



Complete Chloroplast genome of *Mentha aquatica* reveals hypervariable regions and resolves phylogenetic position within the genus *Mentha*

Aboozar Soorni¹ · Mohammad Mehdi Golchini¹

Received: 13 May 2025 / Accepted: 30 June 2025
© The Author(s), under exclusive licence to Springer Nature B.V. 2025

Abstract

Background The genus *Mentha* (Lamiaceae) encompasses economically and medicinally important aromatic herbs, yet its taxonomy remains complex due to frequent hybridization, polyploidy, and morphological plasticity. Chloroplast (cp.) genomes emerged as powerful tools for resolving phylogenetic relationships, but no complete cp. genome of *Mentha aquatica* from Iran has been available.

Methods and results In this study, we sequenced and assembled the first complete cp. genome of *M. aquatica* from Iran using Illumina sequencing. The resulting circular genome measured 152,077 bp and exhibited a typical quadripartite structure, consisting of a large single-copy region (83,212 bp), a small single-copy region (17,665 bp), and two inverted repeats (25,600 bp each). Annotation revealed 132 functional genes, including 87 protein-coding genes, 37 tRNAs, and 8 rRNAs. Comparative analyses with other *Mentha* species showed conserved gene content but detected structural variations at IR-LSC/SSC boundaries, particularly in the positioning of *rps19* and *trnN*. Nucleotide diversity (Pi) analysis identified hypervariable regions, with *ycf1* and **rpl2-trnH** displaying the highest polymorphism, suggesting their potential as DNA barcodes. Phylogenetic reconstruction based on complete cp. genomes placed *M. aquatica* in a strongly supported clade with *M. canadensis*, indicating recent divergence, while the broader *Mentha* lineage formed a monophyletic group distinct from related genera. The *ycf1* locus demonstrated high discriminatory power, generating phylogenies consistent with whole-genome analyses, whereas *rpl14* provided limited resolution.

Conclusions This study established a foundational genomic resource for *M. aquatica*, advancing phylogenetic and biogeographic research within *Mentha*, and highlighted the utility of cp. genomes and hypervariable loci for species identification and evolutionary studies in this taxonomically challenging genus.

Keywords Chloroplast DNA (cpDNA) · Molecular markers · *Mentha* species · IR boundary shift · Hotspot regions

Introduction

The genus *Mentha* L. (family Lamiaceae) comprises a diverse group of aromatic herbaceous plants, with approximately 25–30 species and numerous hybrids classified into five taxonomic Sects. [1–4]. These species are distributed worldwide, particularly in temperate and sub-temperate regions [5, 6], and are cultivated on five continents for their

considerable economic and medicinal value. The genus is taxonomically complex, largely due to frequent interspecific hybridization, polyploidy, and widespread morphological plasticity [7], which complicate the identification and classification of *Mentha* taxa, especially for breeding and conservation purposes [1].

Species within *Mentha*, including *Mentha aquatica*, produce essential oils (EOs) rich in monoterpenes such as menthol, carvone, and limonene, which are biosynthesized in glandular trichomes [8]. These compounds are responsible for the distinctive aroma and biological properties of mints and have applications in the pharmaceutical, food, cosmetic, and agricultural industries [5, 9, 10]. Numerous studies have demonstrated the therapeutic potential of *Mentha*

✉ Aboozar Soorni
soorni@iut.ac.ir

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

EOs, reporting their carminative [11], antispasmodic [12], analgesic [13], anti-inflammatory [14], antioxidant [15], antiviral [16], antibacterial [14], antifungal [17], and even anti-SARS-CoV-2 activities [18]. Hassanpouraghdam et al. (2022) studied Iranian ecotypes of *M. aquatica* from the Hyrcanian hotspot, finding major EO constituents such as menthofuran (13.21–52.46%), 1,8-cineole (12.42–25.55%), and (E)-caryophyllene (3.18–15.43%) [19]. Reflecting this demand, the global essential oil market is projected to grow significantly, from \$10.8 billion USD in 2020 to over \$24.7 billion USD by 2030, with mint EO alone expected to exceed \$330 million USD by 2025 [20].

Despite their global significance, efforts to resolve the evolutionary relationships among *Mentha* species have faced significant challenges. Traditional taxonomic approaches relying on morphological, cytological, and chemotypic traits have shown limited resolution due to high variability and overlapping features among species [21–24]. While some progress has been made using molecular markers such as RAPD, ISSR, and gene-based approaches like limonene synthase (*MsLS*) sequencing [25], complex hybridization and maternal inheritance patterns continue to obscure clear phylogenetic delineation [1, 26].

In this context, the cp. genome has emerged as a valuable resource for plant systematics, population genetics, and species identification. Owing to its uniparental inheritance, conserved structure, and relatively low recombination and mutation rates, cpDNA has been widely applied to resolve complex phylogenies, detect cryptic diversity, and develop reliable DNA barcodes. The cp. genome typically exhibits a quadripartite circular structure consisting of a large single-copy (LSC) region, a small single-copy (SSC) region, and two inverted repeats (IRs), and contains genes related to photosynthesis, transcription, and translation. Advancements in next-generation sequencing (NGS) technologies have enabled the rapid sequencing and assembly of complete cp. genomes, facilitating comparative genomic analyses within and across species. In the genus *Mentha*, several cp. genomes have been sequenced in recent years, providing critical insights into phylogenetic relationships and genome structure. For instance, the cp. genome of *Mentha spicata* was sequenced by Wang et al. (2017), revealing its structural organization and aiding in the conservation of this species [27]. Filimban et al. (2022) analyzed the cp. genomes of *M. longifolia* and *Mentha × verticillata*, identifying structural variations and unexpected phylogenetic placements that underscored the need for further research in *Mentha* systematics [28]. Huaizhu et al. (2019) found that *M. canadensis* was closely related to *M. spicata* and *M. longifolia*, based on cp. genome phylogenies [29]. He et al. (2020) showed that *M. haplocalyx* clustered with *M. spicata* and *Dracocephalum*, supporting the monophyly of certain lineages within

Lamiaceae [30]. Moreover, Chen et al. (2012) demonstrated cpDNA diversity and maternal inheritance patterns in Chinese *Mentha* populations, linking chloroplast variation with essential oil chemotypes [6]. A more recent study by Furan et al. (2025) compared five Lamiaceae cp. genomes, including *Mentha × piperita*, to reveal conserved genome structures, codon usage patterns, and selection pressures, thus contributing to our broader understanding of cp. genome evolution in this family [31].

Despite the medicinal, ecological, and industrial importance of *M. aquatica*, no complete cp. genome sequence from Iranian populations has been reported to date. Given the high levels of morphological and chemical variability within the genus and the prevalence of hybridization, region-specific genome data are essential to support identification, conservation, and genetic improvement efforts.

The present study aims to fill this gap by sequencing and analyzing the complete cp. genome of *M. aquatica* collected from Iran. Our objectives include: (i) assembling and annotating the complete cp. genome; (ii) comparing its structural and genetic features with related *Mentha* species; and (iii) evaluating its potential as a DNA barcode for species identification and phylogenetic studies. This work contributes a valuable genomic resource for future studies on *Mentha* taxonomy, evolution, and molecular breeding.

Materials and methods

Plant material collection and DNA extraction

The collection of plant material and the extraction of DNA involved the use of a fresh leaf sample from *M. aquatica*, which is taxonomically unique and sourced from Sari in Mazandaran Province, for genomic analysis. Total genomic DNA, including chloroplast-enriched fractions, was extracted from 100 mg of leaf tissue that had been flash-frozen, utilizing the DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany), in accordance with the manufacturer's instructions. The quality of the DNA was evaluated through two complementary techniques: (1) electrophoresis on a 1% agarose gel to verify the integrity of high-molecular-weight DNA, and (2) UV spectrophotometric analysis (NanoDrop 2000c, Thermo Fisher Scientific, USA), where absorbance ratios of $A_{260}/A_{280} \geq 1.8$ and $A_{260}/A_{230} \geq 2.0$ were deemed acceptable. Subsequently, the high-quality DNA was used for Illumina TruSeq library preparation, targeting insert sizes of 350 bp. Paired-end sequencing (2×150 bp) was conducted on an Illumina HiSeq 2000 platform (Illumina Inc., USA).

