

RESEARCH ARTICLE

Comparative analysis of plastid genomes within the Campanulaceae and phylogenetic implications

Chun-Jiao Li¹, Ruo-Nan Wang², De-Zhu Li¹ *

1 Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China, **2** College of Life Sciences, Northwest University, Xi'an, China

* dzl@mail.kib.ac.cn

OPEN ACCESS

Citation: Li C-J, Wang R-N, Li D-Z (2020) Comparative analysis of plastid genomes within the Campanulaceae and phylogenetic implications. PLoS ONE 15(5): e0233167. <https://doi.org/10.1371/journal.pone.0233167>

Editor: Tzen-Yuh Chiang, National Cheng Kung University, TAIWAN

Received: March 9, 2020

Accepted: April 29, 2020

Published: May 14, 2020

Copyright: © 2020 Li et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: De-Zhu Li received the funding. This research was funded by Chinese Academy of Sciences, grant number XDB31000000. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

The conflicts exist between the phylogeny of Campanulaceae based on nuclear ITS sequence and plastid markers, particularly in the subdivision of Cyanantheae (Campanulaceae). Besides, various and complicated plastid genome structures can be found in species of the Campanulaceae. However, limited availability of genomic information largely hinders the studies of molecular evolution and phylogeny of Campanulaceae. We reported the complete plastid genomes of three Cyanantheae species, compared them to eight published Campanulaceae plastomes, and shed light on a deeper understanding of the applicability of plastomes. We found that there were obvious differences among gene order, GC content, gene compositions and IR junctions of LSC/IRa. Almost all protein-coding genes and amino acid sequences showed obvious codon preferences. We identified 14 genes with highly positively selected sites and branch-site model displayed 96 sites under potentially positive selection on the three lineages of phylogenetic tree. Phylogenetic analyses showed that *Cyananthus* was more closely related to *Codonopsis* compared with *Cyclocodon* and also clearly illustrated the relationship among the Cyanantheae species. We also found six coding regions having high nucleotide divergence value. Hotpot regions were considered to be useful molecular markers for resolving phylogenetic relationships and species authentication of Campanulaceae.

Introduction

The three closely related families, Campanulaceae, Cyphiaceae, and Lobeliaceae are sometimes treated as subfamilies of the broadly delimited Campanulaceae which consists of more than 2300 species with nearly cosmopolitan distribution [1]. Campanulaceae *sensu stricto* (*s.str.*) primarily distributes in the temperate regions and is centered in East Asia, incorporating three groups of the Platycodonoids, Wahlenbergioids, and Campanuloids based on the capsule dehiscent mode and location of carpel and calyx lobes [2]. Later, Hong and Wang combining the data from palynology, external morphology and DNA fragments, established a classification with three tribes for Campanulaceae *s.str.*, i.e., Cyanantheae, Wahlenbergieae and Campanuleae [3, 4].

Many Cyanantheae species are important traditional medicines, such as *Platycodon grandiflorus* and *Codonopsis pilosula* showing anti-epileptic, anti-oxidative, anti-viral, and anti-inflammatory properties and some species e.g., *Cyananthus incanus* and *Cyananthus formosus* with ornamental values [5–8]. However, less attention has been paid to this group; there are a few taxonomic and phylogenetic studies apart from the research of medicinal value [9, 10]. The Cyanantheae is distinct from other two tribes by colpate or colporate pollen with elongate apertures and a loculicidal capsule or a berry. The subdivision of this group is still controversial since *Codonopsis*, the largest genus among the Cyanantheae is polyphyletic [4, 11]. The controversies mainly exist in the relationship of *Codonopsis* and its allies. *Codonopsis* is mainly distributed in the Himalayas and southwest China. Studying this genus will be helpful to clarify the phylogenetic relationships of Cyanantheae. In the past years, the nuclear ribosomal ITS and several plastid genome regions (such as *atpB*, *matK*, *rbcL*, *petD*) or their combinations had been frequently used in the study of molecular systematics of Cyanantheae [9, 11]. The selected loci failed to provide sufficient systematic information among Cyanantheae species. Some important branches still show the low supported value and are undefined [4, 9, 11, 12]. As a result, it is necessary to seek other methods for rebuilding the classification of Cyanantheae. Whole plastid genome or hyper-variable regions are urgently needed. The broadly definition of this clade comprises *Platycodon*, *Canarina*, *Cyclocodon*, *Echinocodon*, *Codonopsis* and *Cyananthus* et al. [4]. Except the *Canarina*, other genera are only found in East Asia. Obviously, the species of East Asia play a vital role in analyzing the genome evolution and demonstrating the phylogenetic relationship of Cyanantheae. *Cyclocodon* and *Cyananthus* are noteworthy in the flora of the Himalayas and adjacent areas. Alpine species of *Cyananthus* endemic to the Himalaya-Hengduan Mountains, have been used to study the distributional responses to climate change [13]. For the species of *Cyclocodon*, calyx lobes are stripe or strip-lanceolate and have dentate margin or rarely entire. *Cyananthus* is a distinctive member of Campanulaceae due to the superior calyx and corolla, which illustrate that this genus appears earlier [14]. Plastid genomes of these floras remain not to be elucidated. What's more, the plastid genome evolution in Cyanantheae is still blank.

In recent years, based on genomic resources, such as complete plastid sequences, there is a good chance to study the genomic evolution and interspecific relationships of organisms [15–18]. Chloroplasts are small organelles inside the cells of plants with the function of providing photosynthetic machinery and producing essential energy. The majority of the plastid genomes of land plants have highly-conserved compositions, with respect to the gene content and gene order [19–22]. Nevertheless, many rearrangements are the rare evolutionary events and often have certain phylogenetic significance [23]. Various plastid genome structures can be found in the Campanulaceae species because of numerous rearrangements [9, 24–26]. However, the research on plastome structures of Campanulaceae has been relatively scarce [24, 27]. Besides, the conflicts still exist between the phylogeny of Campanulaceae based on ITS and based on plastid markers [4, 11]. Until now, there are few studies of constructing Campanulaceae phylogeny based on the plastomes. Therefore, using the plastid genome structures will be helpful to identify the uncertainty phylogenetic relationships and clarify the structural evolution. Plastid markers and genetic information of more complete plastid genomes of Campanulaceae will also further contribute to the conservation strategy and utilization of this family.

Here, we report newly sequenced complete plastid genomes of *Cyananthus flavus*, *Cyclocodon parviflorus*, and *Codonopsis hongii* using next-generation sequencing technology and genomic comparative analysis with other eight published plastome sequences of Campanulaceae download from the NCBI. The main objectives of this study are to (1) assemble and annotate the genome structures of three Cyanantheae species, (2) reveal structural and size variation in the plastomes of Campanulaceae, and trace the evolutionary pattern of IR

expansion/contraction, (3) identify divergence hotspots of plastome regions for further evolutionary and systematic study of Campanulaceae and determine signatures of positive selection, and (4) test the applicability of plastid phylogenomics in resolving phylogenetic relationships of Campanulaceae *s.str.*, especially within the Cyanantheae.

Materials and methods

Plant material, DNA extraction, and sequencing

There is no specific permits required for obtaining the healthy and fresh leaves of *Cyananthus flavus*, *Cyclocodon parviflorus*, and *Codonopsis hongii*, since they are not endangered or protected species and were collected from the fields that are not privately owned or protected. The plant materials of *Cyananthus flavus*, *Cyclocodon parviflorus*, and *Codonopsis hongii* were collected at Lijiang City (27°0′24.4″N, 100°10′31.1″E, alt. 3439 m), Cangyuan Wa Autonomous County (23°14′39″N, 98°56′55″E, alt. 946 m), Gongshan Derung and Nu Autonomous County (27°43′44.1″N, 98°21′34.4″E, alt. 1660 m) of Yunnan, China, respectively.

The voucher specimens of three species were deposited at Herbarium of Kunming Institute of Botany, Chinese Academy of Sciences (KUN). The voucher numbers are KUN 1379897 (*Cyananthus flavus*), KUN 1380108 (*Cyclocodon parviflorus*), and GLGS21262 (*Codonopsis hongii*). Total genomic DNA was isolated from silica-gel-dried leaves by using a CTAB protocol [28]. The quality and concentration of DNA were evaluated via agarose gel electrophoresis and spectrophotometry (NanoDrop-2000, Thermo Fisher Scientific). We used an ultrasonicator to randomly fragment the extracted genomic DNA into 400–600bp following manufacturer's manual (Illumina). DNA libraries with 500-bp insert size were constructed by the NEBNext® Ultra™ II DNA Prep Kit for illumina. Sequencing of paired-end 150 bp read lengths was run on Illumina HiSeq X TEN at Plant Germplasm and Genomics Center of Kunming Institute of Botany. The sequencing quantity of all newly sequenced species is more than 1 Gigabyte.

Plastid genome assembly and annotation

Complete plastid genome of *Codonopsis lanceolata* (KP889213) as reference, the paired-end reads were filtered and assembled into a complete plastome using GetOrganelle (<https://github.com/Kinggerm/GetOrganelle>) [29]. The final assembly graph was viewed and checked by Bandage [30] to confirm the paths of the plastomes. In addition, the four junctions between the IR (inverted repeat) regions and LSC (large single copy region)/ SSC (small single copy region) were reconfirmed by PCR and Sanger sequencing. The primers were designed based on the reference genome (*Codonopsis lanceolata* MH018574) through the Primer3 algorithm (<http://frodo.wi.mit.edu/primer3/>) with the default setting and displayed in the [S1 Appendix](#) which also showed the PCR reactions. Sanger sequencing was finished in the BioSune company after purify the them through precipitation with 95% ethanol and 3-sodium acetate. Geneious 8.0.2 [31] was used to align the sanger sequences and assembled genomes for checking any differences. The assembled plastid genome was automatically annotated using PGA [32], then manually adjusted in Geneious. Circular plastid genome maps of *Cyananthus flavus*, *Cyclocodon parviflorus*, and *Codonopsis hongii* (Figs 1, 2 and 3) were drawn using OGDRAW tool [33] with default settings and checked manually. The sequence of plastomes generated in this study was submitted to the NCBI database with the GenBank accession number (Table 1).

