

RESEARCH

Open Access



Plastome structure of 8 *Calanthe* s.l. species (Orchidaceae): comparative genomics, phylogenetic analysis

Consolata Nanjala^{1,2,3}, Vincent Okelo Wanga^{1,2,3}, Wyclif Odago^{1,2,3}, Elizabeth Syowai Mutinda^{1,2,3}, Emmanuel Nyongesa Waswa^{1,2,3}, Millicent Akinyi Oulo^{1,2,3}, Elijah Mbandi Mkala^{1,2,3}, Josiah Kuja⁴, Jia-Xin Yang^{1,2,3}, Xiang Dong^{1,2,3}, Guang-Wan Hu^{1,2,3*} and Qing-Feng Wang^{1,2,3}

Abstract

Background: *Calanthe* (Epidendroideae, Orchidaceae) is a pantropical genus distributed in Asia and Africa. Its species are of great importance in terms of economic, ornamental and medicinal values. However, due to limited and confusing delimitation characters, the taxonomy of the *Calanthe* alliance (*Calanthe*, *Cephalantheropsis*, and *Phaius*) has not been sufficiently resolved. Additionally, the limited genomic information has shown incongruences in its systematics and phylogeny. In this study, we used illumina platform sequencing, performed a *de novo* assembly, and did a comparative analysis of 8 *Calanthe* group species' plastomes: 6 *Calanthe* and 2 *Phaius* species. Phylogenetic analyses were used to reconstruct the relationships of the species as well as with other species of the family Orchidaceae.

Results: The complete plastomes of the *Calanthe* group species have a quadripartite structure with varied sizes ranging between 150,105bp-158,714bp, including a large single-copy region (LSC; 83,364bp- 87,450bp), a small single-copy region (SSC; 16,297bp -18,586bp), and a pair of inverted repeat regions (IRs; 25,222bp - 26,430bp). The overall GC content of these plastomes ranged between 36.6-36.9%. These plastomes encoded 131-134 differential genes, which included 85-88 protein-coding genes, 37-38 tRNA genes, and 8 rRNA genes. Comparative analysis showed no significant variations in terms of their sequences, gene content, gene order, sequence repeats and the GC content hence highly conserved. However, some genes were lost in *C. delavayi* (*P. delavayi*), including *ndhC*, *ndhF*, and *ndhK* genes. Compared to the coding regions, the non-coding regions had more sequence repeats hence important for species DNA barcoding. Phylogenetic analysis revealed a paraphyletic relationship in the *Calanthe* group, and confirmed the position of *Phaius delavayi* in the genus *Calanthe* as opposed to its previous placement in *Phaius*.

Conclusion: This study provides a report on the complete plastomes of 6 *Calanthe* and 2 *Phaius* species and elucidates the structural characteristics of the plastomes. It also highlights the power of plastome data to resolve phylogenetic relationships and clarifies taxonomic disputes among closely related species to improve our understanding of their systematics and evolution. Furthermore, it also provides valuable genetic resources and a basis for studying evolutionary relationships and population genetics among orchid species.

Keywords: *Calanthe*, Chloroplast genome, Genome comparison, *Phaius*, Phylogeny

*Correspondence: guangwanhu@wbgcas.cn

³ Sino-Africa Joint Research Center, Chinese Academy of Sciences, Wuhan 430074, China

Full list of author information is available at the end of the article

Background

Calanthe is the largest genus in tribe Collabieae (Epidendroideae; Orchidaceae), with more than 220 species [1], distributed across tropical and subtropical



Asia, Australia, Madagascar, Africa, Central and South America, and the Caribbean [2–4]. *Calanthe* species are evergreen or deciduous plants, terrestrial (rarely epipetric or epiphytic) with thick roots, small oval pseudobulbs, highly ridged leaves, and upright, occasionally arching flowering stems [5]. Their flowers arise from the basal leaf with showy, white, yellow, or pink colors with a resupinate opening, ranging from small, medium and large [6]. They often turn dark blue after damage or during senescence [7]. *Calanthe* is the first orchid species to be artificially used by humans for hybridization purposes [8]. Its species have numerous ornamental and medicinal values and were popular ornamental house plants during the Victorian era [9]. In traditional systems of medicine such as Chinese Traditional Medicine (TCM) and Indian Ayurveda, *Calanthe* has diverse uses, including detoxification and body cooling, resolving hard lumps, promoting blood circulation, treatment of arthritis, rheumatism, ulcers, common colds, and traumatic injuries. In addition, some species are used as tonics and as aphrodisiacs [10, 11].

Calanthe has undergone a series of intrageneric taxonomic revisions for many centuries since its establishment in 1821 [12]. The genus was first subdivided into two subgenera and various sections by Schlechter in 1914, and most authors have observed this subgeneric classification in the subsequent years in their studies [13]. Subgenus *Preptanthe* (Rchb.f.) Schltr. is characterized by swollen pseudobulbs and annual leaves, whereas subgenus *Calanthe* lacks prominent pseudobulbs and has evergreen leaves. The *Calanthe* group, a well-defined group of orchids in tribe Collabieae of subfamily Epidendroideae, was identified to include the genera *Calanthe* R. Br., *Cephalantheropsis* Guillaumin, and *Phaius* Lour. [14]. The three genera have been shown to have a close relationship hence leading to delimitation challenges, especially in the genera *Calanthe* and *Phaius*. Generally, species in this group are characterized by plicate leaves, simple, widely spreading sepals and petals, fused lip base and column, and eight waxy pollinia [6]. Morphologically, *Cephalantheropsis* is characterized by a spurless labellum, free from the column, and pollinia growing directly on the globose viscidium, while the *Phaius* labellum has a spur, grows at the column base but lacks adnation with column wings with pollinia attached by short caudicles. On the other hand, *Calanthe* is characterized by its labellum adnate to column wings forming a tube and spurred base having pollinia bound by conspicuous or inconspicuous caudicles, adherent to a sticky viscidium [15]. However, adnation of the lip to the column has been shown to have evolved several times independently, and some species, such as *Phaius delavayi*

(Finet) P.J.Cribb & Perner, have an intermediate column type between these two states hence taxonomic incongruences [16].

In terms of molecular studies, the family Orchidaceae generally has been subjected to two classification systems (i.e., Dressler 1993 and Chase et al. (1994)) [17–21] that try to infer its phylogeny and evolution from genus to subfamily levels. Within the Subfamily Epidendroideae, three genera; *Calanthe*, *Cephalantheropsis* and *Phaius*, form an independent alliance known as the *Calanthe* alliance, which can be easily distinguished from other taxa within the subfamily [1, 4, 15]. However, the phylogenetic relationships and affinities within the *Calanthe* alliance remain unresolved. Previous molecular studies conducted on Epidendroideae treated two lineages of the traditional *Calanthe*, namely: *Preptanthe* Rchb. f. and *Styloglossum* Breda, as distinct genera [3, 5]. Additionally, other molecular studies on the *Calanthe* group reported that *Calanthe* is a polyphyletic genus that clusters with its relatives *Cephalantheropsis* and *Phaius*, forming an independent alliance within Epidendroideae (Orchidaceae) named the *Calanthe* alliance [3, 15, 16]. This alliance can be differentiated from other taxonomic groups within the family based on plicate leaves, similar sepals and petals, basal and lateral inflorescence, resupinate flowers with free sepals and petals, spurred lips, and eight waxy pollinia forming two groups [15]. However, determining the phylogenetic and taxonomic relationships within this alliance is difficult. This is due to the uncertainties in the precise delimitation characters and the incongruent molecular phylogeny within the *Calanthe* alliance, that has led to poor taxonomic classification; thus, the phylogenetic and taxonomic relationships within the *Calanthe* alliance remain inconclusive. To better understand their phylogenetic relationships, it is necessary for us to identify discrepancies in the genetic information of the major clade the *Calanthe* alliance.

Genus *Phaius* contains ca. 40 known species, out of which 9 occur in China [6, 21]. The species in this genus are also characterized by a labellum growing at the base of the column having a spur but lacks adnation to column wings, and a pollinium usually attached by short caudicles [15]. Based on morphological data, this genus is separated into two types: bract caducous or persistent [6]. *Calanthe*, on the other hand, is characterized by a labellum which is adnate to column wings forming a tube and spurred base, and pollinium having conspicuous or inconspicuous caudicles, usually adhering to a sticky viscidium [6, 14, 21]. Previously, the genera *Calanthe* and *Phaius* were distinguished by the lip being adnate to or almost free from the column, respectively, but all intermediate conditions exist [22]. However, there are limited genetic studies on these species, and the aforementioned

characteristics are ambiguous, thus not sufficient to distinguish among the genera or infrageneric taxa of the *Calanthe* alliance. Therefore, there is a need for more in-depth studies to resolve these relationships.

Systematics and phylogeny, since its establishment, has promoted classification and interpretation of the evolutionary relationships among angiosperms via genomic analysis [23]. Chloroplasts are essential in photosynthesis and form part of the primary genetic system together with the nucleus and the mitochondria [24]. Plastome (Chloroplast genome) sizes range from 120 to 170 kb in many angiosperms [25]. The plastome is relatively conserved in terms of the size of the gene, gene content, arrangement of the genes, and genome structure [26]. Compared to the nuclear genome, the chloroplast genome undergoes very few nucleotide substitutions and gene rearrangements; hence has been a perfect model to study genetic change and phylogeny in complex terrestrial plants [27].

In the current study, we sequenced, assembled and annotated the complete chloroplast genome of 8 species from the two genera in the *Calanthe* alliance, namely: *Calanthe* and *Phaius*. The aim of this study was to; 1.) Understand the genetic structure and variation within the plastomes; 2.) Identify and describe the characteristics of the cp genome structure, sequence divergence, mutational hotspot regions, and repeat regions across the plastomes and, 3.) Evaluate the phylogenetic relationships between the genera *Calanthe* and *Phaius*, which may be useful for further species evolution studies.

Results

Chloroplast genome organization of the *Calanthe* group species

The complete chloroplast genomes of 8 species of the *Calanthe* group display a common quadripartite structure consisting of two Inverted Repeat (IR) regions (IRa and IRb), a Large Single Copy (LSC) region, and a Small Single Copy (SSC) region. Their sizes range as follows: IRs (25,222bp-26,430bp), LSC (83,364bp-87,450bp), and an SSC (16,297bp-18,586bp) (Fig. 1; Table 1).

The GC content was varied within the LSC, SSC, and IR regions, ranging between 34.2-34.6%, 29.4-29.8%, and 43.0-43.1%, respectively, in the regions (Table 1).

Most of the genomes encoded 134 differential genes, containing 88 CDS and 38 tRNA. However, only 131 and 133 genes, 85 and 88 CDs, and 38 and 37 tRNA genes were recorded in *P. delavayi* and *P. flavus*, respectively. All of these species encoded eight rRNA genes. The 8 *Calanthe* group plastomes had identical numbers, order, and names except for the three genes lost in *P. delavayi*, namely, *ndhC*, *ndhF*, and *ndhK*. In addition, the *Calanthe* group plastome contained 6 tRNA genes (*trnA-UGC*,

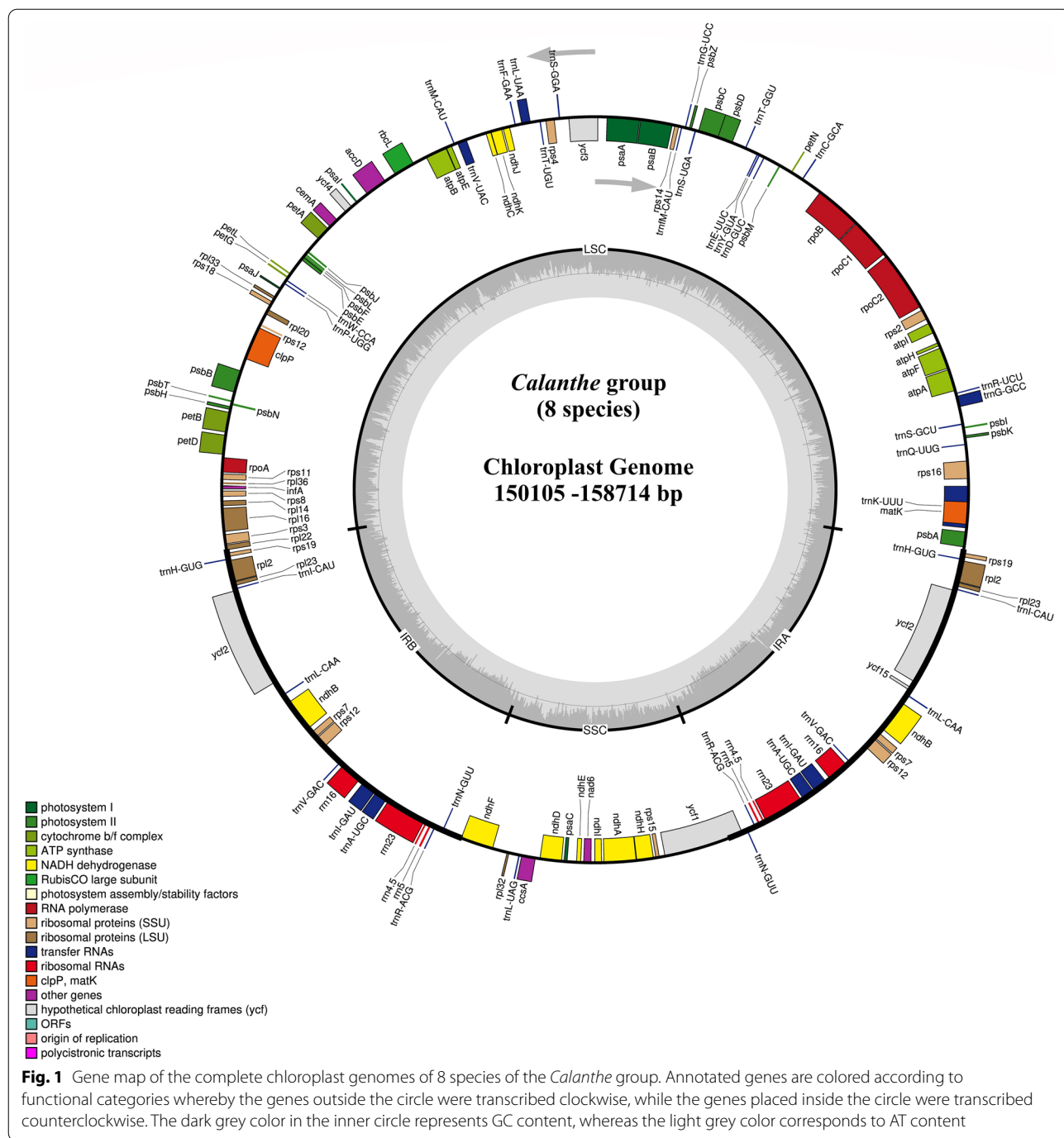
trnI-GAU, *trnG-UCC*, *trnK-UUUU*, *trnL-UAA* and *trnV-UAC*) and nine protein-coding genes (*rpl2*, *rpl16*, *rps16*, *rpoC1*, *ndhA*, *ndhB*, *atpF*, *petB* and *petD*) having one intron and three genes (*ycf3*, *clpP*, and *rps12*) containing two introns. A total of 19 genes were duplicated in the IR regions, including three types of genes, namely, seven coding genes (*rps12*, *rps19*, *rps7*, *rpl2*, *rpl23*, *ndhB*, *ycf2*), eight tRNA (*trnH-GUG*, *trnI-CAU*, *trnL-CAA*, *trnV-GAC*, *trnI-GAU*, *trnA-UGC*, *trnR-ACG*, *trnN-GUU*) genes and four rRNA (*rrn16*, *rrn23*, *rrn4.5*, *rrn5*) genes (Table 2). The *rps12* gene was trans-spliced, overlapping two regions in the cp genome whereby the 5'-end exon was found in the LSC region and the intron, 3'-end exon located in the IR region. Moreover, three pairs of genes, *trnK-UUUU/matK*, *atpE/atpB*, and *psbD/psbC* had overlapping sequences.

