Comparative Chloroplast Genome Analysis of Medicinally Important Veratrum (Melanthiaceae) in China: Insights into Genomic Characterization and Phylogenetic Relationships

Ying-min Zhang
Yunnan University of Traditional Chinese Medicine
Li-jun Han
Yunnan University of Traditional Chinese Medicine
Ying-Ying Liu
Yunnan provincial food and drug evaluation and inspection center
Cong-wei Yang
Yunnan University of Traditional Chinese Medicine
Xing Tian
Yunnan University of Traditional Chinese Medicine
Zi-gang Qian
Yunnan University of Traditional Chinese Medicine
Guodong Li (gammar116@163.com)
Yunnan University of Chinese Medicine https://orcid.org/0000-0002-9108-5454

Research

Keywords: Veratrum, Chloroplast genome, Sequences variations, Medicine-herb, Phylogeny

Posted Date: December 7th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-117897/v1

Version of Record: A version of this preprint was published at Plant Diversity on June 1st, 2021. See the published version at https://doi.org/10.1016/j.pld.2021.05.004.

Abstract

Background: *Veratrum* is a genus of perennial herbs that are widely used as traditional Chinese medicine for emetic, resolving blood stasis and relieve pain. However, the species classification and the phylogenetic relationship of the genus *Veratrum* have long been controversial due to the complexity of morphological variations. Knowledge on the infrageneric relationships of the genus *Veratrum* can be obtained from their chloroplast genome sequences and increase the taxonomic and phylogenetic resolution.

Methods: Total DNA was extracted from ten species of *Veratrum* and subjected to next-generation sequencing. The cp genome was assembled by NOVOPlasty. Genome annotation was conducted using the online tool DOGMA and subsequently corrected by Geneious Prime. Then, genomic characterization of the *Veratrum* plastome and genome comparison with closely related species was analyzed by corresponding software. Moreover, phylogenetical trees were reconstructed, based on the 29 plastomes by maximum likelihood (ML) and Bayesian inference (BI) methods.

Results: The whole plastomes of *Veratrum* species possess a typical quadripartite structure, ranging from 151,597 bp to 153,711 bp in size and comprising 135 genes. The gene order, content, and genome structure were nearly identical with a few exceptions across the *Veratrum* chloroplast genomes. The total number of simple sequence repeats (SSRs) ranged from 31 to 35, and of large sequence repeats (LSRs) ranged from 65 to 71. Seven highly divergent regions (*rpoB-trnC*, *trnT-trnL*, *trnS-trnG*, *psbC-psbZ*, *psbI*, *ycf1*, and *ndhF*) were identified that can be used for DNA barcoding in the genus of *Veratrum*. Phylogenetic analyses based on 29 plastomes strongly supported the monophyly of *Veratrum*. The circumscription and relationships of infrageneric taxa of *Veratrum* were well evaluated with high resolutions.

Conclusions: Our study identified and analyzed the cp genome features of ten *Veratrum* species, and suggested high effectivity of chloroplast complete genome in resolving generic circumscription in *Veratrum*. These results will facilitate the identification, taxonomy, and utilization of *Veratrum* plants as well as the phylogenetic study of Melanthiaceae simultaneously.

Background

The genus *Veratrum* L. is one of the most important groups in the family Melanthiaceae (Liliflorae), with approximately 17–45 species of perennial herbaceous plants [1–5]. It is widely distributed in the temperate to arctic zones of the Northern Hemisphere, the majority of which are native to Eastern Asia [1, 2, 6]. There are roughly 13 species and one variety of *Veratrum* plants in China, several species were used in medicine for more than 1700 years, including *V. nigrum*, *V. schindleri*, *V. maackii* and others [4, 5, 7, 8]. For instance, the dried roots and rhizomes of *V. nigrum* (Li-lu in Chinese) have been used to treat aphasia arising from apoplexy, wind-type dysentery, scabies, jaundice, and chronic malaria [7, 8]. Other species *V. taliense*, *V. stenophyllum*, *V. mengtzeanum*, and *V. grandiflorum* that are the source of folk medicine “Pimacao” in China, has been used for treating bruises, rheumatic pain and wound hemostasis [8–10]. Furthermore, “Pimacao” has a high economic value and is one of the main components of Yunnan Baiyao, as well as the principal drug of Yilizhitong pill [10, 11]. The major active ingredient isolated from roots of the above *Veratrum* plants is steroidal alkaloids, which pharmacological activities mainly focus on decreasing blood pressure, anti-platelet aggregation and anti-thrombosis, and have anti-inflammatory, analgesic and antitumor effect [11–14]. However, *Veratrum* plants also contain toxic components such as cevanine-ester alkaloids [13–15]. In human nausea, bradycardia, hypotension and apnea develop shortly after ingestion, in some cases resulting in death [12, 13, 15]. Our previous investigation found that some closely related species are occasionally mixed with *Veratrum* medicinal varieties as adulterants undermining the security and efficacy of the *Veratrum* containing medicinal products. Thus, securing accurate identification is urgent for the utilization of *Veratrum* medicinal materials safely and effectively.

Veratrum has a controversial taxonomic history and has been combined with *Melanthium* totally or in part [1, 16–19]. Since Linnaeus (1753) first circumscription of *Verarum*, many subgenera, sections, and subsections have been proposed [1, 16–27]. The first subgeneric division of *Veratrum* was suggested by Baker (1880) [21]. Using perianth coloration as a diagnostic trait, he divided the genus into informal groups: stirps *V. albi* (perianth white-green) and stirps *V. nigri* (perianth purple-black) [21]. Loesener split the genus *Veratrum* into three subgenera and two sections including 48 species consisting of elements of *Melanthium* [22–24]. Nakai modified Loesener's infrageneric classification into two sections, omitting the rank of subgenera, and divided each section into two subsections [25, 26]. Recently, *Veratrum* has been circumscribed broadly (including *Melanthium*) and divided into two sections (sect. *Veratrum* and sect. *Fuscoveratrum*) and two subsections (subsect. *Pseudoanti* and subsect. *Asiaveratrum*) [1]. This modified infrageneric classification established a framework for resolving phylogenetic relationships within *Veratrum* [1, 6].

Morphologically, the key characteristics of *Veratrum* vary greatly depending on habitat, environment, and developmental stages, and the range of those variations often overlap among the taxa [1, 22–24, 28]. As a result, some species and its closely related species constitute a taxonomically complex group that is difficult to be clearly distinguished based on morphological traits alone such as *V. maackii* complex group [1, 6]. *V. maackii* has been divided into numerous varieties by different authors, including *V. japonicum* (broader leaves; relatively large flowers), *V. schindleri* (broader leaves; short-pedicellate flowers) and so on [1, 4, 5, 25, 26, 29–31]. Shimizu pointed that *V. japonicum* was a variety of *V. maackii* and was recognized by WCSP (World Check List of Selected Plant Families; http://www.theplantlist.org). In Flora Reipublicae Popularis Sinicae (FRPS), *V. japonicum* and *V. schindleri* were treated as two distinct species from *V. maackii* [4]. In addition, *V. japonicum* was treated as a synonym of *V. schindleri* in Flora of China (FOC) [5]. At the same time, because of its similar characters, such as black-purple perianth, basal leaves, and bulb layers disintegrating into reticulated fibers, *V. maackii* has been historically placed either in the *V. nigrum* or treated as its subspecies [32].

Veratrum species were a subject of many molecular analyses [1, 2, 6, 33–35]. Kim et al. conducted maximum parsimony and Bayesian inference based on ITS and cpDNA regions (*matK*, *psbA-trnH*, *rpl16*, and *trnS-G*)to re-examine the taxonomic status and phylogenetic relationships within *Veratrum* in Korea and Japan [33]. An analysis using the sequences of ITS, *trnL-F*, and *atpB-rbcL* indicated that *Veratrum* possibly originated in East Asia and radiated across the Northern Hemisphere, but most of the species were not well distinguished [6]. Previous studies have significantly advanced the phylogeny and taxonomy of genus *Veratrum*. However, due to limited resolution of molecular phylogeny and insufficient sampling of Asiatic species, phylogenetic relationships among the

species of the main clades (sections and subsections), particularly among the species from East Asia, are still poorly understood [1, 2, 6]. These shortcomings motivated our study.

The chloroplast genome of higher plants is relatively conservative in its structure, being a double-stranded circular molecule of 120–160 kb in length and comprising a large single-copy (LSC) region and a small single-copy (SSC) region, separated by two identical copies of inverted repeats (IRs) regions [34–36]. The cp genome has been widely used for evolutionary, taxonomic and species diversity studies due to such features as highly conserved genome structure, maternal inheritance, low to moderate evolutionary rate and low effective population sizes [36, 37]. The cp genomes have been showed to be effective in resolving problematic phylogenetic relationships at different taxonomic levels [38–41]. Up to now, however, the plastomes of only a few *Veratrum* species have been sequenced, and the data accumulated are still deficient for the clarification of the internal relationships of the family [35, 42–44]. Hence, we attempted to explore more *Veratrum* and its related species phylogenetic relationships with chloroplast genomics.

Here, we sequenced, assembled and annotated the complete cp genomes of ten *Veratrum* species using the next-generation sequencing platform, and performed the first comprehensive analysis of *Veatrum* species from China. This study aimed to: (1) establish and characterize the newly sequenced plastomes of ten *Veratrum* species; (2) examine the variations of SSRs and LSRs among these plastomes plus two previously published plastomes of *Veratrum*; (3) discover the most variable regions that could be used as DNA barcodes for *Veratrum*; (4) and reconstruct phylogenetic relationships among the *Veratrum* species using the plastome sequences.

