Postprint: Intraspecific Variation in the Complete Chloroplast Genome of the Rare and Endangered *Paphiopedilum parishii*

Authors: Gao Xinzhen, Tang Lu, Wang Yu, Shao Shicheng, Luo Yan

Date: 2023-07-13T00:00:00+00:00

Abstract

Paphiopedilum parishii has a narrow distribution range, with only sparse populations in China, Myanmar, Thailand, and Laos. In recent years, habitat destruction and human overexploitation have caused extreme reduction of wild populations of P. parishii. Genetic diversity represents the ability of a species to adapt to the environment and survive and develop; preserving the genetic diversity of rare and endangered species is an important objective of species conservation, yet the genetic information of P. parishii remains unclear. This study compared chloroplast genome sequences obtained from four wild individuals of P. parishii through sequencing, assembly, and annotation with previously published complete chloroplast genome sequences of another two individuals to analyze intraspecific variation. The results showed: (1) The chloroplast genome of P. parishii possesses a typical circular quadripartite structure of angiosperm chloroplast genomes, with a genome length of 154,403–154,809 bp, encoding a total of 129 genes, including 78 protein-coding genes, 39 tRNA genes, 8 rRNA genes, and 4 pseudogenes. (2) A total of 103–107 SSRs (simple sequence repeats) loci were detected in the chloroplast genomes of six individuals of P. parishii, among which 21 SSR loci were polymorphic. Additionally, 60 long sequence repeats were detected in the chloroplast genomes of the six individuals, including 17–21 forward repeats, 18–29 reverse repeats, 9–16 palindromic repeats, and 4–9 complementary repeats. (3) By comparing nucleotide diversity among the chloroplast genome sequences of six individuals, a total of 70 variations were identified, including 10 SNPs (single nucleotide polymorphism) and 60 insertions/deletions (InDels). Among these, three SNP loci underwent nonsynonymous substitutions, resulting in changes to amino acids in encoded functional genes; 19 insertions/deletions exhibited high polymorphism and have potential for development as molecular markers. (4) By calculating nucleotide diversity values (Pi), eight variable regions were identified, with Pi values ranging from

0 to 0.00632; among these, regions with greater variation were rps3-rpl22, trnL-UAC-rpl32, rpoB-trnC-GCA, and ycf4, and these highly variable regions could be developed as molecular markers for evaluating genetic diversity in P. parishii. (5) Phylogenetic analysis results indicated that chloroplast genome sequences of six individuals of P. parishii clustered together, forming a sister group with Paphiopedilum dianthum. In summary, SSRs, long sequence repeats, SNPs, InDels, and nucleotide sequences of the P. parishii chloroplast genome exhibit sufficient intraspecific diversity and can be developed as molecular markers for phylogenetic and conservation biology research on this species.

Full Text

Intraspecific Genetic Variation within the Chloroplast Genome of a Rare and Endangered Species *Paphiopedilum parishii* (Orchidaceae)

Gao Xinzhen[1,2], Tang Lu[1], Wang Yu[1,3], Shao Shicheng[1], Luo Yan[1]*

[1]Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan 666303, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]School of Landscape Architecture, Southwest Forestry University, Kunming 650224, China

Abstract: *Paphiopedilum parishii* is a rare and endangered orchid species with a narrow distribution, found only in small, fragmented populations across China, Myanmar, Thailand, and Laos. In recent years, habitat destruction and over-harvesting have caused severe declines in its wild populations. Genetic diversity underpins a species' capacity to adapt and survive, making its preservation a critical conservation priority. However, the genetic information for *P. parishii* remains largely unknown. This study sequenced, assembled, and annotated the complete chloroplast genomes of four wild individuals and compared them with two publicly available chloroplast genomes to characterize intraspecific variation. The results revealed: (1) The *P. parishii* chloroplast genome exhibits the typical circular quadripartite structure of angiosperms, with a length of 154,403–154,809 bp, encoding 129 genes comprising 78 protein-coding genes, 39 tRNA genes, 8 rRNA genes, and 4 pseudogenes. (2) Across the six individuals, 103–107 simple sequence repeat (SSR) loci were detected, of which 21 were polymorphic. Additionally, 60 long repeats were identified, including 17–21 forward repeats, 18–29 reverse repeats, 9–16 palindromic repeats, and 4–9 complement repeats. (3) Comparative analysis identified 70 variants, including 10 single nucleotide polymorphisms (SNPs) and 60 insertions/deletions (InDels). Three SNPs caused non-synonymous substitutions altering amino acids in functional genes, while 19 InDels showed high polymorphism and potential for molecular marker development. (4) Nucleotide diversity analysis revealed eight variable regions with Pi values ranging from 0 to 0.00632, with the most divergent regions being *rps3-rpl22*, *trnL-UAC-rpl32*, *rpoB-trnC-GCA*, and *ycf4*, which could

be developed as molecular markers for assessing genetic diversity. (5) Phylogenetic analysis showed that all six *P. parishii* individuals formed a monophyletic group, with *P. dianthum* as its sister species. In conclusion, the chloroplast genome of *P. parishii* harbors sufficient intraspecific diversity in SSRs, long repeats, SNPs, InDels, and nucleotide sequences to support the development of molecular markers for evolutionary and conservation biology studies.

