

RESEARCH ARTICLE

Open Access



Application of chloroplast genome in the identification of Traditional Chinese Medicine *Viola philippica*

Dong-Ling Cao[†], Xue-Jie Zhang[†], Shao-Qiu Xie, Shou-Jin Fan^{*} and Xiao-Jian Qu^{* ID}

Abstract

Background: *Viola philippica* Cav. is the only source plant of “Zi Hua Di Ding”, which is a Traditional Chinese Medicine (TCM) that is utilized as an antifebrile and detoxicant agent for the treatment of acute pyogenic infections. Historically, many *Viola* species with violet flowers have been misused in “Zi Hua Di Ding”. *Viola* have been recognized as a taxonomically difficult genera due to their highly similar morphological characteristics. Here, all common *V. philippica* adulterants were sampled. A total of 24 complete chloroplast (cp) genomes were analyzed, among these 5 cp genome sequences were downloaded from GenBank and 19 cp genomes, including 2 “Zi Hua Di Ding” purchased from a local TCM pharmacy, were newly sequenced.

Results: The *Viola* cp genomes ranged from 156,483 bp to 158,940 bp in length. A total of 110 unique genes were annotated, including 76 protein-coding genes, 30 tRNAs, and four rRNAs. Sequence divergence analysis screening identified 16 highly diverged sequences; these could be used as markers for the identification of *Viola* species. The morphological, maximum likelihood and Bayesian inference trees of whole cp genome sequences and highly diverged sequences were divided into five monophyletic clades. The species in each of the five clades were identical in their positions within the morphological and cp genome tree. The shared morphological characters belonging to each clade was summarized. Interestingly, unique variable sites were found in *ndhF*, *rpl22*, and *ycf1* of *V. philippica*, and these sites can be selected to distinguish *V. philippica* from samples all other *Viola* species, including its most closely related species. In addition, important morphological characteristics were proposed to assist the identification of *V. philippica*. We applied these methods to examine 2 “Zi Hua Di Ding” randomly purchased from the local TCM pharmacy, and this analysis revealed that the morphological and molecular characteristics were valid for the identification of *V. philippica*.

Conclusions: This study provides invaluable data for the improvement of species identification and germplasm of *V. philippica* that may facilitate the application of a super-barcode in TCM identification and enable future studies on phylogenetic evolution and safe medical applications.

Keywords: *Viola philippica*, Identification, Chloroplast genome, Unique variable site, Morphological characteristics, Phylogeny

Background

As the largest genus of Violaceae, *Viola* is mainly distributed in temperate and tropical regions, and China is one of the distribution centers [1–3]. More than 20 *Viola* species have been recorded as medicinal plants with definite efficacy and indications [4–7]. *V. philippica* (Herba

[†]Dong-Ling Cao and Xue-Jie Zhang are co-first author.

*Correspondence: fansj@sdu.edu.cn; quxiaojian@sdu.edu.cn

Shandong Provincial Key Laboratory of Plant Stress Research, College of Life Sciences, Shandong Normal University, Ji'nan 250014, China



Violae) possesses significant and unique efficacy in clinical antiviral therapy. Research has shown that *V. philippica* extract possess a wide range of pharmacological and biological activities, including antiviral, antifungal, anticoagulant and anticancer functions [8, 9]. Notably, cyclotides from *V. philippica* extracts are remarkably stable and tolerate harsh thermal, chemical, and enzymatic conditions with high biological activity, including insecticidal, cytotoxic, and neurotensin antagonistic activities. These characteristics make *V. philippica* ideal for potential agrochemical or pharmaceutical applications [10–12]. *V. philippica* is a small perennial herb with violet flowers. The dried whole plant (including the roots) is an important TCM named “Zi Hua Di Ding” in Chinese [13, 14]. “Zi Hua Di Ding” was the only *Viola* species described in the Chinese Pharmacopoeia in 1977, in which *V. philippica* is the only source of “Zi Hua Di Ding”; it has been widely used ever since [15]. Furthermore, the anti-HIV activity of *V. philippica* as traditional medicine was described in the 1989 World Health Organization (WHO) bulletin [16, 17]. Importantly, recent studies have demonstrated that its extracts exhibit high inhibitory activity against HIV-1 and respiratory syncytial virus (RSV) in vitro, with potential for efficacious clinical applications [18–21].

Viola is traditionally considered to be a morphologically difficult genus to classify [1, 2, 22]. Due to its wide distribution, frequent hybridization, and both open and closed flowers (both sexual and asexual), the morphological features of the *Viola* species exhibit high variability [23–26]. The infrageneric and interspecific relationships are confused due to similar morphological characteristics. The phylogenetic positions of some *Viola* species are difficult to identify. Unclear phylogenetic relationships and highly similar morphological characteristics among *Viola* species restrict the breeding and development of *V. philippica* germplasm resources. One or a more chloroplast DNA molecular markers, the nuclear intergenic transcribed spacer (ITS), or the inter-simple sequence repeat (ISSR) have previously been used to infer the phylogenetic relationships within *Viola* [27–29]. However, the interspecies phylogenetic relationships of *Viola* are still controversial due to insufficient informative sites and incomplete taxon sampling. Most prior studies have focused on the *Viola* species of North America, Korea, and Japan, while there are limited studies regarding *Viola* species from China. Flora of China proposed that the *Viola* species in China are divided into 14 sections based on morphological characteristics, including sections *Adnatae*, *Bilobatae*, *Brevicalcaratae*, *Caudicaules*, *Diffusae*, *Erectae*, *Longicalcaratae*, *Noverculae*, *Pinnatae*, *Plagiostigma*, *Serpentes*, *Trigonocarpaceae*, *Vaginatae*, and *Viola* [3, 30]. The range of sections

and their identification depends only on morphological characteristics that are often disputed by taxonomists. Taxonomic and phylogenetic analyses of *Viola* based on molecular inference are frequently inconsistent with the conclusions of traditional views. The delimitation of sect. *Adnatae* and the phylogenetic position of the species within it are widely disputed. *V. philippica* belongs to the sect. *Adnatae*, which is a group of about 35 species in China. It is very difficult to accurately identify *V. philippica* for use in TCM by morphological characteristics or organoleptic methods. Ensuring the authenticity of source plants is a key issue in the use of herbs. Some species have been misidentified due to their violet-colored flowers being the same color as those of *V. philippica*, e.g., *Corydalis bungeana* or *Gueldenstaedtia verna*. Common adulterants in “Zi Hua Di Ding” are often *Viola* species, i.e., *V. patrinii*, *V. inconspicua*, *V. yezoensis*, *V. phalacrocarpa*, and *V. prionantha*. In practice, some micromorphological features are difficult to identify, especially those concentrated in the floral organs. The result is that many *Viola* species with violet flowers were often treated as *V. philippica*. Folk lacks botanical knowledge, confusing the safety and effectiveness of drug use. Consumers are often unable to verify original voucher specimens of herbs when they purchase them in pharmacies. Therefore, accurate and efficient identification of the source plants of “Zi Hua Di Ding” is necessary.

DNA barcoding is an emerging technology of molecular identification and classification; it significantly enhances the safety and efficacy of medicinal herbs [31–33]. DNA barcoding is not restricted by morphological characteristics or physiological conditions, allowing for species authentication without professional taxonomic knowledge [34, 35]. Chloroplast, the vital photosynthesis tissue, is an important organelle in green plants. The plastomic sequence is commonly described as a quadripartite structure with a large single-copy (LSC), a small single-copy (SSC), and a pair of inverted-repeat (IR) regions [36]. The angiosperm cp genomes are generally of moderate size, ranging from 120 to 160 kb [37–39]. With the rapid development of high throughput sequencing technology, the cp genome (chloroplast genome) is widely applied as a super-barcode, which could provide effective information for resolving phylogenetic relationships and identification of medicinal plants [40–42]. In the present study, we set out to analyze the cp genome to authenticate *Viola* species, in particular *V. philippica*. Furthermore, molecular markers in highly divergent cp genome regions were screened for the identification of *Viola* and phylogenetic studies.

Herein, a total of 22 *Viola* samples were obtained, for which cp genome sequences of three species were downloaded from GenBank and the cp genomes of 19 *Viola*

samples were newly sequenced. The common adulterants in “Zi Hua Di Ding” belonging to the *Viola* were covered. The purpose of this study was to 1) compare the chloroplast genome structures of the sampled *Viola* species; 2) find an effective means to by which to distinguish *V. philippica* from other *Viola* species; and 3) resolve the infrageneric relationships within *Viola* species, especially section *Adnetae*, *Pinnatae*, and *Bilobatae*, using complete plastomic sequences and highly diverged sequences. This study provides invaluable data for species identification and improvement in germplasm generation of *V. philippica*; it will facilitate the application of super-barcode cp genomics in TCM identification, allowing for future studies on phylogenetic evolution and safe medical applications of *Viola* species.

Results

Cp genome organization of *Viola*

All *Viola* species analyzed in the current study that have been accurately identified possessed a similar genome structure, gene order, and orientation. Plastome size ranged from 156,483 bp (*V. phalacrocarpa*) to 158,940 bp (*V. acuminata*) (Table 1). The two *V. philippica* plastomes, approximately at 157 kb (Fig. 1). The 17 *Viola* species cp genomes had highly conserved quadripartite structure with the LSC region (85,364–87,250 bp), SSC region (16,558–18,008 bp), and a pair of IRs (26,404–27,404 bp). The overall guanine-cytosine (GC) content was approximately 36%. The GC contents in the LSC and SSC regions of all 17 species were lower than in the IRs. Overall, 128 genes were annotated in the 17 *Viola* species, of which 110 were unique, consisting of 76 protein-coding, 30 tRNA, and four rRNA genes (Table 1). The encoded genes of the cp genomes are divided into four categories based on their functions: photosynthesis genes, self-replication genes, other biosynthesis genes, and some genes of unknown function (Table S1). Among these genes, three genes (*infA*, *rpl32*, *rps16*) were completely degraded. There were 17 intron-containing genes in each of the *Viola* species, of which two PCGs (*ycf3* and *clpP*) had two introns, 9 PCGs (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rpoC1*, and *rps12*) and 6 tRNAs (*trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, and *trnV-UAC*) had a single intron each.

Structural comparison of *Viola* cp genomes

The locations of IR/SC junctions were conserved among all cp genomes (Fig. S1). In general, the *rps19* gene extended 2–73 bp into the IRs at the junction of the LSC/IRb (JLB), resulting in the duplication of 3'-ends of this gene in the IRa region. The *ndhF* gene was located in the SSC at a 3–33 bp distance from the SSC/IRb (JSB) border. Meanwhile, in *V. mongolica*

and *V. yunnanfuensis* the *ndhF* gene traverses the JSB by extending 55 and 51 bp, respectively, into the IRb region. In *V. philippica*, the *ndhF* gene was located in the SSC at a 33 bp distance from the SSC/IRb (JSB) border. In all cp genomes, the SSC/IRa (JSA) junction was located within the *ycf1* gene, with 1091 to 1866 bp of *ycf1* duplicated in the IRb region. In addition, the *trnH* genes were all located in the LSC region, with the distance between *trnH-GUG* and the LSC/IRa (JLA) border varying from 26 to 85 bp.

The number of repeat sequences was calculated and the threshold of repeat length was set to ≥ 30 bp. A total of 729 repeats were detected in the 17 *Viola* cp genomes (Fig. S2A). The results revealed that *V. acuminata*, *V. mirabilis*, and *V. raddeana* had the greatest number of repeats (49), while *V. mongolica* had the least (33). Conversely, such complementary repeats were not detected in the cp genomes of *V. mongolica* and *V. yezoensis*. Additionally, tandem repeats that ranged from 30 to 33 bp were the most abundant, followed by those that ranged from 38 to 41 bp. The number of repeats in *V. philippica* was 46, comprising 19 forward repeats, 8 reverse repeats, 1 complementary repeat, and 18 palindromic repeats (Fig. S2B). A total of 450 SSRs were identified; the most abundant repeats in *Viola* cp genomes were mononucleotide repeats, followed by di-, and tetra- nucleotide repeats (Table S2). The most common mononucleotide repeat type was A/T, and all the dinucleotide repeats were composed of AT/TA. Meanwhile, tetranucleotide repeats were detected only in the cp genome of *V. websteri*. These SSRs, were located most often in the LSC regions. We identified 22 SSRs in the cp genome of *V. philippica*, including 16 mononucleotide repeats (15 A/T repeats and 1 G/C repeats) and 6 dinucleotide repeats (2 AT repeats and 4 TA repeats).

