RESEARCH

Effects of geographic isolation on the *Bulbophyllum* chloroplast genomes

Jiapeng Yang[1,2], Fuwei Zhang[1,2], Yajie Ge[1,2], Wenhui Yu[1,2], Qiqian Xue[1,2], Mengting Wang[1,2], Hongman Wang[1,2], Qingyun Xue[1,2], Wei Liu[1,2], Zhitao Niu[1,2*] and Xiaoyu Ding[1,2*]

Abstract

Background: Because chloroplast (cp) genome has more conserved structures than nuclear genome and mitochondrial genome, it is a useful tool in estimating the phylogenetic relationships of plants. With a series of researches for cp genomes, there have been comprehensive understandings about the cp genome features. The genus *Bulbophyllum* widely distributed in Asia, South America, Australia and other places. Therefore, it is an excellent type genus for studying the effects of geographic isolation.

Results: In this study, the cp genomes of nine *Bulbophyllum* orchids were newly sequenced and assembled using the next-generation sequencing technology. Based on 19 Asian (AN) and eight South American (SA) *Bulbophyllum* orchids, the cp genome features of AN clade and SA clade were compared. Comparative analysis showed that there were considerable differences in overall cp genome features between two clades in three aspects, including basic cp genome features, SSC/IR$_B$ junctions (J$_{SB}$s) and mutational hotspots. The phylogenetic analysis and divergence time estimation results showed that the AN clade has diverged from the SA clade in the late Oligocene (21.50–30.12 mya). After estimating the occurrence rates of the insertions and deletions (InDels), we found that the change trends of cp genome structures between two clades were different under geographic isolation. Finally, we compared selective pressures on cp genes and found that long-term geographic isolation made AN and SA *Bulbophyllum* cp genes evolved variably.

Conclusion: The results revealed that the overall structural characteristics of *Bulbophyllum* cp genomes diverged during the long-term geographic isolation, and the crassulacean acid metabolism (CAM) pathway may play an important role in the *Bulbophyllum* species evolution.

Keywords: Geographic isolation, Chloroplast genome, *Bulbophyllum*, Species evolution

Background

Chloroplasts (cp) are semi-autonomous organelles with their own unique genomes. The cp genomes of terrestrial plants are relatively conservative, with sizes ranging from 120 to 170 kbp, and can be divided into four regions: large single copy (LSC) region, small single copy (SSC) region and two inverted repeat (IR$_B$ and IR$_A$) regions [1, 2]. Compared with nuclear genome and mitochondrial genome, cp genome is characterized by conserved structures, which made it a useful tool in estimating the phylogenetic relationships of plants [3, 4]. Although the cp gene content and composition are highly conserved, e.g., containing about 120 genes, mainly consisting of three categories: self-replication genes, photosynthesis genes, and RNA genes [5], the cp genomes have undergone structural changes in the long evolutionary process [6], such as IR contraction [7], inversion [8, 9], loss of genes and introns [10].

*Correspondence: niuzhitaonj@163.com; dingxynj@263.net
[1] College of Life Sciences, Nanjing Normal University, Nanjing 210023, China
Full list of author information is available at the end of the article

Yang *et al. BMC Plant Biology*   (2022) 22:201

Page 2 of 14

In recent years, as the sequencing costs continue to decrease, more and more complete cp genomes were sequenced and assembled, which is helpful for the plant phylogenetic studies. Orchids have become the focus of phylogenetic studies due to rich species, wide distributions and epiphytic habits. Based on 124 orchid cp genomes, Kim et al. (2020) cleared out the relationships of five subfamilies within Orchidaceae (Apostasioideae, Vanilloideae, Cypripedioideae, Orchidoideae and Epidendroideae) [4]. Moreover, there have been comprehensive understandings about the orchid cp genome features. For example, (i) cp genome substitution rates are considerably lower in IR regions than in SC regions [11, 12]; (ii) the repeat sequences played an important role in cp genome rearrangements [13–15]; (iii) mutational hotspots are genus specific [16]. However, the effects of geographic isolation on the cp genomes have not been investigated.

The genus *Bulbophyllum* Thou. belongs to the family Orchidaceae Juss., mainly distributed in Asia, South America, Australia and other places [17]. It is one of the largest genera of flowering plants, comprising about 2,000 species, which are predominantly restricted to rainforest habitats [18]. And it is characterized by epiphytic growth, CAM metabolism and fleshy flowers. There are about 105 species in China, which are widely distributed in Guizhou, Guangxi, Yunnan and Fujian provinces [19]. The wide distributions make the genus *Bulbophyllum* an excellent mode for studying the effects of geographic isolation on the cp genomes. There have been previous studies on the *Bulbophyllum* phylogenetic relationships. Gamisch et al. (2015) used internal transcribed spacer (ITS) sequences and four short cp genome fragments to study the phylogenetic relationships and estimate divergence times of Madagascar *Bulbophyllum* [20]. Gamisch et al. (2019) analyzed the evolutionary relationships of *Bulbophyllum* between different continents with ITS sequences, showing that Asian (AN) clade first separated from other *Bulbophyllum* orchids, and then South American (SA) clade separated from African *Bulbophyllum* [18]. Therefore, the *Bulbophyllum* is an excellent type genus to study the effects of geographical isolation on cp genomes.

In this study, the cp genomes of nine AN *Bulbophyllum* orchids were newly sequenced and assembled using the next-generation sequencing technology. The comprehensive comparative analyses were performed, based on 27 complete cp genomes of AN and SA *Bulbophyllum*. The aims of this study were: (1) to uncover the differences in cp genome features between AN clade and SA clade; (2) to assess the effects of long-term geographic isolation on cp genome structures; (3) to investigate the direct causes of the variable selective pressures on cp genes. We believed that this study will give us a comprehensive understanding of the effects of geographic isolation on the complete cp genomes.

Results

Chloroplast genome features

The nine newly sequenced *Bulbophyllum* cp genomes were circular with the typical quadripartite structure, including a large single copy (LSC), a small single copy (SSC) and a pair of inverted repeats (i.e., IR_A and IR_B) (Fig. 1). We combined the published cp genomes of ten AN and eight SA *Bulbophyllum* orchids with this study's nine AN *Bulbophyllum* orchids to compare the basic cp genome features of two clades. As shown in Table 1, the AN cp genome sizes were ranged from 144,380 bp to 158,524 bp, with GC contents of 36.66–38.04%. The cp genomes were variable in LSC and SSC regions, with 77,088 to 86,200 bp and 12,140 to 18,632 bp, while conserved in IR regions, with sizes ranging from 25,824 to 26,919 bp. The AN cp genome GC contents varied slightly in LSC (34.20–36.05%), SSC (29.00–30.93%) and IR (43.03–43.38%) regions. Notably, except for IR regions, the GC contents of AN clade were higher than SA clade (Mann–Whitney two-sided, $P < 0.05$) [17], which indicated that there were significant cp genome differences between two clades. Moreover, the *Bulbophyllum* cp genomes were well-conserved in gene number and gene order, with 74 protein-coding genes, 30 tRNA genes and 4 rRNA genes.

The levels of the nine newly sequenced cp genome sequence variability were estimated and plotted using mVISTA, with *Dendrobium huoshanense* as the reference (Fig. S1). As expected, the non-coding regions (intergenic spacers and introns) were more divergent than coding regions, and the SC regions exhibited higher divergence levels than IR regions. Comparative cp genomic analysis of *Bulbophyllum* revealed a minor organizational variation, which covered an inversion of *rpl32*.

