

RESEARCH

Open Access



Characteristics of plastid genomes in the genus *Ceratostigma* inhabiting arid habitats in China and their phylogenomic implications

Yu-Juan Zhao^{1,2,3}, Jian Liu^{1,2,3}, Gen-Shen Yin⁴ and Xun Gong^{1,2,3*}

Abstract

Background *Ceratostigma*, a genus in the Plumbaginaceae, is an ecologically dominant group of shrubs, subshrub and herb mainly distributed in Qinghai-Tibet Plateau and North China. *Ceratostigma* has been the focal group in several studies, owing to their importance in economic and ecological value and unique breeding styles. Despite this, the genome information is limited and interspecific relationships within the genus *Ceratostigma* remains unexplored. Here we sequenced, assembled and characterized the 14 plastomes of five species, and conducted phylogenetic analyses of *Ceratostigma* using plastomes and nuclear ribosomal DNA (nrDNA) data.

Results Fourteen *Ceratostigma* plastomes possess typical quadripartite structures with lengths from 164,076 to 168,355 bp that consist of a large single copy, a small single copy and a pair of inverted repeats, and contain 127–128 genes, including 82–83 protein coding genes, 37 transfer RNAs and eight ribosomal RNAs. All plastomes are highly conservative and similar in gene order, simple sequence repeats (SSRs), long repeat repeats and codon usage patterns, but some structural variations in the border of single copy and inverted repeats. Mutation hotspots in coding (Pi values > 0.01: *matK*, *ycf3*, *rps11*, *rps3*, *rpl22* and *ndhF*) and non-coding regions (Pi values > 0.02: *trnH-psbA*, *rps16-trnQ*, *ndhF-rpl32* and *rpl32-trnL*) were identified among plastid genomes that could be served as potential molecular markers for species delimitation and genetic variation studies in *Ceratostigma*. Gene selective pressure analysis showed that most protein-coding genes have been under purifying selection except two genes. Phylogenetic analyses based on whole plastomes and nrDNA strongly support that the five species formed a monophyletic clade. Moreover, interspecific delimitation was well resolved except *C. minus*, individuals of which clustered into two main clades corresponding to their geographic distributions. The topology inferred from the nrDNA dataset was not congruent with the tree derived from the analyses of the plastid dataset.

Conclusion These findings represent the first important step in elucidating plastome evolution in this widespread distribution genus *Ceratostigma* in the Qinghai-Tibet Plateau. The detailed information could provide a valuable resource for understanding the molecular dynamics and phylogenetic relationship in the family Plumbaginaceae. Lineage genetic divergence within *C. minus* was perhaps promoted by geographic barriers in the Himalaya and Hengduan Mountains region, but introgression or hybridization could not be completely excluded.

Keywords *Ceratostigma*, Plastid genome, Comparative analysis, Interspecific relationship, Plumbaginaceae

*Correspondence:

Xun Gong
gongxun@mail.kib.ac.cn

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



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Cerotastigma Bunge is a small genus falling within the tribe Plumbagineae (Plumbaginaceae), with eight species disjunctively distributed in tropical East Africa and East Asia [1]. South Tibet is the present distribution center for the genus [2]. There are five morphologically distinct species distributed in China, consisting of perennial shrubs, subshrubs and herbs, most of which are typically dominant species in arid environment in the Qinghai-Tibet Plateau *sensu lato* (QTPsl) [3]. For example, the geographic distribution of shrub *C. minus* shows a pattern that extends along xeric valleys of Yarlung Zangbo River and Hengduan Mountains across major gradients of elevation, *C. willmottianum* is an emblematic subshrub endemic to dry valleys mainly in Yunnan and Sichuan, and the cushion shrub *C. ulicinum* grows in open shrubland with a restrictive distribution range on the west QTPsl [2, 4]. As constructive and zonal species, they play crucial roles in sustaining the fragile arid ecosystems [2]. Additionally, all species are of great value in landscapes for their features of drought resistance, salinity tolerance as well as ease of cultivation [5]. And owing to their blue and violet flowers and long flowering time in autumn or winter, some are introduced as garden ornaments outside China [6]. *Cerotastigma* also contain a variety of chemical active ingredients such as plumbagin, which exhibits antitumor, anticancer and antibacterial activities [7, 8]. Therefore, *Cerotastigma* has important ecological and economic value.

To date, the genetic background and resources for *Cerotastigma* remain scarce. Several authors have studied the genus from the following aspects. Li proposed that the genus perhaps had an origin and evolved near Tethys during the early Tertiary [2]. From the breeding system perspective, the results of pollination experiments illustrated that *C. willmottianum* having distyly does not exhibit precise reciprocal herkogamy and was partially self-compatible but primarily outcrossing [9]. In our field survey, we found that the five species of *Cerotastigma* all have heterostyly as many species of Plumbaginaceae and the ratios of long-styled morphs and short-styled morphs exist in different populations with slight differences, which may reflect similar pollination strategies among the different species. Concerning the origins of distyly in Plumbaginaceae, Barrett et al. [10] constructed several models. Their results support the ideas of the more recent selfing avoidance model by D. & B. Charlesworth [11], in which distyly evolves from self-incompatible ancestors other than reciprocal herkogamy. Moreover, the medicinal and horticultural properties place the genus as target studies regarding establishing rapid propagation systems for obtaining plumbagin and developing polyploids used in landscapes [6]. Lastly, as we mentioned above, Plumbaginaceae including *Cerotastigma* species could survive

in harsh environmental conditions like high salt media or heavy-metal-rich soils, and some researchers found a strong correspondence between these capacities and the secretory structures such as salt glands secreting a range of ions, which might have arisen as a means to avoid the toxicity or regulate ion concentrations within leaves [12]. However, up till now, the genome information of *Cerotastigma* is still limited, and no attempt has yet made to explore the interspecific relationships within this genus. Although several studies have incorporated molecular data to study the phylogeny of Plumbaginaceae, in which monophyly of *Cerotastigma* was confirmed by phylogeny of some representative genera or families [13–15], these inferences were derived from limited taxon of *Cerotastigma* or molecular sampling only comprising several plastid fragments, and thus limiting a comprehensive understanding of the interspecific relationships and divergence of *Cerotastigma*.

The plastid is a core organelle in plants and the utilization of complete plastome sequences has been widely accepted to resolve phylogenetic relationships at different taxonomic levels, such as all flowering plant families [16, 17], gymnosperms [18], tribe Cinnamomeae [19] and genus *Rhododendron* [20]. Recently, this strategy has worked for a range of taxa such as those having closely relationships with Plumbaginaceae (e.g., *Calligonum* and *Rheum* within Polygonoideae) [21, 22] and several representatives of *Limonium* belonging to Plumbaginaceae [23]. Despite this, for the species-rich and highly diverse family (27–29 genera) [24, 25], the plastid genomes information of representatives are less well-documented and need to be substantially complemented. To date, only one plastid genome of *Cerotastigma* plus six species from two other Plumbaginaceae genera *Limonium* [15, 23, 26–28] and *Plumbago* [15, 29] have been reported, which account for less than ca. 1% of the currently described ca. 650 Plumbaginaceae species [13, 24, 25]. Thus, to better understand the plastome structure and phylogenomics, it is essential to conduct the comparative analyses for *Cerotastigma* plastid genomes and assess the phylogenetic relationships based on the plastid genomes sequences. In this study, we generated 14 plastomes data of *Cerotastigma* species. To obtain a comprehensive understanding of infrageneric phylogeny, the nuclear ribosomal DNA data was also used to construct phylogenetic tree. Our specific goals were as follows: (1) to compare the plastid structures within *Cerotastigma*; (2) to identify the variable hotspots as potential DNA markers for genetic variation studies; (3) to infer the phylogenetic relationships among *Cerotastigma* species within Plumbaginaceae. These results could provide new sources of information for phylogenetic analyses of Plumbaginaceae and assist with the further population genomics of *Cerotastigma*.

Results

Chloroplast genome feature

The newly sequenced *Cerostasigma* plastid genomes were circular with the typical quadripartite structure with a large single copy (LSC), a small single copy (SSC) and two copies of inverted repeats (IR) regions (Fig. S1). The genome sizes were ranged from 164,076 bp in *C. minus* to 168,355 bp in *C. ulicinum*, and the IR regions were more variable (30,516–32,788 bp) than LSC (88,756–89,988 bp) and SSC regions (13,457–13,534 bp). The overall GC content was 37.3–37.7%, while the IR regions (40.7–41.4%) exhibited higher GC contents than those of both LSC (35.6–36.0%) and SSC regions (31.9–32.2%). Moreover, the five *Cerostasigma* plastoms were conserved in gene numbers and gene orders, with 82–83 protein-coding genes, 37 tRNA genes and eight rRNA genes (Table 1, Table S1). *C. ulicinum* had one more *rpl2* gene than four other *Cerostasigma* species in IRA (Fig. S1).

