Plastome structure and adaptive evolution of *Calanthe* s.l. species and phylogenetic relationships within Epidendroideae (#49446)

First submission

Guidance from your Editor

Please submit by **20 Jun 2020** for the benefit of the authors (and your $200 publishing discount) .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Custom checks
Make sure you include the custom checks shown below, in your review.

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the materials page.

7 Figure file(s)
13 Table file(s)

Custom checks

DNA data checks

- Have you checked the authors data deposition statement?
- Can you access the deposited data?
- Has the data been deposited correctly?
- Is the deposition information noted in the manuscript?

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Speculation is welcome, but should be identified as such.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Plastome structure and adaptive evolution of *Calanthe* s.l. species and phylogenetic relationships within Epidendroideae

Yanqiong Chen [1, 2], **Hui Zhong** [1, 2], **Ya-Ting Zhu** [1, 2], **Yuan-Zhen Huang** [1, 2], **Sha-Sha Wu** [1, 2], **Zhong-Jian Liu** [1, 2], **Si-Ren Lan** [Corresp., 1, 2], **Jun-Wen Zhai** [Corresp. 1, 2]

[1] Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization at College of Landscape Architecture, Fujian Agriculture and Forestry University, Fuzhou, Fujian, China

[2] Fujian Ornamental Plant Germplasm Resources Innovation & Engineering Application Research Center, Fujian Agriculture and Forestry University, Fuzhou, Fujian, China

Corresponding Authors: Si-Ren Lan, Jun-Wen Zhai
Email address: lkzx@fafu.edu.cn, zhai-jw@163.com

~~Chloroplast (cp) genomics can improve our understanding of plant biology and evolution. Orchidaceae, as the largest family of flowering plants, has a complex adaptive evolutionary history that has been concerned for centuries.~~ Illumina sequencing followed by *de novo* assembly was used in this study, and the ~~cp~~ genetic information of *Calanthe* s.l. were used to investigate the adaptive evolution of this taxon. Herein, the complete cp genomes of five *Calanthe* s.l. species (*Calanthe davidii*, *Styloglossum lyroglossa*, *Preptanthe rubens*, *Cephalantheropsis obcordata*, and *Phaius tankervilliae*) were determined to examine the evolutionary pattern of plastome in the alliance. Seven *Calanthe* s.l. plastomes were compared and the results indicated that they were relatively conserved, ranging from 150,181 to 159,014 bp in length and were all mapped as a circular structure, except for the three *ndh* genes (*ndhC*, *ndhF*, and *ndhK*) lost in *C. delavayi*. The remaining six species contain identical gene orders and numbers. Each genome contains 115 unique genes, of which 80 code for proteins, 30 for tRNAs, and 4 for rRNAs. We screened nine hotspot regions, including five coding regions (*accD*, *psbK*, *ycf1*, *rpl22*, and *matK*) and four non-coding regions (*trnF*-GAA-*ndhJ*, *rps16*-*trnQ*-UUG, *trnS*-GCU-*trnG*-GCC, and *psbB*-*psbT*). Furthermore, 18 SSRs were screened as candidate DNA barcodes. As for the adaptive evolution investigation, six genes were under positive selection, namely *accD*, *ndhB*, *ndhD*, *rpoC2*, *ycf1*, and *ycf2*. Phylogenetic analysis of Epidendroideae based on 62 plastomes indicates a close relationship between *Calanthe* s.l. and *Eria*, and the phylogenetic position of 11 tribes of Epidendroideae were clarified. These results, including the new plastomes, provide resources for the comparative cp genomics, breeding, and plastid genetic engineering of orchids and flowering plants.

Plastome structure and adaptive evolution of *Calanthe* s.l. species and phylogenetic relationships within Epidendroideae

Yan-Qiong Chen[1,2], Hui Zhong[1,2], Ya-Ting Zhu[1,2], Yuan-Zhen Huang[1,2], Sha-Sha Wu[1,2], Zhong-Jian Liu[1,2], Si-Ren Lan[1,2] and Jun-Wen Zhai [1,2*]

[1] Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization at College of Landscape Architecture, Fujian Agriculture and Forestry University, Fuzhou, China

[2] Fujian Ornamental Plant Germplasm Resources Innovation & Engineering Application Research Center, Fujian Agriculture and Forestry University, Fuzhou, China

Corresponding Author:

Jun-Wen Zhai[1,2]

No.15 Shangxiadian Road, Cangshan District, Fuzhou City, Fujian Province, 350002, China

Email address: zhai-jw@163.com

Abstract

Chloroplast (cp) genomics can improve our understanding of plant biology and evolution. Orchidaceae, as the largest family of flowering plants, has a complex adaptive evolutionary history that has been concerned for centuries. Illumina sequencing followed by *de novo* assembly was used in this study, and the cp genetic information of *Calanthe* s.l. were used to investigate the adaptive evolution of this taxon. Herein, the complete cp genomes of five *Calanthe* s.l. species (*Calanthe davidii*, *Styloglossum lyroglossa*, *Preptanthe rubens*, *Cephalantheropsis obcordata*, and *Phaius tankervilliae*) were determined to examine the evolutionary pattern of plastome in the alliance. Seven *Calanthe* s.l. plastomes were compared and the results indicated that they were relatively conserved, ranging from 150,181 to 159,014 bp in length and were all mapped as a circular structure, except for the three *ndh* genes (*ndhC*, *ndhF*, and *ndhK*) lost in *C. delavayi*. The remaining six species contain identical gene orders and numbers. Each genome contains 115 unique genes, of which 80 code for proteins, 30 for tRNAs, and 4 for rRNAs. We screened nine hotspot regions, including five coding regions (*accD*, *psbK*, *ycf1*, *rpl22*, and *matK*) and four non-coding regions (*trnF*-GAA-*ndhJ*, *rps16*-*trnQ*-UUG, *trnS*-GCU-*trnG*-GCC, and *psbB*-*psbT*). Furthermore, 18 SSRs were screened as candidate DNA barcodes. As for the adaptive evolution investigation, six genes were under positive selection, namely *accD*, *ndhB*, *ndhD*, *rpoC2*, *ycf1*, and *ycf2*. Phylogenetic analysis of Epidendroideae based on 62 plastomes indicates a close relationship between *Calanthe* s.l. and *Eria*, and the phylogenetic position of 11 tribes of Epidendroideae were clarified. These results, including the new plastomes, provide resources for the comparative cp genomics, breeding, and plastid genetic engineering of orchids and flowering plants.

Introduction

~~Chloroplasts (cp) play an essential role in sustaining life on earth. The processes of photosynthesis and oxygen release in chloroplasts can convert solar energy into carbohydrates (Daniell et al., 2016). Since the first cp genome was obtained from tobacco (*Nicotiana tabacum*) in 1986 (Shinozaki et al., 1986), over 3,000 complete cp genome sequences of plants have been made available in the National Center for Biotechnology Information (NCBI). We can gain insight from cp genomes, enhancing our understanding of plant biology and diversity. Complete cp genome sequences have been broadly utilized in reconstructing phylogenetic relationships, revealing considerable variation within and between plant species (Luo et al.,2014; Song et al., 2017; Dong et al., 2018; Gao et al., 2019).~~

Orchidaceae (orchids) is the largest family in angiosperms, with its fascinating biodiversity attracting the interest of numerous botanists. The first plastid genome of orchids (*Phalaenopsis aphrodite*) was published in 2006 (Chang et al., 2006). In total, over 200 cp genomes of orchids have been ~~assembled~~, including five subfamilies and 51 genera. In general, cp genomes are relatively conserved in land plants, but huge divergence was found among

different orchid species. Orchids encompass all life forms of plants, including both heterotrophic and autotrophic species, namely terrestrial, epiphytic, and saprophytic (also called mycoheterotrophic). The evolutionary direction of the life forms of orchids can be roughly outlined from terrestrial to epiphytic, while saprophytic species independently evolved several times (Sosa et al., 2016). Recent cp genome analysis of orchid species has focused on comparisons with partially and fully mycoheterotrophic species (Feng et al., 2016; Barrett and Kennedy, 2018; Unruh et al., 2018; Yuan et al., 2018). We chose *Calanthe*, which has a life form transition that is the opposite to that of the orchid family (from obligate epiphytic to hemi-epiphytic and then terrestrial), as material to investigate the cp genome structure, sequence variation, and adaptive evolution.

Calanthe, the largest genus of the Collabieae tribe (Epidendroideae, Orchidaceae), has a pantropical distribution, being widely distributed in tropical and subtropical Asia, Australia, Madagascar, Africa, Central and South America, and the Caribbean (Chen et al., 2009; Clayton and Cribb, 2013). Since the establishment of *Calanthe* in 1821 (Ker Gawler, 1821), the genus has undergone multiple intra-generic taxonomic revisions lasted for centuries. For example, Lindley (1855) set two subgenera according to the spur length, Bentham and Pridgeon defined the genus in different criterion based on the pseudobulb and floral characteristics (Bentham, 1881; Pridgeon et al., 2005). For the taxonomic classification, morphological data are limited, so genetic information suggests that the origin of *Calanthe* is polyphyletic and includes its relatives (*Cephalantheropsis* and *Phaius*) (Yukawa and Ishida, 2008; Xiang et al., 2014; Zhai et al., 2014). The redefinition of the three genera seem inevitable, and *Calanthe* may be subdivided into three genera, *Calanthe*, *Styloglossum*, and *Preptanthe* (Yukawa and Cribb, 2014; Zhai et al., 2014). These genera, *Calanthe*, *Phaius*, *Styloglossum*, *Cephalantheropsis*, and *Preptanthe*, form an independent alliance (which we refer to as *Calanthe* s.l.) within Epidendroideae (Orchidaceae). This alliance can be easily distinguished from other taxa in this family, characterised by plicate leaves, similar sepals and petals, and eight waxy pollinia. However, the latest Orchidaceae classification retained the independent genus status of polyphyly *Calanthe* (Chase et al., 2015). In *Calanthe* s.l., we only see three *Calanthe* species (Yang et al., 2014; Dong et al., 2018; Chen et al., 2019) from the cp genome assembly, while its relatives remain vacant. To better understand their phylogenetic relationship, it is necessary for us to identify discrepancies in the genetic information of the major clade of *Calanthe* s.l.