Materials And Methods

Plant material and DNA extraction

Ten species, *V. dahuricum*, *V. grandiflorum*, *V. japonicum*, *V. maackii*, *V. nanchuanense*, *V. nigrum*, *V. oblongum*, *V. schindleri*, *V. stenophyllum* and *V. taliense* were collected in their native habitats in Yunnan, Fujian, Liaoning and Sichuan of China and used in this study. Fresh leaves were taken from healthy, mature individuals and dried with allochroic silicagel. From 3 to 5 individuals per species with flowers or fruits were collected and preserved as voucher specimens for morphological analysis and taxonomic identification. All voucher specimens were deposited in the Herbarium of Yunnan University of Chinese Medicine (YNUCM), China (Table S1). Total genomic DNA extraction was carried out using a modified CTAB protocol [45]. The quality and quantity of the DNA extracts for next-generation sequencing were determined using 1.0% agarose gel electrophoresis, and Nanodrop™ 2000 spectrophotometer (Thermo Fisher Scientific, USA), respectively.

Genome sequencing, assembly and annotation

Paired-end library was constructed by purified DNA and then sequenced on the Illumina HiSeq 2500-PE platform (Illumina, Inc., United States). High quality reads were obtained by removing the low-quality reads and connector sequence of using NGS QC Toolkit with default parameters [46]. Filtered reads of ten *Veratrum* species were assembled de novo using NOVOPlasty with cp genome of its close relative species, *V. patulum* (NC_022715), as the reference [43,47]. Annotation was performed with the online annotation tool DOGMA [48] and subsequently corrected using Geneious Prime® 2020.1.1 (Biomatters Ltd., Auckland, New Zealand). The chloroplast genome maps were drawn through OGDRAW [49]. Finally, the annotated chloroplast genomes of the ten *Veratrum* species were submitted to Genbank (Table 1).

Repeat Sequence Analysis

Simple sequence repeats (SSRs) were detected using IMEx with the search parameters set to 10, 5, 4, 3, 3, and 3 for mono-, di, tri-, tetra-, penta- and hexa-nucleotides, respectively [51]. REPuter software was used for the identification of palindromic, forward, reverse and complement repeats present in the genome, whereby the Hamming distance was set as 3 and the minimum repeat size was 30 bp [50].

Genome comparison and structural analysis

Relative synonymous codon usage (RSCU) in these genes was assessed using MEGA X with default parameters [52,53]. Twelve species were compared and visualized using mVISTA with the Shuffle-LAGAN mode [54,55]. The borders of large single copy (LSC), small singles copy (SSC), and inverted repeat (IR) regions in the genomes of the twelve species were compared and examined. In addition, DnaSP v.5.0 was employed to analyze nucleotide variability among the twelve species of *Veratrum* [56]. The step size was set to 200 bp, with a 600 bp window length.

Phylogenetic analyses

To examine the phylogenetic relationships of *Veratrum* species, 27 plastomes representing phylogenetic diversity in the family Melanthiaceae and two species of Liliales (*Lilium henryi* and *Smilax china*) used as outgroup were included in the phylogenetic analysis (Table S1). All the plastomes were aligned using MAFFT integrated with Geneious Prime. A maximum-likelihood (ML) tree was constructed by RAxML using the general time-reversible (GTR) with gamma distribution (+G) nucleotide substitution model and 1,000 bootstrap replicates for each branch node [57]. Bayesian inference (BI) analyses were conducted using MrBayes v 3.2.6 [58]. The Markov Chain Monte Carlo (MCMC) algorithm was run for 1,00,000 generations and the trees were sampled every 1000 generations. The remaining analysis parameters were set as defaults. The first 25% of the trees were discarded as a burn-in and remaining trees were used to generate the consensus tree, including clade posterior probability (PP).

Results

Chloroplast Genome Features

A total of 12 chloroplast genomes representing majority of *Veratrum* species distributed in China were analyzed (Table 1, Table S1). Among them, 10 plastomes were newly sequenced and assembled in this study. *V. mengtzeanum* (MN589932) and *V. patulum* (NC_022715) (synonym of *V. oxysepalum*) were used as supplemental species for comparative analysis. The complete chloroplast genome of *Veratrum* species ranged from 151,597 bp (*V. maackii*) to 153,711 bp (*V. grandiflorum*) in length (Table 1, Fig. 1). The LSC region ranged 81,822-83,372 bp and SSC region ranged 17,473-17,628 bp, which were separated by two IR regions (26,145-26,359 bp) (Table 1). The overall Guanine-Cytosine (GC) content of twelve *Veratrum* cp genomes ranged 35.7-35.8%, which is similar to that of other Melanthieae species [38,43,44]. And the GC content of the IRs (42.9-43.0%) was higher than that of LSC (35.7-35.8%) and SSC (31.3-31.5%) regions in twelve *Veratrum* species (Table 1).

Table 1
Summary of complete chloroplast genomes of twelve *Veratrum* species

Genome Characteristic	Accession number	Length (bp)				Gene number				GC content (%)			
		Full	LSC	SSC	IR	Full	Protein-coding	tRNA	rRNA	Full	LSC	SSC	IR
V. dahuricum	MN699635	153,688	83,363	17,607	26,359	135	83	38	8	37.7	35.7	31.4	42.9
V. grandiflorum	MN613592	153,711	83,367	17,628	26,358	135	83	38	8	37.7	35.7	31.4	42.9
V. japonicum	MN613594	151,842	81,885	17,533	26,212	135	84	38	8	37.7	35.8	31.3	42.9
V. maackii	MN613590	151,597	81,822	17,473	26,151	135	84	38	8	37.7	35.7	31.4	42.9
V. nanchuanense	MN613591	153,692	83,368	17,608	26,358	135	83	38	8	37.7	35.7	31.4	42.9
V. nigrum	MN613595	151,599	81,823	17,474	26,151	135	84	38	8	37.7	35.7	31.4	42.9
V. oblongum	MN613593	151,767	81,997	17,480	26,145	135	84	38	8	37.7	35.7	31.3	42.9
V. schindleri	MN613588	151,768	81,862	17,530	26,188	135	84	38	8	37.7	35.8	31.4	42.9
V. stenophyllum	MN613589	152,028	82,120	17,558	26,175	135	84	38	8	37.8	35.8	31.5	43.0
V. taliense	MN604405	152,040	82,122	17,558	26,180	135	84	38	8	37.8	35.8	31.5	43.0
V. mengtzeanum	MN589932	152,051	82,111	17,544	26,198	135	84	38	8	37.8	35.8	31.5	42.9
V. patulum	NC_022715	153,699	83,372	17,607	26,360	135	83	38	8	37.7	35.7	31.4	42.9
New sequences from this study are in boldface.													

The gene content, order, and structure were similar in the twelve *Veratrum* cp genomes. Most of the examined plastomes have 84 protein-coding genes, 38 tRNAs and eight rRNAs (Table 1). However, there were 83 protein-coding genes in *V. grandiflorum*, *V. nanchuanense*, *V. nigrum*, and *V. patulum* because of the complete loss of *infA* gene. The LSC region contains 61 protein-coding genes and 21 tRNA genes, whereas the SSC region contains 12 protein-coding genes and one tRNA gene. Six protein-coding and eight tRNA genes are located in the IR regions. Seventeen of the genes contained introns, eight coding genes (*rpl2*, *rpl16*, *rpoC1*, *ndhA*, *ndhB*, *petB*, *petD*, *atpA*) and six tRNA genes (*trnA UGC*, *trnG UCC*, *trnI GAU*, *trnK UUU* and *trnV UAC*) with a single intron, and three coding genes (*rps12*, *clpP ycf3*) with two introns (Table 2). The *rps12* gene was a unique trans-spliced gene with the two duplicated 3' end exons in the IR regions and a 5' end exon in the LSC region. Both *ycf15* and *ycf68* in twelve plastomes of *Veratrum* species contain many internal stop codons, indicating that these sequences represent pseudogenes.

Table 2
Genes in the chloroplast genome of *Veratrum* species

Category	Group of Genes	Name of Genes
Self-replication	Large subunit of ribosomal	*rpl2*[a,c]*、14、16*[a]*、20、22、23*[c]*、32、33、36*
	Small subunit of ribosomal	*rps2、3、4、7*[c]*、8、11、12*[b,c,d]*、14、15、18、19*
	DNA dependent RNA polymerase	*rpoA、rpoB、rpoC1*[a]*、rpoC2*
	rRNA genes	*rrn4.5*[c]*、rrn5*[c]*、rrn16*[c]*、rrn23*[c]
	tRNA genes	*trnA-UGC*[a,c]*、trnC-GCA、trnD-GUC、trnE-UUC、trnF-GAA、trnfM-CAU、trnG-GCC、trnG-UCC*[a]*、trnH-GUG*[c]*、trnI-CAU*[c]*、trnI-GAU*[a,c]*、trnK-UUU*[a]*、trnL-CAA*[c]*、trnL-UAA*[a]*、trnL-UAG、trnM-CAU、trnN-GUU*[c]*、trnP-UGG、trnQ-UUG、trnR-ACG*[c]*、trnR-UCU、trnS-GCU、trnS-GGA、trnS-UGA、trnT-GGU、trnT-UGU、trnV-GAC*[c]*、trnV-UAC*[a]*、trnW-CCA、trnY-GUA*
Photosynthesis	Photosystem I	*psaA、B、C、I、J*
	Photosystem II	*psbA、B、C、D、E、F、H、I、J、K、L、M、N、T、Z*
	NadH oxidoreductase	*ndhA*[a]*、B*[a,c]*、C、D、E、F、G、H、I、J、K*
	Cytochrome b6/f complex	*petA、B*[a]*、D*[a]*、G、L、N*
	ATP synthase	*atpA*[a]*、B、E、F、H、I*
	Rubisco	*rbcL*
Other genes	Maturase	*matK*
	Translational	*infA*
	Protease	*clpP*[b]
	Envelop membrane protein	*cemA*
	Subunit of acetyl-CoA	*accD*
	c-type cytochrome synthesis gene	*ccsA*
Unknown	Conserved Open reading	*ycf1、ycf2*[c]*、ycf3*[b]*、ycf4、ycf15*[c]*、ycf68*[c]

[a]Gene containing a single intron; [b]Gene containing two introns; [c]Gene with two copies; [d]Trans-splicing gene.