Identification of hypervariable regions

High levels of sequence similarity was identified across all 17 *Viola* species. However, it was observed that the LSC and SSC regions were more divergent than IRs. In addition, intergeneric regions exhibit greater divergence than coding regions (Fig. 2A). The nucleotide diversity (Π) value was calculated, and their values varied from 0 to 0.06602 (Fig. 2B). High sequence divergence was detected in the following genomic regions: *matK*, *ndhF*, *ycf1*, *rpl22*, *rps15*, *ndhA*, *petN-psbM*, *petA-psbJ*, *ccsA-ndhD*, *trnG-UCC-2-trnR-UCU*, *rps8-rpl14*, *trnD-GUC-trnY-GUA*, *trnG-GCC-trnfM-CAU*, *trnH-GUG-psbA*, *psbZ-trnG-GCC*, and *rbcL-accD*. These divergent regions could be candidates for the development of critical molecular markers for phylogenetic analyses of *Viola* species.

Table 1 The cp genome characteristics of 17 *Viola* species

Species	Raw reads (pair)	Genome Size (bp)	LSC region (bp)	IR region (bp)	SSC region (bp)	GC content (%)			Number of Genes				
						Over all	LSC	IR	SSC	Total	CDS	tRNAs	rRNAs
<i>Viola acuminata</i>	9,709,090	158940	87250	27159	17372	36.1	33.6	42.1	29.8	110	76	30	4
<i>V. chaerophylloides</i>	8,709,947	157113	85577	27154	17228	36.3	34.0	42.2	29.9	110	76	30	4
<i>V. collina</i>	7,447,895	157890	86495	27102	17191	36.3	33.8	42.2	30.1	110	76	30	4
<i>V. dissecta</i>	7,178,430	157113	85557	27154	17228	36.3	34.0	42.2	29.9	110	76	30	4
<i>V. inconspicua</i>	6,817,359	156624	85822	27122	16558	36.3	33.8	42.1	30.1	110	76	30	4
<i>V. mirabilis</i>	-	158162	86565	27123	17351	36.2	33.8	42.2	29.9	110	76	30	4
<i>V. monbeigii</i>	7,091,420	156580	85726	26423	18008	36.3	33.8	42.6	29.6	110	76	30	4
<i>V. mongolica 1</i>	9,917,465	156870	85364	27184	17138	36.4	34.0	42.1	30.0	110	76	30	4
<i>V. mongolica 2</i>	7,985,972	156904	85373	27192	17147	36.4	34.0	42.1	30.0	110	76	30	4
<i>V. patinii</i>	8,016,998	156508	85694	27404	18006	36.3	33.8	42.6	29.6	110	76	30	4
<i>V. phalacrocarpa</i>	1,630,516	156483	85669	27404	18006	36.3	33.8	42.6	29.6	110	76	30	4
<i>V. philippica 1</i>	8,268,704	156599	85751	26422	18004	36.2	33.8	42.6	29.6	110	76	30	4
<i>V. philippica 2</i>	5,611,822	157298	85756	27195	17146	36.2	33.8	42.1	29.9	110	76	30	4
<i>V. pionantha</i>	8,175,231	156509	85693	26404	18008	36.3	33.8	42.6	29.6	110	76	30	4
<i>V. raddeana</i>	-	157597	86460	26924	17289	36.2	33.8	42.3	29.9	110	76	30	4
<i>V. variegata</i>	7,200,026	156582	85769	26411	17991	36.3	33.8	42.6	29.6	110	76	30	4
<i>V. websteri</i>	-	158111	86588	27166	17191	36.2	33.7	42.1	29.9	110	76	30	4
<i>V. yezoensis</i>	4,636,893	156537	85706	26417	17997	36.3	33.8	42.6	29.6	110	76	30	4
<i>V. yunnanfuensis 1</i>	6,179,277	156974	85975	27160	16679	36.3	33.8	42.1	30.2	110	76	30	4
<i>V. yunnanfuensis 2</i>	4,905,794	156931	85858	27208	16663	36.3	33.8	42.1	30.2	110	76	30	4

LSC Large single copy, SSC Small single copy, IR Inverted repeat, rRNA transfer RNA, rRNA ribosomal RNA

Chloroplast genome for identification of *V. philippica*

The complete cp genome and 16 highly divergent regions (*matK*, *ndhF*, *ycf1*, *rpl22*, *rps15*, *ndhA*, *petN-psbM*, *petA-psbJ*, *ccsA-ndhD*, *trnG-UCC-2-trnR-UCU*, *rps8-rpl14*, *trnD-GUC-trnY-GUA*, *trnG-GCC-trnfM-CAU*, *trnH-GUG-psbA*, *psbZ-trnG-GCC*, and *rbcL-accD*) generated phylogenetic trees with strong support (Fig. S3). Zi Hua Di Ding 1 and Zi Hua Di Ding 2 clustered with all *Viola* species, not *Corydalis tomentella* or *Oxytropis arctobia*. The results showed that “Zi Hua Di Ding 1” formed a clade with *V. yezoensis* and not *V. philippica*. Zi Hua Di Ding 2 is nested between *V. philippica* 1 and *V. philippica* 2.

In this study, four species-specific variable sites were present in *ndhF*, *rpl22*, and *ycf1* of *V. philippica* in comparison to other *Viola* species. The first unique locus is at positions 1612 of *ndhF* in *V. philippica*, where a G is located instead of a T. The second unique locus is A at position 270 of *rpl22* in *V. philippica*, while in other species there is a C. The other two unique sites are G and T at positions 2839 and 4217 of *ycf1* in *V. philippica*, in other species these sites are populated by T and G, respectively (Fig. 3). The nucleotide sequences of four pairs of specific primers used for PCR validation in 14 species are shown in Fig. 4. The amplified products for all individuals were approximately 350 bp (Fig. 4). Nucleotide sequences were consistent between Sanger sequencing results and next-generation sequencing results.

Infrageneric relationships of *Viola*

Comparison of morphological characteristics such as lobe vs entire, length of stipule adnate with petioles, the shape of the leaf blade, stigma type, and fruit shape, demonstrated that the 17 species could be classified into five clades including sections *Viola*, *Pinnatae*, *Adnatae*, *Trigonocarpaceae*, and *Bilobatae* (Fig. 5). Based on the topologies of all ML and BI trees, all 17 species were further divided into these five monophyletic groups (Fig. 6). Notably, when comparing the morphological and the cp genome tree, the species in all five clades were identical. We summarized shared morphological characteristics of species belonging to the same clade by comparative morphological analysis. The shared morphological characteristics for each of the five clades were shown next to each species in hand-drawn illustrations, and the dry leaves of “Zi

Hua Di Ding” purchased from the TCM pharmacy were shown next to Zi Hua Di Ding 1, 2 (Fig. 6). For sections *Viola*, *Trigonocarpaceae*, *Bilobatae*, *Pinnatae*, and *Adnatae*, the specific globose capsule, shared beak stigma, shared 2-lobed stigma, unique dissected leaf blade, and stipule adnate with petioles longer than one-half its length were the main morphological characteristics of each section, respectively. Importantly, the dimorphic leaf blade during the flowering period, i.e., smaller triangular-ovate for the lower leaf blades, longer oblong-ovate for the upper leaf blades, and fine tubular calcar with slightly downward curved ends were the most important morphological character for the identification of *V. philippica* (Fig. 1A, B).

The topologies of the ML and BI trees were highly concordant for complete cp genome sequences and highly diverged sequences with high support (Fig. 6, S5). In the present study, *Viola* was monophyletic and formed five clades. For section *Viola*, *V. collina* occupied the most basal position. The clade comprising *V. acuminata*, *V. mirabilis*, and *V. websteri* was located within section *Trigonocarpaceae*. For section *Bilobatae*, the species *V. mongolica* and *V. yunnanfuensis* were sisters to *V. raddeana*. For section *Pinnatae*, *V. chaerophylloides* is sister to *V. dissecta*. In section *Adnatae*, *V. inconspicua* and *V. variegata* firstly diverged, and *V. phalacrocarpa*, *V. patrinii*, and *V. prionantha* were sisters to *V. yezoensis*, *V. monbeigii*, and *V. philippica*.

Discussion

Cp genome structural changes in *Viola*

The cp genomes of land plants are highly conserved and can therefore provide important phylogenetic data. The cp genomes of 17 *Viola* species exhibited a typical quadripartite structure, with an identical number of protein-coding genes, tRNAs, and rRNAs (Table 1), consistent with the previously reported *Viola* cp genome [43, 44]. Furthermore, we identified that the genes *infA*, *rpl32*, and *rps16* were not present in 17 *Viola* cp genomes (Table S1). Studies suggest that the cause of the loss of these genes may be due to parallel loss of chloroplast DNA during the evolution of angiosperms [45–47]. There were 17 intron-containing genes in each *Viola* species analyzed (Table S1). Introns have been reported to increase the transcriptional efficiency of numerous genes

(See figure on next page.)

Fig. 1 Morphology and cp genome map of *Viola* species. Photo credits: Dong-Ling Cao. **A** The morphological character of *V. philippica*. **B** a. root; b. flower; c. upper leaf blade (ventral); d. upper leaf blade (dorsal); e. lower leaf blade (ventral); f. lower leaf blade (dorsal); g. sepal; h. nectar glands; i. petal; j. style; k. capsule; l. capsule longitudinal section; m. capsule transverse section; n. capsule dehiscence. Scale bar = 1cm in a; 50mm in b, c, d, e, f, g, h, i, and n; 5mm in k, l, m; 1mm in j. **C** Chloroplast genome map of *V. philippica* with annotated genes. The genes transcribed clockwise are shown inside of the circle whereas genes transcribed anti-clockwise are shown outside of the circle. The colored bars indicate different functional groups. The darker gray area in the inner circle denotes GC content while the lighter gray corresponds to the AT content of the genome. LSC: large single copy, SSC: small single copy, IR: inverted repeat

