Comparative chloroplast genome analyses of *Paraboea* (Gesneriaceae): Insights into adaptive evolution and phylogenetic analysis

Yifei Wang¹,2†, Fang Wen³†, Xin Hong⁴,5, Zhenglong Li⁴,5, Yaolei Mi⁶* and Bo Zhao¹,2,3*

¹Department of Pharmacognosy, Guilin Medical University, Guilin, China, ²Department of Pharmacy, Guilin Medical University, Guilin, China, ³Guangxi Key Laboratory of Plant Conservation and Restoration Ecology in Karst Terrain, Guangxi Institute of Botany, Guangxi Zhuang Autonomous Region and Chinese Academy of Sciences, Guilin, China, ⁴Anhui Provincial Engineering Laboratory of Wetland Ecosystem Protection and Restoration, School of Resources and Environmental Engineering, Anhui University, Hefei, China, ⁵Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Populations, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China, ⁶Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China

*These authors have contributed equally to this work

*Correspondence
Bo Zhao
122017017@glmc.edu.cn
Yaolei Mi
xiaomi20063@sina.com

Paraboea (Gesneriaceae) distributed in the karst areas of South and Southwest China and Southeast Asia, is an ideal genus to study the phylogeny and adaptive evolution of karst plants. In this study, the complete chloroplast genomes of twelve *Paraboea* species were sequenced and analyzed. Twelve chloroplast genomes ranged in size from 153166 to 154245 bp. Each chloroplast genome had a typical quartile structure, and relatively conserved type and number of gene components, including 131 genes which are composed of 87 protein coding genes, 36 transfer RNAs and 8 ribosomal RNAs. A total of 600 simple sequence repeats and 389 non-overlapped sequence repeats were obtained from the twelve *Paraboea* chloroplast genomes. We found ten divergent regions (trnH-GUG-psbA, trnM-CAU, trnC-GCA, atpF-atpH, ycf1, trnK-UUU-rps16, rps15, petL, trnS-GCU-trnR-UCU and psaJ-rpl33) among the twelve *Paraboea* species to be potential molecular markers. In the phylogenetic tree of 31 Gesneriaceae plants including twelve *Paraboea* species, all *Paraboea* species clustered in a clade and confirmed the monophyly of *Paraboea*. Nine genes with positive selection sites were detected, most of which were related to photosynthesis and protein synthesis, and might played crucial roles in the adaptability of *Paraboea* to diverse karst environments. These findings are valuable for further study of the phylogeny and karst adaptability of Gesneriaceae plants.

**KEYWORDS**
Paraboea, phylogenetic, positive selection, chloroplast, genome
Introduction

Limestone areas have diverse and unique regional microhabitat and special island habitats (Xin et al., 2021). These special habitats (such as stone mountains, karst caves, Tiankeng, etc.) provide very favorable conditions for species isolation and differentiation. After a long period of evolution and natural selection, limestone areas have bred a high degree of species diversity and significant endemism. With the development of the economy, due to frequent human activities (such as tourism development, etc.), the fragile ecological balance of these diversified environments is easily destroyed, and the survival of karst plants is threatened. Studying the adaptive evolution of karst plants has extremely important practical significance for the protection of karst plants (Tao et al., 2016).

Paraboea, an important genus of Gesneriaceae, contains about 144 species (Xu et al., 2012; Wen et al., 2013; Puglisi and Phutthai, 2017). The genus is mainly distributed in karst areas of south China, southwest China and southeast Asia, China has about 29 species, of which 19 are endemic (Guo et al., 2016). Among these species, some are widely distributed and attached to the rock gaps of stone mountains in direct sunlight (such as P. rufescens and P. swinhoei), some are distributed in dark and wet caves (such as P. filipes), and some grow on limestone rocks or inter stone soil in the dark place under the dense forest from the hillside to the top of the mountain (such as P. barbatipes) (Zhou et al., 2003; Gao et al., 2006). Considering the distribution of Paraboea species in diverse microhabitats, it becomes an excellent group to study the adaptive evolution of karst plants. However, the phylogenetic study of Paraboea is mainly based on a small number of chloroplast or nuclear gene markers, and the phylogenetic relationship has not been completely solved, which greatly limits the discussion of adaptive evolution of genes and traits. Based on nuclear ITS (internal transcribed spacer) sequences and chloroplast genome sequences (trnH-psbA spacer), the phylogenetic relationship of Paraboea was reconstructed, and the taxonomic boundaries among some related species were clarified (Li and Wang, 2007; Puglisi et al., 2011; Xin et al., 2019; Guo et al., 2020). However, the existing chloroplast genome sequences cannot completely solve the phylogenetic relationship of Paraboea. It is necessary to add faster and more suitable molecular sequences to reconstruct its phylogenetic relationship.

The chloroplast genome is one of the important molecular tools to study plant adaptive evolution. The challenging environment may impose selective pressure on genes related to photosynthesis, leaving the footprints of natural selection on genes. The main protein coding genes of the chloroplast genome include those controlling genetic and photosynthetic systems as well as genes encoding other functions. Photosynthetic system genes are genes related to photosynthesis, which are responsible for encoding members of ATP synthase, Rubisco large subunit, NADPH dehydrogenase and photosystem I and II (Zhang et al., 2018). The adaptive evolution analysis showed that chloroplast genes related to photosynthesis generally had positive selection sites in plants living in various extreme environments, and these gene regions might play a crucial role in plant adaptation to different environments (Chen et al., 2021).

The chloroplast genome sequences not only provide full-length protein coding sequences for the adaptive evolution of genes related to photosynthesis under the selection pressure of different environments, but also screen suitable hypervariable regions to solve the phylogenetic relationship of plants (Yang et al., 2020). The size of chloroplast genome in terrestrial plants is 120-160 kb, encoding 110-130 unique genes. Because of the slow evolutionary rate of change, maternal inheritance, less recombination and satisfactory collinearity between the sequences of various plant groups, the chloroplast genome sequences were suitable for molecular markers (Zhai et al., 2021). With the development of Next-generation sequencing technology, a large number of chloroplast genome data can be easily obtained. Based on chloroplast comparative genomics analyses, the high variation regions were located to develop specific molecular markers of groups or species for applying to the research of phylogenetic analysis and species identification (Chen et al., 2022; Song et al., 2022).

So far, there was no scientific research related to the complete chloroplast genome of Paraboea. In this study, we sequenced, assembled and analyzed the chloroplast genomes of twelve Paraboea species, and constructed the phylogenetic relationship of 31 species belonging to 12 genera of Gesneriaceae based on protein coding sequences. We also calculated selective pressures to investigate whether the coding protein genes in Paraboea species were under purifying selection or positive selection. Comprehensive insights into the character and evolution of the chloroplast genomes, provided a theoretical basis for the protection and rational utilization of germplasm resources of Paraboea plants in karst areas.

