scientific reports

Check for updates

OPEN

The complete chloroplast genomes of four *Aspidopterys* species and a comparison with other Malpighiaceae species

He Liao[1,2], Huayuan Chen[1,2] & Shinan Liu[1]✉

The genus *Aspidopterys* has multiple functions in medicine, food and ecological restoration. Due to the similar morphological characteristics of some species and the limited genomic information hinder the studies on germplasm identification and molecular phylogeny analysis. In this study, we compared and explored the six complete chloroplast (cp) genomes including four *Aspidopterys* species (*A. glabriuscula*, *A. concava*, *A. cavaleriei*, *A. obcordata*), *Banisteriopsis caapi* and *Bunchosia argentea*. Their cp genomes in length were 158,473 to 161,091 bp, displaying the high conserved degree in the structure, gene arrangement and GC content. Moreover, 57–80 long repeats and 61–92 SSRs were identified, most of which were forward or palindromic repeats and mononucleotides, respectively. Eleven non-coding regions and 12 coding regions, especially *ndhH_ndhA*, *rpl32_ndhF* and *ycf1*, had the higher nucleotide diversity values that could can be regarded as DNA barcodes of Malpighiaceae species. In addition, the 9 genes (like *accD*, *atpE*, *atpF*, *clpP*) were conducted positive selection (Ka/Ks > 1). As indicated by phylogenetic analysis, those four *Aspidopterys* were clustered into single clade with other Malpighiaceae species and were closely related to *B. caapi* and *B. argentea*. This study sheds more lights on further phylogenetic, evolutionary and genetic diversity studies on the genus *Aspidopterys* and even the Malpighiaceae species.

Keywords *Aspidopterys*, Chloroplast genome, Comparative analysis, Phylogeny

The chloroplast (cp) plays a critical role in various crucial biological and chemical processes such as photosynthesis of plants[1]. Cp genomes in many land plants are 107 to 218 kb long[2]. Relative to the nuclear and mitochondrial genomes, genetic information in cp genomes is relatively conserved[2,3]. The cp genome shows the representative quadripartite structure, which includes a large single-copy (LSC) region, a small single-copy (SSC) region as well as two inverted repeats (IR) regions[4–6], with LSC and SSC regions being separated by two IRs. Length discrepancy in the cp genome length is mostly caused by IR/SC junction contraction/expansion[7]. Notably, cp genome is characterized by easy structure, multiple copies and small molecule, which is suitable for the study on plant phylogeny[8,9]. The hypervariable regions in cp genomes are recognized to be the DNA molecular markers for studying the phylogenetic relationships at the interspecific or intraspecific level like *Artemisia* L.[10], *Eriocaulon* L.[11], and *Pholidota* Lindl.[12]. As next-generation sequencing develops, genomes can be obtained in an easier, faster and more cost-effective manner[13]. Therefore, the cp genome-scale data can be extensively adopted for inferring plant phylogenetic and evolutionary relations in different families, different genera of the same family or even different species of the same genera[14,15].

Malpighiaceae are widely distributed in tropical and subtropical forests with ~1250 species of 65 genera including trees, shrubs and vines[16,17]. Though, the monophyly of Malpighiaceae has been confirmed by morphological and molecular data[18–22], the relationships among groups within the family or genera are not figured out. In China, approximately 23 endemic species are assigned to the genera *Aspidopterys* A. Juss., *Hiptage* Gaertn., *Ryssopterys* Blume and *Tristellateia Thouars*, which are all woody vines. *Aspidopterys* comprises 10 species distributed in the southwest and south of China. Therein, the *Aspidopterys obcordata* (Wall.) A. Juss. species raises more awareness because more than 50 compounds are found in this plant as a folk medicine for treating urolithiasis[23–25]. Besides, *Aspidopterys concava* (Wall.) A. Juss. is only endemic to Guangxi province and has been eaten as vegetables by local residents. Moreover, the researchers also find that it is rich in nutrient

[1]Guangxi Colleges and Universities Key Laboratory for Cultivation and Utilization of Subtropical Forest Plantation, College of Forestry, Guangxi University, Nanning 530004, China. [2]He Liao and Huayuan Chen have contributed equally to this work. ✉email: gxlxylsn2024@gxu.edu.cn

elements and essential amino acids, which is superior to *Erythropalum scandens* Bl., *Sauropus androgynus* (Linn.) Merr. and *Lycium barbarum* Linn. that have been recognized as the edible woody vegetables[26–28]. From the above, it is supposed that *Aspidopterys* can be used as the candidate for exploitation and utilization, so the report of their cp genomes sequences will be beneficial for conservation.

Until now, several cp genomes from *Banisteriopsis* C. B. Rob., *Bunchosia* Kunth, *Brysonima* Rich. ex Kunth, and *Galphimia* Cav. in the Malpighiaceae family have been published[29]. *Aspidopterys* belongs to the same subfamily with *Banisteriopsis* and *Bunchosia*. However, if the cp genome data can figure out the phylogenetic relationships among *Aspidopterys* species and different genus of Malpighiaceae family was unclear. So, this study performed the comparative analysis on the three cp genomes to be sequenced and three published genomes from *Aspidopterys*, *Banisteriopsis* and *Bunchosia*. Our goal was presented as following: (1) enrich the GenBank database to supplement the genetic information for chloroplast genome research of Malpighiaceae; (2) analyze the highly variable regions and select molecular markers for identifying Malpighiaceae and exploring DNA barcoding; and (3) discuss their phylogenetic relationships in Malpighiaceae.

Materials and methods

Plant materials and DNA isolation

Fresh and young leaf samples from *Aspidopterys glabriuscula* (Wall.) A. Juss, *Aspidopterys cavaleriei* (Wall.) A. Juss., and *A. concava* were collected in Guangxi Zhuang Autonomous Region of China (coordinates: 22°45′20.16″N, 107°05′32.14″E; 22°15′14.47″N, 107°00′14.34″E; 22°15′32.95″ N, 106°49′40.06″E), respectively. The plants were identified by Dr. Shengxiang Yu (State Key Laboratory of Plant Diversity and Specialty Crops, Institute of Botany, Chinese Academy of Sciences). The specimens were kept at the College of Forestry, Guangxi University (GAC). Their voucher specimen numbers of *A. glabriuscula*, *A. concava* and *A. cavaleriei* were LSN0120230808, LSN0320230808 and LSN0220230808, respectively. These three species are common species, so there is no need for permission. Total genomic DNA was isolated with using Plant DNAzol Reagent (Invitrogen) following the corresponding instruction. DNA quality was evaluated by using 1% agarose gels and NanoDropTM One spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

Cp genome sequencing, assembly and annotation

We constructed the library with an average length of 350 bp using the Nextera XT DNA Libraries Preparation Kit (Illumina, San Diego, CA). The libraries were then sequenced on the Illumina Novaseq 6000 platform. Raw sequence reads underwent quality control processing using the Trimmomatic[30]. The raw data totaled 6.54G–8.96G, and the clean data totaled 6.53G–8.93G after quality control processing. The Q30 value was > 90%, indicating that the quality of mitochondrial genome sequencing and assembly results was very high. The high-quality clean reads were assembled into the complete chloroplast genome using de novo assembler SPAdes v.3.14.1 software with a K-mer 93, 95, 97, 103, 105, 107 and 115[31]. Finally, it was annotated by PGA with *A. obcordata* (MT590775) as a reference genome. OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) was employed for drawing circular representation for these sequences[32].

Repeat sequencing analysis

Microsatellites (mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats) were detected by Perl script MISA upon the thresholds below (unit size, min repeats): 10, 5, 4, 3, 3 and 3 repeat units for mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide SSRs, separately[32,33]. In addition, we employed REPuter[34] (http://bibiserv.techfak.unibielefeld.de/reputer/) was used to analyze repeat sequences, including complement, forward, reverse, tandem and palindromic repeats, the minimal length and edit distance were 30 bp and ≤ 3 bp.