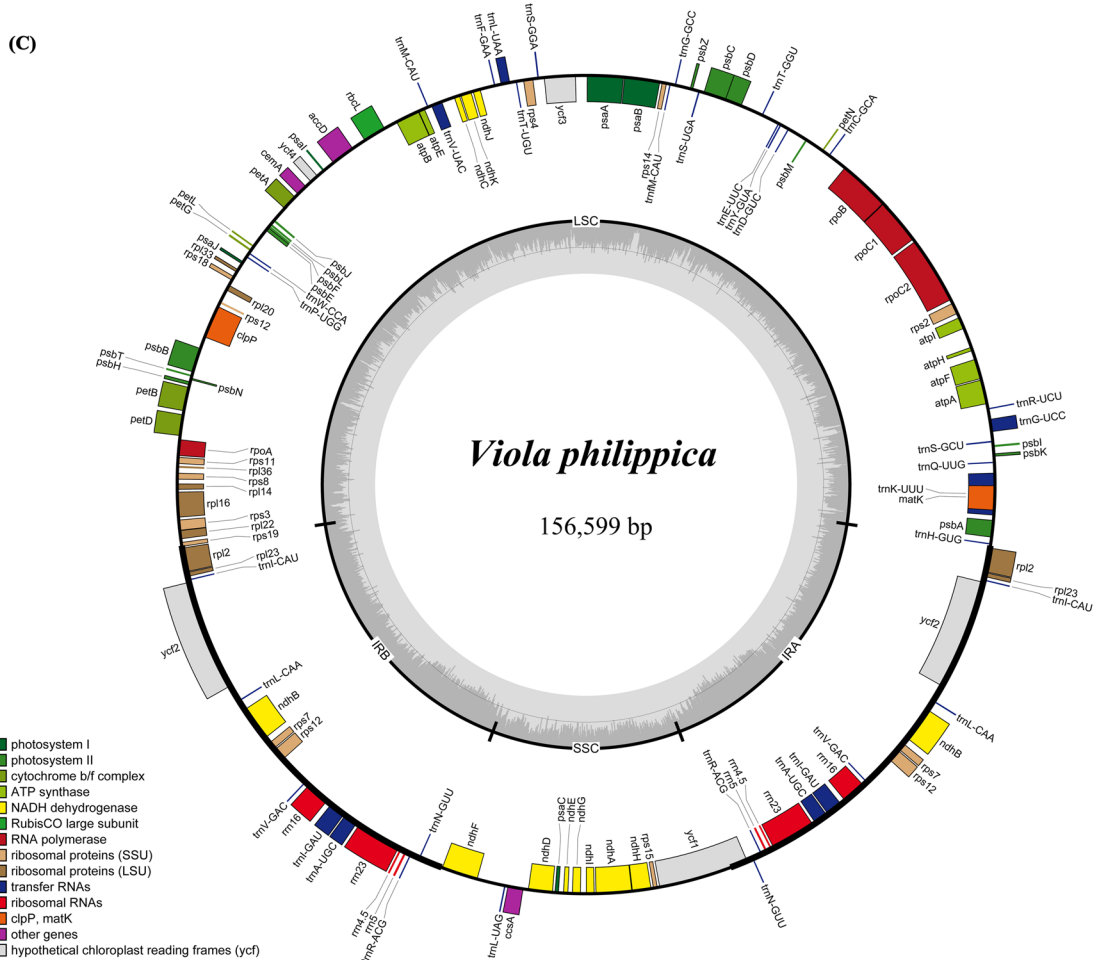
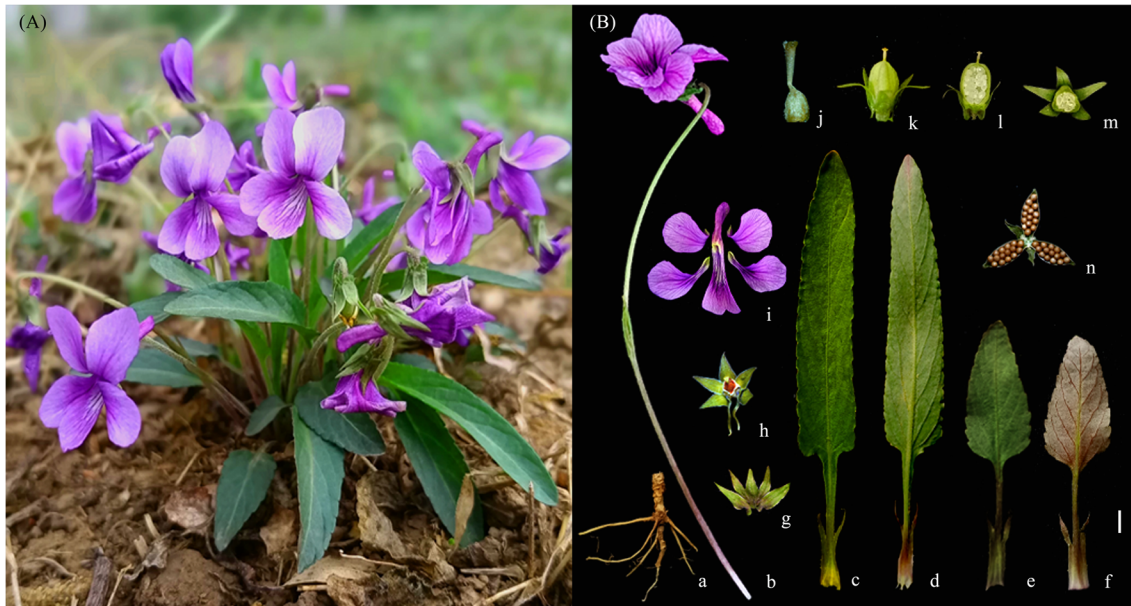


Fig. 1 (See legend on previous page.)

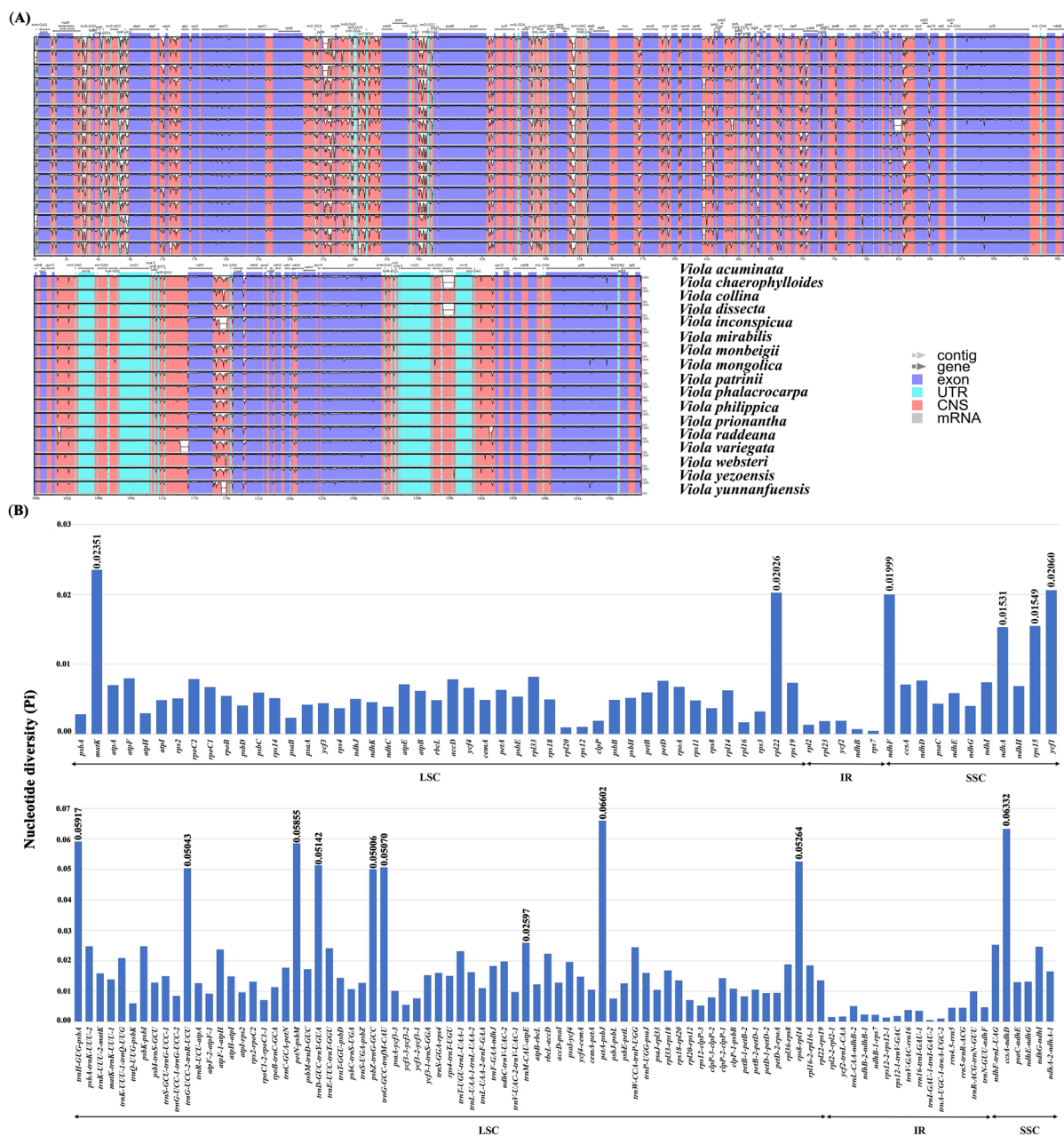


Fig. 2 Sequence polymorphism among 17 *Viola* cp genomes. **A** Sequence alignment of *Viola* cp genomes, with *Viola acuminata* as the reference using mVISTA. The position and transcriptional direction of each gene are indicated by gray arrows. The vertical scale indicates the percentage of identity, ranging from 50 to 100%. The horizontal axis indicates the coordinates within the chloroplast genome. Genome regions are color-coded as exons, introns, and intergenic spacer (IGS). **B** The nucleotide diversity (Pi) values of the 17 *Viola* cp genomes; Pi values of coding genes and Pi values of IGS

in a variety of organisms [48, 49]. Although generally IRs are highly conserved, the expansion and contraction of the IRs is a common characteristic of cp genomes and is thought to be the main cause of their variability in size [50, 51]. These changes have been associated with gene

duplication at the junction between the IRs and the LSC and SSC regions, which results in gene content variation between species. The IR expansion of *V. mongolica* and *V. yunnanfuensis* caused the *ndhF* gene into the IRb region. Repeat sequences are not only hotspots for mutations

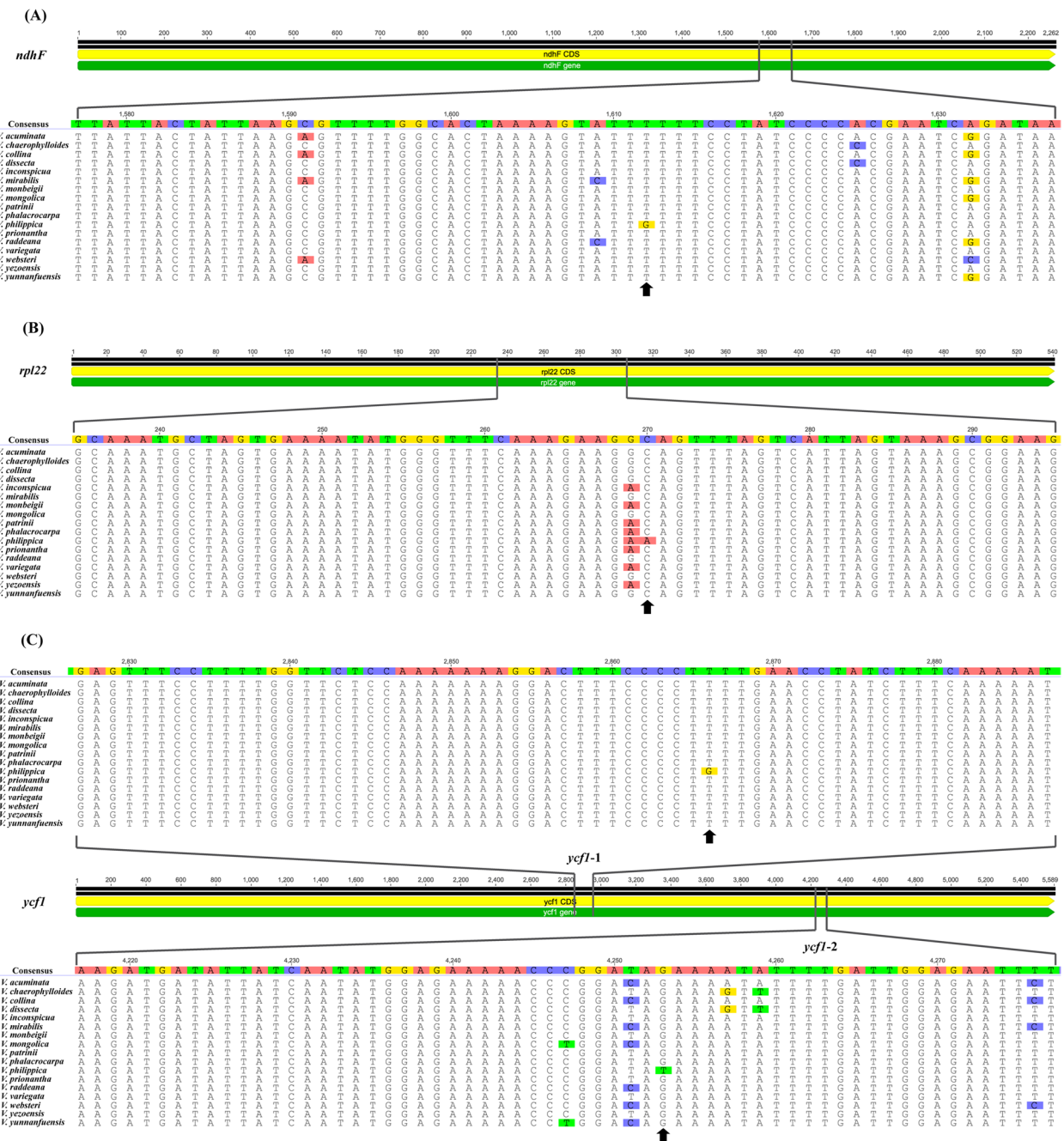


Fig. 3 Unique variation sites in *ndhF*, *rpl22*, and *ycf1* of *Viola philippica*. Alignment of four fragment sequences of 17 *Viola* species. **A** *ndhF*, **(B)** *rpl22*, **(C)** *ycf1*. The black arrow at the bottom indicated four species-specific variation sites in *Viola philippica*

(See figure on next page.)