Materials and methods

Plant materials and DNA extraction

The 12 species of Paraboea in China and Vietnam used in the study were identified, collected and finally cultivated in the Guangxi Institute of Botany (Table 1). Fresh green leaves were sampled, washed, dried and stored at -80°C till DNA extraction (Feng et al., 2020). The total genomic DNA was extracted according to the modified CTAB method (Doyle and Doyle, 1987).
Genome sequencing and assembling

Qualified DNA fragments were obtained by mechanical fracture method, and were sequenced after purification, terminal repair and other processing. The 350 bp fragment was screened by agarose gel electrophoresis and amplified by PCR to construct a sequence library. Paired-end (PE) reads were obtained using the Illumina HiSeq 2000 sequencer (Illumina Biotechnology Company, San Diego, CA, USA) (Gu et al., 2018).

De novo genome assembly from the clean data was accomplished utilizing NOVOPlasty v2.7.2 (Dierckxsens et al., 2017), with a k-mer length of 39 bp and the chloroplast genome of Primulina huaijiensis (NC_036413) as the reference sequence. The correctness of the assembly was confirmed by manually editing and mapping all the raw reads to the assembled genome sequence using Bowtie2 (v2.0.1) (Langmead et al., 2009) under the default settings. Finally, the complete chloroplast genome sequences of twelve Paraboea species were obtained.

Repeat sequences and SSR analysis

The Perl script MISA (http://pgrc.ipk-gatersleben.de/misa/) (Beier et al., 2017) was used with the filter thresholds set to detect SSRs. The specific parameters were set at repeat units $\geq 8$ for mononucleotides, repeat units $\geq 4$ for dinucleotides and trinucleotides, and repeat units $\geq 3$ for tetranucleotides, pentanucleotides and hexanucleotides. To identify complex repetitive sequences such as forward, reverse, complement and palindromic, REPuter online software (Kurtz et al., 2001) was used with a minimum repeat size of 30 bp and 90% sequence identity (Hamming distance of 3).

Boundary regions and genome comparative analysis

In order to better display the expansion/contraction events of the IR region, the connecting regions of IR-LSC and IR-SSC in the chloroplast genomes of twelve Paraboea species were compared by using IRscope (https://irscope.shinyapps.io/irapp/) online software (Amiryousefi et al., 2018). To identify interspecific variations, the mVISTA online software was used to compare the chloroplast genomics of twelve Paraboea plants (Frazer et al., 2004). The comparative analysis was carried out by using the shuffle-LAGAN mode in mVISTA with the annotation of P. sinensis as reference, and the sequence alignment was visualized in an mVISTA plot. We used MEGA v6.0 (Tamura et al., 2013) to calculate the percentage of variable sites in the protein-coding genes. We also used DnaSP v6.0 (Rozas et al., 2017) to calculate the nucleotide polymorphism (Pi) among the twelve Paraboea species. When calculating the Pi value, set the windows length to 100 sites and the step size to 25 sites.

Genome annotation and sequence characterization

Functional annotation of the chloroplast genome includes coding gene prediction and non-coding RNA (rRNA and tRNA) annotation. Using CPGAVAS2 (Shi et al., 2019), the twelve complete chloroplast genomes were annotated with a reference genome (Primulina huaijiensis, GenBank: NC036413). Meanwhile, tRNA scan-SE version 1.21 (Schattner et al., 2005) was used to identify and confirm tRNA genes. The twelve circular chloroplast genome maps were constructed using the OrganellarGenomeDRAW (OGDRAW) v.1.3.1 tool followed by manual modification (Greiner et al., 2019). And the whole twelve sequences were submitted to GenBank (Table 1).

### TABLE 1 Sources of material from twelve Paraboea species.

| Taxon             | Voucher  | GenBank accession number | Location                  | Habitat                          |
|-------------------|----------|--------------------------|---------------------------|----------------------------------|
| P. clavipespula   | ZBPC202100061 | MZ465381                | Jingxi Guangxi, China     | Limestone; ca. 800 m             |
| P. dictyoneura    | ZBPD202100062 | MZ465383                | Yingde Guangdong, China   | Rocks in forests; 100-800 m      |
| P. dolomitica     | ZBPD202100063 | MZ465376                | Shihing Guizhou, China    | rock faces of dolomite karst area, ca. 650-855 m |
| P. filipes        | ZBPF202100064 | MZ465379                | Lianzhou Guangdong, China | Limestone cliffs; ca.100-300 m    |
| P. glutinosus     | ZBPG202100065 | MZ465382                | Caobang, Vietnam          | Rocks of slopes; ca. 400-1400 m   |
| P. guilinensis    | ZBPG202100066 | MZ465377                | Guilin Guangxi, China     | Limestone cliffs                  |
| P. martini        | ZBPM202100067 | MZ465385                | Napo Guangxi, China       | limestone under the hillside forest; ca. 1220-1260 m |
| P. peltifolia     | ZBPP202100068 | MZ465386                | Mashan Guangxi, China     | Limestone; ca. 300-400 m          |
| P. rufescens      | ZBPR202100069 | MZ465384                | Napo Guangxi, China       | On rocks of limestone hills and valley forests; ca. 200-1500 m |
| P. sinensis       | ZBPS202100070 | MZ465380                | Longzhou Guangxi, China   | Crevices of rocks or on cliffs in forests; ca. 600-2500 m |
| P. swinhoei       | ZBPS202100071 | MZ465378                | Rongshui Guangxi, China   | Shady and damp rocks under forests; ca. 300-1000 m |
| P. wenshanensis   | ZBPW202100072 | MZ465375                | Wenshan Yunnan, China     | moist shady cliffs of limestone hills, ca. 1500 m |
Phylogenetic analysis

The complete chloroplast protein-coding genes of 31 Genieriaceae species (12 Paraboae species in this study and 20 other species from NCBI) were aligned using MUSCLE v3.8.31 (Edgar, 2004), and then aligned in MAFFT (version 7.222) using the default parameters (Kazutaka and Standley, 2013). The final two sequence alignment results are consistent. The aligned sequences were used to construct the phylogenetic trees using the maximum likelihood (ML) method implemented in RAxML 7.0.4 (Stamatakis, 2006) with 1000 replicates under the GTR + CAT model.

