

RESEARCH

Open Access



Decoding the chloroplast genomes of five Iranian *Salvia* species: insights into genomic structure, phylogenetic relationships, and molecular marker development

Amir Mohammad Akrami¹, Sepehr Meratian Esfahani¹ and Aboozar Soorni^{1*}

Abstract

Background The genus *Salvia*, a prominent member of the Lamiaceae family, is renowned for its ecological, medicinal, and economic significance. Despite its importance, molecular data, particularly chloroplast (cp.) genome information, remain scarce for many native Iranian *Salvia* species. In this study, we sequenced and analyzed the complete cp. genomes of five Iranian *Salvia* species (*S. aethiopsis*, *S. sclarea*, *S. glutinosa*, *S. verticillata*, and *S. officinalis*) to elucidate their genomic structure, evolutionary relationships, and potential for biotechnological applications.

Results The cp. genomes of the five *Salvia* species exhibited a conserved quadripartite structure, with sizes ranging from 151,163 to 151,662 bp, and a GC content of 38%. Each genome contained 132 or 131 genes, comprising 86 or 87 protein-coding, 8 rRNA, and 37 tRNA genes, with duplications in *rpl2*, *rpl23*, and *rps12*. Minor variations in gene content were observed, such as the absence of *trnS-CGA* in *S. glutinosa*. Comparative analysis of IR boundaries showed subtle expansions in *S. officinalis* and *S. sclarea*, while *S. glutinosa* remained stable. Trans-splicing of the *rps12* gene was observed in all species, with complex structures in *S. glutinosa* and *S. sclarea*. Codon usage analysis revealed a preference for A/U-ending codons, with *S. verticillata* displaying unique patterns. Nucleotide diversity (Pi) identified highly variable regions, such as *rpl14-rpl16* and *psbK-psbI*, as potential molecular markers. Phylogenetic analysis resolved distinct clades, with *S. aethiopsis* and *S. sclarea* forming a close group, *S. glutinosa* clustering with *S. chanryoenica*, and *S. officinalis* showing genetic homogeneity with Mediterranean species. *S. verticillata* exhibited an earlier divergence, highlighting the genus's evolutionary complexity.

Conclusions This study provides critical genomic resources for species identification, phylogenetic studies, and the development of molecular markers, facilitating the conservation of native *Salvia* species and their utilization in breeding programs for medicinal and aromatic traits.

Keywords Lamiaceae, Plastome, Phylogenomics, Barcoding, Diversity

*Correspondence:

Aboozar Soorni
soorni@iut.ac.ir

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



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

The genus *Salvia*, belonging to the Lamiaceae family, is one of the largest and most diverse genera of flowering plants, comprising approximately 1,000 species distributed across tropical, temperate, and subtropical regions worldwide [1–3]. Taxonomically, *Salvia* is characterized by its unique floral morphology, particularly the specialized staminal lever mechanism, which facilitates pollination [4, 5]. *Salvia* species are renowned for their rich chemical diversity, producing a wide array of secondary metabolites, including terpenoids, flavonoids, and phenolic acids, which contribute to their medicinal, aromatic, and ecological significance [6–10]. These bioactive compounds underpin the extensive use of *Salvia* in traditional medicine for treating ailments such as cardiovascular diseases, diabetes, and neurodegenerative disorders [11, 12]. Additionally, *Salvia* essential oils and extracts are widely utilized in the pharmaceutical, food, and cosmetic industries due to their antimicrobial, antioxidant, and anti-inflammatory properties [13].

The *Salvia* genus is divided into three major distribution centers: Central and South America (CASA), Central Asia/Mediterranean (CAM), and East Asia (EA), with each region harboring distinct species adapted to local ecological conditions [2, 4, 14–16]. Iran, as a major center for sage diversity in Asia, boasts 61 *Salvia* species [17, 18], distributed across diverse ecoregions from the humid Caspian coast to the arid central plateau and mountainous regions of the Alborz and Zagros ranges, highlighting its significance as a hotspot for the conservation and study of this ecologically and medicinally important genus. Among the diverse *Salvia* species native to Iran, several are of particular ecological and medicinal importance, each distributed across distinct provinces and regions.

S. aethiopsis is specifically found within Golestan National Park, as well as in Dalamper and Khoy. *S. sclarea* is distributed in Golestan Province, notably in Golestan National Park, and extends to Kahkaraan village in Sepidan, Fars Province. *S. glutinosa* is also predominantly located in Golestan National Park. *S. verticillata* is widely distributed across multiple regions, including Khodkavand village in Taleghan, Alborz Province, as well as Mavana, Qasemlu, Marmisho, Qurdik-Khoy, and Daryan. Lastly, *S. officinalis* is widespread across several provinces, including West Azerbaijan, Fars, and Alborz, and is also present in Golestan National Park [7, 8, 17–19]. Among these species, *S. sclarea* has been demonstrated to exhibit high antioxidant activity in its roots and significant dry matter content in its leaves, making it a valuable source of medicinal compounds. *S. officinalis* also shows notable dry matter content in its leaves and a high essential oil yield, highlighting its importance for pharmaceutical and aromatic applications [20]. *S. aethiopsis* has been

shown to exhibit significant antioxidant properties, further supporting its potential use in health and wellness products according to literature reviews [8]. According to the literature, *S. verticillata* essential oil (EO) exhibits a significant inhibitory effect on cholinesterases, with reported inhibition rates of 20.4% against acetylcholinesterase (AChE) and 1.8% against butyrylcholinesterase (BChE) [7].

Although significant chemical and biochemical information is available for these species, molecular data, particularly cp. genome information, remains scarce or entirely lacking for many of these species. Chloroplasts are vital organelles in plant cells, playing a central role in photosynthesis by converting light energy into chemical energy, a process essential for plant growth and development [21, 22]. The cp. genome is a circular, multicopy DNA molecule characterized by a highly conserved quadripartite structure. This structure includes two inverted repeat (IR) regions (IRa and IRb), which flank a large single-copy (LSC) region and a small single-copy (SSC) region [23–25]. The cp. genome encodes more than 120 genes, which are integral to critical cellular processes such as transcription, translation, photosynthesis, and the biosynthesis of amino acids and fatty acids [26]. Variations in the size of cp. genomes among species are primarily attributed to the expansion or contraction of the IR regions [24]. Due to its uniparental inheritance (typically maternal in angiosperms), low mutation rate, and high conservation, the cp. genome has emerged as a powerful tool for species identification, phylogenetic analysis, and genetic engineering [27–31]. Specific cpDNA sequences, such as *psbA-trnH*, *matK*, and *rbcL*, are widely employed as DNA barcodes for species identification. Furthermore, the complete cp. genome is increasingly recognized as a “super barcode” for resolving phylogenetic relationships at lower taxonomic levels, providing higher resolution and accuracy [32–37]. Beyond its applications in phylogenetic, the cp. genome is also utilized in genetic transformation and molecular breeding, offering significant potential for enhancing medicinal and agricultural traits [26].

Since the initial publication of the first cp. genome of *Salvia* nearly a decade ago [38], there has been a significant increase in the number of *Salvia* cp. genomes deposited in GenBank in recent years. To date, over 100 cp. genomes of *Salvia* have been reported, including multiple entries for some species. For instance, the cp. genomes of *S. multiorrhiza*, *S. przewalskii*, *S. bulleyana*, and *S. japonica* have been sequenced and analyzed, revealing conserved quadripartite structures and providing insights into their phylogenetic relationships and genetic diversity [38, 39]. Similarly, the cp. genomes of *S. bowleyana*, *S. splendens*, and *S. officinalis* have been characterized, identifying hypervariable regions such as

rps16-trnQ-UUG and *trnL-UAA-trnF-GAA* as potential DNA barcodes for species discrimination [40]. Furthermore, comparative studies of *S. sect: Drymospace* have highlighted the utility of cp. genomes in resolving inter-specific relationships and detecting evolutionary patterns [41].

Despite significant advancements in *Salvia* genomics, the chloroplast genomes of several economically and medicinally important species native to Iran remain poorly characterized. While partial cpDNA data exist for some species (*S. sclarea*, *S. glutinosa*, and *S. officinalis*), complete cp. genomes are lacking for ecologically significant taxa like *S. aethiopsis* and *S. verticillata*, and the unique Iranian populations of all five species remain genomically unexplored. These species were strategically selected to represent: (1) Iran's remarkable biogeographic diversity (including endemic, semi-endemic and widely distributed taxa), (2) distinct ecological adaptations (from *S. aethiopsis* in arid highlands to *S. glutinosa* in humid montane forests), and (3) cultural importance in traditional Persian medicine (particularly *S. sclarea* for digestive remedies). To address these critical knowledge gaps and enable future biotechnological applications, we sequenced and analyzed the complete chloroplast genomes of these five Iranian *Salvia* species, performing comprehensive structural characterization, comparative genomics, and phylogenetic analyses to elucidate their evolutionary relationships and conservation priorities. We also performed comparative genomic analyses to identify conserved and variable regions, such as inverted IRs, LSC, and SSC regions, as well as hypervariable regions which are valuable for species identification and phylogenetic studies. Additionally, we analyzed repetitive sequences, codon usage patterns, and RNA editing sites to gain insights into the evolutionary dynamics and functional adaptations of these genomes. Phylogenetic analyses were conducted to resolve the evolutionary relationships among these species and other members of the *Salvia* genus. The genomic resources developed in this study enable: (1) precise authentication of *Salvia* herbal products using *rpl14-rpl16* and *psbK-psbI* barcodes, (2) conservation prioritization of genetically distinct populations through chloroplast DNA markers - addressing critical needs for both sustainable utilization and protection of Iran's *Salvia* diversity.

