BMC Plant Biology

RESEARCH

Open Access

Plastome comparison and phylogenomics of Chinese endemic *Schnabelia* (Lamiaceae): insights into plastome evolution and species divergence

Shengnan Wei[1], Jianan Ying[1], Mengxia Lu[2], Jie Li[1], Yanbo Huang[3], Zhenming Wu[1], Paul Nevill[4], Pan Li[5], Xinjie Jin[6*] and Qixiang Lu[1*]

Abstract

Background *Schnabelia* species, herbaceous perennial plants within the Lamiaceae family, possess medicinal value and are endemic to China. While previous studies have focused on morphological classification, molecular systematics, and medicinal components, there has been limited research on phylogenomics. To reveal their plastid genome characteristics and phylogenetic relationships, we sequenced and assembled the plastomes of all five *Schnabelia* species (*S. oligophylla*, *S. tetrodonta*, *S. nepetifolia*, *S. terniflora*, *S. aureoglandulosa*), conducted comparative genomic analyses, and constructed a phylogenetic tree incorporating closely related taxa in subfamily Ajugoideae, as well as conducting divergence time estimation.

Results Plastome size of the five species ranged from 155,733 bp to 156,944 bp, encompassing 115 unique genes, with a GC content of 37.8% same for all species. Five intergenic spacer regions (*trn*H-*GUG*-*psb*A, *trn*K-*UUU*-*mat*K, *pet*B-*pet*D, *ndh*D-*psa*C, *ndh*A-*ndh*H) were identified as divergence hotspots. Gene selection pressure analysis demonstrated that all genes were under negative selection. Phylogenetic relationship of Ajugoideae species based on plastomes confirmed the monophyly of *Schnabelia*. Two clades within *Schnabelia* were supported, one containing two original species and the other comprising three species transferred from *Caryopteris*. The stem age of the *Schnabelia* is estimated to be approximately 30.24 Ma, with the split of two Sections occurring around 12.60 Ma.

Conclusions We revealed plastid genome evolutionary features for five species within the genus *Schnabelia*. The identified highly variable regions can provide a tool for future identification of these medicinal plants. The diversification of *Schnabelia* during middle Miocene and the Quaternary suggests that historical geological and climatic shifts facilitated species differentiation. These findings enhance our understanding of *Schnabelia*'s evolution and support future research on chloroplast diversity, aiding conservation and sustainable use.

Keywords Chloroplast genome, Comparative genomics, Lamiaceae, Phylogenomics, *Schnabelia*

*Correspondence:
Xinjie Jin
20210347@wzu.edu.cn
Qixiang Lu
lqxdeer@zstu.edu.cn

Full list of author information is available at the end of the article

Background

Schnabelia Hand.-Mazz., a genus within the family Lamiaceae Martinov, comprises five herbaceous perennial plant species endemic to central, south, southeast, and southwest China [1]. *Schnabelia* belongs to one of the 12 subfamilies of Lamiaceae, Ajugoideae, the third-largest subfamily of Lamiaceae with 23 genera and about 760 species, predominantly distributed in tropical regions, with a few species endemic to China [2]. The genus has undergone significant taxonomic revisions over time. Initially classified under Verbenaceae by Handel-Mazzetti in 1921, *Schnabelia* is now recognized as a constituent of Lamiaceae [3–5]. Originally, it included only two species, *S. oligophylla* Hand.-Mazz. and *S. tetrodonta* (Y. Z. Sun) C. Y. Wu & C. Chen, both characterized by their 4-winged stems. Later, Cantino et al. (1992; 1999) expanded the genus by transferring three species of *Caryopteris* [3, 4] (*C. aureoglandulosa*, *C. nepetifolia*, and *C. terniflora*) into *Schnabelia*, a classification supported by molecular evidence [1, 2, 6]. They exhibit notable geographical variations while maintaining morphological similarities. The five species of *Schnabelia* exhibit distinct distributions in China. *S. oligophylla* has a wide range, spanning Fujian, Jiangxi, Hunan, Guangdong, Guangxi, and Sichuan provinces, while the endangered *S. tetrodonta* is restricted to the Sichuan Basin, northern Guizhou, and the Jinfoshan area in Chongqing [7]. Morphologically, *S. oligophylla* has five calyx teeth, while *S. tetrodonta* has four. These two species are notable within Lamiaceae for their unique 4-winged stems, setting them apart from the other three species in the genus [2]. *S. nepetifolia* is found mainly in eastern China, including Zhejiang, Jiangsu, Anhui, and Fujian provinces, characterized by solitary flowers in leaf axils. *S. terniflora* has a broader range, extending from central to southwestern China, including regions like Shaanxi, Henan, Hubei, and Sichuan. *S. aureoglandulosa*, primarily found in southwestern Sichuan, Guizhou, and Yunnan, resembles *S. terniflora* but differs by having blunt leaf margins with dense pubescence on young leaves, which later becomes glabrous [8, 9].

In addition to their botanical interest, all five species of *Schnabelia* contain active chemical compounds with medicinal value and are widely used in traditional Chinese medicine [8, 10–12]. For instance, *S. oligophylla* and *S. tetrodonta* are employed for heat clearing, detoxification, blood circulation promotion, dampness removal, wind dispelling, and pain alleviation due to the alkaloids, flavones and saponins they contain [13–15]. Chemical constituents of diterpenoids in *S. terniflora*, *S. nepetifolia*, and *S. aureoglandulosa* have been investigated and demonstrated to possess antitumor and antibacterial activities [8, 10, 11, 16, 17]. Given the complex medicinal properties and morphological similarities of the five *Schnabelia* species, accurate species identification and differentiation are crucial for both scientific research and practical applications in traditional Chinese medicine. The species' overlapping geographical ranges and morphological similarities further emphasize the need for reliable molecular markers to ensure precise identification. Although previous research has mainly focused on morphological classification, molecular systematics, and the medicinal components of *Schnabelia*, phylogenomic studies have been scarce, largely due to the lack of effective genetic markers. Therefore, developing such markers is essential for advancing our understanding of these species.

Previous studies have utilized plastid fragments such as *rbc*L, *ndh*F, *mat*K, and nuclear ITS to resolve the systematic placement and classification of *Schnabelia* [1, 2, 4, 6]. Cladistic analysis using morphological data and plastid gene sequences *rbc*L and *ndh*F by Cantino et al. (1999) firstly challenged the monophyly of *Caryopteris* and transferred three *Caryopteris* species to *Schnabelia* [4]. Similar conclusions were obtained by Huang (2002) using *ndh*F [6]. Shi et al. (2003), for the first time, used both the plastid fragment *mat*K and nuclear ITS sequences, reconfirming that *S. oligophylla* and *C. terniflora* formed one clade [1]. However, the sampling coverage of the above phylogenetic study did not include all *Schnabelia* species. Until 2018, Xiang's (2018) phylogenetic study of the subfamily Ajugoideae, which used two nuclear genes (ITS and ETS) and five plastid fragments (*mat*K, *rbc*L, *rps*16, *trn*L-*trn*F, *trn*H-*psb*A), provided the first comprehensive revelation of *Schnabelia*, and the results showed that the expanded *Schnabelia* is monophyletic, and consists of two morphologically distinct clades, recognized as sect. *Cylindricaulis* and *Schnabelia* [2]. In recent years, phylogenomic studies using chloroplast genomes have focused on the delimitation of subfamilies and tribes under the family Lamiaceae, with limited sampling of *Schnabelia* species. For example, Li et al. (2016) proposed three new subfamilies within Lamiaceae according to the phylogeny results based on plastome data, using only *S. oligophylla* [18]. Zhao et al. (2021) updated the tribes within the 12 subfamilies of Lamiaceae, and only *S. oligophylla* was analyzed [19]. There have been no studies on the phylogeny of the *Schnabelia* using plastome data, even though complete chloroplast genomes are necessary to fully characterize the phylogenetic relationships of species within the genus.

Considering classification challenges, plastomes offer a promising approach for resolving the phylogenetic relationships and identifying genetic markers for species classification. By examining the plastomes in detail, we aim to deepen our understanding of the evolutionary dynamics within *Schnabelia* and establish a foundation for more comprehensive molecular studies. Plastids

serve as crucial organelles in plant cells, housing an independent genome encoding a set of genes associated with photosynthesis and carbon fixation [20]. In angiosperms, the plastomes typically exhibit a quadripartite structure, comprising a large single-copy region (LSC) and a small single-copy region (SSC) separated by two inverted repeat regions (IRa and IRb) [21]. Angiosperm plastomes consist of 120–130 genes, varying in size from 115 to 165 kb across different species [22]. Compared to nuclear genomes, plastomes are relatively conserved and inherited uniparentally, offering an effective tool for plant identification and phylogenetics, particularly among closely related taxa [23, 24]. However, a more robust phylogenetic study based on the entire plastome has yet to be conducted for *Schnabelia*. Additionally, the complete plastome sequence harbors abundant genetic information, including molecular markers for subsequent barcode development, systematic evolution, and population genetic studies.

To date, only the plastome of *S. tetrodonta* is published, which reported the structure and characterization of the plastid genome [25]. No comprehensive studies have been conducted focusing on plastomic comparative analysis, including genome structure, gene content, and genomic variation among *Schnabelia* species. In this study, we sequenced the plastomes of five *Schnabelia* species, conducted comparative analyses, and constructed a phylogenetic tree incorporating closely related taxa in subfamily Ajugoideae, as well as conducting divergence time estimation. Our study aims to: (1) elucidate the characteristics of *Schnabelia* plastomes, including gene content, IR variation, codon usage patterns, and selective pressure; (2) identify rapidly evolving plastid fragments for potential barcoding markers and application in population genetics analyses; (3) fully clarify the phylogenetic relationships within Ajugoideae; and (4) estimate the divergence time of *Schnabelia* species, and infer the possible influence of geological and historical climate on the species differentiation. Our study will contribute to the expansion of the genomic resources available for *Schnabelia* and provide critical information to support the identification and phylogenetic analysis of *Schnabelia* species. Moreover, the research findings will also have practical applications for the population genetics and conservation study of *Schnabelia*.

Materials and methods

Plant materials

The five *Schnabelia* species were collected from the following locations: Chenshan Botanical Garden in Shanghai (*S. oligophylla*); Qingchengshan in Chengdu, Sichuan Province (*S. tetrodonta*); Tianmushan in Lin'an, Zhejiang Province (*S. nepetifolia*); Maoxian County in Sichuan Province (*S. terniflora*); and Jinfoshan in Chongqing (*S. aureoglandulosa*).Plant specimens were identified by Dr. Pan Li of Zhejiang University, and the Voucher specimens were deposited in the Zhejiang University Herbarium (HZU, Hangzhou, China) with Accession number LP186180, LP 011579, LP161315, LP20120124 and LP186194, corresponding to the order above. For each specimen, 2–3 fresh leaves were collected and dried using silica gel.