Adaptive evolution analysis

In order to detect the positive selection of chloroplast genes in Paraboae, the non-synonymous (DN) and synonymous (DS) substitution rates of protein-coding genes and the DN/DS (ω) values of protein-coding genes were calculated. All of the CDS sequences were extracted from chloroplast genome, and then the single-copy CDS sequences common to all species were selected and aligned with the codon model. We used EasyCodeML v1.21 (Gao et al., 2019b) to identify positive selection sites. A total of 76 CDSs presented in all the analysed species, and were used for identification of positive selection using the site model (seqtype = 1, model = 0, NSsites = 0, 1, 2, 3, 7, 8). In addition, Bayes Empirical Bayes (BEB) method (Huelsenbeck and Ronquist, 2001) was used to calculate the posterior probabilities for amino acid sites that were potentially under positive selection. The results showed that the amino acid sites with posterior probability of more than 0.95 were positively selected. Moreover, the logarithmic likelihood value was positive selected. The results showed that the amino acid sites with posterior probability of more than 0.95 were potentially under positive selection. The logarithmic likelihood value of protein-coding genes was calculated. All of the CDS and IR length ranged from 84761 bp (P. clavisepala) to 85488 bp (P. wenshanensis), and from 25272 bp (P. dolomitica) to 25334 bp (P. wenshanensis). In all twelve Paraboae species, the chloroplast genomes of P. filipes and P. wenshanensis had the lowest total GC content (37.45%), while the chloroplast genome of P. martini had the highest total GC content (37.72%). Gene annotation showed that each chloroplast genome contained 131 genes in conserved order and orientation, which contained 8 ribosomal RNA (rRNA) genes, 36 transfer RNAs (tRNAs), and 87 protein-coding genes (Table 3). Fifteen genes (10 protein coding genes and 5 tRNA genes) with introns were identified. Among them, the clpP and ycf3 genes had two introns, respectively, while the other 13 genes had one intron.

IR expansion and contraction in the twelve Paraboae chloroplast genomes

There were 4 borders between LSC, IRb, IRA and SSC in the cpGenome: LSC/IRb border (JLB line), IRb/SSC border (JSB line), SSC/IRA border (JSA line), IRA/LSC border (JLA line). The borders of the twelve Paraboae chloroplast genomes were compared (Figure 2). The LSC/IRb border and IRA/LSC border were relatively conservative. The rpl2 gene located at the LSC/IRb border, and the distances between rps2 and the JLB line ranged from 41 bp to 95 bp. The trnH-GUG noncoding gene located on the right side of the JLA line with a distance of 0 to 9 bp. At the IRb/SSC border, the ndhF encoding gene located at the IRB-SSC boundary. In the chloroplast genome of P. sinensis, P. swinhoei, P. peltoph, P. filipes and P. guilinensis, the ndhF gene had the length of 72 bp (P. guilinensis) to 138 bp (P. sinensis) in the IRB region. In the other seven Paraboae chloroplast genomes, the ndhF gene spanned the IRb/SSC border and had the length of 113 bp to 124 bp. At the SSC/IRA border, the ycf1 gene spanned the SSC-IRB boundary. Due to the special position of the ycf1 gene, there were seven Paraboae chloroplast genomes in the IRA region with ycf1 pseudogenes, the corresponding length ranged from 842 bp to 863 bp in the IRA region. And in the other Paraboae species chloroplast genomes, the ycf1 gene was located on the SSC-IRA boundary, which made the corresponding pseudogene take place in the IRB region with the length of 799 bp to 833 bp.

Repeat sequence analysis

The twelve Paraboae chloroplast genomes contained 600 SSRs (Figure 3A and Supplementary Table S2). In the chloroplast genome of P. rafescens, 41 SSRs were detected, which was the least of the 12 chloroplast genomes. And in the chloroplast genome of P. sinensis, a total of 57 SSRs were identified, which was the most of the 12 chloroplast genomes. For each Paraboae specie, mononucleotide repeats were the
most common, with numbers ranging from 19 to 34; followed by tetranucleotides ranging from 10 to 16; dinucleotides ranging from 6 to 14; trinucleotides ranging from 1 to 4; pentanucleotides ranging from 0 to 2 and hexanucleotide ranging from 0 to 2 (Supplementary Figure S1).

Non-overlapped sequence repeats including forward repeats, reverse repeats, palindromic repeats and complement repeats were detected in twelve chloroplast genomes. A total of 389 non-overlapped sequence repeats were detected in twelve chloroplast genomes of _Paraboea_ plants (Figure 3B and Supplementary Table 2).

**TABLE 2 Summary of the chloroplast genomes of twelve Paraboea species.**

| Species          | Genome Length (bp) | LSC Length (bp) | SSC Length (bp) | IR Length (bp) | GC (%) | Total Genes | CDS | tRNA | rRNA |
|------------------|--------------------|-----------------|-----------------|---------------|--------|-------------|-----|------|------|
| _P. clavisepala_ | 153398             | 84761           | 18045           | 25296         | 37.58% | 131         | 87  | 36   | 8    |
| _P. dictyoneura_ | 153406             | 84829           | 17999           | 25289         | 37.52% | 131         | 87  | 36   | 8    |
| _P. dolomitica_  | 153510             | 84885           | 18081           | 25272         | 37.45% | 131         | 87  | 36   | 8    |
| _P. filipes_     | 153486             | 84851           | 18001           | 25317         | 37.45% | 131         | 87  | 36   | 8    |
| _P. glutinosa_   | 153505             | 85303           | 17656           | 25273         | 37.66% | 131         | 87  | 36   | 8    |
| _P. guilinensis_ | 153166             | 84819           | 17759           | 25294         | 37.57% | 131         | 87  | 36   | 8    |
| _P. martinii_    | 153580             | 84978           | 18032           | 25285         | 37.72% | 131         | 87  | 36   | 8    |
| _P. petrifolia_  | 153459             | 84784           | 18043           | 25316         | 37.55% | 131         | 87  | 36   | 8    |
| _P. rufescens_   | 153352             | 85098           | 17694           | 25280         | 37.69% | 131         | 87  | 36   | 8    |
| _P. sinensis_    | 153453             | 84669           | 18028           | 25278         | 37.61% | 131         | 87  | 36   | 8    |
| _P. swinhoei_    | 153564             | 85160           | 17800           | 25302         | 37.65% | 131         | 87  | 36   | 8    |
| _P. wenshanensis_| 154245             | 85488           | 18089           | 25334         | 37.70% | 131         | 87  | 36   | 8    |

LSC, large single-copy; SSC, small single-copy; IR, inverted repeat.
Table S3). The number of non-overlapped sequence repeats varied from 28 in *P. dictyoneura* to 37 in *P. sinensis*. Among these non-overlapped repeats, palindromic repeats were the most common with 207, followed by forward repeats with 160; reverse repeats with 15 and complement repeats with 7 (Supplementary Figure S2). The repeat sequence analysis would provide help for the study of genetic variation in *Paraboea*.