Results

Genome characteristics

The cp. genomes of five *Salvia* species were sequenced and analyzed, revealing highly conserved structural and compositional features. The cp. genome sizes varied slightly, ranging from 151,163 bp (*S. officinalis*) to 151,662 bp (*S. glutinosa*), with all species exhibiting the characteristic quadripartite structure of cp. genomes

(Fig. 1). This structure included a pair of IR regions, a SSC region, and a large single-copy LSC region. The IR regions ranged from 51,128 bp (*S. aethiopsis* and *S. sclarea*) to 51,224 bp (*S. officinalis*), while the SSC regions varied between 17,508 bp (*S. officinalis*) and 17,642 bp (*S. glutinosa*). The LSC regions showed minor differences, ranging from 82,428 bp (*S. officinalis*) to 82,807 bp (*S. glutinosa*).

The cp. genomes of the five *Salvia* species (*S. aethiopsis*, *S. sclarea*, *S. glutinosa*, *S. verticillata*, and *S. officinalis*) revealed a highly conserved set of protein-coding genes. These included 5 genes associated with Photosystem I (*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*), 15 genes for Photosystem II (*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*), and 6 genes encoding subunits of ATP synthase (*atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI*). Additionally, 11 genes were identified for large ribosomal proteins (*rpl2*, *rpl14*, *rpl16*, *rpl20*, *rpl22*, *rpl23*, *rpl32*, *rpl33*, *rpl36*), and 12 genes for small ribosomal proteins (*rps2*, *rps3*, *rps4*, *rps7*, *rps8*, *rps11*, *rps12*, *rps14*, *rps15*, *rps16*, *rps18*, *rps19*). Notably, the genes *rpl2*, *rpl23*, and *rps12* were duplicated in all species. Other functional genes identified included those for NADPH dehydrogenase (*ndhA*, *ndhB*, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*), cytochrome b/f complex (*petA*, *petB*, *petD*, *petG*, *petL*, *petN*), and the large subunit of Rubisco (*rbcL*). Ribosomal RNA genes (*rrn16S*, *rrn23S*, *rrn4.5 S*, *rrn5S*) and a comprehensive set of transfer RNA genes were also present across all species. Furthermore, genes encoding essential functions such as acetyl-CoA carboxylase (*accD*), c-type cytochrome synthesis (*ccsA*), envelope membrane protein (*cemA*), protease (*clpP*), maturase (*matK*), and translation initiation factor (*infA*) were consistently identified. Hypothetical reading frames (*ycf1*, *ycf2*, *ycf3*, *ycf4*, *ycf15*) were also conserved among the species.

The characterization of cp. genomes in five *Salvia* species also revealed the presence of both *cis*-splicing and *trans*-splicing genes. A total of 10–13 protein-coding genes, including *rps16*, *atpF*, *rpoC1*, *ycf3*, *clpP*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, and *ndhA*, were identified as *cis*-splicing genes, each containing 1–2 introns (Fig S1-S5). The positions of these genes and their intron-exon boundaries were conserved across species, with minor variations in genomic coordinates. Additionally, the *rps12* gene was found to undergo *trans*-splicing in all five species (Fig. 2), with its exons distributed across different regions of the cp. genome. Specifically, the 5'-end exon was located in LSC region, while the 3'-end exon was situated in IR regions. In *S. glutinosa* and *S. sclarea*, the *rps12* gene exhibited a more complex structure, with exons present in both IRa and IRb regions.

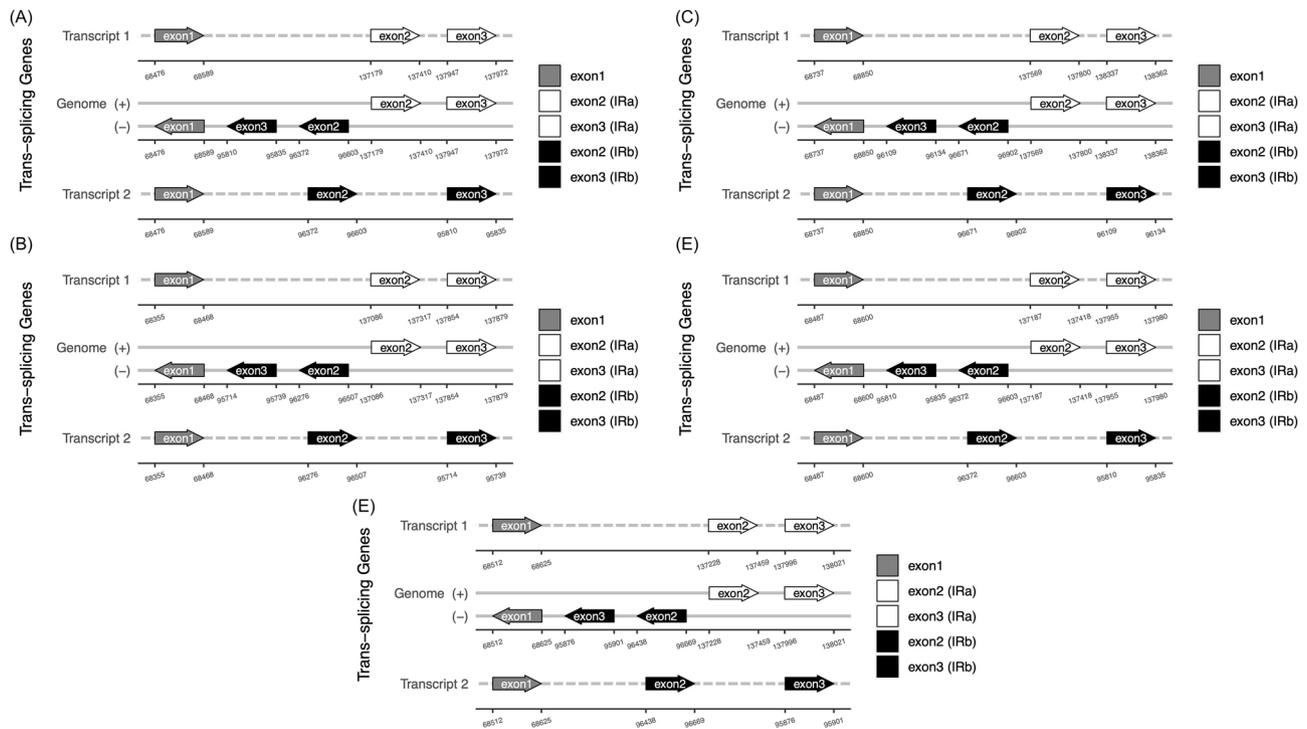


Fig. 2 Trans-splicing of the *rps12* gene in five *Salvia* species: **(A)** *S. aethiopsis*, **(B)** *S. glutinosa*, **(C)** *S. officinalis*, **(D)** *S. sclarea*, and **(E)** *S. verticillata*. The 5'-end exon is located in the LSC region, while the 3'-end exon is situated in the IR regions

Codon usage analysis in *Salvia* species

The codon usage patterns of five *Salvia* species were analyzed, revealing both conserved and species-specific trends. The total number of codons ranged from 26,004 in *S. verticillata* to 26,433 in *S. officinalis*, indicating slight variations in coding sequence length among the species (Fig. 3). According to amino acid frequency results, leucine (Leu) was the most frequently encoded amino acid across all species, with the codon UUA showing the highest RSCU values (ranging from 1.61 to 1.87). In contrast, cysteine (Cys) was the least frequent, with UGU being the predominant codon (RSCU values ranging from 1.05 to 1.52). The standard initiation codon AUG (Met) was universally conserved across all species, with an RSCU value of 1.0. However, variations were observed in stop codon usage. While UAA was the most frequently used stop codon in all species (RSCU values ranging from 1.07 to 1.66), *S. verticillata* exhibited a higher preference for UGA (RSCU = 1.31), which was less pronounced in the other species (RSCU values ranging from 0.62 to 0.66). In general, approximately half of the codons (30 out of 64) showed a usage bias (RSCU > 1.0), with the majority ending in A or U. For example, the codons AGA (Arg), GGA (Gly), and UCU (Ser) consistently exhibited high RSCU values across all species, indicating a strong preference for these codons. In contrast, codons such as CGC (Arg) and GGC (Gly) were less preferred, with RSCU values below 1.0. The synonymous codon usage analysis

revealed a strong bias toward codons ending with adenine (A) or uracil (U) at the third position, consistent with the general trend observed in cp. genomes of higher plants. For example, the codons UUU (Phe), AUU (Ile), and GUU (Val) were highly preferred, with RSCU values exceeding 1.3 in most species. Conversely, codons ending with cytosine (C) or guanine (G), such as UUC (Phe), AUC (Ile), and GUC (Val), were less frequently used, with RSCU values below 1. Among species, *S. verticillata* displayed distinct codon usage patterns compared to the other species. For instance, the codon UUG (Leu) had a higher RSCU value (1.29) in *S. verticillata* compared to the other species (ranging from 1.21 to 1.23). Similarly, the codon CUU (Leu) showed an elevated RSCU value (1.49) in *S. verticillata*, while it ranged between 1.28 and 1.29 in the other species. These differences suggest potential species-specific adaptations in codon usage.