DNA extraction, sequencing, assembly and annotation

Total genomic DNA was extracted from dried leaves using DNA Plantzol Reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer's protocol. The quality and concentration of DNA were assessed using agarose gel electrophoresis. Paired-end sequencing libraries were constructed according to the Illumina standard protocol (Illumina, San Diego, CA, USA) and sequenced on the Illumina HiSeq2500 platform at the Beijing Genomics Institute (BGI, Shenzhen, China). Raw reads were processed with Fastp v 0.23.2 to remove adapters and low-quality bases ($Q < 20$ over a 10-bp sliding window), retaining reads with $> 80\%$ high-quality bases and $< 5\%$ ambiguous nucleotides [26]. The obtained clean reads were de novo assembled using the GetOrganelle program for each specimen [27]. Subsequently, the quadripartite structure was validated and examined using Bandage [28]. The assembled plastomes were annotated using CPGAVAS2 (http://47.96.249.172:16019/analyzer/annotate, accessed on 10 August 2023), with the published plastome sequence of *S. oligophylla* as a reference. Final annotations were cross-validated using GeSeq [29]. Circular maps illustrating the plastome structure were generated using OGDRAW [30]. The plastome data for the five *Schnabelia* species were submitted to GenBank under accession numbers PP503204-PP503208, respectively.

Repeat and SSRs analyses

Repeat sequences in the plastomes, including forward repeats, reverse repeats, palindromic sequences, and complement repeats, were identified using REPuter (https://bibiserv.cebitec.unibielefeld.de/reputer, accessed on 17 August 2023). The parameter settings were a Minimal Repeat Size of 30, Hamming Distance of 3, and Maximum Computed Repeats of 100. Simple Sequence Repeats (SSRs), consisting of small repetitive sequences of 1–6 bp, were analyzed using MISA [31] (http://pgrc.ipk-gatersleben.de/misa, accessed on 18 August 2023), with the following parameter settings for single nucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide SSRs: 10, 6, 5, 5, 5, and 5, respectively.

Codon usage bias analyses

Codon usage bias (CUB) refers to the uneven usage frequency of synonymous codons encoding the same amino acid, which varies among species [32, 33]. The codon usage bias parameters, including ENC (effective number of codons), RSCU (relative synonymous codon usage), and GC3 (GC content at the third position of codons), were estimated using the CHIPS and CUSP sections in the EMBOSS online website (https://codonw.sourceforge.net/, accessed on 12 August 2023) and CodonW V1.4.2 (https://codonw.sourceforge.net/, accessed on 15 August 2023). ENC reflects the deviation of codon usage from random selection, while RSCU represents the relative probability of specific codon usage among synonymous codons. GC3 was used as a parameter of codon preference due to lower selection pressure on the third position of codons compared to the first two positions.

Selective pressure analyses

To quantify selective pressure, we used the Ka-Ks module in the online tool developed by Nanjing Gene Pioneer Biotechnologies Inc. (http://cloud.genepioneer.com:9929/#/tool/alltool?type=4, accessed on 20 August 2023) to calculate non-synonymous (Ka) and synonymous (Ks) nucleotide substitutions and the Ka/Ks ratio ω for each plastid gene. Values of $\omega = 1$, $\omega > 1$, and $\omega < 1$ indicate neutral, positive, and purifying selection, respectively [34].

Sequence divergence analysis and nucleotide diversity analyses

Global comparison of the five *Schnabelia* plastomes was conducted using the online tool mVISTA [35] (Berkeley, CA, USA) (https://genome.lbl.gov/vista/mvista/submit.shtml, accessed on 24 August 2023) with the Shuffle-LAGAN mode and default parameters for sequence alignment. Polymorphic hotspot regions in the plastomes of *Schnabelia* were identified using DnaSP v6.0 [36]. Using *S. oligophylla* as the reference, a sliding window approach was applied with a step size of 200 bp and a window length of 600 bp to analyze nucleotide polymorphism values (Pi) [37].

Phylogenetic analyses

A total of 34 plastomes from 8 genera in subfamily Ajugoideae were obtained for phylogenetic tree reconstruction. *Colquhounia vestita* (NC_058331) and *Cymaria dichotoma* (NC_058333) were used as outgroups. Sequences were aligned using MAFFT V7 [38], and phylogenetic analyses were conducted using both Maximum likelihood (ML) and Bayesian inference (BI) methods based on whole plastomes. ML analysis was performed using IQ-TREE V.1.6.8 [39] with the GTR + F + R4 nucleotide substitution model and 1,000 bootstrap replicates. BI analyses were performed using MrBayes v3.2 [40] with the nucleotide substitution model TVM + I + G. Posterior probabilities were estimated using Markov Chain Monte Carlo (MCMC) chains (20 million generations and sampled every 1,000 generations), with the first 25% of sampled trees discarded as burn-in.

Divergence time estimation

Currently, no fossil records of *Schnabelia* have been reported. We selected two fossils as calibration points and modeled them with a log-normal distribution [41]. The first, *Ocimum* pollen fossils from the Early Eocene of India, were placed in the crown group node of the subfamily Nepetoideae [42, 43]. A minimum age of 47.8 MYA and a maximum age of 107 MYA were set, with a mean of 2.6 and a standard deviation (SD) of 0.5. The second calibration point was the *Stachys laticarpa* fossil from Germany dated back to the Middle Miocene [44, 45], which was used to constrain crown group of *Stachys*. A calibration age of 13.8–11.6 Ma was set, with a mean of 1.5 and an SD of 0.5.

The shared protein-coding sequences (CDS) of 27 species, representing 12 subfamilies of Lamiaceae, were extracted with Python and aligned to be 60,921 bp using MAFFT V7.520 [38], and phylogenetic trees were constructed using IQ-TREE V1.6.8 [39] with the GTR + F + R3 model. Divergence times were estimated by MCMCTree V4.8 [46] using an independent rates clock model (clock = 2), the HKY85 substitution model and a gamma rate model. Samples were collected every 100 generations for a total of 100,000 generations, with 20% of the initial samples discarded as burn-in. Tracer software was used to confirm an effective sample size (ESS) > 200 [41]. The phylogenetic tree and divergence times were visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/) and Chiplot (www.chiplot.online, accessed 24 November 2024).

Results

Plastome structure and features

Sequencing of the plastomes of five *Schnabelia* species yielded raw data ranging from 1.61 G (*S. terniflora*) to 30.4 G (*S. oligophylla*). Following assembly, the size of these genomes ranged from 155,733 bp for *S. nepetifolia* to 156,944 bp for *S. oligophylla*. Each plastome comprises four regions: the large single-copy region (LSC), small single-copy region (SSC), and two inverted repeat regions (IRa and IRb) (Fig. 1). LSC lengths varied from 82,652 bp (*S. nepetifolia*) to 83,756 bp (*S. aureoglandulosa*), while SSC lengths ranged from 9,948 bp (*S. aureoglandulosa*, *S. terniflora*) to 9,959 bp (*S. tetrodonta*). The IR region spanned from 31,563 bp (*S. nepetifolia*) to 31,710 bp (*S. tetrodonta*, *S. oligophylla*). The GC content across all species was consistent at 37.8%, with the IR region exhibiting

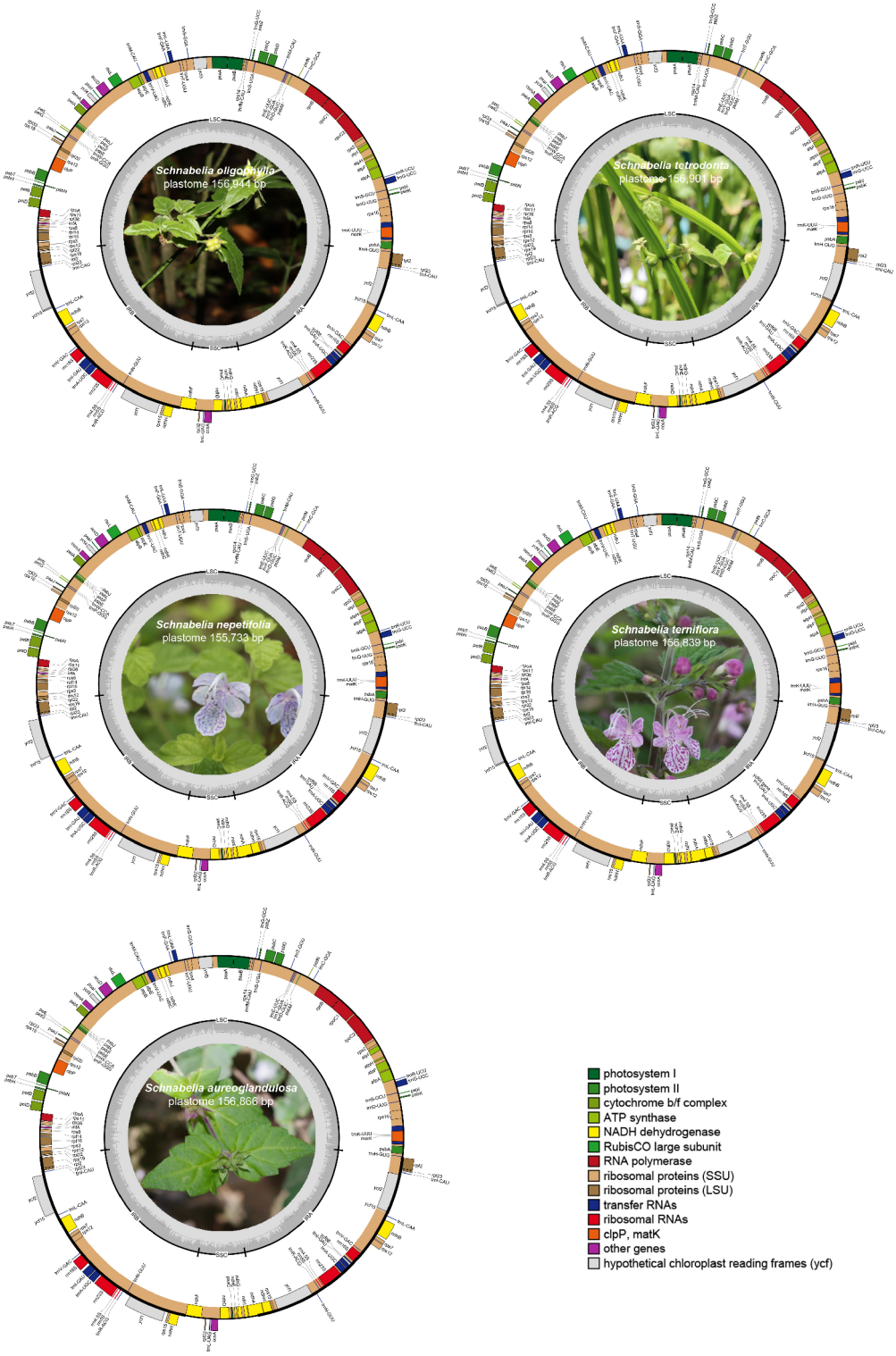

Fig. 1 Plastome maps of five *Schnabelia* species. The genes transcribed clockwise are shown inside of the circle while genes transcribed anti-clockwise are shown outside. The function of genes is color-coded. The dark grey color shows the GC content whereas the light grey color indicates AT content

Table 1 The basic characteristics of five *Schnabelia* plastomes

Species name	Genome Size (bp)	LSC length (bp) (GC content)	SSC length (bp) (GC content)	IR length (bp) (GC content)	Gene	PCGs	rRNA	tRNA	Duplicated genes	GC content (%)
S. oligophylla	156,944	83,567 (36.2%)	9,957 (32.8%)	31,710 (40.8%)	115	81	4	30	21	37.8
S. tetrodonta	156,901	83,522 (36.2%)	9,959 (32.7%)	31,710 (40.8%)	115	81	4	30	21	37.8
S. nepetifolia	155,733	82,952 (36.2%)	9,955 (32.6%)	31,563 (40.8%)	115	81	4	30	21	37.8
S. terniflora	156,839	83,627 (36.1%)	9,948 (32.6%)	31,632 (40.8%)	115	81	4	30	21	37.8
S. aureoglandulosa	156,866	83,756 (36.1%)	9,948 (32.6%)	31,581 (40.8%)	115	81	4	30	21	37.8