Comparisons of border and divergent hotspot identification analysis

The differences between inverted repeats and single-copy (IR/SC) borders among the five *Cerostasigma* plastoms (GenBank accession numbers: OP954207 for *C. griffithii*, OP954208 for *C. minus*, OP954204 for *C. plumbaginoides*, OP921766 for *C. ulicinum*, OP967034 for *C. willmottianum*) were examined by comparative analyses (Table 1). As shown in Fig. 1, the plastome sequences were highly similar within the genus *Cerostasigma*. Except *C. ulicinum*, the IR/SC junctions were conserved in four other species. The LSC/IRB border in *C. ulicinum* plastome was positioned within *rps19* (with 105 bp located at IRB) and *rpl2* genes and the IRA/LSC border was *rpl2/trnH* genes. For the other *Cerostasigma* species, the LSC/IRB and IRA/LSC borders were *rpl2/trnI* genes and *trnI/trnH* genes, respectively. All species had the same IRB/SSC junctions, in which the *ycf1* genes and *ndhF* genes were 124–126 bp and 88–141 bp away from IRB/SSC borders. The *rps15* genes, crossing the SSC/IRA borders, were located at SSC and IRA regions with 272–278 bp and 1–10 bp, respectively.

The mVISTA results showed that the sequences in non-coding regions were more divergent than those in coding regions (Fig. 2). For the non-coding regions, high variation were found in *trnH-GUG-psbA*, *trnK-UUU-rps16*, *rps16-trnQ-UUG*, *atpF-atpH*, *atpI-atpH*, *trnE-UUC-trnT-GGU*, *trnT-UGU-trnL-UAA*, *trnL-UAA-trnF-GAA*, *atpB-rbcL*, *psaI-ycf4*, *petA-psbJ*, *ndhF-rpl32* and *rpl32-trnL*; and the main divergence for the coding regions were *matK*, *rpoC1*, *ycf3*, *rps11*, *rpl36*, *rps3*, *rps19*, *ndhF* and *rps15*. Moreover, we extracted all coding genes, intergenic and intronic loci from the five species, and calculated the nucleotide diversity (Pi) of loci with length longer than 350 bp. In the protein coding regions, the Pi

Table 1 Plastid genome characteristics of the five *Cerostasigma* species

Species	Locality ID	Total length (bp)	LSC (bp)	SSC (bp)	IRs (bp)	Total GC content (%)	LSC GC content (%)	SSC GC content (%)	IRs GC content (%)	No. of protein coding genes	No. of tRNA	No. of rRNA	GenBank Accession numbers
<i>C. griffithii</i>	ZN	164,622	89,828	13,530	30,632	37.6	35.9	32.0	41.2	82	37	8	OP954207*
<i>C. griffithii</i>	JC2	164,348	88,877	13,457	31,007	37.6	35.9	32.1	41.2	82	37	8	OP954206
<i>C. minus</i>	DQ	164,076	89,837	13,497	30,371	37.6	35.9	32.2	41.4	82	37	8	OP967035
<i>C. minus</i>	MK	164,987	89,988	13,477	30,761	37.7	36.0	32.2	41.3	82	37	8	OP967036
<i>C. minus</i>	NML	164,786	89,806	13,454	30,763	37.5	35.9	32.1	41.0	82	37	8	OP954210
<i>C. minus</i>	JC1	164,369	89,826	13,511	30,516	37.5	35.9	32.0	41.2	82	37	8	OP967033
<i>C. minus</i>	LOZ	164,488	89,558	13,520	30,700	37.5	35.8	32.1	41.1	82	37	8	OP954209
<i>C. minus</i>	BR	164,789	89,569	13,518	30,851	37.5	35.9	31.9	41.1	82	37	8	OP954208*
<i>C. minus</i>	MZ	164,182	89,418	13,464	30,650	37.6	35.9	32.1	41.2	82	37	8	OP967032
<i>C. plumbaginoides</i>	XW1	164,744	89,650	13,534	30,780	37.5	35.8	31.9	41.2	82	37	8	OP954204*
<i>C. plumbaginoides</i>	XW2	164,744	89,650	13,534	30,780	37.5	35.8	31.9	41.2	82	37	8	OP954205
<i>C. ulicinum</i>	AR	168,355	89,263	13,516	32,788	37.3	35.6	31.9	40.7	83	37	8	OP921765
<i>C. ulicinum</i>	JZ	167,576	88,756	13,492	32,664	37.3	35.6	32.0	40.8	83	37	8	OP921766*
<i>C. willmottianum</i>	KR	164,164	89,532	13,482	30,575	37.5	35.9	32.1	41.2	82	37	8	OP967034*

Note: The five plastid genome used for compared analyses were indicated by asterisks with GenBank accession numbers

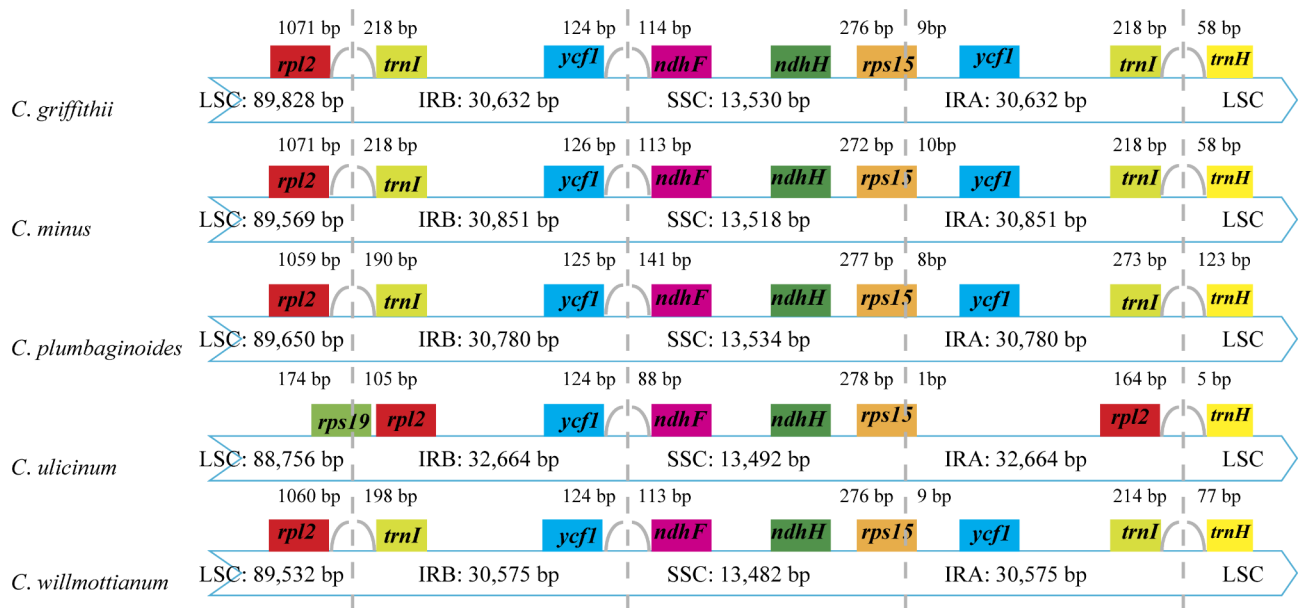


Fig. 1 Comparison of the LSC, IR and SSC borders of five *Ceratostigma* plastomes

values for each locus ranged from 0.00026 to 0.01613 and had an average value of 0.00603 (Table S2). Among these regions, six regions (*matK*, *ycf3*, *rps11*, *rps3*, *rpl22*, and *ndhF*) exhibited remarkably high variation with $P_i > 0.01$ (Fig. 3A, Table S2). The nucleotide diversity in intergenic and intronic regions ranged from 0.0006 to 0.03664, and showed higher average nucleotide diversity (0.0110) than that of coding regions. Four regions including *trnH-GUG-psbA*, *rps16-trnQ-UUG*, *ndhF-rpl32* and *rpl32-trnL* showing relatively high nucleotide diversity values larger than 0.02 were identified (Fig. 3B, Table S2).

Analyses of repeat sequences

A total of 231 simple sequence repeats (SSRs) were detected in the five *Ceratostigma* plastid genomes, in which *C. minus* contained the most SSRs, while *C. willmottianum* contained the least (Table S3). Specifically, five types of SSRs were identified (mononucleotide, dinucleotide, tetranucleotide, pentanucleotide and hexanucleotide), but trinucleotide repeats were not detected in all five species. Among these types, mononucleotide repeats were the most frequent in all five species, with ratios ranged from 0.6600 for *C. ulicinum* to 0.7381 for *C. griffithii*, followed by the ratios of tetranucleotide repeats ranging from 0.1000 for *C. ulicinum* to 0.1273 for *C. minus*, hexanucleotide repeats ranging from 0.0476 for *C. griffithii* to 0.1200 for *C. ulicinum*, dinucleotide repeats accounting for the lower ratios (0.0545–0.0750), and pentanucleotide repeats accounting for the lowest (0.0227–0.0600) (Fig. 4A, Table S3). The majority SSRs were distributed in the intergenic regions (IGS) (71.84%) and less in protein coding regions (15.92%) and the introns (11.84%) (Fig. 4B, Table S3). Among the

mononucleotide repeats, A/T repeats accounted for the most part (96.97%) of all mononucleotide repeat types among the five plastid genomes. In the dinucleotide repeats, the AT/AT and TA/TA repeats were observed more frequently, with 43.75% and 40.63% of all dinucleotide repeats, respectively. In the tetranucleotide repeats, AAGT/TATT repeats were the most abundant type, with 22.73% of all tetranucleotide repeats types in the five plastid genomes (Fig. 4C, Table S3). Additionally, a total of 726 long non-overlapped repeats, including forward, palindromic, reverse, and complementary repeats with range 30–179 bp, were also detected in the five *Ceratostigma* plastid genomes using the online program REPuter. Collectively, repeat numbers and length varied from one species to another. *C. ulicinum* possessed the greatest number of repeats with the number of 176 and *C. griffithii* had the lowest (117). Most abundant were forward repeats ranging from 68 in *C. plumbaginoides* to 102 in *C. ulicinum*; complementary repeats were the least abundant repeats, ranging from zero in *C. ulicinum* to 3 in *C. minus* (Fig. 5A). As for the repeat length, the long repeats with 30–45 bp were found to be the most common, and those with length of 45–60 bp and >75 bp were the second and third abundant types, respectively (Fig. 5B).

