The chloroplast genome of two medicinal species (Veronica anagallis-aquatica and Veronica unulata) and its comparative analysis with related Veronica species

Yonglin Hai
Dali University
Yan Qian
Dali University
Meihua Yang
Dali University
Yue Zhang
Dali University
Huimei Xu
Dali University
Yongcheng Yang
Dali University
Conglong Xia (long7484@126.com)
Dali University

Article

Keywords:

Posted Date: January 5th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3802816/v1

Additional Declarations: No competing interests reported.

Abstract

Veronica anagallis-aquatica L and *Veronica unulata* Wall are widely used ethnomedicinal plants in China. The two species have different clinical efficacies, while their extremely similar morphology and unclear interspecific relationship makes it difficult to accurately identify them, leading to increased instances of mixed usage. This article reports on the complete genome sequence of chloroplasts of these two species and their related *veronica* species to conduct a comparative genomics analysis and phylogenetic construction. The results showed that the chloroplast (cp) genomes of *Veronica* exhibited typical circular tetrad structures, with total lengths of 149,386 to 152,319 bp, and GC content of 37.9 to 38.1%, and the number of genes was 129–134. The total number of simple sequence repeats (SSRs) in *V. anagallis-aquatica* and *V. unulata* is 37 and 36, while *V. arvensis* had the highest total number of SSRs (56), mainly consisting of A/T single bases. The vast majority of long repeat sequence types are forward repeats and palindromic repeats. Selective pressureanalysis showed that 3 genes were under positive selection. Sequence differences often occur in the non-coding regions of the large single-copy region (LSC) and small single-copy region (SSC), with the lowest sequence variation in the inverted repeat regions (IR). Seven highly variable regions (*trnT-GGU-psbD*, *rps8-rpl16*, *trnQ-UUG*, *trnN-GUU-ndhF*, *petL*, *ycf3*, and *ycf1*) were detected, which may be potential molecular markers for identifying *V. anagallis-aquatica* and *V. unulata* The phylogenetic tree indicates that there is a close genetic relationship between the genera *Veronica* and *Neopicrorhiza*, and *V. anagallis-aquatica* and *V. unulata* are sister groups. The chloroplast genome data of nine *Veronica* species provides important insights into the characteristics and evolution of the chloroplast genome of this genus, as well as the phylogenetic relationship of the genus *Veronica*.

Introduction

Veronica anagallis-aquatica L is a genus of *Veronica* in the family Plantaginaceae, with a long history of ethnomedicinal use, widely distributed in China throughout its provinces north and southwest of the Yangtze River, and mainly in temperate regions of Asia and Europe[1]. Its fruit is abnormally swollen due to insect parasitism, and the whole plant is harvested, washed, and used fresh or dried before the parasitic insects escape. This plant with insect galls is used clinically in traditional Mongolian medicine (TMM) to treat edema and arthralgia due to its diuretic and edema-reducing effects[2]. *Veronica undulata* Wall is widely distributed throughout China. It has blood-activating and pain-relieving effects as mentioned in the Processing Standards of Traditional Chinese Medicine (TCM) Decoction Pieces in Shanghai (2018 Edition), and is clinically used for treating hemoptysis, stomach pain, rheumatic pain, dysmenorrhea, carbuncles, and swelling[3]. Previous studies have shown that *V. anagallis-aquatica* and *V. undulata* have different efficacies and applications, but have very similar morphological characteristics. Additionally, traditional morphological identification methods cannot effectively identify them. Ancient materia medica make limited distinctions between these two species, which cause confusion since they are closely related species in the same family and genus (Fig. 1).

Approximately 250 species of plants in the genus *Veronica* exist worldwide, mainly distributed throughout Eurasia. Moreover, 61 of these species are spread throughout China, mainly in its southwestern

mountains, while up to 27 species are currently recorded as having medicinal properties[4]. Research on the chemical constituents of this genus has resulted in phytochemists isolating and identifying over 260 compounds from its plants, including iridoid glycosides, p-Phenylethanoside, and flavonoids[5]. Research on modern pharmacological activity shows that this genus has anti-inflammatory, antibacterial, antioxidant, anti-tumor, *in vitro* coagulation, antitussive, and liver protective effects[6–8]. The traditional identification method of *Veronica* plants is mainly based on the morphological characteristics of their leaf shape, petiole, flower stalk, fruit stalk length, flower color and size, fruit shape, petiole root structure, as well as the number of stomata on the upper and lower epidermis of its leaves[9,10]. However, each distribution area has a different environment since some species are widely distributed. To adapt to local ecological environments, the related traits of these species are discontinuous between each group. In addition, due to the short time of species differentiation and gene exchange between closely related species, some species have a vague interspecific relationship. Therefore, it is difficult to accurately identify the species of this genus only by relying on traditional morphological methods. In addition, the Angiosperm phylogeny classification of flowering plants (APGIV)[11] system removed *Veronica* from Scrophulariaceae and moved it to Plantaginaceae, Many scholars have studied and analyzed it by considering many aspects such as biochemistry, molecular biology, and pharmacology. However, the determination of the phylogenetic position of the genus *Veronica* remains very controversial, and scholars have expressed diverse views on this issue. *Veronica* plants have significant potential ecological[12,13], environmental[14] and medicinal[5,15–17] value. Therefore, their taxonomic issues must be further explored.

As sequencing technology, molecular biology, molecular systematics, and other disciplines develop, an effective method to solve the issue of systematic interspecific relationship and classification is to use suitable molecular markers to analyze the interspecific and intraspecific sequences of related species at the DNA level[18–20]. Chloroplasts (cp) are an essential organelle in plant cells intended for photosynthesis and other functions, including the synthesis of starch, fatty acids, pigments, and amino acids[21]. Recently, the cp genome has provided more information about genetic variations and significantly improved the phylogenetic resolution of species. It has been widely applied when developing molecular markers to classify medicinal plants and implemented in phylogenetic studies[22,23]. In most angiosperms, this genome is usually double-stranded and circular[24]. It consists of a small single copy region (SSC), a large single copy region (LSC), and a pair of inverted repeat regions (IR), which separate the SSC and LSC[25]. The cp genome has a highly conserved structure, gene content, and typically low level of DNA sequence variation compared to nuclear and mitochondrial genomes, and many studies have reported that it is expected to serve as a super barcode to identify closely related species[26,27].

Our study aims to: (i) contribute a fully-sequenced cp genome of *Veronica* species, of which *V. anagallis-aquatica* was reported for the first time, and expand our understanding of the overall structure of these genomes. (ii) *V. anagallis-aquatica* and *V. unulata* have very similar morphologies. By comparing and analyzing the structural characteristics of the whole cp genome of these two species, potential molecular markers were screened to identify them. (iii) To explain the phylogenetic evolution of the two species and

their genus, and to provide a reference for the traditional taxonomy as well as further application and conservation of the medicinal plants in *Veronica*.

Results

The structure and characteristics of the cp genome of Veronica

V. undulata (OQ564497.1) was selected as the reference gene of the *V. anagallis-aquatica* and *V. undulata* for annotation. The complete cp genomes of these species were annotated to obtain 130 genes, including 85 protein-coding genes, 37 tRNA genes, and 8 rRNA genes, which were mainly categorized as photosynthesis genes, replication genes, unknown functional genes, and acetyl-CoA carboxylase subunits (*accD*), mature enzyme genes (*matK*), transcription initiation factors (*infA*), and other genes. In addition, 17 genes were found to contain introns in the cp genomes of both species (Table S1), of which nine protein-coding genes and six tRNA-coding genes contained one intron, while the *ycf3* and *clpP* protein-coding genes contained two introns. The length difference between *Veronica* is significant, ranging from 149386 to 152319 bp, and the cpDNA showed a typical tetrad ring structure (Fig. 2). These both consisted of a pair of inverted repeats (IR) and a large single copy region as well as a small single-copy. The difference in total GC content among the 9 species of the *Veronica* genus is relatively small, ranging from 37.9–38.1%, and the number of genes was 129–134. Significantly, The GC content in the IR region of 9 species is higher than that in the LSC and SSC regions, with the SSC region having the smallest content (Table 1).

Analysis of codon preference

The analysis of codon bias revealed that *V. anagallis-aquatica* and *V. undulatae* had 21,283 and 18,965 codons (including stop codons), respectively. Among all amino acid codons, the largest number of codons encoding leucine (Leu, L) was 2277 and 2281, followed by isoleucine (Ile) with 1793 and 1755, respectively; Cysteine (Cys, C) had the lowest number of codons with 226 and 227, respectively. There were 30 codons with RSCU > 1 in the cp genomes of both species, accounting for about 50% of the total codons. Except for codon UUG, all of them ended with A or U bases. The codon preference was also shown for 18 amino acid codons with RSCU ≥ 1.5, such as leucine (Leu) with a preference for codon UUA (RSCU = 1.95), arginine (Arg) with a preference for codon AGA (RSCU = 1.73), and termination codon with a preference for UAA (RSCU = 1.81). There were 32 codons with RSCU < 1, 29 of which were G/C-terminated codons. These results suggest that the cp genomes of the two species showed a preference for A/U-terminated codons over G/C-terminated codons. Both codons encoding methionine (Met, M) and tryptophan (Trp, W) had only one codon with a RSCU value of 1 and no preference for use (Table S2). In addition, the cp genome codon usage of medicinal *V. anagallis-aquatica* and *V. undulata* was similar. Compared with the other eight species, there was a significant difference in the use of cp genome codons in *V. polita* (Fig. 3).

Repeated sequences and SSRs analysis

Repeated sequences of length greater than or equal to 30 bp are called long repeats. These facilitate cp genome rearrangement and increase a population's genetic diversity. The distribution of the number of long repeats showed that 33–40 long repeats were identified in the cp genomes of 9 species. *V. eriogyne* had the largest number of long repeat sequences, while *V. arvensis* had the smallest number of long repeat sequences. The number of long repeat sequences of *V. anagallis-aquatica* and *V. undulata* is the same, both of which were 37 (Fig. 4C, D). The vast majority of long repeat sequence types are forward repeats and palindromic repeats. In 9 species, most of the long repeat sequences were 30–39 bp in length, and both had at least 1 long repetitive sequence more than 60 bp, respectively. No reverse repeat or complementary repeat sequence was identified in other species except for *V. eriogyne*, *V. nakaiana*, and *V. ovata subsp*. kiusiana (Table S3). SSR analysis showed that 36 SSR loci were detected in *V. anagallis-aquatica*, including 21 mononucleotides, 4 dinucleotides, 7 trinucleotides, 3 tetranucleotides, and 1 pentanucleotide, while 37 SSR loci were detected in *V. undulata*, including 22 mononucleotides, 4 dinucleotides, 7 trinucleotides, 3 tetranucleotides, and 1 pentanucleotide. *V. arvensis* had the highest total number of SSRs (56). No hexanucleotide repeats were found in the nine species, and single nucleotides often end in A/T (Fig. 4A, B and Table S4).

