogeographic inference. The question addressed here is therefore narrower: do *matK* and *rbcL* contain diagnostic sequence variation in the reference set, particularly for *P. fraxinifolia*, and how do those marker-level patterns compare with the broader genomic evidence already available?

Where morphology alone does not reliably separate closely related taxa, DNA barcoding can provide an additional line of evidence for species discrimination and taxonomic assessment (CBOL Plant Working Group, 2009). Chloroplast sequence data can also place such barcode variation in a broader evolutionary context by informing genetic divergence, phylogenetic relationships, and historical biogeography. *Pterocarya* is especially suitable for this combined perspective because it couples taxonomic difficulty with a distinctive biogeographic history and includes the relict *P. fraxinifolia lineage* in the Hyrcanian forests.

The chloroplast loci *rbcL* and *matK* are widely used as core plant DNA barcodes (CBOL Plant Working Group, 2009). *rbcL* is relatively conserved and readily aligned, making it useful across broader taxonomic scales, whereas *th*e faster-evolving *matK* often provides more variation among closely related taxa (CBOL Plant Working Group, 2009; Hollingsworth, 2011). Their status as universal core barcodes, however, does not guarantee adequate resolution within a genus; discriminatory performance still needs to be tested empirically and with sufficient intraspecific sampling.

Previous work in Juglandaceae illustrates this limitation. In an analysis of 196 samples representing nine genera and 54 species, *rbcL* and *matK* showed high amplification and sequencing success, while *matK* discriminated seven of the nine genera examined. Although ITS was relatively variable, its discriminatory power was limited within *Juglans* and *Pterocarya*, and *matK* was proposed as a useful first step in a tiered barcoding strategy (Xiang et al., 2011). Liu et al. (2025) later found that identification success in Juglandaceae was highest with whole chloroplast genomes, followed by taxon-specific barcodes, while universal DNA barcodes performed less well. Together, these results argue for treating *rbcL* and *matK* as baseline markers whose performance must be demonstrated rather than assumed in *Pterocarya*.

Accordingly, our primary objective was to determine whether *rbcL* and *matK* contain diagnostic sequence characters among the *Pterocarya* reference sequences included here, with particular emphasis on *P. fraxinifolia*. As a secondary objective, we used the same chloroplast data to examine phylogenetic relationships and reconstruct the historical biogeographic pattern of the genus, including the position of the Hyrcanian lineage. Because multiple individuals were not available for every species, we treat the barcode component as a reference-sequence assessment rather than a definitive validation of species-level identification. Broader intraspecific sampling, together with plastome-scale and taxon-specific marker comparisons, will be needed to quantify barcode reliability across *Pterocarya*.

Materials and methods

DNA extraction, plastome sequencing, assembly, and reference dataset construction

Fresh leaves from the Sari-3 individual of *Pterocarya fraxinifolia* were used to generate the focal plastome reference. Total genomic DNA was extracted using a modified cetyltrimethylammonium bromide (CTAB) protocol. DNA quantity and integrity were checked before library preparation with a Qubit 2.0 fluorometer and an Agilent 2100 Bioanalyzer. Whole-genome sequencing was performed by Wuhan Benagen Tech Solutions Company Limited (Wuhan, China) on the DNBSEQ platform using 150-bp paired-end reads. Read quality was assessed with FastQC v0.12.1 before and after filtering. Trimmomatic v0.39 was used to remove adapters and low-quality bases with a four-base sliding window and a minimum mean Phred score of 15; reads shorter than 50 bp after trimming were discarded.

Quality-filtered reads were assembled de novo with GetOrganelle v1.7.7.0 using the embryophyte plastid database (embplant_pt) and a maximum of 15 extension rounds (-R 15). No *Pterocarya* plastome was imposed as an assembly template. To verify read support across the assembly, filtered reads were mapped back to the assembled plastome with Bowtie2 v2.5.4. Annotation was performed with GeSeq v2.03. The resulting *P. fraxinifolia* plastome was 160,086 bp long and was deposited in GenBank under accession PV791734.1.

For the interspecific reference dataset, complete plastomes of five East Asian *Pterocarya* taxa were retrieved from GenBank: *P. hupehensis* (NC_046431), *P. stenoptera* (NC_046428), *P. rhoifolia* (OR099227), *P. tonkinensis* (NC_046427), and *P. macroptera var. delavayi* (MW194257). Sampling and voucher information, where available, together with sequence source and accession numbers, are summarised in Supplementary Table S1. The *matK* and *rbcL* regions were extracted from these five plastomes and from the newly generated *P. fraxinifolia* plastome. *Juglans regia* was included as the outgroup for phylogenetic reconstruction and divergence-time estimation. The loci were aligned separately in MEGA and checked for sequence orientation and homologous boundaries. The final alignments comprised 1529 bp for *matK* and 1449 bp for *rbcL;* a 2,978-bp concatenated *matK* + *rbcL* matrix was then constructed for the combined-marker analyses.

Plastome-wide comparisons of genome structure, IR boundaries, repeats, nucleotide diversity, and highly variable regions across Hyrcanian and East Asian *Pterocarya* were reported previously by Sabbagh et al. (2025) and were therefore not repeated here. The present study instead focuses on the individual and combined barcode loci, a model-based phylogenetic check, divergence-time estimation, and historical biogeographic reconstruction.