Genome structure analyses and genome comparison

Six plastomes of Campanulaceae *s.str.* available in GenBank (Table 1) were included as closely related groups. Among of these, three species are the Campanuleae plants. Additionally,

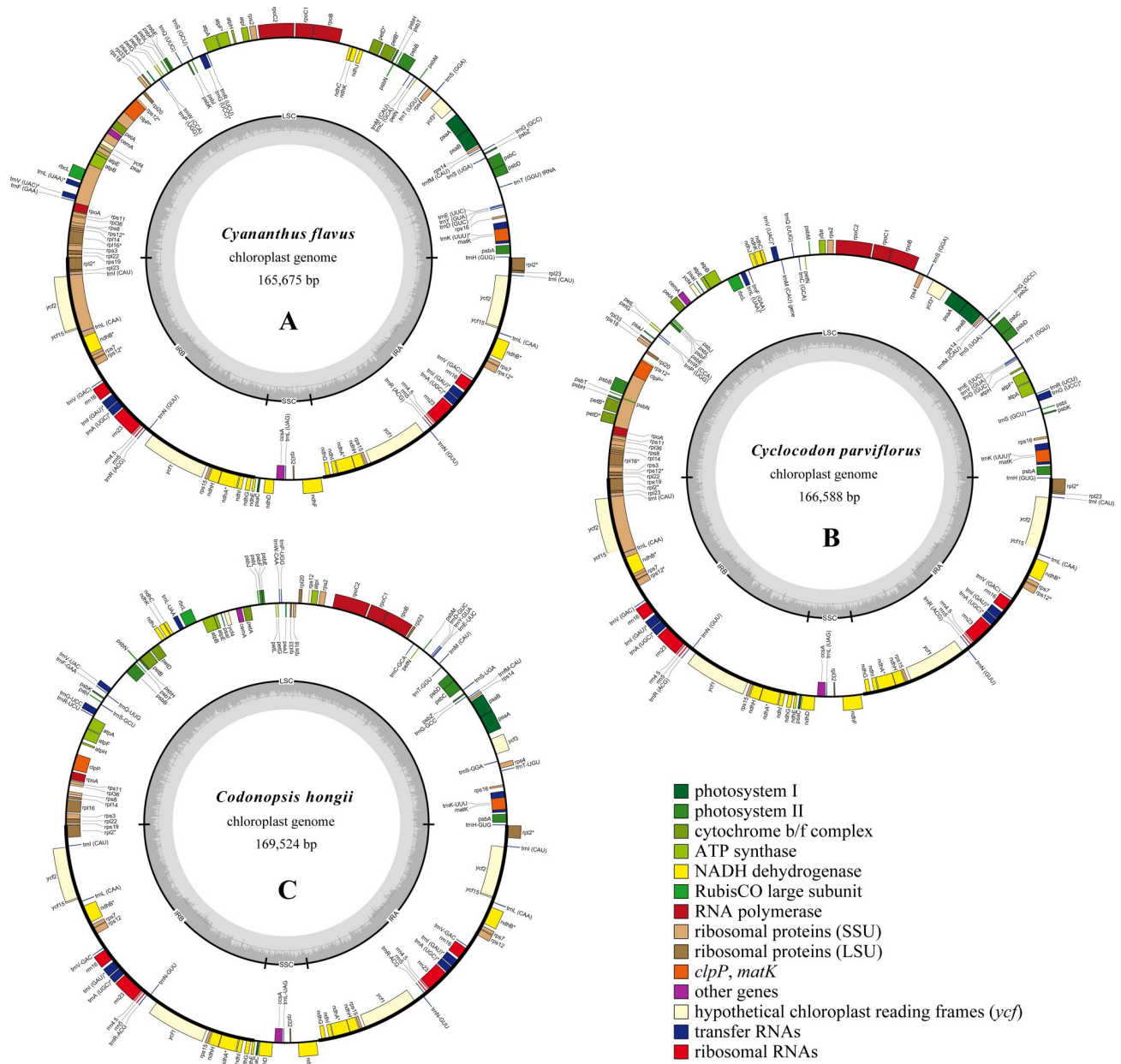


Fig 1. Plastid genomes of *Cyananthus flavus* (A), *Cyclocodon parviflorus* (B), and *Codonopsis hongii* (C). Genes inside the circle are transcribed clockwise, and genes outside the circle are transcribed counter-clockwise. The dark-gray inner circle corresponds to the GC content, and the light-gray represents the AT content.

<https://doi.org/10.1371/journal.pone.0233167.g001>

Lobelia erinus (Lobelioideae) (MF770635) and *Cyphia crenata* (Cyphioideae) (MF770625) were assigned as the out-group to reconstruct phylogenetic relationships. The whole plastid genomes of eleven species, including the three newly sequenced Cyanantheae species in this study were performed using Mauve [34]. We calculated the ORFs (opening reading frame) >300 bp in the IRa regions of each species in the Geneious. The boundaries between the IR and SSC regions, IR and LSC regions, plus the different contents of IR were compared and analyzed. In total, 76 protein coding genes of all studied species were compiled into a single file and aligned with MAFFT [35] and manually adjusted with Geneious. In addition, the *rp123*

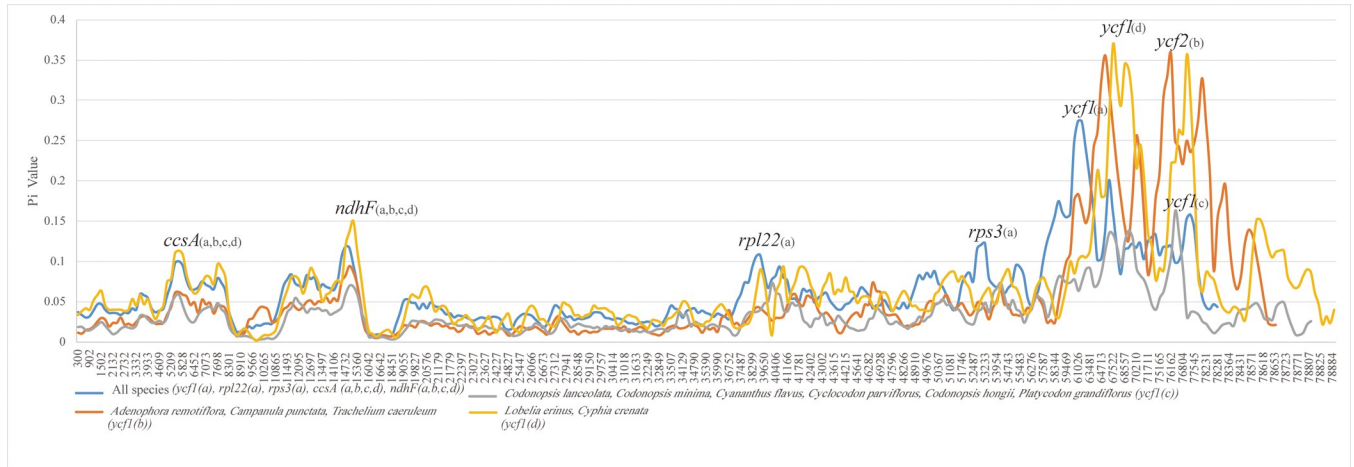


Fig 2. Percentages of variable sites in protein-coding regions. The blue line indicates the comparison of eleven species among the family Campanulaceae; the gray line indicates the comparison of six Cyanantheae species; the orange line indicates the Campanuleae species; the yellow line indicates the out-group. X axis: position of the midpoint of a window. Y axis: nucleotide diversity of each window.

<https://doi.org/10.1371/journal.pone.0233167.g002>

and *infA* genes were excluded from the data matrix, since there being too many losses there. To compare nucleotide diversity (π) in different groups, we divided the eleven samples into the groups of all species, the Cyanantheae, the Campanuleae, and the out-group. The π throughout the coding regions with 200 bp step size and 600 bp window length was determined via the DnaSP version 6 [36] software.

The distribution of codon usage with the relative synonymous codon usage (RSCU) value and the GC content were calculated using the software MEGA 6.0 [37]. RSCU represents the ratio of the observed frequency of a codon to the expected frequency and is a good indicator of codon usage bias [38, 39]. When the RSCU value is less than 1, synonymous codons are used less frequently than expected; otherwise, the value is greater than 1 [40]. The visualization of codon usage in the form of heatmaps of Campanulaceae species and a histogram were conducted with R language with an RSCU value.

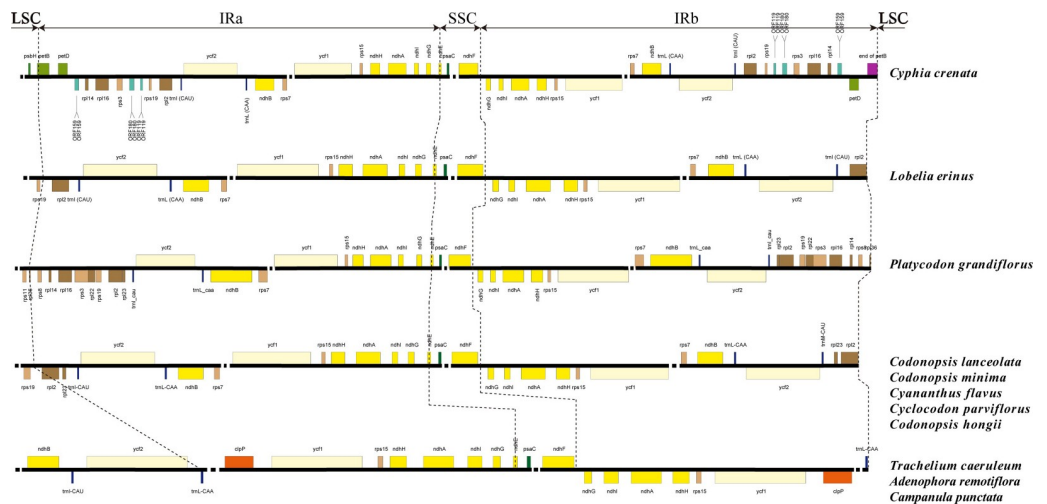


Fig 3. Comparison of the borders between IR and LSC/SSC regions and the gene composition of IR regions.

<https://doi.org/10.1371/journal.pone.0233167.g003>

Table 1. Comparison of plastome features of Campanulaceae species.

Species	<i>Cyananthus flavus</i>	<i>Cyclocodon parviflorus</i>	<i>Codonopsis hongii</i>	<i>Platycodon grandiflorus</i>	<i>Codonopsis lanceolata</i>	<i>Codonopsis minima</i>	<i>Adenophora remotiflora</i>	<i>Campanula punctata</i>	<i>Trachelium caeruleum</i>	<i>Lobelia erinus</i>	<i>Cyphia crenata</i>
Accession number	MIT074354	MIT074353	MIN849357	KX352464	MH018574	KY587457	KP889213	KU198434	EU090187	MF770635	MF770625
Genome size (bp)	165675	166588	169524	171818	169447	169321	171724	169341	162321	166019	178956
LSC length (bp)	82501	83994	85326	79112	85253	85506	105555	102323	100110	81503	79041
SSC length (bp)	8120	8014	7912	7840	8060	8067	11295	7744	7661	7792	8085
IR length (bp)	37527	37290	38143	42433	38067	37874	27437	29637	27276	38362	45915
Coding sequences (bp)	88242	88587	88296	90144	88761	87332	76245	78771	71508	89784	92511
Perscent of coding sequences(%)	53.26	53.18	52.08	52.46	52.38	51.58	44.39	46.52	44.05	54.08	51.69
Non-coding sequences (bp)	77433	78001	81228	81674	80686	81989	95479	90570	90813	76235	86445
Number of genes	136(25)	135(25)	135(24)	146(32)	138(24)	133(24)	140(21)	134(18)	147(22)	139(25)	150(33)
Number of protein-coding genes	89(13)	91(14)	90(13)	95(20)	95(19)	86(12)	82(7)	83(7)	83(7)	89(13)	99(21)
Number of tRNA genes	37(7)	36(7)	37(7)	36(7)	38(7)	37(7)	37(5)	36(5)	44(7)	36(7)	36(7)
Number of rRNA genes	8(4)	8(4)	8(4)	8(4)	8(4)	8(4)	8(4)	8(4)	8(4)	8(4)	8(4)
Total GC content (%)	38.1	37.9	38.2	38.1	38.2	38.3	38.8	38.8	38.3	39.0	36.8
GC content in LSC(%)	36.6	36.7	36.9	37.2	36.9	36.9	37.5	37.8	37.1	38.4	35.8
GC content in SSC (%)	31.1	31.3	32.0	31.0	32.4	32.4	34.9	32.6	32.2	33.5	31.0
GC content in IR (%)	40.4	40.0	40.4	36.9	40.4	40.5	42.0	41.4	41.4	40.2	38.2

<https://doi.org/10.1371/journal.pone.0233167.t001>

Repeat sequence analyses

REPuter [41] was hired to identify dispersed repeats, including forward (F), reverse (R), palindrome (P), and complement (C) repeats. The repeat sizes were limited to a minimum of 50 bp and the maximum computed repeats were detected less than 100, with a Hamming distance of 3. The IRb of each plastome was removed before the repeat detection, and then the location of repeats in IRb as manually found based on those detected in IRa. We used online Tandem repeats finder (<http://tandem.bu.edu/trf/trf.html>) to identify tandem repeats sequences with default parameters. Simple sequence repeats (SSR or microsatellites) in the eleven genomes were detected by A Perl script MISA [42]. Tandem repeats (1–6 nucleotides) were viewed as microsatellites, with the minimal repeat number set to 12, 6, 5, 5, 5 and 5 for mono-, di-, tri-, tetra-, penta-, and hexa- nucleotides, respectively. All of the repeats were manually verified. We also counted the repeat numbers in the regions of LSC, SSC and IRa.

