PLOS ONE

RESEARCH ARTICLE

The complete plastid genome of the endangered shrub *Brassaiopsis angustifolia* (Araliaceae): Comparative genetic and phylogenetic analysis

Zhanghong Dong[1☯], **Ruli Zhang**[2☯], **Ming Shi**[2], **Yu Song**[3], **Yaxuan Xin**[1], **Feng Li**[1], **Jianzhong Ma**[4]*, **Peiyao Xin**[1]*

1 Southwest Research Center for Landscape Architecture Engineering, National Forestry and Grassland Administration, Southwest Forestry University, Kunming, China, **2** Sympodial Bamboos Technological and Engineering Research Center, National Forestry and Grassland Administration, Southwest Forestry University, Kunming, China, **3** Key Laboratory of Ecology of Rare and Endangered Species and Environmental Protection (Ministry of Education), Guangxi Normal University, Guilin, China, **4** Yunnan Academy of Forestry and Grassland, Kunming, China

☯ These authors contributed equally to this work.
* majianzhong2000@sohu.com (JM); xpytgyx@163.com (PX)

Abstract

Brassaiopsis angustifolia K.M. Feng belongs to the family Araliaceae, and is an endangered shrub species in southwest China. Despite the importance of this species, the plastid genome has not been sequenced and analyzed. In this study, the complete plastid genome of *B. angustifolia* was sequenced, analyzed, and compared to the eight species in the Araliaceae family. Our study reveals that the complete plastid genome of *B. angustifolia* is 156,534 bp long, with an overall GC content of 37.9%. The chloroplast genome (cp) encodes 133 genes, including 88 protein-coding genes, 37 transfer RNA (tRNA) genes, and eight ribosomal RNA (rRNA) genes. All protein-coding genes consisted of 21,582 codons. Among the nine species of Araliaceae, simple sequence repeats (SSRs) and five large repeat sequences were identified with total numbers ranging from 37 to 46 and 66 to 78, respectively. Five highly divergent regions were successfully identified that could be used as potential genetic markers of *Brassaiopsis* and Asian Palmate group. Phylogenetic analysis of 47 plastomes, representing 19 genera of Araliaceae and two related families, was performed to reconstruct highly supported relationships for the Araliaceae, which highlight four well-supported clades of the *Hydrocotyle* group, Greater *Raukaua* group, *Aralia-Panax* group, and Asian Palmate group. The genus *Brassaiopsis* can be divided into four groups using internal transcribed spacer (ITS) data. The results indicate that plastome and ITS data can contribute to investigations of the taxonomy, and phylogeny of *B. angustifolia*. This study provides a theoretical basis for species identification and future biological research on resources of the genus *Brassaiopsis*.

Introduction

The genus *Brassaiopsis* Decne. & Planch. in the family Araliaceae includes nearly 45 species, with most of the species in the Himalayas, western China, Indochina, and the Malay Peninsula [1, 2], with southwestern China and northern Indochina as the main center of diversity [1]. The genus is part of the core Asian clade of Araliaceae [1, 3–5], and the species delimitation and infrageneric classification of *Brassaiopsis* has been highly controversial [1]. *Brassaiopsis angustifolia* K.M. Feng is an endangered shrub species. There is an urgent need to manage and conserve the natural resource of this shrub. However, information on its genetic and genomic background is limited. At present, the systematic position of *B. angustifolia* is not clear in the genus of *Brassaiopsis*. Therefore, knowledge of the genomics and phylogeny of its populations is essential to formulating effective protective measures.

During the past 20 years, molecular data have greatly improved our understanding of the phylogenetic relationships within the Araliaceae family and within the Apiales order [6]. Previous phylogenetic studies based on nuclear ribosomal DNA [1–5, 7–12] and chloroplast (cp) DNA [5, 7–13] sequence data have provided important clues about the evolution and diversity of Araliaceae plants. However, traditional phylogenies based on analysis of multiple genes have failed to well solve the relationship among species of Araliaceae. In addition, the early divergences of the Asian Palmate group have been clarified by using the chloroplast genome, but the backbone of its core is not totally resolved [14]. The family of Araliaceae, divided into four main clades by molecular systematics: Greater *Rauakaua*, *Polyscias–Pseudopanax*, *Aralia–Panax*, and Asian Palmate, respectively [3, 5, 10, 14]. Although several recent studies have made progress, sampling from Araliaceae has remained limited largely in most studies because of a focus on problems at other phylogenetic levels, individual geographic regions, or questions dedicated to a single genus [1, 2, 7, 8, 12]. The genus *Brassaiopsis* is part of the core Asian clade of Araliaceae. Based on the comprehensive analysis of internal transcribed spacer (ITS) and cp DNA data, the sister group relationship of *Brassaiopsis* and *Trevesia* were supported [1–3, 5, 7–11]. Because of low sequence quality, the few samples, and rapid divergence in early evolutionary history, the phylogenetic relationship among genera and species has not yet been well solved of Araliaceae [2–5, 9, 11, 15]. In addition, the species delimitation and infrageneric classification of *Brassaiopsis* are highly controversial. Therefore, it is necessary to use the existing cp genome data to construct a robust phylogenetic tree to clarify the phylogenetic relationships within Araliaceae.

In plants, the cp is the main locus of photosynthesis and carbon fixation [16, 17]. The cp genome of higher plants is a double-stranded circular DNA molecule ranging in size from 72 to 217 kb, containing about 130 genes. The cp genome has a typical tetrad structure, including a large single copy (LSC) region, a small single copy (SSC) region, and a pair of inverted repeats (IRs) in most plants [17–19]. Compared with nuclear genomes, the uniqueness of the cp is evident in its maternal inheritance, small size, simple structure, and conserved sequences [20, 21]. The cp genome sequence reveals the phylogenetic relatedness visually at different taxonomic levels and provides an understanding of the evolution of a plant's structure and function [22, 23]. Therefore, the genome sequence is widely used for cp inheritance, domestication studies, phylogeny, and adaptive evolution.

Here, we report for the first time, to our knowledge, the complete chloroplast genome sequence of *B. angustifolia* and characterize the structure, gene content, and organization of its genome. Then, we establish its codon usage frequencies, simple sequence repeats (SSRs), repeats, regions of high sequence divergence, nucleotide variability values, and the expansion and contraction of its IRs. Finally, we evaluated the phylogenetic position by comparative analysis based on 42 entire plastid genomes sequences of Araliaceae species and conducted a

phylogenetic analysis of the *Brassaiopsis* genus using nuclear ribosomal ITS data. The results of this study can provide clues for species classification of *Brassaiopsis* and help to clarify the evolution and phylogenetic relationships of the species and genera of the Araliaceae family. Furthermore, this newly developed genomic resource will help further conserve the genetics of this endangered species.

Materials and methods

Sampling, DNA extraction and sequencing

We collected fresh and young leaves of *B. angustifolia* from a cultivated tree at the Southwest Forestry University, Kunming, China (102˚45.489′ E, 25˚3.639′ N). Total genomic DNA was extracted using the modified cetyltrimethylammonium bromide (CTAB) method [24]. Long-range polymerase chain reaction was performed following Zhang et al. [25], with 15 pairs of universal primers. The entire cp genome of *B. angustifolia* was sequenced via 250 bp paired-end sequencing on a HiSeq 2500 Platform (Illumina, Nanjing, China).

Analysis of cp genome assembly, annotation, and relative synonymous codon usage

Raw reads were filtered to remove low-quality reads, and de novo assembly of circular plastome and ITS sequence were carried out using GetOrganelle software [26]. We used Bandage software [27] to examine and screen the assembled cp genome of *B. angustifolia*. The cp genome was adjusted and annotated with Geneious software [28], and use the Organelle Genome DRAW software [29] to drawn a circular structure diagram of the entire genome. The entire annotated chloroplast genome and the ITS sequences were submitted to the National Center for Biotechnology Information (cp genome GenBank accession: OK638200; ITS GenBank accession: OL352055). Relative synonymous codon usage (RSCU) and codon usage were examined with CodonW (version 1.4.4) [30].

SSRs and identification of repeats

Using the MicroSatellite (MISA) identification tool [31], the minimum repeat number of mononucleotides was 10; for dinucleotides, it was five, and for trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide repeat motifs, it was four. The SSRs of 1–10 units of the nine complete cp genomes of the Asian Palmate group were detected. Complement repeats, forward repeats, palindromic repeats, and reverse repeats in non-SSR of the nine cp genomes were detected with the online software REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer) [32], using the parameter default values. Tandem repeats in non-SSR of nine cp genomes were detected with online software TRF (http://tandem.bu.edu/trf/trf.basic.submit.html) [33].

Comparative genome analysis and sequence divergence

MAFFT software [34] was used to align the nine cp genomes of the Asian Palmate group, which were then manually adjusted with BioEdit software [35]. Sequence identity analysis was visualized with the mVISTA program in Shuffle-LAGAN mode [36], with the annotation for *B. angustifolia* as a reference. We performed a sliding-window analysis in the DnaSP software (version 6) [37] to evaluate the variability(π) of the plastomes. The window length was set to 600 bp and the step size to 200 bp. The contraction and expansion of the IR boundaries of nine species in the Asian Palmate group were visualized with the IRscope software [38].