Population-level validation and DNA barcode assessment

To test the intraspecific stability of the reference-level diagnostic states, we analysed 100 *P. fraxinifolia* individuals sampled from 20 natural populations in Iran, with five individuals from each population. Nineteen populations represented the west-east Hyrcanian forest belt, whereas the Lorestan population was geographically isolated in the

Zagros Mountains. Trees were sampled at least 50 m apart whenever stand structure allowed. Young leaves were dried in silica gel, DNA was extracted using the CTAB procedure described above, and chloroplast sequences were recovered from the whole-genome data with the same quality-control and plastome-recovery workflow. Only *matK* and *rbcL* sequences from these 100 individuals were used in the intraspecific validation reported here.

Reference-sequence variation was examined separately for *matK*, *rbcL*, and the concatenated *matK* + *rbcL* matrix. DnaSP (Rozas et al., 2003) was used to identify haplotypes and variable nucleotide positions in each alignment. A nucleotide state was considered a candidate diagnostic state when, at a given alignment position, it occurred in only one *Pterocarya* reference taxon. Taxa with identical complete marker sequences were considered unresolved by that marker. Discriminatory performance was therefore described by the number of haplotypes and the number of taxa represented by unique sequences for each locus and for the concatenated matrix. Bootstrap support from the phylogenetic analyses was not used as a measure of barcode discrimination.

To visualise the reference haplotype structure in Fig. 2, we also evaluated combinations of variable positions in R and identified the smallest SNP sets that reproduced the complete-marker haplotype partition among the six *Pterocarya* reference taxa. One representative minimum key was selected for each locus: positions 314, 494, and 924 for *matK*, and positions 1098 and 639 for *rbcL*. This minimum-key analysis was used only for the reference-level figure and was kept separate from the *P. fraxinifolia*-specific candidate diagnostic states tested in the 100-individual population sample.

The candidate *P. fraxinifolia* states were then examined at homologous positions in all 100 population-level sequences and scored as fixed or polymorphic. Haplotype number and the number of variable sites were also calculated separately for the population-level *matK* and *rbcL* datasets. Comparable replicate sampling was not available for the five East Asian taxa. This analysis therefore tests intraspecific stability only in *P. fraxinifolia* and should not be interpreted as genus-wide population-level validation of the barcodes.

Phylogenetic reconstruction and model-based analysis

Separate marker trees were reconstructed from the *matK* and *rbcL* alignments using maximum likelihood (ML) in MEGA7 (Kumar et al., 2016) under the best-fitting T92 model identified by BIC. Node support was assessed with 1000 bootstrap replicates, and *Juglans regia* was used to root both trees. These trees provide marker-specific phylogenetic context for the barcode results and are shown in Fig. 1.

We also performed a model-based analysis of the concatenated dataset. In MEGA6, 24 nucleotide-substitution models were compared using the Bayesian information criterion (BIC). T92 had the lowest BIC for *matK* (4478.232), *rbcL* (4218.936), and the concatenated dataset (8633.386). The 2,978-bp concatenated alignment was analysed by maximum likelihood under T92 with uniform rates, complete deletion of sites containing gaps or missing data (2974 positions retained), 1000 bootstrap replicates, and nearest-neighbour-interchange (NNI) branch swapping. This tree was used as an independent model-based check of the main relationships, not as a measure of barcode discrimination.

Divergence time estimation

Divergence times were estimated from the final seven-taxon dataset in BEAST v1.7.5, with XML files prepared in BEAUti v1.7.5 (Drummond and Rambaut, 2007; Drummond et al., 2012). The *matK* and *rbcL* alignments were retained as separate partitions but shared a common tree. Because T92 is not available directly in BEAUti v1.7.5, each partition was assigned the more general HKY model with empirical base frequencies. We did not specify gamma-distributed rate heterogeneity or a separate class of invariant sites. Each partition used an uncorrelated lognormal relaxed clock, relative rates were estimated across partitions, and a Yule process was used as the tree prior.

Crown *Pterocarya* was calibrated directly from the fossil record. Following the dating framework of Lu et al. (2024), in which the earliest *Pterocarya* fossil constrains the crown group, we assigned a Uniform 28–34 Ma prior to the crown node. The six *Pterocarya* taxa were constrained to be monophyletic, with *Juglans regia* retained as the outgroup. MCMC sampling was initialised from a time-scaled, user-specified starting tree compatible with the fossil bounds. The topology of this starting tree was not fixed; apart from *Pterocarya* monophyly, no topological constraint was imposed.

Two independent MCMC runs were conducted for 100,000,000 generations each, with parameters and trees sampled every 10,000 generations. The first 10% of each run was discarded separately as burn-in. Convergence and effective sample sizes were assessed in Tracer, and all key parameters had ESS values above 200. The retained samples from the two runs were then combined, and TreeAnnotator v1.7.5 was used to summarise a maximum-clade-credibility (MCC) tree with mean node heights. This fossil-calibrated MCC tree, rather than either marker-specific ML tree or the concatenated ML tree, provided the dated topology used in the biogeographic analyses.