Cp genome assembly, annotation, and feature characterization

Raw paired-end reads of 150 base pairs underwent quality assessment through FastQC v0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Subsequently, adapter sequences and low-quality reads were eliminated utilizing Trimmomatic v0.39 [32]. The resultant filtered reads were then assembled into complete cp. genomes with the aid of GetOrganelle v1.7.7.1 [33], and the integrity of the assembly was confirmed using Bandage [34]. For annotation purposes, CPGAVAS2 [35], GeSeq [36], and the Plastid Genome Annotator (PGA) [37] were employed. Circular genome visualizations were produced using OGDRAW [38]. The identification of simple sequence repeats (SSRs) was conducted with MISA [39], applying minimum repeat thresholds of eight for mononucleotides, four for di- and trinucleotides, and three for tetra-, penta-, and hexanucleotides. The Relative Synonymous Codon Usage (RSCU) was computed using MEGA6 [40] and visualized through the RSCU-Plot Shiny application (<https://pcg-lab.shinyapps.io/RSCU-Plot/>). Nucleotide diversity (Π) was evaluated using CPStools [41].

Phylogenetic investigation and assessment of hypervariable markers

In order to deduce the evolutionary connections within the *Mentha* genus, a maximum likelihood (ML) phylogenetic tree was constructed utilizing coding sequences (CDS) from newly sequenced species alongside additional cp. genomes from the Lamiaceae family obtained from NCBI, with *Ajuga bracteosa* serving as the outgroup. Common coding sequences were extracted through CPStools [41] and aligned using MUSCLE v3.8.1551 [42]. The phylogenetic reconstruction was executed in IQ-TREE [43] employing the GTR+ Γ model, with nodal support evaluated through 1000 bootstrap replicates. The resultant phylogenetic tree was visualized with the Interactive Tree of Life [44]. To assess the phylogenetic applicability of potential barcode regions, target loci were extracted from the comparative dataset, aligned, and analyzed utilizing standardized phylogenetic methodologies.

Results

Cp genome assembly and annotation

The sequencing and assembly of the *M. aquatica* cp. genome resulted in a circular DNA molecule of 152,077 base pairs (bp) in length (Fig. 1), exhibiting the typical

quadripartite structure observed in most angiosperms. This structure comprised the LSC region spanning 83,212 bp (position 1–83,212), the SSC region of 17,665 bp (position 108,813–126,477), and a pair of IR (IRa and IRb), each measuring 25,600 bp (IRb: 83,213–108,812; IRa: 126,478–152,077). Genome annotation of identified a total of 132 genes, comprising protein-coding genes, transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), and hypothetical open reading frames (ORFs). These genes were functionally categorized based on their involvement in essential chloroplast processes such as photosynthesis, transcription, translation, and metabolic regulation. The cp. genome harbored five genes related to Photosystem I (Table 1), and 15 genes associated with Photosystem II. The ATP synthase complex was encoded by six genes, while 12 genes (*ndhA* to *ndhK*) encoded subunits of the NADH dehydrogenase complex, some of which were duplicated or contained introns. Six genes were assigned to the cytochrome b6/f complex, and the large subunit of RuBisCO was encoded by a single gene (*rbcL*). The cp. genome also included a complete set of ribosomal protein genes, with 14 small subunit genes and 11 large subunit genes, several of which had duplicated copies or introns. The chloroplast-encoded RNA polymerase subunits were represented by four genes. A full complement of ribosomal RNA genes was identified, including two copies each of *rrn16S*, *rrn23S*, *rrn4.5 S*, and *rrn5S*, totaling eight rRNA genes. The genome also encoded 37 tRNA genes, collectively recognizing all 20 amino acids. Several tRNAs occurred in multiple copies or contained introns. In addition, single-copy genes with specialized functions were annotated, including *accD* (acetyl-CoA carboxylase), *ccsA* (cytochrome c synthesis), *cemA* (envelope membrane protein), *clpP* (protease), and *matK* (maturase). A total of seven hypothetical ORFs (*yef1*, *yef2*, *yef3*, *yef4*, *yef15*) were also detected, with some present in multiple copies or containing introns. Among annotated genes, several protein-coding and ribosomal genes in the contained introns, including *atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rpoC1*, *rps12*, *rps16*, and *clpP*. In addition, six tRNA genes, *trnK-UUU*, *trnG-UCC*, *trnL-UAA*, *trnV-UAC*, *trnI-GAU*, and *trnA-UGC*, were also found to contain introns.

Genomic features

A comprehensive analysis of SSRs in the *M. aquatica* cp. genome identified 26 SSRs, revealing distinct patterns of distribution and variation. Mononucleotide repeats accounted for the majority of SSR (84.6%, $n=22$), with poly(T) and poly(A) tracts ranging from 10 to 16 bp in length. Notably, the longest homopolymer was a (T)₁₆ repeat. Dinucleotide repeats (11.5%, $n=3$) were exclusively composed of AT/TA motifs, including one (TA)₇ repeat. Compound SSRs (7.7%,

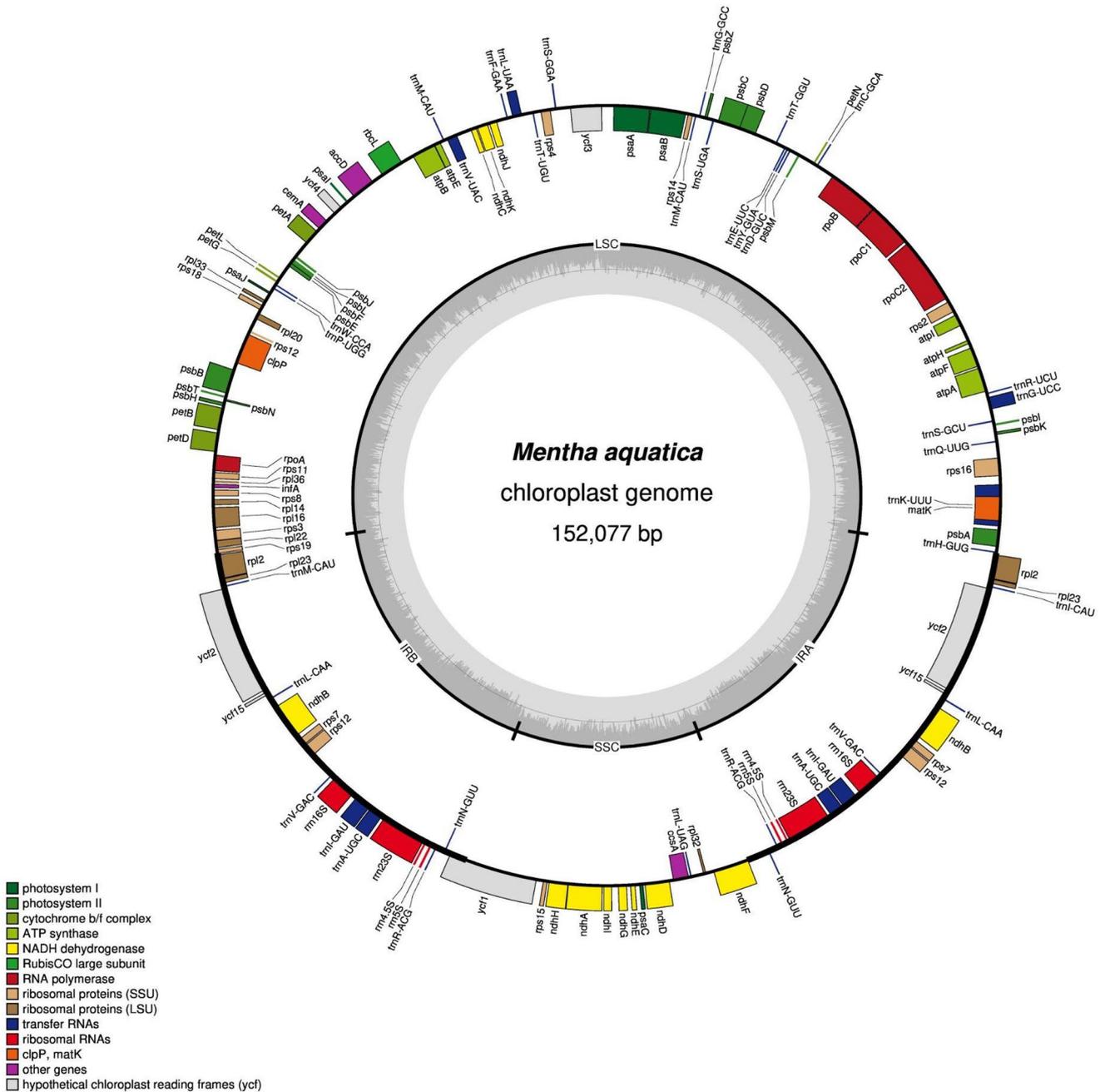


Fig. 1 Circular map of *M. aquatica* cp. genome. The diagram, generated using OGDRAW, highlights key genomic features

$n=2$) were less frequent, with the largest being a 43-bp complex motif featuring interrupted (TA)₆ and (AT)₆ repeats.