Comparative analysis of plastome IR boundaries

Analysis of boundary regions exhibited variation in the length of the inverted repeat (IR) regions across the *Salvia* species examined in this study. The newly assembled cp. genomes displayed IR lengths of 25,564 bp in *S. aethiopsis*, 25,564 bp in *S. sclarea*, 25,605 bp in *S. glutinosa*, 25,597 bp in *S. verticillata*, and 25,612 bp in *S. officinalis*. For comparative analysis, these sequences were evaluated alongside previously published cp. genomes of *S. sclarea*, *S. officinalis*, and *S. glutinosa* available in NCBI,

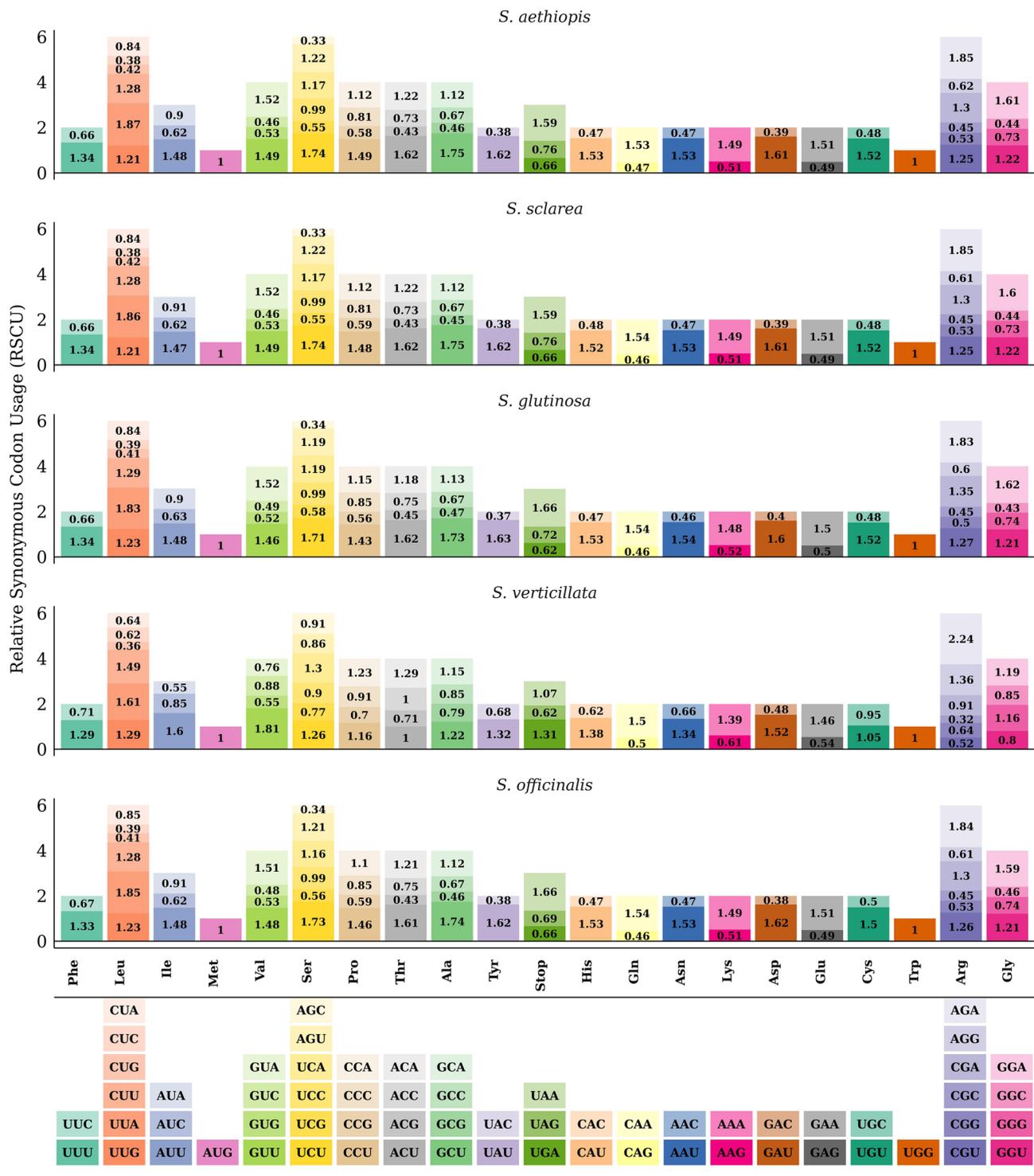


Fig. 3 The bar plot of the Relative Synonymous Codon Usage (RSCU) values for each amino acid, grouped by species. Codons are color-coded, and their corresponding amino acids are labeled below the plot

which exhibited IR lengths of 25,560 bp, 25,591 bp, and 25,605 bp, respectively. These results indicate subtle IR expansion in *S. officinalis* and *S. sclarea*, while *S. glutinosa* maintained structural stability. A detailed comparison of IR boundaries at the four junctions (JLB, JSB,

JSA, and JLA) between the IR regions (IRa and IRb) and the single-copy regions (LSC and SSC) revealed notable structural variations (Fig. 4). The *rps19* gene was consistently located at the LSC/IRb junction, with 233–237 bp positioned within the LSC and 42–43 bp extending into

Inverted Repeats

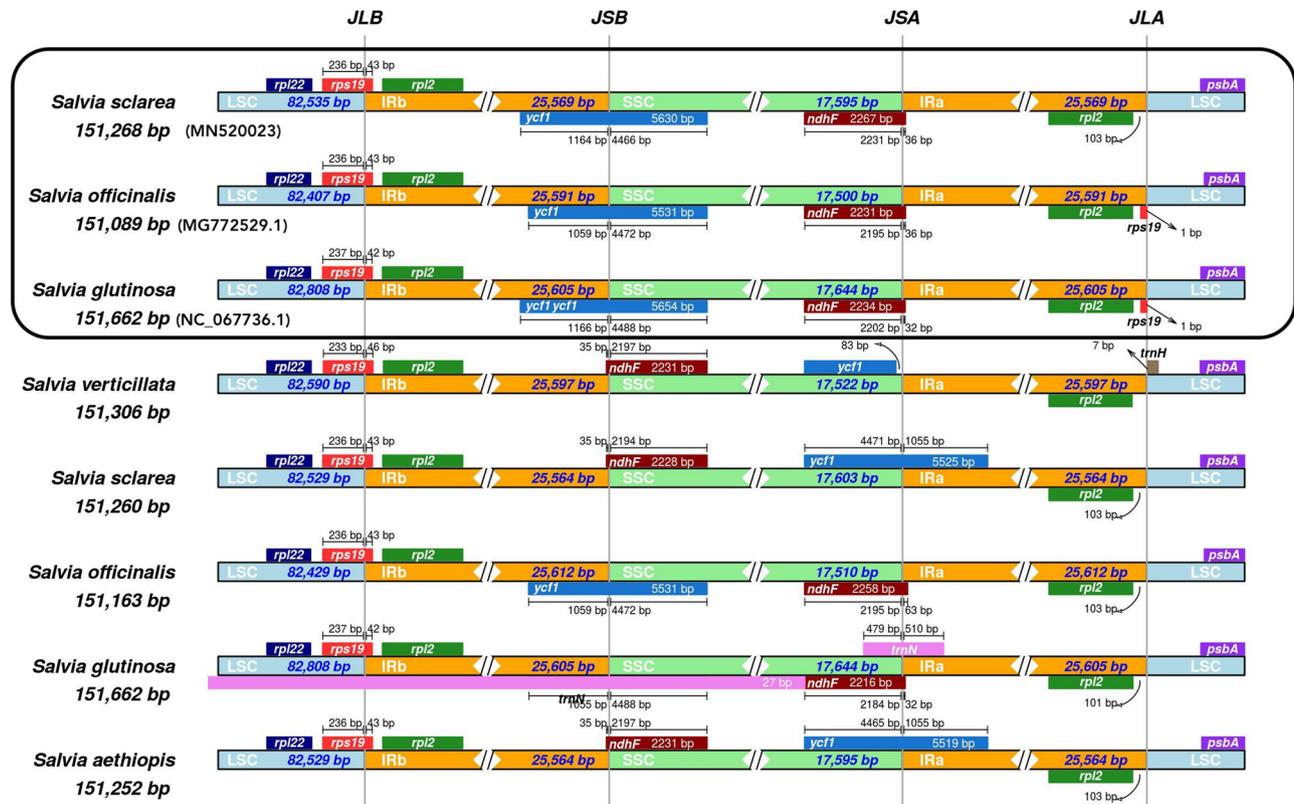


Fig. 4 Comparison of the boundaries of the large single-copy (LSC), small single-copy (SSC), and inverted repeat (IR) regions among five *Salvia* species and three top-related species obtained from the database. Key junctions are labeled: JLB (LSC/IRb), JSB (SSC/IRb), JSA (SSC/IRa), and JLA (LSC/IRa), high-lighting structural variations at the IR-LSC and IR-SSC borders

IRb across all species. The *ndhF* gene, situated in the SSC region, exhibited variability in its positioning relative to the IRb/SSC boundary. While *ndhF* remained within the SSC in all species, its proximity to the boundary differed. In some species, such as *S. glutinosa* and *S. officinalis*, *ndhF* extended slightly into IRa, whereas in others, it extended marginally into IRb. Additionally, the *ycf1* gene spanned the SSC/IR junction in all species, though minor positional shifts were observed between the previously sequenced and newly assembled cp. genomes. Notably, *ycf1* extended further into IRa in *S. glutinosa* (sequenced in this study) and *S. aethiopsis*, while it appeared slightly contracted in *S. verticillata*. In other species, the *ycf1* gene extended into IRb, further highlighting the structural diversity among the cp. genomes of *Salvia* species.

