BMC Plant Biology

RESEARCH Open Access

Comprehensive analysis of the pan-plastome in *Panax*: implications for interspecies divergence and shade tolerance

Xiaolei Wang[1], Shuai Gao[2], Yuanyuan Duan[3], Xiaoju Su[1], Yichun Zuo[1], Shaopeng Yi[3], Jinkun Liu[2], Wei Zhang[1], Hongxia Yu[3*] and Shenglong Kan[1*]

Abstract

Background *Panax*, including medicinally important species such as *P. ginseng*, *P. quinquefolius*, and *P. notoginseng*, exhibits diverse pharmacological properties. However, morphological traits are lost and DNA is often degraded during processing, complicating species identification. As photosynthetic organelles, plastomes offer both stable molecular markers and a direct window into light adaptation. Unlike their sun-loving relatives (*Aralia*), *Panax* species are shade-tolerant, making them ideal for studying shade adaptation.

Results Here, a total of 188 plastomes from *Panax* and its close relatives were annotated and analyzed, including 31 newly assembled plastomes of *P. quinquefolius* from diverse geographic regions. Through a deep comparative genomic analysis, we obtained the following findings: (1) *Panax* plastomes are highly conserved, with only minor shifts at IR-SC boundaries, and 3,397 variants identified across species. (2) All individuals of *P. ginseng*, *P. quinquefolius*, and *P. notoginseng* formed monophyletic clades, and two species-specific SNP markers were identified for accurate discrimination of these three key medicinal taxa. (3) Nine plastid genes showed significantly different selection patterns between *Panax* and *Aralia*, with intensified selection likely linked to shade adaptation.

Conclusion This study provides two single-nucleotide markers for authenticating processed medicinal materials and reveal plastome-level signatures of shade adaptation, offering valuable resources for species identification, cultivation and management.

Keywords Plastid genome, Adaptation, Identification, *Panax*, Araliaceae

*Correspondence:
Hongxia Yu
120903421@qq.com
Shenglong Kan
kanshenglong@sdu.edu.cn
[1]Marine College, Shandong University, Weihai 264209, China
[2]Shandong AnRan Nanometer Industry Development Co., Ltd, 264205 Weihai, China
[3]Weihai Wendeng District Daodi Ginseng Industry Development Co., Ltd, 264407 Weihai, China

Background

Panax belongs to Araliaceae and comprises approximately 18 species, of which 12 are currently recognized as accepted species in The Plant List [1–4]. Species of the genus *Panax* exhibit a disjunct distribution between eastern Asia and eastern North America. Except for *P. trifolius* and *P. quinquefolius*, which occur in North America, all other species are distributed in eastern Asia [1]. Most species within *Panax* are valued in traditional Chinese medicine due to their rich content of ginsenosides [3]. Among them, *P. ginseng* (Asian ginseng), *P. quinquefolius* (American ginseng), and *P. notoginseng* (Sanqi) are the most widely used medicinal species, although their pharmacological effects differ significantly [5]. According to data from UN Comtrade Database [6], global exports of ginseng root products reached 3,466 tons in 2024, with a total export value of 97,480,459 USD. Despite its significance, the taxonomy and phylogenetic relationships within *Panax* remain controversial, attracting considerable scientific attention. Recent studies have attempted to reconstruct the phylogeny of 18 *Panax* species using both nuclear and plastid genome (plastome) [1]. However, these reconstructions have revealed substantial incongruence between nuclear and plastid phylogenies, particularly for species distributed in the Himalayas and adjacent regions. A common limitation of these studies is their reliance on single-individual sampling for most species, which undermines the reliability and representativeness of the phylogenetic inferences.

Robust phylogenetic relationships are fundamental for accurate species identification. Plastids, one of the organelles harboring their own genomes, possess highly conserved genomes in land plants, typically organized into a circular quadripartite structure comprising a large single-copy (LSC) region, a small single-copy (SSC) region, and a pair of inverted repeats (IRs), with sizes ranging from 115 to 165 kilobases (kb) [7]. The plastid genome is generally multi-copy within the cell, maternally inherited, non-recombinant, and exhibits a moderate mutation rate [8–12]. These features make it widely used in phylogenetic studies across various taxonomic levels, including species, genus, family, and higher taxonomic category [13–15]. With advances in sequencing technologies, an increasing number of plastomes and high-throughput sequencing datasets for *Panax* have become available, providing an unprecedented opportunity to enhance sampling representativeness in phylogenetic analyses. Moreover, the plastome serves as an important source of molecular markers, for instance, *matK* and *rbcL* are recognized as core DNA barcodes in plants [16]. Whole-plastome barcoding has also demonstrated substantial potential for species identification, particularly in cases where traditional single-locus markers lack sufficient discriminatory power [17, 18]. However, in certain practical contexts, such as the identification of traditional Chinese medicinal materials that have undergone processing treatments like baking and steaming, morphological traits are often lost and DNA is degraded [19]. Under such conditions, traditional morphological methods and conventional DNA barcoding approaches become ineffective. Developing single-nucleotide diagnostic markers (SNDMs) thus represents a promising strategy for identifying materials with degraded DNA [20]. SNDMs are designed to amplify and target extremely short DNA fragments (often < 200 bp), whereas conventional barcoding relies on amplifying much longer conserved regions (e.g., ITS2, ~400 bp; *psbA-trnH*, ~450 bp). In severely degraded DNA, the probability of preserving a short, targetable fragment is significantly higher than that of a long, intact barcode region [21, 22].

On the other hand, plastids are the primary site of photosynthesis, where plants capture light energy and convert it into chemical energy, functioning as both photoreceptors and energy converters [23, 24]. The plastome encodes approximately 85 unique genes, most of which are involved in photosynthesis, including subunits of photosystems I and II, components of the electron transport chain, and proteins associated with replication and translation within the plastid [25]. As such, the plastome is the organellar genome most closely linked to plant adaptation to varying light environments. Species of *Panax* are perennial, shade-tolerant herbs that typically grow in forest understories and require shade management even under cultivation (e.g., *P. ginseng* and *P. quinquefolius*) [26–28]. In contrast, most species of *Aralia*, the closest related genus to *Panax*, are sun-loving trees that thrive under direct sunlight [29]. This ecological divergence makes *Panax* and *Aralia* an ideal natural comparative system for studying plastome adaptations to varied light habitats.

The advancement of high-throughput sequencing technologies has led to the generation of a large volume of sequencing data. As of December 2024, we downloaded all publicly available plastomes of *Panax* and its close relatives from the Nucleotide database of the National Center for Biotechnology Information (NCBI) and obtained raw sequencing data from the Genome Sequence Archive (GSA) at the China National Center for Bioinformation (CNCB). We reassembled 31 plastomes of *P. quinquefolius* individuals from different locations, re-annotated 157 published plastomes, and conducted comprehensive comparative genomic analyses. Our objectives were to: (1) perform a pan-plastome analysis of all accepted *Panax* species (except *P. sokpayensis*), focusing on gene content, repeats, IR-SC boundary, and generated a comprehensive landscape of plastome variant in *Panax*; (2) reconstruct the phylogenetic relationships among *Panax* and its related species based on plastomes, and identify

species-specific markers to distinguish *P. ginseng*, *P. quinquefolius*, and *P. notoginseng* at single-nucleotide level; and (3) use the natural shade-tolerant/sun-loving contrast between *Panax* and *Aralia*, in combination with site model, branch model, branch-site model, and RELAX analysis, to explore the evolutionary mechanisms of plastome adaptation to differing light environments. This study offers novel insights into the genetic basis of shade adaptation in *Panax* and provides valuable makers for species identification and conservation, supporting improved management and utilization of economically important ginseng resources.

Results

Characteristics of pan-plastomes in *Panax*

To investigate the interspecific relationships within *Panax*, identification of economically important species, and adaptive evolutionary mechanisms under shaded environments, we re-annotated and manually curated the plastomes of 188 accessions from *Panax* (154), *Aralia* (26), and outgroups (8). This dataset included 31 *de novo* assembled plastomes of *P. quinquefolius*, as well as 47 accessions of *P. ginseng* and 16 accessions of *P. notoginseng* (Additional file 1: Table S1). All plastomes exhibited a conserved quadripartite structure, consisting of a pair of IR regions separating a LSC region and a SSC region.

The average length of *Panax* plastomes was 156,238 bp, ranging from 155,984 to 156,548 bp (Fig. 1; Table 1). The average LSC region length was 86,135 bp (range: 86,071–86,322 bp), SSC region 18,036 bp (range: 17,934 – 19,364 bp), and IR region 26,033 bp (range: 25,393 – 26,136 bp). The average GC content of *Panax* plastomes was 38.1% (range: 38.0%-38.1%). The GC content of the LSC, SSC, and IR regions ranged from 36.2% to 36.3%, 32.0%-32.3%, and 42.9%-43.1%, respectively. A total of 133 genes (115 unique) were annotated across all 154 *Panax* samples, including 88 protein-coding genes (PCGs, 81 unique), 37 tRNA genes (30 unique), and 8 rRNA genes (4 unique). Seven PCGs (*rpl2*, *rpl23*, *ndhB*, *rps12*, *rps7*, *ycf2*, *ycf15*) were duplicated due to their presence in the IR regions (Additional file 1: Table S1 and S2). Specifically, due to shifts in the IR–SSC boundary, *ycf1* was fully retained only at the IRa-SSC junction, whereas at the IRb-SSC junction only a truncated fragment remains, which was not counted as an independent PCG. Ten PCGs (*rps16*, *atpF*, *rpoC1*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, *rps12*, *ndhA*) each contained a single intron, while two PCGs (*ycf3* and *clpP*) contained two introns each. Only *rps12* contained a trans-spliced intron. The total alignment length across all samples was 159,863 bp (Fig. 1; Additional file 1: Table S2). A total of 3,568 variant sites were identified in *Panax*, including 171 indels (4.79%) and 3,397 SNPs (95.21%). Of these, 2,318 (64.97%) variants were located in the LSC, 609 (17.07%) in the SSC, and 848 (23.77%) in the IR regions (Additional file 1: Table S3). In *P. ginseng*, 154 variant sites were detected, including 11 indels (7.14%) and 143 SNPs (92.86%). In *P. quinquefolius*, 59 variant sites were identified, including 18 indels (30.51%) and 41 SNPs (69.49%). *P. notoginseng* exhibited 39 variant sites, with only one indel (2.56%) and 38 SNPs (97.44%).

Relative synonymous codon usage (RSCU) analysis was conducted for all examined plastomes (Additional file 2: Fig. S1). Overall, a clear bias toward A/U-ending codons was observed across most amino acids. For example, codons such as UUU (Phe), AUU (Ile), and GUU (Val) were generally preferred over their C/G-ending synonymous counterparts. Similarly, UUA and UUG were more frequently used among leucine codons. This A/U-ending preference was consistent across all examined *Aralia* and *Panax* species, with only minor interspecific variations in RSCU values. In contrast, codons ending with G or C were generally underrepresented, indicating a conserved codon usage pattern in *Panax* plastomes.

Plastid phylogenomics analysis in *Panax*

We reconstructed the phylogenetic relationships of *Panax* and its closely related genera based on both complete plastomes and concatenated PCGs. The final aligned and quality-controlled matrices were 149,100 bp in length for the full plastome and 63,736 bp for the 81 shared PCGs. To clarify interspecific relationships within *Panax*, multiple accessions were included for each species except *Panax pseudoginseng*. The two phylogenetic trees exhibited consistent topologies across major clades (Fig. 2; Additional file 2: Fig. S2). *Panax* was recovered as a monophyletic group, with *P. trifolius* occupying the most basal position, representing the earliest diverging lineage. This was followed by the divergence of *P. stipuleanatus*, after which two major clades emerged. One clade comprised *P. ginseng* and *P. quinquefolius*, while the other included the early-diverging *P. notoginseng* along with *P. pseudoginseng*, *P. bipinnatifidus*, *P. japonicus*, *P. wangianus*, *P. vietnamensis*, and *P. zingiberensis*. Notably, the wild and cultivated accessions of *P. quinquefolius* clustered together without forming separate monophyletic clades, suggesting limited differentiation in the plastome between cultivated and wild populations. In contrast, the genus *Aralia* was not monophyletic and split into two major clades: Clade A, which included both woody and herbaceous species, and Clade B, consisting exclusively of woody species, which exhibited a closer phylogenetic relationship with *Panax*.