SSRs and LSRs of *Veratrum* chloroplast genomes

Numerous simple sequence repeats (SSRs) loci were found through the IMEx analysis in twelve *analyzed plastomes,* ranging from 63 SSRs in *V. taliense* to 78 SSRs in *V. oblongum* (Fig. 2A, Table S2). The most common types of SSRs were mono-nucleotide repeats (65%), followed by di-nucleotide (18%), tetra-nucleotide repeats (8%) and tri-nucleotide repeats (5%). No hexa-nucleotide repeats were found in *any* of the analyzed plastomes (Fig. 2A). Among these, the type number of mono-, di-, tri-, tetra-, penta-, and hexa- nucleotides SSRs are 3, 2, 5, 8, 7, and 2, respectively (Fig. 2B). Most SSRs (97%) were located in LSC and SSC regions, while the inverted regions had very few repeats (3%) (Fig. 2C). The SSRs were more abundant in non-coding than in coding regions (64% vs. 36%) (Fig. 2D).

A total of 403 long sequence repeats (LSRs) including 125 forward, 188 palindromic and 22 reverse repeats were detected in the twelve plastomes (Fig. 3A, Table S3). Repeat numbers varied from 31 in *V. oblongum* to 35 in *V. grandiflorum*, *V. nanchuanense*, *V. patulum*, and *V. dahuricum* (Fig. 3A). LSRs mainly ranged 30-47 bp in length, whereas only one palindromic repeat was longer than 54 bp in *V. dahuricum*, *V. grandiflorum*, *V. nanchuanense* and *V. patulum* (Fig. 3B). Further analysis revealed that most of the LSRs were located in the LSC region, with a small portion distributed throughout the SSC and IR regions (Fig. 3C). In the first location, 39% of the repeats are in the non-coding region (Fig. 3D).

Codon Usage

The frequency of the codon usage present in the plastome of 27 Melanthiaceae species and two Liliales species was computed using the nucleotide sequence of protein-coding genes. The results showed the genes in the plastome are encoded by 23,963 (*Smilax china*) to 29,006 (*Paris rugosa*) codons with the RSCU (relative synonymous codon usage) values ranging from 0.297 (AGC) to 1.988 (AGA) (Fig. 4, Table S4). No codon bias (RSCU=1) can be shown by methionine (AUG) and tryptophan (UGG), encoded by only one codon. Leucine (10.0–10.4%) and cysteine (1.2% of each species) are the most and the least abundant amino acids except for stop codons (0.3%-0.4%) (Table S5). We found that the coding sequence of a protein-coding gene in the *Veratrum* cp DNA is preference A□T base, the same as the third position of the codon.

Comparative analysis of twelve *Veratrum* chloroplast genomes

The alignment of the sequenced *Veratrum* plastomes having 154,858 bp in length revealed high sequence similarity, ranging from 96.46% (*V. nigrum* and *V. patulum*) to 99.99% (*V. stenophyllum* and *V. taliense*) (Table 3). The visualization analysis of the alignment using mVISTA showed that the genomic structure, order, and orientation of these plastomes were highly conserved. Notably, IR regions exhibited less divergence than the SSC and LSC regions (Fig. 5). In addition, the non-coding region was more variable than the coding region, the highly divergent regions among the twelve plastomes appeared in the intergenic spacers, such as *trnK UUU- trnQ UUG*, *trnS GCU-trnG UCC*, *psbC-psbZ*, *rps7-trnV GAC* and *ndhF-rpl32* (Fig. 5). Among coding regions, *ycf1*, *rpl22* and *petD* genes were relatively divergent (Fig. 5).

Table 3
Sequence identity in twelve complete chloroplast genomes of *Veratrum* species

	V. dahuricum	*V. grandiflorum*	*V. japonicum*	*V. maackii*	*V. nanchuanense*	*V. nigrum*	*V. oblongum*	*V. schindleri*	*V. stenophyllum*	*V. taliense*
V. dahuricum										
V. grandiflorum	99.906									
V. japonicum	96.797	96.786								
V. maackii	96.516	96.501	98.543							
V. nanchuanense	99.925	99.964	96.799	96.511						
V. nigrum	96.513	96.498	98.544	99.985	96.508					
V. oblongum	96.615	96.602	98.663	99.529	96.613	99.529				
V. schindleri	96.834	96.818	99.725	98.551	96.831	98.552	98.675			
V. stenophyllum	97.005	96.991	98.73	98.431	97.002	98.429	98.543	98.762		
V. taliense	97.012	96.998	98.737	98.437	97.009	98.436	98.55	98.77	99.989	
V. mengtzeanum	97.013	96.998	98.742	98.436	97.008	98.435	98.551	98.771	99.84	99.847
V. patulum	99.869	99.871	96.738	96.466	99.887	96.463	96.558	96.769	96.947	96.954

Nucleotide diversity of highly variable regions was calculated with a sliding window to estimate the divergence level of different regions in the analyzed plastomes. Of these, the SSC region exhibited the highest divergence levels (0.01469) and IR regions had the lowest (0.00218). Meanwhile, all highly divergent fragments were found in the SC regions whereas no highly variable loci were found in the IR regions. Seven regions with the highest variability, including 4 intergenic regions (*trnS GCU-trnG UCC*, *rpoB-trnC GCA*, *psbC-psbZ*, and *trnT UGU-trnL UAA*) and three gene regions (*psbI*, *ndhF*, and *ycf1*), were identified as most promising gene fragments for species identification and studying *Veratrum* phylogeny (Fig. 6). The *rpoB-trnC GCA* intergenic region was the most variable (Pi=0.03576), followed by *pabC-psbZ* (Pi=0.03477), *trnT UGU-trnL UAA* (Pi=0.03278), *psbI* (Pi=0.02914), *ycf1* (Pi=0.02896), *ndhF* (Pi=0.02097), and *trnS GCU-trnG UCC* (Pi=0.02533) (Table S6).

Although the IR region of the twelve chloroplast genomes was highly conserved, genomic structure and size varied in the twelve *Veratrum* cp genomes and the IR/SC border regions of these species were also different (Fig. 7). *V. dahuricum*, *V. grandiflorum*, *V. nanchuanense and V. patulum* exhibited larger than the other species in plastome size due to the increased IR length. Six genes (*rpl22*, *rps19*, *trnH*, *rpl2*, *ycf1*, and *ndhF*) were found in the LSC/IR and SSC/IR borders of the twelve plastomes. The *rps19* gene was found to be 3 bp away from the JLB junction in *V. dahuricum, V. grandiflorum, V. patulum* and *V. nanchuanense*, while it was across the IRb/SSC border in other species. The SSC/IRb junction is located in the *ycf1* region in the chloroplast genomes of all *Veratrum* species and extends a different length (5,372-5,408 bp) into the SSC region in all genomes; the IRb region includes 959 to 991 bp of the *ycf1* gene. In addition, the *trnH* gene is located in the LSC region, 141-144 bp away from the IRa/LSC border in the twelve *Veratrum* chloroplast genomes species.

Phylogenetic relationship

Plastome-based phylogenetic analysis showed identical topology for ML and BI methods (Figure 8). All the nodes in the phylogenetic tree were highly supported (BS ≥ 98% and PP ≥ 0.86). Consistent with the previous analyses, Melanthiaceae was monophyletic with five strongly supported groups corresponding to previously circumscriped tribes Parideae (Trillieae), Xerophylleae, Heloniadeae, Chionographideae and Melanthieae [1,2]. Within Melanthiaceae, Melanthieae was sister to the rest of the family, and Parideae (Trilliaceae) and Xerophylleae comprised a clade sister to the Heloniadeae and Chionographideae clades (BS=100, PP=1).

As shown in the cladogram, thirteen species of *Veratrum* formed one clade (Fig. 8, clade A) with highly support (BS=100%, PP=1) and this clade was subdivided into two sub-clades, labeled B and C. Clade B, corresponding to Loesener's section *Veratrum*, included *V. grandiflorum*, *V. nanchuanense*, *V. patulum* and *V. dahuricum* [22-24]. Clade C, corresponding to Loesener's section *Fuscoveratrum*, consisted of two subclades D and E, which were in accordance with the elements of Zomlefer's two subsections [1,2,22-24]. Within clade D, *V. taliense*, *V. stenophyllum* and *V. mengtzeanum* were clustered together with high support. Clade E included remaining six species of *Melanthium* s.l. sensu Bodkin, the taxa of the *V. maackii*, *V. nigrum*, and relatives [18]. Two accessions of *V. japonicum*, one from South Korea (NC_041306) and another one from Yunnan, China didn't form a monophyletic clade while were clustered together with *V. schindleri* (Fig. 8, clade G) (Table S1).

Discussion

Chlorplast Genome Features

The *Veratrum* plastomes showed typical structural characteristics and genetic properties of the angiosperm plastome. The plastomes of *Veratrum* taxa are approximately 151 kb in length on average and organized into quadripartite regions with no structural variation among taxa. In total, 135 genes were identified, including 83-84 protein-coding genes, 38 tRNA genes, and 8 rRNA genes. Confirming previous reports [38,43] we have found the losses of the *rps16* gene in *Veratrum* species, which might be a specific feature of the *Veratrum* genus compared to the other genera of Liliales. In addition, it was found that *ycf68* and *ycf15* are pseudogenes not encoding any protein and apparently non-functional. We detected *infA* in four *Veratrum* plastomes (*V. grandiflorum*, *V. nanchuanense*, *V. nigrm* and *V. patulum*) and it contained an internal stop codons indicating that this sequence is from a pseudogene. It is believed that the most common gene loss in angiosperm, *infA* loss could result from transferring of the gene to the nucleus [59,60].