Repeat and SSR analysis

The Longer dispersed repeats (LDRs) were detected using REPuter and shown in Fig. S2A. The results showed that, (i) the LDRs distribution densities in SSC region were higher than LSC and IR regions (Mann–Whitney two-sided, $P < 0.05$). (ii) The LDRs in SSC regions were highly varied between two clades (T-test for Independent Samples: $P < 0.05$). The analysis of simple sequence repeats (SSRs) also revealed the identical trends (Fig. S2B). Considering the higher LDRs and SSRs distribution densities with the lower GC contents in SSC regions, we conducted correlation tests to see if there were correlative relationships among them. Indeed, significant negative correlations were detected among them (LDRs vs GC contents: Pearson's $r = -0.529$; SSRs vs GC contents:
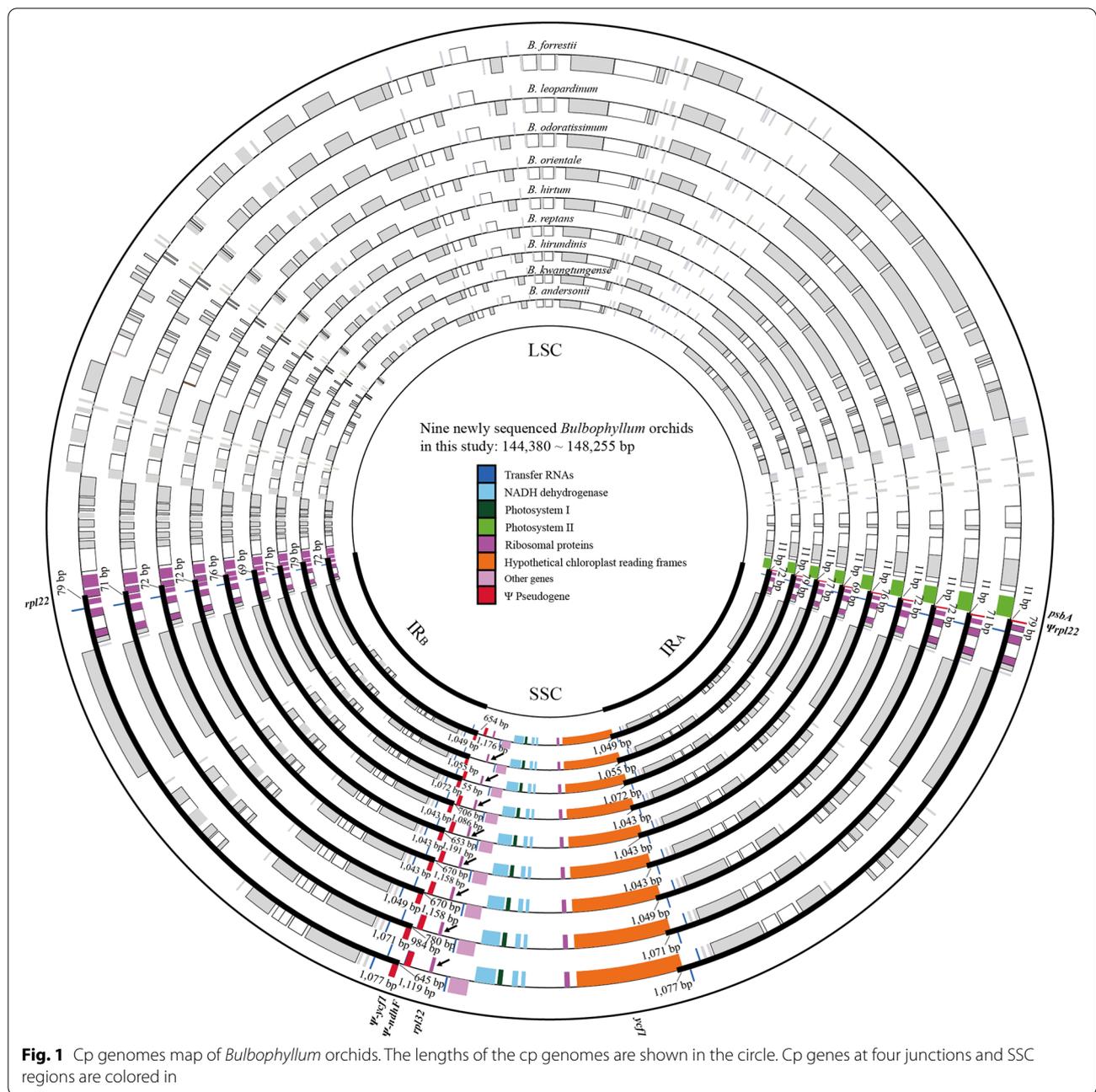
Fig. 1 Cp genomes map of *Bulbophyllum* orchids. The lengths of the cp genomes are shown in the circle. Cp genes at four junctions and SSC regions are colored in

Pearson's $r = -0.977$, $P < 0.05$), which indicated that low GC contents may lead to the SSC regions more variable.

Comparison of sequences flanking IR/SC junctions between Asian and South American clades

The sequences flanking IR/SC junctions in cp genomes were compared. As shown in Fig. 1 and Fig. S3, the *Bulbophyllum* IR/SC junctions were highly conserved. The SSC/IR$_A$ junctions (J$_{SA}$s) was located in the 5' end of *ycf1*, which resulted in a pseudogene fragment *Ψycf1* (33

to 1,434 bp) in the SSC/IR$_B$ junctions (J$_{SB}$s). Meanwhile, the LSC/IR$_B$ junctions (J$_{LB}$s) were located in *rpl22*, which leaded *Ψrpl22* (10 to 79 bp) in the LSC/IR$_A$ junctions (J$_{LA}$s). Remarkably, IR$_B$ expanded progressively from J$_{SB}$s to *ndhF*. The main differences between the two clades were in J$_{SB}$, which can be divided into three types, (A) a relatively long *Ψycf1* (796 to 1,434 bp) with a *ΨndhF* (*B. forrestii, B. leopardinum, B. odoratissimum, B. orientale, B. hirtum, B. hirundinis, B. reptans, B. menghaiense, B. andersonii, B. affine, B. pectinatum, B. funingense, B.*