SSR and long repeats analysis

SSR analysis performed on five species of the genus *Salvia* revealed a total of 18, 28, 24, 20, and 20 SSRs in *S. aethiopsis*, *S. glutinosa*, *S. officinalis*, *S. sclarea*, and *S. verticillata*, respectively. Mononucleotide repeats

dominated the SSR profiles, constituting 100% of the total SSRs in *S. aethiopsis*, *S. officinalis*, *S. sclarea*, and *S. verticillata*, and 96% (27/28) in *S. glutinosa*. The A/T repeats were the most prevalent, accounting for all SSRs in *S. aethiopsis* (18/18), *S. officinalis* (24/24), and *S. sclarea* (20/20), and the majority in *S. glutinosa* (27/28) and *S. verticillata* (20/20). Dinucleotide repeats were rare, with only one TA repeat detected in *S. glutinosa*. The majority of SSRs across all species were 10–13 bp in length, with a smaller number of longer repeats (up to 16 bp). Compound microsatellites were infrequent, being observed only in *S. glutinosa* (1) and *S. verticillata* (2). The analysis of long repeats in the cp. genomes of five *Salvia* species revealed a conserved pattern in the distribution and types of repeats. In *S. aethiopsis*, 31 long repeats were identified, comprising 16 palindromic, 13 forward, and 1 reverse repeat, with 94% of repeats falling within the 30–40 bp range and 6% in the 40–50 bp range. *S. glutinosa* exhibited 46 long repeats, including 21 palindromic and 25 forward repeats, with 91% in the 30–40 bp range and 8% in the 40–50 bp range. *S. officinalis* contained 41

long repeats, consisting of 21 palindromic, and 20 forward, with 88% in the 30–40 bp range and 22% in the 40–50 bp range. Similarly, *S. sclarea* had 31 long repeats, with 17 palindromic, 13 forward, and 1 reverse repeat, and a length distribution of 74% in the 30–40 bp range and 26% in the 40–50 bp range. Finally, *S. verticillata* displayed 38 long repeats, including 20 palindromic, and 18 forward, with 92% in the 30–40 bp range and 8% in the 40–50 bp range. Across all species, palindromic repeats were the most abundant, followed by forward repeats, while reverse repeats were rare, and no complement repeats were observed. The majority of long repeats were 30–40 bp in length, indicating a highly conserved organization of long repeats in *Salvia* cp. genomes, likely contributing to genome stability and evolutionary adaptation.

Plastome analysis and divergence

To investigate the extent of sequence divergence and nucleotide diversity among the five *Salvia* species, the mVISTA tool was employed to align the cp. genomes, using *S. officinalis* as the reference (Fig S6). The results demonstrated that LSC and SSC regions exhibited significantly higher sequence divergence compared to IR regions, which were more conserved. The coding regions, particularly those involved in essential photosynthetic and metabolic functions, were generally more conserved, while the non-coding regions, including intergenic spacers (IGS) and introns, showed higher variability. Several IGS regions were identified as highly divergent, including *trnH-GUG-psbA*, *trnK-UUU-matK*, *rps16-trnQ-UUG*, *trnQ-UUG-psbK*, *trnS-UGA-psbZ*, *trnG-GCC-trnM-CAU*, and *petG-trnW-CCA*. These regions exhibited significant sequence variation among the five *Salvia* species, making them potential hotspots for species-specific divergence. For example, the *trnH-GUG-psbA* spacer, located at the beginning of the LSC region, showed high variability, which is consistent with its frequent use as a DNA barcode in phylogenetic studies. Similarly, the *rps16-trnQ-UUG* and *trnG-GCC-trnM-CAU* spacers also displayed notable divergence, further highlighting the utility of non-coding regions in identifying species-specific markers. Additionally, the *petG-trnW-CCA* spacer, located in the SSC region, exhibited high variability, suggesting its potential for distinguishing closely related species within the genus.

Pi analysis further quantified the variability across the cp. genomes of the five *Salvia* species (Fig. 5). Among the coding regions, several genes exhibited exceptionally high variability, including *trnF-GAA* (Pi=0.4162), *ycf1* (Pi=0.03393), *trnQ-UUG* (Pi=0.02778), *rpl22* (Pi=0.02467), *rps15* (Pi=0.02454), *matK* (Pi=0.0232), *ndhF* (Pi=0.02154), and *trnS-UGA* (Pi=0.02151). These genes, with Pi values significantly higher than the

genome-wide average, are strong candidates for molecular markers due to their high variability and potential utility in phylogenetic studies and species identification. In contrast, genes like *rbcl* (Pi=0.00837), *psbA* (Pi=0.0085), and *rpoB* (Pi=0.00809) exhibited lower Pi values, reflecting their conserved nature. The *rps12* gene, often used as a marker in cp. studies, showed relatively low variability (Pi=0.00215), while *rpl32* (Pi=0.01605) displayed moderate divergence. These findings highlight the contrasting levels of variability across different regions and genes, with the highly divergent coding genes and IGS regions being particularly promising for marker development. Among IGS, several regions exhibited exceptionally high Pi values, including *trnH-GUG-psbA* (Pi=0.063), *trnG-GCC-trnM-CAU* (Pi=0.06), *petG-trnW-CCA* (Pi=0.044), *rpl14-rpl16_2* (Pi=0.043), *rps16_1-trnQ-UUG* (Pi=0.043), *trnS-UGA-psbZ* (Pi=0.042), *trnK-UUU_1-rps16_2* (Pi=0.041), *psbK-psbI* (Pi=0.04). These regions, with Pi values significantly higher than the genome-wide average, are strong candidates for molecular markers due to their high variability. In contrast, some IGS regions, such as *trnR-UCU-atpA* (Pi=0.02286) and *psbZ-trnG-GCC* (Pi=0.03358), showed moderate variability, while others like *trnL-UAA_1-trnL-UAA_2* (Pi=0.01826) and *psbA-trnK-UUU_2* (Pi=0.02124) exhibited lower divergence.

Phylogenetic analysis

The phylogenetic analysis of the *Salvia* genus revealed intricate evolutionary relationships among the studied species (Fig. 6). *S. aethiopsis* clustered closely with *S. sclarea*, indicating a strong evolutionary relationship and suggesting a recent divergence from a shared ancestor; both species grouped within a larger clade that included other *Salvia* species, such as *S. deserta* and *S. merjamie*, further supporting their genetic affinity. In contrast, *S. glutinosa* formed a distinct clade, showing a more distant relationship to *S. aethiopsis* and *S. sclarea*, and instead exhibited closer genetic ties to species like *S. chanryoenica* and *S. glabrescens*, highlighting its unique phylogenetic position within the genus. *S. officinalis* demonstrated high genetic homogeneity, as evidenced by its well-supported clade that included multiple isolates and vouchers, and was phylogenetically distinct from *S. aethiopsis* and *S. glutinosa*, instead showing closer relationships to other Mediterranean species such as *S. rosmarinus*. *S. verticillata* was placed in a separate clade alongside *S. nilotica* and *S. forsikkaolei*, indicating an earlier divergence from the other studied species and a distinct evolutionary trajectory. The inclusion of database sequences revealed that *S. verticillata* shared a more distant relationship with *S. officinalis* and *S. glutinosa*, further emphasizing its unique phylogenetic placement.