Phylogenetic analysis based on cp genome and ITS data

Two different data matrices were assembled and analyzed using both maximum likelihood and Bayesian inference (BI) methods. Matrix I contained 42 taxa with complete plastid genomes available (S1 Table), including members of *Aralia* L. (5 spp.), *Brassaiopsis* Decne. & Planch. (2 spp.), *Cheirodendron* Nutt. ex Seem. (1 sp.), *Chengiopanax* C. B. Shang & J. Y. Huang (1 sp.), *Dendropanax* Decne. & Planch. (4 spp.), *Eleutherococcus* Maxim. (5 spp.), *Fatsia* Decne. & Planch. (2 spp.), *Hedera* L. (2 spp.), *Heptapleurum* Gaertn. (3 spp.), *Heteropanax* Seem. (1 sp.), *Hydrocotyle* L. (1 sp.), *Kalopanax* Miq. (1 sp.), *Macropanax* Miq. (1 sp.), *Merrilliopanax* H. L. Li (1 sp.), *Oplopanax* Miq. (1 sp.), *Panax* L. (6 spp.), *Raukaua* Seem. (3 spp.), *Schefflera* J. R. Forst. & G. Forst. (1 sp.), and *Tetrapanax* (K. Koch) K. Koch (1 sp.) (S1 Table). *Angelica keiskei* (Miq.) Koidz. (GenBank accession: MW125613), *Diplopanax stachyanthus* Hand.-Mazz. (GenBank accession: MG524991), *Diplopanax stachyanthus* Hand.-Mazz. (GenBank accession: KP318983), and *Ostericum grosseserratum* (Maxim.) Kitag. (GenBank accession: KT852844), were sampled as outgroups. Matrix II included 24 species of *Brassaiopsis* and eight species of *Trevesia*, and included 34 ITS sequences (S2 Table).

The complete cp genome and ITS matrix were aligned with MAFFT software (version 7) [34] and then manually edited with BioEdit 7.2.5 [35]. Bayesian inference was undertaken using MrBayes (version 3.2.6) [39]. Used jModelTest (version 2.1.10) [40] to select the most suitable replacement DNA model for phylogenetic reconstruction. The Markov chain Monte Carlo (MCMC) algorithm was run for 10,000,000 generations. The first 25% of the trees was discarded as aging trees and the remaining trees were used to generate a consensus tree for majority-rule. When the average standard deviation of split frequencies was less than 0.01, we considered the operation to be stable. Nodes with posterior probability (PP) values of 0.95 or greater were considered statistically significant. ML analysis was performed IQ-TREE (version 1.6.7) [41]. ML bootstrap support (BS) values of 70% or greater were considered well supported, and ML BS of less than 50% were considered poorly supported or unresolved. The best-fit DNA substitution models for matrix I and matrix II, respectively, were chosen as "TPM1uf + I + G" (freqA = 0.3138, freqC = 0.1857, freqG = 0.1807, freqT = 0.3197, R (a) [AC] = 1.0000, R(b) [AG] = 1.6623, R(c) [AT] = 0.4054, R (d) [CG] = 0.4054, R(e) [CT] = 1.6623, R (f) [GT] = 1.0000, p-inv = 0.5260, and gamma shape = 0.9030) and "TrNef + G" (R(a) [AC] = 1.0000, R(b) [AG] = 4.0619, R (c) [AT] = 1.0000, R(d) [CG] = 1.0000, R(e) [CT] = 7.9724, R(f) [GT] = 1.0000, and gamma shape = 0.1830) to construct the phylogenetic tree. Results of all phylogenetic analysis are displayed with FigTree (version 1.4.3) (http://tree.bio.ed.ac.uk/publications/).

Results

Chloroplast genome characteristics of *B. angustifolia*

The size of the *B. angustifolia* cp genome was 156,534 bp, and exhibited the usual quadripartite structure, featuring a LSC region (86,509 bp), a SSC region (18,061 bp), and a pair of IRs (25,982 bp) (Fig 1). The guanine-cytosine (GC) content was 37.9% in whole cp, 36.1% and 31.9% in the LSC and SSC regions, respectively, whereas higher rates (43.0%) were distributed in the IR regions. A total of 133 genes were annotated in the sequenced *B. angustifolia* cp genome, containing 88 protein-coding genes, 37 transfer RNAs (tRNAs), and eight ribosomal RNAs (rRNAs) (S3 Table). These genes belong to several categories with different functions; among which 17 duplicated genes were located in the IR regions, including six protein-coding genes (*ndhB*, *rpl2*, *rpl23*, *rps7*, *ycf2*, and *ycf15*), seven tRNAs (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*), and four rRNAs (*rrn4.5S*, *rrn5S*, *rrn16S*,
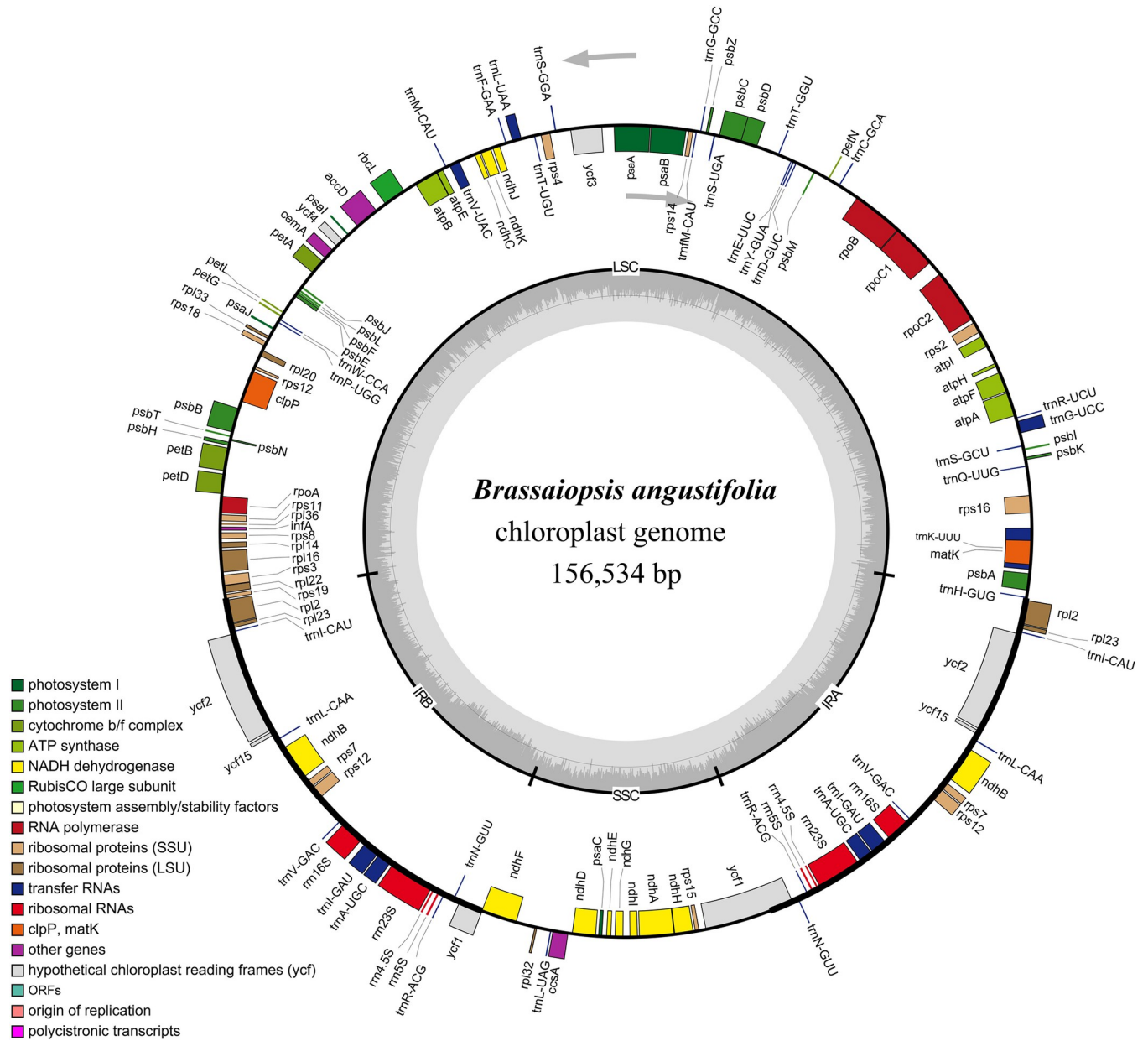
Fig 1. Circular map of chloroplast genome of *Brassaiopsis angustifolia* with annotated genes. Genes shown inside and outside of the circle are transcribed in clockwise and counterclockwise directions, respectively. Genes belonging to different functional groups are color-coded. The GC and AT content are denoted by the dark gray and light gray colors in the inner circle, respectively.

and *rrn23S*). The transcription regulation of genes was believed to be affected by introns and exons. Nineteen genes (11 protein-coding genes and eight tRNAs) contained at least one intron, and three genes (*clpP*, *ycf3, and rps12*) had two introns (S3 Table).

