Planta (2026) 264:37
https://doi.org/10.1007/s00425-026-05052-5

ORIGINAL ARTICLE

Comparative plastome analysis reveals structural variation, selection, and phylogenetic relationships in *Verbascum* species

Motahareh Akbari[1] · **Aboozar Soorni**[1]

Received: 11 February 2026 / Accepted: 10 June 2026

Main conclusion **The chloroplast genomes of Verbascum speciosum and V. sinuatum reveal lineage-specific structural variations and hypervariable loci (ycf3_1–trnS-GGA, rps15, matK) that serve as powerful markers for species identification. Signatures of positive selection in ndhB, ycf3, and ycf4 suggest adaptive evolution in plastid genes, while phylogenomic analyses confirm that plastome-scale data are essential for resolving species-level relationships in this medicinally valuable genus.**

Abstract The genus *Verbascum* (Scrophulariaceae) is both species-rich and notoriously difficult to classify, with many members valued for their medicinal properties but hampered by vague morphological boundaries and a lack of reliable molecular data. To address this, we sequenced and assembled the complete chloroplast genomes of two Iranian medicinal species, *Verbascum speciosum* and *Verbascum sinuatum*. The two plastomes measured 153,325 bp and 153,038 bp, respectively, each containing the typical quadripartite structure and 131 genes. Comparative work uncovered lineage-specific shifts at the inverted repeat boundaries, most notably involving the rpl23 gene, as well as clear differences in simple sequence repeat abundance, with *V. sinuatum* showing far fewer SSRs than its relative. Scanning across the genomes, we identified several hypervariable spots (*rps15*, *rps16*, *matK*, *ycf1*, and even some tRNA genes) that could serve as useful barcoding markers. While most coding regions are under strong purifying selection, we found signs of positive selection at specific codon sites within *ndhB*, *ycf3*, and *ycf4*. Phylogenomic analysis using whole plastome data strongly supported the monophyly of *Verbascum* and confidently placed both new species within the genus. In contrast, standard single-locus barcodes failed to resolve species-level relationships. This work provides the first complete cp genomes for these two species and demonstrates that plastome-scale data offer a real step forward for untangling *Verbascum* systematics, identifying informative markers, and guiding future taxonomic work.

Keywords Scrophulariaceae · Plastome evolution · Simple sequence repeats (SSRs) · Positive selection · DNA barcoding markers

Introduction

The Scrophulariaceae family is an important plant family, with reported numbers ranging from 62 genera and approximately 1830 recognized species (Christenhusz 2016; Yılmaz and Genç 2024) to 200 genera and 2500 species in total (Jamshidi-Kia et al. 2019). Within this family, *Verbascum* is the largest genus, comprising approximately 360 species globally. *Verbascum* species exhibit notable morphological diversity, ranging from annual, biennial, or perennial herbs to, rarely, small shrubs, with a characteristic hairy indumentum and flowers typically arranged in terminal racemes or panicles. The genus has a wide but uneven distribution, with its center of diversity located in Turkey and Iran, spanning across Europe, Asia, and Northeast Africa (Fischer 2004; Christenhusz 2016; Yılmaz and Genç 2024). This distribution has led to its naturalization in other regions, including the American continent (Yılmaz and Genç 2024). Turkey represents a primary hotspot for the genus, harboring 248 species and 129 hybrids, while other significant regional floras include 48 species in Flora Iranica and 51 species in the flora of the former USSR (Fedchenko et al. 1955; Güner and Aslan 2012; Karavelioğulları 2012; Sotoodeh et al. 2016).

Communicated by Stefan de Folter.

✉ Aboozar Soorni
soorni@iut.ac.ir

[1] Department of Biotechnology, College of Agriculture, Isfahan University of Technology, Isfahan 84156-83111, Iran

Iran is another key center of diversity, with 43 species and three hybrids, of which 20 are endemic (Sotoodeh et al. 2016; Jamshidi-Kia et al. 2019).

Phytochemical analysis confirms that *Verbascum* species synthesize a wide spectrum of secondary metabolites, including flavonoids, iridoid glycosides (e.g., aucubin, catalpol), phenylethanoid glycosides (e.g., verbascoside), saponins, monoterpene glycosides, and alkaloids. These compounds form the chemical basis for their medicinal uses (Tatli et al. 2003; Grigore et al. 2013; Jamshidi-Kia et al. 2019). Specifically, the flowers of Iranian *Verbascum* species are abundant in phenolic and flavonoid compounds, with high-performance liquid chromatography (HPLC) analysis identifying key constituents, such as chlorogenic acid, rutin, quercetin, caffeic acid, gallic acid, and apigenin, the concentrations of which vary significantly between species (Karamian and Ghasenlou 2015; Nofouzi 2015; Halimi and Nasrabadi 2018; Zengin et al. 2021; Amini et al. 2022). For instance, *Verbascum erianthum* was found to possess the highest amounts of chlorogenic acid, caffeic acid, and apigenin, while *Verbascum saccatum* was richest in gallic acid, rutin, and quercetin (Amini et al. 2022). Focusing on two other prominent species, targeted HPLC–DAD analysis reveals that *V. speciosum* contains notable levels of chlorogenic acid (0.838 g kg^{-1} DW) and quercetin (0.526 g kg^{-1} DW), whereas *V. sinuatum*, while possessing a lower total flavonoid content, was characterized by a significant concentration of quercetin (0.393 g kg^{-1} DW) but lower amounts of other quantified phenolics like gallic acid (0.015 g kg^{-1} DW) and chlorogenic acid (0.154 g kg^{-1} DW) (Amini et al. 2022). These specific phenolic compounds, particularly quercetin and chlorogenic acid, are critically linked to the plants' strong antioxidant capacity, as demonstrated by assays such as FRAP (Ferric-Reducing Antioxidant Power), where significant positive correlations were observed between antioxidant activity and total flavonoid content (TFC) as well as specific compounds like quercetin (Amini et al. 2022). This antioxidant activity is a key mechanism underpinning many of the reported pharmacological effects, including anti-inflammatory and antimicrobial actions (Jamshidi-Kia et al. 2019). The broader phytochemical profile for *V. speciosum* includes reports of macrocyclic spermine alkaloids like verbaspinoside and significant levels of saponins, while *V. sinuatum* is also noted for its high total phenolic content, which correlates with potent metal-chelating activities and a confirmed presence of antimicrobial saponins (Karamian and Ghasemlou 2013; Jamshidi-Kia et al. 2019). The pharmacological activities documented for these species and the genus broadly include expectorant, demulcent, anti-inflammatory, and diuretic effects traditionally used for respiratory, urinary, and skin disorders, which are substantiated by modern studies showing antibacterial, antifungal, and wound-healing properties (Jamshidi-Kia et al. 2019). However, the effective study and application of this valuable phytochemical diversity are fundamentally dependent on accurate species identification and a stable taxonomic framework.

Accurate species delimitation within the genus *Verbascum* (Scrophulariaceae) has long been hindered by complex taxonomic histories and reliance on labile morphological characters. Historical classifications controversially segregated genera, such as Celsia, Staurophragma, and Rhabdotosperma, based on traits like stamen number and seed morphology, leading to inconsistent and contested generic boundaries. While morphological studies, including analyses of indumentum and seed characteristics, have provided a foundational framework, they have proven insufficient for resolving deep-seated phylogenetic uncertainties and clarifying species statuses across the polymorphic genus (Hartl 1977; Grabias et al. 1991; Juan et al. 1997; Fischer 2004). Early molecular phylogenetic studies using the nuclear ITS region and a limited set of cp markers, such as *trnY-T, psbA-trnH*, *rbcL*, *matK*, and *trnL*, provided initial support for the monophyly of a broadly defined *Verbascum*, including the reintegration of Rhabdotosperma, but offered limited resolution for infrageneric relationships and species-level taxonomy (Remal Puigmal 2014; Ghahremaninejad et al. 2015; Sotoodeh 2015; Riahi and Ghahremaninejad 2019; Alzahrani et al. 2024). Consequently, to address these persistent limitations and robustly test taxonomic hypotheses, higher-resolution genomic approaches have become essential.

In this regard, cp genome sequencing has emerged as a powerful tool for elucidating phylogenetic relationships and genome evolution across species, genera, and families. It offers a rich source of informative characters compared to short standard loci. Initial cp genome reports for individual *Verbascum* species, such as *Verbascum chinense*, *Verbascum phoeniceum*, and *Verbascum thapsus*, confirmed the utility of plastome data for phylogenetic placement, consistently recovering a sister relationship between *Verbascum* and *Scrophularia* within Scrophulariaceae (Bi et al. 2020; He et al. 2020; Zhang et al. 2022). A landmark comparative phylogenomic study by Dong et al. (2022) leveraged 80 protein-coding genes from nine Verbascum plastomes, including seven newly sequenced species (*V. sinaiticum*, *Verbascum brevipedicellatum*, *V. thapsus*, *Verbascum chaixii*, *Verbascum songaricum*, *Verbascum phoeniceum*, and *Verbascum blattaria*) and two from public databases, to construct a robust phylogenetic framework. This study not only confirmed the monophyly of *Verbascum* and its sister relationship to *Scrophularia*, but also resolved interspecific relationships, revealing *V. blattaria* as sister to a clade containing the remaining species. Comparative genomic analysis identified highly variable regions (e.g., *trnH-psbA*, *rpl32-trnL*) suitable for DNA barcoding, detected signatures of positive selection in genes, such as *ycf1* and *ndhF*. They critically provided the phylogenetic evidence to reinstate *V.*

brevipedicellatum within *Verbascum*, rejecting its previous placement in the segregate genus Rhabdotosperma (Dong et al. 2022).