Keywords: chloroplast genome, *Paphiopedilum parishii*, sequence divergence, polymorphic DNA markers, InDels, SSRs

---

Genetic diversity forms the foundation of biodiversity and represents a species' capacity to adapt and survive. Preserving the genetic diversity of rare and endangered species is a primary objective of conservation efforts (Guerrant & Pavlik, 1998; Huang, 2018). Chloroplasts are essential semi-autonomous organelles responsible for photosynthetic carbon fixation and stress responses in plant cells. The chloroplast genome is a circular DNA molecule with uniparental inheritance, typically ranging from 107 to 218 kb in size and maintaining stable gene structure and composition (Zhang & Li, 2011; Ivanova et al., 2017). Most chloroplast genomes exhibit a quadripartite structure comprising two inverted repeat regions (IR), a large single copy region (LSC), and a small single copy region (SSC) (Wolfe et al., 1987; Tian & Li, 2002; Jansen et al., 2005; Zhang & Li, 2011; He et al., 2019). These genomes encode 110–130 genes, primarily involved in photosynthesis and chloroplast gene expression (Zhang & Li, 2011). Despite their relatively small size, chloroplast genomes contain substantial genetic information, displaying both evolutionary divergence and conservation at the nucleotide sequence level and in structural rearrangements. Coding and non-coding regions evolve at different rates, providing molecular markers for distinguishing interspecific relationships and assessing genetic variability. Consequently, chloroplast genomes have been widely applied in population genetics and phylogenetic studies (Wolfe et al., 1987; Zhang & Li, 2011; He et al., 2019; Li et al., 2022). Recent studies utilizing intraspecific variation from multiple chloroplast genomes have provided fundamental genetic information for evolutionary research, germplasm utilization, and endangered plant conservation (Ishizuka et al., 2017; Muraguri et al., 2020; Zhang et al., 2020b).

The genus *Paphiopedilum* belongs to the subfamily Cypripedioideae of Orchidaceae, comprising over 80 species distributed primarily in tropical and subtropical limestone mountainous regions across Asia to Pacific islands (Cribb, 1998; Liu et al., 2009). Known as "slipper orchids" due to their pouch-like labella, *Paphiopedilum* species possess exceptional ornamental value (Yang et al., 2021). In recent years, natural habitats have been severely degraded by human activities, and combined with over-collection, wild populations have declined dramatically, with some species facing extinction in the wild (Luo et al., 2003; Yang et al., 2021). All wild *Paphiopedilum* species are listed in the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), prohibiting international trade, and all Chinese species are included in

the National Key Protected Wild Plants List (implemented September 2021). Urgent research is needed to understand the endangerment mechanisms and develop conservation strategies for Chinese *Paphiopedilum* species.

Paphiopedilum parishii is an epiphytic orchid that grows on tree trunks or rocks at elevations of 1,000–1,100 m and is classified as a first-class nationally protected plant in China (Figure 1 [Figure 1: see original paper]). The International Union for Conservation of Nature (IUCN) assesses *P. parishii* as Endangered, with only a few wild populations remaining in southern Yunnan, China, Thailand, Myanmar, and Laos (Rankou & Averyanov, 2015). Despite its high ornamental value and cultivation in botanical gardens, wild populations have become extremely rare. Chen et al. (2012) documented a unique automatic self-pollination mechanism in this species in southern Yunnan. Phylogenomic studies based on complete chloroplast genomes have placed *P. parishii* within subgenus *Paphiopedilum* (Guo et al., 2021). However, genetic information for *P. parishii* remains incomplete, and molecular markers for assessing genetic diversity and intraspecific variation are lacking, hindering comprehensive conservation efforts.

This study sequenced four wild individuals of *P. parishii* using low-coverage genome sequencing, assembled and annotated their complete chloroplast genomes, and compared them with two publicly available chloroplast genomes (Guo et al., 2021; Kao et al., 2021). We analyzed intraspecific variation to understand the evolutionary dynamics and genetic diversity of *P. parishii*, with the goal of developing polymorphic molecular markers for conservation biology research and providing foundational genetic data for investigating endangerment mechanisms and conservation strategies.

1.1 Plant Material, DNA Extraction, and Sequencing

Four *P. parishii* individuals were collected from two wild populations in Mengla County (P. parishii_1 and P. parishii_2) and Lancang County (P. parishii_3 and P. parishii_4), Yunnan Province (Figure 1). Fresh leaves were dried in silica gel. Total genomic DNA was extracted using a plant genomic DNA extraction kit (TIANGEN, Beijing, China). Sequencing libraries were constructed with the Illumina TruSeq kit (San Diego, CA, USA) and sequenced on the Illumina HiSeq 2500 platform (Shanghai Personal Biotechnology Co., Ltd.) using low-coverage genome sequencing. Additionally, two published chloroplast genome sequences of *P. parishii* (P. parishii_5, MW528213 and P. parishii_6, MN587822) were downloaded from GenBank for comparative genomic analysis.

1.2 Chloroplast Genome Assembly and Annotation

Raw sequencing data were filtered using Fastp to obtain high-quality clean data. The chloroplast genomes were assembled de novo from the sequencing data using GetOrganelle v1.6.3 (Jin et al., 2018). Assembled chloroplast genomes were aligned using the MAFFT tool in Geneious Prime v2021.2.2, and protein-coding regions, ribosomal RNAs (rRNAs), and

transfer RNAs (tRNAs) were annotated using the online tools CPGAVAS2 (http://47.90.241.85:16019/analyzer/annotate2) (Shi et al., 2019), GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html) (Tillich et al., 2017), and tRNAscan-SE v2.0.3 (Lowe & Chan, 2016). Annotations were corrected by referencing published *Paphiopedilum* chloroplast genomes. The chloroplast genome map was drawn using the OGDRAW online tool (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) (Lohse et al., 2007). Geneious Prime 2021.2.2 was used to analyze the basic structure of the six chloroplast genomes, including lengths of LSC, SSC, and IR regions, GC content, and numbers of protein-coding genes, tRNA genes, and rRNA genes.

