scientific reports

Check for updates

OPEN

Characterization of the chloroplast genome of a relict tree, *Pterocarya fraxinifolia* (Juglandaceae), and its comparative analysis

Seyedeh Alemeh Sabbagh[1], Hamed Yousefzadeh[2]✉, Mohammad Vatanparast[3]✉, Mohammad Reza Bakhtiarizadeh[4], Gregor Kozlowski[5,6,7] & Yi-Gang Song[7]✉

The rare, vulnerable relict species *Pterocarya fraxinifolia* is among the last surviving tree species growing in small, scattered populations in the southern Caucasus region; *P. fraxinifolia* grows up to 1000 m in plain forests and is threatened by habitat loss and environmental changes. Here, we sequenced and annotated the chloroplast genome of *P. fraxinifolia* from Hyrcanian forests and compared it to the chloroplast genomes of five other *Pterocarya* species. The evolutionary relationships of *P. fraxinifolia* were subsequently evaluated using the chloroplast genomes and individual chloroplast loci. The chloroplast genome of *P. fraxinifolia* was 160,086 bp in length, comprising 128 genes and a typical quadripartite structure. A comparative analysis of the six *Pterocarya* species revealed limited nucleotide diversity and structural variations in genes. The bulk of the 68 loci identified by SSR analysis comprised A/T repeats. Codon bias analysis revealed strong purifying selection, with the *ndhF* gene showing the highest Ka/Ks ratio. Our phylogenetic analysis revealed *Pterocarya* as a sister to the genus *Juglans* and a distinct subclade within *Pterocarya*.

Keywords Comparative genomics, Conservation, Hyrcanian forests, Plastome evolution

Relict species have always excited evolutionary biologists and biogeographers who consider these species 'living fossils' or relics of prehistoric periods[1,2]. These species have great value as research models for the geographical distribution of intercontinental rifts and as species that ensure biodiversity and ecosystem balance. Relict species also provide relevant information about the adaptation of species to specific environmental changes, as well as the impact of climate change on the animal and plant kingdoms[3].

Hyrcanian forests are hotspots for biodiversity and are home to numerous relict species[4], including 280 endemic and subendemic species[5–7]. The genus *Pterocarya* Kunth (Juglandaceae), commonly referred to as wingnuts, has a disjunct distribution in East Asia and the Caucasus region with its most recent common ancestor present 40 Ma[8]. *Pterocarya* comprises six species, which are classified into two sections, *Pterocarya* (*P. fraxinifolia*, *P. hupehensis*, *P. stenoptera*, and *P. tonkinensis*) and *Platyptera* (*P. macroptera* and *P. rhoifolia*), on the basis of the presence or absence of scales on the terminal buds[9]. *P. fraxinifolia* is the only species in western Asia[10]. The remaining species of *Pterocarya* occur in eastern Asia, such as China and Japan[11–15]. Recently, a series of studies have focused on the phylogeny, biogeography, population genetics, and landscape genetics of species in this genus[14,15]. However, resources regarding the chloroplast genome in this genus are insufficient, and more research is still needed.

Pterocarya fraxinifolia is a deciduous tree that can reach 20–25 m in height and 1.8 m in trunk diameter and is wind-pollinated to produce wing-nut fruits[12]. This species is among the last surviving trees growing in small scattered populations in the southern Caucasus region, which includes northern Iran, Georgia, Armenia,

[1]Faculty of Natural Sciences, Department of Environment Science, Tarbiat Modares University, Tehran, Iran. [2]Faculty of Natural Sciences, Department of Forestry, Tarbiat Modares University, Tehran, Iran. [3]Royal Botanic Gardens Kew, Richmond, Surrey, UK. [4]Department of Animal and Poultry Science, College of Aburaihan, University of Tehran, Tehran, Iran. [5]Department of Biology and Botanic Garden, University of Fribourg, Chemin du Musée 10, Fribourg CH-1700, Switzerland. [6]Natural History Museum Fribourg, Chemin du Musée 6, Fribourg CH-1700, Switzerland. [7]Eastern China Conservation Centre for Wild Endangered Plant Resources, Shanghai Chenshan Botanical Garden, Shanghai 201602, China. ✉email: h.yousefzadeh@modares.ac.ir; M.Vatanparast@kew.org; cherish-faith@163.com

Azerbaijan, and the Anatolian region in Turkey[12,15]. However, less than two decades ago, small populations were first recorded in western Iran in the provinces of Lorestan and Ilam in the Zagros Mountains[16].

The chloroplast genome is widely used in phylogenetic studies because of its relatively conserved structure[17,18] and uniparental inheritance[19,20],. Chloroplast genomes can provide important information about the adaptation of species to different environmental conditions[21–23]. Despite the slow evolutionary rates of chloroplast genomes, coding and noncoding regions are useful for the identification of closely related species[24–26] and for detecting genome-scale evolutionary patterns. Comparisons of the structure and sequence of these regions across different species within a genus can reveal important evolutionary phenomena such as gene transfer, deletion, or duplication. Recently, with the continuous application of high-throughput sequencing techniques, chloroplast DNA sequences have become readily available[13,27,28]. However, there is no annotated chloroplast genome available for *P. fraxinifolia*, which hinders the understanding of the evolution of the chloroplast genome of this species from West Asia[14,15].

In this study, we aim to (1) assemble and annotate the chloroplast genome of the relict species *P. fraxinifolia* from Hyrcanian forests; (2) perform comparative genomics of the chloroplast genomes of six *Pterocarya* species; and (3) assess the systematic affinity of *P. fraxinifolia* using phylogenetic analysis of the assembled chloroplast genomes.

Materials and methods

Leaf material for the *P. fraxinifolia* sample was collected from a wild population in Mazandaran, Iran (Fig. 1). The voucher samples were deposited at the Herbarium of the Nowshahr Botanical Garden (HNBG) under voucher number 12,876.

Genomic DNA was extracted using the CTAB method, and its quality and quantity were checked using a Qubit 2.0 and Agilent 2100 Bioanalyzer. Libraries were created and sequenced at Wuhan Benagen Tech Solutions Company Limited, Wuhan, China, using the DNBSEQ platform (paired-end 150 bp). SOAPnuke v1.3.0 was used to filter the raw data, yielding 20 GB of clean data[29].