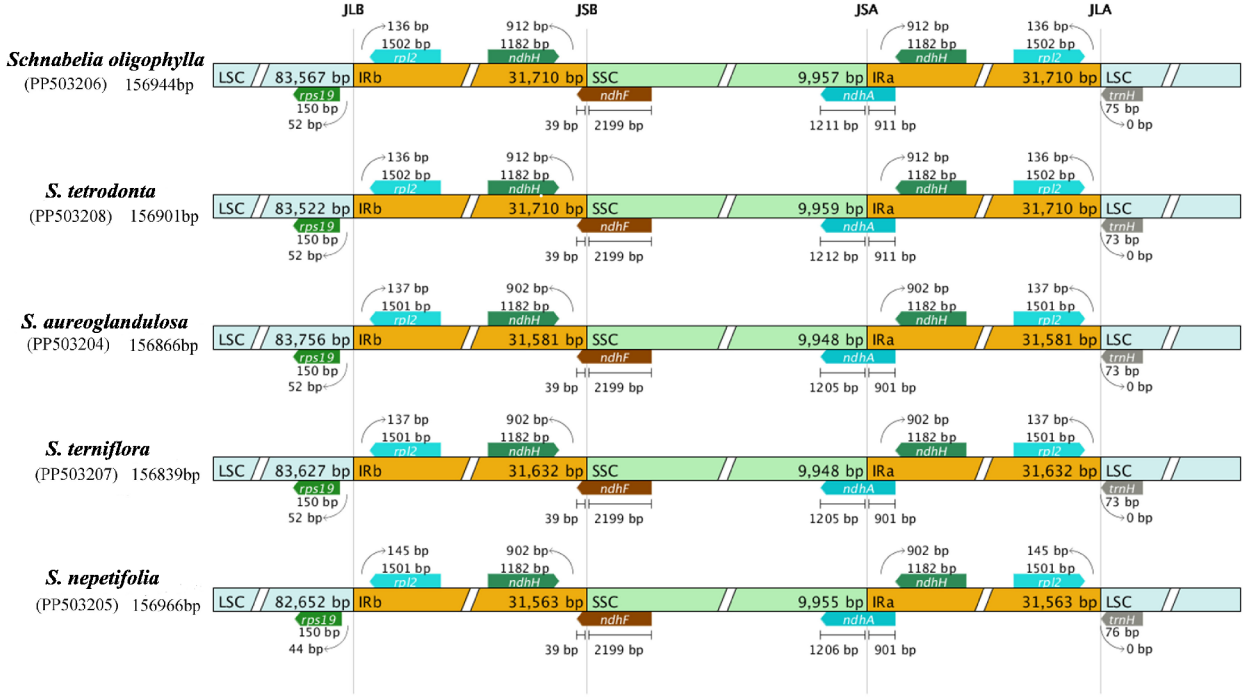

Fig. 2 Comparison of the border positions of LSC, SSC, and IR regions in the plastomes in five *Schnabelia* species. JLB, LSC/IRb junction; JSB, SSC/IRb junction; JLA, LSC/IRa junction; JSA, SSC/IRa junction

a higher GC content (40.8%) compared to the LSC (36.1–36.2%) and SSC (32.6–32.8%) regions (Table 1).

The plastomes of *Schnabelia* species demonstrated a high conservation in gene content and order. All five species shared 115 unique genes, comprising 81 CDS genes, 4 ribosomal RNA genes (rRNA), and 30 transfer RNA genes (tRNA). Among these, 21 genes were duplicated in the IR regions, including 10 CDSs, 4 rRNAs, and 7 tRNAs (Fig. 1). Nine protein-coding genes (*ndh*A, *ndh*B, *pet*B, *pet*D, *atp*F, *rpl*16, *rpl*2, *rps*16, *rpo*C1) contained one intron, while six tRNA genes (*trn*A-*UGC*, *trn*G-*UCC*, *trn*I-*GAU*, *trn*K-*UUU*, *trn*L-*UAA*, *trn*V-*UAC*) and three genes (*ycf*3, *clp*P, *rpl*12) contained two introns. Additionally, partial genes *ndh*F and *ndh*A formed at the borders of IR regions due to IR duplication (Table S1).

Contraction and expansion of inverted repeats

Variations were evident at the IR/LSC and IR/SSC borders across the five *Schnabelia* plastomes (Fig. 2). The LSC/IRb junction (JLB) was positioned between the *rps*19 and *rpl*2 genes, with a consistent distance of 52 bp from the boundary to *rps*19 in all species, except *S. nepetifoli*a (44 bp), while the distance from the boundary to *rpl*2 ranged from 126 bp to 145 bp. The *ndh*F gene spanned the IRb/SSC junction (JSB), covering 2,199 bp in the SSC region and 39 bp in the IRb region, forming a partial gene in IRa. The distance from JSB to *ndh*H was 912 bp in *S. oligophylla* and *S. tetrodonta*, and 902 bp in the other species. The *ndh*A gene at the SSC/IRa junction (JSA) had 905/911 bp in the IRa region and 1,205–1,112 bp in the SSC. The adjacent boundary to *trn*H (JLA) possessed a distance of 136–145 bp to the *rpl*2 gene, consistent with the IRb (136 bp in *S. oligophylla* and

S. tetrodonta; 145 bp in *S. nepetifolia*; 137 bp in *S. terniflora* and *S. aureoglandulosa*) (Fig. 2).

Repeat sequence analyses and SSRs

Using REPuter analysis, three types of repeat sequences were identified, with 446 repeats detected, comprising 234 forward repeats, 209 palindromic repeats, and 3 reverse repeats (Table S2). Forward and palindromic sequences were the most common. *S. aureoglandulosa* exhibited the highest number of forward repeats (42), while *S. oligophylla* displayed the most palindromic sequences (40). Repeat sequences of 30–39 bp predominated, accounting for 59.75% across the five species (Fig. 3). MISA analysis identified 171 SSRs, with *S. terniflora* harboring the most (38) and *S. tetrodonta* the fewest (31). A/T repeats were most abundant, constituting 86.55% of the total, with 93.57% of SSRs being mononucleotide repeats. SSRs were predominantly located in the intergenic spacer (IGS) region (56%), followed by CDSs (22%) (Fig. 4; Table S3). Distribution analysis revealed 120 SSRs in the LSC region, 5 in the SSC region, and 46 in the IR region.

Codon usage analyses

The average codon numbers for *Schnabelia* species ranged from 51,911 to 52,314, with effective codon numbers ranging from 55.28 (*S. tetrodonta*) to 56.01 (*S. aureoglandulosa*). Leucine was the most prevalent amino acid (~9.80%), while tryptophan was least abundant (~1.38%). Relative synonymous codon usage (RSCU) analysis revealed variations across species, with *S. oligophylla* and *S. terniflora* displaying 29 codons with RSCU values greater than 1, while *S. aureoglandulosa*, *S. nepetifolia*, and *S. tetrodonta* exhibited 31. The AGA codon, encoding arginine (Arg), had the highest RSCU value in all species (Fig. 5; Table S4). Plastid genes had enriched A/T bases at the third codon position compared to G/C, with

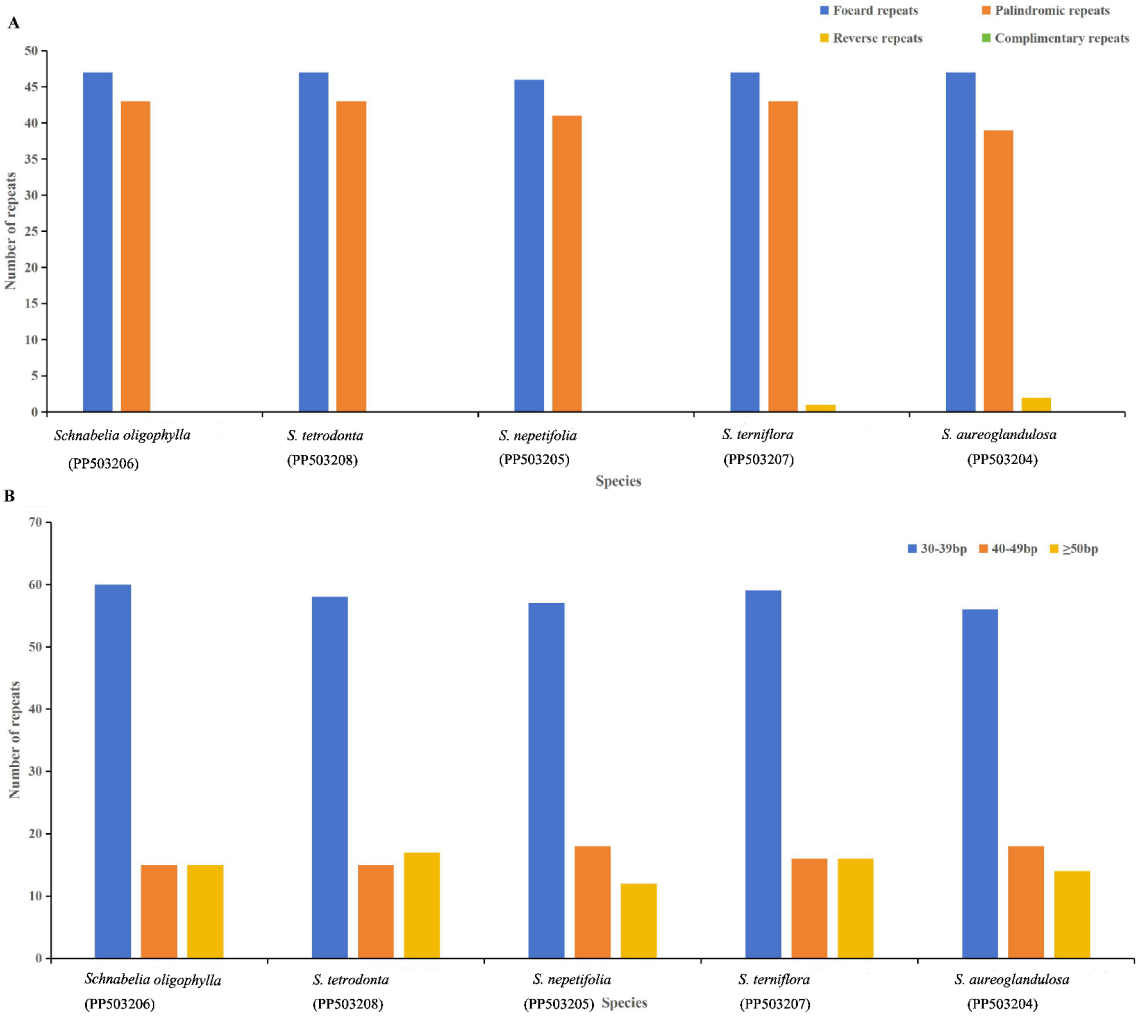

Fig. 3 Repeat analysis in five plastomes of *Schnabelia*. (**A**) Frequency of repeat types. (**B**) Frequency of repeats by length