Contraction and expansion of IR regions

The chloroplast genome structure and the junction positions between IR regions among the eight species exhibited several structural variations in the LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC borders (Fig. 2). Three different occurrences were observed in the LSC/IRb border. First, in *Calanthe ecarinata* and *C. tricarinata*, the *rpl22* gene was found in the LSC region 22bp away from the IRb region. Secondly, in 5 species, namely: *C. brevicornu*, *C. alpina*, *P. flavus*, *P. delavayi*, and *C. nipponica*, the *rpl22* gene overlapped in the LSC/IRb region by 52-60 bp in the IRb region. The third occurrence was observed in *C. taibaishanensis* whereby the *rps19* gene was 24 bp away from the LSC/IRb instead of *rpl22*. The IRb/SSC junction regions were relatively conserved in 7 species whereby the *ndhF* gene crossed over to the IRb region by 51-70 bp except in *P. delavayi* due to its *ndhF* gene loss. In this regard, the nearest gene *trnN* in IRb, was 367 bp away from the SSC region in *P. delavayi*. Both the SSC/IRa and IRa/LSC are well conserved among the 8 *Calanthe* group genomes whereby the *ycf1* gene crossed over the SSC/IRa boundary having 42-1035 bp into the IRa section. Furthermore, in the IRa/LSC junction of 7 species, the *psbA* gene is found in the LSC region, 106-154 bp away from the IRa. The IRa/LSC junction of *C. taibaishanensis* is distinct in that the *rps19* gene occurs in the IRa, 25 bp away from the LSC.

Comparative genomic analysis

The mVISTA-based identity plot revealed the DNA sequence and gene synteny conservation across the eight plastomes and showed the regions with increased genetic variations (Fig. 3). The number, order, and orientation of genes were relatively conserved. Distinct sequence variations were recorded in several gene regions including *psbA-trnK-UUUU*, *rps16-trnQ-UUG*, *matK-trnK-UUUU-rps16*,



trnS-GCU-trnG-GCC, *rpoB-trnC-GCA*, *petN-psbM*, *psbM-trnD-GUC*, *trnE-UUC-trnT-GGU*, *trnT-GGU-psbD*, *ndhK-trnM-CAU*, *atpB-rbcL*, *rbcL-accD*, *accD-psaI*, *petA-psbJ*, *psbE-petL*, *trnV-GAC-rps12*, *ccsA-ndhD*, *trnL-UAA*, *trnL-GAU*, *ndhF*, *ndhI*, *rps15*, *trnP-UGG*, *rpl33*, *clpP*, *psbT*, *rpl16*, *rpl14*, *rps8* and *rpl32*. Higher genetic variability was observed in the LSC and SSC regions than in the IR regions and in non-coding regions than in the conserved protein-coding

regions. The rRNA genes were highly conserved with almost no variation in terms of their numbers among the plastomes. Moreover, greater variation was recorded in the IGS regions than those in the gene regions.

The sliding window analysis identified three highly variable regions in the 8 *Calanthe* group plastomes with a nucleotide diversity (P_i) cut-off point set at $P_i \geq 0.03$ (Fig. 4). The highly variable regions were mainly found

Table 1 Basic features of the chloroplast genome of the 8 *Calanthe* group species

Species	<i>Calanthe alpina</i>	<i>Calanthe brevicornu</i>	<i>Calanthe ecarinata</i>	<i>Calanthe nipponica</i>	<i>Calanthe taibaishanensis</i>	<i>Calanthe tricarinata</i>	<i>Phaius delavayi</i>	<i>Phaius flavus</i>
Accession no.	OL322023	OL348396	OL348397	OL348398	OL351366	OL351367	OL351368	OL351369
Total Length (bp)	156591	158384	158329	158714	157959	158343	150105	158556
LSC Size (bp)	85489	87155	87293	87450	87129	87285	83364	87216
LSC GC%	34.3	34.3	34.3	34.2	34.2	34.3	34.5	34.6
LSC Length (%)	54.6	55.0	55.1	55.1	55.2	55.1	55.5	55.0
SSC Size (bp)	18436	18531	18500	18404	18424	18522	16297	18586
SSC GC%	29.5	29.6	29.7	29.7	29.6	29.7	29.4	29.8
SSC Length (%)	13.1	11.7	11.7	11.6	11.7	12.0	11.0	11.7
IR Size (bp)	26333	26349	26268	26430	26203	26268	25222	26377
IR GC%	43.1	43.1	43.1	43.0	43.1	43.1	43.3	43.0
IR Length (%)	16.2	16.6	16.6	16.7	16.6	16.4	16.8	16.6
PCGs region size (bp)	78678	79083	79083	79275	79251	79083	73473	79353
tRNA size (bp)	2870	2870	2870	2870	2870	2870	2870	2812
rRNA size (bp)	9042	9042	9042	9042	9042	9042	9042	9042
GC content (%)	36.7	36.7	36.7	36.6	36.6	36.7	36.9	36.8
No. of PCGs	88	88	88	88	88	88	85	88
No. of tRNA	38	38	38	38	38	38	38	37
No. of rRNA	8(4)	8(4)	8(4)	8(4)	8(4)	8(4)	8(4)	8(4)
No. of genes	134	134	134	134	134	134	131	133

in the LSC and SSC region compared to IR regions and in non-coding regions than coding regions. The highly variable regions were identified as follows; *trnS-GCU-trnG-GCC*, *rpoB-trnC-GCA*, *trnE-UCC-trnT-GGU*, *rpl32-trnL-UAG*, *ccsA-ndhD* and *psbL*, *clpP* and *rpl32* genes of the chloroplast genomes. These findings were consistent with the mVISTA results, whereby the variation in the IR regions of the chloroplast genomes was relatively lower than that in the LSC and SSC sections.

Sequence repeats

A total of 507 SSRs were recorded in this study, and the chloroplast genomes of the eight species contained nearly similar numbers of SSRs (57-76) (Fig. 5). *Calanthe alpina* had the highest number of simple repeats (76), while *Phaius delavayi* had the least SSRs (57). Additionally, a base preference was recorded in the base composition of the repeating motifs from mononucleotide SSRs to trinucleotide SSRs, mainly A-T rich repeating motifs. Mononucleotide repeats were the most abundant SSRs (28-45), while hexanucleotide repeats were the least (1-2) in the 8 cp genomes (Table S4). A/T repeats were the most abundant mononucleotide repeats (287), followed by dinucleotide repeat sequences (92) which predominantly

consisted of AT/AT repeats and all trinucleotide repeats were AAT/ATT. The least abundant repeats were AAG/CTT, AAATAT/ATATTT, AAGTAT/ACTTAT, ACATAT/ATATGT, and AGATAT/ATATCT (Table S4). The LSC of the eight complete genomes had the highest number of SSRs (385). Mononucleotide repeats were the most abundant ranging from 18-33 (LSC), 8-10 (SSC), and (1-2) IR. Hexanucleotides were the least SSRs in all the regions, with none occurring in the IRs of all the species (Figs. 6, 7, 8, and Table S4).

Tandem repeats were detected and are classified as forward (F), palindrome (P), reverse (R), or complement (C), with each repeat having a length of ≥ 30 bp sequence similarity of $\geq 90\%$. A total of 28-40 repeat sequences were identified, and the highest number of repeats were recorded in *P. delavayi* (Fig. 9). Palindromic repeats were the most abundant in all the 8 *Calanthe* group plastomes (17-25), whereas complement repeats were the least abundant (1-5). No complement repeats were found in the chloroplast genomes of *C. alpina* and *P. flavus*. The length of the repeat sequences detected predominantly varied between 31 to 50 bp. Additionally, there were no complement repeat sequences in all the eight cp genomes having 51-70 bp in length (Fig. 10 and Table S5). Overall,

Table 2 Group of genes encoded in the complete cp genome of the 8 *Calanthe* group species

Category for Genes	Group of Genes	Name of Genes
Self-replication	transfer RNAs	<i>trnK-UUU^a, trnQ-UUG, trnS-GCU, trnG-GCC, trnR-UCU, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnT-GGU, trnS-UGA, trnG-UCC^a, trnI-M-CAU, trnS-GGA, trnT-UGU, trnL-UAA^a, trnF-GAA, trnV-UAC^a, trnM-CAU, trnW-CCA, trnP-UGG, trnH-GUG^a, trnI-CAU^a, trnL-CAA^a, trnV-GAC^a, trnI-GAU^{a*}, trnA-UGC^{a*}, trnR-ACG^a, trnN-GUU^a, trnL-UAG</i>
	ribosomal RNAs	<i>rrn16^a, rrn23^a, rrn4.5^a, rrn5^a</i>
	RNA polymerase	<i>rpoA, rpoB, rpoC1^a, rpoC2</i>
	Small subunit of ribosomal proteins (SSU)	<i>rps11, rps12^{a*}, rps14, rps15, rps16^a, rps18, rps19^a, rps2, rps3, rps4, rps7^a, rps8</i>
	Large subunit of ribosomal proteins (LSU)	<i>rpl14, rpl16^a, rpl2^{a*}, rpl20, rpl22, rpl23^a, rpl32, rpl33, rpl36</i>
	Subunits of NADH-dehydrogenase	<i>ndhA^a, ndhB^{a*}, ndhC^b, ndhD, ndhE, ndhF^b, ndhH, ndhI, ndhJ, ndhK^b</i>
	Subunits of photosystem I	<i>psaA, psab, psac, psal, psaj</i>
Genes for photosynthesis	Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ,</i>
	Subunits of cytochrome b/f complex	<i>petA, petB^a, petD^a, petG, petL, petN</i>
	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF^a, atpH, atpI</i>
	Large subunit of rubisco	<i>rbcl</i>
	Translational initiation factor	<i>infA</i>
	Protease	<i>clpP^a</i>
	Maturase	<i>matK</i>
Other genes	Subunit of Acetyl-CoA-carboxylase	<i>accD</i>
	Envelope membrane protein	<i>cemA</i>
	C-type cytochrome synthesis gene	<i>ccsA</i>
	hypothetical chloroplast reading frames (<i>ycf</i>)	<i>ycf1, ycf^a, ycf3^a, ycf4, ycf15</i>

Note:

^a Genes containing introns

^b Genes lost in *P. delavayi*

* Duplicated genes

the SSRs and tandem repeats in the 8 *Calanthe* group cp genomes showed no significant differences (Kruskal-Wallis, $P < 0.05$; Table S6). The numbers, types and sizes of SSRs and tandem repeats, however, varied greatly across different structural and functional regions of the cp genomes whereby these repeats were abundant in non-coding regions than in the coding regions (Table S3).

Codon Usage

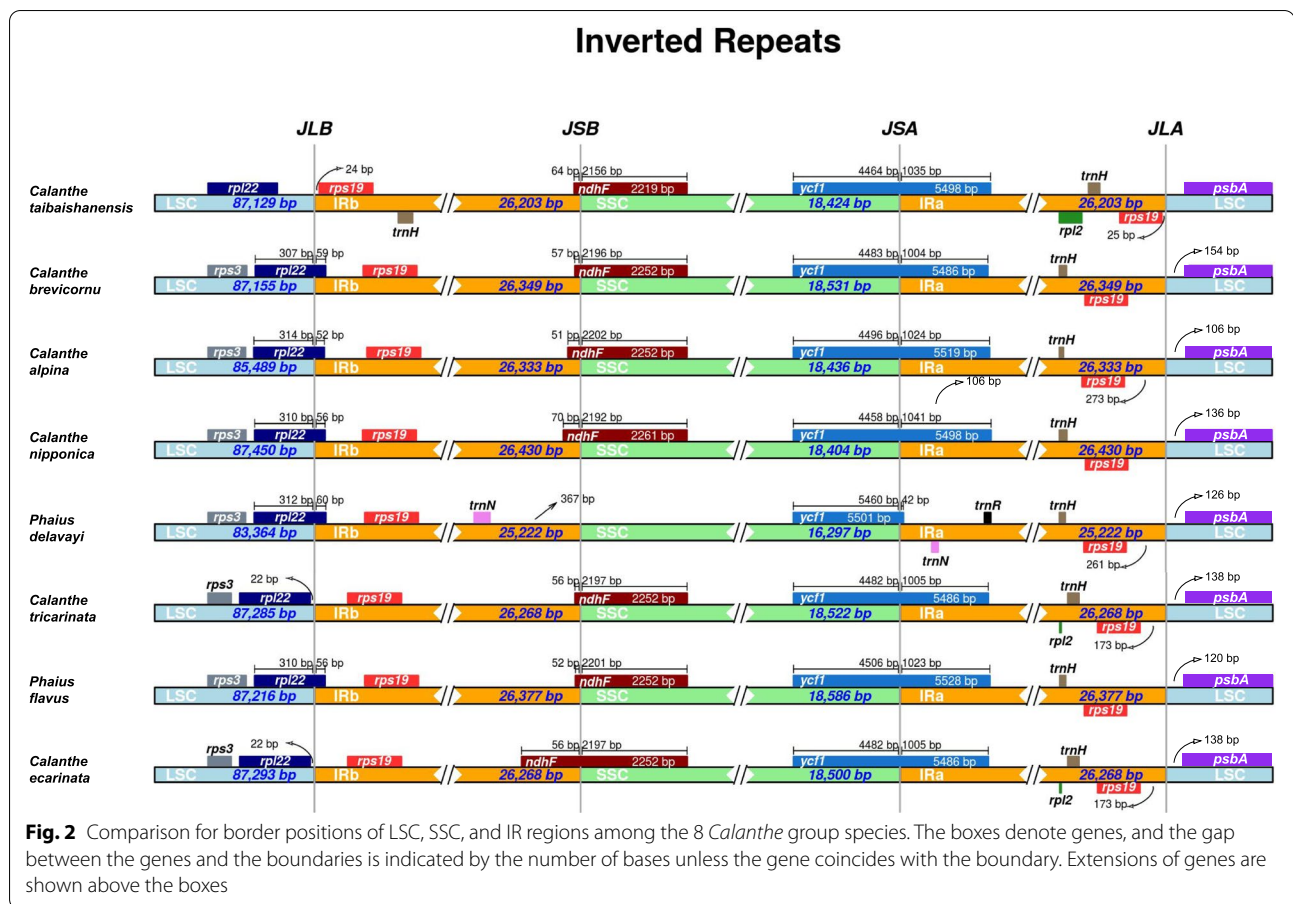
The RSCU of the chloroplast genomes of the 8 *Calanthe* group species was calculated using all protein-coding genes, and a total of 50,035–52,904 codons were recorded. The RSCU values for each species displayed an identical codon preference in the 64 codons of the protein-coding genes. In this regard, 30, 31, 32, and 33 codons from *Calanthe taibaishanensis*; *C. alpina*, *C. brevicornu*, *Phaius delavayi*, *Phaius flavus*, and *Calanthe ecarinata*; *C. tricarinata*; and *C. nipponica* respectively exhibited greater preference (RSCU > 1). Two of them, tryptophan (Trp) and methionine (Met), displayed no preferences (RSCU = 1) in all the species. The rest of the codons were least preferred (RSCU < 1). There were no rare codons (RSCU

< 0.1) recorded in the CDS genes of the 8 cp genomes of the *Calanthe* group (Fig. 11).