Yang *et al. BMC Plant Biology*   (2022) 22:201

Page 4 of 14

Table 1 The features of *Bulbophyllum* cp genomes

No	Species	Accession no	Reference	Cp genome length (bp)	LSC length (bp)	IR length (bp)	SSC length (bp)	GC content (%)			
								Total	LSC	IR	SSC
1	B. disciflorum	LC498826	Tang et al. (2019) [21]	148,554	79,001	26,378	16,797	37.94	35.80	43.38	30.93
2	B. gedangense	MW161053	Tang et al. (2021) [22]	158,524	86,200	26,846	18,632	36.76	34.47	43.05	29.26
3	B. lingii	MW161052	Tang et al. (2021) [22]	156,669	84,607	26,919	18,224	36.77	34.35	43.09	29.30
4	B. menghaiense	MW161050	Tang et al. (2021) [22]	156,550	84,663	26,891	18,105	36.66	34.20	43.03	29.26
5	B. pentaneurum	MW161051	Tang et al. (2021) [22]	156,142	84,240	26,838	18,226	36.81	34.44	43.05	29.40
6	B. pingnanense	MW822749	Zhang et al. (2021) [23]	151,224	86,017	25,855	13,497	36.98	34.45	43.21	29.25
7	B. inconspicuum	NC_046811	Kim et al. (2020) [4]	149,548	85,760	25,824	12,140	37.00	34.41	43.17	29.00
8	B. forrestii	LC642719	This study	148,067	78,167	26,414	17,072	37.91	35.86	43.29	30.63
9	B. hirtum	LC642724	This study	147,382	77,587	26,333	17,129	37.96	35.93	43.34	30.60
10	B. hirundinis	LC642721	This study	144,380	77,399	26,437	14,107	38.02	35.79	43.22	30.76
11	B. kwangtungense	LC642722	This study	145,092	77,192	26,262	15,376	37.98	35.93	43.31	30.03
12	B. leopardinum	LC642723	This study	147,514	77,762	26,378	16,996	38.04	36.05	43.37	30.58
13	B. odoratissimum	LC642720	This study	147,494	77,664	26,290	17,250	37.88	35.87	43.32	30.34
14	B. orientale	LC642725	This study	147,388	77,392	26,395	17,206	38.01	36.05	43.29	30.65
15	B. reptans	LC642726	This study	146,928	77,088	26,401	17,038	37.98	35.98	43.29	30.58
16	B. andersonii	LC703293	This study	148,255	78,074	26,366	17,449	37.83	35.84	43.32	30.19
17	B. affine	LC556091	Yang et al. (2020) [24]	148,230	78,178	26,386	17,280	37.86	35.83	43.30	30.38
18	B. pectinatum	LC556092	Yang et al. (2020) [24]	147,169	77,478	26,081	17,529	38.01	36.01	43.37	30.87
19	B. funingense	LC556093	Yang et al. (2020) [24]	147,464	77,564	26,403	17,094	37.92	35.89	43.25	30.68
20	B. epiphytum	NC_048486	Zavala-Páez et al. (2020) [17]	147,546	82,354	25,847	13,498	36.73	34.06	43.20	28.20
21	B. exaltatum	NC_048480	Zavala-Páez et al. (2020) [17]	150,410	83,335	25,847	15,381	36.78	34.21	43.22	29.09
22	B. granulosum	NC_048481	Zavala-Páez et al. (2020) [17]	151,112	84,492	25,465	15,690	36.72	34.26	43.13	29.13
23	B. mentosum	NC_048482	Zavala-Páez et al. (2020) [17]	150,217	83,642	26,338	13,899	36.74	34.14	43.11	28.28
24	B. plumosum	NC_048479	Zavala-Páez et al. (2020) [17]	146,401	83,260	26,025	11,091	36.56	33.72	43.00	27.62
25	B. regnellii	NC_048483	Zavala-Páez et al. (2020) [17]	151,493	84,868	25,542	15,541	36.66	34.17	43.09	29.11
26	B. steyermarkii	NC_048484	Zavala-Páez et al. (2020) [17]	146,720	83,488	25,953	11,326	36.74	34.20	42.80	27.72
27	B. weddellii	NC_048485	Zavala-Páez et al. (2020) [17]	151,355	83,450	25,927	16,051	36.64	34.09	43.16	28.90

pentaneurum, B. disciflorum, B. gedangense, B. lingii, B. steyermarkii); (B) a short *Ψycf1* (156 to 360 bp) with a *ΨndhF*, including *B. mentosum* and *B. epiphytum*; (C) no *ΨndhF*, including *B. kwangtungense, B. inconspicuum, B. pingnanense, B. regnellii, B. exaltatum, B. granulosum, B. plumosum* and *B. weddellii*. However, there was no type B in AN clade. Meanwhile, there were more type C in SA clade. These results revealed that the evolution of IR/SC junctions between two clades were diversified.

Diverse mutational hotspots between the Asian and South American clades

We extracted a total of 72 intergenic and intronic loci (conserved sites > 100 bp) from two clades and calculated their sequence variabilities (SVs), with *D. huoshanense* as the reference. As shown in Fig. 2, the SVs in SC regions were significantly higher than IR regions (Mann–Whitney two-sided, *P* < 0.05). In the AN clade, the top-ten loci (SV = 0.4309 to 0.6177) mainly located in LSC region

(eight in LSC and two in SSC) (Fig. 2A). And in the SA clade, only two mutational hotspots were identical with the AN clade (*ccsA-ndhD* and *clpP-ex1-psbB*) (Fig. 2B), which suggested that the evolution of non-coding regions between two clades were disproportional.

Geographic isolation verification

The aligned cp genomes and ITS sequences were connected to test hybridization. The ten AN *Bulbophyllum* orchids were set as the BAN branch, the seven SA *Bulbophyllum* orchids were set as the BSO branch, two *D. huoshanense* species were set as the DHU branch, four *D. moniliforme* species were set as the DMO branch and five *D. officinale* species were set as the DOF branch (Fig. S4). The JML results showed that no hybridization occurred between BAN branch and BSO branch. These results proved that there was no gene flow between two clades, which may due to the long periods of geographic isolation.

Fig. 2 Sequence variabilities of non-coding loci among AN and SA clades. Dashed lines indicate top ten hotspots. Blue points indicate common top ten hotspots of two clades

Yang *et al. BMC Plant Biology*   (2022) 22:201

Page 6 of 14

Phylogenetic analysis and divergence time estimation

The phylogenetic relationships among 27 *Bulbophyllum* species were reconstructed by 61 angiosperm cp genomes (Table S1). The ML tree exhibited identical topology with the BI tree. As shown in the Fig. 3B, the phylogenetic trees revealed a monophyletic relationship for the orchids with strongly support (bootstrap values = 100/100). Meanwhile, the *Bulbophyllum* achieved a monophyly with 100/100 bootstrap values, and was sister to *Dendrobium*. As expected, the 27 *Bulbophyllum* orchids were divided into the AN clade and SA clade. These results were consistent with previous studies [4, 18, 25].

ITS sequences and three cp genes (*psbA*, *matK* and *rbcL*) were used to reconstructed the phylogenetic relationships of 27 orchids (including *Bulbophyllum*, *Dendrobium* and *Phalaenopsis*) (Fig. S5). The results supported that 27 *Bulbophyllum* orchids were strictly divided into two clades. In addition, we found that neither the ITS tree nor the ITS + *psbA* + *matK* + *rbcL* tree was comparable to cp genome tree in bootstrap values. Obviously, the cp genomes have a better performance in studying plant phylogenetic relationships.

The divergence times of 61 orchids were estimated using BEAST2. The topologies of the BEAST tree were almost consistent to ML and BI trees except for *B. affine* and *B. reptans*, which may be because reconstructing phylogenetic relationships at the angiosperm-scale hid SNPs that can distinguish between close species (Fig. 3A). In this study, *Bulbophyllum* was inferred to have diverged from *Dendrobium* at 36.77 (32.15–41.39) million years ago (mya), and the *Bulbophyllum* AN clade was inferred to have diverged from SA clade at 25.81 (21.5–30.12) mya. Notably, these results were slightly higher than previous studies [18, 25], mainly because cp genome sequences are longer than ITS sequences and other short fragments.

Cp genome structural variation analysis

The insertions and deletions (InDels) were identified in 27 *Bulbophyllum* cp genomes, with *D. huoshanense* as outgroup. There were more deletions (5,314 to 14,498 bp) than insertions (4,216 to 10,652 bp) in *Bulbophyllum* (Fig. 4A). In addition, the distribution of InDels were different between AN and SA clades. For example, (i) the InDels lengths were varied between two clades. The total InDels lengths of AN clade (5,540 to 10,652 bp and 5,314 to 14,498 bp) were higher than SA clade (4,216 to 6,801 bp and 8,466 to 12,642 bp) (Fig. 4A). (ii) Except for partial AN *Bulbophyllum*

Fig. 3 Chronogram and Phylogenetic relationships of 61 angiosperms. **a** Molecular dating results of 61 angiosperms by BEAST2; **b** ML tree topology with ML and BI bootstrap values. The first one represents ML bootstrap value and the second one represents BI bootstrap value. * represents 100 bootstrap value

Fig. 4 InDels analysis results of *Bulbophyllum* orchids. **a** Total InDels lengths with reference *D. huoshanense*; **b** InDels distribution densities in different regions of *Bulbophyllum* orchids; **c** InDels occurrence rates of different branches in *Bulbophyllum*

species, the InDels distribution densities in SSC region were higher than in LSC and IR regions (Fig. 4B), which indicated that SSC region structure was diversified among two clades. (iii) In addition to InDels distribution densities in different regions, the occurrence rates of InDels between AN and SA clades were also inconsistent (Fig. 4C). The InDels occurrence rates of AN clade (473.01 and 246.90 bp/myr) were higher than SA clade (127.38 and 213.95 bp/myr). These results revealed that the evolution of cp genome structures between two clades were asymmetry.