Positive selection analysis

In order to detect the protein-coding genes under selection within the species of Campanulaceae, we used Muscle (codon) implemented in MEGA to align the each gene. We analyzed all CDS gene regions, except the *rpl23* and *infA*. A Maximum likelihood phylogenetic tree based on CDS regions was constructed using RAxML [43]. The codon substitution models were performed for calculating the non-synonymous (dN) and synonymous (dS) substitution rates, along with their ratios ($\omega = dN/dS$), which were implemented in the Codeml program, PAML3.15 [44]. We used the site-specific model of M0, M1a, M2a, M3, M7, and M8. This model allowed ω ratio to vary among sites with a fixed ω ratio in all branches. M1a (neutral) vs. M2a (positive selection), M7 (β) vs. M8 (β and ω), and M0 (one-ratio) vs. M3 (discrete), were calculated in order to detect positive selection, by comparing the site-specific model [45]. Likelihood ratio test (LRT) of the above comparison was conducted respectively to evaluate the selection strength and the p -values of Chi square (χ^2) smaller than 0.05 was thought as significant.

The branch-site model with difference ω among branches (labeled foreground-lineages) of the phylogeny and sites, were also used to test which sites were influenced by the positive selection in the foreground-branch and conducted using the CODEML algorithm [44] executed in EasyCodeML [45, 46]. We took three main lineages of Cyanantheae, Campanuleae and out-group as the foreground branch separately and calculated the positive selection occurred on the aboved branches by using 76 protein-coding genes individually. If the LRT p -values were significant (<0.05), Bayes Empirical Bayes (BEB) method [47] was implemented to calculate posterior probabilities for finding sites under positive selection on the three branches [48].

Phylogenetic analyses

A total of 76 common protein-coding genes shared in the plastomes of Campanulaceae were aligned with MAFFT [33] and were manually adjusted. *Lobelia erinus* and *Cyphia crenata* were selected as the out-group (Table 1). Maximum likelihood (ML) analysis were implemented using RAxML with a bootstrap of 1000 repetitions [42], and the best tree in a single run were found by using the GTR+G model consulted from the RAxML instruction. The jModelTest 2.0 program [49] was used to determine the best-fitting model for dataset based on the Bayesian information criterion (BIC). Regarding Bayesian inference (BI), two independent chains (burnin = 1000) was performed using the program MrBayes v3.2 (Ronquist et al. 2012) at the CIPRES Science Gateway website (<http://www.phylo.org/>) [50], with the GTR+I+G model determined by jModelTest in the unpartitioned dataset. The Markov chain Monte Carlo (MCMC) analysis was run for 2×1000,000 generations, with trees sampled every 1,000

generations. The first twenty-five per cent of trees calculated were removed as burn-in and the tree of a majority rule consensus would be generated from the remaining trees. The average standard deviation of split frequencies equal to or less than 0.01 would be considered the convergence of the MCMC chains. Figtree v1.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) was used to visualize and annotate trees.

Results and discussion

General features of the plastid genomes

In this study, we first determined the whole plastid genomes of three Cyanantheae species. The mean coverages of *Cyananthus flavus*, *Cyclocodon parviflorus* and *Codonopsis hongii* were 679x, 483x and 1000x, respectively, and the clean reads of the above species were 2,926,584 to 8,710,738. The complete plastid genomes of *Cyananthus flavus*, *Cyclocodon parviflorus*, and *Codonopsis hongii* displayed a typical quadripartite structure and were circular molecular 165,675bp-169,524bp in size (Fig 1 and Table 1). A total of seven protein-coding genes and six tRNA genes contained one intron, whereas three genes (*rps12*, *clpP*, *ycf3*) contained two introns, as shown in Table 2. *Ycf3* gene expression result in stable accumulation of photosystem I complexes [51].

The size of the *Cyphia crenata* plastid genome (178,956bp) was the longest, and that of the *Trachelium caeruleum* plastid genome (162,321bp) was the shortest. Interestingly, the LSC

Table 2. List of genes present in three newly sequenced plastomes.

Category of genes	Group of gene	Name of gene
Self-replication	Ribosomal RNA genes	<i>rrn16</i> ^(×2) , <i>rrn23</i> ^(×2) , <i>rrn4.5</i> ^(×2) , <i>rrn5</i> ^(×2)
	Transfer RNA genes	<i>trnA</i> -UGC ^{*, (×2)} , <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC, <i>trnF</i> -GAA, <i>trnG</i> -CAU, <i>trnG</i> -GCC, <i>trnG</i> -UCC [*] , <i>trnH</i> -GUG, <i>trnI</i> -CAU, <i>trnI</i> -GAU ^{*, (×2)} , <i>trnK</i> -UUU [*] , <i>trnL</i> -CAA, <i>trnL</i> -UAA [*] , <i>trnM</i> -CAU, <i>trnN</i> -GUU, <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnR</i> -ACG, <i>trnR</i> -UCU, <i>trnS</i> -GCU, <i>trnS</i> -GGA, <i>trnS</i> -UGA, <i>trnT</i> -GGU, <i>trnT</i> -UGU ^(a,b) , <i>trnV</i> -GAC, <i>trnV</i> -UAC [*] , <i>trnW</i> -CCA, <i>trnY</i> -GUA
	Small subunit of ribosome	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> ^(×2) , <i>rps8</i> , <i>rps11</i> , <i>rps12</i> ^{*, (×2)} , <i>rps14</i> , <i>rps15</i> ^(×2) , <i>rps16</i> , <i>rps18</i> , <i>rps19</i>
	Large subunit of ribosome	<i>rpl2</i> ^{*, (×2)} , <i>rpl14</i> , <i>rpl16</i> [*] , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> ^{(a,c×2)(d×1)} , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
	DNA-dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>rpoC2</i>
Genes for photosynthesis	Subunits of NADH-dehydrogenase	<i>ndhA</i> ^{*, (×2)} , <i>ndhB</i> ^{*, (×2)} , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> ^(×2) , <i>ndhH</i> ^(×2) , <i>ndhI</i> ^(×2) , <i>ndhJ</i> , <i>ndhK</i>
	Subunits of photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i> , <i>ycf3</i> ^{**} , <i>ycf4</i>
	Subunits of photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>
	Subunits of cytochrome b/f complex	<i>petA</i> , <i>petB</i> [*] , <i>petD</i> [*] , <i>petG</i> , <i>petL</i> , <i>petN</i>
	Subunits of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> [*] , <i>atpH</i> , <i>atpI</i>
Other genes	Subunits of rubisco	<i>rbcL</i>
	Maturase	<i>matK</i>
	Protease	<i>clpP</i> ^{**}
	Envelope membrane protein	<i>cemA</i>
Genes of unknown function	C-type cytochrome synthesis gene	<i>ccsA</i>
	Conserved open reading frames	<i>ycf1</i> ^(×2) , <i>ycf2</i> ^(×2)

a gene is in *Cyananthus flavus*; c gene is in *Cyclocodon parviflorus*; d gene is in *Codonopsis hongii*;

* gene contains one intron;

** gene contains two introns; (×2) indicates that the number of the repeat unit is 2.

<https://doi.org/10.1371/journal.pone.0233167.t002>

region (79,041bp) of the *Cyphia crenata* was the shortest, while its IR region (45,915bp) and the coding region (92,511bp) were the longest among the studied species, which might be related to the expansion of the border positions between the LSC and IR regions [52, 53]. The length of LSC regions of Campanuleae species was 100,110bp-105,555bp, which were longer than the other species, whereas this group had the shortest IR, with length of 27,276bp-29,637bp, which might be caused by the contraction between the LSC and IR regions. The size of plastid genome was similar among the six species of Cyanantheae (Table 1), apart from *Platycodon grandiflorus* with the longest IR region and shortest LSC and SSC region among the species of this group.

As shown in Table 2, 44.05%-46.52% sequences of plastid were responsible for coding among the Campanuleae species, but more than half sequences being in charge of coding among the other studied species. The GC contents of the LSC and SSC regions in all studied species (except for *P. grandiflorus*) were slightly lower than those of the IR regions. The *Lobelia erinus* plastid genome had the highest GC content (39.0%), while the *Cyphia crenata* plastid genome had the lowest GC content (36.38%). For the Campanuleae species, they showed more GC content in the IR region (41.1% or 142.0%). The overall GC content is a significant species indicator [54]. In addition, 80 or 83 protein-coding genes were identified in the Campanuleae species, with 7 genes in the IR regions. 86–95 were identified in the Cyanantheae species, with 13–21 genes located in the IR regions. 99 protein-coding genes were found in the *Cyphia crenata*, with 21 genes in the IR regions. Four conserved rRNAs were checked in every species. The *T. caeruleum* plastome encodes 44 types of tRNAs, whereas other species encodes 36–38 (Table 1).

The plastid genome structure comparison using MAUVE software revealed that the plastomes of all the accessions were not conserved, and many rearrangements of gene organization had occurred (S2 Appendix). We identified some obvious differences, such as gene composition, gene order, GC content, IR junction in the plastomes of the Campanulaceae, although the plastid genomes of land plants are commonly supposed to be highly conserved [55].

On the other side, we divided the eleven species within Campanulaceae into four groups according to the phylogenetic results of this study, they were the groups of all species, the Campanuleae, the Cyanantheae and the out-group of *Lobelia erinus* and *Cyphia crenata*. The nucleotide diversity (π) value of four groups was calculated to evaluate the sequence divergence among the 76 protein-coding genes of plastomes (Fig 2 and S1 Table), with the mean value of 0.06649 in the out-group, 0.05687 in the Campanuleae species, and 0.03394 in the Cyanantheae species. The analysis revealed that all four groups exhibited the high levels of divergence in the *ccsA* and *ndhF* gene of the SSC regions, which indicated that the SSC region might be undergoing rapid nucleotide substitution in species of family Campanulaceae and contain variable information for species authentication and phylogenetic analysis. *Ycf1* and *ycf2* gene were the hotspot regions for each group. Furthermore, we also identified two hotspot regions (*rpl22* and *rps3* gene, $\pi > 0.1$) for the group of all species, while the other three groups did not show the high divergence in above two genes. Many fragments of coding genes, such as *atpB*, *matK*, *ndhF*, have been used for phylogenetic reconstructions at various taxonomic levels [56–58]. We could use the hotspot regions acquired from this study to develop the potential markers, which would be helpful not only in identifying the species, but also in the reconstruction of phylogeny within different groups of Campanulaceae in further studies.