Genome comparison and sequence divergence analysis

We compared the cp genome boundaries of *A. glabriuscula*, *A. concave*, *A. cavaleriei, A. obcordata*, *Banisteriopsis caapi* (Spruce ex Griseb.) Morton (NC_037945) and *Bunchosia argentea* (Jacq.) DC (NC_041491) for LSC, SSC, and IRs based on corresponding annotations with IRscope (https://irscope.shi-nyapps.io/irapp/)[35]. The software mVISTA (http://genome.lbl.gov/vista/mvista/submit.shtml) was used in the Shuffle-LAGAN mode[36] to compare the six cp genomes with the *A. glabriuscula* genome as the reference. It was necessary to manually adjust overlapping gene names in the software results. In the resulting figures, the ordinate represents the similarity between each sample and the control sample, with the displayed interval ranging from 50 to 100%. Therefore, regions of unevenness in the graph indicate significant differences between the sample and the control. The nucleotide variability (Pi) of the cp genome was assessed using DnaSP version 5.1 software through sliding window analysis, with a step size and window length of 100 and 600 base pairs, respectively[37].

Ka/Ks and codon usage analysis

The synonymous (Ks) and non-synonymous (Ka) substitution rates were analyzed to investigate the molecular evolutionary process of the Malpighiaceae family. The protein-coding genes (PCGs) of *A. glabriuscula* were compared with those of *A. concava*, *A. obcordata*, *A. cavaleriei*, *B. caapi* and *B. argentea* respectively using ParaAT 2.0 with the default parameters[38]. Subsequently, the Ka/Ks values for each gene were calculated using KaKs_Calculator 2.0 with the YN method[39]. We further elucidated selective pressure on protein-coding genes using additional codon models such as fast unconstrained Bayesian AppRoximation (FUBAR) with HyPhy using the Datamonkey server[40]. FUBAR was used with posterior probability of > 0.9 to identify episodic/diversifying selection on codons sites. Additionally, the relative synonymous codon usage (RSCU) values in PCGs of these six cp genomes were assessed by CodonW v1.3[41].

Phylogenetic analysis

A total of 17 cp genome sequences were selected for the phylogenetic analysis, and their GenBank accession numbers were presented in Table S1. *Averrhoa carambola* L. (KU569488) was selected as the outgroup. The multiple alignments of the complete cp genome, LSC, SSC, and IR sequences were aligned by MAFFT v7.429[42]. 66 orthologous genes in each sample with a copy number were identified by OrthoFinder v2.3.14. Then the best-fit models of Maximum-likelihood (ML) analysis and Bayesian inference (BI) analysis were selected by IQ-TREE v1.6.1 and PhyloSuite, respectively. Based on the complete cp genome sequences and 66 single-copy protein-coding genes (PCGs), both ML trees were respectively constructed using GTR + G and GTR + F + I + G4 model under 1000 bootstrap replicates[42,43]. Differently, ML trees of LSC and SSC were constructed by TVM + F + R2 model, while ML tree of IR sequence was constructed by GTR + F + G4 model under same replicates[43]. Moreover, according to above different datasets, both BI trees of the entire cp genome and 66 cp genes were respectively formed in MrBayes using GTR + F + G4 and GTR + F + I + G4 model, with 1,000,000 generations and sampling once every 1000 generations. BI trees of LSC, SSC, and IR sequences were formed using GTR + F + G4 model with 900,000, 300,000, and 500,000 generations, respectively. The first 25% of trees from all runs were discard as burn-in, and the remaining trees were employed to construct a majority-rule consensus tree.

Results

Cp genome analysis

This study obtained approximately 21.8 M to 59.7 M raw reads of *A. cavaleriei*, *A. glabriuscula* and *A. concava*, which were further assembled and annotated. They all had the representative tetrad structure with one large single region (LSC, 88,475–89,112 bp) and one small single-copy region (SSC, 18,032–18,156 bp) together with two inverted repeats (IRs, 26,914–26,923 bp) (Fig. 1). The cp genome sizes in *A. cavaleriei*, *A. glabriuscula* and *A. concava* were 160,805 bp, 160,341 bp and 161,091 bp, with the coverage depth of 1006×, 2957×, and 1642×,

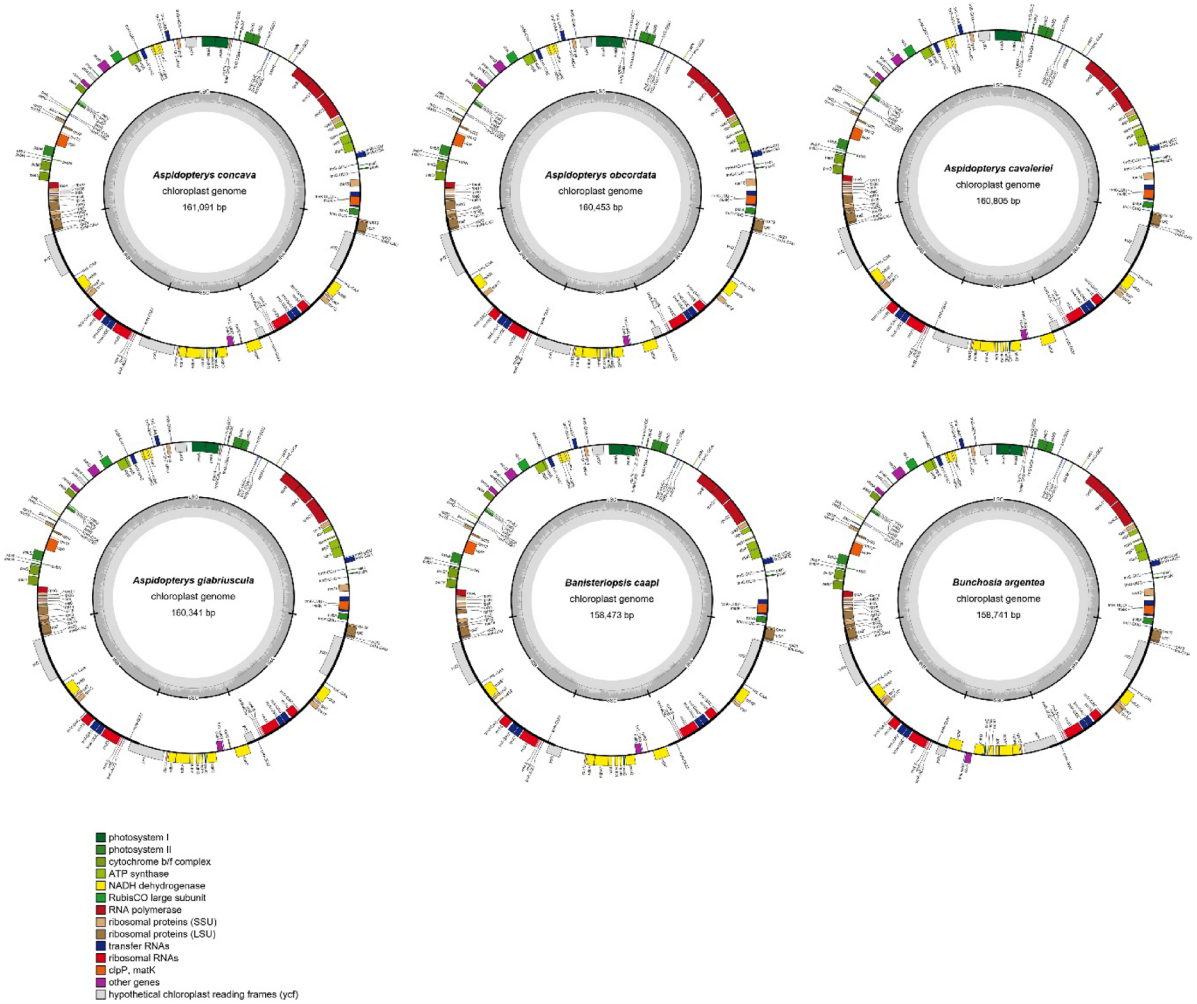

Fig. 1. Chloroplast (cp) genome map for the *A. glabriuscula*, *A. cavaleriei*, *A. concava*, *A. obcordata*, *B. caapi* and *B. argentea*. Genes inside and outside the circle are transcribed clockwise and counter -clockwise separately. Diverse colors represent various functions shown in left corner of bottom panel. The lighter gray and darker gray represent the AT and GC contents, separately.