Codon usage pattern and adaptive selection analyses

The codon usages of the protein-coding genes in the plastomes from the five *Ceratostigma* species were analyzed. The total sequenced sizes of the protein-coding genes for codon analysis were 86,982–87,465 bp in the five *Ceratostigma* plastomes. The number of encoded codons ranged from 18,734 to 18,802 in plastomes.

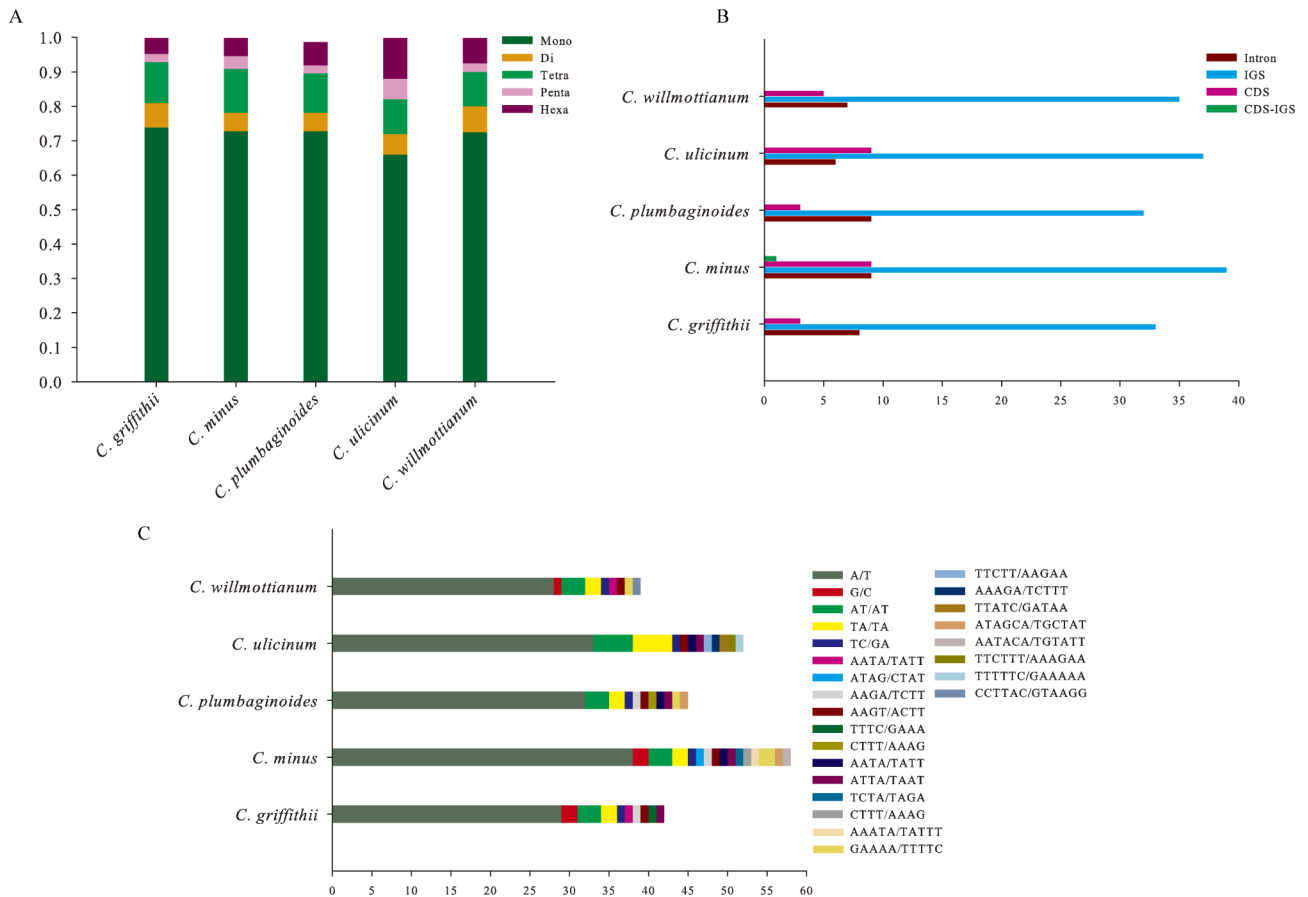


Fig. 4 Analysis of simple sequence repeats (SSRs) in the five *Ceratostigma* plastomes. **(A)** the ratios of different types detected in the *Ceratostigma* plastomes within each species; **(B)** the number of SSRs in the introns, intergenic regions (IGS), protein-coding genes (CDS) and CDS-IGS (partly in CDS and partly in IGS); **(C)** the types and number of each identified SSR in the five *Ceratostigma* plastomes

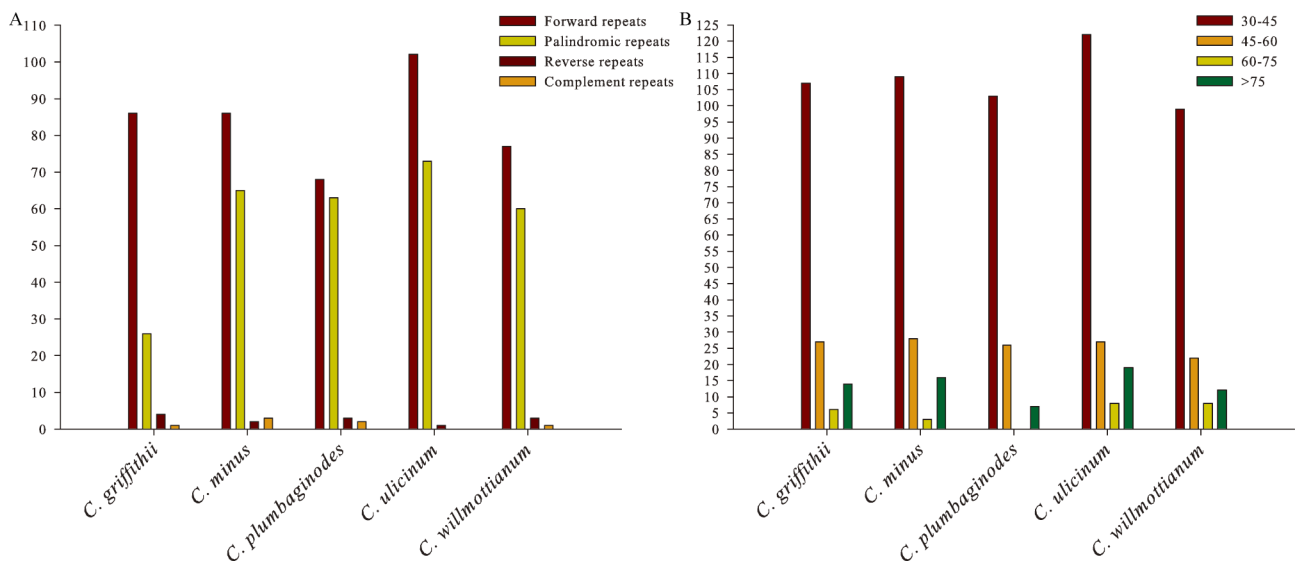


Fig. 5 Long repeat sequences among the five *Ceratostigma* plastomes **(A)** total numbers of four repeat types (forward, palindromic, reverse and complement) detected; **(B)** numbers of long repeat sequences by length

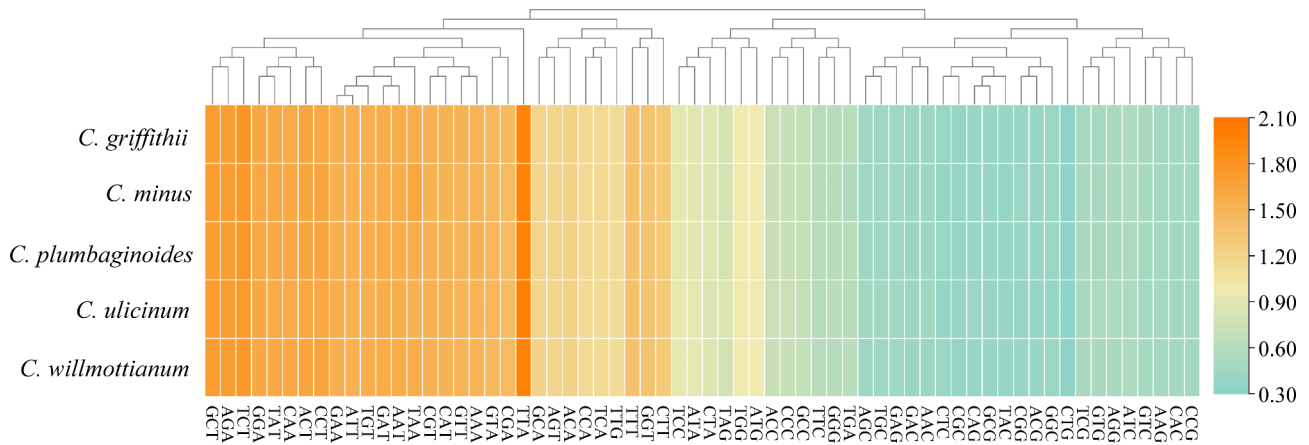


Fig. 6 The codon usages of all protein-coding genes for five *Ceratostigma* plastomes. Orange colors indicate higher values of relative synonymous codon usage (RSCU) and blue color indicate lower RSCU values