IR contraction and expansion

It is well known that the IR regions facilitated the stability of the other regions of the genome by intramolecular recombination, thus limiting recombination between the LSC and SSC

regions [59, 60]. The expansion and contraction of IR regions at the borders are considered the major reasons for genome size differences, and are best to study the phylogeny and the plastid genome evolution history of plants [61–63]. We checked the differences of the borders among the IR, LSC and SSC regions of 9 genera. The differences of genes located in the IR region were also examined. Detailed comparisons of the boundaries among the studied plastomes were presented in Fig 3. The *ndhE* gene crossed the IRa and SSC regions for the Campanulaceae species, and the boundary between SSC and IRb regions was in the *ndhF-ndhG* spacer (Fig 3). The *ndhF* gene was complete in the SSC region, more than 200bp away from the IRb region.

Cyanantheae species and *Lobelia erinus* had the same IRa/LSC borders: the *rps19* gene in the LSC region and the *rpl2* gene in the IRa region. The IR regions contained the *rpl2*, *trnI-CAU*, *ycf2*, *trnL-CAA*, *ndhB*, *rps7*, *ycf1*, *rps15*, *ndhH*, *ndhA*, *ndhI*, *ndhG*, and part of *ndhE* genes. It was worth mentioning that *P. grandiflorus* had the IRa/LSC boundary spanning the *rpl36* gene. Besides, this species had the similar gene contents to the other species of Cyanantheae, coupled with the *rps8*, *rpl14*, *rpl16*, *rps3*, *rpl22*, *rps19*, and part of *rpl36* gene. The species of this group showed no IR expansion and contraction, which were canonical IR and similar to *L. erinus*.

For the Campanuleae species, the IRa/LSC boundary was located between the *trnL-CAA* gene and the *ycf2* gene. There were only eight complete genes, *trnL-CAA*, *clpP* genes, etc., in the IR region. The *ycf2* genes appeared in the LSC regions. The length of the IRa regions of three Campanuleae species, varying in the range of 27,276–29,637 bp, was shorter than the eight other species, which varied from 37,290–45,915 bp (Table 1). Species of Campanuleae occurred the IR-contracted out of LSC, and the large IR contractions have been rarely reported, and the most plausible explanation is considered as illegitimate recombination [64–66].

Plastome of *Cyphia crenata* experienced IR-expanded into LSC, which lead to the largest plastome of studied Campanulaceae (Fig 3). The *petB* gene of *Cyphia crenata* crossed the IRa/LSC region, with 187bp located in the LSC region and 2,595 bp in the IRa region. IR region of *Cyphia crenata* had the part of *petB* gene, *petD* gene, ORF 159, ORF 180 and ORF119, which did not show in the IR regions of the other studied species. We also calculated the ORFs >300 bp in the IRa region, among the eleven species, and the results illustrated that there were five ORFs appearing in the IRa regions of *Cyphia crenata*, with total length of 2,211bp. However, other species had 1–3 ORFs, with length of 324 to 1,230bp (S2 Table). *Cyphia crenata* was the only species indicating the IR region expanded into LSC region. It was hypothesized that the longer sequences of ORFs appearing in the IRa regions might be closely associated with IR expansion. Additionally, the IR region of *Cyphia crenata* had more tandem and dispersed repeats compared with the LSC region and SSC region. Previous studies have suggested that the intramolecular recombination, the occurrence of many various repeat sequences, and the insertion-deletions may interpretate the variety of the IR boundary region sequences [59, 67–69], which could also be applied to explain the large IR expansion of *Cyphia crenata*.

The IR expansion and contraction of this study provided new evidence for the classification of Campanulaceae *s.str.* at the genome level. Based on the species included in this study, the group of Cyanantheae species with canonical IR was sister to the Campanuleae species having the IR-contracted out of LSC regions, which was consistent with previous studies about the subdivision of Campanulaceae *s.str.* [4, 10, 11]. In addition, the IRa/LSC boundary and the IR contents of Cyanantheae species were similar but different from Campanuleae species, with the exception of *Platycodon grandiflorum*.

Overall, the junction positions of LSC/IRa regions varied slightly in the plastid genomes of Campanulaceae, and the genes existed in the IRa region were also different in the studied groups. Whereas, the boundary of IRa/SSC of all species had the similar pattern. The events of

IR expansion and contraction are helpful to research subdivision of *Campanulaceae s.str.* and the genome evolution among the *Campanulaceae* species.

Codon usage bias

The plastid genome of *Campanulaceae* was detected for its codon usage frequency according to sequences of protein-coding genes and relative synonymous codon usage (RSCU). RSCU refers to the relative probability of a codon encoding a corresponding amino acid synonymous codon, which eliminates the effect of amino acid composition on codon usage [70]. The pattern of the codon preference has the vital role in studying species evolution [71–73]. The analytic varieties provided by statistical analyses of all 76 protein-coding cpDNA and amino acid sequences demonstrated obvious codon preferences. It showed the similarity of protein codons in the *Campanulaceae* species, of which AGA had the highest frequencies, and CGC had the least occurrence frequencies (Figs 4 and 5). 64 codon preferences were identified, with 20 amino acids and one stop codon involved. The standard ATG codon was typically the start codon for nearly all protein-coding genes. All three stop codons were present, with UAA being the most frequent stop codon in all eleven plastomes. RSCU values of methionine (AUG) and tryptophan (UGG) were equal to one and encoded by only one codon, indicating no codon bias for these two amino acids. All the protein-coding genes were composed of 42,552–48,095 codons as shown in S3 Table.

As shown in Fig 5, the result of the distributions and the visualization of codon usage in the form of heatmaps of *Campanulaceae* species showed that approximately half of the codons were not frequently used. These codons had the RSCU value of >1 , and most of these (25/28, 89.3%) ended with base A or U, resulting in the bias for A/T bases. About half of codons had the RSCU value of <1 , and most of those (27/34, 79.41%) ended with base C or G. The third codon shows a high A/U preference, which is a common phenomenon in plastid genomes of higher plant [74–76]. The high RSCU value is possibly caused by the function of the amino acid or the structure of the peptide to avoid mistakes in transcription [77].

Analysis of repeats

This analysis of repeats was only taken one IR into account. In the majority of the studied species, the most dispersed repeats were forward, then palindromic, and the least reverse. The

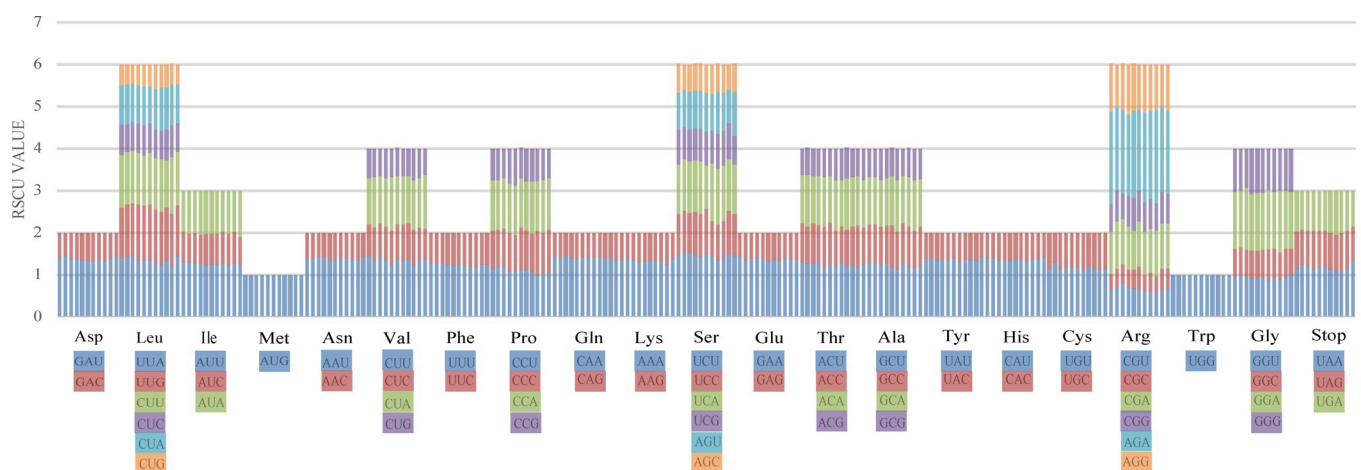


Fig 4. Codon contents of 20 amino acids and stop codons in all protein-coding genes of the *Campanulaceae* plastomes. The color of the histogram corresponds to the color of codons.

<https://doi.org/10.1371/journal.pone.0233167.g004>

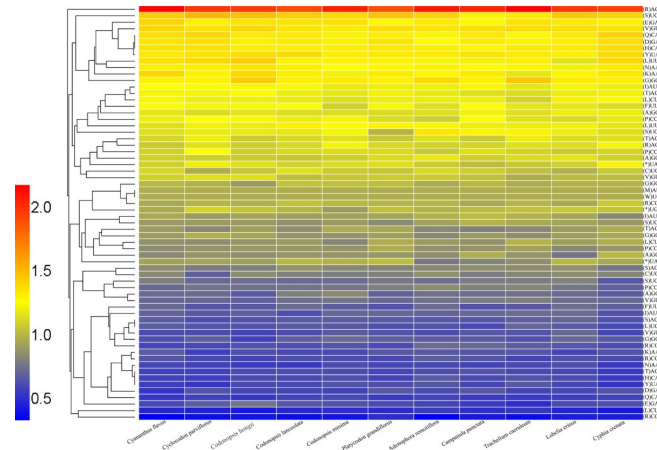


Fig 5. Heatmap analysis for codon distribution of all protein-coding genes of all considered species. Colour key: Higher red values indicate higher RSCU values, and lower blue values indicate lower RSCU values.

<https://doi.org/10.1371/journal.pone.0233167.g005>

comparison analyses (Fig 6) revealed that most of the forward repeats were 50–69 bp, and the longest repeats with length of 1,009 bp, were detected in the *T. caeruleum*, followed by *Campanula punctata* of 640 bp length, and *Adenophora remotiflora* of 620 bp length, which were much longer than other species studied. Besides, in the group of Campanuleae species, dispersed repeats were mainly distributed in non-coding regions (IGS) (S4 Table). Long repeat sequences may be useful to do phylogenetic analysis and increase plastid genome rearrangements [73, 78, 79].

The results also displayed that among the tandem repeats, the repeats located in the spacer of *rpl2-trnI CAU* had appeared in the clade of Cyanantheae and the out-group, but not shown in the Campanuleae species (S5 Table) which had the IR contraction and did not show the *ycf2* gene in the IR region. It indicated that the lack of repeats in *rpl2-trnI CAU* might be linked to the IR contraction. Most and variable tandem repeats (except for species of *Codonopsis minima*, *Trachelium caeruleum* and *Cyphia crenata*) were located in the CDS regions, which might accelerate evolution of coding and regulatory sequences [80].

A large proportion of SSRs was found in the non-coding regions (IGS). We identified A/T/G mononucleotide repeats (p1), while the majority of the dinucleotide repeat sequences (p2) were comprised of AT/TA repeats, and the TG, CA, AC and GT repeats were also found. Furthermore, A and T were the most frequent bases in all SSR types, which resulted in the bias for

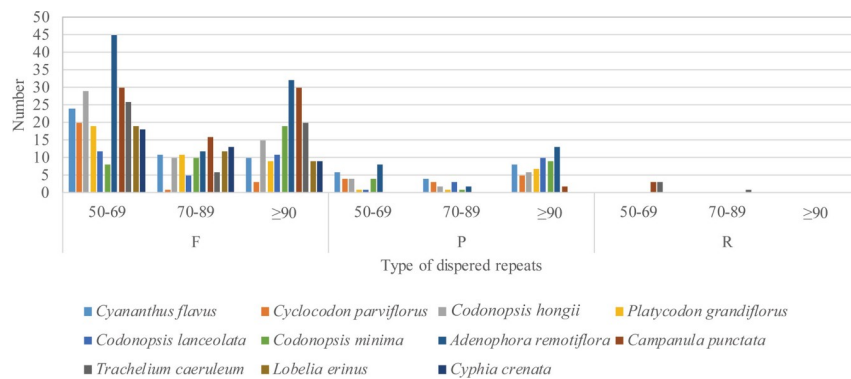


Fig 6. Frequency of three types of dispersed repeats by length. (F: forward, P: palindrome, R: reverse).