Genome features	*A. glabriuscula*	*A. concava*	*A. cavaleriei*	*A. obcordata*	*B. caapi*	*B. argentea*
Genome size (bp)	16,0341	161,091	160,805	160,453	158,473	158,741
LSC size (bp)	88,475	89,112	88,821	88,491	87,646	88,345
SSC size(bp)	18,032	18,133	18,152	18,156	18,931	16,816
IR size(bp)	26,917	26,923	26,914	26,905	25,948	26,790
Total GC content (%)	36.7	36.6	36.6	36.6	36.8	36.5
GC content within LSC (%)	34.4	34.25	34.27	34.33	30.93	34.13
GC content within SSC (%)	30.35	30.55	30.5	30.49	34.57	30.23
GC content within IR (%)	42.49	42.48	42.48	42.47	42.79	42.31
Total gene number	132	132	131	126	125	127
Protein-coding gene number	87	87	86	83	80	82
tRNA gene number	37	37	37	37	37	37
rRNA gene number	8	8	8	6	8	8

Table 1. Six chloroplast genomes features.

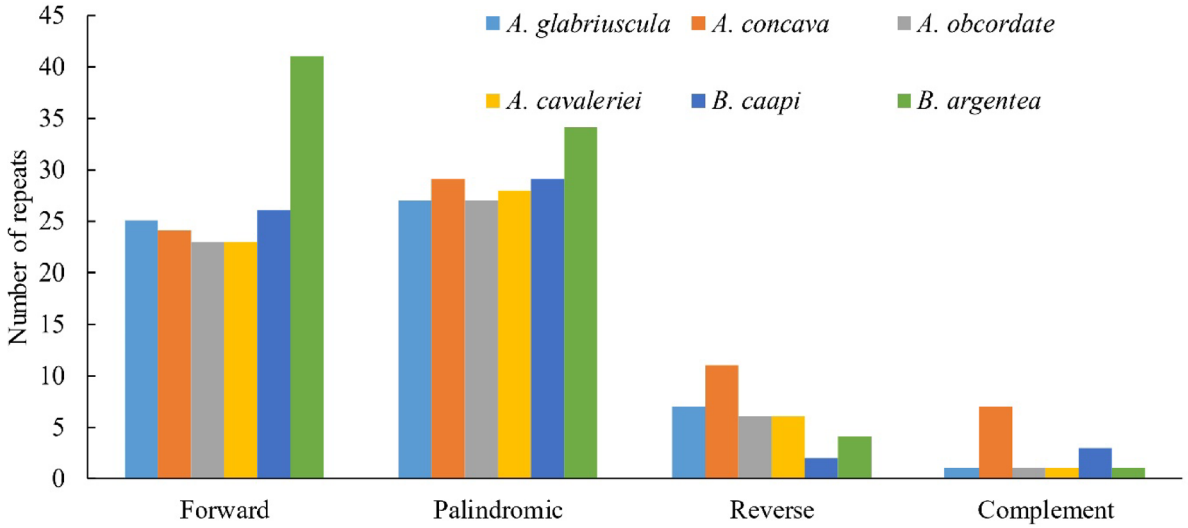

Fig. 2. Long repeat numbers and types in *A. glabriuscula*, *A. cavaleriei*, *A. concava*, *A. obcordata*, *B. caapi* and *B. argentea*.

respectively. The other three published cp genomes (*A. obcordata*, *B. caapi* and *B. argentea*) are also presented in Table 1. The total GC levels were 36.5–36.8%, so there was no significant difference among these six species. With regard to the *B. caapi* cp genome, the order of GC content in three regions was IR > SSC > LSC, while that for the rest five cp genomes was IR > LSC > SSC. The cp genomes in *A. glabriuscula* and *A. concava* contained 132 functional genes, consisting of 87 PCGs, and 8 ribosomal RNA genes (rRNAs) as well as 37 transfer RNA genes (tRNAs). Compared with the above two species, the lower PCGs numbers were observed within cp genomes in *A. cavaleriei*, *B. caapi* and *B. argentea*, while the less PCGs and tRNA numbers were found within cp genome in *A. obcordata* (Table 1). Among those genes (Table S2), 17 harbored introns, including 3 (*rpl2*, *clpP* and *ycf3*) possessing 2 introns, while others (*ndhA*, *ndhB*, *petB*, *atpF*, *rpl16*, *rpl2*, *rps16*, *rpoC1*, *trnA-UGC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, *trnS-CGA*, *trnV-UAC*) only had one intron. Moreover, 19 genes were multi-copy genes, of which *trnM-CAU* contained four copies, while the remaining 18 genes had two copies. Moreover, *rpl32* gene was absent within SSC region in cp genomes of *A. obcordata*, *A. cavaleriei* and *B. argentea*. The cp genome of *A. obcordata* lost *infA* in the LSC region, as well as *rps7* and *rrn4.5* within IR region, while that of *B. caapi* lost *rps16* in the LSC region. In addition, *infA*, *rps7* and *ycf1* genes were pseudogenized in *B. caapi* and *B. argentea*.

Dispersed repeats and SSRs analysis

As revealed by the analysis, cp genomes in *A. glabriuscula*, *A. cavaleriei*, *A. concava*, *A. obcordata*, *B. caapi* and *B. argentea* all had 4 repeat types, namely, the palindromic, forward, reverse and complement repeats, with forward and palindromic types being dominant (Fig. 2). *A. obcordata* had least repeats (n = 57), while the highest number was 80 from *B. argentea* (Fig. 2). Of every type, the repeat length was ≥ 30 bp, while the sequence similarity was greater than or equal to 90%. Most of them were 30–39 bp in size (Table S2). Meanwhile, the repeats length > 49 bp increased within the cp genome in *B. caapi* compared with other five species (Table S3).

Through MISA analysis, the cp genomes in 6 species contained 61–92 SSRs (*B. argentea*: 92; *A. concava*: 88; *A. obcordata*: 86; *A. cavaleriei*: 81; *A. glabriuscula*: 76; *B. caapi*: 61) (Fig. 3). Among the total SSRs, the mononucleotide repeats occupied the highest proportion of 56.6%–71.7%, followed by dinucleotides and tetranucleotides. Meanwhile, there were slightly more trinucleotides than pentanucleotides. Most of these SSRs were the mononucleotide ademine (A) and thymine (T) repeats, accounting for 55.3% in *A. glabriuscula*, 59.1% in *A. concava*, 61.7% in *A. cavaleriei*, 67.2% in *B. caapi*, 68.6% in *A. obcordata* and 71.7% in *B. argentea*. Compared with the four *Aspidopterys* species, *B. caapi* and *B. argentea* did not have cytosine (C) or guanine (G) units in the mononucleotide repeats. Moreover, there were some unique units in trinucleotides, tetranucleotides and penta-nucleotides of some species such as AAT, TAA, TTGA, TTTC, AAATT and TCCAT in *B. argentea*, ATT, TTC, TAAT in *A. concava* and AAAT in *A. glabriuscula* (Table S4).

Codon usage

The cp genomes of *A. glabriuscula*, *A. concava*, *A. obcordata*, *A. cavaleriei*, *B. caapi* and *B. argentea* had 23,952–26,788 codons, all which encoded 1 stop codon together with 20 amino acids (Fig. 4). The number of codon or each amino acid of *Aspidopterys* was higher than those of *B. caapi* and *B. argentea*. Among these codons, the leucine had the highest frequency in the above species, then isoleucine and serine, whereas stop codon exhibited the lowest frequency. AUU (encoding isoleucine) and UGA (encoding translational stop). 64 relative synonymous codon usages (RSCU) could be identified within genomes in these six species (Table S5). 30 codons were found with an RSCU > 1 in *A. glabriuscula*, *A. concava*, *A. obcordata* and *B. argentea*, while 31 codons were observed with an RSCU > 1 in *A. cavaleriei* and *B. caapi*, including the A/U-ending codons except UUG. There were 34 codons whose RSCU ≤ 1 in *A. glabriuscula*, *A. concava*, *A. obcordata* and *B. argentea*, while 33 codons were found with an RSCU ≤ 1 in *A. cavaleriei* and *B. caapi*, including the C/G-ending codons with the exception