In the current study, we assembled and annotated the plastid genomes of five *Calanthe* s.l. species for the first time. Additionally, we published and compared *Calanthe* s.l. cp genomes with the aim to: (1) understand the genetic variation within the *Calanthe* s.l. cp genomes, (2) identify the characteristics of the plastomes structure, sequence divergence, variant hotspot regions, repeat regions, and exam them as candidate molecular marker for species classification and further species evolution studies, (3) assess the selective pressure among *Calanthe* s.l. species by identifying genes underlying positive selection, and (4) evaluate the phylogenetic relationship within Epidendroideae.

Materials & Methods

Plant Materials, DNA Extraction and Sequencing

Specimens of five species, *Calanthe davidii*, *Stylo[illegible]ssum lyroglossa*, *Preptanthe rubens*, *Phaius tankervilliae*, and *Cephalantheropsis obcordate*, were growing in the Fujian Agriculture and Forestry University, Fujian province, China. The modified version of the CTAB method (Doyle and Doyle, 1987) was applied to extracting the total genomic DNA. We constructed the short-insert (500 bp) pair-end (PE) library and the sequencing was conducted by the Beijing Genomics Institute (Shenzhen, China) on Illumina Hiseq 2500 platform, with a read length of 150 bp. At least 5 Gb clean data were obtained for each species.

Genome Assembling and Annotation

We used the GetOrganelle toolkit (Jin et al., 2019), with 18 orchid species as references, to assemble the cp genomes and then used Bandage (Wick et al., 2015) to manually finalize the complete target genomes. Geneious primer v2019.03 (Kearse et al.,2012) was used to annotate the cp genomes. OrganellarGenomeDRAW (OGDRAW) version 1.3.1 (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) (Greiner et al., 2019) was used to visualize the structural features of the seven species.

Genome Comparison and Analysis

Combined with two published *Calanthe* cp genomes (*C. ~~triplicate~~*, KF753635 and *C. delavayi*, MK388860), seven complete cp genomes of *Calanthe* s.l. provide the possibility of comparative analysis within the relatives. Mauve Alignment (Darling et al., 2004) was employed to analyse the cp DNA rearrangement of seven species. The junction regions between the IR, SSC, and LSC of seven species were compared using the online program IRscope (https://irscope.shinyapps.io/irapp/)(Amiryousefi et al., 2018). Meanwhile, the sequence identity of the seven species were compared and plotted using the program mVISTA (http://genome.lbl.gov/vista/mvista/submit.shtml) with Shuffle-LAGAN (Brudno et al., 2003; Frazer et al., 2004). Then, region with greater level of variation was selected as mutational hotspots. The coding sequences (CDS) and intergenic spacers (IGS) were extracted by PhyloSuit (Zhang et al., 2019) and the DNA sequence variation was calculated by IQ-tree (Nguyen et al., 2015), including variable and parsimony informative sites (pi).

Repeat Sequence Analysis

The online software REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) was used to identify the repeat sequences (Kurtz, 2001), including palindromic direct and reverse repeats. The parameters were setting as: (1) The maximum and minimum computed repeat sizes were limited to 50 and 30, respectively, (2) Hamming distance of 3. Tandem repeat sequences were identified with the tandem repeats finder (Benson, 1999). The alignment parameters set as 2 match, 7 mismatch, and 7 indels. Identify repeat with condition in 80 minimum alignment score, maximum period sizes in 500 bp, and maximum TR array sizes in 2 millions. Perl script MISA (MIcroSAtellite identification tool) applied to detecting the simple sequence repeats (SSRs) loci

of cp genome, with a threshold of moni-, di-, tri-, tetra-, penta-, and hexa-nucleotide SSRs, respectively (Thiel et al., 2003).

Gene Selection Pressure Analysis

We analyzed all CDS gene regions (77 genes), except for *ndhC*, *ndhF*, and *ndhK*, as these regions were lost by *C. delavayi*. The maximum likelihood tree was constructed based on the aligned concatenated CDS genes dataset of the seven species using IQ-tree (Nguyen et al., 2105). We calculated the non-synonymous (dN) substitution, synonymous (dS) substitution rates, and dN/dS ratio (ω) in order to estimate the selection pressure with site-specific model (the option of the analyses was set to seqtype = 1, model = 0, and Nssites = 0, 1, 2, 3, 7, and 8) with CodeML program in PAML 4.9 (Yang and Nielsen, 2002; Yang, 2007). The likelihood ratio tests (LRTs) P-values of under three pairs of site models were calculated to detect positive selection ($p < 0.01$), including: M0 (one-ratio) vs. M3 (discrete), M1 (nearly neutral) vs. M2 (positive selection), and M7 (β) vs. M8 (β and ω).

Phylogenomic Analysis

We downloaded 55 Epidendroideae species complete cp genome sequences representing 11 of 16 tribes and two Orchidoideae subfamily species from NCBI. These cp sequences (Accession number: **Table S1**) combined with five *Calanthe* s.l. complete cp genome sequences generated from this study were used to construct phylogenetic relationships. We limited our date to protein-coding genes, primarily on account of their relatively slow evolution rate. PhyloSuit (Zhang et al., 2019) was implemented to extract the CDS gene regions and the alignment of these 68 protein-coding genes was conducted by MAFFT v7.407 (Katoh and Standley, 2013). A maximum likelihood tree was constructed using IQ-tree with 1000 bootstrap replicates. The optimal nucleotide substitution model was found with ModelFinder module in IQ-tree (Nguyen et al., 2105).

Results

Features of the Chloroplast Genome

We reassembled and annotated the plastid genome of *C. davidii* (we use the abbreviations D1 (MN708353) and D2 (MG925365) to represent the different versions of the *C. davidii* cp genome). In the present study, the reassembled version of *C. davidii* (D1) is hugely different to the published plastid genome. A comparison between these two genomes indicated that the gene numbers and orders are identical, while huge differences were found in the genome size, gene length, and GC content (**Tables 1 and 2**). We used the D1 version of *C. davidii* for the following analysis.

The genome size of D1 is 5,385 bp longer than D2, with the difference mainly concentrated in the IR region, especially *ndh* genes, which encode the subunits of the nicotinamide adenine dinucleotide (NADH) dehydrogenase-like complex proteins (Yamori and Shikanai, 2016). The

assembling and annotating of D2 used *Dendrobium nobile* as the reference (Dong et al., 2018), which has a discrepancy with *C. triplicata* we used as reference in this study, especially *ndh*, e.g., the *ndhC*, *ndhI*, *ndhK*, and *nad6* genes were lost in *D. nobile* but exist in *C. triplicata*. Moreover, *D. nobile* showed a distant relationship with *C. davidii*, *Dendrobium* in the Malaxideae tribe were far away from the Collabieae tribe, to which *C. davidii* belongs (Chase et al., 2015). Considering the conservatism of the cp genome, and the reasons mentioned above, we may infer that the results of this study are more accurate.

The plastome of the seven species of *Calanthe* s.l. ranged from 150,181 (*C. delavayi*) to 159,014 bp (*C. davidii,* D1), with a GC content of 36.60%–37.00% (**Table 1**). They all shared the common structure: a pair of IRs (IRa and IRb) (25,216–26,617 bp), separated by LSC region (83,411–87,857 bp) and SSC region (16,338–18,589 bp) (**Table 1** and **Fig. 1**). The gene numbers, orders, and names were very similar among *Calanthe* s.l., except for the three gene losses of *C. delavayi*, namely *ndhC*, *ndhF*, and *ndhK* (**Fig. 2, 3, 4**).

The genomes encoded 133–136 genes, of which 112–115 were unique genes, containing 77–80 protein-coding genes, 30 tRNA genes, and four rRNA genes. The 18 genes were found duplicated in the IR regions, including three type of genes, namely coding genes, tRNA genes and rRNA genes (**Table 3**). In the seven cp genomes, 15 genes contained one intron (containg six tRNA and nine protein-coding genes) and three genes (*ycf3*, *clpP*, and *rps12*) contained two introns. There has a trans-spliced gene, *rps12*. The gene across the two areas with the 5'-end exon lie in the LSC region and the intron and 3'-end exons located in the IR region. Overlapping sequences were found in three pair of genes: *trnK*-UUU/*matK*, *atpE*/*atpB*, and *psbD*/*psbC*.

We examed the expansion and contraction of junction area between the single-copy regions and the pair of IR regions for the seven *Calanthe* s.l. species (**Fig. 5**). Among the seven species, the gene positions of four borders (LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC) have different types. The LSC/IRb border has three situations. Firstly, in *C. triplicata*, *C. delavayi*, *P. tankervilliae*, and *C. obcordata* the *rpl22* gene overlapped in the LSC/IRb region; secondly, in *C. davidii* and *S. lyroglossa* the *rpl22* gene was located in the LSC region, 9–47 bp away from the IRb region; the third situation is in *P. rubens* where the *rps19* gene overlapped the border instead of the *rpl22* gene. While the IRb/SSC junction regions were relatively stable in the seven species, the *ndhF* gene crossed the border of six of the seven species, except for *C. delavayi*, due to its *ndhF* gene loss and the nearest gene (*trnN* (in IRb)) is 361 bp away from the SSC region. SSC/IRa and IRa/LSC are both very conserved among the seven plastomes. The *ycf1* gene strode the SSC/IRa bounder, having 35–1042 bp into the IRa region. The distance between *psbA* and the IRa/LSC junction ranged from 91 to 239 bp.

Repeats in cp genomes were detected. *P. rubens* had the greatest number of long repeat and *C. tripliacata* had the greatest number of tandem repeat region (**Tables S2 and S3**). A total number of 49–73 SSRs were found in the seven cp genomes. Mononucleotide, dinucleotide, trinucleotide, tetranucleotide, and pentanucleotide SSRs were all discovered in seven *Calanthe* s.l. species (**Fig. 6**). Hexanucleotide SSRs were found except in *C. davidii* and *P. rubens*. In all seven species, mononucleotide repetitions were accounts for more than half of all (57.89%,

56.67%, 54.79%, 61.22%, 54.68%, 53.73%, and 60%). The IGS region contains the largest number of SSRs, with 231 identified, while 175 were identified in the CDS and 64 in the coding sequence introns. In particular, all mononucleotide SSRs belonged to A or T types and the richness of A or T were found in the major of dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide SSRs (**Table S4**).