Chloroplast genome assembly and annotation

Raw reads were filtered using Trimmomatic v0.39[30] with a quality cutoff of 15 in a 4-base sliding window; any reads that were less than 50 bp were removed, and the adapters were filtered out. The quality of the reads before and after trimming was tested using FASTQC v0.12.1. We used GetOrganelle[31] v.1.7.7.0 for chloroplast genome assembly, with the embplant_pt database used as a reference and maximum extension rounds of 15 (-R). GetOrganelle produced two isomers of the whole chloroplast genome of *P. fraxinifolia*, and each genome had a distinct relative orientation for the small single-copy (SSC) region[32]. A Python script from GetOrganelle was used along with Bowtie2 v2.5.4[33] to determine the average read coverage throughout the chloroplast genome. GeSeq v2.03[34] was used for the initial chloroplast genome annotation of *P. fraxinifolia*, and the output from GeSeq was imported into Geneious Prime 2025.0.3 for an additional annotation check via the "Transfer Annotation" function. Chloroplot[35] was used to produce a circular representation of the plastome.

Comparative analyses of the Chloroplast genomes

Because the flanking inverted repeat (IR) regions of the chloroplast genome often vary among species, we used CPJSdraw[36] to compare the IR regions of the six species. We used CUSP from EMBOSS v6.6.0.0 to calculate relative synonymous codon usage (RSCU) for protein-coding genes of *P. fraxinifolia*. To identify simple sequence repetitions (SSR), we used a Perl script from the Microsatellite Identification tool (MISA)[37]. The settings were adjusted to ten, five, and four repeats for mononucleotides, dinucleotides, and trinucleotides, respectively. Forward, reverse, palindrome, and complementary sequences with a minimum repeat length of eight bp and a maximum computed repeat of 50% were analyzed using REPuter[38]. The complete chloroplast genome sequences of the six *Pterocarya* species were aligned with Fast Statistical Alignment v1.15.9[38] to perform the nucleotide

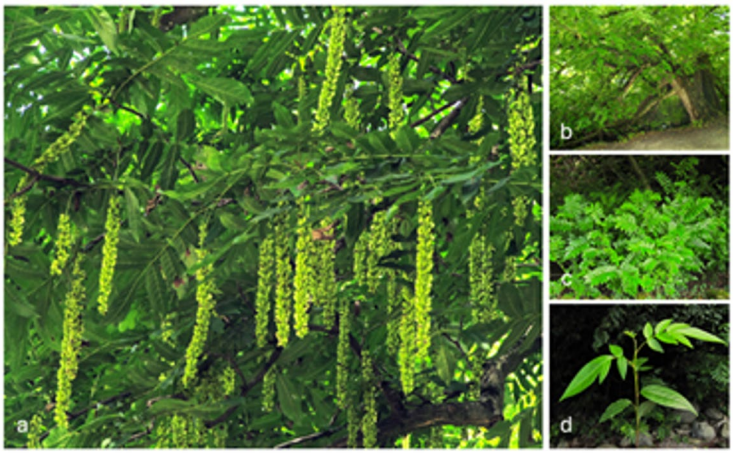


Fig. 1. (**a**) Fruits in pendulous form; (**b**) A mature tree; (**c**) Regeneration under tree canopy; (**d**) Seedling.

diversity analysis. We used a Perl script (https://github.com/xul962464/perl-Pi-nucleotide-diversity) to estimate the nucleotide diversity (PI) with a sliding window analysis with a step size of 200 bp and a window length of 800 bp.

The selection pressure on chloroplast protein-coding genes (CDSs) was evaluated by aligning the nonredundant genes from six species using MAFFT v7.526[39]. We ran ParaAT.pl v2.0[40] to compute synonymous substitution rates (Ks), nonsynonymous substitution rates (Ka), and Ka/Ks. Each CDS pair of one-to-one species combinations is used as a homolog with genetic code 11. We estimated Ka, Ks, and Ka/Ks among the six *Pterocarya* species with KaKs_Calculator v2.0[41].

Phylogenetic analysis

We constructed a maximum likelihood (ML) phylogenetic tree to understand the relationships of *Pterocarya* species. Chloroplast genome sequences were acquired from GenBank for the other *Pterocarya* and related genera in the Juglandaceae family. The multiple sequence alignment contained a total of 21 taxa. We performed our phylogenetic analysis using the full chloroplast genome alignment, treating it as a standard coalescent gene[41]. The chloroplast genomes were aligned using Fast Statistical Alignment v1.15.9[42] and then trimmed with trimAL v1.5[43] with the following settings: -automated1 -res overlap 0.7, -seqoverlap 65. To overcome the alignment issues, we also employed TAPER v1.0.047 with the -m N -a N parameters.

Using RAxML-NG v1.2.1[44], we constructed the GTR + G model and the ML tree with 500 bootstrap repetitions. The phylogenetic tree was rooted using *Engelhardia roxburghiana* Wall. as an outgroup. The tree was drawn using FigTree v1.4.4 (https://github.com/rambaut/figtree). To determine the genetic distance between the six *Pterocarya* species, the HKY85 model[45] was used, and a phylogenetic network was generated using the NeighborNet approach in SplitsTree CE v6.0.0[46].

Results

Chloroplast genome assembly and annotation

The total numbers of raw and trimmed reads for *P. fraxinifolia* in this study were 143,190,876 and 141,927,817 base pairs (bp), respectively. The number of matched mapped pairs across the chloroplast genome was 393.42 ± 82.15 (Fig. S1). The complete chloroplast genome of *P. fraxinifolia* has a typical quadripartite structure that is 160,086 bp in length with a large single-copy region (LSC) of 89,582 bp, a small single-copy region (SSC) of 18,398 bp, and a pair of inverted repeat regions (IRs) of 26,053 bp (Fig. 2). A total of 148 genes were annotated in the chloroplast genome of *P. fraxinifolia*, including 103 protein-coding genes, 37 transfer RNA (tRNA) genes, and eight ribosomal RNA (rRNA) genes (Table 1 and Table S1). The GC content of the chloroplast genome was 36.17%. The annotated complete chloroplast genome of *P. fraxinifolia* was deposited in GenBank (accession number PV791734).

Comparative analyses of the Chloroplast genome and nucleotide diversity

According to a comparative analysis of the chloroplast genomes of *Pterocarya* species, the locations of eight genes in the chloroplast maps differed among species. The *rps19* gene starts at position zero of the LSC region for *P. fraxinifolia*, but its position has shifted three times into the IRb region in the others. However, in other species of *Pterocarya*, a small portion of the genes were located in the IRb region. The *ndhF* gene in *P. fraxinifolia*, *P. stenoptera*, *P. macroptera*, and *P. rhoifolia is* located inside the SSC and is 2226 bp in length, whereas in *P. tonkinensis* and *P. hupehensis*, it spans 69 and 145 bp, respectively, into the IRb region (Fig. 3a).

The average nucleotide diversity (π) value was 0.001492, with a range of 0 to 0.00556 (Fig. 3B). The CDSs with the highest π values, which were greater than 0.0031, were *ndhF*, *infA*, *ycf1*, *rps15*, and *matK*. The *ycf1* gene is found in the SSC area, whereas *ndhF*, *infA*, *rps15*, and *matK* are found in the LSC region. Nucleotide diversity decreased in both IR zones. Furthermore, 35 CDSs had a π value of zero among the six *Pterocarya* species, indicating that they were conserved (Table S1).

