Research paper

Chloroplast genomic diversity in Bulbophyllum section Macrocaulia (Orchidaceae, Epidendroideae, Malaxideae): Insights into species divergence and adaptive evolution

Hanqing Tang a,1, Lu Tang a,1, Shicheng Shao b, Yulan Peng c, Lu Li d,**, Yan Luo a,e,*

a Gardening and Horticulture Department, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglung, Mengla, 666303, Yunnan, China
b College of Forestry, Shanxi Agricultural University, Taigu, Jinzhong, 030800, Shanxi, China
c Key Laboratory of Mountain Ecological Restoration and Bioresource Utilization & Ecological Restoration Biodiversity Conservation, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, Sichuan, 610041, China
d Department of Biodiversity Conservation, Southwest Forestry University, Kunming, 650224, Yunnan, China
e Gardening and Horticulture Department, Core Botanical Gardens, Chinese Academy of Sciences, Menglung, Mengla, 666303, Yunnan, China

A R T I C L E   I N F O

Article history:
Received 29 October 2020
Received in revised form 7 January 2021
Accepted 12 January 2021
Available online 28 January 2021

Keywords:
Chloroplast genome
Sequence divergence
DNA markers
SSR
Bulbophyllum
Orchidaceae

A B S T R A C T

Bulbophyllum is the largest genus in Orchidaceae with a pantropical distribution. Due to highly significant diversifications, it is considered to be one of the most taxonomically and phylogenetically complex taxa. The diversification pattern and evolutionary adaptation of chloroplast genomes are poorly understood in this species-rich genus, and suitable molecular markers are necessary for species determination and phylogenetic analysis. A natural Asian section Macrocaulia was selected to estimate the interspecific divergence of chloroplast genomes in this study. Here, we sequenced the complete chloroplast genome of four Bulbophyllum species, including three species from section Macrocaulia. The four chloroplast genomes had a typical quadripartite structure with a genome size ranging from 156,182 to 158,524 bp. The chloroplast genomes included 113 unique genes encoding 79 proteins, 30 tRNAs and 4 rRNAs. Comparison of the four chloroplast genomes showed that the three species from section Macrocaulia had similar structure and gene contents, and shared a number of indels, which mainly contribute to its monophyly. In addition, interspecific divergence level was also great. Several exclusive indels and polymorphism SSR loci might be used for taxonomical identification and determining interspecific polymorphisms. A total of 20 intergenic regions and three coding genes of the most variable hotspot regions were proposed as candidate effective molecular markers for future phylogenetic relationships at different taxonomical levels and species divergence in Bulbophyllum. All of chloroplast genomes in four Bulbophyllum species were under purifying selection, while 13 sites within six genes exhibited site-specific selection. A whole chloroplast genome phylogenetic analysis based on Maximum Likelihood, Bayesian and Parsimony methods all supported the monophyly of section Macrocaulia and the genus of Bulbophyllum. Our findings provide valuable molecular markers to use in accurately identifying species, clarifying taxonomy, and resolving the phylogeny and evolution of the genus Bulbophyllum. The molecular markers developed in this study will also contribute to further research of conservation of Bulbophyllum species.

Copyright © 2021 Kunming Institute of Botany, Chinese Academy of Sciences. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Bulbophyllum Thouars is probably the largest pantropical genus of Orchidaceae, with ca. 2200 species (Pridgeon et al., 2014). The Paleotropical region is the richest in species of this genus, with ca. 1700 species occurring in Asian (sub) tropical area, followed by
et al., 2007; Smidt et al., 2011). The orchid family (Gamisch and Comes, 2019). Bulbophyllum species are notoriously difficult to classify taxonomically, due to highly morphological variation (Seidenfaden, 1979). Many independent and closely related genera established then combined. Numerous sectional classifications remained controversial (Pridgeon et al., 2014).

Several studies have been reported in recent years examining the phylogeny of Bulbophyllum using molecular data (Chase et al., 2015; Hosseini et al., 2012, 2016; Gravendeel et al., 2014; Fischer et al., 2007; Smidt et al., 2011). The finding of these works well supported the monophyly of continental groups, e.g., Asian, African and Neotropical clade, but contributed little to elucidating the relationships at infragenic, interspecific level. Very recently, Hu et al. (2020) combined DNA and floral characters to reconstruct the phylogenetic relationship in Asian Cirrhopetalum alliance, and further indicated that the majority of traditional sectional classifications were shown to be highly artificial. The elucidation of the phylogenetic relationships within Bulbophyllum still requires more detailed study.