**Comparative chloroplast genome analysis**

Taking *P. sinensis* as a reference, multiple alignments of twelve *Paraboea* chloroplast genomes were conducted, and the results suggested that the non-coding sequences showed more divergence than the coding regions (Figure 4). According to the comparative analysis, the main divergent sequences for the noncoding regions were *atpH-atpI, atpF-atpH, rps16-trnQ-UUG, trnK-UUU-trnS*, *trnH-GUG-psbA, trnS-GCU-trnR-UCU* and *psaA-ycf3*, and the strongly divergent sequences for the coding regions were *matK, petL, ycf1, ycf2 and ndhF*, which might be good candidates for *Paraboea* species identification. To quantify the levels of DNA polymorphism, we calculated the Pi values of above twelve regions, the Pi values of these regions were calculated ranged from 0.01569 (*psaA-ycf3*) to 0.08362 (*trnH-GUG-psbA*) (Figure 5A). The highest average Pi value of the coding regions was calculated in the SSC region, followed by the coding regions of the LSC and IR region (Figures 5B–D). The ten coding genes with the highest polymorphism in descending order include: *ycf1, rps15, petL, matK, rpl22, ndhF, rps3, rpoB, psaI* and *ccsA*. The Pi values of tRNA and rRNA genes were also

**TABLE 3** Genes in the chloroplast genome of twelve *Paraboea* species.

| Category | Gene group | Gene name |
|----------|------------|-----------|
| Protein synthesis and DNA-replication | Ribosomal RNA genes | *trnA-UGC, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnH-GUG, trnI-CAU, trnL-GAU*, trnL-CAA, trnL-UAA*, trnL-UAG, trnM-CAU, trnN-GUU, trnP-UUG, trnQ-UUG, trnR-UCU, trnR-ACG, trnS-GCU, trnS-GGA, trnT-GUA, trnV-GAC, trnV-UGA, trnW-CCA, trnY-GUA* |
| Ribosomal protein genes (larger subunit) | *trpl2*, *trpl4*, *trpl6*, *trpl20*, *trpl22*, *trpl32*, *trpl33*, *trpl36* |
| Ribosomal protein genes (smaller subunit) | *trp2*, *trp3*, *trp4*, *trp7*, *trp8*, *trpl4*, *trpl5*, *trpl6*, *trpl8* |
| Photosynthesis | RNA polymerase | *psaA, psaB, psaC, psaD, psaE*
| Photosystem I | *psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbJ, psbL, psbM, psbN, psbT* |
| Photosystem II | *petA, petB, petD, petG, petL, petN* |
| Cytoskeleton b/f complex | *atpA, atpB, atpF*, *atpH, atpI* |
| ATP synthase | *rbCL* |
| Rubisco large subunit | *ndhA*, *ndhA*, *ndhB*, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK* |
| Miscellaneus group | ATP-dependent protease | *clpP*** |
| Maturase | *matK* |
| Acetyl-CoA carboxylase | *accD* |
| Cytochrome c biogenesis | *ccsA* |
| Inner membrane protein | *cemA* |
| Pseudogene unknown function | Hypothetical chloroplast reading frames (ycf) | *ycf1*, *ycf2, ycf3***, *ycf4, ycf15* |
| Other Gene | *LhBA* |

*** indicates the presence of one intron.
** indicates the presence of two introns.
calculated, and the results showed that trnC-GCA and trnM-CAU had high Pi values, 0.0676 and 0.068, respectively. We finally screened out ten divergent regions with the highest value, which were trnH-GUG-psbA, trnM-CAU, trnC-GCA, atpF-atpH, ycf1, trnK-UUU-rps16, rps15, petL, trnS-GCU-trnR-UCU and psaJ-rpl33. These divergent regions may be the best candidate marker for DNA barcoding.

**Phylogenetic relationship**

In order to study the phylogenetic position of *Paraboea*, ML tree were constructed using 76 protein coding genes of the chloroplast genomes for 31 Gesneriaceae species, including 12 *Paraboea* species (Figure 6). Among the 31 Gesneriaceae species, except for 12 *Paraboea* species, the chloroplast genomes of the remaining species were obtained from NCBI (Supplementary Table S1). In the phylogenetic tree, all nodes were supported with bootstrap values greater than 60%, and each genus clustered together into a clade (100% bootstrap values). The 12 *Paraboea* species clustered into a clade, and then clustered with *Dorcoceras* hygrometrica (100% bootstrap values). *Paraboea* clade were divided into two major small clades with 100% bootstrap support value. In one major small clade, *P. clavisepala* and *P. dolomitica* form a clade, and then sequentially formed clades with *P. peltifolia*, *P. dictyoneura*, *P. guilinensis* and *P. filipes*. In another major small clade, the clade formed by *P. sinensis* and *P. wenshanensis*, clustered with *P. rufescens* and *P. glutinosa*, and then shared a sister relationship with *P. martii* and *P. swinhoei*.

**Adaptive evolution analysis**

The 76 chloroplast protein coding genes of twelve *Paraboea* species were tested, and positive selection was found in nine genes (*lhbA*, *matK*, *ndhF*, *psbK*, *rbcL*, rpl22, *rps12*, *rps18* and *ycf1*) with a high posterior probability (>95%) using the BEB test (Figure 7 and Supplementary Table S4). One amino acid site (the 39th codon) was identified to be under positive selection in *lhbA* gene (Figure 7A). The spatial analysis of LhbA protein under positive selection indicated that the site was located in the α-helix (Figure 8A). Four amino acid sites (the 81th, 116th, 284th and 353th codons) under positive selection were detected in Maturase coded by *matK* gene (Figure 7B). The spatial analysis indicated that two sites were located in α-helix, and the other sites were located at β-sheet and random coil, respectively (Figure 8B). In addition, three of nine genes were coding genes for photosynthesis: the *ndhF* gene for NADH dehydrogenase subunit F (NDHF), the *psbK* gene for Photosystem II subunits K (PsbK), and the *rbcL* gene for rubisco large subunit (RBCL). Three amino acid sites (463th, 651th and 729th) under positive selection in NDHF were located in the random coil, α-helix and α-helix respectively (Figure 7C and Supplementary Figure S3). Based on the protein structure prediction, one amino acid site...
(34th) under positive selection in PsbK was located in α-helix (Figure 7D and Figure 8C). Three amino acid sites (464th, 470th and 479th) under positive selection in RBCL were located in the random coil (Figure 7E and Figure 8D).

Meanwhile, other three genes were coding genes for protein synthesis: rps12 and rps18 genes for Ribosomal protein smaller subunit (RPS), and rpl22 gene for Ribosomal protein larger subunit 22 (RPL22). One positive selection site was identified in RPS12 and RPS18 protein, respectively (Figures 7G, H; Figures 8F, G). Four positive selection sites were identified in RPL22 (Figure 7F and Figure 8E). Finally, seven sites were identified in YCF1 (Hypothetical chloroplast reading frame 1) coded by ycf1 gene (Figure 7I and Supplementary Figure S4). Based on the protein structure prediction, most of these positive selection sites were located in the α-helix, followed by random coil and β-sheet (Figure 8).