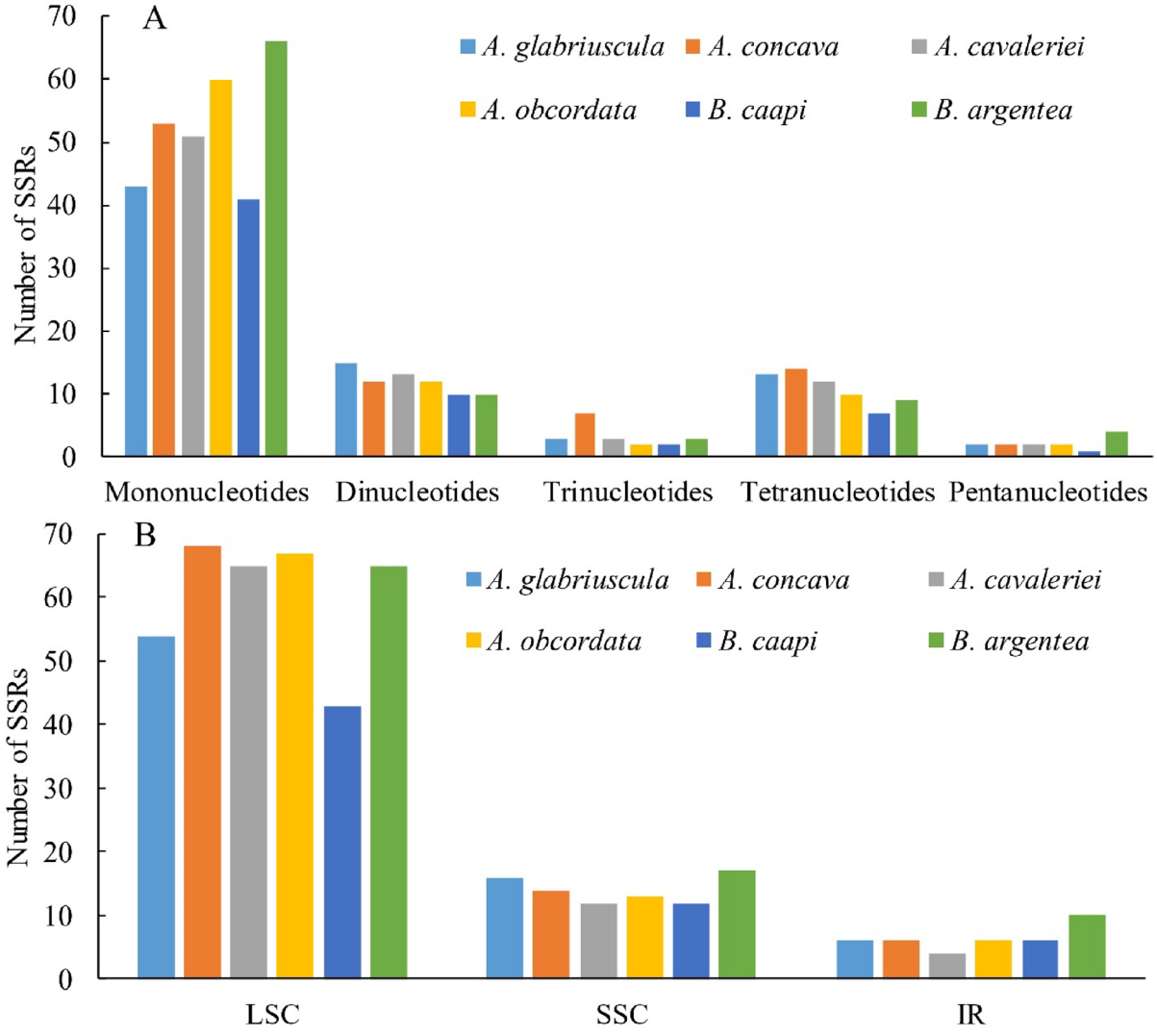

Fig. 3. The simple sequence repeat (SSR) numbers and types within cp genomes in *A. glabriuscula*, *A. cavaleriei*, *A. concava*, *A. obcordata*, *B. caapi* and *B. argentea*. (**A**) Diverse repeat type numbers. (**B**) Repeat type frequencies within the LSC, SSC, and IR regions.

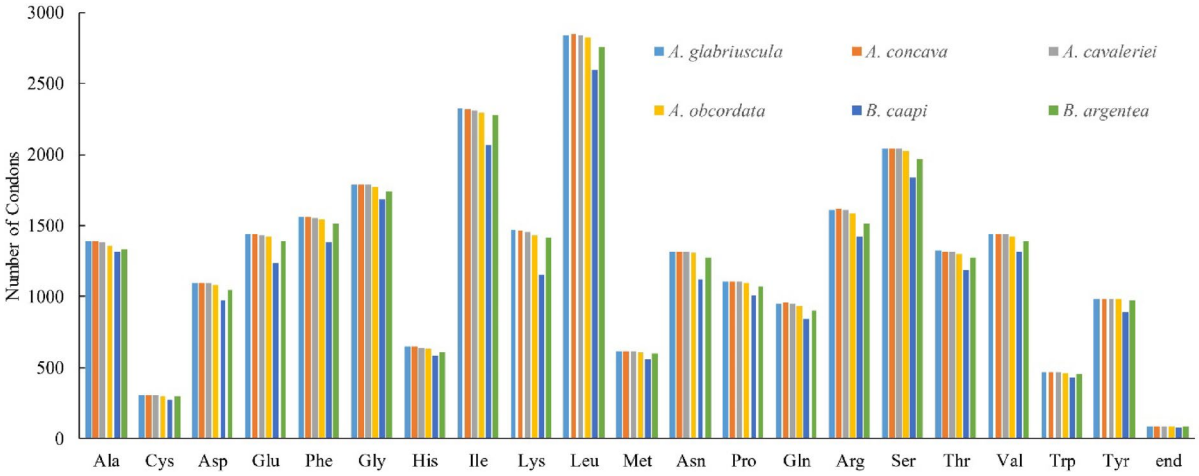

Fig. 4. Dispersed repeat numbers and types in *A. glabriuscula*, *A. cavaleriei*, *A. concava*, *A. obcordata*, *B. caapi* and *B. argentea*.

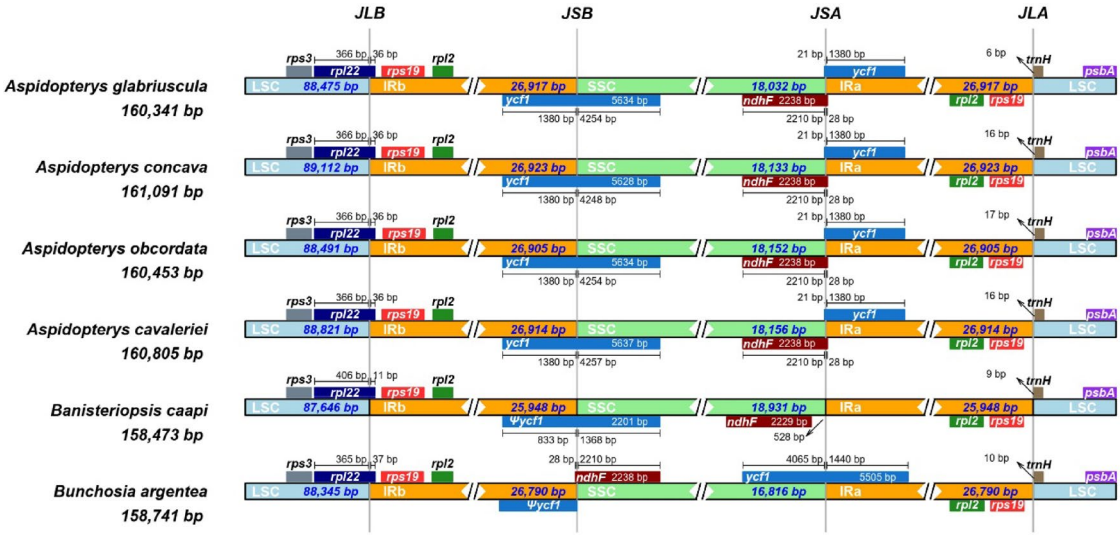

Fig. 5. Comparisons of border region of LSC, SSC and IR region in cp genomes of *A. glabriuscula*, *A. cavaleriei*, *A. concava*, *A. obcordata*, *B. caapi* and *B. argentea*.

of AUA, CUA, CCA and UGA. AGA (arginine, 1.9), AGA (arginine, 1.9), UAA (leucine, 1.88), AGA (arginine, 1.89), UUA (leucine, 1.86), UUA (leucine, 1.92) of *A. glabriuscula*, *A. concava*, *A. obcordata*, *A. cavaleriei*, *B. caapi* and *B. argentea* had greatest RSCU values, respectively.