Despite this significant progress, published *Verbascum* plastomes remain sparse and are predominantly limited to short reports or studies with a strong sampling bias toward Eurasian taxa, including the recent comprehensive study by Dong et al. (2022) which focused on species from China (Dong et al. 2022). A critical manifestation of this deficit is the lack of comprehensive comparative plastomic data for species from major biodiversity hotspots, such as Iran, where taxonomic complexity is high and species-level relationships are poorly understood. To directly address this gap and advance the systematics of two key, yet genomically unexplored, Iranian medicinal species, we sequenced the complete cp genomes of *V. speciosum* and *V. sinuatum*. This study aimed to: (i) assemble, annotate, and characterize the first complete cp genomes for these species; (ii) perform a detailed comparative genomic analysis to elucidate patterns of codon usage bias, examine structural dynamics at inverted repeat boundaries, and characterize the distribution of simple sequence repeats (SSRs); (iii) assess genome-wide nucleotide diversity to identify mutation hotspots and screen protein-coding genes for signatures of positive selection; and (iv) reconstruct a high-resolution phylogenomic framework to determine the precise evolutionary placement of *V. speciosum* and *V. sinuatum* within the genus *Verbascum*. Furthermore, we evaluated the phylogenetic utility of individual candidate barcode regions (including *rps15* and *rps16*) against whole-genome sequence data to assess their efficacy for species identification and future population-level studies.

Material and methods

Plant material and DNA extraction

Fresh leaf specimens of *V. speciosum* and *V. sinuatum* were collected from the same collection sites and using the same voucher information as previously reported by Amini et al. (2021): *V. speciosum* from Bukan (36°47′11.06"N, 45°49′32.18"E; voucher UHDH-203) and *V. sinuatum* from Mahabad (36°44′16.87"N, 45°37′21.07"E; voucher UHDH-208), West Azerbaijan Province, Iran (Amini et al. 2022). Genomic DNA was isolated from 100 mg of flash-frozen leaf tissue employing the DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany). The quality of the extracted DNA was rigorously assessed through dual verification: integrity was confirmed via 1% agarose gel electrophoresis. Purity and concentration were quantified using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, USA), with acceptable thresholds set at A260/A280 ≥ 1.8 and A260/A230 ≥ 2.0. High-purity DNA meeting these criteria was used to construct Illumina TruSeq libraries with an average insert size of 350 bp. Sequencing was subsequently performed on an Illumina HiSeq 2000 platform, generating 2 × 150 bp paired-end reads.

Cp genome assembly and annotation

The raw sequencing reads were subjected to a quality control pipeline. Initial quality assessment was conducted with FastQC (v0.11.9), followed by the removal of adapter sequences and low-quality bases using Trimmomatic v 0.39 (Bolger et al. 2014). The resulting high-quality reads were de novo assembled into complete cp genomes using GetOrganelle v1.7.7.1 (Jin et al. 2020). The assembly graphs were visually inspected and validated for circularity and completeness using Bandage (Wick et al. 2015). The assembled genomes were annotated utilizing a dual-software approach with CPGAVAS2 (Shi et al. 2019) and GeSeq (Tillich et al. 2017) to ensure comprehensive feature prediction. Finally, circular graphical maps of the annotated cp genomes were generated using OGDRAW (Greiner et al. 2019).

Analysis of codon usage bias

Variation in synonymous codon use was assessed by calculating Relative Synonymous Codon Usage (RSCU) values. This analysis was performed using MEGA6 (Tamura et al. 2013), and the resulting codon bias patterns were visualized using the RSCU-Plot Shiny application (https://pcg-lab.shinyapps.io/RSCU-Plot/) (Akrami et al. 2025; Diani Gohar and Soorni 2025; Hejazi et al. 2025; Soorni and Golchini 2025; Ghotbi and Soorni 2026; Hejazi and Soorni 2026).

Comparative cp genome analysis and IR boundary assessment

The cp genomes of two *Verbascum* species were compared with those of eight additional species retrieved from the NCBI database (NC_085504, NC_064357, NC_085505, NC_051533, NC_050920, NC_064358, NC_085506, and MW751819) (Bi et al. 2020; He et al. 2020; Dong et al. 2022; Zhang et al. 2022), bringing the total number of *Verbascum* species included in our comparative analyses to ten. To investigate structural variations at the genomic level, the contraction and expansion dynamics of the inverted repeat (IR) regions, specifically the large single-copy (LSC), small single-copy (SSC), IRb, and IRa regions, were analyzed and visualized using IRScope (Amiryousefi et al. 2018). This included a detailed examination of the four junction points (JLB, JSB, JSA, and JLA) between the IR and single-copy regions.

Identification of simple sequence repeats (SSRs)

Microsatellite identification was conducted using CPStools (Huang et al. 2024) with the following minimum repeat unit thresholds: mononucleotides ≥ 10, dinucleotides ≥ 6, trinucleotides ≥ 5, and tetra-, penta-, and hexanucleotides ≥ 4.

Assessment of nucleotide diversity and selective pressure

Sequence polymorphism across protein-coding regions was evaluated by computing nucleotide diversity (Pi) values for all annotated genes using CPStools (Huang et al. 2024). To infer patterns of selective pressure, site models were employed within EasyCodeML (Gao et al. 2019) to compare four model pairs (M0 vs. M3, M1a vs. M2a, M7 vs. M8, and M7a vs. M8a). Likelihood ratio tests (LRTs) were conducted with a significance threshold of $p < 0.05$. The nonsynonymous-to-synonymous substitution rate ratio ($\omega = Ka/Ks$) was calculated for each gene. Signatures of positive selection acting on specific amino acid sites were inferred from significantly elevated ω values in conjunction with LRT outcomes.

Phylogenetic reconstruction and evaluation of marker resolution

Phylogenetic relationships within *Verbascum* were reconstructed based on established methodologies (Akrami et al. 2025; Diani Gohar and Soorni 2025; Hejazi et al. 2025; Soorni and Golchini 2025; Golchini and Soorni 2026). Whole cp coding sequences from the sequenced taxa were combined with publicly available genomes from the representative Scrophulariaceae family, using *Salvia species* as the outgroup (Zhang et al. 2022). Sequence alignment was performed with MUSCLE v3.8.1551 (Edgar 2004) under default parameters, followed by refinement using trimAl v1.4 (Capella-Gutiérrez et al. 2009) to eliminate ambiguously aligned regions (retention threshold: 95% site occupancy; minimum sequence similarity: 0.1%). The curated alignments were concatenated into a supermatrix using SequenceMatrix (Vaidya et al. 2011). Maximum likelihood inference was implemented in IQ-TREE (Nguyen et al. 2015) with the GTR + Γ model, selected via ModelFinder, and nodal support was assessed with 1000 ultrafast bootstrap replicates. The resulting topology was visualized and annotated in iTOL (Letunic and Bork 2021). To evaluate the discriminatory power of hypervariable markers, the *rps15* and *rps16* loci were extracted from the same taxonomic sampling and analyzed using identical alignment and phylogenetic protocols to assess their phylogenetic resolution.

Results

Cp genome assembly and annotation

A comparative analysis of cp genomes of *V. speciosum* and *V. sinuatum* was conducted to examine their structural features and functional gene composition. Both species retain the characteristic quadripartite architecture of angiosperm plastomes, consisting of a LSC region, a SSC region, and two inverted repeats (IRa and IRb) (Fig. 1). However, detailed boundary assessments revealed notable size variations in these genomic components. The plastome of *V. speciosum* measured 153,325 base pairs (bp), which is 287 bp longer than the 153,038 bp plastome of *V. sinuatum*. This difference was distributed across all genomic regions: the LSC measured 84,562 bp in *V. speciosum* versus 84,305 bp in *V. sinuatum* (a difference of 257 bp), the SSC measured 17,875 bp versus 17,821 bp (54 bp difference), and each IR measured 24,744 bp compared with 24,456 bp (288 bp difference per IR). These variations indicate that expansion or contraction events have occurred across the plastome, contributing to the overall size divergence. Both plastomes exhibited a GC content of 38%, and Bandage-based evaluation revealed mean coverage depths of ~ 174 × and ~ 162 × for *V. speciosum* and *V. sinuatum*, respectively.

Genome annotation revealed that both plastomes harbor an identical set of 131 genes (Table 1), which were assigned to well-defined functional groups. Among these, 45 genes encode components of the photosynthetic machinery, including 15 photosystem II subunits, five photosystem I subunits, six cytochrome b6/f complex subunits, 12 NADH dehydrogenase subunits, six ATP synthase subunits, and the large subunit of Rubisco (*rbcL*). The genetic system for self-replication is represented by 74 genes, encompassing ten large ribosomal subunit proteins, 14 small ribosomal subunit proteins, four RNA polymerase subunits, eight ribosomal RNAs (duplicated within the IRs), and 37 transfer RNAs. The remaining 12 genes, with diverse roles, include *matK*, *clpP*, *cemA*, *accD*, *ccsA*, *infA*, and seven conserved open reading frames (*ycf* genes). The strict conservation of gene content and functional categorization, despite detectable structural length polymorphisms, suggests that genome size differences between *V. speciosum* and *V. sinuatum* are restricted to non-coding intergenic spacer regions, while coding regions remain highly conserved under strong evolutionary constraint.