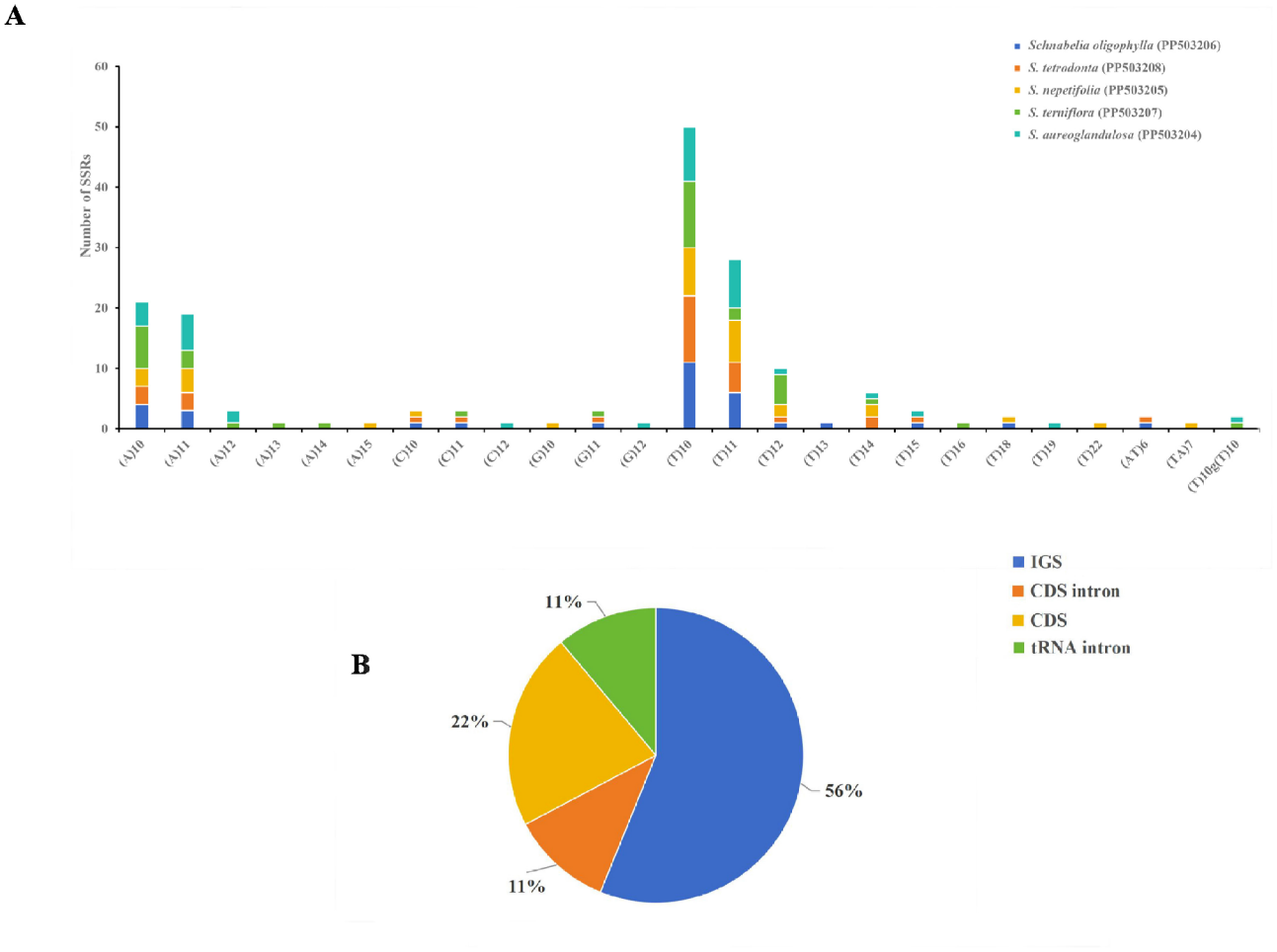

Fig. 4 Analysis of simple sequence repeats (SSRs) in the five *Schnabelia* plastomes. (**A**) Number of SSRs by length. (**B**) Distribution of SSR loci. CDS, coding DNA sequences; IGS, intergenic spacer region

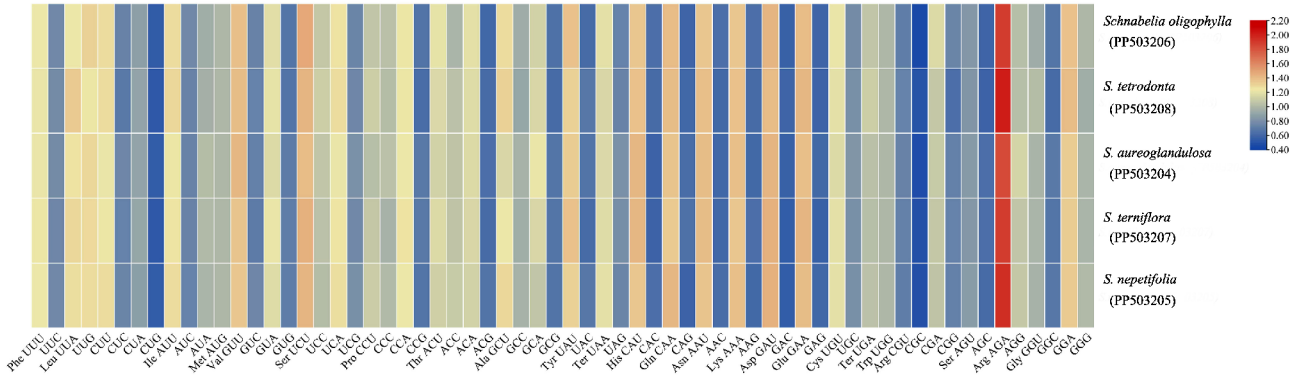

Fig. 5 Heatmap of codon usage bias in five *Schnabelia* plastomes based on Relative Synonymous Codon Usage (RSCU) values. The color gradient (red to blue) represents normalized RSCU values, with darker red indicating stronger bias

the G/C content ranging from 35.6% (*S. tetrodonta*) to 36.5% (*S. aureoglandulosa*).

Selective pressure analyses

Ka, Ks, and Ka/Ks ratio values were computed for the genes of the five plastomes (Table S5). For the 81 protein-coding genes, the highest Ka/Ks ratio of 0.84462 was detected in *S. terniflora*. None of the protein-coding genes had a Ka/Ks ratio greater than 1, indicating that all genes in these plants have undergone purifying selection (Table S5; Fig. 6).

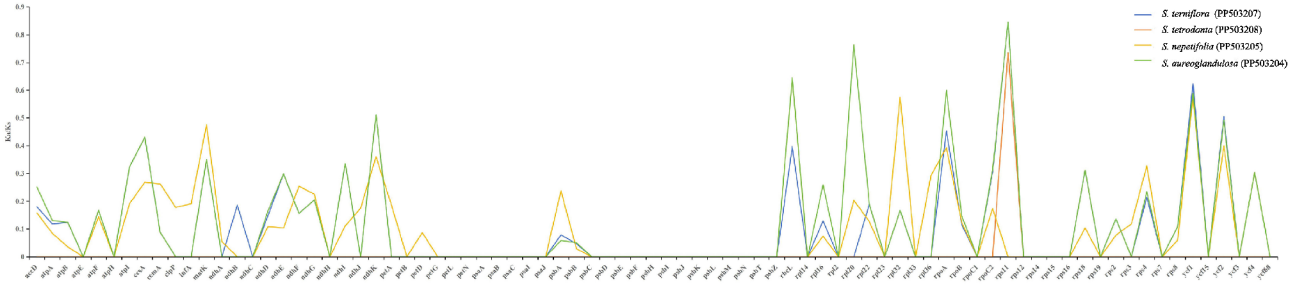

Fig. 6 The Ka/Ks value of the 81 CDS regions in the five *Schnabelia* plastomes

Fig. 7 Sequence identity plots among the five *Schnabelia* plastomes with *S. oligophylla* as a reference. Annotated genes are displayed along the top. The vertical scale represents the percent identity between 50 and 100%

Sequence polymorphism analyses

The mVISTA analysis provided insights into the global variation among the five *Schnabelia* plastomes, utilizing *S. oligophylla* as the reference genome. Results indicated a degree of similarity among these species, with no discernible evidence of gene rearrangements. Most sequence variations were concentrated in the large single-copy (LSC) and small single-copy (SSC) regions, while the inverted repeats (IRs) were comparatively more conserved (Fig. 7). Nucleotide polymorphism values (Pi), obtained from DnaSP v6.0, facilitated the identification of divergence hotspots among the five individuals, ranging from 0 to 0.045 (Table S6). Eleven divergence hotspots (Pi > 0.022) were delineated, including *trn*H-*GUG-psb*A, *trn*K-*UUU-mat*K, *trn*K-*UUU*, *trn*G-*UCC* intron, *psb*M, *trn*M-*CAU*, *trn*L-*UAA* intron, *pet*B-*pet*D,

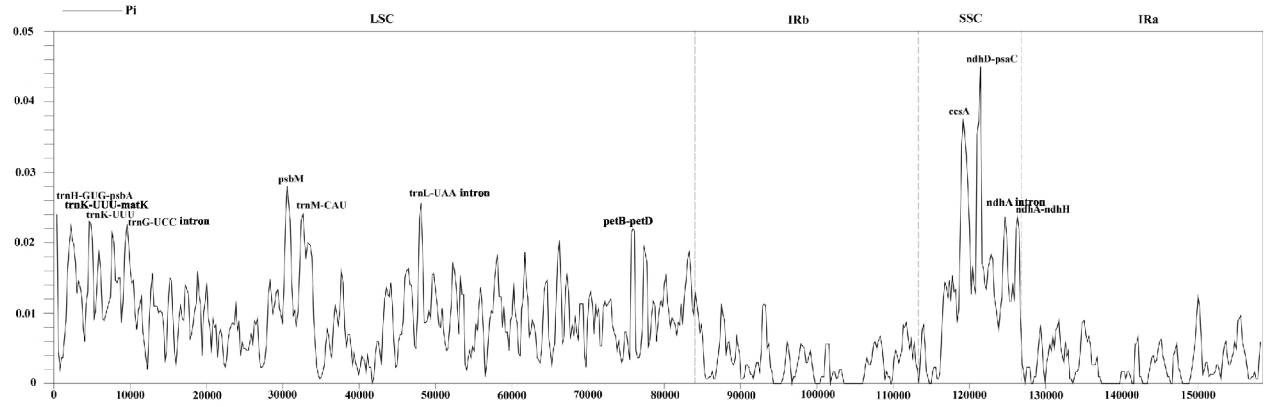

Fig. 8 The nucleotide variation (Pi) values of five *Schnabelia* plastomes were compared, with nucleotide diversity at the midpoint of the window on the x axis and within each window on the y axis (window length of 600 bp, step length of 200 bp)