Evolution analysis of protein-coding genes

The synonymous (dS) and non-synonymous (dN) substitution rates for each protein-coding genes were estimated using codonML, with *D. huoshanense* as reference. The values of dS (0–0.1973) were higher than dN (0–0.0451) in *Bulbophyllum*. Although most of ω values between two clades were similar (Fig. 5A), there were some significantly different protein-coding genes, such as *atpA*, *psbZ*, *rpl16* and *atpE* (Fig. 5B, Table S2). These genes, which mainly contained photosynthesis and self-replication functions, were mostly distributed

Yang *et al. BMC Plant Biology*   (2022) 22:201

Page 8 of 14

Fig. 5 Selective pressure analysis results in *Bulbophyllum*. **a** Scatter diagram for dN and dS of 63 cp protein-coding genes from 27 *Bulbophyllum* orchids with reference *D. huoshanense*; **b** Scatter diagram for dN and dS of *atpA*, *psbZ*, *rpl16* and *atpE*; **c** ML tree with positively selective genes in *Bulbophyllum* branch, AN branch and SA branch, respectively

in SC regions (14 in SC regions, two in IR regions) (Table S3). These results indicated that a part of protein-coding genes between two clades were under different selective pressures.

The Branch model of codeML were used to investigate the possible role of positive selection in driving protein-coding gene evolution among three branches (*Bulbophyllum* branch, AN branch and SA branch). The comparison showed that a part of genes was under different selective pressures. For example, there were six genes (*psbB*, *psbF*, *rbcL*, *rpl23*, *ycf3* and *ycf2*) were found with positive selection in the *Bulbophyllum* branch; and five genes (*matK*, *psbJ*, *ycf2*, *psbZ* and *psbF*) were found in the AN branch; and three genes (*rpl2*, *rpl36* and *ycf2*) were found in the SA branch (Fig. 5C). Most of them were in the LSC region (Table S4). The positively selective protein-coding genes between AN and SA branches were different,

which suggested that different environments diversified the evolution of the protein-coding genes.

Discussion

Chloroplast genomes of *Bulbophyllum* Asian and South American clades were different

Comparative cp genome analyses have revealed that cp genomes are conserved in genome sizes, structures, and gene contents [26–28]. Because evolutionary rates of cp genomes are slower than nuclear and mitochondrial genomes [3, 29], complete cp genomes are useful resources for phylogenetic analysis and species identification [30–32]. With the rapid development of cp genome researches, there have been comprehensive understandings about the cp genome characteristics. For example, (i) cp genome substitution rates are considerably lower in IR regions than in SC regions [3, 11, 12]; (ii) the repeat

Yang *et al. BMC Plant Biology*   (2022) 22:201

Page 9 of 14

sequences played an important role in rearrangements of cp genomes [13–15]; (iii) mutational hotspots are genus specific [33, 34]. However, the differences of the cp genomes in different continents have not been analyzed. The genus *Bulbophyllum*, which is widely distributed in Asia, south America and other places [17, 18], is an excellent source for the comparison of cp genome features, cp genome junctions and mutational hotspots.

We compared the overall structures and variations of *Bulbophyllum* cp genomes. All results showed that there were considerable differences in overall cp genome features between AN and SA clades. Firstly, they have different cp genome features. For example, (i) the overall, LSC and SSC GC contents of AN clade are higher than SA clade; (ii) part of the *rpl32* orientations in AN clade were different from SA clade, which caused by an inversion; (iii) except LSC LDRs and IR SSRs distribution densities, the LDRs and SSRs distribution densities in SA clade were higher than in AN clade. Secondly, J_{SB}s between two clades varied significantly. AN and SA clades had different types of J_{SB}s. There was no type B J_{SB}s in AN clade and there were more type C J_{SB}s in SA clade. Thirdly, the distributions and kinds of mutational hotspots between two clades were significantly different. Niu et al. (2017) found specific barcodes searched by 8–10 closely related cp genomes from specific genus can uncover their cp genome features [33]. Therefore, the mutational hotspots of two clades estimated in this study can reflect the variations of non-coding loci. However, there were only two hotspots shared between two clades, which indicated that the AN and SA clades had discordant evolution in noncoding regions. In conclusion, AN and SA clades had considerable differences in overall cp genome features.

The cp genome structures between Asian and South American clades were different under geographic isolation
Hybridization, a driver of plant evolution, is common among terrestrial orchids [35]. Although dust-like seeds have the capacity to travel long distance [36–38], the results of this study showed that *Bulbophyllum* AN clade has diverged completely from SA clade in the late Oligocene (21.50–30.12 mya) [18]. Because the results of the JML software have ruled out the possibility of hybridization between the two clades [39], *Bulbophyllum* AN and SA clades have evolved independently for at least 25.81 million years. Our subsequent analyses revealed significant cp genome structural variations between two clades.

The insertions and deletions (InDels) are important structural variations of cp genomes [32]. In order to evaluate the structural variations of the cp genomes during the evolution of *Bulbophyllum*, we calculated the total InDels lengths of each *Bulbophyllum* by the aligned complete cp genome sequences, with *D. huoshanense* as outgroup. There were 5,314 to 14,498 bp deletions and 4,216 to 10,652 bp insertions in *Bulbophyllum*. Obviously, the lengths of deletions were longer than insertions, which is consistent with previous studies [16, 40]. To obtain more detailed information about InDels, we calculated the distribution densities of InDels in different cp genome regions. The distribution densities of insertions in LSC, IR, and SSC regions were 22.7–48.9 bp/kbp, 4.8–46.1 bp/kbp, and 29.4–286.9 bp/kbp, respectively. And the deletions distribution densities in LSC, IR and SSC regions were 37.8–130.5 bp/kbp, 16.2–92.3 bp/kbp, and 24–389.1 bp/kbp, respectively. The results revealed that the distributions of InDels are dependent on their locations in cp genomes [16, 41, 42], e.g., associated with low GC contents.

To estimate the pace of structural variations in the *Bulbophyllum* cp genomes, we dated the occurrence rates of InDels by located them at the BEAST tree branches. The results showed that, during the past ~10.96 million years since the late Eocene, the insertions and deletions occurred at the speeds of 67.15 bp/myr and 194.07 bp/myr. Then, AN clade and SA clade diverged at ~25.81 mya. Our results further revealed that InDels remarkably varied across different lineages [40]. Nevertheless, the most InDels occurrence rates of AN *Bulbophyllum* were higher than SA *Bulbophyllum*. We speculated that there were two main reasons for the higher occurrence rates of InDels in the AN clade: (1) the results of Gamisch et al. (2021) indicated that Asia was the original continent of the *Bulbophyllum* orchids [43]; (2) Of all the continents where *Bulbophyllum* orchids grow, Asia has the largest suitable area for speciation and niche expansion of the genus *Bulbophyllum* [43]. In conclusion, the cp genome structures between two clades were different under geographic isolation. And for further study, InDels may be an important indicator of species evolution.

Crassulacean acid metabolism (CAM) pathway may be the inducement of cp gene differentiation
Crassulacean acid metabolism (CAM) is the typical plant photosynthetic pathway that improves water use efficiency of plants [44]. About 10% of orchid species have been confirmed to be CAM plants [45–47], and 50% of Orchidaceae species were anticipated to be CAM plants [48]. CAM has evolved multiple independent times in the Orchidaceae [49]. Phylogenetic analyses of the neotropical orchids indicated that CAM has evolved independently at least 10 times [48]. For example, CAM in *Dendrobium*, the sister genus of *Bulbophyllum*, has arisen for at least eight times [49]. The *Bulbophyllum* orchids are typical CAM orchids. Although the C_3 pathway is the main photosynthetic

Yang *et al. BMC Plant Biology*   (2022) 22:201

Page 10 of 14

pathway of *Bulbophyllum* presently, the CAM pathway plays an important role in adapting to new ecological niches and increasing species diversity [43].