SSRs (1–6 repeating sequences) were distributed throughout the plastomes, have been used for analysis of population genetics due to their high variability and stable reproducibility [61-63]. Here, 762 SSRs were identified in the *Veratrum* species, which will be useful developing lineage-specific cpSSR markers and population genetics of the *Veratrum* species. With the increasing number of chloroplast genome sequences, many researchers have demonstrated that large repeat sequences play an important role in insertion/deletion mismatches and recombination of genomic variation [63]. Additionally, these repeat motifs might provide some informative sources to develop genetic markers for analysis on population genetics [64].

The structural integrity of the whole plastome is highly linked to the IR structure, and the changes in plastome structure are often associated with IR expansions and contractions [65]. In this study, a detailed comparison of four junctions of two IRs between twelve species of *Veratrum* were performed. There were some variations in the size and distribution of *rps19*, *ndhF*, *ycf1*, and *trnH* genes among the four boundaries. Overall the two IR regions are highly conserved and structure variation was not significant in the twelve *Veratrum* plastomes

Identification of highly variable regions

DNA barcodes and gene markers derived from divergence hotspots in the cp genome are reported to be efficient in species identification of closely related species in the plant kingdom [66]. The plastomes of twelve *Veratrum* species were relatively conserved, though each of them contained unique genetic information and variant sites. Previously two fragments (*ndhF*, and *trnS-trnG*) were suggested to be useful for the study of *Veratrum* phylogeny [33,42]. We detected several variant sites, namely genes *psbI*, *ycf1*, and *ndhF* and four intergenic spacer regions *rpoB-trnC*, *trnT-trnL*, *trnS-trnG*, and *psbC-psbZ* useful as potential barcoding regions. Future studies will clarify the efficacy of the above barcoding regions for species delimitation and identification in *Veratrum*.

Phylogenetic *inferences*

The topology of our phylogenetic tree for 27 species of Melanthiaceae (circumscribed based on APG II) was identical with the one of Zomlefer et al. and Kim et al. [35,42,67], and had improved resolution for previously defined five tribes. The monophyly of *Veratrum s.l.* (clade A) including *Melanthium* was confirmed with high support. Our results also suggest that broad circumscription of *Veratrum* advocated previously is appropriate [1,2].

The phylogenetic topology for *Veratrum* is consist of previous infrageneric circumscriptions with clade B responding to sect. *Veratrum* and clade C responding to sect. *Fuscoveratrum* [1,22-24]. Morphologically, species of sect. *Veratrum* has cauline leaves, styles terminal on the fruiting carpels and bulb layers disintegrating into fibers. All species of clade B exhibited these characters. The species composition of the sect. *Veratrum* was compatible with the circumscriptions proposed by Nakai and Loesener [22-26]. *Meanwhile, V. nanchuanense*, with the unclear systematic position since its nomenclature [4,5], clustered with *V. grandiflorum* closely (Fig 8, clade B). Its morphological characteristics, such as leafy stems and pubescent many-branched inflorescences, are identical with the circumscriptions of Loseners's sect. *Veratrum* [22-24]. Therefore, we suggested *V. nanchuanense* was a member of the sect. *Veratrum*. Sect. *Fuscoveratrum* differs from the sect. *Veratrum* by its leaves that are mostly basal and reduced upward on the stem, styles sublaterally attached to the fruiting, and bulb layers disintegrating into reticulated fibers [25,26]. In this study, sect. *Fuscoveratrum* (Fig. 8, clade C) were separated into two subsections: subsect. *Pseudoanticlea* (Fig. 8, clade D) and subsect. *Asiaveratrum* (Fig. 8, clade E), and which is consistent with the infrageneric circumscription of Zomlefer [1].

V. stenophyllum, *V. taliense*, *V. mengtzeanum* and *V. grandiflorum* had been used as the original plant of traditional Chinese medicine namely “Pimacao” to treat bruise, rheumatic pain, wound hemostasis, which listed in the Standard of Chinese Herbal Pieces in Yunnan Province [9]. Phylogenetic analysis indicated that *V. grandiflorum* belongs to sect. *Veratrum*, while *V. stenophyllum*, *V. taliense* and *V. mengtzeanum* belong to sect. *Fuscoveratrum*. These species were placed into two distinct clades (Fig. 8, clades B and C). In general, closely related species also had similar chemical composition and efficacy, conversely, there could be different effects [68]. Thus, additional studies should be imperative to investigate the chemical composition and pharmacological effect of *V. grandiflorum* to solve the safety of medication concerns.

Till now, the taxonomic treatment of subsect. *Asiaveratrum* is uncertain, due to either the morphological synapomorphies or lack of resolution from molecular phylogenetic analysis based on sequences of the internal transcribed spacers (ITS) of the nuclear ribosomal DNA [1,6,33]. In previous studies, all species of subsect. *Asiaveratrum* formed an unresolved polytomy among *V. maackii*/ *V. nigrum*, *V. japonicum* and *V. schindleri*. In this study, high resolutions were detected for these taxa. *V. maackii* and *V. nigrum* were closely related, and they formed the sister clade of *V. oblongum* with strong support (BS=100%, PP=0.859). Hence, we suggested the taxonomic treatment of *V. nigrum* to be a distinct species from *V. maackii*. The cladogram exhibited two accessions of *V. japonicum* formed paraphyly *with one accession* clustered together with *V. schindleri* (Fig. 8, clade G). Therefore, we advocate treating *V. japonicum* as the synonym of *V. schindleri* corresponding to the Flora of China [5].

Conclusions

In this study, the complete chloroplast genome was sequenced, assembled, and annotated for ten *Veratrum* species from China. The ten plastomes exhibited a typical circular quadripartite structure and shared a high similarity in gene order and genomic structure. SSRs, LSRs, and the seven highly variable loci were identified across the *Veratrum* plastid genomes, which could serve as potential markers for future study on phylogenetic and population genetic studies. The monophyly of *Veratrum* s.l. including *Melanthium* was confirmed and two sections were exhibited in the phylogenetic analysis. The circumscription and relationships of infrageneric taxa of *Veratrum* were well evaluated, too. We suggest *V. nanchuanense* was a member of the sect. *Veratrum, and V. japonicum* was the synonym of *V. schindleri. In addition, V. nigrum* and *V. maackii should be treated as* distinct species. Overall, the comparative complete chloroplast genome analysis in this study provides valuable insight for species clarification, phylogenetic construction for the genus *Veratrum*, and will also conduce to the phylogenetic studies of Melanthiaceae.

Abbreviations

cp: chloroplast; LSC: large single-copy; SSC: small single-copy; IR: inverse repeat; NCBI: National Center for Biotechnology Information database; tRNA: transfer RNA; rRNA: ribosomal RNA

Declarations

Ethics approval and consent to participate

Not applicable

Consent to publish

Not applicable

Availability of data and materials

All data generated or analyzed during the course of this study are included in this document or obtained from the appropriate author(s) at reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

The present study was partially supported by the High-end talent of Yunnan province ‘Youth Top-notch Talent Support Program’ (2020) to Guo-dong Li; the Reserve Talent for Academic and Young Technical Leaders of Yunnan Province (2020) to Guo-dong Li; the Yunnan Provincial Science and Technology Department - Applied Basic Research Joint Special Funds of Yunnan University of Traditional Chinese Medicine [no. 2018FF001(-039)], and the major project of science and technology in Yunnan Province [2017ZF002].

Authors’ contributions

ZQ and GL conceived and designed the research. GL, CY, YZ and XT collected the leaf materials. YZ, LH and YL performed the experiments; YZ, YL and CY analyzed the data. YZ, LH and XT drafted the manuscript. GL revised the manuscript. All authors read and approved the final manuscript

Acknowledgements

We are very grateful to prof. Yong-Hong Zhang for comments, and technical advice.