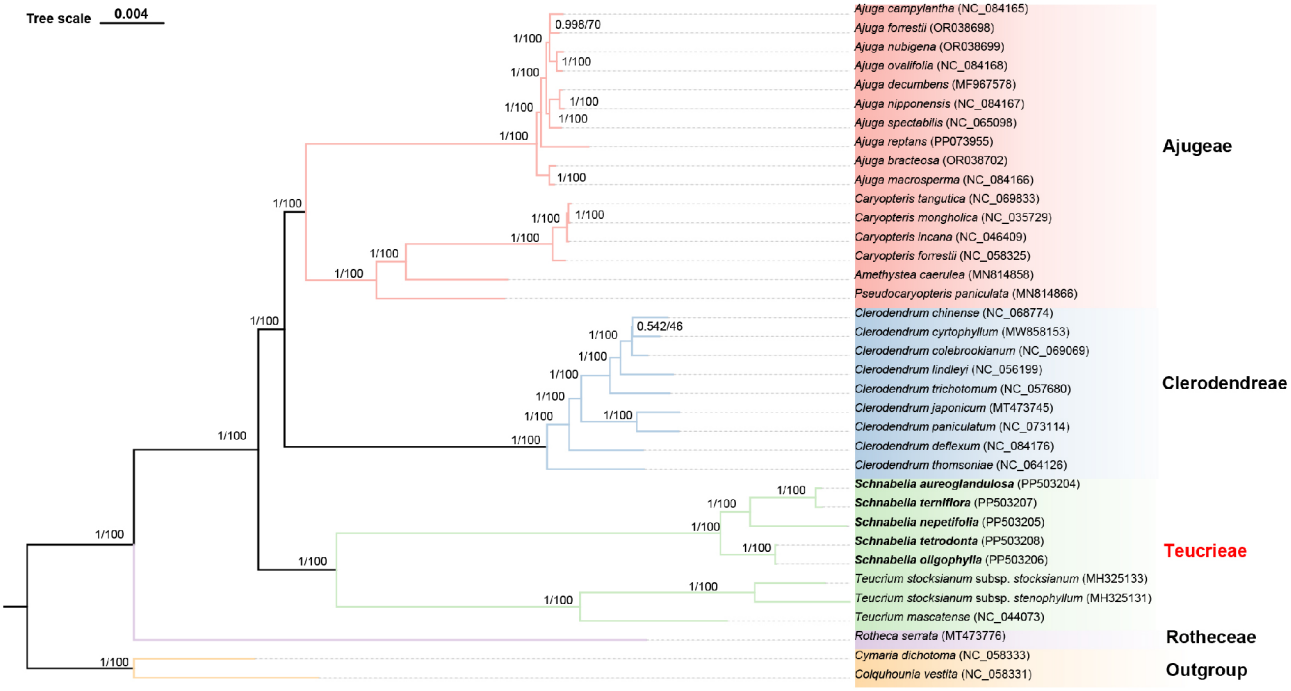

Fig. 9 Phylogenetic tree of Ajugoideae based on 34 plastomes inferred using the Maximum Likelihood (ML) and Bayesian Inference (BI) method. Support values were labeled on each branch as PP (posterior probability)/BS (bootstrap support)

*ccs*A, *ndh*D-*psa*C, *ndh*A intron, and *ndh*A-*ndh*H (Fig. 8). Among these, five IGS regions (*trn*H-*GUG*-*psb*A, *trn*K-*UUU*-*mat*K, *pet*B-*pet*D, *ndh*D-*psa*C, *ndh*A-*ndh*H) could prove valuable in the development of genetic markers. Our nucleotide polymorphism analysis indicated that high-variation hotspots are primarily situated in the SSC region.

Phylogenetic analyses

A phylogenetic tree encompassing 34 species of the subfamily Ajugoideae was reconstructed using two methods: maximum likelihood (ML) and Bayesian inference (BI). Phylogenetic trees generated by both methods exhibited congruent topology and garnered robust support values. Bootstrap (BS) values and Bayesian posterior probabilities (PP) were assigned to each node (Fig. 9). The subfamily Ajugoideae segregated into four main clades, representing four tribes, Ajugeae, Clerodendreae, Teucrieae and Rotheceae. The phylogenetic analysis demonstrated that Ajugeae initially formed a monophyletic clade with Clerodendreae (PP/BS = 1/100), which subsequently merged with Teucrieae to establish a higher-level lineage (PP/BS = 1/100). This expanded group ultimately grouped with Rotheceae, completing

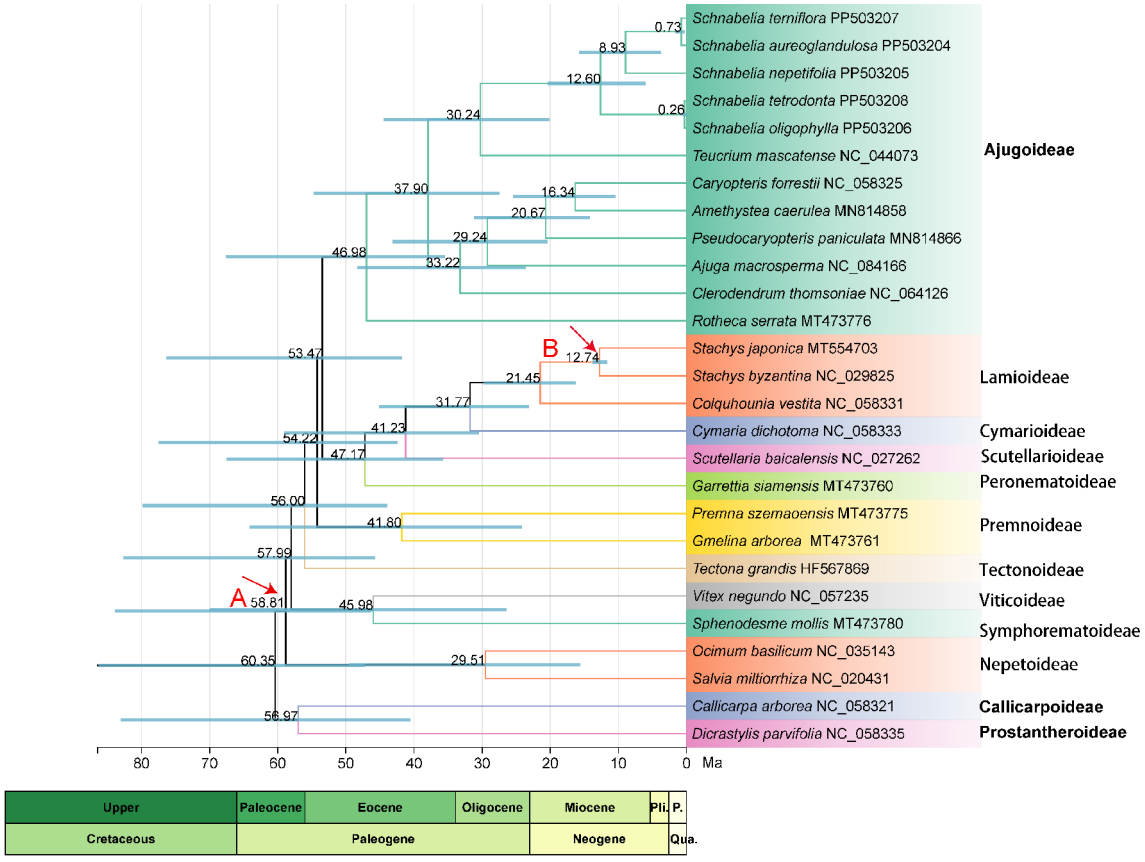

Fig. 10 Divergence time of *Schnabelia* species estimated using plastid CDS sequences. Mean divergence time (Ma) and 95% high posterior density (HPD) are shown at the branches. Arrows mark two calibration points. Node A and Node B represent the fossil constraints of subfamily Nepetoideae and genus *Stachys*

a fully resolved monophyletic topology with maximal statistical support (PP/BS = 1/100) across all nodes. The tribe Teucrieae comprises *Teucrium* and *Schnabelia*. *Schnabelia* forms a monophyletic clade and shares the closest relationship with *Teucrium* within the sampled extent. Within *Schnabelia*, two clades were discerned: one comprising *S. oligophylla* and *S. tetrodonta*, the two original species based on traditional classification, and the other clade encompassing three species transferred from *Caryopteris*. Within *Teucrium*, the two subspecies of *T. stocksianum* and *T. mascatense* formed a monophyletic clade. While Rotheceae was represented by a single taxon in our sampling, Clerodendreae formed a strongly supported monophyletic clade (PP/BS = 1/100). The tribe Ajugeae segregated into four distinct genus-level lineages (*Ajuga*, *Caryopteris*, *Amethystea*, and *Pseudocaryopteris*). A strongly supported sister relationship (PP/BS = 1/100) was recovered between *Caryopteris* and *Amethystea*, which subsequently formed a sister clade with *Pseudocaryopteris* (PP/BS = 1/100). This tripartite lineage ultimately emerged as sister to *Ajuga* with maximal nodal support (PP/BS = 1/100). While *Amethystea* and *Pseudocaryopteris* were each represented by single taxa in our analysis, *Caryopteris* and *Ajuga* comprised multiple accessions. Both multi-sampled genera formed strongly supported monophyletic groups (*Caryopteris*: PP/BS = 1/100; *Ajuga*: PP/BS = 1/100).

Fossil-calibrated molecular dating

The divergence time of *Schnabelia* was estimated based on plastid CDS sequences (Fig. 10). The split between *Schnabelia* ancestor and its sister genus (*Teucrium*) was dated to the early Oligocene, approximately 30.24 million years ago (Ma) (95% highest posterior density (HPD): 19.03–42.78 Ma). The crown age of *Schnabelia* was inferred to be around 12.60 Ma (95% HPD: 5.52–19.70 Ma). Within Sect. *Cylindricaulis*, the divergence time between *S. nepetifolia* and (*S. aureoglandulosa* + *S. terniflora*) is estimated to be during the Late Miocene to Pliocene, approximately 8.93 Ma (95% HPD: 3.47–15.35 Ma), the diversification within Sect. *Schnabelia* was dated back to 0.26 Ma (95% HPD: 0.042–0.50 Ma).

Discussion

China possesses high species diversity and endemism [47]. *Schnabelia*, a genus endemic to China, contain five species each with distinct distribution, hinting at China's role as a "museum" or "cradle" [48]. However, there are few studies on the evolutionary history of *Schnabelia*, and the available molecular markers and genetic resources are limited. All five species of the *Schnabelia* have medicinal value. Recent studies of the bioactivities of crude polysaccharides from leaves and flowers of *S. terniflora* showed promising value in cosmetics industry [49]. Climatic changes, as well as anthropogenic collection for medicinal use, pose a potential threat to their populations, especially those of limited-distribution species, such as *S. tetrodonta* and *S. aureoglandulos*. Here, we sequenced the chloroplast genome of five *Schnabelia* species, revealed their characteristics through comparative analysis, and used plastome data for evolutionary analyses. The polymorphic loci that we obtained will provide useful tools for population genetics and conservation study of *Schnabelia* in the future.

Plastome features

While plastomes are commonly utilized for plant classification, population genetics, and phylogenetic analyses, research on the plastomes of *Schnabelia* is still scarce [25], with no comparative genomic analysis conducted on the genus. In this study, we sequenced and assembled genomes of all five *Schnabelia* species (*S. oligophylla*, *S. tetrodonta*, *S. nepetifolia*, *S. terniflora*, and *S. aureoglandulosa*) and compared the plastomic features and variations among them. While gene loss may occur in some plants to adapt to specific environments or changes in nutritional modes [50, 51], no gene rearrangements or losses were detected in *Schnabelia*, indicating a high degree of conservation within the genus, likely due to their similar morphological characteristics and habitat preferences.