Leucine (Leu), encoded by UUA, UUG, CUU, CUC, CUA, and CUG was the most abundant amino acid, with a proportion of 9.70–10.56 %, which consist 4,993–5,374 of the total number of codons. Serine (Ser), encoded by UCU, UCC, UCA, UCG, AGU, and AGC was the second most plentiful amino with a proportion of 8.93–9.81 % (4,661–5,187 codons), whereas tryptophan (Trp), encoded by UGG was the least abundant amino acid encoded, with a proportion of 1.22–1.47 % (641–759 codons) (Table S7). Statistical analysis of the RSCU values in the 8 *Calanthe* group cp genomes did not vary significantly (Kruska-Wallis, $P < 0.05$; Table S8).

Phylogenetic analysis

The application of high-throughput sequencing technology has enhanced the availability of whole plastid genomes, leading to the resolution of closely related taxa using plastome sequences [24]. The phylogenetic positions of the eight newly sequenced *Calanthe* and *Phaius* species were inferred using a matrix of 64,593 characters (nucleotides). These characters represent the



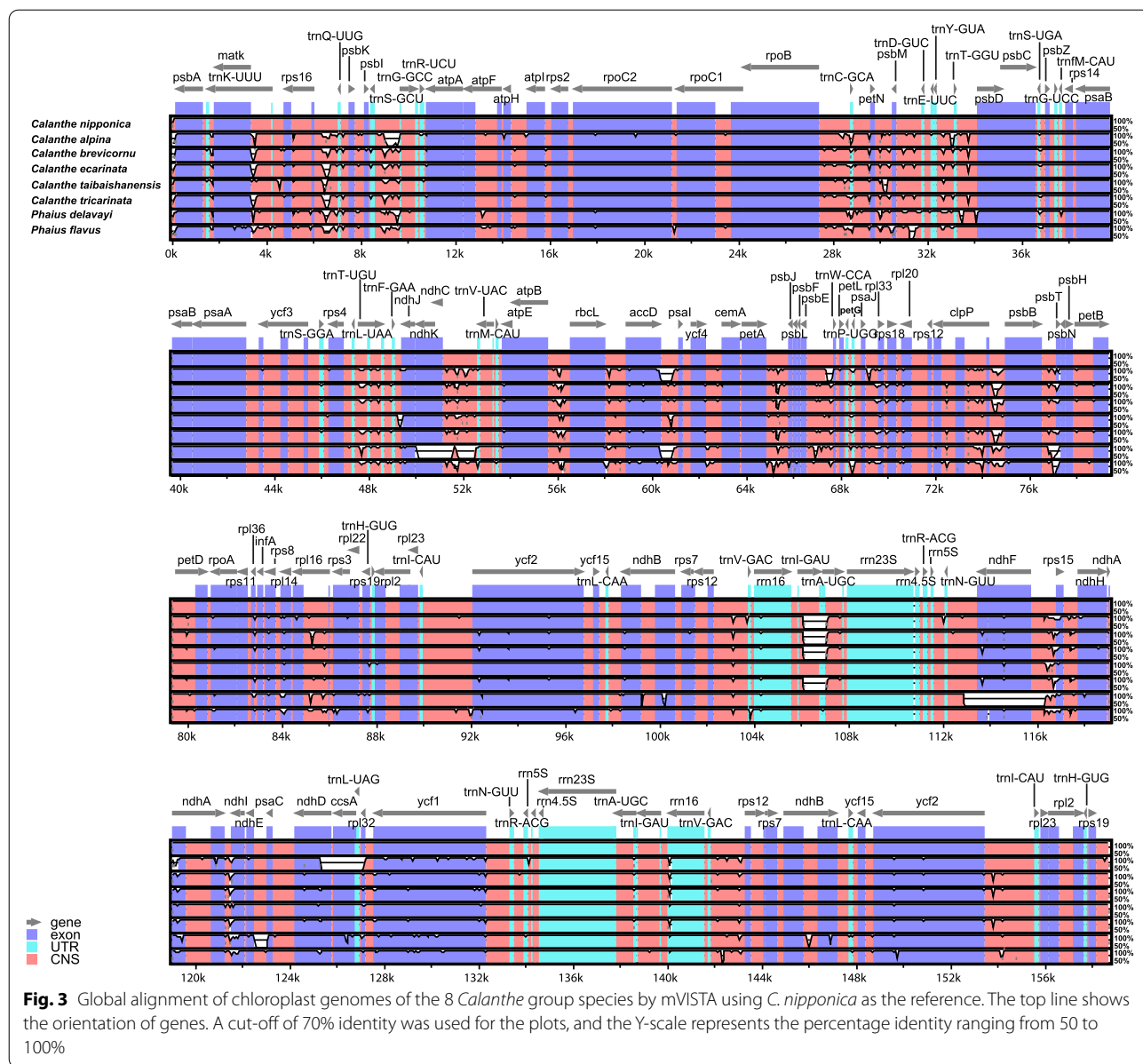
73 protein-coding genes shared among the eight species in the *Calanthe* group, combined with 14 species of the *Calanthe* alliance, for which their complete chloroplast genome sequences had been officially published in the NCBI database. The ML and BI trees exhibited similar phylogenetic topologies with high bootstrap values and posterior probabilities (Fig. 12; Figure S1). *Phaius* species (excluding *P. delavayi*) form two clades in the phylogenetic tree (Fig. 12; Figure S1). The first clade includes *P. tankervilleae*, and *P. hainanensis* [BP_(ML) = 100%, PP = 1.00]. The second clade consists of only *P. flavus* [BP_(ML) = 75.6/57%, PP = 0.9851], closely related to *Cephalantheropsis*, and *Styloglossum*. The third clade was composed of two sister groups: *Cephalantheropsis* (*C. obcordata*) and *Styloglossum* (*C. lyroglossa*) [BP_(ML) = 100%, PP = 1.00]. The fourth clade consisted of two species, *P. delavayi*/*C. delavayi* [BP_(ML) = 100%, PP = 1.00], while the last clade was made of the rest of the species of section *Calanthe*. All the species of sect. *Calanthe* clustered together in a super clade [BP_(ML) = 100%, PP = 1.00], consisting of three lineages and includes 14 *Calanthe* group species. The first lineage includes only *C. alpina* [BP_(ML) = 98.2/95%, PP = 1.00], and this taxon is

sister to a clade containing the remaining species of this section. The rest of the species formed the other two lineages of this section which include (*C. triplicata*, *C. sylvatica*, *C. davidii*) and (*C. taibaishanensis*, *C. nipponica*, *C. arcuata*) [BP_(ML) = 98.9/96%, PP = 1.00] as well as (*C. griffithii*, *C. tricarinata*, *C. ecarinata*, *C. brevicornu*) and (*C. henryi*, *C. bicolor* and *C. aristulifera*) [BP_(ML) = 100%, PP = 1.00].

Discussion

Comparison of the chloroplast genomes of the 8 *Calanthe* group species

Complete chloroplast sequencing and genomic analyses have revealed that orchid plastomes are highly conserved in terms of size, structure, gene order and content [28–31]. These findings are congruent with results from our study on the 8 plastomes of the *Calanthe* group which revealed that the cp genome of the 8 *Calanthe* group species is a quadripartite structure that varied in size among the species ranging between 150,105 bp (*P. delavayi*) and 158,714 bp (*C. nipponica*). The plastome is divided into four regions consisting of an LSC (83,364bp–87,450bp), IRs (25,222bp–26,430bp), and an SSC (16,297–18,586bp).



The inferred structure and contents are consistent with previous research on orchids [32, 33]. The chloroplast genome in angiosperms has a conserved genome structure [34], including two inverted repeats (IRs) which separate a large single-copy section (LSC) and a small single-copy section (SSC). Furthermore, when compared to nuclear and certain plant mitochondrial genomes, chloroplast genomes are smaller and less prone to recombination, providing unique data for studying genome size variation and evolutionary status [35, 36]. These characteristics are useful for comparative studies because they allow researchers to investigate genome divergences across a wide range of evolutionary time, from early land

plants [37] to recently domesticated plants, and to detect selection signals of genome size evolution [38].

The genome sizes of the 8 *Calanthe* group species varied in size among the species between 150,105 bp (*P. delavayi*) and 158,714 bp (*C. nipponica*). Previous studies on seed plants have proposed three important factors that cause variation in chloroplast genome size: (1) intergenic region variations, which mainly affects variation in chloroplast genome size within a genus [39, 40]; (2) variation of an IR region [41, 42]; and (3) gene loss, which is an important reason for the shrinking of chloroplast genome size in some plants [41, 42]. The length corresponds to the size range of the cp genomes of most

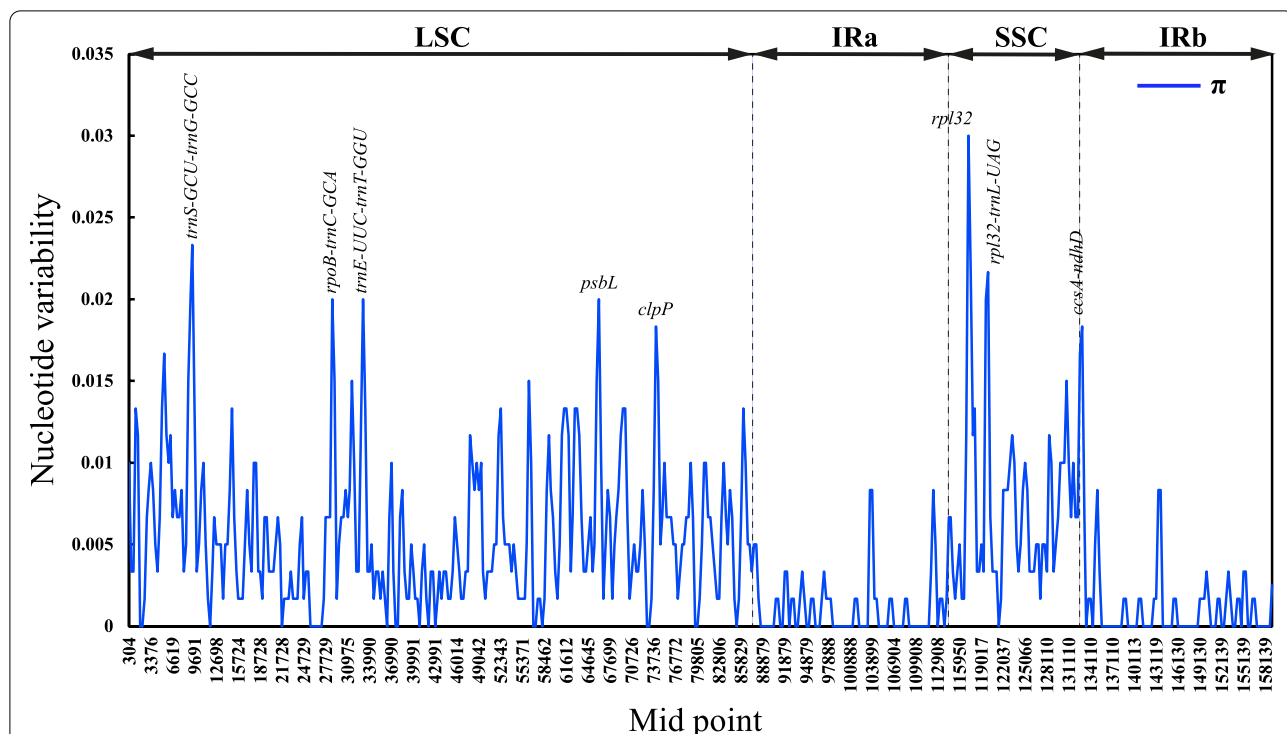


Fig. 4 Comparative analysis of the nucleotide diversity values among the 8 *Calanthe* group chloroplast genomes. The X-axis represents the position of the midpoint of a window (kb), while the Y-axis indicates the nucleotide diversity of each window

