

RESEARCH

Open Access



Comparison of plastid genomes and ITS of two sister species in *Gentiana* and a discussion on potential threats for the endangered species from hybridization

Jiuyang Mao¹, Yuze Liang¹, Xue Wang¹ and Dequan Zhang^{1,2*}

Abstract

Background *Gentiana rigescens* Franchet is an endangered medicinal herb from the family Gentianaceae with medicinal values. *Gentiana cephalantha* Franchet is a sister species to *G. rigescens* possessing similar morphology and wider distribution. To explore the phylogeny of the two species and reveal potential hybridization, we adopted next-generation sequencing technology to acquire their complete chloroplast genomes from sympatric and allopatric distributions, as along with Sanger sequencing to produce the nrDNA ITS sequences.

Results The plastid genomes were highly similar between *G. rigescens* and *G. cephalantha*. The lengths of the genomes ranged from 146,795 to 147,001 bp in *G. rigescens* and from 146,856 to 147,016 bp in *G. cephalantha*. All genomes consisted of 116 genes, including 78 protein-coding genes, 30 tRNA genes, four rRNA genes and four pseudogenes. The total length of the ITS sequence was 626 bp, including six informative sites. Heterozygotes occurred intensively in individuals from sympatric distribution. Phylogenetic analysis was performed based on chloroplast genomes, coding sequences (CDS), hypervariable sequences (HVR), and nrDNA ITS. Analysis based on all the datasets showed that *G. rigescens* and *G. cephalantha* formed a monophyly. The two species were well separated in phylogenetic trees using ITS, except for potential hybrids, but were mixed based on plastid genomes. This study supports that *G. rigescens* and *G. cephalantha* are closely related, but independent species. However, hybridization was confirmed to occur frequently between *G. rigescens* and *G. cephalantha* in sympatric distribution owing to the lack of stable reproductive barriers. Asymmetric introgression, along with hybridization and backcrossing, may probably lead to genetic swamping and even extinction of *G. rigescens*.

Conclusion *G. rigescens* and *G. cephalantha* are recently diverged species which might not have undergone stable post-zygotic isolation. Though plastid genome shows obvious advantage in exploring phylogenetic relationships of some complicated genera, the intrinsic phylogeny was not revealed because of matrilineal inheritance here; nuclear genomes or regions are hence crucial for uncovering the truth. As an endangered species, *G. rigescens* faces serious threats from both natural hybridization and human activities; therefore, a balance between conservation and utilization of the species is extremely critical in formulating conservation strategies.

Keywords *Gentiana rigescens*, Medicinal plant, Chloroplast genome, Phylogenetic relationship, Genetic introgression

*Correspondence:
Dequan Zhang
zhangdeq2008@126.com

¹ College of Pharmacy, Dali University, Dali 671000, Yunnan, China

² Yunnan Key Laboratory of Screening and Research on Anti-pathogenic Plant Resources from Western Yunnan (Cultivation), Dali 671000, Yunnan, China



© 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

Gentiana is a famous but extremely complex genus in the family Gentianaceae, comprising approximately 362 species worldwide. Alpine and subalpine regions in southwestern China are the centre of the highest concentration of species in *Gentiana* [1]. *Gentiana rigescens* Franchet is a perennial herb with beautiful flowers and is mainly distributed in Yunnan, Sichuan, and Guizhou [1, 2]. According to Chinese Pharmacopoeia (2020), the species is one of the original species of famous *Gentianae Radix et Rhizoma*, the so-called “Jianlongdan” [3]. The root and rhizome of the species are traditionally used to clear damp-heat and quench the fire of liver and gall bladder, and have recently been shown to possess neuroprotective and anti-Alzheimer effects [4, 5]. The medicine contains medicinal components, such as iridoids and flavonoids [6, 7]. Owing to over-harvesting and habitat destruction, wild resources of *G. rigescens* have decreased sharply over the last few decades. In addition, the species is narrowly distributed in altitude gradients, probably due to its rigid requirements for biotope, according to our field investigation and a previous study [8]. As a result, *G. rigescens* has been listed in the third class of National Key Protected Wild Medicinal Materials in China [9–11].

Gentiana cephalantha Franchet is a sister species to *G. rigescens* with similar morphology, but with better adaptation capacity in diverse habitats (Fig. 1). In Yunnan and adjacent regions, *G. cephalantha* is often used as a medicinal substitute for *G. rigescens* because of its rich resources and confusing morphology [12]. According to Flora of China (1995), both *G. rigescens* and *G. cephalantha* belong to the sect. *Kudoa* [2], but since have been adjusted to sect. *Monopodiae* in the latest taxonomic classification [1]. In terms of morphology, *G. cephalantha* has well-developed rosette leaves and violet corolla, whereas *G. rigescens* does not possess well-developed rosette leaves and carries blue corolla (Fig. 1). Although the two species are widely distributed in SW China with an obvious overlap in geographic ranges, they dominate different altitude gradients. According to a previous field investigation, *G. rigescens* narrowly grows in low elevation areas, but *G. cephalantha* occupies a wide range and diverse habitats in middle and high elevation regions (Table 1). They are generally separated by geography, except for some sympatric distributions. Under such conditions, are *G. rigescens* and *G. cephalantha* conspecific or independent ones in phylogeny? Does nature hybridization occur between the two species (Fig. S1), and will the process pose a threat to the endangered species?

Over the last few decades, molecular techniques have been widely adopted to study species evolution and phylogeny [13, 14], but phylogenetic relationships among *Gentiana* species are not well understood because of the

complexity of this genus, insufficient sampling, and low resolution of traditional molecular markers [15–20]. The complete chloroplast genome could provide more useful genetic information than the cpDNA fragments for exploring the phylogeny of complicated genera and discriminating closely related species [21–23]. Dong et al. [24] used plastid genomes to clarify the phylogenetic relationships between 20 complex species in *Lagerstroemia*. Chen et al. [25] also adopted plastid genomes to perform analysis of species discrimination and phylogeny for 21 *Fritillaria* species in China and revealed that the chloroplast genome could efficiently identify closely related species and resolve complicated phylogeny. Ji et al. [26] used complete chloroplast genomes to clarify the phylogenetic relationships among 29 *Paris* species, providing important molecular evidence for clarifying long-standing taxonomic problems in this genus. However, plastid genomes or regions might not really reveal the phylogeny of certain groups due to maternal inheritance; therefore, genetic information from other genomes, such as nrDNA ITS and nuclear genes, could play a vital role here, as well as explore potential hybridization and its threats to endangered species [27, 28].

In the present study, both the plastid genome and ITS were analysed to: (1) compare the basic characteristics of plastid genomes of *G. rigescens* and *G. cephalantha* from sympatric and allopatric distributions; (2) reveal phylogenetic relationships between the two species in *Gentiana*; and (3) explore natural hybridization and genetic introgression between the two species, as well as potential threats to *G. rigescens*. Undoubtedly, this study would be beneficial to understanding the phylogeny of *Gentiana*, as well as the conservation of *G. rigescens* and the pharmacophylogeny of *Gentianae Radix et Rhizoma*.

Results

Characteristics of plastid genomes for *G. rigescens* and *G. cephalantha*

The plastid genomes of *G. rigescens* and *G. cephalantha* were highly similar in their lengths, ranging from 146,795 to 147,001 bp and 146,856 to 147,016 bp, respectively. The intergenic spacer regions (IGS) in *G. cephalantha* were significantly longer than those in *G. rigescens*. Moreover, all chloroplast genomes of the two species had typical circular quadripartite structures, consisting of a large single-copy region (LSC), a small single-copy region (SSC) and a pair of inverted repeat regions (IR) with similar lengths (Table S1, Fig. 2). In total, 116 genes were annotated in these plastid genomes, including 78 protein-coding genes, 30 tRNA genes, four rRNA genes and four pseudogenes (*ψinfA*, *ψycf1*, *ψrps16*, *ψrps19*). Most genes were located in the LSC and SSC regions. Nineteen genes were duplicated