Repeats and structural variations in *Panax* and relative genus

We conducted a comparative analysis of repeats across the plastomes of *Panax* and *Aralia*, using *Osmoxylon*

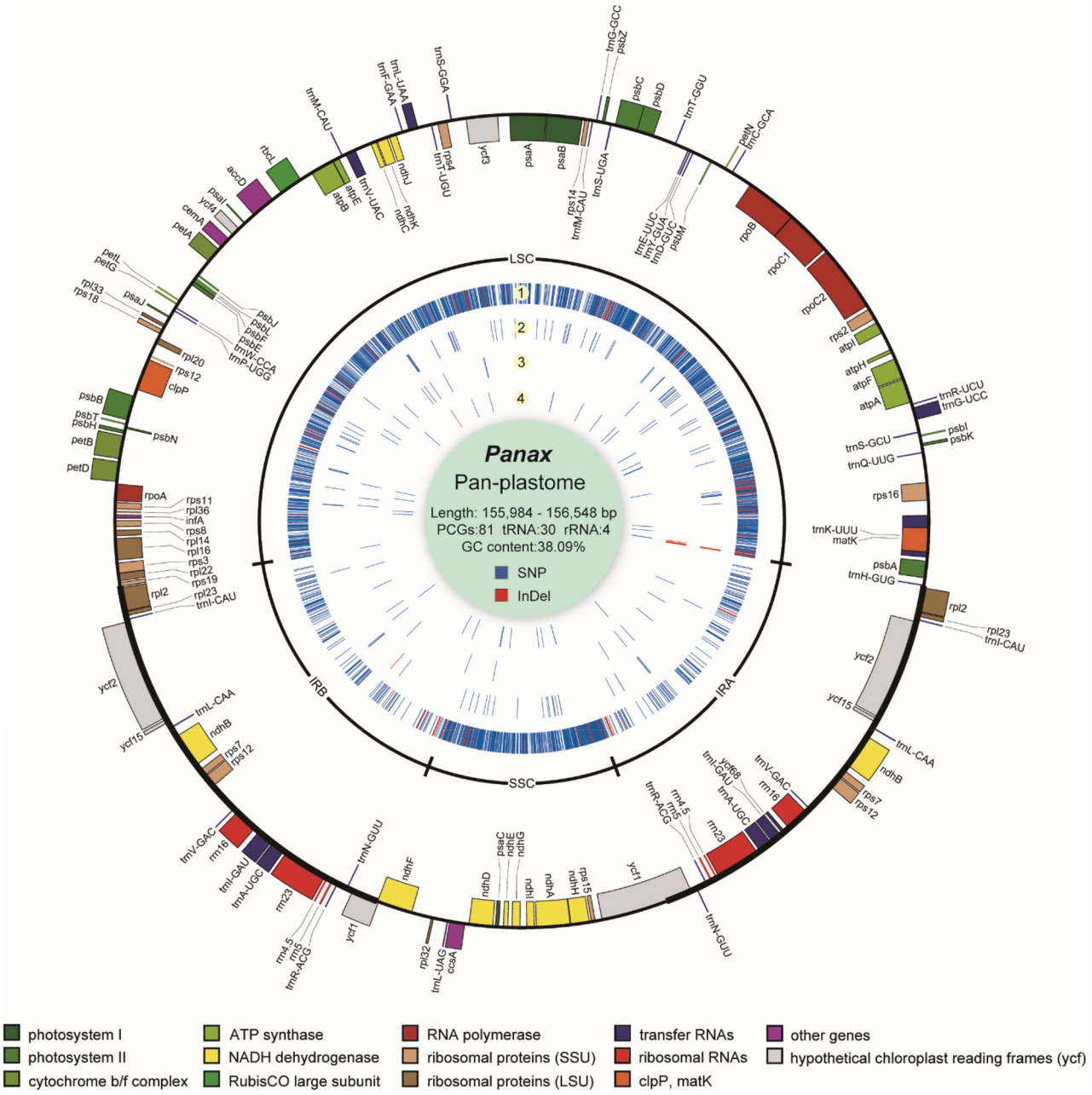

Fig. 1 Pan-plastome and variants distribution in *Panax*. From the outside to the inside, the first circle exhibited the annotation information of the plastome. Gene name on the outside of the circle is transcribed clockwise, while those on the inside are transcribed counterclockwise. Gene colors correspond to functional categories indicated in the legend below. The second circle shows the boundaries of LSC, SSC, and IRs. The inner four heatmap tracks exhibited the distribution of variants across the plastomes. Blue lines indicate single nucleotide polymorphisms (SNPs), and red lines indicate insertions/deletions (InDels). From outer to inner, the four heatmap tracks correspond to: ① all *Panax* plastomes, ② *P. ginseng*, ③ *P. quinquefolius*, and ④ *P. notoginseng*

novoguineense (MK943807.1) as the baseline (Fig. 3). The repeats were categorized into tandem repeats and dispersed repeats. Tandem repeats were further divided into simple sequence repeats (SSRs, repeat units of 1–6 bp) and general tandem repeats (repeat units > 6 bp). Among the SSRs in the plastomes of *Panax* and *Aralia*, mononucleotide repeats were the most abundant, followed by di-, tetra-, tri- and pentanucleotide repeats. Hexanucleotide repeats were exclusively detected in *Panax ginseng*, *P. quinquefolius*, *P. notoginseng*, *P. wangianus*, *P. zingiberensis*, and *P. vietnamensis* (Fig. 3A; Additional file 1: Table S4). To further investigate the functional distribution of SSRs, all identified SSR loci were classified into coding sequences (CDS), introns, and intergenic spacers (IGS) based on genome annotation (Additional file 1: Table S5 and S6). Across all sampled species, SSRs were

Table 1 General information of studied plastomes

Species	Number of accessions	Length (bp)				Gene content			
		LSC	SSC	IR	Total	PCGs	tRNA	rRNA	Total
Panax ginseng	47	86134.11 ± 29.89	18079.87 ± 16.72	26084.55 ± 40.03	156383.09 ± 76.95	88(81)	37(30)	8(4)	133(115)
Panax japonicus	2	86199.00 ± 0.00	18013.00 ± 0.00	25988.00 ± 0.00	156188.00 ± 0.00	88(81)	37(30)	8(4)	133(115)
Panax bipinnatifidus	16	86162.81 ± 43.78	18002.69 ± 10.42	26014.88 ± 38.84	156195.25 ± 109.04	88(81)	37(30)	8(4)	133(115)
Panax notoginseng	16	86142.63 ± 28.95	18123.50 ± 358.11	26032.06 ± 182.10	156330.25 ± 53.59	88(81)	37(30)	8(4)	133(115)
Panax quinquefolius	32	86080.97 ± 7.77	17993.00 ± 0.00	26001.78 ± 10.08	156077.53 ± 20.91	88(81)	37(30)	8(4)	133(115)
Panax stipuleanatus	7	86112.86 ± 36.49	18155.43 ± 18.92	25885.86 ± 3.02	156043.86 ± 35.88	88(81)	37(30)	8(4)	133(115)
Panax vietnamensis	9	86183.44 ± 19.48	17946.89 ± 16.41	25985.11 ± 63.64	156100.56 ± 144.48	88(81)	37(30)	8(4)	133(115)
Panax trifolius	2	86322.00 ± 0.00	18047.00 ± 0.00	25894.00 ± 0.00	156157.00 ± 0.00	88(81)	37(30)	8(4)	133(115)
Panax zingiberensis	13	86137.62 ± 20.65	17967.69 ± 2.84	26089.08 ± 30.64	156283.46 ± 78.95	88(81)	37(30)	8(4)	133(115)
Panax wangianus	9	86184.44 ± 6.50	17966.33 ± 1.00	26042.22 ± 30.16	156235.22 ± 58.49	88(81)	37(30)	8(4)	133(115)
Panax pseudoginseng	1	86124.00 ± NA	18008.00 ± NA	25971.00 ± NA	156074.00 ± NA	88(81)	37(30)	8(4)	133(115)
Aralia elata	4	86030.00 ± 266.74	18105.25 ± 7.23	25923.00 ± 0.00	155981.25 ± 273.96	88(81)	37(30)	8(4)	133(115)
Aralia spinifolia	1	86384.00 ± NA	18068.00 ± NA	25874.00 ± NA	156200.00 ± NA	88(81)	37(30)	8(4)	133(115)
Aralia echinocaulis	3	86066.67 ± 7.51	18121.00 ± 0.00	25941.00 ± 0.00	156069.67 ± 7.51	88(81)	37(30)	8(4)	133(115)
Aralia continentalis	2	86049.00 ± 0.00	18052.00 ± 0.00	25949.00 ± 0.00	155999.00 ± 0.00	88(81)	37(30)	8(4)	133(115)
Aralia undulata	2	86025.00 ± 0.00	18092.00 ± 0.00	26108.00 ± 0.00	156333.00 ± 0.00	88(81)	37(30)	8(4)	133(115)
Aralia chinensis	2	86154.50 ± 171.83	18087.00 ± 19.8	25943.00 ± 29.70	156120.00 ± 142.84	88(81)	37(30)	8(4)	133(115)
Aralia atropurpurea	3	86242.67 ± 20.21	18072.00 ± 6.93	25963.00 ± 0.00	156241.33 ± 26.56	88(81)	37(30)	8(4)	133(115)
Aralia cordata	3	86099.33 ± 2.89	18052.00 ± 0.00	25967.00 ± 0.00	156085.33 ± 2.89	88(81)	37(30)	8(4)	133(115)
Aralia caesia	3	86295.00 ± 0.00	18060.00 ± 0.00	25923.00 ± 0.00	156201.00 ± 0.00	88(81)	37(30)	8(4)	133(115)
Aralia parasitica	2	86171.00 ± 0.00	18055.00 ± 0.00	25926.00 ± 0.00	156078.00 ± 0.00	88(81)	37(30)	8(4)	133(115)
Aralia thomsonii	1	86026.00 ± NA	18123.00 ± NA	25938.00 ± NA	156025.00 ± NA	88(81)	37(30)	8(4)	133(115)
Schefflera heptaphylla	1	86610.00 ± NA	18147.00 ± NA	25964.00 ± NA	156685.00 ± NA	86(80)	37(30)	8(4)	131(114)
Schefflera delavayi	2	86072.50 ± 67.18	18069.50 ± 108.19	26073.00 ± 49.50	156205.50 ± 191.63	87(80)	37(30)	8(4)	132(114)
Schefflera actinophylla	1	86576.00 ± NA	18157.00 ± NA	25968.00 ± NA	156675.00 ± NA	86(80)	37(30)	8(4)	131(114)
Schefflera digitata	1	86540.00 ± NA	18154.00 ± NA	25927.00 ± NA	156548.00 ± NA	86(80)	37(30)	8(4)	131(115)
Heteropanax fragrans	2	86567.50 ± 41.72	18117.50 ± 37.48	25949.00 ± 38.18	156583.00 ± 2.83	88(81)	37(30)	8(4)	133(115)
Osmoxylon novoguineense	1	86695.00 ± NA	18544.00 ± NA	25521.00 ± NA	156281.00 ± NA	88(81)	37(30)	8(4)	133(115)

Number indicates mean and standard deviation

The numbers in parenthesis represented the number of unique genes in plastome

LSC large single copy region, *SSC* small single copy region, *IR* one of a pair of inverted repeat region, *PCGs* protein-coding genes, *tRNA* transfer RNA *rRNA* ribosomal RNA

predominantly located in IGS regions, with counts ranging from 23 to 29 per species, accounting for the majority of SSR loci. In contrast, relatively fewer SSRs were detected in CDS regions, typically ranging from 9 to 14, while intronic regions harbored the lowest numbers (3–9 SSRs).

For general tandem repeats, the predominant repeat unit size ranged from 10 to 30 bp, with only a few species exhibiting a small number of repeat units < 10 bp or > 30 bp. The accumulated length of tandem repeats varied among species, ranging from 856 bp to 1677 bp, with *Aralia undulata* accumulated the longest cumulative repeat length (Fig. 3B; Additional file 1: Table S7 and S8). According to the orientation of the repeat units, dispersed repeats can be classified into four types: forward (F), reverse (R), complement (C), and palindromic (P) repeats. In general, forward (F) and palindromic (P) repeats are predominant in the plastomes of *Panax* and *Aralia*, reverse (R) repeats are relatively rare, whereas complement (C) repeats were detected only in *A. elata* and *P. zingiberensis* (Fig. 3C; Additional file 1: Table S9). On the other hand, based on the repeat length, dispersed repeats were categorized into four groups: <40 bp, 40–60 bp, 60–80 bp, and > 80 bp, with the < 40 bp being the most prevalent (Fig. 3D; Additional file 1: Table S10). Overall, the number of dispersed repeats ranged from 29 to 87 across species, with total accumulated lengths varying from 738 bp to 3,172 bp. *P. zingiberensis* exhibited the highest number, and displayed the greatest accumulated repeat length, reaching 3,466 bp.