Comparative analysis of codon usage bias

To elucidate the patterns of codon usage bias in cp genomes of two *Verbascum* species, we calculated RSCU values for all protein-coding genes (Fig. 2). The analysis

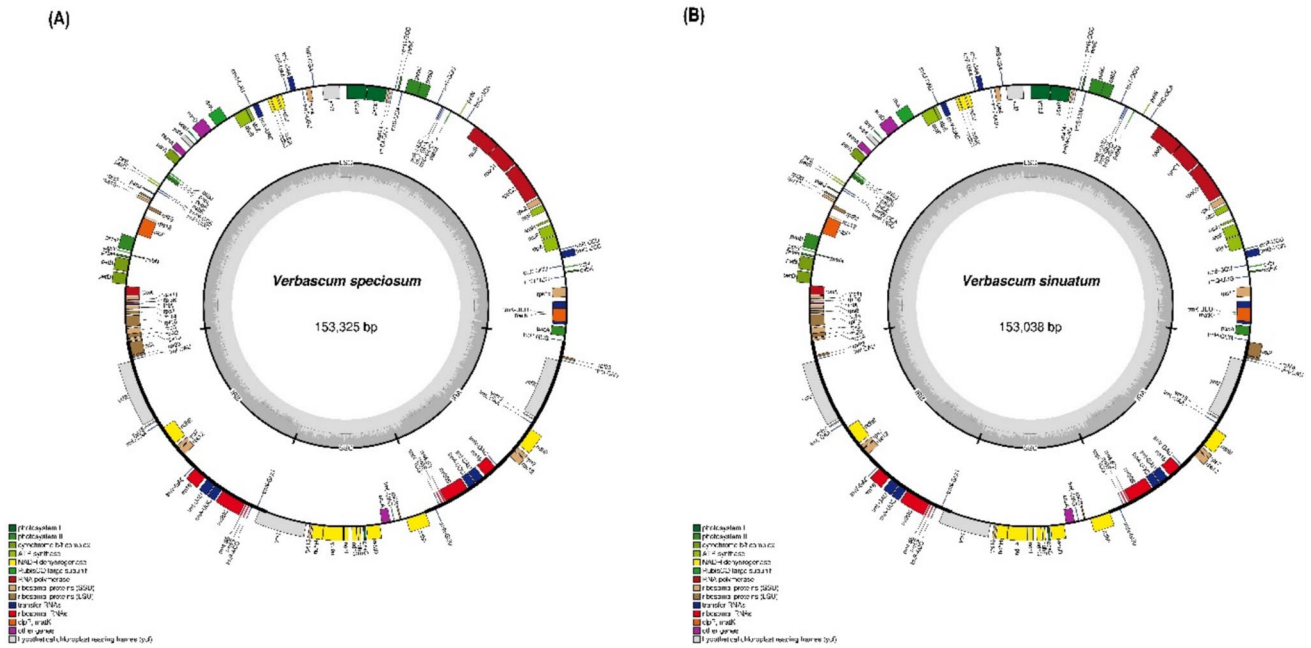

Fig. 1 Circular gene map of the chloroplast genomes of two *Verbascum* species, *V. speciosum* (**A**) and *V. sinuatum* (**B**). The diagram, generated using OGDRAW, illustrates the conserved quadripartite structure. From the center outward: the first circle shows the two inverted repeat regions (IRa and IRb) in light gray, which separate the large single-copy (LSC) and small single-copy (SSC) regions. The second circle represents the GC content, where darker peaks indicate higher GC content. The third circle visualizes the GC skew, using green and purple colors to denote positive and negative skew, respectively. The outermost circle displays the gene features, with genes color-coded by their functional groups (e.g., photosynthesis, self-replication). Genes shown on the inside of this circle are transcribed clockwise, while those on the outside are transcribed counter clockwise

revealed a highly conserved and similar codon usage pattern between *V. speciosum* and *V. sinuatum*, indicating a strong evolutionary constraint on the cp genome within this genus. A total of 64 codons were analyzed, encompassing 31 amino acid and stop codons. The RSCU values showed a pronounced bias toward codons with A or T at the third codon position. This A/T-ending preference was consistent across the vast majority of synonymous codon families in both species. A direct comparison demonstrated near-perfect concordance, with the RSCU values for 34 codons being entirely identical, including high-usage codons, such as GCT (Ala), TTA (Leu), and CCT (Pro), as well as low-use C/G-ending codons like GCC (Ala) and AGC (Ser). The minimal divergences observed were negligible, not exceeding 0.02 RSCU units, with *V. speciosum* exhibiting marginally higher values for codons including GCA (Ala) and ACC (Thr), while *V. sinuatum* showed slightly elevated values for others, such as ATT (Ile) and ACT (Thr).

Structural variation at IR boundaries in *Verbascum* cp genomes

The comparative analysis of IR boundaries among ten *Verbascum* species, including the newly sequenced *V. speciosum* and *V. sinuatum*. The results revealed a conserved genomic structure punctuated by species-specific variations in gene length and regional distribution. The *rps19* gene, located at the junction of LSC region and the IRb, exhibited a total length of 273 or 279 bp, with a consistent 236 bp segment in the LSC and a variable extension of 37 to 47 bp within the Irb (Fig. 3). A notable distinction was the presence of the *rpl23* gene (282 bp), which was identified exclusively within the IRb of *V. sinuatum* and the IRa of *V. speciosum*, being absent in the other eight species. Similarly, the *rpl2* gene demonstrated a consistent total length of 1492 bp across the genus, yet its distribution varied; in most species, 107 bp were located within an IR (either IRa or IRb), while the remainder spanned the junction into the SSC. The *trnN* gene length was predominantly 72 bp, with *V. brevipedicellatum*, *V. chinense*, *V. phoeniceum*, *V. sinaiticum*, and *V. thapsus* exhibiting a 73 bp length, and it was consistently duplicated across the IRb and IRa regions.

At the opposite end of the genome, the boundaries between SSC region and the IRs were characterized by the substantial *ycf1* and *ndhF* genes. The *ycf1* gene was the largest and most variable, with total lengths ranging from 5358 to 5379 bp. Its distribution was consistently bipartite, with the majority of its sequence (4533–4549 bp) located in the SSC and a significant segment (824–830 bp) extending into

Table 1 Gene content and functional classification of two *Verbascum* cp genomes. Genes are categorized by functional groups, with intron-containing genes (#) and multi-copy genes (*n*) indicated

Category	Gene group	Gene name	*V. speciosum*	*V. sinuatum*
Photosynthesis	Subunits of photosystem I	*psaB, psaA, psaI, psaJ, psaC*	5	5
	Subunits of photosystem II	*psbA, psbK, psbI, psbM, psbD, psbC, psbZ, psbJ, psbL, psbF, psbE, psbB, psbT, psbN, psbH*	15	15
	Subunits of NADH dehydrogenase	*ndhJ, ndhK, ndhC, ndhB(2)#, ndhH, ndhA#, ndhI, ndhG, ndhE, ndhD, ndhF*	12	12
	Subunits of cytochrome b/f complex	*petN, petA, petL, petG, petB#, petD#*	6	6
	Large subunit of rubisco	*rbcL*	1	1
	Subunits of ATP synthase	*atpA, atpF#, atpH, atpI, atpE, atpB*	6	6
Self-replication	Proteins of large ribosomal subunit	*rpl33, rpl20, rpl36, rpl14, rpl16#, rpl22, rpl23(2), rpl32, rpl2#*	10	10
	Proteins of small ribosomal subunit	*rps12(2)##, rps16#, rps2, rps14, rps4, rps18, rps11, rps8, rps3, rps19, rps7(2), rps15*	14	14
	Subunits of RNA polymerase	*rpoC2, rpoC1#, rpoB, rpoA*	4	4
	Ribosomal RNAs	*rrn16(2), rrn23S(2), rrn4.5S(2), rrn5S(2)*	8	8
	Transfer RNAs	*trnH-GUG, trnK-UUU#, trnQ-UUG, trnS-GCU, trnG-UCC#, trnR-UCU, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnT-GGU, trnS-UGA, trnG-GCC, trnfM-CAU, trnS-GGA, trnT-UGU, trnL-UAA#, trnF-GAA, trnV-UAC#, trnM-CAU, trnW-CCA, trnP-UGG, trnI-CAU(2), trnL-CAA(2), trnV-GAC(2), trnI-GAU(2)#, trnA-UGC(2)#, trnR-ACG(2), trnN-GUU(2), trnL-UAG*	37	37
Other genes	Maturase	*matK*	1	1
	Protease	*clpP##*	1	1
	Envelope membrane protein	*cemA*	1	1
	Acetyl-CoA carboxylase	*accD*	1	1
	c-type cytochrome synthesis gene	*ccsA*	1	1
	Translation initiation factor	*infA*	1	1
	Conserved open reading frames	*ycf3##, ycf4, ycf2(2), ycf15(2), ycf1*	7	7
Total			131	131

an IR (either IRb or IRa). The ndhF gene, with a constant total length of 2229 bp, was primarily situated in the SSC; however, in all species except *V. chinense* and *V. phoeniceum*, a very short segment of 2 bp was found to protrude into the adjacent IRa or IRb. Finally, the *trnH* gene was uniformly 74 bp in length and situated entirely within the LSC, adjacent to the IRa boundary. The cumulative effect of these gene-specific variations in length and regional distribution underpins the slight size divergence observed in the overall cp genomes, which range from 153,014 to 153,618 bp, illustrating the dynamic nature of IR boundary evolution in *Verbascum*.

cpSSR distribution and IGS variation in *Verbascum* species

The comparative analysis of cpSSRs across ten *Verbascum* species revealed significant insights into their plastomic diversity and evolutionary dynamics. A total of 44 and 34 cpSSRs were identified in the newly sequenced plastomes of *V. speciosum* and *V. sinuatum*, respectively. Mononucleotide repeats (A/T) constituted the most abundant class across all species, comprising 36 (81.8%) of the repeats in *V. speciosum*, a figure consistent with congeners like *V. songaricum* (39, 83.0%). In contrast, *V. sinuatum* exhibited a notably lower count of 27 mononucleotide repeats, the lowest in the study, resulting in a distinct profile. Both species shared a commonality of 2 trinucleotide repeats, while differing in dinucleotide counts; *V. speciosum* contained 6, a trait shared with *V. blattaria*, whereas *V. sinuatum* possessed 5, a more common number within the genus.