<https://doi.org/10.1371/journal.pone.0233167.g006>

the studied plastomes. About half of the species had the compound repeats (S6 Table and Fig 7). Most simple sequence repeats (SSRs) are widely used for species authentication, phylogenetic analysis, and population genetics because of their high levels of polymorphism [81–84]. Microsatellites have a great influence on the genome recombination and rearrangement by their wide distribution across the entire genome [85–87]. The other types of mono-, di-, tri-, tetra- and penta- nucleotide were identified at a much lower frequency among the Campanulaceae species and other plants [88–90].

The total plastome regions of all Campanulaceae possessed the highest number of tandem, dispersed and SSR repeats (S4–S6 Tables and Fig 8), and SSC regions had the lowest number of these repeats. SSR repeats of LSC regions contained higher number of repeats compared with IRa and SSC regions. Tandem repeats of IRa regions had more repeats than LSC and SSC regions in some species, while less in species of Campanuleae. However, Campanuleae had more tandem repeats in LSC regions, which may be guessed that this phenomena is relevant to the IR contraction [91–93].

There was nearly no dispersed repeats in the SSC regions, except the *Adenophora remotiflora* with more than 89 dispersed repeats. The results showed that dispersed repeats of IR regions of Cyanantheae appeared more frequently than in LSC regions, except for *Cyananthus flavus*. The presence of all types of repeats demonstrated that the locus was a crucial hotspot for genome reconfiguration [94–97]. Moreover, the repeats of plastid genomes will be helpful for identifying polymorphisms at the species level for deducing distant phylogenetic relationships among Campanulaceae species. Repeats were previously inferred to associate with plastome structural variation [98–101]. In this study, the plastomes of all studied species possessed large amount of repeats and longer repeats, and presented the structural variations. These together supposed that repeats might also affect size variation in the Campanulaceae plastomes.

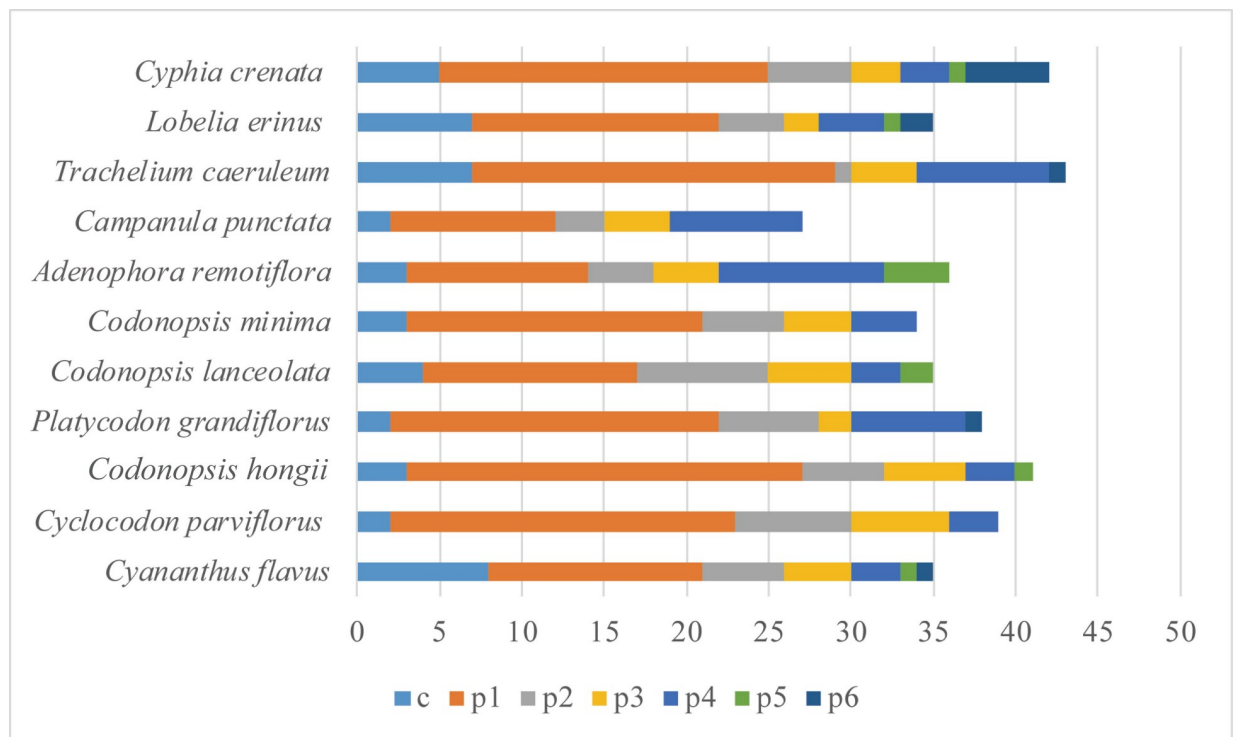


Fig 7. The distribution maps of simple sequence repeats (SSR). Classification of SSRs by repeat types. p1, mononucleotides (mono-); p2, dinucleotides (di-); p3, trinucleotides (tri-); p4, tetranucleotides (tetra-); p5, pentanucleotides (penta-); p6, hexanucleotides (hex-); c, compound.

<https://doi.org/10.1371/journal.pone.0233167.g007>

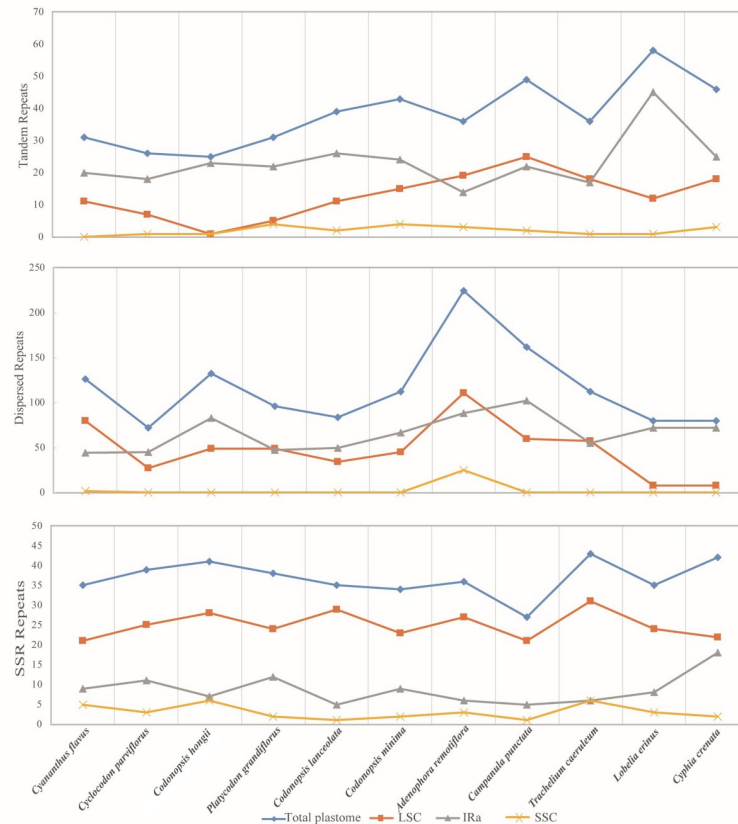


Fig 8. Repeat number in the different regions of Campanulaceae plastomes, including Tandem repeats, Dispersed repeats and SSR repeats. The yellow line refers to SSC regions, the gray lines refers to IRa regions, the orange line refers to LSC regions, and the blue line refers to the total plastome.

<https://doi.org/10.1371/journal.pone.0233167.g008>

Positive selection analysis

We compared the ratio of non-synonymous (dN) and synonymous (dS) substitution for 76 protein-coding genes among the newly sequenced species with other eight species. We focused on the Bayes Empirical Bayes (BEB) analysis of Paml and the highly positively selected sites of $P > 99\%$ (**), because one slightly positive selection had more than 10 positive sites of $P > 95\%$. Finally, fourteen genes with highly positively selected sites within the Campanulaceae family were identified (S7 and S8 Tables). Those genes contained one subunit of Protease (*clpP*), two NADH-dehydrogenase subunit genes (*ndhD*, *ndhI*), two photosystem II subunit genes (*psbL*, *psbN*), one ribosome large subunit gene (*rpl16*), six ribosome small subunit genes (*rps3*, *rps4*, *rps8*, *rps11*, *rps12*, *rps18*), and the *ycf1*, *ycf2* gene. According to the M2 and M8 models, *ndhI*, *psbI* and *rps3* only had one sites under highly positive selection. The gene *ycf1* and *ycf2* harbored more than 30 highly positive selections, followed by *clpP* (7,11), *ndhD* (10, 0), *psbN* (0,2), *rpl16* (3,4), *rps4* (3,6), *rps8* (0,2), *rps11* (1,1), *rps12* (18, 22). Likelihood ratio tests (M0 vs. M3, M1 vs. M2 and M7 vs. M8) supported the presence of highly positively selected codon sites (S8 Table). Some studies have indicated that *ycf1* is required for plant viability and encodes Tic214, which is a vital component of the TIC complex in *Arabidopsis* [102–104]. Most genes under positive selection have the functions in genetic system or photosynthesis, which demonstrate that the functional genes of plastid have important significance during the plant evolution [105–108].

There existed limitation in the study of natural selection by using branch and site modes separately because for the majority of genes in a specified branch, only few sites were under positive selection, however, branch-site model allowed us to detect the various selective pressure on the labeled foreground lineage against the remanent background branches [48]. After the analysis of BEB, we found 96 sites under potentially positive selection in the 76 protein-coding genes with posterior probabilities more than 0.95 and 10 sites greater than 0.99 (S9 Table). The branches of Cyanantheae, Campanuleae and out-group all showed there were positively selected sites in *ycf1* and *ycf2*, and there were more detected on the branch of Cyanantheae for *ycf2*. Campanuleae lineage demonstrated the positively selected sites in *rpl16* but did not reveal the positively selected sites in *rps2*, *rps3*, *rps4*, *rps11* and *rps15* although the LRT *p*-value was less than 0.05. The out-group branch showed one positively site in *ndhI*. *rpoA* gene also did not have positively selected sites in the branch of Cyanantheae. It has been shown that the high rate of molecular evolution existing in numerous genes following genome duplication actuates the functional changes [109, 110]. Besides, the positive selection is concerned with the shift of function and environment [109, 111]. Therefore, positively selected sites detected in this study may drive the protein-coding genes allowing occupation of diverse habitats [48, 109].

Phylogenetic analysis

In recent years, more plastid genome database provides an important basis for the determination of the evolutionary, taxonomic, and phylogenetic studies of plants [51, 112–116].