The examination of RSCU in *M. aquatica* revealed a notable codon bias, characterized by distinct preferences for certain synonymous codons associated with various amino acids (Fig. 2). Codons ending in A/T were predominantly preferred, as demonstrated by elevated RSCU values for UUU (Phe, 1.27), AUU (Ile, 1.34), AAU (Asn, 1.33), AAA (Lys, 1.32), GAU (Asp, 1.40), and GAA (Glu, 1.41), while their G/C-ending counterparts showed diminished usage

(UUC: 0.73; AUC: 0.70; AAC: 0.67; AAG: 0.68; GAC: 0.60; GAG: 0.59). This pattern likely reflects the influence of mutational or translational selection pressures favoring A/T-rich sequences, consistent with overall genomic composition observed in many plant species. Notably, exceptions were identified, such as the strong preference for CGA (Arg, 1.08) and AGA (Arg, 1.76) compared to other Arg codons (CGC: 0.43; CGG: 0.78), along with moderate usage of GGU (Gly, 1.03) and GGA (Gly, 1.29) despite their G-ending nature, indicating the presence of additional

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

Group of genes	Gene names	Number of genes
Photosystem I	<i>psaB, psaA, psaI, psaJ, psaC</i>	5
Photosystem II	<i>psbA, psbK, psbI, psbM, psbD, psbC, psbZ, psbJ, psbL, psbF, psbE, psbB, psbT, psbN, psbH</i>	15
Subunits of ATP synthase	<i>atpA, atpF#, atpH, atpI, atpE, atpB</i>	6
NADH dehydrogenase	<i>ndhJ, ndhK, ndhC, ndhB(2)#, ndhH, ndhA#, ndhI, ndhG, ndhE, ndhD, ndhF</i>	12
Cytochrome b/f complex	<i>petN, petA, petL, petG, petB#, petD#</i>	6
Large subunit of rubisco	<i>rbcL</i>	1
Small ribosomal subunits	<i>rps12(2)##, rps16#, rps2, rps14, rps4, rps18, rps11, rps8, rps3, rps19, rps7(2), rps15</i>	14
Large ribosomal subunits	<i>rpl33, rpl20, rpl36, rpl14, rpl16#, rpl22, rpl2(2)#, rpl23(2), rpl32</i>	11
DNA-dependent RNA polymerase	<i>rpoC2, rpoC1#, rpoB, rpoA</i>	4
Ribosomal RNA	<i>rrn16S(2), rrn23S(2), rrn4.5 S(2), rrn5S(2)</i>	8
Transfer RNA	<i>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, trnM-CAU(3), trnS-GGA, trnT-UGU, trnL-UAA#, trnF-GAA, trnV-UAC#, trnW-CCA, trnP-UGG, trnL-CAA(2), trnV-GAC(2), trnI-GAU(2)#, trnA-UGC(2)#, trnR-ACG(2), trnN-GUU(2), trnL-UAG, trnI-CAU</i>	37
Acetyl-CoA carboxylase	<i>accD</i>	1
c-type cytochrome synthesis	<i>ccsA</i>	1
Envelope membrane protein	<i>cemA</i>	1
Protease	<i>clpP##</i>	1
Maturase	<i>matK</i>	1
Hypothetical ORFs	<i>ycf3##, ycf4, ycf2(2), ycf15(2), ycf1</i>	7
Total number of genes		132

#: Intron number,
(n): Gene copy number

selective constraints. The usage of stop codons was nearly balanced (UAA: 1.00; UAG: 1.11; UGA: 0.90), while single-codon amino acids (Met: AUG, 1.00; Trp: UGG, 1.00) exhibited neutral RSCU values, as anticipated.

Analysis of nucleotide diversity across coding genes and IGS of *M. aquatica* cp. genome revealed substantial heterogeneity in sequence variation, with intergenic regions exhibiting significantly higher Pi values compared to genic regions (Fig. 3). The most polymorphic IGS was *rpl2_2-trnH-GUG* (Pi=0.02901), followed by *psbI-trnS-GCU* (0.01726) and *petL-petG* (0.01709), suggesting potential hotspots for evolutionary divergence. Among coding regions, *trnA-UGC* (Pi=0.00685) and *ycf1* (0.00598) showed the highest diversity, while ribosomal protein genes (*rpl14*=0.00452, *rps3*=0.00385) and photosystem components (*psbK*=0.00448, *ndhF*=0.00396) displayed intermediate variability. In contrast, essential housekeeping genes (*rpoB*=0.00176, *rbcL*=0.00108, *psbA*=0.00084) were highly conserved, reflecting strong functional constraints. The *ndh* gene cluster exhibited uniformly low diversity (*ndhB*=0.00065, *ndhG*=0.00084), consistent with purifying selection acting on genes involved in photosynthetic electron transport chain.

Junction structure and cis-splicing of *M. aquatica* cp. genome

The analysis of *M. aquatica* cp. genome revealed specific patterns of gene arrangement at the boundaries of the four major regions (LSC, IRb, SSC, and IRa). The *rps19* gene was located across the LSC/IRb junction, with 236 bp situated in the LSC region and 43 bp extending into IRb, a configuration shared with most *Mentha* species examined, except *M. canadensis*, in which *rps19* entirely located within the LSC (Fig. 4). The *trnN* gene was present in two copies, each positioned at the SSC/IR boundaries, with a shift of 1,392 bp into both IRa and IRb, a structure to the arrangement observed in *M. villosa*. In contrast, other species displayed only a single copy of *trnN* located within the IRa region. The *rpl2* gene was duplicated, with both copies partially spanning the LSC/IRb and IRa/LSC junctions, exhibiting a 103 bp extension into both IRb and IRa. This structure was conserved among most species, except *M. canadensis* and *M. longifolia*, which possessed only a single copy within IRb. Additionally, *ycf1* and *ndhF* were found to be shifted across the IRb/SSC and SSC/IRa junctions, respectively, indicating minor expansions and contractions at these boundaries in *M. aquatica*.

The cp. genome analysis identified several genes that demonstrated cis-splicing, characterized by non-contiguous exon arrangements (Fig. 5). These included *rps16* (4800–5026, 5903–5942), *rpoC1* (11757–12166, 12861–13005), *ycf3* (20768–22392, 23150–23579), *clpP* (41919–42071, 42801–43028, 43732–43860), *petB* (69303–69528, 70146–70439, 71165–71235), *petD* (74156–74161, 74890–75531), *rpl2* (dual loci: 80355–80753/81631–81639