A detailed examination of IGS regions (Fig. 4), which were confirmed as the primary reservoirs for SSR proliferation, uncovered both conserved hotspots and species-specific variations. The trnS-GCU-trnG-UCC IGS emerged as a major variable locus. This region was particularly

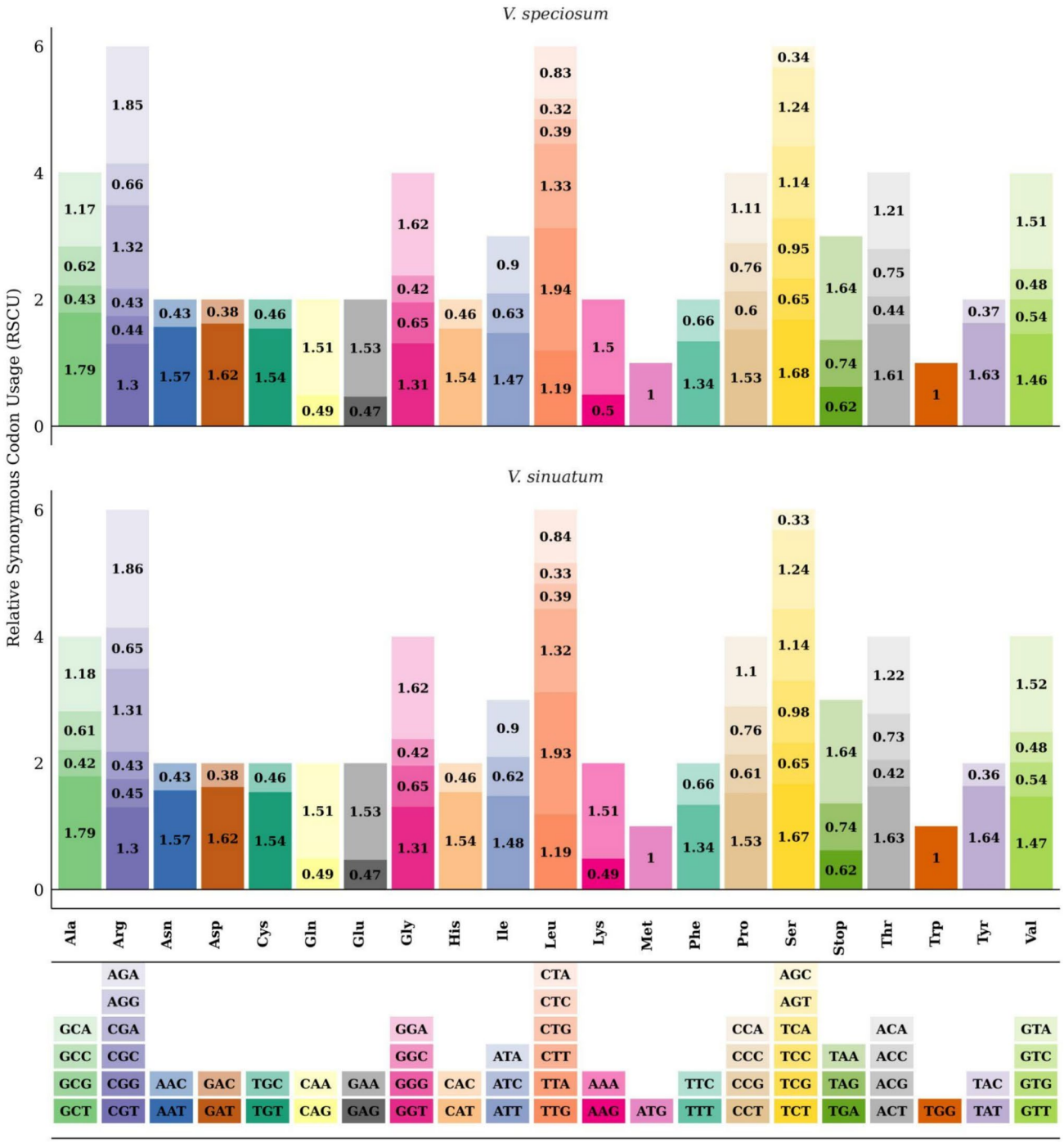

Fig. 2 Bar plot of Relative Synonymous Codon Usage (RSCU) values for each amino acid, comparing two species (*V. speciosum* and *V. sinuatum*). RSCU values were calculated from all protein-coding genes. Codons are color-coded, and their corresponding amino acids are labeled below the plot. RSCU values

SSR-rich in *V. speciosum*, harboring four distinct repeats (TA, T, A, T), a complexity mirrored in *V. thapsus* and *V. songaricum*. Conversely, *V. sinuatum* was found to entirely lack SSRs in this specific spacer, a distinctive feature shared only with *V. phoeniceum*, highlighting a significant structural divergence.

Furthermore, the rps4-trnT-UGU IGS was identified as another key variable region. Here, both *V. speciosum* and *V. sinuatum*, along with most species, featured complex di-nucleotide (TA) repeats. However, the architecture differed; *V. speciosum* possessed two adjacent TA repeats (lengths 6 and 7), a pattern also observed in *V. phoeniceum*,

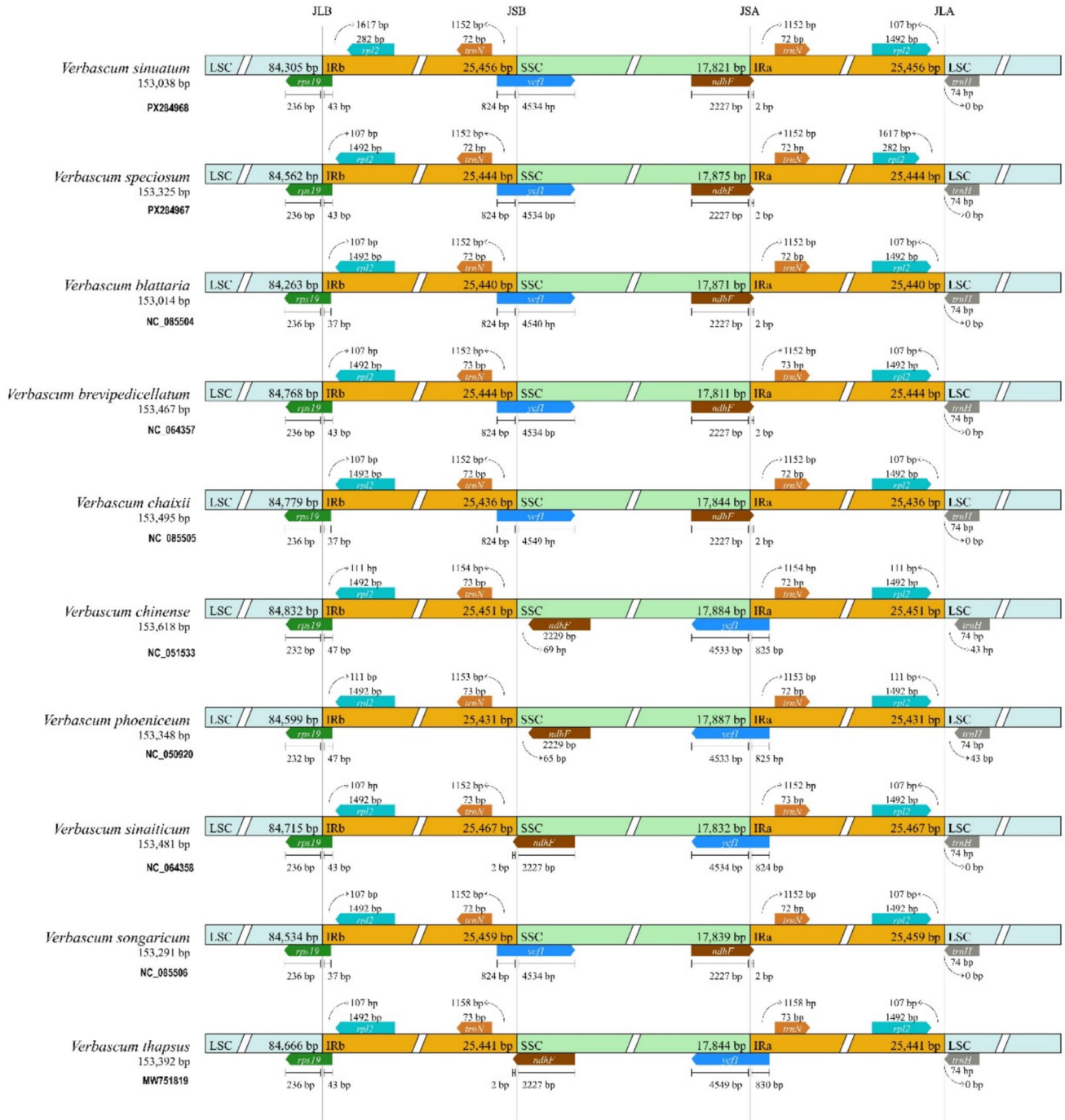

Fig. 3 Comparison of IR/LSC/SSC boundaries among ten *Verbascum* species (two newly sequenced: *V. speciosum* and *V. sinuatum*; eight retrieved from NCBI). Key junctions are labeled: JLB (LSC/IRb), JSB (SSC/IRb), JSA (SSC/IRa), and JLA (LSC/IRa). For each species, the distance (in base pairs) from each junction to the nearest gene is shown. The *rpl23* gene was present within IR regions exclusively in *V. speciosum* and *V. sinuatum* ($n=2$ species), while absent in the other eight species

whereas *V. sinuatum* displayed a more complex organization with a TA repeat followed by an AT motif. This suggests that while the locus is a conserved hotspot, the specific nature of the repeats is lineage-specific.