The chloroplast (cp) genomes are considered to be useful and informative genetic resources for studies on evolutionary relationships in the plant kingdom at various taxonomic levels (Asaf et al., 2017; Wu et al., 2020), because of their uniparentally inherited and relatively conserved genome structure and higher evolutionary rate (Shaw et al., 2014; Wicke et al., 2011). Especially in Orchidaceae, the size and gene content of cp genome vary greatly. The size of the cp genome within orchids is ranging from 10 to 170 kb (Dong et al., 2018; Lin et al., 2015; Kim et al., 2015). The most reduced chloroplast genomes have been found in several mycoheterotrophic genera, e.g., Epipogium, Gastrodia, and Aphyllorchis (Feng et al., 2016; Schkelunov et al., 2015; Yuan et al., 2018), which have lost many photosynthetic genes (Graham et al., 2017). There are more than 100 orchid cp genomes, including Denrobiun 40 spp., Holcoglossum 11 spp., Corallorhiza 10 spp., Cymbidium 9 spp., Neottia 7 spp., have been reported and available in the recent years (Barrett and Davis, 2012; Kim et al., 2020a; Li et al., 2019; Logacheva et al., 2011; Niu et al., 2017; Shao et al., 2020; Yang et al., 2013; Zhu et al., 2018). However, the cp genomes are still little known in Bulbophyllum. Only nine cp genomes of Bulbophyllum species (eight neotropical and one Asian) were reported until now (Kim et al., 2020a; Zavala-Páez et al., 2020). Comparative cp genome studies at different taxonomical level are needed to clarify the diversification pattern in this species-rich group, and then to develop suitable molecular markers to study speciation, evolutionary adaptation and phylogeny. For taxonomists, uncertainty in taxonomic status of the Bulbophyllum species is quite often, which also requires suitable molecular markers for taxonomic revision (Hosseini et al., 2016; Zavala-Páez et al., 2020).

The Asian section Macroaulia (Bl.) Aver. is one of acceptable natural sections in Bulbophyllum supported both by morphological characters and molecular evidence, and it is a medium section with more or less 68 species, mainly distributed in Southeastern Asia (Averyanov and Komarov, 1994; Vermeulen, 2014). There are six species recorded in (sub) tropical regions of China, which are the most northern parts for section Macroaulia (Chen and Vermeulen, 2009; Jin et al., 2019; Huang et al., 2020). The section Macroaulia is characterized by mini-miniature sized, epiphytes with creeping, close set pseudobulbs carrying a single, apical, usually persistent leaf that blooms on a single flower that have morphological diverse labellum and sepals. Based on the nrITS, the section Macroaulia is monophyletic and regarded as the basal group sister to the rest of Asian Bulbophyllum clade (Gravendeel et al., 2014).

In this study, the complete chloroplast genomes of three species from section Macroaulia were sequenced, assembled and characterized, then to investigate chloroplast genome diversification pattern and examine species divergence and evolutionary adaptation. The main objectives of the present study were to (1) understand the interspecific variation pattern in the cp genomes of closely related species in the genus of Bulbophyllum; (2) test the monophyly of section Macroaulia using chloroplast genome; (3) identify genes underlying positive selection as potentially genes for adaptive evolution, radiation and speciation; (4) examine hotspot regions as candidate sequences for species delimitation and phylogenetic analysis in the genus Bulbophyllum.

2. Materials and methods

2.1. Plant material

Three species, Bulbophyllum lingii M.Z. Huang, G.S. Yang & J.M. Yin, B. pantenurum Seid., and B. menghaiense Z.H. Tsi, are closely related species from section Macroaulia. B. lingii is newly found (Huang et al., 2020) and B. pantenurum is a new record (unpublished) in China recently. Bulbophyllum gedangense Y. Luo, J. P. Deng & Jian W. Li, which is also newly found in China recently, might belonging to section Desmosanthes (Bl.) J. J. Smith (Luo et al., 2020). Fresh leaves of four Bulbophyllum species were collected from the living collections at the greenhouse of Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences and Yunnan Fengchunfang Biotechnology Co. Ltd. The voucher specimens were deposited in the herbarium of Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences (HTBCC) (Table 1). In addition, we downloaded the available complete cp genomes of nine other Bulbophyllum species from GenBank (Table 1) to analyze phylogenetic relationships.

2.2. DNA extraction and sequencing

Total genomic DNA from fresh leaves was extracted by using TiangenDNA kit (TIANGEN, China). An Illumina paired-end DNA library was constructed using the IlluminaTruSeq Library Preparation Kit (San Diego, CA, USA) following the manufacturer’s instructions. The library was sequenced by the Illumina Hiseq 2500 sequencing platform (Illumina, CA, USA) at Personal Biotechnology Co., Ltd (Shanghai, China).

2.3. Chloroplast genome assembly and annotation

The total 3 G data of sequencing reads were trimmed to high quality reads (Table 2), and then GetOrganelle pipeline were used for de novo chloroplast genome assembly (Jin et al., 2020). CGAPAVAS2 (Shi et al., 2019) and GeSeq (Tillich et al., 2017) were used to annotate the assembled chloroplast genome using default parameters to predict protein coding, tRNA, and tRNA genes. The annotation results were checked plus manual corrections and codon positions were adjusted by comparing them with the more similar six Bulbophyllum reference species present in the database using Geneious prime. The genome map of the species was illustrated with the help of OGDRAW (Greiner et al., 2019), and the annotated chloroplast genome sequences were submitted to NCBI (Table 1). GC content was calculated using Geneious prime.

2.4. Simple repeat sequence analysis

We identified simple sequence repeats (SSRs) of Bulbophyllum with MISA (Beier et al., 2017) by setting the minimum number of
repeats to 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides, respectively.

2.5. Comparative genome analysis

The complete chloroplast genomes of four Bulbophyllum species were compared using the program mVISTA (Frazer et al., 2004). We also compared the borders of large single copy (LSC), small single copy (SSC), and inverted repeat (IR) regions in the genomes of the four species. The nucleotide diversity value (Pi) of the cp genomes of all four Bulbophyllum species was calculated using DnaSP v5.10 (Librado and Rozas, 2009). The number of mutations and indels events was calculated by DnaSP v5.0 (Rozas et al., 2004), and their rate was calculated per 100 bp. Due to indels having higher levels of homoplasly than mutations, indels were considered as events instead of sites in the alignment (Ingvarsson et al., 2003).