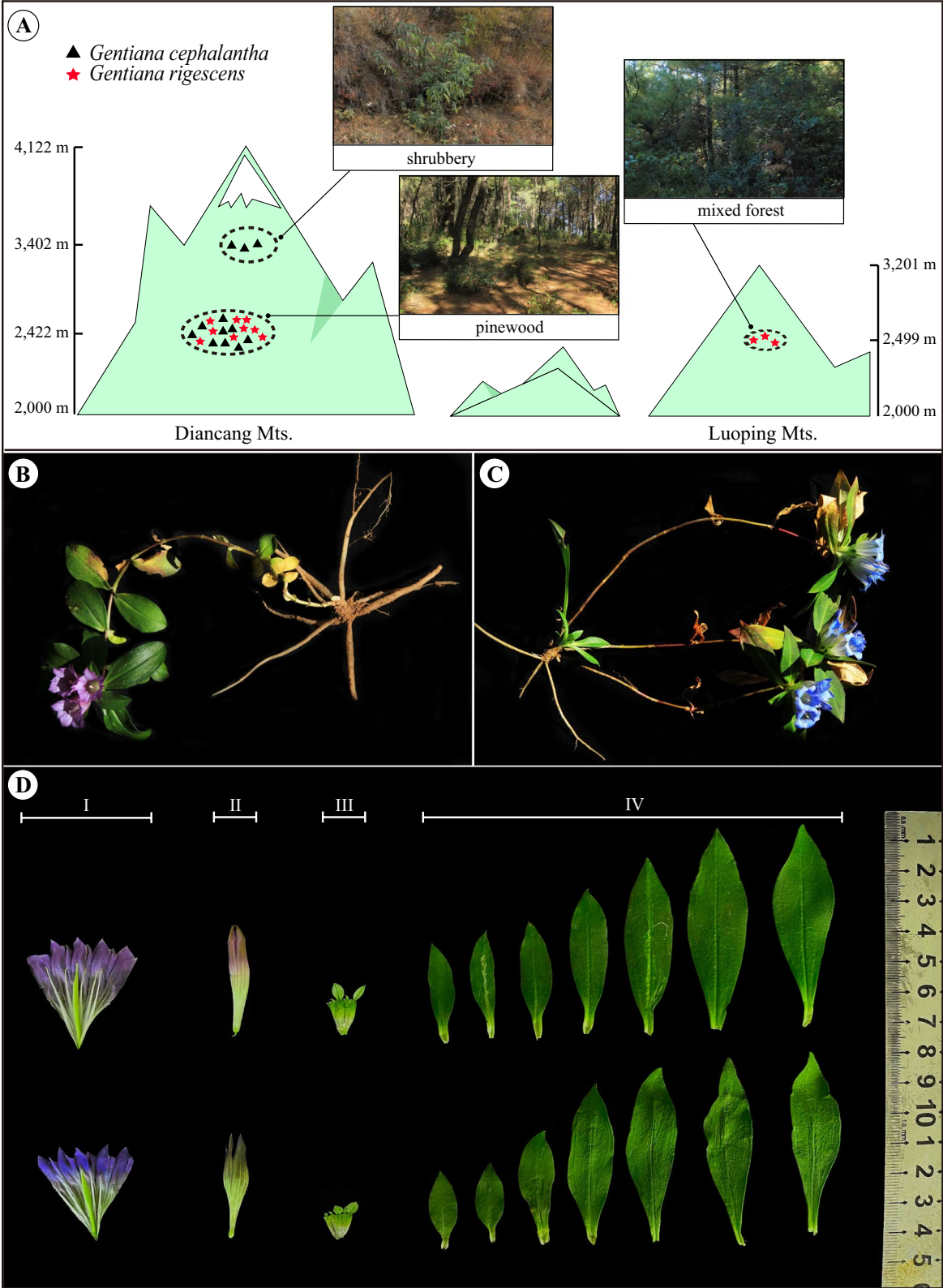


Fig. 1 Morphology of *Gentiana rigescens* and *G. cephalantha*, as well as their habitats. **(A)** Habitats of sympatric and allopatric distributions of *G. rigescens* and *G. cephalantha*; **(B)** whole plant of typical *G. rigescens* without rosette leaves; **(C)** whole plant of typical *G. cephalantha* with rosette leaves; **(D)** comparison of floral and leaf morphology between *G. rigescens* (upper) and *G. cephalantha* (under) in the sympatric distribution

Table 1 Information of sample collection for *Gentiana rigescens* and *G. cephalantha*

Species	Number	Locality	Longitude	Latitude	Altitude	Habitats	Voucher specimen	Accession number of cpDNA	Accession number of nrITS
<i>G.cephalantha</i> (Allopatric)	GcA01	Diancang Mts., Dali, China	100°06'23.16"E	25°40'56.38"N	3,402 m	shrubbery	ZDQ21038	OM961150	ON820192
	GcA02							OM961165	ON820193
	GcA03							OM961153	ON820194
<i>G.cephalantha</i> (Sympatric)	GcS19	Diancang Mts., Dali, China	100°06'50.72"E	25°43'08.25"N	2,422 m	pinewood	ZDQ21048	OM961154	ON820195
	GcS20							OM961164	ON820196
	GcS21							OM961144	ON820197
	GcS22							OM961148	ON820198
	GcS23							OM961160	-
	GcS24							OM961152	ON820200
	GcS25							OM961171	ON820204
<i>G.rigescens</i> (Sympatric)	GrS26	Diancang Mts., Dali, China	100°06'50.72"E	25°43'08.25"N	2,422 m	pinewood	ZDQ21044	OM961173	ON820205
	GrS27							OM961168	ON820206
	GrS28							OM961172	ON820207
	GrS29							OM961174	ON820208
	GrS30							OM961169	ON820209
<i>G.rigescens</i> (Allopatric)	GrA31	Luoping Mts., Dali, China	99°53'56.11"E	26°01'36.89"N	2,499 m	mixed forest	ZDQ21045	OM961167	ON820201
	GrA32							OM961170	ON820202
	GrA33							OM961175	ON820203

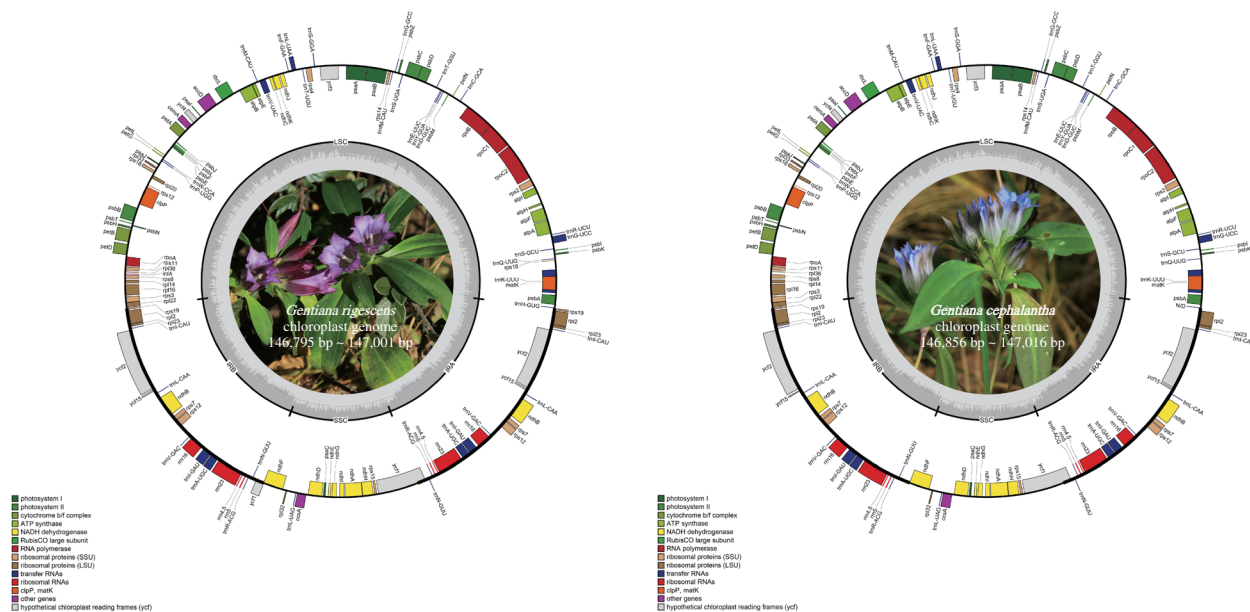


Fig. 2 Plastid genome maps of *Gentiana rigescens* and *G. cephalantha*. Genes drawn inside the circle are transcribed clockwise, while those outside the circle are transcribed counter-clockwise. The inner dark gray circle corresponds to GC content and the inner light gray circle corresponds to the AT content. Different colors represent of distinctive genes within separate functional groups

in the IR regions, including eight protein-coding genes, seven tRNA genes and four rRNA genes. In addition, 11 protein-coding genes and six tRNA genes were found to contain introns (Table 2).

Comparative analysis of re-sequencing revealed that the types and numbers of SSRs were highly similar between *G. rigescens* and *G. cephalantha*. This study identified five types of SSRs (mono-, di-, tri-, tetra-, and

Table 2 Information of the complete chloroplast genomes of *Gentiana rigescens* and *G. cephalantha*

Species	Number	Total (bp)	Large single copy (LSC, bp)	Small single copy (SSC, bp)	Inverted Repeat (IR, bp)	Coding sequence region (CDS, bp)	Intergenic spacer region (IGs, bp)	GC (%)	Total genes	PCG	tRNA	rRNA
<i>G. cephalantha</i>	GcA01	147,016	79,439	17,025	25,276	78,444	39,677	37.8	116	78	30	4
	GcA02	147,016	79,439	17,025	25,276	78,444	39,677	37.8	116	78	30	4
	GcA03	147,002	79,385	17,041	25,288	78,444	39,734	37.8	116	78	30	4
	GcS19	147,002	79,385	17,041	25,288	78,444	39,734	37.8	116	78	30	4
	GcS20	146,760	79,373	16,849	25,269	78,492	39,436	37.8	116	78	30	4
	GcS21	146,856	79,353	17,028	25,237	78,492	39,547	37.8	116	78	30	4
	GcS22	146,935	79,352	17,027	25,278	78,492	39,600	37.8	116	78	30	4
	GcS23	146,856	79,353	17,027	25,238	78,492	39,547	37.8	116	78	30	4
	GcS24	146,866	79,364	17,026	25,238	78,492	39,548	37.8	116	78	30	4
	GrS25	146,901	79,336	17,027	25,269	78,417	38,479	37.8	116	78	30	4
	GrS26	146,990	79,404	17,026	25,280	78,417	38,552	37.8	116	78	30	4
	GrS27	146,759	79,372	16,849	25,269	78,549	38,198	37.8	116	78	30	4
	GrS28	146,759	79,372	16,849	25,269	78,417	38,198	37.8	116	78	30	4
<i>G. rigescens</i>	GrS29	147,001	79,384	17,041	25,288	78,417	38,580	37.8	116	78	30	4
	GrS30	147,001	79,384	17,041	25,288	78,417	38,580	37.8	116	78	30	4
	GrA31	146,901	79,336	17,027	25,269	78,417	38,479	37.8	116	78	30	4
	GrA32	146,901	79,336	17,027	25,269	78,417	38,479	37.8	116	78	30	4
	GrA33	146,901	79,336	17,027	25,269	78,417	38,479	37.8	116	78	30	4

penta-nucleotides). Among them, mono-nucleotides occurred more frequently in the genomes of the two species (63.6%), followed by tetra-nucleotide repeats (19.6%) and tri-nucleotide repeats (11.3%) (Fig. 3A and B). Most of these SSRs were distributed in the LSC region (60.0%), and a few were distributed in the IR regions (5.6%) (Fig. 3C). In addition, the distribution of SSRs in the intergenic spacer regions (51.8%) was more abundant than that in coding genes (38.6%) or in intronic regions (9.6%) (Fig. 3D).

Comparison of the plastid genome structures of *G. rigescens* and *G. cephalantha*

To compare the plastid genome structures of *G. rigescens* and *G. cephalantha*, IR region expansion and contraction were analyzed using IRscope. The results showed that the boundaries of the genomes were highly consistent between the two species. The *rps19* gene had a 145 bp expansion in the IRb region, and the SSC/IRb junction was located in the overlapping region of *ycf1* and *ndhF* genes. In addition, because *ycf1* gene crossed IRa/SSC, most of its sequences were located in the SSC region

(4,404 bp), and the remaining sequences were located in IRa (912 bp), with the same length repetition in IRb (Fig. 4).