The overall GC content (37.8%) of the five plastomes is nearly identical, which closely aligns with values reported in other Lamiaceae plastomes [52–54]. The SSC region consistently exhibits the lowest GC content, which can contribute to significant plastome structure variability [55]. The GC content of the IR region was higher than that of the LSC and SSC regions, a phenomenon commonly observed in other plants [56]. The length of the IR regions was relatively conserved and varied between 31,563 bp and 31,710 bp across the five plastomes. Although the contraction and expansion of the IR region leading to length variation has been detected in many angiosperm plastomes [57, 58], the total length variation due to IR change was not obvious in *Schnabelia*.

Repetitive sequences significantly correlated with gene rearrangements, recombination, or inversion in plastomes [59]. In *Schnabelia*, three types of repeats were identified at similar levels: forward, reverse, and palindromic, with forward repeats being the most abundant and no gene rearrangements detected. Simple sequence repeats (SSRs) are widespread in the genomes of eukaryotic organisms [60], and SSRs, as molecular markers, play an important role in population genetics analysis, such as population genetic structure and dynamic history inference [61–63]. In this study, a total of 171 SSRs were identified in the five genomes. Most SSRs were found in the IGS region of the LSC region. Among the detected SSRs, A/T-type mononucleotide repeats were the most abundant, consistent with other plant plastomes [64]. These SSRs could provide valuable information for detecting polymorphisms within and among *Schnabelia* species in future population genetics studies.

Coding sequences are more conserved and are often used in phylogenetic analyses above the genus level [21, 65]. In our study, genetic variation mainly occurred in non-coding regions, which are widely used to investigate taxonomy and molecular phylogeny at the interspecific level [23, 66–68]. Sliding window analysis using DnaSP revealed 5 highly variable IGS regions, including *trn*H-*GUG*-*psb*A, *trn*K-*UUU*-*mat*K, *pet*B-*pet*D, *ndh*D-*psa*C, *ndh*A-*ndh*H. Insertion/deletion is prevalent in *trn*H-*GUG*-*psb*A [69], and researchers have used this region to study closely related genera and species such as *Corythophora* (Lecythidaceae) [70], Saxifragaceae [71]. In addition, *trn*K-*UUU*-*mat*K was also commonly used for phylogenetic analysis [72]. These regions could be targeted as DNA barcode development to assist with identification of morphologically similar *Schnabelia* species (e.g. *S. terniflora* and *S. aureoglandulosa*) or after processing of plant material for medicinal use. Extensive research is focused on developing DNA barcoding from plastomes for identifying medicinal plants. For example, Guo et al. (2023) identified 8 highly divergent hotspot regions from the plastome for the medicinal herb *Scrophularia ningpoensis* and its common adulterants discrimination [73]. *S. oligophylla* and *S. tetrodonta* are among the eight most effective medicines in the folklore of Jinfoshan, Chongqing, but long-term over-exploitation has led to the decline of wild resources [74], and it is proposed that they be listed as Grade III wild endangered protected plants [75]. To promote conservation and sustainable utilization of plant resources, future studies could also use plastid variation hotspot regions (e.g., *trn*H-*GUG*-*psb*A, *trn*K-*UUU*-*mat*K, *ndh*A-*ndh*H) to assess the genetic diversity of *Schnabelia* populations in fragmented habitats. These data can inform prioritization for in situ conservation, especially for endemic species with high medicinal potential but limited distribution, such as *S. tetrodonta* and *S. oligophylla*.

Codon usage bias is likely a result influenced by selection and mutation, which can help us better understand the mechanisms of gene evolution [76]. More than 90% of the codons typically terminate with A and/or T (U), exhibiting high RSCU scores across the five plastomes. The AGA codon encoding arginine (Arg) has the highest RSCU value in all five species. GC3s is correlated with codon bias to evaluate codon usage patterns [77]. The GC3s values ranged from 35.6 to 36.5% in *Schnabelia*, indicating a strong bias towards A/U-ending codons. Previous research showed that CUB of the plastid genes is related to organisms' gene expression level and adaptability to the environment [78, 79]. During the evolutionary process of plastomes, most genes undergo purifying selection, while some genes experience positive selection due to environmental adaptation [80]. In selective pressure analysis, all the Ka/Ks ratios of 81 protein-coding genes were identified as less than 1 across the five *Schnabelia* species, suggesting that despite the different distribution areas among the five species, there hasn't been significant adaptive evolution in the plastid genes. All plastid genes appear to be essential in their current habitats and life forms, subject to strong purifying selection.

Phylogenetic relationships

Previous studies have examined the taxonomic position of *Schnabelia* from morphology and anatomy [81, 82]. However, due to the similarity of *Schnabelia* to certain genera within the Verbenaceae and Lamiaceae, its taxonomic status was once ambiguous [4]. For example, *Schnabelia* shows similarities to *Ajuga* and *Teucrium* from Lamiaceae in pollination, ovary cleavage and axillary solitary flower structure. At the same time, its calyx characteristics are much closer to *Caryopteris* from Verbenaceae [83]. Although the phylogeny and classification of *Schnabelia* have been determined by researchers using plastid gene fragments [1, 2], the use of whole chloroplast genome sequences to ultimately confirm their robust evolutionary relationships is still necessary. Nowadays, Plastomes have been extensively employed for reconstructing phylogenetic relationships instead of plastid fragments in recent years [84]. Phylogenetic studies of Lamiaceae using plastomes have been conducted for subfamily and clade circumscription, but only *S. oligophylla* was analyzed [18, 19]. Our present study focused on *Schnabelia* and included all the species to reconfirm their phylogenetic relationships. Our phylogenetic tree yielded by ML and BI methods displayed the same topology consistent with previous research [19], with all clades being almost fully supported. Our results indicated that the subfamily Ajugoideae contained four main clades, representing four tribes of Ajugeae, Clerodendreae, Teucrieae and Rotheceae recently updated by Zhao et al. (2021). Our results show that *Schnabelia* is encompassed within the tribe Teucrieae, with the species being most closely related to *Teucrium* within the present sampling extent. However, under a larger-scale sampling, a dataset of five plastid genes demonstrated that *Rubiteucris* and *Schnabelia* share the closest relationship, followed by *Teucrium* [18]. Furthermore, analyses by Zhao et al. (2021), based on 79 shared plastid genes, corroborated the sister relationship between *Rubiteucris* and *Schnabelia*. In the present study, the five species of *Schnabelia* formed a monophyletic group, which consisted of two clades: one including the original two species (*S. oligophylla* and *S. tetrodonta*), and the other incorporating three species (*S. aureoglandulosa*, *S. nepetifolia*, and *S. terniflora*) that were transferred from *Caryopteris*. Our results are consistent with a previous study using seven DNA regions by Xiang et al. (2018). The two clades represent the two sections of Sect. *Schnabelia* and Sect. *Cylindricaulis*, between which obvious morphological differences exist. Specifically, Sect. *Schnabelia* has winged stems, caducous leaves, a relatively smooth nutlet surface, and an absence of peltate glandular trichomes on the leaves. Conversely, Sect. *Cylindricaulis* is characterized by a distinctly reticulate nutlet surface, with the calyx partially enclosing the nutlet, and the presence of peltate glandular trichomes.

Divergence time of *Schnabelia*

Based on the plastid CDS sequences, the root age of Lamiaceae we obtained (60.35 Ma) was slightly younger than the 63.9/65.45 Ma proposed by Fonseca et al. (2021) [85]and Yao et al. (2016) [48]. This discrepancy may be due to the difference in the selection of fossil calibration sites, use of plastome data and molecular clock models [86, 87]. The major diversification within Lamiaceae occurred after the Paleocene-Eocene Thermal Maximum (PETM, ~56 Ma), a global warming event that likely facilitated the northward expansion of plant distributions and triggered diversification in numerous plant lineages, such as *Koenigia* [88], Styracaceae [41], and Blechnaceae [89, 90].

Our results also indicated that the divergence of the main Lamiaceae lineages occurred mainly after the Cretaceous-Paleoproterozoic (K-Pg) boundary, particularly in the 40–50 Ma time frame [91]. This suggested that the rapid cooling of the climate following the climatically optimal warm conditions of the Eocene had a profound effect on the origin of the modern diversity of Lamiaceae [86]. The timing of Lamioideae divergence (41.23 Ma, 95% HPD: 29.32–56.78 Ma) coincides with the climate cooling in the early Eocene, a period of habitat fragmentation that may have facilitated species differentiation [92, 93]. The age of the crown group of Ajugoideae (46.98 Ma, 95% HPD: 33.94–64.72 Ma), on the other hand, supported the expansion of the warm temperate

flora in the Eocene, corroborating the findings of Rose et al. (2022).

The origin of *Schnabelia* was dated back to the early Oligocene (30.24 Ma), accompanied by the uplift of the Tibetan Plateau and Cenozoic climatic cooling events, which may have facilitated the adaptive radiation of the species through the creation of new ecological niches and increased habitat heterogeneity [47, 90]. Furthermore, the divergence of Sect. *Schnabelia* and Sect. *Cylindricaulis* can be traced back to 12.60 Ma in the middle Miocene, corresponding to the uplift of the southeastern margin of the Tibetan Plateau (~15–10 Ma) [94, 95]. Climate change and topographic complexity triggered by the uplift of the Tibetan Plateau may have provided diverse microclimatic ecological niches for the ancestral species of *Schnabelia*. During the Quaternary, the interplay of glacial-interglacial cycles and monsoonal climate played an important role in shaping plant diversity. During this period, many plant taxa have experienced a high degree of diversification at the intraspecific level [96]. This increased monsoon variability and climatic cooling also facilitated species formation and adaptive differentiation of subtropical taxa, such as *Dysosma versipellis* [97] and *Tetracentron sinense* [98]. The most recent differentiation of Sect. *Schnabelia* (0.26 Ma) may be associated with several warm and prolonged interglacials after the mid-Pleistocene (1.2–0.8 Ma), which provided favorable conditions for population expansion [96].

Conclusions

This study investigated the plastomes features and evolution of five *Schnabelia* species, revealing a certain degree of conservation in their gene content, repeats and SSRs numbers, and IR/SC boundary variation. Codon usage bias across all species showed similar patterns, and no positive selective pressure were detected on all the protein-coding genes despite their different distribution areas. However, five IGS regions (*trn*H-*GUG*-*psb*A, *trn*K-*UUU*-*mat*K, *pet*B-*pet*D, *ndh*D-*psa*C, *ndh*A-*ndh*H) were identified as divergence hotspots which can serve as potential molecular markers for species identification. The phylogenetic tree constructed based on the complete plastomes confirmed the systematic position of *Schnabelia* within the subfamily Ajugoideae, showing a close relationship with *Teucrium* and supporting two internal lineages: one containing the original two species and the other involving three species transferred from *Caryopteris*. The middle Miocene crown age of *Schnabelia* (12.60 Ma) implied that the divergence of Sect. *Schnabelia* and Sect. *Cylindricaulis* could be associated with uplift of the Tibetan Plateau and global climate cooling events. Our study provided polymorphic loci and genetic resources for future species identification, population genetics research, genetic diversity estimation, and conservation of *Schnabelia*. The divergence time estimation of *Schnabelia* provided insights into *Schnabelia*'s endemism and evolutionary history in China. However, to infer the genetic diversity and a more informative evolutionary history of this Chinese endemic genus, further population genetics studies based on large-scale sampling are necessary in the future.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06647-y.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Acknowledgements
Not applicable.

Author contributions
Q.L., P.L. and X.J. conceived and designed the study. P.L. and M.L. collected the sample. S.W., J.L., J.Y., Y.H. and Z.W. analyzed the data. S.W. and J.L. wrote the manuscript. Q.L., X.J. and P.N. revised the paper. All authors reviewed the manuscript.