2.6. Selective pressure analysis

In order to detect the protein coding genes under selection in four Bulbophyllum species, the synonymous (dS), non-synonymous (dN) nucleotide substitution rates and the dN/dS ratio were calculated using the model in DnaSP v5.0. (Librado and Rozas, 2009). To explore the selection patterns and identify positive selection on the protein coding genes, six models, model M0, M3, M1a, M2a, M7 and M8, implemented in the PAMLX Codeml program (Xu and Yang, 2013). The codon frequencies were determined by the F3 × 4 model. 79 common genes were examined. The Bayes empirical Bayes (BEB) method was used to statistically identify sites under positive selection with posterior probabilities >0.95. Pairwise distances and branch lengths were computed using the maximum likelihood (ML) criterion under a codon model. The likelihood ratio test (LRT) was used to estimate the quality of each model (Yang and Nielsen, 2002; Yang et al., 2005).

2.7. Phylogenetic analysis

The complete chloroplast genomes of the four newly sequenced Bulbophyllum species, together with one Asian (from Japan) and eight neotropical Bulbophyllum species, and two species from Dendrobium (Table 1), which download from GenBank, were selected to reconstruct phylogenetic tree. Phylogenetic analyses were performed for two datasets: complete plastid genome sequences and the 68 protein coding sequences. The chloroplast DNA sequences were aligned with MAFFT v.7 (auto strategy) (Katoh and Standley, 2013). Maximum parsimony (MP), Maximum Likelihood (ML) and Bayesian (BI) analyses were run for each data set. Parsimony analyses were performed with MegaX (Kumar et al., 2008). Heuristic searches were conducted with tree bisection-reconnection (TBR) branch swapping, and random stepwise addition with 1000 replications. ML trees with 1000 BS were produced.
with IQ-tree (Triﬁnopoulous et al., 2016). Bootstrap analyses were used to evaluate the support for individual clades with 1000 replicates. BI tree used MrBayes 3.2.4 (Ronquist et al., 2012) in PhyloSuite (Zhang et al., 2020). The Markov chain Monte Carlo (MCMC) algorithm was run with two independent chains with a random starting tree and default priors for $2 \times 10^6$ generations, with trees sampled every 100 generations. The first 25% of all generations were excluded, and a consensus phylogenetic tree was obtained based on majority rule from the remaining trees.

3. Results

3.1. Characteristics of Bulbophyllum chloroplast genome

The cp genomes of three species of section Macrocaulia ranged in size from 156,182 bp (Bulbophyllum pentaneurum) to 156,689 bp (B. lingii), and that of B. gedangense is 158,524 bp (Fig. 1, Table 2). The genome length differences between the largest and smallest genomes were the 2342 bp. All of the cp genomes had a typical quadripartite structure, similar to those of most land plants, consisting two inverted repeats (IRs, 26,838 to 26,919 bp) separated by large (LSC, 84,240 to 86,200 bp) and small (SSC, 18,105 to 18,632 bp) single copy regions, respectively (Fig. 1, Table 2). GC content was similar (36.6–36.8%; Table 2). In addition, the gene composition and order in the cp genomes of four Bulbophyllum species are nearly identical (Fig. 1, Table 2). Four each cp genome, 113 unique genes were annotated, including 79 protein-coding genes, 30 tRNAs, and 4 rRNAs (Table 2). Most of genes occur as a single copy in LSC or SSC, while 20 genes are duplicated in the IRs, including eight protein coding genes, eight tRNAs and four rRNAs (Table 3). Eight protein coding genes and six tRNA genes included one intron, while three genes (ycf3, clpP, rps12) included two introns (Table 3).

3.2. Characters of simple sequence repeats

A total of 45 (Bulbophyllum pentaneurum) to 58 (B. menghaiense) SSRs were detected in the cp genome of four Bulbophyllum species, and six categories of SSRs (mono-, di-, tri-, tetra-, penta-, hexa-...
nucleotide repeats) were found with lengths ranged from 10 to 26 bp (Fig. 2, Table 4). In the four *Bulbophyllum* species, the most abundant SSRs were mono-nucleotide repeats, accounting for 63.64%–78.57%, and number varies from 28 to 44, followed by 4 (6.9%) to 8 (14.29%) di-nucleotide repeats, 0 to 4 (6.9%) tri-nucleotide repeats, 2 (3.57%) to 10 (17.24%) tetra-nucleotide repeats, 0 to 2 (3.45%) penta-nucleotide repeats, and 0 to 2 (3.45%) hexa-nucleotide repeats (Fig. 2A, Table 4). Interestingly, all mono-nucleotide SSRs belonged to A or T type and the majority of di-, tri-, tetra-, penta-, and hexa-nucleotide SSRs were especially rich in A or T (Table S1). Of the SSR regions, 31 to 40 SSRs were identified in intergenic spaces, while 5 to 13 SSRs were located in the coding DNA sequence and 3 to 11 SSRs in the coding sequence introns (Fig. 2B).