Repeated sequence analysis

The six *Pterocarya* chloroplast genomes have an average of 72.6 SSR loci (Fig. 4A), with *P. rhoifolia* having the most SSR loci (85) of the six species (Table S2). A thorough examination of the chloroplast genome of *P. fraxinifolia* revealed 68 microsatellites, comprising 63 mononucleotides, four dinucleotides, and one trinucleotide simple sequence repeat. The five types of sequence repeat motifs—forward, reverse, complementary, palindromic, and tandem—are summarized in Table S3 and Fig. 4B. The analysis also revealed that the number of repetitive sequences differed across the six *Pterocarya* chloroplast genomes. Approximately 96.82% of the mononucleotide repeats found in *P. fraxinifolia* were classified as A/T (61), and 3.18% (2 repeats) were classified as C/G. In contrast, approximately 88.2% of the repeats found in *P. rhoifolia* were classified as A/T (75), and 3.52% (3 repeats) were classified as C/G (Fig. 4C). Dinucleotide repeats (6) for *P. rhoifolia* and (4) for *P. fraxinifolia* were the next most prevalent type of SSR. This investigation revealed no repeats of tetranucleotides, pentanucleotides, or hexanucleotides.

Ka/Ks ratio and codon bias analysis

Strong purifying selection and functional limitations are indicated by the very low Ka/Ks ratios found in most CDS regions among *Pterocarya* species (Fig. 5A). With the exception of *P. tonkinensis* and *P. stenoptera*, the highest Ka/Ks ratio was detected in the chloroplast NADH dehydrogenase F (*ndhF*) gene. The GC contents for the first, second, and third codon locations were 45.30%, 38.25%, and 30.36%, respectively, whereas the overall coding GC content was 37.97%. The greatest frequencies were 42.361 for the ATT codon and 37.605 for the GAA codon. The only two codons with an RSCU value of 1 were tryptophan (TGG) and methionine (ATG) (Fig. 5B). Every codon ending in A or T had an RSCU value greater than 0.5.

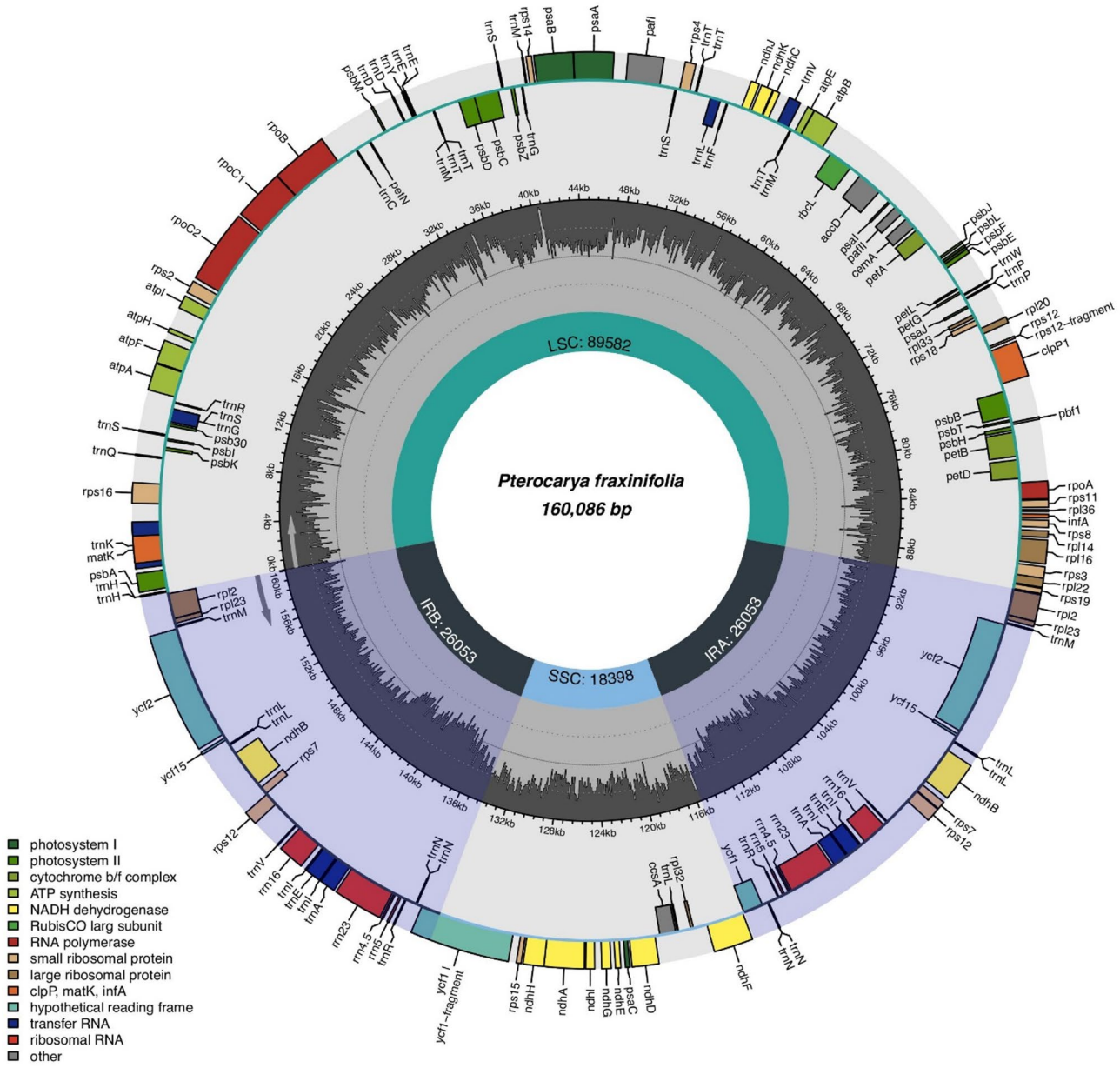


Fig. 2. Schematic map of overall features of the chloroplast genome of *P. fraxinifolia*. From the center outward, the first track shows the small single-copy (SSC), inverted repeat (IRa and IRb), and large single-copy (LSC) regions. The GC content along the genome is plotted on the second track. The genes are shown on the third track. Genes are color-coded by their functional classification. The transcription directions for the inner and outer genes are clockwise and anticlockwise, respectively. The functional classification of the genes is shown in the bottom left corner.