For the plastid genomes, nucleotide diversity (Pi) of *G. cephalantha* was higher than that of *G. rigescens*. Most of the highly variable sites were concentrated in the LSC and SSC regions. In contrast, the Pi value was relatively low in the IR regions (Table 3). Furthermore, the Pi values were calculated for the coding genes and intergenic spacer regions of the two species. A total of 28 coding genes and 24 intergenic regions in the plastid genomes of *G. rigescens* showed genetic variability within species, as well as 32 coding genes and 30 intergenic regions in those of *G. cephalantha*. Among these, five coding genes and seven intergenic regions possessed high nucleotide diversity (Pi > 0.001; Fig. 5).

The mVISTA analysis showed that the alignments of plastid genomes were highly similar between *G. rigescens* and *G. cephalantha*. Hypervariable regions (HVR) were mainly distributed in the LSC region, but few were distributed in the IR regions. Moreover, the genetic variation of the non-coding regions was greater than that

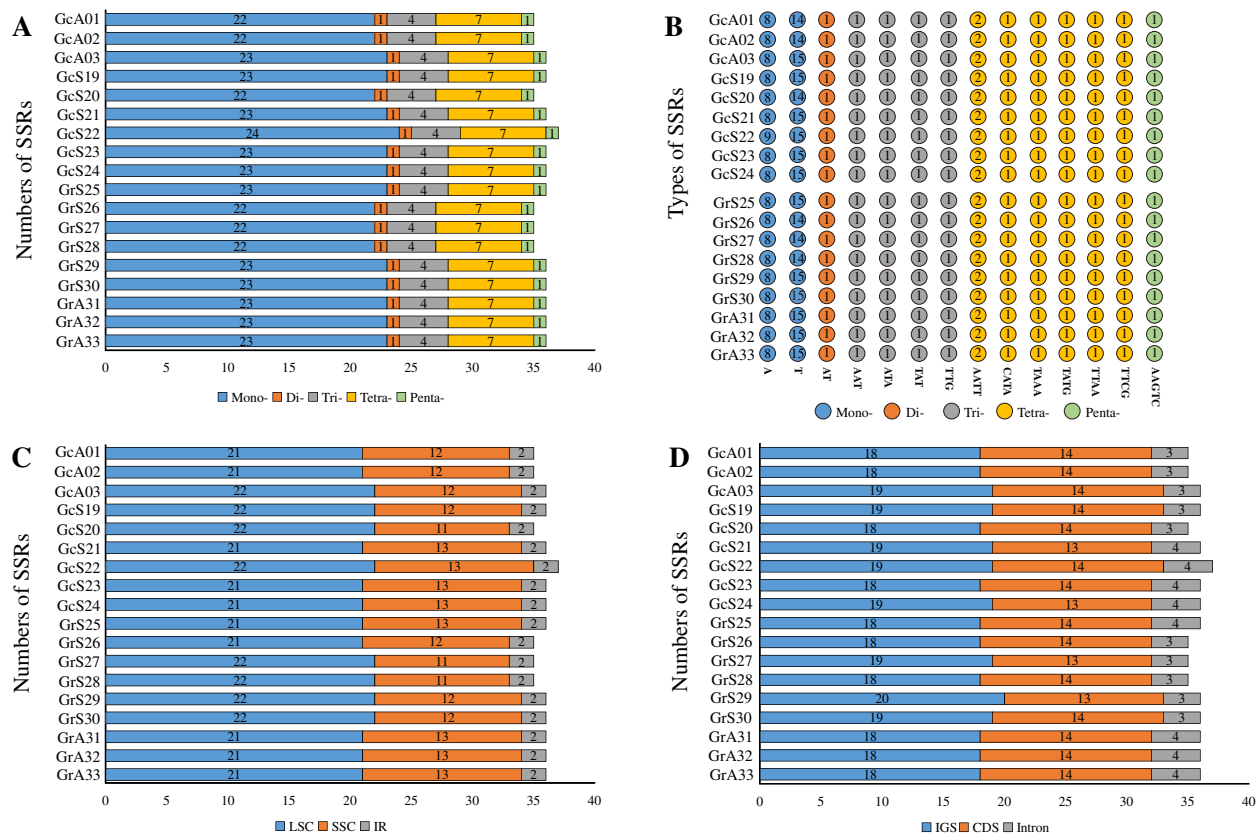


Fig. 3 Analysis of simple repeat sequences (SSRs) in complete chloroplast genomes of *Gentiana rigescens* and *G. cephalantha*. **(A)** Numbers of SSRs of five types; **(B)** type of shared SSRs among the eighteen plastid genomes; **(C)** numbers of SSRs in LSC, SSC, and IR regions (IGS), and intronic regions; **(D)** numbers of SSRs in the coding regions (CDS), intergenic spacer regions

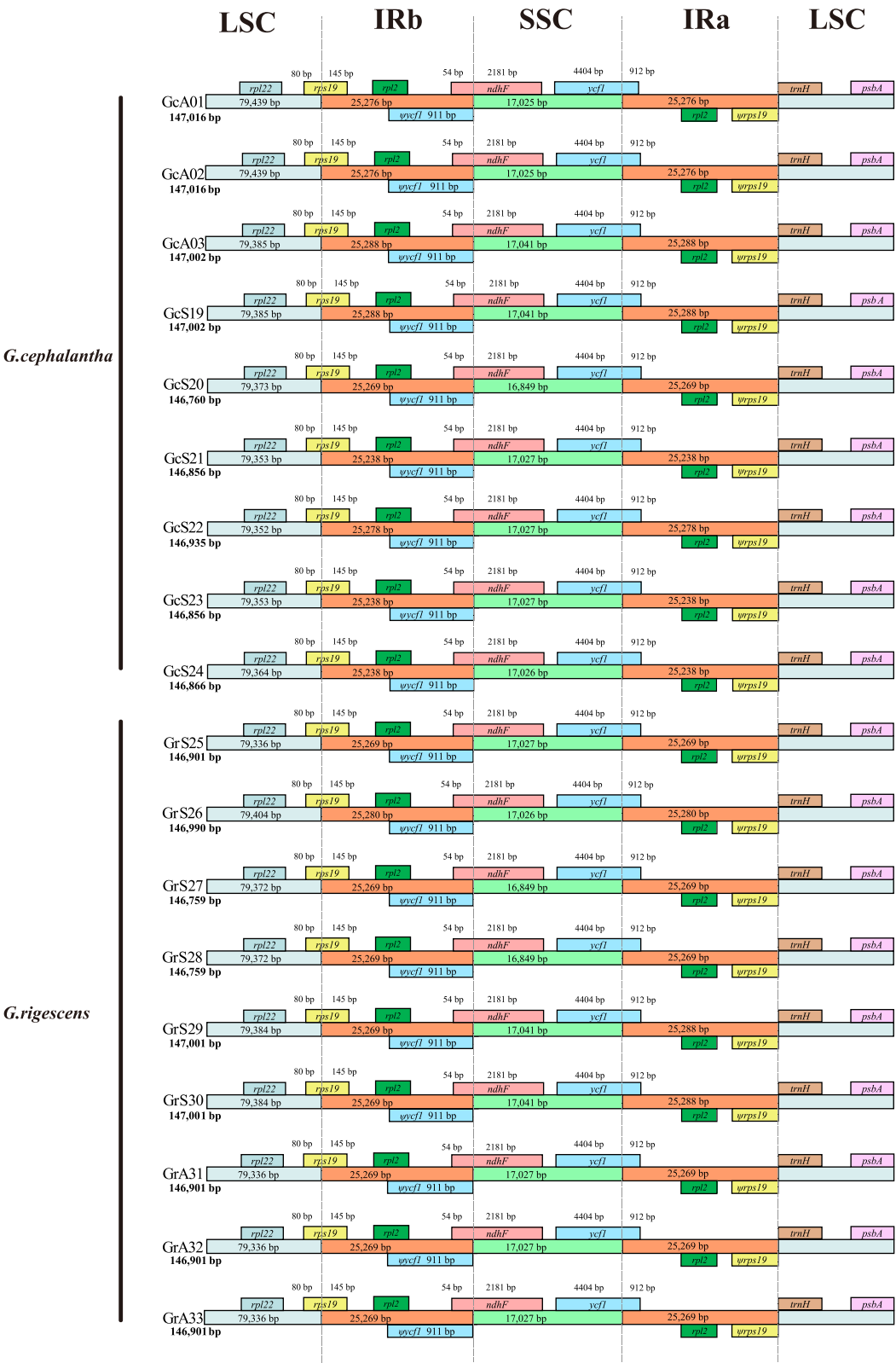


Fig. 4 Comparison of LSC, SSC, and IR border regions among the plastid genomes of *Gentiana rigescens* and *G. cephalantha*. Colored boxes for genes represent the gene position

Table 3 Variable sites in the chloroplast genomes of *Gentiana rigescens* and *G. cephalantha*

	Numbers of sites	Parsimony informative sites	Variable sites	Total numbers of InDel sites	Nucleotide diversity (Pi)
<i>G. rigescens</i>					
Genome	147,109	74	89	441	24×10^{-6}
LSC	79,460	60	71	171	35×10^{-6}
IR	25,299	0	1	30	0.88×10^{-6}
SSC	17,051	14	16	210	39×10^{-6}
Protein coding genes	78,495	24	31	78	16×10^{-6}
<i>G. cephalantha</i>					
Genome	147,161	63	123	581	30×10^{-6}
LSC	79,501	49	94	240	43×10^{-6}
IR	25,304	4	5	66	7×10^{-6}
SSC	17,052	6	21	209	36×10^{-6}
Protein coding genes	78,492	15	40	48	16×10^{-6}

of the coding regions. Therefore, most HVR, including *trnH-GUG-psbA*, *atpH-atpI*, *trnY-GUA-trnT-GGU*, *psbL-psbF*, *psbB-psbT*, *rpl32-trnL-UAG*, and *rps7-ndhB*, were located in the non-coding regions, in addition to *psbT*, *ndhB*, *rpoC2*, and *ycf1* present in the coding regions (Fig. 6).

Analysis of heterozygotes in ITS sequences

In this study, 18 individuals from *G. rigescens* and *G. cephalantha* were used to acquire ITS sequences by the Sanger sequencing. Subsequently, DNA sequences of good quality from 17 individuals were used for final analysis, except that of GcS23, because of confused overlapping peaks. The total length of the alignment of the ITS sequences was 626 bp, including six informative sites (Table 4). Heterozygotes occurred intensively in individuals from the sympatric distribution, especially in those belonging to *G. rigescens* without rosette leaves. Furthermore, GcS21, GcS24, GrS27, and GrS33 possessed complete heterozygosity across all the variable sites.