Positive Selection Analysis of Protein Sequences

The selective pressure on seven species of *Calanthe* s.l. were assessed by using the site model in the PAML program. We compared the rates of non-synonymous (dN) and synonymous (dS) substitutions and calculated the dN/dS ratio (ω) for 77 common protein-coding genes between the seven species. We identified the six genes under positive selection (**Tables S5 and S6**). These genes included *accD*, *ndhB*, *ndhD*, *rpoC2*, *ycf1*, and *ycf2*.

Phylogenomic Analysis of Epidendroideae

We used 68 chloroplast protein coding genes to evaluate the phylogenetic relationships within Epidendroideae, the alignment information was shown in **Table S7 and S8**. The best substitution model used for this dataset was GTR+F+R4. Almost all nodes of the ML tree were strongly supported by the bootstrap values, only Vandeae-Cymbidieae clade with the weak bootstrap value in 49 (**Fig. 7**). Independent positions of 11 tribes were detected. *Calanthe* s.l. has closely relationship with *Eria.* The seven *Calanthe* s.l. species were classified into three major clades, with all the *Calanthe* s.l. species composing a monophyly. *Preptanthe* is ~~located at the basal position and is the~~ sister ~~taxon~~ to the ~~clade formed by the~~ remaining ~~other~~ four genera. *Cephalantheropsis* and *Styloglossum* are clustered into a sister clade to *Phaius*, which forms the sister group to *Calanthe*.

Discussion

Chloroplast Sequence Variation

The whole cp genomes of five *Calanthe* s.l. species were determined, which had never been done for *Phaius*, *Cephalantheropsis*, *Styloglossum* and *Preptanthe* and was a reconstruction for *C. davidii*. Comparative analyses of the cp genomes of seven *Calanthe* s.l. species showed highly conserved structures and genes. The published Orchidaceae cp genome size ranges from 35,304 (*Gastrodia elata*) to 178,131 bp (*Cypripedium formosanum*). This discrepancy in the cp genome size can be explained by the different life forms, as the heterotrophic plants that do not photosynthesise have lost the photosynthesis-related genes. While it is common to see a lack of functional *ndh* genes in some autotrophic orchids (Yang et al., 2013; Lin et al., 2017; Roma et al.,2018). In the present study, only *C. delavayi* had lost the *ndh* genes (*ndhC*, *F*, and *K*) in *Calanthe* s.l.. Although the transfer of the *ndh* genes from the cp genome to the mitochondrial (mt) genome was detected in some orchids, there is no direct evidence show these transfers were linked to the losses of *ndh* in the cp genome (Lin et al., 2015). The mechanisms for the complex

deletion and truncation of the genes encoding NADH dehydrogenase subunits in orchids remain unclear.

Although the overall genomic structures and gene orders of *Calanthe* s.l. are highly conserved, like other orchids, significant differences at the IR/SSC junction area were detected. The different expansion and contraction situation of the IR junction area will cause the genome size differentiation in plant lineages. In present study, among the four boundaries, only LSC/IRb shows three types in the seven species, while the remaining three (IRb/SSC, SSC/IRa, and IRa/LSC) are conservative and stable. Because of the *ndhF* gene loss in *C. delavayi*, IR contraction was detected. Previous research has pointed out that the deletion of the *ndh* genes has great influence on the instability of the IR/SSC boundary in orchids (Kim et al.,2015; Niu et al., 2017). The position of the boundary, especially the expansion and contraction of the region, could shed light on the evolution of the lineage. However, in *Calanthe* s.l., our observations would not provide the information required to elucidate the evolutionary relationships of the taxa, and whether it can benefit from adaptation require further investigation, thus additional sampling of *Calanthe* spp. and related genera will allow for explicit tests.

Molecular Markers and Hotspot Regions

~~Coding regions and conserved sequences of the plastome are widely used for phylogenetic infer at higher taxonomic levels (family or genus) (Givnish et al., 2015; Jansen et al., 2007).~~ Plastid genomes are ideal resources for selecting the mutational hotspots of various lineages and used for intraspecies discrimination and phylogenetic studies at the species level (Ahmed et al., 2013; Liu et al., 2018). At present, some plastid DNA fragments are used in the taxonomy of *Calanthe* s.l., for instance *matK*, *rbcL*, and intergenic spacer *trnL-trnF* (Zhai et al., 2014; Guo et al., 2107), but they could not provide sufficient phylogeny signal to establish the high-resolution phylogeny relationship for classification of related taxa, especially some infrageneric taxa whose taxonomic classification status are unclear. Our alignment screened the top nine loci that most likely contain the highest degrees of genetic variability in *Calanthe* s.l., namely five CDS (*accD*, *psbK*, *ycf1*, *rpl22*, and *matK*) and four IGS regions (*trnF*-GAA-*ndhJ*, *rps16*-*trnQ*-UUG, *trnS*-GCU-*trnG*-GCC, and *psbB*-*psbT*), which can be useful in species-level phylogenetic studies of *Calanthe* s.l..

Simple sequence repeats are short (1–5 bp) repeat motifs that are tandemly repeated for varying numbers of times (Kantety et al.,2002). Because of its extensively dispersal in genomes, they are widely utilized in population genetics and molecular evolution studies (Guichoux et al., 2002). SSRs can provide interspecific polymorphisms, which are effective markers in population genetic analysis. We identified 18 SSRs as polymeric SSRs between *Calanthe* s.l. species (**Table 4**). Hotspot regions and SSRs derived from the cp genome can serve as valuable tools. These candidate DNA barcodes are helpful in elucidating the evolutionary relationships and plant identification of *Calanthe* s.l.

Adaptive Evolution Analysis

Calanthe s.l. are pantropical in their distribution, with high geographic and ecological diversity, from obligate epiphytic to hemi-epiphytic and terrestrial, ranging from sea level to alpine

mountain areas. However, the majority live under tropical ~~woods and~~ forests, ~~very~~ often in deep shade (Clayton and Cribb, 2013; Stone and Cribb, 2017). To better understand the evolutionary history of these groups, the analysis of the genetic diversity and adaptive evolution of *Calanthe* s.l. was essential. Positive selection genes played an important part in the adaption to various environments. Six genes were under positive selection in seven Calanthe s.l. species. Out of these genes one function as subunit of acetyl-CoA-carboxylase (*accD*), two NADH dehydrogenase subunit genes (*ndhB* and *ndhD*), one RNA polymerase subunit (*rpoC2*), and two genes whose function was uncertain (ycf1 and ycf2). Expect for the *ndh* genes, the remaining four positively selected genes were also detected in orchid species (Dong et al., 2018), and these four genes may has played a significant role in the adaptive evolution history of *Calanthe* s.l. and specific roles need to be further studied.

The function of the plastid *accD* gene has been reported as an essential for plant leaf development (Kode et al.,2005). We detected the two positively selected sites in *accD* genes for *Calanthe* s.l.. The *ndh* gene family in the cp genome is involved in photosynthesis. The *ndhB* and *ndhD* genes possessed 12 and 10 positively selected sites, respectively. Although there is the viewpoint that NDH activity may not be required in some plants (Yang et al., 2013), these two genes may play an important role in the adaptation of *Calanthe* s.l. species to deep shade environments. In addition, the *rpoC2* (RNA polymerase subunit C2) gene was crucial for gene transcription (Xie et al., 1989), and only one site was positively detected in our study. Cp genomes contain a number of uncertain genes. The *ycfs* (hypothetical cp open reading frame) gene has great application potential for elucidating plant phylogeny (Neubig et al., 2009; Huang et al., 2010; Dong et al., 2015). As the largest genes in *ycfs*, *ycf1* and *ycf2* have shown positive selection in one and eight sites, respectively, and the phenomenon extends to the many plant lineages, including orchids (Greiner et al., 2008; Huang et al., 2010; Carbonell-Caballero et al., 2015).

Phylogenetic Analysis

We reconstructed the relationship at the tribe level of the epidendroids to infer a more robust phylogenetic tree. The subfamily Epidendroideae account for 70% genera of the whole family, including approximately 650 genera and 18,000 species and classified into 16 tribes and 28 subtribes (Pridgeon et al., 2005; Chase et al., 2015). Whereas much progress has been made resolving the composition of tribe and subtribe of epidendroids , understanding relationships among these has remained challenging. In this study, we retrieved 55 plastome genome sequences to present 11 tribe of Epidendroideae, four in “Lower Epidendroideae” (Neottieae, Sobralieae, Gastroideae, Nervillieae) and seven in “Higher Epidendroideae” (Arethuseae, Malaxideae, Podochileae, Collabieae, Epidendreae, Cymbidieae and Vandeae) (**Fig. 7**).

Within the lower epidendroids, support for the relationships of these four tribes are very high. Neottieae in the basal node and with the strongly supported as sister to the remainder tribes which consist with the previous studies (Xiang et al., 2012; Freudenstein et al., 2015). Sobralieae is the only tribe contains epiphytes in “Lower Epidendroideae” sister to the clade formed by tribe Nervilieae and Gastrodieae. In higher epidendroids, we inclued all seven tribes in present

ananlysis, the topology of the basal tribe (Arethuseae and Malaxideae) is consistene with previous molecular analyses (Freudemstein et al., 2015; Li et al., 2016). The remaining five tribes have various topologies in different dataset. Chase et al (2015), with the topology in (Cymbidieae(Epidendreae(Collabidea(Podochileae,Vandeae)))), while Freudenstein et al (2015) in ((Cymbidieae,Epidendrea),((Vandeae,Podochileae),Collabieae)) by using *rbcL* and *matK*. In present study we obtain the same topology with the result of Xiang et al(2012), (((Vandeae,Cymbidieae),Epidendreae),(Collabieae,Podochileae)). The genus *Calanthe, Phaius*, *Paraphaius*, *Cephalantheropsis, Styloglossum* and *Preptanthe* were formed a clade and placed in the tribe Collabieae of Epidendroideae, our result were consistent with result of previous study (Pridgeon et al., 2005; Zhai et al., 2014; Yukawa and Cribb, 2014; Chase et al., 2015). Although we retrieved the highly resolution phylogeny of the Epidendroideae, to obtain a reliable inference a comprehensive sampling of the subfamily is necessary as limited taxon sampling can result in different tree topologies (Puslednik and Serb, 2008).