1.3 Analysis of Repeat Sequences and SSRs

Simple sequence repeats (SSRs) in the chloroplast genomes were identified using the MISA online tool (https://webblast.ipk-gatersleben.de/misa) (Beier et al., 2017) with minimum repeat thresholds set to 10 for mononucleotides, 5 for dinucleotides, 4 for trinucleotides, and 3 for tetra-, penta-, and hexanucleotides. Other repeat sequences were identified using the REPuter online tool (https://bibiserv.cebitec.uni-bielefeld.de/reputer) (Kurtz et al., 2001) with parameters set to a minimum repeat length of 30 bp and a Hamming distance of 3 (Liang et al., 2019).

1.4 Comparative Analysis of Chloroplast Genomes

The mVISTA online tool (https://genome.lbl.gov/vista) (Frazer et al., 2004) was used to compare coding regions, intergenic spacers, and introns of the six chloroplast genomes, using P. parishii_1 as the reference. IRscope scripts in R v4.0.2 (Amiryousefi et al., 2018) were used to visualize and compare the junction boundaries among the six individuals, assessing expansion and contraction patterns. DnaSP v5.10.0.1 was used to identify SNPs and InDels and calculate nucleotide diversity (Pi) values among the six chloroplast genomes.

1.5 Phylogenetic Analysis

Complete chloroplast genome sequences of 15 *Paphiopedilum* species were downloaded from GenBank and analyzed together with the four newly sequenced individuals and two published *P. parishii* genomes, using three *Cypripedium* species as outgroups (accession numbers shown in Figure 5 [Figure 5: see original paper]). PhyloSuite v1.2.2 was used to determine the optimal nucleotide substitution model with ModelFinder (Zhang et al., 2020a). Complete chloroplast genome sequences were aligned using MAFFT, and a maximum likelihood (ML) phylogenetic tree was constructed using IQ-TREE (Trifinopoulos et al., 2016). Branch support was assessed via bootstrap analysis with 5,000 replicates and 1,000 iterations.

2.1 Basic Characteristics of the *Paphiopedilum parishii* Chloroplast Genome

After filtering raw data, the four samples yielded 2.4–3.1 Gb of high-quality clean data each. The assembled and annotated complete chloroplast genome sequences have been deposited in GenBank (accession numbers OP604356–OP604359). Comparative analysis of the four newly sequenced individuals and two published genomes revealed that all six chloroplast genomes share the typical circular quadripartite structure, consisting of LSC, SSC, and two IR regions (Figure 2 [Figure 2: see original paper]). Genome length ranged from 154,403 to 154,809 bp, with LSC regions of 86,581–86,983 bp, SSC regions of 2,436–2,446 bp, and IR regions of 32,690–32,693 bp. Total GC content was 35.9%, with 33.4% in the LSC, 29.1–29.2% in the SSC, and 39.5% in the IR regions (Table 1).

All six individuals contained 129 genes, including 78 protein-coding genes, 39 tRNA genes, 8 rRNA genes, and 4 pseudogenes (*ndhJ*, *ndhD*, and two copies of *ycf15*) (Appendix 1). Twenty-two genes were duplicated in the IR regions, present in two copies: nine protein-coding genes (*psaC*, *ndhB*, *rps7*, *rps15*, *rps19*, *rpl2*, *rpl23*, *ycf1*, *ycf2*), nine tRNA genes (*trnA-UGC*, *trnH-GUG*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnL-UAG*, *trnN-GUU*, *trnR-ACG*, *trnV-GAC*), and four rRNA genes (*rrn16*, *rrn23*, *rrn4.5*, *rrn5*). Eighteen genes contained introns: fifteen with a single intron (nine protein-coding genes: *petB*, *petD*, *atpF*, *ndhB*, *rpoC1*, *rps16*, *rpl2*, *rpl16*, *accD*; and six tRNA genes: *trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, *trnV-UAC*), two genes with two introns (*clpP*, *ycf3*), and *rps12* as a trans-spliced gene (Appendix 1).

2.2 SSR Analysis

MISA analysis identified 103–107 SSR loci in each of the six chloroplast genomes, including 47–51 mononucleotide repeats (predominantly A/T), 20 dinucleotide repeats, 15 trinucleotide repeats, 14 tetranucleotide repeats, 2 pentanucleotide repeats, and 5 hexanucleotide repeats (Figure 3 [Figure 3: see original paper]A, Appendix 2). Most SSRs were located in the LSC region (84–88), followed by 18–19 in the IR regions, with none in the SSC region (Figure 3B). SSRs were primarily distributed in non-coding intergenic spacers (72–76), with 17 in introns and 14 in coding regions (Figure 3C, 3D).

Detailed comparison of homologous SSR regions among the six individuals identified 21 polymorphic SSR loci with the same repeat type but different copy numbers, suitable for intraspecific marker development (Table 2).