Comparative analysis of the cp genomes of Veronica

The mVISTA analysis showed that the variation of non-coding regions in 9 cp genome sequences was higher than that of conserved protein coding regions (Fig. 5). The variation in the LSC and SSC regions was significantly greater than that in the IR region, while the rRNA genes (*rrn4.5*, *rrn5*, *rrn16*, *rrn23*) were highly conserved and had almost no variation. The sequence similarity between *V. anagallis-aquatica* and *V. undulata* was high, while the remaining seven species all differed significantly from the reference sequence. The highly differentiated regions in the 9 cpDNAs were mainly located in the non-coding regions (IGS). For example, highly variable *trnT-GGU-psbD*, *ycf3*, and *rps8-rpl16* fragments could be used as potential molecular markers to identify *V. anagallis-aquatica* and *V. undulata*.

Homologous collinearity analysis was performed on the chloroplast genome sequences of 9 *Veronica* species, and the results showed that the purple region is a rearrangement region. The same colored blocks connected by the same line represent similar gene fragments of different species in the *Veronica* genus. All genes fragments are located above the axis, indicating that gene inversion has not occurred. The gene arrangement order of the chloroplast genomes of 9 species is consistent, with good collinearity, and there is no structural inversion phenomenon inside (Fig. 6). In terms of length, *V. undulata* is related to *V. agrestis* and *V. subsp Kiusiana* is relatively close. In terms of location, *V. anagallis-aquatica*, *V. undulata, V arvensis*, and *V. eriogyne* has 5 common regions. It is worth noting that there were no significant deletions of gene fragments in the species of the *Veronica* genus, indicating that the cp genome sequences of *Veronica* plants are highly similar and no inversion or gene rearrangement of large fragments was detected. Overall, these chloroplast genomes are essentially conserved and collinear.

IR boundary analysis of cp genome in Veronica

The study of the shrinkage and expansion of the cp genome IR, LSC, and SSC boundary regions of nine species of *Veronica* indicates that the length of its cp genome is significantly different (149386 bp-152319 bp). The *rpl2*, *rps19*, and *rpl22* genes were distributed in the boundary LSC/IRb, and these nine species did not expand in this boundary. The *ycf1* and *ndhF* genes were distributed in the border IRb/SSC, and both were expanded. Notably, no *ycf1* gene was detected at the IRb/SSC boundary in *V. anagallis-aquatica, V. undulata*, and *V. agrestis* (Fig. 7). The three genes of *rpl2*, *trnH*, and *psbA* were distributed in the boundary IRa/LSC, and no genes crossed this region. The SSC/IRa boundary of all 10 sequences was the *ycf1* gene, and expansion occurred. The length of the *ycf1* gene of *V. anagallis-aquatica* and *V. undulata* at this boundary was the same (5304 bp), while the length of the SSC region of the other seven species of *Veronica* was 4007–4093 bp. Similarly, the *rps19* gene expansion occurred in the LSC/IRb boundary for *V. ovata subsp. kiusiana* and *V. nakaiana*, while the other seven species of *Veronica* were located in the LSC region, and the distance from the LSC/IRb boundary was 14 bp.

Nucleotide polymorphism analysis of the cp genome

The nucleotide polymorphism (Pi) results (Fig. 8A) showed that the Pi values of *V. anagallis-aquatica* and *V. undulata* were detected in the range of 1-151418 bp (Table S5). The LSC and SSC regions showed higher variability compared to the IR region, which was consistent with the results of mVISTA analysis, indicating that the IR region behaved conservatively. In addition, the variation of these genomes was detected within a certain interval. Three highly variable regions were obtained, with Pi values greater than 0.007, including two intergenic regions (*trnQ-UUG* and *trnN-GUU-ndhF*) and two coding protein region (*ycf1* and *petL*), which can be applied as hypervariable fragments to identify the above two species. The Pi of nine species of *Veronica* (Fig. 8B) showed that 6 highly variable fragments with pi values greater than 0.066, including three intergenic regions (*ndhK-ndhC*, *psal-ycf4*, and *ndhA-ndhH*), and four genes (*rps16*, *ycf3*, *ndhF*, and *ycf1*). These regions can be potential molecular markers to identify nine different species of the genus *Veronica*.

To determine whether the protein coding genes of the genomes of the other eight species of the *Veronica* genus are under selection pressure, non synonymous (Ka) and synonymous (Ks) substitution rates and Ka/Ks values were calculated using the reference sequence of the *V. undulata* (OR187860). The selection pressure analysis results indicate that the average Ka/Ks values of protein genes in the eight genomes is 0.12309. Among all genes, except for *ndhC*, *rps12*, and *rps7* (Ka/Ks > 1) in some species, the Ka/Ks values of other genes are less than 1, indicating that the genes are in purification selection (Fig. 9) (Ka/Ks = 1, neutral segment, Ka/Ks < l, purification selection, and Ka/Ks > 1 positive selection). The Ka/Ks values of most genes are less than 0.6, indicating a clear purification selection pattern. Among them, up to 12 genes have Ka/Ks values of 0, indicating that these genes may be under strong purification selection pressure (Table S6)

Phylogenetic and genetic distance analysis

In the present study, we constructed a ML tree containing 37 whole cp genome sequences to determine the phylogenetic position of *V. anagallis-aquatica* and *V. undulata* (Fig. 10). As shown in this figure, most of the nodes received high support values in the ML tree, which is consistent with the support values of previous studies. The genus *Veronica* can be divided into three clades: (i) clade A, including *V. anagallis-aquatica* and *V. undulata*; (ii) clade B, including *V. polita* (NC_060687.1), *V. persica* (OQ564496.1), *V. agrestis* (NC_068050.1), *V. chamaedrys* (NC_068051.1), *V. arvensis* (ON461915.1), and *V. eriogyne* (NC_058571.1); (ii) clade C, including *V. nakaiana* (NC_031153.1) and *V. ovata subsp. Kiusiana* (MT671999.1). In addition, *Veronica* is clustered into one large clade with related species of the family Plantaginaceae, such as *Neopicrorhiza*, *Lagotis*, and *Veronicastrum*, revealing the close relationship between their genera. The relationship between the genera *Veronica* and Scrophulariaceae is far from *Scrophularia* and *Verbascum*, but close to the five genera of Plantaginaceae.

MEGA X was used to analyze the genetic distance of the whole chloroplast genome of *Veronica* (Table 2). The results showed that the average genetic distance between the *Veronica* species was 0.00040 ~ 0.042169, indicating that the genetic relationship between the samples of *Veronica* was quite different. We found that the average genetic distance among *V. agrestis* and *V. polita* was the smallest, indicating that their genetic relationships were close. The average genetic distance difference between *V. anagallis-aquatica* and *V. undulata* is also small, indicating a close genetic relationship between them. Except for the four species of plants in the genus *Veronica* mentioned above, there is a significant difference in the average genetic distance between other species.

Discussion

The cp genome structure and comparative analysis of the Veronica genus

The classification and identification of plants in the genus *Veronica* has tended to consider their morphological characteristics. However, geographical and ecological factors may continually change these characteristics in species of the genus *Veronica*[28,29]. Recent taxonomy studies have utilized cp genomes to assess the genetic relationship between related species[30–33]. The basic groups of nine species in this genus range in size variations from 149386 to 152319bp, indicating that each of these species has significantly different cp gene sizes[34–37]. A comparative analysis of the cp genomes in *V. anagallis-aquatica* and *V. undulata* showed that they contain highly conserved structures and genes as well as the same protein-coding genes, tRNA, and rRNA. When using Illumina high-throughput sequencing and comparative genomics to analyze these cp genomes, the results demonstrated that some highly variable regions of these genomes can be correlational study-specific DNA barcodes for identifying species and analyzing phylogenetic relationships.

GC content is an important indicator for determining phylogenetic interspecific relationships[38]. The total GC content, GC content in the IR, LSC, and SSC regions of *V. anagallis-aquatica* and *V. undulata* were consistent, while the other seven types of GC content had obvious differences. In addition, the *rps19* and *trnH* genes of both species were 14 and 1 bp away from the LSC/IR boundary, the *ndhF* and *ycf1* genes

were both expanded, and the gene length was consistent, revealing the close relationship between these two species. In addition, different IR boundaries exist between other species of *Veronica*, and the fluctuation of these boundaries is the main reason for interspecific differences[39]. Therefore, the GC content and boundary of the IR region to a high extent indicate that there is a close genetic relationship between *V. anagallis-aquatica* and *V. undulata*, but a distant phylogenetic relationship with other species of the genus *Veronica*.

Repeat sequence and codon bias analysis

SSRs are repetitive sequences of 1–6 bp in length in the genome. These sequences are highly polymorphic, and those in non-coding regions are more variable than those in coding regions[40–42]. Currently, SSR analysis of cp genomes is widely applied in plant classification, biogeography, and population genetics[43,44]. We detected a total of 36 and 37 SSRs in the cp genomes of *V. anagallis-aquatica* and *V. undulata*, respectively. Most of the SSRs in both species are A/T mononucleotides. These results indicate that the cp genome SSRs consist mostly of adenine (A) or thymine (T) repeats, and few contain tandem guanine (G) orcytosine (C) repeats[45]. This conclusion resembles the results of studies on three Veroniceae species (*Veronica nakaiana*, *Veronica persica*, *Veronicastrum sibiricum*), most of which are A/T single nucleotide repeats[37]. Long repeat analysis of both species revealed only two repeat types (forward and palindromic), with the same number of repeats. It is noteworthy that *V. anagallis-aquatica* has one more forward repeat (F) than *V. undulata*, while *V. undulata* has one more palindrome repeat than *V. anagallis-aquatica*.

It is widely known that codon usage bias occurs in different organisms. Codon preferences are influenced by natural selection, base mutations, and genetic drift, are the result of long-term evolution of species in response to environmental selection, and affect the expression of mRNA and protein levels in the genome[46–50]. The most abundant amino acid in both species is leucine (Leu), with an average of 2,279, which is also common in other angiosperm species[51,52]. Meanwhile, based on the results of other angiosperm cp genomes, our study demonstrated that most codons ending in A/U have RSCU values higher than 1, which may be caused by a composition bias towards high A/T ratios[53,54]. Therefore, in the future, when designing the exogenous gene vector in Boragina spp. cp genetic engineering, the optimal codon ending with an A/U base is selected, which can appropriately improve the expression efficiency of exogenous genes. In addition, codon usage preferences were found as very similar in the cp genomes of these two species. These results may indicate that these two species experienced considerable environmental stress during their evolution.