Conclusions

In this study, We assembled and analyzed five new complete chloroplast genome sequences of *Calanthe* s.l. and compared them with other *Calanthe* species for the first time. The annotation and comparision within *Calanthe* s.l. species showed conservative in gene sequence, GC content and genomic composition. The repeated sequences,18 microsatellites and nine highly hotspot regions (*accD*, *psbK*, *ycf1*, *rpl22*, *matK*, *trnF*-GAA-ndhJ, *rps16-trnQ*-UUG, *trnS*-GCU-*trnG*-GCC, and *psbB-psbT*) were identified in *Calanthe* s.l. chloroplast genome. Six positive selected genes were detected, these genes will lead to understanding the adaptation of *Calanthe* s.l. speices to deep shade environments.The Phylogenetic relationships of the Epidendroideae species inferred from chloroplast genomes obtained high support. The study will help to resolve the phylogenetic relationship and understand the adaptive evolution of *Calanthe*. It will also provide genomic resources and potential markers suitable for future species identification and speciation studies of the genus.

References

Ahmed, I., Matthews, P. J., Biggs, P. J., Naeem, M., McLenachan, P. A., and Lockhart, P. J. (2013). Identification of chloroplast genome loci suitable for high-resolution phylogeographic studies of *Colocasia esculenta* (L.) Schott (Araceae) and closely related taxa. *Mol. Ecol. Resour*. 13, 929–937.

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

Barrett, C. F., and Kennedy, A. H. (2018). Plastid genome degradation in the endangered, mycoheterotrophic, North American orchid *Hexalectris warnockii*. *Genome Biol. Evol*. 10, 1657–1662.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 27, 573–580 .

Bentham, G. (1881). Notes on Orchidceae. *Bot. J. Linn. Soc*. 18, 281–360.

Bi, G., Mao, Y., Xing, Q., and Cao, M. (2018). HomBlocks: A multiple-alignment construction pipeline for organelle phylogenomics based on locally collinear block searching. *Genomics*. 110, 18–22.

Brudno, M., Malde, S., Poliakov, A., Do, C. B., Couronne, O., Dubchak, I., et al. (2003). alignment: finding rearrangements during alignment. *Bioinformatics*. 19, i54–i62.

Carbonell-Caballero, J., Alonso, R., Ibañez, V., Terol, J., Talon, M., and Dopazo, J. (2015). A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. *Mol. Biol. Evol.* 32, 2015–2035.

Chang, C. C., Lin, H. C., Lin, I. P., Chow, T. Y., Chen, H. H., Chen, W. H., et al. (2006). The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol. Biol. Evol*. 23, 279–291.

Chase, M. W., Cameron, K. M., Freudenstein, J. V., Pridgeon, A. M., Salazar, G., Berg, C., et al. (2015). An updated classification of Orchidaceae. *Bot. J. Linn. Soc.* 177, 151–174.

Chen, Y. Q., Lan, S. R., Liu, Z. J., and Zhai, J. W. (2019). The complete chloroplast genome sequence of *Calanthe* delavayi (Orchidaceae), an endemic to China. *Mitochondrial DNA Part B*. 4, 1562–1563.

Clayton, D., Cribb, P. (2013) *The Genus Calanthe*, Natural History Publications.

Daniell, H., Lin, C. S., Yu, M., and Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17, 134.

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403.

Dong, W. L., Wang, R. N., Zhang, N. Y., Fan, W. B., Fang, M. F., and Li, Z. H. (2018). Molecular evolution of chloroplast genomes of Orchid species: insights into phylogenetic relationship and adaptive evolution. *Int. J. Mol. Sci.* 19, 716.

Dong, W. P., Xu, C., Li, C. H., Sun, J. H., Zuo, Y., Shi, S., et al. (2015). *ycf1*, the most promising plastid DNA barcode of land plants. *Sci. Rep.* 5, 1–5.

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

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

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

403 Freudenstein, J. V., & Chase, M. W. (2015). Phylogenetic relationships in Epidendroideae
404 (Orchidaceae), one of the great flowering plant radiations: progressive specialization and
405 diversification. *Annals of botany*, 115(4), 665–681.
406 Gao, C. M., Deng, Y. F., and Wang, J. (2019). The complete chloroplast genomes of
407 *Echinacanthus* Species (Acanthaceae): phylogenetic relationships, adaptive evolution, and
408 screening of molecular markers. *Front. Plant Sci.* 9, 1989.
409 Givnish, T. J., Spalink, D., Ames, M., Lyon, S. P., Hunter, S. J., Zuluaga, A., et al. (2015).
410 Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proc R Soc*
411 *B.* i, 20151553.
412 Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version
413 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids*
414 *Res.* 47, W59–W64.
415 Greiner, S., Wang, X., Rauwolf, U., Silber, M. V., Mayer, K., Meurer, J., et al. (2008). The
416 complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*,
417 subsection *Oenothera* : I. Sequence evaluation and plastome evolution. *Nucleic Acids Res.* 36,
418 2366–2378.
419 Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., et al. (2002). Current
420 trends in microsatellite genotyping. *Mol. Ecol. Resour.* 11, 591–611.
421 Guo, M., Zhai, J.W., Wu, X.Y., Li, S.X., Wang, M., Li, Z.J., et al. (2017). *Calanthe*
422 *taibaishanensis* , a new orchid species from china: evidence from morphological and
423 molecular analyses. *Phytotaxa*. 327, 184–190.
424 Huang, J. L., Sun, G. L., Zhang, D. M. (2010). Molecular evolution and phylogeny of the
425 angiosperm ycf2 gene. *J. Syst. Evol.* 48, 240–248.
426 Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., Depamphilis, C. W., Leebens-Mack, J., et al.
427 (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms
428 and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci.* 104, 19369–19374.
429 Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., de Pamphilis, C. W., Yi, T. S., et al. (2019).
430 GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes.
431 *bioRxiv*. 256479
432 Kantety, R. V., Rota, M. L., Matthews, D. E., and Sorrells, M. E. (2002). Data mining for simple
433 sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat.
434 *Plant Mol. Biol.* 48, 501–510.
435 Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7:
436 improvements in performance and usability. *Mol. Biol. Evol.* 30(4), 772-780.
437 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012).
438 Geneious Basic: An integrated and extendable desktop software platform for the organization
439 and analysis of sequence data. *Bioinformatics*. 28, 1647–1649.
440 Ker Gawler, J. B. (1821). *Botanical Register, Consisting of Coloured Figures of Exotic Plants*
441 *Cultivated in British Gardens, with their History and Mode of Treatment*, In, sub t, London,
442 Vol. 7, p. 573.

443 Kim, H. T., Kim, J. S., Moore, M. J., Neubig, K. M., Williams, N. H., Whitten, W. M., et al.
444 (2015). Seven new complete plastome sequences reveal rampant independent loss of the *ndh*
445 gene family across orchids and associated instability of the inverted repeat/small single-copy
446 region boundaries. *PLOS ONE*. 10, e0142215.
447 Kode, V., Mudd, E. A., Iamtham, S., and Day, A. (2005). The tobacco plastid *accD* gene is
448 essential and is required for leaf development. *Plant J.* 44, 237–244.
449 Kurtz, S. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale.
450 *Nucleic Acids Res*. 29, 4633–4642.
451 Li, M. H., Zhang, G. Q., Lan, S. R., Liu, Z. J., & China Phylogeny Consortium. (2016). A
452 molecular phylogeny of Chinese orchids. Journal of Systematics and Evolution, 54(4), 349–
453 362.
454 Lin, C. S., Chen, J. J. W., Chiu, C. C., Hsiao, H. C. W., Yang, C. J., Jin, X. H., et al. (2017).
455 Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in
456 some orchids. *Plant J.* 90, 994–1006.
457 Lin, C. S., Chen, J. J. W., Chiu, C. C., Hsiao, H. C. W., Yang, C. J., Jin, X. H., et al. (2015). The
458 location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. *Sci.*
459 *Rep.* 5, 9040.
460 Lindley, J. (1855). *Folia orchidacea, an enumeration of the known species of orchids,* Vol. 1.
461 London.
462 Liu, L. X., Wang, Y. W., He, P. Z., Li, P., Lee, J. K., Soltis, D. E., et al. (2018). Chloroplast
463 genome analyses and genomic resource development for epilithic sister genera *Oresitrophe*
464 and *Mukdenia* (Saxifragaceae), using genome skimming data. *BMC Genomics*. 19, 235.
465 Luo, J., Hou, B.W., Niu, Z. T., Liu, W., Xue, Q. Y., and Ding, X. Y. (2014). Comparative
466 chloroplast genomes of photosynthetic orchids: Insights into evolution of the Orchidaceae and
467 development of molecular markers for phylogenetic applications. *PLOS ONE.* 9, e99016.
468 Neubig, K. M., Whitten, W. M., Carlsward, B. S., Blanco, M. A., Endara, L., Williams, N. H., et
469 al. (2009). Phylogenetic utility of *ycf1* in orchids: a plastid gene more variable than *matK*.
470 *Plant Syst. Evol*. 277, 75–84.
471 Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A fast and
472 effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. *Mol. Biol.*
473 *Evol.* 32, 268–274.
474 Niu, Z. T., Xue, Q. Y., Zhu, S. Y., Sun, J., Liu, W., and Ding, X. Y. (2017). The complete
475 plastome sequences of four orchid species: Insights into the evolution of the Orchidaceae and
476 the utility of plastomic mutational hotspots. *Front. Plant Sci.* 8, 715
477 Puslednik, L., & Serb, J. M. (2008). Molecular phylogenetics of the *Pectinidae* (Mollusca:
478 Bivalvia) and effect of increased taxon sampling and outgroup selection on tree topology.
479 *Molecular Phylogenetics and Evolution*, 48(3), 1178–1188.
480 Pridgeon, A. M., Cribb, P. J., Chase, M. W., and Rasmussen, F. N. (2005). *Genera*
481 *Orchidacearum Volume 4: Epidendroideae*, Oxford University Press.