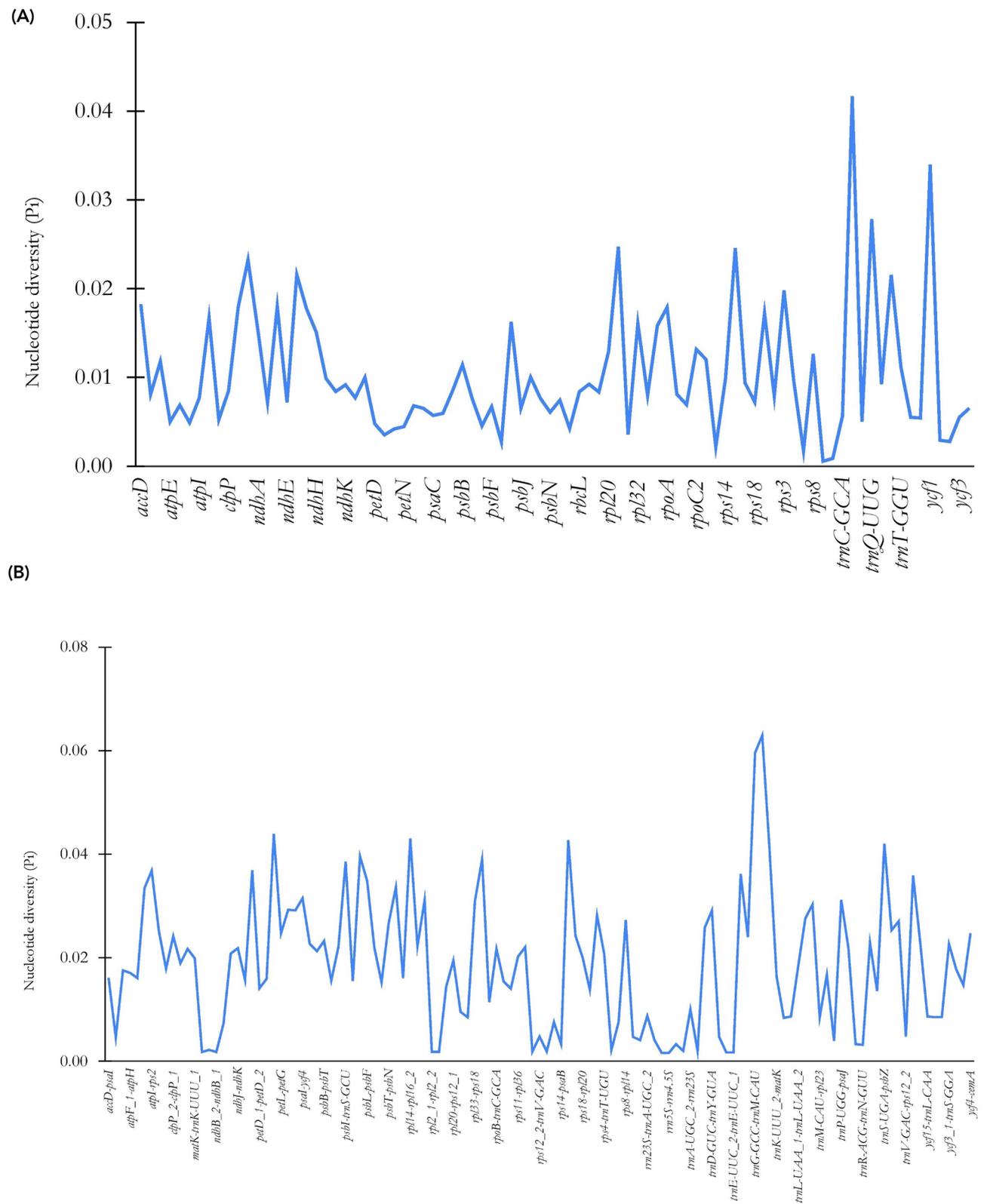


Fig. 5 Nucleotide diversity (Pi) across **(A)** coding and **(B)** intergenic spacer (IGS) regions in the cp. genomes of five *Salvia* species

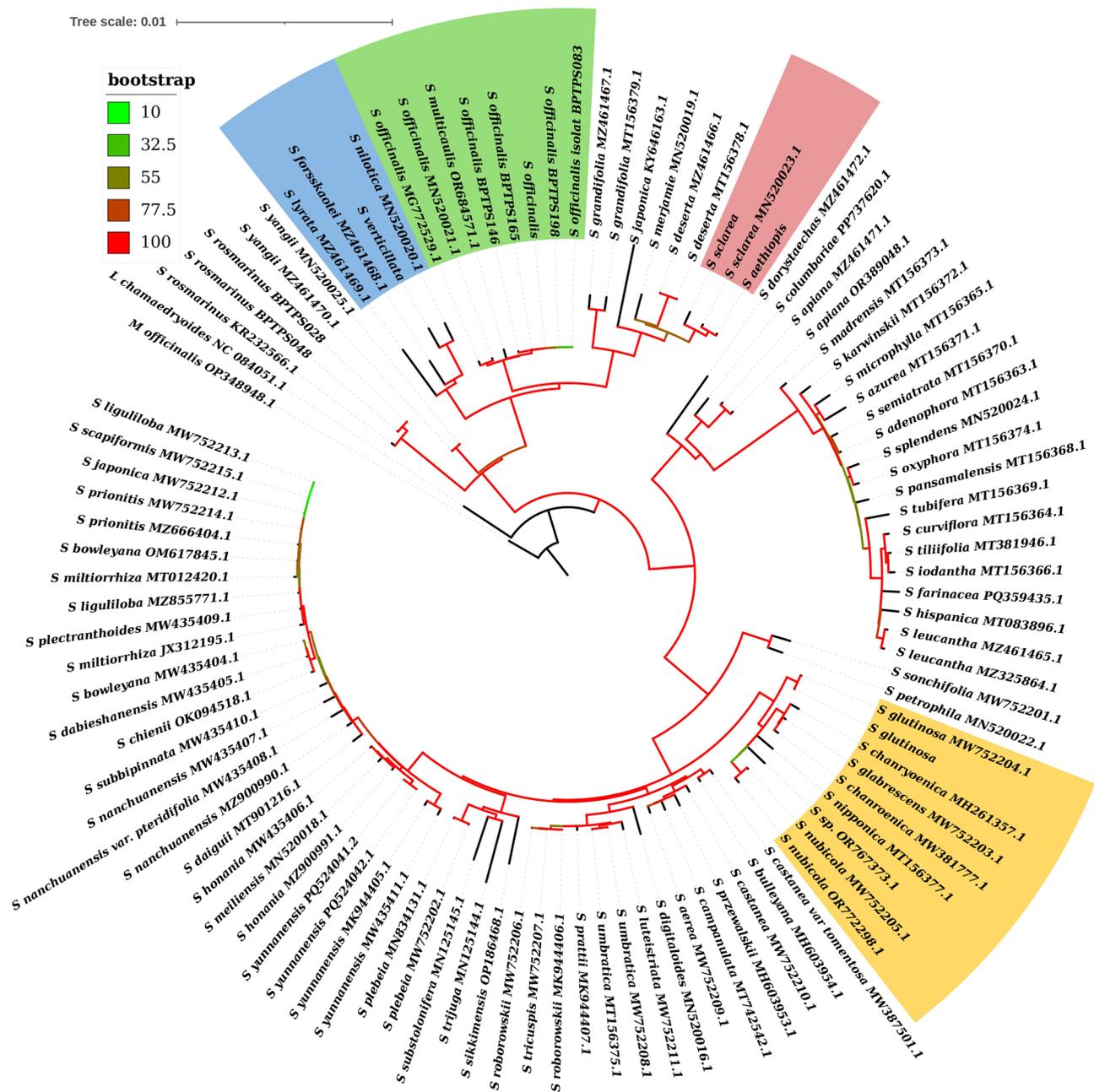


Fig. 6 Phylogenetic tree of 104 cp. genomes, including 72 *Salvia* species and two outgroups (*Melissa officinalis* and *Lepechinia chamaedryoides*), reconstructed using maximum likelihood in IQ-TREE under the GTR + Gamma model. The highlighted region emphasizes the position and relationships of the sample sequences from this study, illustrating their clustering with closely related species. Branch support was assessed with 1000 bootstrap replicates

Phylogenetic analysis based on IGS regions

Phylogenetic analysis based on the hypervariable IGS regions *rpl14-rpl16* (Fig. 7A) and *psbK-psbI* (Fig. 7B) revealed a tree topology highly congruent with that inferred from whole chloroplast genomes, demonstrating the reliability of these markers for resolving evolutionary relationships within *Salvia*. The IGS phylogeny consistently recovered major species groupings observed in the whole-genome analysis, with *S. aethiopsis* and *S. sclarea* forming a strongly supported clade (bootstrap $\geq 95\%$)

alongside *S. deserta* and *S. merjamie*, reflecting their close genetic affinity. Similarly, *S. glutinosa* clustered distinctly with *S. chanryoenica* and *S. glabrescens*, while *S. officinalis* isolates formed a monophyletic group closely related to *S. rosmarinus*, mirroring their shared Mediterranean origin. The early-diverging position of *S. verticillata*, which grouped with *S. nilotica* and *S. forskkaolei*, was also replicated in the IGS tree. Overall, the strong topological concordance between the two phylogenetic approaches validates the utility of these IGS regions as