Phylogenetic analysis of the two species

To explore the phylogenetic relationship between *G. rigescens* and *G. cephalantha*, maximum likelihood (ML) and Bayesian inference (BI) were mainly performed based on chloroplast genomes, coding sequences (CDS), and ITS sequences; and then phylogenetic analysis using representative HVR was also carried out. The results showed that topological structures between the ML and BI trees were highly identical for plastid datasets and showed little inconformity for the ITS samples (Fig. 7), but they were obviously inconsistent for the HVR sequences (Fig. S2). However, there were significant differences among the trees based on plastid genomes, CDS, and ITS sequences. For all the three datasets, both *G. rigescens* and *G.*

cephalantha were clustered into a monophyletic branch; however, the phylogeny between the chloroplast genomes and ITS sequences of the two species were completely different (Fig. 7). Individuals of *G. rigescens* and *G. cephalantha* were intermingled in the trees of plastid genomes, as well as those of HVR sequences, but could be weakly distinguished by the CDS dataset. On the contrary, the phylogenetic trees based on ITS sequences showed that the two species were well discriminated, except for the potential hybrids, namely GcS21, GrS27, and GrS33 in BI, as well as GcS21 and GcS24 in the ML tree. Moreover, the ML and BI analysis using ITS sequences covering nine sections of *Gentiana* (14 sect.) supported the independence of sect. *Monopodiae* separated from sect. *Kudoa*.

Discussion

Comparison on the plastid genomes between *G. rigescens* and *G. cephalantha*

In this study, the structures of the plastid genomes between *G. rigescens* and *G. cephalantha* were highly similar. The total lengths of the genomes were from 146,759 bp to 147,001 bp in *G. rigescens* and from 146,856 bp to 147,016 bp in *G. cephalantha*, which were shorter than those of most of the other species in *Gentiana*, as well as other genera in Gentianaceae [29, 30]. Four common pseudogenes were found in the plastid genomes of *G. rigescens* and *G. cephalantha*. Among them, the existence of *ψycf1* and *ψrps19* could be attributed to boundary effects, whereas *ψinfA* and *ψrps16* are probably caused by gene transfer and loss during evolution, which are also common in species of Gentianaceae and other family [29, 31–35]. In addition, deletion of *ndh* gene has been frequently reported in *Gentiana*, but we could not detect it in either of the two species [29].

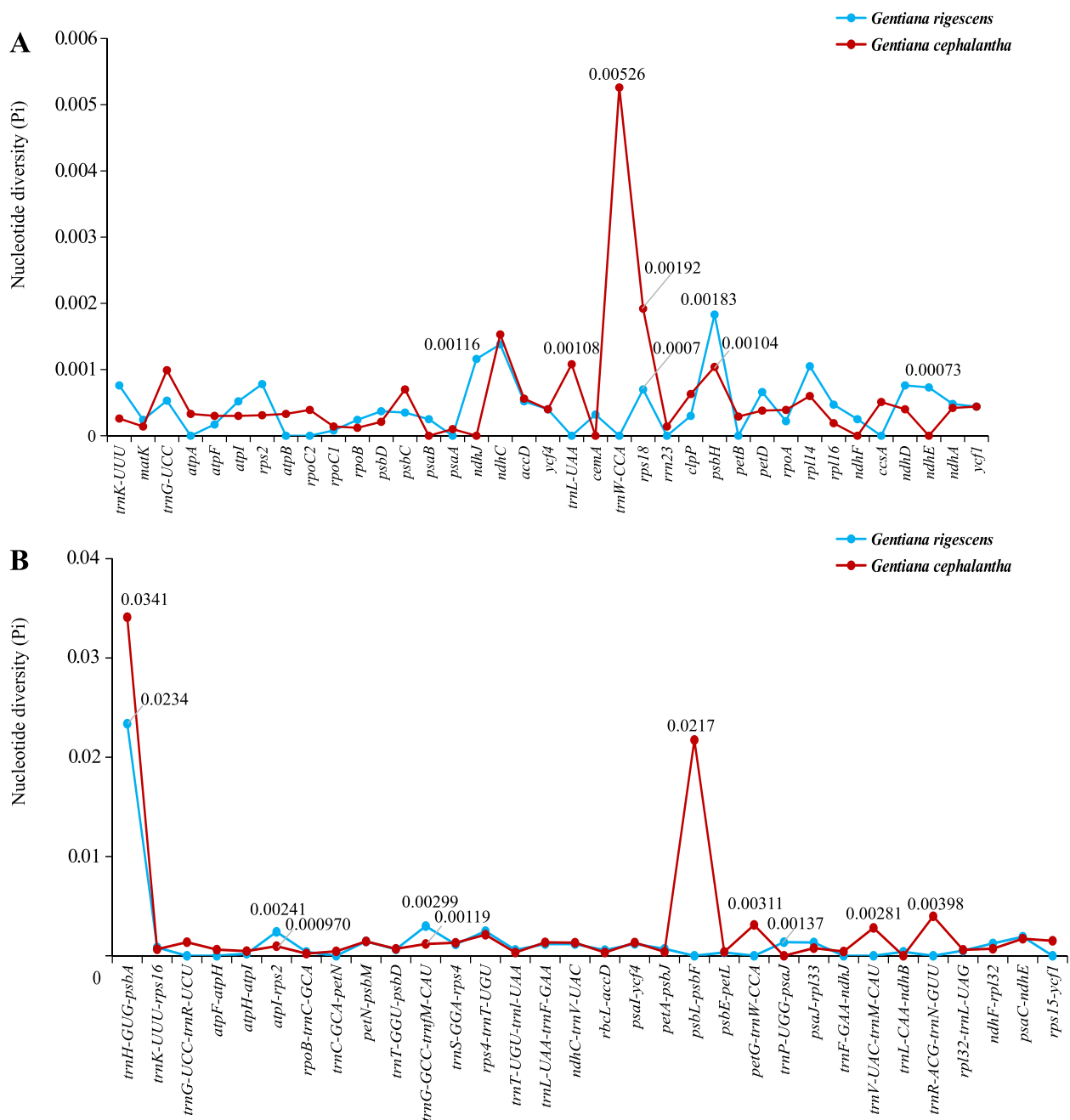


Fig. 5 Comparative analysis of nucleotide diversity (Pi) values among the plastid genomes of *Gentiana rigescens* and *G. cephalantha*. **(A)** Nucleotide diversity (Pi) values of coding genes; **(B)** nucleotide diversity (Pi) values of intergenic regions

Simple sequence repeats (SSRs) are useful tools for evaluating population genetic diversity and structure of species and are thus widely adopted in studies on the conservation of endangered species and evolution of complicated groups [36, 37]. The number of SSRs was 35 – 37 in the two species, and mono-nucleotide repeats were the most abundant, especially A/T repeat units. These results were consistent with those of our previous

study [30]. In addition, most of the SSRs were distributed in the intergenic spacer region, followed by the coding genes (*rpoC1*, *psaB*, *atpB*, *ndhF*, *ycf1*), but the least in the intronic region, which was also similar to those reported for other *Gentiana* species [35].

Hypervariable regions (HVR) not only resolve phylogeny and identify species at the species level but also provide critical information for exploring species

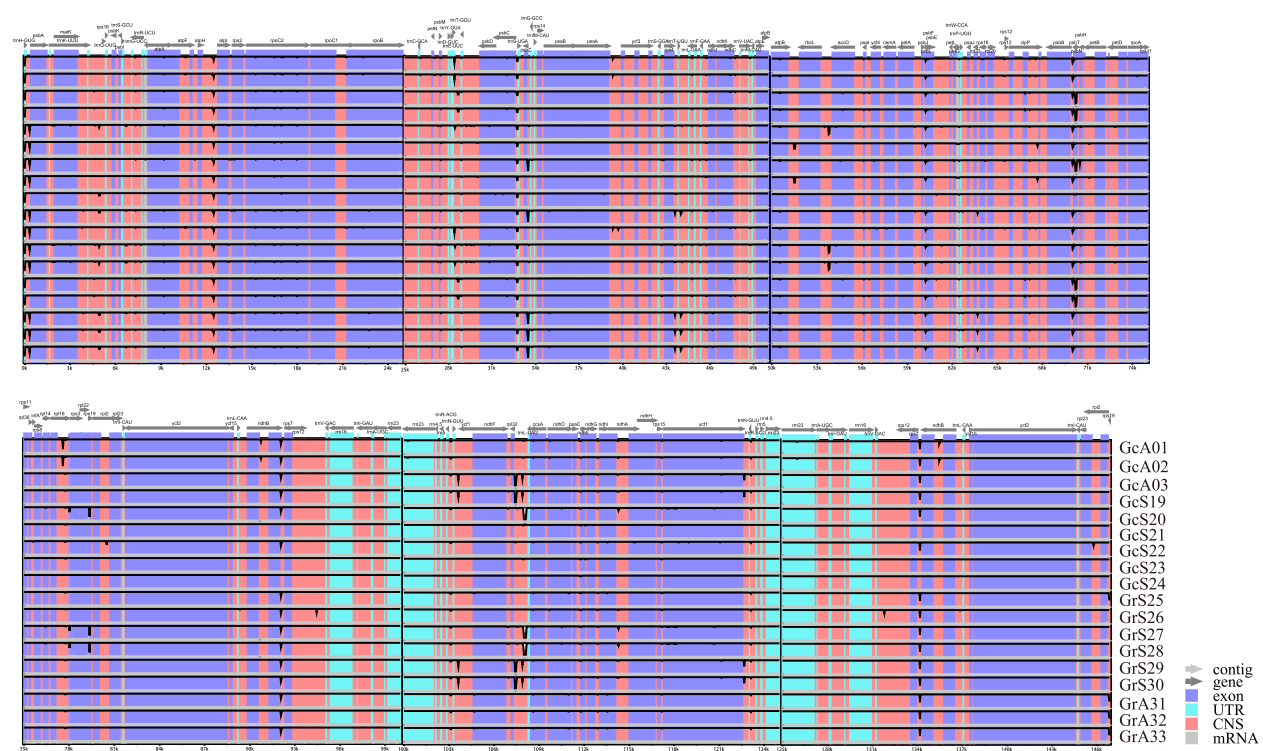


Fig. 6 Visualization alignment of the plastid genomes of *Gentiana rigescens* and *G. cephalantha*