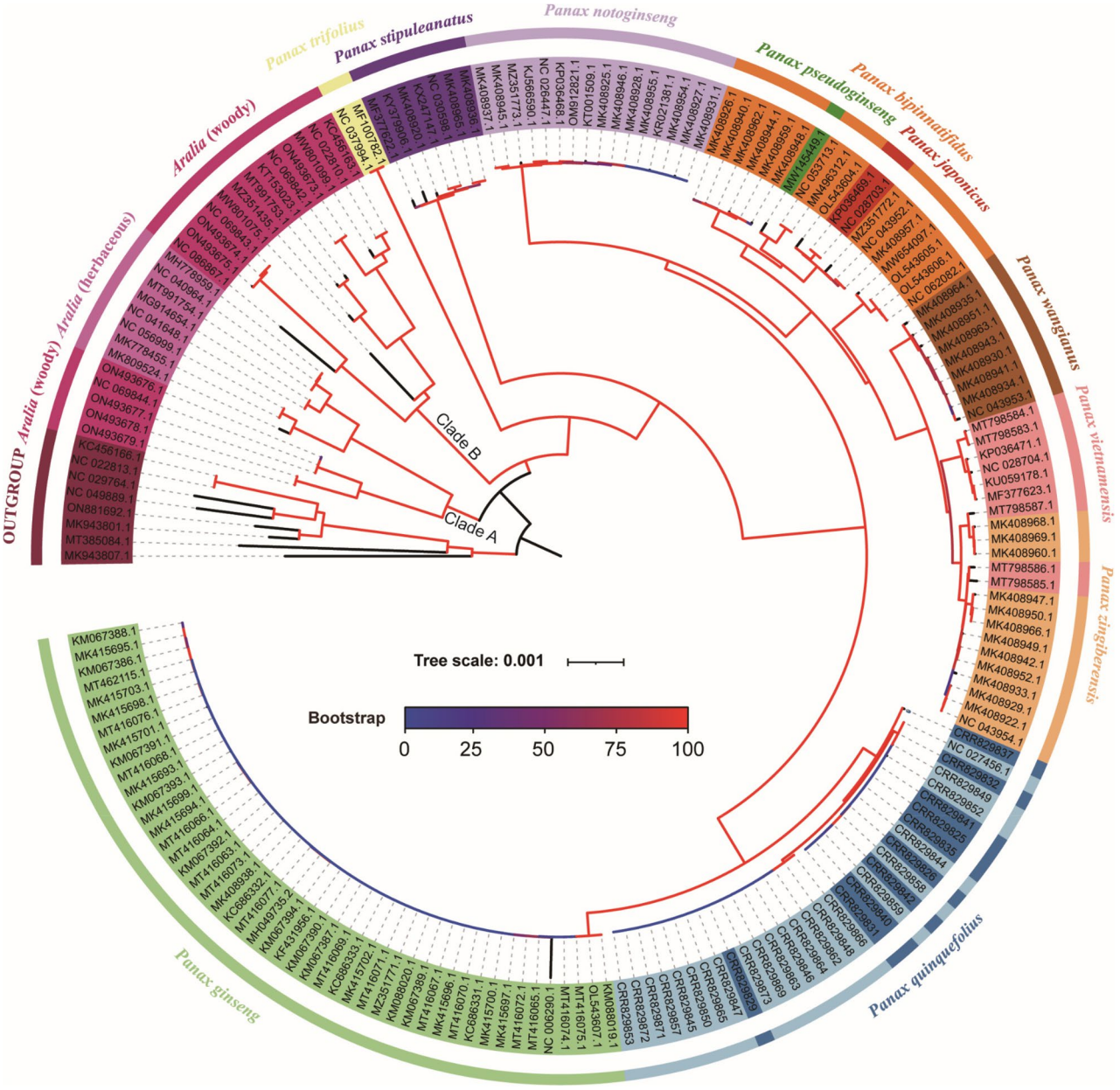

Fig. 2 Maximum likelihood tree reconstructed based on whole plastome sequences. Branch colors indicate bootstrap values, while shade color indicates species. Deep blue indicates wild *P. quinquefolius* samples, whereas light blue indicates cultivated *P. quinquefolius* samples

To detect whether there are structural variations in the plastomes of *Panax* and *Aralia*, we used mVISTA to visualize the plastomes across different individuals within each species, and no structural variation was observed within each species. Therefore, one representative per species was selected for visualization, and no recombination or rearrangement was detected, only shifts in the boundaries between IR and SC was detected (Fig. 4; Additional file 2: Fig. S3). The *rps19* is located at the boundary between the LSC and IRb. In *P. trifolius*, 233 bp of the *rps19* is located in the LSC and 46 bp in the IRb. In *P. stipuleanatus*, the gene has shifted 7 bp further into the IRb, whereas in other *Panax* species, the shift into IRb is 5 bp. In contrast, *Aralia* species exhibit greater variation than *Panax*, with the *rps19* of *A. spinifolia* entirely located in the LSC. The *ndhF* is positioned at the IRb/SSC boundary. In *P. trifolius*, *P. stipuleanatus*, *P. ginseng*, *P. pseudoginseng*, *P. bipinnatifidus*, and *P. japonicus*, the *ndhF* is entirely contained within the SSC, whereas in other *Panax* species, it spans both the IRb and SSC. Among *Aralia* species, only *A. chinensis* shows *ndhF* spanning the IRb and SSC. The *ycf1* gene located at the SSC/IRa boundary and is shared between SSC and IR regions in both *Panax* and *Aralia*. In *Panax*, the portion

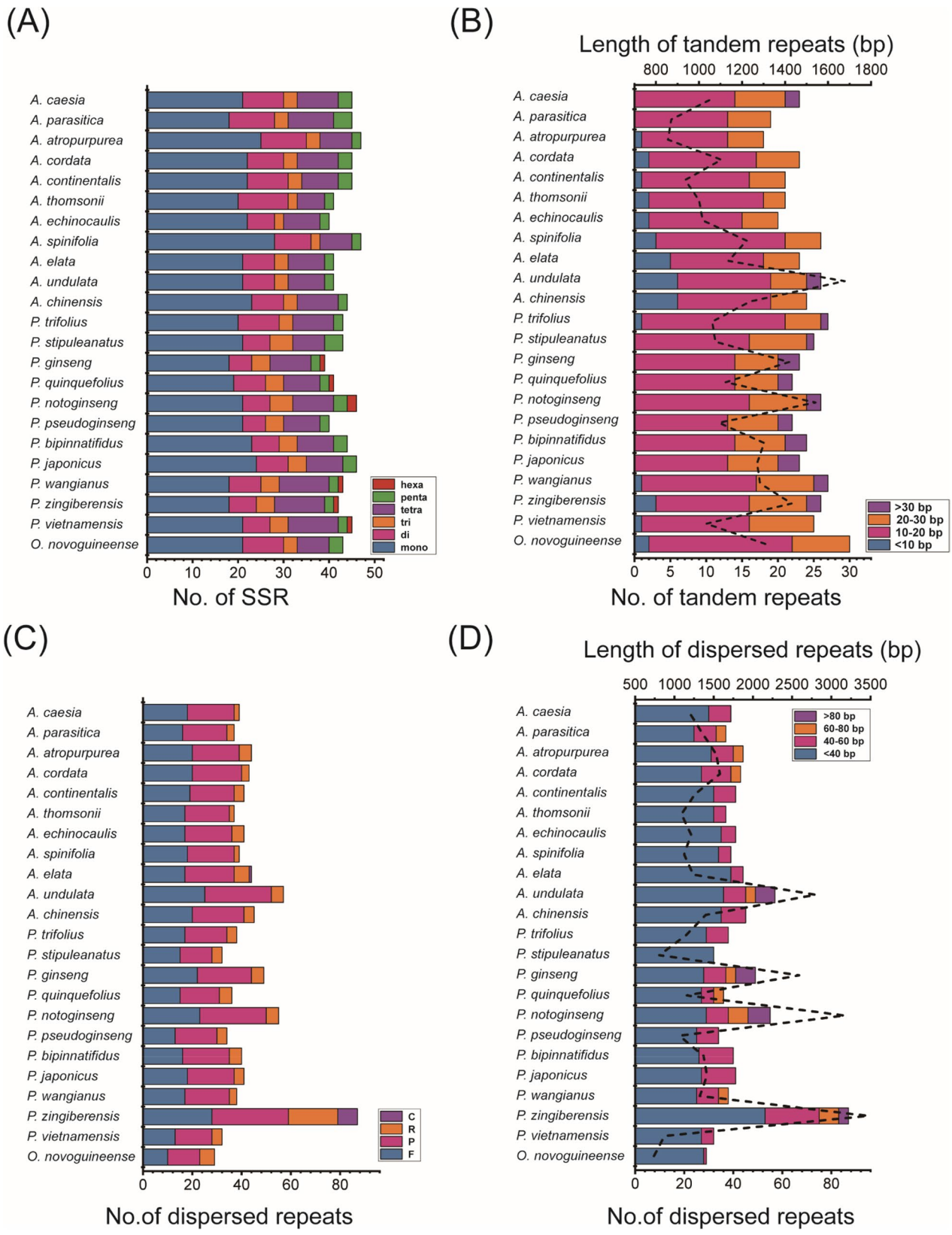

Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Statistics on various types of repeats in the sampled plastomes. **A** Number of SSR with different lengths. "mono", "di", "tri", "tetra", "penta" indicated simple mono-, di-, tri-, tetra- and pentanucleotide microsatellites, respectively. **B** Number and accumulated length of tandem repeats with different lengths. The line indicates the accumulated length, while the bar indicates the number of repeats. **C** Number of dispersed repeats classified by type. "F" indicates the forward repeats, "R" indicates the reverse repeats, "C" indicates the complement repeats, and "P" indicates the palindromic repeats. **D** Number and accumulated length of dispersed repeats with different length. The line indicates the accumulated length, and the bar indicates the number of repeats

of *ycf1* located in the SSC region ranges from 4,051 to 4,150 bp. The IRa/LSC boundary is located between the *psbA* and *rpl2*. In the plastome of *P. trifolius*, *psbA* is located 99 bp from the boundary in the IRa, and *rpl2* is 489 bp from the boundary in the LSC. In *P. stipuleanatus*, *psbA* is 106 bp and *rpl2* is 484 bp from the boundary. In other *Panax* species, *psbA* is 104 bp from the boundary, and *rpl2* ranges from 483 to 493 bp away. Slight shifts in the positions of *psbA* and *rpl2* were also observed in the plastomes of *Aralia*.

Evolutionary pattern in *Panax* and related species

All *Panax* species are herbs typically found in canopy-shaded habitats with low light availability. Even key medicinal species such as *P. ginseng*, *P. quinquefolius*, and *P. notoginseng* require shading during cultivation. In contrast, the closely related genus *Aralia* consists predominantly of woody species that are exposed to direct sunlight, particularly those lineages most closely related to *Panax* (Fig. 2). To elucidate the evolutionary mechanisms underlying plastid adaptation to shade environments, this study conducted a comparative analysis of plastomes, which responsible for light capture and photosynthesis in plants.

Firstly, we employed a pairwise model to calculate selection pressures of the concatenated PCGs matrix among species, revealing that all d_N/d_S were all below 1 - even below than 0.5 - thereby indicating that plastid PCGs are generally subject to purifying selection and remain relatively conserved (Additional file 2: Fig. S4).

Due to structural and functional constraints on intact genes that may obscure detectable signals of selection, we further employed a site model to identify sites potentially under selection. Comparison between the M0 and M3 models revealed significant variation in ω across sites in 11 genes (*accD*, *cemA*, *clpP*, *matK*, *ndhF*, *psbD*, *rbcL*, *rpoA*, *ycf1*, *ycf15*, and *ycf2*) (Table 2). Subsequent analyses using the M8 vs. M7 models, combined with the Bayesian Empirical Bayes (BEB) method, identified 39 positively selected sites across 8 PCGs. These included two genes involved in photosynthesis (*ndhF* and *rbcL*), one associated with plastid self-replication (*rpoA*), and five with other functions (*clpP*, *matK*, *ycf1*, *ycf15*, and *ycf2*), comprising 7 sites from photosynthesis genes, 2 from replication-related genes, and 30 from other categories. Furthermore, we conducted protein structural and coevolutionary analyses on these eight genes harboring positively selected sites, revealing extensive residue - residue coevolution networks (Fig. 5; Additional file 1: Table S11). The y*cf1* exhibited the most complex network, with 128 unique residues forming 949 coevolving residue pairs, followed by *ndhF* (171 pairs), *ycf15* (120 pairs), *ycf2* (38 pairs), *matK* (31 pairs), *clpP* (9 pairs), *rbcL* (9 pairs), and *rpoA* (3 pairs). Notably, certain residues participated in multiple coevolving pairs, often spanning distinct protein domains and contributing to the structural complexity of the coevolutionary networks. For instance, in *clpP*, although residues V23 and I75 were not directly linked, both exhibited coevolution with D40 (V23 - D40 and D40 - I75). Remarkably, all 120 coevolving pairs in *ycf15* were derived from just 15 residues, suggesting the presence of localized coevolutionary hotspots. Mapping the positively selected and coevolving residues onto predicted 3D structures further revealed that, with the exception of *rpoA*, all analyzed genes harbored residues subjected to both evolutionary forces, highlighting potential sites of functional and evolutionary significance.