Species	Genome Size (bp)	LSC size (bp)	SSC size (bp)	IR size (bp)	GenBank number	GC%	Number of genes			
							Total	CDS	rRNA	tRNA
P. fraxinifolia	160,086	89,582	18,398	52,106	TBC	36.17	153	91	8	54
P. hupehensis	159,770	89,229	18,504	26,018	NC046431	36.24	137	89	8	40
P. stenoptera	160,202	89,727	18,432	26,021	NC046428	36.17	137	89	8	40
P. tonkinensis	160,096	89,600	18,481	26,007	NC046427	36.21	137	89	8	40
P. macroptera	160,168	89,701	18,453	26,007	MW194257	36.17	136	88	8	40
P. rhoifolia	160,315	89,814	18,458	26,021	ON380923	36.15	138	90	8	40

Table 1. Summary of the genome of *Pterocarya* species.

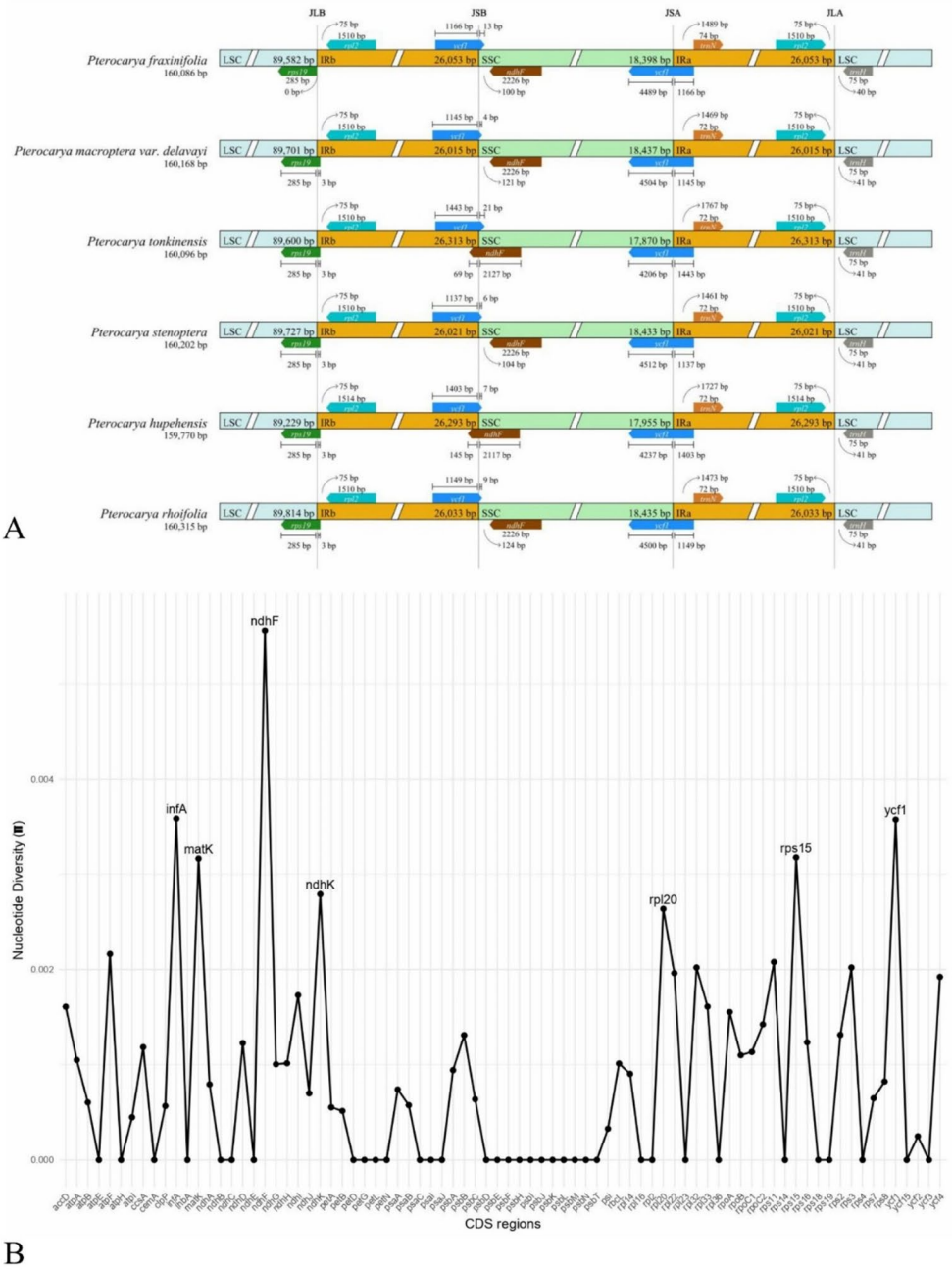
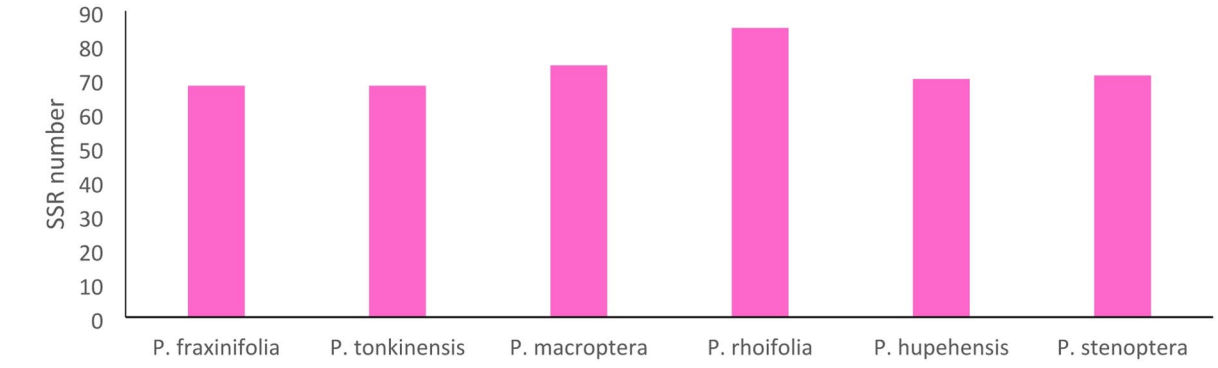


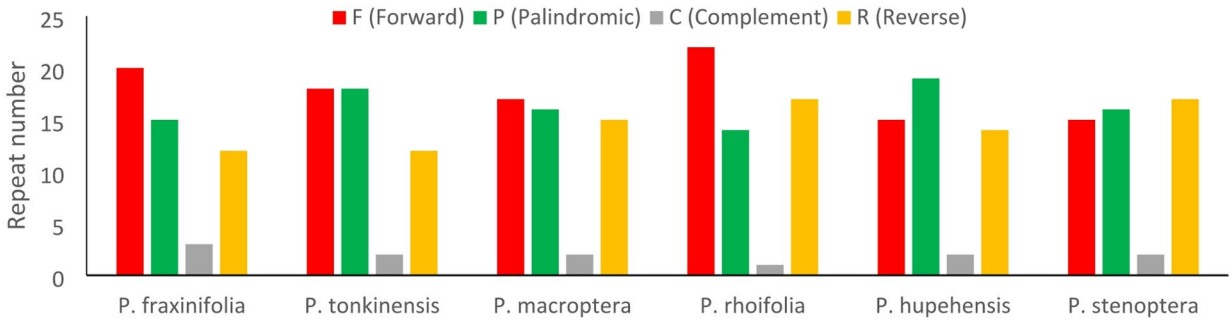
Fig. 3. (**A**) Comparisons of LSC, SSC, and IR region boundaries among six *Pterocarya* species; (**B**) Nucleotide diversity (π) of CDS regions.