Table 4 Variable sites of ITS in *Gentiana cephalantha* and *G. rigescens*

Species	Numbers							Accession number
		2	4	4	4	4		
		5	0	1	1	2	4	
		0	1	7	8	7	9	
<i>G.cephalantha</i> (Allopatric)	GcA01	A	C	T	T	A	A	ON820192
	GcA02	A	C	T	T	A	A	ON820193
	GcA03	A	C	T	T	A	A	ON820194
<i>G.cephalantha</i> (Sympatric)	GcS19	A	C	T	T	A	A	ON820195
	GcS20	A	C	T	T	A	A	ON820196
	GcS21	R	M	Y	K	W	R	ON820197
	GcS22	A	C	T	T	A	A	ON820198
<i>G.rigescens</i> (Sympatric)	GcS24	R	M	Y	K	W	R	ON820200
	GrS25	G	A	C	G	T	G	ON820204
	GrS26	R	M	Y	K	T	R	ON820205
	GrS27	R	M	Y	K	W	R	ON820206
	GrS28	G	A	C	G	T	G	ON820207
<i>G.rigescens</i> (Allopatric)	GrS29	G	M	Y	K	T	R	ON820208
	GrS30	G	M	Y	K	T	R	ON820209
	GrA31	G	A	C	G	T	G	ON820201
	GrA32	G	A	C	G	T	G	ON820202
	GrA33	R	M	Y	K	W	R	ON820203

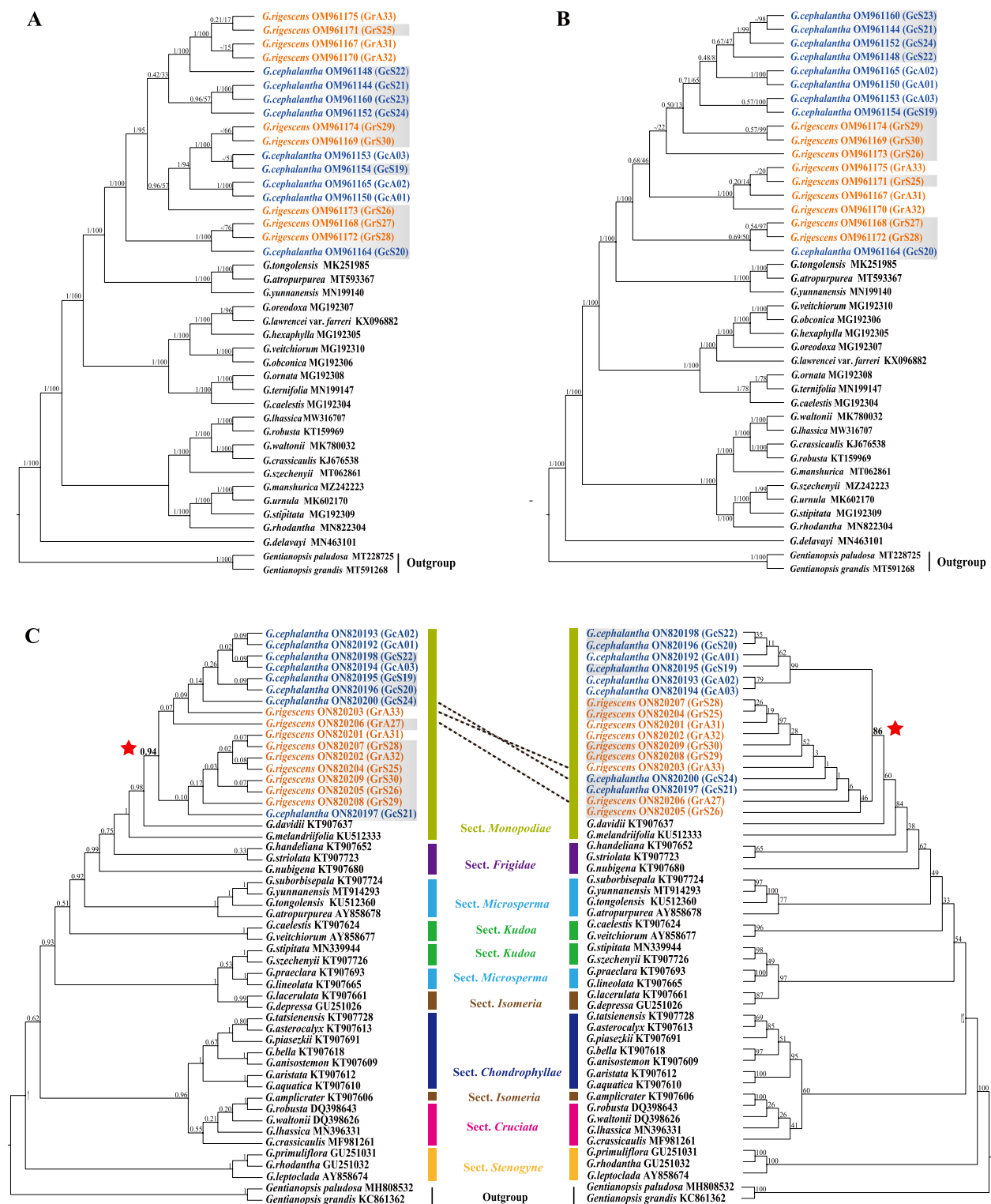


Fig. 7 Phylogenetic relationship of *Gentiana rigescens* and *G. cephalantha*. (A) complete chloroplast genomes; (B) CDS regions; (C) BI (left) and ML (right) phylogeny based on the ITS sequences. Number above nodes are support values with BI posterior probabilities (PP) values on the left and ML bootstrap (BS) values on the right. Orange fonts represent *G. rigescens*, dark blue fonts represent *G. cephalantha*, and grey shades indicate the individuals located in the sympatric distribution

differentiation and genetic structure at the population level [38]. The results of mVISTA and the sliding window have been used for screening the HVR [39, 40]. Herein, we detected eight HVR from non-coding (*trnH-GUG-psbA*, *atpH-atpI*, *petG-trnW-CCA*, *rpl32-trnL-UAG*, *rps7-ndhB*) and coding regions (*ndhB*, *rpoC1*, *psbH*), which have been widely adopted for studies on phylogenetic analysis and DNA barcodes in angiosperms [25, 41–45]. In contrast, *matK* and *rbcL* as core DNA barcodes showed no genetic variation in *G. rigescens* and *G. cephalantha*, further supporting the results of our previous study on DNA barcodes in *Gentiana* [19]. It should be noted that the *trnH-psbA*, as an efficient barcode for identifying species, showed the highest Pi value in the sliding window. A possible reason is that *Gentiana* species are prone to base inversion in this region, leading to overestimation of genetic variation within the species [46]. Therefore, this region was deemed unsuitable to identify *Gentiana* species, as it has been reported before [19, 20]. Expansion and contraction of the IR regions are important factors leading to changes in plastid genome size and play an important role in the stability and evolution of the genome structure [47–49]. The results showed that the genetic compositions of the four junctions in the chloroplast genomes were highly identical between the two species, similar to the findings from studies on other species belonging to the Gentianaceae [29, 31]. The expansion and contraction of *rps19*, *ndhF*, and *ycf1* located at the boundaries of plastid genomes were consistent between *G. rigescens* and *G. cephalantha*, probably due to conservatism during plastid evolution of *Gentiana* [30].

Phylogenetic analysis and species definition of *G. rigescens* and *G. cephalantha*

Currently, plastid genomes have been widely used to reveal the phylogeny of complicated genera or closely related species and developed as field known as phylogenomics [24, 26, 50]. *Gentiana* is an extremely complicated genus in Gentianaceae with approximately 362 species worldwide, including a series of confusing species such as *G. rigescens* and *G. cephalantha* [2]. Although the two species can be discriminated by the basal rosettes of *G. cephalantha* and its blue corolla, they are still easily confused because of their similar but variable morphology, especially in sympatric distributions [51]. In our study, the two species were clustered into a monophyletic group using plastid genomes, CDS, and ITS datasets, which showed that the two species were closely related in terms of phylogeny. According to the ML and BI analysis, complete chloroplast genomes, coding sequences, and HVR regions could

not correctly discriminate individuals from *G. rigescens* and *G. cephalantha*; on the contrary, ITS could gather the two species into distinct independent clades after removal of potential hybrids (Fig. 7). First, in the phylogenetic tree based on the plastid genomes, all individuals were heavily mixed which was also supported by comparative analysis of the chloroplast genomes of the two *Gentiana* species. Comparatively, the tree constructed using CDS showed better resolution, in which most individuals of *G. cephalantha* (except GcS20) formed a monophyletic group. The present result also supports that from the previous report in which protein-coding sequences were adopted to reveal phylogenetic relationships of species in *Gentiana* sect. *Kudoa* and evaluate divergence times [45]. On the contrary, *G. rigescens* and *G. cephalantha* were clustered into two independent clades based on ITS sequences, regardless of the possible hybrids (Fig. 7). Therefore, both *G. cephalantha* and *G. rigescens* are closely related but independent species, according to their morphological and molecular evidences. The present results further verified the value of ITS as shown in previous phylogenetic analyses of *Gentiana* species [52]. Moreover, molecular evidence from the current study also supported the adjustment of the two species into sect. *Monopodiae* from sect. *Kudoa* [1, 2, 16].

It is well known that the chloroplast genome could provide much richer genetic information and a better solution than DNA regions for revealing the phylogeny of complicated genera and discriminating closely related species [53, 54]. In our previous study, the plastid genome was confirmed to be a DNA super barcode that could efficiently discriminate most species in *Fritillaria* in China, but the results from ITS and other universal DNA barcodes analyses were disappointing [25, 53]. Zhang et al. [55] adopted plastid phylogenomics to reveal the deep phylogenetic relationships and diversification history of Rosaceae. Li et al. [56] also reported that phylogenomic analyses of *Fagopyrum* supported the division of the cymosum and urophyllum groups, and resolved the systematic position of subclades within the urophyllum group. However, it should be noted that plastid genomes could enhance species discrimination and reveal phylogeny, but they are still not powerful enough to resolve all species in complicated genera and recently diverged lineages, such as *Paris*, *Berberis*, and *Rhododendron* [26, 57, 58]. This research provides a case study in which only the plastid genome resulted in incorrect assessment of phylogenetic relationships and species discrimination; therefore, nuclear genes should be adopted alongside plastid genomes to reveal intrinsic phylogeny and trace species boundaries.