References

1. Zomlefer WB, Whitten WM, Williams NH, Judd WS. An Overview of *Veratrum* s.l. (Liliales: Melanthiaceae) and an Infrageneric Phylogeny Based on ITS Sequence Data. Syst Bot. 2003;28(2):250–69.
2. Zomlefer WB, Whitten WM, Williams NH, Judd WS. Generic Circumscription and Relationships in the Tribe Melanthieae (Liliales, Melanthiaceae), with Emphasis on Zigadenus: Evidence from ITS and *trnL-F* Sequence Data. Am J Bot. 2001;88(9):1657–69.
3. An update of the Angiosperm Phylogeny Group classification for the. orders and families of flowering plants: APG IV. Bot J Linn Soc. 2016;181(1):1–20.
4. Flora of China editorial Committee. Flora Reipublicae Popularis Sinicae (*Veratrum*). In: Wu ZY, Raven PH, Hong DY, editors. Beijing, and St. Vol. 14. Louis: Science Press and Missouri Botanical Garden Press; 1980. pp. 19–30.
5. Flora of China editorial Committee. Flora of China (*Veratrum*). In: Chen XQ, Hiroshi T, editors. Beijing, and St. Vol. 24. Louis: Science Press and Missouri Botanical Garden Press; 2000. 24: 82–5.
6. Liao WJ, Yuan YM, Zhang DY. Biogeography and evolution of flower color in *Veratrum* (Melanthiaceae) through inference of a phylogeny based on multiple DNA markers. Plant Syst Evol. 2007;267(1–4):177–90.
7. An Editorial Committee of the Administration Bureau of Traditional Chinese Medicine. Chinese Materia Medica (Zhonghua Bencao), Shanghai Science & Technology Press: Shanghai, China. 1999; Vol. 8: 183.
8. Nanjing University of Traditional Chinese Medicine. Dictionary of traditional Chinese Medicine. Shanghai Science and Technology Press. 2006; 3801–3804.
9. Yunnan Provincial Department of health. Drug standard of Yunnan Province. 1998; 67–68.
10. Yin ZL, Yin ZL, Xie H, Zhang J. A Preliminary Study of Biological Characteristics of *Veratrum nigrum*. Yunnan journal of traditional Chinese medicine and Materia Medica. 2014; Vol.4; 54–57.
11. Li Q, Yang KX, Zhao YL, Qin HJ, Yang XW, Liu L, Liu YP, Luo XD. Potent anti-inflammatory and analgesic steroidal alkaloids from *Veratrum taliense*. Journal of Ethnopharmacology, 2016, 179.
12. Chandler CM, McDougal OM. Medicinal history of North American *Veratrum*. Phytochemistry reviews: proceedings of the Phytochemical Society of Europe, 2014, 13(3): 671–694.
13. Cong Y, Wu YT, Shen S, Liu XP, Guo JG. A Structure-Activity Relationship between the *Veratrum* Alkaloids on the Antihypertension and DNA Damage Activity in Mice. Chem Biodivers. 2020;17(2):e1900473.
14. Turner MW, Rossi M, Campfield V, French J, Hunt E, Wade E, McDougal OM. Steroidal alkaloid variation in *Veratrum californicum* as determined by modern methods of analytical analysis. Fitoterapia. 2019;137:104281.
15. Kikkawa HS, Tsuge K, Kubota S, Aragane M, Ohta H, Sugita R. Species identification of white false hellebore (*Veratrum album* subsp. *oxysepalum*) using real-time PCR. Forensic Sci Int. 2017;275:160–6.
16. Zimmerman JH. A Monograph of *Veratrum*. Ph. D. thesis. University of Wisconsin, Madison, WI, USA. 1958.
17. Kupchan SM, Zimmerman JH, Afonso A. The alkaloids and taxonomy of *Veratrum* and related genera. Lloydia. 1961;24:1–26.
18. Bodkin NL A revision of North American *Melanthium* L. (Liliaceae). Ph.D. dissertation. University of Maryland, College Park. 1978.
19. Zomlefer WB. The Genrea of Melanthiacea in the southeastern united states. Harvard Papers in Botany,1997, 2(2): 133–177.
20. Linnaeus C, Stockholm, Holmiae IL. Salvii. Species Plantarum. 1753; Vol. 1 and 2.
21. Baker JG. A synopsis of Colchicaceae and the aberrant tribes of Liliaceae. Botanical Journal of Linnean Society. 1879; 17: 405–510.
22. Loesener O. Studien über die Gattung *Veratrum* und ihre Verbreitung. Verhandlungen des botanischen Vereins der Provinz Brandenburg. 1926;68:105–66.
23. Loesener O. Übersicht über Arten der Gattung *Veratrum*, Teil I. Repertorium Specierum Novarum Regni Vegetabilis. 1927;24:61–72.
24. Loesener O. Übersicht über Arten der Gattung *Veratrum*. Schluss Repertorium Specierum Novarum Regni Vegetabilis. 1928;25:1–10.
25. Nakai T. Species generic Veratri in regione Manshurico- Koreano sponte nascentes. Rep Inst Sci Res Manchoukuo I (9):325–344.
26. Nakai T. Japanese species of *Veratrum*. Journal of Japanese Botany. 1937a;13:631–45, 701–713.
27. Tamura MN. Melanthiaceae. In: Kubitzki K, editor. The families and genera of vascular plants. Flowering plants: monocotyledons. Vol3. Berlin: Springer; 1998. pp. 369–80.
28. Mathew B. A Review of *Veratrum*. Plantsman. 1989;11:35–61.
29. Ohwi J. Flora of Japan. Shibundo. Tokyo. 1965; 284–285.
30. Lee NS. A taxonomic study of Korean *Veratrum*. Korean Journal of Plant Taxonomy. 1985;15:49–65.
31. Takada J, Kawanobe H. Distribution and variation of *Veratrum maackii* in Japan. Japanese Journal of Botany. 1996;71:1–28.
32. Satake Y. Veratrum. In: Satake Y, Ohwi J, Kitamura S, Watari S, Tominari T, editors. Wild flowers of Japan, herbaceous plants. Vol. 1. Tokyo: Heibonsha; 1982. pp. 28–9.
33. Kim JO, Tamura MN, Fuse S, Lee NS. Taxonomic status and phylogeny of *Veratrum* section *Veratrum* (Melanthiaceae) in Korea and Japan based on chloroplast and nuclear sequence data. Plant Systematics and Evolution,2014,300: 75–89.

34. Kikuchi R, Park JH, Takahashi H, Maki M. Disjunct distribution of chloroplast DNA haplotypes in the understory perennial *Veratrum album* ssp. *oxysepalum* (Melanthiaceae) in Japan as a result of ancient introgression. New Phytol. 2010;188:879–91.
35. Zomlefer WB, Whitten WM, Williams NH, Judd WS. Infrageneric phylogeny of *Schoenocaulon* (Liliales: Melanthiaceae) with classification of cryptic species based on ITS sequence data and geographical distribution. Am J Bot. 2006;93:1178–92.
36. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc Natl Acad Sci USA. 2010;107:4623–8.
37. Sugiura M. The chloroplast genome. Plant Mol Biol. 1992;19:149–57.
38. Yang LF, Yang ZY, Liu CK, He ZS, Zhang ZR, Yang J, Liu HY, Yang JB, Ji YH. Chloroplast phylogenomic analysis provides insights into the evolution of the largest eukaryotic genome holder, Paris japonica (Melanthiaceae). BMC plant biology. 2019;19(1):293.
39. Gu CH, Ma L, Wu ZQ, Chen K, Wang YX. Comparative analyses of chloroplast genomes from 22 Lythraceae species: inferences for phylogenetic relationships and genome evolution within Myrtales. BMC plant biology. 2019;19(1):281.
40. Xie DF, Yu HX, Price M, Xie C, Deng YQ, Chen JP, Yu Y, Zhou SD, He XJ. Phylogeny of Chinese *Allium* Species in Section Daghestanica and Adaptive Evolution of *Allium* (Amaryllidaceae, Allioideae) Species Revealed by the Chloroplast Complete Genome. Frontiers in plant science. 2019;10:460.
41. Yang Z, Zhao TT, Ma QH, Liang LS, Wang GX. Comparative Genomics and Phylogenetic Analysis Revealed the Chloroplast Genome Variation and Interspecific Relationships of *Corylus* (Betulaceae) Species. Frontiers in plant science. 2018;9:927.
42. Kim SC, Kim JS, Mark W, Chase MF, Fay, Kim JH. Molecular phylogenetic relationships of Melanthiaceae (Liliales) based on plastid DNA sequences. Botanical Journal of the Linn Society. 2016;181:567–84.
43. Do HDangK, Kim JS, Kim JH. Comparative genomics of four Liliales families inferred from the complete chloroplast genome sequence of *Veratrum patulum* O. Loes (Melanthiaceae). 2013;530(2):229–35.
44. Han LJ, Liu YY, Zhang YM, Yang CW, Qian ZG, Li GD. The complete chloroplast genome and phylogenetic analysis of *Veratrum mengtzeanum* Loes. F. (Liliaceae). 2019, 4(2): 4170–4171.
45. Doyle JJ, Davis JI, Soreng RJ, Garvin D, Anderson MJ. Chloroplast DNA inversions and the origin of the glass family (Poaceae). PNAS CTAB. 1992;89:7722–6.
46. Patel RK, Jain M. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. PLos One. 2012;7:e30619.
47. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017;45:e18.
48. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 2004;20(17):3252–5.
49. Greiner S, Lehwark P, Bock R. Organellar Genome DRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47:W59–64.
50. Stefan K, Choudhuri JV, Enno O, Chris S, Jens S, Robert G. Reputer: the manifold applications of repeat analysis on a genomic scale. Nuclc Acids Research.2001. (22), 4633–4642.
51. Suresh B. Mudunuri, Nagarajaram HA. IMEx: Imperfect Microsatellite Extractor Bioinformatics. 2007;23(10):1181–7.
52. BehunaS K, Severson DW. Codon usage bias: causative factors, quantification methods and genome-wide patterms: with emphasis on insect genomes. Biol Rev. 2013;88(1):49–61.
53. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9.
54. Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, Batzoglou S. Glocal Alignment: Finding Rearrangements During Alignment, 2003. Bioinformatics, 19S1: i54-i62.
55. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue): W273-9.
56. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. Mol Biol Evol. 2017;34:3299–302.
57. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30(9): 1312–3.
58. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogeny. Bioinformatics. 2001;17:754–5.
59. Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao TN, Heggie L, Kavanagh AT, Hibberd JM, Gray JC, Morden CW, Calie PJ, Jermiin LS, Wolfe HK. Many Parallel Losses of *infA* from Chloroplast DNA during Angiosperm Evolution with Multiple Independent Transfers to the Nucleus. Plant Cell. 2001;13(3):645–58.
60. Cummings HS, Hershey JW.. Translation initiation factor IF1 is essential for cell viability in *Escherichia coli*. J Bacteriol. 1994;176:198–205.
61. Powell W, Morgante M, Mcdevitt R, Vendramin GG, Rafalski JA. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc Natl Acad Sci USA. 1995;92:7559–763.
62. Pauwels M, Vekemans X, Godé C, Frérot H, Castric V, Saumitoulaprade P. Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, *Arabidopsis halleri* (Brassicaceae). New Phytol. 2012;193:916–28.
63. Asaf S, Khan AL, Khan MA, Shahzad R, Lubna, Kang SM, Alharrasi A, Alrawahi A, Lee IJ. Complete chloroplast genome sequence and comparative analysis of *Loblolly pine* (Pinus taeda L.) with related species. PLOS ONE. 2018;13(3):e0192966.
64. Nie XJ, Lv SZ, Zhang YX, Du XH, Wang L, Biradar SS, Tan XF, Wan FH, Song WN. Complete chloroplast genome sequence of a major invasive species, *Crofton weed* (Ageratina adenophora). PLOS ONE. 2012;7:e36869.