Roma, L., Cozzolino, S., Schlüter, P. M., Scopece, G., and Cafasso, D. (2018). The complete plastid genomes of *Ophrys iricolor* and *O. sphegodes* (Orchidaceae) and comparative analyses with other orchids. *PLOS ONE*. 13, e0204174.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., et al. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* 5, 2043–2049.

Song, Y., Yu, W. B., Tan, Y., Liu, B., Yao, X., Jin, J., et al. (2017). Evolutionary comparisons of the chloroplast genome in Lauraceae and insights into loss events in the magnoliids. *Genome Biol. Evol.* 9, 2354–2364.

Sosa, V., Cameron, K. M., Angulo, D. F., and Hernández-Hernández, T. (2016). Life form evolution in epidendroid orchids: Ecological consequences of the shift from epiphytism to terrestrial habit in Hexalectris. *TAXON*. 65, 235–248.

Stone, J., and Cribb, P. (2017). *LADY TANKERVILLE ' S LEGACY*, *A History and Monographic Review of Phaius and Gastrorchis*, Royal Botanic Gardens KEW.

Thiel, T., Michalek, W., Varshney, R., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet*. 106, 411–422.

Unruh, S. A. McKain, M. R., Lee, Y. I., Yukawa, T. McMcormick, M. K., Shefferson, R. P., et al. (2018). Phylotranscriptomic analysis and genome evolution of the Cypripedioideae (Orchidaceae). *Am. J. Bot.* 105, 631–640.

Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics*. 31, 3350–3352.

Xiang, X. G., Li, D. Z., Jin, W. T., Zhou, H. L., Li, J. W., & Jin, X. H. (2012). Phylogenetic placement of the enigmatic orchid genera *Thaia* and *Tangtsinia*: evidence from molecular and morphological characters. *Taxon*, 61(1), 45–54.

Xiang, X. G., Jin, W. T., Li, D. Z., Schuiteman, A., Huang, W. C., Li, J. W., et al. (2014). Phylogenetics of tribe Collabieae (Orchidaceae, Epidendroideae) based on four chloroplast genes with morphological appraisal. *PLoS ONE*. 9, e87625.

Xie, W. Q., Jäger, K., and Potts, M. (1989). Cyanobacterial RNA polymerase genes *rpoC1* and *rpoC2* correspond to rpoC of Escherichia coli. J. *Bacteriol*. 171, 1967–1973.

Yamori, W., and Shikanai, T. (2016). Physiological functions of cyclic electron transport around photosystem I in sustaining photosynthesis and plant growth. *Annu*. *Rev*. *Plant Biol*. 67, 81–106.

Yang, J. B., Li, D. Z., and Li, H. T. (2014). Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol*. *Ecol*. *Resour*. 14, 1024–1031.

Yang, J. B., Tang, M., Li, H. T., Zhang, Z. R., and Li, D. Z. (2013). Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evol*. *Biol*. 13, 84.

521 Yang, Z. H. (2007). PAML 4: Phylogenetic analysis by Maximum Likelihood. *Mol*. *Biol*. *Evol*.
522 24, 1586–1591.
523 Yang, Z. H., and Nielsen, R. (2002). Codon-substitution models for detecting molecular
524 adaptation at individual sites along specific lineages. *Mol*. *Biol*. *Evol*. 19, 908–917.
525 Yuan, Y., Jin, X. H., Liu, J., Zhao, X., Zhou, J. H. Wang, X.. et al. (2018). The *Gastrodia elata*
526 genome provides insights into plant adaptation to heterotrophy. *Nat*. *Commun*. 9, 161.
527 Yukawa, T., and Ishida G. (2008). Phylogeny and *systematics* of *Calanthe* and allied genera.
528 Abstracts retrieved from 7th Annual Meeting of Japanese Society for Plant Systematics. abstr.
529 32.
530 Yukawa, T., and Cribb, P. (2014). Nomenclatural changes in the genus *Calanthe* (Orchidaceae).
531 *Bul. of the Nat. Muse. of Nat. & Sci*. 40, 145–151.
532 Zhai, J. W., Zhang, G. Q., Li, L., Wang, M., Chen, L. J., Chung, S. W., et al. (2014). A new
533 phylogenetic analysis sheds new light on the relationships in the *Calanthe* alliance
534 (Orchidaceae) in China. *Mol*. *Phylogenet*. *Evol*. 77, 216.
535 Zhang, D., Gao, F. L., Jakovli□, I., Zhou, H., Zhang, J., Li, W.X., et al. (2019). PhyloSuite: an
536 integrated and scalable desktop platform for streamlined molecular sequence data management
537 and evolutionary phylogenetics studies. *Mol*. Ec*o*l. *Resour*. 00, 1–8.

Figure 1

Plastid genome map of *Calanthe* s.l..

Genes inside the circle are transcribed clockwise, and those outside are transcribed counterclockwise. Genes of different functions are color-coded. The darker gray in the inner circle shows the GC content, while the lighter gray shows the AT content.

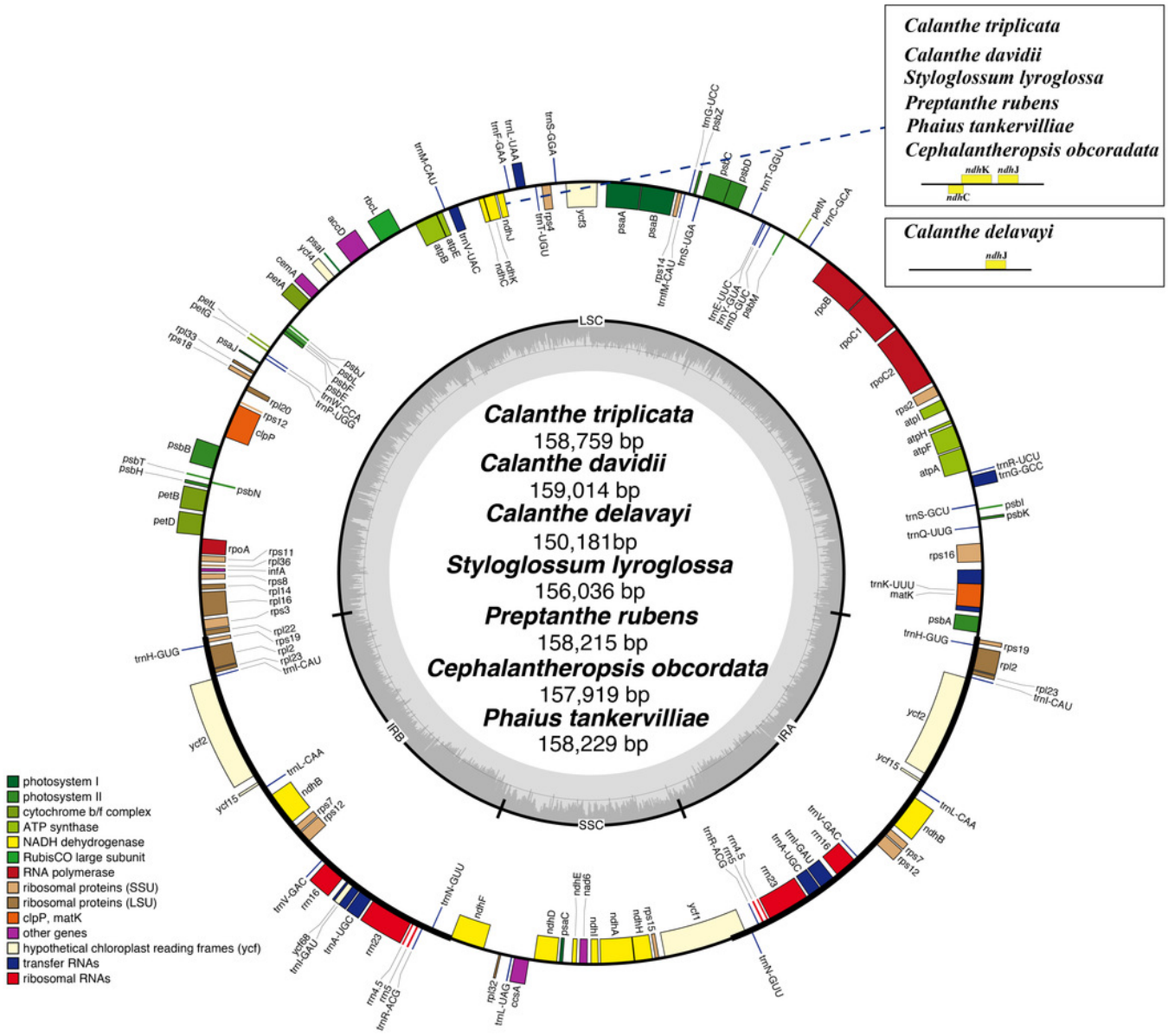

Figure 2

Mauve (Multiple Alignment of Conserved Genomic Sequence With Rearrangements) alignment of plastid genomes of seven species of *Calanthe* s.l. The *Calanthe triplicata* genome is shown at the top as the reference genome.