Discussion

Chloroplast genome features

In this study, the chloroplast genomes of twelve Paraboea species were characterized (Figure 1; Table 2). The twelve
chloroplast genomes also showed a highly conserved feature in terms of structures, gene orders, gene numbers (protein-coding genes, rRNAs and tRNAs) and intron number. The chloroplast genomes of twelve *Paraboea* plants ranged from 153166 to 154245 bp in length. The chloroplast genomes of angiosperms have a highly conserved feature, but the contraction and expansion of the boundary between the IR and SC region is considered to be the main reason for the size change of the chloroplast genome (Zhang et al., 2016). The same phenomenon also existed in the twelve *Paraboea* chloroplast genomes. Despite the twelve *Paraboea* chloroplast genomes having well-conserved genomic structures including gene number and order, length variation of the whole sequences comprising IR, LSC and SSC regions was detected among these chloroplast genomes (Figure 2). In particular, *ycf1* and *ndhF* genes located at the SSC/IR border had the greatest variation in position and length in the twelve *Paraboea* chloroplast genomes. These sequence variations might be the result of boundary contraction and expansion between the SSC/IR regions in plants (Wang and Messing, 2011).

**Repeat sequence analysis**

SSRs have been used as molecular markers for determining a high degree of variation in similar species and are helpful to explore population genetics and polymorphisms (Zhao et al., 2015). In total, 600 SSRs were detected in the twelve chloroplast genomes, 315 of which were mononucleotide repeats, accounting for the majority of all SSRs (52.50%) (Figure 3A and Supplementary Figure S1). Among the twelve chloroplast genomes, the number of mononucleotide repeats was the largest.
FIGURE 5
The nucleotide variability (Pi) value in the 12 Paraboea chloroplast genomes. (A) The Pi value of divergent sequences. (B) The Pi value of LSC region. (C) The Pi value of SSC region. (D) The Pi value of IR region.

FIGURE 6
The ML phylogenetic tree of 31 Gesneriaceae species. Supporting values of > 50% for ML were shown on the branch.
In angiosperm chloroplast genomes, lots of similar results were also reported previously (Gandhi et al., 2010; Bessega et al., 2013). The results also demonstrated that the SSRs identified in the chloroplast genome were mostly made up of polyadenine (Poly-A) or polythymine (Poly-T) repeats, and the contents of guanine (Poly-G) and cytosine (Poly-C) repeats were low, which was consistent with the general SSR characteristics of chloroplast genomes in angiosperms (Ebert and Peakall, 2009; Asaf et al., 2020).

Moreover, 389 non-overlapped sequence repeats were identified in twelve chloroplast genomes (Figure 3B and Supplementary Figure S2), including the most non-overlapped sequence repeats (37) in *P. sinensis* and the least non-overlapped sequence repeats (28) in *P. dictyoneura*. Among the 389 repeats,

**FIGURE 7**
Nine genes of positive selection of amino acid sequences in site model tests.

**FIGURE 8**
Spatial location of the positively selected sites in proteins of *P. clavisepala*. (A) Spatial location of the positively selected sites in the *lbaA* protein of *P. clavisepala*. *A*. *B* Spatial location of the positively selected sites in the *matK* protein of *P. clavisepala*. *A*. *C* Spatial location of the positively selected sites in the *psbK* protein of *P. clavisepala*. *A*. *D* Spatial location of the positively selected sites in the *rbcL* protein of *P. clavisepala*. *A*. *E* Spatial location of the positively selected sites in the *rpl22* protein of *P. clavisepala*. *A*. *F* Spatial location of the positively selected sites in the *rps12* protein of *P. clavisepala*. *A*. *G* Spatial location of the positively selected sites in the *rps18* protein of *P. clavisepala*.
there were four types: forward repetition, reverse repetition, complement repetition and palindromic repetition. Palindromic repetition and forward repetition accounted for the highest proportion. The same conclusion was obtained in the analysis of repetitive sequences in the chloroplast genomes of other Gesneriaceae plants (Gu et al., 2020). All of these SSRs, together with non-overlapped sequence repeats, are useful sources to develop markers for genetic diversity analysis of Paraboea species (Supplementary Table S2 and Supplementary Table S3).

Phylogenetic relationship

Most of the studies on the molecular phylogeny of Paraboea were based on chloroplast trnL-F sequences and nuclear ITS sequences (Puglisi et al., 2011; Xin et al., 2019; Guo et al., 2020). Strict consensus tree based on combined ITS and trnL-trnF sequences of 53 samples showed that Paraboea samples formed three major clades (Puglisi et al., 2011). The major clade 1 contains Chinese and Thai Paraboea species, some small clades of which were with low or no branch support. The existing two chloroplast sequences didn’t completely solve the phylogenetic relationship of P. sinensis, P. rufescens, P. glanduliflora and P. swinhoei.

In recent years, phylogeny based on the complete chloroplast genome has been widely used in plants (Feng et al., 2017; Kyalo et al., 2020; Tian and Wariss, 2021). 31 species belonging to 12 genera of Gesneriaceae were used to construct the ML tree in this study. All nodes were supported with bootstrap values greater than 60%, and the 12 Paraboea species clustered into a clade with 100% bootstrap values (Figure 6), supporting the monophyly of Paraboea. The topology of the phylogenetic tree was more resolved than found in combined ITS and trnL-trnF datasets of previous studies (Puglisi et al., 2011; Xin et al., 2019; Guo et al., 2020). The clade formed by P. sinensis and P. wenshanensis, clustered with P. rufescens and P. glutinosa, and then shared a sister relationship with P. martinii and P. swinhoei. The sequences of chloroplast genome sequences could completely solve the phylogenetic relationship of Paraboea, and chloroplast genome data could provide more genetic information on the evolutionary relationships and phylogeny among species of Gesneriaceae.

Adaptive evolution analysis

The plants of Paraboea are mainly distributed in karst areas of south and southwest China and southeast Asia. Karst is a unique and fragile ecological environment and the rocks forming karst landforms mainly consist of limestone, dolomite and other soluble carbonate rocks (Li et al., 2019). Because of the thin soil layer, low water holding capacity and strong permeability of rocks, the stress of frequent alternation of dry and wet is common in karst habitats. Facing frequent temporary drought, karst plants generally appear enhanced photosynthetic capacity and light protection mechanisms (Liu et al., 2021). The changes in ground temperature, air temperature, light intensity and atmospheric relative humidity are quite different in different microhabitats (such as rocky hills with direct sunlight, forests with weak light and dark caves, etc.) (Ou et al., 2020). These challenging karst environments may impose selective pressure on genes, which could leave a footprint of natural selection in genes of chloroplast involved in adaptation to the environment. In this study, among the chloroplast genes of twelve Paraboea species, nine genes (hlhA, mATk, ndhF, psbK, rbcL, rpl2, rps12, rps18 and ycf1) were identified under positive selection using a site model (Figure 7, Figure 8 and Supplementary Table S4).