65. Kim KJ, Lee HL. Complete chloroplast genome sequences from Korean ginseng (Panax schinseng Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res. 2004;11:247–61.
66. Li XW, Yang Y, Henry RJ, Rossetto M, Wang YT, Chen SL. Plant DNA barcoding: from gene to genome. Biol Rev. 2015;90:157–66.
67. APG II. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants. Bot J Linn Soc. 2003;141:399–436.
68. Ma HF, Sima YK, Zhang D, Yang JY, Xu T. Chemical Constituents of the Volatile Oils from the Leaves of Thre Michelia Species. Journal of Northwest Forestry University. 2019;34(4):212–6.

Figures

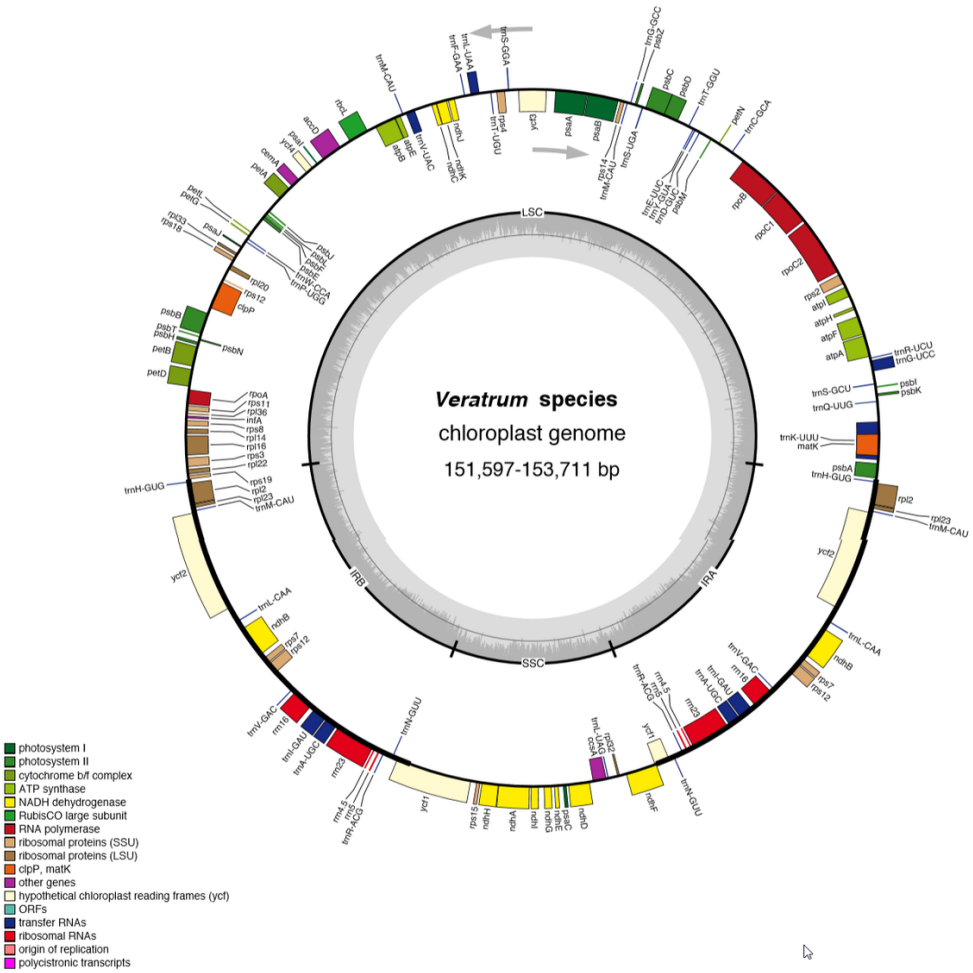

Figure 1

Gene map of the complete chloroplast genomes from Veratrum species. Genes outside the outer circle are transcribed clockwise and those inside are counter-clockwise. The colored bars indicate different functional group. The dark gray inner circle corresponds to the GC content, the light-gray to the AT content.

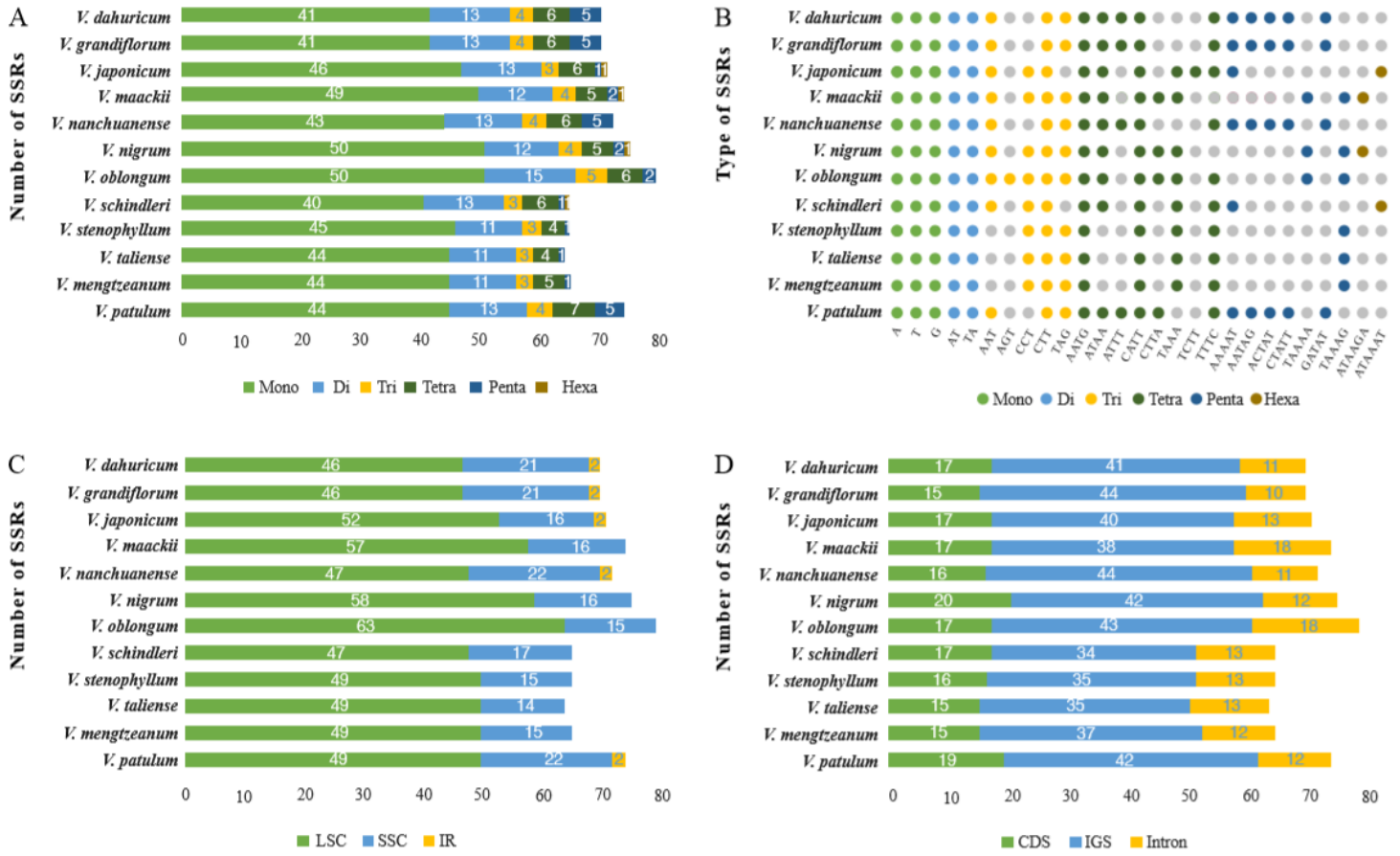

Figure 2

Analysis of simple repeat sequences (SSR) in twelve Veratrum chloroplast genomes. (A) Number of six SSRs repeat. (B) Type of shared SSRs among the twelve cp genomes. (C) Number of SSRs in LSC, SSC, and IR regions. (D) Number of SSRs in IGS, CDS, and Intron.