Funding
The work was supported by the National Natural Science Foundation of China (31700321), Science Foundation of Zhejiang Sci-Tech University (grant No. 19042144-Y), and the Open Fund of Shanghai Key Laboratory of Plant Functional Genomics and Resources (grant no. PFGR202402).

Data availability
The annotated sequences of the newly generated plastomes of Schnabelia species were deposited in the National Center for Biotechnology Information (NCBI) GenBank database under the accession numbers: PP503204-PP503208.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Zhejiang Province Key Laboratory of Plant Secondary Metabolism and Regulation, College of Life Sciences and Medicine, Zhejiang Sci-Tech University, Hangzhou 310018, China
[2]Jiashan Lige Ecological Technology Co. Ltd, Jiashan 314113, China
[3]Shanghai Key Laboratory of Plant Functional Genomics and Resources, Shanghai Chenshan Botanical Garden, Shanghai 201602, China
[4]Minesite Biodiversity Monitoring with eDNA Research Group, Trace and Environmental DNA (TrEnD) Laboratory, School of Molecular and Life Sciences, Curtin University, Bentley, WA, Australia
[5]Laboratory of Systematic & Evolutionary Botany and Biodiversity, College of Life Sciences, Zhejiang University, Hangzhou 310058, China
[6]College of Life and Environmental Science, Wenzhou University, Wenzhou 325035, China

Received: 2 March 2025 / Accepted: 29 April 2025
Published online: 07 May 2025

References

1. Shi SH, Du YQ, Boufford DE, Gong X, Huang Y, He HH, Zhong Y. Phylogenetic position of *Schnabelia*, a genus endemic to China: evidence from sequences of cpDNA*mat*K gene and NrDNA ITS regions. Chin Sci Bull. 2003;48(15):1576–80.
2. Xiang CL, Zhao F, Cantino PD, Drew BT, Li B, Liu ED, Soltis DE, Soltis PS, Peng H. Molecular systematics of *Caryopteris* (Lamiaceae) and its allies with reference to the molecular phylogeny of subfamily Ajugoideae. Taxon. 2018;67(2):376–94.
3. Cantino PD. Evidence for a polyphyletic origin of the Labiatae. Ann Mo Bot Gard. 1992;79(2):361–79.
4. Cantino PD, Wagstaff SJ, Olmstead RG. *Caryopteris* (Lamiaceae) and the conflict between phylogenetic and pragmatic considerations in botanical nomenclature. Syst Bot. 1999;23(3):369–86.
5. Wagstaff SJ, Hickerson L, Spangler R, Reeves PA, Olmstead RG. Phylogeny in Labiatae S. l., inferred from CpDNA S.quences. Plant Syst Evol. 1998;209(3):265–74.
6. Huang M. Systematics of *Trichostema* L. (Lamiaceae) and phylogenetic relationships with its disjunct taxa in Asia. Columbus, Ohio, U.S.A.: Ohio State University; 2002.
7. Li HW, Hedge IC. Lamiaceae. In: *Flora of China.* Edited by Wu CY, Raven PH, vol. 17: Beijing: Science Press;St. Louis: Missouri Botanical Garden Press, 1994:269–291.
8. Chen Y. Studies on the chemical composition of the folk medicines *Caryopteris aureoglandulosa* and *Caryopteris terniflora*. (Master. Thesis)Shanghai University of Traditional Chinese Medicine, Shanghai, China, (2020).
9. Chen LH, Liu YQ, Zhang JS, Yu YQ, Weng ZJ, Huang WC, Zheng YZ, Zhang ZX. Progress in the study of *Caryopteris*. Agric Technol. 2023;43:54–7.
10. Zhang YH, Wang YL, Wei QY, Cai YJ, Wang Q, Liu ZL. Diterpenoids from the Chinese herb *Caryopteris Terniflora* and their antibacterial and antitumor activity. Pharmazie. 2005;60(7):551.
11. Zhang CG, Chou GX, Mao XD, Yang QS, Zhou JL. Nepetaefolins A-J, cytotoxic Chinane and abietane diterpenoids from *Caryopteris Nepetaefolia*. J Nat Prod. 2017;80(6).
12. Xu HT, Zhang CG, He YQ, Shi SS, Chou GX. Phenylethanoid glycosides from the *Schnabelia nepetifolia* (Benth.) P.D.Cantino promote the proliferation of osteoblasts. Phytochemistry. 2019;164:111–21.
13. Huang Y, Quan J, Cao HQ, Xiao Z, Han F, Yi SR. Biological characteristics and artificial domestication and cultivation techniques of *Schnabelia oligophylla*. Mod Chin Med. 2012;14:34–7.
14. Cheng P, Chen F, He XY, Zhou MQ, Ming FX. Comparative study on anti-sports fatigue effect of herb *Schnabelia tetrodonta* in mice. In: *Asia-Pacific Traditional Medicine,* 2014;13: 9–11.
15. Ouyang K, He XY, Chen F, Huang YR, Zhou MQ, Cheng P. Pharmacognosy identification of *Schnabelia tetrodonta*. J Chin Med Mater. 2016;09:1997–2000.
16. Zheng XK, Yan H, Li DD, Li M, He JL, Feng WS. Chemical constituents of *Caryopteris Terniflora* Maxim. Chin Pharm J. 2013;48:1997–2001.
17. Mao XD. Studies on chemical constituents of two medical plants from *Caryopteris* genus and synthesis towards bioactive diterpenoids and their derivatives. In: Doctor Thesis, Shanghai University of Traditional Chinese Medicine, Shanghai, China. 2020.
18. Li B, Cantino PD, Olmstead RG, Bramley GL, Xiang CL, Ma ZH, Tan YH, Zhang DX. A large-scale Chloroplast phylogeny of the Lamiaceae sheds new light on its subfamilial classification. Sci Rep. 2016;6:34343.
19. Zhao F, Chen YP, Salmaki Y, Drew BT, Wilson TC, Scheen AC, Celep F, Bräuchler C, Bendiksby M, Wang Q, et al. An updated tribal classification of Lamiaceae based on plastome phylogenomics. BMC Biol. 2021;19(1):2.
20. Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011;76(3–5):273–97.
21. Shahzadi I, Abdullah, Mehmood F, Ali Z, Ahmed I, Mirza B. Chloroplast genome sequences of *Artemisia maritima* and *Artemisia absinthium*: comparative analyses, mutational hotspots in genus *Artemisia* and phylogeny in family Asteraceae. Genomics. 2020;112(2):1454–63.
22. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17(1):134.
23. Shaw J, Lickey EB, Schilling EE. L. SR: Comparison of whole Chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot. 2007.
24. Liu J, Feng YQ, Chen C, Yan J, Bai XY, Li HR, Lin C, Xiang YN, Tian W, Qi ZC, et al. Genomic insights into the clonal reproductive *Opuntia cochenillifera*: mitochondrial and Chloroplast genomes of the cochineal cactus for enhanced Understanding of structural dynamics and evolutionary implications. Front Plant Sci. 2024;15:1347945.
25. Ren FM, Wang LQ, Zhuo W, Lu SG, Zhu XF, Liu C, Yang MS. First complete Chloroplast genome of the rare medicinal plant *Schnabelia tetrodonta*. Mitochondrial DNA Part B Resour. 2021;6(10):2993–4.
26. Chen SF, Zhou YQ, Chen YR, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinf (Oxford England). 2018;34(17):i884–90.
27. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS. GetOrganelle: a fast and versatile toolkit for accurate de Novo assembly of organelle genomes. Genome Biol. 2020;21(1):241.
28. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de Novo genome assemblies. Bioinf (Oxford England). 2015;31(20):3350–2.
29. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–11.
30. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47(W1):W59–64.
31. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinf (Oxford England). 2017;33(16):2583–5.
32. Hershberg R, Petrov DA. Selection on codon bias. Annu Rev Genet. 2008;42:287–99.
33. Li J, Zhou J, Wu Y, Yang SH, Tian DC. GC-Content of Synonymous Codons Profoundly Influences Amino Acid Usage. *G3 (Bethesda, Md)* 2015;5(10):2027–2036.
34. Wang RH, Gao J, Feng JY, Yang ZP, Qi ZC, Li P, Fu CX. Comparative and phylogenetic analyses of complete Chloroplast genomes of *Scrophularia Incisa* complex (Scrophulariaceae). Genes. 2022;13(10):1691.
35. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res 2004;32(Web Server issue):W273–279.
36. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34(12):3299–302.
37. Li QQ, Chen X, Yang DF, Xia PG. Genetic relationship of *Pleione* based on the Chloroplast genome. Gene. 2023;858:147203.
38. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
39. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.
40. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–42.
41. Jian X, Wang Y, Li Q, Miao Y. Plastid phylogenetics, biogeography, and character evolution of the Chinese endemic genus *Sinojackia* Hu. Divers Distrib. 2024;16(5):305.
42. Kar R. On the Indian origin of *Ocimum* (Lamiaceae): A palynological approach. J Palaeosciences 1993.
43. Cai HM, Liu X, Wang WQ, Ma ZH, Li B, Bramley GL, Zhang DX. Phylogenetic relationships and biogeography of Asia *Callicarpa* (Lamiaceae), with consideration of a long-distance dispersal across the Pacific Ocean—insights into divergence modes of Pantropical flora. Front Plant Sci. 2023;14:1133157.
44. Roy T, Chang TH, Lan T, Lindqvist C. Phylogeny and biogeography of new world Stachydeae (Lamiaceae) with emphasis on the origin and diversification of Hawaiian and South American taxa. Mol Phylogenetics Evol. 2013;69(1):218–38.
45. Yao G, Drew BT, Yi TS, Yan HF, Yuan YM, Ge XJ. Phylogenetic relationships, character evolution and biogeographic diversification of *Pogostemon* Sl (Lamiaceae). Mol Phylogenetics Evol Dev. 2016;98:184–200.
46. Yang ZH. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biology Evol. 2007;24(8):1586–91.