Phylogenetic analysis and exploration of phylogenetic position

The cp genome is a valuable molecular tool in phylogenetic studies since it has a relatively conserved gene number and sequence. Furthermore, its taxonomic and evolutionary biological advantages, such as less susceptibility to recombination, have revitalized traditional fields such as taxonomy and evolutionary

biology, and have extensively developed biological taxonomy[55]. For example, Zhang et al. (2018) applied the cp genome sequence of *Leonurus japonicus* to analyze its phylogenetic position, and their results showed that the genus *Leonurus* is more closely related to the genus *Stachys* in *Lamioideae*[56]. Cui et al. explored the phylogenetic position of *Zingiber officinale* in *Zingiberaceae* based on the published cp genome sequences of *Zingiberaceae* species. The results showed that *Zingiber* is a sister branch of *Kaempferia*[57]. Guo et al. first reported the cp genome sequence of *Schisandra chinensis*, and developed ML and BI phylogenetic trees to indicate that the genus *Schisandra* forms a sister group with the genus *Illicium*[58].

Veronica is considered a relatively complex taxonomic group at both the morphological and molecular levels due to its large number of species, morphological similarities, and wide distribution areas. Previously, Wang et al. applied phylogeny to comparatively study perennial and annual plants of *Veronica*, and found that the ancestors of *Veronica* were likely perennial plants[59]. Albach et al. used ribosomal and plastid DNA sequences to reveal the phylogenetic relationships of the genus *Veronica* and its subgenera, finding that the Asian and European perennial species are monophyletic sister groups[60]. Ellmouni et al. phylogenetically analyzed the genus *Veronica* by using plastid and ribosomal DNA[61]. Xu et al. (2018) used the cp gene matK to divide 16 species of *Veronica* into two independent parts by establishing a phylogenetic tree, among which *Veronica hederifolia* was the most primitive[62]. In this study, phylogenetic trees were constructed by establishing the ML as well as maximum parsimony (MP) (Fig. S1), along with implementing the neighbor joining (NJ) method (Fig. S2). The three phylogenetic trees showed the same evolutionary relationships, of which the genetic relationship between *Veronica* and *Neopicrorhiza* is the closest, followed by *Lagotis* and *Veronicastrum*, while that with related genus species under Scrophulariaceaes is the farthest. Therefore, based on the phylogenetic relationship of cp genome, it is reasonable to classify Veronica into Plantaginaceae. In addition, two closely related species of *V. anagallis-aquatica* and *V. undulata* clustered in one clade with a bootstrap support value of 100%. This indicates that the cp genome can be used as a super barcode to identify *V. anagallis-aquatica* and *V. undulata*. Nevertheless, in this study, only 1 individuals were detected for each species, and further sample collection is needed for validation in the future to evaluate the discriminative power and effectuality of these locies in distinguishing diverse *Veronica* species.

Conclusion

The two species of *V. anagallis-aquatica* and *V. undulata* have different functions and very similar morphologies. It is difficult to identify them by relying on traditional morphological classifications. In the present study, we compared and analyzed the cp genome characteristics of 9 species of the genus *Veronica*. The results indicate that the cp genome of the nine *Veronica* species exhibits a typical tetrad structure, with total lengths of 149,386 to 152,319 bp, and GC content of 37.9 to 38.1%. Comparative analysis showed that the cp genome sequences of *V. angalallis-aquatica* and *V. undulata* were less different, while the sequences of these two species were significantly different from those of the other seven genera of *Veronica*. Seven highly variable loci (*trnT-GGU-psbD*, *rps8-rpl16*, *trnQ-UUG*, *trnN-GUU-ndhF*,

petL, *ycf3*, and *ycf1*) were identified as potential molecular markers to discriminate between these two species. The phylogenetic tree showed that *V. anagallis-aquatica* and *V. undulata* had a close genetic relationship and were sister groups. In conclusion, this study expanded the genomic resources of the genus *Veronica* as well as provided valuable information to support the phylogenetic analysis of this genus along with the identification of *V. anagallis-aquatica* and *V. undulata*.

Materials and Methods

Materials, DNA extraction and sequencing

The plant materials used in this study were obtained from the field and were licensed for sample collection. Collection of plant materials is also in accordance with institutional, national or international guidelines. Fresh leaves of *V. angagallis-aquatica* (OR187859) were collected from Shuhe Town, Lijiang City, Yunnan Province, China (100°12'39"E, 26°55'52"N; elevation: 2,403m), and those of *V. undulata* (OR187860) were collected from Nanjian City, Yunnan Province, China (100°29'59.4"E, 24°55'8.08"N; elevation: 1,927m). The samples were identified by Professor Cong-long Xia (College of Pharmacy, Dali University) and Professor Wen-guang Yang (Kunming Institute of Botany, Chinese Academy of Sciences). The voucher specimens were preserved in the Herbarium of Dali University (Voucher Code of *V. angagallis-aquatica*: BSKM202201 and Voucher Code of *V. undulata*: SKM202202). Total DNA was extracted using the E.Z.N.A® Plant DNA kit (OMEGA). The chloroplast genome data of other *veronica* species were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/) Extracted DNA was checked for quality and integrity by 1% agarose gel electrophoresis, and also for concentration and content using TBS380 Picogreen (Invitrogen). Then, this DNA was fragmented to 300–500 bp using Covaris M220 sonication. The fragments were then purified using a TruSeq™ Nano DNA Sample Prep Kit for trimming, 3'-end adenylation, and ligation index adaptation. The sequencing library was created by conducting PCR amplification of appropriately-sized fragments, whose library was sequenced with paired-end (2×150 bp) using the Illumina NovaSeq 6000 platform (Shanghai Biozeron Biotech Co, Ltd).

Genome assembly and annotation

Afterwards, the quality of paired-end Illumina[63] reads was assessed with FastQC, (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and the low-quality reads were removed by using Fastp. NOVO-Plasty[64] was used to assemble the original reads into a complete cp genome. BWA was implemented to align high-quality reads to the cp genome sequence, and manual inspection was performed in the IGV to ensure that the assembly was correct. The complete cp genome sequence was annotated using CPGAVAS[65] and GeSeq[66], while the tRNA gene was identified and manually corrected by running tRNAscan-SE[67]. The cp genome map was drawn online using IRscope (https://irscope.shinyapps.io/Chloroplot/).

Analysis of codon preference

Codons with length < 300 bp and duplicates were removed, and CodonW (http://codonw.sourceforge.net) was applied to analyze codon usage[68]. Preference values of codons were obtained by calculating relative synonymous codon usage (RSCU). A codon RSCU value equal to 1.00 indicated that the codon lacked a preference, while an RSCU value greater than 1.00 indicated that the codon was used more frequently, and vice versa.

Characterization of Repeat Sequences and SSRs

We used the online software REPuter[69] to analyze the repeat sequences of cp genomes of the *Veronica* species. The parameter settings were as follows: The maximum calculated repeats were set to 5,000, the minimal repeat size was set to 30, the Hamming distance was set to 3, and other default parameters were set. MISA detected markers of SSRs in the *Veronica* species, setting parameters as 10 for mononucleotide SSRs, 5 for dinucleotide SSRs, 4 for trinucleotide SSRs, as well as 3 each for tetranucleotide, pentanucleotide, and hexanucleotide SSRs[70].

Comparative analysis of cp genomes

By operating the online program mVISTA (https://genome.lbl.gov/vista/mvista/submit.shtml) to explore the differences between cp genomes within the genus *Veronica*, using the *V. undulata* (OQ564497.1) as the reference sequence and setting the model of Shuffle-LAGAN to visualize and analyze the mVIST structural variation map for nine *Veronica* species. The collinearity analysis of chloroplasts is a global alignment using the Mauve plug-in[71] in Geneious9.0.2 to perform collinearity analysis to detect gene rearrangements in the genome. The IRscope online tool (https://irscope.shinyapps.io/irapp/) was implemented to compare and analyze the structural variations of the IR region. The nucleotide diversity index Pi was calculated by sliding window analysis while using DNAsp software for *V. anagallis-aquatica*, *V. undulata*, and other *Veronica* species. The interspecific high variation sequences (hotspots) were screened based on the analysis results. The window length was set to 600 bp, and the step size was 200 bp[72]. KaKs_Calculator software[73] was used to calculate the substitution rate (Ka and Ks) of protein-coding genes.

Phylogenetic and genetic distance analysis

Three species of *Veronicastrum*, 6 species of *Plantago*, 1 species of *Neopicrorhiza*, 1 species of *Bacopa*, 2 species of *Lagotis*, 9 species of *Veronica*, 6 species of *Scrophularia*, and 5 species of *Verbascum* were selected from the NCBI database (https://www.ncbi.nlm.nih.gov). Achimenes cettoana (NC_050917.1) and Achimenes erecta (NC_051524.1) were chosen as outgroups, and all sequences were subjected to multiple alignment of sequences by using MAFFT 7.0[74]. Phylogenetic analysis was carried out using three tree-building methods, Neighbor-joining (NJ), Maximum likelihood (ML), and Maximum Parsimony (MP). Using MEGAX[75]software to establish NJ and MP trees, parameter settings are: bootstrap repeats 1000 times, NJ tree model selection Kimura2-parameter, MP tree search method selection Subtree-Pruning-Regrafting (SPR). At the same time, MEGA X software was used to calculate the average genetic distance between species by Kimura 2-parameter (K2-P) model. The ML phylogenetic tree of *Veronica*

and other species was constructed using IQtree software[76]. The best nucleotide substitution model was screened using IQtree's built-in model finder. The bootstrap repetition value was 1000. After the program was completed, the software MEGAX was used to process the image of the ML tree.

Declarations

Ethics declarations

This article does not contain any studies with animals performed by any of the authors. The plant materials used in this study were sourced from the wild in Dali, Yunnan Province, and the collection process complied with the regulations of the college and relevant local laws and regulations.

Data availability

The complete chloroplast genomes reported in this paper have been submitted to the NCBI database under the following accession numbers: OR187859 for *V. angagallis-aquatica* and OR187860 for *V. undulata*.

Acknowledgments: We are grateful to Zhen Li, Yuze Liang and Xue Wang for their valuable discussions. We also thank the Shanghai Biozeron Biotechnology company for sequencing. Our sincere thanks are also to the anonymous reviewers for their comments and suggestions.

Author contributions: C.X. and Y.Y. designed the research. Y.Q. and Y.Z. collected the data. Y.H., M.Y. and H.X. conducted data analysis. Y.H. prepared the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Major Projects of Science and Technology Plan of Dali state [D2019NA03] and Li Jian Expert Workstation of Yunnan Province [2020-1].

Competing interests

Te authors declare no competing interests.

Additional information

Supplementary Information Te online version contains supplementary material available at

Correspondence and requests for materials should be addressed to Y.Y. or C.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional afliations.