Fig. 4 The primer sequences and electrophoretogram for four sequence fragments with unique variation sites in 14 newly sequenced *Viola* species. **A** *ndhF*, **(B)** *rpl22*, **(C)** *ycf1* primer pairs. Left and right arrows indicate forward and reverse primers, respectively. M: Marker. 1–14: *V. acuminata*; *V. chaerophylloides*; *V. collina*; *V. dissecta*; *V. inconspicua*; *V. monbeigii*; *V. mongolica*; *V. patrinii*; *V. phalacrocarpa*; *V. philippica*; *V. prionantha*; *V. variegata*; *V. yezoensis*; *V. yunnanfuensis*

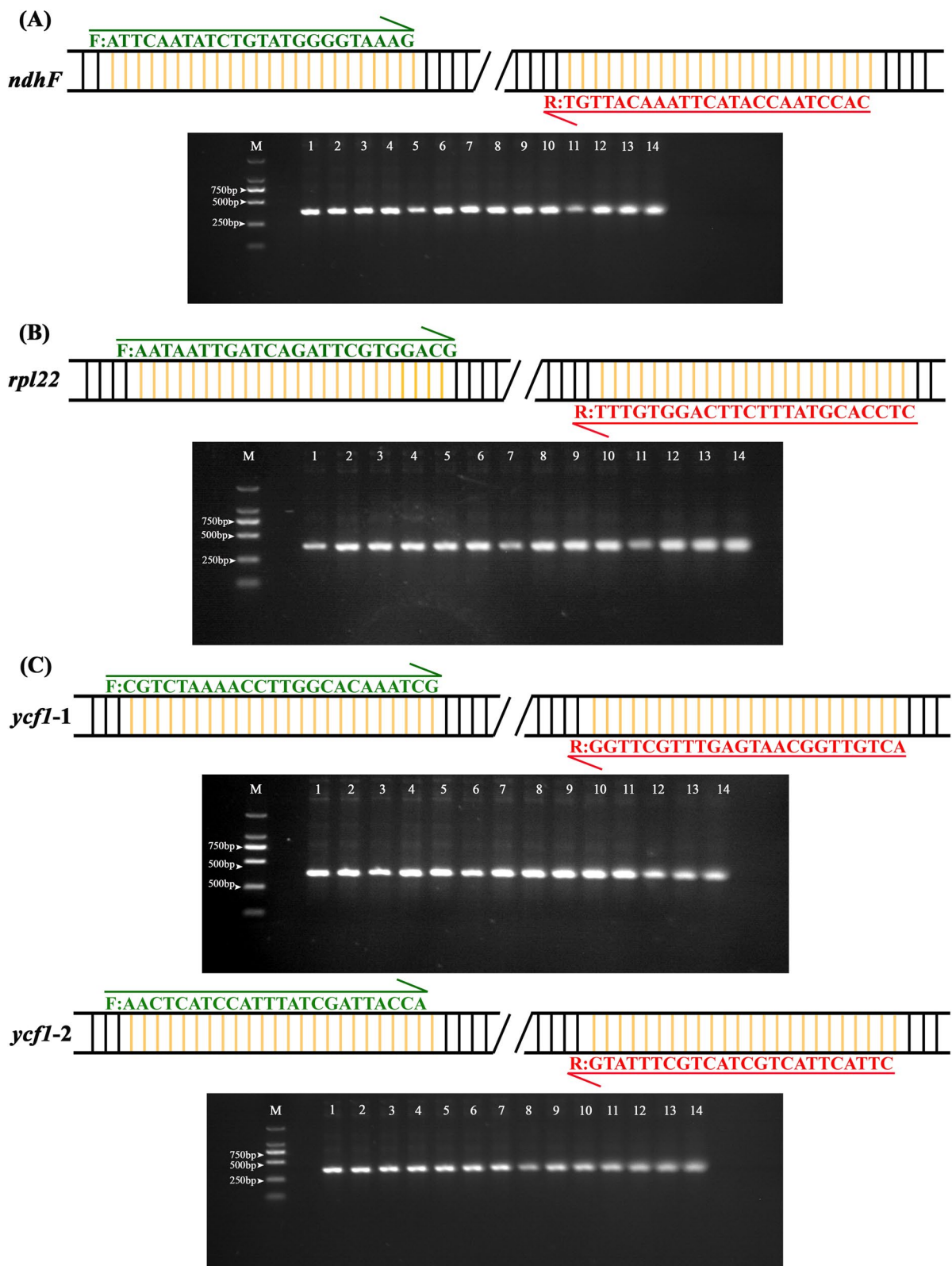
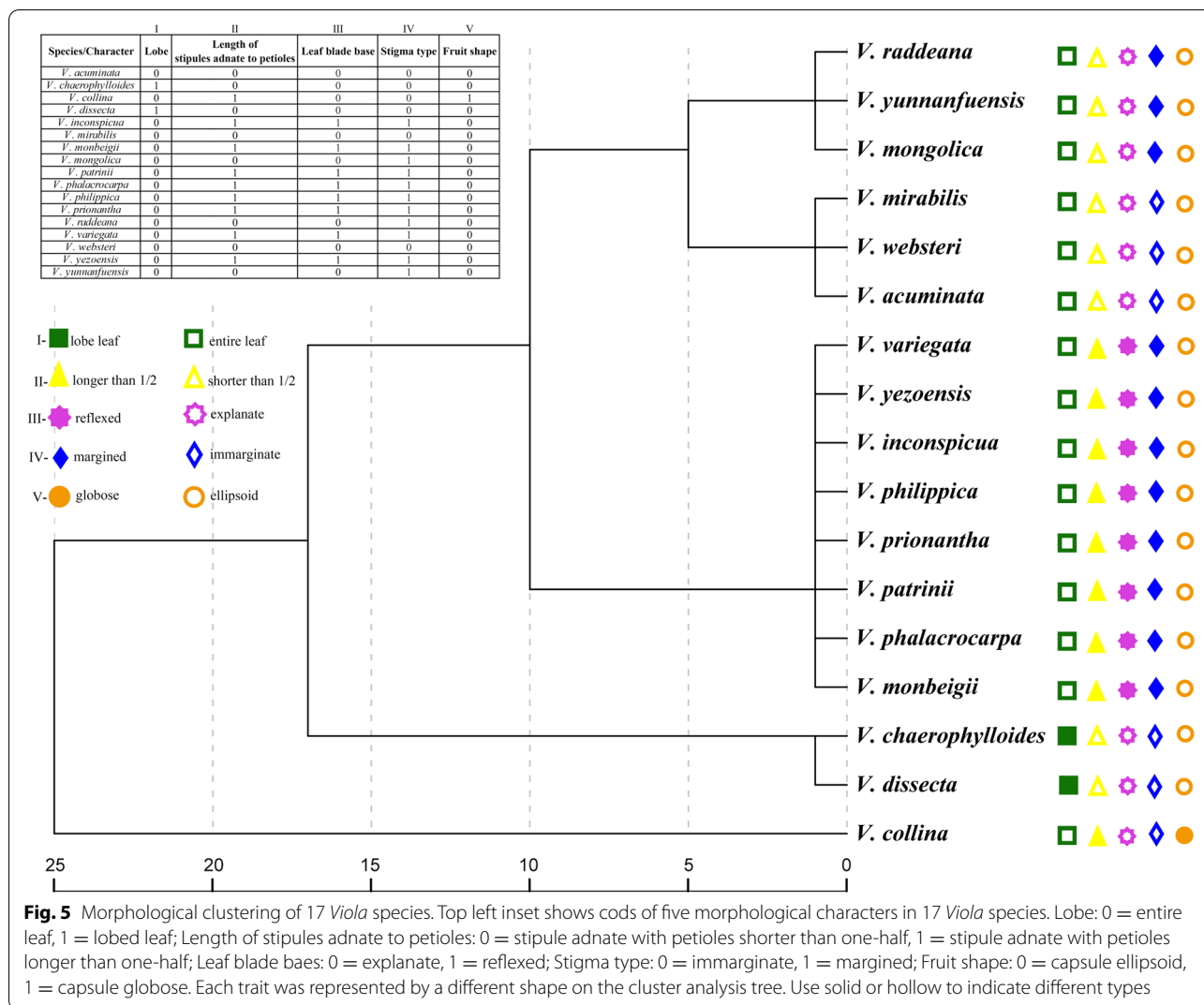


Fig. 4 (See legend on previous page.)

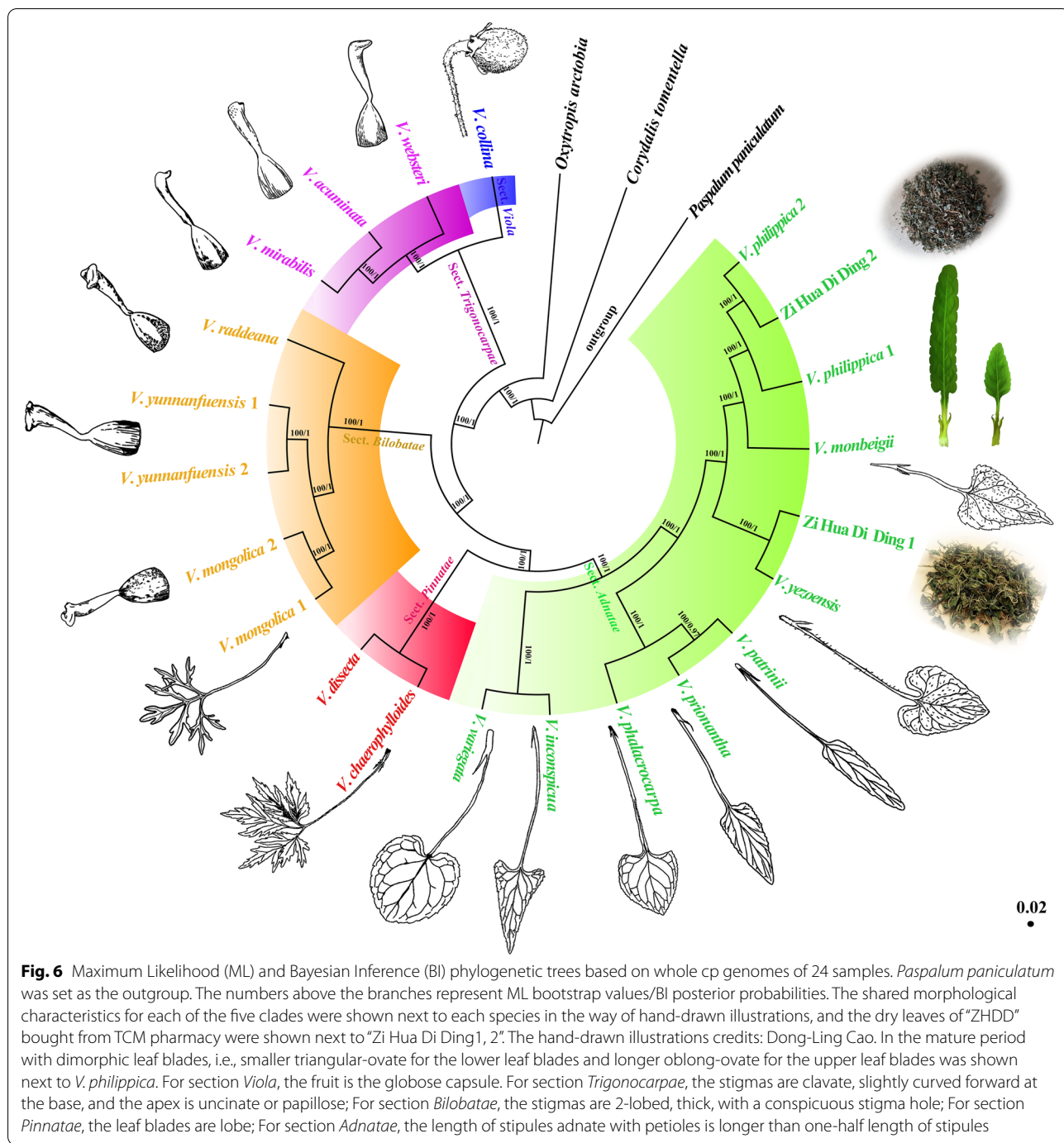


such as nucleotide substitutions, insertions and deletions, but are also important in phylogenetic studies [52, 53]. The number of repeats in *V. philippica* was 46, with 19 forward repeats, 8 reverse repeats, 1 complementary repeat, and 18 palindromic repeats. These data will provide a basis for studying the phylogeny of *V. philippica*. Notably, most mononucleotides and dinucleotides are composed of A and T, which may contribute to a bias in base composition [54].