Species evolution in *Gentiana* and potential hybridization between *G. rigescens* and *G. cephalantha*

According to previous studies, both *G. rigescens* and *G. cephalantha* are widely distributed in Yunnan, Sichuan, Guizhou and adjacent regions with obvious geographical overlap [1, 2]. Field investigations revealed that the two species are generally distributed in allopatric regions. *G. rigescens* narrowly grows at lower altitudes, but *G. cephalantha* can occur in a much wider range at higher altitude gradients (Fig. 1). It is well known that the Qinghai-Tibet Plateau (QTP) and its adjacent mountains are key regions for evolution of alpine plants, and *Gentiana* was revealed to originate from QTP in the Eocene and then spread to other distributions from the late Miocene onwards, a phenomenon named as the "out of Tibet" hypothesis [59]. According to molecular clock dating using pollen and seed fossils, divergence time of *Gentiana* was deduced to be 29 million years, and the divergence of *G. rigescens* and *G. cephalantha* happened about 0.51 Ma and supposedly is the youngest node of differentiation. Obviously, *G. cephalantha* is more adaptable to the cooling climate and diverse habitats, so it colonizes a much broader territory than *G. rigescens* in altitude gradients (Fig. 1). The two species generally grow in allopatric areas at altitudes, except for a few sympatric distributions that are generally located in the border zones between *G. rigescens* and *G. cephalantha*. In the present study, potential hybridization was observed in sympatric distributions in the Diancang Mountains, which showed a visible intermediate corolla color compared to typical *G. rigescens* and *G. cephalantha* (Fig. S1); meanwhile, possible hybrids were preliminarily identified based on the ITS sequences (Table 4). Natural hybrids between *Gentiana straminea* and *G. siphonantha* have been confirmed based on molecular evidence [60]. Considering the overlapping of flowering, geographic isolation might be the main factor for the two species, but possible physiological factors should be explored using pollination biology and other evidences [61].

Threats to *G. rigescens* from hybridization and the balance between conservation and utilization of the endangered species

Natural hybridization occurs widely among species and evolutionary lineages in plants and plays critical role in species evolution [62, 63]. Hybridization probably leads to phenotypic novelty and results in the formation of new species [64]. Meanwhile, the same process may cause the breakdown of species integrity and ultimately drive rare species to extinction through genetic swamping, where the rare form is replaced by hybrids or demographic swamping [63, 65]. *G. rigescens*, a rare species, is threatened by hybridization with its common

congeneric species. According to the ITS sequences of the 17 individuals of the two species (Table 4 and Fig. 7), hybridization was massively detected in the sympatric distribution and sporadically in the territory of *G. rigescens*, but seemingly not in the allopatric distribution of *G. cephalantha*. Therefore, asymmetric introgression had a significantly more serious impact on *G. rigescens* than *G. cephalantha*. Hybridization, as well as subsequent backcrossing between hybrids and the common species, constantly dilutes the genetic loci of the rare one [65–67], which finally results in genetic erosion of *G. rigescens*. As two sister species, no steady reproductive barrier exists between *G. rigescens* and *G. cephalantha*, in addition to geographic isolation along altitude gradients. Habitat change along with human activities provides more possibilities for hybridization between the newly divergent species and affords new niches to hybrids [63]. As a result, the rare species may be replaced by hybrids, even faced with extinction due to genetic swamping or demographic swamping from hybridization [63, 66–68].

As an endangered species possessing medicinal value, balance between efficient conservation and rational utilization of *G. rigescens* is an urgent problem to be resolved. First, strict conservation of the species should be adopted for basic measurements, so as to avoid serious harvesting and destruction. Then, exploring new original species of Gentianae Radix et Rhizoma by pharmacophylogeny could be an efficient method to satisfy medicinal demands, such as *G. cephalantha* [69]. Although both species are used as traditional Chinese medicine "Jianlongdan" in Yunnan and adjacent regions, further evaluation of the medicinal quality and efficacy of medicines between the two species is necessary. Moreover, developing artificial cultivation for *G. rigescens* would be beneficial for balancing conservation and utilization. It should be noted that screening pure individuals without genetic pollution is critically important to ensure the quality of the medicine in consideration of hybridization and genetic introgression. Finally, protecting habitat diversity could provide different niches for *G. rigescens* and *G. cephalantha*, and geographic isolation might be an efficient reproductive barrier to maintaining the independence of the species [66, 68].

Conclusion

In this study, the next-generation sequencing and Sanger sequencing were used to investigate plastid genomes and ITS sequences of *G. rigescens* and *G. cephalantha* from sympatric and allopatric distributions. Comparative analysis of the plastid genomes showed that the two *Gentiana* species possess highly similar genome structures. *G. rigescens* and *G. cephalantha* were clustered into a monophyletic group during phylogenetic analyses based on the

plastid genome, CDS, and ITS regions, supporting their close relationship. Moreover, the two species were mixed in the trees constructed using plastid genomes, CDS, and HVR, but they could form obvious monophyly in phylogenetic analysis based on ITS, excluding the potential hybrids. The present study supports the current treatment of *G. rigescens* and *G. cephalantha* as independent but closely related species in *Gentiana* based on the present morphological and molecular evidences. Natural hybridization occurs intensively between the two species, especially in sympatric distribution. Asymmetric introgression has a significantly more serious impact on *G. rigescens* than *G. cephalantha*, and this process probably leads to constant dilution of genetic loci and even extinction of the rare species. Conserving wild resources, exploring alternative species, developing artificial cultivation, and maintaining habitat diversity are efficient measurements for ensuring a balance between conservation and utilization. In summary, the complete chloroplast genome possesses rich genetic information; thus, it can be used to explore the phylogenetic relationships of closely related species or complicated genera. However, phylogenetic trees might not reveal real phylogeny when we only use plastid genomes or genes because of matrilineal inheritance, and datasets from nuclear genomes or regions are crucial for exploring the truth.

Methods

Materials collection

Dali (Tali) in China is the origin of the type specimens of *G. rigescens* and *G. cephalantha*. In the present study, fresh and clean leaves of adult individuals were collected from the Diancang and Luoping Mountains in Dali, China (Fig. 1 and Table 1). The leaves were directly dried using allochroic silica gel during fieldwork and were used as molecular materials. Herein, we sampled six accessions of each species from their sympatric distribution in the Diancang Mountains (2,422 m). Considering the morphological similarity between the two species, rosette leaves and blue corolla were regarded as distinctive features for *G. cephalantha* that was discriminated from *G. rigescens* (Fig. 1). Three individuals of *G. cephalantha* were also sampled from an allopatric distribution of alpine shrubbery in the alpine region of the same mountain (3,402m), and three individuals representing *G. rigescens* were collected from one allopatric distribution in the Luoping Mountains (Fig. 1). Eight individuals representing the two species were used for DNA sequencing. To avoid sampling individuals from the same female parent, geographic distances among individuals were above 30 m. Meanwhile, 2-3 mature individuals with flowers from each site were excavated and used as the voucher specimens (Table 1). The collection of molecular

materials and specimens was approved by the Forestry and Grassland Administration in Dali Prefecture, China (Grant no. 2021-137). All specimens were identified by Professor Dequan Zhang according to Ho's classification system (1995) and Flora of China (2001) [1, 2] and then deposited at the Herbarium of Medicinal Plants and Crude Drugs of the College of Pharmacy, Dali University.

Molecular experiments

Total genomic DNA was extracted from the dried leaves using a modified CTAB method [70]. DNA quality and concentration were detected using 1.2% agarose gel electrophoresis and spectrophotometer (Bio-Rad, Hercules, CA, USA). The DNA was then sheared to yield approximately 500 bp long fragments for library construction. The library was sequenced on an Illumina Hiseq 2500 platform according to the standard protocol of manufacturer's instructions. About 2-4 Gb raw paired-end reads (2×150 bp) were obtained for each individual of the two species. Next-generation sequencing were performed by Novogene Bioinformatics Technology Co. Ltd., Beijing, China.

The amplification reaction system of ITS sequence was: 2×Taq PCR MasterMix (10 µl), two ITS primers (0.3 µl), template DNA (1 µl) and ddH₂O (8.4 µl). The PCR amplification procedure was as follows: pre-denaturation at 94°C for 2 min; 94°C for 40 s, 55°C for 45 s, and 72°C for 55 s, after 35 cycles; 72°C extension for 7 min primer sequence: ITS4: 5'-TCC TCC GCT TAT TGA TAT GC-3'; ITS5: 5'-GGA AGT AAA AGT CGT AAC AAG G-3'. The entire ITS sequence was sequenced by the Kunming Institute of Botany, Chinese Academy of Sciences.

Assembly, annotation, and submission of plastid genomes and ITS sequence

Raw data were filtered using Trimmomatic v.0.32 with default settings [71], and paired-end reads in the clean data were assembled into contigs using GetOrganelle.py [72]. After assembly, the *de novo* assembly graphs were visualized and edited using Bandage, and then a whole or nearly whole circular plastid genome was generated [73]. Using *G. rigescens* (MT062862) downloaded from the National Center for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov/>) as the reference sequence, MAFFT was performed in the Geneious v.11.1.4 [74, 75], and annotation, modification and manual correction were performed. Circular genome visualization was generated using the online program OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>).

Original ITS DNA sequences were assembled using the SeqMan program (DNASTAR, Lasergene) [76] and then aligned in MEGA v.7.0.26 [77]. The boundaries of

ITS1, 5.8S, and ITS2 were determined using the reference sequence of *G. cephalantha* (KT907627) from NCBI. All annotated plastid genomes and ITS sequences were confirmed, edited using Sequin software, and submitted to the GenBank database (Table 1).