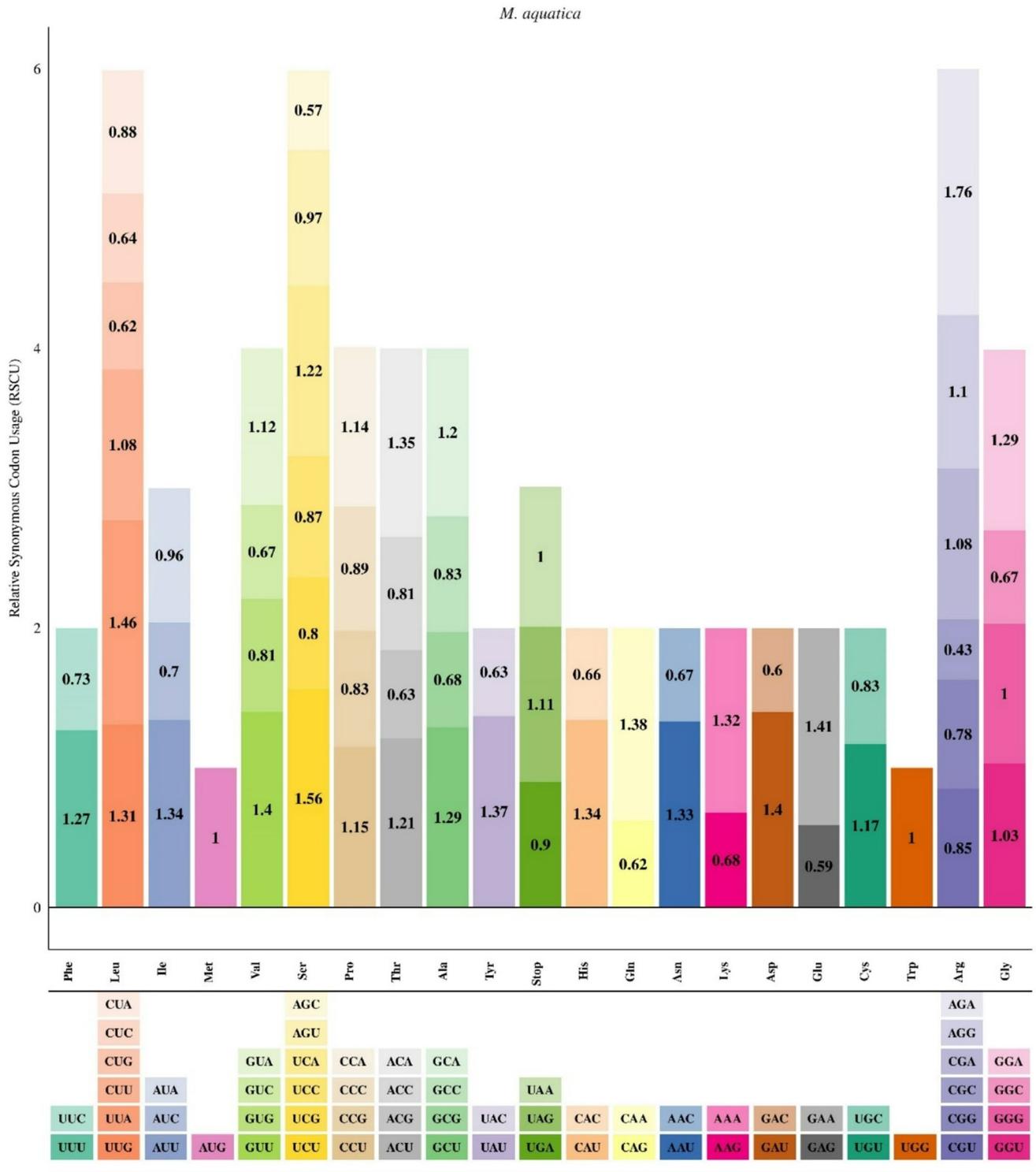


Fig. 2 Relative Synonymous Codon Usage (RSCU) analysis of *M. aquatica* chloroplast genes. The bar graph displays codon preference patterns across all protein-coding sequences

and 139615–140389/141065–141822), *ndhB* (83316–83749/84415–84805 and 115263–115815/116828–117366), and *ndhA* (93468–94225, 94901–95675). The splicing patterns, particularly the complex multi-exon configurations

observed in *clpP* and *rpl2*, suggest the present of sophisticated post-transcriptional regulatory mechanisms that are essential for proper gene function. The affected genes encoded critical components of photosynthesis (*petB*,

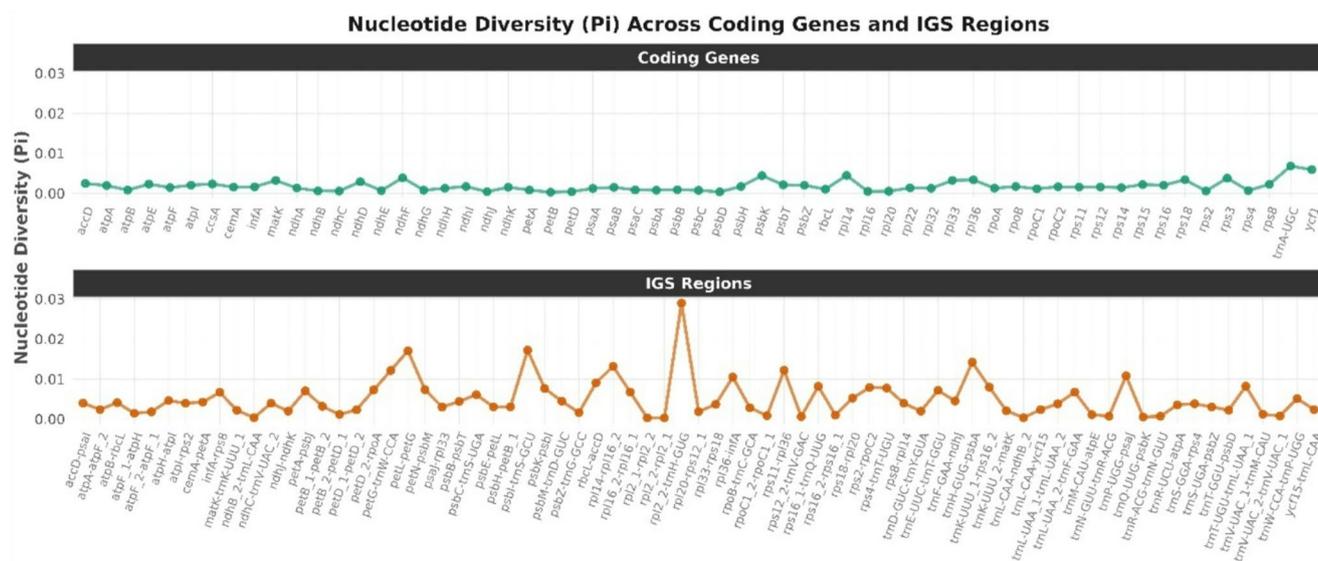


Fig. 3 Comparative analysis of nucleotide diversity (Pi) in coding genes and intergenic spacer (IGS) regions. Points represent individual genomic regions, with colors distinguishing coding genes (blue) from IGS regions (orange). The y-axis displays Pi values

petD), transcription (*rpoC1*), and ribosomal proteins (*rps16*, *rpl2*), underscoring the evolutionary conservation of *cis*-splicing in maintaining chloroplast proteostasis. Although *atpF* exhibited partial positional data, its inclusion indicated potential splicing activity requiring experimental validation.

Phylogenetic analysis

Phylogenetic reconstruction using complete cp. genomes placed *M. aquatica* within a strongly supported clade (BS=100) alongside *M. canadensis*, with a notably short branch length, indicating recent divergence (Fig. 6). This *M. aquatica*-*M. canadensis* lineage formed a sister group to a larger *Mentha* radiation comprising *M. longifolia* and *M. spicata*-*M. verticillata*-*M. piperita* complex, the latter exhibiting minimal cp. genome divergence consistent with known hybridization events. The genus *Mentha* as a whole was monophyletic and distinct from related genera (*Thymus*, *Origanum*, *Salvia*), though *Thymus serpyllum* appeared nested within *Mentha* with weak support, potentially reflecting incomplete lineage sorting or plastid capture. Topological congruence with prior studies was observed in the clear separation of *Mentha* from the *Salvia/Lavandula* and *Origanum/Thymus* lineages, reinforcing the utility of cp. genomes for resolving deep Lamiaceae relationships.

To evaluate the utility of two hypervariable chloroplast regions (*ycf1* and *rpl14*) as phylogenetic markers, phylogenetic trees were constructed based on these loci across selected species of the Lamiaceae family and compared with the tree derived from complete chloroplast genome sequences (Fig. 7). The tree based on the *ycf1* locus showed strong concordance with the whole cp. genome phylogeny,

accurately resolving *M. aquatica* and *M. canadensis* as sister taxa with minimal genetic divergence. Moreover, *ycf1* effectively distinguished the hybrid species *M. piperita* from its parental taxa, offering robust support for the relationships inferred from complete cp. genomes. In contrast, the *rpl14* locus failed to resolve intra-genus relationships within *Mentha*, demonstrating limited phylogenetic resolution and utility at the species level. These findings highlight the *ycf1* region as a reliable alternative to complete cp. genome analysis for resolving recent divergences and hybrid origins within *Mentha*, whereas the highly conserved nature of *rpl14* restricts its applicability to broader, genus-level classifications.

Discussion

The complete cp. genome of *M. aquatica* was successfully sequenced and annotated, providing valuable insights into the genomic architecture, phylogenetic relationships, and potential DNA barcode regions within the genus *Mentha*. As the first complete cp. genome reported for *M. aquatica*, this study fills a notable gap in genomic resources for this medicinally and economically important species. The cp. genome exhibited the typical quadripartite structure with a total length of 152,077 bp, comprising a LSC region (83,212 bp), a SSC region (17,665 bp), and two IR regions (25,600 bp each). These dimensions are consistent with other *Mentha* cp. genomes, including *M. spicata* (152,132 bp) [27], *M. longifolia* (152,078 bp) [28], and *M. canadensis* (152,154 bp) [29], reinforcing the high conservation of cp. genome structure within the genus.