Additional IGS regions provided further phylogenetic signals. The atpF_1-atpH IGS consistently contained a T-repeat across all species except *V. phoeniceum*, indicating a high degree of conservation. The rps2-rpoC2 IGS often contained adjacent mono-nucleotide (T and A)

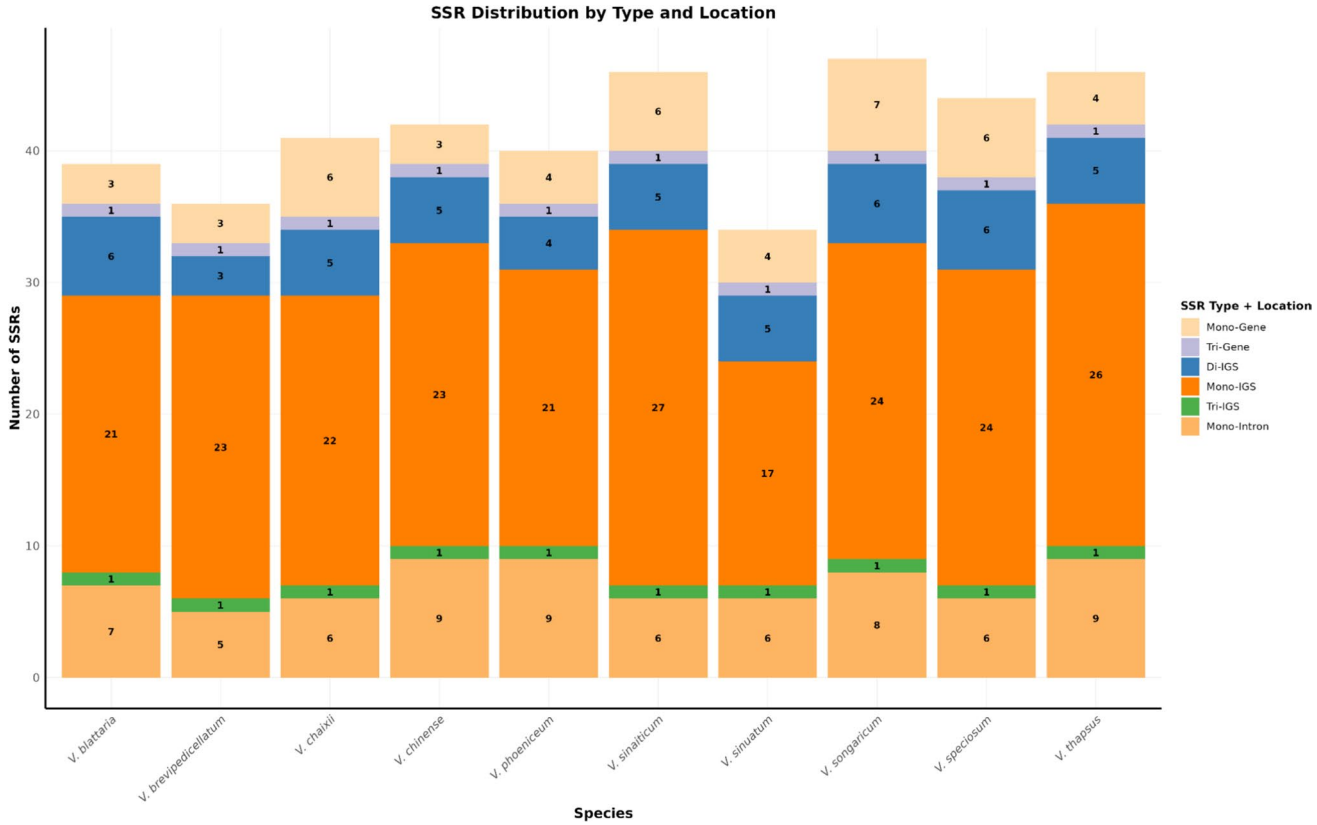

Fig. 4 Distribution of simple sequence repeats (SSRs) along the chloroplast genomes of ten *Verbascum* species, including two newly sequenced (*V. speciosum* and *V. sinuatum*) and eight species retrieved from NCBI

repeats, as seen in *V. speciosum*, a pattern conserved in *V. thapsus* and *V. sinaiticum*. Notably, *V. sinuatum* lacked the A-repeat in this locus, presenting only a T-repeat, which again underscores its divergent character. The ycf1-rps15 IGS near the SSC/IR boundary was another variable region, with *V. sinuatum* and *V. speciosum* both possessing TA repeats, though of different lengths and complexities compared to each other and to other species like *V. blattaria*.

In summary, the distribution of cpSSRs provides a robust framework for understanding plastomic evolution in *Verbascum*. The profile of *V. speciosum* aligns it with a group of species exhibiting high SSR abundance and complexity in common IGS hotspots like trnS-GCU-trnG-UCC_1. In stark contrast, *V. sinuatum* is delineated as a distinct outlier not only by its depauperate overall SSR content but also by the conspicuous absence of repeats in key variable regions and the possession of unique compound motifs in others.

Site-specific analysis elucidates targets of positive selection across plastid genes

The results of the site-specific positive selection analysis revealed that, while pervasive signals of selection were observed across the plastid genome, only three genes, including *ndhB*, *ycf4*, and *ycf3*, exhibited statistically significant evidence of codon-level positive selection under stringent likelihood ratio test comparisons (M8 vs. M7 and/or M2a vs. M1a). For ndhB, a glycine (G) at site 259 was strongly supported (posterior probability = 0.975) by the M8 model (LRT P = 0.00544). In ycf4, multiple sites were identified under the M8 model (LRT P = 4.40×10^{-5}), with site 1688, an aspartate (D), receiving the highest confidence (posterior probability = 1.000). Similarly, ycf3 also displayed a positively selected aspartate (D) at position 1688 (posterior probability = 0.977), supported by the M8 model (LRT P = 0.00947).

Landscape of nucleotide diversity across *Verbascum* cp genomes

The analysis of Pi across cp genomes of ten *Verbascum* species revealed a distinct landscape of sequence variation,

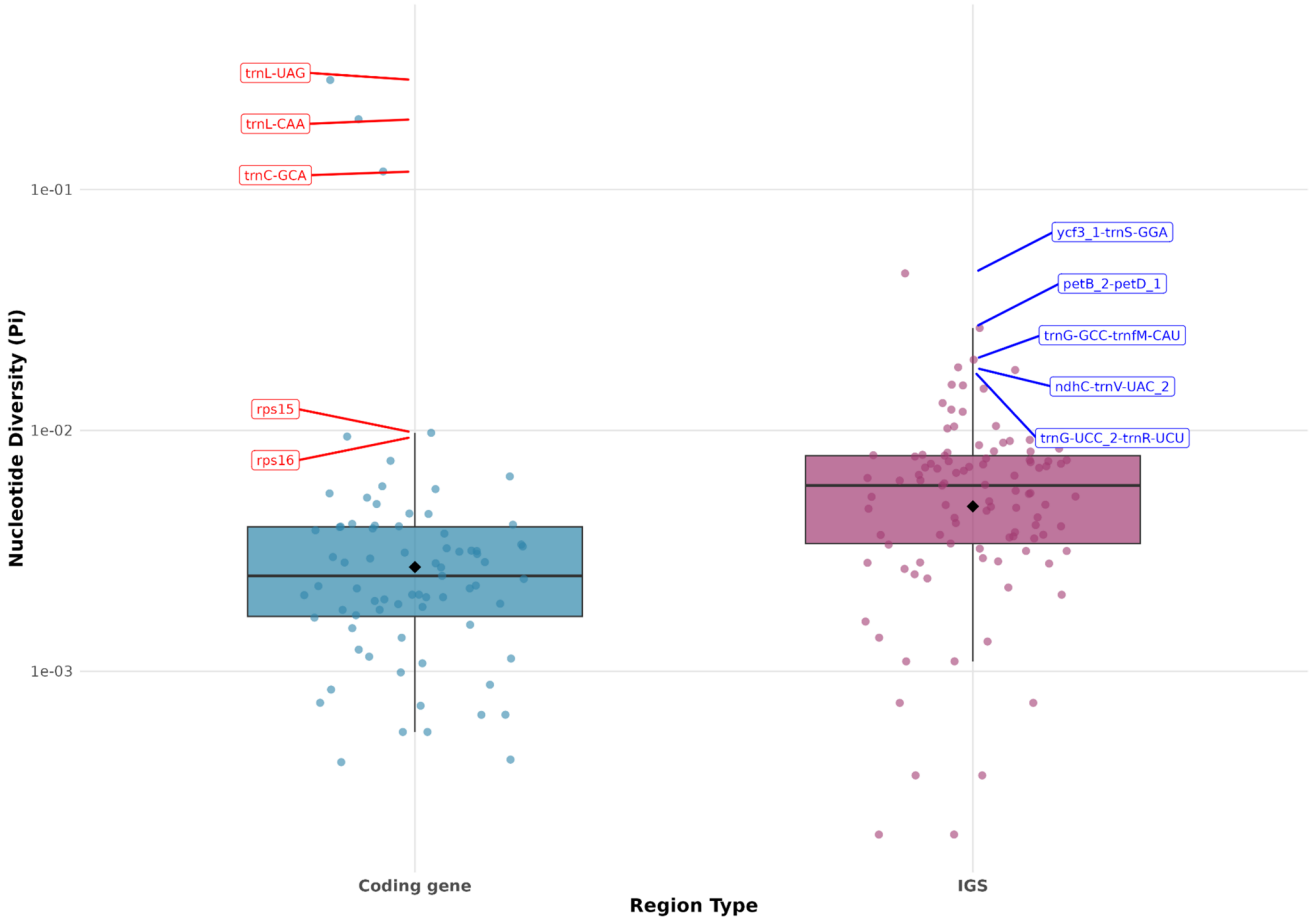

Fig. 5 Nucleotide diversity (Pi) calculated from protein-coding genes and intergenic spacer regions across ten *Verbascum* species (two newly sequenced: *V. speciosum* and *V. sinuatum*; eight retrieved from NCBI). Boxplots depict Pi values for coding genes (blue) and intergenic spacer regions (IGS; orange), with jittered points showing individual values. The top and bottom three most extreme values for each region type are labeled with feature names and exact Pi values

characterized by both profound conservation and exceptional localized diversity (Fig. 5). A substantial suite of regions was found to be entirely invariant (Pi = 0), indicating perfect sequence conservation across the studied species. This set included numerous tRNA genes (e.g., trnA-UGC, trnD-GUC, trnF-GAA, trnI-GAU, trnL-UAA, trnN-GUU, trnQ-UUG, trnV-GAC), the coding genes *psaC*, *psbE*, *psbL*, *psbN*, *rpl23*, and *petN*, alongside several intergenic spacers, such as ndhB_1-ndhB_2, psaB-psaA, psbF-psbE, and rpoC1_1-rpoB.

Beyond these invariant loci, the spectrum of nucleotide diversity in variable regions was remarkably broad. The most striking finding was the extraordinarily high Pi values associated with several tRNA genes, which emerged as the most variable coding sequences in the plastome. The genes trnL-UAG (Pi = 0.28502), trnL-CAA (Pi = 0.19570), and trnC-GCA (Pi = 0.11884) displayed diversity levels in order of magnitude greater than the most variable protein-coding genes. Among standard protein-coding genes, *rps15* (Pi = 0.00977) and *rps16* (Pi = 0.00944) were the most polymorphic, followed by *matK* (Pi = 0.00748) and *ycf1* (Pi = 0.00644). In contrast, genes like *rps7* (Pi = 0.00043), *ycf3* (Pi = 0.00042), *ycf2* (Pi = 0.00066), and ndhB (Pi = 0.00074) were among the most conserved, reflecting strong purifying selection.