According to the RSCU analysis (Fig 2), all protein-coding genes consisted of 21,582 codons. Among them, leucine (2,278 codons, 10.56%) was the most abundant amino acid, and isoleucine (1,824 codons, 8.45%) was the second. Each amino acid corresponded to at least one codon (tryptophan) and up to six (arginine, leucine, methionine, and serine). Trp had only one codon (UGG), which meant there was no codon usage bias. In addition, with the
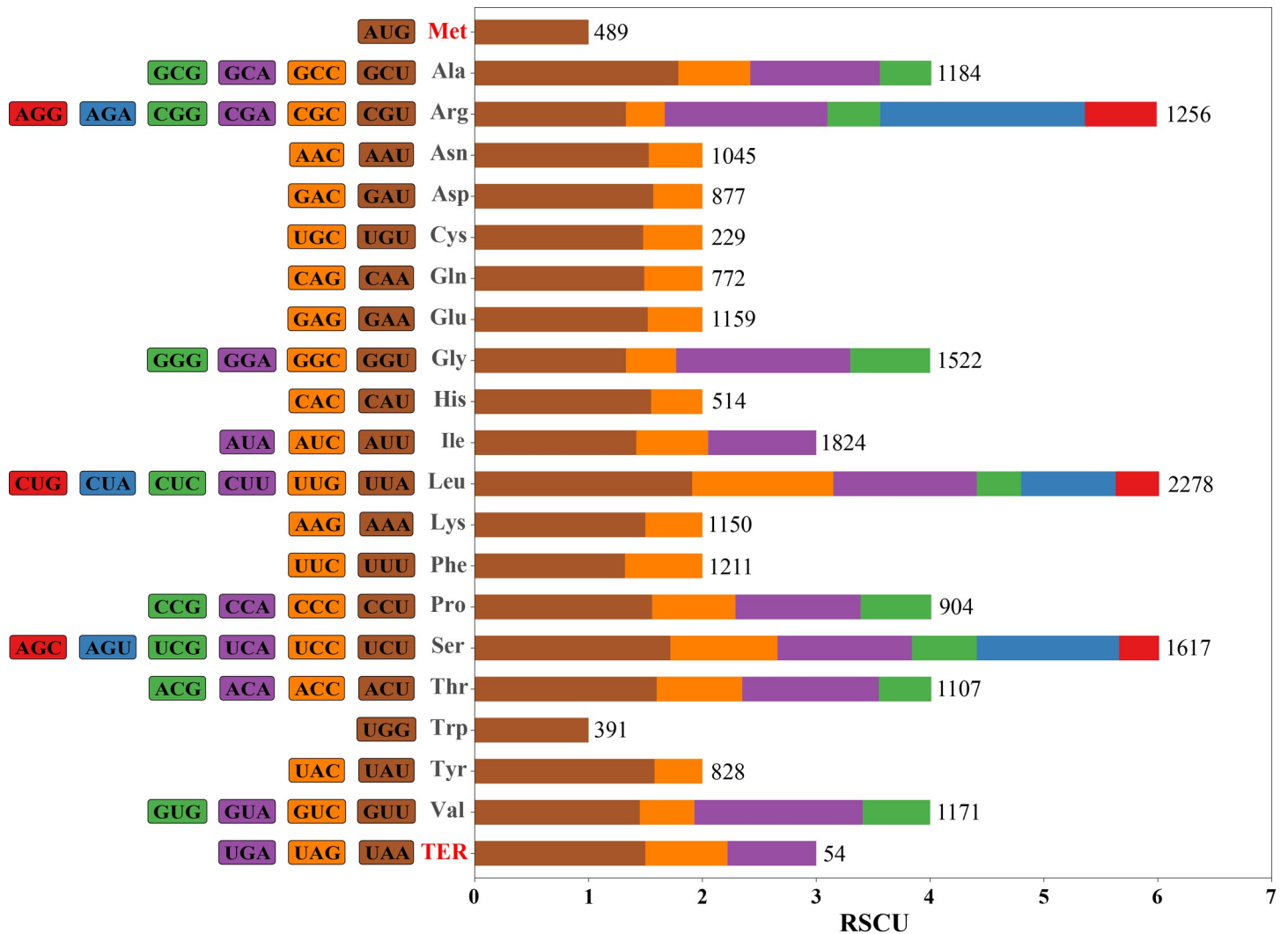
Fig 2. Relative synonymous codon usage (RSCU) analysis of 20 amino acids and three termination codons in *Brassaiopsis angustifolia*. Bar diagram in different colors reflects codon usage bias (brown, yellow, purple, green, blue, and red, correspond to the proportion of different codons in descending order). Numbers represent codon quantities.

exception of methionine, most amino acids tended to use codons that ended in A/U rather than in C/G.

Analysis of SSR and repeats for the nine cp genomes of the Asian Palmate group

MISA analysis of 9 species cp genomic sequences from the Asian Palmate group revealed numerous SSR loci. In total, six types of SSRs (mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide repeats) were detected based on the comparison of nine genomes in the Asian Palmate group. A total of 41 perfect SSRs were found in *B. angustifolia* (Fig 3A). Similarly, 38, 42, 37, 45, 42, 40, 46, and 35 SSRs were detected in *Brassaiopsis hainla* (Buch.-Ham.) Seem., *Eleutherococcus brachypus* Nakai, *Eleutherococcus trifoliatus* (L.) S. Y. Hu, *Eleutherococcus gracilistylus* (W. W. Sm.) S. Y. Hu, *Eleutherococcus senticosus* Maxim., *Eleutherococcus sessiliflorus* (Rupr. & Maxim.) S. Y. Hu, *Kalopanax*
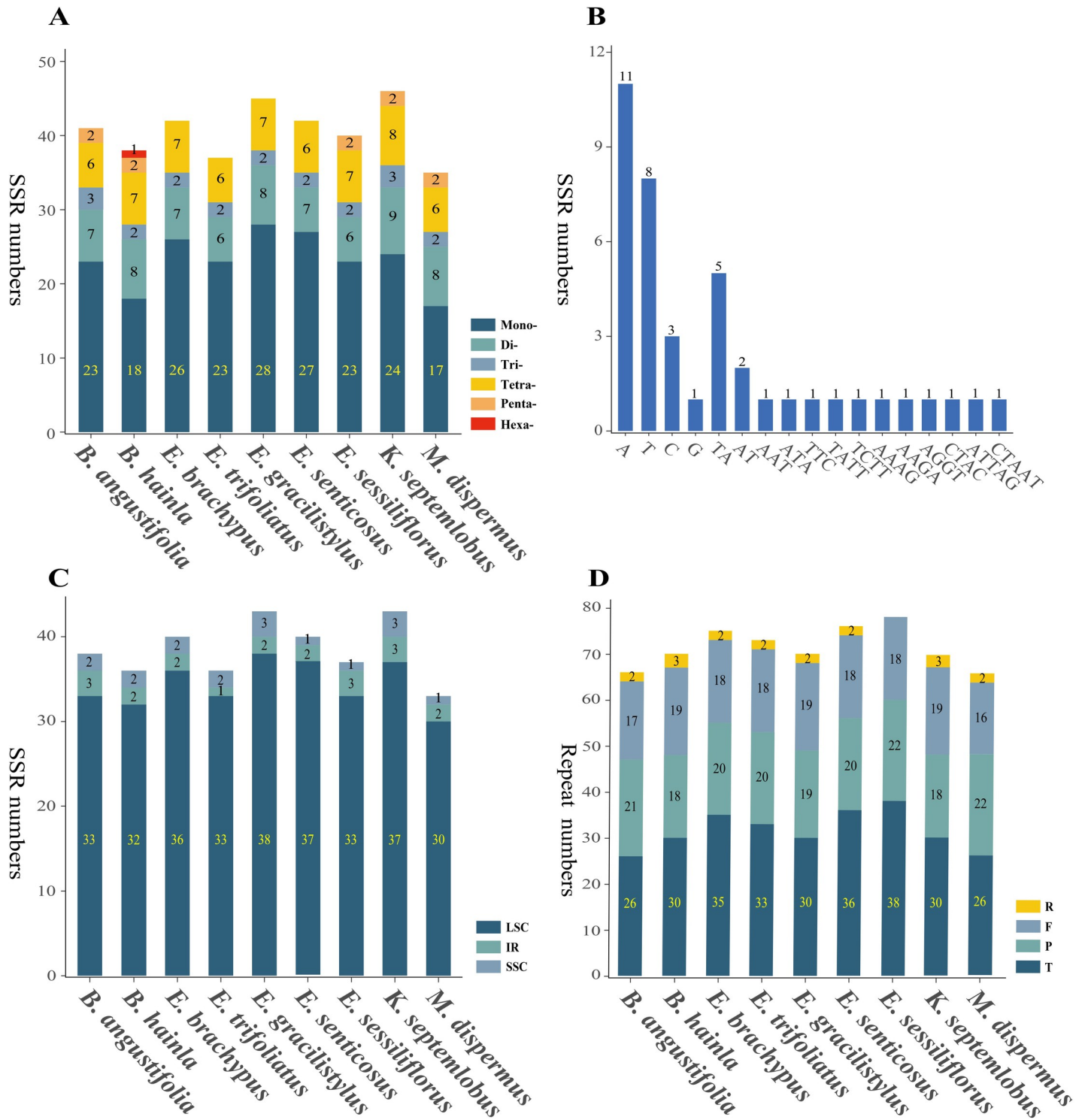
Fig 3. The simple sequence repeat (SSR) elements in the chloroplast genome of nine species in the Asian Palmate group. (A) Number of six SSR types. (B) Number of identified SSR motifs of *Brassaiopsis angustifolia*. (C) Frequency of identified SSRs in large single copy (LSC) regions, small single copy (SSC) regions, and inverted repeat (IR) regions. (D) Total of four repeat types.

https://doi.org/10.1371/journal.pone.0269819.g003

septemlobus Koidz., and *Macropanax dispermus* Kuntze. The most abundant type of SSR was mononucleotide repeats, which ranged from 17 bp in *M. dispermus* to 28 bp in *E. gracilistylus*, followed by dinucleotide, tetranucleotide, trinucleotide, pentanucleotide, and hexanucleotide repeats (Fig 3A). In the cp genome of *B. angustifolia*, most mononucleotide SSRs had A (47.83%) and T (34.78%) motifs, whereas all dinucleotide repeats were composed of TA (74.43%) and AT (28.57%) motifs (Fig 3B). Further analysis shows that most of the microsatellites were in the LSC region and a small proportion in the SSC and IR regions (Fig 3C). A total of 66 repeats were found in the *B. angustifolia* chloroplast genome, including tandem, palindromic, forward, inverted, and complement repeats. Similarly, 70, 75, 73, 70, 76, 78, 70, and 66 repeats were detected in *B. hainla*, *E. brachypus*, *E. trifoliatus*, *E. gracilistylus*, *E. senticosus*, *E. sessiliflorus*, *K. septemlobus*, and *M. dispermus*, respectively (Fig 3D).