Identification of *V. philippica* by morphology, cp genome phylogeny, and species-specific variable sites

The safe use of TCM conventionally relies on correct identification and TCM-guided clinical prescription. In the trade of herbal medicines, consumers often do not verify the identity original voucher specimen. Therefore, ensuring the authenticity of raw materials used as herbs is particularly important [55]. The traditional

identifications of herbs rely upon morphology, odor, or flavor and is performed by experts. Even now, morphological characteristics are still an important basis for identification [56–58]. However, it is difficult to train a person to acquire the required professional skills. Moreover, for some specimens with very similar morphological characteristics, it is difficult even for experts to accurately identify. It is well known that *V. philippica* and closely related species have highly similar morphological characters. The identification of common herbs methods including morphological observation, thin layer chromatography (TLC), high-performance liquid chromatography (HPLC), near-infrared spectroscopy (NIRS), and metabolomic approaches [59–62]. However, these methods are often complicated and costly. DNA barcoding can achieve rapid, accurate, and automated species identification [63–65]. ITS and cp genome fragment analysis have been used to distinguish *V. philippica* from closely related



species previously, however, fewer adulterants of *V. philippica* were sampled and strong support for some nodes in the phylogenetic tree was not acquired [28, 30]. *Viola* exhibit a low level of genetic differentiation as revealed by the ITS analysis [66]. However, ITS and plastid datasets did not provide substantial phylogenetic information, and the phylogenetic position of species within the genus is uncertain. Due to the rapid development of

sequencing techniques and bioinformatics, the complete cp genomes of plants can be rapidly acquired at low cost. Cp genomes have a moderate rate of nucleotide evolution, which results in their suitability for species identification and phylogenetic studies at different taxonomic levels [67, 68]. Cp genomes are proposed as potential super-barcode for species identification [69–71]. The *V. philippica* were analyzed by Blast online comparison in

the NCBI database. The results of gene sequence similarity comparison showed that the *V. philippica* sequences in this study had highest homology with *V. philippica* sequences that had been registered on GenBank. The similarity results reach more than 98%. This suggests that validation of the chloroplast genome is effective. The use of genetic distances using standardized gene regions (DNA barcodes) has provided complementary or alternative support for species identification, which is especially useful when distinct morphological characters are scarce or subtle [72]. Here, the genetic distance results based on the complete chloroplast genome showed that genetic distance is small with 17 *Viola* species. Among them, *V. philippica* and *V. monbeigii* had the smallest genetic distance, indicating that they are both closely related, which is consistent with the phylogenetic tree results. The identity of genes between populations within species is quite high, and consequently, genetic distance is small (Table S3). In our study, all common adulterants of *V. philippica* were sampled. The topologies of ML and BI trees were highly concordant for the complete cp genome sequence and the highly diverged sequences. Our phylogenetic results indicated strong support for *Viola* species in all sample, and the cp genomes sequences could be used as a super-barcode for authentication of *V. philippica* (Fig. 6). There are four specific variable sites in *ndhF*, *rpl22*, and *ycf1* of *V. philippica*, which were identified when we compared the alignment matrix of these genes for all sampled *Viola* species. Furthermore, Sanger sequencing results validated the four unique variable sites of *V. philippica* (Fig. 3), suggesting that they can be selectively amplified to distinguish *V. philippica* from other sampled *Viola* species, especially its adulterants. In short, the original source plant of “Zi Hua Di Ding” can be accurately identified by these species-specific variable sites. *Viola* species are traditionally morphologically difficult taxon to classify due to their very similar morphological characteristics. The wide distribution coupled with frequent hybridization increases the difficulty of species identification and phylogenetic analysis. Geographical differences and frequent hybridization may cause the occurrence of interspecific mutations in some genes of *V. philippica* (Fig. S4). In practice, these four unique variable loci should be considered simultaneously to ensure the greatest possible accuracy of identification. In addition, important morphological characteristics and the construction of a phylogenetic tree are presented to aid in the accurate identification of *V. philippica*. We applied these methods to examine two “Zi Hua Di Ding” purchased randomly from a local TCM pharmacy. The results showed the “Zi Hua Di Ding 1” and *V. yezoensis* clustered together, but not with *V. philippica*. “Zi Hua Di Ding 2” is nested between *V. philippica* 1 and *V. philippica* 2. There is a

high probability that the “Zi Hua Di Ding 1” purchased from TCM pharmacy in this study was not *V. philippica*, and considered should be *V. philippica* adulterants. “Zi Hua Di Ding 2” should be considered as genuine *V. philippica*.

In addition, morphological comparative analysis of the dimorphic leaf blade during the flowering period, i.e., smaller triangular-ovate for the lower leaf blades and longer oblong-ovate for the upper leaf blades in addition to fine tubular calcar with slightly downward curved ends were the most important morphological characteristic for the identification of *V. philippica*. In addition, we observed slight morphological changes in *V. philippica* individuals sampled during different growth periods. For example, some flowers become pale in color and the base of leaves occasionally widen. Therefore, we propose the flowering period to be the optimal time to morphologically identify *V. philippica*. According to reports regarding *V. philippica* extracts, the medicinal components vary at different times of the year, and the appropriate time of collection can be chosen according to medicinal needs [73]. Overall, morphological characteristics, cp genome phylogeny and species-specific variable sites analysis can be applied to distinguish *V. philippica* from other sample *Viola* species. For the authentication of “Zi Hua Di Ding”, a combination of method can be chosen, so that *V. philippica* can be accurately distinguished from its common adulterants. The identification of *V. philippica* is carried out according to the actual situation, combined with the distinct methods.

Infrageneric relationships of *Viola* (section *Adnetae*, *Pinnatae*, *Bilobatae*) based on morphological characters and cp genomes

Chloroplast genome sequences are invaluable for understanding plant evolution and phylogeny [74, 75]. One or a more of several chloroplast molecular markers (*atpB-rbcL*, *matk*, *petG-trnW*, *psbA-trnH*, *psbZ-trnG*, *psbK-I*, *rps19-trnH*, *rpl16-rps3*, *rpl2-23*, and *trnL-F*) and nuclear ITS were used to infer the phylogeny of *Viola*, however, the majority of interspecies relationships are not currently well resolved [27–29, 76]. The short branches in the phylogenetic tree in our study show consistency with previous studies and are the result of rapid divergence (Fig. S5) [66]. Their sequence identity is therefore likely to reflect explosive radiation, and not simply a recent origin. *Viola* species are traditionally morphologically difficult taxon to classify due to their very similar morphological characteristics, and there is some synapomorphy in *Viola* species. This may be why phylogenetic analyses of *Viola* based on molecular inference are often inconsistent with the results of the traditional morphological study. Complete cp genomes contain a wealth of genetic variation

and is an ideal source of data to study phylogeny among species [77–79]. Based on the whole cp genome, the infrageneric phylogenetic relationships of *Viola* were resolved in this study, and both ML and BI trees were strongly supported. Therefore, we suggested that the complete cp genome can be used as a super-barcode to distinguish closely related *Viola* species. Furthermore, relationships among sampled *Viola* species were also resolved based on only 16 molecular markers with high Pi values, their topologies are highly consistent with the complete cp genome. Therefore, these identified regions (*matK*, *ndhF*, *ycf1*, *rpl22*, *rps15*, *ndhA*, *petN-psbM*, *petA-psbJ*, *ccsA-ndhD*, *trnG-UCC-2-trnR-UCU*, *rps8-rpl14*, *trnD-GUC-trnY-GUA*, *trnG-GCC-trnfM-CAU*, *trnH-GUG-psbA*, *psbZ-trnG-GCC*, and *rbcl-accD*) could be used as markers for elucidating phylogenetic relationship within *Viola* species. These findings provide additional information for the selection of effective molecular markers to detect intra- and interspecific genetic polymorphisms (Fig. S3).

Viola are known as one of the taxonomically difficult groups to define since *Viola* species possessing many morphologically similar characteristics and intermediate forms that occur freely due to interspecific hybridization [23–25]. Morphological data are necessary to identify species and to infer their relationships in phylogenetic studies [80–82]. Interestingly, in this study the position of species in each of the five clades were identical between the morphological tree and the cp genome tree. The results revealed that species with the same morphological characteristics cluster together with a higher internal resolution (Fig. 5). Therefore, we advocated that the cp genomes should be combined with morphological characteristics in analyze of the phylogenetic position and identification of *V. philippica*. Previous research reported that *V. chaerophylloides* and *V. dissecta* are distantly related and that section *Pinnatae* should be treated as a subsection under section *Adnatae* [29, 30]. However, the leaf blade of *V. chaerophylloides* and *V. dissecta* were pinnatifid and parted, respectively, and it has been demonstrated that the lobe is not a taxonomic characteristic in the *Viola*. In this study, we identified a close relationship between *V. chaerophylloides* and *V. dissecta*; they form a clade with strong node support. This data indicates that the lobe can be used as a classification characteristic for *Viola*. These data support the taxonomic position of section *Pinnatae*. Previous study suggested that *V. mongolica* and *V. yunnanfuensis* belong to the section *Adnatae* according to the morphological characters of the stipules [29]. However, in our study, the clade composed of *V. mongolica* and *V. yunnanfuensis* was a sister to that *V. raddeana*, and they share the bilobate stigma. Therefore, we suggest that *V. mongolica* and *V. yunnanfuensis* are

positioned within section *Bilobatae*. In addition, we must note that some species in section *Adnatae*, whose stipules are adnate to the petiole should be carefully examined for their stigma in future studies. Phylogenetic analysis based on the cp genomes successfully resolved the relationship between *Viola* sampled species. Due to the overlapped taxonomic characters of sections *Adnatae*, *Bilobatae*, and *Pinnatae*, the ranges need to be further delimited. It is worth noting that a comprehensive consideration of morphological characters is necessary for the phylogenetic study of *Viola*. The monophyly of sections *Adnatae* and related taxa and its taxonomic position need to be further analyzed, to generate more data we plan to conduct further investigations with broad sampling and further more morphological evidence.

Conclusion

In the current study, all *Viola* cp genomes share a highly similar gene content and order. The topologies of ML and BI trees are highly concordant for both complete cp genome sequences and highly diverged sequences. Phylogenetic analysis revealed highly supported for interspecies relationships. Morphological characteristics, cp genome phylogeny, and species-specific variable sites can be applied to distinguish *V. philippica*, the only source plant of “Zi Hua Di Ding”, from other *Viola* species, in particular its adulterants. Furthermore, we propose that the most favorable time for accurate identification of *V. philippica* is the flowering period. This study provides invaluable data for the improvement of species identification and germplasm of *V. philippica* that may facilitate the application of a super-barcode in TCM identification and enable future studies on phylogenetic evolution and safe medical applications.

Methods

Sample collection, DNA extraction, and sequencing

The plastomes of 19 samples were newly sequenced in this study, plus 5 plastid genomes already available from GenBank (<https://www.ncbi.nlm.nih.gov>), for a total of 24 individuals. Plant materials used in this study were collected and deposited at the herbarium of the College of Life Sciences, Shandong Normal University. The sampling newly sequenced species were collected from Shandong Province, China. FSJ and ZXJ undertook the formal identification of the samples (Table 2). No specific permissions were required for the relevant locations/activities and met local policy requirements. Table 2 indicates the detailed voucher and locality information for the newly sequenced species. Total genomic DNA was extracted using a modified cetyltrimethylammonium bromide (CTAB) method [83]. The quality and concentration of the genomic DNA were checked using

Table 2 Summary of the sequencing data for 17 *Viola* species

Species	Collecting locations	GPS	Voucher specimen number	GenBank accession number
<i>Viola acuminata</i> Ledeb.	Zibo, Shandong, China	117°50'E, 36°29'N	SDNU 042910	MW802528
<i>V. chaerophylloides</i> (Regel) W. Becker	Zibo, Shandong, China	117°52'E, 36°19'N	SDNU 042915	MW802529
<i>V. collina</i> Besser	Zibo, Shandong, China	117°51'E, 36°29'N	SDNU 042917	MW802530
<i>V. dissecta</i> Ledeb.	Zibo, Shandong, China	117°53'E, 36°25'N	SDNU 042922	MW802531
<i>V. inconspicua</i> Blume	Zibo, Shandong, China	117°51'E, 36°20'N	SDNU 091832	MW802532
<i>V. mirabilis</i> L.	-	-	-	NC041582
<i>V. monbeigii</i> W.Becker	Qingdao, Shandong, China	120°28'E, 36°06'N	SDNU 3157	MW802533
<i>V. mongolica</i> Franch. 1	Zibo, Shandong, China	117°51'E, 36°29'N	SDNU 050113	MW802534
<i>V. mongolica</i> Franch. 2	Jinan, Shandong, China	117°15'E, 36°55'N	SDNU 706	ON548135
<i>V. patrinii</i> Ging.	Zibo, Shandong, China	117°51'E, 36°29'N	SDNU 042913	MW802535
<i>V. phalacrocarpa</i> Maxim.	Yantai, Shandong, China	121°16'E, 37°29'N	SDNU 890210	MW802536
<i>V. philippica</i> Cav. 1	Jinan, Shandong, China	116°50'E, 36°32'N	SDNU 89588	MZ343563
<i>V. philippica</i> Cav. 2	Jinan, Shandong, China	117°04'E, 36°65'N	SDNU 704	ON548136
<i>V. prionantha</i> Bunge	Zibo, Shandong, China	117°51'E, 36°29'N	SDNU 091825	MW802538
<i>V. raddeana</i> Regel	-	-	-	NC041584
<i>V. variegata</i> Fisch. ex Link	Zibo, Shandong, China	117°51'E, 36°29'N	SDNU 042916	MW802539
<i>V. websteri</i> Hemsl.	-	-	-	NC041585
<i>V. yezoensis</i> Maxim.	Qingdao, Shandong, China	120°28'E, 36°06'N	SDNU 3076	MW802540
<i>V. yunnanfuensis</i> W. Becker 1	Qingzhou, Shandong, China	118°28'E, 36°41'N	SDNU 0022	MW802541
<i>V. yunnanfuensis</i> W. Becker 2	Jinan, Shandong, China	117°04'E, 36°62'N	SDNU 705	ON548137