IR contraction and expansion

Comparisons of IR/SC junctions among *A. glabriuscula*, *A. concava*, *A. cavaleriei*, *A. obcordata*, *B. caapi* and *B. argentea* cp genomes were illustrated in Fig. 5. Overall, the gene size, category and order in cp genomes were highly similar among the four *Aspidopterys* species, but some differences existed among three genera. The results revealed that LSC/IRb junction in the six species was located within *rpl22* gene, which extended lengths of 36 bp, 11 bp and 37 bp in the IRb region of four *Aspidopterys* species, *B. caapi* and *B. argentea*, respectively. *ycf1* gene was positioned within IRb/SSC and SSC/IRa boundary of four *Aspidopterys* species, with 1380 bp overlapping with the IRb region and 21 bp overlap with SSC region. Differently, *ycf1* gene was positioned within the SSC/IRa region, with 1440 bp overlapping with the IRa in *B. argentea* but not presented in the same region of *B. caapi*; while the pseudogene ($ycf1^{\Psi}$) was found to be positioned in the IRb/SSC boundary or IRb region in both species. Besides, *ndhF* gene was also positioned within SSC region, with 28 bp overlapping with IRa in four *Aspidopterys* species, meanwhile, it was positioned within SSC region, with 28 bp overlapping with IRb of *B. argentea*, and was only presented within SSC region in *B. caapi*.

Divergence of cp genome sequences

We also compared six cp genomes with mVISTA. The most variable regions were detected within in SSC region, and then LSC region. Relative to the LSC and SSC regions, IR regions showed the lower divergence. In total, cp genome sequences for 6 species were strongly divergent among different genera than within the same genus.

The nucleotide diversity values were determined for assessing the sequence divergence level[7]. For LSC regions, Pi values were 0–0.0433 (mean, 0.01465) (Fig. S1). For the SSC region, Pi values were 0.00657–0.24161 (mean, 0.19075), meanwhile, 90.8% of this region had the values >0.1 (Fig. S1). For the IRs regions, Pi values were 0–0.0927 (mean, 0.00410) (Fig. S1). The coding regions showed higher conserved degree than the non-coding gene regions. Additionally, the higher divergent genes included *ndhH*, *ndhA*, *ndhE*, *ndhI*, *ndhG*, *ndhD*, *psaC*, *ccsA*, *rpl32*, *ndhF*, *rps15*, and *ycf1* for coding regions, and those were *rpl32_ndhF*, *ndhE_psaC*, *trnL-UAG_rpl32*, *ndhI_ndhG*, *psaC_ndhD*, *ndhD_ccsA*, *ndhG_ndhE*, *ccsA_trnL-UAG*, *ndhA_ndhI*, *rps15_ndhH*, and *ycf1_rps15* in non-coding regions (Fig. 6). These genes were all positioned within the SSC region, which could serve as candidate markers used to reconstruct phylogeny and identify species in this subgenus.

Selective pressure during *A. glabriuscula* evolution

We determined synonymous (Ks) and non-synonymous (Ka) values and Ka/Ks ratio for 55 PCGs and compared them within six cp genomes with *A. glabriuscula* as a reference. Ka values were 0–0.116, while Ks values were 0–0.175 in all genes. As shown in Table S6, we found that there were 9 genes (*accD*, *atpE*, *atpF*, *clpP*, *rpl20*, *rpl22*, *rpl32*, *rpoB*, *ycf2*) impacted by positive selection (Ka/Ks > 1), whereas those rest 46 genes could be impacted by the purifying selection (Ka/Ks < 1). The only one gene (*rpoB*) in *A. glabriuscula* vs *A. obcordate* and two genes (*atpE* and *rpl32*) in *A. glabriuscula* vs *B. caapi* were affected by positive selection. Compared with *A. glabriuscula* vs *A. obcordate* and *A. glabriuscula* vs *B. caapi*, there were more genes (*accD*, *atpE*, *clpP*, *rpl20*, *rpl22*, *ycf2*) in *A. glabriuscula* vs *B. argentea* that were affected by positive selection. The Ka/Ks ratio peaked at 15.689 in the *clpP* gene. In addition, gene selection pressure was estimated by FUBAR. The results showed that 16 PCGs were under positive selection (Table S7). The *rbcL* had most positive selective sits (5), followed by *atpF* (4), *rpoC1* (4), *ccsA* (3), *clpP* (2), and *ycf4* (2), while the rest genes possessed one positive selective site (Table S7). Thus, the results of both methods were not completely consistent except for the *atpF*, *clpP*, and *rpl20* genes.

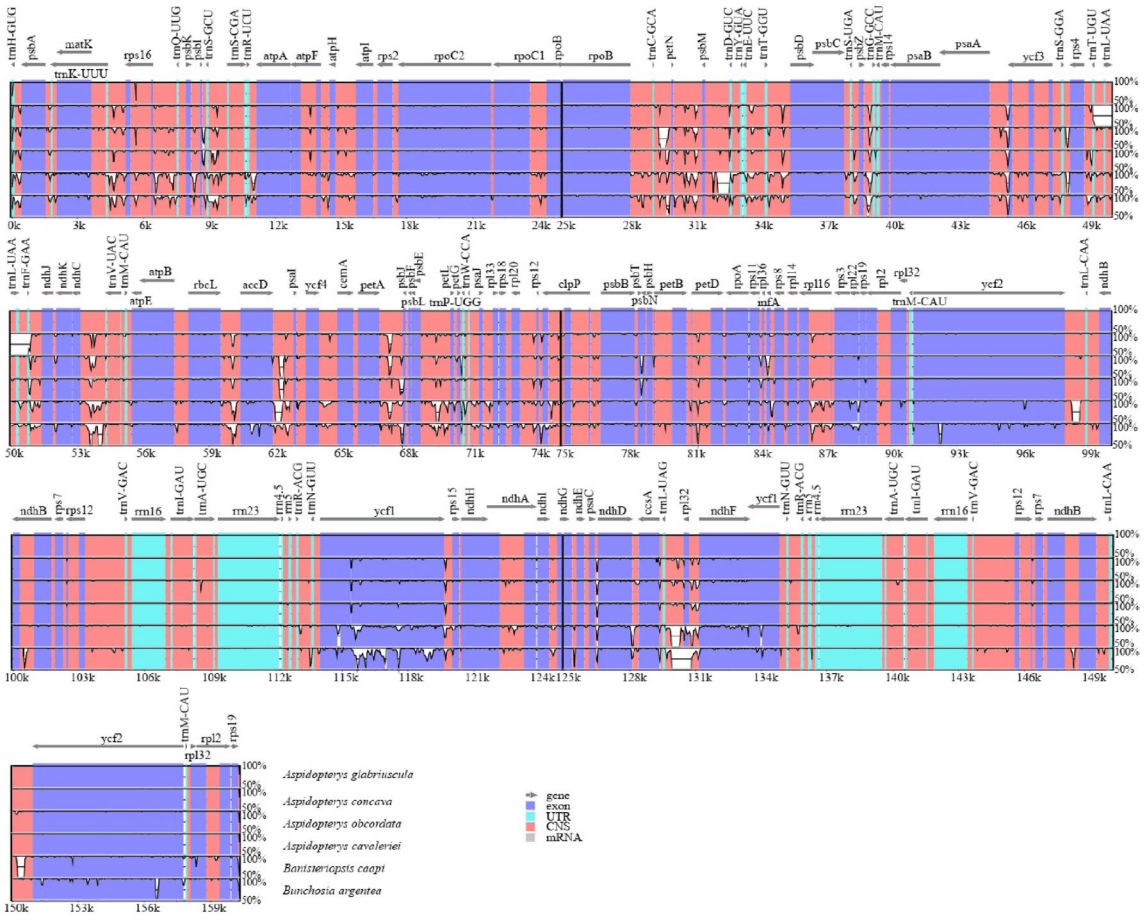

Fig. 6. Visualized alignment of the chloroplast genomes among *A. glabriuscula*, *A. cavaleriei*, *A. concava*, *A. obcordata*, *B. caapi* and *B. argentea* and was performed with *A. glabriuscula* being the control. Vertical scale represent identity the percentage of 50–100%.