References

1. Wu, S. X. *Mongolian Medicine of Chinese Materia Medica*, Shanghai Science and Technology Press (Shanghai, 2004).
2. Xiao, Q. X., Peng, X. J. & Li, Y. B. Optimize the Extraction Technology of Total Favonlids in Veronica Anagallis-Aquatica L.by Orthogonal Experiment Method. Hub Agricultural Ences. 55, 5588–5590 (2016).
3. Administration, Shanghai City Drug. Processing Standard of TCM Decoction Pieces in Shanghai, 1st ed. Shanghai Science and Technology Press: shanghai, pp. 327. ISBN 978-7-5478-4333-8 (2019).
4. Chen, J., Yang, C. & Ling, W. Research Progress On Pharmacological Activity of Medicinal Plants of *Veronica*. Research and Practice On Chinese Medicines. 27, 77–80 (2013).
5. Xue, H. B., Chen, K. X., Zhang, L. Q. & Li, Y. M. Review of the Ethnopharmacology, Phytochemistry, and Pharmacology of the Genus *Veronica*. The American Journal of Chinese Medicine. 47, 1–29 (2019).
6. Jensen, S. R., Gotfredsen, C. H., Harput, U. S. & Saracoglu, I. Chlorinated Iridoid Glucosides From Veronica Longifolia and their Antioxidant Activity. J. Nat. Prod. 73, 1593–1596 (2010).
7. Zhang, R. B. & Dou, D. L. Research Progress On Medicinal Plants of *Veronica* in China. Science & Technology Information. 224–242 (2009).
8. Saracoglu, I., Harput, U. S., Inoue, M. & Ogihara, Y. New Phenylethanoid Glycosides From *Veronica Pectinata* Var. *glandulosa* and their Free Radical Scavenging Activities. Chem. Pharm. Bull. 50, 665–668 (2002).
9. Liu, G. & Li, Q. Morphological Characteristics and Anatomical Structures of Vegetative Organs of *Veronica*. Guizhou Agricultural Sciences. 04, 55–56 (2011).
10. Taskova, R. M., Albach, D. C. & Grayer, R. J. Phylogeny of *Veronica–A* Combination of Molecular and Chemical Evidence. Plant Biol. 6, 673–682 (2004).
11. None. An Update of the Angiosperm Phylogeny Group Classification for the Orders and Families of Flowering Plants: APG IV. Bot. J. Linnean Soc. 181, 1–20 (2016).
12. Kroflic, A., Germ, M., Golob, A. & Stibilj, V. Does Extensive Agriculture Influence the Concentration of Trace Elements in the Aquatic Plant *Veronica Anagallis-Aquatica*? Ecotox. Environ. Safe. 150, 123–128 (2018).
13. Saygideger, S. Sorption of Cadmium and their Effects On Growth, Protein Contents, and Photosynthetic Pigment Composition of *Veronica Anagallis-Aquatica* L. And Ranunculus Aquatilis L. Bull. Environ. Contam. Toxicol. 65, 459–464 (2000).
14. Ahmad, A., Hadi, F., Ali, N., & Jan, A. U. Enhanced Phytoremediation of Cadmium Polluted Water through Two Aquatic Plants *Veronica Anagallis-Aquatica* and *Epilobium Laxum*. Environmental Science and Pollution. 23, 17715–17729 (2016).
15. Salehi, B. et al. *Veronica* Plants-Drifting From Farm to Traditional Healing, Food Application, and Phytopharmacology. Molecules. 24, 2454 (2019).

16. Shim, K. S. et al. Ethanol Extract of *Veronica Persica* Ameliorates House Dust Mite-Induced Asthmatic Inflammation by Inhibiting Stat-3 and Stat-6 Activation. Biomed. Pharmacother. 152, 113264 (2022).
17. Nazlic, M., Akrap, K., Kremer, D. & Dunkic, V. Hydrosols of *Veronica* Species-Natural Source of Free Volatile Compounds with Potential Pharmacological Interest. Pharmaceuticals. 15, 1378 (2022).
18. Chen, J. et al. Molecular Identification and Phylogenetic Analysis of the Traditional Chinese Medicinal Plant Kochia Scoparia Using its2 Barcoding. Interdiscip. Sci. 13, 128–139 (2021).
19. Liu, M., Li, X. W., Liao, B. S., Luo, L. & Ren, Y. Y. Species Identification of Poisonous Medicinal Plant Using DNA Barcoding. Chin. J. Nat. Med. 17, 585–590 (2019).
20. Zhu, S. et al. Phylogenetic Analysis of Uncaria Species Based On Internal Transcribed Spacer (ITS) Region and ITS2 Secondary Structure. Pharm. Biol. 56, 548–558 (2018).
21. Liu, L. X. et al. The Complete Chloroplast Genome of Chinese Bayberry (*Morella Rubra*, Myricaceae): Implications for Understanding the Evolution of Fagales. Front. Plant Sci. 8, 968 (2017).
22. Jun, H. K. et al. Diversity and Evolution of Major Panax Species Revealed by Scanning the Entire Chloroplast Intergenic Spacer Sequences. Genet. Resour. Crop Evol. 60, (2013).
23. Sun, J. L., Han, Y., Cui, X. M. & Liu, Y. Development and Application of Chloroplast Molecular Markers in Panax Notoginseng. Zhongguo Zhong Yao Za Zhi. 45, 1342–1349 (2020).
24. Lee, S. R., Kim, K., Lee, B. Y. & Lim, C. E. Complete Chloroplast Genomes of All Six *Hosta* Species Occurring in Korea: Molecular Structures, Comparative, and Phylogenetic Analyses. Bmc Genomics. 20, 833 (2019).
25. Iram, S., Hayat, M. Q., Tahir, M., Gul, A. & Ahmed, I. Chloroplast Genome Sequence of Artemisia Scoparia: Comparative Analyses and Screening of Mutational Hotspots. Plants-Basel. 8, (2019).
26. Nock, C. J. et al. Chloroplast Genome Sequences From Total DNA for Plant Identification. Plant Biotechnol. J. 9, 328–333 (2011).
27. Saina, J. K. et al. The Complete Chloroplast Genome Sequence of *Dodonaea Viscosa*: Comparative and Phylogenetic Analyses. Genetica. 146, 101–113 (2017).
28. Kishore, A., Sodhi, M., Mukesh, M., Mishra, B. P. & Sobti, R. C. Sequence Analysis and Identification of New Variations in the 5'-Flanking Region of αS2-Casein Gene in Indian Zebu Cattle. Mol. Biol. Rep. 40, 4473–4481 (2013).
29. Cui, Y., Chen, X., Nie, L., Sun, W. & Yao, H. Comparison and Phylogenetic Analysis of Chloroplast Genomes of Three Medicinal and Edible *Amomum* Species. International Journal of Molecular Sciences. 20, 4040 (2019).
30. Du Q et al. Complete Chloroplast Genomes of Two Medicinal *Swertia* Species: The Comparative Evolutionary Analysis of *Swertia* Genus in the Gentianaceae Family. Planta. 256, 73 (2022).
31. Du, Q. et al. Comparative Genomics and Phylogenetic Analysis of the Chloroplast Genomes in Three Medicinal *Salvia* Species for Bioexploration. Int. J. Mol. Sci. 23, 12080 (2022).

32. Zhao, K. et al. Comparative Analyses of Chloroplast Genomes From Six *Rhodiola* Species: Variable DNA Markers Identification and Phylogenetic Relationships within the Genus. Bmc Genomics. 23, 577 (2022).
33. Wang, R. et al. Comparative and Phylogenetic Analyses of Complete Chloroplast Genomes of *Scrophularia Incisa* Complex (Scrophulariaceae). Genes. 13, 1691 (2022).
34. Liu, K. & Fan, S. J. Characterization and Phylogenetic Analysis of the Complete Plastome of *Veronica* Undulata (Plantaginaceae). Mitochondrial Dna Part B-Resour. 6, 2706–2707 (2021).
35. Jia, S. et al. Characterization of the Complete Chloroplast Genome of the Medicinal Herb *Veronica Polita* Fr. 1819 (Lamiales: Plantaginaceae). Mitochondrial Dna Part B-Resour. 7, 1078–1080 (2022).
36. Liu, H. et al. Characterization of the Complete Chloroplast Genome of *Veronica* Arvensis and its Phylogenomic Inference in Plantaginaceae. Mitochondrial Dna Part B-Resour. 7, 1928–1932 (2022).
37. Choi, K. S., Chung, M. G. & Park, S. The Complete Chloroplast Genome Sequences of Three Veroniceae Species (Plantaginaceae): Comparative Analysis and Highly Divergent Regions. Front. Plant Sci. 7, 355 (2016).
38. Li, H., Guo, Q. Q. & Zheng, W. L. Characterization of the Complete Chloroplast Genomes of two Sister Species of Paeonia: Genome Structure and Evolution. Conserv. Genet. Resour. 10, 209–212 (2018).
39. Ravi, V., Khurana, J. P., Tyagi, A. K. & Khurana, P. An Update On Chloroplast Genomes. Plant Systematics & Evolution. 271, 101–122 (2008).
40. Kangkang, M. et al. Application of Molecular Markers and Transcriptomics Research Progress in Olive. Molecular Plant Breeding. 14, 3469–3478 (2016).
41. Zhang, Q., Li, J., Zhao, Y. B. & Han, X. P. Evaluation of Genetic Diversity in Chinese Wild Apple Species Along with Apple Cultivars Using SSR Markers. Plant Mol. Biol. Rep. 30, 539–546 (2011).
42. Gandhi, S. G., Awasthi, P. & Bedi, Y. S. Analysis of SSR Dynamics in Chloroplast Genomes of Brassicaceae Family. Bioinformation. 5, 16–20 (2010).
43. Zhao, Y. et al. The Complete Chloroplast Genome Provides Insight Into the Evolution and Polymorphism of Panax Ginseng. Front. Plant Sci. 5, 696 (2014).
44. Bayly, M. J. et al. Chloroplast genome analysis of Australian eucalypts-*Eucalyptus*, *Corymbia*, *Angophora*, *Allosyncarpia* and *Stockwellia* (Myrtaceae). *Mol. Phylogenet. Evol.* 69, 704–716 (2013).
45. Qian, J. et al. The Complete Chloroplast Genome Sequence of the Medicinal Plant *Salvia Miltiorrhiza*. Plos One. 8, e57607 (2013).
46. Zhou, M. et al. Non-Optimal Codon Usage Affects Expression, Structure and Function of Clock Protein FRQ. Nature. 495, 111–115 (2013).
47. Lyu, X. & Liu, Y. Non-Optimal Codon Usage is Critical for Protein Structure and Function of the Master General Amino Acid Control Regulator CPC-1. Cold Spring Harbor Laboratory. 11, e02605-e02620 (2020).
48. Jian, H. Y. et al. The Complete Chloroplast Genome of a Key Ancestor of Modern Roses, Rosa Chinensis Var. Spontanea, and a Comparison with Congeneric Species. Molecules. 23, 389 (2018).