2.3 Repeat Sequence Analysis

REPuter analysis identified 60 repeat sequences in each of the six chloroplast genomes, including 17–21 forward repeats, 18–29 reverse repeats, 4–9 complement repeats, and 9–16 palindromic repeats (Figure 4 [Figure 4: see original

paper]A). Repeat lengths varied substantially but were concentrated in the 30–40 bp range (Figure 4B).

2.4 Interspecific Chloroplast Genome Diversity

Using P. parishii_1 as the reference, mVISTA alignment of the six genomes revealed high overall sequence similarity, with slightly higher diversity in the LSC region compared to IR and SSC regions (Appendix 3). The most divergent fragments were located in non-coding regions such as *atpH-atpI*, *psaA-ycf3*, *atpB-rbcL*, *accD-psaI*, and *trnP-UGG-psaJ*. Coding regions showed lower variation, though the *ycf1* gene exhibited relatively high divergence, while tRNA and rRNA genes were highly conserved.

Comparison of junction boundaries showed that *rpl22*, *trnL-UAG*, *rps19*, and *psbA* were located at or near the LSC/IRb, IR/SSC, IRa/LSC boundaries, respectively. The *rpl22* gene spanned the LSC/IRb boundary with 58 bp in the IRb region; the duplicated *trnL-UAG* gene was located in the IR region, 176–186 bp from the IR/SSC boundary; the duplicated *rps19* gene was in the IR region, 282 bp from the IR/LSC boundary; and *psbA* was in the LSC region, 96 bp from the IRa/LSC boundary (Appendix 4).

2.5 Nucleotide Diversity in *Paphiopedilum parishii*

Analysis of nucleotide diversity identified 10 SNPs and 60 InDels among the six chloroplast genomes, distributed across intergenic spacers (4 SNPs, 43 InDels), introns (1 SNP, 13 InDels), and coding regions (5 SNPs, 4 InDels) (Table 3). Among the 10 SNPs, three were non-synonymous: an A→C substitution in *rpoC1* changed isoleucine to serine; a C→T substitution in *rpoB* changed glycine to arginine; and a G→C substitution in *ycf4* changed methionine to isoleucine. Two SNPs were synonymous, and the remainder were in non-coding regions. Analysis of InDels >3 bp identified 19 polymorphic loci (Table 4), such as the *psaA-ycf3* intergenic spacer, which contained insertions in P. parishii_1, _3, and _4 but was absent in P. parishii_2, _5, and _6. Two InDels in the *accD* coding region and intron also showed intraspecific polymorphism.

Nucleotide diversity (Pi) values calculated for coding and intergenic regions revealed that coding regions had Pi values of 0–0.00061, with only five genes showing Pi > 0, the highest being *ycf4*. Intergenic regions showed Pi values of 0–0.00632, with only three regions exceeding Pi > 0, the highest being *rps3-rpl22* (Appendix 5). These results indicate low overall nucleotide diversity within *P. parishii*, with relatively higher diversity in intergenic regions. Most variable sites were located in the LSC region, while IR and SSC regions were more conserved.

2.6 Phylogenetic Relationships of *Paphiopedilum parishii*

PhyloSuite identified TVM+F+R2 as the optimal substitution model for phylogenetic reconstruction. The ML analysis of 16 *Paphiopedilum* species, using

three *Cypripedium* species as outgroups, resolved three highly supported clades (bootstrap support = 100%). *Paphiopedilum delenatii*, *P. armeniacum*, *P. emersonii*, and *P. micranthum* formed a basal clade within subgenus *Parvisepalum*. *Paphiopedilum wenshanense*, *P. concolor*, and *P. bellatulum* formed a monophyletic group within subgenus *Brachypetalum*. The six *P. parishii* individuals formed a single clade with *P. dianthum* as sister species (Figure 5 [Figure 5: see original paper]).

3 Discussion and Conclusion

This study employed next-generation sequencing to obtain low-coverage genome data from wild *P. parishii* populations, successfully assembling and annotating complete chloroplast genomes. All six individuals shared identical gene content and order without genome rearrangements or inversions. The chloroplast genomes exhibited the typical quadripartite structure of IR, LSC, and SSC regions (Jansen et al., 2005). Genome length (154,403–154,809 bp) is consistent with previously reported chloroplast genomes of Orchidaceae and *Paphiopedilum* (Kim et al., 2014; Guo et al., 2021). Length variation among individuals occurred primarily in the LSC region (86,581–86,983 bp), while SSC and IR regions were relatively stable (2,436–2,446 bp and 32,690–32,693 bp, respectively). The IR regions showed significant expansion and the SSC region substantial contraction, containing only two genes (*rpl32* and *ccsA*), which aligns with characteristic features of *Paphiopedilum* chloroplast genomes (Guo et al., 2021). Compared to other orchids such as *Bulbophyllum lingii* (18,244 bp SSC), *Dendrobium wangliangii* (18,373 bp SSC), and *Cypripedium japonicum* (21,911 bp SSC) (Kim et al., 2020; Shao et al., 2020; Tang et al., 2021), the SSC region in *Paphiopedilum* is severely contracted, resulting in a much smaller overall genome size. Extensive gene transfer from the SSC to IR regions is common in *Paphiopedilum*; in *P. armeniacum*, typical SSC genes such as *ycf1*, *psaC*, and *ndhD* have been relocated to the IR region (Kim et al., 2015; Lin et al., 2015; Niu et al., 2017). Larger IR regions may enhance plastome stability through homologous recombination-based DNA repair, potentially benefiting gene expression (Palmer & Thompson, 1982; Wicke et al., 2011). This gene transfer creates high interspecific diversity in SSC region length (524–5,913 bp) and gene content, while maintaining intraspecific stability, making it a promising source of molecular markers for species identification (Guo et al., 2021).