In order to evaluate the effects of CAM pathway on cp gene evolution in the genus *Bulbophyllum*, Selective pressure analysis of functional genes has been employed [50–52]. Non-synonymous (dN), synonymous (dS) nucleotide substitution rates and non-synonymous/synonymous rate ratio ($\omega = dN/dS$) are widely used as indicators of adaptive evolution [40]. The values of dN and dS on all protein-coding genes between two clades were compared, and we found three photosynthesis genes, ten self-replication genes and three other genes, which have significantly different ω values. The three photosynthesis genes (*psbZ*, *atpE* and *atpA*) encode fractions of the photosystems and the thylakoid ATP synthase complex. In addition, positively selective genes (ω values > 1) through the *Bulbophyllum* evolution were screened. There were six, five and three positively selective cp genes in the *Bulbophyllum* branch, AN branch and SA branch, respectively. And these genes were mostly distributed in the LSC regions, which is consistent with previous studies [40]. Because there were many photosynthesis genes, which are components of Photosystem II, among two clades under different selective pressures, the CAM pathway may play an important role in adaptive evolution of *Bulbophyllum*. In addition, Hu et al. (2022) assessed the origin and loss of CAM in the *Bulbophyllum* since middle Miocene using two cp genes (*matK* and *psbA*) and two nuclear sequences (ITS and *Xdh*) [53]. They found that although CAM increased the adaptability of the *Bulbophyllum* plants to specific conditions, it was less competitive than C$_3$ in the long term. Predictably, the role of CAM in the species evolution of *Bulbophyllum* can be further explored using the complete cp genomes.

Before the ancestors of the *Bulbophyllum* diverged from *Dendrobium* at late Eocene, CAM and epiphytism in orchids are likely to have originated about 65 mya [54, 55]. Then, some of the *Bulbophyllum* photosynthetic genes (*psbB*, *psbF* and *rbcL*) had evolved to adapt to the changeable environment and expand habitat region. At late Oligocene, AN clade diverged from other *Bulbophyllum* clade, through which SA clade had no gene flow with AN clade. The cp genomes of two clades evolved to different features under different environments. In this process, the CAM pathway affects the selective pressures of photosynthetic genes. Therefore, in subsequent studies, CAM studies and cp genome studies will be more closely linked to explore species evolution of *Bulbophyllum*.

Conclusions

In this study, the nine *Bulbophyllum* cp genomes were newly sequenced and assembled. Comparative analyses showed that there were considerable differences in overall cp genome features between AN and SA clades in three aspects, including basic cp genome features, JS$_B$s and mutational hotspots. Then we found that the cp genome structures between two clades were different under geographic isolation by estimating the occurrence rates of the InDels. Finally, the results showed that long-term geographic isolation put AN and SA *Bulbophyllum* cp genes under different selective pressures. All results indicated that the overall structural characteristics of *Bulbophyllum* cp genomes diverged during the long-term geographic isolation, and the CAM pathway may play an important role in the cp genes evolution of *Bulbophyllum*.

Methods

Plant materials and DNA extraction

In this study, nine medicinal *Bulbophyllum* orchids sequenced (*B. forrestii* Seidenf. (voucher specimen: Yang202101), *B. leopardinum* (Wall.) Lindl. (voucher specimen: Yang202102), *B. odoratissimum* Lindl. (voucher specimen: Yang202103), *B. orientale* Seidenf. (voucher specimen: Yang202104), *B. hirtum* (J. E. Smith) Lindl. (voucher specimen: Yang202105), *B. reptans* Lindl. (voucher specimen: Yang202106), *B. kwangtungense* Schltr. (voucher specimen: Yang202107), *B. hirundinis* (Gagnep.) Seidenf. (voucher specimen: Yang202108) and *B. andersonii* (Hook. f.) J. J. Smith (voucher specimen: Yang202201)) were collected in Yunnan and Fujian provinces, China, and stored in College of Life Sciences, Nanjing Normal University, Nanjing, China. All the samples were identified by Prof. Dr. Xiaoyu Ding. Total genomic DNA of each sample was extracted from 0.2 g fresh leaves using Dneasy Plant Mini Kits (QIAGEN, Germany). DNA samples that met the quality requirements (A260/280 ratio = 1.8–2.0, A260/230 ratio > 1.7, and DNA concentration > 300 ng/μL) were used for sequencing. In addition, ten published AN *Bulbophyllum* (*B. disciflorum*, *B. gedangense*, *B. lingii*, *B. menghaiense*, *B. pentaneurum*, *B. pingnanense*, *B. inconspicuum*, *B. affine*, *B. pectinatum* and *B. funingense*) and eight SA *Bulbophyllum* (*B. epiphytum*, *B. exaltatum*, *B. granulosum*, *B. mentosum*, *B. plumosum*, *B. regnellii*, *B. steyermarkii* and *B. weddellii*) cp genomes were selected for subsequent comparative genome analysis.

DNA sequencing, assembly and annotation

The total genomic DNA of nine *Bulbophyllum* orchids was sequenced using Illumina Hiseq4000 platform

(Illumina Inc, San Diego, CA, USA). Approximately 5.0 Gb of raw data were generated with 150 bp paired-end reads for each sample. The fragments with coverage < 50 × were deleted and filtered paired-end reads were assembled on CLC Genomics Workbench v8.5.1 (CLC Bio, Aarhus, Denmark), with reference *Dendrobium huoshanense* C. Z. Tang et S. J. Cheng. The four junctions between LSC/SSC and IRs were validated with PCR amplification. The assembled cp genomes were annotated using DOGMA v1.2 and tRNAscan-SE v1.21 [56, 57]. The boundaries of annotated genes were confirmed by multiple sequence alignment.

Comparative analysis of cp genomes

The gene location information was extracted, and the IR/SC junction information was collated with the gene location information. The GC contents of 27 *Bulbophyllum* orchids were analyzed. The nine newly sequenced *Bulbophyllum* cp genomes were compared using online mVISTA (https://genome.lbl.gov/vista/mvista/submit.shtml) on LAGAN model, with reference *D. huoshanense* [58].

Characterization of repeat sequences and simple sequence repeats

The REPuter program (https://bibiserv.cebitec.uni-bielefeld.de/reputer) was used to identify the four different types of repeats (forward, reverse, complement and palindromic repeats), with sequence identity ≥ 90%, minimum repeat size ≥ 30 bp, and Hamming distance = 3 [59]. Simple sequence repeats (SSRs) were discovered using the online program MISA (https://webblast.ipk-gatersleben.de/misa/) with the following parameters: 8 for mononucleotide motifs, 5 for polynucleotide motifs [60]. Then the distribution densities of repeat sequences and SSRs on LSC, SSC and IR regions were compared.

IR/SC junctions map

Refer to the method of Zhu et al. (2018) [42], the IR/SC junctions map of 19 AN *Bulbophyllum* orchids, eight SA *Bulbophyllum* orchids and *D. huoshanense* was drawn.

Mutational hotspots

Based on 19 AN *Bulbophyllum* cp genomes, the top ten hotspots in AN clade were screened. Based on eight SA *Bulbophyllum* cp genomes, the top ten mutational hotspots in SA clade were screened. The gene location information was transformed into the non-coding region location information, and the sequence of each non-coding region was extracted (*ycf1-rpl32* was taken as a whole fragment due to the loss of *ndhF*). Each non-coding region was aligned with MEGA 5.0 [61], and gaps at both ends were deleted. SNPs, InDels, and conserved sites were calculated by DnaSP v5 [62]. Non-coding regions, of which conserved sites were less than 100 bp, were removed. Sequence variabilities were calculated as follows: Sequence Variabilities = (SNPs + InDels) / (SNPs + InDels + Conserved sites) × 100% [34, 63].