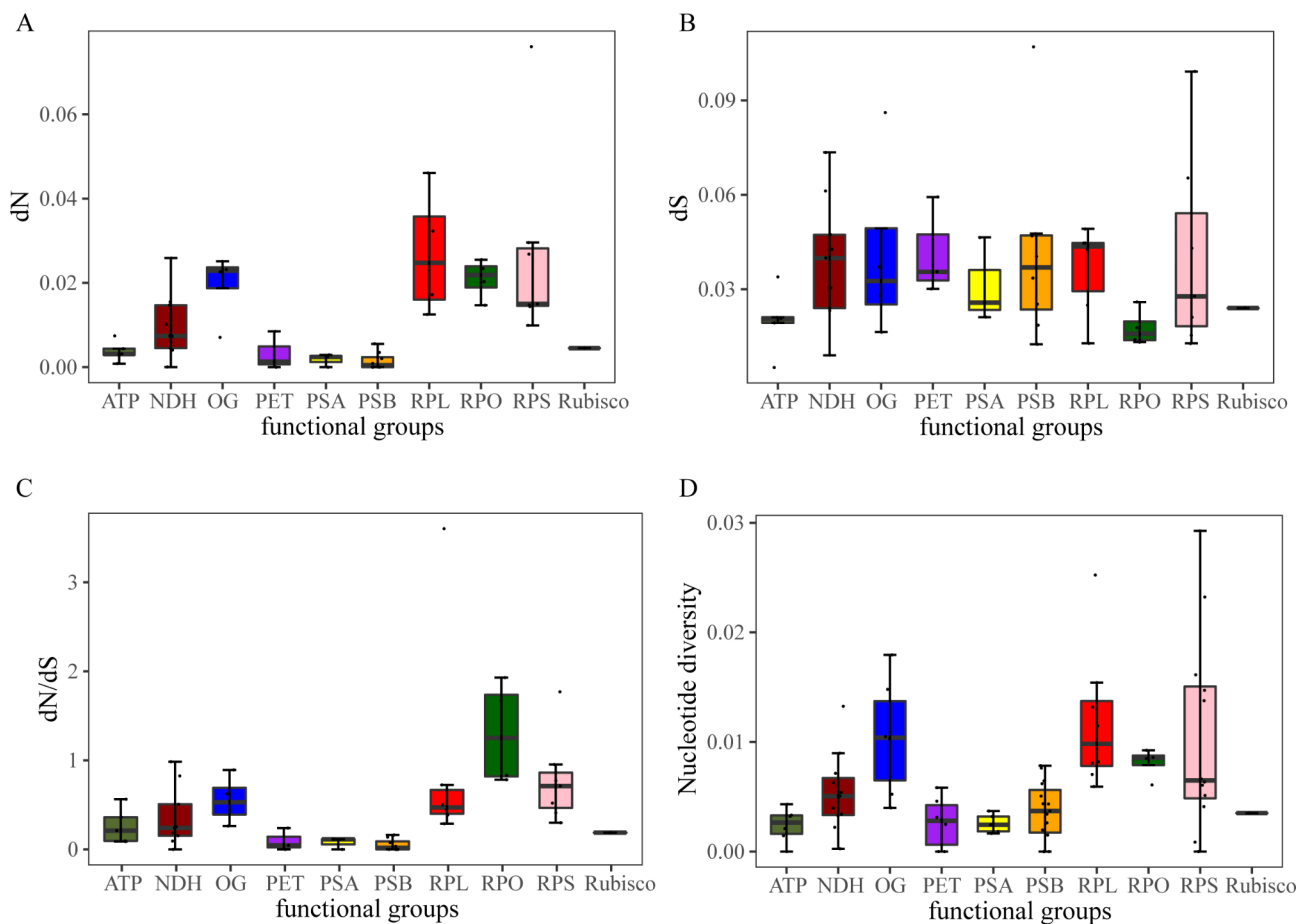


Fig. 7 The results of selective pressure analysis in *Ceratostigma* plastomes. **(A)** estimates of non-synonymous nucleotide substitution (dN); **(B)** estimates of synonymous nucleotide substitution (dS); **(C)** estimates of dN/dS; **(D)** nucleotide diversity of different gene functional groups

the combined datasets of nrDNA and plastomes (Fig. S4) largely resembled the tree topology of plastome data, and four species were recovered as monophyletic with high support, but individuals of *C. minus* still failed to form a unique cluster, which showed a trend of geographical

clustering visible in the ML and BI trees. When compared with each other, the plastome phylogeny was better supported than that of nrDNA, and resolution and node support values were significantly improved by the combined dataset.

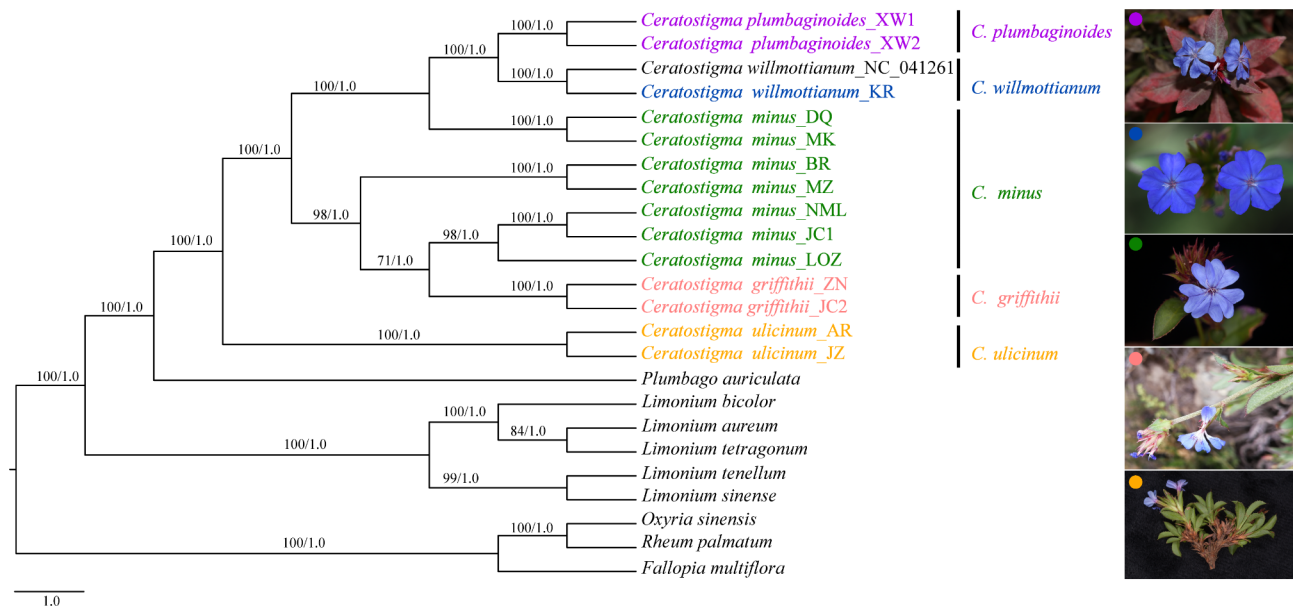


Fig. 8 The phylogenetic trees of all accessions based on chloroplast genome sequences. The number above lines indicates bootstrap values for maximum likelihood (ML) and posterior probabilities for Bayesian inference (BI) analyses of the phylogenetic analysis for each clade. The five species are shown by different colors that are used in the sampling map (Fig. S5). A picture of species is shown on the right, with each pie color corresponding to that of each species in the tree

Discussion

Plastome structure and sequence variation

The plastome structure is generally conserved in most angiosperm and comparative analyses have been used to examine the differences in plastome evolution. In this study, 14 *Ceratostigma* plastomes were generated. We found that the plastomes in *Ceratostigma* are highly conserved, exhibiting little differences in terms of gene number (127–128) and IR/SC borders. Comparatively, genome size among the five species and within different individuals of the same species showed some variations (Table 1), which have been reported in other species, such as *Dipelta floribunda* and *Dipelta yunnanensis* [31] and *Colligonum* [22]. These differences may be attributed to the extraction or expansion in Inverted Repeats regions, which were regarded as a common evolutionary phenomenon in plastome evolution [32]. The detailed comparison of the five IR/SC junctions of the *Ceratostigma* plastomes showed similar characteristics but expansion in IR region with shifts in gene positions of *C. ulicinum*, resulting in its longest plastome length (Fig. 2). The changes in position of IR/SC border were also observed in plastid genomes in other species belonging to Plumbaginaceae, due to the duplication to the part of the *ycf1* gene [33]. Additionally, the GC content of *Ceratostigma* (37.3–37.7%) was comparable to that of *Limonium* (36.7–37.1%) [23] and *Plumbago* (37.1%) [15] published in Plumbaginaceae (Table 1). The IR regions had higher GC content (40.7–41.4%) than those of LSC region (35.6–36.0%) and SSC region (31.9–32.2%). These

results reported here have similarity among other species, most likely due to the high GC content in rRNA genes [34, 35].