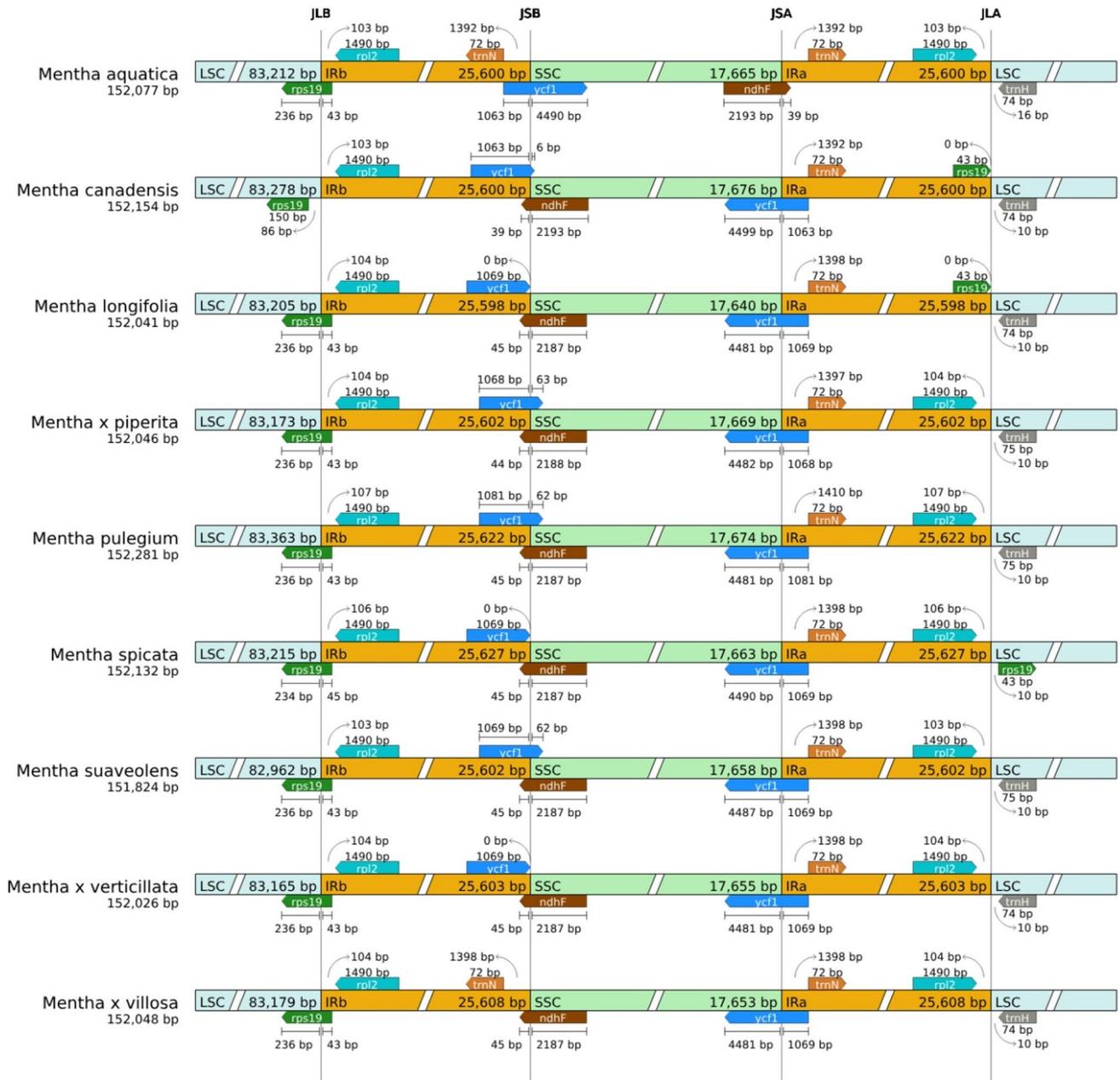


Fig. 4 Comparative analysis of cp. genome architecture across nine *Mentha* species, showing length variations in LSC, SSC, and IR regions along with junction positions (JLB, JSB, JSA, JLA) at IR-LSC/SSC boundaries

The complete cp. genome of *Mentha aquatica* sequenced in this study comprises 132 functional genes, including 87 protein-coding genes, 37 tRNAs, and 8 rRNAs, a genomic structure broadly conserved across the *Mentha* genus yet exhibiting subtle interspecific variations that shed light on evolutionary dynamics. Comparative analyses with other *Mentha* species reveal that *M. spicata* shares an identical gene complement [27], whereas *M. longifolia* and *M. canadensis* harbor one and two additional protein-coding genes, respectively [28, 29]. These modest differences in

gene content are attributable to a combination of biological and technical factors. First, lineage-specific gene loss or pseudogenization, particularly among the *ndh* gene family, has been frequently reported in Lamiaceae and contributes to interspecies variability [45]. Second, expansions or contractions of IR regions can alter gene content by duplicating or truncating genes at the IR boundaries, such as *rps19* and *ycf1* [46]. Third, hybrid speciation events common in *Mentha*, such as those giving rise to *Mentha* × *piperita*, may lead to gene content reduction through genomic rearrangements

[31]. In addition to gene presence/absence variation, structural differences at the intron level, such as those observed in the *clpP* and *ycf3* genes across species like *Salvia* and *Teucrium*, may also reflect underlying evolutionary pressures or annotation inconsistencies [47, 48]. Although not as extensively characterized in *Mentha*, such polymorphisms in intron size or splicing patterns may influence post-transcriptional regulation and merit further investigation. Furthermore, discrepancies in gene annotation pipelines, particularly regarding the classification and retention of hypothetical chloroplast open reading frames (*ycf* genes), may partly explain inconsistencies in reported gene counts across studies [49, 50].

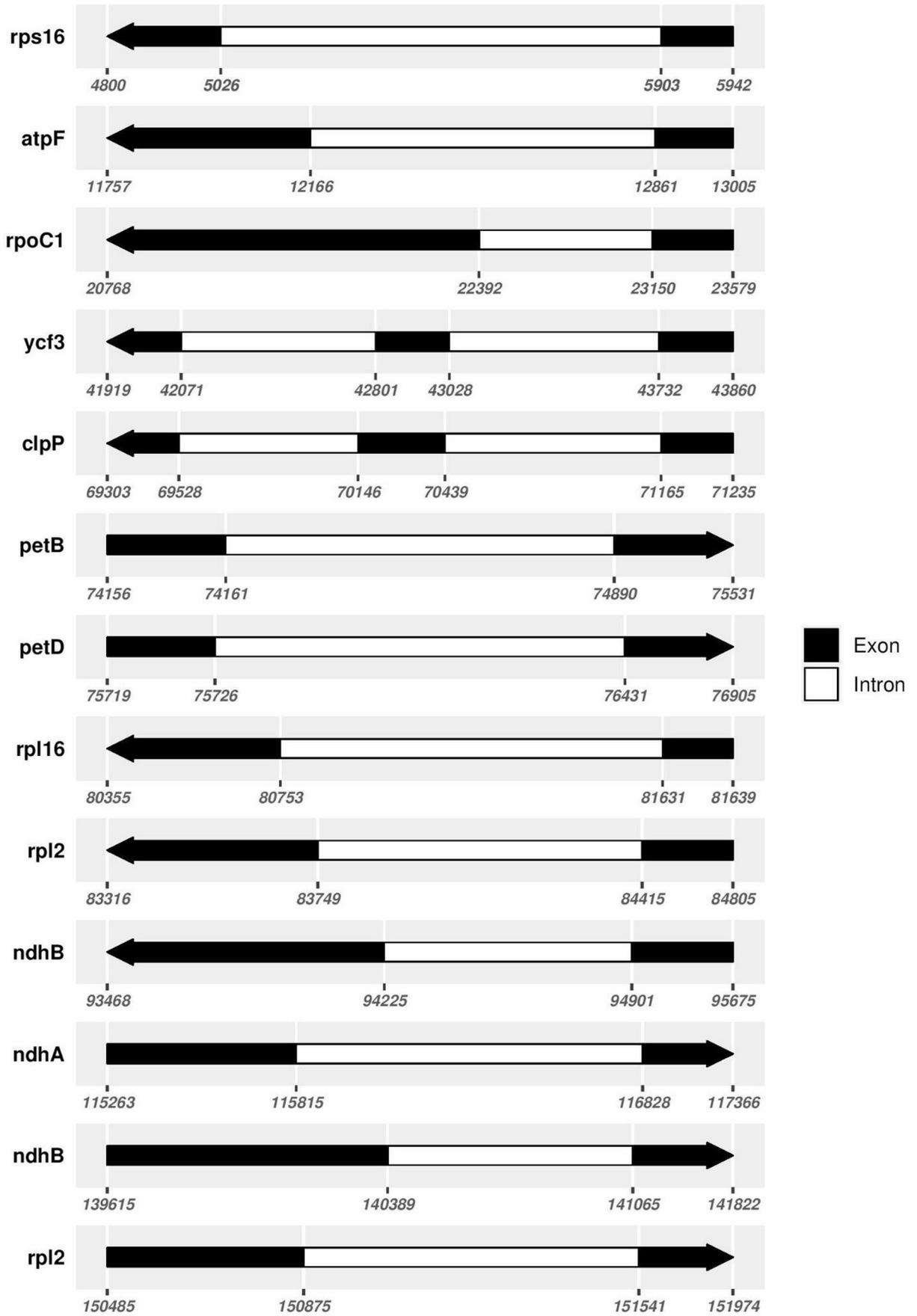
The identification of 26 SSRs in the *M. aquatica* cp. genome has significant implications for molecular marker development and genetic diversity studies within the genus *Mentha*. These chloroplast-derived SSRs represent valuable genetic resources for several applications due to their unique characteristics and maternal inheritance pattern [51, 52]. The AT/TA-rich dinucleotide repeats and the extensive poly(T) and poly(A) tracts identified in this study can be developed into PCR-based molecular markers for taxonomic investigation and authentication of *Mentha* species. This is particularly valuable in a genus characterized by frequent hybridization, polyploidy, and morphological plasticity, where traditional identification methods often prove insufficient. Furthermore, these markers could facilitate population genetic studies, assessment of genetic erosion in wild populations, and support conservation strategies for *Mentha* species [53, 54]. Compared to other marker systems, chloroplast SSRs typically demonstrate high transferability across closely related species, allowing for comparative studies across the entire *Mentha* genus. When used in combination with nuclear markers, these chloroplast SSRs could provide a more comprehensive understanding of genetic diversity patterns and evolutionary processes within this economically and medicinally important genus.