According to the criteria of SSRs, i.e., presence of the same repeat units, being located in the homologous regions, and having different number of the repeat units, 35 SSRs were identified as polymorphic SSRs between *Bulbophyllum* species (Table 5). Most of SSR loci were specific for each species. *B. menghaiense* had 10 exclusive SSRs, e.g., (TTA)_4 in trnS-trnQ and atpB-rbcL regions; *B. lingii* had six exclusive SSRs, e.g., (AT)_5 in trnL-UAA and (TTA)_3 in clpP genes; *B. gedangense* had six exclusive SSRs e.g., (AT)_6 in trnT-trnl and (TA)_3 in trnE-trnl regions; *B. pentanurum* had five exclusive SSRs, e.g., (TA)_5 in psbM-trnD and (CCT)_4 in rpl14-rpl16 regions (Table 5).

### Table 3

| Group of genes     | Gene names          |
|--------------------|---------------------|
| 1 Photosystem I    | psaA, psaB, psaC, psal, psal |
| 2 Photosystem II   | psbA, psbB, psbC, psbD, psbE, psbF, psbT, psbT, psbT |
| 3 Cytochrome b/f complex | petA, petB, petC, petF, petH, petN |
| 4 ATP synthase     | atpA, atpB, atpE, atpF, atpH, atpI |
| 5 NADH dehydrogenase | ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhK |
| 6 Rubisco large subunit | rbcL |
| 7 RNA polymerase   | rps2, rps3, rps4, rps7 (\(\times 2\)), rps8, rps11, rps12 (\(\times 2\)), rps14, rps15, rps16 (\(\times 2\)), rps18, rps19 (\(\times 2\)) |
| 8 Ribosomal proteins (SSU) | rpl2a (\(\times 2\)), rpl14, rpl16 (\(\times 2\)), rpl20, rpl22, rpl23 (\(\times 2\)), rpl32, rpl33, rpl36 |
| 10 Other genes     | clpP, matK, accD, ccsA, infA, cemA |
| 11 Proteins of unknown function | rps12b, rps14, rps15, rps16 (\(\times 2\)), rps8, rps11, rps12, rps13, rps14, rps15, rps16 (\(\times 2\)), rps18, rps19 (\(\times 2\)) |
| 12 Ribosomal RNAs  | rmn4.5 (\(\times 2\)), rmn55 (\(\times 2\)), rmn16S (\(\times 2\)), rmn23S (\(\times 2\)) |
| 13 Transfer RNAs   | trnA-UGC (\(\times 2\)), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC, trnH-GUG (\(\times 2\)), trnI-CAU (\(\times 2\)), trnK-UUU, trnL-CAU (\(\times 2\)), trnL-UAA-A, trnL-UAC, trnM-CAU, trnN-GUU (\(\times 2\)), trnQ-UUG, trnR-AGG (\(\times 2\)), trnR-UCU, trnS-GGU, trnT-UGC, trnT-GGU, trnT-GGU, trnV-GAC (\(\times 2\)), trnV-UAC (\(\times 2\)), trnW-CAU, trnY-GUA, trnM-CAU |

**a** Gene containing one intron. **b** Gene containing two introns, (\(\times 2\)) Gene has two copies.

---

**3.3. Comparative genomic analysis**

**3.3.1. Sequence divergence**

The divergence of sequences in the cp genomes of four *Bulbophyllum* species was plotted by mVISTA program using the annotated *B. gedangense* sequence as a reference (Fig. 3). A whole genome alignment revealed an overall high sequence similarity among the four *Bulbophyllum* species, and also showed lower sequence divergence between three species from section *Macrocaulia*. Furthermore, LSC and SSC regions were more divergent compared with IR regions. The regions with highest level of divergence were rps16-trnQ, rnl16, ndhF-rpl32, psbl-trn5, psaA-ycf3, trnS-rps4, ndhF-rpl32, and rps15-ycf1 (Fig. 3).

**3.3.2. Mutations and indels**

The number of mutations and indels events were compared among these four *Bulbophyllum* cp genomes. There were 4600 mutations and 943 indels calculated. The intergenic regions had the most 2320 mutations and 710 indels respectively (Table 5). We further analyzed the details of indels (indel size≥10 bp) of the four *Bulbophyllum* species based on multiple whole cp genome alignment. 189 positions showed DNA indels in at least one of the cp genomes. The three species from section *Macrocaulia* shared 67 deletions, mostly located in intergenic spaces.
intergenic regions, which showed low similarity to *B. gedangense* (Table S2). These regions are related to large deletions, which become obvious in a multiple alignment of related whole cp genome sequences. The largest deletion, which is in the intron of *trnL*·UAG, is in *B. menghaiense* (237 bp), *B. pentaneurum* (189 bp) (Table S2). In all cp genomes of the four *Bulbophyllum* species, the genes *rpl22*, *psbT*, *ndhF*, *ycf1*, *rps19*, and *psbA* were located at the junction of the LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC borders (Fig. 4). The expansion and contraction of the border regions of cp genomes were similar in the four *Bulbophyllum* species. There were 10–78 bp extension of *rpl22* gene into IRb region. Moreover, the IRb/SSC borders were well conserved among the four cp genomes, of which the IRb region expanded into the gene *ndhF* with 70 bp. The *ycf1* gene crossed the SSC/IRa junction expanding 1059 bp to IRb regions in three species of section *Macrocaulia*, while 1074 bp in *B. gedangense*. There were 206–267 bp between *rps19* and the LSC/IRa border, meanwhile, the *psbA* generated a distance of 81–136 bp to another LSC/IRa junction.