Identification of mutational hotspots among nine species of the Asian Palmate group

Sequence identity analysis was performed using mVISTA (Fig 4). Multiple sequence alignment revealed high similarity among the nine chloroplast genomes of the Asian Palmate group, which suggests that they are highly conserved; specifically, the results showed that the divergence of the single-copy (SC) region was larger than that of the IR region, and the divergence of the non-coding region was larger than that of the coding region.

The software DnaSP 6.0 was used to calculate the nucleotide variation value (π) within 600 bp of the cp genome of *B. angustifolia*, *B. hainla*, *E. brachypus*, *E. gracilistylus*, *E. senticosus*, *E. sessiliflorus*, *E. trifoliatus*, *K. septemlobus*, and *M. dispermus*. The difference between the two *Brassaiopsis* species the value varied from 0 to 0.02333, with an average of 0.00395, suggesting that their genomic differences were small. However, five highly variable loci with much higher π values ($\pi > 0.015$), including the *psbI*, *petN-psbM-trnD-GUC*, *trnT-GUU-psbD*, *ndhF-rpl32*, and *ycf1*, which were precisely located (Fig 5A). Among seven species of the Asian Palmate group and the two *Brassaiopsis* species, the π values varied from 0 to 0.02657 with a mean of 0.00461, indicating that the differences among species of the Asian Palmate group were larger than those between congeneric species. Five highly variable loci including the *trnK-UUU-rps16*, *trnE-UUC-trnT-GGU*, *psbE-petL*, *ndhF*, and *ycf1* were precisely located in the nine species of the Asian Palmate group ($\pi > 0.015$; Fig 5B).

Expansion and contraction of the inverted repeat regions

The IR boundaries of *B. angustifolia* and eight species of the Asian Palmate group were compared, and the possible expansion or contraction of IR regions was analyzed (Fig 6). The expansion and contraction of the boundaries of chloroplast genomic IR regions in nine plants of the Asian Palmate group were revealed, and the four junctions of two IRs between *B. angustifolia* and its related species were compared in detail (Fig 6). By comparing the plastids of the Asian Palmate group species, we found that the IR/LSC connections of IRb were mainly located between *rpl2* and *rps19* genes (Fig 6). In addition, the overlap of *ycf1* genes appeared in different positions among the Asian Palmate group species, that is, the SSC region of *B. angustifolia* and the IRB/SSC junction of other seven species. *M. dispermus* did not contain the *ycf1* gene. The *ycf1* genes sited at the SSC/IRa boundary and the length of *ycf1* ranged from 5,520 to 5,649 bp. The *trnH* genes of the four species in the Asian Palmate group were located in the LSC region, 2–5 bp away from the IRa-LSC border. In *B. angustifolia*, the expansion of the LSC region leads to the *ndhF* gene being at the IRb/LSC junction.
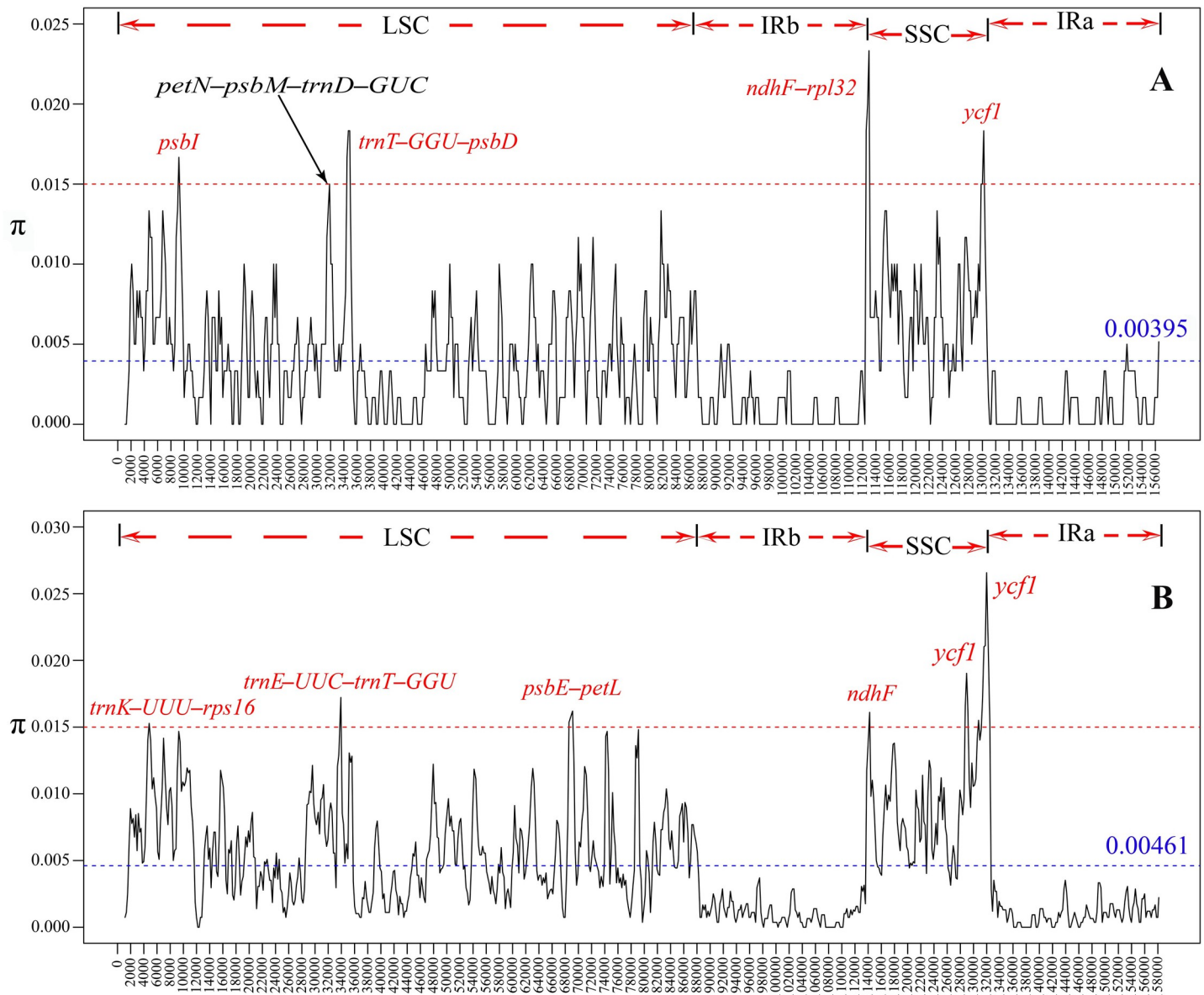
Fig 4. Complete chloroplast genome alignments for nine species of the Asian Palmate group using the mVISTA program, with the *Brassaiopsis angustifolia* chloroplast genome as a reference sequence. The y-scale indicates identity from 50% to 100%. Gray arrows indicate the position and direction of each gene. Red indicates conserved non-coding sequences (CNSs). Blue indicates the exons of protein-coding genes (exon).

Phylogenetic relationships among the Araliaceae genera

In order to better understand the phylogenetic relationship among the 42 sequenced taxa from 19 genera of the Araliaceae, we have downloaded the corresponding sequences from GenBank (S1 Table). The BI and ML analysis of the complete plastid sequences of the main clades and most genera indicate that most phylogenetic relationships have a high degree of internal support (Fig 7). Araliaceae was divided into four groups: the *Hydrocotyle* group (BI-PP = 1.00, ML-BS = 100%), which included the *Hydrocotyle*. The Greater *Raukaua* Group (BI-PP = 0.9, ML-BS = 100%) included *Schefflera*, *Cheirodendron*, and *Raukaua*. The *Aralia-Panax* Group (BI-PP = 0.8, ML-BS = 98%) included *Aralia* and *Panax*. The Asian Palmate Group (BI-PP = 0.7, ML-BS = 100%) can be divided into four small clades (A, B, C, and D). Clade A (BI-

Fig 5. Comparison of the nucleotide variability (π) values of the two *Brassaiopsis* plastomes (A) and nine plastomes of the Asian Palmate group (B). (Window length, 600 bp, step size, 200 bp); x-axis: position of the midpoint of a window; y-axis: nucleotide diversity of each window.

PP = 0.7, ML-BS = 100%) included *Oplopanax*. Clade B (BI-PP = 0.9, ML-BS = 100%) included *Heptapleurum*, *Heteropanax*, and *Tetrapanax*. Clade C (BI-PP = 1.00, ML-BS = 99%) included *Chengiopanax* plus *Dendropanax*. Clade D (BI-PP = 0.6, ML-BS = 92%) included *Fatsia*, *Hedera*, *Merrilliopanax*, *Brassaiopsis*, *Eleutherococcus*, *Kalopanax*, and *Macropanax*.