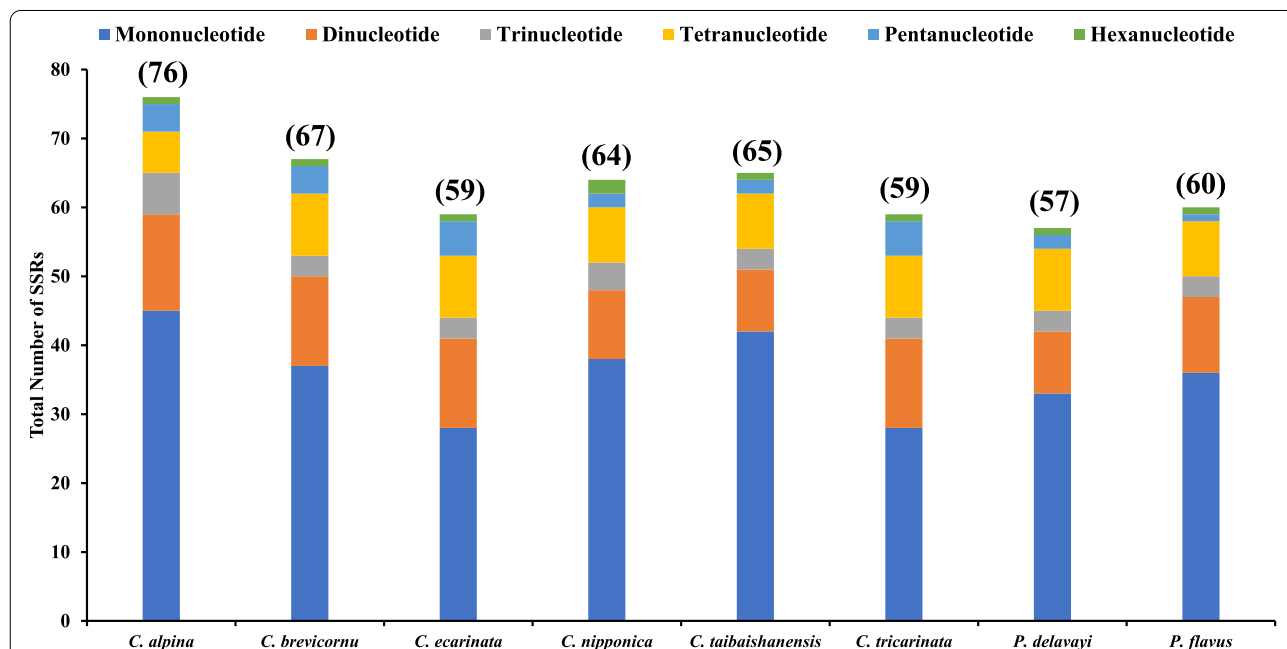
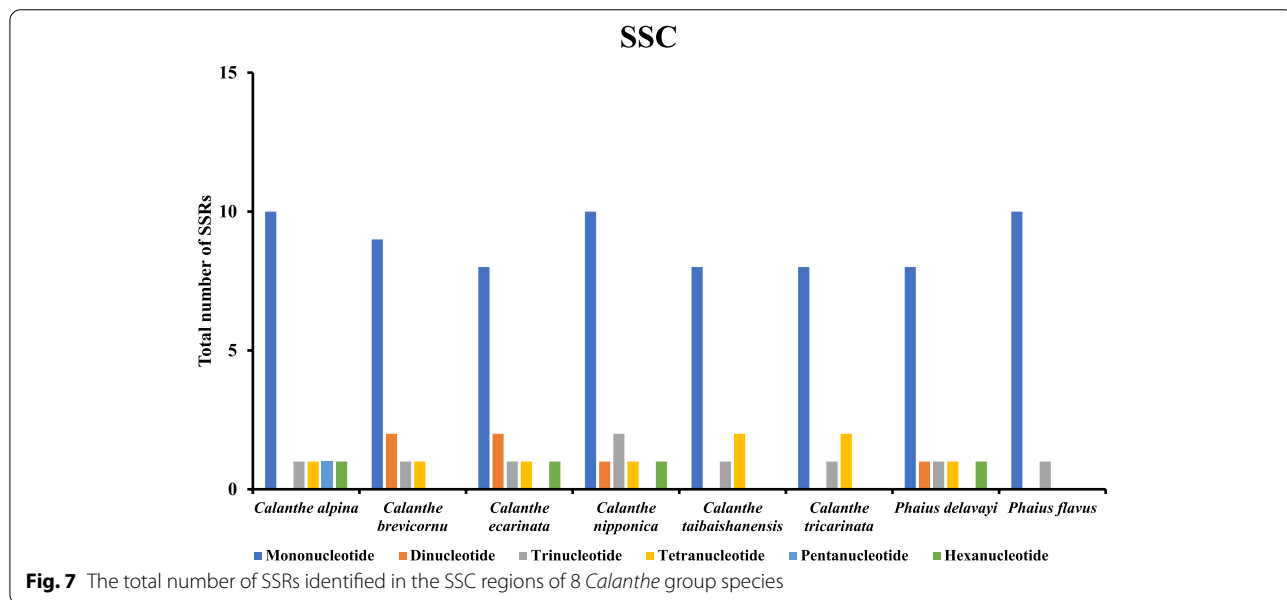
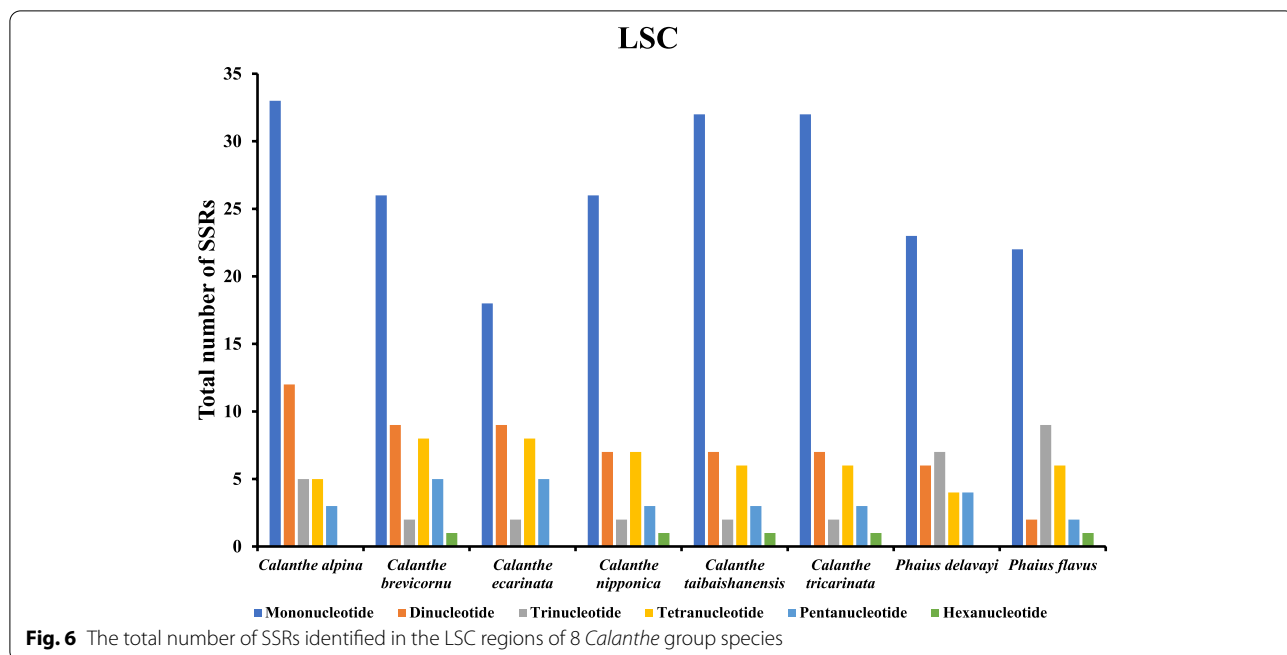


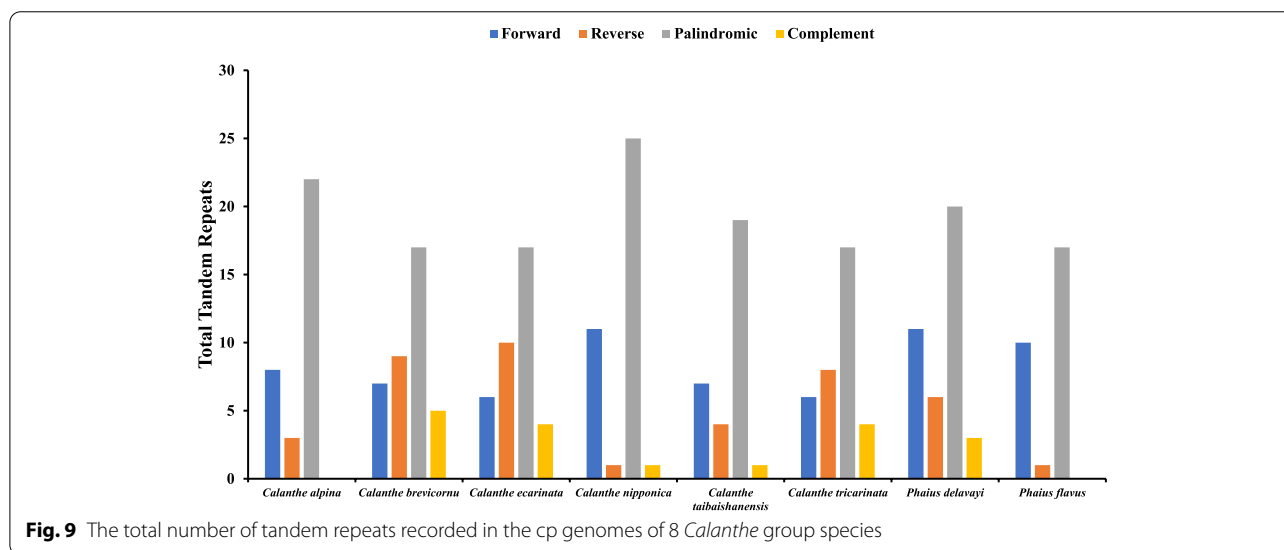
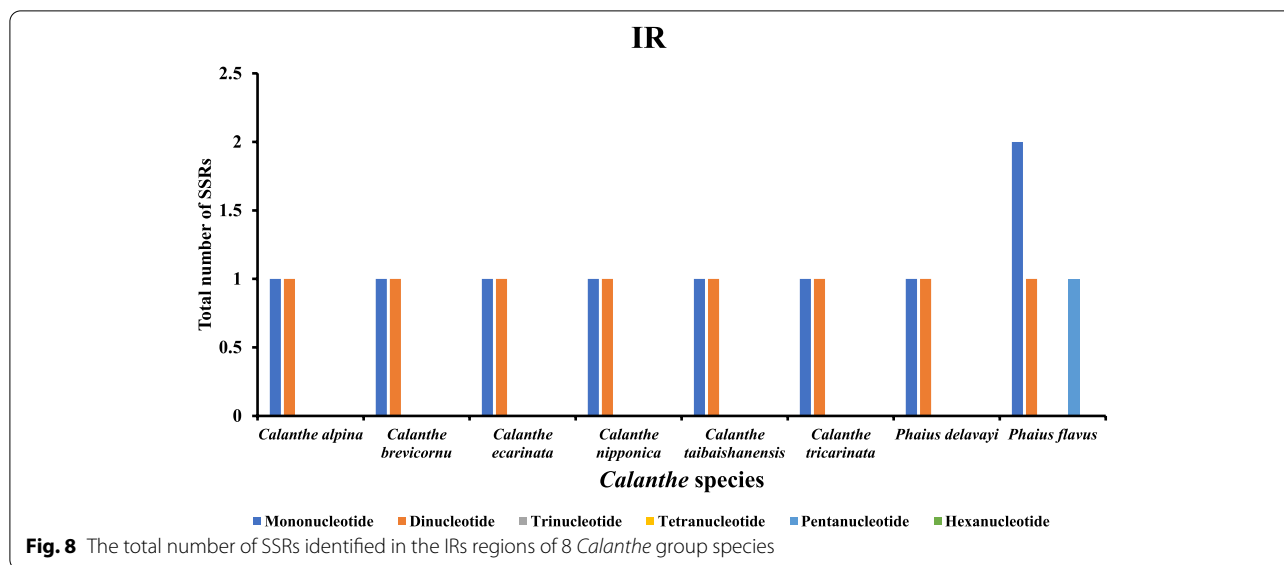
Fig. 5 The total number of SSRs recorded in the cp genomes of 8 *Calanthe* group species



angiosperms [43]. However, the variation in size among cp genomes in orchids been linked to the contraction and expansion of both the inverted repeat regions [4, 28, 44].

Angiosperm plastomes have comparatively little variation in gene content, despite their differences in size range [45], similar to findings from our study which displayed sequence similarity in gene order and arrangement across the 8 *Calanthe* group plastomes. The plastomes' characteristics and sequence variabilities

have been linked species phylogenetic relationships and evolution; thus, closely related species are more likely to have similar plastome sizes and characteristics [46]. A previous study on the evolution of flowering plants plastome architecture revealed that the cumulative influence of transposable elements proliferation greatly dwarfs the impacts of tandem or dispersed gene duplication in increasing genomic DNA content, and the process of long-term genomic fractionation,

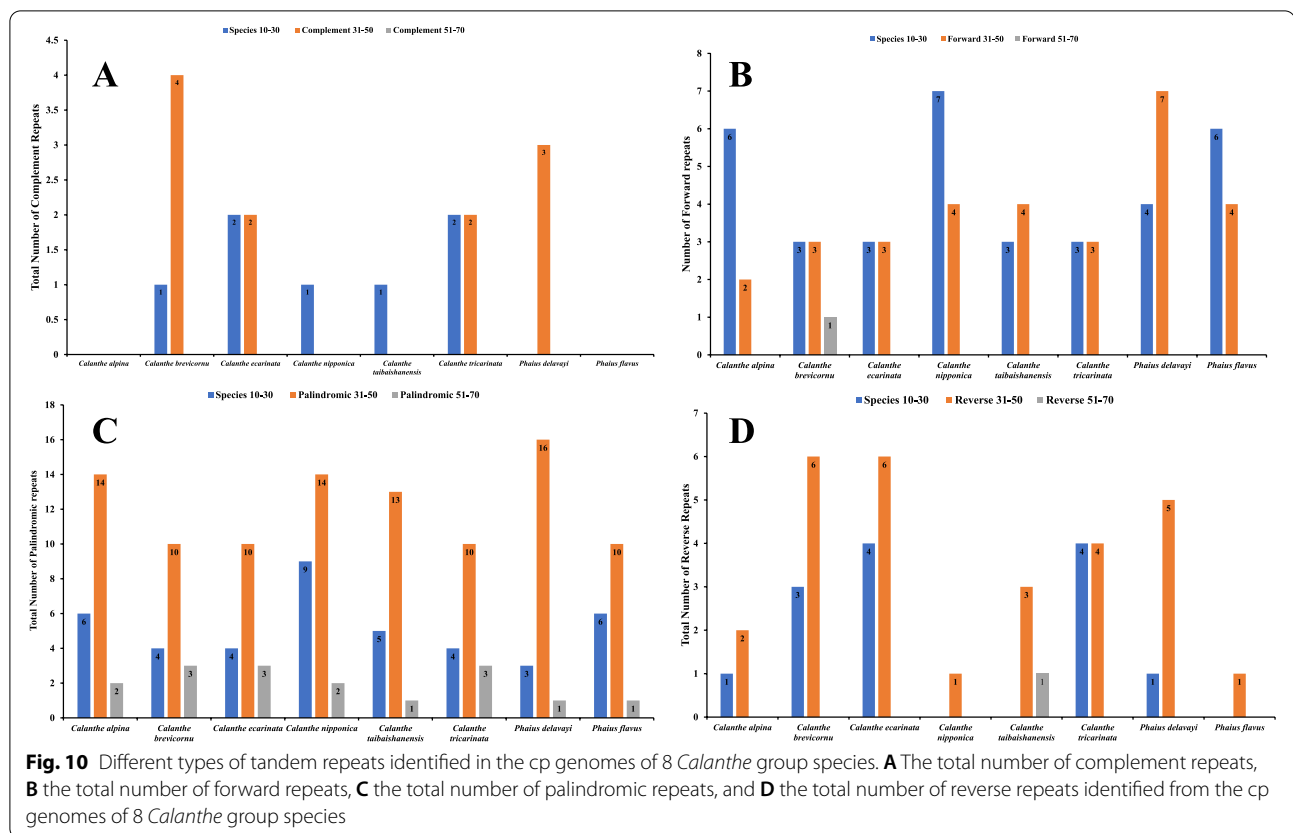


which is associated with the loss of most gene duplications after whole genome duplication [47]. Transposable elements have been implicated as important factors in gene regulation and adaptation, particularly because gene content is fairly consistent across plants and transposable elements accumulate and degrade rapidly [48–50]. Although this pattern is now known, the underlying causes of constancy of genic content in related orchid genera despite the rapid diversification rate in the family Orchidaceae are far less well understood.

Additionally, the GC contents of the LSC and SSC regions in the 8 *Calanthe* group species were lower compared to that of the IR regions. This occurrence

was possibly due to the four rRNA genes, *rrn16*, *rrn23*, *rrn4.5*, and *rrn5* sequences in the IR regions [33].