Phylogenetic analysis

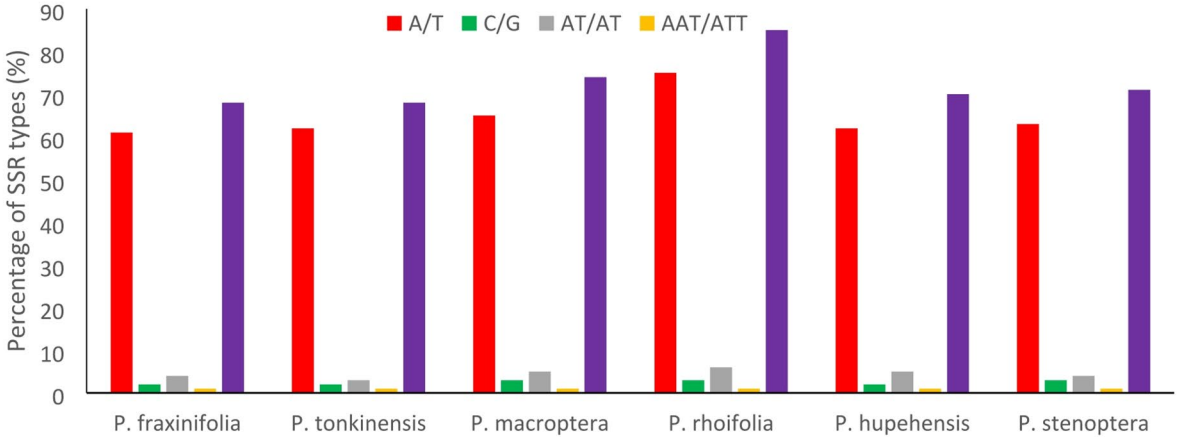
The aligned multiple sequence alignment for the phylogenetic analysis consisted of 158,422 bp across 21 accessions, with 0.21% gaps and 96.19% invariant sites. The phylogenetic tree revealed *Pterocarya* as a sister genus to *Juglans* L. with 100% bootstrap support (Fig. 6A). The ML phylogenetic tree confirmed the monophyly of the genus *Pterocarya* with 100% bootstrap support with two subclades. *P. fraxinifolia* is a sister to a monophyletic subclade that include *P. tonkinensis* and *P. macroptera* and a sister to another subclade that includes *P. rhoifolia*, *P. stenoptera*, and *P. hupehensis*. The network analysis of the six *Pterocarya* species revealed a topology similar to that of the ML tree, with *P. tonkinensis* clustering with *P. macroptera* and *P. rhoifolia* clustering with *P. stenoptera* and *P. hupehensis*, while *P. fraxinifolia* branched off independently. In this study, the efficiency of two barcode regions, *matK* and *ycf1*, in the phylogeny of the genus *Pterocarya* was evaluated (Fig. 6B and C). The results revealed that the phylogenetic tree based on the *matK* region was identical to the phylogenetic tree derived from the complete chloroplast genome sequence. Pairwise distance analysis using the HKY85 method revealed that *P. fraxinifolia* is distantly related to Asiatic *Pterocarya* species (Fig. S2). The genetic distances between *P. macroptera* and *P. tonkinensis* (0.000259) and between *P. stenoptera* and *P. hupehensis* (0.000526) were the lowest,



(a)



(b)



(c)

Fig. 4. Analysis of perfect simple sequence repeats (SSRs) in six *Pterocarya* chloroplast genomes. (**A**) The frequency of identified SSRs in large single-copy (LSC), inverted repeat (IR,) and small single-copy (SSC) regions; (**B**) The number of SSR types detected in the nine sequenced chloroplast genomes; (**C**) The frequency of identified SSR motifs in different repeat class types.

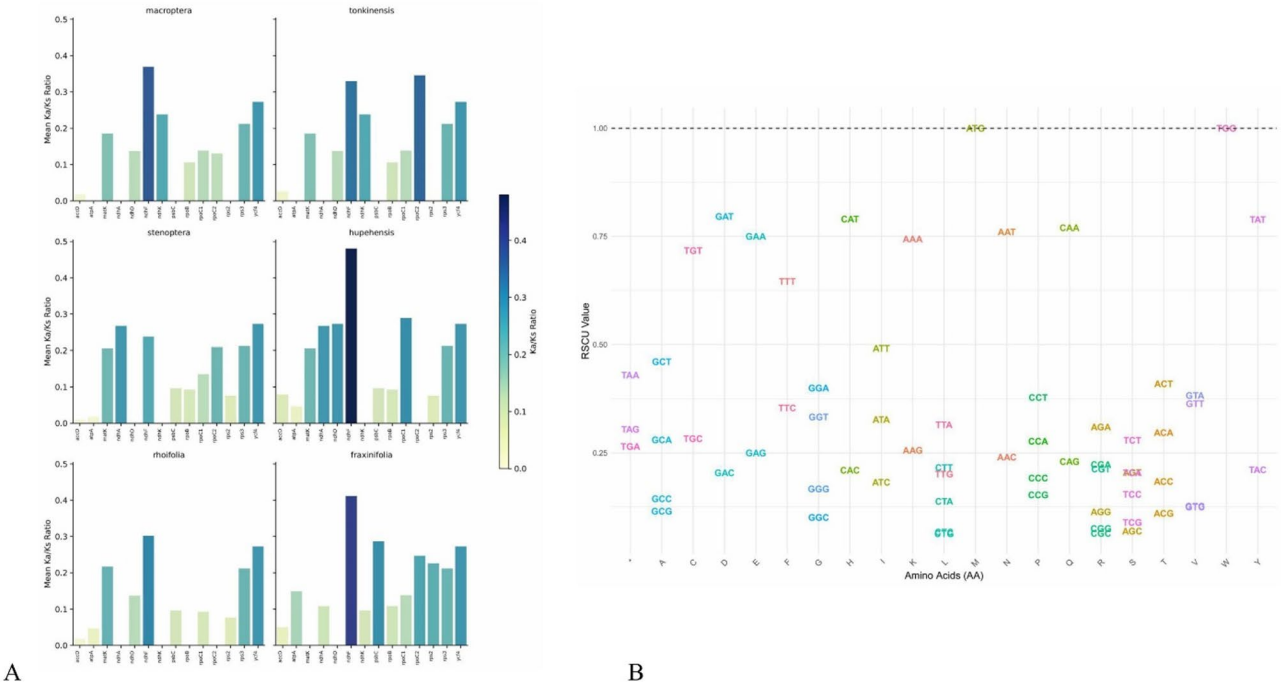


Fig. 5. Ka/Ks ratios of chloroplast protein-coding sequences across six *Pterocarya* species. (**A**) The X-axis is selected CDS with Ka/Ks ratios above 0.001. The Y-axis shows the mean Ka/Ks ratio for each gene. (**B**) Relative Synonymous Codon Usage (RSCU) value for each codon.