The junctional variations, widely recognized as indicators of structural rearrangements and divergence in angiosperms [55–57], demonstrate both conserved features and species-specific adaptations across *Mentha*. A notable conserved trait is the positioning of the *rps19* gene across the LSC/IRb boundary in *M. aquatica*, where 236 bp lie in the LSC and 43 bp extend into IRb, a configuration shared by most *Mentha* species. However, *M. canadensis* deviates from this pattern, with *rps19* entirely contained within the LSC, suggesting either a lineage-specific contraction of the IR region or a recent boundary shift during evolutionary divergence, similar to patterns observed in other angiosperm lineages [56, 58]. In *M. longifolia*, *rps19* partially extends into IRb (43 bp), reflecting subtle but evolutionarily significant structural plasticity. These variations align with

trends seen in other Lamiaceae genera, such as *Teucrium*, where the *rps3* gene exhibits variable positioning (21–38 bp shifts) across species [59], and *Salvia*, where IR dynamics correlate with ecological adaptations [60]. The presence of two *trnN* gene copies at the SSC/IR boundaries in *M. aquatica*, each extending 1,392 bp into IRa and IRb, mirrors the arrangement in *M. villosa*, implying either a shared ancestral IR expansion event or parallel evolution in these species. In contrast, other *Mentha* species retain only a single *trnN* copy, likely indicating IR contraction events, a phenomenon previously documented in Lamiaceae and other plant families [47, 61, 62]. Such expansions and contractions at the SSC/IR boundary are common in plant cp. genomes and serve as hotspots for structural evolution [63]. The duplicated *rpl2* gene, partially spanning the LSC/IRb and IRa/LSC junctions in *M. aquatica*, further underscores the prevalence of IR-associated duplication events. The 103 bp overlap of *rpl2* into both IR regions represents a structural pattern conserved among most *Mentha* species, while its absence in *M. canadensis* and *M. longifolia* (which retain only one *rpl2* copy in IRb) suggests species-specific IR boundary shifts or deletion events [64]. Further evidence of IR flexibility comes from the partial duplication of *ycf1* and the shifting of *ndhF* across the IRb/SSC and SSC/IRa junctions in *M. aquatica*, indicating minor IR expansion and SSC contraction. The partial duplication of *ycf1* within the IRs is a common structural feature in angiosperms and often contributes to IR size variation [65], while the shifting of *ndhF* reflects dynamic SSC boundaries influenced by selection pressures or genome stabilization mechanisms [66]. Notably, *Mentha* species exhibit relatively stable positioning of *ndhF* and *ycf1* compared to genera like *Salvia*, where these genes frequently span IR-SSC boundaries [46]. This difference may reflect genus-specific evolutionary rates or selective pressures acting on IR stability. The conservation of certain IR features (e.g., *rps19* extension in *M. longifolia*) alongside species-specific variations suggests a balance between functional constraints and adaptive potential in IR boundary configurations. The moderate IR plasticity observed in *Mentha* parallels patterns in *Salvia*, where conserved IR lengths coexist with subtle junctional changes [45], reinforcing the idea that IR boundary dynamics are a common driver of structural evolution in Lamiaceae.

Comparative analysis of nucleotide diversity across cp. genome of *Mentha* species revealed distinct patterns of sequence variation, highlighting the potential of specific genomic regions for DNA barcoding and phylogenetic inference. As observed in this study, intergenic spacers demonstrated significantly higher nucleotide diversity than coding regions, a trend that corroborates earlier reports across the Lamiaceae family. Among coding regions, *ycf1* and *trnA-UGC* exhibited the highest diversity, consistent with

Cis-splicing Genes



◀ **Fig. 5** Cis-splicing gene organization in cp. genome of *M. aquatica*

findings from comparative plastome studies in Lamiaceae that identify *ycf1* as the most rapidly evolving chloroplast gene [60, 67, 68].

The *ycf1* gene is consistently recognized as one of the most hypervariable regions across a wide range of plant taxa [65]. This pronounced variability, however, coexists with its critical role in chloroplast function, presenting an apparent paradox. This phenomenon may be attributed to a molecular mosaic pattern within *ycf1*, wherein specific domains are highly conserved to maintain functional integrity, while others display marked sequence divergence. Previous studies have demonstrated that *ycf1* is subject to heterogeneous selection pressures across plant lineages, with signatures of positive selection reported in *Astragalus* species [69], *Krascheninnikovia ceratoides* [70], and Coelogyninae orchids [71], suggesting a potential role in environmental adaptation. Despite its high sequence variability, *ycf1* retains its essential function as a component of the translocon at the inner chloroplast membrane (TIC) complex, which facilitates protein import, indicating the presence of evolutionary constraints that balance functional necessity with adaptive flexibility. Additionally, lineage-specific deletions of *ycf1* at plastome boundaries, such as those observed in *Lagerstroemia* species [72], further underscore its dynamic evolutionary trajectory.

Although the absolute Pi values of these coding genes were lower than those of the most variable intergenic spacers, their informativeness in phylogenetic reconstruction is underscored by their alignment consistency and phylogenetic signal. By contrast, canonical barcoding genes such as *rbcL* and *psbA* showed very low sequence variation, reaffirming their limited utility for resolving relationships at lower taxonomic levels due to functional conservation associated with core photosynthetic processes. The phylogenetic performance of *ycf1* was particularly notable; this gene produced tree topologies congruent with those derived from whole plastomes, confirming its high discriminatory power at the species level. Previous research has similarly advocated for *ycf1* as a universal plant barcode due to its extensive length, high substitution rate, and low saturation [65].