To further investigate genes potentially associated with shade tolerance, we designated *Panax* (representing shade-adapted taxa, marked as #1) and *Aralia* Clade B (composed primarily of sun-exposed taxa, marked as #2) as foreground branches for comparative evolutionary analysis. After excluding genes with extremely low d_S values that resulted in inflated ω estimates, a total of 59 PCGs were retained, comprising 32 photosynthesis-related genes, 17 self-replication genes, and 10 genes of other functional categories (Additional file 1: Table S12). Among these, 26 genes exhibited completely contrasting selection pressures between the two clades (Fig. 5; Additional file 1: Table S12). Specifically, 20 genes showed $\omega > 1$ in the *Panax* and $\omega < 1$ in *Aralia* Clade B, suggesting positive selection in *Panax* but purifying selection in *Aralia*. These included 9 photosynthesis-related genes, 6 self-replication genes, and 5 from other category. Conversely, 6 genes exhibited $\omega < 1$ in *Panax* and $\omega > 1$ in *Aralia* Clade B, comprising 4 photosynthesis-related genes, 1 self-replication gene, and 1 from other category. These contrasting patterns suggest divergent selective trajectories across functional gene categories between *Panax* and *Aralia* Clade B. Further analysis using RELAX revealed 9 genes (*atpB*, *ndhJ*, *petD*, *psaA*, *rpl20*, *rps8*, *ccsA*, *clpP*, and *ycf1*) with significantly intensified selection ($K > 1$) in the *Panax* or/and *Aralia* Clade B, all of which correspond to the subset with $\omega > 1$ in *Panax* and $\omega < 1$ in *Aralia*,

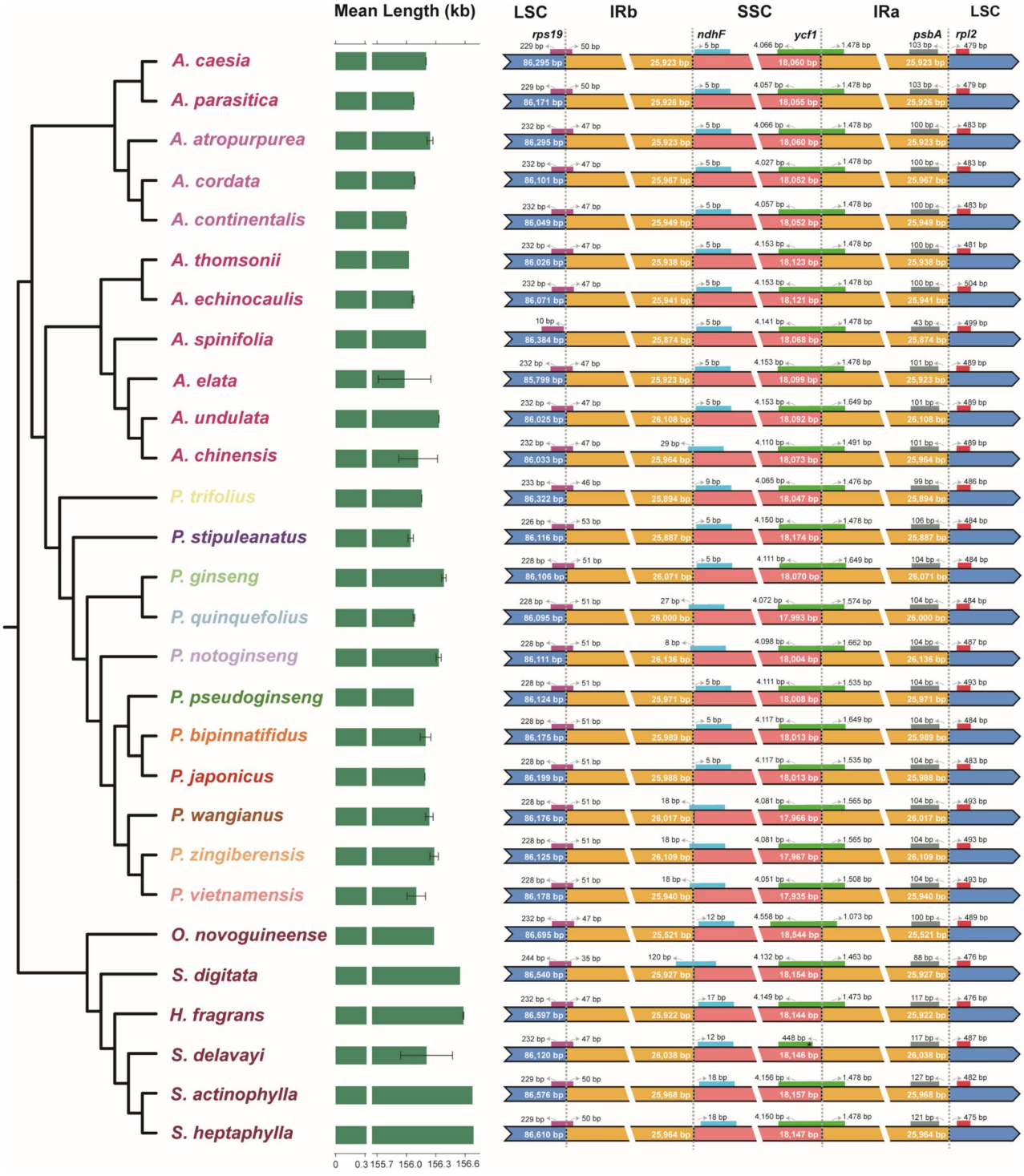

Fig. 4 Plastome structure of *Panax* and related species. Boxes represent boundary genes, with gene names labeled above. Arrows and numbers indicate the distance between each junction and its adjacent gene. Genes marked with an asterisk (*) are pseudogenes. The bar graph displays the mean length and standard error of sampled plastomes

reinforcing the hypothesis of adaptive divergence in response to contrasting light environments. In particular, *ndh* genes encode components of the NAD(P)H dehydrogenase complex, which plays a crucial role in optimizing photosynthetic efficiency under varying light intensities in terrestrial plants [30]. The *psaA* encodes a core subunit of Photosystem I and has been reported to be positively selected in both shade-tolerant and sun-loving

Table 2 Parameter estimation of the site model tests

Gene Name	Model	d.f.	2△lnL	LRT *P*-value	Positive selection sites (BEB)
accD	M2a vs. M1a	2	2.3986	0.3014	Dismiss
	M8 vs. M7	2	3.1076	0.2114	Dismiss
	M8 vs. M8a	1	2.3982	0.1220	Not allowed
cemA	M2a vs. M1a	2	10.1863	0.0061	Dismiss
	M8 vs. M7	2	10.0871	0.0065	Dismiss
	M8 vs. M8a	1	9.4115	0.0022	Not allowed
clpP	M2a vs. M1a	2	23.4009	0	A38**, V39**, D40**
	M8 vs. M7	2	22.1734	0	A38**, V39**, D40**
	M8 vs. M8a	1	22.0556	0	Not allowed
matK	M2a vs. M1a	2	5.6310	0.0599	R77*
	M8 vs. M7	2	7.4971	0.0236	R77*
	M8 vs. M8a	1	5.4382	0.0197	Not allowed
ndhF	M2a vs. M1a	2	4.2298	0.1206	F737*
	M8 vs. M7	2	14.2584	0.0008	A667*, F737*
	M8 vs. M8a	1	4.2256	0.0398	Not allowed
psbD	M2a vs. M1a	2	7.9034	0.0192	Dismiss
	M8 vs. M7	2	13.4001	0.0012	Dismiss
	M8 vs. M8a	1	7.9034	0.0049	Not allowed
rbcL	M2a vs. M1a	2	24.3466	0	Y226*, A328**, T439**
	M8 vs. M7	2	30.1167	0	E30*, Y226**, M320*, A328**, T439**
	M8 vs. M8a	1	24.1097	0	Not allowed
rpoA	M2a vs. M1a	2	3.3821	0.1843	Dismiss
	M8 vs. M7	2	5.9909	0.0500	E256*, N264*
	M8 vs. M8a	1	3.3523	0.0671	Not allowed
ycf1	M2a vs. M1a	2	93.7486	0	K205**, Y206**, K518**, G587*, L645**, L661*, L662*, D888*, D1480**, G1500*, T1532*
	M8 vs. M7	2	110.7480	0	K205**, Y206**, K518**, G587*, L645**, L661*, L662*, D888*, D1249*, D1480**, G1500*, T1532**
	M8 vs. M8a	1	93.1606	0	Not allowed
ycf15	M2a vs. M1a	2	28.4161	0	I64**, I65*, F67**, R69*, S72*, F75*, G76**, V78*, H83**
	M8 vs. M7	2	28.0286	0	I64**, I65**, F67**, R69*, S72**, H74*, F75*, G76**, V78*, H83**
	M8 vs. M8a	1	27.6975	0	Not allowed
ycf2	M2a vs. M1a	2	39.5836	0	S466*, S841*, Y850*, F1414**
	M8 vs. M7	2	30.7887	0	S466**, S841**, Y850**, F1414**
	M8 vs. M8a	1	30.7755	0	Not allowed

d.f. difference in the number of parameters in the ω distribution, *lnL* log-likelihood value, *LRT* likelihood ratio test, *BEB* Bayes empirical Bayes analysis

* *P* > 95%; ** *P* > 99%

rice cultivars [31]. The *rpl20* and *rps8* encode ribosome responsible for the translation of plastid-encoded genes, and their expression is closely linked to plant adaptation to different light environments [32, 33].

Additionally, to identify specific sites under positive selection in *Panax* and *Aralia* Clade B, we applied the branch-site model using each clade as the foreground branch. Notably, significant support for Model A was detected only when *Aralia* Clade B was designated as the foreground branch, with the gene *accD* showing evidence of episodic positive selection; however, no individual positively selected sites were identified under the Bayesian Empirical Bayes (BEB) analysis (Additional file 1: Table S13 and S14).

Collectively, we identified three genes (*matK*, *clpP*, and *ycf1*) that harbor positively selected sites, suggesting their potential roles in adaptive evolution. Moreover, *Panax* and *Aralia* Clade B exhibited opposing selection pressures on these genes. Notably, the divergence in selective patterns for *clpP* and *ycf1* between the two clades appeared to be increasingly pronounced, indicating their possible involvement in adaptation to contrasting light environments.