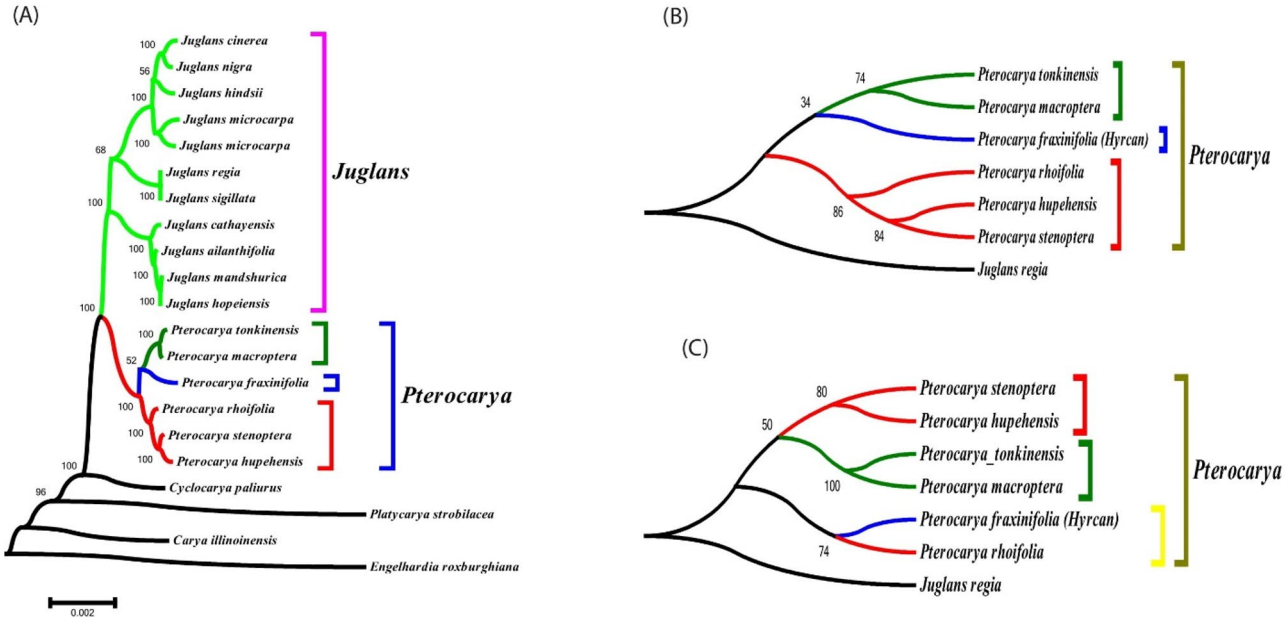


Fig. 6. Comparison of three phylogenetic trees based on different chloroplast sequences: (**a**) Whole chloroplast genome, (**b**) *matK* gene region, and (**c**) *ndhF* gene region.

whereas the genetic distances between *P. fraxinifolia* and *P. hupehensis* (0.002153) and between *P. fraxinifolia* and *P. tonkinensis* (0.001928) were more than eightfold greater (Table S4). In the MatK dataset, *P. fraxinifolia* had three unique character states that differentiated it from other species of *Pterocarya (*Table S5).

Discussion

Chloroplast genomes are useful tools for studying the evolutionary relationships among species because of their preserved structure and uniparental inheritance (usually maternal in angiosperms[47,48]. Considering mechanisms of plant evolution[49,50] and that the evolutionary history of chloroplasts is normally different from that of nuclear markers[51,52], the use of genetic information from chloroplasts could reflect how seed dispersal affects the genetic makeup of wild populations and species.

This study is the first to annotate the chloroplast genome of *P. fraxinifolia* and compare it to that of other species. We found that the positions of eight markers, namely, *rps19*, *rpl2*, *ycf1* (IRa and IRb), *ndhF*, *trnN*, *rpl2*, and *trnH*, varied among the six *Pterocarya* chloroplast genomes. This implies that the expansion and contraction of the IR, LSC, and SSC areas are the primary sources of fluctuations in chloroplast genome size[53,54]. Between 68 and 85 SSRs were found among the chloroplast genomes of the six *Pterocarya* species. While the number of poly(G)/(C) repeats was shown to be greater in other angiosperms, the number of poly(A)/(T) repeats was significantly greater in *Pterocarya.*

Five genes, *ndhF*, *infA*, *ycf1*, *rps15* and *matK*, presented the greatest nucleotide variability (above 0.003). The *matK* and *ycf1* genes have been suggested to function as barcode regions in plants[55]. The *matK* gene encodes the maturase protein, which facilitates the splicing of group II introns in several chloroplast genes and is considered a core barcode for land plants[50,51]. The *ycf1* gene, which encodes the TIC214 protein that is essential for plant viability, is the second largest in the chloroplast genome and has recently been assessed for its DNA barcoding potential[50–52], showing higher variability than the existing chloroplast candidate barcodes (such as *rbcL*, *matK* and *trnH-psbA*). Therefore, the *ycf1* gene might be potentially useful as a DNA barcode for the *Pterocarya* genus[56]. With the exception of the *matK* region, none of the seven recommended barcode candidate genes in chloroplast genomes[50] have the potential for barcoding of the *Pterocarya* genus because of a lack of nucleotide variation. Surprisingly, the accuracy of the *matK* region in resolving the phylogeny of the genus *Pterocarya* was identical to that of the complete chloroplast genome. Therefore, the *matK* gene alone is sufficient for reconstructing the phylogenetic relationships within the genus *Pterocarya*, eliminating the need for the additional time and financial resources required for whole-chloroplast-genome sequencing.