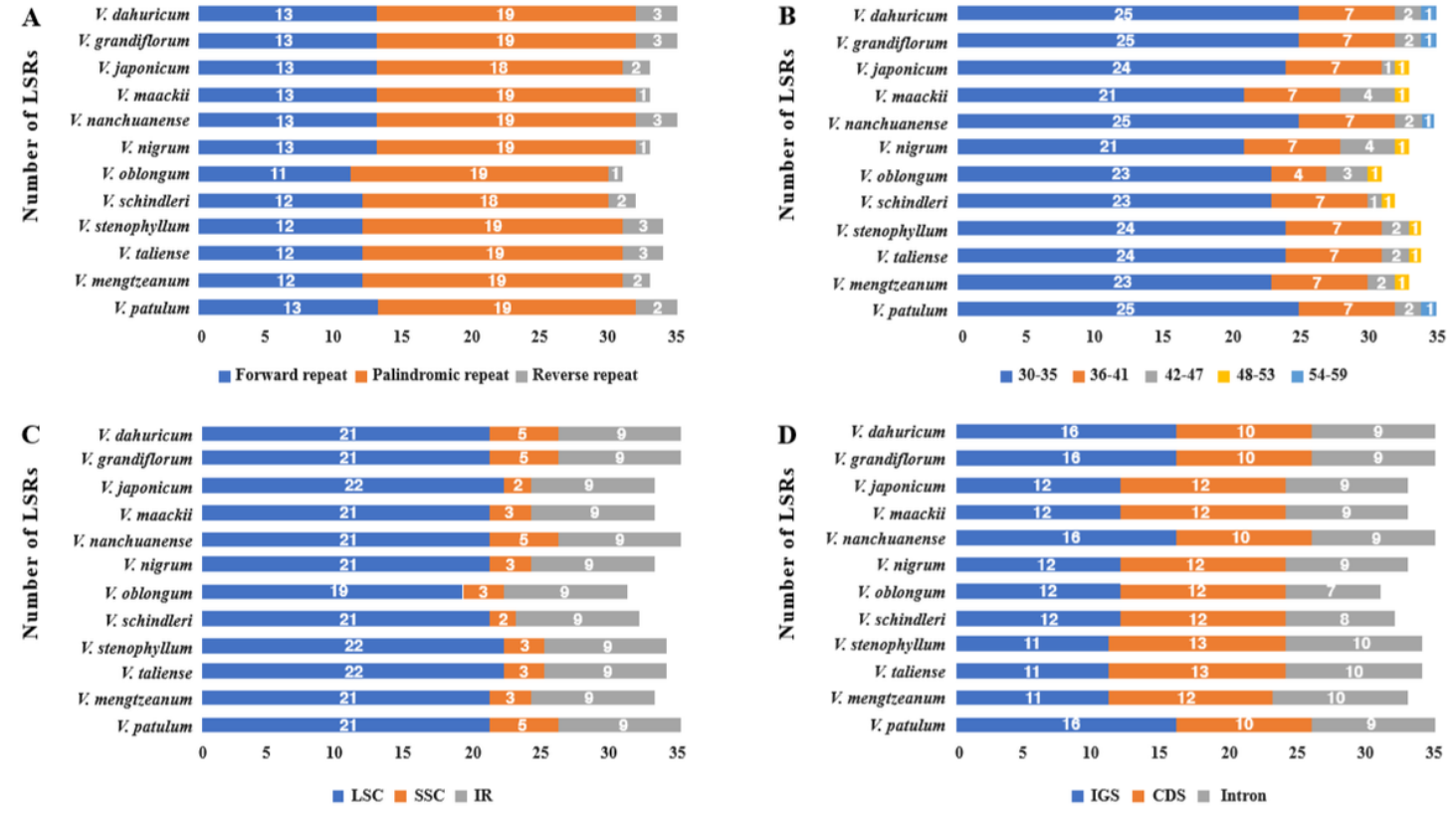

Figure 3

Analysis of large repeat sequences (LSR) in twelve Veratrum chloroplast genomes. (A) Number of three LSRs repeat. (B) Number of LSRs by lengths. (C) Number of LSRs in LSC, SSC, and IR. (D) Number of LSRs in IGS, CDS, and Intron.

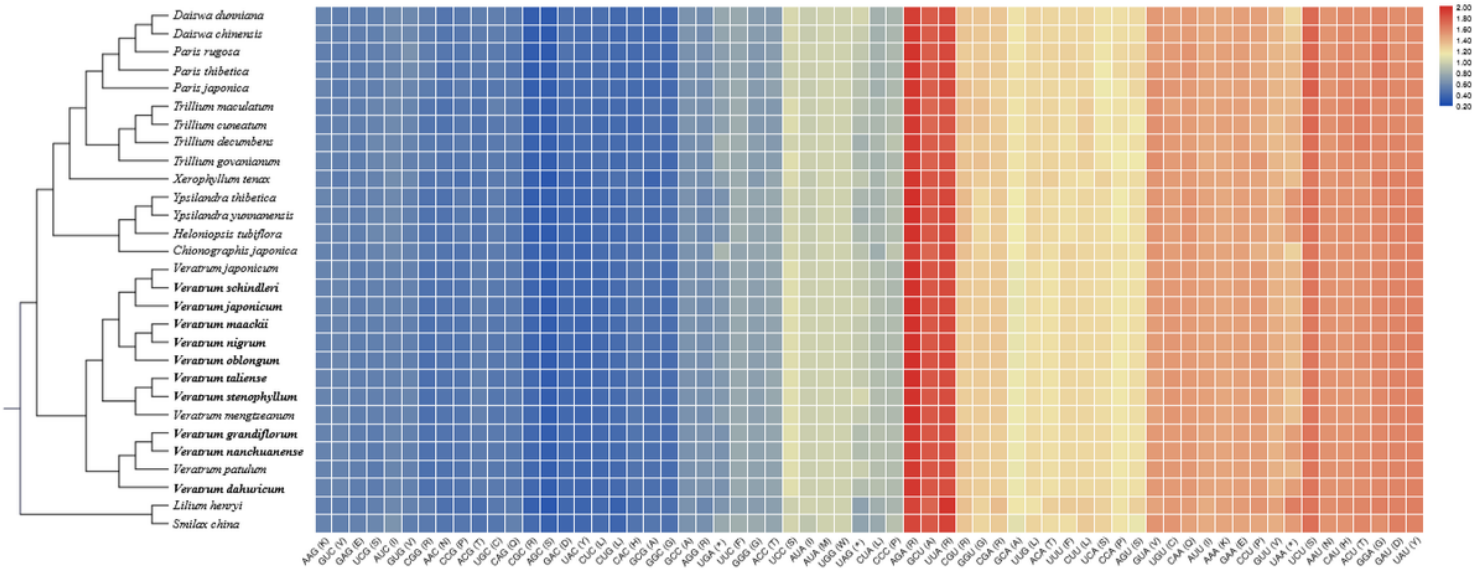

Figure 4

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

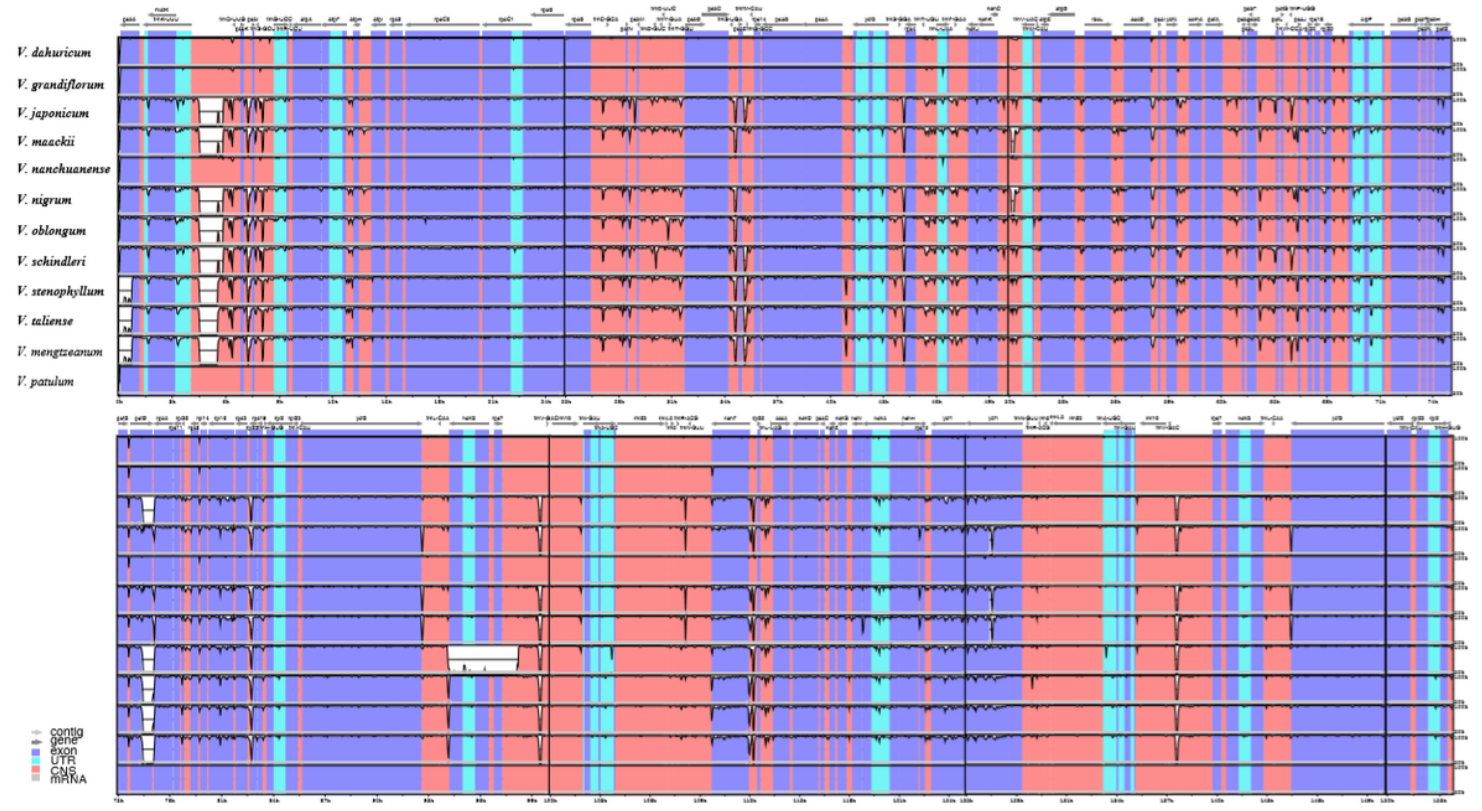

Figure 5

Comparison of chloroplast genomes of twelve Veratrum species. Gray arrow above the alignment indicate the direction of gene. The dark blue regions represent exons, the light-blue regions represent untranslated regions (UTRs), the pink regions represent Conserved Non-Coding Sequences (CNS). The y-axis represents the percent identity ranging from 50% to 100%.