Historical biogeographic reconstruction

Historical biogeographic analyses were performed in RASP v4.4

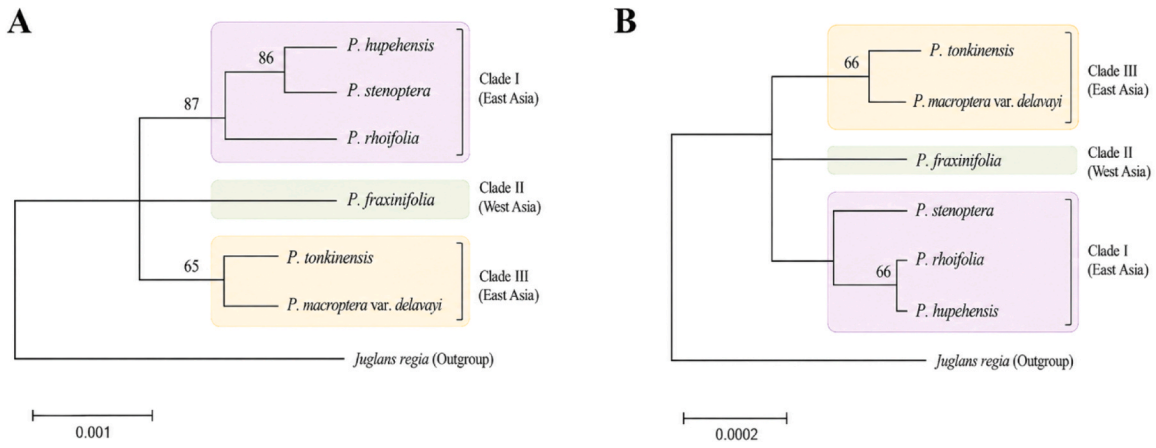


Fig. 1. Phylogenetic relationships among *Pterocarya* species inferred from maximum-likelihood (ML) analyses of the chloroplast markers **(A)** *matK* and **(B)** *rbcL*. Bootstrap support values > =50% are shown at the nodes. *Juglans regia* was used as the outgroup. Clade I represents an East Asian lineage, Clade II the West Asian lineage, and Clade III a second East Asian lineage. The scale bar indicates substitutions per site.

(build 20251114; Yu et al., 2020). We defined two geographic areas: A, West Asia, representing the Hyrcanian/Caucasus range of *P. fraxinifolia*; and B, East Asia, representing the five sampled East Asian *Pterocarya* taxa. *Juglans regia* was retained for rooting and dating but removed from the MCC tree before the RASP analyses. This restricted the geographic coding to *Pterocarya* and avoided assigning an artificial A/B state to the outgroup. The resulting six-taxon tree retained branch lengths in millions of years.

DEC was run in RASP through the BioGeoBEARS v1.1.1 module. The maximum range size was set to two, allowing A, B, and AB as ancestral states. A single time interval from 0 to 30.807 Ma was used. Dispersal multipliers were symmetric and fixed at 1.0 in both directions between A and B, and no additional geographic fossil or MRCA constraints were imposed in RASP.

For BBM, the maximum number of areas was likewise set to two and null distributions were not allowed. The analysis used 50,000 cycles, 10 chains, a sampling frequency of 100 cycles, a burn-in of 100 sampled states, and a chain temperature of 0.1. State frequencies were fixed under the JC option and among-site rates were set to equal. Two independent BBM runs produced highly similar results (run-to-run distance = 0.0049), and their combined reconstruction was used for the reported node probabilities. DEC and BBM were treated as alternative reconstructions rather than combined into a single history. Because the analysis included only two areas and a single extant West Asian species, all ancestral-range probabilities were interpreted within this simplified extant-taxon framework.

The three tree-based outputs therefore have distinct roles. Fig. 1 presents the marker-specific ML trees, the concatenated ML analysis provides an independent model-based check, and the RASP analyses use the fossil-calibrated BEAST MCC tree. Differences among these topologies reflect the different analytical frameworks rather than repeated use of a single tree.

Results

Plastome assembly and reference dataset

The Sari-3 assembly yielded a 160,086-bp *P. fraxinifolia* plastome, deposited in GenBank as PV791734.1. After homologous loci were extracted from this plastome and the five East Asian reference plastomes, the six-taxon ingroup alignments comprised 1529 bp for *matK* and 1449 bp for *rbcL*. The concatenated *matK* + *rbcL* matrix was 2978 bp long. Inclusion of *Juglans regia* as the outgroup produced the seven-taxon datasets used for phylogenetic reconstruction and divergence-time estimation.

Population-level validation and DNA barcode assessment

The two markers differed in the amount of reference-level variation they resolved. For *matK*, the six taxa formed four haplotypes: *P. fraxinifolia* and *P. rhoifolia* each had a unique sequence, *P. hupehensis* shared a haplotype with *P. stenoptera*, and *P. tonkinensis* shared a haplotype with *P. macroptera var. delavayi*. *rbcL* produced three haplotypes. Only *P. fraxinifolia* had a unique sequence; *P. rhoifolia*, *P. hupehensis*, and *P. stenoptera* shared one haplotype, whereas *P. tonkinensis* and *P. macroptera var. delavayi* shared another. The concatenated *matK* + *rbcL* matrix also produced four haplotypes and two uniquely represented taxa, so combining the loci did not improve on *matK*. Thus, unique reference sequences were recovered for two of six taxa with *matK*, one of six with *rbcL*, and two of six with the combined matrix.

Across the six-taxon reference alignments, *matK* contained 11 variable positions and *rbcL* contained three. Three *matK* sites were sufficient to reproduce the four complete-marker haplotypes; the representative key shown in Fig. 2 uses positions 314, 494, and 924. For *rbcL*, two sites were sufficient to reproduce the three complete-marker haplotypes, and Fig. 2 uses positions 1098 and 639. These minimum discrimination keys summarise the reference haplotype structure and are separate from the *P. fraxinifolia*-specific candidate diagnostic states evaluated across the 100 individuals.