Phylogenetic analysis was performed by ML and BI nucleic acid analyses based on the 76 aligned sequences of plastomes (Fig 9). *Lobelia erinus* and *Cyphia crenata* were used as out-group. The two typologies showed similar phylogenetic patterns. The ML tree revealed that Campanulaceae *s.str.* formed a monophyletic clade, and Cyanantheae and Campanuleae were also monophyletic. The bootstrap value of previous researches on the phylogenetic relationships of Cyanantheae was relatively low by using ITS sequence and several plastid markers [4, 11]. However, the relationships of Cyanantheae species were well supported in this study. All nodes in the phylogenetic tree were strongly supported, with 100% bootstrap (BP) values and 1.00 Bayesian posterior probabilities (PP). From phylogenetic analysis, Cyanantheae

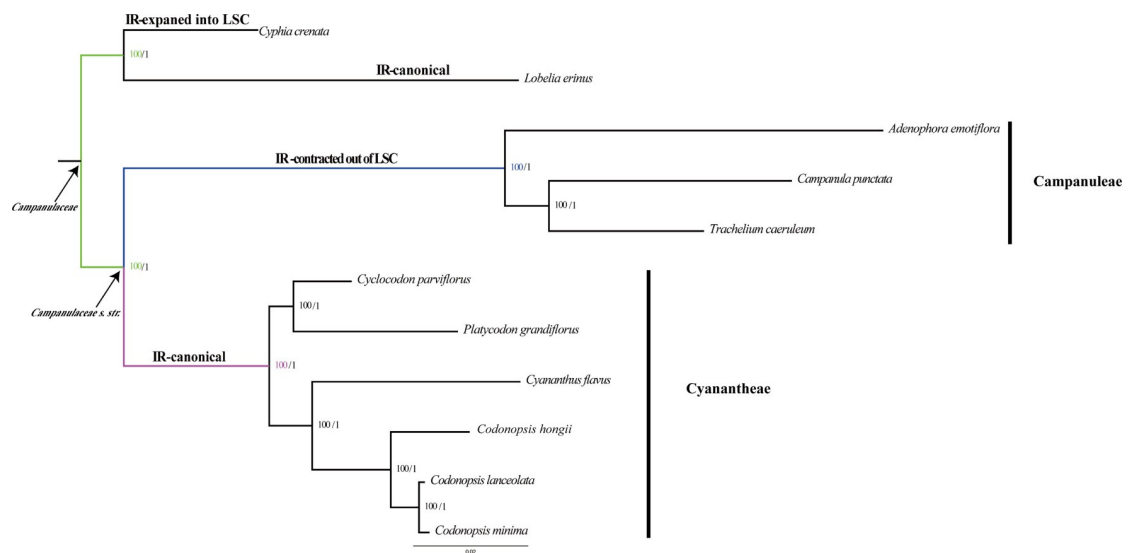


Fig 9. Phylogenetic relationship of all Campanulaceae species by using the 76 protein-coding genes, based on the Maximum likelihood (ML) analysis and Bayesian inference (BI) analysis.

<https://doi.org/10.1371/journal.pone.0233167.g009>

species were divided into two clades. One clade consisting of *Cyclocodon parviflorus* and *P. grandiflorus* was the earlier diverging lineage in the group of Cyanantheae. The other clade was composed of *Cyananthus flavus*, *Codonopsis hongii*, *Codonopsis lanceolate* and *Codonopsis minima*. *Codonopsis hongii* was a sister species to other *Codonopsis* species. *Cyclocodon parviflorus* had a close relationship with *Platycodon grandiflorus*. Previous studies had demonstrated that *Cyclocodon* was restored as the separate genus only based on the morphology of pollen and seed coat, plus the gross morphological characters [12, 117]. In this study, *Cyclocodon* was not closely related to *Codonopsis* and had different structures of plastid genomes compared with *Codonopsis* species (Fig 3), which supply the extra evidence for confirming *Cyclocodon* at the generic rank. *Cyananthus* were treated as the generic rank by the former researches, but the phylogenetic relationships between *Codonopsis* and *Cyananthus* were weakly supported [12]. Nevertheless, *Cyananthus flavus* being related to all studied *Codonopsis* species was demonstrated in our study with strong supports based on the 76 protein coding genes. Therefore, successful phylogenetic construction for eleven Campanulaceae species studied here imply that plastid genome database will be a potentially useful resource for molecular phylogeny studies within the order Cyanantheae.

The results also indicated that it was helpful to illustrate phylogenetic analysis of species in the family Campanulaceae. The phylogenetic tree constructed in this study showed that Cyanantheae formed a sister clade to Campanuleae clade (Fig 9), which is consistent with the previous studies [9, 11]. Therefore, it is hypothesized that Cyanantheae had an earlier divergence among the Campanulaceae from a common ancestor than Campanuleae species because Campanuleae had a unique IR contraction structure (Fig 3). The phylogenetic relationships of Campanuleae have been explored by using the coding regions of plastomes [24].

Conclusions

We first reported the complete plastid genome sequences of three Asian Cyanantheae species (*Cyananthus flavus*, *Cyclocodon parviflorus*, and *Codonopsis hongii*) and compared these to published species in the family Campanulaceae. The results of the genome structural comparison indicated the large amount of rearrangements and various repeat sequences. The junctions between the LSC region and IRa region manifested the diverse locations in different clades. IR contraction/expansion might be explained by the multiple repeat sequences, the indels and the recombination. Fourteen genes with highly positively selected sites within the Campanulaceae family had been identified, and most of them were genetic system or photosynthesis related genes. Branch-site model revealed many positively selected sites in certain genes on the specified branches, which may offer the important significances for the adaptive evolution. We also discussed the type of the codon preference, which had the vital roles in studying species' evolution. Six coding-regions (*ccsA*, *ndhF*, *rpl22*, *rps3*, *ycf1* and *ycf2*) in the highly variable regions will be utilized as potential molecular markers for constructing the phylogenetic relationships of the family Campanulaceae. Phylogenetic analysis indicated that *Cyananthus* was more closely related to *Codonopsis* compared with *Cyclocodon* and clearly showed the relationship among the Cyanantheae species. The plastid genomes will contribute to the development of genetic resources in resolving the phylogenetic analysis and species authentication of Campanulaceae, and in facilitating the exploration their structural differences. Nevertheless, only limited species were shown in this study, and thus, we believe that further studies that include various species having the information of plastomes, are needed to clarify the molecular evolution and phylogenetic relationships of Campanulaceae.

Supporting information

S1 Table. The midpoint and pi value among the different groups of Campanulaceae species.

(XLSX)

S2 Table. ORFs (more than 300bp) showing in the plastomes.

(XLSX)

S3 Table. Putative preferred codons in the Campanulaceae plastid genomes.

(XLSX)

S4 Table. Dispersed repeats found in the plastomes.

(XLSX)

S5 Table. Tandem repeats among the studied species.

(XLSX)

S6 Table. SSRs showing in the plastomes of Campanulaceae species.

(XLSX)

S7 Table. Maximum likelihood parameter estimates for the 76 genes of Campanulaceae species.

(XLSX)

S8 Table. Likelihood ratio test (LRT) of the variable ω ratio using site model.

(XLSX)

S9 Table. Parameter estimates and likelihood values for 76 protein-coding genes inferred using branch-site model.

(XLSX)

S1 Appendix. Primers used for assembly validation.

(DOCX)

S2 Appendix. Mauve result of the plastid genomes of eleven Campanulaceae species.

(DOCX)

Acknowledgments

We appreciate the assistance of Prof. Jun-Bo Yang (Kunming Institute of Botany, Chinese Academy of Sciences) during the experiments.

Author Contributions

Conceptualization: Chun-Jiao Li, De-Zhu Li.

Data curation: Chun-Jiao Li, Ruo-Nan Wang.

Formal analysis: Chun-Jiao Li, Ruo-Nan Wang.

Funding acquisition: De-Zhu Li.

Investigation: Chun-Jiao Li.

Methodology: Chun-Jiao Li, Ruo-Nan Wang.

Project administration: De-Zhu Li.

Resources: Chun-Jiao Li.

Software: Chun-Jiao Li, Ruo-Nan Wang.

Supervision: De-Zhu Li.

Writing – original draft: Chun-Jiao Li.

Writing – review & editing: Chun-Jiao Li, De-Zhu Li.

References

1. Lammers TG. World checklist and bibliography of Campanulaceae. Richmond: Royal Botanic Gardens, Kew; 2007b. pp. 1–675.
2. Schönland S. Campanulaceae. In: Engler A, Prantl K, editors. Die Natürlichen Pflanzenfamilien. Leipzig, IV; 1889. pp. 40–70.
3. Cellinese N, Smith SA, Edwards EJ, Kim ST, Haberle RC, Avramakis M, et al. Historical biogeography of the endemic Campanulaceae of Crete. *J Biogeogr.* 2009; 36(7): 1253–1269.
4. Hong DY, Wang Q. A new taxonomic system of the Campanulaceae s.s.: system of Campanulaceae s.s. *J Syst Evol.* 2014; 53(3): 203–209.
5. Lee KJ, You HJ, Park SJ, Kim YS, Jeong HG. Hepatoprotective effects of *Platycodon grandiflorum* on acetaminophen-induced liver damage in mice. *Cancer Lett.* 2002; 174(1): 73–81.
6. Li Z, Zhu L, Zhang H. Protective effect of a polysaccharide from stem of *Codonopsis pilosula* against renal ischemia/reperfusion injury in rats. *Carbohydr Polym.* 2012; 90(4): 1739–1743. <https://doi.org/10.1016/j.carbpol.2012.07.062> PMID: 22944441
7. Zhang L, Wang Y, Yang D, Zhang C, Li M, Liu Y. *Platycodon grandiflorus*-an ethnopharmacological, phytochemical and pharmacological review. *J Ethnopharmacol.* 2015; 164: 147–161. <https://doi.org/10.1016/j.jep.2015.01.052> PMID: 25666431
8. Jiang YP, Liu YF, Guo QL, Jiang ZB, Shi JG. C₁₄-Polyacetylene glucosides from *Codonopsis pilosula*. *J Asian Nat Prod Res.* 2015; 17(6): 601–614. <https://doi.org/10.1080/10286020.2015.1041932> PMID: 26009940
9. Haberle RC, Dang A, Lee T, Penaflor C, Cortes BH, Oestreich A, et al. Taxonomic and biogeographic implications of a phylogenetic analysis of the Campanulaceae based on three chloroplast genes. *Taxon.* 2009; 58(3): 715–734.
10. Crowl AA, Miles NW, Visger CJ, Hansen K, Ayers T, Haberle RC, et al. A global perspective on Campanulaceae biogeographic, genomic, and flora evolution. *Am J Bot.* 2016; 103(2): 233–245. <https://doi.org/10.3732/ajb.1500450> PMID: 26865121
11. Eddie WMM, Shulkina T, Gaskin J, Haberle RG, Jansen RK. Phylogeny of Campanulaceae s.str. inferred from ITS sequences of nuclear ribosomal DNA. *Ann Mo Bot Gard.* 2003; 90(4): 554–575.
12. Wang Q, Zhou SL, Hong DY. Molecular phylogeny of the platycodonoid group (Campanulaceae s.str.) with special reference to the circumscription of *Codonopsis*. *Taxon.* 2013; 62(3): 498–504.
13. He X, Burgess K, Gao LM, Li DZ. Distributional responses to climate change for alpine species of *Cyananthus* and *Primula* endemic to the Himalaya-Hengduan Mountains. *Plant Diversity.* 2019; 41(001): 26–32.
14. Hong DY, Song G, Lammers TG, Klein LL. Campanulaceae. In: Wu ZY, Raven PH, Hong DY, editors. *Flora of China*. Science Press, Beijing, and Missouri Botanical Garden Press, St. Louis; 2011. pp. 505–563.
15. Lemieux C, Otis C, Turmel M. Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. *Nature.* 2000; 403(6770): 649–652. <https://doi.org/10.1038/35001059> PMID: 10688199
16. Moore MJ, Soltis PS, Bell CD, Gordon B, Soltis DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc Natl Acad Sci.* 2010; 107(10): 4623–4628. <https://doi.org/10.1073/pnas.0907801107> PMID: 20176954
17. Knox EB. The dynamic history of plastid genomes in the Campanulaceae *sensu lato* is unique among angiosperms. *Proc Natl Acad Sci.* 2014; 111(30): 11097–1102. <https://doi.org/10.1073/pnas.1403363111> PMID: 25024223
18. Knox EB, Li CJ. The East Asian origin of the giant lobelias. *Am J Bot.* 2017; 104(6): 924–938. <https://doi.org/10.3732/ajb.1700025> PMID: 28645921