Phylogenetic relationships among *Brassaiopsis* species

To better understand the phylogenetic relationships among *Brassaiopsis* species the ITS sequences of 24 *Brassaiopsis* species were used in the evaluation, with eight *Trevesia* species as the outgroup. The phylogenetic tree divided the genus into four main groups (Fig 8). Group I (BI-PP = 1.00, ML-BS = 100%) included *Brassaiopsis griffithii* C. B. Clarke and *Brassaiopsis simplicifolia* C. B. Clarke. Group II (BI-PP = 1.00, ML-BS = 94%) included *Brassaiopsis elegans*
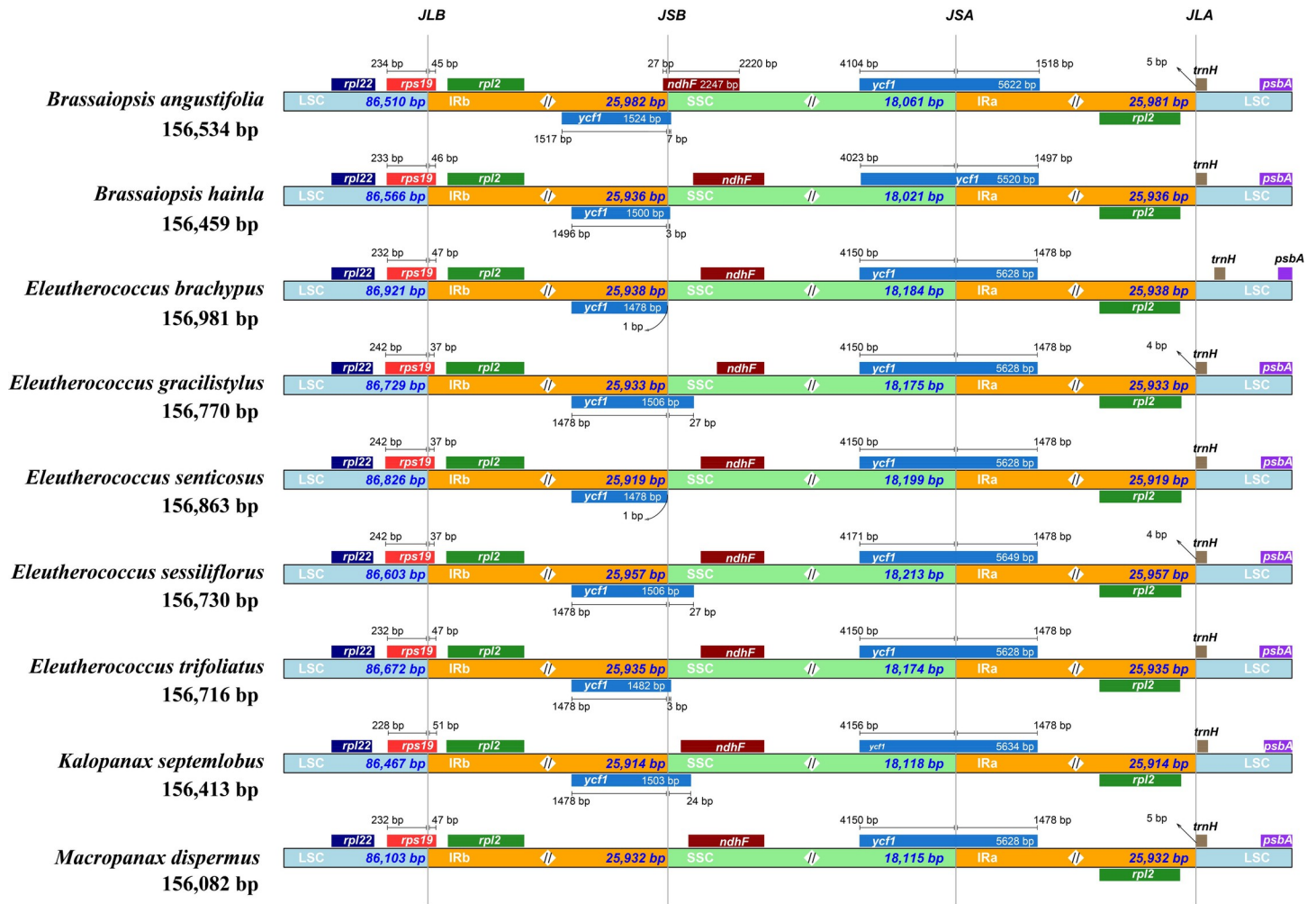
Fig 6. Comparison of the borders of the inverted repeat (IR), small single copy (SSC), and large single copy (LSC) regions among nine chloroplast genomes of the Asian Palmate group. JLB, JSB, JSA, and JLA represent the junctions of LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC, respectively. The Fig is not drawn to scale based on sequence length but shows only the relative change near or at the IR/SC junctions.

Ridl., *Brassaiopsis malayana* J. Wen & Frodin, ined., *Brassaiopsis simplex* (King) B. C. Stone, and *Brassaiopsis sumatrana* Ridl.. Group III (BI-PP = 1.00, ML-BS = 95%) included *Brassaiopsis ficifolia* Dunn, *Brassaiopsis grushvitzkyi* J. Wen, Lowry & T. H. Nguyên, *Brassaiopsis moumingensis* (Y. R. Ling) C. B. Shang, *Brassaiopsis phanrangensis* C. B. Shang, *Brassaiopsis producta* (Dunn) C. B. Shang, and *Brassaiopsis stellata* K. M. Feng. Group IV (BI-PP = 0.66, ML-BS = 83%) included *Brassaiopsis aculeata* Seem., *B. angustifolia*, *Brassaiopsis ciliata* Dunn, *Brassaiopsis ferruginea* (H. L. Li) G. Hoo, *Brassaiopsis glomerulata* (Blume) Regel, *Brassaiopsis gracilis* Hand.-Mazz., *B. hainla*, *Brassaiopsis hispida* Seem., *Brassaiopsis mitis* C. B. Clarke, *Brassaiopsis shweliensis* W. W. Sm., and *Brassaiopsis tripteris* (H. Lév.) Rehder.

Discussion and conclusions

Comparison of cp genomes in the Asian Palmate group species

This study revealed the entire cp genome of an endangered shrub, namely *B. angustifolia* in the family Araliaceae. The plastome with a length of 156,534 bp was larger than the published
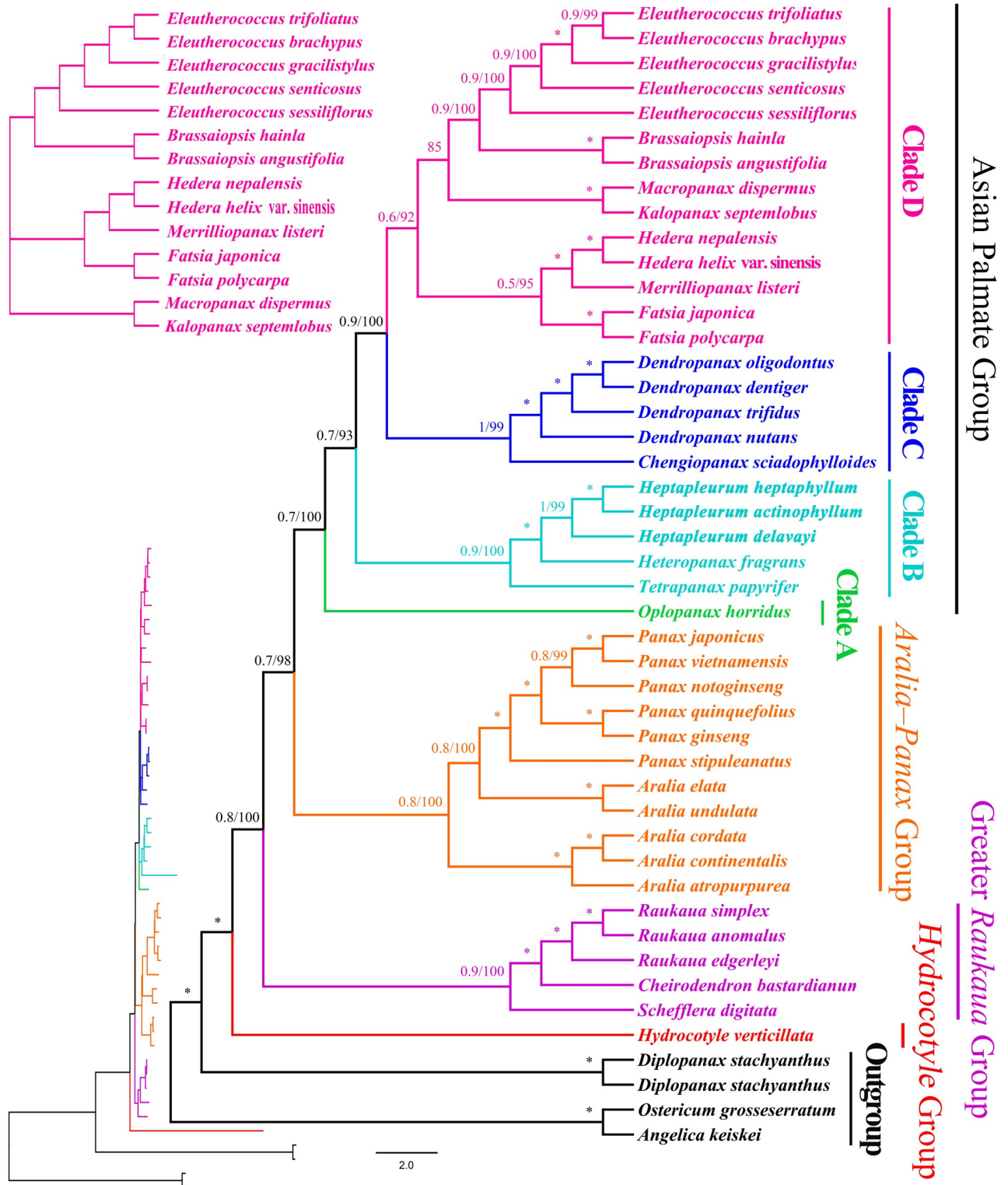
Fig 7. Molecular phylogenetic tree of 42 taxa of Araliaceae based on plastome sequences using unpartitioned Bayesian inference (BI) and maximum likelihood (ML). Numbers at each node are bootstrap support values. The tree is rooted in the plastome sequences of *Angelica keiskei*, *Ostericum grosseserratum*, and *Diplopanax stachyanthus*. Numbers associated with the branches are BI posterior probabilities (PP) and ML bootstrap value (BS), and asterisks (*) indicate BS/PP of 100/1.00.