### Table 4
Types and amounts of simple sequence repeats (SSRs) in the chloroplast genomes of four *Bulbophyllum* species.

| SSR Type | *B. menghaiense* | *B. pentaneurum* | *B. lingii* | *B. gedangense* |
|----------|------------------|------------------|-------------|-----------------|
| Mono-    | 38               | 43               | 33          | 44              |
| Di-      | 4                | 6                | 7           | 8               |
| Tri-     | 4                | 6                | 3           | 5               |
| Tetra-   | 10               | 17               | 7           | 2               |
| Penta-   | 2                | 3.45             | 2           | 3.85            |
| Hexa-    | 0                | 0.00             | 0           | 0.00            |
| Total    | 58               | 100.00           | 52          | 100.00          |

### Table 5
The polymorphic simple sequence repeats (SSRs) in the chloroplast genomes of four *Bulbophyllum* species. Number indicated SSR repeats.

| SSR Type | *B. menghaiense* | Location | Regions |
|----------|------------------|----------|---------|
| AT       | 5/6/5/0          | atpH-atp | LSC     |
| AT       | 0/0/5/0          | trnL-UAA | LSC     |
| AT       | 0/0/5/0          | trnL-UAG | LSC     |
| AT       | 0/0/5/0          | psa1-trp3 | LSC |
| AT       | 0/0/5/0          | psbB-psbT | LSC |
| TA       | 0/5/0/0          | psbM-trnB-GUC | LSC |
| TA       | 0/0/0/13         | trnL-UAG | LSC     |
| TA       | 0/0/5/0          | ctp-psbB | LSC     |
| CT       | 0/0/5/0          | trnL-UAA | LSC     |
| CT       | 0/4/0/0          | rpl14-rpl16 | LSC |
| TTA      | 4/0/0/0          | trnL-UAG | LSC     |
| TTA      | 3/3/3/0          | atpF-atpH | LSC     |
| TTA      | 3/0/0/0          | atpB-atl | LSC     |
| TTC      | 4/0/0/0          | trnL-UAG | LSC     |
| TTA      | 4/0/0/0          | trnL-UAG | LSC     |
| TTA      | 0/0/0/0          | rpl16    | LSC     |
| AAGT     | 3/0/0/0          | rpoB-trnC-GCA | LSC     |
| AATA     | 0/3/0/0          | atpF     | LSC     |
| AATA     | 3/3/3/0          | ndhG-ndhI | SSC |
| ATGCA    | 3/0/0/0          | trnN-GCU-trnR-ACG | IR |
| ATTA     | 3/0/0/0          | trnL-UAA | LSC     |
| GGGCA    | 3/0/0/0          | trnL-UAG | LSC     |
| GGGCA    | 3/0/0/0          | trnL-UAG | LSC     |
| GGGCA    | 3/0/0/0          | trnL-UAG | LSC     |
| TCC      | 0/0/0/5          | trnL-UAG | LSC     |
| ATCA     | 3/0/0/0          | trnL-UAG | LSC     |
| TCA      | 3/0/0/0          | trnL-UAG | LSC     |
| TCT      | 0/0/0/0          | rpl16    | LSC     |
| TGA      | 3/0/0/0          | ndhE     | SSC     |
| TTATA    | 0/0/0/3          | rpl33-tps3 | LSC |
| TTATA    | 0/0/0/3          | rpl33-tps3 | LSC |
| ATCAA    | 3/0/0/0          | trnN-GGA-tp4 | LSC |
| ATCAA    | 3/0/0/0          | trnN-GGA-tp4 | LSC |
| TACTG    | 0/0/0/0          | matK-trp16 | LSC |
| TCTTA    | 0/0/0/0          | rpoA     | LSC     |
| TTATA    | 0/0/0/0          | ctp      | LSC     |
| TTA TCA  | 0/0/0/0          | trnL-UAG | LSC     |
| TCTGTC   | 0/0/0/0          | rps7-rtnV-GAC | IR |

### 3.3.4. Nucleotide diversity

The nucleotide diversity value (Pi) of the coding regions and intergenic regions were calculated using DnaSP for understanding the sequence divergence among these four *Bulbophyllum* species. The coding regions were more conserved with the Pi value ranging from 0 to 0.06329, compared to the intergenic spacer with the Pi value ranging from 0 to 0.12543 (Fig. 5). Most of highly variable sites were located in the LSC regions, followed SSC regions. The sequences in IR regions exhibited lower nucleotide diversity value. Among the coding regions, the highest nucleotide diversity (Pi = 0.06329) was in *trnL-UAA* gene. A total of 12 genes were detected as highly variable coding regions, e.g., *trnL-UAA*, *rpl16*, *trnK-UUU* (Pi > 0.03), and *ycf1*, *psbM*, *matK*, *ndhA*, *petD*, *clpP*, *rps16*, *rps15*, *ndhA* (Pi > 0.02) (Fig. 5A). Among the non-coding regions, the highest nucleotide diversity (Pi = 0.12543) was in *psbB-psbT* region. A total of 20 intergenic regions, *psbB-psbT*, *rps16-trnQ*, *psaJ-rpl16*, *rps4*, *ndhF* were recognized as hotspot regions with nucleotide diversity >0.04 (Fig. 5B).