Intergenic spacers in *Verbascum* also contained significant hotspots of diversity. The IGS ycf3_1-trnS-GGA (Pi = 0.04488) was the most variable non-coding region identified. Other highly polymorphic spacers included petB_2-petD_1 (Pi = 0.02661), trnG-GCC-trnfM-CAU (Pi = 0.01966), and ndhC-trnV-UAC_2 (Pi = 0.01829). The commonly used barcoding region trnH-psbA also showed considerable diversity (Pi = 0.01221). Conversely, some IGS regions, particularly those like rps12_2-rps12_3 (Pi = 0.00037) and those associated with inverted repeats, such as trnI-GAU_1-trnI-GAU_2 (Pi = 0.00021), were highly conserved.

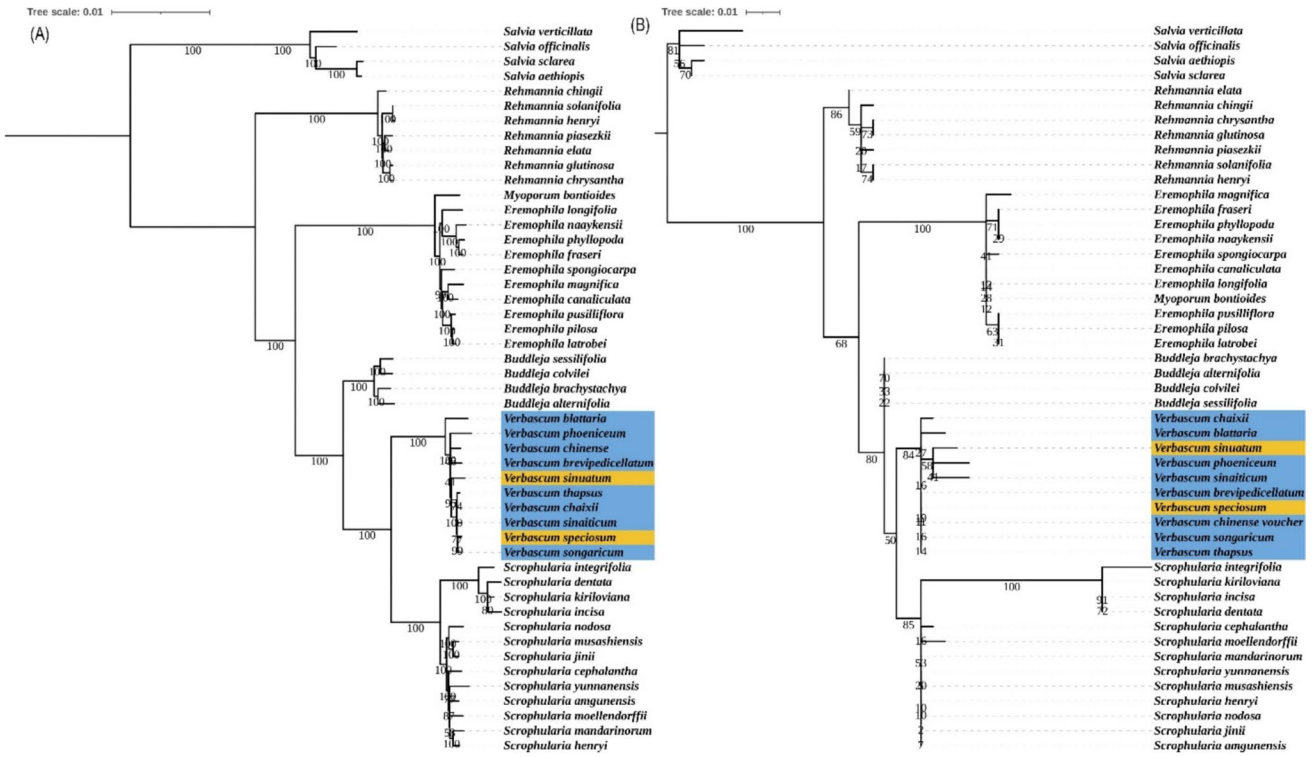

Fig. 6 Phylogenetic tree of cp genomes, including 10 *Verbascum* species ((two newly sequenced: *V. speciosum* and *V. sinuatum*; eight retrieved from NCBI), other species from Scrophulariaceae family, and some outgroups (*Salvia*), reconstructed by whole cp genome (**A**), and *rps15* (**B**) using maximum likelihood in IQ-TREE under the GTR + Gamma model. The highlighted region emphasizes the position and relationships of the sample sequences from this study, illustrating their clustering with closely related species. Branch support was assessed with 100 bootstrap replicates

Phylogenetic reconstruction based on complete cp genomes and barcode regions

The phylogenetic relationships of *Verbascum* species within Scrophulariaceae were inferred using maximum likelihood analysis based on complete cp genome sequences (Fig. 6A). According to the results, both *V. speciosum* and *V. sinuatum* were robustly placed within a strongly supported, monophyletic *Verbascum* clade, confirming their taxonomic assignment within Scrophulariaceae. Within this clade, *V. speciosum* was resolved as the sister species to *V. songaricum* with high bootstrap support, and this pair formed a well-supported subgroup closely related to *V. sinaiticum*. In contrast, *V. sinuatum* occupied a more derived position within *Verbascum*, being resolved as sister to the clade comprising *V. sinaiticum, V. songaricum*, and *V. speciosum,* indicating a slightly more distant evolutionary relationship compared with *V. speciosum*. Additional internal relationships within *Verbascum* were also recovered. *V. chaixii* and *V. thapsus* formed a distinct sister pair with moderate bootstrap support, while other species, such as *V. brevipeduncu-latum*, *V. chinense*, and *V. phoeniceum*, were successively resolved within the genus, although some deeper nodes showed lower support. At the family level, the *Verbascum* clade was recovered as sister to a monophyletic Scrophularia clade with strong bootstrap support, reinforcing the close evolutionary relationship between these two genera within Scrophulariaceae.

The phylogeny inferred from the *rps15* gene showed general congruence with the complete cp genome tree at the genus level, as *Verbascum*, *Scrophularia*, *Rehmannia*, and *Buddleja* were each recovered as monophyletic groups (Fig. 6B). However, the internal relationships within *Verbascum* differed between the two analyses, and several nodes exhibited low bootstrap support, leading to inconsistent placement of closely related species, including *V. speciosum* and *V. sinuatum*. Compared with the genome-wide tree, the *rps15*-based phylogeny provided lower resolution among *Verbascum* species. These results indicated that *rps15* alone contained insufficient phylogenetic signal for reliable species-level discrimination within *Verbascum*.

Discussion

Chloroplast genomes have been widely applied to resolve evolutionary relationships and clarify taxonomic boundaries, particularly among closely related plant species, by providing a stable and informative genetic framework. However, comprehensive *Verbascum* plastome studies remain scarce, largely limited to Dong et al. (2022), who sequenced seven species from China and Kenya (Dong et al. 2022), but left critical gaps, no data from Iran (a primary diversity center), unexamined structural variation, and overlooked selection signatures. Here, we address these gaps with three advances: (i) first plastomes from Iranian *Verbascum* species; (ii) novel positively selected genes (*ndhB*, *ycf3*, *ycf4*); and (iii) *rpl23* IR rearrangement.

In this study, the complete plastome sequences of *V. speciosum* (153,325 bp) and *V. sinuatum* (153,038 bp) were characterized, thereby expanding available genomic resources for the genus *Verbascum*. Both plastomes displayed the typical quadripartite organization comprising an LSC, SSC, and a pair of IR regions. This conserved structural architecture was consistent with previously reported plastomes of *V. chinense* (153,618 bp), *V. thapsus* (153,338 bp), and *V. phoeniceum* (153,348 bp), indicating a highly conserved plastome organization across the genus (Bi et al. 2020; He et al. 2020; Zhang et al. 2022).

The minor differences in genome size observed among *Verbascum* species were relatively small, with the plastomes of *V. speciosum* and *V. sinuatum* falling well within the previously reported size range of approximately 153,014–153,618 bp. Such limited size variation appears to represent a common feature of cp genome evolution in *Verbascum* and is consistent with patterns reported for other angiosperm lineages (Dong et al. 2022). Notably, an absolute conservation of gene content was observed between *V. speciosum* and *V. sinuatum*, with both plastomes containing an identical complement of 131 functional genes. This gene number differed slightly from those reported in earlier studies, where plastomes of *V. chinense* and *V. phoeniceum* were annotated with 114 unique genes, and *V. thapsus* with 135 genes (Bi et al. 2020; He et al. 2020; Zhang et al. 2022). These inconsistencies likely reflect differences in annotation strategies, particularly in the treatment of duplicated genes within the IR regions, rather than true biological variation. The high level of gene content conservation observed here suggests strong purifying selection acting on essential photosynthetic and genetic functions, a pattern commonly reported in recently diverged plant taxa (Yao et al. 2016; Ge et al. 2018).

Despite this overall structural stability, cp genomes are known to accumulate subtle variations, most notably through expansions and contractions at the junctions between the single-copy and inverted repeat regions. Such boundary shifts have been recognized as a primary driver of plastome size variation in angiosperms (Wicke et al. 2011; Wicke and Naumann 2018). A comparative analysis across ten *Verbascum* species demonstrated that the overall gene order remained highly conserved; however, a detailed examination of IR/SC boundaries revealed informative microstructural differences. For instance, although the *rps19* gene consistently spanned the LSC/IRb boundary, the extent of its incorporation into the IR region varied slightly among species. A distinctive structural feature identified in this study was the placement of the *rpl23* gene within the IR regions of *V. speciosum* and *V. sinuatum*, a configuration not reported in the plastomes of V. chinense, *V. phoeniceum*, or *V. thapsus* (Bi et al. 2020; He et al. 2020; Zhang et al. 2022). This unique positioning is most likely attributable to lineage-specific IR expansion/contraction events, wherein an expansion of the IR boundary into the adjacent LSC region led to the incorporation and subsequent duplication of *rpl23* within the IR, a phenomenon commonly observed in angiosperm plastome evolution. These lineage-specific modifications at plastome boundary regions could also likely contribute to modest genome size variation while maintaining the overall conserved architecture of the plastome. Comparable patterns of fine-scale structural variation have been reported in a wide range of plant families, highlighting the evolutionary significance of IR boundary dynamics (Lei et al. 2016; Asaf et al. 2020).