Population-level sampling revealed substantially more variation within *P. fraxinifolia* than was apparent from the single reference sequence. Across the 100 individuals, *matK* contained six haplotypes defined by nine variable sites, whereas *rbcL* contained two haplotypes defined by one variable site. Four *matK* states appeared diagnostic for *P. fraxinifolia* in the six-taxon reference comparison: 208 (C), 244 (A), 265 (C), and 314 (C). At positions 208, 244, and 265, however, the reference state was polymorphic in the population sample, with a 77:23 split between the reference and alternative states at each position. Only position 314 remained fixed, with C present in all 100 individuals. The candidate *rbcL* state at position 1098 was also polymorphic, with 55 individuals carrying T and 45 carrying C. Thus, four of the five candidate diagnostic states identified from the reference comparison were not fixed within *P. fraxinifolia*; only the *matK* state at position 314 was invariant across the full population sample.

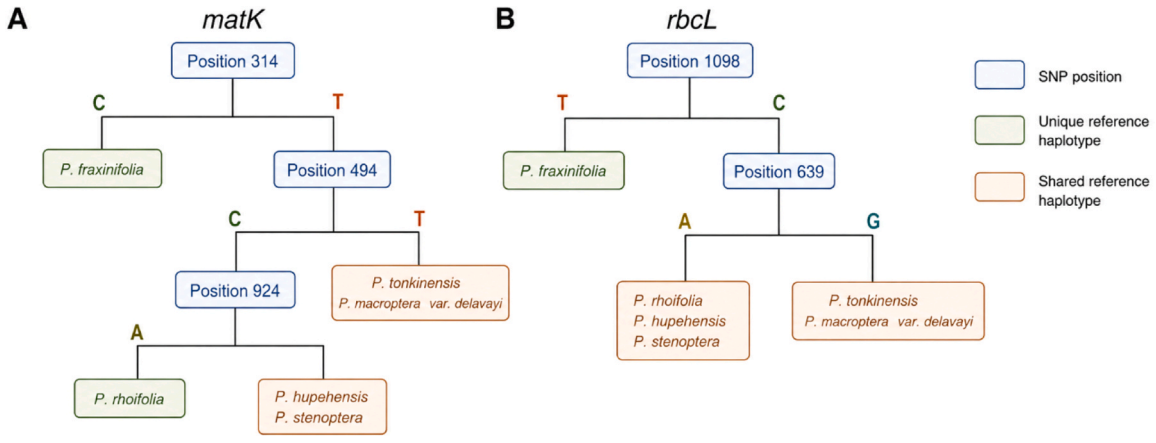


Fig. 2. SNP-based discrimination among *Pterocarya* reference sequences using the chloroplast DNA barcode regions **(A)** *matK* and **(B)** *rbcL*. Each panel shows one representative minimum SNP key sufficient to reproduce the complete-marker haplotype partition (*matK*: positions 314, 494, and 924; *rbcL*: positions 1098 and 639). Blue boxes mark SNP positions, nucleotide states are shown on branches, green terminal boxes denote unique reference haplotypes, and beige terminal boxes denote shared reference haplotypes. These reference-level keys are distinct from the *P. fraxinifolia*-specific candidate diagnostic states evaluated in the 100-individual validation.

Phylogenetic reconstruction and model-based analysis

The marker-specific ML trees recovered several East Asian relationships, although their details differed (Fig. 1). In the *matK* tree, *P. hupehensis* and *P. stenoptera* formed a pair with BS = 86, and *P. rhoifolia* joined this pair at BS = 87. *P. tonkinensis* and *P. macroptera var. delavayi* formed a second East Asian pair (BS = 65), while *P. fraxinifolia* remained a separate West Asian lineage. In the *rbcL* tree, *P. tonkinensis* + *P. macroptera var. delavayi* was supported at BS = 66, and *P. rhoifolia* + *P. hupehensis* formed a pair at BS = 66; *P. stenoptera* joined the latter lineage without bootstrap support > =50%. In both marker trees, bootstrap support for the deeper relationships among the three broad lineages remained below 50%.

T92 had the lowest BIC for all three datasets: 4478.232 for *matK*, 4218.936 for *rbcL*, and 8633.386 for the concatenated alignment. After complete deletion, 2974 positions remained in the concatenated ML analysis. The resulting tree recovered *P. tonkinensis* + *P. macroptera var. delavayi* with BS = 85, *P. hupehensis* + *P. stenoptera* with BS = 86, and *P. rhoifolia* + (*P. hupehensis* + *P. stenoptera*) with BS = 95. Bootstrap support for the deeper backbone involving *P. fraxinifolia* remained below 50%. The concatenated analysis therefore strengthened support for several terminal East Asian relationships without resolving the deeper placement of *P. fraxinifolia*.

Divergence time estimation

The two fossil-calibrated BEAST runs produced closely matching posterior results, and all key parameters had ESS values above 200 after burn-in. In the combined posterior, crown *Pterocarya* (N11) was dated to 30.807 Ma (95% HPD: 28.000–33.638 Ma). Because this node was directly constrained by the 28–34 Ma fossil prior, we regard the estimate as fossil-anchored rather than as an independent molecular date. The uncalibrated *Juglans-Pterocarya* root had a mean age of 51.868 Ma, but a much wider 95% HPD (29.392–80.534 Ma).

Within the *Pterocarya* MCC tree, N7 (*P. macroptera var. delavayi* + *P. tonkinensis*) was dated to 3.424 Ma (95% HPD: <0.001–10.151 Ma), N8 (*P. fraxinifolia* + N7) to 24.659 Ma (95% HPD: 15.147–33.043 Ma), N9 (*P. hupehensis* + *P. stenoptera*) to 2.280 Ma (95% HPD: <0.001–6.765 Ma), and N10 (*P. rhoifolia* + N9) to 12.893 Ma (95% HPD: 4.200–21.941 Ma) (Table 1). Posterior support was low for N8 (PP = 0.434), whereas N7, N9, N10, and the crown node N11 each had PP = 1.00 in the combined tree sample. The value of approximately 0.43 therefore refers to support for the N8 topology and not to an ancestral-area probability.