Species-specific sites and phylogenetic evaluation of hypervariable regions in *Panax*

To identify candidate molecular markers for distinguishing *Panax* species, we assessed the nucleotide diversity

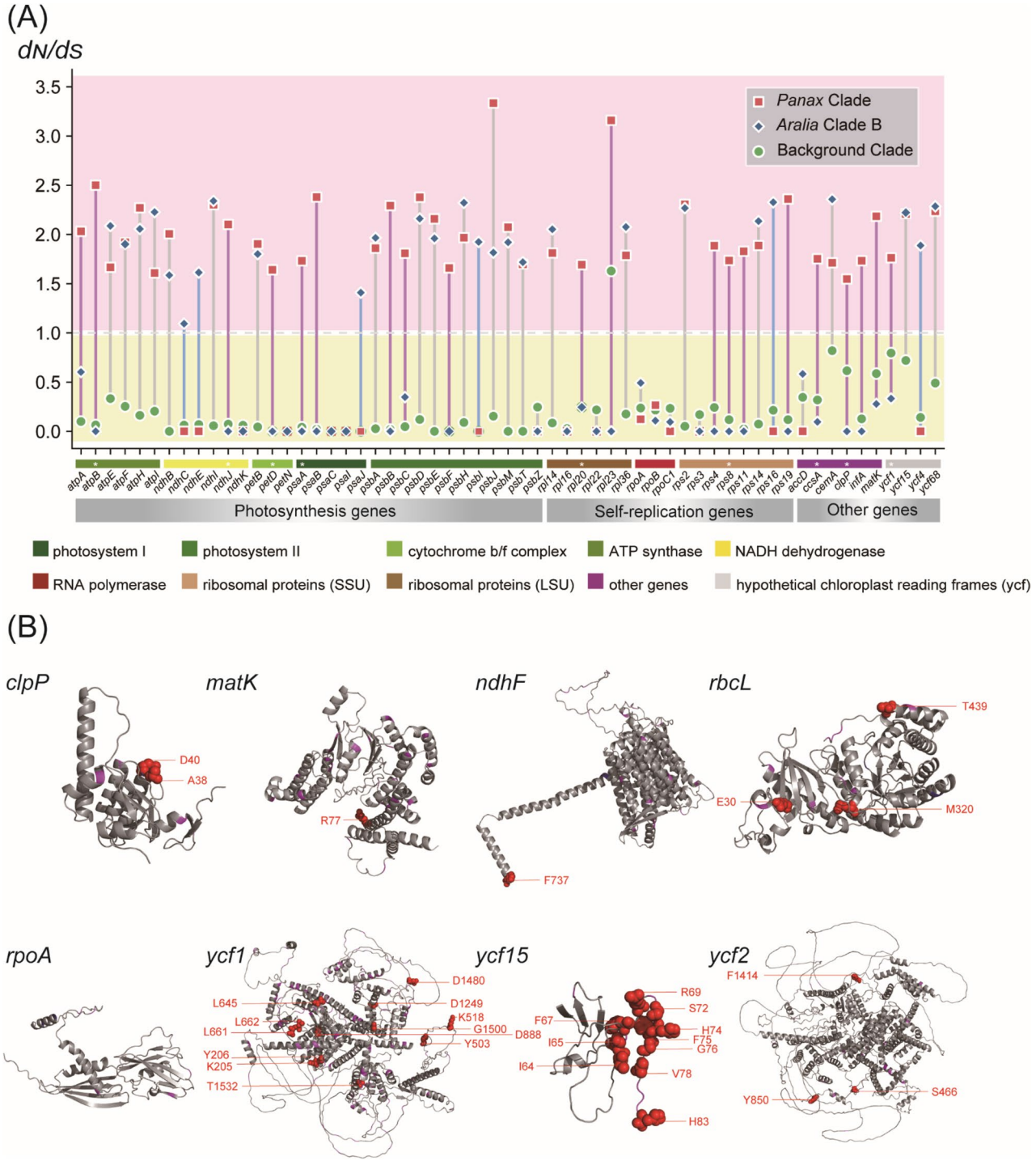

Fig. 5 Selective pressure in *Panax* plastomes estimated using branch model and site model. **A** The d_N/d_S values of retained protein-coding genes across different foreground clades under the branch model. Pink and yellow backgrounds represent positive and purifying selection, respectively, based on d_N/d_S values. A purple stick indicates that the gene is under positive selection in *Panax* but under purifying selection in *Aralia*. Among these genes under intensified selection are marked with an asterisk (*) in the color blocks below. A blue stick indicates that the gene is under positive selection in *Aralia* but under purifying selection in *Panax*. The color blocks at the bottom represent gene functions consistent with those shown in Fig. 1. **B** Distribution of positively selected and co-evolved residues in the *clpP*, *matK*, *ndhF*, *rbcL*, *rpoA*, *ycf1*, *ycf15*, and *ycf2* proteins. All tertiary protein structures were predicted by homology modeling based on *Osmoxylon novoguineense*. Magenta indicates co-evolved residues, blue indicates positively selected residues, and red marks residues under both, with appropriate labels

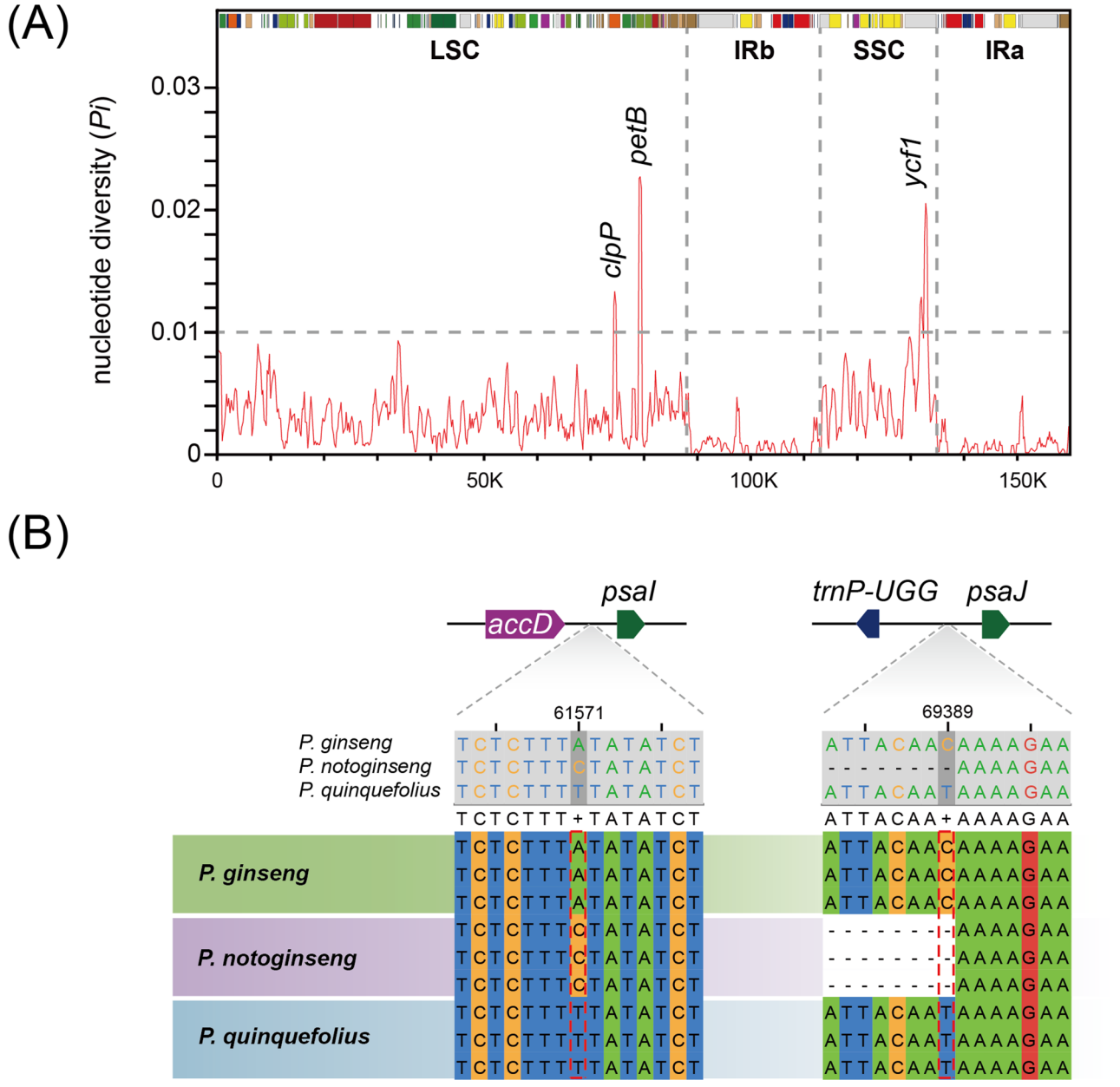

Fig. 6 Species-specific sites in three important medicinal *Panax* species. **A** Nucleotide diversity across *Panax* plastomes. Colored blocks above indicate gene locations and functional categories, consistent with Fig. 1. **B** Two species-specific sites among *P. ginseng*, *P. quinquefolius*, and *P. notoginseng*

(*Pi*) across the plastid genomes of 154 *Panax* accessions (Fig. 6A). The averaged nucleotide diversity values for the LSC, SSC, and IR regions were 0.0033, 0.0053, and 0.0009, respectively. Notably, the *clpP*, *petB*, and *ycf1* genes and their flanking regions exhibited relatively high variability ($Pi > 0.01$), suggesting their potential as candidate regions for molecular marker development in *Panax*. Given that *P. ginseng*, *P. quinquefolius*, and *P. notoginseng* are widely used traditional medicinal plants, often subjected to processing that leads to DNA degradation, short DNA barcodes are particularly valuable for their authentication. By comparing plastid genomes from 47 accessions of *P. ginseng*, 32 of *P. quinquefolius*, and 16 of *P. notoginseng* from diverse geographic origins, we obtained two single-nucleotide sites that reliably differentiate all three species (Fig. 6B). The first, located at position 61,571 (between *accD* and *psaI*), is an A in *P. ginseng*, a C in *P. notoginseng*, and a T in *P. quinquefolius*. The second site, at position 69,389 (between *trnP*-UGG and *psaJ*), is a C in *P. ginseng*, gap in *P. notoginseng*, and a T in *P. quinquefolius*. These loci provide robust species-specific markers for the authentication of medicinal *Panax* products.

To further assess the phylogenetic utility of these hypervariable regions, we reconstructed maximum likelihood trees based on each locus (*clpP*, *petB*, and *ycf1*) as well as their concatenated dataset (Additional file 2: Fig. S5). Substantial differences in sequence characteristics were observed among loci (Additional file 1: Table S15), with *ycf1* (5586 bp) containing a markedly higher number of parsimony-informative sites (612) compared to *clpP* (33) and *petB* (14).

Consistent with this pattern, trees based on *clpP* and *petB* showed extremely low bootstrap support (mean BS = 11.2 and 7.6, respectively), with fewer than 10% of nodes supported at ≥ 70, resulting in poorly resolved topologies (Additional file 1: Table S15). In contrast, *ycf1* exhibited markedly improved support (mean BS = 40.9), and the concatenated dataset showed similar performance. Notably, both *ycf1* and the concatenated dataset provided strong resolution at the species level, with most interspecific relationships receiving high bootstrap support (often close to 100) and showing overall topological congruence with the plastome-based phylogeny.

Discussion

In this study, 31 plastomes of *Panax quinquefolius* were newly assembled, and a total of 123 plastomes representing all currently accepted species of the genus *Panax* [4], except the narrowly distributed and endangered *P. sokpayensis* from Sikkim [34], were re-annotated. Each species was represented by at least two individuals except *P. pseudoginseng*, including 47 individuals of *P. ginseng*, 32 of *P. quinquefolius*, and 16 of *P. notoginseng*, all of which are widely used traditional Chinese medicine. A pan-plastome analysis was conducted across *Panax*. The plastomes exhibited the typical quadripartite circular structure characteristic of angiosperms, comprising a large single-copy (LSC) region, a small single-copy (SSC) region, and a pair of inverted repeats (IRs) [7]. Genome lengths ranged from 155,984 to 156,548 bp, with a variation of only 564 bp and a consistent GC content of approximately 38%. A total of 3,568 variant sites were identified in the pan-plastome of *Panax*, which is fewer than those reported in the herbaceous *Petrocosmea* (4,139) [35], but more than in *Gossypium* (1,809) and *Capsicum* (756) [15, 36]. The observed preference for A/U-ending codons is consistent with the overall AT-rich nature of plastid genomes. Such codon usage bias is commonly attributed to mutational pressure and/or selection acting on translational efficiency. This pattern is consistent with codon usage bias reported in other angiosperm plastomes [37–39]. The conserved RSCU patterns across species further suggest a stable evolutionary constraint on codon usage in *Panax* plastomes. No large-scale structural rearrangements or recombination events were detected; only minor shifts at the boundaries between IR and SC regions were observed, indicating that the plastomes of *Panax* are relatively conserved. SSRs were mainly distributed in IGS regions, with fewer loci in CDS and introns. This pattern likely reflects stronger functional constraints on CDS regions and relatively relaxed selection in IGS regions [40], and is conserved across all examined species. Additionally, 26 plastomes of the closely related genus *Aralia*, which composed mostly of woody species exposed to direct sunlight [29], were re-annotated. Comparative plastomics between *Panax* and *Aralia* were performed to investigate the potential evolutionary impact of shade tolerance on plastome evolution in *Panax*.