1.5% agarose gel electrophoresis and the NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific Inc., USA). The total genomic DNA was used for library preparation and paired-end (PE) sequencing by the Illumina Novaseq instrument at Novogene (Beijing, China). The raw data is approximately 2Gb and the insert library size is approximately 350bp. In addition, two individuals for each species of *V. philippica*, *V. mongolica* and *V. yunnanfuensis* were sampled from different locations and labeled *V. philippica* 1, *V. philippica* 2, *V. mongolica* 1, *V. mongolica* 2, *V. yunnanfuensis* 1, and *V. yunnanfuensis* 2 (Fig. 6).

Cp genomes assembly and annotation

We assembled the cp genomes by Organelle Genome Assembler (OGA; <https://github.com/quxiaojian/OGA>) [84]. Annotation was performed by using Plastid Genome Annotator (PGA; <https://github.com/quxiaojian/PGA>) [85]. Geneious v8.0.2 was used for annotation correction [86]. The circular maps for newly sequenced cp genomes were generated using the OGDRAW v1.3.1 [87]. All chloroplast genomes assembled in this study have been deposited in GenBank under accession numbers of MW802528 - MW802536, MW802538 - MW802541, MZ343563, and ON548135 - ON548137. Complete

plastome of three *Viola* species were downloaded from GenBank, including *V. mirabilis* L. (NC_041582), *V. raddeana* Regel (NC_041584), and *V. websteri* Hemsl. (NC_041585) (Table 2). The genome size, GC content, gene number, and intron number of the 17 complete plastome were summarized by using Geneious v8.0.2.

Expansion and contraction of IRs

The expansion and contraction of IRs were analyzed by IRscope (<https://irscope.shinyapps.io/irapp/>) [88], coupled with manual modification. In this study, IR borders and neighboring genes were compared for 17 *Viola* species.

Characteristics of repeat sequences and SSRs

SSR markers are valuable in study of genetic diversity and molecular marker selection. The size and position of the repeat sequences were detected using REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer/>) [89], including forward, reverse, complement, and palindromic repeats within the cp genomes. The following settings were used: (1) Hamming distance of 3; (2) 90% or greater sequence identity; (3) a minimum repeat size of 30bp. Simple sequence repeats (SSRs) in cp genomes were detected using MISA [90], with repeat units set to

≥ 10 for mononucleotide, ≥ 6 for dinucleotide, and ≥ 5 for trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide.

Comparative analysis and divergence hotspot identification

mVISTA (<http://genome.lbl.gov/vista/index.shtml>) [91] is a commonly used comparative cp genome map-drawing web application, but the input file of mVISTA needs to meet the format requirements. A custom Perl script (https://github.com/quxiaojian/Bioinformatic_Scripts/get_mVISTA_format_from_GenBank_annotation.pl) was used to convert GenBank annotation files to mVISTA format files. Then we aligned the complete cp genomes using the Shuffle-LAGAN mode of mVISTA, with *V. acuminata* as reference.

Single nucleotide polymorphism (SNP) mainly refers to DNA sequence polymorphism caused by single nucleotide variation at the genome level. The percentage of parsimony information sites (Pi) of the coding and intergenic regions were calculated by using DnaSP v6.0 [92]. The screening conditions were as follows: (1) sequence length > 200 bp; (2) variable sites and parsimony information sites > 0.

Identification of *V. philippica* and phylogenetic analysis

We blast *V. philippica* with the NCBI database (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) by Blast online. The *Viola* genetic distances were calculated in MEGA, using Kimura's 2-parameter model [93]. The coding and intergenic sequences were individually aligned using MAFFT v7.313 [94] with default parameters, and then manually edited using Geneious v8.0.2. Some species have been misidentified mainly because their violet-colored flowers are the same color as those of *V. philippica*, e.g., *Corydalis bungeana* or *Gueldenstaedtia verna*. To improve the reliability and accuracy of the identification results, we added closely related species of both, *Corydalis tomentella* (NC_060366) and *Oxytropis arctobia* (NC_050861). *Paspalum paniculatum* (MF563367) was set as the outgroup. The ML trees were reconstructed by RAxML v8.0.26 with the GTRGAMMA substitution model and 1000 bootstrap replicates [95]. Bayesian inference of phylogeny was explored using the MrBayes v3.1.2 [96]. Phylogenetic trees were reconstructed based on the following data sets: 1) complete cp genome sequences; 2) highly diverged sequences (*matK*, *ndhF*, *ycf1*, *rpl22*, *rps15*, *ndhA*, *petN-psbM*, *petA-psbJ*, *ccsA-ndhD*, *trnG-UCC-2-trnR-UCU*, *rps8-rpl14*, *trnD-GUC-trnY-GUA*, *trnG-GCC-trnfM-CAU*, *trnH-GUG-psbA*, *psbZ-trnG-GCC*, and *rbcL-accD*).

Selection of species-specific variable sites in protein-coding genes of *V. philippica*

The specific variable sites of protein-coding genes that can distinguish *V. philippica* from other *Viola* species were screened by examining the alignment matrix of 76 genes for 17 *Viola* species. To ensure the accuracy of the four selected specific variation sites in *V. philippica*, Sanger sequencing of PCR amplicons was performed on 14 newly sequenced *Viola* species. Primers were designed using Primer3 v0.4.0 [97]. The coding regions of the *V. philippica* cp genome were used as the templates for primer design. The forward primers for these four fragments were ATTCAATATCTGTATGGGGTAAAG, AATAATTGATCAGATTCGTGGACG, CGTCTAAAA CCTTGGCACAAATCG and AACTCATCCATTTATCGATTACCA and the reverse primers were TGTTACAAATTCATACCAATCCAC, TTTGTGGACTTCTTTATGCACCTC, GGTTCGTTTGAGTAACGGTTGTCA and GTATTTTCGTCATCGTCATTCATTC, respectively (Fig. 4, S6). The target fragments that span these four specific sites were about 300 bp, respectively. The PCR amplification was conducted in a total volume of 50 μ L reaction system containing 3 μ L genomic DNA, 2 μ L forward primer, 2 μ L reverse primer, 4 μ L dNTPs (0.4 mM), and 1 μ L 1.5 units mix including the high-fidelity polymerase ExTaq (TaKaRa). The amplification was carried out with an initial denaturation step at 95 $^{\circ}$ C for 3 min followed by 35 cycles of denaturation for 30 s at 95 $^{\circ}$ C, primer annealing for the 40s at 55 $^{\circ}$ C, and then product extension for 3 min at 72 $^{\circ}$ C. A final extension step was at 72 $^{\circ}$ C for 5 min. The PCR products were verified by gel electrophoresis on a 1.5% agarose gel. All sequences were deposited to GenBank (Table S4).

Morphological anatomy and clustering

We summarized previous studies on the detailed morphological classification of the *Viola* and conducted extensive trait comparisons and statistical analyses, which are presented by hand-drawn ink line diagrams. We mainly focus on traits that were more controversial in previous studies. For example, the lobe or not, length of stipules adnate to petioles, the shape of the leaf blade, stigma type, and fruit shape. The details of plant materials used were observed and measured using microscopes. The statistical analysis was performed using IBM SPSS Statistics v22.0 (SPSS Inc., Chicago, IL, USA). We coded five-character, including lobe, length of stipules adnate to petioles, leaf blade baes, stigma type, and fruit shape. lobe: 0 = entire leaf, 1 = lobed leaf; length of stipules adnate to petioles: 0 = stipule adnate with petioles shorter than 1/2, 1 = stipule

adnate with petioles longer than 1/2; leaf blade baes: 0 = explanate, 1 = reflexed; stigma type: 0 = immarginate, 1 = margined; fruit shape: 0 = capsule ellipsoid, 1 = capsule globose. We mapped these traits onto a cluster analysis tree and represented each trait with a different shape (Table S5).

Abbreviations

TCM: Traditional Chinese Medicine; RSV: Respiratory syncytial virus; ITS: Nuclear intergenic transcribed spacer; ISSR: Inter-simple sequence repeat; LSC: Large single-copy; SSC: Small single copy; IR: Inverted repeat; CTAB: Cetyltrimethylammonium bromide; SNP: Single nucleotide polymorphism; SSRs: Simple sequence repeats; Pi: Nucleotide diversity.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08727-x>.

Additional file 1: Fig. S1. Comparisons of LSC, SSC, and IR region borders among 17 chloroplast genomes. **Fig. S2.** Repeat sequences analysis of 17 cp genomes. **Fig. S3.** Maximum Likelihood (ML) and Bayesian Inference (BI) phylogenetic trees are based on 16 highly diverged regions. **Fig. S4.** The variable sites in *ndhF*, *rpl22*, and *ycf1* of *Viola philippica*. **Fig. S5.** Maximum Likelihood (ML) and Bayesian Inference (BI) phylogenetic trees are based on complete chloroplast genome. **Fig. S6.** Original electrophoretogram for four sequence fragments with unique variable sites in 14 newly sequenced *Viola* species.

Additional file 2: Table S1. A list of genes found in the cp genomes of 17 *Viola* species. **Table S2.** SSR distributed situation in the 17 *Viola* cp genomes. **Table S3.** Results of genetic distance analysis. **Table S4.** Summary of amplified nucleotide sequences and GenBank accession numbers. **Table S5.** Morphological characteristics for analysis.

Acknowledgments

The authors thank Mrs. Xiu-Xiu Guo, Miss Cai-Cai Zhai, Miss Xin-Xin Zhu, Miss Xin Zhang, and Miss Ben-Xia Lou for sample collection.

Authors' contributions

SJF and XJQ conceived and designed the research framework; SJF and XJQ collected and identified the sample; DLC and SQX performed the experiments; DLC analyzed the data, and wrote the paper, with contributions from XJQ. All authors approved the manuscript.

Funding

The study was financially supported by the National Natural Science Foundation of China (31470298), Shandong Provincial Natural Science Foundation (ZR2020QC022), the Shandong Agricultural Science and Technology Fund Project (2019LY002), and the Survey of Herbaceous Plant Germplasm Resources of Shandong Province (2021001). The cost of sample collection and sequencing analysis was funded by this funding source.

Availability of data and materials

The data sets supporting the results of this article are included within the manuscript and its additional files. The complete plastome and amplified nucleotide sequences of 14 *Viola* species was submitted to GenBank (<https://www.ncbi.nlm.nih.gov/>) (accession numbers: MW802528 - MW802536, MW802538 - MW802541, MZ343563, ON548135 - ON548137, and MZ407852 - MZ407907; see Table 2, S4).