The genus *Pterocarya* consists of six species and is closely related to *Juglans* in terms of pollen morphology, wood anatomy and molecular phylogenetics[8,9]. Our phylogenetic results confirm the sister relationship of *Pterocarya* to *Juglans*. Two sections for *Pterocarya* have been proposed on the basis of the presence or absence of scales on the terminal buds[9,13,50]. *P. fraxinifolia*, *P. hupehensis*, *P. stenoptera*, and *P. tonkinensis* belong to the section *Pterocarya*, while *P. macroptera* and *P. rhoifolia* belong to the section *Platyptera*. According to our chloroplast genome-based phylogeny, this suggested morphological classification is not supported, and the Caucasian wingnut (*P. fraxinifolia*) is in a distant subclade from the Chinese wingnut (*P. stenoptera*) and the Japanese wingnut (*P. rhoifolia*).

The pairwise genetic distance between the Caucasian wingnut and other Asiatic *Pterocarya* species is greater. This distance might reflect the prolonged isolation and considerable geographic distance between Caucasian wingnut and East Asian species. Recent divergence time analyses based on fossil calibrations estimated the age of *P. fraxinifolia* between 9.4 and 18.4 Ma from the Miocene period and suggested the westward dispersal of *Pterocarya* from East Asia[8]. Wingnut fruit structure could facilitate the dispersal of these species by wind and water[57]. In this study, we collected *P. fraxinifolia* materials from its natural habitat in Hyrcanian forests. Our initial phylogenetic results revealed that the publicly available *P. fraxinifolia* in GenBank (NC046430) is not a *P. fraxinifolia* and is most likely a misidentified voucher that could be *P. stenoptera* (data not shown).

Toward conservation of *P. fraxinifolia*

P. fraxinifolia is classified as a vulnerable relict species on the IUCN Red List[12]. Our phylogenetic tree, which was constructed on the basis of chloroplast genome analysis, indicates that this species is completely distinct from other species of the genus originating from China and Japan. This distinction might highlight the species' unique evolutionary path and specialized ecological environments. Recent studies have shown that the potentially suitable ranges of *P. fraxinifolia* will increase under future climate scenarios[8,58], and the rapid loss of its habitat, combined with growing threats such as drought and the destruction of riparian ecosystems in Hyrcanian forests, will result in its conservation an urgent priority.

Data availability

The annotated complete chloroplast genome of *P. fraxinifolia* was deposited in GenBank, under accession number PV791734.1.

Received: 17 June 2025; Accepted: 3 October 2025
Published online: 19 December 2025

References

1. Grandcolas, P., Nattier, R. & Trewick, S. Relict species: a relict concept? *Trends Ecol. Evol.* **29**, 655–663 (2014).
2. Ian Milne, R. Northern hemisphere plant disjunctions: A window on tertiary land bridges and climate change? *Ann. Bot.* **98**, 465–472 (2006).
3. Raposo, M. & Pinto-Gomes, C. Dynamics of vegetation and climate change. *Environments* **9**, 36 (2022).

4. Zarandian, A. et al. Anthropogenic decline of ecosystem services threatens the integrity of the unique hyrcanian (Caspian) forests in Northern Iran. *Forests* **7**, 51 (2016).
5. Akhani, H., Malekmohammadi, M., Mahdavi, P., Gharibiyan, A. & Chase, M. W. Phylogenetics of the Irano-Turanian taxa of *Limonium* (Plumbaginaceae) based on ITS NrDNA sequences and leaf anatomy provides evidence for species delimitation and relationships of lineages: phylogenetics of IRano-TUranian *Limonium*. *Bot. J. Linn. Soc.* **171**, 519–550 (2013).
6. Noroozi, J. et al. Endemic diversity and distribution of the Iranian vascular flora across phytogeographical regions, biodiversity hotspots and areas of endemism. *Sci. Rep.* **9**, 12991 (2019).
7. Naqinezhad, A. et al. The combined effects of climate and canopy cover changes on understorey plants of the hyrcanian forest biodiversity hotspot in Northern Iran. *Glob Change Biol.* **28**, 1103–1118 (2022).
8. Yan, H. et al. Biogeographic history of *Pterocarya* (Juglandaceae) inferred from phylogenomic and fossil data. *J. Syst. Evol.* **62**, 1165–1176 (2024).
9. Manning, W. E. The classification within the Juglandaceae. *Ann. Mo Bot. Gard*. **65**, 1058 (1978).
10. Usher, M. B. & Edwards, M. A dipteran from South of the Antarctic circle: belgica Antarctica (Chironomidae) with a description of its larva. *Biol. J. Linn. Soc.* **23**, 19–31 (1984).
11. Manchester, S. R. *The Fossil History of the Juglandaceae* (Missouri Botanical Garden, 1987).
12. Kozlowski, G., Bétrisey, S. & Song, Y. G. *Wingnuts (Pterocarya) & Walnut Family: Relict Trees: Linking the Past, Present and Future*. (Natural History Museum Fribourg (NHMF), Department of Education, Culture and Sport of the State of Fribourg, Switzerland, Fribourg, (2018).
13. Song, Y. G. et al. Phylogeny, Taxonomy, and biogeography of Pterocarya (Juglandaceae). *Plants* **9**, 1524 (2020).
14. Wang, T. R. et al. Adaptive divergence and genetic vulnerability of relict species under climate change: a case study of *Pterocarya macroptera*. *Ann. Bot.* **132**, 241–254 (2023).
15. Lu, Z. J. et al. Phylogeography of *Pterocarya hupehensis* reveals the evolutionary patterns of a cenozoic relict tree around the Sichuan basin. *For. Res.* **4**, e008 (2024).
16. Akhani, H. & Salimian, M. An extant disjunct stand of *Pterocarya fraxinifolia (Juglandaceae)* in the central Zagros Mountains, W Iran. *Willdenowia* **33**, 113–120 (2003).
17. Yao, X. et al. The first complete Chloroplast genome sequences in actinidiaceae: genome structure and comparative analysis. *PLOS ONE*. **10**, e0129347 (2015).
18. Cui, G. et al. Complete Chloroplast genome of hordeum brevisubulatum: genome organization, synonymous codon usage, phylogenetic relationships, and comparative structure analysis. *PLOS ONE*. **16**, e0261196 (2021).
19. Ni, Z., Zhou, P., Xin, Y., Xu, M. & Xu, L. A. Parent–offspring variation transmission in full-sib families revealed predominantly paternal inheritance of Chloroplast DNA in Pinus massoniana (Pinaceae). *Tree Genet. Genomes*. **17**, 36 (2021).
20. Villanueva-Corrales, S. et al. The complete Chloroplast genome of Plukenetia volubilis provides insights into the organelle inheritance. *Front. Plant. Sci.* **12**, 667060 (2021).
21. Pottosin, I. & Shabala, S. Transport across Chloroplast membranes: optimizing photosynthesis for adverse environmental conditions. *Mol. Plant*. **9**, 356–370 (2016).
22. Xia, L. et al. Chloroplast Pan-Genomes and comparative transcriptomics reveal genetic variation and temperature adaptation in the cucumber. *Int. J. Mol. Sci.* **24**, 8943 (2023).
23. Li, Y. et al. New insights on the phylogeny, evolutionary history, and ecological adaptation mechanism in cycle-cup Oaks based on Chloroplast genomes. *Ecol. Evol.* **14**, e70318 (2024).
24. Lim, L. W. K., Chung, H. H. & Hussain, H. Complete Chloroplast genome sequencing of Sago palm (Metroxylon Sagu Rottb.): molecular structures, comparative analysis and evolutionary significance. *Gene Rep.* **19**, 100662 (2020).
25. Turudić, A. et al. Variation in Chloroplast genome size: biological phenomena and technological artifacts. *Plants* **12**, 254 (2023).
26. Daniell, H., Lin, C. S., Yu, M. & Chang, W. J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* **17**, 134 (2016).
27. Han, H. et al. Analysis of Chloroplast genomes provides insights into the evolution of agropyron. *Front. Genet.* **13**, 832809 (2022).
28. Long, L. et al. Complete Chloroplast genomes and comparative analysis of ligustrum species. *Sci. Rep.* **13**, 212 (2023).
29. Chen, Y. et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience* 7, (2018).
30. Bolger, A. et al. The genome of the stress-tolerant wild tomato species solanum pennellii. *Nat. Genet.* **46**, 1034–1038 (2014).
31. Jin, J. J. et al. GetOrganelle: a fast and versatile toolkit for accurate de Novo assembly of organelle genomes. *Genome Biol.* **21**, 241 (2020).
32. Palmer, J. D., Shields, C. R., Cohen, D. B. & Orton, T. J. Chloroplast DNA evolution and the origin of amphidiploid brassica species. *Theor. Appl. Genet.* **65**, 181–189 (1983).
33. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with bowtie 2. *Nat. Methods*. **9**, 357–359 (2012).
34. Tillich, M. et al. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **45**, W6–W11 (2017).
35. Zheng, S., Poczai, P., Hyvönen, J., Tang, J. & Amiryousefi, A. Chloroplot: an online program for the versatile plotting of organelle genomes. *Front. Genet.* **11**, 576124 (2020).
36. Li, H. et al. CPJSdraw: analysis and visualization of junction sites of Chloroplast genomes. *PeerJ* **11**, e15326 (2023).
37. Thiel, T., Michalek, W., Varshney, R. & Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L). *Theor. Appl. Genet.* **106**, 411–422 (2003).
38. Kurtz, S. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **29**, 4633–4642 (2001).
39. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
40. Zhang, Z. et al. ParaAT: A parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* **419**, 779–781 (2012).
41. Zhang, Z. et al. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteom. Bioinf.* **4**, 259–263 (2006).
42. Bradley, R. K. et al. Fast statistical alignment. *PLoS Comput. Biol.* **5**, e1000392 (2009).
43. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
44. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
45. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).
46. Bryant, D. & Huson, D. H. NeighborNet: improved algorithms and implementation. *Front. Bioinforma.* **3**, 1178600 (2023).
47. Brock, J. R., Mandáková, T., McKain, M., Lysak, M. A. & Olsen, K. M. Chloroplast phylogenomics in *Camelina* (Brassicaceae) reveals multiple origins of polyploid species and the maternal lineage of *C. sativa*. *Hortic. Res.* **9**, uhab050 (2022).
48. Yang, L. et al. Phylogenomic analyses reveal an allopolyploid origin of core Didymocarpinae (Gesneriaceae) followed by rapid radiation. *Syst. Biol.* **72**, 1064–1083 (2023).
49. Wu, F. Y., Ma, S. C., Ye, P. M., Ye, H. & Ma, J. L. The complete Chloroplast genome sequence of camellia Zhaiana (Theaceae), a critically endangered species from China. *Mitochondrial DNA Part. B*. **6**, 2425–2426 (2021).
50. Wei, F. et al. The complete Chloroplast genome sequence of the medicinal plant sophora tonkinensis. *Sci. Rep.* **10**, 1–13 (2020).