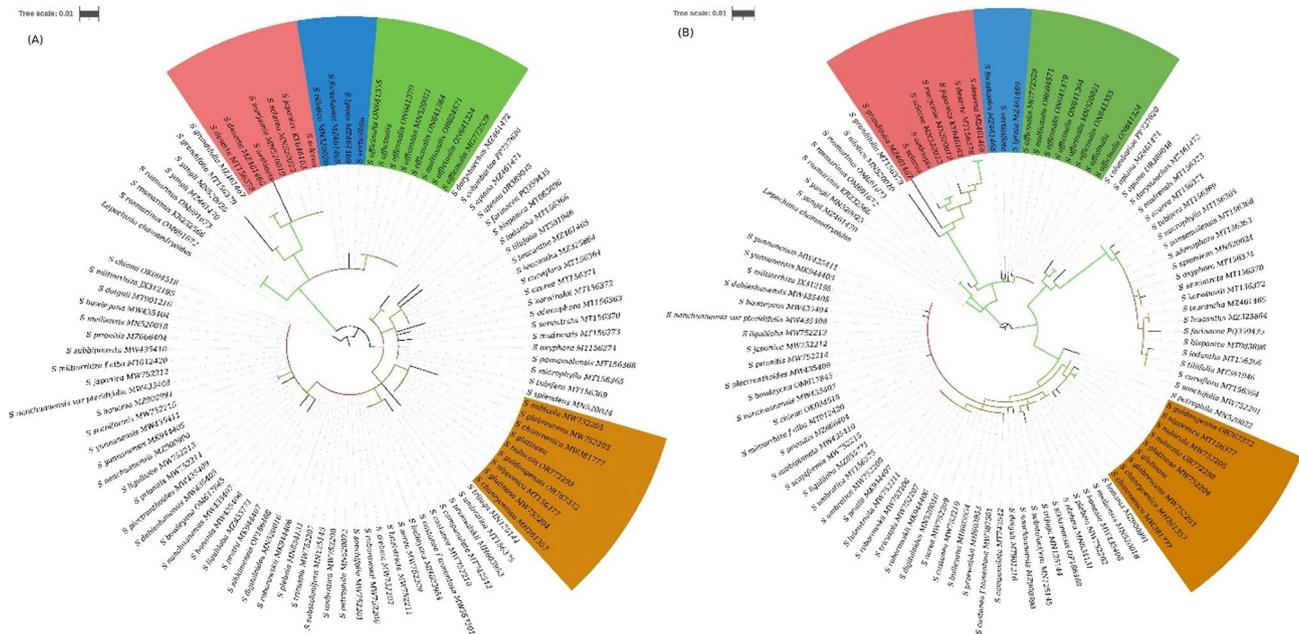


Fig. 7 Phylogenetic tree of 101 cp. genomes, including 100 *Salvia* samples and one outgroups (*Lepechinia chamaedryoides*), reconstructed using IGS regions *rpl14-rpl16* (A) and *psbK-psbI* (B) and maximum likelihood in IQ-TREE under the GTR+Gamma model. The highlighted region emphasizes the position and relationships of the sample sequences from this study, illustrating their clustering with closely related species. Branch support was assessed with 1000 bootstrap replicates

robust, cost-effective markers for delineating species relationships and inferring evolutionary histories within *Salvia*.

Discussion

The complete chloroplast (cp.) genomes of five Iranian native *Salvia* species (*S. aethiopsis*, *S. sclarea*, *S. glutinosa*, *S. verticillata*, and *S. officinalis*) were sequenced and analyzed, revealing highly conserved structural and functional features while also identifying species-specific variations. These findings contribute to the growing body of knowledge on *Salvia* cp. genomes, particularly for species native to Iran, which have been underrepresented in previous genomic studies. The cp. genomes of the five species sequenced here exhibited a typical quadripartite structure, with sizes ranging from 151,163 bp (*S. officinalis*) to 151,662 bp (*S. glutinosa*), consistent with the genome sizes reported for other *Salvia* species [39, 41–45]. Indeed, to date, 173 *Salvia* cp. genomes are available in the NCBI database, representing 72 unique species, with sequence lengths ranging from 150,604 bp to 156,047 bp. Among these, *S. officinalis* exhibits a length range of 151,089 bp to 151,164 bp, while *S. sclarea* and *S. glutinosa* have single entries of 151,268 bp and 151,662 bp, respectively. Notably, *S. aethiopsis* and *S. verticillata* lack complete cp. genome data in the NCBI database, underscoring the novelty of this study. Previous studies have demonstrated that the variability in cp. genome length among species is primarily driven by the

expansion or contraction of IR regions, IR loss, variations in intergenic spacers and intron lengths, parasitic adaptations, polyphyly, and potential technical artifacts arising from sequencing and assembly methodologies. These factors collectively underscore the dynamic nature of cp. genome evolution. Additionally, genes such as *atpA*, *accD*, and *ycf1* have been shown to contribute approximately 13% to genome size variation, playing a significant role in shaping cp. genome evolution. The elevated K_a/K_s ratios (> 1) observed in these genes suggest that positive selection has influenced their evolution, potentially impacting energy generation and ecological strategies in seed plants [46, 47]. These findings highlight the complex interplay of structural, functional, and evolutionary forces driving cp. genome diversity.

The chloroplast genomes of *Salvia* species demonstrate a high degree of conservation, yet variability in unique gene content offers valuable evolutionary insights. In this study, the chloroplast genomes of the analyzed *Salvia* species encoded a conserved set of genes, including 86 or 87 protein-coding, 8 rRNA, and 37 tRNA genes, with duplications in *rpl2*, *rpl23*, and *rps12*, consistent with typical angiosperm chloroplast structure. Comparisons with previous studies revealed variability in reported gene counts of some *Salvia* species such as *S. miltiorrhiza*, *S. hispanica*, *S. yangii*, *S. leucantha*, *S. daiguii*. This variability, driven by tRNA and protein-coding genes, indicate potential gene losses or gains, possibly reflecting evolutionary divergence or annotation differences. The

duplication of *rpl2*, *rpl23*, and *rps12* genes in the inverted repeat regions was a conserved feature across all *Salvia* species examined in this study and mirrored findings from prior research, reinforcing a shared evolutionary mechanism for genomic stability despite fluctuations in unique gene numbers [38, 40, 43, 44, 48, 49].

The variations in IR lengths and boundary dynamics observed in this study are consistent with patterns reported in previous *Salvia* and Lamiaceae research, pointing to a common evolutionary mechanism within the genus [24, 26, 39, 41]. Earlier studies, such as Qian et al. (2013) on *S. miltiorrhiza* [38] and Yu et al. (2023) across various *Salvia* species [44], reported IR lengths typically ranging from 25,500 to 25,700 bp, aligning with our observed expansions in *S. officinalis* and *S. sclarea* and stability in *S. glutinosa*. Du et al. (2022) similarly documented IR boundary variability in three medicinal *Salvia* species (*S. bowleyana*, *S. splendens*, and *S. officinalis*), attributing these shifts to small insertions, deletions, or gene duplications near junctions, which corroborates our findings [40]. The consistent positioning of the *rps19* gene at the LSC/IRb junction, with a portion extending into the IR region, reflects a highly conserved feature of cp. genome architecture across *Salvia* species. This conservation is likely driven by the functional importance of *rps19* in ribosome [26] assembly and protein synthesis, as evidenced by its critical role in maintaining chloroplast function [39]. However, the variability in the positioning of the *ndhF* and *ycf1* genes relative to the IR boundaries highlights the dynamic nature of cp. genome evolution. For instance, the extension of *ndhF* into IRa in *S. glutinosa* and *S. officinalis*, and its marginal extension into IRb in other species, suggests that IR boundary shifts may play a role in optimizing gene expression and functional efficiency. These variations could influence the regulation of *ndhF*, which is involved in photosynthetic electron transport, potentially affecting energy production and stress responses [26]. Similarly, the positional shifts in *ycf1*, particularly its extension into IRa in *S. glutinosa* and *S. aethiopsis*, and contraction in *S. verticillata*, underscore the potential for IR boundary adjustments to influence genome stability and gene regulation. The *ycf1* gene, though its exact function remains unclear, is known to play a role in protein translocation and chloroplast biogenesis, and its duplication status may impact these processes [50]. These findings align with studies in other angiosperms, where IR boundary shifts have been linked to adaptive evolution and functional diversification [51–53]. The observed variability in *ndhF* and *ycf1* positioning may also reflect species-specific adaptations to ecological niches. For example, the extension of *ndhF* into IRa in *S. glutinosa* and *S. officinalis* could enhance the stability and expression of this gene, potentially conferring adaptive advantages in specific environments. Conversely, the

contraction of *ycf1* in *S. verticillata* might indicate a more streamlined genome structure, possibly linked to its ecological specialization. These structural variations could serve as valuable phylogenetic markers for resolving taxonomic relationships within *Salvia*, as IR boundary shifts are often species-specific and can provide insights into evolutionary divergence [41].

The observed patterns of variability, particularly in non-coding regions such as *trnH-GUG-psbA*, *rps16-trnQ-UUG*, *rpl14-rpl16* and *psbK-psbI* align with their established use as DNA barcodes in plant taxonomy [32, 33]. These regions, characterized by high nucleotide diversity, are particularly effective for distinguishing closely related species due to their rapid evolutionary rates and low functional constraints. The utility of these non-coding regions as molecular markers is further supported by their frequent application in phylogenetic studies across diverse plant taxa, including *Salvia* [34, 35]. In addition to non-coding regions, several coding genes, such as *ycf1* and *matK*, exhibited high variability, making them strong candidates for phylogenetic markers. These genes, known for their high substitution rates, are particularly valuable for resolving species-level relationships and understanding evolutionary divergence within the genus [26]. In contrast, conserved genes like *rbcl* and *psbA*, which are under strong selective pressure to maintain photosynthetic function, are less suitable for species discrimination but remain essential for deeper phylogenetic analyses and functional genomics. The contrasting levels of variability between coding and non-coding regions underscore the importance of selecting appropriate markers based on the specific goals of the study. Non-coding regions, with their higher mutation rates, are ideal for species identification and resolving recent divergences, while conserved coding regions are better suited for understanding broader evolutionary relationships [41].

Conclusion

This study provides a comprehensive genomic resource for five native Iranian *Salvia* species, significantly advancing our understanding of their cp. genome architecture, evolutionary relationships, and potential biotechnological applications. By sequencing and analyzing the complete cp. genomes of *S. aethiopsis*, *S. sclarea*, *S. glutinosa*, *S. verticillata*, and *S. officinalis*, we have filled critical gaps in the genomic data available for these ecologically and medicinally important species. The findings underscore the conserved nature of cp. genome structure across the genus while revealing species-specific variations that offer valuable insights into their evolutionary dynamics and functional adaptations. The identification of highly variable regions, such as IGS regions *rpl14-rpl16* and *psbK-psbI*, provides robust molecular markers

for species discrimination and phylogenetic studies. These markers, coupled with the observed patterns of codon usage and IR boundary dynamics, highlight the potential for leveraging cp. genome data in resolving taxonomic uncertainties and understanding the evolutionary history of *Salvia*. The genomic data generated here can be instrumental in developing breeding programs to enhance medicinal and aromatic traits, as well as in guiding conservation strategies for native species. Ultimately, these findings not only enrich our understanding of *Salvia* genomics but also provide a foundation for applied research in biotechnology, agriculture, and conservation. In summarize, the identified molecular markers and genomic features in this study enable three key applications: (1) authentication of *Salvia* herbal products using the highly variable *rpl14-rpl16* and *psbK-psbI* to combat adulteration in commercial markets, (2) targeted breeding of stress-tolerant cultivars by leveraging the conserved stress-response genes in the *S. aethiopsis* - *S. sclarea* clade, and (3) conservation prioritization of genetically distinct populations like *S. glutinosa* through cpDNA barcoding.