Conclusion

This study represents the first comprehensive characterization of the complete cp. genome of *M. aquatica*, providing critical insights into the genomic architecture and evolutionary relationships within this economically and medicinally important genus. Through comparative analysis with other

Mentha species, we demonstrated that while the overall quadripartite structure and gene content of the *M. aquatica* plastome (152,077 bp) remains highly conserved, significant structural variations occur at IR-LSC/SSC boundaries, particularly in the positioning of *rps19* and *trnN* genes. Our nucleotide diversity analysis revealed *ycf1* and *rpl2-trnH* as the most polymorphic regions, with *ycf1* emerging as a particularly robust molecular marker for species-level discrimination in *Mentha*, outperforming conventional barcoding loci. Phylogenetic reconstruction robustly resolved *M. aquatica* and *M. canadensis* as sister taxa, supporting their recent divergence, while simultaneously clarifying broader relationships within the *Mentha* clade. The identification of dynamic IR expansions/contractions and conserved cis-splicing mechanisms in genes such as *clpP* and *rpl2* provides new insights into cp. genome evolution in Lamiaceae. Beyond its taxonomic implications, this work establishes valuable genomic resources for future studies of hybridization patterns, biogeographic history, and potential correlations between plastome variation and essential oil chemotypes in *Mentha*. The markers and analytical approaches developed here will facilitate more accurate species identification, conservation planning, and breeding programs for this taxonomically complex genus. These findings not only advance our understanding of *Mentha* evolution but also demonstrate the continued value of cp. genomics in resolving phylogenetic relationships and uncovering patterns of diversification in plants with complex evolutionary histories. Future research integrating nuclear genomic data with the chloroplast resources presented here will be particularly valuable for elucidating the full extent of reticulate evolution in this ecologically and economically significant group.

Tree scale: 0.01

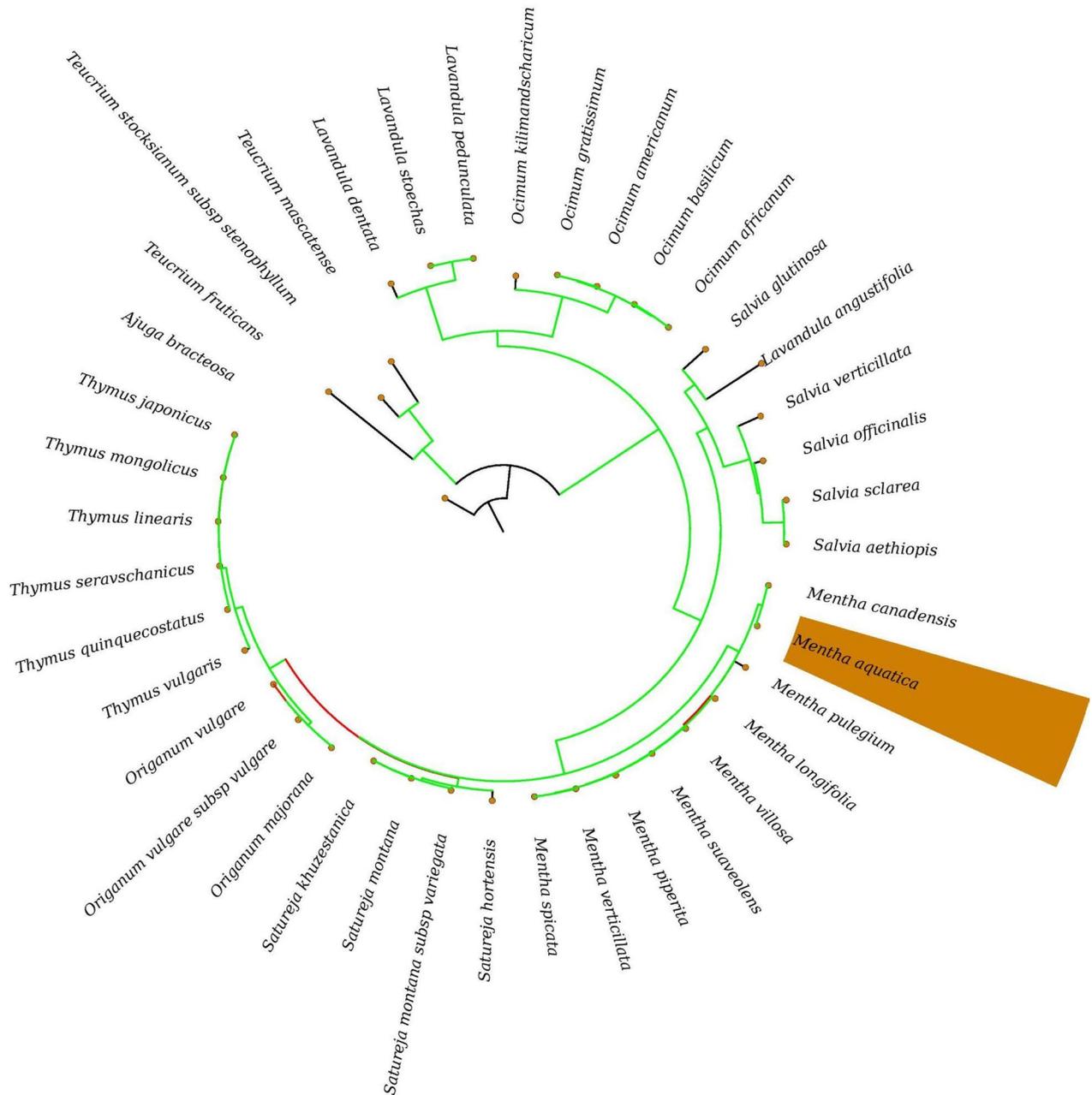


Fig. 6 Maximum likelihood phylogenetic of 40 Lamiaceae cp. genomes reconstructed using IQ-TREE (GTR+ Γ model). The analysis included representative species from nine genera (*Salvia*, *Thymus*, *Mentha*, *Lavandula*, *Origanum*, *Ocimum*, *Satureja*, *Teucrium*) with *A.*

bracteosa as the outgroup. The highlighted clade shows the phylogenetic placement of *M. aquatica* (this study) within its genus context. Node support values were calculated from 100 bootstrap replicates, with values $\geq 70\%$ shown at key branches

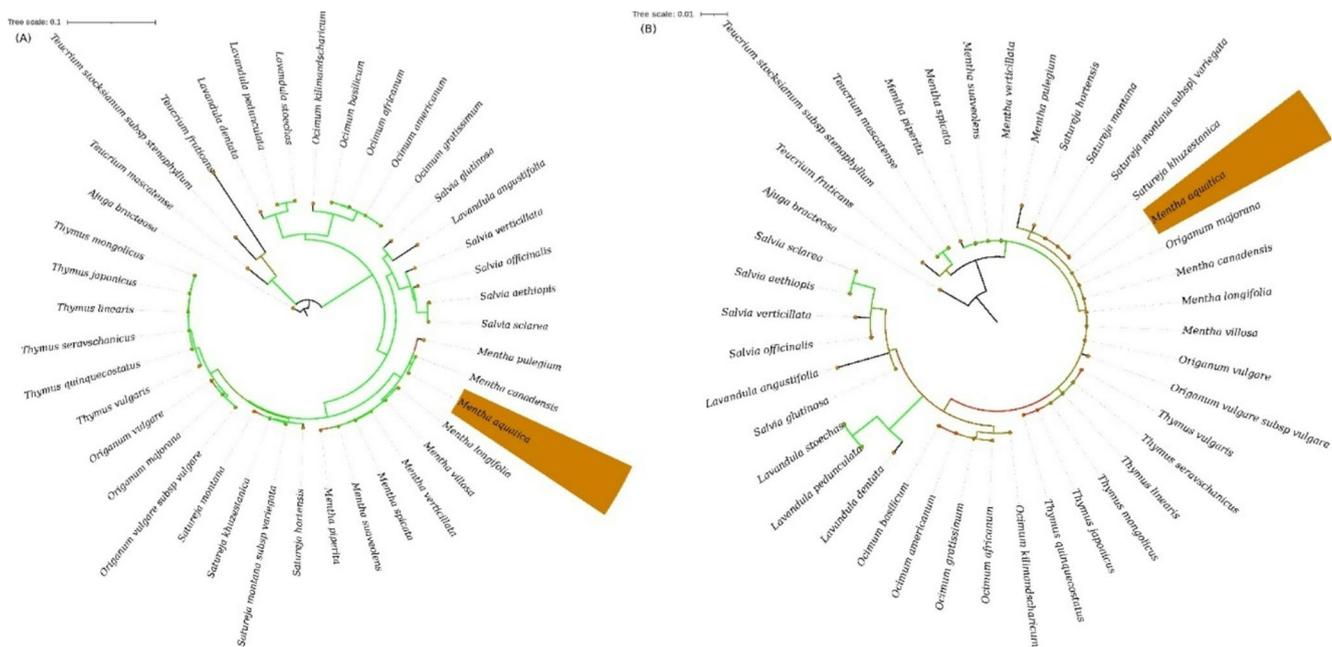


Fig. 7 Comparative chloroplast phylogenies of 40 Lamiaceae species based on hypervariable regions (a) *ycf1* and (b) *rpl14*, reconstructed using maximum likelihood in IQ-TREE (GTR+ Γ model). Analyses

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11033-025-10789-5>.

Acknowledgements Not applicable.