The maturase encoded by matK gene is involved in splicing of introns of trnK, trnL, atpF and other genes, which is important for maintaining the normal function of chloroplast (Moran et al., 1994; Vogel et al., 1999; Lambowitz and Zimmerly, 2004; Stern et al., 2010; Zoschke et al., 2010). There were four positive selection sites in matK gene of Paraboea species, and matK gene also had undergone adaptive evolution in Lycopodiaceae, Bryophyta and other plants (Hao et al., 2010a; Hao et al., 2010b). Adaptive evolution of matK may fine-tune its function to optimize its performance in various environmental conditions.

Three genes under positive selection were related to photosynthesis, namely psbK, ndhF and rbcL gene. The psbK gene encodes Photosystem II subunits K. Photosystem II is the first link in the chain of photosynthesis, and captures photons and uses the energy to extract electrons from water molecules (Ferreira et al., 2004). PSBK is not necessary for the assembly or activity of photosystem II complex, but is essential for optimal photosystem II function. The psbK gene was detected under positive selection in Echinacanthus (Acanthaceae) and Calligonum Mongolicum (Polygonaceae), and speculated to play an important role in plant adaptation evolutionary process to the diverse environment (Gao et al., 2019a; Duan et al., 2020). The ndhF gene encodes NADH dehydrogenase subunit protein (Kubicki et al., 1996). In previous studies on plant adaptive evolution, ndhF genes were often under positive selection pressure (Liu et al., 2020; Li et al., 2021; Wen et al., 2021). The NADH dehydrogenase complex of higher plants not only participated in photosynthetic electron transport (Joet et al., 2001; Joet, 2002), but also acted as an electron transport carrier for chloroplast respiration (Casano et al., 2000). The adaptive evolution of the ndhF gene may affect energy transformation and resistance to photodissociative stress in different environments. The rbcL gene encoded the gene coding for the rubisco large subunit protein of Rubisco, which was an important part of the photosynthesis electron transport regulator (Piot et al., 2018). The rbcL gene was often under positive selection because of being the target of selection diverse environment factors related to the changes in temperature, drought and carbon dioxide concentration (Fan et al., 2018). NADH-dehydrogenase subunits
and Photosystem subunits were essential in the electron transport chain for the generation of ATP and light energy utilization, which were all indispensable parts for photosynthesis of plants (Yamori and Shikanai, 2016; Peltier et al., 2016). Therefore, the signature of positive selection in three genes related to photosynthesis suggests that they might have been involved in adaptation to diversified environments for Paraboea species in karst habitats.

Meanwhile, positive selection sites were also identified in lhbA, rpl22, rps12 and rps18 genes. The specific function of lhbA gene has not been fully studied (Wu et al., 2020). The rps genes encode small ribosomal subunit proteins, and rpl genes encode large ribosomal subunit proteins (Muto and Ushida, 1995). The mutation of genes encoded in ribosomal proteins under the pressure of the natural environment may affect the translation of chloroplast ribosome (Ramundo et al., 2013).

Seven sites were detected under positive selection in the ycf1 gene. Positive selection of ycf1 was also found to be involved in the adaptation of the genus Panax (Jiang et al., 2018). Being one of the largest chloroplast genes, the ycf1 gene encoding a component of the chloroplast’s inner envelope membrane protein translocon, has become a useful gene for assessing sequence variations and evolutionary processes in plants (Huang et al., 2010; Kikuchi et al., 2013). The function and the adaptive evolutionary analysis of the ycf1 gene would better understand the evolutionary mechanism of plants in the future.

Because of environmental pressure, adaptive evolution of chloroplast genomes is a common phenomenon, especially for genes involved in photosynthesis. Genes associated with photosynthesis are more likely to evolve adaptively in plants distributed in extreme environments, such as shade plants or aquatic plants (Xie et al., 2018). In karst areas, there are great differences in environmental factors such as light intensity, soil water content and nutrient availability, which might have exerted strong selective forces on plant evolution (Ai et al., 2015). In this study, nine chloroplast genes under positive selection, most of which were related to photosynthesis and protein synthesis, may possibly contribute to the diverse evolution and adaptation of Paraboea species to karst extreme environments.

Conclusion

This is the first report of the complete chloroplast genome sequence of Paraboea species. In this study, the newly sequenced chloroplast genomes of twelve Paraboea species were reported and compared. The genome annotation and comparative analysis showed that each chloroplast genome was a typical quadruplicate structure like traditional angiosperms, and the GC content, gene number and order were similar to each other. The chloroplast genomes of the twelve Paraboea species were similar in structure, composition and gene order. In the twelve Paraboea chloroplast genomes, a total of 600 SSRs and 389 non-overlapped sequence repeats were identified, which were informative sources for developing markers for genetic diversity analysis of Paraboea species. In addition, we found that 10 different regions (trnH-GUG-psbA, trnM-CAU, trnC-GCA, atpF-atpH, ycf1, trnK-UUU-rps16, rps15, petL, trnSGCU-trnR-UCU and psaJ-rpl33) were potential molecular markers in twelve Paraboea species. The phylogenetic tree based on 76 protein coding genes clearly demonstrated the genetic and evolutionary relationships of 31 species belonging to 12 genera of Gesneriaceae. Adaptive evolution analysis detected positive selection signals in nine chloroplast genes (i.e., lhbA, matK, ndhF, psbK, rbcL, rpl22, rps12, rps18 and ycf1). The evolution of Paraboea to adapt to extreme habitats in karst environments may be linked to changes in these positive selection sites. These analyses of chloroplast genomes will provide preparations for the development and utilization of Paraboea species germplasm resources and the formulation of conservation strategies.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

BZ, FW, YW, YM and XH conceived and designed the study. ZL, YW collected and analyzed the data. BZ, FW, YW, ZL and XH wrote the manuscript. All authors have directly contributed to this manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study is supported by the Key Sci. & Tech. Research and Development Project of Guangxi (Guike AD20159091 &
Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1019831/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The comparison of each type of SSRs among 12 Paraboea chloroplast genomes.

SUPPLEMENTARY FIGURE 2
Number of the four complex repeat types (forward, palindrome, reverse, and complement) in the twelve Paraboea chloroplast.

SUPPLEMENTARY FIGURE 3
Protein secondary structure of ndhF.