Materials and methods

Plant material and DNA extraction

Leaves from five native Iranian *Salvia* species were collected during their flowering stage from four distinct provinces in Iran. These species included *S. aethiopsis*, *S. sclarea*, *S. glutinosa*, *S. verticillata*, and *S. officinalis*. No specific permissions were required for the collection of these plant samples. The plant materials were formally identified by Dr. Soorni and are available at the Bu-Ali Sina University Herbarium (BASU). Table S1 gives details about the studied *Salvia* species, collection sites, and herbarium voucher numbers of the specimens. Since the study did not involve the description of a novel species, no new voucher specimens were deposited in the herbarium collection. To obtain cp. genome sequence, total genomic DNA was extracted from 100 mg of fresh leaf tissue using the DNeasy Plant Mini Kit (QIAGEN, Germany). The quality and concentration of the extracted DNA were assessed through agarose gel electrophoresis and a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific Inc., USA). DNA samples meeting the required quality standards were subsequently used for library preparation, following the manufacturer's guidelines. Sequencing was performed on the Illumina HiSeq 2000 platform (Illumina Inc., USA) using a paired-end 150 bp read configuration.

Cp genome assembly and annotation

For each accession, quality of raw data was assessed using FastQC software version 0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads

containing adapters or exhibiting low quality were filtered and trimmed using Trimmomatic version 0.39 [54]. The resulting high-quality clean reads were subsequently assembled into cp. genomes using GetOrganelle version 1.7.7.1 [55]. The assembly graphs were visually inspected and validated using Bandage software [56] to ensure the completeness of the cp. genomes. The average coverage of the final genome assembly was 128X. Subsequently, the cp. genomes were annotated automatically using two tools: CPGAWAS2 [57] and GeSeq [58]. Circular representations of the cp. genomes were generated using OGDRAW, a tool specifically designed for creating organelle genome maps [59]. Besides, The gene composition of the cp. genome of *Salvia* species was characterized using CPGView [60].

Codon usage analysis

Patterns of codon usage and nucleotide composition provide a theoretical framework for the genetic modification of cp. genomes. In this study, we assessed the variation in synonymous codon usage by calculating the relative synonymous codon usage (RSCU) values. This analysis was conducted using the MEGA6 software [61], which facilitated the determination of RSCU metrics. The resulting RSCU patterns were visualized using an interactive RSCU plot generated with the RSCU-Plot Shiny app (<https://pcg-lab.shinyapps.io/RSCU-Plot/>).

Analysis for boundary regions of CP-genomes

The cp. genomes of five *Salvia* species, along with the cp. genomes of three other species (*S. sclarea*, *S. officinalis*, and *S. glutinosa*) retrieved from NCBI (NC_050900, NC_038165, NC_067736), were thoroughly compared and analyzed at the genome level. In detail, the contraction and expansion of inverted repeat (IR) regions among the large single-copy (LSC), IRb, small single-copy (SSC), and IRa regions were visualized using IRScope [62]. This analysis included a detailed examination of the four junction points (JLB, JSB, JSA, and JLA) between the IR regions and single-copy regions.

Genome comparison and nucleotide variation analysis

The comparative analysis of cp. genomes among five *Salvia* species was conducted using a series of bioinformatics tools. Initially, pairwise alignments were performed using the online software mVISTA with the Shuffle-LAGAN algorithm [63], employing the annotated cp. genome of *S. officinalis* as the reference sequence. Subsequently, the five *Salvia* cp. regions were aligned using MUSCLE version v3.8.1551 [64], followed using trimAl v1.4 with the parameter “-gt 0.95 -st 0.001” [65] to remove poorly aligned regions and gaps. To assess nucleotide diversity (Pi) within the *Salvia* cp. genomes, DnaSP v6 was utilized [66]. A sliding window analysis was implemented,

with parameters set to a step size of 200 bp and a window length of 800 bp, to evaluate genetic diversity across the genomes.

Phylogenetic analysis

To investigate the phylogenetic relationships within the *Salvia* genus, we collected coding sequences from our sequenced species, along with 99 additional cp. genomes from the NCBI database, representing 67 *Salvia* species. In total, the dataset comprised 104 cp. genomes belonging to 72 *Salvia* species. Following *Salvia* species, we selected *Melissa officinalis* and *Lepechinia chamaedryoides* as outgroups. Each gene sequence was aligned using MUSCLE version v3.8.1551 [64] with default parameters to ensure accurate multiple sequence alignment. The aligned sequences were then trimmed using trimAl v1.4 with the parameter “-gt 0.95 -st 0.001” [65] to remove poorly aligned regions and gaps, ensuring high-quality alignments for downstream analyses. The trimmed gene alignments were concatenated into a single sequence matrix using SequenceMatrix [67], creating a comprehensive dataset for phylogenetic reconstruction. Maximum likelihood phylogenetic analysis was performed using IQ-TREE [68] with the GTR+Gamma model of nucleotide substitution. Branch support was assessed using 1000 bootstrap replicates to ensure robustness of the inferred tree topology. The final phylogenetic tree was visualized and annotated using the iTOL web server [69].

Validation of hypervariable IGS markers

To evaluate the phylogenetic utility of the candidate intergenic spacer (IGS) regions *rpl14-rpl16* and *psbK-psbI*, these loci were extracted across a comprehensive dataset comprising 100 *Salvia* samples and one outgroup (*L. chamaedryoides*). Sequence alignment, quality trimming, and phylogenetic reconstruction were performed using standardized bioinformatics pipelines, as detailed in preceding methodological sections.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11729-0>.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

AMA: investigation; methodology; formal analysis, writing-review. SME: Investigation; methodology; funding acquisition; writing-review and editing. 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

All data were deposited in the NCBI database under the project PRJNA1236251 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1236251>). The assembled and annotated genomes are accessible in NCBI database under the research accessions PV340578, PV340579, PV340580, PV340581, and PV340582.

Declarations

Ethics 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.

Competing interests

The authors declare no competing interests.

Received: 16 March 2025 / Accepted: 19 May 2025

Published online: 30 May 2025

References

- Li Q, Li M, Yuan Q, Cui Z, Huang L, Xiao P. Phylogenetic relationships of *Salvia* (Lamiaceae) in China: evidence from DNA sequence datasets. *J Syst Evol.* 2013;51:184–95.
- Drew BT, González-Gallegos JG, Xiang C-L, Kriebel R, Drummond CP, Walked JB, et al. *Salvia* united: the greatest good for the greatest number. *Taxon.* 2017;66:133–45.
- Hu G-X, Takano A, Drew BT, Liu E-D, Soltis DE, Soltis PS, et al. Phylogeny and staminal evolution of *Salvia* (lamiaceae, nepetoideae) in East Asia. *Ann Bot.* 2018;122:649–68.
- Walker JB, Sytsma KJ. Staminal evolution in the genus *Salvia* (Lamiaceae): molecular phylogenetic evidence for multiple origins of the staminal lever. *Annals Bot.* 2007;100:375–91.
- Kriebel R, Drew BT, Drummond CP, González-Gallegos JG, Celep F, Mahdjoub MM, et al. Tracking Temporal shifts in area, biomes, and pollinators in the radiation of *Salvia* (sages) across continents: leveraging anchored hybrid enrichment and targeted sequence data. *Am J Bot.* 2019;106:573–97.
- Lu Y, Foo LY. Antioxidant activities of polyphenols from Sage (*Salvia officinalis*). *Food Chem.* 2001;75:197–202.
- Gharehbagh HJ, Ebrahimi M, Dabaghian F, Mojtavavi S, Hariri R, Saeedi M, et al. Chemical composition, cholinesterase, and α -glucosidase inhibitory activity of the essential oils of some Iranian native *Salvia* species. *BMC Complement Med Ther.* 2023;23:184.
- Moshari-Nasirkandi A, Iaccarino N, Romano F, Graziani G, Alirezalu A, Alipour H, et al. Chemometrics-based analysis of the phytochemical profile and antioxidant activity of *Salvia* species from Iran. *Sci Rep.* 2024;14:17317.
- Topçu G. Bioactive triterpenoids from *Salvia* species. *J Nat Prod.* 2006;69:482–7.
- Jažo Z, Glumac M, Paštar V, Bektić S, Radan M, Carev I. Chemical composition and biological activity of *salvia officinalis* L. Essential oil. *Plants.* 2023;12:1794.
- Huang M, Wang P, Xu S, Xu W, Xu W, Chu K, et al. Biological activities of Salvi-anolic acid B from *Salvia miltiorrhiza* on type 2 diabetes induced by high-fat diet and streptozotocin. *Pharm Biol.* 2015;53:1058–65.
- Ho JH-C, Hong C-Y. Salvi-anolic acids: small compounds with multiple mechanisms for cardiovascular protection. *J Biomed Sci.* 2011;18:1–5.
- Ghorbani A, Esmaeilzadeh M. Pharmacological properties of *Salvia officinalis* and its components. *J Tradit Complement Med.* 2017;7:433–40.
- Walker JB, Sytsma KJ, Treutlein J, Wink M. *Salvia* (Lamiaceae) is not monophyletic: implications for the systematics, radiation, and ecological specializations of *Salvia* and tribe mentheae. *Am J Bot.* 2004;91:1115–25.
- Hao D-C, Xiao P-G. Genomics and evolution in traditional medicinal plants: road to a healthier life. *Evol Bioinforma.* 2015;11:EBO–S31326.
- Hu J, Zhao M, Hou Z, Shang J. The complete Chloroplast genome sequence of *Salvia miltiorrhiza*, a medicinal plant for preventing and treating vascular dementia. *Mitochondrial DNA Part B.* 2020;5:2460–2.