Comparative analysis of the plastid genomes

Using the website IRscope (<https://irscope.shinyapps.io/irapp/>), the IR boundaries of 18 plastid genomes of *G. rigescens* and *G. cephalantha* were analyzed [78], followed by manual editing. The lengths of large single copy (LSC), small single copy (SSC), inverted repeats (IRs), protein-coding genes, and intergenic spacer regions for each of genomes were counted by Geneious v.11.1.4. The online program mVISTA (<https://genome.lbl.gov/vista/mvista/instructions.shtml>) was used for sequence alignment and variation analysis of plastid genomes in Shuffle-LAGAN mode [79, 80], with the annotation of *G. cephalantha* (MN199135) as a reference. In addition, parsimony informative sites and variable sites were analyzed, and a sliding window was used to analyze the nucleotide diversity (Pi) of complete chloroplast genomes and protein-coding genes in *G. rigescens* and *G. cephalantha* by DnaSP v.6.11 (with a window length of 600 bp and a step size of 200 bp) [81]. MISA was adopted to evaluate SSRs in plastid genomes, among which the SSRs parameters of mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide motifs were set as 10, 5, 4, 3, 3, and 3 [82].

Phylogenetic analysis for *G. rigescens* and *G. cephalantha*

A total of 18 new plastid genomes from *G. rigescens* and *G. cephalantha*, and 23 published plastid genomes for other *Gentiana* species from NCBI were used for phylogenetic analysis (Table S2). In addition, phylogenetic trees were constructed based on the 17 new ITS sequences (excluding GcS23) and 32 reported ITS sequences from other *Gentiana* species (Table S3). Two *Gentianopsis* species were selected as outgroups for constructing phylogenetic trees based on the plastid genomes, CDS, and ITS datasets; moreover, phylogenetic analysis using HVR was additionally performed for the two species. The most appropriated model of sequence substitution for plastid genomes (GTR + G + I), CDS (GTR + G + I), HVR (GTR or GTR+G), and ITS (GTR + G + I) were screened by MEGA v.7.0.26 [78]. Phylogenetic analysis was performed using Maximum likelihood (ML) and Bayesian inference (BI). ML analysis was performed using the RAxML v.8.2.10. The local bootstrap (BS) probability of each branch was calculated with 1,000 repetitions [83]. BI analysis was performed using MrBayes v.3.2.6 [84]. The Markov Chain

Monte Carlo (MCMC) algorithm was calculated for 1,000,000 generations with a sampling of trees every 1,000 generations. The first 25% of the generations were discarded as burn-in, and posterior probability (PP) values were determined from the remaining trees to evaluate the support rate of each branch. The state was considered to have been reached when the average standard deviation of the split frequency was < 0.01. Finally, all the methods were performed in accordance with relevant guidelines and regulations.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-023-04088-z>.

Additional file 1: Figure S1. Phenotype of potential hybrids between *Gentiana rigescens* and *G. cephalantha*.

Additional file 2: Figure S2. Phylogenetic relationships of *Gentiana rigescens* and *G. cephalantha* based on each HVR regions.

Additional file 3: Table S1. Gene contents of the plastid genomes of *Gentiana rigescens* and *G. cephalantha*.

Additional file 4: Table S2. The complete chloroplast genomes of *Gentiana* species downloaded from NCBI.

Additional file 5: Table S3. The nrITS sequences of *Gentiana* species downloaded from NCBI.

Acknowledgements

We thanks Miss Haisu Hu and Qian Zhang in Dali University for their assistance in field work for sample collection and data analysis of plastid genomes.

Authors' contributions

JM performed molecular experiments, analyzed the data, and drafted the manuscript. YL and XW participated in field work and data analysis. DZ conceived this study, collected molecular materials and specimens, and revised the manuscript finally. All authors have read and approved the manuscript.

Funding

This research was supported by National Natural Science Foundation of China (32060091, 31660081), Reserve Talents Project for Young and Middle-Aged Academic and Technical Leaders of Yunnan Province (202105AC160063).

Availability of data and materials

All sequences (plastid genomes and ITS sequences) used in this study have been submitted to the National Center for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov/>) with accession numbers (OM961144, OM961148, OM961150, OM9652-OM961154, OM961160, OM961164, OM961165, and OM961167-OM961175 for plastid genomes; ON820192-ON820203 for ITS sequences) (Table 1). All the sequences will be available after publication of this manuscript.

Declarations

Ethics approval and consent to participate

The collection of molecular materials and specimens had been approved by the Forestry and Grassland Administration in Dali Prefecture, Yunnan, China (Grant no. 2021-137).

Consent for publication

Not applicable

Competing interests

The authors declare no competing interests.

Received: 1 August 2022 Accepted: 25 January 2023
Published online: 20 February 2023

References

- Ho TN, Liu SW. A worldwide monograph of *Gentiana*. Beijing: Science Press; 2001.
- Ho TN, Pringle JS. *Gentiana* L. In: Wu ZY, Raven PH, editors. Flora of China. Beijing: Science Press; 1995. p. 15–98.
- Chinese Pharmacopoeia Commission. Chinese Pharmacopoeia (Part I). Beijing: Chinese Medical Science and Technology Press; 2020. p. 99.
- Cheng LH, Osada H, Xing TY, Yoshida M, Xiang L, Qi JH. The insulin receptor: a potential target of amarogentin isolated from *Gentiana rigescens* Franch that induces neurogenesis in PC12. *Biomedicines*. 2021;9(5):581.
- Disasa D, Cheng LH, Manzoor M, Liu Q, Wang Y, Xiang L, et al. Amarogentin from *Gentiana rigescens* Franch exhibits antiaging and neuroprotective effects through antioxidative stress. *Oxid Med Cell Longev*. 2020;2020:3184019.
- Xu LL, Liu C, Han ZZ, Han H, Yang L, Wang ZT. Microbial biotransformation of iridoid glycosides from *Gentiana rigescens* by *Penicillium brasilianum*. *Chem Biodivers*. 2020;17(12):e2000676.
- Xu LL, Ling XF, Zhao SJ, Wang RF, Wang ZT. Distribution and diversity of endophytic fungi in *Gentiana rigescens* and cytotoxic activities. *Chin Herb Med*. 2020;12(3):297–302.
- Shen T, Yu H, Wang YZ. Geographical distribution and bioclimatic characteristics of the wild *Gentiana rigescens* resources. *Chin J Appl Ecol*. 2019;30(7):2291–300.
- Zhou YH, Xu ZL. Discussion about range of national key-protected wild medicinal materials. *Chin Tradit Herb Drugs*. 2016;47(7):1061–73.
- Tang RP, Su HL. Introduction and reproductive technique of endangered medicinal plant *Gentiana rigescens*. *Agric Sci Technol*. 2014;15(8):1326–7+334.
- Zhang J, Zhang ZX, Wang ZX, Zuo YM, Cai CT. Environmental impact on the variability in quality of *Gentiana rigescens*, a medicinal plant in southwest China. *Glob Ecol Conserv*. 2020;24:e01374.
- Han D, Zhao ZL, Liu WH, Li YH, Li HF. Primary discussion for quality evaluation of medicinal plants of *Gentiana cephalantha* in Bai Nationality. *J Chin Med Mater*. 2016;39(11):2549–53.
- Bouetard A, Lefeuvre P, Gigant R, Bory S, Pignal M, Besse P, et al. Evidence of transoceanic dispersion of the genus *Vanilla* based on plastid DNA phylogenetic analysis. *Mol Phylogenet Evol*. 2010;55(2):621–30.
- Hollingsworth PM. Refining the DNA barcode for land plants. *Proc Natl Acad Sci U S A*. 2011;108(49):19451.
- Favre A, Yuan YM, Küpfer P, Alvarez N. Phylogeny of subtribe Gentianinae (Gentianaceae): Biogeographic inferences despite limitations in temporal calibration points. *Taxon*. 2010;59(6):1701–11.
- Favre A, Pringle JS, Heckenbauer J, Kozuharova E, Gao QB, Lemmon EM, et al. Phylogenetic relationships and sectional delineation within *Gentiana* (Gentianaceae). *Taxon*. 2020;69(6):1221–38.
- Mishiba K, Yamane K, Nakatsuka T, Nakano Y, Yamamura S, Abe J, et al. Genetic relationships in the genus *Gentiana* based on chloroplast DNA sequence data and nuclear DNA content. *Breed Sci*. 2009;59(2):119–27.
- Shi DL, Wang MH, Chen SY, Xu L, Kang YG. DNA barcoding identification of *Gentiana* plants and herbs based on ITS2 sequences. *J Chin Med Mater*. 2018;41(1):79–83.
- Liu J, Yan HF, Ge XJ. The use of DNA barcoding on recently diverged species in the genus *Gentiana* (Gentianaceae) in China. *PLoS One*. 2016;11(4):e0153008.
- Whitlock BA, Hale AM, Groff PA. Intraspecific inversions pose a challenge for the trnH-psbA plant DNA barcode. *PLoS One*. 2010;5(7):e11533.
- Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol*. 2016;17(1):134.
- Zeb U, Dong WL, Zhang TT, Wang RN, Shahzad K, Ma XF, et al. Comparative plastid genomics of *Pinus* species: insights into sequence variations and phylogenetic relationships. *J Syst Evol*. 2020;58(2):118–32.
- Shen ZF, Lu TQ, Zhang ZR, Cai CT, Yang JB, Tian B. Authentication of traditional Chinese medicinal herb “Gusuibu” by DNA-based molecular methods. *Ind Crops Prod*. 2019;141:111756.
- Dong WP, Xu C, Liu YL, Shi JP, Li WY, Suo ZL. Chloroplast phylogenomics and divergence times of *Lagerstroemia* (Lythraceae). *BMC Genomics*. 2021;22(1):434.
- Chen Q, Hu HS, Zhang DQ. DNA Barcoding and phylogenomic analysis of the genus *Fritillaria* in China based on complete chloroplast genomes. *Front Plant Sci*. 2022;13:764255.
- Ji YH, Yang LF, Chase MW, Liu CK, Yang ZY, Yang J, et al. Plastome phylogenomics, biogeography, and clade diversification of Paris (Melanthiaceae). *BMC Plant Biol*. 2019;19(1):543.
- Pardo C, Cubas P, Tahir H. Molecular phylogeny and systematics of *Genista* (Leguminosae) and related genera based on nucleotide sequences of nrDNA (ITS region) and cpDNA (trnL-trnF intergenic spacer). *Plant Syst Evol*. 2004;244:93–119.
- Ottensburghs J. The genic view of hybridization in the Anthropocene. *Evol Appl*. 2021;14(10):2342–60.
- Dong BR, Zhao ZL, Ni LH, Wu JR, Danzhen ZG. Comparative analysis of complete chloroplast genome sequences within Gentianaceae and significance of identifying species. *Chin Tradit Herbal Drugs*. 2020;51(6):1641–9.
- Hu HS, Zhang DQ. DNA super-barcoding of several medicinal species in *Gentiana* from Yunnan Province. *China J Chin Mater Med*. 2021;46(20):5260–9.
- Zhang Y, Yu JY, Xia MZ, Chi XF, Khan G, Chen SL, et al. Plastome sequencing reveals phylogenetic relationships among *Comastoma* and related taxa (Gentianaceae) from the Qinghai-Tibetan Plateau. *Ecol Evol*. 2021;11(22):16034–46.
- Kong HH, Liu WZ, Yao G, Gong W. A comparison of chloroplast genome sequences in *Aconitum* (Ranunculaceae): a traditional herbal medicinal genus. *PeerJ*. 2017;5:e4018.
- Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, et al. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell*. 2001;13(3):645–58.
- Chen Q, Wu XB, Zhang DQ. Phylogenetic analysis of *Fritillaria cirrhosa* D. Don and its closely related species based on complete chloroplast genomes. *PeerJ*. 2019;7:e7480.
- Zhou T, Wang J, Jia Y, Li WL, Xu FS, Wang XM. Comparative chloroplast genome analyses of species in *Gentiana* section *Cruciata* (Gentianaceae) and the development of authentication markers. *Int J Mol Sci*. 2018;19(7):1962.
- Kaila T, Chaduvla PK, Rawal HC, Saxena S, Tyagi A, Mithra SVA, et al. Chloroplast genome sequence of clusterbean (*Cyamopsis tetragonoloba* L.): genome structure and comparative analysis. *Genes*. 2017;8(9):212.
- 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.
- Du YP, Bi Y, Yang FP, Zhang MF, Chen XQ, Xue J, et al. Complete chloroplast genome sequences of *Lilium*: Insights into evolutionary dynamics and phylogenetic analyses. *Sci Rep*. 2017;7(1):5751.
- Dong WP, Liu J, Yu J, Wang L, Zhou SL. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS One*. 2012;7(4):e35071.
- Ren FM, Wang LQ, Li Y, Zhuo W, Xu ZC, Guo HJ, et al. Highly variable chloroplast genome from two endangered Papaveraceae lithophytes *Corydalis tomentella* and *Corydalis saxicola*. *Ecol Evol*. 2021;11(9):4158–71.
- Hong Z, Wu ZQ, Zhao KK, Yang ZJ, Zhang NN, Guo JY, et al. Comparative analyses of five complete chloroplast genomes from the genus *Pterocarpus* (Fabaceae). *Int J Mol Sci*. 2020;21(11):3758.
- CBOL Plant Working Group. A DNA barcode for land plants. *Proc Natl Acad Sci U S A*. 2009;106(31):12794–7.
- Njuguna AW, Li ZZ, Saina JK, Munywoki JW, Gichira AW, Gituru RW, et al. Comparative analyses of the complete chloroplast genomes of *Nymphoides* and *Menyanthes* species (Menyanthaceae). *Aquat Bot*. 2019;156:73–81.
- Li BC, Liu T, Ali A, Xiao Y, Shan N, Sun JY, et al. Complete chloroplast genome sequences of three Aroideae species (Araceae): lights into selective pressure, marker development and phylogenetic relationships. *BMC Genomics*. 2022;23(1):218.
- Sun SS, Fu PC, Zhou XJ, Cheng YW, Zhang FQ, Chen SL, et al. The complete plastome sequences of seven species in *Gentiana* sect. *Kudoa* (Gentianaceae): Insights into plastid gene loss and molecular evolution. *Front Plant Sci*. 2018;9:493.