Plastid phylogenomics, species-specific sites, and marker evaluation in *Panax*

Two rounds of whole-genome duplication, along with geographic isolation, ecological shifts, and a recent radiation event in *Panax* centered in the Hengduan Mountains of western China, have promoted species diversification within this genus. Furthermore, due to the prevalence of morphological intermediates in rhizomes and leaflets [41], species delimitation and phylogenetic relationships within *Panax* remain controversial. Therefore, this study reconstructed the phylogenies within the *Panax* using complete plastomes and concatenated plastid PCGs. The analysis revealed that *P. trifolius* represents the earliest diverging lineage within *Panax*, followed by *P. stipuleanatus*, *P. ginseng* and *P. quinquefolius* forming a well-supported monophyletic clade. *P. notoginseng* appears to be a distinct species that diverges prior to the diversification of the remaining taxa, which is consistent with the findings of Xia et al. [42], which inferred a similar topology based on 11 plastomes. However, the authors sampling included only one individual per species, limiting the representativeness of their results. In contrast, our expanded sampling revealed that *P. pseudoginseng*, *P. bipinnatifidus*, and *P. japonicus* could not be clearly distinguished based on plastid data, nor could *P. zingiberensis* and *P. vietnamensis*. These findings suggest that *Panax* underwent a rapid radiation [43], leading to insufficient plastome divergence to resolve certain closely related species. Previous studies have also reconstructed the phylogenetic relationships within *Panax* using 358 nuclear orthologous loci and complete plastome [1]. Consistently, *P. trifolius* was identified as the earliest diverging lineage. However, the phylogenetic positions of *P. japonicus* and *P. pseudoginseng* in both nuclear and plastid trees differ markedly from those reconstructed in our study, potentially due to limited sampling or misidentification of species. Despite differences in sampling, these studies collectively support the notion that, aside from the well-resolved clades comprising *P. trifolius*, *P. stipuleanatus*, *P. ginseng*, and *P. quinquefolius*, the relationships among

the remaining *Panax* species remain complex and are likely confounded by historical gene flow [1]. Although discrepancies exist between plastid and nuclear phylogenies, all available evidence supports the recognition of *P. ginseng*, *P. quinquefolius*, and *P. notoginseng* as distinct and well-defined species. This, in turn, provides a solid foundation for mining species-specific molecular markers from the plastome for accurate identification.

The high economic and medicinal value of ginseng has led to frequent occurrences of adulteration and counterfeiting in the marketplace, with even toxic species such as *Mirabilis jalapa* and *Phytolacca acinosa* being misrepresented as ginseng substitutes [44, 45]. Although previous studies on *Panax* taxonomy and identification have identified numerous candidate molecular markers [42, 46, 47], practical application remains challenging. *Panax* species are typically sold in processed forms such as dried roots, powders, and liquid extracts, and DNA degradation during storage, decoction, or heating in traditional medicine preparations often compromises the effectiveness of conventional molecular markers. Additionally, significant price differences among *Panax* species have led to widespread substitution even within this genus. Despite all containing ginsenosides, species such as *P. ginseng*, *P. quinquefolius* and *P. notoginseng* differ markedly in their pharmacological properties [3, 42]. Therefore, in this study, we focused on identifying highly specific, single-nucleotide molecular markers to distinguish these three major medicinal species. Based on a comprehensive sampling of 47 individuals of *P. ginseng*, 32 of *P. quinquefolius*, and 16 of *P. notoginseng* from diverse geographic origins, we identified two species-specific sites in the plastome that allow complete differentiation among the three species. Notably, at position 61,571, the base type of all *P. ginseng* samples was A, the base type of *P. notoginseng* samples was C, and the base type of *P. quinquefolius* samples was T, demonstrating that a single nucleotide site can serve as a reliable marker for precise species identification.

Although several hypervariable regions were identified based on nucleotide diversity, their phylogenetic performance differed substantially. The short loci (*clpP* and *petB*), despite relatively high variability, provided limited phylogenetic resolution and low bootstrap support, likely due to their small number of informative sites. In contrast, the long *ycf1* region showed strong species-level resolution and high support, largely consistent with the plastome phylogeny. The concatenated dataset further improved overall support, highlighting the importance of increasing total phylogenetic signal [48, 49]. These results indicate that both sequence variability and sufficient length are critical for effective marker selection, and that *ycf1* represents a promising candidate for species discrimination in *Panax*.

Adaptation to shade environment of *Panax* plastomes

Photosynthesis is a fundamental biological process that converts light energy into chemical energy, serving as the primary source of food and oxygen for nearly all life forms [50]. It is intricately linked to key global issues such as food security, energy production, ecological stability, and environmental sustainability [50]. Plastids, as the primary organelles responsible for photosynthesis, function both as light energy receptors and as converters, making them central to the study of plant responses to light environments [51, 52]. *Panax*, which typically grows in understory forest habitats and requires shade when cultivated artificially, represents an important shade-tolerant species [28]. Light intensity not only influences its growth but also plays a critical role in the biosynthesis of ginsenosides, directly affecting both yield and medicinal quality [26, 27]. Consequently, *Panax* species provide valuable models for investigating the mechanisms underlying plant adaptation to low-light environments. In this study, the phylogenetic relationships within *Panax* and its closely related genus *Aralia*, reconstructed using plastome data, align with previous findings. Notably, *Aralia* is not monophyletic and is divided into two major clades: clade A, comprising both woody and herbaceous species, and clade B, consisting exclusively of woody species that are more closely related to *Panax* [29, 53, 54]. Therefore, clade B of *Aralia* serves as a natural control for exploring the evolutionary mechanisms underlying shade tolerance in *Panax*.

We compared the ω values of all PCGs concatenated matrix between any two species of *Panax* and *Aralia*, and found that all values were below 0.5. This pattern is consistent with findings from plastome analyses across 773 angiosperms [55], supporting the notion that plastid genes are generally subject to purifying selection due to their fundamental roles in photosynthesis. In this study, we identified eight PCGs contained positively selected sites, including two photosynthesis-related genes (*ndhF* and *rbcL*), one gene involved in self-replication (*rpoA*), and five genes of other or unknown functions (*clpP*, *matK*, *ycf1*, *ycf15*, and *ycf2*). The *ndhF* gene encodes a subunit of the NDH complex, which mediates a secondary electron transport pathway in the PSI cycle and is integral to the overall photosynthetic process [30]. Functional studies in *Nicotiana tabacum* (tobacco) mutants have shown that *ndhF* is responsive to light intensity and plays a role in photosynthetic efficiency [56]. The *rbcL* gene, encoding the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), is not only a widely used molecular marker in plant phylogenetics but also directly involved in carbon fixation during photosynthesis [57, 58]. Previous studies have shown that *rbcL* is subject to positive selection in aquatic plants such as *Potamogeton*, likely as an adaptive response to low-light

underwater environments [59]. Moreover, *rbcL* has been implicated in the broad ecological adaptation of *Ilex* across a wide range of habitats, from tropical regions to temperate alpine zones [60]. In addition, *matK* and *ycf2* have been identified as positively selected during the adaptation of *Chrysosplenium* to shaded habitats [61], while *ycf1* has been linked to light adaptation in alpine plants such as *Saxifraga* [62]. Therefore, the eight PCGs containing positively selected sites identified in this study may contribute to the light habitat differentiation between *Panax* and *Aralia*, either directly through their involvement in photosynthetic processes or indirectly via related functional pathways.

To further investigate the impact of light habitat on plastome evolution, we separately designated the *Panax* and *Aralia* Clade B as foreground branches and conducted a comparative selection analysis, following a strategy similar to that used by Iverson et al. [63], which focused on mitochondrial genome evolution in high- and low-altitude animals. In total, we identified 26 genes exhibiting completely opposite selection pressures between *Panax* and *Aralia* Clade B, encompassing genes related to photosynthesis, self-replication, and other functions. Among these, 20 genes were positively selected in *Panax*, while 6 showed positive selection in *Aralia* Clade B. Notably, nine genes (*atpB*, *ndhJ*, *petD*, *psaA*, *rpl20*, *rps8*, *ccsA*, *clpP*, and *ycf1*) were found to be under intensified selection. These genes are fundamental to the maintenance of plastid functions and have also been implicated in light-environment adaptation. Collectively, these nine genes may be involved in the adaptive divergence between the shade-tolerant *Panax* and the light-preferring *Aralia* Clade B, and may exert a lasting influence on their respective light habitat differentiation.

Conclusion

In this study, we re-annotated 123 plastomes representing all currently accepted species of the genus *Panax*, except the narrowly distributed and endangered *P. sokpayensis* from Sikkim. We conducted a comparative analysis of the plastomes and found that the plastomes of *Panax* are highly conserved. We reconstructed the phylogenetic relationships of multiple accessions and confirmed that the three economically important species (*P. ginseng*, *P. quinquefolius*, and *P. notoginseng*) are genetically distinct and well-defined. We identified two single-nucleotide species-specific markers that can reliably distinguish *P. ginseng*, *P. quinquefolius*, and *P. notoginseng*. In addition, three highly variable regions were identified as potential candidates for molecular marker development in *Panax*. When we use the natural contrast between the shade-tolerant *Panax* and the sun-loving *Aralia*, we found nine genes potentially involved in the adaptive divergence, and eight genes harboring positively selected sites along with extensive residue-residue coevolution networks in *Panax*. Therefore, these findings serve as a foundation for molecular-based screening, accurate identification, and conservation of valuable medicinal species. They also provide new insights into how plant responses to variable light conditions, supporting improved management and utilization of economically important ginseng resources.

Materials and methods

Data recourse

According to The Plant List [4], *Panax* comprises 12 recognized species and one variety: *P. bipinnatifidus*, *P. bipinnatifidus* var. *angustifolius*, *P. ginseng*, *P. japonicus*, *P. notoginseng*, *P. pseudoginseng*, *P. quinquefolius*, *P. sokpayensis*, *P. stipuleanatus*, *P. trifolius*, *P. vietnamensis*, *P. wangianus*, and *P. zingiberensis*. In this study, we collected data from 11 recognized species, excluding *P. sokpayensis*, which is narrowly distributed in Sikkim, India (part of the Himalayan region), and currently classified as endangered. To robustly resolve interspecific relationships within *Panax*, multiple accessions were included for each species, except for *P. pseudoginseng*, which was represented by a single accession due to limited availability. Sampling was particularly intensive for three economically significant species: 47 accessions of *P. ginseng*, 32 accessions of *P. quinquefolius*, and 16 accessions of *P. notoginseng*. The genome sequencing data for the 31 *P. quinquefolius* accessions were downloaded from globally distributed sources, including both cultivated and wild populations. To further investigate potential adaptation of *Panax* species to shade environments, we also downloaded 26 plastomes from the closely related genus *Aralia*. *Aralia* is the genus most closely related to *Panax* and comprises predominantly sun-loving, large rhizomatous perennials. This relationship provides an ideal natural comparative system for investigating plastome adaptations to contrasting light environments. Additionally, based on the phylogeny of Araliaceae, we selected eight samples from the genera *Schefflera*, *Heteropanax*, and *Osmoxylon*—all of which are closed to both *Aralia* and *Panax*—to serve as outgroups [64, 65]. Sample accession numbers and related information are provided in Additional file 1: Table S1.

Plastome assembly and annotation

The whole-genome resequencing data (150 bp paired-end, Illumina platform) for 31 *P. quinquefolius* accessions were obtained from the National Genomics Data Center [66]. The raw data were first subjected to quality assessment using FastQC v0.12.1 [67]. Low-quality bases and adapter were trimmed by Trimmomatic v0.39 with the following parameters: ILLUMINACLIP: TruSeq3-PE.fa:2:30:10, HEADCROP:3, LEADING:20, TRAILING:24, CROP:147, SLIDINGWINDOW:4:15, AVGQUAL:28,

MINLEN:36, TOPHRED64 [68]. After quality control, the clean data were *de novo* assembled using GetOrganelle v1.7.7.1 [69]. Assembly graphs were manually inspected and adjusted using Bandage v0.8.1, based on sequencing depth and connectivity [70]. To facilitate comparison of plastomes between *Panax* and its closely related genera, all plastid genomes were annotated using PGA (Plastid Genome Annotator), with the plastome of *Amborella trichopoda* (AJ506156.2) as the reference [71]. CDS sequences were extracted using CPStools v2.0.2 [72], and gene annotations were manually refined, following multiple sequence alignment performed with MAFFT v7.505 to ensure consistency and accuracy [73]. Finally, the reliable annotation information of each plastome was obtained and submitted to figshare [74].

Phylogeny reconstruction

We reconstructed the phylogenetic relationships among *Panax* and *Aralia* species based on complete plastome and coding sequences (CDS), respectively. CPStools v2.0.2 was used to automatically extract and merge shared CDS regions from the annotated file [72]. The merged CDS and full-length sequences were aligned separately using MAFFT v7.505, and gap regions were removed using Gblocks v0.91b [75]. Phylogenetic trees were reconstructed using RAxML-NG [76] with the GTR + G model and 1000 bootstrap replicates. The resulting trees were visualized using iTOL [77].

Comparative plastome analysis

To assess plastome-wide variation within *Panax*, DNA polymorphism was analyzed across all ginseng samples. Specifically, complete plastomes from all *Panax* species were first aligned using MAFFT v7.505 [73]. In addition to the genus-wide analysis, intra-specific polymorphisms were separately calculated for multiple individuals of *P. ginseng*, *P. quinquefolius*, and *P. notoginseng*. Variants were identified using DnaSP v6 [78], and their distribution across the plastome was visualized with the circlize package in R v4.4.1 [79]. The Relative Synonymous Codon Usage (RSCU) was computed using CondonW (https://codonw.sourceforge.net/) and visualized through the RSCU-Plot Shiny application (https://pcg-lab.shinyapps.io/RSCU-Plot/).