A few differences were recorded in the protein-coding genes of the complete cp genome of the 8 *Calanthe* group species, despite land plants being generally considered highly conserved [51]. We revealed that protein-coding genes in the *ndh* family differed between *Calanthe* group species. The genes *ndhC*, *ndhF*, and *ndhK* were lost in *P. delavayi* but they were retained in the other species. The loss of these three NADH dehydrogenase subunits is common in orchids and was first reported in this species by Chen in 2020 [4]. In higher plants, the cp genomes contain 11 *ndh* genes (*ndhA-ndhK*) that encode nicotinamide-adenine dinucleotide (NADH) dehydrogenase



subunits that associate with nuclear-encoded subunits to form the NADH dehydrogenase-like (NDH) complex, which is involved in cyclic electron flow around photosystem I (PSI) and chlororespiration [52, 53]. Although the chloroplast NDH complex mediates cyclic electron transport in PSI, no negative effects in *ndh*-deficient mutants or transgenics have been observed under suitable growing conditions [29], suggesting that chloroplast *ndh* genes may be unnecessary in autotrophic plants. Evidently, loss or pseudogenization of plastid *ndh* genes has been observed in a variety of photoautotrophic seed plant lineages [54, 55] including *Cymbidium*, *Dendrobium*, *Phalaenopsis*, and *Ophrys* [29, 56–58]. These studies also showed that different orchid species exhibited a variable loss or retention of the genes; for instance, *Cymbidium* encodes the *ndhE*, *ndhJ*, and *ndhC* genes [59] while *Oncidium* only encodes the *ndhB* gene [31]. The loss of the *ndh* genes has been linked to evolutionary processes whereby several studies inferred that orchids' ancestral protein-coding *ndh* genes might have been transferred to the nucleus [28, 60]. Fungal symbionts have also been attributed to the lack of functional *ndh* genes; thus, homologous genes from these resources have been presumed to perform the functions of the lost chloroplast-encoded *ndh* genes in some orchids [32,

60]. Nevertheless, this hypothesis is yet to be tested, and the mechanisms underlying the variable loss or retention of *ndh* genes in orchid species are worthy of further research.

Comparative analysis

DNA barcode technology has been widely used in identifying species, resource management, and phylogenetic and evolutionary studies [61, 62]. The comparative analysis of the 8 *Calanthe* group chloroplast genomes using mVISTA revealed the DNA sequence similarities among related species. No definitive rearrangements or gene inversions were recorded, indicating that the *Calanthe* group plastome was highly conserved [28]. The size of the genome and organization of the intergenic spacers correspond to previously observed variations in the size of the *Calanthe* chloroplast genomes [4].

In line with findings from other studies [63] and those from mVISTA, the LSC and SSC regions were more variable than the IR regions. The mVISTA results revealed the following highly variable regions across the 8 plastomes: *psbA-trnK-UUU*, *rps16-trnQ-UUG*, *matK-trnK-UUU-rps16*, *trnS-GCU-trnG-GCC*, *rpoB-trnC-GCA*, *petN-psbM*, *psbM-trnD-GUC*, *trnE-UUC-trnT-GGU*, *trnT-GGU-psbD*, *ndhK-trnM-CAU*, *atpB-rbcL*, *rbcL-accD*, *accD-psaI*,

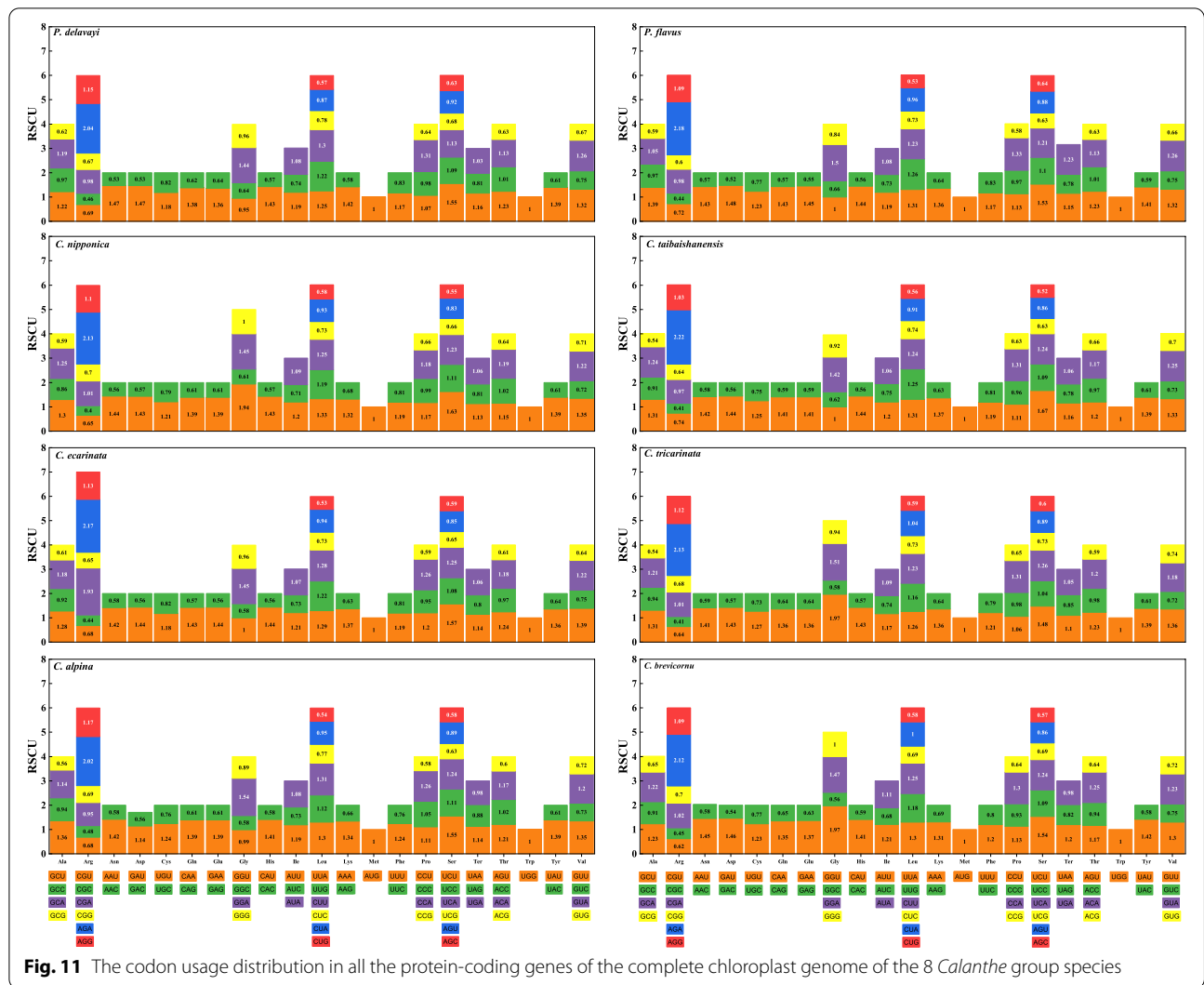
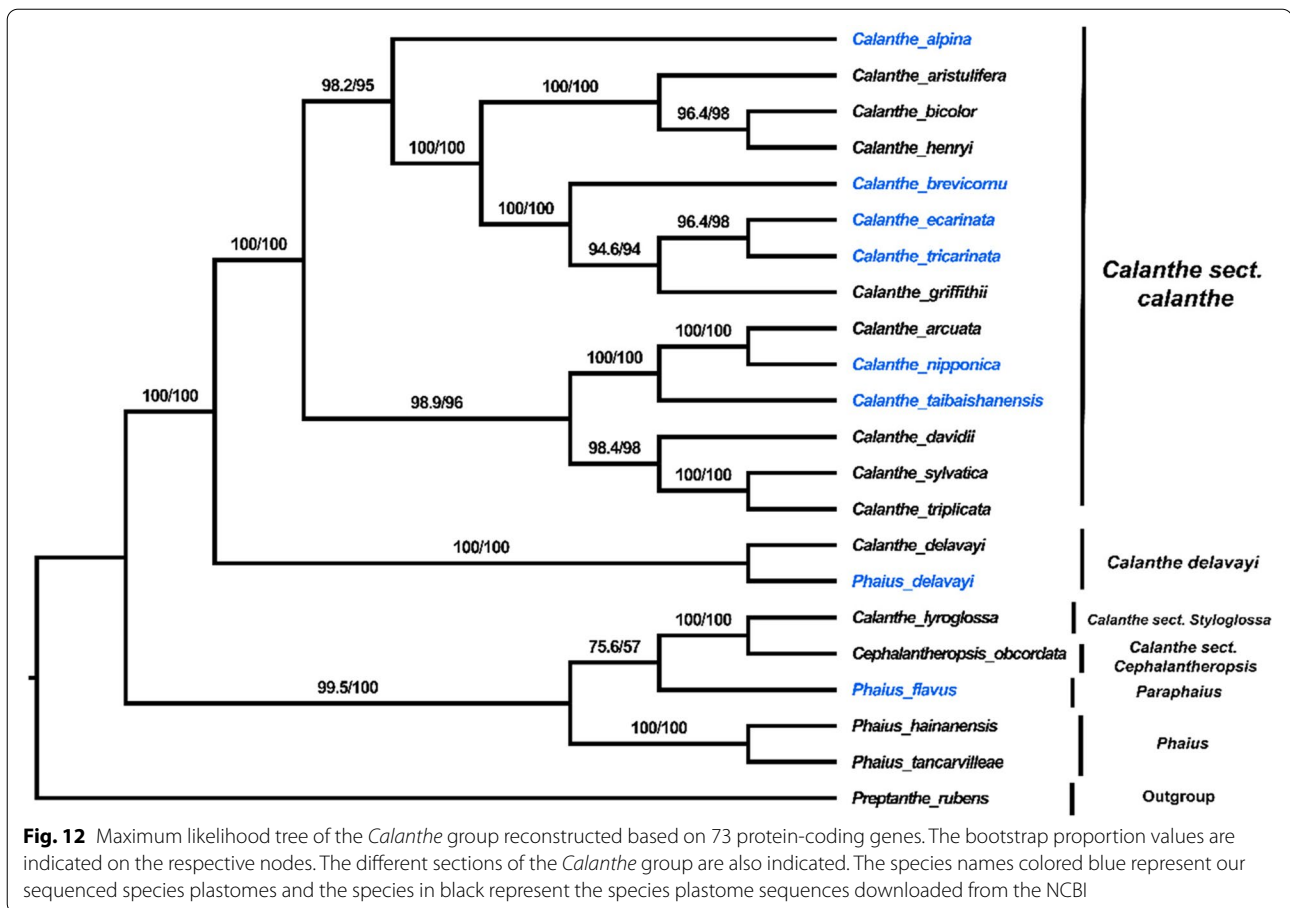


Fig. 11 The codon usage distribution in all the protein-coding genes of the complete chloroplast genome of the 8 *Calanthe* group species

petA-psbJ, *psbE-petL*, *trnV-GAC-rps12*, *ccsA-ndhD*, *trnL-UAA*, *trnL-GAU*, *ndhF*, *ndhI*, *rps15*, *trnP-UGG*, *rpl33*, *clpP*, *psbT*, *rpl16*, *rpl14*, *rps8* and *rpl32*. Interestingly, these highly variable regions were mostly similar to the mutational hotspots identified in other species of the *Calanthe* alliance [4], suggesting that these variable loci can be used as important references for future studies on the evolution and diversity in the *Calanthe* alliance. The nucleotide diversity was higher in the LSC and SSC compared to SSC regions and identified the following hypervariable regions across the *Calanthe* group plastome: *trnS-GCU-trnG-GCC*, *rpoB-trnC-GCA*, *trnE-UCC-trnI-GGU*, *rpl32-trnL-UAG*, *ccsA-ndhD* and protein-coding genes *psbL*, *clpP* and *rpl32*. The markedly high divergence observed in these genes and intergenic regions is similar to that observed in other angiosperms [4, 64, 65] and may be attributed to rapid genome

evolution due to higher mutation rates compared to other regions [66]. The IR regions are relatively conserved compared to the SSC and LSC regions in the *Calanthe* group plastomes. Significant variation was only observed in the LSC/IRb junction, which displayed three occurrences in the eight species. At the same time, the remaining three (IRb/SSC, SSC/IRa, and IRa/LSC) are conservative and stable. Contraction in the IR was detected due to the loss of the *ndhF* gene in *P. delavayi*. Previous studies have highlighted that the loss of *ndh* genes significantly contributes to the instability of the IR/SSC borders in orchids [58, 67]. The variation in size and evolutionary events in different plants may also be linked to the expansion and contraction of the junctions in the different regions of the chloroplast



plastomes [23, 68, 69]. The location of the boundary, particularly the expansion and contraction, has been successfully used to infer phylogenetic relationships and provide insights on the evolution of the lineages in Apiaceae [70], ferns [71], Poaceae [72], Pinaceae [28], and many monocots [73]. Nevertheless, even though overall genomic structures and gene orders are highly conserved, orchid plastomes exhibited clear differences at the IR/SSC boundaries, which cannot readily be used in a phylogenetic study. Furthermore, the *ndh* genes in SSC regions have been lost independently across orchid genera [58, 67], corroborating the findings by Kim et al. (2015) [30] which proposed that the instability of orchid IR/SSC junctions was highly related to the loss of the *ndhF* gene. Even so, the mechanism underlying the variations in the sequences flanking the IR/SSC junctions of orchid plastomes remains unknown. Therefore, our findings from the present study on the IR boundary does not provide the necessary information to elucidate the evolutionary relationships within the *Calanthe* group, thus, additional sampling of

Calanthe spp. and related genera will allow for clear and specific tests [74].

Molecular Markers

Simple sequence repeats have distinct features that make them efficient genetic markers such as abundance in number, highly repetitive, a simple structure, maternal inheritance of chloroplast genomes, and relatively conserved [75]. SSRs and repeat sequences have been extensively used in identifying species, phylogenetic analysis, population evolution studies, and system geography of various species [76]. In this regard, the variation in the number and distribution of SSRs and tandem repeats in all the 8 *Calanthe* group genomes and different regions of the whole plastomes were detected. Repeats were widespread in the non-coding regions compared to the coding regions, consistent with previous reports on other species [30, 77]. The chloroplast genome rearrangement and nucleotide substitution can be attributed to the differential distribution of these repeats [78].

Additionally, the SSRs were mainly distributed in LSC, compared to the SSC and IR regions illustrating that the distribution of SSRs was dependent on their locations in the chloroplast genome [79]. These repeats can therefore be used to develop genetic markers for phylogenetic studies. The identified SSR and tandem repeats can also be used to investigate the genetic structure, diversity, phylogeny, and differentiation of species in the *Calanthe* alliance and other orchid species.

Relative Synonymous Codon Usage

The RSCU value is the ratio of the usage frequency of a specific codon to the expected frequency and can eradicate the influence of amino acid composition on the codon usage [80]. Additionally, RSCU promotes the detection of synonymous codons [81]. Most codons with RSCU values greater than 1 ended with A or U, whereas those ending with a C or G had RSCU values of less than 1. These findings are consistent with previous studies [82, 83].

Compositional constraints and translational selection are presumed as the main factors that result in the codon usage variation among protein-coding genes in and across the plastomes [84]. Moreover, compositional bias has been shown to determine the codon usage variation amidst genes in most AT or GC-rich organisms [85]. Analysis of RSCU may provide a basis for studying specific mechanisms causing biased preference of synonymous codons in different species [86]. In addition, it plays a crucial role in both practical and theoretical studies on the basics of molecular biology [87].