To further assess the extent of structural and sequence conservation across *Verbascum* plastomes and to identify loci potentially subject to adaptive evolution, a site-specific positive selection analysis was performed. Despite the strong overall evolutionary constraint observed at the plastome-wide level, this analysis detected signatures of selection acting on a restricted subset of plastid genes, indicating a more nuanced evolutionary pattern than would be inferred from genome-wide conservation alone. Significant signatures of positive selection were detected in only three genes, including *ndhB*, *ycf3*, and *ycf4*. The identification of positive selection in ndhB, which encodes a core subunit of cp NADH dehydrogenase-like (NDH) complex, is particularly informative. This result is concordant with previous findings by Dong et al. (2022), who reported adaptive signals in the closely related *ndhF* gene across a broader representation of *Verbascum* species (Dong et al. 2022). Together, these results suggest that components of the NDH complex may represent recurrent targets of adaptive evolution within the genus. Given the role of the NDH complex in cyclic electron flow around photosystem I and in plant responses to environmental stress, selection acting on these genes may reflect adaptive adjustments to heterogeneous ecological conditions experienced by different *Verbascum* taxa. In addition, the detection of positive selection in *ycf3* and *ycf4*, both

of which function as assembly factors for photosystem I, indicates potential adaptive fine-tuning of the photosynthetic machinery. Although these genes are essential and generally highly conserved, the presence of positively selected codons suggests that specific amino acid residues may have evolved to optimize photosystem I assembly, stability, or functional efficiency. This pattern extends current understanding of plastome evolution in *Verbascum* by implicating not only auxiliary electron transport components but also core photosystem assembly pathways as targets of adaptive change.

The analysis of Pi across *Verbascum* plastomes delineated a clear genomic framework for species discrimination and marker selection. Consistent with plastome-wide surveys across angiosperms, IGS regions exhibited substantially higher variability than protein-coding genes, reflecting relaxed selective constraints and rapid accumulation of substitutions in non-coding cp DNA (Dong et al. 2012, 2015; Samarina et al. 2025). Such regions have repeatedly been identified as the most informative plastid loci for resolving closely related species and intraspecific diversity, particularly where universal coding barcodes show limited resolution (Kress and Erickson 2007; Kress et al. 2009; Pang et al. 2012; Bell et al. 2017).

Among all loci examined, the IGS ycf3_1–trnS-GGA displayed the highest nucleotide diversity, exceeding even the widely used plant barcode *trnH–psbA*. This finding is concordant with comparative plastome analysis demonstrating that lineage-specific intergenic regions frequently outperform standard universal barcodes in intrageneric discrimination (Dong et al. 2022). Similar patterns have been reported in diverse angiosperm families, including Lamiaceae (Akrami et al. 2025), Orchidaceae, and Asteraceae (Asaf et al. 2020), where ycf- and tRNA-flanking spacers rank among the most variable plastid regions. These results further support the growing consensus that effective plastid barcoding requires empirical identification of hypervariable regions rather than exclusive reliance on traditionally recommended loci.

Protein-coding genes showed overall lower Pi values relative to IGS regions, a pattern widely documented across plastome-based barcoding studies (Wicke et al. 2011; Dong et al. 2015; Li et al. 2015; Wicke and Naumann 2018). Nevertheless, *matK* exhibited comparatively elevated nucleotide diversity among coding loci, consistent with its recognized role as a core plant DNA barcode (Kress et al. 2005, 2009; Kress and Erickson 2007; Scriver et al. 2015; Bruno et al. 2025). However, the lower Pi of *matK* relative to the most variable IGS regions in *Verbascum* corroborates previous findings that *matK* alone is often insufficient for fine-scale species discrimination, particularly in recently diverged or taxonomically complex plant groups (Burgess et al. 2011; Li et al. 2015; Daniell et al. 2016). This limitation is further compounded by well-documented challenges in primer universality and amplification efficiency for *matK* across land plants (Dunning and Savolainen 2010; Heckenhauer et al. 2016).

Accordingly, our results support a multilocus barcoding strategy combining both coding (*matK*) and non-coding markers (e.g., ycf3_1–trnS-GGA, trnH–psbA), as advocated by plastome-scale barcoding frameworks and primer performance studies (Bell et al. 2017; Su et al. 2022; Samarina et al. 2025).

A particularly striking result of this study is the exceptionally high nucleotide diversity observed in several tRNA genes, especially trnL-UAG. This level of variability exceeds that of all protein-coding genes and contrasts with the classical plastome model in which loci such as *ycf1* or *matK* represent the primary reservoirs of variation (Dong et al. 2015). However, recent plastome-wide analysis has demonstrated that tRNA genes and their adjacent non-coding regions frequently act as hotspots of sequence divergence, largely due to relaxed functional constraints outside the mature tRNA structure and frequent inclusion of flanking sequences within annotated regions (Dong et al. 2012; Pang et al. 2012; Samarina et al. 2025). The pronounced variability detected in trnL-UAG, therefore, highlights its potential as a powerful, genus-specific phylogenetic marker in *Verbascum*.

Given that even the most variable barcoding loci failed to resolve species-level relationships in *Verbascum*, our whole plastome-based phylogeny offers a straightforward answer: the evolutionary signal is there, but we cannot see it without enough genomic data. The strongly supported monophyly of *Verbascum* and the clear placement of *V. speciosum* and *V. sinuatum* give us, for the first time, a stable evolutionary framework for asking deeper taxonomic questions. This matters because, for decades, identifying *Verbascum* species has been frustratingly difficult—morphological characters shift continuously across environments, and traditional markers have repeatedly fallen short. Our findings show that single-locus barcodes are simply not up to the task for this genus, likely because many species have too recently diverged or share complex evolutionary histories. We believe it is time to embrace plastome-scale data as a new baseline for species delimitation in *Verbascum* bring nuclear evidence into the picture.

Finally, as a conclusion, we have delivered new genomic resources for *V. speciosum* and *V. sinuatum*, but connecting these plastome data to their medicinal properties remained a limitation. The phenolic and flavonoid compounds behind their antioxidant activity are primarily produced by nuclear genes, not plastid ones. Our analysis alone could not explain why these two species differ in their phytochemical profiles. What we could offer instead was an indirect link: the three positively selected plastid genes we found, *ndhB*, *ycf3*, and *ycf4*, are all involved in stress responses. Because environmental stress often triggers antioxidant production, adaptive

changes in these plastid genes might indirectly influence medicinal chemistry. Whether a direct connection exists between plastome evolution and medicinal chemistry in *Verbascum* remains to be evaluated in future studies.

Conclusion

This study presents the first complete cp genomes of *V. speciosum* and *V. sinuatum*, two medicinally important Iranian species. Both plastomes retained the typical quadripartite structure and gene content seen across angiosperms, but they also carried subtle yet telling signs of change, such as lineage-specific shifts at IR boundaries, including an unusual placement of *rpl23*. Along the way, we uncovered several hypervariable spots (*rps15*, *rps16*, *matK*, *ycf1*, and even tRNA genes like trnL-UAG) that could serve as promising candidates for DNA barcoding. We also found evidence of positive selection in *ndhB*, *ycf3*, and *ycf4*, hinting that even conserved plastid genes can sometimes adapt. Phylogenomic analysis placed both species firmly within *Verbascum* with high confidence, while the single-locus barcode fell short of providing reliable species-level resolution. Taken together, our findings showed how much plastome-scale data can clarify relationships in a group as morphologically challenging as *Verbascum*. Looking ahead, broader sampling across Iran's rich *Verbascum* diversity, ideally combined with nuclear genomic data, will be key to fully understanding the evolutionary history of this fascinating and medicinally valuable genus. Despite these advances, the present study is limited by the relatively small number of plastomes currently available for the genus and by its focus on cp genomes alone, which reflect only the maternally inherited evolutionary history. Expanded taxon sampling across the geographic and morphological diversity of *Verbascum*, combined with nuclear genomic data, will be essential to fully resolve complex evolutionary relationships, hybridization patterns, and species boundaries within this diverse lineage.

Acknowledgements Not applicable.

Author contributions MA: investigation; methodology; formal analysis, writing-review, AS: Conceptualization; funding acquisition; investigation; project administration; methodology; formal analysis; validation; writing-original draft; writing-review and editing.

Funding No funding was received for conducting this study.

Data availability The complete chloroplast genome sequences of *V. speciosum* and *V. sinuatum* have been deposited in the NCBI database and are publicly accessible under the accession numbers PX284967 and PX284968, respectively.

Declarations

Conflict of interest The authors declare no conflicts of interest.

Ethical approval and consent to participate The authors declared that experimental research works on the plants described in this paper comply with institutional, national and international guidelines. This article does not involve any endangered or protected species.

Consent for publication Not applicable.