The divergent hotspot regions dispersed throughout the plastomes could provide variable information and contribute to further researches exploring angiosperm relationships and relating to population genetics [36, 37]. In accordance with most previous studies on the plastomes of angiosperms, we also found that the nucleotide sequence diversity of the non-coding regions was higher than that of the coding regions (Fig. 3). Unlike previous markers such as *matK*, *rbcL* and *trnL-F* employed in Plumbaginaceae [13, 23], more highly variable regions (*ycf3*, *rps11*, *rps3*, *rpl22*, and *ndhF*, $P_i > 0.01$) were identified in the coding regions; four intergenic regions including *trnH-GUG-psbA*, *rps16-trnQ-UUG*, *ndhF-rpl32* and *rpl32-trnL* had high nucleotide diversity values ($P_i > 0.02$). These regions could serve as potential molecular markers for further phylogenetic and population genetics studies.

Besides genes or intergenic fragments, plastid simple sequence repeats (cpSSRs) markers possess unique and important variations, and are also potential tools to investigate population genetic variation and phylogeographic patterns [38, 39]. A total of 40 (*C. willmottianum*) to 55 (*C. minus*) cpSSRs were found and vastly distributed in the intergenic region (IGS) of five species (Fig. 4). Most of the SSRs types are mononucleotide repeats (A/T), whereas di- or trinucleotide repeats are rare. This result was similar to the repeat characteristics of other reported

gymnosperms and angiosperm plastomes [21, 40–42] and it is suggested that more abundant A/T motifs could keep a more stable framework compared to polyC and polyG [43]. In addition to short sequence repeats, abundant long repeat sequences were also identified, which play an important role in promoting plastome rearrangement and sequence divergence [44–46]. Among the five congeneric species, repeat numbers varied from one species to another (117 for *C. griffithii* and 176 for *C. ulicinum*) (Fig. 5). The type of 30–45 bp repeat accounted for the largest number, and this observation is congruent with the results of unrearranged plastomes [37, 47].

We further analyzed the codon usage and relative synonymous codon usage frequency (RSCU) in the five *Cerotastigma* plastomes. In this investigation, the more frequent amino acid was leucine, followed by cysteine identified as a rarest amino acid. The findings were comparable to those observed in *Limonium* [23]. In the 64 mutant codons including three stop codons, 32 codons possessed A/T at the third nucleotide positions and 30 of which showed RSCU values larger than one, indicating the higher content of A/T used in the codons and especially at the third codon position (Fig. 6). In general, codons with a high AT content are used more often in plastomes [48]. Although several exceptions were also reported such as in Geraniaceae with high GC content of protein-coding genes [49], the same tendencies of favored codon usage pattern followed for many other plastomes, which might be correlated with the high proportion of A/T in plastid genomes or induced by adaptation evolution of the plastid genomes [49, 50].

Plastome evolution

Furthermore, to investigate the possible genomic evolution, we estimated the selective pressures of common protein-coding genes in *Cerotastigma* plastomes based on the useful measuring selection pressure parameter (dN/dS) at the protein level. In most genes, synonymous nucleotide substitution generally occurred more frequently than non-synonymous nucleotide substitution ones, and consequently, most genes evolved under purifying selection or displayed neutral evolution [51]. We found that most of the dN/dS values for the genes were less than one, representing that the genes in *Cerotastigma* plastomes were subjected to purifying selection. When grouping genes into different functional groups, the functional groups RPO and RPS had higher median dN and dN/dS values, suggesting that accelerated substitution rates were detected in genes involved in transcription and translation when compared with those participated in photosynthesis systems (Table S1). These results were similar to what have been found in some land and aquatic plants [52, 53]. In our comparisons, specifically, dN/dS values of two genes *rpoC1* and *rps15* were larger than one

and positive selection sites were identified (Fig. 7). The gene *rpoC1* encodes DNA dependent RNA polymerase involving in transcription and *rps15* belongs to ribosome small subunit gene encoding ribosomal protein, and both play important roles in transcription and translation. Such functional genes with signature of positive selections were also detected in *Rheum* experienced rapid radiations in the QTP [21], and in grasses dominated as savannas belonging to several sub-families such as Panicoideae and Arundinoideae [54–56]. It is proposed that *Cerotastigma* probably have a tropical origin [2], and according to our field investigation, some of five species are distributed parapatrically and possess similar dry habitats in the QTPsl [4, 57]. Therefore, the narrow spectrum of candidate gene functional classes affected may on the one hand, reflect the typically conservative plastid genome across most angiosperms; on the other hand, might mirror adaptation of species in the genus *Cerotastigma* to similar but different environments with gradually increased elevations and longitude [58, 59].

Phylogenetic relationships

The availability of plastid genome sequence has provided many new insights into gymnosperms and angiosperm phylogenetic relationships at the ordinal, familial, tribal and lower levels [16, 18, 20, 60]. In Plumbaginaceae, the plastid genomes remained very limited, and only several species were sequenced. In this study, we obtained 14 plastid genomes of five *Cerotastigma* species distributed in China and multiple individuals from different populations per species were included in our analyses. By incorporating the sequenced plastomes in Plumbaginaceae, we were able to gain some insights into the interspecific relationships of the small genus of *Cerotastigma*. Our result suggested that *Cerotastigma* was monophyletic and closely related to *Plumbago*, as identified in phylogenomic works of Caryophyllales [15] and molecular phylogenetic results of Plumbaginaceae [13]. Within the genus *Cerotastigma*, the available evidence suggested that *C. ulicinum* was monophyletic with high support and formed a clade sister to the other clades composed of four species (Fig. 8). *C. ulicinum* is a cushion shrub, and has the most restricted distribution only in the South Tibet of China. The species has unique morphological characteristics, i.e., rigid and linear to almost needlelike bud scales [1], which could be explicitly distinguished from other species within *Cerotastigma*. Thus, in accordance with the morphological differentiation, our data supported the species was highly genetically differentiated from other species across the sampled distribution.

Additionally, the subshrub *C. willmottianum* with main extant distribution in Hengduan Mountains was retrieved as sister to the herb *C. plumbaginoides*, whose life form is not dominant in arid areas in the QTPsl and distribution

range is more northward. *C. griffithii* and *C. minus* are morphologically more similar and also have parapatric distributions in the QTPs, and this explains their close relationship in the plastomic phylogeny. In particular, currently, *C. minus* was not recognized as a monophyletic group, but had two distinctly main lineages divided by Mekong-Salween Divide (MSD) [61]. These results indicate that groups of species within *Ceratostigma* tend to exhibit a high degree of geographic endemism corresponding to their clade affiliation, in which samples from different populations per species seem to be clustered by geography other than species. Taking the distribution range into consideration, it is perhaps reasonable to pinpoint that geographic isolation is one of the major precursors to promote interspecific/intraspecific genetic divergence of *Ceratostigma*. Mekong-Salween Divide composed by large mountains like Nushan Mountains has been invoked to explain genetic discontinuities or cryptic speciation in a wide range of different species in the Himalaya and Hengduan Mountains regions, such as *Sinopodophyllum hexandrum* [62], *Taxus wallichiana* [63] and *Marmoritis complanatum* from the subnival belt [64]. Yet such geographical structure of *C. minus* may also reflect genealogical processes including hybridization and introgression [65]. Especially when hybridization and subsequent backcrossing occurred, the plastome of one species might be captured by the other [66], the phenomenon of plastid capture could occur frequently with closely related species with sympatric distribution and reproductive compatibility [67–69]. Within *Ceratostigma*, interspecific hybridization has been observed and inferred, for example, morphological intermediates of leaves between *C. minus* and *C. griffithii* are described in Yunnan and adjacent areas in Sichuan province [1], suggesting hybridization or introgression may occur among at least some populations of these species. When considering the parapatric distribution of *C. griffithii* and *C. minus* such as in Lhoka regions in Tibet, the possibility of gene flow between them may not be completely excluded. However, we also noticed that the plastome-based phylogeny of the *Ceratostigma* represents one aspect of the overall evolutionary history of the group. It is essential to compare plastid phylogenies against those from nuclear genome to assess the influence of multiple processes on the phylogenetic relationships.

Unfortunately, the results inferred from the nrDNA dataset could not provide solid resolution for the phylogenetic relationships of *Ceratostigma*, perhaps due to its short length and not enough parsimony-informative characters. Especially, the systematic positions of individuals of *C. minus* and *C. griffithii* varied from those of plastome and combined datasets, and similar results were also found in other phylogenetic studies [70, 71]. This phenomenon could be better explained by different

important determinants including incomplete lineage sorting, hybridization/introgression and other genetic processes [72, 73]. Currently, except *C. willmottianum*, more than one individual of the species were sampled and included in the phylogenetic reconstruction, and it should be noticed that genome skimming data discards a large portion of the nuclear genome, thus may limit the power of discrimination at, or below, the species level [74]. Moreover, gene flow could leave similar traces in the genome to those created by incomplete lineage sorting [75], consequently, based on available data, pinpointing cases due to either phenomenon is difficult. It is therefore important to use more phylogenetic markers from genome for reconstruction of true relationships among the species. Further studies involving population genomic data and analyzed approaches could help to better elucidate their evolutionary histories.