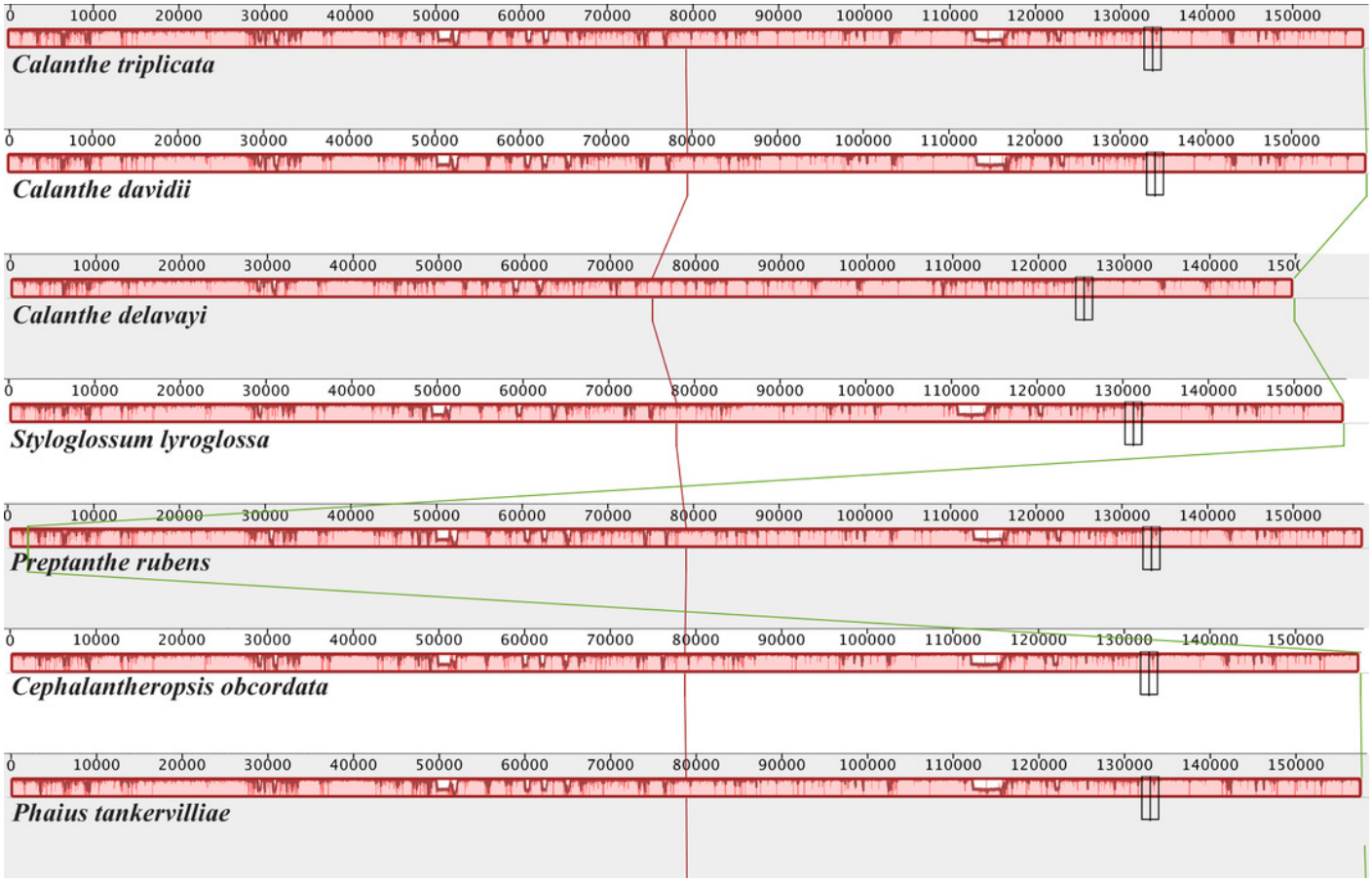

Figure 3

Comparison of seven *Calanthe* s.l. chloroplast genomes using mVISTA program, taking the annotation of *Calanthe triplicata* as a reference.

The top line shows the genes in order. A cut-off of 70% identity was used for the plots and the Y-scale represents the percent identity between 50 and 100%. Genome regions are color-coded as exon and conserved non-coding sequences (CNS).

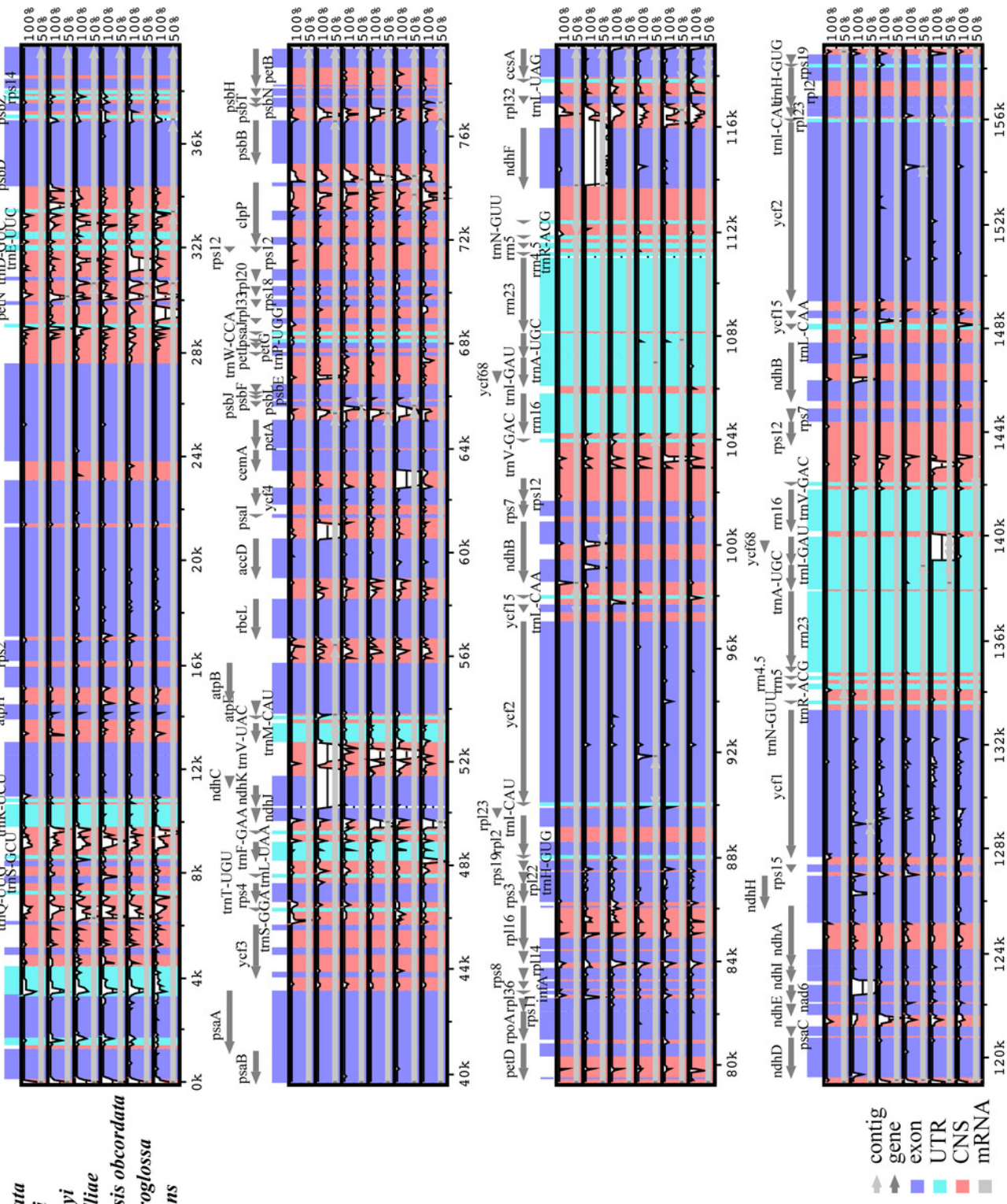
Calanthe triplicata
Calanthe davidii
Calanthe delavayi
Phaius tankervilliae
Cephalantheropsis obcordata
Stylogloss um lyroglossa
Preptanthe rubens
contig
gene
exon
UTR
CNS
mRNA

Figure 4

Information of gene and the intergenic spacer of the seven Calanthe s.l. with complete chloroplast genomes. (A) The intergenic spacers (IGS); and (B) protein coding sequences (CDS).

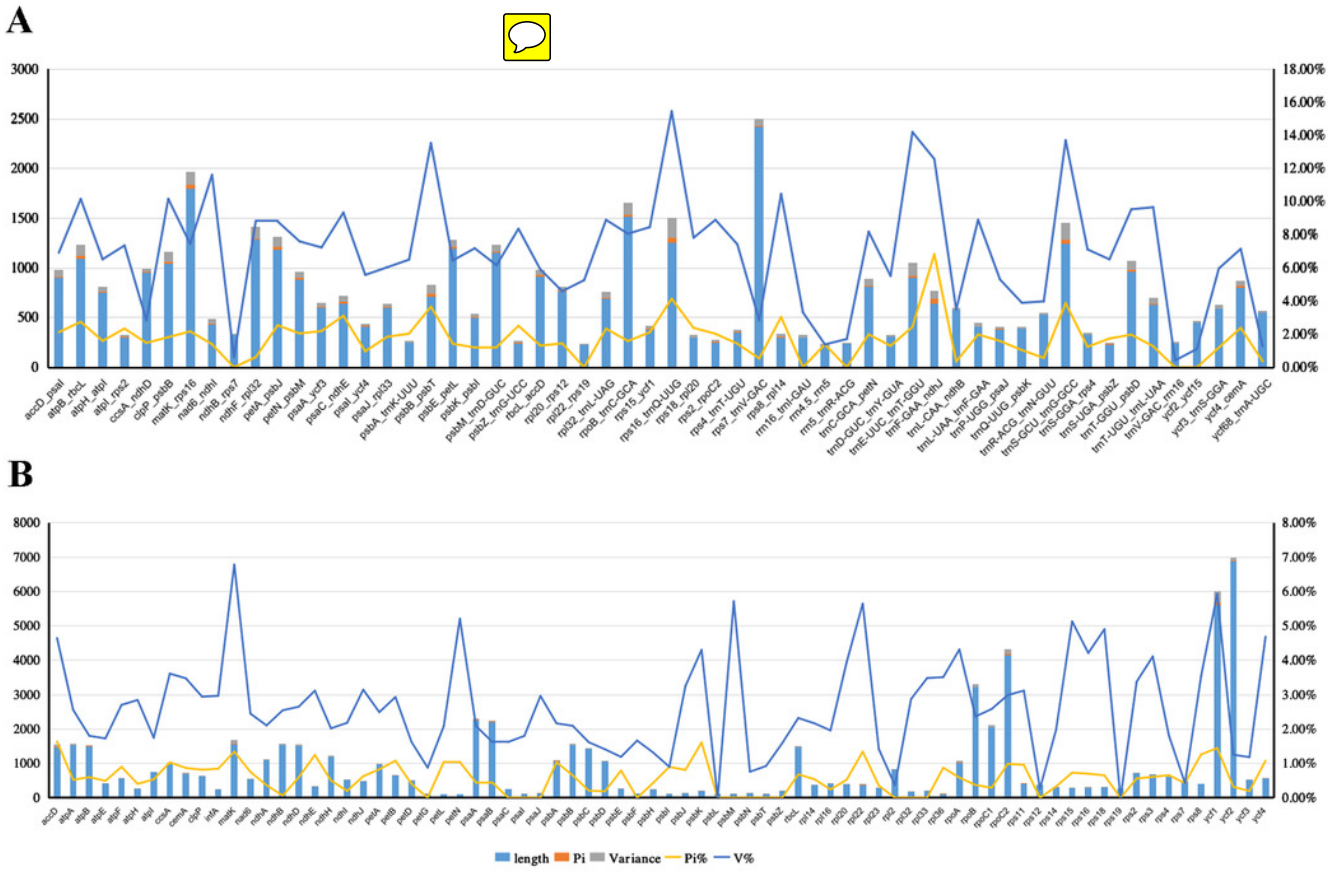

Figure 5

Comparison of the borders of LSC, SSC, and IR regions in seven Calanthe s.l. complete chloroplast genomes. JLB (IRb /LSC), JSB (IRb/SSC), JSA (SSC/IRa) and JLA (IRa/LSC) denote the JSs between each corresponding region in the genome.