SUPPLEMENTARY FIGURE 4
Protein secondary structure of ycf1.

SUPPLEMENTARY TABLE 1
Chloroplast genome sequences from GenBank used in this study.

SUPPLEMENTARY TABLE 2
Summary of SSRs in twelve Paraboea chloroplast genomes.

SUPPLEMENTARY TABLE 3
Summary of complex repeats in twelve Paraboea chloroplast genomes.

SUPPLEMENTARY TABLE 4
Positive selection sites identified in the chloroplast genomes of twelve Paraboea species.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ai, B., Gao, Y., Zhang, X., Tao, J., Kang, M., and Huang, H. (2015). Comparative transcriptome resources of eleven Primulina species, a group of ‘stone plants’ from a biodiversity hot spot. Mol. Ecol. Resour. 15, 619–632. doi: 10.1111/1755-0998.12333

Amirouysefi, A., Hyvonen, J., and Poczaï, P. (2018). iRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34, 3030–3033. doi: 10.1093/bioinformatics/bty220

Asaf, S., Jan, R., Khan, A. L., and Lee, I. J. (2020). Complete chloroplast genome characterization of Ocimum corniculatum and its comparison with related species from family oxalidaceae. Plants 9, 928. doi: 10.3390/plants9080928

Beter, S., Thiel, T., Munch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. Bioinformatics 33 (16), 2583–2585. doi: 10.1093/bioinformatics/btx198

Bessenga, C. F., Pometti, C. L., Miller, J. T., Watts, R., Saidman, B. O., and Vilardi, J. C. (2013). New microsatellite loci for Prosopis alba and P. chilensis (Fabaceae). Appl. Plant Sci. 1 (5), 1200324. doi: 10.3732/apps.1200324

Buchan, W. A., Minneci, F., Nugent, C. O., Bryson, K., and Jones, D. T. (2013). Scalable web services for the PISPRED protein analysis workbench. Nucleic Acids Res. 41, W349–W357. doi: 10.1093/nar/gkt381

Casano, L. M., Zapata, J. M., Martin, M., and Sabater, B. (2000). Chlororespiration and poising of cyclic electron transport. plastoquinone as electron transporter between thylakoid NADH dehydrogenase and peroxidase. J. Biol. Chem. 275 (2), 942–948. doi: 10.1074/jbc.275.2.942

Chen, H. M., Chen, Z. E., Du, Q., Jiang, M., Wang, B., and Liu, C. (2022). Complete chloroplast genome of Calligonum mongolicum (Thunb.) schum and Primula ora (Oxalidaceae). Mol. Biol. Rep. 49 (4), 3085–3098. doi: 10.1007/s11033-022-07139-0

Chen, J., Zang, Y., Shang, S., Liang, S., Zhu, M., Wang, Y., et al. (2021). Comparative chloroplast genomes of Zosteraceae species provide adaptive evolution insights into seagrass. Front. Plant Sci. 12. doi: 10.3389/fpls.2021.741152

Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45 (4), e18. doi: 10.1093/nar/gkw935

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11–15.

Duan, H. R., Zhang, Q., Yang, H., Tian, F. P., Hu, Y., Wang, C. M., et al. (2020). Complete chloroplast genome of Calligonum mongolicum: Genome organization, codon usage pattern, phylogenetic relationships, comparative structure and adaptive evolution analysis. Research Square. doi: 10.21203/rs.3.rs-49271/v1

Ebert, D., and Peakall, R. (2009). Chloroplast simple sequence repeats (cpSSRs): Technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. Mol. Ecol. Res. 9, 673–690. doi: 10.1111/j.1755-0998.2008.02319.x

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32 (5), 1792–1797. doi: 10.1093/nar/gkh340

Fan, W. B., Wu, Y., Yang, J., Shahzad, K., and Li, Z. H. (2018). Comparative chloroplast genomics of Dipacales species: Insights into sequence variation, adaptive evolution and phylogenetic relationships. Front. Plant Sci. 9. doi: 10.3389/fpls.2018.00689

Feng, C., Xu, M., Feng, C., and Kang, M. (2017). The complete chloroplast genome of Primulina and two novel strategies for development of high polymorphic loci for population genetic and phylogenetic studies. BMC Evol. Biol. 17 (1). doi: 10.1186/s12862-017-1130-0

Feng, S. G., Zheng, K. X., Jiao, L. K., Cai, Y. C., Chen, C. L., Mao, Y. Y., et al. (2020). Complete chloroplast genomes of four Phyllis species (Solanaeae): lights into genome structure, comparative analysis, and phylogenetic relationships. BMC Plant Biol. 20 (1), 242. doi: 10.1186/s12870-020-02429-w

Ferrera, K. N., Iverson, T. M., Magliauoa, K., Barber, J., and Iwata, S. (2004). Architecture of the photosynthetic oxygen-evolving center. Science 303 (5665), 1831–1838. doi: 10.1126/science.1093087

ZY21195050), the Guangxi Natural Science Foundation (2020GXNSFBA297049), the capacity-building project of SBR of CAS (KJF-BRP-017-68), the Anhui Provincial Natural Science Foundation (1908085QC1), the Fund of Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Populations (PSESP2021F007).

Impact of interest

No potential conflict of interest was reported by the authors.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
phylogenomics: Analysis of membrane.

plastoquinone oxidoreductase is involved in cyclic electron transport.

plastome-encoded ndh genes in mesophyll and bundle-sheath chloroplasts of the phylogenetic analysis. of the vulnerable doi: 10.3897/phytokeys.153.50933

Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale.

Sensitivity of photosynthesis to antimycin induced by inactivation of the phylogeny of the angiosperm phylogeny of the angiosperm plastid genomes of four Pila (Urticaceae) species: insight into interspecific plastid genome diversity in pilea. BMC Plant Biol. 21 (1), 25. doi: 10.1186/s12870-020-02793-7

Liu, C., Huang, Y., Wu, F., Liu, W., Ning, Y., Huang, Z., et al. (2021). Plant adaptability in karst regions. J. Plant Res. 134 (5), 889–906. doi: 10.1002/jphr.20213-01530-3

Liu, Q., Li, X., Li, M., Xu, W., and Heslop-Harrison, J. S. (2020). Comparative chloroplast genome analyses of Avena: insights into evolutionary dynamics and phylogeny. BMC Plant Biol. 20 (1), 406. doi: 10.1186/s12870-020-02621-y

Liu, J. M., and Wang, Y. Z. (2007). Phylogenetic reconstruction among species of Chrysopsis and Chirita sect. gibbosus (Gesneriaceae) based on nrDNA ITS and cpDNA trnl-f sequences. Syst. Botany 32 (4), 888–899. doi: 10.1660/0366440733890764