Methods

Plant material sampling and genomic DNA extraction

In total, we included 14 samples representing all five extant species of *Ceratostigma* distributed in China (Fig. S5, Table S6). Among the five species, four had more than one individual sampled except *C. willmottianum*, whose plastome had been reported. Fresh leaves were collected from the field in Yunnan, Tibet and Henan, and then dried in silica-gel immediately. Because the *Ceratostigma* species we collected from field were currently not protected species, no permission was required during the sampling process. The formal identification of *Ceratostigma* species in this study was undertaken by Yajuan Zhao with help provided by Prof. Heng Li (Kunming Institute of Botany, Chinese Academy of Sciences), the author of the study of areas of the genus *Ceratostigma* [2]. Voucher specimens were deposited at the herbarium of Kunming Institute of Botany, Chinese Academy of Sciences under voucher specimens numbers GX008, GX025-026, GX047, ZYJ001, ZYJ006-007, ZYJ013-015, ZYJ034, ZYJ044 and XW001 (Table S6). The genomic DNA was extracted from silica-dried leaves using a modified CTAB method [76].

Plastid genome sequencing, assembly and annotation

In total, 250–500 ng DNA from each sample was used to prepare libraries. Sequencing was conducted on an Illumina HiSeq X Ten platform with a paired-end of 150 bp reads. We checked the sequencing quality of raw reads with FASTQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) [77], which were filtered by removing reads with low sequencing qualities such as duplicate reads and adapter-contaminated reads. The obtained clean reads were then assembled de novo using GetOrganelle toolkit [78], and the main parameters were adjusted according to the assembled results. The plastid

genome sequence of *C. willmottianum* (MK397862) was utilized as reference [15]. Plastid genomes were annotated using the online program GeSeq [79] and adjusted manually in Geneious v.9.0.1 [80] by comparing these with previously published plastomes of Plumbaginaceae, and the physical maps were drawn and illustrated using Organellar Genome Draw (OGDRAW) (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) [81]. The final annotated plastomes were deposited in GenBank with accessions listed in Table 1.

Genome comparison and divergent hotspot identification analysis

The plastomes structure variation of the five species was examined by comparing the positions of SC/IR junctions and their adjacent genes to assess the expansion/contraction of the IR regions. All five complete plastomes were also aligned and compared with mVISTA in shuffle-LAGAN model [82], using *C. minus* as a reference. In order to identify regions of high genetic divergence between *Ceratostigma* species that could be potentially be informative for phylogenetic studies within this genus, the coding genes and intergenic regions (including introns) longer than 350 bp referred to Zhang et al. [83] were extracted from the plastomes using R scripts and the nucleotide diversities (Π) within these regions were calculated in DnaSP v5.0 [84].

Codon usage and repeat sequences analyses

The protein-coding genes were extracted from the five *Ceratostigma* plastomes in Geneious v9.0.1 [80]. The parameter of codon usage and relative synonymous codon usage (RSCU) were estimated in CodonW v1.4.2 program [85]. When RSCU values are larger than 1, these codons are more often used than expected. Simple sequence repeats (SSRs) were identified in these plastome sequences using GMATA v2.3 [86], with minimum repeat numbers of 10 for mononucleotide (mono-) repeats, 5 for dinucleotide (di-) repeats and trinucleotide (tri-) repeats, and 3 for tetranucleotide (tetra-), pentanucleotide (penta-), and hexanucleotide (hexa-) repeats, respectively. Moreover, REPuter program [87] was employed to identify long repeats including forward, palindromic, reverse, and complementary repeats. The parameters were set as follows: Hamming distance of 3, a repeat identity of more than 90% and minimum repeat size of 30 bp. All overlapping repeat sequences within the plastomes were not considered in the statistical analyses.

Selective pressure analyses

To detect the protein-coding genes under selection in the five *Ceratostigma* species, the ratio (ω) of non-synonymous (dN) to synonymous (dS) nucleotide substitution rates were calculated utilizing the CodeML algorithm

implemented in PAML v4.9j [88]. In these analyses, each single-copy CDS was extracted and aligned separately using MUSCLE Alignment in Geneious v9.0.1 [80] and the stop codons were deleted. The codon frequency was determined by the F3 \times 4 model. Positive selection was determined when the value of ω was larger than 1. Then we further compared three pairs of site-specific model (M0 vs. M3, M1 vs. M2 and M7 vs. M8) to analyze the significances, in which the likelihood ratio test (LRT) values under the different model was compared [30] and the P values were calculated using Chi-square [89] to confirm the quality of the sets.

Phylogenetic analysis

Phylogenetic relationships within Plumbaginaceae were reconstructed using the plastomes of species sampled in our comparative analysis and other species with plastomes publicly available in NCBI. We downloaded three plastomes of the sister family Polygonaceae (*Oxyria sinensis*, GenBank accession number: MK397882; *Rheum palmatum*, GenBank accession number: NC_027728; *Fallopia multiflora*, GenBank accession number: MK330002) as outgroups (Table S7) [15]. Multiple alignments of these matrices were made using MAFFT v.7.310 with default settings [90]. To evaluate the phylogenetic effects of character inclusion/exclusion and to minimize systematic error due to poor alignment, the conserved loci were selected by Gblocks v0.91b [91] with three different gap positions treated methods (none, half and all) and other parameters were set as default. In addition, two more datasets were analyzed to explore the phylogenetic relationships within *Ceratostigma*, one using the nuclear ribosomal DNA (nrDNA) data and the other with all plastome and nrDNA data combined. Reads containing nrDNA sequences from the genome skimming dataset of all *Ceratostigma* samples were collected and assembled using the same method as assembling plastomes. Regions composed of 18 S, ITS1 (internal transcribed spacer), 5.8 S, ITS2 and 26 S were well aligned and used for phylogenetic analysis. Three nrDNA sequences (GenBank accession numbers: MZ366771, MZ366779 and MZ366785) of *Opuntia* belonging to Caryophyllales were downloaded from NCBI to be used as outgroups (Table S7). Maximum likelihood (ML) and Bayesian inference (BI) methods were used to infer the phylogenetic relationships. The best substitution model determined by jModeltest v1.5 [92] were selected in BI analyses, and Marko chain Monte Carlo (MCMC) algorithm was run for one million generations, with one tree sampled every 1000 generations implemented in MrBayes v3.2 [93]. GTR+G model with 1000 rapid bootstrap replications for each matrix were set for ML analyses performed with RAxML v8.3 [94].

Conclusions

In conclusion, we determined the complete plastome sequences of five species including 14 samples from different populations in China. The comparative analysis of these plastomes exhibited high similarities in terms of the overall structure, SSR, long repeat sequence and codon usage, but expansion in IR region with shifts in gene positions of *C. ulicinum* was detected. The highly polymorphic regions identified in the current study might be suitable for the phylogenetic analysis and resolving taxonomic discrepancies at the genus level. Moreover, our data resolves the phylogenetic relationships of *Cerotastigma* and establishes monophyly of the genus *Cerotastigma*. However, interspecific delimitation and relationships of four species were well resolved except *C. minus*, which was shown to be non-monophyletic, indicating that lineage genetic divergence perhaps was promoted by geographic barriers in Himalaya and Hengduan Mountains regions, but hybridization or introgression may not be excluded. These findings represent the first important step in elucidating plastome evolution and phylogenetic relationship in this widespread distribution genus *Cerotastigma* in the QTPsl. In the future, additional evidence from genomic data is needed to comprehensively uncover the evolutionary history of *Cerotastigma*.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-023-04323-7>.

Supplementary Material 1
 Supplementary Material 2
 Supplementary Material 3
 Supplementary Material 4
 Supplementary Material 5
 Supplementary Material 6
 Supplementary Material 7
 Supplementary Material 8
 Supplementary Material 9
 Supplementary Material 10
 Supplementary Material 11
 Supplementary Material 12
 Supplementary Material 13

Acknowledgements

We would like to thank Ningning Zhang, Yang Lu, Huihui Xi, Zhaochun Wang, Shuai Long and other members in the research group for field samplings and data analyses. We are grateful to Professor Heng Li from Kunming Institute of Botany, Chinese Academy of Sciences for help identifying the species.

Authors' contributions

X.G. and Y.J.Z. conceived and designed the work. Y.J.Z. performed the experiment, analysed the data and wrote the manuscript. J.L. and G.S.Y.

reviewed the manuscript. All authors approved the final version of the manuscript.

Funding

This work was supported by the Second Tibetan Plateau Scientific Expedition and Research (STEP) Program (2019QZKK0502). The funders were not involved in the design, sample collection, analysis and interpretation of data and in writing the manuscript.

Data Availability

All annotated chloroplast sequences data in this study have been submitted to NCBI (<https://www.ncbi.nlm.nih.gov/>) with accession numbers OP954204-OP954210, OP967032-OP967036 and OP921765-OP921766 shown in Table 1. Other chloroplast genomes for phylogenetic analysis can be obtained from NCBI and their accession numbers are listed in Table S7. All voucher specimens were deposited in the herbarium of Kunming Institute of Botany, Chinese Academy of Sciences.