49. Somaratne, Y., Guan, D. L., Wang, W. Q., Zhao, L. & Xu, S. Q. The Complete Chloroplast Genomes of two *Lespedeza* Species: Insights Into Codon Usage Bias, Rna Editing Sites, and Phylogenetic Relationships in Desmodieae (Fabaceae: Papilionoideae). *Plants-Basel*. 9, 51 (2019).
50. Jiang, W., Lü, B., He, J., Wang, J. & Tang, X. Codon Usage Bias in the Straw Mushroom Volvariella Volvacea. Chinese Journal of Biotechnology. 30, 1424–1435 (2014).
51. Chen, M. M., Zhang, M., Liang, Z. S. & He, Q. L. Characterization and Comparative Analysis of Chloroplast Genomes in Five *Uncaria* Species Endemic to China. Int. J. Mol. Sci. 23, (2022).
52. Dong, S. 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. 22, 880 (2021).
53. Dong, F., Lin, Z., Lin, J., Ming, R. & Zhang, W. Chloroplast Genome of Rambutan and Comparative Analyses in Sapindaceae. *Plants-Basel*. 10, 283 (2021).
54. Yu, X. et al. Comparative Analysis of Chloroplast Genomes of Five *Robinia* Species: Genome Comparative and Evolution Analysis. Gene. 689, 141–151 (2019).
55. Jie, W., Wenchuang, H., Kunli, X., Zhiqiang, W. & Cuihu, G. Advances in Plant Phylogeny in the Genome Era. Journal of Zhejiang a & F University. 40, 10 (2023).
56. Hui, Z., Shuaibing, H., Fande, K. & Haimei, C. Sequence of Chloroplast Genome and the Phyletic Evolution within Leonurus Artemisia. Information On Traditional Chinese Medicine. 35, 21–27 (2018).
57. Cui, Y. et al. Comparative and Phylogenetic Analyses of Ginger (*Zingiber Officinale*) in the Family Zingiberaceae Based On the Complete Chloroplast Genome. Plants-Basel. 8, 283 (2019).
58. Guo, H. et al. Complete Chloroplast Genome Sequences of Schisandra Chinensis: Genome Structure, Comparative Analysis, and Phylogenetic Relationship of Basal Angiosperms. Science China. Life Sciences. 60, 1286–1290 (2017).
59. Wang, J. C., Pan, B. R. & Albach, D. C. Evolution of Morphological and Climatic Adaptations in *Veronica* L. (Plantaginaceae). *Peerj*. 4, e2333(2016).
60. Albach, D. C., Jensen, S. R., Zg Kce, F. & Grayer, R. J. Veronica: Chemical Characters for the Support of Phylogenetic Relationships Based On Nuclear Ribosomal and Plastid Dna Sequence Data. Biochem. Syst. Ecol. 33, 1087–1106 (2005).
61. Ellmouni, F. Y., Karam, M. A., Ali, R. M. & Albach, D. C. Molecular and Morphometric Analysis of Veronica L. Section Beccabunga (Hill) Dumort. Aquat. Bot. 136, 95–111 (2017).
62. Xu, J. & Zhao, Y. M. Analysis of Genetic Relationship of some Plants in *Veronica* based on matk Sequence. Shaanxi Journal of Agricultural Sciences. 65, 5–10 (2019).
63. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. Bioinformatics. 30, 2114–2120 (2014).
64. Dierckxsens, N., Mardulyn, P. & Smits, G. Unraveling Heteroplasmy Patterns with NOVOPlasty. Nar Genom. Bioinform. 2, lqz11 (2020).

65. Liu, C. et al. CpGAVAS, an Integrated Web Server for the Annotation, Visualization, Analysis, and Genbank Submission of Completely Sequenced Chloroplast Genome Sequences. Bmc Genomics. 13, 715 (2012).
66. Tillich, M. et al. Geseq - Versatile and Accurate Annotation of Organelle Genomes. Nucleic Acids Res. 45, W6-W11 (2017).
67. Schattner, P., Brooks, A. N. & Lowe, T. M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33, W686-W689 (2005).
68. Zhang, T. et al. Codon Usage Bias in Pomegranate Transcriptome. Acta Horticulturae Sinica. 44, 675–690 (2017).
69. Kurtz, S. et al. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 29, 4633–4642 (2001).
70. Beier, S., Thiel, T., Munch, T., Scholz, U. & Mascher, M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 33, 2583–2585 (2017).
71. Kearse M, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28, 1647–1649 (2012).
72. Rozas, J. et al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol. Biol. Evol. 34, 3299–3302 (2017).
73. Wang, D. Zhang, Y. Zhang, Z. Zhu, J. & Yu, J. KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom. Bioinform.* 8, 77–80 (2010) <uvertical-align:super;>.</uvertical-align:super;>
74. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. 30, 772–780 (2013).
75. Nakamura, T., Yamada, K. D., Tomii, K. & Katoh, K. Parallelization of MAFFT for Large-Scale Multiple Sequence Alignments. Bioinformatics. 34, 2490–2492 (2018).
76. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol. Biol. Evol. 32, 268–274 (2015).

Tables

Tables 1-2 is available in the Supplementary Files section.

Figures

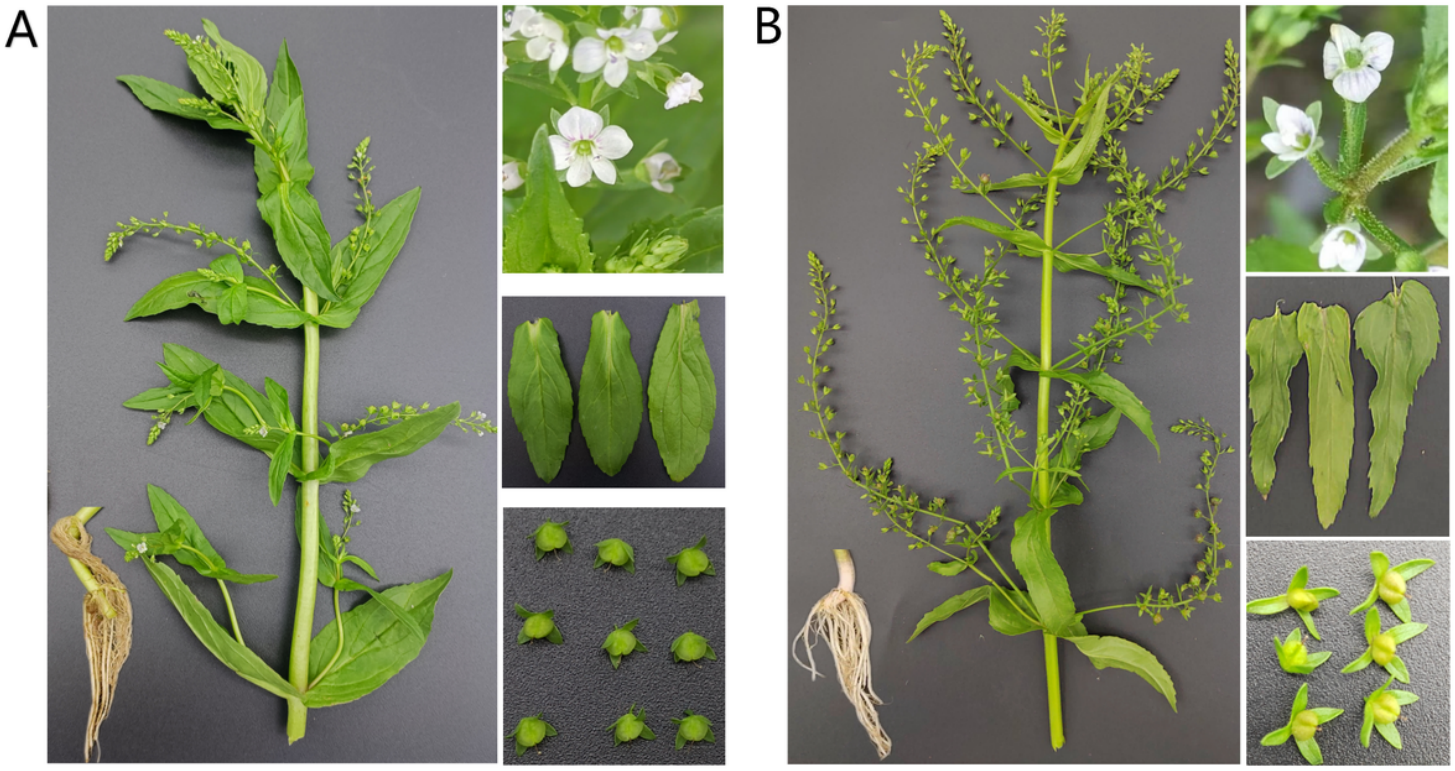

Figure 1

Plant morphology of *V.anagallis-aquatica* (**A**) and *V.undulata*(**B**)