Historical biogeographic reconstruction

DEC reconstructed area B at N7, N9, and N10, with probability 1.00 at each node, and AB at both N8 and the crown node N11, again with probability 1.00 (Fig. 3B; Table 1). At N8, the inferred event was vicariance along the route AB -> A|B. At N11, DEC instead placed *a* dispersal event along AB -> AB^B -> B|AB. Both event routes had probability 1.0000. No extinction event was inferred, giving global DEC totals of one dispersal, one vicariance, and zero extinctions (Table 2).

The two BBM runs produced highly similar reconstructions (run-to-run distance = 0.0049), and the combined result is shown in Fig. 3A. In the final RASP graphic, N7, N9, and N10 are represented by area B, whereas N8 and N11 are displayed as mixed A/B states (Fig. 3A). The numerical BBM output nevertheless assigned the highest probability to B at N8 (85.07%) and N11 (96.70%) (Table 1), indicating that these deeper nodes retain alternative-range signal despite B being the modal state. N8 was the only node with non-zero biogeographic events: BBM inferred one dispersal and one vicariance along B -> AB -> A|B, with an event-route probability of 0.8247. No extinction event was inferred. Global BBM totals were therefore one dispersal, one vicariance, and zero extinctions, matching the DEC totals but not their placement on the tree. The two models agreed on B at N7, N9, and N10 but differed at the deeper nodes N8 and N11: DEC reconstructed AB, whereas the BBM graphic retained mixed A/B states, with B having the highest numerical probability at both nodes. These probabilities are conditional on the simplified two-area reconstruction based on extant taxa.

Discussion

DNA barcoding and phylogenetic context

Both marker-specific trees recovered three broad lineages, but relationships within them differed from the better-resolved *Pterocarya* phylogenies based on larger datasets. Xiang et al. (2014), using five chloroplast markers, likewise recovered *P. fraxinifolia* as a distinct lineage and placed *P. stenoptera*, *P. tonkinensis*, and forms of *P. macroptera* within another major clade. With broader taxon sampling and genome-wide RAD-seq data, Song et al. (2020) obtained a well-resolved phylogeny that supported the two traditional sections and differed from our marker trees in the positions of *P. rhoifolia* and *P. tonkinensis*. Wang et al. (2026) further showed that *Pterocarya* plastome topologies can shift substantially depending on whether analyses use whole plastomes, coding regions, hypervariable regions, or sliding-window datasets. In our concatenated ML tree, support was strong for *P. tonkinensis* + *P. macroptera var. delavayi* (BS = 85), *P. hupehensis* + *P. stenoptera* (BS = 86), and *P. rhoifolia* + (*P. hupehensis* + *P. stenoptera*) (BS = 95), but bootstrap support for the deeper placement of *P. fraxinifolia* remained below 50%. This pattern is consistent with the limited phylogenetic signal carried by the two barcode loci, the small reference set, and the particular history represented by chloroplast DNA. We therefore use the *matK*, *rbcL*, and concatenated ML trees as context for the barcode analysis rather than as alternatives to the better-resolved RAD-seq and broader plastome phylogenies.

The population-level analysis is central to interpreting the barcode results. Among the six reference taxa, *matK* produced four haplotypes and two unique sequences, whereas *rbcL* produced three haplotypes and only one unique sequence, that of *P. fraxinifolia*. Combining the loci did not improve on *matK*, and the sets of taxa left unresolved differed between markers. More importantly, the 100-individual *P. fraxinifolia*

Table 1
Fossil-calibrated divergence-time estimates and ancestral-range reconstructions for the revised *Pterocarya* topology. The *Juglans-Pterocarya* root is reported for temporal context but was excluded from RASP after outgroup pruning. For N8, the age interval is conditional on the clade being present in the posterior tree sample (BEAST PP = 0.434). The ancestral-range probabilities reported here are distinct from the event-route probabilities reported in Table 2.

Node	Clade	Mean age (Ma)	95% HPD (Ma)	BBM ancestral range	BBM range probability	DEC ancestral range (probability)
Root	*Juglans* + *Pterocarya*	51.868	29.392–80.534	-	-	-
N11	Crown *Pterocarya*	30.807	28.000–33.638	B	0.9670	AB (1.0000)
N8	*P. fraxinifolia* + N7	24.659	15.147–33.043	B	0.8507	AB (1.0000)
N10	*P. rhoifolia* + N9	12.893	4.200–21.941	B	0.9950	B (1.0000)
N7	*P. macroptera var. delavayi* + *P. tonkinensis*	3.424	< 0.001–10.151	B	0.9825	B (1.0000)
N9	*P. hupehensis* + *P. stenoptera*	2.280	< 0.001–6.765	B	0.9986	B (1.0000)