Declarations

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Collection of all samples completely complies with the Regulations on the Protection of Wild Plants of the People's Republic of China, the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora. Plant samples in this study were not included in the list of national key protected plants and not collected from national reserve. According to the Regulations on the Protection of Wild Plants of the People's Republic of China and the list of national key protected wild plants, no specific permissions were required for collecting these plants. Experimental researches with *Cerotastigma* species comply with Kunming Institute of Botany, Chinese Academy of Sciences guidelines (<http://in.kib.ac.cn/RegulationList.aspx>), preserving the genetic resources of the species used. It does not require ethical approval. All voucher specimens were deposited in the herbarium of Kunming Institute of Botany, Chinese Academy of Sciences. Professor Heng Li from Kunming Institute of Botany, Chinese Academy of Sciences identified the species.

Consent for publication

Not applicable.

Author details

¹CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China

²Key Laboratory of Economic Plants and Biotechnology, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

³Yunnan Key Laboratory for Wild Plant Resources, Kunming, Yunnan 650201, China

⁴Institute of Agriculture and Life Sciences, Kunming University, Kunming 650214, China

Received: 10 January 2023 / Accepted: 29 May 2023

Published online: 07 June 2023

References

- Peng ZX, Rudolf VK. Plumbaginaceae. In: Flora of China. Edited by Wu ZY, Raven PH, vol. 15. Beijing/St. Louis: Science Press/Missouri Botanical Garden Press; 1996: 190–204.
- Li H. The study of areas of the genus *Cerotastigma* Bunge. Acta Bot Yunnanica. 1981;3(1):49–55.
- Mao KS, Wang Y, Liu JQ. Evolutionary origin of species diversity on the Qinghai-Tibet Plateau. J Syst Evol. 2021;59(6):1142–58.
- Jin ZZ. Floristic features of dry-hot and dry-warm valleys in Yunnan and Sichuan. Kunming: Yunnan Science and Technology Press; 2002.

5. Niu GH, Rodriguez DS. Relative salt tolerance of selected herbaceous perennials and groundcovers. *Sci Hortic*. 2006;110(4):352–8.
6. Shi L, Gao S, Lei T, Duan Y, Yang L, Li J, et al. An integrated strategy for polyploidization of *Cerastostigma willmottianum* Stapf based on tissue culture and chemical mutagenesis and the carbon dioxide fixation ability of tetraploids. *Plant Cell Tiss Org*. 2022;149(3):767–82.
7. Hu J, Gao SP, Liu SL, Hong MT, Zhu Y, Wu YC, et al. An aseptic rapid propagation system for obtaining plumbagin of *Cerastostigma willmottianum* Stapf. *Plant Cell Tiss Org*. 2019;137(2):369–77.
8. Yue JM, Xu J, Zhao Y, Sun HD, Lin ZW. Chemical components from *Cerastostigma willmottianum*. *J Nat Prod*. 1997;60(10):1031–3.
9. Gao S, Li W, Hong M, Lei T, Shen P, Li J, et al. The nonreciprocal heterostyly and heterotypic self-incompatibility of *Cerastostigma willmottianum*. *J Plant Res*. 2021;134(3):543–57.
10. Barrett SCH. A most complex marriage arrangement: recent advances on heterostyly and unresolved questions. *New Phytol*. 2019;224(3):1051–67.
11. Charlesworth D, Charlesworth B. A model for the evolution of distyly. *Am Nat*. 1979;114(4):467–98.
12. Caperta AD, Rois AS, Teixeira G, Garcia-Caparrós P, Flowers TJ. Secretory structures in plants: lessons from the Plumbaginaceae on their origin, evolution and roles in stress tolerance. *Plant Cell Environ*. 2020;43(12):2912–31.
13. Koutroumpa K, Theodoridis S, Warren B, Jimenez A, Celep F, Dogan M, et al. An expanded molecular phylogeny of Plumbaginaceae, with emphasis on *Limonium* (sea lavenders): taxonomic implications and biogeographic considerations. *Ecol Evol*. 2018;8(24):12397–424.
14. Ding G, Zhang D, Yu Y, Zhao L, Zhang B. Phylogenetic relationship among related genera of Plumbaginaceae and preliminary genetic diversity of *Limonium sinense* in China. *Gene*. 2012;506(2):400–3.
15. Yao G, Jin JJ, Li HT, Yang JB, Mandalá VS, Croley M, et al. Plastid phylogenomic insights into the evolution of Caryophyllales. *Mol Phylogenet Evol*. 2019;134:74–86.
16. Li HT, Luo Y, Gan L, Ma PF, Gao LM, Yang JB, et al. Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biol*. 2021;19(1):232.
17. Zhao F, Chen YP, Salmaki Y, Drew BT, Wilson TC, Scheen AC et al. An updated tribal classification of Lamiaceae based on plastome phylogenomics. *BMC Biol*. 2021;19(1).
18. Yang Y, Ferguson DK, Liu B, Mao KS, Gao LM, Zhang SZ, et al. Recent advances on phylogenomics of gymnosperms and a new classification. *Plant Divers*. 2022;44(4):340–50.
19. Xiao TW, Ge XJ. Plastome structure, phylogenomics, and divergence times of tribe Cinnamomeae (Lauraceae). *BMC Genomics*. 2022;23(1):642.
20. Mo ZQ, Fu CN, Zhu MS, Milne RI, Yang JB, Cai J, et al. Resolution, conflict and rate shifts: insights from a densely sampled plastome phylogeny for *Rhododendron* (Ericaceae). *Ann Bot*. 2022;130(5):687–701.
21. Zhou T, Zhu HH, Wang J, Xu YC, Xu FS, Wang XM. Complete chloroplast genome sequence determination of *Rheum* species and comparative chloroplast genomics for the members of Rumiceae. *Plant Cell Rep*. 2020;39(6):811–24.
22. Song F, Li T, Burgess KS, Feng Y, Ge XJ. Complete plastome sequencing resolves taxonomic relationships among species of *Calligonum* L. (Polygonaceae) in China. *BMC Plant Biol*. 2020;20(1).
23. Darshetkar AM, Maurya S, Lee C, Bazzaraghaa B, Batdelger G, Janchiv A et al. Plastome analysis unveils inverted repeat (IR) expansion and positive selection in Sea Lavenders (*Limonium*, Plumbaginaceae, Limonioidaeae, Limoniaceae). *PhytoKeys*. 2021(175):89–107.
24. Kubitzki K. Plumbaginaceae. In: Flowering Plants - Dicotyledons: Magnoliid, Hamamelid and Caryophyllid Families. Edited by Kubitzki K, Rohrer JG, Bittrich V. Berlin, Heidelberg: Springer Berlin Heidelberg; 1993:523–530.
25. Hernandez-Ledesma P, Berendsohn WG, Borsch T, Von Mering S, Akhiani H, Arias S, et al. A taxonomic backbone for the global synthesis of species diversity in the angiosperm order Caryophyllales. *Willdenowia*. 2015;45(3):281–383.
26. Kim Y, Xi H, Park J. The complete chloroplast genome of *Limonium tetragonum* (Plumbaginaceae) isolated in Korea. *Korean J Plant Taxon*. 2021;51(3):337–44.
27. Zhang XY, Xu Y, Liu X. Complete plastome sequence of *Limonium aureum*, a medicinal and ornamental species in China. *Mitochondrial DNA Part B-Resour*. 2020;5(1):333–4.
28. Li JF, Xu B, Yang Q, Wang T, Zhu QY, Lin YN, et al. The complete chloroplast genome sequence of *Limonium sinense* (Plumbaginaceae). *Mitochondrial DNA Part B-Resour*. 2020;5(1):556–7.
29. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc Natl Acad Sci U S A*. 2010;107(10):4623–8.
30. Yang ZH, Wong WSW, Nielsen R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol*. 2005;22(4):1107–18.
31. Peng FF, Zhao Z, Xu B, Han J, Yang Q, Lei YJ et al. Characteristics of organellar genomes and nuclear internal transcribed spacers in the tertiary relict genus *Dipelta* and their phylogenomic implications. *Front Genet*. 2020; 11.
32. Kim KJ, Lee HL. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res*. 2004;11(4):247–61.
33. Logacheva MD, Penin AA, Valiejo-Roman CM, Antonov AS. Structure and evolution of junctions between inverted repeat and small single copy regions of chloroplast genome in non-core Caryophyllales. *Mol Biol*. 2009;43(5):757–65.
34. Asaf S, Khan AL, Khan AR, Waqas M, Kang SM, Khan MA et al. Complete chloroplast genome of *Nicotiana otophora* and its comparison with related species. *Front Plant Sci*. 2016;7.
35. Jiang H, Tian J, Yang JX, Dong X, Zhong ZX, Mwachala G et al. Comparative and phylogenetic analyses of six Kenya *Polystachya* (Orchidaceae) species based on the complete chloroplast genome sequences. *BMC Plant Biol*. 2022; 22(1).
36. Menezes APA, Resende-Moreira LC, Buzatti RSO, Nazareno AG, Carlsen M, Lobo FP et al. Chloroplast genomes of *Byrsonima* species (Malpighiaceae): comparative analysis and screening of high divergence sequences. *Sci Rep*. 2018; 8.
37. Liu LX, Wang YW, He PZ, Li P, Lee J, Soltis DE et al. Chloroplast genome analyses and genomic resource development for epilithic sister genera *Oresitrophe* and *Mukdenia* (Saxifragaceae), using genome skimming data. *BMC Genomics*. 2018; 19.
38. Powell W, Morgante M, Andre C, McNicol JW, Machray GC, Doyle JJ, et al. Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Curr Biol*. 1995;5(9):1023–9.
39. Ebert D, Peakall R. Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. *Mol Ecol Resour*. 2009;9(3):673–90.
40. Yi X, Gao L, Wang B, Su YJ, Wang T. The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. *Genome Biol Evol*. 2013;5(4):688–98.
41. Liu J, Lindstrom AJ, Gong X. Towards the plastome evolution and phylogeny of *Cycas* L. (Cycadaceae): molecular-morphology discordance and gene tree space analysis. *BMC Plant Biol*. 2022; 22(1).
42. Cozzolino S, Cafasso D, Pellegrino G, Musacchio A, Widmer A. Molecular evolution of a plastid tandem repeat locus in an orchid lineage. *J Mol Evol*. 2003;57:41–549.
43. Gragg H, Harfe BD, Jinks-Robertson S. Base composition of mononucleotide runs affects DNA polymerase slippage and removal of frameshift intermediates by mismatch repair in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 2002;22(24):8756–62.
44. Ogihara Y, Terachi T, Sasakuma T. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proc Natl Acad Sci U S A*. 1988;85(22):8573–7.
45. McDonald MJ, Wang WC, Huang HD, Leu JY. Clusters of nucleotide substitutions and insertion/deletion mutations are associated with repeat sequences. *PLoS Biol*. 2011;9(6).
46. Timme RE, Kuehl JV, Boore JL, Jansen RK. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. *Am J Bot*. 2007;94(3):302–12.
47. Ren T, Li ZX, Xie DF, Gui LJ, Peng C, Wen J et al. Plastomes of eight *Ligusticum* species: characterization, genome evolution, and phylogenetic relationships. *BMC Plant Biol*. 2020;20(1).
48. Shimada H, Sugiura M. Fine-structural features of the chloroplast genome - comparison of the sequenced chloroplast genomes. *Nucleic Acids Res*. 1991;19(5):983–95.
49. Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Mol Biol Evol*. 2011;28(1):583–600.
50. Ravi V, Khurana JP, Tyagi AK, Khurana P. An update on chloroplast genomes. *Plant Syst Evol*. 2008;271(1–2):101–22.