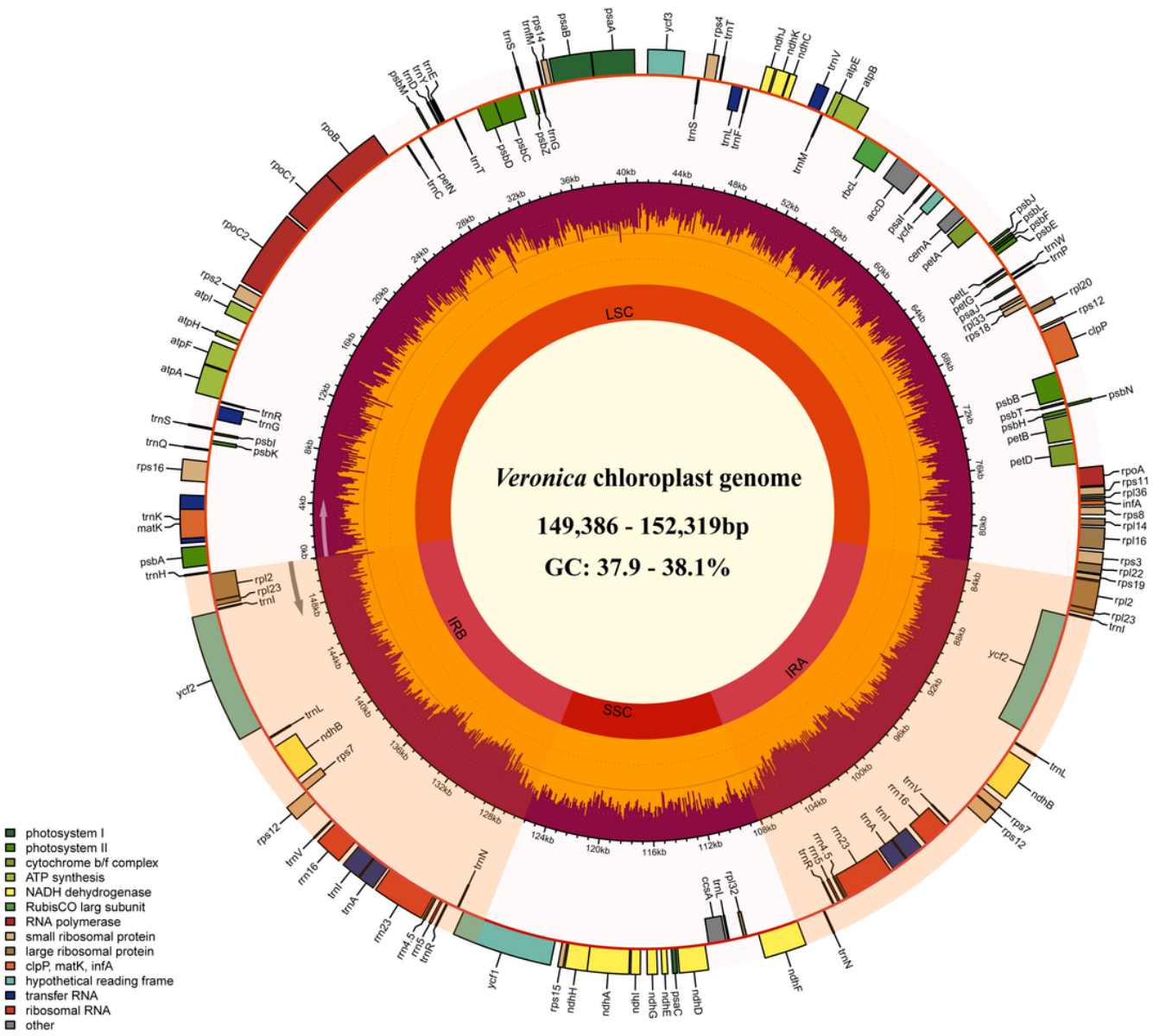

Figure 2

Physical map of cp genome of 9 *Veronica* specie

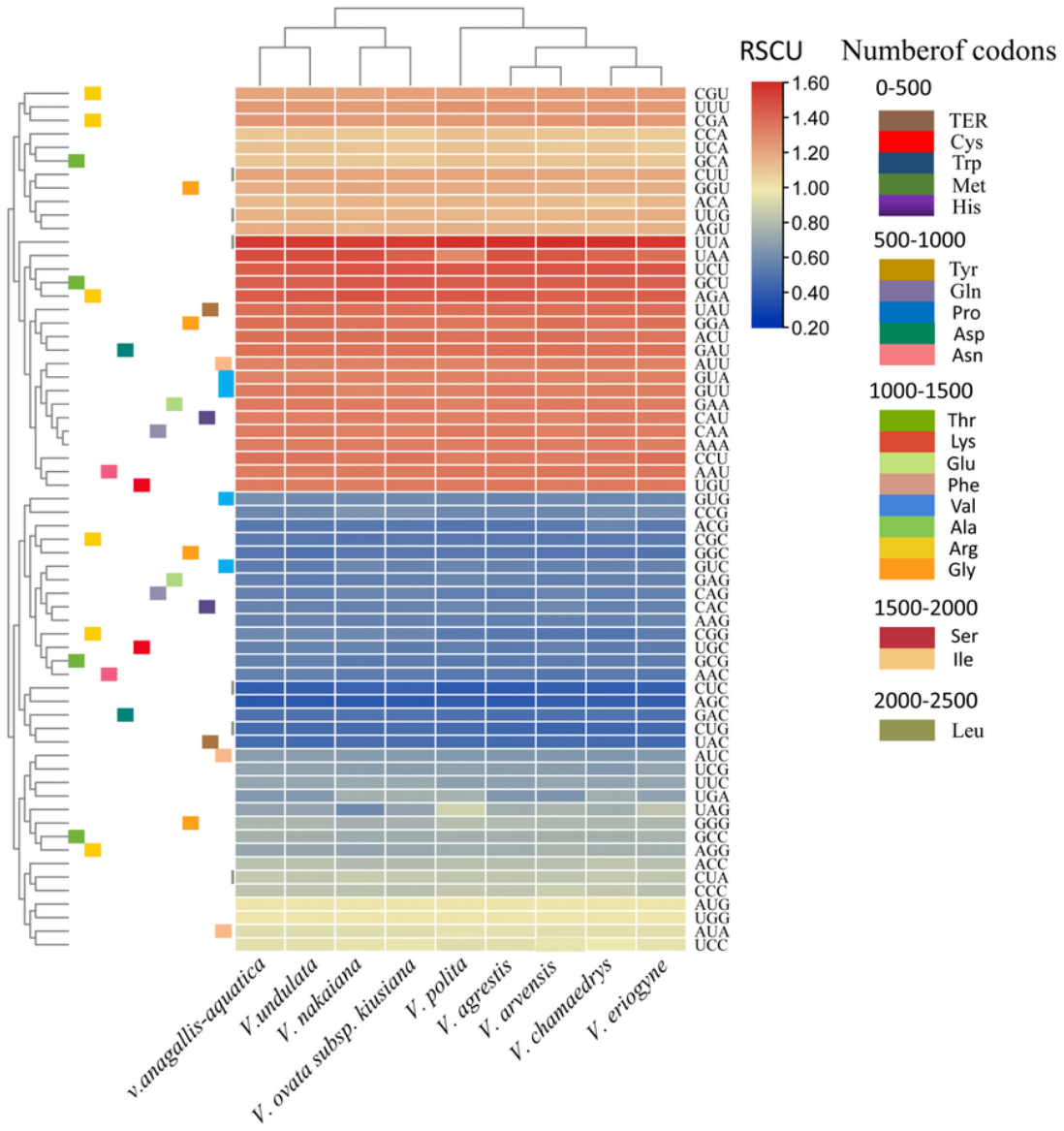

Figure 3

Heat map of RSCU values for nine species of 9 *Veronica* specie

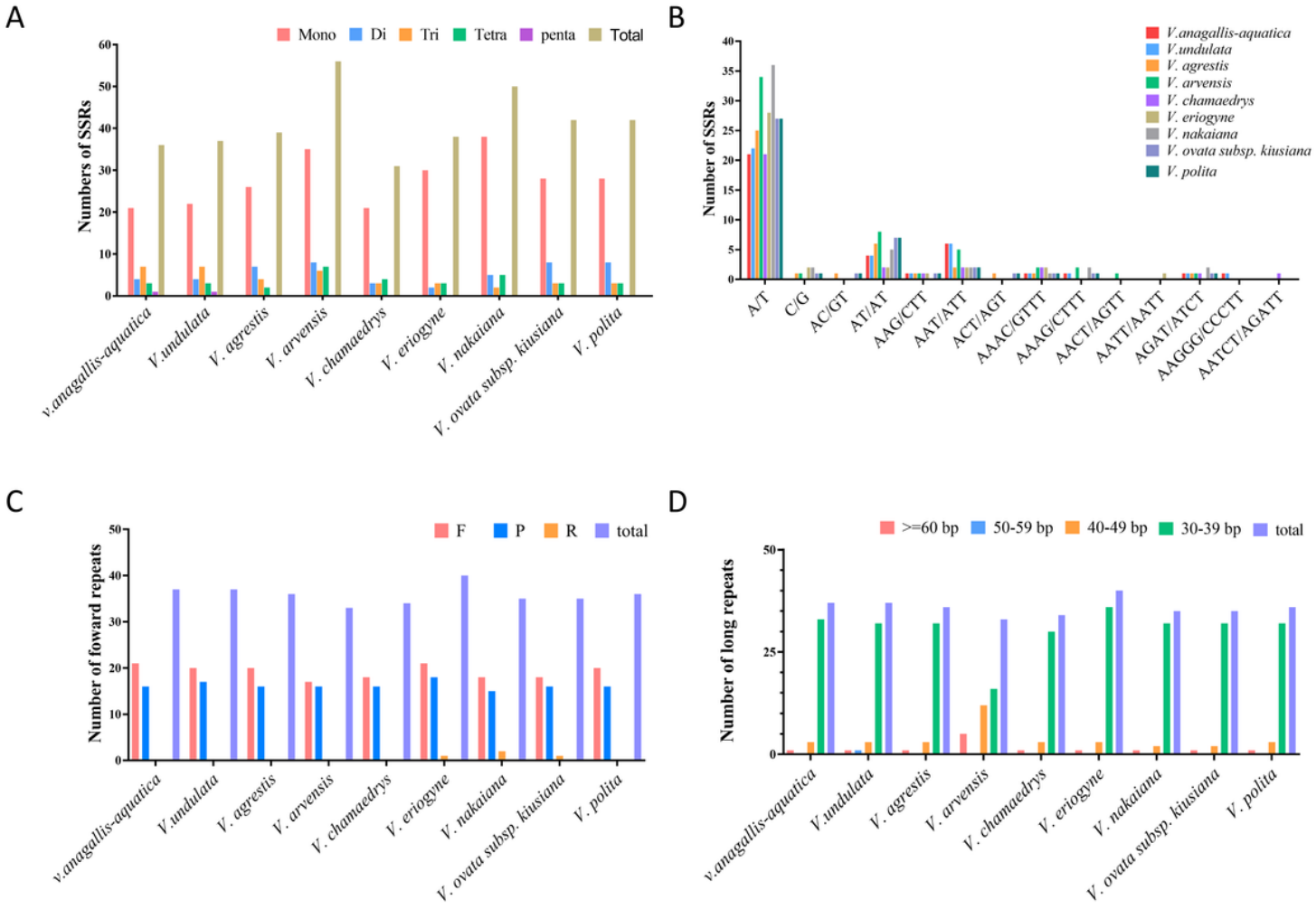

Figure 4

Analysis of SSR and long repeat sequences of 9 *Veronica* species

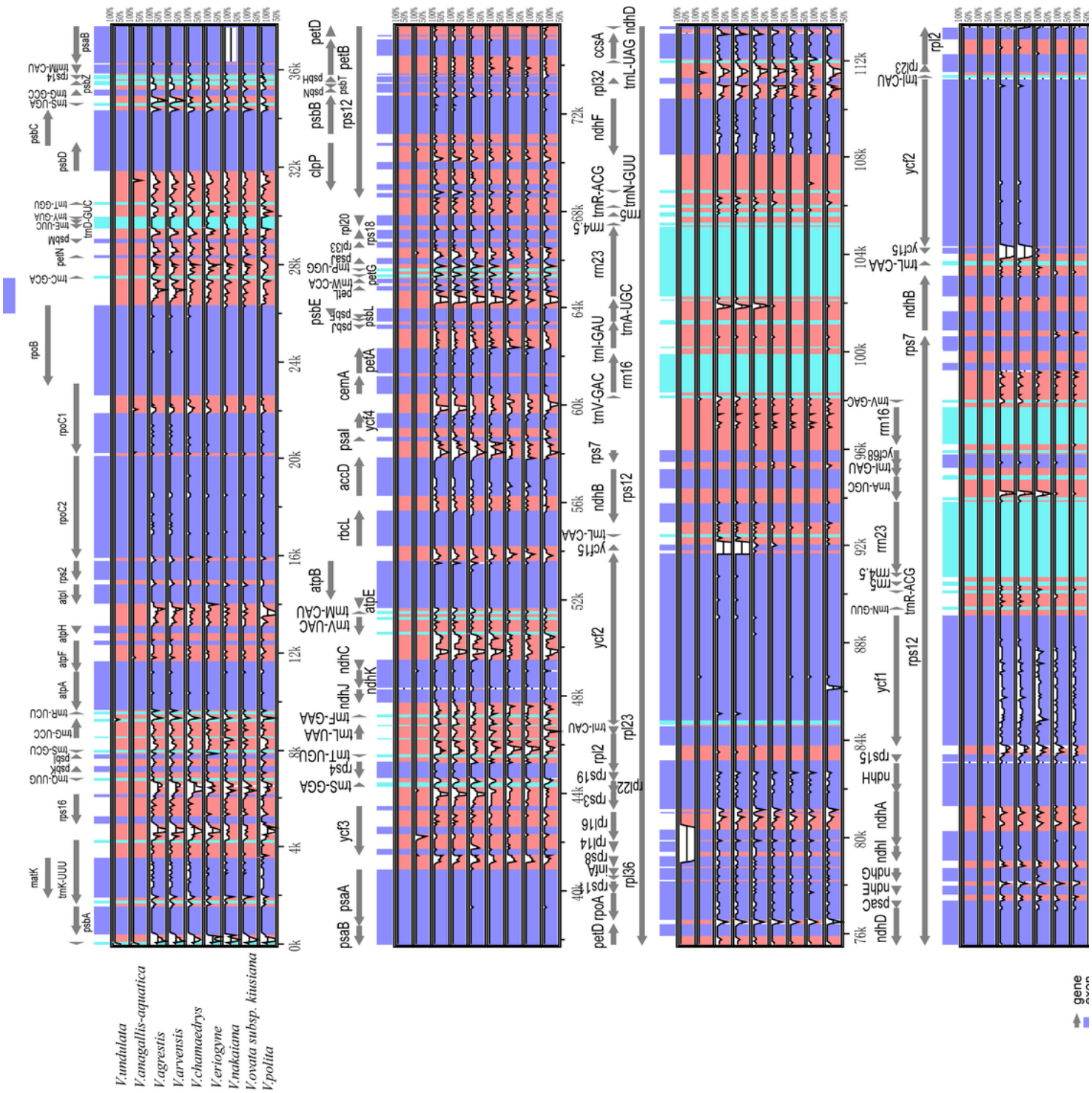

Figure 5