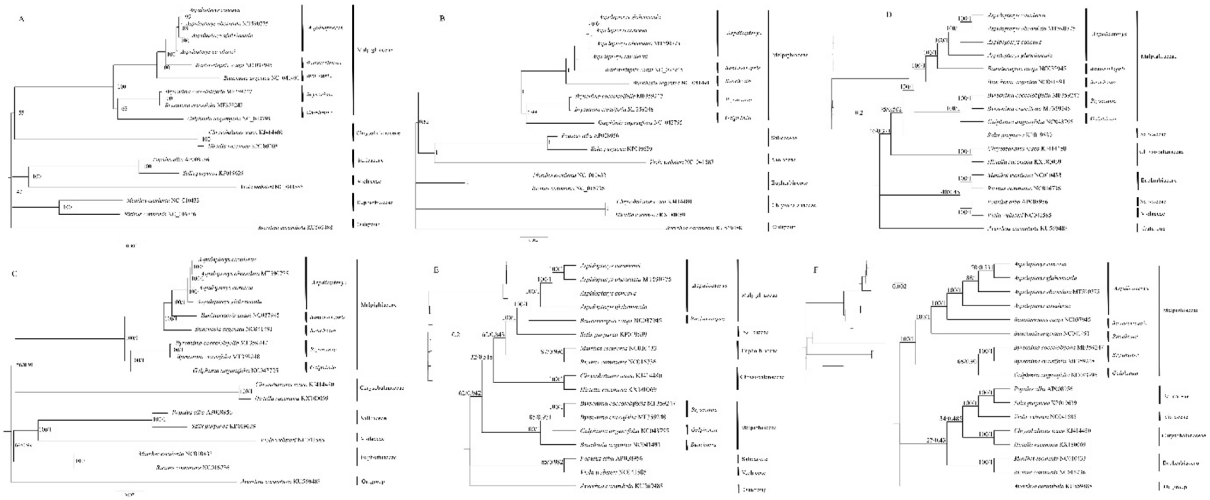

Fig. 7. Phylogenetic relationships among 17 species detected by maximum likelihood (ML) analysis and Backyesian inference (BI) analysis constructed by complete cp genome sequences, 66 single-copy PCGs, LSC, SSC, and IR regions. (**A**)The phylogenetic tree was constructed according to the complete cp genome sequences by ML methods. (**B**) The phylogenetic tree was constructed according to the complete cp genome sequences by BI method. (**C**) The phylogenetic tree was constructed according to the 66 PCGs by ML and BI methods. (**D**) The phylogenetic tree was constructed according to the LSC region by ML and BI methods. (**E**) The phylogenetic tree was constructed according to the SSC region by ML and BI methods. (**F**) The phylogenetic tree was constructed according to the IR region by ML and BI methods.

Phylogenetic analysis

For determining the positions of these *Aspidopterys* species, the other 13 species in NCBI were chosen for constructing a phylogenetic tree according to complete cp genomes, LSC, SSC, IR sequences, and 66 single-copy PCGs by using ML and BI approaches. Based on the complete genomes, PCGs, and IR sequence, the constructed trees showed that Malpighiaceae, Chrysobalanceae, Salicaceae, Violaceae, and Euphorbiacea grouped into a single clade (Fig. 7A, B, C, F). Except for the formed tree according to the SSC region (Fig. 7E), others trees based on four datasets indicated that nine Malpighiaceae species clustered together; moreover, four *Aspidopterys* species, *B. cappi* and *B. argentea* formed a subclade, while *Brysonima coccolobifolia* Kunth, *B. crassifolia* (L.) Kunth, and *G. anustifolia* formed another subclade (Fig. 7A, B, C, D, F). Furthermore, the results showed that *A. concava* is more closely related to *A. obcordata* by the ML analysis based on the complete cp genome (Fig. 7A). Meanwhile, BI tree indicated that *A. concava* had a closer relationship with *A. glabriuscula* based on the complete cp genome, same as the results of ML/BI trees by being constructed according to the IR region (Fig. 7B, F). In addition, the phylogenetic topologies derived from the ML and BI analysis of the PCGs, LSC, and SSC datasets were highly similar, indicating that *A. obcordata* was closely related to *A. cavaleriei* (Fig. 7C, D, E).

Discussion

The cp genomes in angiosperms have a greatly conserved size, structure and content[44,45]. Their cp genomes are generally 120–170 kb in length, and encode 120–130 genes[46]. Like other angiosperms, the cp genome of four *Aspidopterys* cp genomes also displayed a characteristic quadripartite structure, including LSC, SSC and IRs regions with the size of 160,341 bp–161,091 bp, consistent with other published Malpighiaceae cp genomes such as *B. caapi* and *B. argentea*. The above six Malpiphiaceae cp genomes ranged 125 to 132 unique genes, including 80 to 87 PCG genes, 6 to 8 tRNA and 37 rRNA genes. Notably, *A. glabriuscula*, *A. concava* and *A. cavaleriei* possessed more than 130 unique genes, significantly more than those observed in *A. obcordata*, *B. caapi* and *B. argentea*. Besides, the cp genomes in *B. coccolobifolia* and *B. crassifolia* both contained 8 tRNA and 37 rRNA genes[29]. These results showed that the number of rRNA genes was more conserved than those of PCG and tRNA genes in the Malpighiaceae family. The sequential stability is determined by its GC content. The sequence owns the higher GC content, indicating that it is more stable and vice versa[47]. In our study, the total GC contents were similar among the six cp genomes, consistent with those of *B. coccolobifolia* and *B. crassifolia*. However, the GC content varied significantly among the three regions. Except *B. caapi*, the GC content in other 5 species followed the order of IR > LSC > SSC. Moreover, the Pi value within IR region was far lower than in the LSC (0.01465) and SSC (0.19075) regions. Based on these results, the IR region showed a higher conservation degree in the Malpighiaceae family. This point is similar to other angiosperms.

Although IR region is generally conserved, IR boundary contraction and expansion can be detected within cp genomes in many species[48]. These variations are a key factor in influencing cp genome size, which is crucial for maintaining structural stability and driving the evolution of the cp genome[49]. In our study, we noticed that the IR/SC boundary genes were *rpl22*, *ycf1*, *ndhF* and *trnH*, which are consistent with observations in *Byrsonima* from the same family[29], and *Artemisia* from different families[10]. The IRb/LSC boundary gene in all Malphighiales species was *rpl22* with different extensions (Fig. 5, S2). In four *Aspidopterys* species, the IRb/SSC junction was

ycf1 gene, while the SSC/IRa boundary was *ndhF* and *ycf1* genes (Fig. 5). Conversely, the IRb/SSC boundary in *B. argentea*, *B. coccolobifolia*, *B. crassifolia*, and *G. angustiflolia* was the truncated of *ycf1* (a pseudogene) and *ndhF* genes, whereas the SSC/IRa junction was the truncated of *ycf1* gene (Fig. 5, S2). Different from other species, in *B. caapi*, the IRb/SSC junction was the *ycf1* pseudogene, while *ndhF* was observed to present entirely in the SSC region (Fig. 5). Kim et al. reported that *ndhF* gene was associated with the IR/SSC junction stability[50]. Notably, the *ycf1* gene was longer in the SSC/IRa boundary that had switched the position with *ndhF* gene in the IRb/SSC boundary in *B. argentea*, *B. coccolobifolia*, *B. crassifolia*, and *G. angustiflolia* (Fig. 5, S2). Therefore, the results suggested that the expansion or contraction of the IR boundary in Malpighiaceae species may attribute to the variations of *ycf1* and *ndhF* genes.

Generally, gene gain or loss event always occurs within the cp DNA in angiosperms plants during the evolution[9,51,52]. Menezes et al. found that variations among the Malphighiales species were attributed to 3 PCGs, *rps16*, *rpl32* or *infA*[29]. In Malpighiaceae, the gene *infA* was found to be absent in the *B. coccolobifolia* and *B. crassifolia*[29]. This phenomenon was also observed in *A. obocordata*, *B. caapi*, and *G. angustifolia* (Table S2, Fig. S3). *rpl32* gene was also found to be completely lost in *A. cavaleriei*, *A. obcordata*, and *B. argentea* (Fig. S3), contrasting with the previously and our own findings in other cp genomes from the same family. Moreover, compared to other species, *A. obcordata* and *B. caapi* completely lost *rps7* and *rps16*, respectively. Besides, *ycf15* gene only existed in *Byrsonima* genus (Table S2, Fig. S3). These results further suggested that the not only *rpl32*, *infA,* and *rps16* but also *rps7* and *ycf15* may represent variable genes among the Malpighiaceae species. Similar loss events have been documented in other plant species: *rps16* gene in Gentianaceae[53], *ycf15* gene in Loranthaceae[54], *rps7* gene in *Passiflora* L., *infA* gene in Rosids[55], and *rpl32* gene in Euphorbiaceae and Rhizophoraceae[56,57]. Studies have indicated that the loss of *rps7*, *rpl32* and *infA* genes may be due to their transfer to the nuclear genome[55,58], while the loss of *rps16* may be correlated with substitution by nuclear-encoded homologs[54,59]. However, the consequences of *ycf15* loss remained to be explored. Further researches are warranted to investigate the losing reasons of these genes in Malpighiaceae. In addition, the chloroplast genomes of *B. argentea, B. coccolobifolia, B. crassifolia,* and *G. angustiflolia* lacked a copy of the *rps19* gene, while that of *B. caapi* also lost a copy of *rps12* (Table S2, Fig. S3). Taken together, it is speculated that the absence of genes and the variation in copy number may be the primary reasons for the differences in gene count, which in turn may lead to size variations.