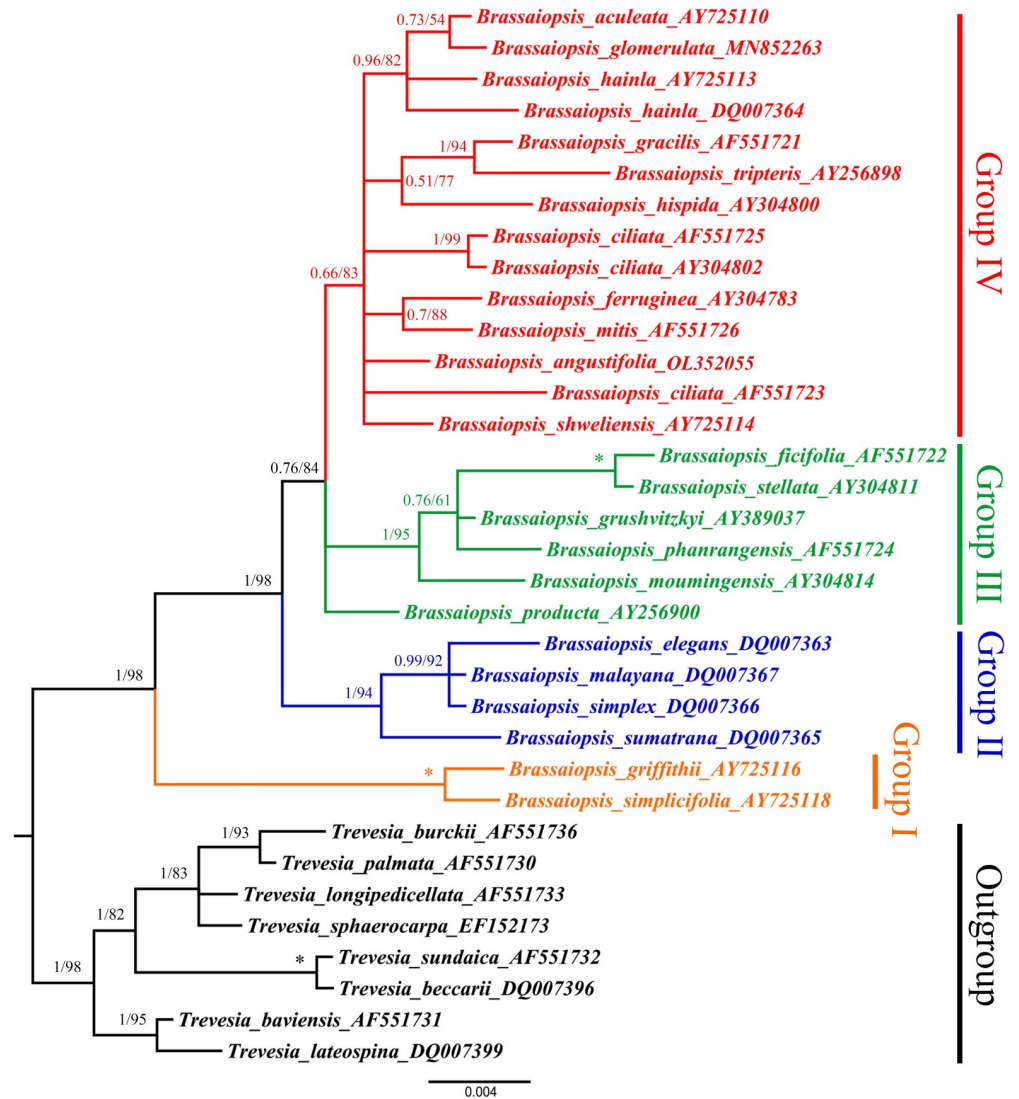https://doi.org/10.1371/journal.pone.0269819.g007

Fig 8. Bayesian inference (BI) and maximum likelihood (ML) strict tree illustrating the phylogeny of the genus *Brassaiopsis* based on nrDNA ITS datasets. Numbers associated with the branches are BI posterior probabilities (PP) and ML bootstrap values (BS), and asterisks (*) indicate BS /PP of 100/1.00.

https://doi.org/10.1371/journal.pone.0269819.g008

plastome of *B. hainla* [13], which both possess the typical angiosperm quadripartite structure (Fig 1). The cp genomes are highly conserved for mostland plants. The difference in the size of the IRs and intergenic spacers leads to differences between the cp genomes [42]. Reportedly, the *ycf1* and *ycf2* genes are located on the border between the IR region and the LSC and SSC regions, and there is incomplete replication of these two genes [13, 43]. Unlike *ycf2*, the lengths of the truncated *ycf1* genes were different among *B. angustifolia*, *B. hainla*, *E. brachypus*, *E. gracilistylus*, *E. senticosus*, *E. sessiliflorus*, *E. trifoliatus*, *K. septemlobus*, and *M. dispermus* (Fig 5). The gene *ycf1* passes through the SSC-IRb region, and the truncated *ycf1* gene was located in the LSC-IRa region. The change in length of the truncate *ycf1* gene directly caused the shrinkage of the IR region in the plastomes of *E. gracilistylus* and *K. septemlobus*. In addition, in the *M. dispermus*, there is a truncated *ycf1* gene in the LSC-IRa region; the same phenomenon exists in some species of other angiosperms [44]. Except for *B. angustifolia*, the *ndhF* gene is

completely located in the SSC region in the other eight species, but the distance from the IRB/SSC boundary is different, which is consistent with the results obtained by Li et al. [13] in a study of seven species of Araliaceae.

Analysis of mutational hotspots

Not all genetic mutation events are random, with some clustering as hotspots [45, 46]. These mutation dynamics created highly variable regions in the genome [43]. In *B. angustifolia* and *B. hainla* plastomes, we identified five highly variable loci, which included *psbI*, *petN-psbM-trnD-GUC*, *trnT-GUU-psbD*, *ndhF-rpl32*, and *ycf1* (Fig 5A). In the Asian Palmate group, we identified five highly variable loci, which included *trnK-UUU-rps16*, *trnE-UUC-trnT-GGU*, *psbE-petL*, *ndhF*, and *ycf1* (Fig 5B). Compared with other highly variable regions, *ycf1* had the greatest genetic divergence among the nine sequenced plastid genomes of the Asian Palmate group (Fig 5B); from which, three highly variable loci were identified. These highly variable loci can be used for phylogenetic studies of the Araliaceae DNA barcode and at the species level. These results are partially congruent with those of Dong et al. [47] and Song et al. [48]. Therefore, Therefore, these associated regions can serve as barcodes as potential markers to reconstruct the phylogenetic relationship of Araliaceae.

The values of the nucleotide variability from the complete cp genomes among species of *B. angustifolia* and *B. hainla* were only 0.39%, and the nine species of the Asian Palmate group were 0.46%, which is similar to the nucleotide variability of two *Panax* L. species (0.40%) [49]; six *Persea* Mill., *Machilus* Nees, *Phoebe* Nees, and *Cinnamomum* Schaeff. species (0.32%) [50]; nine *Lindera* Thunb. species (0.48%) [51]; and 40 *Populus* L. species (0.36%) [52], which were much greater than the values of the two *Phoebe* Nees species (0.1%) [53], and three *Alseodaphne* Nees species (0.12%) [48], and were much less than the sequence divergence among the six *Cymbidium* Sw. species (3.70%) [54] and the five *Epimedium* L. species (3.97%) [55].

Phylogenetic analysis of Araliaceae and *Brassaiopsis*

Previous molecular markers provided limited information to elucidate the relationship among Araliaceae plants. Many studies have attempted to resolve the relationships within Araliaceae using molecular markers of ITS or plastid-region data, but the relationship of Araliaceae was still unclear [1–3, 5, 7–9, 11]. In addition, studies show that increased sampling of taxa can greatly improve overall phylogenetic accuracy [56–58]. Our study and Valcárcel et al. [14] had found that using the whole plastomes, with an appropriate sampling, can greatly aid the relationship between the deep pedigrees of the Araliaceae. The family of Araliaceae, grouped in four main clades in molecular phylogenies (Greater *Rauakaua*, *Polyscias–Pseudopanax*, *Aralia–Panax*, and Asian Palmate) [3, 5, 10, 14]. From the results of our study, four branches of phylogenetic meaningful have been identified in deep lineages of the Araliaceae. The topological backbone of the phylogenetic genome obtained in this work agrees with those previously published [5, 10, 14]. Moreover, the problems of several major branches of the Araliaceae were solved. In the *Hydrocotyle* group, the genus of *Hydrocotyle* is located within Araliaceae and is a relatively primitive genus (Fig 7). In the Greater *Rauakaua* Group, our phylogenetic analysis revealed a sister group containing three *Rauakaua* species, one *Cheirodendron* species, and the *Schefflera digitata* species (Fig 7), with strong support, as in the previously published phylogenetic tree constructed with combinations of plastid markers of *trnL-trnF* [5], *trnD-trnT* plus *rpl16* [6], nuclear (ITS + external transcribed spacers [ETS]) plus plastid markers (*ndhF-rpl32*, *rpl32-trnL*, *trnK-rps16*, *trnH-psbA*) [7], and ITS plus plastid markers (*ndhF*, *trnL–F*, *rps16*, *atpB–rbcL*, *rpl16* and *psbA–trnH*) [10]. In the *Aralia-Panax* Group, our phylogenetic analysis revealed that *Aralia* is nested among the members of *Panax* (Fig 7), which is consistent with