Phylogenetic and taxonomic implications

Phylogenetic analyses using chloroplast genome data have been used successfully to infer the evolutionary relationships among angiosperms [30, 81, 83]. Phylogenetic studies of Orchidaceae using complete plastomes are in a rather early stage due to paucity of plastome sequences. However, the relationships among major orchid lineages determined using whole plastomes (species tree) agree well to the large-scale phylogenetic studies of Orchidaceae using two or three genes (gene tree). Therefore, by sequencing more Orchidaceae complete plastomes can help resolve the pressing phylogenetic problem. Molecular datasets comprising of protein-coding genes, non-coding regions, and hypervariable regions have been used to infer major phylogenetic relationships between major orchids clades [88]. However, there are numerous uncertainties about the phylogenetic placement of several subtribes and genera. This knowledge gap is caused by a lack of both taxonomic and genomic sampling efforts required to cover all major orchid clades (subtribes/groups of genera) [89]. In this study the relationships

among the *Calanthe* alliance genera included in our phylogenetic assessment are generally consistent with recent studies [6, 15], although there are a few differences.

Previous studies on the *Calanthe* group based on morphological characteristics (adnation of the lip to the column) recognized *Calanthe* and *Phaius* as paraphyletic [21, 90, 91]. In addition, *P. delavayi*, which was previously included in genus *Phaius* based on its floral morphology by Pridgeon [14], was later classified as a member of genus *Calanthe* based on molecular evidence (ITS and cpDNA) by Zhai [15]. These findings are in agreement with results from our study as further discussed in the subsequent section.

In the present study, *Phaius* species (excluding *P. delavayi*) form two clades. The first clade includes *Phaius*: *P. tankervilleae* and *P. hainanensis*, while the second divergent clade comprised only one species of *Phaius*: *P. flavus*. These results are consistent with those of Zhai [15], who were the first to report the divergence within *Phaius*, excluding *C. delavayi*/*P. delavayi* based on ITS and cpDNA data. Therefore, we strongly support the proposal by Zhai's study that *Phaius* is restricted to the lineage that includes species such as *P. tankervilleae* and *P. hainanensis*. The clade consisting of *P. tankervilleae* and *P. hainanensis* is characterized by caducous floral bracts and eight pollinia in two groups separated from each other. In contrast, the second clade, which includes species such as *P. flavus*, has distinct features, including persistent floral bracts, pollinia which occur in two categories attached to a sticky substance by caudicles [21, 92]. Zhai's study suggested the inclusion of a new genus, *Paraphaius*, to encompass the lineage consisting of *P. flavus*. Results from our study on the phylogenetic position of *P. flavus* [$BP_{(ML)} = 75.6/57\%$, $PP = 0.9851$] are consistent with Zhai's in 2014. This species was clustered together with the clade supporting *Calanthe lyroglossa* of *Calanthe* sect. *Styloglossa* and *Cephalantheropsis obcordata* of *Calanthe* sect. *Cephalantheropsis*, supporting Zhai's proposal of recognizing subsection *Paraphaius* to encompass *P. flavus*. However, more sampling is required to help further resolve the phylogeny of this species adequately.

Presently, *Phaius delavayi* has become a vital species in the taxonomic studies of the *Calanthe* alliance due to its complex taxonomic history between *Calanthe* and *Phaius* [1, 17]. Previous studies have identified *Phaius delavayi* as a link between *Calanthe* and *Phaius* [16]. It is morphologically identical to *Calanthe* due to its relatively small individual, basal leaves, elongated column, and inconspicuous pseudobulbs [22]. Nevertheless, it has similar morphological features characterized in *Phaius* because of its long labellum embracing its column

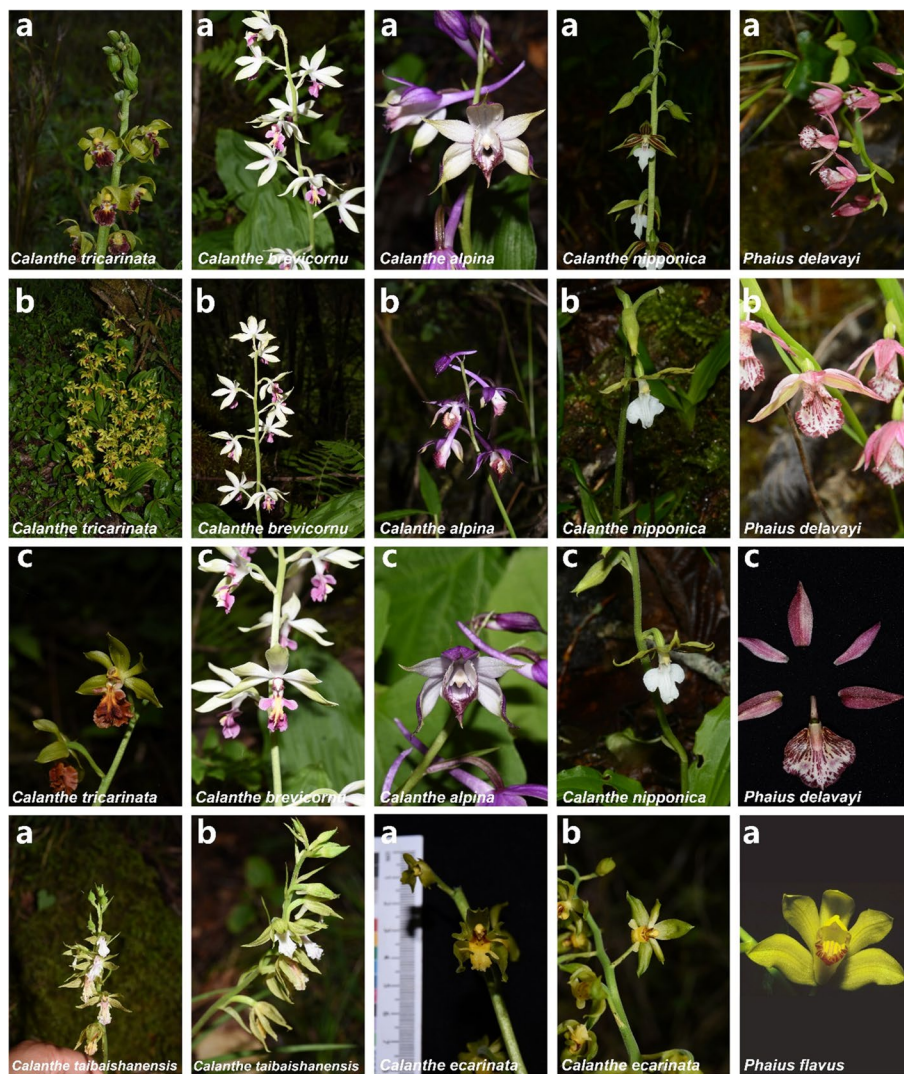


Fig. 13 Morphological characteristics of the *Calanthe* group species. Photos taken by Guangwan Hu. The picture of *P. flavus* was taken by Lourens Grobler (<http://www.orchidspecies.com/phaiusflavus.html>)

(Fig. 13) [15]. Previously, based on morphological characteristics, Chen in 1999 [21], treated this species as a member of the *sect. Calanthe*, although in their work, the Flora of China project in 2009 [91], they accepted a taxonomic placement of this species within *Phaius* as earlier proposed by Perner and Cribb [90]. However, Zhai's study on the phylogenetic relationships in the *Calanthe* alliance in China suggested that *P. delavayi* should be retained within *Calanthe* rather than within *Phaius*. In addition, this study proposed the inclusion of a new section, *Alpinocalanthe*, to accommodate this unique taxon due to its phylogenetic placement and distinct morphological characteristics, namely: small plants, persistent bracts in flowers; labellum adnate to column wings at the

basal area, a slender column; a somewhat 3-lobed labelum, circumjacent column and a disk-shaped labellum having three shortly hairy ridges. Based on the ML and BI trees using coding sequences, our results indicate *P. delavayi* [$BP_{(ML)} = 100\%$, $PP = 1.00$], is closely related to *Calanthe* than to *Phaius*, consistent with findings by Zhai [15]. In this regard, we also support the use of the name *Calanthe delavayi* instead of *Phaius delavayi*.

Calanthe sect. Calanthe is the largest infrageneric group of the genus *Calanthe*, including approximately 140 species worldwide, out of which 50 species occur in China [15, 93]. In the current study, an independent super clade encompassing all the 14 *Calanthe* group species of this section was identified. The results revealed

that the primary relationship was consistent with other studies among the *Calanthe section Calanthe* group [4, 5, 94]. The results obtained from the Maximum Likelihood (ML)/Bayesian Inference (BI) analyses revealed that the genus *Calanthe* forms a high support clade as a paraphyletic group [$BP_{(ML)} = 100\%$, $PP = 1.00$].

Our conclusion, however, on the paraphyly of the *Calanthe* group differs from that of the latest study by Chen et al. (2020) on the plastome structure and adaptive evolution of *Calanthe* s.l. which revealed the monophyly of *Calanthe* s.l. species [4]. This rigorous but taxonomically under-sampled study completely distinguished the seven species of the *Calanthe* group in their phylogenetic inference with high bootstrap support values. The discrepancies noted between this study and our present study is most likely due to large variation in the number of characters and taxa included [95]. Regardless of the fact that our matrix is character-rich and thus less prone to error caused by individual genes [96], we must highlight that our taxon sampling is highly fragmentary, and supplementary plastome sequences from poorly sampled and/or unsampled genera in the *Calanthe* alliance may result in topological changes.

Based on protein coding genes shared among all the target *Calanthe* group species, our study on the *calanthe* group chloroplast genome provides valuable genetic information on the eight newly sequenced species, highlights the power of using plastome data to resolve phylogenetic relationships between closely related species, and will facilitate future phylogenetic studies on orchids.

Conclusions

In conclusion, the complete plastomes can provide relevant information for resolving evolutionary disputes between closely related taxa. In this study, the complete chloroplast genome of 8 *Calanthe* group species were sequenced and compared. In addition, phylogenetic relationships in the *Calanthe* group were resolved with high or moderate support values. The highly divergent genes and regions of cp genomes identified in this study can be used as effective DNA barcodes in genetic diversity studies and in phylogenetic analyses. Further chloroplast genome sequencing of orchids is necessary to clarify the diversity of complete plastomes and to facilitate species identification, phylogenetic analysis, and elucidate evolutionary relationships within orchid species.

Materials and methods

Sample collection

Collection permits for sample collection were granted by the Sichuan Forestry and Grassland Administration Sichuan province, China and Yunnan Forestry and

Grassland Administration, Yunnan province, China. Fresh leaves from 8 *Calanthe* group species were collected from the Sichuan and Yunnan provinces of China (Table S1). Guang-Wan Hu, Jiixin Yang and Xiang Dong performed formal identification of the samples after collection whereby the leaf samples of *Calanthe tricarinata*, *Calanthe alpina*, *Calanthe nipponica*, *Calanthe taibaishanensis*, *Calanthe ecarinata*, *Calanthe brevicornu*, *Phaius delavayi*, and *Phaius flavus* and stored in seal-bags containing silica gel before DNA extraction. The sample specimen of each species was then deposited at the Herbarium of Wuhan Botanical Garden (HIB) with specific voucher numbers (Table S1).

Chloroplast Genome Sequencing and Assembly

The genomic DNA was extracted from about 100 micrograms of the leaves using a modified cetyltrimethylammonium bromide (CTAB) method [97]. Genome sequencing was performed using the Illumina platform at Novo gene Company (Beijing, China), followed by filtration of low-quality data and adaptors and assembly of the clean data was obtained using GetOrganelle-1.6.2 software [98], using *Calanthe triplicata* (NC_024544) as the reference genome. Bandage software was then used to check the final results of the assembled genome after manual corrections. The optimal result was selected, after which manual adjustment of these results was also made. Lastly, inverted repeat regions were identified using Geneious Prime 2019.2.1 (<https://www.geneious.com>).

Genome annotation

The annotation of the assembled genomes was performed using GeSeq online tool with default settings [99], followed by confirmation of tRNAs annotations using the tRNAscan-SE [100]. The Plastid Genome Annotator (PGA), a standalone command-line annotation tool, validated the *Calanthe* cp genomes [101]. The gene map of the complete cp genomes was drawn using OrganelleGenome DRAW software [102] (Fig. 1). The annotated complete chloroplast genomes were submitted to the GenBank database with accession numbers as follows: *Calanthe alpina* (OL322023), *C. brevicornu* (OL348396), *C. ecarinata* (OL348397), *C. nipponica* (OL348398), *C. taibaishanensis* (OL351366), *C. tricarinata* (OL351367), *Phaius delavayi* (OL351368), and *P. flavus* (OL351369).

Genome comparison and sequence divergence

The IRscope [103] was used in the comparison of the border junctions of inverted repeat (IR), small single copy (SSC), and large single copy (LSC) regions. Using Shuffle-LAGAN mode, the mVISTA software [104] was used to

compare and visualize the complete chloroplast genomes of the eight species with *C. nipponica* as the reference. Additionally, all the 8 *Calanthe* alliance cp genome sequences were aligned using MAFFT v7.409 [105]. Further, we performed a sliding window analysis to evaluate the variability (Pi) over the plastomes using DnaSP v5.10 [106] at 600 base pairs window length and 200 base-pair step size.

Repeat structure and Simple Sequence Repeats (SSRs) analysis

The visualization of forward, palindrome, reverse, and complement repeats in the *Calanthe* group genome was conducted using REPuter [107], with the minimum repeat size being set at 30 bp, maximum at 50 bp and sequence identity of no less than 90% (hamming distance= 3). Identification of simple sequence repeats (SSRs) was performed using MISA (<https://webblast.ipkgatersleben.de/misa/index.php>) [108], with the minimum number of repeats as follows: 10 for mono-, 5 for di-, 4 for tri-, and 3 for tetra-, 3 for penta-, and 3 for hexanucleotide SSRs.

Relative synonymous codon usage

All the protein-coding genes for the combined genomes were extracted using MEGA 7 [109] software which was then used to calculate the relative synonymous codon usage (RSCU) ratio. RSCU values >1 represent frequently used codons than expected, while values <1 signify the opposite. Codons having no preference value are set to 1.00.

Phylogenetic analysis

Phylogenetic relationship analysis was conducted using 73 PCGs extracted from the complete cp genome sequences of the 8 *Calanthe* group taxa mentioned above, with one outgroup, *Preptanthe rubens* (NC_050869) and thirteen previously sequenced members of the *Calanthe* alliance downloaded from the NCBI database (Table S2). Multiple sequence alignment of the 22 complete cp genome sequences was performed using MAFFT with default parameters. The best fit model was identified using the Model Finder program [110] integrated into Phylosuite. The best-fit models for the phylogenetic analysis were GTR+G, GTR+I+G, and setting (rcluster) for the concatenated alignment as implemented in ModelFinder. Phylogenetic reconstructions were performed using the maximum likelihood (ML) method using the IQ-Tree integrated in Phylosuite [111]: a GUI-based software written in python 3.6.7. The analyses were run with 1000 bootstrap replicates. Phylogenies were then

inferred by Bayesian Inference using MrBayes 3.2.6 [112] under the GTR+G+F model (2 parallel runs, 10,000,000 generations and sampled at a frequency of 1000 generations), in which the first 25% of the sampled trees were discarded as burn-in. The remaining trees were used to build a majority rule consensus tree and establish posterior probability values for each branch. Finally, the trees were refined and visualized using FigTree v1.4.4 and later combined using AI software.