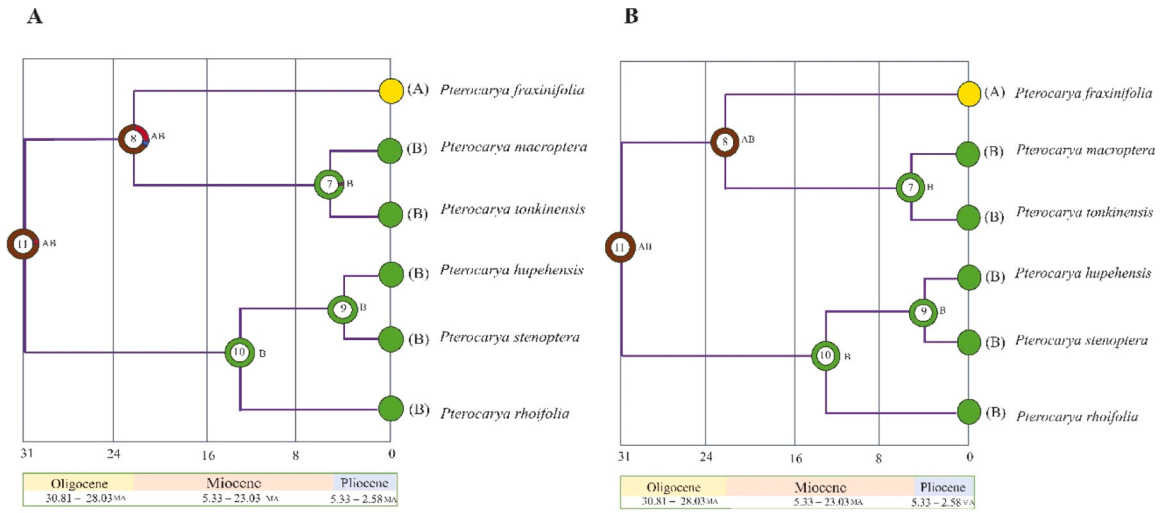


Fig. 3. Time-calibrated biogeographic reconstruction of *Pterocarya* lineages using **(A)** Bayesian Binary MCMC (BBM) and **(B)** dispersal-extinction-cladogenesis (DEC) in RASP 4.4; DEC was implemented through the BioGeoBEARS module. A = West Asia, B = East Asia, and AB = both areas. Node states show inferred ancestral ranges, while event symbols mark dispersal and vicariance: BBM places both events at N8, whereas DEC places vicariance at N8 and dispersal at N11; neither analysis infers extinction. The RASP reconstruction is restricted to the six *Pterocarya* taxa after pruning *Juglans regia*. The time axis spans approximately 32 Ma to the present, from the Oligocene to the Quaternary.

Table 2
Non-zero biogeographic events inferred from the revised BBM and DEC reconstructions. Event-route probability is reported separately from the ancestral-state probability in Table 1. Both methods yielded global totals of one dispersal, one vicariance, and zero extinctions.

Method	Node	Ancestral range	Event route	Dispersal	Vicariance	Extinction	Event-route probability
BBM	N8	B	B -> AB -> A|B	1	1	0	0.8247
DEC	N8	AB	AB -> A|B	0	1	0	1.0000
DEC	N11	AB	AB -> AB^B -> B|AB	1	0	0	1.0000

dataset showed that four of the five states that appeared diagnostic in the reference comparison were actually polymorphic; only one *matK* state remained fixed throughout the sample. A nucleotide state observed in a single reference sequence therefore cannot be assumed to represent a stable species-level barcode character. This result is consistent with the broader pattern in Juglandaceae, where whole plastomes, taxon-specific barcodes, and read-based approaches generally achieve higher identification success than universal plastid loci (Liu et al., 2025). The fixed *matK* state may still prove useful for *P. fraxinifolia*, but its specificity cannot be established until comparable population-level sampling is available for the East Asian congeners. On the present evidence, *matK* and *rbcL* are better regarded as limited reference-level markers than as validated stand-alone species barcodes for *Pterocarya*.

Biogeography

The fossil-calibrated analysis dated crown *Pterocarya* to 30.807 Ma (95% HPD: 28.000–33.638 Ma), in the early Oligocene. Because the crown node was directly constrained by the 28–34 Ma fossil prior, this is not an independent molecular estimate and is better treated as fossil-anchored. The ancestral range of the crown, however, was model-dependent. The BBM graphic retained a mixed A/B state at the crown, although B had the highest numerical probability (96.70%), whereas DEC reconstructed a widespread West-East Asian range (AB; 100%).

Our data therefore do not distinguish decisively between an ancestor restricted to East Asia and one already distributed across both regions. Even so, East Asia remains prominent in both reconstructions because the principal internal East Asian lineages were assigned to B by both models, consistent with its recognised role as a major persistence centre for Northern Hemisphere relict woody floras (Milne and Abbott, 2002; Tang et al., 2018).

The fossil record also shows why this extant-only reconstruction requires caution. Fossil-informed work on Juglandaceae supports a Boreotropical origin and highlights the importance of Europe, intercontinental land bridges, range contraction, and extinction in the family's history (Zhang et al., 2021). For *Pterocarya*, Yan et al. (2024) combined plastome and morphological data from living and extinct taxa and compiled 83 fossil records. Their tip-dating analysis placed the most recent common ancestor in East Asia at 40.46 Ma and, when fossils were included, inferred six dispersal, six vicariance, and 11 local-extinction events. By contrast, their extant-only reconstruction recovered only one dispersal and one vicariance event. They also inferred exchange between East Asia and North America through the Bering Land Bridge and between East Asia and Europe after closure of the Turgai Strait. This contrast is directly relevant to our results. Neither BBM nor DEC inferred extinction in our analysis, but that result reflects the information available to a two-area model containing only six surviving *Pterocarya* lineages; it should not be read as evidence that extinction was historically unimportant.