The sequential discrepancy among cp genomes of different species can be identified and further applied in phylogenetic analysis to distinguish closely related species[60]. By calculating the nucleotide diversity levels, we recognized highly variable regions including 11 non-coding regions and 12 coding regions. Notably, the regions *ndhH_ndhA*, *rpl32_ndhF* and *ycf1* exhibited significantly higher Pi values than others. These regions have also been found to be highly variable in several other species such as Asteraceae, Magnoliaceae, Anacardiaceae and Hydrangeaceae[61–63]. Moreover, they have been widely employed in phylogenetic studies across diverse taxa, including Hydrangeeae, Magnoliaceae, *Pinus* L. and *Cerasus* Mill.[64–67]. Therefore, these fragments hold great potential as DNA barcodes and molecular markers for phylogenetic and taxonomic applications.

The complete cp genome is widely recognized as a reliable resource for inferring evolutionary and phylogenetic relationships among species[68]. In our study, the phylogenetic tree based on the complete cp genomes, cp genes, LSC, and IR sequences (with moderate to high support) except for SSC sequence (with low support) showed that *Aspidopterys* clustered together with *Banisteriopsis* and *Bunchosia*, while *Brysonima* and *Galphimia* formed a separate clade, conforming to prior findings. For instance, Cameron et al. used *matK* and *rbcL* sequences to demonstrate the close relationship between *Brysonima* and *Galphimia*[69], consistent with the result obtained by Davis et al. using *ndhF* and *trnL-F* sequences[21]. Meanwhile, Davis et al. also showed that *Aspidopterys* was more closely related to *Banisteriopsis* than to *Bunchosia* based on the noncoding *trnL-F* sequence[21]. This conclusion was in agreement with our study. Besides, the stamen characteristics supported the close affinity of *Aspidopterys* with *Banisteriopsis* and *Bunchosia*[19], further corroborating our genomic findings. Furthermore, the relationships of four *Aspidopterys* species in different trees based on five datasets were inconsistent. *A. concava* was closely related to *A. obcordata* by ML analysis according to the entire cp genome. However, it had a closer relationship with *A. glabriuscula* by BI analysis based the entire cp genome and by ML and BI analysis based on the IR region. The rest of trees based on other datasets showed that *A. obcordata* was closely related to *A. cavaleriei* with high support. Based on morphology, *A. obcordata* and *A. cavaleriei* were also found to be most closely related in *Aspidopterys* genus, which was supported by the phylogeny inferred from the three different datasets of PCGs, LSC, and SSC regions using the ML and BI methods. Taken together, with a view to the classification of families and genus of these selected species, these findings collectively suggested that the PCGs, and LSC regions were high effective in resolving the phylogenetic relationships within the Magnoliaceae family and the *Aspidopterys* genus.

Conclusions

This study sequenced the complete cp genomes in *A. glabriuscula*, *A. cavaleriei* and *A. concava* and compared them with the published cp genomes of *A. obcordata*, *B. caapi* and *B. argentea* from Malpighiaceae family. The cp genomes of these six species showed highly conserved sequence sizes, GC content, and gene numbers. However, some variations in repeat structures and gene expansions were observed in the IR-SC boundary regions among different species, particularly between genera. Besides, the highly variable regions (*ndhH_ndhA*, *rpl32_ndhF* and *ycf1*) in LSC were recognized as potential DNA markers for investigating genetic diversity and resolving taxonomical discrepancies within *Aspidopterys*, or even across the Malpighiaceae family.

Data availability

The chloroplast genome datasets of A. glabriuscula, A. cavaleriei and A. concava have been deposited in the GenBank of NCBI under the accession number PQ112543, PQ112544 and PQ112545, respectively.

Received: 22 July 2024; Accepted: 7 May 2025
Published online: 23 May 2025

References

1. Neuhaus, H. E. & Emes, M. J. Nonphotosynthetic metabolism in plastids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51**, 111–140 (2000).
2. Daniell, H., Lin, C. S., Yu, M. & Chang, W. J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* **17**, 1–29 (2016).
3. Yang, J. et al. Development of chloroplast and nuclear DNA markers for Chinese oaks (*Quercus subgenus Quercus*) and assessment of their utility as DNA barcodes. *Front. Plant Sci.* **8**, 816 (2017).
4. Abdullah, et al. Correlations among oligonucleotide repeats, nucleotide substitutions, and insertion–deletion mutations in chloroplast genomes of plant family Malvaceae. *J. Syst. Evol.* **59**, 388–402 (2021).
5. Bock, R. Plastid biotechnology: Prospects for herbicide and insect resistance, metabolic engineering and molecular farming. *Curr. Opin. Biotech.* **18**, 100–106 (2007).
6. Henriquez, C. L. et al. Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae). *Genomics* **112**, 2349–2360 (2020).
7. Mo, Z. H. et al. The chloroplast genome of *Carya illinoinensis*: genome structure, adaptive evolution, and phylogenetic analysis. *Forests* **11**, 207 (2020).
8. Dong, W. L. et al. Molecular evolution of chloroplast genomes of orchid species: Insights into phylogenetic relationship and adaptive evolution. *Int. J. Mol. Sci.* **19**, 716 (2018).
9. Xu, J. H. et al. Dynamics of chloroplast genomes in green plants. *Genomics* **106**, 221–231 (2015).
10. Chen, C. et al. Complete chloroplast genomes of eight *Artemisia* species: comparative analysis, molecular identification, and phylogenetic analysis. *Plant Biol.* **26**, 257–269 (2024).
11. Li, E. et al. Insights into the phylogeny and chloroplast genome evolution of *Eriocaulon* (Eriocaulaceae). *BMC Plant Biol.* **23**, 32 (2023).
12. Li, L. et al. Comparative analyses and phylogenetic relationships of thirteen *Pholidota* species (Orchidaceae) inferred from complete chloroplast genomes. *BMC Plant Biol.* **23**, 269 (2023).
13. Alkan, C., Sajjadian, S. & Eichler, E. E. Limitations of next-generation genome sequence assembly. *Nat. Methods* **8**, 61–65 (2011).
14. Du, Y. et al. Complete chloroplast genome sequences of *Lilium*: Insights into evolutionary dynamics and phylogenetic analyses. *Sci. Rep.* **7**, 1–10 (2017).
15. Yang, Y. et al. Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Front. Plant Sci.* **7**, 959 (2016).
16. Charles, C. et al. Phylogeny of Malpighiaceae: Evidence from chloroplast *ndhF* and *trnL-F* nucleotide sequences. *Am. J. Bot.* **88**, 1830–1846 (2001).
17. Davis, C. C., Bell, C. D., Mathews, S. & Donoghue, M. J. Laurasian migration explains Gondwanan disjunctions: Evidence from Malpighiaceae. *PNAS* **99**, 6833–6837 (2002).
18. Anderson, W. R. Floral conservatism in Neotropical Malpighiaceae. *Biotropica* **11**, 219–223 (1979).
19. Anderson, W. R. The origin of the Malpighiaceae: The evidence from morphology. *Mem. N. Y. Bot. Gard.* **64**, 210–224 (1990).
20. Araújo, J. S., Azevedo, A. A., Silva, L. C. & Meira, R. M. S. A. Leaf anatomy as an additional taxonomy tool for 16 species of Malpighiaceae found in the Cerrado area (Brazil). *Plant Syst. Evol.* **286**, 117–131 (2010).
21. Davis, C. C., Anderson, W. R. & Donoghue, M. J. Phylogeny Malpighiaceae: Evidence from chloroplast *ndhF* and *trnL-F* nucleotide sequences. *Am. J. Bot.* **88**, 1830–1846 (2001).
22. Davis, C. C. & Anderson, W. R. A complete generic phylogeny of Malpighiaceae inferred from nucleotide sequence data and morphology. *Am. J. Bot.* **97**, 2031–2048 (2010).
23. Hu, M. et al. New polyoxypregnane glycosides from *Aspidopterys obcordata* vines with antitumor activity. *Fitoterapia* **129**, 203–209 (2018).
24. Khan, A., Bashir, S. & Khan, S. R. Antiurolithic effects of medicinal plants: Results of in vivo studies in rat models of calcium oxalate nephrolithiasis—a systematic review. *Urolithiasis* **49**, 95–122 (2021).
25. Sun, P. et al. Aspidoptoids A–D: Four new diterpenoids from *Aspidopterys obcordata* vine. *Molecules* **25**, 529 (2020).
26. Arif, T. Therapeutic potential and traditional uses of *Sauropus androgynous*: A review. *J. Pharmacogn. Phytochem.* **9**, 2131–2137 (2020).
27. Mocan, A. et al. Polyphenolic content, antioxidant and antimicrobial activities of *Lycium barbarum* L. and *Lycium chinense* Mill. Leaves. *Molecules* **19**, 1005–10073 (2014).
28. Yang, T. et al. Evaluating the genetic diversity of *Erythropalum scandens* based on using inter-simple sequence repeat markers. *Genet. Resour. Crop Ev.* **70**, 2377–2390 (2023).
29. Menezes, A. P. A. et al. Chloroplast genomes of *Byrsonima* species (Malpighiaceae): Comparative analysis and screening of high divergence sequences. *Sci. Rep.* **8**, 2210 (2018).
30. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2020 (2014).
31. Bankevich, A. et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
32. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **47**, W59–W64 (2019).
33. Thiel, T., Michalek, W., Varshney, R. K. & Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **106**, 411–422 (2003).
34. Kurtz, S. et al. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **29**, 4633–4642 (2001).
35. Amiryousefi, A., Hyvönen, J. & Poczai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **34**, 3030–3031 (2018).
36. Frazer, K. A., et al. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **32**(suppl_2): W273–W279 (2004).
37. Librado, P. & Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
38. Zhang, Z. et al. ParaAT: A parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* **419**, 779–781 (2012).
39. Wang, D. et al. KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom. Bioinform.* **8**, 77–80 (2010).
40. Murrell, B. et al. FUBAR: A fast, unconstrained bayesian approximation for inferring selection. *Mol. Biol. Evol.* **30**(5), 1196–1205 (2013).
41. Sharp, P. M. & Li, W. H. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* **24**, 28–38 (1986).
42. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* **20**, 1160–1166 (2019).