Comparison of cp genome sequences of 9 *Veronica* species

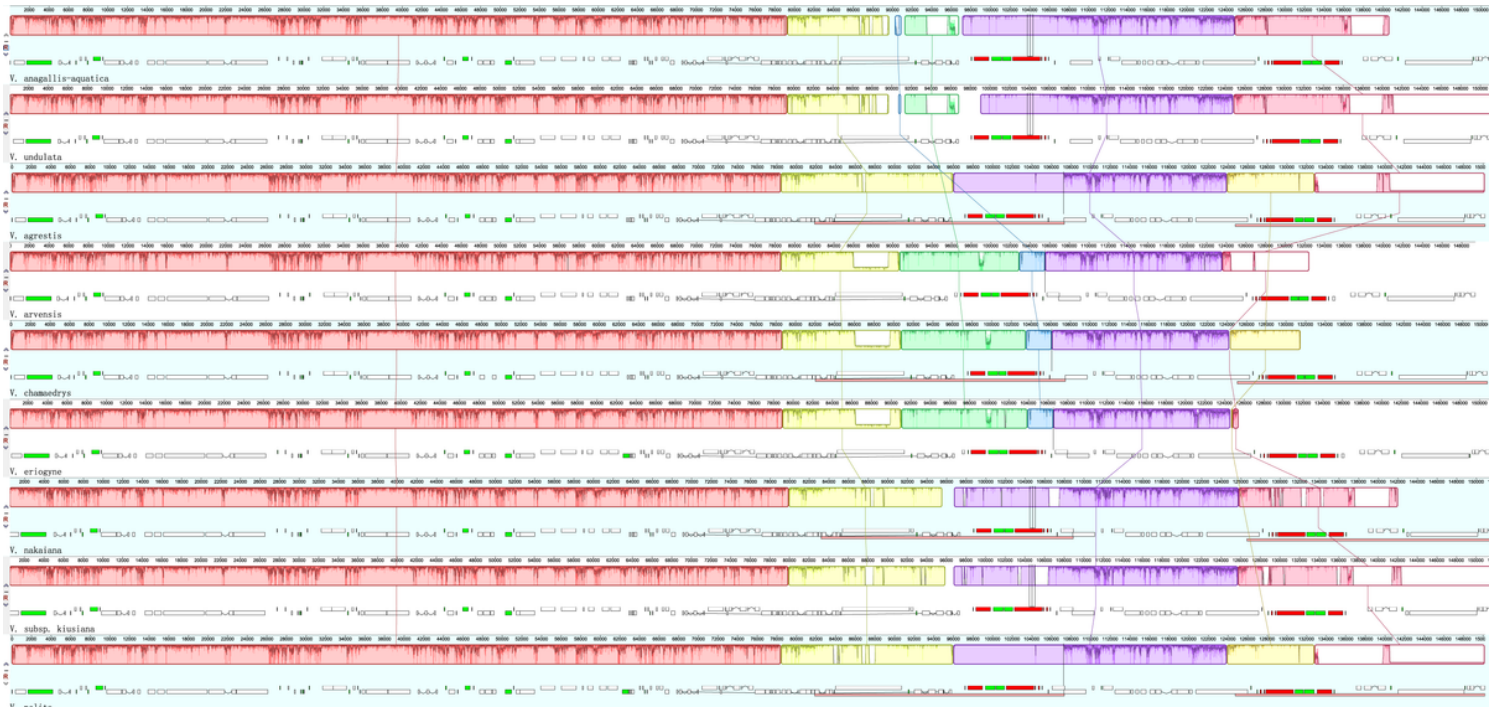

Figure 6

Collinearity analysis of the cp genomes of *Veronica*

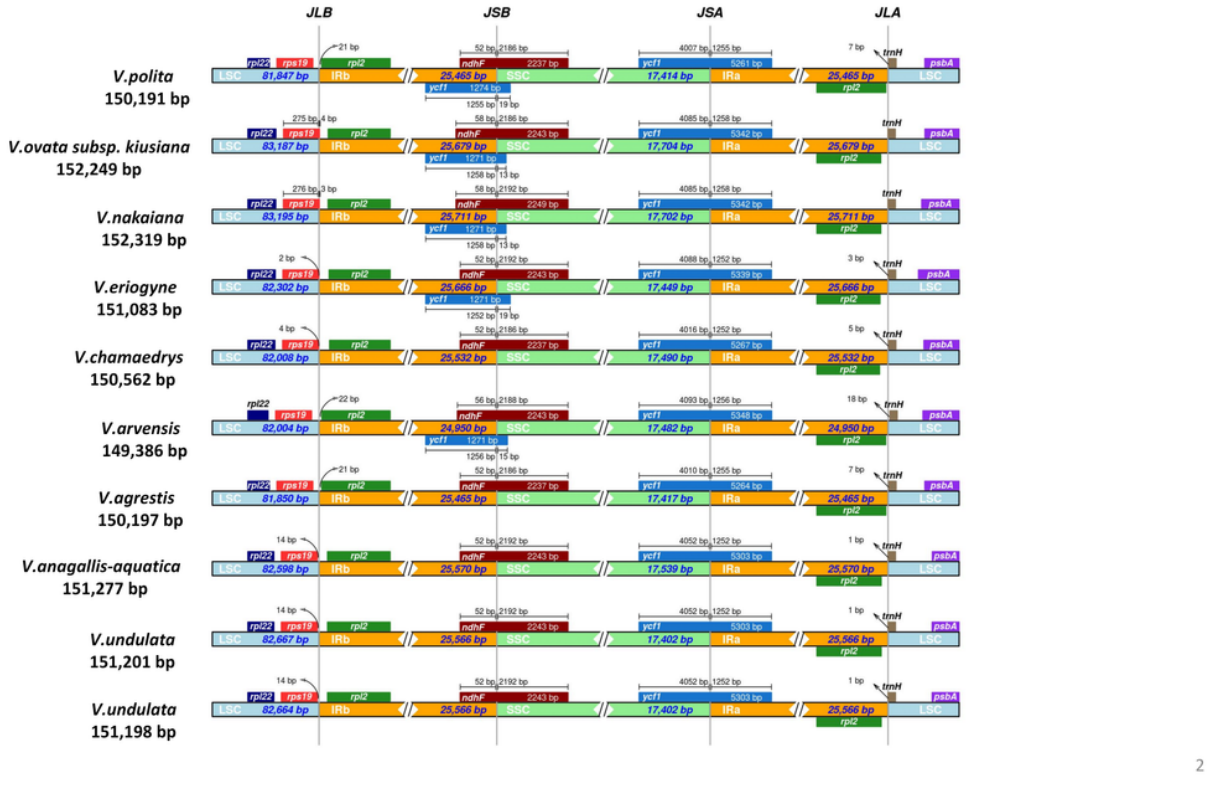

Figure 7

Comparison of the boundaries of LSC, SSC and IR regions in the cp genomes of 9 *Veronica* species. JLB connects LSC and IRb; jSB connects IRb and SSC; jSA connects SSC and IRa; jLA connects IRa and LSC