In the dated MCC tree, N8 joins *P. fraxinifolia* with the lineage containing *P. macroptera var. delavayi* and *P. tonkinensis*. Its mean age was 24.659 Ma (95% HPD: 15.147–33.043 Ma), but the clade itself was weakly supported (PP = 0.434). Both the date and the biogeographic interpretation of this split therefore depend on an uncertain deep relationship. The broad HPD spans a long interval of Cenozoic cooling, increasing continentality, and environmental reorganisation across Eurasia. These changes may have contributed to the fragmentation of temperate forest connections, but the present data cannot associate the split with any single climatic or tectonic event. The persistence of *P. fraxinifolia* in the Hyrcanian region is consistent with the well-known refugial role of these forests for Tertiary relict taxa (Akhani et al., 2010), although our reconstruction does not establish when or by what route the western lineage became isolated.

BBM and DEC yielded the same global totals-one dispersal, one

vicariance, and no extinction-but placed those events differently on the tree. Both models strongly supported area B at N7, N9, and N10. At N8 and N11, however, DEC reconstructed AB with probability 1.00, whereas the BBM graphic retained mixed A/B states; B was nevertheless the highest-probability numerical state at N8 (85.07%) and N11 (96.70%). Their event histories also differed: BBM placed both dispersal and vicariance at N8 along B -> AB -> A|B, whereas DEC placed vicariance at N8 (AB -> A|B) and dispersal at N11. This contrast is not surprising because the two approaches represent geographic change differently. DEC explicitly models range evolution along branches and range inheritance at cladogenesis, whereas BBM estimates ancestral geographic states within a Bayesian binary-state framework. In our dataset, the difference is also likely influenced by the strongly unbalanced tip coding-five East Asian taxa and one West Asian taxon-the use of only two geographic areas, and the weak posterior support for N8. We therefore interpret the disagreement as model uncertainty rather than evidence that either reconstruction is uniquely correct.

Taken together, the two analyses do not provide an unambiguous reconstruction of the deepest ancestral range of *Pterocarya*. DEC is consistent with subdivision of a widespread West-East Asian ancestor, whereas BBM retains a predominantly East Asian signal but shows mixed A/B states at the deeper nodes, consistent with uncertainty over whether westward range expansion preceded geographic separation. The more stable signal across models is the predominance of area B within the internal East Asian lineages and the historical differentiation of the lineage leading to P. fraxinifolia, rather than an unambiguous reconstruction of the deepest ancestral range. Fossil-based studies place the modern East Asia-Hyrcanian disjunction in a much broader context, suggesting that it represents the surviving remnant of a more extensive Northern Hemisphere history involving dispersal, vicariance, and regional extinction (Zhang et al., 2021; Yan et al., 2024). The absence of extinction in our RASP output should therefore be viewed as a consequence of the extant-only, two-area state space rather than as evidence that extinction was unimportant in *Pterocarya* evolution. In this respect, our reconstruction complements rather than replaces fossil-based accounts of the genus.

Conclusion

The two chloroplast barcodes provided only limited discrimination in *Pterocarya*, and their apparent performance depended strongly on sampling scale. In the six-taxon reference set, *matK* resolved more haplotypes than *rbcL*, while combining the loci added no resolution beyond *matK*. Population-level sampling changed the interpretation of the candidate diagnostic states: among 100 *P. fraxinifolia* individuals, four of five were polymorphic and only one *matK* state remained fixed. Neither marker, alone or combined, can therefore be regarded as a validated genus-wide barcode without comparable intraspecific sampling of the East Asian taxa. The phylogenetic analyses recovered several well-supported terminal East Asian relationships, but the deeper placement of *P. fraxinifolia* remained uncertain. Crown *Pterocarya* was dated to 30.81 Ma under the fossil calibration, and the inferred ancestral crown range depended on the biogeographic model: DEC favoured AB, whereas the BBM graphic retained a mixed A/B crown state, with B having the highest numerical probability. The extant-only analysis therefore does not resolve a single geographic origin for the crown. Considered together with fossil-inclusive studies, the modern East Asia-Hyrcanian disjunction is better interpreted as the surviving remnant of a broader Cenozoic history shaped by dispersal, vicariance, and regional extinction. This study provides a conservative barcode baseline and a model-dependent biogeographic framework that can be tested with broader population sampling, additional plastome-scale data, and fossil-inclusive analyses.

CRediT authorship contribution statement

Seyedeh Alemeh Sabbagh: Conceptualization. **Hamed Yousefzadeh:** Supervision. **Mohammad Vatanparast:** Methodology, Data curation. **Mohammad Reza Bakhtiaryzadeh:** Data curation. **Gregor Kozlowski:** Investigation. **Yi- Gang Song:** Project administration, Funding acquisition.

Funding

This work was supported by a grant from the Iranian National Science Foundation (INSF), project No 4024068.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Mansour Pouramin, Ali Besati, and Mohsen Yousefzadeh for their assistance with field sampling and data collection. Their contributions were important to the completion of this study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.hisbio.2026.100048.

Data availability

Data will be made available on request.

References

Akhani, H., Djamali, M., Ghorbanalizadeh, A., Ramezani, E., 2010. Plant biodiversity of Hyrcanian relict forests, N Iran: an overview of the flora, vegetation, palaeoecology and conservation. Pak. J. Bot. 42 (1), 231–258.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolut. Biol. 7 (1), 214. https://doi.org/10.1186/1471-2148-7-214.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29 (8), 1969–1973. https://doi.org/10.1093/molbev/mss075.

Hollingsworth, P.M., 2011. Refining the DNA barcode for land plants. In: Proceedings of the National Academy of Sciences, 108, pp. 19451–19452. https://doi.org/10.1073/pnas.1116812108.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33 (7), 1870–1874. https://doi.org/10.1093/molbev/msw054.