51. Birky, C. W. Uniparental inheritance of mitochondrial and Chloroplast genes: mechanisms and evolution. *Proc. Natl. Acad. Sci.* **92**, 11331–11338 (1995).
52. Xia, M. et al. Comparative Chloroplast genome study of Zingiber in China sheds light on plastome characterization and phylogenetic relationships. *Genes* **15**, 1484 (2024).
53. Kim, K. J. Complete Chloroplast genome sequences from Korean ginseng (Panax schinseng Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* **11**, 247–261 (2004).
54. Guo, L., Zhai, J. & Gu, Y. The complete Chloroplast genome sequence of *Isoetes Baodongii* (Isoetaceae). *Mitochondrial DNA Part. B.* **9**, 667–671 (2024).
55. Zhu, S., Liu, Q., Qiu, S., Dai, J. & Gao, X. DNA barcoding: an efficient technology to authenticate plant species of traditional Chinese medicine and recent advances. *Chin. Med.* **17**, 1–17 (2022).
56. Dong, W. et al. ycf1, the most promising plastid DNA barcode of land plants. *Sci. Rep.* **5**, 1–5 (2015).
57. Maharramova, E. et al. Phylogeography and population genetics of the riparian relict tree Pterocarya fraxinifolia (Juglandaceae) in the South Caucasus. *Syst. Biodivers.* **16**, 14–27 (2018).
58. Song, Y., Feng, L., Alyafei, M. A. M., Jaleel, A. & Ren, M. Function of chloroplasts in plant stress responses. *Int. J. Mol. Sci.* **22**, 13464 (2021).

Acknowledgements

Most data analysis was performed at the Smithsonian Institution Hydra cluster https://doi.org/10.25572/SIHPC. M.V. thanks the support of Rebecca B. Dikow, Matthew Kweskin, and Eric Schuettpelz.

Author contributions

H. Y. conceived and designed this study. M. V. and S. A. S conducted a formal analysis. M. B. contributed to the analytical methods. S. A. S, H. Y., and M. V. wrote the original draft. G. K. and Y. G. S. edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

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

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

The plant material of *P. fraxinifolia* was collected from natural populations in northern Iran under a PhD research project approved by Tarbiat Modares University, the Ministry of Science, Research and Technology of Iran. According to national regulations, the collection of plant material for academic research within Iran does not require additional permits when conducted as part of an approved university project. All sampling was done in compliance with institutional and national guidelines. We fully acknowledge the importance of adhering to the IUCN Policy Statement on Research Involving Species at Risk of Extinction as well as the Convention on the Trade in Endangered Species of Wild Fauna and Flora (CITES). We are committed to ensuring that our research complies with these guidelines and supports the conservation of endangered species.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-23028-5.

Correspondence and requests for materials should be addressed to H.Y., M.V. or Y.-G.S.