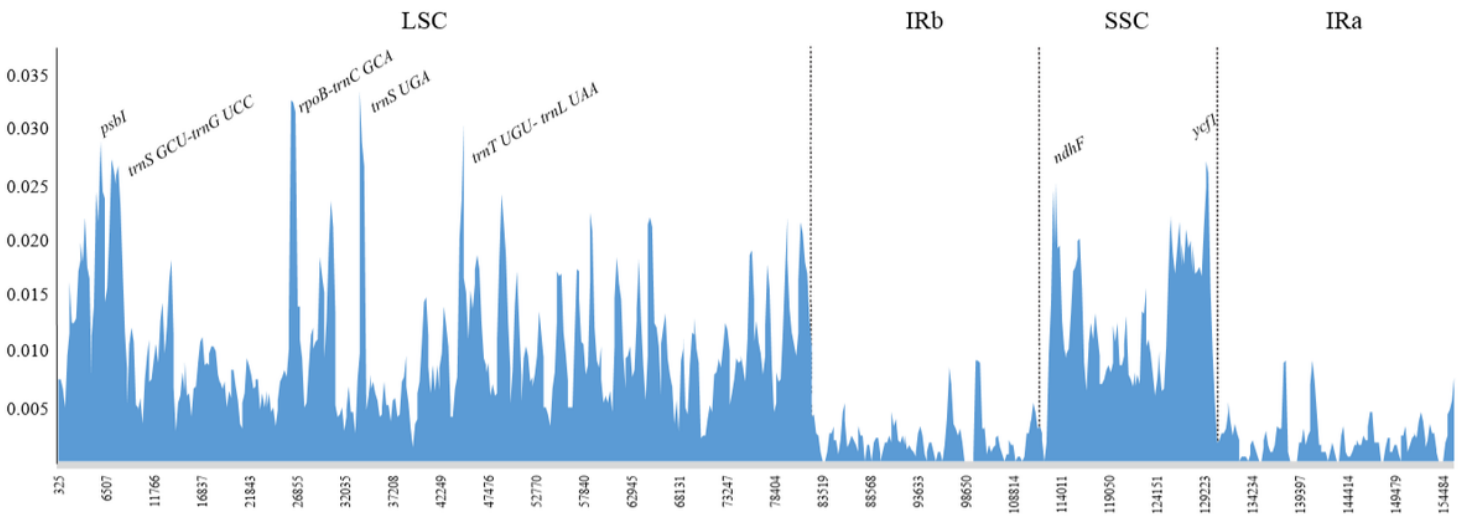

Figure 6

Sliding-window analysis on the cp genomes for twelve Veratrum species. X-axis: position of the midpoint of a window; Y-axis: nucleotide diversity (Pi) of each window.

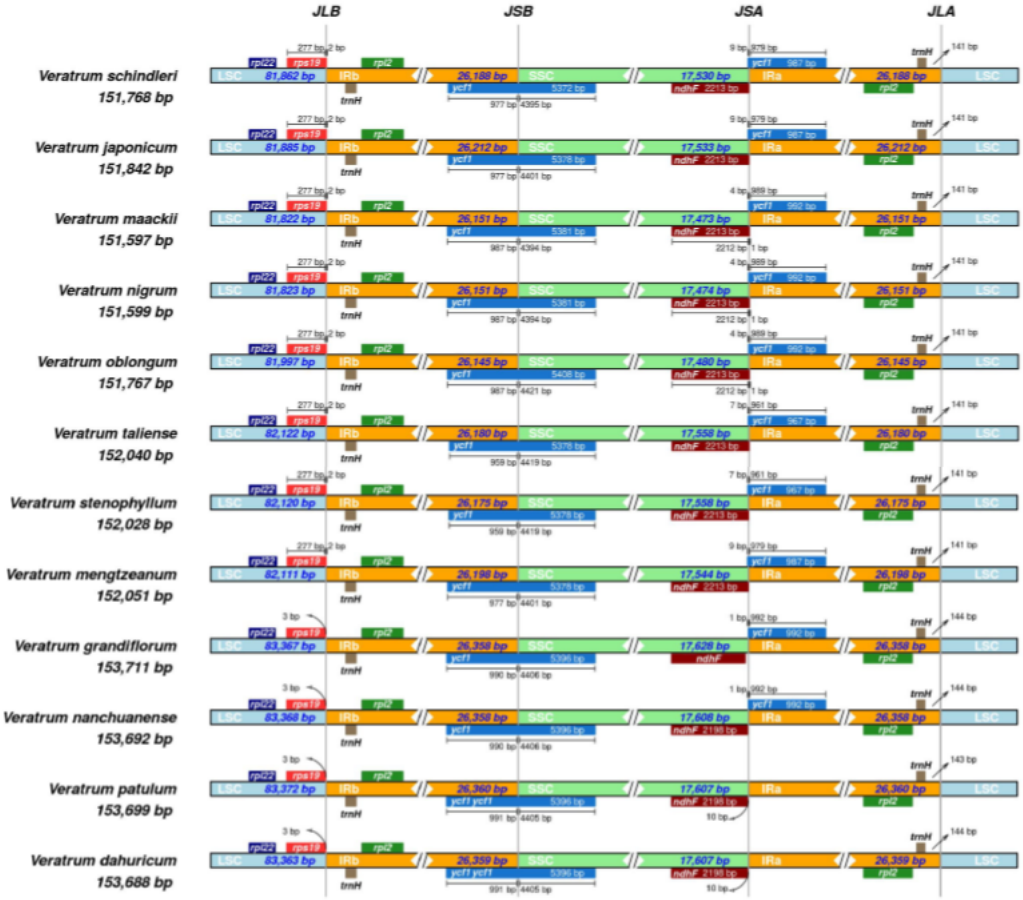

Figure 7

Comparison of the LSC/IRb/SSC/IRa junctions among the chloroplast genomes of Veratrum species.

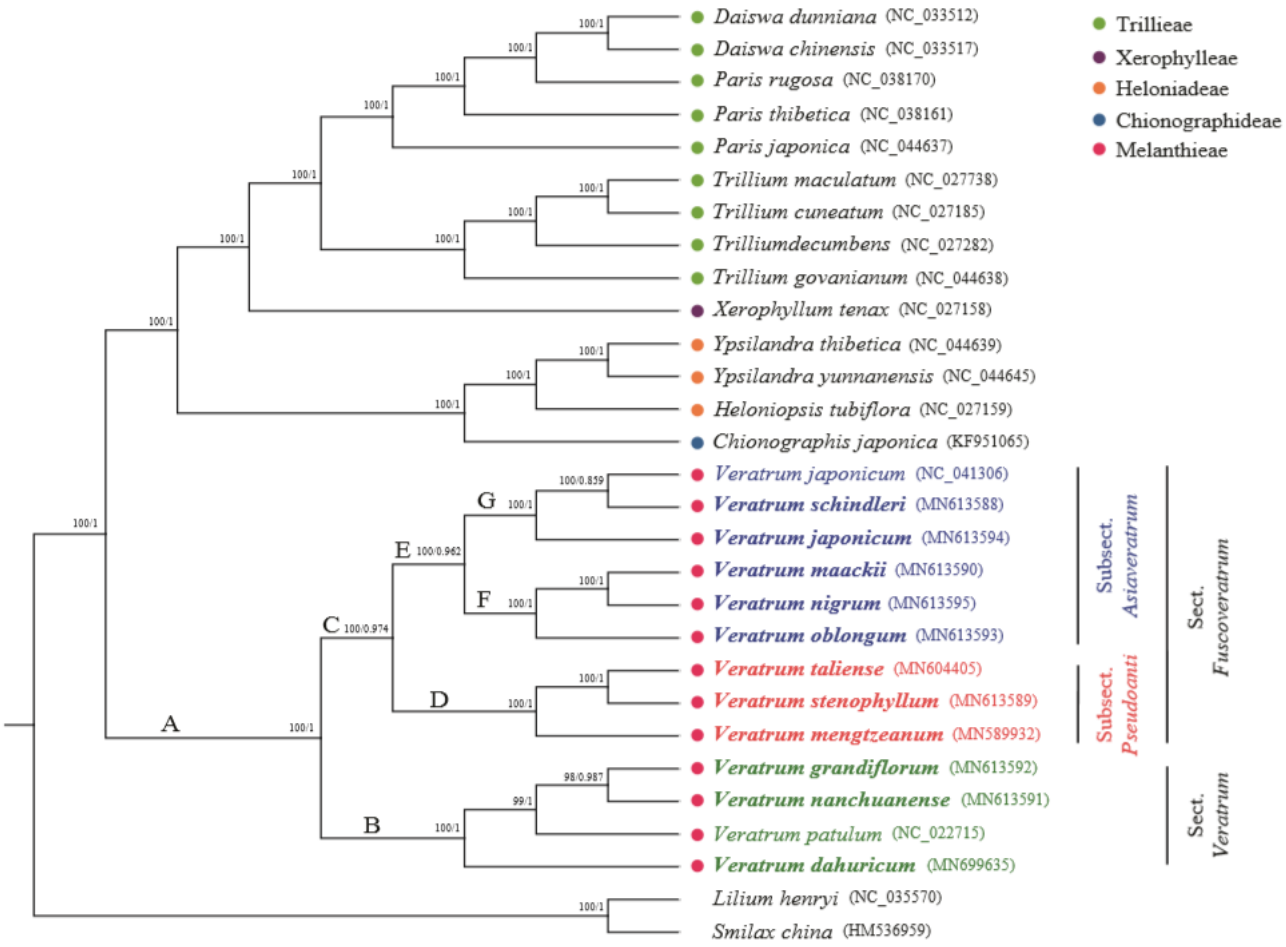

Figure 8

Phylogenetic relationships of Veratrum species with related 29 species based on the whole cp genomes. ML tree topology is shown and ML bootstrap values and BI posterior probability are indicated on the nodes. Colored lines and braces at the right of the tree indicate section and subgenus names of Veratrum. The color of the dot represents different tribes.

Supplementary Files

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

- TableS1.samplecollection.xlsx
- TableS2.SSRs.xlsx
- TableS3.LSRs.xlsx
- TableS4.usag.xlsx
- TableS5.Freq.xlsx
- TableS6.xlsx