Liu, Y., Chen, K., Wang, L., Yu, X., Xu, C., Suo, Z., Zhou, S., Shi, S., Dong, W., 2025. Assembly-free reads accurate identification (AFRAID) approach outperforms other methods of DNA barcoding in the walnut family (Juglandaceae). Plant. Divers. 47 (1), 115–126. https://doi.org/10.1016/j.pld.2024.10.002.

Lu, A.-M., 1982. On the geographical distribution of Juglandaceae. Acta Phytotaxon. Sin. 20 (3), 257–274.

Lu, Z.-J., Wang, T.-R., Zheng, S.-S., Meng, H.-H., Cao, J.-G., Song, Y.-G., Kozlowski, G., 2024. Phylogeography of Pterocarya hupehensis reveals the evolutionary patterns of a Cenozoic relict tree around the Sichuan Basin. For. Res. 4, e008. https://doi.org/10.48130/forres-0024-0005.

Manchester, S.R., 1987. The fossil history of the Juglandaceae (Vol. 21). Missouri Botanical Garden. https://doi.org/10.5962/bhl.title.154222.

Manning, W.E., 1978. The Classification Within the Juglandaceae. Ann. Mo. Bot. Gard. 65 (4), 1058. https://doi.org/10.2307/2398782.

Milne, R.I., Abbott, R.J., 2002. The origin and evolution of tertiary relict floras. In: Advances in Botanical Research, 38. Elsevier, pp. 281–314. https://doi.org/10.1016/S0065-2296(02)38033-9.

Plant Working Group, C.B.O.L., 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci. 106 (31), 12794–12797. https://doi.org/10.1073/pnas.0905845106.

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19 (18), 2496–2497. https://doi.org/10.1093/bioinformatics/btg359.

Sabbagh, S.A., Yousefzadeh, H., Vatanparast, M., Bakhtiarizadeh, M.R., Kozlowski, G., Song, Y.-G., 2025. Characterization of the chloroplast genome of a relict tree,

Pterocarya fraxinifolia (Juglandaceae), and its comparative analysis. Sci. Rep. 15, 44153. https://doi.org/10.1038/s41598-025-23028-5.

Song, Y.-G., Li, Y., Meng, H.-H., Fragnière, Y., Ge, B.-J., Sakio, H., Yousefzadeh, H., Bétrisey, S., Kozlowski, G., 2020. Phylogeny, Taxonomy, and Biogeography of Pterocarya (Juglandaceae). Plants 9 (11), 1524. https://doi.org/10.3390/plants9111524.

Tang, C.Q., Matsui, T., Ohashi, H., Dong, Y.-F., Momohara, A., Herrando-Moraira, S., Qian, S., Yang, Y., Ohsawa, M., Luu, H.T., Grote, P.J., Krestov, P.V., Ben LePage, Werger, M., Robertson, K., Hobohm, C., Wang, C.-Y., Peng, M.-C., Chen, X., López-Pujol, J., 2018. Identifying long-term stable refugia for relict plant species in East Asia. Nat. Commun. 9 (1), 4488. https://doi.org/10.1038/s41467-018-06837-3.

Tiffney, B.H., Manchester, S.R., 2001. The Use of Geological and Paleontological Evidence in Evaluating Plant Phylogeographic Hypotheses in the Northern Hemisphere Tertiary. Int. J. Plant. Sci. 162 (S6), S3–S17. https://doi.org/10.1086/323880.

Wang, S., Liu, Q., Li, Y., Wang, T., Dai, X., Kozlowski, G., Xu, J., Song, Y., 2026. Cautions on using chloroplast genome-based phylogeny for species identification and biogeography: A case study of Pterocarya. Biodivers. Sci. 34 (5), 25475. https://doi.org/10.17520/biods.2025475.

Xiang, X., Zhang, J., Lu, A., Li, R., 2011. Molecular identification of species in Juglandaceae: A tiered method. J. Syst. Evol. 49 (3), 252–260. https://doi.org/10.1111/j.1759-6831.2011.00116.x.

Xiang, X.-G., Wang, W., Li, R.-Q., Lin, L., Liu, Y., Zhou, Z.-K., Li, Z.-Y., Chen, Z.-D., 2014. Large-scale phylogenetic analyses reveal fagalean diversification promoted by the interplay of diaspores and environments in the Paleogene. Perspect. Plant. Ecol. Evol. Syst. 16 (3), 101–110. https://doi.org/10.1016/j.ppees.2014.03.001.

Yan, H., Zhou, P., Wang, W., Ye, J.-F., Tan, S.-L., Guo, C.-C., Zhang, W.-G., Zhu, Z.-W., Liu, Y.-Z., Xiang, X.-G., 2024. Biogeographic history of Pterocarya (Juglandaceae) inferred from phylogenomic and fossil data. J. Syst. Evol. 62 (6), 1165–1176. https://doi.org/10.1111/jse.13055.

Yu, Y., Blair, C., He, X., 2020. RASP 4: Ancestral State Reconstruction Tool for Multiple Genes and Characters. Mol. Biol. Evol. 37 (2), 604–606. https://doi.org/10.1093/molbev/msz257.

Zhang, Q., Ree, R.H., Salamin, N., Xing, Y., Silvestro, D., 2021. Fossil-informed models reveal a Boreotropical origin and divergent evolutionary trajectories in the walnut family (Juglandaceae). Syst. Biol. 71 (1), 242–258. https://doi.org/10.1093/sysbio/syab030.

Zhou, H., Hu, Y., Ebrahimi, A., Liu, P., Woeste, K., Zhao, P., Zhang, S., 2021. Whole genome based insights into the phylogeny and evolution of the Juglandaceae. BMC Ecol. Evol. 21 (1), 191. https://doi.org/10.1186/s12862-021-01917-3.