Hybridization testing

The cp genomes and ITS sequences of 17 *Bulbophyllum* orchids, two *D. huoshanense*, four *D. moniliforme* and five *D. officinale* were aligned using MAFFT 7.220 [64], and were connected together. Under the rule of AIC (Akaike information criterion), the optimum base substitution model calculated by Modeltest 3.7 was GTR + I + Γ [65]. The species tree was constructed using BEAST2 [66]. The Markov Chain Monte Carlo (MCMC) chains were performed to check convergence with 10,000,000 generations. The probability of hybridization was calculated using JML [67]. The significance level was 0.01.

Phylogenetic analysis

Phylogenetic relationships were analyzed based on 61 complete cp genomes, including one dicotyledonous plant, eight monocotyledonous non-orchids, and 52 orchids (Table S1). The cp genome sequences were aligned using MAFFT 7.220 [64]. The conserved loci were selected by Gblocks v.0.91b with the following parameters: minimum number of sequences for a conserved position: 37, minimum number of sequences for a flank position: 62, maximum number of contiguous non-conserved positions: 8, minimum length of a block: 10, allowed gap positions: none [68]. Under the rule of AIC (Akaike information criterion), the optimum base substitution model calculated by Modeltest 3.7 was GTR + I + Γ [65]. The Maximum Likelihood (ML) phylogenetic tree was constructed using RAxML (version 7.4.2) with 1000 rapid bootstrap inferences [69]. The outgroup was *Arabidopsis thaliana*. The BI analysis was made by using MrBayes 3.2.7 with 1,000,000 generations [70]. Trees were sampled every 1,000 generations, and the first 25% of these were discarded. The remaining trees were used to build the Bayesian tree of posterior probabilities. In addition, ITS, *psbA*, *matK*, *rbcL* and cp genomes of 27 orchids were used to reconstructed ML trees with 1000 rapid bootstrap inferences.

Divergence time estimation

Divergence times were estimated by BEAST2 [66]. The following constraints were used for time calibrations: (1) The root age was set as 194 million years ago (Mya) (prior distribution: normal, mean: 194, sd: 8) [71–73]. (2) The stem age of Lemnoideae was set as 62 Mya (prior distribution: normal, mean: 62, sd: 3) [74]. (3) Based on the fossil record of *Dendrobium*, the Crown age of *Dendrobium* (the divergence time between the

Yang *et al. BMC Plant Biology*   (2022) 22:201

Page 12 of 14

Dendrobium Asian clade and the Australian clade) was set at 21 Mya (prior distribution: normal, mean: 21, sd: 5) [25, 75]. Three independent MCMC were performed to check convergence, each with 10,000,000 generations. LogCombiner was used to integrate the data results of three independent runs and remove the top 10% of unstable data.

Structural variation analysis of *Bulbophyllum* cp genomes

The cp genome sequences of 27 *Bulbophyllum* orchids and *D. huoshanense* were aligned using MAFFT 7.220 [64]. The InDels of each *Bulbophyllum* cp genome were counted, with *D. huoshanense* as reference. Then, the common InDels of 27 *Bulbophyllum* cp genomes were counted to calculate the InDels occurrence rates of AN, SA and *Bulbophyllum* branches. The above common InDels were excluded to calculate the recent occurrence rates of InDels in each *Bulbophyllum* cp genome.

Nucleotide substitutions rates and positive selection analysis of protein-coding genes

The molecular evolution of 63 cp protein-coding genes were investigated from 27 *Bulbophyllum* orchids and *D. huoshanense.* The values of dN, dS and ω were estimated by the codonML in the PAML software (version 4.4) with reference *D. huoshanense* [76, 77]. Chloroplast protein-coding genes, whose ω values were different between two clades (T-test for Independent Samples) significantly, were screened. EasyCodeML was used to detect positively selective genes in the *Bulbophyllum* branch, AN branch and SA branch under Branch model with 5% significance level, respectively [40].

Abbreviations

Cp: Chloroplast; AN: Asian; SA: South American; InDels: Insertions and Deletions; CAM: Crassulacean Acid Metabolism; LSC: Large Single Copy; SSC: Small Single Copy; IR: Inverted Repeat; ITS: Internal Transcribed Spacer; LDRs: Longer Dispersed Repeats; SSRs: Simple Sequence Repeats; J_{SA}s: SSC/IR_A junctions; J_{SB}s: SSC/IR_B junctions; J_{LB}s: LSC/IR_B junctions; J_{LA}s: LSC/IR_A junctions; SVs: Sequence Variabilities; mya: Million Years Ago; ML: Maximum Likelihood.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-022-03592-y.

Additional file 1: Fig S1. mVISTA analysis results of nine newly sequenced *Bulbophyllum* orchids with reference *D. huoshanense*. Chloroplast coding regions are indicated in blue, non-translation regions in cyan, andnon-coding regions in red.

Additional file 2: Fig S2. Distribution densities of LDRs and SSRs in 27 *Bulbophyllum* orchids and *D. huoshanense*. **a** Distribution density of LDRs; **b** Distribution density of SSRs.

Additional file 3: Fig S3. IR/SC junctions map of 19 AN *Bulbophyllum*, eight SA *Bulbophyllum* and *D. huoshanense*. Yellow represents the

---

rpl22 gene, blue represents the *ycf1* gene, red represents the *ndhF* pseudogene and green represents the *psbA* gene.

Additional file 4: Fig S4. The species tree constructed by BEAST2.

Additional file 5: Fig S5. Phylogenetic relationships of 27 orchids. **a** ML tree reconstructed by ITS sequences; **b** MLtree reconstructed by ITS sequences, *psbA*, *matK* and *rbcL*; **c** ML tree reconstructed by cp genomes. * represents 100 bootstrap value.

Additional file 6: Table S1. The species information of 61 angiosperms used in the phylogenetic analysis

Additional file 7: Table S2. dN and dS of 16 screened protein-coding genes in *Bulbophyllum*.

Additional file 8: Table S3. The basic information of 16 screened protein-coding genes.

Additional file 9: Table S4. The basic information of positively selective genes.

Declarations

Ethics approval and consent to participate
The collected *Bulbophyllum* species are distributed in Yunnan, China. Experimental researches with *Bulbophyllum* species comply with Nanjing Normal University guidelines (http://bwc.njnu.edu.cn/info/1085/1433.htm), preserving the genetic background of the species used. It does not require ethical approval.

Conflicts of interest
The authors have no conflicts of interest to declare.

Author details
[1]College of Life Sciences, Nanjing Normal University, Nanjing 210023, China. [2]Jiangsu Provincial Engineering Research Center for Technical Industrialization for Dendrobiums, Nanjing 210023, China.

References
1. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: from gene to genome. Biol Rev Camb Philos Soc. 2015;90(1):157–66.
2. Feng S, Zheng K, Jiao K, Cai Y, Chen C, Mao Y, Wang L, Zhan X, Ying Q, Wang H. Complete chloroplast genomes of four *Physalis* species

Yang *et al. BMC Plant Biology*   (2022) 22:201

Page 13 of 14

(Solanaceae): lights into genome structure, comparative analysis, and phylogenetic relationships. BMC Plant Biol. 2020;20(1):242.

3. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci U S A. 1987;84(24):9054–8.

4. Kim YK, Jo S, Cheon SH, Joo MJ, Hong JR, Kwak M, Kim KJ. Plastome Evolution and Phylogeny of Orchidaceae, With 24 New Sequences. Front Plant Sci. 2020;11:22.

5. Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal JR, Kuehl JV, Boore JL. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci U S A. 2007;104(49):19369–74.

6. Dyall SD, Brown MT, Johnson PJ. Ancient invasions: from endosymbionts to organelles. Science. 2004;304(5668):253–7.

7. Palmer JD, Nugent JM, Herbon LA. Unusual structure of geranium chloroplast DNA: A triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. Proc Natl Acad Sci U S A. 1987;84(3):769–73.

8. Kim KJ, Lee HL. Widespread occurrence of small inversions in the chloroplast genomes of land plants. Mol Cells. 2005;19(1):104–13.

9. Greiner S, Wang X, Rauwolf U, Silber MV, Mayer K, Meurer J, Haberer G, Herrmann RG. The complete nucleotide sequences of the five genetically distinct plastid genomes of Oenothera, subsection Oenothera: I. sequence evaluation and plastome evolution. Nucleic Acids Res. 2008;36(7):2366–78.

10. Magee AM, Aspinall S, Rice DW, Cusack BP, Sémon M, Perry AS, Stefanović S, Milbourne D, Barth S, Palmer JD, Gray JC, Kavanagh TA, Wolfe KH. Localized hypermutation and associated gene losses in legume chloroplast genomes. Genome Res. 2010;20(12):1700–10.

11. Smith DR. Mutation rates in plastid genomes: they are lower than you might think. Genome Biol Evol. 2015;7(5):1227–34.

12. Wu CS, Chaw SM. Evolutionary Stasis in Cycad Plastomes and the First Case of Plastome GC-Biased Gene Conversion. Genome Biol Evol. 2015;7(7):2000–9.

13. Asaf S, Khan AL, Khan AR, Waqas M, Kang SM, Khan MA, Lee SM, Lee IJ. Complete Chloroplast Genome of *Nicotiana otophora* and its Comparison with Related Species. Front Plant Sci. 2016;7:843.

14. Choi KS, Chung MG, Park S. The Complete Chloroplast Genome Sequences of Three Veroniceae Species (Plantaginaceae): Comparative Analysis and Highly Divergent Regions. Front Plant Sci. 2016;7:355.

15. Zhang Y, Du L, Liu A, Chen J, Wu L, Hu W, Zhang W, Kim K, Lee SC, Yang TJ, Wang Y. The Complete Chloroplast Genome Sequences of Five *Epimedium* Species: Lights into Phylogenetic and Taxonomic Analyses. Front Plant Sci. 2016;7:306.

16. Zhitao N, Shuying Z, Jiajia P, Ludan L, Jing S, Xiaoyu D. Comparative analysis of *Dendrobium* plastomes and utility of plastomic mutational hotspots. Sci Rep. 2017;7(1):2073.

17. Zavala-Páez M, Vieira LDN, de Baura VA, Balsanelli E, de Souza EM, Cevallos MC, Chase MW, Smidt EC. Comparative Plastid Genomics of Neotropical *Bulbophyllum* (Orchidaceae; Epidendroideae). Front Plant Sci. 2020;11:799.

18. Gamisch A, Comes HP. Clade-age-dependent diversification under high species turnover shapes species richness disparities among tropical rainforest lineages of *Bulbophyllum* (Orchidaceae). BMC Evol Biol. 2019;19(1):93.

19. Ya JD, Guo YJ, Liu C, Cai J, Dong GJ, Jiang H, Li DZ. *Bulbophyllum reflexipetalum* (Orchidaceae, Epidendroideae, Malaxideae), a new species from Xizang. China PhytoKeys. 2019;130:33–9.

20. Gamisch A, Fischer GA, Comes HP. Multiple independent origins of auto-pollination in tropical orchids (*Bulbophyllum*) in light of the hypothesis of selfing as an evolutionary dead end. BMC Evol Biol. 2015;15:192.

21. Tang Y, Yang J, Niu Z, Ding X. The complete chloroplast genome sequence of a traditional Chinese medicine plant *Bulbophyllum disciflorum* Rolfe (Orchidaceae). Mitochondrial DNA B Resour. 2019;5(1):59–60.

22. Tang H, Tang L, Shao S, Peng Y, Li L, Luo Y. Chloroplast genomic diversity in *Bulbophyllum* section *Macrocaulia* (Orchidaceae, Epidendroideae, Malaxideae): Insights into species divergence and adaptive evolution. Plant Divers. 2021;43(5):350–61.

23. Zhang S, Zeng MY, Gao XY, Li MH, Chen SP. The complete plastome of *Bulbophyllum pingnanense* (Orchidaceae: Dendrobiinae). Mitochondrial DNA B Resour. 2021;6(11):3193–5.

24. Yang J, Zhu Z, Fan Y, Zhu F, Chen Y, Niu Z, Ding X. Comparative plastomic analysis of three *Bulbophyllum* medicinal plants and its significance in species identification. Acta Pharmaceutica Sinica. 2020;55(11):2736–45.

25. Xiang X, Mi X, Zhou H, Li J, Chung S, Li D, Huang W, Jin W, Li Z, Huang L, Jin X. Biogeographical diversification of mainland Asian *Dendrobium* (Orchidaceae) and its implications for the historical dynamics of evergreen broad-leaved forests. J Biogeogr. 2016;43(7):1310–23.

26. Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, Chen WH, Cheng CH, Lin CY, Liu SM, Chang CC, Chaw SM. The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. Mol Biol Evol. 2006;23(2):279–91.

27. Jheng CF, Chen TC, Lin JY, Chen TC, Wu WL, Chang CC. The comparative chloroplast genomic analysis of photosynthetic orchids and developing DNA markers to distinguish *Phalaenopsis* orchids. Plant Sci. 2012;190:62–73.

28. Yang JB, Tang M, Li HT, Zhang ZR, Li DZ. Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. BMC Evol Biol. 2013;13:84.

29. Drouin G, Daoud H, Xia J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol Phylogenet Evol. 2008;49(3):827–31.

30. Tonti-Filippini J, Nevill PG, Dixon K, Small I. What can we do with 1000 plastid genomes? Plant J. 2017;90(4):808–18.

31. Carbonell-Caballero J, Alonso R, Ibañez V, Terol J, Talon M, Dopazo J. A Phylogenetic Analysis of 34 Chloroplast Genomes Elucidates the Relationships between Wild and Domestic Species within the Genus *Citrus*. Mol Biol Evol. 2015;32(8):2015–35.

32. Cho KS, Yun BK, Yoon YH, Hong SY, Mekapogu M, Kim KH, Yang TJ. Complete Chloroplast Genome Sequence of Tartary Buckwheat (Fagopyrum tataricum) and Comparative Analysis with Common Buckwheat (F esculentum). PLoS One. 2015;10(5):e0125332.

33. Niu Z, Xue Q, Zhu S, Sun J, Liu W, Ding X. The Complete Plastome Sequences of Four Orchid Species: Insights into the Evolution of the Orchidaceae and the Utility of Plastomic Mutational Hotspots. Front Plant Sci. 2017;8:715.

34. Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, Morris AB. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. Am J Bot. 2014;101(11):1987–2004.

35. Piett S, Hager HA, Gerrard C. Characteristics for evaluating the conservation value of species hybrids. Biodivers Conserv. 2015;24(8):1931–55.

36. Li X, Ding X, Chu B, Zhou Q, Ding G, Gu S. Genetic diversity analysis and conservation of the endangered Chinese endemic herb *Dendrobium officinale* Kimura et Migo (Orchidaceae) based on AFLP. Genetica. 2008;133(2):159–66.

37. Fay MF, Pailler T, Dixon KW. Orchid conservation: making the links. Ann Bot. 2015;116(3):377–9.

38. Niu Z, Hou Z, Wang M, Ye M, Zhang B, Xue Q, Liu W, Ding X. A comparative plastomics approach reveals available molecular markers for the phylogeographic study of *Dendrobium huoshanense*, an endangered orchid with extremely small populations. Ecol Evol. 2020;10(12):5332–42.