References

Akrami AM, Meratian Esfahani S, Soorni A (2025) Decoding the chloroplast genomes of five Iranian Salvia species: insights into genomic structure, phylogenetic relationships, and molecular marker development. BMC Genom 26:545

Alzahrani AM, Brehm JM, Ghazanfar SA, Maxted N (2024) DNA barcoding of the genus Verbascum (Scrophulariaceae) in the Arabian Peninsula. Taxon 73:547–555

Amini S, Hassani A, Alirezalu A, Maleki R (2022) Phenolic and flavonoid compounds and antioxidant activity in flowers of nine endemic Verbascum species from Iran. J Sci Food Agric 102:3250–3258

Amiryousefi A, Hyvönen J, Poczai P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34:3030–3031

Asaf S, Khan AL, Khan A, Al-Harrasi A (2020) Unraveling the chloroplast genomes of two prosopis species to identify its genomic information, comparative analyses and phylogenetic relationship. Int J Mol Sci 21:3280

Bell KL, Loeffler VM, Brosi BJ (2017) An rbcL reference library to aid in the identification of plant species mixtures by DNA metabarcoding. Appl Plant Sci 5:1600110

Bi Y, Deng P, Liu L (2020) The complete chloroplast genome sequence of purple mullein (Verbascum phoeniceum L.). Mitochondrial DNA Part B 5:819–820

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bruno MC, Posik DM, Olinera LH et al (2025) Food-metabarcoding: identification of multiple animal and plant species in foodstuffs using target GBS assay. Food Control. https://doi.org/10.1016/j.foodcont.2025.111938

Burgess KS, Fazekas AJ, Kesanakurti PR et al (2011) Discriminating plant species in a local temperate flora using the rbcL+ matK DNA barcode. Methods Ecol Evol 2:333–340

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Christenhusz M (2016) The number of known plants species in the world and its annual increase. Phytotaxa. https://doi.org/10.11646/phytotaxa.261.3.1

Daniell H, Lin C-S, Yu M, Chang W-J (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol 17:1–29

Diani Gohar S, Soorni A (2025) Plastome diversity and phylogenomic analysis of Satureja (Lamiaceae): uncovering evolutionary patterns and diagnostic markers. Planta 262:149

Dong W, Liu J, Yu J et al (2012) Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS ONE 7:e35071

Dong W, Xu C, Li C et al (2015) ycf1, the most promising plastid DNA barcode of land plants. Sci Rep 5:8348

Dong X, Mkala EM, Mutinda ES et al (2022) Taxonomy, comparative genomics of Mullein (Verbascum, Scrophulariaceae), with

implications for the evolution of Verbascum and Lamiales. BMC Genom 23:566
Dunning LT, Savolainen V (2010) Broad-scale amplification of matK for DNA barcoding plants, a technical note. Bot J Linn Soc 164:1–9
Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340
Fedchenko BA, Schischkin BK, Bobrow EG (1955) Verbascum l. Flora USSR 22:132–197
Fischer E (2004) Scrophulariaceae. Flowering plants· dicotyledons: Lamiales (except Acanthaceae including Avicenniaceae). Springer, pp 333–432
Gao F, Chen C, Arab DA et al (2019) EasyCodeML: a visual tool for analysis of selection using CodeML. Ecol Evol 9:3891–3898
Ge J, Cai L, Bi G-Q et al (2018) Characterization of the complete chloroplast genomes of buddleja colvilei and b sessilifolia: implications for the taxonomy of buddleja l. Molecules 23:1248
Ghahremaninejad F, Riahi M, Babaei M et al (2015) Monophyly of verbascum (scrophularieae: scrophulariaceae): evidence from nuclear and plastid phylogenetic analyses. Aust J Bot 62:638–646
Ghotbi FS, Soorni A (2026) Chloroplast phylogenomics and barcode discovery in medicinal stachys species reveal evolutionary relationships and adaptive signatures. Ecol Evol 16:e73618
Golchini MM, Soorni A (2026) Structural and phylogenetic insights from complete chloroplast genomes of seven vicia species. PLoS ONE 21:e0340630
Grabias B, Świątek L, Świętosławski J (1991) The morphology of hairs in Verbascwn L. species. Acta Soc Bot Pol 60:191–208
Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res 47:W59–W64
Grigore A, Colceru-Mihul S, Litescu S et al (2013) Correlation between polyphenol content and anti-inflammatory activity of Verbascum phlomoides (mullein). Pharm Biol 51:925–929
Güner A, Aslan S (2012) Türkiye bitkileri listesi:(damarlı bitkiler). Nezahat Gökyiğit Botanik Bahçesi Yayınları
Halimi M, Nasrabadi M (2018) Isolation and identification macrocyclic spermine alkaloid (protoverbine) from verbascum speciosum. Iran Chem Commun 6:143–147
Hartl D (1977) RHABDOTOSPERMA, EINE NEUE, AUS GLIEDERN VON VERBASCUM L. UND CELSIA L. GEBILDETE GATTUNG DER SCROPHULARIACEEN.
He Y, Ma Y, Li Z et al (2020) The complete chloroplast genome of Verbascum chinense (L.) Santapau. Mitochondrial DNA Part B 5:3021–3022
Heckenhauer J, Barfuss MHJ, Samuel R (2016) Universal multiplexable matK primers for DNA barcoding of angiosperms. Appl Plant Sci 4:1500137
Hejazi FA, Soorni A (2026) Structural and evolutionary analysis of Nepeta nuda plastid genome reveals lineage-specific structure variation and signature of positive selection. Plant Gene. https://doi.org/10.1016/j.plgene.2026.100581
Hejazi FA, Mohammadi P, Soorni A (2025) Comparative chloroplast genomics of Teucrium species reveals genome evolution, phylogenetic relationships, and candidate molecular markers. Sci Rep 15:44318
Huang L, Yu H, Wang Z, Xu W (2024) CPStools: a package for analyzing chloroplast genome sequences. iMetaOmics 1:e25
Jamshidi-Kia F, Lorigooini Z, Asgari S, Saeidi K (2019) Iranian species of Verbascum: a review of botany, phytochemistry, and pharmacological effects. Toxin Rev 38(4):1–8
Jin J-J, Yu W-B, Yang J-B et al (2020) GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol 21:1–31
Juan R, Fernandez I, Pastor J (1997) Systematic consideration of microcharacters of fruits and seeds in the genusverbascum (scrophulariaceae). Ann Bot 80:591–598
Karamian R, Ghasemlou F (2013) Total phenolic content, antioxidant and antibacterial activities of three Verbascum species from Iran. J Med Plants by-Prod 2:43–51
Karamian R, Ghasenlou F (2015) Study of saponin content in aerial parts and roots of three Verbascum L species. J Plant Res Iran J Biol 28:596–606
Karavelioğulları FA (2012) Verbascum L. Türkiye Bitk List (Damarlı Bitk Nezahat Gökyiğit Bot Bahçesi ve Flora Araştırmaları Derneği Yayını, İstanbul 850–870
Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS ONE 2:e508
Kress WJ, Erickson DL, Jones FA et al (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proc Natl Acad Sci 106:18621–18626
Kress WJ, Wurdack KJ, Zimmer EA et al (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci 102:8369–8374
Lei W, Ni D, Wang Y et al (2016) Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of Astragalus membranaceus. Sci Rep 6:21669
Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. https://doi.org/10.1093/nar/gkab301
Li X, Yang Y, Henry RJ et al (2015) Plant DNA barcoding: from gene to genome. Biol Rev 90:157–166
Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300
Nofouzi K (2015) Study on the antioxidant activity and in vitro antifungal activity of Verbascum speciosum methanolic extract. J Mycol Res 2:97–103
Pang X, Liu C, Shi L et al (2012) Utility of the trnH–psbA intergenic spacer region and its combinations as plant DNA barcodes: a meta-analysis. PLoS ONE 7:e48833
Remal Puigmal S (2014) Approche morphologique et moléculaire du genre Verbascum L.
Riahi M, Ghahremaninejad F (2019) The tribe scrophularieae (scrophulariaceae): a review of phylogenetic studies. Hacquetia. https://doi.org/10.2478/hacq-2019-0003
Samarina LS, Koninskaya NG, Shkhalakhova RM et al (2025) DNA-barcoding for cultivar identification and intraspecific diversity analysis of agricultural crops. Int J Mol Sci 26:6808
Scriver M, Marinich A, Wilson C, Freeland J (2015) Development of species-specific environmental DNA (eDNA) markers for invasive aquatic plants. Aquat Bot 122:27–31
Shi L, Chen H, Jiang M et al (2019) CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res 47:W65–W73
Soorni A, Golchini MM (2025) Complete chloroplast genome of Mentha aquatica reveals hypervariable regions and resolves phylogenetic position within the genus Mentha. Mol Biol Rep 52:677
Sotoodeh A (2015) Histoire biogéographique et évolutive des genres Verbascum et Artemisia en Iran à l'aide de la phylogénie moléculaire
Sotoodeh A, Attar F, Civeyrel L (2016) A new species of Verbascum L. (Scrophulariaceae) from the Gilan province (Iran), based on morphological and molecular evidences. Adansonia 38:127–132
Su T, Geng Y-F, Xiang C-L et al (2022) Chloroplast genome of salvia sect. Drymosphace: comparative and phylogenetic analysis. Diversity 14:324

Tamura K, Stecher G, Peterson D et al (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Tatli R, Akdemir SZ, Bedir E, Khan IA (2003) Search for antifungal compounds from some verbascum species growing in Turkey. FABAD J Pharm Sci 28:137–140

Tillich M, Lehwark P, Pellizzer T et al (2017) GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45:W6–W11

Vaidya G, Lohman DJ, Meier R (2011) SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27:171–180. https://doi.org/10.1111/j.1096-0031.2010.00329.x

Wick RR, Schultz MB, Zobel J, Holt KE (2015) Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350–3352

Wicke S, Naumann J (2018) Molecular evolution of plastid genomes in parasitic flowering plants. Advances in botanical research. Elsevier, pp 315–347

Wicke S, Schneeweiss GM, Depamphilis CW et al (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol 76:273–297

Yao X, Tan Y-H, Liu Y-Y et al (2016) Chloroplast genome structure in Ilex (Aquifoliaceae). Sci Rep 6:28559

Yılmaz M, Genç GE (2024) Overview of ethnobotanical, phytochemical and biological activity relations of verbascum species in worldwide. Turkish J Biodivers 7:131–154

Zengin G, Mahomoodally MF, Sinan KI et al (2021) Evaluation of chemical constituents and biological properties of two endemic Verbascum species. Proc Biochem 108:110–120

Zhang Y, Guan Y-H, Pu C-Z et al (2022) The complete chloroplast genome of Verbascum thapsus L. (Scrophulariaceae) and its phylogenetic affinities. Mitochondrial DNA Part B 7:1583–1585