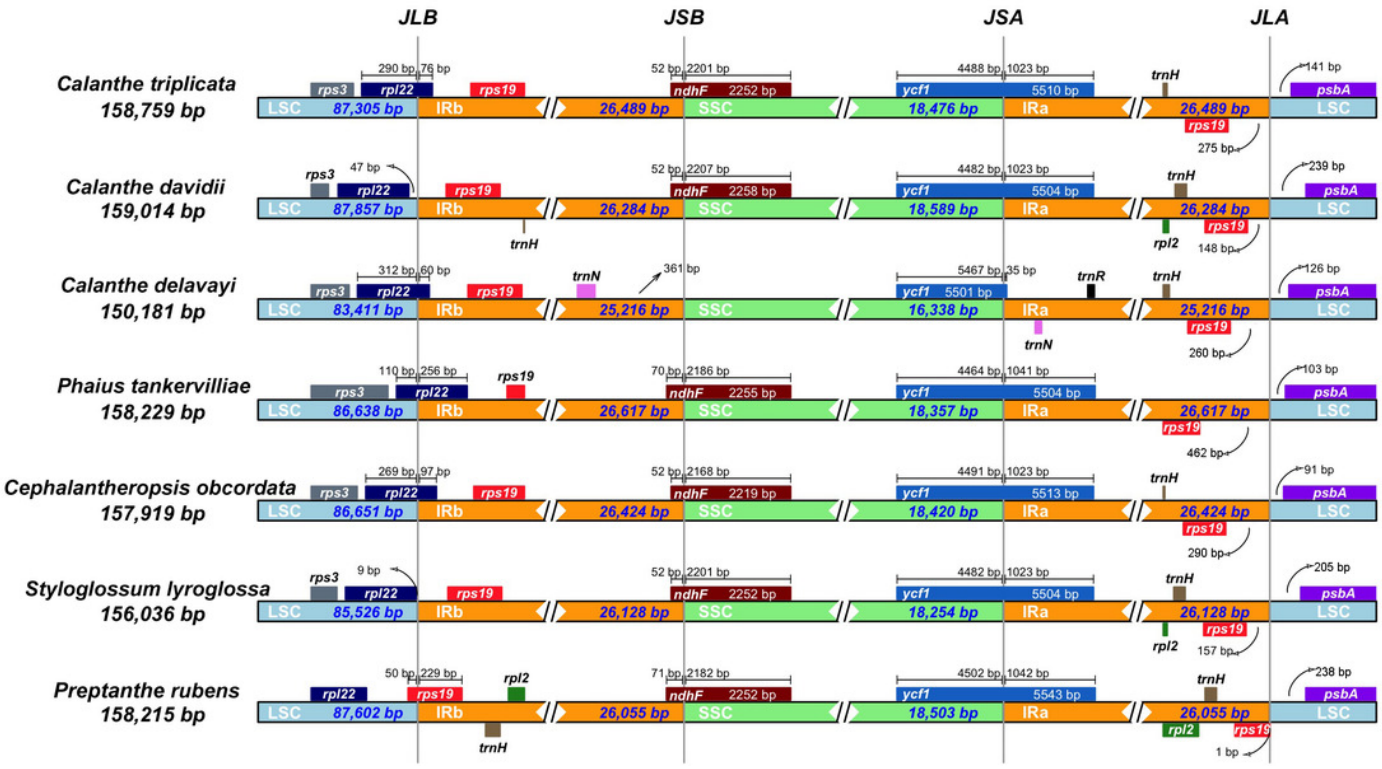

Figure 6

Maps of repeat sequence analyses, repeat sequence in seven *Calanthe* s.l. species chloroplast genome.

(A) Classification of SSRs by repeat type. mono-, mononucleotides; di-, dinucleotides; tri-, trinucleotides; tetra-,tetranucleotides; penta-, pentanucleotides; and hexa-, hexanucleotides. compound-, compound formation. (B) Classification of SSRs in seven species. IGS, intergenic spacer; CDS, coding sequence, CDS-IGS, part in CDS and part in IGS. (C) Number of the four repeat types, F, P, R, and C, indicate the long repeat type (F: forward, P: palindrome, R: reverse, and C: complement, respectively). (D) SSRs locus distribution among three different regions.

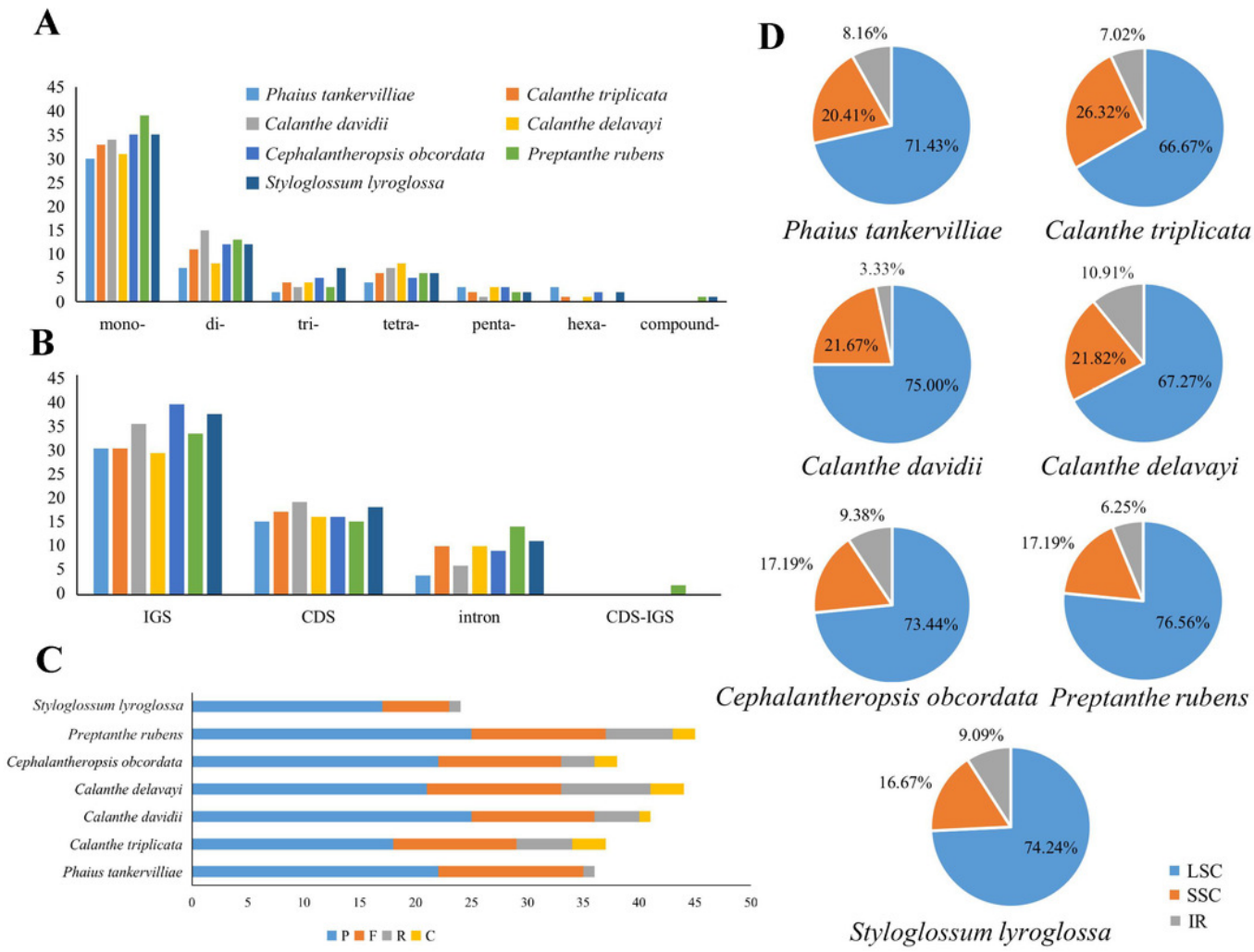
A
Phaius tankervilliae
Calanthe triplicata
Calanthe davidii
Calanthe delavayi
Cephalantheropsis obcordata
Preptanthe rubens
Styloglossum lyroglossa
mono-
di-
tri-
tetra-
penta-
hexa-
compound-
B
IGS
CDS
intron
CDS-IGS
C
Styloglossum lyroglossa
Preptanthe rubens
Cephalantheropsis obcordata
Calanthe delavayi
Calanthe davidii
Calanthe triplicata
Phaius tankervilliae
P F R C
D
8.16%
20.41%
71.43%
Phaius tankervilliae
7.02%
26.32%
66.67%
Calanthe triplicata
3.33%
21.67%
75.00%
Calanthe davidii
10.91%
21.82%
67.27%
Calanthe delavayi
9.38%
17.19%
73.44%
Cephalantheropsis obcordata
6.25%
17.19%
76.56%
Preptanthe rubens
9.09%
16.67%
74.24%
Styloglossum lyroglossa
LSC
SSC
IR

Figure 7

Maximum-likelihood (ML) tree based on 68 plastid protein-coding genes of 60 Epidendroideae species, with two Orchidoideae subfamily species (*Goodyera rosulacea* and *Chamaegastrodia shikokiana*) as the outgroup.