Osmoxylon novoguineense (MK943807.1), a species closely related to both *Panax* and *Aralia*, was used as the reference genome. One plastome from each species was selected to analyze structural variation across *Panax* and *Aralia*. In addition, sequence conservation among these plastomes was assessed using mVISTA, with the Shuffle-LAGAN algorithm [80].

Repeats were examined in the representative plastomes of *Panax* and *Aralia*. Simple sequence repeats (SSRs) were identified using the SSR subcommand in CPStools v2.0.2 [72], with minimum repeat thresholds set to 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexanucleotide motifs, respectively. Tandem repeats were detected using Tandem Repeats Finder v4.09 with default parameters [81]. Dispersed repeats were identified using REPuter [82], with the minimum repeat length set to 20 bp.

All comparative analyses were conducted based on whole-genome sequence alignments and were independent of annotation transfer from the reference genome.

Evolutionary analysis

To investigate the adaptive evolution of plastomes in *Panax* to shaded habitats, we conducted a comprehensive comparative analysis between *Panax* and its sun-exposed close relatives in the genus *Aralia* across the following six aspects: (1) genome-wide genetic diversity analysis of plastomes; (2) pairwise comparisons of non-synonymous (d_N), synonymous (d_S) substitution rates, and the d_N/d_S ratio (ω) for protein-coding genes (PCGs) between *Panax* and *Aralia*; (3) detection of positive selection for each PCG among different lineages using PAML v4.9j under the branch model; (4) identification of positively selected sites using PAML v4.9j under the site model; (5) detection of positive selection acting on specific sites along foreground branches using PAML v4.9j under the branch-site model; and (6) assessment of relaxed or intensified natural selection on foreground branches using HYPHY v2.5.42 under the RELAX model. The detailed procedures are as follows:

(1) These plastomes were aligned using MAFFT v7.505 [73], and highly variable regions were filtered out using Gblocks v0.91b [75]. Then, the Nucleotide diversity (*Pi*) was calculated using DnaSP v6 with window size: 600 bp and step size: 200 bp [78].
(2) We first used CPStools v2.0.2 to extract coding region of PCGs from 23 representative plastomes of *Panax*, *Aralia*, and *Osmoxylon novoguineense* [72]. The extracted sequences were aligned using the codon alignment (translation align) mode in PhyloSuite v1.2.3, and stop codons were removed [83]. Gap regions were further filtered manually to generate reliable alignments for each PCG. These individual alignments were concatenated into a supermatrix using FASconCAT-G v1.05.1 [84], which was then converted to PAML format using EasyCodeML v1.41 [85]. Finally, pairwise values of d_N, d_S, and the d_N/d_S ratio (ω) were calculated using the pairwise mode in PAML (runmode = -2) [86].
(3) Detection of positive selection among different lineages. Foreground and background branches were specified in the phylogenetic tree using EasyCodeML v1.41 [85], with the *Panax* and *Aralia* Clade B

designated as branch #1 and branch #2, respectively. For each PCG, d_N, d_S, and the d_N/d_S ratio (ω) were estimated using the codeml subprogram in PAML v4.9j under the branch model (runmode = 0) [86]. The M0 model (model = 0) assumes a uniform ω ratio across all clades, while the M2 model (model = 2) allows for different ω values between foreground and background clades. Model comparisons were assessed using likelihood ratio tests (LRTs) based on the χ^2 distribution.

(4) Identification of positively selected sites. Each PCG was initially analyzed using PAML v4.9j under the M0 and M3 models to evaluate variation in the d_N/d_S ratio (ω) among sites [86]. The M0 model (NSsites = 0) served as the null hypothesis, assuming a uniform ω across all sites, while the M3 model (NSsites = 3) allowed ω to vary discretely among sites. A significantly better fit of M3 compared to M0 ($P < 0.05$) indicated that selective pressure varies across sites [86]. Three pairs of models (M2a vs. M1a, M8 vs. M7, and M8a vs. M8) were also analyzed using PAML v4.9j [86]. Among these, M2a and M8 are models that allow for positive selection, whereas M1a, M7, and M8a serve as their respective null models. A significantly better fit of the positive selection model over its corresponding null model ($P < 0.05$) was taken as evidence that the gene is subject to positive selection. Furthermore, we performed structural and coevolutionary analyses of genes under positive selection. For PCGs identified as being under positive selection, three-dimensional (3D) protein structures were predicted using AlphaFold3 [87], based on the reference sequence MK943807.1. Coevolutionary analyses integrating the predicted 3D structures, phylogenetic tree, and aligned matrix were conducted using CAPS v2.0 with default parameters [88]. These protein structures were visualized using the educational version of PyMOL [89], where positively selected sites, coevolving residues, and their overlapping regions were mapped and displayed.

(5) Detection of positive selection on specific sites in foreground lineages. Positive selection acting on specific codon sites in foreground branches of each PCG was assessed using the branch-site model in the codeml implemented in PAML v4.9j [86]. Model A (model = 2, NSsites = 2, fix_omega = 0, omega = 2) assumes that a subset of sites is under positive selection ($\omega > 1$) in the foreground branch, while all other sites and background branches are under purifying selection or neutrality. The corresponding null model (model = 2, NSsites = 2, fix_omega = 1, omega = 1) constrains ω in the foreground branch to 1, thereby disallowing positive selection. A significantly better fit of Model A compared to its null hypothesis ($P < 0.05$), based on likelihood ratio tests (LRTs) evaluated against a chi-square (χ^2) distribution, indicates the presence of positive selection in the foreground lineage. Codon sites under positive selection were further identified using the Bayesian Empirical Bayes (BEB) method, with reference to sequence MK943807.1.

(6) Assessment of relaxed or intensified natural selection on foreground branches using the RELAX model in HYPHY v2.5.42 [90]. *Panax* and *Aralia* Clade B were designated as foreground clades FG1 and FG2, respectively. The analysis involved a likelihood ratio test (LRT) comparing the null model (K = 1), which assumes equal selection strength between foreground and background lineages, with alternative models where $K \neq 1$. A significant *P*-value ($P < 0.05$) indicates a deviation from the null model: $K < 1$ suggests relaxed selection, whereas $K > 1$ indicates intensified selection in the foreground branch.

Identification of species-specific sites and phylogenetic evaluation of hypervariable regions in *Panax*

Panax ginseng, *P. quinquefolius*, and *P. notoginseng* are three major species widely used in traditional Chinese medicine. To identify species-specific sites in *P. ginseng*, all plastomes of the three species were aligned using MAFFT v7.505 [73]. Polymorphic sites were then detected using DnaSP v6 [78]. After manual curation, we retained only those sites that showed complete differentiation among the three species. The identified species-specific sites were subsequently visualized using Jalview2 [91]. We quantified the number of informative sites in the alignment matrices of individual hypervariable genes and concatenated datasets using AMAS [92], and reconstructed their phylogenetic trees using RAxML-NG [76] under the GTR + G model with 1,000 bootstrap replicates; the resulting trees were visualized in iTOL [77].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-026-08908-w.

Additional file 1: Table S1. Samples used and general information of plastomes in this study. Table S2. Protein-coding genes and functional classification encoded by *Panax* plastomes. Table S3. Nucleotide variants in plastomes of all samples in *Panax*. Table S4. The number of different lengths SSR (simple sequence repeat) in 23 representative samples. Table S5. Detailed annotation of simple sequence repeats (SSRs) identified in 23 representative samples. Table S6. Summary of the distribution of SSRs across CDS, introns, and IGS regions in 23 representative samples. Table S7. The number of different lengths tandem repeats and the accumulated length in 23 representative samples. Table S8. The general information of tandem repeats in 23 representative samples. Table S9. The number of the different type dispersed repeats in 23 representative samples. Table S10. The number of the different length dispersed repeats and the accumulated length in 23 representative samples. Table S11. Information about intra-molecular co-evolved residue pairs for protein sequences containing,

and overlapping residues in co-evolution and adaptive evolution analyses. Table S12. Parameter estimation of the branch model tests and the RELAX selection tests. Table S13. Parameter estimation of the branch-site model tests. Table S14. The positive selection sites identified by branch-site model. Table S15. Comparison of sequence characteristics and phylogenetic performance of selected hypervariable loci

Additional file 2: Fig. S1 The bar plot of the Relative Synonymous Codon Usage (RSCU) values for each amino acid, grouped by species. Codons are color-coded, and their corresponding amino acids are labeled below the plot.Fig. S2 Maximum likelihood tree reconstructed based on coding sequences. Fig. S3 Sequence conservation and identity among the representative plastomes. Fig. S4 Pairwise selective pressure among *Panax* and Aralia plastomes. Fig. S5 Maximum likelihood phylogenetic trees reconstructed from individual hypervariable loci and their concatenated dataset. (A) clpP, (B) petB, (C) ycf1, and (D) concatenated dataset (clpP + petB + ycf1)

Authors' contributions

XLW wrote the initial draft, performed the formal analysis, developed the methodology, and created the visualizations. SG contributed to the formal analysis and provided resources. YYD, XJS, YCZ, and SPY conducted the formal analysis; SPY additionally provided resources. JKL contributed resources. WZ conceived the study concept and contributed to manuscript review and editing. HXY conceptualized the study, supervised the research, and participated in manuscript review and editing. SLK designed the study and methodology, curated the data, secured funding, conducted the investigation, and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (Grant numbers 32200187 and 82173936).

Data availability

All plastid genome sequences used in this study are publicly available at Figshare (https://doi.org/10.6084/m9.figshare.29588627.v1).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 5 March 2026 / Accepted: 30 April 2026
Published online: 04 May 2026