Abbreviations

AT: Adenine-Thymine; Bp: Base Pair; CTAB: Cetyl Trimethyl Ammonium Bromide; cp: Chloroplast; cpDNA: Chloroplast DNA; DNA: Deoxyribonucleic Acid; GUI-based software: Graphical User Interface based software; GC: Guanine-Cytosine; IRs: Inverted Repeats; LSC: Large Single Copy; MAFFT: Multiple Alignment using Fast Fourier Transform; MEGA: Molecular Evolutionary Genetic Analysis; MISA: Microsatellite Identification Tool; ML: Maximum Likelihood; NGS: Next Generation Sequencing; NCBI: National Centre for Biotechnology Information; PCG: Protein Coding Gene; RSCU: Relative Synonymous Codon Usage; rRNA: ribosomal RNA; RNA: Ribonucleic Acid; SSC: Small Single Copy; SSR: Simple Sequence Repeat; tRNA: transfer RNA.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-022-03736-0>.

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Additional file 5.

Additional file 6.

Additional file 7.

Additional file 8.

Additional file 9.

Acknowledgments

We are grateful to Yusuf Hassan Hamza, Peninah Cheptoo Rono, and John Mulinge for their valuable suggestions.

Authors' contributions

G-WH and Q-FW-Conceptualization, Funding acquisition, Project administration, Supervision and Resources. CN, VOW, WO, ESM, and ENW- Formal analysis, Investigation and Data curation. CN, VOW, JK, EMM, MAO, XD and JXY- Methodology, Software and Validation. CN wrote the manuscript. EMM, JK, and MAO revised the manuscript. All the authors read and approved the final version of the manuscript.

Funding

This work was funded by the International Partnership Program of Chinese Academy of Sciences (151853KYSB20190027), the National Natural Science Foundation of China (31970211), Special Project of Orchid Survey of National Forestry and Grassland Administration (2019073008), and Sino-Africa Joint Research Center, CAS (SAJC202101).

Availability of data and materials

The complete chloroplast genomes generated in this study were submitted to the NCBI database (<https://www.ncbi.nlm.nih.gov/>) with GenBank accession numbers as follows; *Calanthe alpina* (OL322023), *C. brevicornu* (OL348396), *C. ecarinata* (OL348397), *C. nipponica* (OL348398), *C. taibaihanensis* (OL351366), *C. tricarinata* (OL351367), *Phaius delavayi* (OL351368), and *P. flavus* (OL351369).

All other data and material analyzed in the current study are included in the manuscript and the supplementary information files.

Declarations

Ethics approval and consent to participate

We complied with all relevant institutional, national and international guidelines with permissions from Wuhan Botanical Garden, Chinese Academy of Sciences.

Consent for publication

Not applicable

Competing interests

The authors declare no conflict of interest.

Author details

¹CAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China. ²University of Chinese Academy of Sciences, Beijing 100049, China. ³Sino-Africa Joint Research Center, Chinese Academy of Sciences, Wuhan 430074, China. ⁴Department of Biology, University of Copenhagen, Copenhagen, Denmark.

Received: 6 December 2021 Accepted: 29 June 2022

Published online: 03 August 2022

References

- Clayton D, Cribb PJ. The genus *Calanthe*. Natural History Publications (Borneo); 2013.
- Chase MW. Classification of Orchidaceae in the Age of DNA data. *Curtis Bot Mag*. 2005;22(1):2–7.
- Yukawa T, Cribb P. Nomenclatural Changes in the Genus *Calanthe* (Orchidaceae). *Bull*. 2014;40(4):145–51.
- Chen Y, Zhong H, Zhu Y, Huang Y, Wu S, Liu Z, et al. Plastome structure and adaptive evolution of *Calanthe* s.l. species. *PeerJ [Internet]*. 2020 [cited 2021 Oct 8];8. Available from: <https://pubmed.ncbi.nlm.nih.gov/33083127/>.
- Kurzweil H, Ormerod P. A review of the *Calanthe* group (Orchidaceae) in Myanmar. *Thai For Bull*. 2019;47(2):196–225.
- Kurzweil H. A precursory study of the *Calanthe* group (Orchidaceae) in Thailand. *Adansonia*. 2010;32(1):57–107.
- Kim SH, Lee JS, Lee GJ, Kim JS, Ha BK, Kim DS, et al. Analyses of genetic diversity and relationships in four *Calanthe* taxa native to Korea using AFLP markers. *Hortic Environ Biotechnol*. 2013;54(2):148–55.
- De Lakshman C, Pathak P, Rao AN, Rajeevan PK. Commercial orchids. *Commercial Orchids*. Walter de Gruyter GmbH & Co KG; 2015. 1–301 p.
- Arditti J, Ernst R, Yam TW, Glabe C. The contribution of orchid mycorrhizal fungi to seed germination: a speculative review. *Lindleyana*. 1990;5(4):249–55.
- Teoh ES. Traditional Chinese Medicine, Korean Traditional Herbal Medicine, and Japanese Kanpo Medicine. In: *Medicinal Orchids of Asia*. Springer; 2016. p. 19–31.
- Nanjala C, Ren J, Mutie FM, Waswa EN, Mutinda ES, Odago WO, et al. Ethnobotany, phytochemistry, pharmacology, and conservation of the genus *Calanthe* R. Br. (Orchidaceae). *J Ethnopharmacol [Internet]*. 2021 11 [cited 2021 Nov 17];114822. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0378874121010527>.
- Edwards ST. The Botanical register: consisting of coloured figures of exotic plants cultivated in British gardens with their history and mode of treatment. Vol. 7. 1821. https://books.google.com/hk/books?hl=en&lr=&id=7OoWAAAYAAJ&oi=fnd&pg=PP7&ots=B4uiMGHnAn&sig=kQjRuCXwMQjHf4dDccYeQURCpc&redir_esc=y&hl=zh-CN&sourceid=cnr.
- Schlechter 1872-1925 R (Rudolf). *Die Orchidaceen von Deutsch-Neu-Guinea / von Rud. Schlechter*. Dahlem bei Berlin: Gebruder Borntraeger; 1914. (Repertorium specierum novarum regni vegetabilis. Beihefte ; Bd. 1).
- Pridgeon AM, Cribb PJ, Chase MW, Rasmussen FN. Genera Orchidacearum: Epidendroideae (part one), vol. 4. New York: Oxford University Press Inc.; 2005. p. 89–115.
- Zhai J-W, Zhang G-Q, Li L, Wang M, Chen L-J, Chung S-W, et al. A new phylogenetic analysis sheds new light on the relationships in the *Calanthe* alliance (Orchidaceae) in China. *Mol Phylogenet Evol*. 2014;77:216–22.
- Xiang XG, Jin WT, Li DZ, Schuiteman A, Huang WC, Li JW, et al. Phylogenetics of tribe Collabieae (Orchidaceae, Epidendroideae) based on four chloroplast genes with morphological appraisal. *PLoS One*. 2014;9(1):e87625 Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0087625>.
- Chase MW, Cameron KM, Freudenstein JV, Pridgeon AM, Salazar G, van den Berg C, et al. An updated classification of Orchidaceae. *Bot J Linn Soc*. 2015;177(2):151–74.
- Dressler RL. The orchids: natural history and classification. Harvard University Press; 1981. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Dressler+RL+The+orchids%3A+natural+history+and+classification.+Harvard+University+Press%3B+1981.+&btnG=https://agris.fao.org/agris-search/search.do?recordID=US8110022.
- Dressler RL. Phylogeny and classification of the orchid family. Cambridge University Press; 1993. https://books.google.com/hk/books?hl=en&lr=&id=4_xL5qOVa-sC&oi=fnd&pg=PA7&dq=Dressler+RL.+Phylogeny+and+classification+of+the+orchid+family.+Cambridge+University+Press%3B+1993&ots=pja1AM71lu&sig=oV1abVDaFYVbg-K8J-mRGoJ-RLo&redir_esc=y&hl=zh-CN&sourceid=cnr.
- Freudenstein JV, Chase MW. Phylogenetic relationships in Epidendroideae (Orchidaceae), one of the great flowering plant radiations: progressive specialization and diversification. *Ann Bot*. 2015;115(4):665–81.
- Chen S, Chen S, Chen X, Tsi Z, Luo Y. Native orchids of China in colour: Elsevier; 1999.
- Chase MW, Christenhusz MJM, Schuiteman A. (2782) Proposal to conserve *Calanthe*, nom. cons., against the additional names *Phaius*, *Cyanorkis*, and *Gastorkis* (Orchidaceae, Collabieae). *Taxon*. 2020;69(6):1364–5.
- Zhao J, Xu Y, Xi L, Yang J, Chen H, Zhang J. Characterization of the Chloroplast Genome Sequence of *Acer miaotaiense*: Comparative and Phylogenetic Analyses. *Mol A J Synth Chem Nat Prod Chem [Internet]*. 2018 [cited 2021 Oct 19];23(7).
- Viljoen E, Odeny DA, Coetzee MPA, Berger DK, Rees DJG. Application of Chloroplast Phylogenomics to Resolve Species Relationships Within the Plant Genus *Amaranthus*. *J Mol Evol*. 2018;86(3–4):216–239. Available from: <https://doi.org/10.1007/s00239-018-9837-9>.
- Downie SR, Palmer JD. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: *Molecular systematics of plants*. Springer; 1992. p. 14–35.
- Smith DR. Mutation rates in plastid genomes: they are lower than you might think. *Genome Biol Evol*. 2015;7(5):1227–34.
- Huang H, Shi C, Liu Y, Mao S-Y, Gao L-Z. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol Biol*. 2014;14(1):1–17.
- Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, Chen WH, et al. 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 Available from: <https://tmu.pure.elsevier.com/en/publications/the-chloroplast-genome-of-phalaenopsis-aphrodite-orchidaceae-comp>.
- 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 [cited 2021 Oct 18];13(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/23597078/>.
- Kim K, Lee S-C, Lee J, Yu Y, Yang K, Choi B-S, et al. Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Sci Reports* 2015 51 [Internet]. 2015 28 [cited 2021 Oct 18];5(1):1–13. Available from: <https://www.nature.com/articles/srep15655>.
- Wu F-H, Chan M-T, Liao D-C, Hsu C-T, Lee Y-W, Daniell H, et al. Complete chloroplast genome of *Oncidium Gower* Ramsey and evaluation of