The *P. parishii* chloroplast genome encodes 129 genes, including 78 protein-coding genes, 39 tRNA genes, 8 rRNA genes, and 4 pseudogenes, consistent with previous reports (Guo et al., 2021). The four pseudogenes resulted from large deletions causing loss of function: *ndhJ* lacks a start codon, while *ndhD* and *ycf15* have lost over 50% of their coding sequences. The genome shows extensive loss of *ndh* genes, retaining only *ndhB*, *ndhD*, and *ndhJ*, with the latter two being pseudogenes, contributing to SSC contraction. Loss of *ndh* genes is common in Orchidaceae (Yang et al., 2013; Feng et al., 2016; Niu et al., 2017; Zavala-Páez et al., 2020), though the extent varies; *Bulbophyllum* retains 12 *ndh*

genes (Tang et al., 2021), *Cymbidium* retains 10 (Hu, 2020), while *Dendrobium* shows severe loss (Niu, 2017). Chloroplast *ndh* genes encode components of the NAD(P)H dehydrogenase complex involved in cyclic electron transport, playing important roles in stress tolerance. Besides orchids, *ndh* gene loss has also been reported in gymnosperms (Braukmann et al., 2009; Wu et al., 2009) and may represent an evolutionary transition from autotrophy to heterotrophy (Lin et al., 2017).

Comparative genomic analysis identified 103–107 SSR loci in the six *P. parishii* genomes, with mononucleotide repeats being most abundant, followed by dinucleotide repeats, showing strong A/T bias as observed in other angiosperm chloroplast genomes (Qin et al., 2015; Chen & Yang, 2022). SSRs are important in plant genomes due to their high polymorphism and utility as molecular markers in population genetics and evolutionary studies (Muraguri et al., 2020). We identified 21 polymorphic SSR loci suitable for assessing genetic diversity in *P. parishii*. SSRs were far more abundant in non-coding than coding regions, indicating higher genetic diversity in non-coding regions, likely due to relaxed selective pressure (Shaw et al., 2007). Long repeats (forward, reverse, complement, and palindromic) also showed variation among individuals, possibly arising from insertions/deletions that alter repeat types and may be associated with recombination events (Somaratne et al., 2019).

Insertions, deletions, and mutations causing structural differences represent important sources of genetic variation and reflect adaptive capacity (Han & Xue, 2003). Chloroplast structural variation has been used to study phylogenetic relationships and population genetics at both inter- and intraspecific levels (McCauley, 1995; Kersten et al., 2016). We identified 60 InDels and 10 SNPs among the six *P. parishii* chloroplast genomes, primarily in intergenic spacers. Compared to other species, these numbers are relatively low: Muraguri et al. (2020) found 162 SNPs and 92 InDels in 12 *Ricinus communis* individuals; Zhang et al. (2020b) detected 77 SNPs and 255 InDels in three *Quercus acutissima* individuals; and Alexander et al. (2014) identified 8 SNPs and 45 InDels in four *Q. rubra* individuals. We found 19 polymorphic loci with potential as molecular markers. mVISTA alignment and Pi value calculations confirmed that non-coding regions exhibited substantially higher nucleotide diversity than coding regions, with *rps3-rpl22* showing the highest Pi value (0.00632). These structural variants provide valuable resources for studying inter- and intraspecific relationships and genetic diversity.

Genetic diversity is a crucial indicator of a species' adaptive capacity and resilience, and assessing genetic diversity and structure is prerequisite for effective conservation planning (Zhang et al., 2019). Through comparative analysis of complete chloroplast genomes, we have identified polymorphic molecular markers that provide a genetic foundation for evaluating genetic diversity, phylogeny, and conservation genetics of *P. parishii*.

Acknowledgments

We thank Sven Landrein and Yang Guoping of Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences for assistance with fieldwork, and Zhang Ze of Yunnan Yelantang Biotechnology Co., Ltd. for providing photographs.

References

Alexander, L. W., & Woeste, K. E. (2014). Pyrosequencing of the northern red oak (*Quercus rubra* L.) chloroplast genome reveals high quality polymorphisms for population management. *Tree Genetics & Genomes*, 10, 803–812.

Amiryousefi, A., Hyvönen, J., & Poczai, P. (2018). IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics*, 34(17), 3030–3031.

Beier, S., Thiel, T., Münch, T., Scholz, U., & Mascher, M. (2017). MISA-web: A web server for microsatellite prediction. *Bioinformatics*, 33(16), 2583–2585.

Braukmann, T. W., Kuzmina, M., & Stefanović, S. (2009). Loss of all plastid *ndh* genes in Gnetales and conifers: Extent and evolutionary significance for the seed plant phylogeny. *Current Genetics*, 55(3), 323–337.

Chen, L. J., Liu, K. W., Xiao, X. J., Tsai, W. C., Hsiao, Y. Y., Huang, J., & Liu, Z. J. (2012). The anther steps onto the stigma for self-fertilization in a slipper orchid. *PLOS ONE*, 7(5), e37478.

Chen, M. S., & Yang, Z. Y. (2022). Genealogical structure and differentiation analysis of *Carpinus tientaiensis* based on single nucleotide polymorphism of chloroplast genome. *Guihaia*, 42(10), 1703–1716.