47. Manchester SR, Chen ZD, Lu AM, Uemura K. Eastern Asian endemic seed plant genera and their paleogeographic history throughout the Northern hemisphere. J Syst Evol. 2009;47(1):1–42.
48. Lu LM, Mao LF, Yang T, Ye JF, Liu B, Li HL, Sun M, Miller JT, Mathews S, Hu HH. Evolutionary history of the angiosperm flora of China. Nature. 2018;554(7691):234–8.
49. Zhao Y, Liu YX, Zhou HX, Guo W, Wang WD, Chen HP. Characterisation and skin protection activities of polysaccharides from *Schnabelia Terniflora*. Nat Prod Res. 2024;38(23):4191–5.
50. Chen J, Zang Y, Shang S, Liang S, Zhu M, Wang Y, Tang X. Comparative Chloroplast genomes of Zosteraceae species provide adaptive evolution insights into seagrass. Front Plant Sci. 2021;12.
51. Wu LW, Fan PH, Zhou JG, Li YH, Xu ZC, Lin YL, Wang Y, Song JY, Yao H. Gene losses and homology of the chloroplast genomes of *Taxillus* and *Phacellaria* species. *Genes (Basel).* 2023;14(4).
52. Chen HM, Chen HD, Wang B, Liu C. Conserved Chloroplast genome sequences of the genus *Clerodendrum* Linn. (Lamiaceae) as a super-barcode. PLoS ONE. 2023;18(2):e0277809.
53. Khan A, Asaf S, Khan AL, Khan A, Al-Harrasi A, Al-Sudairy O, Abdulkareem NM, Al-Saady N, Al-Rawahi A. Complete Chloroplast genomes of medicinally important *Teucrium* species and comparative analyses with related species from Lamiaceae. Peer J. 2019;7.
54. Chen H, Zhang X, Zhang G, Zhang Z, Ma G, Sun Z, Liu C, Huang L. The complete Chloroplast genome sequence of *Nepeta bracteata* and comparison with congeneric species. Gene. 2024;893:147919.
55. Li Y, Zhang LN, Wang TX, Zhang CC, Wang RJ, Zhang D, Xie YQ, Zhou NN, Wang WZ, Zhang HM. The complete Chloroplast genome sequences of three lilies: genome structure, comparative genomic and phylogenetic analyses. J Plant Res. 2022;135(6):723–37.
56. Chen MM, Zhang M, Liang ZS, He QL. Characterization and comparative analysis of Chloroplast genomes in five *Uncaria* species endemic to China. Int J Mol Sci. 2022;23(19):11617.
57. Huang H, Shi C, Liu Y, Mao SY, Gao LZ. Thirteen *Camellia* Chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evol Biol. 2014;14(1):151–151.
58. Xiang YN, Wang XQ, Ding LL, Bai XY, Feng YQ, Qi ZC, Sun YT, Yan XL. Deciphering the plastomic code of Chinese hog-peanut (*Amphicarpaea Edgeworthii* Benth., Leguminosae): Comparative genomics and evolutionary insights within the Phaseoleae tribe. Genes (Basel). 2024;15(1).
59. Weng ML, Blazier JC, Govindu M, Jansen RK. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. Mol Biol Evol. 2014;31(3):645–59.
60. Ilyas MZ, Park H, Jang SJ, Cho J, Sa KJ, Lee JK. Association mapping for evaluation of population structure, genetic diversity, and physiochemical traits in drought-stressed maize germplasm using SSR markers. *Plants (Basel).* 2023;12(24).
61. Wang YH, Jiang WM, Comes HP, Hu FS, Qiu YX, Fu CX. Molecular phylogeography and ecological niche modelling of a widespread herbaceous climber, *Tetrastigma Hemsleyanum* (Vitaceae): insights into Plio-Pleistocene range dynamics of evergreen forest in subtropical China. New Phytol. 2015;206(2):852–67.
62. Yu YJ, Lu QX, Lapirov AG, Freeland J, Xu XW: Clear phylogeographical structures shed light on the origin and dispersal of the aquatic boreal plant. 2022;13:1046600.
63. Wang CX, Yap ZY, Wan PL, Chen KQ, Folk RA, Damrel DZ, Barger W, Diamond A, Horn C, Landry GP: Molecular phylogeography and historical demography of a widespread herbaceous species from eastern North America, *Podophyllum peltatum*. Am J Bot. 2023;110(11):e16254.
64. Yin DP, Pang B, Li HB, Liu Q, Zhai YF, Ma N, Chen TT, Shen HJ, Jia QJ, Wang DK. The complete Chloroplast genome of the medical plant *Huperzia crispata* from the Huperziaceae family: structure, comparative analysis, and phylogenetic relationships. Mol Biol Rep. 2022;49(12):11729–41.
65. Zhang ZR, Yang X, Li WY, Peng YQ, Gao J. Comparative Chloroplast genome analysis of *Ficus* (Moraceae): insight into adaptive evolution and mutational hotspot regions. Front Plant Sci. 2022;13:965335.
66. Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, Morris AB. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. Am J Bot. 2014;101.
67. Morris AB, Shaw J. Markers in time and space: A review of the last decade of plant phylogeographic approaches. Mol Ecol. 2018;27(10):2317–33.
68. Abdullah, Henriquez CL, Mehmood F, Hayat A, Sammad A, Waseem S, Waheed MT, Matthews PJ, Croat TB, Poczai P, et al. Chloroplast genome evolution in the *Dracunculus* clade (Aroideae, Araceae). Genomics. 2021;113(1 Pt 1):183–92.
69. Aldrich J, Cherney BW, Merlin E. The role of insertions/deletions in the evolution of the intergenic region between *psb*A and *trn*H in the Chloroplast genome. Curr Genet. 1988;14:137–46.
70. Hamilton MB. Tropical tree gene flow and seed dispersal. Nature. 1999;401(6749):129–30.
71. Soltis DE, Kuzoff RK, Mort ME, Zanis M, Fishbein M, Hufford L, Koontz J, Arroyo MK. Elucidating deep-level phylogenetic relationships in Saxifragaceae using sequences for six chloroplastic and nuclear DNA regions. Ann Mo Bot Gard 2001:669–93.
72. Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. The tortoise and the hare II: relative utility of 21 noncoding Chloroplast DNA sequences for phylogenetic analysis. Am J Bot. 2005;92(1):142–66.
73. Guo L, Wang X, Wang RH, Li P. Characterization and comparative analysis of Chloroplast genomes of medicinal herb *Scrophularia ningpoensis* and its common adulterants (Scrophulariaceae). Int J Mol Sci. 2023;24(12).
74. Zhang J, Lin MX, Cheng YH, Han RG, Shen J, Jin JQ, Wang L, Liu ZY. The folk eight major specific medicine in mount. Jinfo of Chongqing. Chin J Ethnomed Ethnopharmacy. 2016;25(18).
75. Yi SR, Huang Y, Quan J, Pei LR, Lei MY, Cao HQ. Resource investigation and evaluation system establishment of endangered medicinal plants in Jinfo mountain. J Trop Subtropical Bot. 2016:1005–3395.
76. Morton BR. The role of context-dependent mutations in generating compositional and codon usage bias in grass Chloroplast DNA. J Mol Evol. 2003;56(5):616–29.
77. Li TT, Ma Z, Ding TM, Yang YX, Wang F, Wan XJ, Liang FY, Chen X, Yao HP. Codon usage bias and phylogenetic analysis of Chloroplast genome in 36 gracilariaceae species. Funct Integr Genom. 2024;24(2):45.
78. Hao J, Liang YY, Ping JY, Li J, Shi WX, Su YJ, Wang T. Chloroplast gene expression level is negatively correlated with evolutionary rates and selective pressure while positively with codon usage bias in *Ophioglossum vulgatum* L. BMC Plant Biol. 2022;22(1):580.
79. Qian RJ, Ye YJ, Hu QD, Ma XH, Zheng J. Complete Chloroplast genome of *Gladiolus gandavensis* (*Gladiolus*) and genetic evolutionary analysis. Genes (Basel). 2022;13(9).
80. Fan ZF, Ma CL. Comparative Chloroplast genome and phylogenetic analyses of Chinese *Polyspora*. Sci Rep. 2022;12(1):15984.
81. Lu HL, Zhang WC, Zhang SS. Study on comparative morphology of *Schnabelia*, Chinese endemic plant, and its systematic position. V. Comparative anatomy of stem-nodes and petioles of genus *Schnabelia* and its Sibs. Acta Bot Boreali-occidentalia Sinica. 1997;17(2):233–7.
82. Ryding O. Pericarp structure of the *Caryopteris* group (Lamiaceae subfam. Ajugoideae). Nord J Bot. 2009;27(4):257–65.
83. Chen C. The systematic relationship of *Schnabelia* Hand.-Mazz. J Syst Evol. 1964;9(1):1.
84. Du Z, Yuan F, Huang Y, Chen Y, Cheng Y, Wei Y, Yang D. The complete Chloroplast genome of *Salvia Liguliloba* Y. Z. Sun (Lamiaceae). Mitochondrial DNA Part B Resour. 2022;7(7):1355–6.
85. Fonseca LHM. Combining molecular and geographical data to infer the phylogeny of Lamiales and its dispersal patterns in and out of the tropics. Mol Phylogenet Evol. 2021.
86. Rose JP, Xiang CL, Sytsma KJ, Drew BT. A timeframe for mint evolution: towards a better Understanding of trait evolution and historical biogeography in Lamiaceae. Bot J Linn Soc. 2022.
87. Zhu JQ, Huang Y, Chai WG, Xia PG. Decoding the Chloroplast genome of *Tetrastigma* (Vitaceae): variations and phylogenetic selection insights. Int J Mol Sci 2024;25(15).
88. Qu XJ, Zhang XJ, Cao DL, Guo XX, Mower JP, Fan SJ. Plastid and mitochondrial phylogenomics reveal correlated substitution rate variation in *Koenigia* (Polygonoideae, Polygonaceae) and a reduced plastome for *Koenigia delicatula* including loss of all *Ndh* genes. Mol Phylogenetics Evol. 2022;174:107544.
89. Wing SL, Harrington GJ, Smith FA, Bloch JI, Boyer DM, Freeman KH. Transient floral change and rapid global warming at the Paleocene-Eocene boundary. Science. 2005;310(5750):993–996.
90. Testo WL, de Gasper AL, Molino S, Galán JMG, Salino A, Dittrich VAO, Sessa EB. Deep vicariance and frequent transoceanic dispersal shape the evolutionary history of a globally distributed fern family. Am J Bot. 2022;109(10):1579–95.

91. Li P, Qi ZC, Liu LX, Ohi-Toma T, Lee J, Hsieh TH, Fu CX, Cameron KM, Qiu YX. Molecular phylogenetics and biogeography of the mint tribe elsholtzieae (Nepetoideae, Lamiaceae), with an emphasis on its diversification in East Asia. Sci Rep. 2017;7(1):2057.
92. Bendiksby M. Molecular phylogeny, taxonomy, and historical biogeography of Lamiaceae subfamily lamioideae, including surveys of alloploid speciation in two Eurasian genera, Galeopsis and Lamium. AIT Oslo, Norway; 2011.
93. Jiang CX, Wu M, He SS, Lu YX, Zhao C. Phylogeography of a herbal *Pinellia ternata* reveals repeated range expansions and inter/postglacial recolonization routes on the fragmented distribution pattern in China. Ecol Evol. 2024;14(9):e70206.
94. Qiu YX, Fu CX, Comes HP. 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. Mol Phylogenet Evol. 2011;59(1):225–44.
95. Xiao JH, Ding X, Li L, Ma H, Ci XQ, van der Merwe M, Conran JG, Li J. Miocene diversification of a golden-thread Nanmu tree species (*Phoebe zhennan*, Lauraceae) around the Sichuan basin shaped by the East Asian monsoon. Ecol Evol. 2020;10(19):10543–57.
96. Kou YX, Cheng SM, Tian S, Li B, Fan DM, Chen YJ, Soltis DE, Soltis PS, Zhang ZY. The antiquity of *Cyclocarya paliurus* (Juglandaceae) provides new insights into the evolution of relict plants in subtropical China since the late early miocene. J Biogeogr. 2016;43(2):351–60.
97. Qiu YX, Guan BC, Fu CX, Comes HP. Did glacials and/or interglacials promote allopatric incipient speciation in East Asian temperate plants? Phylogeographic and coalescent analyses on refugial isolation and divergence in *Dysosma versipellis*. Mol Phylogenet Evol. 2009;51(2):281–93.
98. Sun YX, Moore MJ, Yue LL, Feng T, Chu HJ, Chen ST, Ji YH, Wang HC, Li JQ. Chloroplast phylogeography of the East Asian Arcto-Tertiary relict *Tetracentron Sinense* (Trochodendraceae). J Biogeogr 2014.

Publisher's note