- molecular markers for identification and breeding in Oncidiinae. BMC Plant Biol 2010 101 [Internet]. 2010;10(1):1–12. Available from: <https://bmcplantbiol.biomedcentral.com/articles/10.1186/1471-2229-10-68>. [cited 2021 Oct 25].
32. Luo J, Hou B-W, Niu Z-T, Liu W, Xue Q-Y, Ding X-Y. Comparative Chloroplast Genomes of Photosynthetic Orchids: Insights into Evolution of the Orchidaceae and Development of Molecular Markers for Phylogenetic Applications. PLoS One [Internet]. 2014 Jun 9 [cited 2021 Oct 25];9(6):e99016. Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0099016>.
 33. Dong W-L, Wang R-N, Zhang N-Y, Fan W-B, Fang M-F, Li Z-H. Molecular Evolution of Chloroplast Genomes of Orchid Species: Insights into Phylogenetic Relationship and Adaptive Evolution. Int J Mol Sci [Internet]. 2018 Mar 2 [cited 2021 Oct 25];19(3).
 34. Palmer JD. Comparative organization of chloroplast genomes. <https://doi.org/10.1146/annurev.ge.19120185001545> [Internet]. 2003 28 [cited 2022 May 9];19:325–54. Available from: <https://www.annualreviews.org/doi/abs/10.1146/annurev.ge.19.120185.001545>.
 35. Ravi V, Khurana JP, Tyagi AK, Khurana P. An update on chloroplast genomes. Plant Syst Evol. 2008;271(1–2):101–22.
 36. Fleischmann A, Michael TP, Rivadavia F, Sousa A, Wang W, Temsch EM, et al. Evolution of genome size and chromosome number in the carnivorous plant genus *Genlisea* (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. Ann Bot [Internet]. 2014 1 [cited 2022 May 9];114(8):1651–63. Available from: <https://academic.oup.com/aob/article/114/8/1651/209026>.
 37. Kugita M, Kaneko A, Yamamoto Y, Takeya Y, Matsumoto T, Yoshinaga K. The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. Nucleic Acids Res [Internet]. 2003 Jan 15 [cited 2022 May 9];31(2):716–21. Available from: <https://pubmed.ncbi.nlm.nih.gov/12527781/>.
 38. Yamane H, Ikeda K, Ushijima K, Sassa H, Tao R. A Pollen-Expressed Gene for a Novel Protein with an F-box Motif that is Very Tightly Linked to a Gene for S-RNase in Two Species of Cherry, *Prunus cerasus* and *P. avium*. Plant Cell Physiol [Internet]. 2003 15 [cited 2022 May 9];44(7):764–9. Available from: <https://academic.oup.com/pcp/article/44/7/764/1855683>.
 39. Shahid Masood M, Nishikawa T, Fukuoka SI, Njenga PK, Tsudzuki T, Kadowaki KI. The complete nucleotide sequence of wild rice (*Oryza nivara*) chloroplast genome: first genome wide comparative sequence analysis of wild and cultivated rice. Gene [Internet]. 2004 29 [cited 2022 May 9];340(1):133–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/15556301/>.
 40. Greiner S, Wang X, Rauwolf U, Silber M V, Mayer K, Meurer J, et al. The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. Sequence evaluation and plastome evolution. Nucleic Acids Res [Internet]. 2008 Apr [cited 2022 May 9];36(7):2366. Available from: <https://pubmed.ncbi.nlm.nih.gov/18556301/>.
 41. Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc Natl Acad Sci U S A [Internet]. 1994 Oct 11 [cited 2022 May 9];91(21):9794. Available from: <https://pubmed.ncbi.nlm.nih.gov/15556301/>.
 42. Wolfe KH, Morden CW, Palmer JD. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc Natl Acad Sci U S A [Internet]. 1992 [cited 2022 May 9];89(22):10648. Available from: <https://pubmed.ncbi.nlm.nih.gov/15556301/>.
 43. Dong W, Xu C, Cheng T, Lin K, Zhou S. Sequencing angiosperm plastid genomes made easy: A complete set of universal primers and a case study on the phylogeny of saxifragales. Genome Biol Evol [Internet]. 2013 1 [cited 2021 Oct 18];5(5):989–97. Available from: <https://pubmed.ncbi.nlm.nih.gov/23595020/>.
 44. Li C, Cai C, Tao Y, Sun Z, Jiang M, Chen L, et al. Variation and Evolution of the Whole Chloroplast Genomes of *Fragaria* spp. (Rosaceae). Front Plant Sci [Internet]. 2021 14 [cited 2022 May 9];12:62293. Available from: <https://www.frontiersin.org/articles/10.3389/fpls.2021.754209/full>.
 45. DePamphilis CW, Palmer JD, Rounsley S, Sankoff D, Schuster SC, Ammiraju JSS, et al. The *Amborella* genome and the evolution of flowering plants. Science. 2013 [cited 2022 May 9];342(6165). Available from: <https://pubmed.ncbi.nlm.nih.gov/24357323/>.
 46. Xiao-Ming Z, Junrui W, Li F, Sha L, Hongbo P, Lan Q, et al. Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. Sci Rep. 2017 Dec 1 [cited 2022 May 9];7(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/3045431534/>.
 47. Wendel JF, Jackson SA, Meyers BC, Wing RA. Evolution of plant genome architecture. Genome Biol 2016 171 [Internet]. 2016 1 [cited 2022 May 9];17(1):1–14. Available from: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0908-1>.
 48. Stapley J, Santure AW, Dennis SR. Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. Mol Ecol [Internet]. 2015 1 [cited 2022 May 9];24(9):2241–52. Available from: <https://pubmed.ncbi.nlm.nih.gov/25611725/>.
 49. Casacuberta E, González J. The impact of transposable elements in environmental adaptation. Mol Ecol [Internet]. 2013 Mar [cited 2022 May 9];22(6):1503–17. Available from: <https://pubmed.ncbi.nlm.nih.gov/23293987/>.
 50. Oliver KR, McComb JA, Greene WK. Transposable elements: powerful contributors to angiosperm evolution and diversity. Genome Biol Evol [Internet]. 2013 1 [cited 2022 May 9];5(10):1886–901. Available from: <https://pubmed.ncbi.nlm.nih.gov/24065734/>.
 51. Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011;76(3):273–97.
 52. Ruhlman TA, Jansen RK. The plastid genomes of flowering plants. Methods Mol Biol [Internet]. 2014 [cited 2022 May 7];1132:3–38. Available from: <https://pubmed.ncbi.nlm.nih.gov/24599844/>.
 53. Ruhlman TA, Jansen RK. Plastid Genomes of Flowering Plants: Essential Principles. Methods Mol Biol [Internet]. 2021 [cited 2022 May 7];2317:3–47. Available from: <https://pubmed.ncbi.nlm.nih.gov/34028761/>.
 54. Blazier CC, Guisinger MM, Jansen RK. Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). Plant Mol Biol [Internet]. 2011 Jul [cited 2022 May 7];76(3–5):263–72. Available from: <https://pubmed.ncbi.nlm.nih.gov/21327834/>.
 55. Kim YK, Jo S, Cheon SH, Joo MJ, Hong JR, Kwak M, et al. Plastome Evolution and Phylogeny of Orchidaceae, With 24 New Sequences. Front Plant Sci. 2020 Feb;21(11):22.
 56. Lin CS, Chen JJW, Chiu CC, Hsiao HCW, Yang CJ, Jin XH, et al. Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. Plant J. 2017;90(5):994–1006 Available from: <https://pubmed.ncbi.nlm.nih.gov/28258650/>.
 57. Roma L, Cozzolino S, Schlüter PM, Scopece G, Cafasso D. The complete plastid genomes of *Ophrys iricolor* and *O. sphegodes* (Orchidaceae) and comparative analyses with other orchids. PLoS One [Internet]. 2018 1 [cited 2021 Oct 18];13(9):e0204174. Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0204174>.
 58. 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 May 3;0:715.
 59. Clegg MT, Gaut BS, Learn GH, Morton BR. Rates and patterns of chloroplast DNA evolution. Proc Natl Acad Sci U S A [Internet]. 1994 19 [cited 2021 Oct 25];91(15):6795–801. Available from: <https://pubmed.ncbi.nlm.nih.gov/8041699/>.
 60. Huang CY, Ayliffe MA, Timmis JN. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature [Internet]. 2003 6 [cited 2021 Oct 26];422(6927):72–6. Available from: <https://pubmed.ncbi.nlm.nih.gov/12594458/>.
 61. Gregory TR. DNA barcoding does not compete with taxonomy. Nat 2005 4347037 [Internet]. 2005 27 [cited 2021 Oct 26];434(7037):1067–1067. Available from: <https://www.nature.com/articles/4341067b>.
 62. Liu X, Chang E-M, Liu J-F, Huang Y-N, Wang Y, Yao N, et al. Complete Chloroplast Genome Sequence and Phylogenetic Analysis of *Quercus bawanglingensis* Huang, Li et Xing, a Vulnerable Oak Tree in China. For 2019, Vol 10, Page 587 [Internet]. 2019 15 [cited 2021 Oct 26];10(7):587. Available from: <https://www.mdpi.com/1999-4907/10/7/587/html>.
 63. Ren T, Yang Y, Zhou T, Liu ZL. Comparative plastid genomes of *Primula* species: Sequence divergence and phylogenetic relationships. Int J Mol Sci [Internet]. 2018 Apr 1 [cited 2021 Oct 26];19(4). Available from: <https://pubmed.ncbi.nlm.nih.gov/3045431534/>.

64. Thode VA, Lohmann LG. Comparative Chloroplast Genomics at Low Taxonomic Levels: A Case Study Using *Amphilophium* (Bignoniaceae, Bignoniaceae). *Front Plant Sci.* 2019;10:796.
65. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, et al. The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol Biol Evol.* 2006;23(11):2175–90.
66. Tyagi S, Jung J-A, Kim JS, Won SY. A comparative analysis of the complete chloroplast genomes of three *Chrysanthemum boreale* strains. *PeerJ [Internet].* 2020 [cited 2021 Oct 26];8(7). Available from: [/pmc/articles/PMC7337036/](https://pmc/articles/PMC7337036/)
67. Kim HT, Kim JS, Moore MJ, Neubig KM, Williams NH, Whitten WM, et al. Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across orchids and associated instability of the inverted repeat/small single-copy region boundaries. *PLoS One [Internet].* 2015;10(11):e0142215 Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0142215>.
68. Wang W, Messing J. High-Throughput Sequencing of Three Lemnoidae (Duckweeds) Chloroplast Genomes from Total DNA. *PLoS One [Internet].* 2011;6(9):e24670 Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0024670>.
69. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics.* 2007 8(1):1–27. Available from: <https://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-8-174>.
70. Downie SR, Jansen RK. A comparative analysis of whole plastid genomes from the apiales: Expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Syst Bot.* 2015;40(1):336–51.
71. Gao L, Yi X, Yang YX, Su YJ, Wang T. Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: Insights into evolutionary changes in fern chloroplast genomes. *BMC Evol Biol [Internet].* 2009;9(1):1–14 Available from: <https://bmcecolol.biomedcentral.com/articles/10.1186/1471-2148-9-130>.
72. Guisinger MM, Chumley TW, Kuehl J V, Boore JL, Jansen RK. Implications of the Plastid Genome Sequence of *Typha* (Typhaceae, Poales) for Understanding Genome Evolution in Poaceae. *J Mol Evol [Internet].* 2010 Feb [cited 2022 May 7];70(2):149. Available from: [/pmc/articles/PMC2825539/](https://pmc/articles/PMC2825539/)
73. Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol Biol [Internet].* 2008 Jan 31 [cited 2022 May 7];8(1):1–14. Available from: <https://doi.org/10.1186/1471-2148-8-36>.
74. Iram S, Hayat MQ, Tahir M, Gul A, Ahmed I. Chloroplast genome sequence of *Artemisia scoparia*: Comparative analyses and screening of mutational hotspots. *Plants [Internet].* 2019 1 [cited 2021 Oct 19];8(11). Available from: <https://pubmed.ncbi.nlm.nih.gov/31698805/>.
75. Wu L, Nie L, Xu Z, Li P, Wang Y, He C, et al. Comparative and Phylogenetic Analysis of the Complete Chloroplast Genomes of Three *Paeonia* Section *Moutan* Species (Paeoniaceae). *Front Genet.* 2020;0:980.
76. Li Y, Dong Y, Liu Y, Yu X, Yang M, Huang Y. Comparative Analyses of *Euonymus* Chloroplast Genomes: Genetic Structure, Screening for Loci With Suitable Polymorphism, Positive Selection Genes, and Phylogenetic Relationships Within Celastrineae. *Front Plant Sci.* 2021;0:2307.
77. Meng D, Xiaomei Z, Wenzhen K, Xu Z. Detecting useful genetic markers and reconstructing the phylogeny of an important medicinal resource plant, *Artemisia selengensis*, based on chloroplast genomics. *PLoS One [Internet].* 2019 Feb 1 [cited 2021 Oct 19];14(2). Available from: <https://pubmed.ncbi.nlm.nih.gov/30716116/>.
78. Weng ML, Blazier JC, Govindu M, Jansen RK. Reconstruction of the ancestral plastid genome in geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol Biol Evol [Internet].* 2014;31(3):645–59 Available from: <https://pubmed.ncbi.nlm.nih.gov/24336877/>.
79. Zhitao N, Shuying Z, Jiajia P, Ludan L, Jing S, Xiaoyu D. Comparative analysis of *Dendrobium* plastomes and utility of plastomic mutational hotspots. *Sci Reports* 2017. 2017;7(1):1–11 Available from: <https://www.nature.com/articles/s41598-017-02252-8>.
80. Rono PC, Dong X, Yang J-X, Mutie FM, Oulo MA, Malombe I, et al. Initial Complete Chloroplast Genomes of *Alchemilla* (Rosaceae): Comparative Analysis and Phylogenetic Relationships. *Front Genet.* 2020;0:1390.
81. Wanga VO, Dong X, Oulo MA, Mkala EM, Yang J-X, Onjalalaina GE, et al. Complete Chloroplast Genomes of *Acanthochlamys bracteata* (China) and *Xerophyta* (Africa) (Velloziaceae): Comparative Genomics and Phylogenomic Placement. *Front Plant Sci.* 2021;0:1135.
82. Lu Q, Ye W, Lu R, Xu W, Qiu Y. Phylogenomic and Comparative Analyses of Complete Plastomes of *Croomia* and *Stemona* (Stemonaceae). *Int J Mol Sci [Internet].* 2018 Aug 13 [cited 2021 Nov 4];19(8). Available from: [/pmc/articles/PMC6122011/](https://pmc/articles/PMC6122011/)
83. Huang S, Ge X, Cano A, Salazar BGM, Deng Y. Comparative analysis of chloroplast genomes for five *Dicliptera* species (Acanthaceae): molecular structure, phylogenetic relationships, and adaptive evolution. *PeerJ [Internet].* 2020;8(2):e8450 Available from: <https://peerj.com/articles/8450>.
84. Liu J, Zhu D, Ma G, Liu M, Wang M, Jia R, et al. Genome-Wide Analysis of the Synonymous Codon Usage Patterns in *Riemerella anatipestifer*. *Int J Mol Sci* 2016, Vol 17, Page 1304 [Internet]. 2016 10 [cited 2021 Nov 4];17(8):1304. Available from: <https://www.mdpi.com/1422-0067/17/8/1304/htm>.
85. Wright F, Bibb MJ. Codon usage in the G+C-rich *Streptomyces* genome. *Gene.* 1992;113(1):55–65.
86. Lü H, Zhao W-M, Zheng Y, Wang H, Qi M, Yu X-P. Analysis of Synonymous Codon Usage Bias in *Chlamydia*. 2005 [cited 2021 Nov 4]; Available from: <http://www.bioinfo-hr.org/inca>.
87. Hassan S, Mahalingam V, Kumar V. Synonymous Codon Usage Analysis of Thirty Two Mycobacteriophage Genomes. *Adv Bioinformatics.* 2009;1(2009):1–11.
88. Serna-Sánchez MA, Pérez-Escobar OA, Bogarín D, Torres-Jimenez MF, Alvarez-Yela AC, Arcila-Galvis JE, et al. Plastid phylogenomics resolves ambiguous relationships within the orchid family and provides a solid timeframe for biogeography and macroevolution. *Sci Reports* 2021 11 [Internet]. 2021 25 [cited 2022 May 7];11(1):1–11. Available from: <https://www.nature.com/articles/s41598-021-83664-5>.
89. Li MH, Zhang GQ, Liu ZJ, Lan SR. Subtribal relationships in cymbidieae (Epidendroideae, Orchidaceae) reveal a new subtribe, dipodiinae, based on plastid and nuclear coding DNA. *Phytotaxa.* 2016;246(1):037–48.
90. Perner H, Cribb P. Orchid wealth. *Alp Gard.* 2002;70(3):285–95.
91. Chen C-L, Liao F-S, Cheng S-H. PHYLOGENETIC ANALYSIS IN THE GENERA PHAIUS AND CEPHALANTHEROPSIS USING *RPL32-TRNL* MARKER. In: *Acta Horticulturae [Internet]. International Society for Horticultural Science (ISHS), Leuven, Belgium; 2010. p. 107–13. Available from: https://doi.org/10.17660/ActaHortic.2010.878.11.*
92. Li D, Chen Y, Wan D, Li ML. Common west Hubei *Tujia* medicinal Orchidaceae plants. *Zhongnan Minzu Daxue Xuebao, Ziran Kexueban.* 2009;28(Copyright (C) 2012 American Chemical Society (ACS). All Rights Reserved.):48–50.
93. Yukawa T, Cribb P. Nomenclatural changes in the genus *Calanthe* (Orchidaceae). *Bull.* 2013;65(163).
94. Zhong H, Shen L-M, Liu H-P, Liu Z-J, Wu S-S, Zhai J-W. The complete chloroplast genome of *Calanthe arcuata*, an endemic terrestrial orchid in China. *Mitochondrial DNA Part B.* 2019;4(2):2629–30.
95. Heath TA, Hedtke SM, Hillis DM. Taxon sampling and the accuracy of phylogenetic analyses. *J Syst Evol.* 2008;46(3):239 Available from: <https://www.jse.ac.cn/EN/10.3724/SP.J.1002.2008.08016>.
96. Rokas A, Carroll SB. More Genes or More Taxa? The Relative Contribution of Gene Number and Taxon Number to Phylogenetic Accuracy. *Mol Biol Evol.* 2005;22(5):1337–44 Available from: <https://academic.oup.com/mbe/article/22/5/1337/1066965>.
97. Doyle J. DNA protocols for plants. In: *Molecular techniques in taxonomy.* Springer; 1991. p. 283–93.
98. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 2020;21(1):1–31.
99. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 2017;45(W1):W6–11.
100. 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.

101. Qu X-J, Moore MJ, Li D-Z, Yi T-S. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 2019;15(1):1–12.
102. Lohse M, Drechsel O, Kahlau S, Bock R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res*. 2013;41(W1):W575–81.
103. Amiryousefi A, Hyvönen J, Pocza P. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* [Internet]. 2018;34(17):3030–1 Available from: <https://academic.oup.com/bioinformatics/article/34/17/3030/4961430>.
104. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004;32(suppl_2):W273–9.
105. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* [Internet]. 2013 Apr [cited 2021 Oct 12];30(4):772–80. Available from: <https://pubmed.ncbi.nlm.nih.gov/23329690/>.
106. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* [Internet]. 2009;25(11):1451–2 Available from: <https://academic.oup.com/bioinformatics/article/25/11/1451/332507>.
107. 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.
108. Thiel T, Michalek W, Varshney R, 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.
109. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33(7):1870–4.
110. Kalyanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587–9.
111. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, et al. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour*. 2020;20(1):348–55.
112. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61(3):539–42.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