39. Barton NH. The role of hybridization in evolution. Mol Ecol. 2001;10(3):551–68.

40. Gao LZ, Liu YL, Zhang D, Li W, Gao J, Liu Y, Li K, Shi C, Zhao Y, Zhao YJ, Jiao JY, Mao SY, Gao CW, Eichler EE. Evolution of *Oryza* chloroplast genomes promoted adaptation to diverse ecological habitats. Commun Biol. 2019;2:278.

41. Ahmed I, Biggs PJ, Matthews PJ, Collins LJ, Hendy MD, Lockhart PJ. Mutational dynamics of aroid chloroplast genomes. Genome Biol Evol. 2012;4(12):1316–23.

42. Zhu S, Niu Z, Xue Q, Wang H, Xie X, Ding X. Accurate authentication of *Dendrobium officinale* and its closely related species by comparative analysis of complete plastomes. Acta Pharm Sin B. 2018;8(6):969–80.

43. Gamisch A, Winter K, Fischer GA, Comes HP. Evolution of crassulacean acid metabolism (CAM) as an escape from ecological niche conservatism in Malagasy *Bulbophyllum* (Orchidaceae). New Phytol. 2021;231(3):1236–48.

Yang *et al. BMC Plant Biology*   (2022) 22:201

Page 14 of 14

44. Gilman IS, Edwards EJ. Crassulacean acid metabolism. Curr Biol. 2020;30(2):R57–62.
45. Winter K, Wallace BJ, Stocker GC, Roksandic Z. Crassulacean acid metabolism in australian vascular epiphytes and some related species. Oecologia. 1983;57(1–2):129–41.
46. Earnshaw MJ, Winter K, Ziegler H, Stichler W, Cruttwell NE, Kerenga K, Cribb PJ, Wood J, Croft JR, Carver KA, Gunn TC. Altitudinal changes in the incidence of crassulacean acid metabolism in vascular epiphytes and related life forms in Papua New Guinea. Oecologia. 1987;73(4):566–72.
47. Kluge M, Vinson B, Ziegler H. Ecophysiological Studies on Orchids of Madagascar: Incidence and Plasticity of Crassulacean Acid Metabolism in Species of the Genus *Angraecum Bory*. Plant Ecol. 1998;135(1):43–57.
48. Silvera K, Santiago LS, Cushman JC, Winter K. Crassulacean acid metabolism and epiphytism linked to adaptive radiations in the Orchidaceae. Plant Physiol. 2009;149(4):1838–47.
49. Li MH, Liu DK, Zhang GQ, Deng H, Tu XD, Wang Y, Lan SR, Liu ZJ. A perspective on crassulacean acid metabolism photosynthesis evolution of orchids on different continents: *Dendrobium* as a case study. J Exp Bot. 2019;70(22):6611–9.
50. Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Stern-berg SM, Margulies EH, Blanchette M, Siepel AC, Thomas PJ, McDowell JC, Maskeri B, Hansen NF, Schwartz MS, Weber RJ, Kent WJ, Karolchik D, Bruen TC, Bevan R, Cutler DJ, Schwartz S, Elnitski L, Idol JR, Prasad AB, Lee-Lin SQ, Maduro VV, Summers TJ, Portnoy ME, Dietrich NL, Akhter N, Ayele K, Benjamin B, Cariaga K, Brinkley CP, Brooks SY, Granite S, Guan X, Gupta J, Haghighi P, Ho SL, Huang MC, Karlins E, Laric PL, Legaspi R, Lim MJ, Maduro QL, Masiello CA, Mastrian SD, McCloskey JC, Pearson R, Stantri-pop S, Tiongson EE, Tran JT, Tsurgeon C, Vogt JL, Walker MA, Wetherby KD, Wiggins LS, Young AC, Zhang LH, Osoegawa K, Zhu B, Zhao B, Shu CL, De Jong PJ, Lawrence CE, Smit AF, Chakravarti A, Haussler D, Green P, Miller W, Green ED. Comparative analyses of multi-species sequences from targeted genomic regions. Nature. 2003;424(6950):788–93.
51. Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, Sninsky J, Adams MD, Cargill M. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 2005;3(6):e170.
52. Sawyer SL, Wu LI, Emerman M, Malik HS. Positive selection of primate TRI-M5alpha identifies a critical species-specific retroviral restriction domain. Proc Natl Acad Sci U S A. 2005;102(8):2832–7.
53. Hu AQ, Gale SW, Liu ZJ, Fischer GA, Saunders RMK. Diversification Slow-down in the *Cirrhopetalum* Alliance (*Bulbophyllum*, Orchidaceae): Insights From the Evolutionary Dynamics of Crassulacean Acid Metabolism. Front Plant Sci. 2022;13:794171.
54. Russell KM. On the evolutionary pathways resulting in C4 photosynthesis and Crassulacean acid metabolism (CAM). Adv Ecol Res. 1989;19:57–92.
55. Ehleringer JR, Monson RK. Evolutionary and ecological aspects of photosynthetic pathway variation. Annu Rev Ecol Syst. 1993;24:411–39.
56. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 2004;20(17):3252–5.
57. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 2016;44(W1):W54–7.
58. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004;32:W273–9.
59. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29(22):4633–42.
60. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet. 2003;106(3):411–22.
61. Kumar S, Nei M, Dudley J, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform. 2008;9(4):299–306.
62. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451–2.
63. Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot. 2005;92(1):142–66.
64. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
65. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinformatics. 1998;14(9):817–8.
66. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, Matschiner M, Mendes FK, Müller NF, Ogilvie HA, du Plessis L, Popinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu CH, Xie D, Zhang C, Stadler T, Drummond AJ. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019;15(4):e1006650.
67. Joly S. JML: testing hybridization from species trees. Mol Ecol Resour. 2012;12(1):179–84.
68. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17(4):540–52.
69. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22(21):2688–90.
70. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19(12):1572–4.
71. Zhang GQ, Liu KW, Li Z, Lohaus R, Hsiao YY, Niu SC, Wang JY, Lin YC, Xu Q, Chen LJ, Yoshida K, Fujiwara S, Wang ZW, Zhang YQ, Mitsuda N, Wang M, Liu GH, Pecoraro L, Huang HX, Xiao XJ, Lin M, Wu XY, Wu WL, Chen YY, Chang SB, Sakamoto S, Ohme-Takagi M, Yagi M, Zeng SJ, Shen CY, Yeh CM, Luo YB, Tsai WC, Van de Peer Y, Liu ZJ. The Apostasia genome and the evolution of orchids. Nature. 2017;549(7672):379–83.
72. Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pè ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quétier F, Wincker P. French-Italian Public Consortium for Grapevine Genome Characterization. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature. 2007;449(7161):463–7.
73. Ramírez SR, Gravendeel B, Singer RB, Marshall CR, Pierce NE. Dating the origin of the Orchidaceae from a fossil orchid with its pollinator. Nature. 2007;448(7157):1042–5.
74. Givnish TJ, Spalink D, Ames M, Lyon SP, Hunter SJ, Zuluaga A, Iles WJ, Clements MA, Arroyo MT, Leebens-Mack J, Endara L, Kriebel R, Neubig KM, Whitten WM, Williams NH, Cameron KM. Orchid phylogenomics and multiple drivers of their extraordinary diversification. Proc Biol Sci. 1814;2015(282):20151553.
75. Conran JG, Bannister JM, Lee DE. Earliest orchid macrofossils: Early Miocene *Dendrobium* and *Earina* (Orchidaceae: Epidendroideae) from New Zealand. Am J Bot. 2009;96(2):466–74.
76. Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol. 1998;15(5):568–73.
77. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.

Publisher's Note