51. Chaw SM, Wu CS, Sudioanto E. Evolution of gymnosperm plastid genomes. In: *Plastid Genome Evolution*. Edited by Chaw SM, Jansen RK, vol. 85. London: Academic Press Ltd-Elsevier Science Ltd; 2018: 195–222.
52. Guisinger MM, Kuehl JNV, Boore JL, Jansen RK. Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proc Natl Acad Sci U S A*. 2008;105(47):18424–9.
53. Ren Y, Yu MJ, Low WY, Ruhlman TA, Hajrah NH, El Omri A et al. Nucleotide substitution rates of diatom plastid encoded protein genes are positively correlated with genome architecture. *Sci Rep*. 2020; 10(1).
54. Piot A, Hackel J, Christin PA, Besnard G. One-third of the plastid genes evolved under positive selection in PACMAD grasses. *Planta*. 2018;247(1):255–66.
55. Aliscioni S, Bell HL, Besnard G, Christin PA, Columbus JT, Duvall MR, et al. New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *New Phytol*. 2012;193(2):304–12.
56. Cotton JL, Wysocki WP, Clark LG, Kelchner SA, Pires JC, Edger PP et al. Resolving deep relationships of PACMAD grasses: a phylogenomic approach. *BMC Plant Biol*. 2015; 15.
57. Favre A, Packert M, Pauls SU, Jahnic SC, Uhl D, Michalak I, et al. The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of tibetan biotas. *Biol Rev*. 2015;90(1):236–53.
58. La Q, Zhaxi CR, Zhu WD, Xu M, Zhong Y. Plant species-richness and association with environmental factors in the riparian zone of the Yarlung Zangbo River of Tibet, China. *Biodivers Sci*. 2014;22(3):337–47.
59. Sun YB, Fu TT, Jin JQ, Murphy RW, Hillis DM, Zhang YP, et al. Species groups distributed across elevational gradients reveal convergent and continuous genetic adaptation to high elevations. *Proc Natl Acad Sci U S A*. 2018;115(45):E10634–41.
60. Guo C, Luo Y, Gao LM, Yi TS, Li HT, Yang JB, et al. Phylogenomics and the flowering plant tree of life. *J Integr Plant Biol*. 2023;65(2):299–323.
61. Ward FK. The Mekong-Salween divide as a geographical barrier. *Geogr J*. 1921;58(1):49–56.
62. Li Y, Zhai SN, Qiu YX, Guo YP, Ge XJ, Comes HP. Glacial survival east and west of the 'Mekong-Salween divide' in the Himalaya-Hengduan Mountains region as revealed by AFLPs and cpDNA sequence variation in *Sinopodophylum hexandrum* (Berberidaceae). *Mol Phylogenet Evol*. 2011;59(2):412–24.
63. Liu J, Möller M, Provan J, Gao LM, Poudel RC, Li DZ. Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. *New Phytol*. 2013;199(4):1093–108.
64. Luo D, Xu B, Li ZM, Sun H. The 'Ward Line-Mekong-Salween divide' is an important floristic boundary between the eastern Himalaya and Hengduan Mountains: evidence from the phylogeographical structure of subnival herbs *Marmoritis complanatum* (Lamiaceae). *Bot J Linnean Soc*. 2017;185(4):482–96.
65. Funk DJ, Omland KE. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu Rev Ecol Syst*. 2003;34:397–423.
66. Rieseberg L, Soltis D. Phylogenetic consequences of cytoplasmic gene flow in plants. *Am J Bot*. 1991; 5.
67. Yang YY, Qu XJ, Zhang R, Stull GW, Yi TS. Plastid phylogenomic analyses of Fagales reveal signatures of conflict and ancient chloroplast capture. *Mol Phylogenet Evol*. 2021;163:107232.
68. Acosta MC, Premoli AC. Evidence of chloroplast capture in South American *Nothofagus* (subgenus *Nothofagus*, Nothofagaceae). *Mol Phylogenet Evol*. 2010;54(1):235–42.
69. Zhang L, Huang YW, Huang JL, Ya JD, Zhe MQ, Zeng CX, et al. DNA barcoding of *Cymbidium* by genome skimming: call for next-generation nuclear barcodes. *Mol Ecol Resour*. 2023;23(2):424–39.
70. Ogoma CA, Liu J, Stull GW, Wambulwa MC, Oyeibanji O, Milne RI et al. Deep insights into the plastome evolution and phylogenetic relationships of the tribe Urticeae (Family Urticaceae). *Front Plant Sci*. 2022; 13.
71. Turner B, Paun O, Munzinger J, Chase MW, Samuel R. Sequencing of whole plastid genomes and nuclear ribosomal DNA of *Diospyros* species (Ebenaceae) endemic to New Caledonia: many species, little divergence. *Ann Bot*. 2016;117(7):1175–85.
72. Rosenberg NA. Discordance of species trees with their most likely gene trees: a unifying principle. *Mol Biol Evol*. 2013;30(12):2709–13.
73. Naciri Y, Linder HP. Species delimitation and relationships: the dance of the seven veils. *Taxon*. 2015;64(1):3–16.
74. Bohmann K, Mirarab S, Bafna V, Gilbert MTP. Beyond DNA barcoding: the unrealized potential of genome skim data in sample identification. *Mol Ecol*. 2020;29(14):2521–34.
75. Pinho C, Hey J. Divergence with gene flow: models and data. *Annu Rev Ecol Syst*. 2010;41(1):215–30.
76. Doyle J. DNA protocols for plants—CTAB total DNA isolation. In: *Molecular techniques in taxonomy*. Edited by Hewitt GM, Johnston A. Berlin: Springer; 1991: 283–293.
77. Andrews S. FastQC: A quality control tool for high throughput sequence data. In: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. 2019.
78. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 2020;21(1).
79. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45(W1):W6–W11.
80. Ripma LA, Simpson MG, Hasenstab-Lehman K. Geneious! Simplified genome skimming methods for phylogenetic systematic studies: a case study in *Oreocarya* (Boraginaceae). *Appl Plant Sci*. 2014; 2(12).
81. Greiner S, Lehwark P, Bock R. OrganelleGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organelle genomes. *Nucleic Acids Res*. 2019;47(W1):W59–W64.
82. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004;32:W273–9.
83. Zhang YJ, Ma PF, Li DZ. High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). *PLoS One*. 2011; 6(5).
84. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009;25(11):1451–2.
85. Peden JF. Analysis of codon usage. Nottingham: University of Nottingham; 1999.
86. Wang XW, Wang L. GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. *Front Plant Sci*. 2016;7.
87. Kurtz S, Schleiermacher C. REPuter: fast computation of maximal repeats in complete genomes. *Bioinformatics*. 1999;15(5):426–7.
88. Yang ZH. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24(8):1586–91.
89. Xu B, Yang ZH. PAMLX: a graphical user interface for PAML. *Mol Biol Evol*. 2013;30(12):2723–4.
90. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80.
91. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 2000;17(4):540–52.
92. Posada D. jModelTest: phylogenetic model averaging. *Mol Biol Evol*. 2008;25(7):1253–6.
93. Ronquist F, Huelsenbeck JP. MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19(12):1572–4.
94. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–3.

Publisher's Note

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