19. Wicke S, Schneeweiss GM, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol*. 2011; 76(3–5): 273–297. <https://doi.org/10.1007/s11103-011-9762-4> PMID: 21424877
20. Cai J, Ma PF, Li HT, Li DZ. Complete plastid genome sequencing of four *Tilia* species (Malvaceae): a comparative analysis and phylogenetic implications. *PLoS One*. 2015; 10(11).
21. Xu C, Dong WP, Li WQ, Lu YZ, Xie XM, Jin XB, et al. Comparative analysis of six *Lagerstroemia* complete chloroplast genomes. *Front Plant Sci*. 2017; 8: 15–26. <https://doi.org/10.3389/fpls.2017.00015> PMID: 28154574
22. Raubeson LA, Jansen RK. Chloroplast genomes of plants. In: Henry RJ, editor. *Diversity and evolution of plants-genotypic and phenotypic variation in higher plants*. Wallingford (UK): CABI publishing; 2005. pp. 45–68.
23. Palmer JD, Jansen RK, Michaels HJ. Chloroplast DNA variation and plant phylogeny. *Ann Mo Bot Gard*. 1988; 75(4): 1180–1206.
24. Kyeong SC, Kyung AK, Ki OY, Xiu QL. The complete chloroplast genome sequences of three *Adenophora* species and comparative analysis with *Campanuloid* species (Campanulaceae). *PLoS One*. 2017; 12(8): e0183652–.
25. Li HT, Yi TS, Gao LM, Ma PF, Zhang T, Yang JB, et al. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat Plants*. 2019; 5: 461–470. <https://doi.org/10.1038/s41477-019-0421-0> PMID: 31061536
26. Cosner ME, Jansen RK, Lammers TG. Phylogenetic relationships in the Campanulales based on *rbcl* sequences. *Plant Syst Evol*. 1994; 190(1–2): 79–95.
27. Hong CP, Park J, Lee Y, Lee M, Park SG, Uhm Y, et al. *AccD* nuclear transfer of *Platycodon grandiflorum* and the plastid of early Campanulaceae. *BMC Genomics*. 2017; 18(1): 607–. <https://doi.org/10.1186/s12864-017-4014-x> PMID: 28800729
28. Doyle JJ. A rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochem Bull*. 1987; 19: 11–15.
29. Jin JJ, Yu WB, Yang JB, Song Y, Yi TS, Li DZ. GetOrganelle: a simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. *BioRxiv*. 2018; 256479.
30. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*. 2015; 31(20): 3350–3352. <https://doi.org/10.1093/bioinformatics/btv383> PMID: 26099265
31. Kearse M, Moir R, Wilson A, Stones HS, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012; 28(12): 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199> PMID: 22543367
32. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 2019; 15(1): 1–12.
33. Lohse M, Drechsel O, Kahlau S, Ralph B. Organellar genome DRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res*. 2013; 41(1): 575–581. <https://doi.org/10.1093/nar/gks1075>
34. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 2004; 14(7): 1394–1403. <https://doi.org/10.1101/gr.2289704> PMID: 15231754
35. Katoh K, Standley DM. MAFFT: iterative refinement and additional methods. *Multiple sequence alignment methods*. Humana Press, Totowa, NJ; 2014. pp. 131–146.
36. Rozas J, Ferrer-mata A, Sánchez-delbarrio JC, Guirao-rico S, Librado P, Ramosonsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large datasets. *Mol Biol Evol*. 2017; 34(12): 3299–3302. <https://doi.org/10.1093/molbev/msx248> PMID: 29029172
37. Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013; 30: 2725–2729. <https://doi.org/10.1093/molbev/mst197> PMID: 24132122
38. Kumar S, Nei M, Dudley J, Tamura K. MEGA: biologist centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform*. 2008; 9(4): 299–306. <https://doi.org/10.1093/bib/bbn017> PMID: 18417537
39. Sharp PM, Li WH. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*. 1987; 15(3): 1281–1295. <https://doi.org/10.1093/nar/15.3.1281> PMID: 3547335
40. Gupta SK, Bhattacharyya TK, Ghosh TC. Synonymous codon usage in *Lactococcus lactis*: mutational bias versus translational selection. *J Biomol Struct Dyn*. 2004; 21(4): 527–536. <https://doi.org/10.1080/07391102.2004.10506946> PMID: 14692797

41. 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): 4633–4642. <https://doi.org/10.1093/nar/29.22.4633> PMID: 11713313
42. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet.* 2003; 106(3): 411–422. <https://doi.org/10.1007/s00122-002-1031-0> PMID: 12589540
43. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 2006; 22(21): 2688–2690. <https://doi.org/10.1093/bioinformatics/btl446> PMID: 16928733
44. Yang ZH. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007; 24:1586–91. <https://doi.org/10.1093/molbev/msm088> PMID: 17483113
45. Gao F, Du Z, Shen J, Yang H, Liao F. Genetic diversity and molecular evolution of *Ornithogalum mosaic* virus based on the coat protein gene sequence. *Peer J.* 2018; 6(11): e4550.
46. Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. EasyCodeML: a visual tool for analysis of selection using CodeML. *Ecol Evol.* 2019; 9(7): 3891–3898. <https://doi.org/10.1002/ece3.5015> PMID: 31015974
47. Yang Z, Wong WSW, Nielsen R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 2005; 22(4): 1107–1118. <https://doi.org/10.1093/molbev/msi097> PMID: 15689528
48. Dhar D, Dey D, Basu S, Fortunato H. Understanding the adaptive evolution of mitochondrial genomes in intertidal chitons. *BioRxiv.org.* 2020.
49. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 2012; 9(8): 772.
50. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *GCE.* 2010; 14: 1–8.
51. Boudreau E, Takahashi Y, Lemieux C, Turmel M, Rochaix JD. The chloroplast *ycf3* and *ycf4* open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. *EMBO J.* 1997; 16(20): 6095–6104. <https://doi.org/10.1093/emboj/16.20.6095> PMID: 9321389
52. Jer JD. Plastid chromosomes: structure and evolution. In *cell culture and somatic cell genetics in plants, the Molecular Biology of Plastids 7A*; Vasil I.K., Bogorad L., Eds.; Academic Press: San Diego, CA, USA; 1991. pp. 5–53.
53. Bendich AJ. Circular chloroplast chromosomes: The grand illusion. *Plant Cell.* 2004; 16(7): 166–1666.
54. Shen X, Wu M, Liao B, Liu Z, Bai R, Xiao S, et al. Complete chloroplast genome sequence and phylogenetic analysis of the medicinal plant *Artemisia annua*. *Molecules.* 2017; 22(8): 1330–.
55. Mower JP, Vickrey TL. Structural diversity among plastid genomes of land plants. *Adv Bot Res.* 2018; 85: 263–292.
56. Li JH. Phylogeny of *Catalpa* (Bignoniaceae) inferred from sequences of chloroplast *ndhF* and nuclear ribosomal DNA. *J Syst Evol.* 2008; 46(3): 341–348.
57. Peterson PM, Romaschenko K, Johnson G. A classification of the Chloridoideae (Poaceae) based on multi-gene phylogenetic trees. *Mol Phylogenet Evol.* 2010; 55(2): 580–598. <https://doi.org/10.1016/j.ympev.2010.01.018> PMID: 20096795
58. Wilson CA. Phylogenetic relationships among the recognized series in *Iris* section *Limniris*. *Syst Bot.* 2009; 34(2): 277–284.
59. Palmer JD. Plastid chromosomes: structure and evolution. In *Molecular Biology of Plastids*; Bogorad L, editors. Academic Press: San Diego, CA, USA; 1991. pp. 5–53.
60. Wang Y, Zhan DF, Jia X, Mei WL, Dai HF, Chen XT, et al. Complete chloroplast genome sequence of *Aquilaria sinensis* (Lour.) Gilg and evolution analysis within the Malvales order. *Front Plant Sci.* 2016; 7: 280. <https://doi.org/10.3389/fpls.2016.00280> PMID: 27014304
61. Yang Y, Dang Y, Li Q, Lu J, Li X, Wang Y. Complete chloroplast genome sequence of poisonous and medicinal plant *Datura stramonium*: organizations and implications for genetic engineering. *PloS One.* 2014; 9(11): e110656-. <https://doi.org/10.1371/journal.pone.0110656> PMID: 25365514
62. Shetty SM, Md Shah MU, Makale K, Mohd YY, Khalid N, Othman RY. Complete chloroplast genome sequence of corroborates structural heterogeneity of inverted repeats in wild progenitors of cultivated bananas and plantains. *Plant Genome.* 2016; 9(2): 2.
63. Yao X, Tang P, Li Z, Li D, Liu Y, Huang H. The first complete chloroplast genome sequences in *Actinidiaceae*: genome structure and comparative analysis. *PloS One.* 2015; 10(6): e0129347. <https://doi.org/10.1371/journal.pone.0129347> PMID: 26046631