46. Loera-Sánchez M, Studer B, Kölliker R. DNA barcode trnH-psbA is a promising candidate for efficient identification of forage legumes and grasses. *BMC Res Notes*. 2020;13(1):35.
47. Dugas DV, Hernandez D, Koenen EJ, 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(1):16958.
48. Asaf S, Khan AL, Khan MA, Shahzad R, Lubna, Kang SM, et al. Complete chloroplast genome sequence and comparative analysis of loblolly pine (*Pinus taeda* L.) with related species. *PLoS One*. 2018;13(3):e0192966.
49. Guo MY, Pang XH, Xu YQ, Jiang WJ, Liao BS, Yu JS, et al. Plastid genome data provide new insights into the phylogeny and evolution of the genus *Epimedium*. *J Adv Res*. 2022;36:175–85.
50. Ren T, Xie DF, Peng C, Gui LJ, Price M, Zhou SD, et al. Molecular evolution and phylogenetic relationships of *Ligusticum* (Apiaceae) inferred from the whole plastome sequences. *BMC Ecol Evol*. 2022;22(1):55.
51. Zheng B. Taxonomic studies of sect. *Monopodiae* and sect. *Kudoa* of *Gentiana* (Tourn.) L. (Gentianaceae) – And the discussions on the systematic status of two gentians in Flora of Hubei. Wuhan Botanical Garden, Chinese Academy of Sciences 2017.
52. Yuan YM, Kupfer P, Doyle JJ. Infrageneric phylogeny of the genus *Gentiana* (Gentianaceae) inferred from nucleotide sequences of the internal transcribed spacers (ITS) of nuclear ribosomal DNA. *Am J Bot*. 1996;83(5):641–52.
53. Chen Q, Wu XB, Zhang DQ. Comparison of the abilities of universal, super, and specific DNA barcodes to discriminate among the original species of *Fritillariae cirrhosae* bulbous and its adulterants. *PLoS One*. 2020;15(2):e0229181.
54. Lee SY, Xu KW, Huang CY, Lee JH, Liao WB, Zhang YH, et al. Molecular phylogenetic analyses based on the complete plastid genomes and nuclear sequences reveal *Daphne* (Thymelaeaceae) to be non-monophyletic as current circumscription. *Plant Divers*. 2022;44(3):279–89.
55. Zhang SD, Jin JJ, Chen SY, Chase MW, Soltis DE, Li HT, et al. Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. *New Phytol*. 2017;214(3):1355–67.
56. Li QJ, Liu Y, Wang AH, Chen QF, Wang JM, Peng L, et al. Plastome comparison and phylogenomics of *Fagopyrum* (Polygonaceae): insights into sequence differences between *Fagopyrum* and its related taxa. *BMC Plant Biol*. 2022;22(1):339.
57. Kreuzer M, Howard C, Adhikari B, Pendry CA, Hawkins JA. Phylogenomic approaches to DNA barcoding of herbal medicines: developing clade-specific diagnostic characters for *Berberis*. *Front Plant Sci*. 2019;10:586.
58. Fu CN, Mo ZQ, Yang JB, Cai J, Ye LJ, Zou JY, et al. Testing genome skimming for species discrimination in large taxonomically difficult genus *Rhododendron*. *Mol Ecol Resour*. 2022;22(1):404–14.
59. Favre A, Michalak I, Chen CH, Wang JC, Pringle JS, Matuszak S, et al. Out-of-Tibet: the spatio-temporal evolution of *Gentiana* (Gentianaceae). *J Biogeogr*. 2016;43(10):1967–78.
60. Li XJ, Wang LY, Yang HL, Liu JQ. Confirmation of natural hybrids between *Gentiana straminea* and *G. siphonantha* (Gentianaceae) based on molecular evidence. *Front Biol China*. 2008;3(4):470–6.
61. Wu JF, Jia DR, Liu RJ, Zhou ZL, Wang LL, Chen MY, et al. Multiple lines of evidence supports the two varieties of *Halenia elliptica* (Gentianaceae) as two species. *Plant Divers*. 2021;44(3):290–9.
62. Payseur BA, Rieseberg LH. A genomic perspective on hybridization and speciation. *Mol Ecol*. 2016;25(11):2337–60.
63. Todesco M, Pascual MA, Owens GL, Ostevik KL, Moyers BT, Hubner S, Heredia SM, et al. Hybridization and extinction. *Evol Appl*. 2016;9(7):892–908.
64. Rieseberg LH, Willis JH. Plant speciation. *Science*. 2007;317(5840):910–4.
65. Čertner M, Kolář F, Schönschetter P, Frajman B. Does hybridization with a widespread congener threaten the long-term persistence of the Eastern Alpine rare local endemic *Knautia carinthiaca*? *Ecol Evol*. 2015;5(19):4263–76.
66. Čertner M, Kolář F, Frajman B, Winkler M, Schönschetter P. Massive introgression weakens boundaries between a regionally endemic allopolyploid and a widespread congener. *Perspect Plant Ecol*. 2020;42:125502.
67. Caeiro-Dias G, Brelford A, Kaliontzopoulou A, Meneses-Ribeiro M, Crochet PA, Pinho C. Variable levels of introgression between the endangered *Podarcis carbonelli* and highly divergent congeneric species. *Heredity*. 2021;126(3):463–76.
68. Tang J, Sun SG, Huang SQ. Experimental sympatry suggests geographic isolation as an essential reproductive barrier between two sister species of *Pedicularis*. *J Syst Evol*. DOI: <https://doi.org/10.1111/jse.12835>.
69. Hao DC, Xiao PG. Pharmaceutical resource discovery from traditional medicinal plants: pharmacophylogeny and pharmacophylogenomics. *Chin Herb Med*. 2020;12(2):104–17.
70. Yang JB, Li DZ, Li HT. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol Ecol Resour*. 2014;14(5):1024–31.
71. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
72. 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):241.
73. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*. 2015;31(20):3350–2.
74. Katoh K, Standley DM. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics*. 2016;32(13):1933–42.
75. Kearse M, Moir R, Wilson A, Stones-Havas S, 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):1647–9.
76. Burland TG. DNASTAR's Lasergene sequence analysis software. *Methods Mol Biol*. 2000;132:71–91.
77. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33(7):1870–4.
78. Amiryousefi A, Hyvönen J, Poccai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics*. 2018;34(17):3030–1.
79. Dubchak I. Comparative analysis and visualization of genomic sequences using VISTA browser and associated computational tools. *Methods Mol Biol*. 2007;395:3–16.
80. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004;32:W273–9.
81. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*. 2017;34(12):3299–302.
82. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet*. 2003;106(3):411–22.
83. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–3.
84. Ronquist F, Teslenko M, Van DMP, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61(3):539–42.

Publisher's Note

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

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

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

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