Cribb, P. (1998). *The genus Paphiopedilum*. Borneo, Malaysia: Natural History Publications.

Feng, Y. L., Wicke, S., Li, J. W., Han, Y., Lin, C. S., Li, D. Z., Zhou, X. T., & Huang, W. C. (2016). Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. *Genome Biology and Evolution*, 8(7), 2184–2196.

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., & Dubchak, I. (2004). VISTA: Computational tools for comparative genomics. *Nucleic Acids Research*, 32, W273–W279.

Guerrant, E. O., & Pavlik, B. M. (1998). Reintroduction of rare plants: Genetics, demography, and the role of ex situ conservation methods. In P. L. Fiedler & P. M. Kareiva (Eds.), *Conservation biology: For the coming decade* (pp. 80–108). Boston, MA: Springer US.

Guo, Y. Y., Yang, J. X., Bai, M. Z., Liu, Z. J., & Li, D. Z. (2021). The chloroplast genome evolution of Venus slipper (*Paphiopedilum*): IR expansion,

SSC contraction, and highly rearranged SSC regions. *BMC Plant Biology*, 21(1), 248.

Han, B., & Xue, Y. B. (2003). Genome-wide intraspecific DNA-sequence variations in rice. *Current Opinion in Plant Biology*, 6(2), 134–138.

He, J., Yao, M., Lyu, R. D., Li, S. W., Zhu, G. P., & Song, Y. (2019). Structural variation of the complete chloroplast genome and plastid phylogenomics of the genus *Asteropyrum* (Ranunculaceae). *Scientific Reports*, 9(1), 15285.

Hu, G. J. (2020). *The complete chloroplast genomes of* Paphiopedilum* and *Cymbidium* (Orchidaceae) species: Comparative genomic and phylogenetic analyses* (Doctoral dissertation). Northwest University, Xi'an.

Huang, H. W. (2018). *Principles and practice of ex situ plant conservation*. Beijing: Science Press.

Ishizuka, W., Tabata, A., Ono, K., & Ubukata, M. (2017). Draft chloroplast genome of *Larix gmelinii* var. *japonica*: Insight into intraspecific divergence. *Journal of Forest Research*, 22(6), 393–398.

Ivanova, Z., Sablok, G., Daskalova, E., Zahmanova, G., Apostolova, E., Yahubyan, G., & Favaretto, P. (2017). Chloroplast genome analysis of resurrection tertiary relict *Haberlea rhodopensis* highlights genes important for desiccation stress response. *Frontiers in Plant Science*, 8, 204.

Jansen, R. K., Raubeson, L. A., Boore, J. L., dePamphilis, C. W., Chumley, T. W., Haberle, R. C., Wyman, S. K., Alverson, A. J., Peery, R., Herman, S. J., & Fourcade, H. M. (2005). Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods in Enzymology*, 395, 348–384.

Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., dePamphilis, C. W., Yi, T. S., & Li, D. Z. (2018). GetOrganelle: A simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. *Genome Biology*, 21(1), 241.

Kao, H., Zhao, Y., Yang, M., & Liu, C. (2021). The complete chloroplast genome sequences of an endangered orchid species *Paphiopedilum parishii* (Orchidaceae). *Mitochondrial DNA Part B*, 6(9), 2521–2522.

Kersten, B., Rampant, P. F., Mader, M., Le Paslier, M. C., Bounon, R., Berard, A., Vettori, C., Schroeder, H., Leplé, J. C., & Fladung, M. (2016). Genome sequences of *Populus tremula* chloroplast and mitochondrion: Implications for holistic poplar breeding. *PLoS ONE*, 11(1), e0147209.

Kim, H. T., Kim, J. S., Moore, M. J., Neubig, K. M., Williams, N. H., Whitten, W. M., & Kim, J. H. (2015). 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(11), e0142215.

Kim, J. S., Kim, H. T., Kim, J. H., Lee, S. H., & Kim, J. H. (2014). The largest plastid genome of monocots: A novel genome type containing AT residue repeats in the slipper orchid *Cypripedium japonicum*. *Plant Molecular Biology Reporter*, 33(5), 1–11.

Kim, Y. K., Jo, S. J., & Choi, S. H. (2020). Plastome evolution and phylogeny of Orchidaceae, with 24 new sequences. *Frontiers in Plant Science*, 11, 22.

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., & Giegerich, R. (2001). REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research*, 29(22), 4633–4642.

Liang, C. L., Wang, L., Lei, J., Jiang, H. B., & Liu, Y. L. (2019). A comparative analysis of the chloroplast genomes of four *Salvia* medicinal plants. *Engineering*, 5(5), 907–915.

Lin, C. S., Chen, J. J., Huang, Y. T., Chan, M. T., Daniell, H., Chang, W. J., Hsu, C. T., Liao, D. C., Wu, F. H., & Lin, S. Y. (2015). The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. *Scientific Reports*, 5(1), 9040.

Lin, C. S., Chen, J. J. W., Chiu, C. C., Hsu, C. Y., Wu, F. H., & Yang, C. H. (2017). Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. *The Plant Journal*, 90(5), 994–1006.

Liu, Z. J., Chen, X. Q., & Chen, L. J. (2009). *The genus Paphiopedilum in China*. Beijing: Science Press.

Li, R. Z., Cai, J., Yang, J. B., Li, D. Z., & Yi, T. S. (2022). Plastid phylogenomics resolving phylogenetic placement and genera phylogeny of Sterculioideae (Malvaceae s.l.). *Guihaia*, 42(1), 25–38.