64. Goulding SE, Olmstead RG, Morden CW, Wolfe KH. Ebb and flow of the chloroplast inverted repeat. *Mol Genet Genomics*. 1996; 252(1–2): 195–206.
65. Blazier J, Guisinger MM, Jansen RK. Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). *Plant Mol Biol*. 2011; 76(3–5): 263–272. <https://doi.org/10.1007/s11103-011-9753-5> PMID: 21327834
66. 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): 336–351.
67. Palmer JD, Osorio B, Aldrich J, Thompson WF. Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements. *Curr Genetics*. 1987; 11(4): 275–286.
68. Li B, Zheng Y. Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae. *Sci Rep*. 2018; 8(1): 9285-. <https://doi.org/10.1038/s41598-018-27453-7> PMID: 29915292
69. Yan M, Zhao X, Zhou J, Huo Y, Ding Y, Yuan Z. The complete chloroplast genomes of *Punica granatum* and a comparison with other species in Lythraceae. *Int J Mol Sci*. 2019; 20(12): 2886.
70. Sharp PM, Li WH. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J Mol Evol*. 1986; 24(1): 28–38.
71. Yu X, Zuo L, Lu D, Lu B, Yang M, Wang J. Comparative analysis of chloroplast genomes of five *Robinia* species: genome comparative and evolution analysis. *Gene*. 2019; 689: 141–151. <https://doi.org/10.1016/j.gene.2018.12.023> PMID: 30576807
72. Barbhuiya PA, Uddin A, Chakraborty S. Genome-wide comparison of codon usage dynamics in mitochondrial genes across different species of amphibian genus *Bombina*. *J Exp Zool B Mol Dev Evol*. 2019; 332(3–4): 99–112. <https://doi.org/10.1002/jez.b.22852> PMID: 31033182
73. Somaratne Y, Guan DL, Wang WQ, Zhao L, Xu SQ. The Complete chloroplast genomes of two *Lespedeza* species: insights into codon usage bias, rRNA editing sites, and phylogenetic relationships in Desmodieae (Fabaceae: Papilionoideae). *Plants*. 2020; 9(1): 51.
74. Nie XJ, Lv SZ, Zhang YX, Du XH, Wang L, Biradar SS, et al. Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PLoS One*. 2012; 7(5): e36869-. <https://doi.org/10.1371/journal.pone.0036869> PMID: 22606302
75. Yi DK, Kim KJ. Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. *PLoS One*. 2012; 7(5): e35872. <https://doi.org/10.1371/journal.pone.0035872> PMID: 22606240
76. Zuo LH, Shang AQ, Zhang S, Yu XY, Ren YC, Yang MS, et al. The first complete chloroplast genome sequences of *Ulmus* species by de novo sequencing: genome comparative and taxonomic position analysis. *PLoS One*. 2017; 12(2): e0171264. <https://doi.org/10.1371/journal.pone.0171264> PMID: 28158318
77. Li Y, Kuang XJ, Zhu XX, Zhu YJ, Sun C. Codon usage bias of *Catharanthus roseus*. *Zhongguo Zhong Yao Za Zhi*. 2016; 41: 4165–4168. <https://doi.org/10.4268/cjcm20162213> PMID: 28933083
78. Park I, Yang S, Choi G, Kim WJ, Moon BC. The complete chloroplast genome sequences of *Aconitum pseudolaevae* and *Aconitum longecassidatum*, and development of molecular markers for distinguishing species in the *Aconitum* Subgenus *Lycocotnum*. *Molecules*. 2017; 22(11): 2012-.
79. Cui Y, Nie L, Sun W, Xu Z, Wang Y, Yu J, et al. Comparative and phylogenetic analyses of ginger (*Zingiber officinale*) in the family Zingiberaceae based on the complete chloroplast genome. *Plants*. 2019; 8(8): 283.
80. Gemayel R, Vences MD, Legendre M, Verstrepen KJ. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu Rev Genet*. 2010; 44(1): 445–477.
81. X JH, W S, Z SL. Polymorphic chloroplast microsatellite loci in *Nelumbo* (Nelumbonaceae). *Am J Bot*. 2012; 99(6): 240–244.
82. Yang AH, Zhang JJ, Yao XH, Huang HW. Chloroplast microsatellite markers in *Liriodendron tulipifera* (Magnoliaceae) and cross-species amplification in *L. chinense*. *Am J Bot*. 2011; 98: 123–126.
83. Park I, Yang S, Kim WJ, Noh P, Lee HO, Moon BC. The complete chloroplast genomes of six *Ipomoea* species and indel marker development for the discrimination of authentic *pharbitidis* semen (Seeds of *I. nil* or *I. purpurea*). *Front Plant Sci*. 2018; 9: 965. <https://doi.org/10.3389/fpls.2018.00965> PMID: 30026751
84. Lee SR, Kim K, Lee BY, Lim CE. Complete chloroplast genomes of all six *Hosta* species occurring in Korea: molecular structures, comparative, and phylogenetic analyses. *BMC Genomics*. 2019; 20: 833. <https://doi.org/10.1186/s12864-019-6215-y> PMID: 31706273
85. Ni L, Zhao Z, Xu H, Chen S, Dorje G. The complete chloroplast genome of *Gentiana straminea* (Gentianaceae), an endemic species to the Sino-Himalayan subregion. *Gene*. 2015; 577(2): 281–288. <https://doi.org/10.1016/j.gene.2015.12.005> PMID: 26680100

86. Rogivue A, Choudhury R, Zoller S, Stéphane J, Gugerli F. Genome-wide variation in nucleotides and retrotransposons in alpine populations of *Arabis alpina* (Brassicaceae). *Mol Ecol Resour.* 2019; 19(3): 773–787. <https://doi.org/10.1111/1755-0998.12991> PMID: 30636378
87. Vu HT, Tran N, Nguyen TD, Vu QL, Bui MH, Le MT, et al. Complete chloroplast genome of *Paphiopedilum delenatii* and phylogenetic relationships among Orchidaceae. *Plants.* 2020; 9(1): 61.
88. Yu XQ, Drew BT, Yang JB, Gao LM, Li DZ. Comparative chloroplast genomes of eleven Schima (Theaceae) species: insights into DNA barcoding and phylogeny. *PLoS One.* 2017; 12(6): e0178026. <https://doi.org/10.1371/journal.pone.0178026> PMID: 28575004
89. Thode VA, Lohmann LG. Comparative chloroplast genomics at low taxonomic levels: a case study using *Amphilophium* (Bignoniaceae, Bignoniaceae). *Front Plant Sci.* 2019; 10: 796. <https://doi.org/10.3389/fpls.2019.00796> PMID: 31275342
90. Li W, Zhang C, Guo X, Liu Q, Wang K. Complete chloroplast genome of *Camellia japonica* genome structures, comparative and phylogenetic analysis. *PLoS One.* 2019; 14(5): e0216645. <https://doi.org/10.1371/journal.pone.0216645> PMID: 31071159
91. Choi KS, Jeong KS, Ha YH, Choi K. Complete chloroplast genome sequences of *Clematis*: IR expansion and relative rates of synonymous substitutions. *Preprints.* 2018; 2018040106.
92. Mehmood F, Abdullah Shahzadi I, Ahmed I, Waheed MT, Mirza B. Characterization of *Withania somnifera* chloroplast genome and its comparison with other selected species of Solanaceae. *Genomics.* 2019; 112(2): 1522–1530. <https://doi.org/10.1016/j.ygeno.2019.08.024> PMID: 31470082
93. Dugas DV, Hernandez D, Koenen EJM, Schwarz E, Straub S, Hughes CE, et al. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clpP*. *Sci Rep.* 2015; 5: 16958. <https://doi.org/10.1038/srep16958> PMID: 26592928
94. Gao L, Yi X, Yang YX, Su Y, Wang T. Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insight into evolutionary changes in fern chloroplast genomes. *BMC Evol Biol.* 2009; 9(1): 130.
95. Asaf S, Waqas M, Khan AL, Khan MA, Kang SM, Imran QM, et al. The complete chloroplast genome of wild rice (*Oryza minuta*) and its comparison to related species. *Front Plant Sci.* 2017; 8: 304. <https://doi.org/10.3389/fpls.2017.00304> PMID: 28326093
96. Shrestha B, Weng ML, Theriot EC, Gilbert LE, Ruhlman TA, Krosnick SE, et al. Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus *Decaloba*. *Mol Phylogenet Evol.* 2019; 138: 53–64. <https://doi.org/10.1016/j.ympev.2019.05.030> PMID: 31129347
97. Zong D, Gan P, Zhou A, Li J, Xie Z, Duan A, et al. Comparative analysis of the complete chloroplast genomes of seven *Populus* species: insights into alternative female parents of *Populus tomentosa*. *PLoS one.* 2019; 14(6).
98. Cosner ME, Jansen RK, Palmer JD, Downie SR. The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. *Curr Genet.* 1997; 31(5): 419–429. <https://doi.org/10.1007/s002940050225> PMID: 9162114
99. Greiner S, Wang X, Rauwolf U, Silber MV, 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.* 2008; 36(7): 2366–2378. <https://doi.org/10.1093/nar/gkn081> PMID: 18299283
100. Ruhlman TA, Zhang J, Blazier JC, Sabir JSM, Jansen RK. Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. *Am J Bot.* 2017; 104: 559–572. <https://doi.org/10.3732/ajb.1600453> PMID: 28400415
101. Choi IS, Jansen R, Ruhlman T. Lost and found: return of the inverted repeat in the legume clade defined by its absence. *Genome Biol Evol.* 2019; 11(4): 1321–1333. <https://doi.org/10.1093/gbe/evz076> PMID: 31046101
102. Shingo K, Jocelyn B, Minako H, Yoshino H, Maya O, Midori I, et al. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science.* 2013; 339(6119): 571–574. <https://doi.org/10.1126/science.1229262> PMID: 23372012
103. Simeone MC, Grimm GW, Papini A, Vessella F, Denk T. Plastome data reveal multiple geographic origins of *Quercus* Group *Ilex*. *Peer J.* 2016; 4(6): e1897.
104. Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, et al. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science.* 2013; 339: 571–574. <https://doi.org/10.1126/science.1229262> PMID: 23372012
105. Hao DC, Chen SL, Xiao PG. Molecular evolution and positive Darwinian selection of the chloroplast maturase *matK*. *J Plant Res.* 2010; 123(2): 241–247. <https://doi.org/10.1007/s10265-009-0261-5> PMID: 19943076

106. Zhang Z, An M, Miao J, Gu Z, Liu C, Zhong B. The Antarctic sea ice alga *Chlamydomonas* sp. ICE-L provides insights into adaptive patterns of chloroplast evolution. *BMC Plant Biol.* 2018; 18(1): 53. <https://doi.org/10.1186/s12870-018-1273-x> PMID: 29614974
107. Jiang P, Shi FX, Li MR, Liu B, Wen J, Xiao HX, et al. Positive selection driving cytoplasmic genome evolution of the medicinally important ginseng plant genus *Panax*. *Front Plant Sci.* 2018; 9: 359. <https://doi.org/10.3389/fpls.2018.00359> PMID: 29670636
108. Heyduk K, Moreno-Villena JJ, Gilman IS, Christin PA, Edwards EJ. The genetics of convergent evolution: insights from plant photosynthesis. *Nat Rev Genet.* 2019; 20(8): 485–493. <https://doi.org/10.1038/s41576-019-0107-5> PMID: 30886351
109. Hu Q, Zhu Y, Liu Y, Wang N, Chen S. Cloning and characterization of wnt4a gene and evidence for positive selection in half-smooth tongue sole (*Cynoglossus semilaevis*). *Sci Rep.* 2014; 4: 7167. <https://doi.org/10.1038/srep07167> PMID: 25418599
110. Delsuc F, Brinkmann H, Philippe H. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet.* 2005; 6: 361–375. <https://doi.org/10.1038/nrg1603> PMID: 15861208
111. Rensch B. Evolution above the species level. New York: Columbia Univ Press; 1960. pp. 124.
112. Yang JB, Tang M, Li HT, Zhang ZR, Li DZ. Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evol Biol.* 2013; 13(1): 84.
113. Raman G, Park SJ. The complete chloroplast genome sequence of *Ampelopsis*: gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. *Front Plant Sci.* 2016; 7: 341. <https://doi.org/10.3389/fpls.2016.00341> PMID: 27047519
114. Zhang Y, Du L, Liu A, Chen J, Wu L, Hu W, et al. The complete chloroplast genome sequences of five *Epimedium* species: lights into phylogenetic and taxonomic analyses. *Front Plant Sci.* 2016; 7: 306. <https://doi.org/10.3389/fpls.2016.00306> PMID: 27014326
115. Kahraman K, Lucas SJ. Comparison of different annotation tools for characterization of the complete chloroplast genome of *Corylus avellana* cv Tombul. *BMC Genomics.* 2019; 20: 874. <https://doi.org/10.1186/s12864-019-6253-5> PMID: 31747873
116. Li X, Zuo Y, Zhu X, Liao S, Ma J. Complete chloroplast genomes and comparative analysis of sequences evolution among seven *Aristolochia* (Aristolochiaceae) medicinal species. *Int J Mol Sci.* 2019; 20(5): 1045.
117. Hong DY, Pan KY. The restoration of the genus *Cyclocodon* (Campanulaceae) and its evidence from pollen and seed-coat. *Acta Phytotax Sin.* 1998; 36: 106–110.