References

1. Zhang MH, Nie ZL, Fairbanks RA, Liu J, Literman R, Johnson G, Handy S, Wen J. Phylogenomic insights into species relationships, reticulate evolution, and biogeographic diversification of the ginseng genus *Panax* (Araliaceae), with an emphasis on the diversification in the Himalayan-Hengduan Mountains. J Syst Evol. 2024;63(1):99–114.
2. Shi FX, Li MR, Li YL, Jiang P, Zhang C, Pan YZ, Liu B, Xiao HX, Li LF. The impacts of polyploidy, geographic and ecological isolations on the diversification of *Panax* (Araliaceae). BMC Plant Biol. 2015;15:297.
3. Hou M, Wang R, Zhao S, Wang Z. Ginsenosides in *Panax* genus and their biosynthesis. Acta Pharm Sin B. 2021;11(7):1813–34.
4. The Plant List. http://www.theplantlist.org/. Accessed 6 Dec 2024.
5. Liu J, Dong Q, Du G, Wang J, An Y, Liu J, Su J, Xie H, Yin J. Identification of metabolites in plasma related to different biological activities of *Panax ginseng* and American ginseng. Rapid Commun Mass Spectrom. 2022;36(4):e9219.
6. UN Comtrade Database. https://comtradeplus.un.org/. Accessed 5 Apr 2025.
7. Tonti-Filippini J, Nevill PG, Dixon K, Small I. What can we do with 1000 plastid genomes? Plant J. 2017;90(4):808–18.
8. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci. 1987;84(24):9054–8.
9. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17(1):134.
10. Drouin G, Daoud H, Xia J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol Phylogenet Evol. 2008;49(3):827–31.
11. Wang J, Kan S, Kong J, Nie L, Fan W, Ren Y, Reeve W, Mower JP, Wu Z. Accumulation of large lineage-specific repeats coincides with sequence acceleration and structural rearrangement in *Plantago* plastomes. Genome Biol Evol. 2024;16(8):evae177.
12. Wang J, Kan S, Liao X, Zhou J, Tembrock LR, Daniell H, Jin S, Wu Z. Plant organellar genomes: much done, much more to do. Trends Plant Sci. 2024;29(7):754–69.
13. Davis CC, Xi Z, Mathews S. Plastid phylogenomics and green plant phylogeny: almost full circle but not quite there. BMC Biol. 2014;12:1–4.
14. Li HT, Yi TS, Gao LM, Ma PF, Zhang T, Yang JB, Gitzendanner MA, Fritsch PW, Cai J, Luo Y, et al. Origin of angiosperms and the puzzle of the Jurassic gap. Nat Plants. 2019;5(5):461–70.
15. Yan XL, Kan SL, Wang MX, Li YY, Tembrock LR, He WC, Nie LY, Hu GJ, Yuan DJ, Ma XF, et al. Genetic diversity and evolution of the plastome in allotetraploid cotton (*Gossypium* spp). J Syst Evol. 2024;62(6):1118–36.
16. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. PLoS ONE. 2011;6(5):e19254.
17. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: from gene to genome. Biol Rev. 2014;90(1):157–66.
18. Zeng CX, Hollingsworth PM, Yang J, He ZS, Zhang ZR, Li DZ, Yang JB. Genome skimming herbarium specimens for DNA barcoding and phylogenomics. Plant Methods. 2018;14:43.
19. Yue J, Zuo Z, Huang H, Wang Y. Application of identification and evaluation techniques for ethnobotanical medicinal plant of genus *Panax*: a review. Crit Rev Anal Chem. 2021;51(4):373–98.
20. Li M, Jiang C, Pui-Hay P, Shaw P-C, Yuan Y. Molecular identification of traditional medicinal materials. In: *Molecular Pharmacognosy*. Edited by Huang LQ. Singapore: Springer; 2019. p. 13–39.
21. Yeo D, Srivathsan A, Meier R. Longer is not always better: optimizing barcode length for large-scale species discovery and identification. Syst Biol. 2020;69(5):999–1015.
22. Gonçalves LT, Françoso E, Deprá M. Shorter, better, faster, stronger? Comparing the identification performance of full-length and mini-DNA barcodes for apid bees (Hymenoptera: Apidae). Apidologie. 2022;53(5):55.
23. de Vries J, Stanton A, Archibald JM, Gould SB. Streptophyte terrestrialization in light of plastid evolution. Trends Plant Sci. 2016;21(6):467–76.
24. Bock R. Engineering plastid genomes: methods, tools, and applications in basic research and biotechnology. Annu Rev Plant Biol. 2015;66:211–41.
25. Forsythe ES, Sharbrough J, Havird JC, Warren JM, Sloan DB. CyMIRA: the cytonuclear molecular interactions reference for *Arabidopsis*. Genome Biol Evol. 2019;11(8):2194–202.
26. Kim NH, Jayakodi M, Lee SC, Choi BS, Jang W, Lee J, Kim HH, Waminal NE, Lakshmanan M, van Nguyen B, et al. Genome and evolution of the shade-requiring medicinal herb *Panax ginseng*. Plant Biotechnol J. 2018;16(11):1904–17.
27. Jung JH, Kim HY, Kim HS, Jung SH. Transcriptome analysis of *Panax ginseng* response to high light stress. J Ginseng Res. 2020;44(2):312–20.
28. Zhang YX, Niu YQ, Wang XF, Wang ZH, Wang ML, Yang J, Wang YG, Zhang WJ, Song ZP, Li LF. Phenotypic and transcriptomic responses of the shade-grown species *Panax ginseng* to variable light conditions. Ann Bot. 2022;130(5):749–62.
29. Liu J, Nie ZL, Ren C, Su C, Wen J. Phylogenomics of *Aralia* sect. *Aralia* (Araliaceae): signals of hybridization and insights into its species delimitations and intercontinental biogeography. Mol Phylogenet Evol. 2023;181:107727.
30. Martín M, Sabater B. Plastid *ndh* genes in plant evolution. Plant Physiol Biochem. 2010;48(8):636–45.
31. Gao LZ, Liu YL, Zhang D, Li W, Gao J, Liu Y, Li K, Shi C, Zhao Y, Zhao YJ, et al. Evolution of *Oryza* chloroplast genomes promoted adaptation to diverse ecological habitats. Commun Biol. 2019;2:278.
32. Schuster M, Gao Y, Schöttler MA, Bock R, Zoschke R. Limited responsiveness of chloroplast gene expression during acclimation to high light in tobacco. Plant Physiol. 2020;182(1):424–35.

33. Liu H, Ye H, Zhang N, Ma J, Wang J, Hu G, Li M, Zhao P. Comparative analyses of chloroplast genomes provide comprehensive insights into the adaptive evolution of *Paphiopedilum* (Orchidaceae). Horticulturae. 2022;8(5):391.
34. Sharma SK, Pandit MK. A new species of *Panax* L.(Araliaceae) from Sikkim Himalaya, India. Syst Bot. 2009;34(2):434–8.
35. Kan S, Su X, Yang L, Zhou H, Qian M, Zhang W, Li C. From light into shadow: comparative plastomes in *Petrocosmea* and implications for low light adaptation. BMC Plant Biol. 2024;24(1):949.
36. Magdy M, Ou L, Yu H, Chen R, Zhou Y, Hassan H, Feng B, Taitano N, van der Knaap E, Zou X, et al. Pan-plastome approach empowers the assessment of genetic variation in cultivated *Capsicum* species. Hortic Res. 2019;6:108.
37. Akrami AM, Meratian Esfahani S, Soorni A. Decoding the chloroplast genomes of five Iranian Salvia species: insights into genomic structure, phylogenetic relationships, and molecular marker development. BMC Genomics. 2025;26(1):545.
38. Diani Gohar S, Soorni A. Plastome diversity and phylogenomic analysis of Satureja (Lamiaceae): uncovering evolutionary patterns and diagnostic markers. Planta. 2025;262(6):149.
39. Hejazi FA, Mohammadi P, Soorni A. Comparative chloroplast genomics of Teucrium species reveals genome evolution, phylogenetic relationships, and candidate molecular markers. Sci Rep. 2025;15(1):44318.
40. Chen Y, Ma T, Zhang T, Ma L. Trends in the evolution of intronless genes in Poaceae. Front Plant Sci. 2023;14:1065631.
41. Wen J. Species diversity, nomenclature, phylogeny, biogeography, and classification of the ginseng genus (*Panax* L., Araliaceae). In: *Proceedings of the International Ginseng Workshop*. Edited by Punja ZK. Vancouver: Simon Fraser University Press; 2021. p. 67–88.
42. Xia P, Huang Y, Zhu J. The complete chloroplast genome sequences of 11 *Panax* species: providing insights for evolution and species identification. Ind Crop Prod. 2025;223:120160.
43. Zhou M, Yang G, Sun G, Guo Z, Gong X, Pan Y. Resolving complicated relationships of the *Panax* bipinnatifidus complex in southwestern China by RAD-seq data. Mol Phylogenet Evol. 2020;149:106851.
44. Nguyen VB, Park HS, Lee SC, Lee J, Park JY, Yang TJ. Authentication markers for five major *Panax* species developed via comparative analysis of complete chloroplast genome sequences. J Agric Food Chem. 2017;65(30):6298–306.
45. Ngan F, Shaw P, But P, Wang J. Molecular authentication of *Panax* species. Phytochemistry. 1999;50(5):787–91.
46. Ji Y, Liu C, Yang Z, Yang L, He Z, Wang H, Yang J, Yi T. Testing and using complete plastomes and ribosomal DNA sequences as the next generation DNA barcodes in *Panax* (Araliaceae). Mol Ecol Resour. 2019;19(5):1333–45.
47. Zuo Y, Chen Z, Kondo K, Funamoto T, Wen J, Zhou S. DNA barcoding of *Panax* species. Planta Med. 2011;77(2):182–7.
48. Ma S, Wu Q, Hu Y, Wei F. Patterns and effects of GC3 heterogeneity and parsimony informative sites on the phylogenetic tree of genes. Gene. 2018;655:56–60.
49. Soorni A, Golchini MM. Complete Chloroplast genome of Mentha aquatica reveals hypervariable regions and resolves phylogenetic position within the genus Mentha. Mol Biol Rep. 2025;52(1):677.
50. Gago J, Carriquí M, Nadal M, Clemente-Moreno MJ, Coopman RE, Fernie AR, Flexas J. Photosynthesis optimized across land plant phylogeny. Trends Plant Sci. 2019;24(10):947–58.
51. Khan N, Choi SH, Lee CH, Qu M, Jeon JS. Photosynthesis: genetic strategies adopted to gain higher efficiency. Int J Mol Sci. 2024;25(16):8933.
52. Garcia A, Gaju O, Bowerman AF, Buck SA, Evans JR, Furbank RT, Gilliham M, Millar AH, Pogson BJ, Reynolds MP, et al. Enhancing crop yields through improvements in the efficiency of photosynthesis and respiration. New Phytol. 2023;237(1):60–77.
53. Li R, Wen J. Phylogeny and diversification of Chinese Araliaceae based on nuclear and plastid DNA sequence data. J Syst Evol. 2016;54(4):453–67.
54. Wen J, Plunkett GM, Mitchell AD, Wagstaff SJ. The evolution of Araliaceae: a phylogenetic analysis based on ITS sequences of nuclear ribosomal DNA. Syst Bot. 2001;26(1):144–67.
55. Robbins EHJ, Kelly S. The evolutionary constraints on angiosperm chloroplast adaptation. Genome Biol Evol. 2023;15(6):evad101.
56. Martin M, Noarbe DM, Serrot PH, Sabater B. The rise of the photosynthetic rate when light intensity increases is delayed in *ndh* gene-defective tobacco at high but not at low CO2 concentrations. Front Plant Sci. 2015;6:34.
57. Pottier M, Gilis D, Boutry M. The hidden face of Rubisco. Trends Plant Sci. 2018;23(5):382–92.
58. Newmaster SG, Fazekas AJ, Ragupathy S. DNA barcoding in land plants: evaluation of *rbcL* in a multigene tiered approach. Can J Bot. 2006;84(3):335–41.
59. Iida S, Miyagi A, Aoki S, Ito M, Kadono Y, Kosuge K. Molecular adaptation of *rbcL* in the heterophyllous aquatic plant *Potamogeton*. PLoS ONE. 2009;4(2):e4633.
60. Yao X, Tan YH, Yang JB, Wang Y, Corlett RT, Manen JF. Exceptionally high rates of positive selection on the *rbcL* gene in the genus *Ilex* (Aquifoliaceae). BMC Evol Biol. 2019;19(1):192.
61. Wu Z, Liao R, Yang T, Dong X, Lan D, Qin R, Liu H. Analysis of six chloroplast genomes provides insight into the evolution of *Chrysosplenium* (Saxifragaceae). BMC Genomics. 2020;21(1):621.
62. Chen Z, Yu X, Yang Y, Wei P, Zhang W, Li X, Liu C, Zhao S, Li X, Liu X. Comparative analysis of chloroplast genomes within *Saxifraga* (Saxifragaceae) takes insights into their genomic evolution and adaption to the high-elevation environment. Genes. 2022;13(9):1673.
63. Iverson ENK, Criswell A, Havird JC. Stronger evidence for relaxed selection than adaptive evolution in high-elevation animal mtDNA. Mol Biol Evol. 2025;42(4):msaf061.
64. Li R, Wen J. Phylogeny and biogeography of Asian *Schefflera* (Araliaceae) based on nuclear and plastid DNA sequence data. J Syst Evol. 2013;52(4):431–49.
65. Gallego-Narbon A, Johnson G, Fernandez-Mazuecos M, Wen J, Valcarcel V. Ancient polyploidization events influence the evolution of the ginseng family (Araliaceae). Front Plant Sci. 2025;16:1595321.
66. Yuan Y. Whole genome resequencing of *P. quinquefolius*. https://ngdc.cncb.ac.cn/gsa/browse/CRA011933. Accessed 2 Dec 2024.
67. FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 3 Dec 2024.
68. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
69. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21(1):241.
70. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of *de novo* genome assemblies. Bioinformatics. 2015;31(20):3350–2.
71. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods. 2019;15:50.
72. Huang L, Yu H, Wang Z, Xu W. CPStools: a package for analyzing chloroplast genome sequences. iMetaOmics. 2024;1(2):e25.
73. Nakamura T, Yamada KD, Tomii K, Katoh K. Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics. 2018;34(14):2490–2.
74. Wang XL. GeneBank Files for plastome sequences. https://doi.org/10.6084/m9.figshare.29588627.v1. Accessed 17 Jul 2025.
75. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17(4):540–52.
76. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35(21):4453–5.
77. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52(W1):W78–82.
78. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34(12):3299–302.
79. Gu Z, Gu L, Eils R, Schlesner M, Brors B. *circlize* implements and enhances circular visualization in R. Bioinformatics. 2014;30(19):2811–2.
80. Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, Batzoglou S. Glocal alignment: finding rearrangements during alignment. Bioinformatics. 2003;19(suppl1):i54–62.
81. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.
82. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29(22):4633–42.
83. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2019;20(1):348–55.
84. Kück P, Longo GC. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Front Zool. 2014;11:1–8.
85. Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. EasyCodeML: a visual tool for analysis of selection using CodeML. Ecol Evol. 2019;9(7):3891–8.

86. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.
87. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630(8016):493–500.
88. Fares MA, McNally D. CAPS: coevolution analysis using protein sequences. Bioinformatics. 2006;22(22):2821–2.
89. PyMOL. http://www.pymol.org. Accessed 12 Nov 2024.
90. Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. RELAX: detecting relaxed selection in a phylogenetic framework. Mol Biol Evol. 2015;32(3):820–32.
91. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–91.
92. Borowiec ML. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ. 2016;4:e1660.

Publisher's note