Lohse, M., Drechsel, O., & Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics*, 52(5/6), 267–274.

Lowe, T. M., & Chan, P. P. (2016). tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research*, 44(1), W54–W57.

Luo, Y. B., Jia, J. S., & Wang, C. L. (2003). A general review of the conservation status of Chinese orchids. *Biodiversity Science*, 11(1), 70–77.

McCauley, D. E. (1995). The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends in Ecology & Evolution*, 10(5), 198–202.

Muraguri, S., Xu, W., Chapman, M., & Morrell, P. L. (2020). Intraspecific variation within castor bean (*Ricinus communis* L.) based on chloroplast genomes. *Industrial Crops and Products*, 155, 112779.

Niu, Z. T., Xue, Q. Y., Zhu, S. Y., Sun, J. T., Liu, X., & Liu, J. X. (2017). The complete plastome sequences of four orchid species: Insights into the evolution

of the Orchidaceae and the utility of plastomic mutational hotspots. *Frontiers in Plant Science*, 8, 715.

Niu, Z. T. (2017). *Comparative plastomic studies of* Dendrobium* species and comparison of the physiological effects and transcriptome responses of *Dendrobium officinale* under different abiotic stresses* (Doctoral dissertation). Nanjing Normal University.

Palmer, J. D., & Thompson, W. F. (1982). Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell*, 29(2), 537–550.

Qin, Z., Wang, Y. P., Wang, Q. M., Li, A., Hou, F. N., & Zhang, L. (2015). Evolution analysis of simple sequence repeats in plant genomes. *PLoS ONE*, 10(12), e0144108.

Rankou, H., & Averyanov, L. (2015). *Paphiopedilum parishii. The IUCN Red List of Threatened Species*, 2015, e.T193512A2240580.

Shao, S. C., Tang, L., & Luo, Y. (2020). The complete chloroplast genome sequence of *Dendrobium wangliangii* (Orchidaceae). *Mitochondrial DNA Part B*, 5(3), 3513–3515.

Shaw, J., Lickey, E. B., Schilling, E. E., & Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *American Journal of Botany*, 94(3), 275–288.

Shi, L. C., Chen, H. M., Jiang, M., Wang, L., Wu, X. Q., Huang, L., & Liu, C. (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Research*, 47(1), W65–W73.

Somaratne, Y., Guan, D. L., Wang, W. Q., Xu, S. Z., Han, X., & Song, Y. (2019). Complete chloroplast genome of *Xanthium sibiricum* provides useful DNA barcodes for future species identification and phylogeny. *Plant Systematics and Evolution*, 305(10), 949–960.

Tang, H. Q., Tang, L., Shao, S. C., & Luo, Y. (2021). Chloroplast genomic diversity in *Bulbophyllum* section *Macrocaulia* (Orchidaceae, Epidendroideae, Malaxideae): Divergence and adaptive evolution. *Plant Diversity*, 43(5), 350–361.

Tian, X., & Li, D. Z. (2002). Application of DNA sequences in plant phylogenetic studies. *Acta Botanica Yunnanica*, 24(2), 170–184.

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., & Greiner, S. (2017). GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, 45(1), W6–W11.

Trifinopoulos, J., Lam-Tung, N., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, 44(1), W232–W235.

Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F., & Quandt, D. (2011). The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Molecular Biology*, 76(3–5), 273–297.

Wolfe, K. H., Li, W. H., & Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences*, 84(24), 9054–9058.

Wu, C. S., Lai, Y. T., Lin, C. P., Wang, Y. N., & Chaw, S. M. (2009). Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: Selection toward a lower-cost strategy. *Molecular Phylogenetics and Evolution*, 52(1), 115–124.

Yang, J. B., Tang, M., Li, H. T., Zhang, Y., & Li, D. Z. (2013). Complete chloroplast genome of the genus *Cymbidium*: Lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evolutionary Biology*, 13(1), 84.

Yang, Y. J., Huang, J. L., Hu, H., & Zhang, Z. M. (2021). Progress on conservation and utilization of *Paphiopedilum* species in China. *Journal of West China Forestry Science*, 50(5), 108–112.

Zavala-Páez, M., Vieira, L. D. N., DeBaura, V. A., & de Oliveira, R. R. (2020). Comparative plastid genomes of neotropical *Bulbophyllum* (Orchidaceae; Epidendroideae). *Frontiers in Plant Science*, 11, 799.

Zhang, D., Gao, F. L., Jakovlić, I., Zou, H., Zhang, J., Li, W. X., & Wang, G. T. (2020a). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular Ecology Resources*, 20(1), 348–355.

Zhang, R. S., Yang, J., Hu, H. L., Yang, H. Q., & Yang, J. B. (2020b). A high level of chloroplast genome sequence variability in the sawtooth oak *Quercus acutissima*. *International Journal of Biological Macromolecules*, 152, 340–348.

Zhang, Y. H., Jia, H. X., Wang, Z. B., Zhang, J., & Wang, J. (2019). Genetic diversity and population structure of *Populus yunnanensis*. *Biodiversity Science*, 27(4), 355–365.

Zhang, Y. J., & Li, D. Z. (2011). Advances in phylogenomics based on complete chloroplast genomes. *Plant Diversity and Resources*, 33(4), 365–375.