Moran, J. V., Mecklenburg, K. L., Sass, P., Belcher, S. M., Mahnke, D., Lewin, A., et al. (1994). Splicing defective mutants of the COXI gene of yeast mitochondrial DNA: initial definition of the maturase domain of the group II intron AI2. Nucleic Acids Res. 22 (11), 2057–2064. doi: 10.1093/nar/22.11.2057

Muto, A., and Ushida, C. (1995). Transcription and translation. Methods Cell Biol. 48, 483. doi: 10.1016/s0076-6875(09)80009-7

Ou, Z., Pang, S., He, Q., Peng, Y., Huang, X., and Shen, W. (2020). Effects of vegetation restoration and environmental factors on understory vascular plants in a typical karst ecosystem in southern China. Sci. Rep. 10 (1), 1–10. doi: 10.1038/s41598-020-68785-7

Pelletier, G., Aro, E. M., and Shikani, T. (2016). NDH-1 and NDH-2 plastozoonene reductases in oxygenic photosynthesis. Ann. Rev. Plant Biol. 67, 55–80. doi: 10.1146/annurev-arplant-043014-114752

Piot, A., Hackel, J., Christin, P. A., and Besnard, G. (2018). One-third of the plastid genes evolved under positive selection in PACMAD grasses. Planta 247, 255–266. doi: 10.1007/s11240-017-2781-x

Puglisi, C., Middleton, D. J., Triboun, P., and Möller, M. (2011). New insights into the relationships between Paraboea, Trisepalum and Phyillebece (Gesneriaceae) and their taxonomic consequences. Taxon 60 (6), 1693–1702. doi: 10.1270/tax.600024

Puglisi, C., and Phuthai, T. (2017). A new species of Paraboea (Gesneriaceae) from Thailand. Edinburgh J. Botany 73 (3), 227–238. doi: 10.1111/1365-3090.12582

Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis software for complex population genomics. BMC Bioinformatics 18, 1–12. doi: 10.1186/s12859-017-1423-8

Rheinheimer, P., and Wang, Y. (2018). Effects of environmental and interspecific factors on the growth and development of Pila bicolor (Urticaceae) in a karst region. J. Syst. Evol. 56 (1), 1–12. doi: 10.7555/JSE.2018.1702.

Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., et al. (2019). Distinct plastid genome diversity in pilea. BMC Plant Biol. 21 (1), 25. doi: 10.1186/s12870-020-02793-7

Stern, D. B., Goldschmidt-Clermont, M., and Hanson, M. R. (2010). Chloroplast genome polymorphism in large data sets. Mol. Biol. Evol. 34 (12), 3299–3302. doi: 10.1093/molbev/mst248

Schattner, P., Brooks, A. N., and Lowe, T. M. (2005). The sRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snRNAs. Nucleic Acids Res. 33, W868–W869. doi: 10.1093/nar/gki366

Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., et al. (2019). CGAVAS, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 47 (W1), W65–W73. doi: 10.1093/nar/gkz345

Song, Y., Zhao, W. J., Xu, J., Li, M. F., and Zhang, Y. J. (2022). Chloroplast genome evolution and species identification of Styx (Styriaceae). BioMed. Res. Int. 21, 5364694. doi: 10.1155/2022/5364694

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690. doi: 10.1093/bioinformatics/btl446

Stern, D. B., Goldchmidt-Clermont, M., and Hanson, M. R. (2010). Chloroplast RNA metabolism. Annu. Rev. Plant Biol. 61 (1), 125–155. doi: 10.1146/annurev-arplant-042808-112242

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kuma, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2745–2749. doi: 10.1093/molbev/msx127

Tao, J., Feng, C., Ai, B., and Kang, M. (2016). Adaptive molecular evolution of the two-pore channel 1 gene TP1C in the karst-adapted genus Primulina (Gesneriaceae). Annu. botany 90 (4), 1677–1688. doi: 10.1007/s10265-016-9200-x
endemic to karst areas in China. *Mitochondrial DNA B Resour.* 6 (3), 833–834. doi: 10.1080/23802359.2021.1884021

Vogel, J., Borner, T., and Hess, W. R. (1999). Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. *Nucleic Acids Res.* 27 (19), 3866–3874. doi: 10.1093/nar/27.19.3866

Wang, W., and Messing, J. (2011). High-throughput sequencing of three lemnoideae (duckweeds) chloroplast genomes from. *PloS One* 6, e24670. doi:10.1371/journal.pone.0024670

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303. doi:10.1093/nar/gky427

Wen, F., Hong, X., Chen, L. Y., Zhou, S. B., and Wei, Y. G. (2013). A new species of Paraboea (Gesneriaceae) from a karst limestone hill in southwestern guangdong, China. *Phytotaxa* 131 (1), 1–8. doi:10.11646/phytotaxa.131.1.1

Wen, F., Wu, X., Li, T., Jia, M., Liu, X., and Liao, L. (2021). The complete chloroplast genome of Stauntonia chinesis and compared analysis revealed adaptive evolution of subfamily lardizabaloideae species in China. *BMC Genomics* 22 (1), 161. doi:10.1186/s12864-021-07484-7

Wu, Z. H., Liao, R., Yang, T. G., Dong, X., Lan, D. Q., Qin, R., et al. (2020). Analysis of six chloroplast genomes provides insight into the evolution of Chrysosplenium (Saxifragaceae). *BMC Genomics* 21 (1), 621. doi:10.1186/s12864-020-07045-4

Xie, D. F., Yu, Y., Deng, Y. Q., Li, J., Liu, H. Y., Zhou, S. D., et al. (2018). Comparative analysis of the chloroplast genomes of the Chinese endemic genus Urophylla and their contribution to chloroplast phylogeny and adaptive evolution. *Int. J. Mol. Sci.* 19 (7), 1847. doi:10.3390/ijms19071847

Xin, Z. B., Chou, W. C., Maciejewski, S., Fu, L. F., and Wen, F. (2021). Primulina papillosa (Gesneriaceae), a new species from limestone areas of guangxi, China. *Phytotaxa* 62 (1), 39–43. doi:10.1007/s12228-010-9175-8

Zhao, Y. B., Yin, J. L., Guo, H. Y., Zhang, Y. Y., Xiao, W., Sun, C., et al. (2015). The complete chloroplast genome provides insight into the evolution and polymorphism of Panax ginseng. *Front. Plant Sci.* 5. doi:10.3389/fpls.2014.00696

Zoschke, R., Nakamura, M., Liere, K., Sugura, M., Borner, T., Schmitz-Zeitz, R., and Mitter, J. (2010). An organelar maturase associates with multiple group II introns. *Proc. Natl. Acad. Sci. U. S. A.* 107 (7), 3245–3250. doi:10.1073/pnas.0909400107