17. Jamzad Z. A survey of Lamiaceae in the flora of Iran. *Rostaniha*. 2013;14:59–67.
18. Askari SF, Avan R, Tayarani-Najaran Z, Sahebkar A, Eghbali S. Iranian *Salvia* species: A phytochemical and Pharmacological update. *Phytochemistry*. 2021;183:112619.
19. Shahhoseini R, Hossaini SM, Nikjouyan MJ. Investigation of the growth and phytochemical variations for different *Salvia* species in terms of their transfer and adaptation in Iran. *J Appl Res Med Aromat Plants*. 2024;38:100526.
20. Esmaili G, Fatemi H, Baghani avval M, Azizi M, Arouiee H, Vaezi J, et al. Diversity of chemical composition and morphological traits of eight Iranian wild *Salvia* species during the first step of domestication. *Agronomy*. 2022;12:2455.
21. Douglas SE. Plastid evolution: origins, diversity, trends. *Curr Opin Genet Dev*. 1998;8:655–61.
22. Brunkard JO, Runkel AM, Zambryski PC. Chloroplasts extend stromules independently and in response to internal redox signals. *Proc Natl Acad Sci*. 2015;112:10044–9.
23. Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol*. 2011;76:273–97.
24. Wang M, Cui L, Feng K, Deng P, Du X, Wan F, et al. Comparative analysis of Asteraceae Chloroplast genomes: structural organization, RNA editing and evolution. *Plant Mol Biol Rep*. 2015;33:1526–38.
25. Dobrogowski J, Adamiec M, Luciński R. The Chloroplast genome: a review. *Acta Physiol Plant*. 2020;42:98.
26. Daniell H, Lin C-S, Yu M, Chang W-J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol*. 2016;17:1–29.
27. Nock CJ, Waters DLE, Edwards MA, Bowen SG, Rice N, Cordeiro GM, et al. Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnol J*. 2011;9:328–33.
28. Yang YC, Kung TL, Hu CY, Lin SF. Development of primer pairs from diverse Chloroplast genomes for use in plant phylogenetic research. *Genet Mol Res*. 2015;14:14857–70.
29. Adem M, Beyene D, Feyissa T. Recent achievements obtained by Chloroplast transformation. *Plant Methods*. 2017;13:1–11.
30. Wu W-G, Dong L-L, Chen S-L. Development direction of molecular breeding of medicinal plants. *Zhongguo Zhong Yao Za zhi = zhongguo Zhongyao zazhi = china. J Chin Mater Med*. 2020;45:2714–9.
31. Soorni A, Borna T, Alemardan A, Chakrabarti M, Hunt AG, Bombarely A. Transcriptome landscape variation in the genus *Thymus*. *Genes (Basel)*. 2019;10:620.
32. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: from gene to genome. *Biol Rev*. 2015;90:157–66.
33. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. *PLoS ONE*. 2011;6:e19254.
34. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci*. 2005;102:8369–74.
35. Kress WJ, Erickson DL. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS ONE*. 2007;2:e508.
36. Chase MW, Cowan RS, Hollingsworth PM, Van Den Berg C, Madrián S, Petersen G, et al. A proposal for a standardised protocol to barcode all land plants. *Taxon*. 2007;56:295–9.
37. 1 CPWG, Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, et al. A DNA barcode for land plants. *Proc Natl Acad Sci*. 2009;106:12794–7.
38. Qian J, Song J, Gao H, Zhu Y, Xu J, Pang X, et al. The complete Chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS ONE*. 2013;8:e57607.
39. Liang C, Wang L, Lei J, Duan B, Ma W, Xiao S, et al. A comparative analysis of the Chloroplast genomes of four *Salvia* medicinal plants. *Engineering*. 2019;5:907–15.
40. Du Q, Yang H, Zeng J, Chen Z, Zhou J, Sun S, et al. Comparative genomics and phylogenetic analysis of the Chloroplast genomes in three medicinal *Salvia* species for bioexploration. *Int J Mol Sci*. 2022;23:12080.
41. Su T, Geng Y-F, Xiang C-L, Zhao F, Wang M, Gu L, et al. Chloroplast genome of *salvia* sect. *Drymosphace*: comparative and phylogenetic analysis. *Diversity*. 2022;14:324.
42. Pei Y, Zhao G, Li X, Yu D, Cui N. The complete Chloroplast genome and phylogenetic analysis of *Salvia oxyphora* Briq. 1896 (Nepetoideae, Lamiaceae). *Mitochondrial DNA Part B*. 2022;7:1342–4.
43. Gao C, Wu C, Zhang Q, Zhao X, Wu M, Chen R, et al. Characterization of Chloroplast genomes from two *Salvia* medicinal plants and gene transfer among their mitochondrial and Chloroplast genomes. *Front Genet*. 2020;11:574962.
44. Yu D, Pei Y, Cui N, Zhao G, Hou M, Chen Y, et al. Comparative and phylogenetic analysis of complete Chloroplast genome sequences of *Salvia* regarding its worldwide distribution. *Sci Rep*. 2023;13:14268.
45. Rose JP, Kriebel R, Kahan L, DiNicola A, González-Gallegos JG, Celep F, et al. Sage insights into the phylogeny of *Salvia*: dealing with sources of discordance within and across genomes. *Front Plant Sci*. 2021;12:767478.
46. Xiao-Ming Z, Junrui W, Li F, Sha L, Hongbo P, Lan Q, et al. Inferring the evolutionary mechanism of the Chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. *Sci Rep*. 2017;7:1555.
47. Turudić A, Liber Z, Grdiša M, Jakše J, Varga F, Šatović Z. Variation in Chloroplast genome size: biological phenomena and technological artifacts. *Plants*. 2023;12:254.
48. Cao M-T, Wu J-J, Wang R-H, Xu L, Qi Z-C, Wei Y-K. The complete Chloroplast genome of Russian Sage *Salvia Yangii* BT drew (Lamiaceae). *Mitochondrial DNA Part B*. 2020;5:2590–1.
49. Zhou Y, Zhang H, Ping H, Ding Y, Hu S, Bi G, et al. Characterization of the complete Chloroplast genome of *Salvia leucantha* (Lamiaceae). *Mitochondrial DNA Part B*. 2021;6:3406–8.
50. Nakai M. The TIC complex uncovered: the alternative view on the molecular mechanism of protein translocation across the inner envelope membrane of chloroplasts. *Biochim Biophys Acta (BBA)-Bioenergetics*. 2015;1847:957–67.
51. Yan K, Lu X, Li W, Sun C, Zhou X, Wang Y. Chloroplast genome diversity and molecular evolution in hypericaceae: new insights from three *Hypericum* species. *Int J Mol Sci*. 2025;26:323.
52. Gu X, Li L, Li S, Shi W, Zhong X, Su Y, et al. Adaptive evolution and co-evolution of Chloroplast genomes in Pteridaceae species occupying different habitats: overlapping residues are always highly mutated. *BMC Plant Biol*. 2023;23:511.
53. Li Q-Q, Zhang Z-P, Aogan, Wen J. Comparative Chloroplast genomes of Argentina species: genome evolution and phylogenomic implications. *Front Plant Sci*. 2024;15:1349358.
54. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
55. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, et al. GetOrganelle: a fast and versatile toolkit for accurate de Novo assembly of organelle genomes. *Genome Biol*. 2020;21:1–31.
56. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de Novo genome assemblies. *Bioinformatics*. 2015;31:3350–2.
57. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res*. 2019;47:W65–73.
58. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45:W6–11.
59. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organelle genomes. *Nucleic Acids Res*. 2019;47:W59–64.
60. Liu S, Ni Y, Li J, Zhang X, Yang H, Chen H, et al. CPGView: A package for visualizing detailed Chloroplast genome structures. *Mol Ecol Resour*. 2023;23:694–704.
61. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30:2725–9.
62. Amirousséfi A, Hyvönen J, Poczai P. Irscope: an online program to visualize the junction sites of Chloroplast genomes. *Bioinformatics*. 2018;34:3030–1.
63. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004;32 Web Server issue:W273–9.
64. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.
65. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972–3.
66. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Osins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*. 2017;34:3299–302.
67. Vaidya G, Lohman DJ, Meier R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*. 2011;27:171–80.

68. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32:268–74.
69. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 2021;49:W293–6.

Publisher's note

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