### 3.4. Selection pressure analysis of evolution

Syntonymous and non-syntonymous nucleotide substitution patterns are very important markers for gene evolution studies. A ratio of dN/dS < 1 indicates purifying selection, dN/dS > 1 denotes probable positive selection, and dN/dS values close to one indicates neutral evolution. We estimated the selective pressure among four *Bulbophyllum* species. 79 protein-coding genes were used to calculate the ratio of synonymous and non-synonymous substitution, and thereby to determine whether the selection pressure has acted on individual sequences. Thus, we found all 79 genes of the dN/dS average ratio were lower than 1.0 might be under purifying selection, and no gene under positive selection in four *Bulbophyllum* species (Fig. 6). We found that ten genes (*matK*, *rps15*, *rps16*, *ndhA*, *petA*, *rpoA*, *rps16*, *clpP*, *ycf1*, and *ycf2*) of the dN/dS average ratio above 0.5, indicating relaxed selection (Fig. 6). Fourteen most conserved genes (*psac*, *psA*, *psbF*, *psbL*, *psbL*, *psbN*, *psbT*, *psbZ*, *petD*, *petG*, *atpH*, *rpl23* and *rps7*) with the dN/dS average...
values of 0.0, suggesting very strong purifying selective pressure. The remaining genes showed the dN/dS average value were ranged from 0.0 to 0.5, suggesting strong purifying selective pressure. In addition, according to the models analysis, we found a total of 6 genes 13 sites exhibiting site-specific selection (Table 7). Of these genes, the ndhK gene harbored 6 sites under positive selection, with three in ycf2, and the other four genes (ndhF, rbcL, rps4, and rps16) each had only one selection site (Table 7).

3.5. Phylogenetic analysis

The protein coding gene sequences and all the complete chloroplast genome sequences of 15 Malaxideae members were used to reconstruct the phylogenetic relationships, including the 13 Bulbophyllum cpDNA sequences, and two Dendrobium species (Table 1, Fig. 7). In our analysis, the topologies obtained based on the maximum likelihood (ML), the parsimony (MP) and Bayesian (BI) analyses were largely consistent with the different two datasets, and the bootstrap values and posterior probabilities were very high for each lineage (Fig. 7). These results suggested that there was no conflict among the entire genome data set and protein coding gene sequences. All the sampled species of Bulbophyllum were clustered into one clade with 100% bootstrap value or the Bayesian posterior probability. Within the Bulbophyllum clade, the Asian tropical and neotropical Bulbophyllum form a monophyletic clade, respectively. Within the Asian tropical clade, B. pentaneurum was sister to B. lingii with B. menghaiense as their closest relative, which strongly supported that section Macrocauliais was monophyletic; In another clade, B. gedangense and B. inconspicuum (native to Japan) grouped together (Fig. 7).

4. Discussion

4.1. Conservation and divergence of chloroplast genome in the section Macrocauliais

Based on our comprehensive comparative genomics analysis, the chloroplast genomes of section Macrocauliais showed both conservative and divergent. In terms of genome size, all of four cp genomes fell within the size of Orchidaceae species (Kim et al., 2015; Lin et al., 2015). The three species of the section Macrocauliais cp genomes were similar and smaller than that of B. gedangense. In addition, although loss of ndh genes has been
previously in some orchids (Feng et al., 2016; Niu et al., 2017; Yang et al., 2013; Zavala-Páez et al., 2020), all the cp genomes of the section Macrocaulia had no evidence of ndh genes loss. The total length of angiosperm cp genomes is influenced by contractions and expansions of the IRs (Green, 2011; Weng et al., 2017). However, the decrease of cp genomes size in three species of section Macrocaulia might have been due to the large deletions. Several large deletions, more than 100 bp, were detected in the cp genomes of these miniature Bulbophyllum species, mostly in intergenic regions (Table S2). Compared with B. gedangense, 84 indels were shared in the three species of section Macrocaulia. The similarity between cp DNA sequences indicated the cp genome are highly conserved in the section Macrocaulia, and contributed to their monophyletic relationships. However, several exclusive large indels were also identified for the three species of section Macrocaulia respectively, e.g., an 87 bp-indel in rpoB-trnC intergenic region for...
B. menghaiense, a 65 bp-indel in the intron of rps16 gene for B. lingii, and a 43 bp indels in petN-psbM intergenic region for B. pentaneurum (Table S2). The differences between cp DNA sequences indicated interspecific differentiation in the section Macracaulia, and might contributed to the species divergence and adaptive evolution. Large indels easily identified in whole cp genome alignments might be used for species-specific molecular marker as recently shown for a Populus tremula-specific marker that has been developed based on a 96 bp-indel (Kersten et al., 2016). The indels identified among the four Bulbophyllum cp genomes in this study might also serve as species-specific molecular marker for species delimitation and identification.