43. Kozlov, A. M. et al. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
44. Guo, Y. Y. et al. Chloroplast genomes of two species of *Cypripedium*: expanded genome size and proliferation of AT-biased repeat sequences. *Front. Plant Sci.* **12**, 609729 (2021).
45. Li, Y. et al. Complete chloroplast genome of an endangered species *Quercus litseoides*, and its comparative, evolutionary, and phylogenetic study with other *Quercus* section *Cyclobalanopsis* species. *Genes* **13**, 1184 (2022).
46. Green, B. R. Chloroplast genomes of photosynthetic eukaryotes. *Plant J.* **66**, 34–44 (2011).
47. Benjamini, Y. & Speed, T. P. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* **40**, e72–e72 (2012).
48. Maréchal, A. & Brisson, N. Recombination and the maintenance of plant organelle genome stability. *New Phytol.* **186**, 299–317 (2010).
49. He, L. et al. Complete chloroplast genome of medicinal plant *Lonicera japonica*: Genome rearrangement, intron gain and loss, and implications for phylogenetic studies. *Molecules* **22**, 249 (2017).
50. Kim, H. T. et al. Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across orchids and associated instability of the inverted repeat/small single-copy region boundaries. *PLoS ONE* **10**, e0142215 (2015).
51. Cao, J. et al. Extreme reconfiguration of plastid genomes in Papaveraceae: Rearrangements, gene loss, pseudogenization, IR expansion, and repeats. *Int. J. Mol. Sci.* **25**, 2278 (2024).
52. Liu, M. et al. Adaptive evolution of chloroplast division mechanisms during plant terrestrialization. *Cell Rep.* **43**, 113950 (2024).
53. Jansen, R. K. et al. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae). *Mol. Phylogenet. Evol.* **48**, 1204–1217 (2008).
54. Nie, L. et al. Gene losses and variations in chloroplast genome of parasitic plant macrosolen and phylogenetic relationships within santalales. *Int. J. Mol. Sci.* **20**(22), 5812 (2019).
55. Millen, R. S. et al. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* **13**, 645–658 (2001).
56. Alqahtani, A. A. & Jansen, R. K. The evolutionary fate of *rpl32* and *rps16* losses in the *Euphorbia schimperi* (Euphorbiaceae) plastome. *Sci. Rep.* **11**, 7466 (2021).
57. Ruang-Areerate, P. et al. Comparative analysis and phylogenetic relationships of *Ceriops* species (Rhizophoraceae) and *Avicennia lanata* (Acanthaceae): insight into the chloroplast genome evolution between middle and seaward zones of mangrove forests. *Biology* **11**, 383 (2022).
58. Park, S., Jansen, R. K. & Park, S. Complete plastome sequence of *Thalictrum coreanum* (Ranunculaceae) and transfer of the *rpl32* gene to the nucleus in the ancestor of the subfamily Thalictroideae. *BMC Plant Biol.* **15**, 40 (2015).
59. Claude, S. J., Park, S. & Park, S. J. Gene loss, genome rearrangement, and accelerated substitution rates in plastid genome of *Hypericum ascyron* (Hypericaceae). *BMC Plant Biol.* **22**, 135 (2022).
60. Gostel, M. R. & Kress, W. J. The expanding role of DNA barcodes: Indispensable tools for ecology, evolution, and conservation. *Diversity* **14**, 213 (2022).
61. Loeuille, B. et al. Extremely low nucleotide diversity among thirty-six new chloroplast genome sequences from Aldama (Heliantheae, Asteraceae) and comparative chloroplast genomics analyses with closely related genera. *Peer J.* **9**, e10886 (2021).
62. Xin, Y. et al. Comparative analyses of 18 complete chloroplast genomes from eleven *Mangifera* species (Anacardiaceae): Sequence characteristics and phylogenomics. *Horticulturae* **9**, 86 (2023).
63. Sun, S. S. et al. The complete plastome sequences of seven species in *Gentiana* sect. *Kudoa* (Gent-ianaceae): Insights into plastid gene loss and molecular evolution. *Front. Plant Sci.* **9**, 493 (2018).
64. De Smet, Y. et al. Molecular phylogenetics and new (infra) generic classification to alleviate polyphyly in tribe Hydrangeeae (Cornales: Hydrangeaceae). *Taxon* **64**, 741–753 (2015).
65. Kim, M. & Kim, T. J. Genetic species identification using *ycf1b*, *rbcL*, and *trnH-psbA* in the Genus *Pinus* as a complementary method for anatomical wood species identification. *Forests* **14**, 1095 (2023).
66. Wan, T. et al. Evolutionary and phylogenetic analyses of 11 *Cerasus* species based on the complete chloroplast genome. *Front. Plant Sci.* **14**, 1070600 (2023).
67. Yang, A. et al. New insight into the phylogeographic pattern of *Liiodendron chinense* (Magnoliaceae) revealed by chloroplast DNA: east-west lineage split and genetic mixture within western subtropical China. *Peer J.* **7**, e6355 (2019).
68. Dobrogojski, J., Adamiec, M. & Luciński, R. The chloroplast genome: A review. *Acta Physiol. Plant.* **42**, 98 (2020).
69. Cameron, K. M. et al. Molecular systematics of Malpighiaceae: Evidence from plastid rbcL and matK sequences. *Am. J. Bot.* **88**, 1847–1862 (2001).

Acknowledgements

This work has been supported by the Guangxi Key Research and Development Project (Grant numbers: GuiKe AB21238014).

Author contributions

H. L. and H. Y. C. collected samples, performed the experiments, analyzed the data. H. L wrote the draft of manuscript. S. N. L conceptualized and designed this research project, reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-01724-6.

Correspondence and requests for materials should be addressed to S.L.