A bootstrap support value with 1000 replicates labelled on each node.

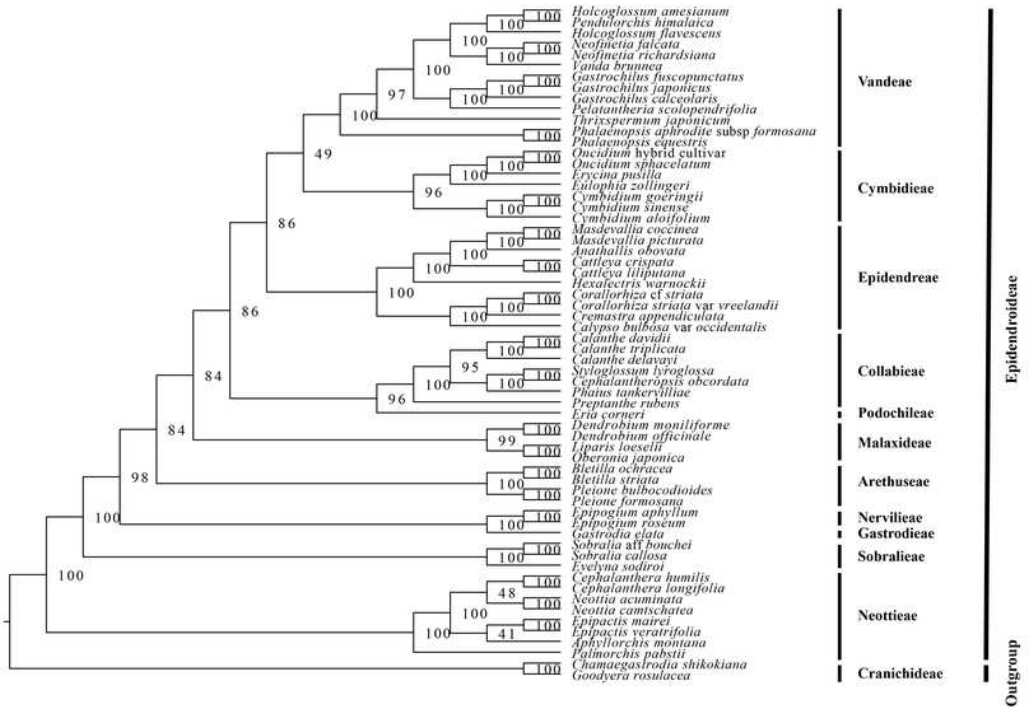

Table 1(on next page)

The basic characteristics of the chloroplast genomes of eight *Calanthe* s.l. species.

1 Table 1 The basic characteristics of the chloroplast genomes of eight *Calanthe* s.l. species.

Species	*Calanthe triplicata*	*Calanthe davidii (D1)*	*Calanthe davidii (D2)*	*Calanthe delavayi*	*Phaius tankervilliae*	*Cephalantheropsis obcordata*	*Styloglossum lyroglossa*	*Preptanthe rubens*
Accession number	KF753635	MN708353	MG925365	MK388860	MN708349	MN708351	MN708350	MN708352
Genome size(bp)	158,759	159,014	153,629	150,181	158,229	157,919	156,036	158,215
LSC length(bp)	87,305	87,857	86,045	83,411	86,638	86,650	85,421	87,498
SSC length(bp)	18,476	18,589	15,672	16,338	18,357	18,420	18,149	18,397
IR length(bp)	26,489	26,284	25,956	25,216	26,617	26,424	26,233	26,160
Coding(bp)	79,578	79,572	72,495	73,731	79,671	79,609	79,109	79,208
Non-Coding(bp)	79,181	79,442	81,134	76,450	78,558	78,310	76,927	79,007
Number of genes	136(115)	136(115)	136(115)	133(112)	136(115)	136(115)	136(115)	136(115)
Number of protein-coding genes	88(80)	88(80)	88(80)	85(77)	88(80)	88(80)	88(80)	88(80)
Number of tRNA genes	38(30)	38(30)	38(30)	38(30)	38(30)	38(30)	38(30)	38(30)
Number of rRNA genes	8(4)	8(4)	8(4)	8(4)	8(4)	8(4)	8(4)	8(4)
GC content (%)	36.70	36.60	36.90	36.90	37.00	36.80	36.90	36.70
GC content in LSC (%)	34.40	34.40	34.50	34.50	34.80	34.50	34.60	34.30
GC content in SSC (%)	29.70	29.60	30.20	29.40	29.90	29.70	29.90	29.70
GC content in IR (%)	43.00	43.10	43.10	43.30	43.00	43.10	43.10	43.30

2

Table 2(on next page)

Genes length (bp) difference between two versions of Calanthe davidii complete chloroplast genome.

1 Table 2 Genes length (bp) difference between two versions of *Calanthe davidii* complete chloroplast genome.

Version	nad6	ndhA	ndhC	ndhD	ndhE	ndhF	ndhI	ndhK	ycf2
D1 (MG925365)	120	869	171	987	306	1,827	297	138	4,650
D2 (MN708353)	531	2,215	363	1,509	321	2,259	504	774	6,813

2

Table 3(on next page)

Gene contents in seven *Calanthe* s.l. species chloroplast genome

Table 3 Gene contents in seven Calanthe s.l. species chloroplast genome

Classfication	Genes
Genetic apparatus	
Large ribosomal subunits	*rpl2*∗(×2), *rpl14*, *rpl16*∗, *rpl20*, *rpl22*, *rpl23*(×2), *rpl32*, *rpl33*, *rpl36*
Small ribosomal subunits	*rps2*, *rps3*, *rps4*, *rps7* (×2), *rps8*, *rps11*, *rps12*∗∗, *rps14*, *rps15*, *rps16*∗, *rps18*, *rps19*(×2)
RNA polymerase subunits	*rpoA*, *rpoB*, *rpoC1*∗, *rpoC2*
DNA dependent RNA polymerase Protease	*clpP*∗∗
Maturase	*matK*
Ribosomal RNAs	*rrn4.5*(×2), *rrn5*(×2), *rrn23*(×2), *rrn16*(×2)
Transfer RNAs	*trnA*-UGC(×2)∗, *trnC*-GCA, *trnD*-GUC, *trnE*-UUC, *trnF*-GAA, *trnfM*-CAU, *trnG*-GCC, *trnG*-UCC∗, *trnH*-GUG(×2), *trnI*-CAU(×2), *trnH*-GUG, *trnI*-GAU(×2)∗, *trnK*-UUU∗, *trnL*-CAA(×2), *trnL*-UAG, *trnL*-UUA∗, *trnM*-CAU, *trnN*-GUU(×2), *trnP*-UGG, *trnQ*-UUG, *trnR*-ACG(×2), *trnR*-UCU, *trnS*-GCU, *trnS*-GGA, *trnS*-UGA, *trnT*-GGU, *trnV*-GAC(×2), *trnV*-UAC∗, *trnW*-CCA, *trnY*-GUA
Light dependent photosynthesis	
Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*, *ycf3*∗∗, *ycf4*
Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
NAD(P)H dehydrogenase complex	*ndhA*∗, *ndhB*∗(×2), *ndhC*†, *ndhD*, *ndhE*, *ndhF*†, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*†
F-type ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*∗, *atpH*, *atpI*
Cytochrome b6/f complex	*petA*, *petB*∗, *petD*∗, *petG*, *petL*, *petN*
Light independent photosynthesis	
Inner membrane protein	*cemA*
Cytochrome C biogenesis protein	*ccsA*
Large subunit of Rubisco	*rbcL*

Subunit of acetyl-CoA-carboxylase	*accD*
Translation initiation factor	*infA*
Function uncertain	*ycf1*, *ycf2*(×2), *ycf15*(×2), *ycf68*(×2)

∗ Gene containing one intron, ∗∗ gene containing two introns, a trans-splinting gene, (×2) shows genes have two copies, †gene lost in *C*. *delavayi*.

Table 4(on next page)

The polymorphic simple sequence repeats in *Calanthe* s.l.

1 Table 4 The polymorphic simple sequence repeats in Calanthe s.l..

Type	*C. triplicata*/*C. davidii*/*C. delavayi*/*S. lyroglossa*/*P. rubens*/C. *obcordata*/*P. tankervilliae*	Location	Region
AT	6/6/0/6/0/0/0	*rpoB*-*trnC*-GCA	LSC
AT	7/7/0/0/5/0/0	*trnE*-UUC-*trnT*-GGU	LSC
AT	0/0/7/6/5/7/0	*trnL*-UAA-*trnF*-GAA	LSC
AT	0/0/7/6/0/7/0	*trnL*-UAA-*trnF*-GAA	LSC
GA	5/5/5/0/5/5/5	*ycf2*	IR
TA	5/7/0/5/0/6/0	*psbB*-*psbT*	LSC
TA	5/5/0/0/6/0/0	*ndhF*-*rpl32*	SSC
TA	0/0/8/5/5/0/5	*clpP*-*psbB*	LSC
TC	5/5/5/0/5/5/5	*ycf2*	IR
TG	5/5/5/0/5/5/5	*rpl33*-*rps18*	LSC
AAT	4/4/4/4/0/4/0	*psaC*-*ndhE*	SSC
AGAA	0/0/3/0/3/3/0	*psbM*-*trnD*-GUC	LSC
AATG	0/3/3/3/0/3/0	*cemA*	LSC
ATTA	3/3/0/3/3/0/0	*psaJ*-*rpl33*	LSC
GTCT	3/3/3/3/0/3/3	*atpA*	LSC
TTGA	3/3/3/3/0/3/3	*ndhE*	SSC
ATCTT	3/3/0/0/3/3/3	*psbK*-*psbI*	LSC
ACAAA	3/0/0/0/0/3/3	*ndhC*-*trnV*-UAC	LSC

2