4.2. Simple sequence repeats diversity

Simple sequence repeats (SSRs) are the most frequent in genome sequences with high mutation rates, and play important roles in plants (Qin et al., 2015). The most abundant SSR type was the mononucleotide repeat, and the majority of SSRs in four Bulbophyllum species was composed of polyT and polyA repeats,
corresponding with the angiosperm cp genome composition (Kim et al., 2015; Qin et al., 2015). In our study, the number and types of SSRs identified from the cp genomes were variable in four Bulbophyllum species (Fig. 2). Of di- to hexa-nucleotide SSRs among four Bulbophyllum species, most of SSRs were specific to each species, and only five SSRs shared in the section Macrocaulia. These findings indicated that SSRs exhibiting highly variable even in closely related species, might have few phylogenetic significance. Most of the previous studies revealed that the richness of SSR types is various in different species, which may contribute to the genetic variations differently among species (Kuang et al., 2011). SSRs are also effective molecular markers for interspecific and intraspecific polymorphisms and extensively used for species identification, speciation, population genetic diversity and phytogeography studies (Powell et al., 1995). In four Bulbophyllum species, SSR loci in the cp genome exhibited high level of interspecific polymorphisms. Several exclusive SSR loci presented in each species, e.g., B. menghaiense with 10 exclusive SSRs, B. lingii with six, and B. pentanurum with five. Thus, these specific SSRs are useful molecular markers for species identification and detecting genetic diversity.

4.3. Hotspot regions

We found the nucleotide sequence diversity was higher in the non-coding regions than the coding regions, which is generally consistent with most previous studies of the chloroplast genomes of angiosperms (Clegg et al., 1994). A number of cpDNA, e.g., matK, rbcL, psbA-trnH, trnS-trnG, trnL-trnF, trnI-ndhF, trnE-trnD are previously used within Bulbophyllum to reconstruct phylogenetic relationships, but have not been shown useful for detecting infragenic or interspecific relationships (Fischer et al., 2007; Hosseini et al., 2012, 2016; Smidt et al., 2011). Smidt et al. (2011), for example, used two non-coding regions of cpDNA, psbA-trnH and trnS-trnG to examine phylogenetic relationships among Neotropical sections of Bulbophyllum, but did not support the monophyly of the neotropical group, offering very little phylogenetic value. Fischer et al. (2007), who studied Madagascan Bulbophyllum using four plastid markers, did not resolved some terminal nodes. In recent years, numerous orchid plastid markers based on the comparative cp genome have been proposed for Orchidaceae phylogeny at different taxonomical level (Dong et al., 2018; Li et al., 2019; Niu et al., 2017; Yang et al., 2013; Zavala-Paz et al., 2020; Zhu et al., 2018).

Based on comprehensive comparative analysis, a total of 20 sequence divergence hotspot regions located in intergenic regions will be helpful to analyze phylogenetic relationships especially between closely related taxa, or where recent divergence, rapid radiation in Bulbophyllum (Daniell et al., 2016; Tonti-Filippini et al., 2017). Compared with the previous study of cp genomes of Neotropical Bulbophyllum (Zavala-Paz et al., 2020), where 10 potential markers proposed for barcoding and phylogenetic studies in Bulbophyllum, there are four regions identical (psbB-psbT, trnS-trnG, trnK-atpA, rpl32-trnL) with our work. In addition, three protein coding genes (trnL-UAA, rpl16, trnK-UUU) also showed remarkably higher Pi values (>0.03). These genes might be undergoing rapid nucleotide substitution at the species level, and might be applied for molecular evolution and adaptive evolution analysis.

4.4. Adaptive evolution

The dn/ds ratios have been widely used toinfer the evolutionary dynamics and identify adaptive signatures among species
The dN/dS value in all of 79 protein coding genes in four *Bulbophyllum* species were lower than 1. The dN/dS values are higher in protein-coding genes as a result of more frequent non-synonymous nucleotide substitutions than the synonymous substitutions, and the adaptation evolution may have contributed to the elevated dN/dS ratios (Makaowski and Boguski, 1998). Thus, the elevated dN/dS ratios indicated some unknown selective forces on *Bulbophyllum* species and results in species divergence (Hurst, 2002), which may increase our understanding of the evolutionary speciation and environment adaptation in *Bulbophyllum* species. We found that ten genes under relaxed selection (Fig. 6) and six genes under site-specific selection (Table 7). Of these genes, *petA* encoding cytochrome f, *ndhF* and *ndhK* encoding subunits of NADH dehydrogenase, *rbcL* encoding the large subunit of ribulose-bisphosphate carboxylase, all functioned in photosynthesis (Green, 2011; Table 3). A previous study showed that *rbcL* is often under positive selection in land plants, and evolved under strong positive selection after the C3–C4 photosynthetic transition in grasses (Kapralov and Filatov, 2007; Pirot et al., 2017). The genes *rpaA*, *rps15*, *rps16*, *rpl16*, *rpl22* all functioned in self-replication. The other genes are *matK* acting as a maturase, *ycf1* and *ycf2* with unknown function (Daniell et al., 2016; Green, 2011). These genes play important role in improving energy conversion efficiency, therefore they might have some evolutionary and adaptive significance for *Bulbophyllum* species. Different from our results, selection analysis in eight Neotropical *Bulbophyllum* species showed that four genes (*ycf1, *ycf2, accD*, and *rps3*) were under site-specific selection (Zavala-Páez et al., 2020). The differences of site-specific selection pattern between Asian and neotropical *Bulbophyllum* indicated that they might each had various adaptive strategies under unpredictable environmental conditions. Our findings from selective pressure analysis could lead to a better understanding of the evolution of *Bulbophyllum* species.