All read data are available at the SRA (<http://www.ncbi.nlm.nih.gov/bioproject/751372>) with the BioProject following accession numbers: PRJNA751372. These data will remain private until the related manuscript has been accepted. All other data generated in this manuscript are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 8 December 2021 Accepted: 29 June 2022

Published online: 27 July 2022

References

- Marcussen T, Blaxland K, Windham MD, Haskins KE, Armstrong F. Establishing the phylogenetic origin, history, and age of the narrow endemic *Viola guadalupensis* (Violaceae). *Am J Bot.* 2011;98(12):1978–88. <https://doi.org/10.3732/ajb.1100208>.
- Ballard HE, Jr., Kenneth J. Sytsma, Robert R. Kowal. Phylogenetic relationships of infrageneric groups in *Viola* (Violaceae) based on internal transcribed spacer DNA sequences. *Syst Bot.* 1999;23(4):439–58. <https://doi.org/10.2307/2419376>.
- Chen YS, Yang QE, Ohba H, Nikitin VV. *Violaceae: Flora of China*, vol. 13. In: *Flora of China Editorial Committee*. Beijing: Science Press; Saint Louis: Missouri Botanical Garden Press; 2007. p. 74–111.
- Daly NL, Rosengren KJ, Craik DJ. Discovery, structure and biological activities of cyclotides. *Adv Drug Deliv Rev.* 2009;61(11):918–30. <https://doi.org/10.1016/j.addr.2009.05.003>.
- Parsley NC, Sadecki PW, Hartmann CJ, Hicks LM. *Viola* "inconspicua" no more: an analysis of antibacterial cyclotides. *J Nat Prod.* 2019;82(9):2537–43. <https://doi.org/10.1021/acs.jnatprod.9b00359>.
- Foreman DJ, Parsley NC, Lawler JT, Aryal UK, Hicks LM, McLuckey SA. Gas-phase sequencing of cyclotides: introduction of selective ring opening at dehydroalanine via ion/ion reaction. *Anal Chem.* 2019;91(24):15608–16. <https://doi.org/10.1021/acs.analchem.9b03671>.
- Dang TT, Chan LY, Huang YH, Nguyen LTT, Kaas Q, Huynh T, et al. Exploring the sequence diversity of cyclotides from vietnamese *Viola* species. *J Nat Prod.* 2020;83(6):1817–28. <https://doi.org/10.1021/acs.jnatprod.9b01218>.
- Zuo J, He H, Zuo Z, Bou-Chacra N, Lobenberg R. Erding formula in hyperuricaemia treatment: unfolding traditional Chinese herbal compatibility using modern pharmaceutical approaches. *J Pharm Pharmacol.* 2018;70(1):124–32. <https://doi.org/10.1111/jphp.12840>.
- Xie C, Kokubun T, Houghton PJ, Simmonds MS. Antibacterial activity of the Chinese Traditional Medicine, *zi hua di ding*. *Phytother Res.* 2004;18(6):497–500. <https://doi.org/10.1002/ptr.1497>.
- Sze SK, Wang W, Meng W, Yuan R, Guo T, Zhu Y, et al. Elucidating the structure of cyclotides by partial acid hydrolysis and LC-MS/MS analysis. *Anal Chem.* 2009;81:1079–88. <https://doi.org/10.1021/ac802175r>.
- Wang CK, Colgrave ML, Gustafson KR, Ireland DC, Goransson U, Craik DJ. Anti-HIV cyclotides from the Chinese medicinal herb *Viola yedoensis*. *J Nat Prod.* 2008;71(1):47–52. <https://doi.org/10.1021/np070393g>.
- Yu Q, Liu M, Xiao H, Wu S, Qin X, Lu Z, et al. The inhibitory activities and antiviral mechanism of *Viola philippica* aqueous extracts against grouper iridovirus infection in vitro and in vivo. *J Fish Dis.* 2019;42(6):859–68. <https://doi.org/10.1111/jfd.12987>.
- Du D, Cheng Z, Chen D. Anti-complement sesquiterpenes from *Viola yedoensis*. *Fitoterapia.* 2015;101:73–9. <https://doi.org/10.1016/j.fitote.2014.12.015>.
- Zhou HY, Hong JL, Shu P, Ni YJ, Qin MJ. A new dicoumarin and anticoagulant activity from *Viola yedoensis* Makino. *Fitoterapia.* 2009;80(5):283–5. <https://doi.org/10.1016/j.fitote.2009.03.005>.
- Commission CP. *Pharmacopoeia of the People's Republic of China*. 2020 edn. Beijing: China Medical Science Press; 2020.
- Anonymous. In vitro screening of traditional medicines for anti-HIV activity: memorandum from a WHO meeting. *World Health Organ.* 1989;67(6):613–8. <https://doi.org/10.1093/annhyg/33.4.653>.

17. Hui FL, Min CL, Hon CC, Cheng CW. Effects of herbal medicinal formulas on suppressing viral replication and modulating immune responses. *Am J Chin Med*. 2010;38(1):173–90. <https://doi.org/10.1142/S0192415X10007749>.
18. Ngan F, Chang RS, Tabba HD, Smith KM. Isolation, purification and partial characterization of an active anti-HIV compound from the Chinese medicinal herb *Viola yedoensis*. *Antiviral Res*. 1988;10:107–16. [https://doi.org/10.1016/0166-3542\(88\)90019-8](https://doi.org/10.1016/0166-3542(88)90019-8).
19. Shuang CM, Jiang D, Paul PHB, Xue LD. Antiviral Chinese medicinal herbs against respiratory syncytial virus. *J Ethnopharmacol*. 2002;79:205–11. [https://doi.org/10.1016/S0378-8741\(01\)00389-0](https://doi.org/10.1016/S0378-8741(01)00389-0).
20. Huang SF, Chu SC, Hsieh YH, Chen PN, Hsieh YS. *Viola yedoensis* suppresses cell invasion by targeting the protease and NF- κ B activities in A549 and Lewis lung carcinoma cells. *Int J Med Sci*. 2018;15(4):280–90. <https://doi.org/10.7150/ijms.22793>.
21. Wang YL, Zhang L, Li MY, Wang L-W, Ma CM. Lignans, flavonoids and coumarins from *Viola philippica* and their α -glucosidase and HCV protease inhibitory activities. *Nat Prod Res*. 2018;33(11):1550–5. <https://doi.org/10.1080/14786419.2017.1423305>.
22. Russell NH. Three field studies of hybridization in the stemless white violets. *Am J Bot*. 1954;41(8):679–86. <https://doi.org/10.2307/2438295>.
23. Marcussen T, Jakobsen KS, Danihelka J, Ballard HE, Blaxland K, Brysting AK, et al. Inferring species networks from gene trees in high-polyploid north american and hawaiian violets (*Viola*, *Violaceae*). *Syst Biol*. 2012;61(1):107–26. <https://doi.org/10.1093/sysbio/syr096>.
24. Marcussen T, Heier L, Brysting AK, Oxelman B, Jakobsen KS. From gene trees to a dated allopolyploid network: insights from the angiosperm genus *Viola* (*Violaceae*). *Syst Biol*. 2015;64(1):84–101. <https://doi.org/10.1093/sysbio/syu071>.
25. Migdalek G, Nowak J, Saluga M, Cieslak E, Szczepaniak M, Ronikier M, et al. No evidence of contemporary interploidy gene flow between the closely related European woodland violets *Viola reichenbachiana* and *V. riviniana* (sect. *Viola*, *Violaceae*). *Plant Biol*. 2017;19(4):542–51. <https://doi.org/10.1111/plb.12571>.
26. Marcussen T. Evolution, phylogeography, and taxonomy within the *Viola alba* complex (*Violaceae*). *Plant Syst Evol*. 2003;237(1–2):51–74. <https://doi.org/10.1007/s00606-002-0254-5>.
27. Toyama H, Yahara T. Comparative phylogeography of two closely related *Viola* species occurring in contrasting habitats in the Japanese archipelago. *J Plant Res*. 2009;122(4):389–401. <https://doi.org/10.1007/s10265-009-0235-7>.
28. Yoo KO, Jang SK. Infrageneric relationships of Korean *Viola* based on eight chloroplast markers. *J Syst Evol*. 2010;48(6):474–81. <https://doi.org/10.1111/j.1759-6831.2010.0102.x>.
29. Liang GX, FW. X. Infrageneric phylogeny of the genus *Viola* (*Violaceae*) based on trnL-trnF, psbA-trnH, rPL16, its sequences, cytological and morphological data. *Acta Bot Yunnan*. 2010;32(6):477–88. <https://doi.org/10.3724/SP.J.1143.2010.10122>.
30. Gong Q, Zhou JS, Zhang YX, Liang GX, Chen HF, Xing FW. Molecular systematics of genus *Viola* L. in China. *J Trop Subtrop Bot*. 2010;18(6):633–42. <https://doi.org/10.3969/j.issn.1005-3395.2010.06.007>.
31. Dong S, Ying Z, Yu S, Wang Q, Liao G, Ge Y, et al. Complete chloroplast genome of *Stephania tetrandra* (*Menispermaceae*) from Zhejiang province: insights into molecular structures, comparative genome analysis, mutational hotspots and phylogenetic relationships. *BMC Genomics*. 2021;22(1):880. <https://doi.org/10.1186/s12864-021-08193-x>.
32. Huang R, Xie X, Chen A, Li F, Tian E, Chao Z. The chloroplast genomes of four *Bupleurum* (*Apiaceae*) species endemic to southwestern China, a diversity center of the genus, as well as their evolutionary implications and phylogenetic inferences. *BMC Genomics*. 2021;22(1):714. <https://doi.org/10.1186/s12864-021-08008-z>.
33. Wang Y, Wang S, Liu Y, Yuan Q, Sun J, Guo L. Chloroplast genome variation and phylogenetic relationships of *Atractylodes* species. *BMC Genomics*. 2021;22(1):103. <https://doi.org/10.1186/s12864-021-07394-8>.
34. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: from gene to genome. *Biol Rev*. 2015;90(1):157–66. <https://doi.org/10.1111/brv.12104>.
35. Mishra P, Kumar A, Nagireddy A, Mani DN, Shukla AK, Tiwari R, et al. DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. *Plant Biotechnol J*. 2016;14(1):8–21. <https://doi.org/10.1111/pbi.12419>.
36. Jansen RK, Raubeson LA, Boore JL, Pamphiliis CW, Chumley TW, Haberle RC, et al. Methods for obtaining and analyzing whole chloroplast genome sequences. *Method Enzymol*. 2005;395:348–84. [https://doi.org/10.1016/S0076-6879\(05\)95020-9](https://doi.org/10.1016/S0076-6879(05)95020-9).
37. Zoschke R, Bock R. Chloroplast translation: structural and functional organization, operational control, and regulation. *Plant Cell*. 2018;30(4):745–70. <https://doi.org/10.1105/tpc.18.00016>.
38. Dong W, Liu Y, Xu C, Gao Y, Yuan Q, Suo Z, et al. Chloroplast phylogenomic insights into the evolution of *Distylium* (*Hamamelidaceae*). *BMC Genomics*. 2021;22(1):293. <https://doi.org/10.1186/s12864-021-07590-6>.
39. Dong W, Xu C, Liu Y, Shi J, Li W, Suo Z. Chloroplast phylogenomics and divergence times of *Lagerstroemia* (*Lythraceae*). *BMC Genomics*. 2021;22(1):434. <https://doi.org/10.1186/s12864-021-07769-x>.
40. Guo LL, Guo S, Jiang XB, He LX, Carlsson JE, Hou XG. Phylogenetic analysis based on chloroplast genome uncover evolutionary relationship of all the nine species and six cultivars of tree peony. *Ind Crop Prod*. 2020;153:112567. <https://doi.org/10.1016/j.indcrop.2020.112567>.
41. Tian S, Lu P, Zhang Z, Wu JQ, Zhang H, Shen H. Chloroplast genome sequence of chongming lima bean (*Phaseolus lunatus* L.) and comparative analyses with other legume chloroplast genomes. *BMC Genomics*. 2021;22(1):194. <https://doi.org/10.1186/s12864-021-07467-8>.
42. Wen F, Wu X, Li T, Jia M, Liu X, Liao L. The complete chloroplast genome of *Stauntonia chinensis* and compared analysis revealed adaptive evolution of subfamily *Lardizabaloideae* species in China. *BMC Genomics*. 2021;22(1):161. <https://doi.org/10.1186/s12864-021-07484-7>.
43. Cheon KS, Kim KA, Kwak M, Lee B, Yoo KO. The complete chloroplast genome sequences of four *Viola* species (*Violaceae*) and comparative analyses with its congeneric species. *PLoS One*. 2019;14(3):e0214162. <https://doi.org/10.1371/journal.pone.0214162>.
44. Guo Y, Lin P, Wang M. The complete chloroplast genome of *Viola philippica*. *Mitochondrial DNA B Resour*. 2021;6(4):1494–5. <https://doi.org/10.1080/23802359.2021.1906176>.
45. Shrestha B, Gilbert LE, Ruhlman TA, Jansen RK. Rampant nuclear transfer and substitutions of plastid genes in *Passiflora*. *Genome Biol Evol*. 2020;12(8):1313–29. <https://doi.org/10.1093/gbe/evaa123>.
46. Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, et al. Characterization and dynamics of intracellular gene transfer in plastid genomes of *Viola* (*Violaceae*) and order *Malpighiales*. *Plant Cell*. 2001;13:645–58. <https://doi.org/10.2307/3871412>.
47. Yang J, Park S, Gil HY, Pak JH, Kim SC. Characterization and dynamics of intracellular gene transfer in plastid genomes of *Viola* (*Violaceae*) and order *Malpighiales*. *Front Plant Sci*. 2021;12:678580. <https://doi.org/10.3389/fpls.2021.678580>.
48. Hir HL, Nott A, Moore MJ. How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci*. 2003;28(4):215–20. [https://doi.org/10.1016/S0968-0004\(03\)00052-5](https://doi.org/10.1016/S0968-0004(03)00052-5).
49. Sun W, Zhou ZZ, Liu MZ, Wan HW, Dong X. Reappraisal of the generic status of *Pteroxygonum* (*Polygonaceae*) on the basis of morphology, anatomy and nrDNA ITS sequence analysis. *J Syst Evol*. 2008;46(1):73–9. <https://doi.org/10.3724/SP.J.1002.2008.06120>.
50. dSL A, GP T, GdS K, dNV L, PG M, ON R, et al. The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of *Linaceae* within *Malpighiales*. *Plant Cell Reports*. 2018;37(2):307–28. <https://doi.org/10.1007/s00299-017-2231-z>.
51. Guo YY, Yang JX, Bai MZ, Zhang GQ, Liu ZJ. The chloroplast genome evolution of *Venus slipper* (*Paphiopedilum*): IR expansion, SSC contraction, and highly rearranged SSC regions. *BMC Plant Biol*. 2021;21(1):248. <https://doi.org/10.1186/s12870-021-03053-y>.
52. Press MO, Carlson KD, Queitsch C. The overdue promise of short tandem repeat variation for heritability. *Trends Genet*. 2014;30(11):504–12. <https://doi.org/10.1016/j.tig.2014.07.008>.
53. Williams AV, Miller JT, Small I, Nevill PG, Boykin LM. Integration of complete chloroplast genome sequences with small amplicon datasets improves phylogenetic resolution in *Acacia*. *Mol Phylogenet Evol*. 2016;96:1–8. <https://doi.org/10.1016/j.ympev.2015.11.021>.
54. Dodsworth S, Chase MW, Kelly LJ, Leitch IJ, Macas J, Novak P, et al. Genomic repeat abundances contain phylogenetic signal. *Syst Biol*. 2014;64(1):112–26. <https://doi.org/10.1093/sysbio/syu080>.
55. Chan K, Zhang H, Lin ZX. An overview on adverse drug reactions to traditional Chinese medicines. *Br J Clin Pharmacol*. 2015;80(4):834–43. <https://doi.org/10.1111/bcp.12598>.