Appendix 1 Gene Contents in the Chloroplast Genomes of Six *Paphiopedilum parishii* Individuals

Group of Genes	Gene Names
Photosystem I	$psaA$, $psaB$, $psaC$ ($\$\times 2$), $*psaI*$, $*psaJ*$
	$\|\|PhotosystemII\|*$ $psbA*$, $*psbB*$, $*psbC*$, $*psbD*$, $*psbE*$, $*psbF*$, $*psbH*$, $*psbI$
	$\|\|Cytochromeb/fcomplex\|*$ $petA*$, $*petB**$, $*petD*$
	$*$, $*petG*$, $*petL*$, $*petN*$
	$\|\|ATPsynthase\|*$ $atpA*$, $*atpB*$, $*atpE*$, $*atpF*$
	$*$, $*atpH*$, $*atpI*$
	$\|\|NADHdehydrogenase\|*$ $ndhB*(\times 2)$, $*ndhD*\Delta$, $*ndhJ*$
	$\Delta\|\|RubisCOlargesubunit\|*$ $rbcL*$
	$\|\|RNApolymerasesubunits\|*$ $rpoA*$, $*rpoB*$, $*rpoC1*$
	$*$, $*rpoC2*$
	$\|\|Ribosomalproteins(smallsubunit)\|*$ $rps2*$, $*rps3*$, $*rps4*$, $*rps7*$
	$(\times 2)$, $*rps8*$, $*rps11*$, $*rps12*$
	$(\times 2)$, $*rps14*$, $*rps15*$
	$(\times 2)$, $*rps16*$
	$*$, $*rps18*$, $*rps19*$
	$(\times 2)\|\|Ribosomalproteins(largesubunit)\|*$ $rpl2**(\times 2)$, $*rpl14*$, $*rpl16*$
	$*$, $*rpl20*$, $*rpl22*$, $*rpl23*$
	$(\times 2)$, $*rpl32*$, $*rpl33*$, $*rpl36*$
	$\|\|Othergenes\|*$ $clpP*$
	$*$, $*matK*$, $*accD*$
	$*$, $*ccsA*$, $*infA*$
	$\|\|Unknownfunctionproteins\|*$ $ycf1*(\times 2)$, $*ycf2*$
	$(\times 2)$, $*ycf3**$, $*ycf4*$, $*ycf15*$
	$(\times 2)\Delta\|\|RibosomalRNAs\|*$ $rrn4.5*(\times 2)$, $*rrn5*$
	$(\times 2)$, $*rrn16*(\times 2)$, $*rrn23*$
	$(\times 2)\|\|TransferRNAs\|*$ $trnA-UGC**(\times 2)$, $*trnC-$
	$GCA*$, $*trnD-$
	$GUC*$, $*trnE-UUC*$, $*trnF-$
	$GAA*$, $*trnfM-$
	$CAU*$, $*trnG-GCC*$, $*trnG-$
	$UCC**(\times 2)$, $*trnH-GUG*$
	$(\times 2)$, $*trnI-CAU*$
	$(\times 2)$, $*trnI-GAU*$
	$*(\times 2)$, $*trnK-UUU*$
	$*$, $*trnL-CAA*(\times 2)$, $*trnL-$
	$UAA**$, $*trnL-UAG*$
	$(\times 2)$, $*trnM-CAU*$, $*trnN-$
	$GUU*(\times 2)$, $*trnP-$
	$UGG*$, $*trnQ-$
	$UUG*$, $*trnR-ACG*$
	$(\times 2)$, $*trnR-UCU*$, $*trnS-$
	$GCU*$, $*trnS-GGA*$, $*trnS-$

Group of Genes	Gene Names

Note: indicates genes containing one intron; ** indicates genes containing two introns; ($×$2) indicates duplicated genes; Δ indicates pseudogenes.*

Appendix 2 Types and Amounts of Simple Sequence Repeats (SSRs) Identified in the Chloroplast Genomes of Six *Paphiopedilum parishii* Individuals

SSR Type	P. parishii_1	P. parishii_2	P. parishii_3	P. parishii_4	P. parishii_5	P. parishii_6
Mononucleotide	47	49	49	51	49	49
Dinucleotide	20	20	20	20	20	20
Trinucleotide	15	15	15	15	15	15
Tetranucleotide	14	14	14	14	14	14
Pentanucleotide	2	2	2	2	2	2
Hexanucleotide	5	5	5	5	5	5
Total	**103**	**105**	**105**	**107**	**105**	**105**

Appendix 3 Alignment of the Chloroplast Genome Sequences of Six *Paphiopedilum parishii* Individuals

With Paphiopedilum parishii_1 as reference. The X-axis corresponds to coordinates within the chloroplast genome. The Y-axis shows percentage identity in the 50–100% range.

Appendix 4 Comparison of LSC, IRs, and SSC Boundaries Among the Six *Paphiopedilum parishii* Plastomes

JLB, JSB, JSA, and JLA refer to junction boundaries between LSC/IRb, SSC/IRb, SSC/IRa, and LSC/IRa, respectively.

Appendix 5 Comparative Analysis of Nucleotide Diversity Values (Pi) Among the Chloroplast Genome Sequences of Six *Paphiopedilum parishii* Individuals

A. Nucleotide diversity (Pi) values of coding genes in LSC, SSC, and IR regions; B. Nucleotide diversity (Pi) values of intergenic regions in LSC, SSC, and IR regions.

Note: Figure translations are in progress. See original paper for figures.