### 4.5. Phylogenetic relationship

Complete chloroplast DNA sequences were valuable for analyzing phylogenetic relationships (Daniell et al., 2016; Tonti-Filippini et al., 2017). Kim et al. (2020a), for example, utilized the chloroplast genomes of 124 species of orchids to evaluate the phylogenomics and relationship within the Orchidaceae family. Moreover, the whole cp genome phylogeny that was generated at different taxonomical level in several groups of Orchidaceae, such as subtribe Aeridineae (Kim et al., 2020b), subtribe Calypsoinae (Li et al., 2020), and Holcoglossum (Li et al., 2019), Considering that most species-level diversity of Orchidaceae, a whole cp genome phylogeny is highly desired for *Bulbophyllum*. We examined phylogenetic relationships among 13 *Bulbophyllum* species from two continental groups in this study. Both the complete cp genome sequences and 65 protein coding genes datasets produced the same topology with similar support. In addition, all the ML, BI and MP phylogenetic results strongly supported that *B. pentaneum* is sister to *B. lingii* with *B. menghaiense* as their closest relative, and all 13 *Bulbophyllum* species were closely clustered into one lineage, which strongly support that section Macrocalus and the genus of *Bulbophyllum* are monophyletic, respectively (Fig. 7). The monophyly of section Macrocalus was also supported by Gravendeel et al. (2014) using nuclear ITS sequence. However, our phylogenetic analysis did not show that section Macrocalus was the basal clade in Asian *Bulbophyllum* group as their work, might due to sampling limited. Furthermore, the phylogenetic relationships among the Neotropical clade were also in accordance with recent phylogeny research (Zavala-Páez et al., 2020). In the future, more complete cpDNA sequences of member individuals of *Bulbophyllum* are needed to exploit the power of the whole cp genome phylogenies, especially for differentiation at lower taxonomic levels.

### Author contributions

YL and LD designed the study and revised the manuscript. HQT, LT drafted the manuscript and prepared tables and figures. HQT, LD and SCS performed comparative genome analysis. YLP performed phylogenetic analysis. All authors discussed the results and commented on the manuscript.

### Declaration of competing interest

The authors declared no conflict of interest.

### Acknowledgements

We are grateful to Wenbin Yu, Yu Song, from Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Fei Zhao from Kunming Institute of Botany, Chinese Academy of Sciences, for their help in genomic data analysis. We also thank Lin Li, from South China Botanical Garden, Chinese Academy of Sciences, Xiaoyun Wang and Zhifeng Xu from the Orchid Conservation Center of Yunnan Fengchunfang Biotechnology Co., Ltd, for kindly providing photos and sample materials. This work was supported by the National Natural Science Foundation of China (No. 31870183, No.U1702235) and Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences (Y4ZK111B01).

### Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pld.2021.01.003.

### References

Asaf, S., Khan, A.L., Aaqil Khan, M., et al., 2017. Comparative analysis of complete plastid genomes from wild soybean (*Glycine soja*) and nine other *Glycine* species. PLoS One 12, e0182281.

Averyanov, L.V., Komarov, V.L., 1994. Identification Guide to Vietnamese Orchids (*Orchidaceae* Juss.). Botanical Institute, Russian Academy of Sciences, St. Petersburg, Russia. p. 432.

Barrett, C.F., Davis, J.L., 2012. The plastid genome of the myco-heterotrophic *Corallirhiza striata* (*Orchidaceae*) is in the relatively early stages of degradation. Ann. J. Bot. 99, 1513–1523.

Beier, S., Thiel, T., Münch, T., et al., 2017. MISA-web, a web server for microsatellite prediction. Bioinformatics 33, 2583–2585.

Chase, M.W., Cameron, K.M., Freudenstein, J.V., et al., 2015. An updated classification of Orchidaceae. Bot. J. Linn. Soc. 177, 151–174.

Chen, X.Q., Vermeulen, J.L., 2009. Bulbophyllum flouars. In: Wu, Z.Y., Raven, P.H., Hong, D.Y. (Eds.), *Flora of China*, vol. 25. Science Press and St. Louis, Missouri Botanical Garden Press, Beijing, China, pp. 404–440.

Clegg, M.T., Gaut, B.S., Leon, G.H., et al., 1984. Rates and patterns of chloroplast DNA evolution. Proc. Natl. Acad. Sci. U.S.A. 81, 6795–6801.

Daniell, H., Lin, C.S., Yu, M., et al., 2016. Chloroplast genomes, diversity, evolution, and applications in genetic engineering. Genome Biol. 17, 134.

Dong, W.L., Wang, R.N., Zhang, N.Y., et al., 2018. Molecular evolution of chloroplast genomes of orchid species, insights into phylogenetic relationship and adaptive evolution. Int. J. Mol. Sci. 19, 716.

Fischer, G.A., Gravendeel, B., Sieder, A., et al., 2007. Evolution in resupination of Malagasy species of *Bulbophyllum* (*Orchidaceae*). Mol. Phylogenet. Evol. 45, 358–376.

Feng, Y.L., Wicke, S., Li, J.W., et al., 2016. Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. Genome Biol. Evol. 8, 2164–2175.

Frazier, K.A., Pachter, L., Polakov, A., et al., 2004. VISTA, computational tools for comparative genomics. Nucleic Acids Res. 1, W273–W279.

Garnisch, A., Comes, H.P., 2019. Cline-age-dependent diversification under high species turnover shapes species richness disparities among tropical rainforest lineages of *Bulbophyllum* (*Orchidaceae*). BMC Evol. Biol. 19, 93.

Graham, S.W., Lam, V.K.Y., Merckx, V.S.F.T., 2017. Plastomes on the edge, the plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. Plant Diversity 43 (2021) 350–361.