the cp data from Nicolas et al. [6], Valcárcel et al. [9], Mitchell et al. [10], Plunkett et al. [11] and Li et al. [13], the nuclear ribosomal DNA data from Wen et al. [3], and the nuclear plus plastid data from Li et al. [8]. In the Asian Palmate Group, *Oplopanax* that appears as sister to the remaining Asian Palmate group, as found in previous studies [7, 10, 14]. *Schefflera* has been shown to be polyphyletic and distributed in the main pedigree of the family [7–11]. With the deepening of the research, there has been a large taxonomic rearrangement of *Schefflera* that resulted in the description of five new or reinstated genera [7, 59–64]. The resurrection of *Heptapleurum* in an Asian clade formerly belonging to *Schefflera* (Araliaceae), and with the completion of these transfers, *Heptapleurum* is now the largest genus in Araliaceae [63, 64]. In the Clade B, our phylogenomic analysis shows that sisterhood contained three *Heptapleurum* species and *H. fragrans*, *Tetrapanax papyrifer* (Hook.) K. Koch and *Oplopanax horridus* (Smith.) Miq. (Fig 7), likewise significant support in the nuclear ribosomal ITS sequences [2, 11], ITS plus plastid markers [9, 10], and the cp genomes data [14]. Furthermore, these clade can be divided into four groups, with strong support, the backbones of the phylogenomic topologies obtained here are consistent with previously published phylogenetic relationships [6, 10, 11, 14].

The phylogeny of *Brassaiopsis* was estimated based on ITS sequences. In the deep lineages of the *Brassaiopsis* species, four branches of phylogenetic importance have been identified. In group I, *B. griffithii* and *B. simplicifolia* were located in the earliest-diverging extant lineage within *Brassaiopsis* (Fig 8). Four species *B. elegans*, *B malayana*, *B. simplex*, and *B. sumatrana*, were located within group II (Fig 8). In group III, *B. ficifolia*, *B. grushvitzkyi*, *B. moumingensis*, *B. phanrangensis*, *B. producta*, and *B. stellata* formed a sister relationship. In the study of Mitchell et al. [1], *B. moumingensis* was found to be a sister with *B. grushvitzkyi*, and *B. stellata*. In group IV, *B. angustifolia* was a sister with the other 11 *Brassaiopsis* species (Fig 8). In addition, our phylogeny supports the following relationships: the sisterhood of the *B. hainla* and *B. aculeata*, flowed *B. glomerulata*, as found in previous studies [1, 2]. Because of the limited mutation sites in the ITS sequence, the systematic relationship among *Brassaiopsis* species has not been solved, and needs to be studied further.

Supporting information

S1 Table. Plant materials for data matrix I.
(DOCX)

S2 Table. Plant materials for data matrix II.
(DOCX)

S3 Table. Annotated genes of the *Brassaiopsis angustifolia* cp genome.
(DOCX)

Author Contributions

Data curation: Zhanghong Dong, Ruli Zhang, Jianzhong Ma, Peiyao Xin.

Formal analysis: Zhanghong Dong, Ruli Zhang, Yu Song.

Funding acquisition: Jianzhong Ma, Peiyao Xin.

Investigation: Zhanghong Dong, Ruli Zhang.

Methodology: Zhanghong Dong, Ruli Zhang, Ming Shi, Yu Song, Yaxuan Xin, Feng Li.

Project administration: Ruli Zhang, Ming Shi, Yaxuan Xin, Feng Li, Jianzhong Ma, Peiyao Xin.

Software: Zhanghong Dong, Ruli Zhang, Yu Song.

Supervision: Jianzhong Ma, Peiyao Xin.

Validation: Jianzhong Ma.

Visualization: Peiyao Xin.

Writing – original draft: Zhanghong Dong, Ruli Zhang, Yu Song.

Writing – review & editing: Yu Song, Jianzhong Ma, Peiyao Xin.

References

1. Mitchell A, Wen J. Phylogeny of *Brassaiopsis* (Araliaceae) in Asia based on nuclear ITS and 5S-NTS DNA sequences. Syst. Bot. 2005; 30(4):872–886. https://doi.org/10.1600/036364405775097761

2. Wen J, Zhu Y-P, Lee C, Widjaja E, Leng GS. Evolutionary relationships of Araliaceae in the Malesian region: a preliminary analysis. Acta Bot Yunnanica. 2008; 30:391–399.

3. 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):144–167. https://doi.org/10.1043/0363-6445-26.1.144

4. Mitchell A, Wen J. Phylogenetic utility and evidence for multiple copies of Granule-Bound Starch Synthase I (GBSSI) in Araliaceae. TAXON. 2004; 53(1):29–44. https://doi.org/10.2307/4135486

5. Plunkett G. M.; Wen J.; Lowry P. P. II Infrafamilial classifications and characters in Araliaceae: insights from the phylogenetic analysis of nuclear (ITS) and plastid (*trn*L-*trn*F) sequence data. Plant Syst. Evol. 2004; 245:1–39. https://doi.org/10.1007/s00606-003-0101-3

6. Nicolas AN, Plunkett GM. The demise of subfamily Hydrocotyloideae (Apiaceae) and the re-alignment of its genera across the entire order Apiales. Mol Phylogenet Evol. 2009; 53(1):134–151. https://doi.org/10.1016/j.ympev.2009.06.010 PMID: 19549570

7. Plunkett GM, Lowry PP, Fiaschi P, Frodin DG, Nicolas AN. Phylogeny, biogeography, and morphological evolution among and within the Neotropical and Asian clades of *Schefflera* (Araliaceae). TAXON. 2019; 68(8): 1278–1313. https://doi.org/10.1002/tax.12177

8. Li R, Wen J. Phylogeny and biogeography of Asian *Schefflera* (Araliaceae) based on nuclear and plastid DNA sequence data: Phylogeny and biogeography of Asian *Schefflera*. J. Syst. Evol. 2014; 52(4):431–449. https://doi.org/10.1111/jse.12052

9. Valcárcel V, Fiz-Palacios O, Wen J. The origin of the early differentiation of Ivies (*Hedera* L.) and the radiation of the Asian Palmate group (Araliaceae). Mol Phylogenet Evol. 2014; 70: 492–503. https://doi.org/10.1016/j.ympev.2013.10.016 PMID: 24184542

10. Mitchell A, Li R, Brown JW, Schönberger I, Wen J. Ancient divergence and biogeography of *Raukaua* (Araliaceae) and close relatives in the southern hemisphere. Aust Syst Bot. 2012; 25: 432. https://doi.org/10.1071/SB12020

11. Plunkett GM, Ii L, Frodin DG, Wen J. Phylogeny and geography of *Schefflera*: Pervasive polyphyly in the largest genus of Araliaceae. Ann Mo Bot Gard. 2005; 92: 202–224.

12. Kim K, Nguyen VB, Dong JZ, Wang Y, Park JY, Lee S-C, et al. Evolution of the Araliaceae family inferred from complete chloroplast genomes and 45S nrDNAs of 10 *Panax*-related species. Sci. Rep. 2017; 7(1):4917. https://doi.org/10.1038/s41598-017-05218-y PMID: 28687778

13. Li R, Ma P-F, Wen J, Yi T-S. Complete sequencing of five Araliaceae chloroplast genomes and the phylogenetic implications. PLoS ONE. 2013; 8(10):e78568. https://doi.org/10.1371/journal.pone.0078568 PMID: 24205264

14. Valcárcel V, Wen J. Chloroplast phylogenomic data support Eocene amphi-Pacific early radiation for the Asian Palmate core Araliaceae. J Syst Evol. 2019; 57: 547–560. https://doi.org/10.1111/jse.12522

15. Yi T, Lowry PP, Plunkett GM, Wen J. Chromosomal evolution in Araliaceae and close relatives. TAXON. 2004; 53:987–1005. https://doi.org/10.2307/4135565

16. Zhao YB, Yin JL, Guo HY, Zhang YY, Xiao W, Sun C, et al. The complete chloroplast genome provides insight into the evolution and polymorphism of *Panax ginseng*. Front. Plant Sci. 2015; 5:696. https://doi.org/10.3389/fpls.2014.00696 PMID: 25642231

17. Daniell H, Lin C-S, Yu M, Chang W-J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016; 17:134. https://doi.org/10.1186/s13059-016-1004-2 PMID: 27339192

18. Bendich A.J. Circular chloroplast chromosomes: the grand illusion. Plant Cell. 2004; 16(7):1661–1666. https://doi.org/10.1105/tpc.160771 PMID: 15235123

19. Sugiura M. The chloroplast genome. Plant Mol Biol. 1992; 19:149–168. https://doi.org/10.1007/BF00015612 PMID: 1600166

20. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci. 2005; 102(23):8369–8374. https://doi.org/10.1073/pnas.0503123102 PMID: 15928076

21. Parks M, Cronn R, Liston A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. 2009; 7(1):84. https://doi.org/10.1186/1741-7007-7-84 PMID: 19954512

22. Jansen RK, Cai Z, Raubeson LA, Daniell H, dePamphilis CW, Leebens-Mack J, et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci. 2007; 104(49):19369–19374. https://doi.org/10.1073/pnas.0709121104 PMID: 18048330

23. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc Natl Acad Sci. 2010; 107(10):4623–4628. https://doi.org/10.1073/pnas.0907801107 PMID: 20176954

24. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987; 19(1):11–15.