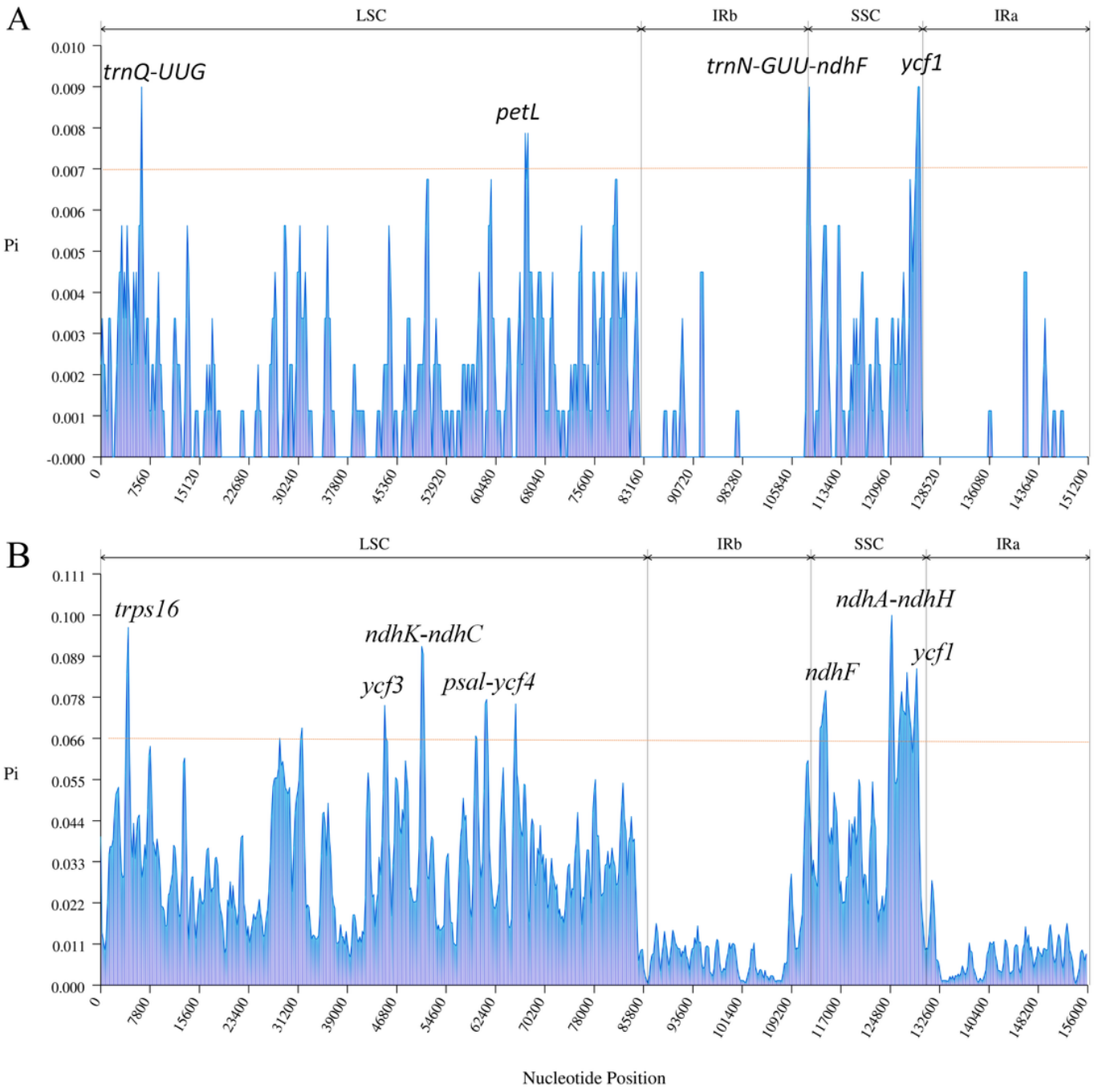

Figure 8

Sliding window analysis of cp genomes of *V. anagallis-aquatica* and *V. undulata* (**A**) and 9 *Veronica* species (**B**) (X axis : the position of the window 's midpoint ; y axis : nucleotide polymorphism Pi value of each window)

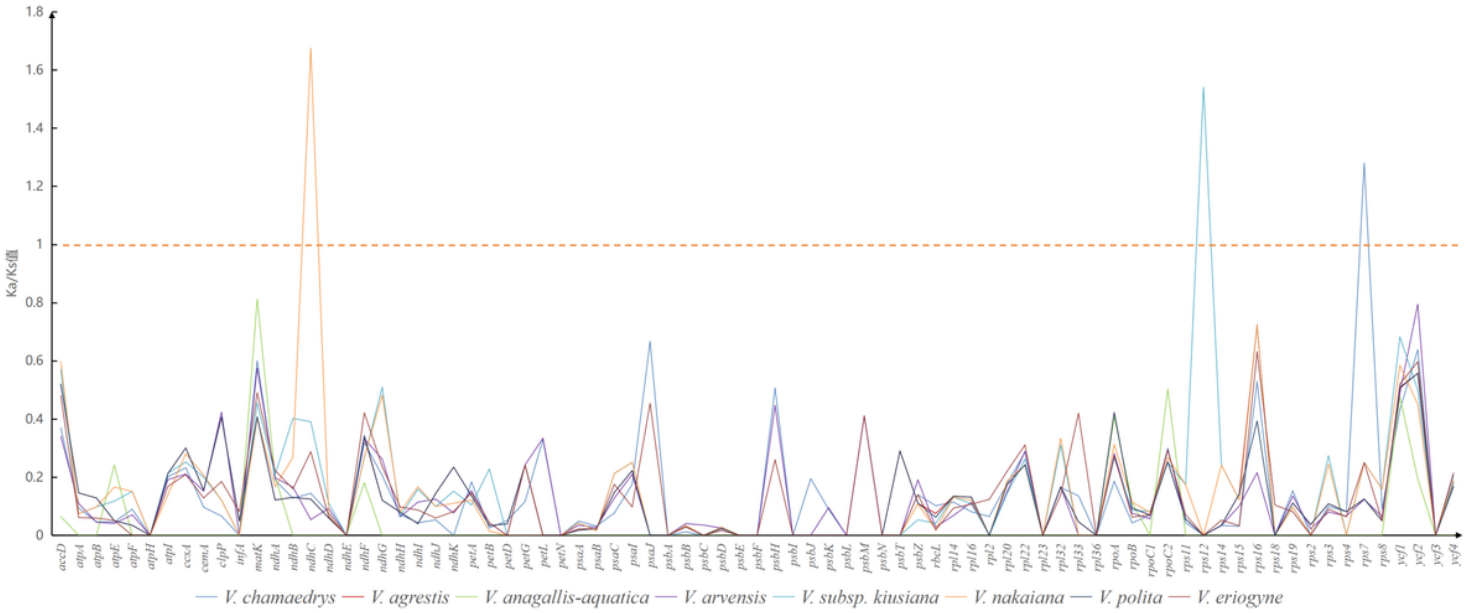

Figure 9

The Ka/Ks ratio of 79 protein-coding genes of 8 cp genomes for comparison with *V.undulata* (OR187860).

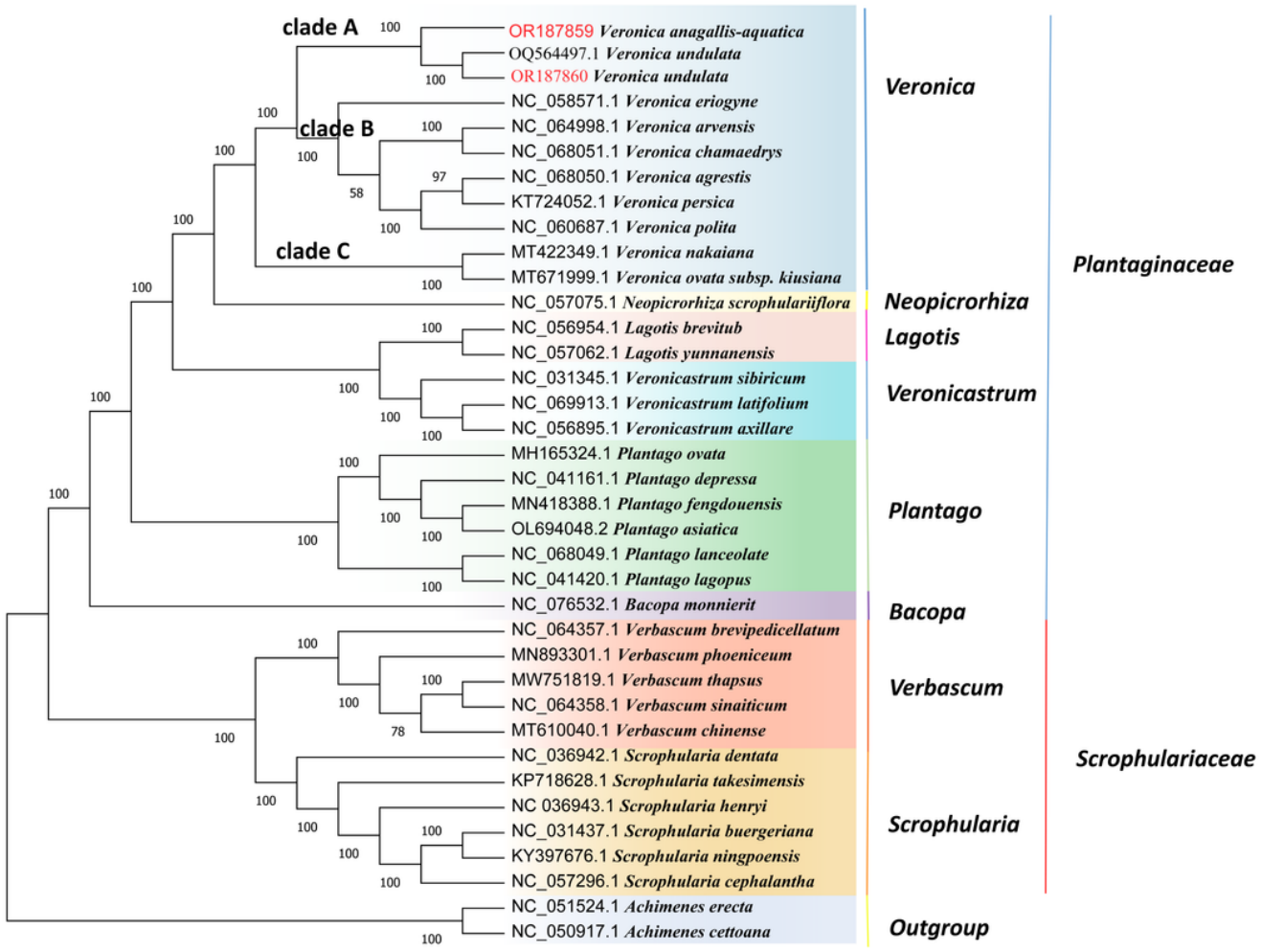

Figure 10

Maximum Likelihood (ML) phylogenetic tree based on complete cp genomes. *Achimenes cettoana* and *Achimenes erecta* were used as outgroups. Numbers at nodes are bootstrap support values.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- tablesofarticle.zip
- supplementarymaterial.zip