56. Xie HQ, Chu SS, Zha LP, Cheng ME, Jiang L, Ren DD, et al. Determination of the species status of *Fallopia multiflora*, *Fallopia multiflora* var. *angulata* and *Fallopia multiflora* var. *var. ciliinervis* based on morphology, molecular phylogeny, and chemical analysis. *J Pharm Biomed Anal.* 2019;166:406–20. <https://doi.org/10.1016/j.jpba.2019.01.040>.
57. Chen S, Pang X, Song J, Shi L, Yao H, Han J, et al. A renaissance in herbal medicine identification: from morphology to DNA. *Biotechnol Adv.* 2014;32(7):1237–44. <https://doi.org/10.1016/j.biotechadv.2014.07.004>.
58. Khalil AAK, Akter KM, Kim HJ, Park WS, Kang DM, Koo KA, et al. Comparative inner morphological and chemical studies on *Reynoutria* species in Korea. *Plants.* 2020;9(2):222. <https://doi.org/10.3390/plants9020222>.
59. Narayani M, Chadha A, Srivastava S. Cycloptides from the Indian medicinal plant *Viola odorata* (banafsha): identification and characterization. *J Nat Prod.* 2017;80(7):1972–80. <https://doi.org/10.1021/acs.jnatprod.6b01004>.
60. Zhang J, Wang LS, Gao JM, Xu YJ, Li LF, Li CH. Rapid separation and identification of anthocyanins from flowers of *Viola yedoensis* and *V. prionantha* by high-performance liquid chromatography-photodiode array detection-electrospray ionisation mass spectrometry. *Phytochem Anal.* 2012;23(1):16–22. <https://doi.org/10.1002/pca.1320>.
61. Zhang C, Su J. Application of near infrared spectroscopy to the analysis and fast quality assessment of traditional Chinese medicinal products. *Acta Pharm Sin B.* 2014;4(3):182–92. <https://doi.org/10.1016/j.apsb.2014.04.001>.
62. Han Y, Sun H, Zhang AH, Yan GL, Wang XJ. Chinmedomics, a new strategy for evaluating the therapeutic efficacy of herbal medicines. *Pharmacol Ther.* 2020;216:107680. <https://doi.org/10.1016/j.pharmthera.2020.107680>.
63. Yu J, Wu X, Liu C, Newmaster S, Ragupathy S, Kress WJ. Progress in the use of DNA barcodes in the identification and classification of medicinal plants. *Ecotoxicol Environ Saf.* 2021;208:111691. <https://doi.org/10.1016/j.ecoenv.2020.111691>.
64. Tnah LH, Lee SL, Tan AL, Lee CT, Ng KKS, Ng CH, et al. DNA barcode database of common herbal plants in the tropics: a resource for herbal product authentication. *Food Control.* 2019;95:318–26. <https://doi.org/10.1016/j.foodcont.2018.08.022>.
65. Newmaster SG, Grguric M, Shanmughanandhan D, Ramalingam S, Ragupathy S. DNA barcoding detects contamination and substitution in North American herbal products. *BMC Med.* 2013;11:222. <https://doi.org/10.1186/1741-7015-11-222>.
66. Yockteng R Jr, Ballard HE, Mansion G, Dajoz I, Nadot S. Relationships among pansies (*Viola* section *Melanium*) investigated using ITS and ISSR markers. *Plant Syst Evol.* 2003;241(3–4):153–70. <https://doi.org/10.1007/s00606-003-0045-7>.
67. Aecyo P, Marques A, Huettel B, Silva A, Esposito T, Ribeiro E, et al. Plastome evolution in the *Caesalpinia* group (Leguminosae) and its application in phylogenomics and populations genetics. *Planta.* 2021;254(2):27. <https://doi.org/10.1007/s00425-021-03655-8>.
68. Li L, Hu Y, He M, Zhang B, Wu W, Cai P, et al. Comparative chloroplast genomes: insights into the evolution of the chloroplast genome of *Camellia sinensis* and the phylogeny of *Camellia*. *BMC Genomics.* 2021;22(1):138. <https://doi.org/10.1186/s12864-021-07427-2>.
69. Nock CJ, Waters DL, Edwards MA, Bowen SG, Rice N, Cordeiro GM, et al. Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnol J.* 2011;9(3):328–33. <https://doi.org/10.1111/j.1467-7652.2010.00558.x>.
70. Shen Z, Lu T, Zhang Z, Cai C, Yang J, Tian B. Authentication of traditional Chinese medicinal herb "Gusuibu" by DNA-based molecular methods. *Ind Crop Prod.* 2019;141:111756. <https://doi.org/10.1016/j.indcrop.2019.111756>.
71. Chen X, Zhou J, Cui Y, Wang Y, Duan B, Yao H. Identification of *ligularia* herbs using the complete chloroplast genome as a super-barcode. *Front Pharmacol.* 2018;9:695. <https://doi.org/10.3389/fphar.2018.00695>.
72. Del-Prado R, Divakar PK, Crespo A. Using genetic distances in addition to ITS molecular phylogeny to identify potential species in the *Parmotrema* reticulatum complex: a case study. *Lichenologist.* 2011;43(6):569–83. <https://doi.org/10.1017/S0024282911000582>.
73. Dong AW, Zhu SW, He Z. Study the concentration of flavonoids of seasonal change in *Viola philippica*. *Inf Tradit Chin Med.* 2004;21(2):27–8. <https://doi.org/10.19656/j.cnki>
74. Li C, Cai C, Tao Y, Sun Z, Jiang M, Chen L, et al. Variation and evolution of the whole chloroplast genomes of *Fragaria* spp. (Rosaceae). *Front. Plant Sci.* 2021;12:754209. <https://doi.org/10.3389/fpls.2021.754209>.
75. Jeong M, Kim JI, Nam SW, Shin W. Molecular phylogeny and taxonomy of the genus *Spumella* (Chrysophyceae) based on morphological and molecular evidence. *Front Plant Sci.* 2021;12:758067. <https://doi.org/10.3389/fpls.2021.758067>.
76. Chervin J, Talou T, Audonnet M, Dumas B, Camborde L, Esquerre-Tugaye MT, et al. Deciphering the phylogeny of violets based on multiplexed genetic and metabolomic approaches. *Phytochemistry.* 2019;163:99–110. <https://doi.org/10.1016/j.phytochem.2019.04.001>.
77. 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;20:261. <https://doi.org/10.1186/s12870-020-02466-5>.
78. Zhang XF, Landis JB, Wang HX, Zhu ZX, Wang HF. Comparative analysis of chloroplast genome structure and molecular dating in Myrtales. *BMC Plant Biol.* 2021;21(1):219. <https://doi.org/10.1186/s12870-021-02985-9>.
79. GP T, Mds G, dSL A, DdO J, MR J, B E, et al. Phylogenetic and evolutionary features of the plastome of *Tropaeolum pentaphyllum* Lam. (Tropaeolaceae). *Planta.* 2020;252(2):17. <https://doi.org/10.1007/s00425-020-03427-w>.
80. Rieppel O. Morphology and phylogeny. *J Hist Biol.* 2020;53(2):217–30. <https://doi.org/10.1007/s10739-020-09600-x>.
81. Talavera M, Balao F, Casimiro-Soriguer R, Ortiz MÁ, Terrab A, Terrab A, et al. Molecular phylogeny and systematics of the highly polymorphic *Rumex bucephalophorus* complex (Polygonaceae). *Mol Phylogenet Evol.* 2011;61(3):659–70. <https://doi.org/10.1016/j.ympev.2011.08.005>.
82. Xi Z, Ruhfel BR, Schaefer H, Amorim AM, Sugumaran M, Wurdack KJ, et al. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proc Natl Acad Sci U S A.* 2012;109(43):17519–24. <https://doi.org/10.1073/pnas.1205818109>.
83. Allen GC, Flores Vergara MA, Krasynanski S, Kumar S, Thompson WF. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat Protoc.* 2006;1:2320–5. <https://doi.org/10.1038/nprot.2006.384>.
84. Qu XJ, Fan SJ, Wicke S, Yi TS. Plastome reduction in the only parasitic gymnosperm parasitaxus is due to losses of photosynthesis but not housekeeping genes and apparently involves the secondary gain of a large inverted repeat. *Genome Biol Evol.* 2019;11(10):2789–96. <https://doi.org/10.1093/gbe/evz187>.
85. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods.* 2019;15:50. <https://doi.org/10.1186/s13007-019-0435-7>.
86. 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. <https://doi.org/10.1093/bioinformatics/bts199>.
87. Greiner S, Lehwark P, Bock R. Organellar GenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 2019;47(W1):W59–64. <https://doi.org/10.1093/nar/gkz238>.
88. Amiryousefi A, Hyvonen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics.* 2018;34(17):3030–1. <https://doi.org/10.1093/bioinformatics/bty220>.
89. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 2001;29(22):4633–42. <https://doi.org/10.1093/nar/29.22.4633>.
90. Beier S, Thiel T, Munch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics.* 2017;33(16):2583–5. <https://doi.org/10.1093/bioinformatics/btx198>.
91. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 2004;32(Web Server issue):W273–9. <https://doi.org/10.1093/nar/gkh458>.
92. Rozas J, Ferrer-Mata A, Sanchez-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):3299–3302. <https://doi.org/10.1093/molbev/msx248>.

93. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* 1980;16(2):111–20.
94. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80. <https://doi.org/10.1093/molbev/mst010>.
95. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30:1312–3. <https://doi.org/10.1093/bioinformatics/btu033>.
96. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 2003;19(12):1572–4. <https://doi.org/10.1093/bioinformatics/btg180>.
97. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3-new capabilities and interfaces. *Nucleic Acids Res.* 2012;40(15):e115. <https://doi.org/10.1093/nar/gks596>.

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