25. Zhang T, Zeng C-X, Yang J-B, Li H-T, Li D-Z. Fifteen novel universal primer pairs for sequencing whole chloroplast genomes and a primer pair for nuclear ribosomal DNAs: universal primer pairs for sequencing whole chloroplast genomes and nuclear ribosomal DNAs. J. Syst. Evol. 2016; 54(3):219–227. https://doi.org/10.1111/jse.12197

26. Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, et al. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020; 21:241, https://doi.org/10.1186/s13059-020-02154-5 PMID: 32912315

27. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: Interactive visualization of de novo genome assemblies. Bioinformatics. 2015; 31(20):3350–3352, https://doi.org/10.1093/bioinformatics/btv383 PMID: 26099265

28. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28(12):1647–1649. https://doi.org/10.1093/bioinformatics/bts199 PMID: 22543367

29. Lohse M, Drechsel O, Kahlau S, Bock R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res. 2013; 41:W575–W581. https://doi.org/10.1093/nar/gkt289 PMID: 23609545

30. Peden JF. Analysis of codon usage. Biosystems. 1999; 5:45–50.

31. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: A web server for microsatellite prediction. Bioinformatics. 2017; 33(16):2583–2585. https://doi.org/10.1093/bioinformatics/btx198 PMID: 28398459

32. Kurtz S, Choudhuri JJ, 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):4633–4642. https://doi.org/10.1093/nar/29.22.4633 PMID: 11713313

33. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999; 27:573–580. https://doi.org/10.1093/nar/27.2.573 PMID: 9862982

34. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 2013; 30:772–780. https://doi.org/10.1093/molbev/mst010 PMID: 23329690

35. Hall TA. BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symp. Ser. 1999; 41(41):95–98. https://doi.org/10.1021/bk-1999-0734.ch008

36. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004; 32:W273–W279. https://doi.org/10.1093/nar/gkh458 PMID: 15215394

37. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large datasets. Mol. Biol. Evol. 2017; 34(12):3299–3302. https://doi.org/10.1093/molbev/msx248 PMID: 29029172

38. Amiryousefi A, Hyvönen J, Poczai P. IRscope: An online program to visualize the junction sites of chloroplast genomes. Hancock J, editor. Bioinformatics. 2018; 34(17):3030–3031. https://doi.org/10.1093/bioinformatics/bty220 PMID: 29659705

39. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003; 19(12):1572–1574. https://doi.org/10.1093/bioinformatics/btg180 PMID: 12912839

40. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 2012; 9(8):772–772. https://doi.org/10.1038/nmeth.2109 PMID: 22847109

41. Nguyen L-T, Schmidt HA, Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 2015; 32(1):268–274. https://doi.org/10.1093/molbev/msu300 PMID: 25371430

42. Mardanov AV, Ravin NV, Kuznetsov BB, Samigullin TH, Antonov AS, Kolganova TV, et al. Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. J. Mol. Evol. 2008; 66(6):555–564. https://doi.org/10.1007/s00239-008-9091-7 PMID: 18463914

43. Song Y, Dong WP, Liu B, Xu C, Yao X, Gao J, et al. Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. Front. Plant Sci. 2015; 6:662. https://doi.org/10.3389/fpls.2015.00662 PMID: 26379689

44. Wu ZH, Liao R, Yang TG, Dong X, Lan DQ, Qin R, et al. Analysis of six chloroplast genomes provides insight into the evolution of *Chrysosplenium* (Saxifragaceae). BMC Genomics 2020; 21:621. https://doi.org/10.1186/s12864-020-07045-4 PMID: 32912155

45. Shaw J, Lickey EB, Schilling EE, Small RL. 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; 94(3):275–288. https://doi.org/10.3732/ajb.94.3.275 PMID: 21636401

46. Worberg A, Quandt D, Barniske A-M, Löhne C, Hilu KW, Borsch T. Phylogeny of basal eudicots: insights from non-coding and rapidly evolving DNA. Org. Divers. Evol. 2007; 7(1):55–77. https://doi.org/10.1016/j.ode.2006.08.001

47. Dong WP, Xu C, Li CH, Sun JH, Zuo YJ, Shi S, et al. *ycf1*, the most promising plastid DNA barcode of land plants. Sci. Rep. 2015; 5(1):8348. https://doi.org/10.1038/srep08348 PMID: 25672218

48. Song Y, Yao X, Liu B, Tan YH, Corlett RT. Complete plastid genome sequences of three tropical *Alseodaphne* trees in the family Lauraceae. Holzforschung. 2018; 72(4):337–345. https://doi.org/10.1515/hf-2017-0065

49. Dong WP, Liu H, Xu C, Zuo YJ, Chen ZJ, Zhou SL. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. BMC Genet. 2014; 15(1):138. https://doi.org/10.1186/s12863-014-0138-z PMID: 25526752

50. Song Y, Yao X, Tan YH, Gan Y, Corlett RT. Complete chloroplast genome sequence of the avocado: gene organization, comparative analysis, and phylogenetic relationships with other Lauraceae. Can. J. For. Res. 2016; 46(11):1293–1301. https://doi.org/10.1139/cjfr-2016-0199

51. Zhao M-L, Song Y, Ni J, Yao X, Tan YH, Xu Z-F. Comparative chloroplast genomics and phylogenetics of nine *Lindera* species (Lauraceae). Sci. Rep. 2018; 8(1):8844. https://doi.org/10.1038/s41598-018-27090-0 PMID: 29891996

52. Zong D, Gan PH, Zhou AP, Zhang Y, Zou XL, Duan AA, et al. Plastome sequences help to resolve deep-level relationships of *Populus* in the family Salicaceae. Front. Plant Sci. 2019; 10:5. https://doi.org/10.3389/fpls.2019.00005 PMID: 30723484

53. Song Y, Yao X, Tan YH. Gan Y, Yang J-B, Corlett RT. Comparative analysis of complete chloroplast genome sequences of two subtropical trees, *Phoebe sheareri* and *Phoebe omeiensis* (Lauraceae). Tree Genet. Genomes. 2017, 13(6):120. https://doi.org/10.1007/s11295-017-1196-y

54. Yang J-B, Li D-Z, Li H-T. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. Mol Ecol Resour. 2014; 14(5):1024–1031. https://doi.org/10.1111/1755-0998.12251 PMID: 24620934

55. Zhang YY, Du LW, Liu A, Chen JJ, Wu L, Hu WM, et al. The complete chloroplast genome sequences of five *Epimedium* species: lights into phylogenetic and taxonomic analyses. Front. Plant Sci. 2016, 7, 306. https://doi.org/10.3389/fpls.2016.00306 PMID: 27014326

56. Dong ZH, Qu SH, Landrein S, Yu W-B, Xin J, Zhao WZ, et al. Increasing taxa sampling provides new insights on the phylogenetic relationship between Eriobotrya and Rhaphiolepis. Front Genet. 2022; 13: 831206. https://doi.org/10.3389/fgene.2022.831206 PMID: 35368713

57. Alvizu A, Eilertsen MH, Xavier JR, Rapp HT. Increased taxon sampling provides new insights into the phylogeny and evolution of the subclass Calcaronea (Porifera, Calcarea). Org Divers Evol. 2018; 18: 279–290. https://doi.org/10.1007/s13127-018-0368-4

58. Pollock DD, Zwickl DJ, McGuire JA, Hillis DM. Increased taxon sampling is advantageous for phylogenetic inference. Syst Biol. 2002; 51: 664–671. https://doi.org/10.1080/10635150290102357 PMID: 12228008

59. Lowry PP II, Plunkett GM, Mora MM, Cano A, Fiaschi P, Frodin DG, et al. Studies in Neotropical Araliaceae. I. Resurrection of the genus *Sciodaphyllum* P. Browne to accommodate most New World species previously included in *Schefflera* J. R. Forst. & G. Forst. Brittonia. 2020; 72: 1–15. https://doi.org/10.1007/s12228-019-09593-w

60. Lowry PP II, Plunkett GM, Neill DA. Studies in Neotropical Araliaceae. II. Resurrection of the Neotropical genus Crepinella for a clade of New World species previously included in Schefflera (Araliaceae). Novon. 2019; 27: 253–261. https://doi.org/10.3417/2019510

61. Fiaschi P, Lowry PP, Plunkett GM. Studies in Neotropical Araliaceae. III. Resurrection of the New World genus *Didymopanax* Decne. & Planch., previously included in *Schefflera* (Araliaceae). Brittonia. 2020; 72: 16–22. https://doi.org/10.1007/s12228-019-09604-w

62. Plunkett GM, Lowry PP, Neill DA. Studies in Neotropical Araliaceae. VII. Two new genera, *Cephalopanax* and *Frodinia*, to accommodate the remaining species of Neotropical *Schefflera*. Brittonia. 2021; 73: 251–261. https://doi.org/10.1007/s12228-021-09660-1

63. Lowry PP II, Plunkett GM. Resurrection of the Genus *Heptapleurum* for the Asian Clade of species previously included in *Schefflera* (Araliaceae). Novon. 2020; 28: 143–170. https://doi.org/10.3417/2020612

64. Esser H-J. *Schefflera* (Araliaceae) in Thailand becomes *Heptapleurum*. Thai For Bull Bot. 2021; 49: 151–154. https://doi.org/10.20531/tfb.2021.49.1.19