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

Funding No funding was received for conducting this study.

Data availability The assembled and annotated genome is accessible in NCBI database under the research accession PV600640.

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.

Conflict of interest The authors declare no conflicts of interest.

included representatives from nine genera (*Salvia*, *Thymus*, *Mentha*, *Lavandula*, *Origanum*, *Ocimum*, *Satureja*, *Teucrium*) with *A. bracteosa* as the outgroup

References

- Jabeen A, Guo B, Abbasi BH et al (2012) Phylogenetics of selected *Mentha* species on the basis of *rps8*, *rps11* and *rps14* Chloroplast genes. *J Med Plants Res* 6:30–36
- Tutin TG (1972) *Flora Europaea*. Diapensiaceae to Myoporaceae
- Harley RM, Brighton CA (1977) Chromosome numbers in the genus *Mentha* L. *Bot J Linn Soc* 74:71–96
- Hassan M, Badr A, Mustafa A et al (2003) Genetic diversity among *Mentha* populations in Egypt as reflected by morphological and protein electrophoretic variations. *J Agri Food Chem* 2:269–286
- Al-Rawashdeh IM (2011) Molecular taxonomy among *Mentha* *spicata*, *Mentha* *longifolia* and *Ziziphora* *tenuior* populations using the RAPD technique. *Jordan J Biol Sci* 4:63–70
- Chen X, Zhang F, Yao L (2012) Chloroplast DNA molecular characterization and leaf volatiles analysis of mint (*Mentha*; Lamiaceae) populations in China. *Ind Crops Prod* 37:270–274
- Šarić-Kundalić B, Fialová S, Dobeš C et al (2009) Multivariate numerical taxonomy of *Mentha* species, hybrids, varieties and cultivars. *Sci Pharm* 77:851–876
- Croteau R, Gershenzon J (1994) Genetic control of monoterpene biosynthesis in mints (*Mentha*: Lamiaceae). Genetic engineering of plant secondary metabolism. Springer, pp 193–229
- Eftekhari A, Khusro A, Ahmadian E et al (2021) Phytochemical and nutra-pharmaceutical attributes of *Mentha* spp.: A comprehensive review. *Arab J Chem* 14:103106
- Lawrence BM (2006) *Mint: the genus Mentha*. CRC
- May B, Köhler S, Schneider B (2000) Efficacy and tolerability of a fixed combination of peppermint oil and Caraway oil in patients suffering from functional dyspepsia. *Aliment Pharmacol Ther* 14:1671–1677
- Heghes SC, Vostinaru O, Rus LM et al (2019) Antispasmodic effect of essential oils and their constituents: A review. *Molecules* 24:1675

13. Yousuf PMH, Noba NY, Shohel M et al (2013) Analgesic, anti-inflammatory and antipyretic effect of *Mentha spicata* (Spearmint). *Br J Pharm Res* 3:854–864
14. Xia X, Lin Z, Shao K et al (2021) Combination of white tea and peppermint demonstrated synergistic antibacterial and anti-inflammatory activities. *J Sci Food Agric* 101:2500–2510
15. Ed-Dra A, Filali FR, Presti V, Lo et al (2020) Chemical composition, antioxidant capacity and antibacterial action of five Moroccan essential oils against *Listeria monocytogenes* and different serotypes of *Salmonella enterica*. *Microb Pathog* 149:104510
16. Minami M, Kita M, Nakaya T et al (2003) The inhibitory effect of essential oils on herpes simplex virus type-1 replication in vitro. *Microbiol Immunol* 47:681–684
17. Piras A, Porcedda S, Falconieri D et al (2021) Antifungal activity of essential oil from *Mentha spicata* L. and *Mentha pulegium* L. growing wild in Sardinia Island (Italy). *Nat Prod Res* 35:993–999
18. Jan J-T, Cheng T-JR, Juang Y-P et al (2021) Identification of existing pharmaceuticals and herbal medicines as inhibitors of SARS-CoV-2 infection. *Proc Natl Acad Sci* 118:e2021579118
19. Hassanpouraghdam MB, Mohammadzadeh A, Morshedloo MR et al (2022) *Mentha aquatica* L. populations from the hyrcanian hotspot: volatile oil profiles and morphological diversity. *Agronomy* 12:2277
20. Chauhan S, Deshmukh R (2022) Essential oils market by type (Orange, Eucalyptus, Cornmint, Peppermint, Citronella, Lemon, Lime, Clove, Spearmint, Others), by application (Food and beverages, medical, cleaning and home, spa and relaxation, others), by distribution channel (Direct distr. Allied Mark Res Available online www.alliedmarketresearch.com/essential-oils-market (accessed March 4, 2022)
21. Harley RM (1963) Taxonomic studies on the genus *Mentha*. D Phil thesis, Oxford Univ
22. Lawrence BM (1978) A study of the monoterpene interrelationships in the genus *Mentha* with special reference to the origin of pulegone and menthofuran
23. Celenk S, Tarimcilar G, Bicakci A et al (2008) A palynological study of the genus *Mentha* L. (Lamiaceae). *Bot J Linn Soc* 157:141–154
24. Liang ChengYuan LC, Li WeiLin LW, Zhou Yong ZY et al (2011) Study on morphological diversity of *Mentha haplocalyx* Briq
25. Wang HT, Yu X, Liu Y et al (2013) Analysis of genetic variability and relationships among *Mentha* L. using the limonene synthase gene. *LS Gene* 524:246–252
26. Clegg MT (1993) Chloroplast gene sequences and the study of plant evolution. *Proc Natl Acad Sci* 90:363–367
27. Wang K, Li L, Hua Y et al (2017) The complete Chloroplast genome of *Mentha spicata*, an endangered species native to South Europe. *Mitochondrial DNA Part B Resour* 2:907–909. <https://doi.org/10.1080/23802359.2017.1413311>
28. Zubair Filimban F, Yarádua SS, Bello A, Choudhry H (2022) Complete Chloroplast genomes of two mint species (Lamiaceae) from Al-Madinah, Saudi arabia: phylogenetic and genomic comparative analyses. *Mitochondrial DNA Part B* 7:1797–1799
29. Huaizhu L, Lin Y, Bai J et al (2020) The complete Chloroplast genome sequence of *Mentha canadensis* (Labiatae), a traditional Chinese herbal medicine. *Mitochondrial DNA Part B* 5:55–56
30. He S-L, Yang Y, Tian Y (2020) Characteristic and phylogenetic analyses of Chloroplast genome for *Mentha Haplocalyx* (Lamiaceae). *Mitochondrial DNA Part B* 5:2099–2100
31. Alp Furan M, Yildiz F, Kaya O (2025) Exploring the complete Chloroplast genomes of key Lamiaceae species uncovers the secrets of evolutionary dynamics and phylogenetic relationships. *J Plant Growth Regul* 1–14
32. 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>
33. 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
34. Wick RR, Schultz MB, Zobel J, Holt KE (2015) Bandage: interactive visualization of de Novo genome assemblies. *Bioinformatics* 31:3350–3352
35. Shi L, Chen H, Jiang M et al (2019) CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res* 47:W65–W73
36. Tillich M, Lehwark P, Pellizzer T et al (2017) GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res* 45:W6–W11
37. Qu X-J, Moore MJ, Li D-Z, Yi T-S (2019) PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* 15:1–12
38. 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
39. Beier S, Thiel T, Münch T et al (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33:2583–2585
40. Tamura K, Stecher G, Peterson D et al (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729
41. Huang L, Yu H, Wang Z, Xu W (2024) CPStools: A package for analyzing Chloroplast genome sequences. *iMetaOmics* 1:e25
42. 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>
43. 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>
44. 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>
45. 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
46. Su T, Geng Y-F, Xiang C-L et al (2022) Chloroplast genome of *salvia* sect. *Drymospace*: comparative and phylogenetic analysis. *Diversity* 14:324
47. Lukas B, Novak J (2013) The complete Chloroplast genome of *Origanum vulgare* L. (Lamiaceae). *Gene* 528:163–169. <https://doi.org/10.1016/j.gene.2013.07.026>
48. Qian J, Song J, Gao H et al (2013) The complete Chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS ONE* 8:e57607
49. Xiao-Ming Z, Junrui W, Li F et al (2017) Inferring the evolutionary mechanism of the Chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. *Sci Rep* 7:1555
50. Turudić A, Liber Z, Grdiša M et al (2023) Variation in Chloroplast genome size: biological phenomena and technological artifacts. *Plants* 12:254
51. Ping J, Feng P, Li J et al (2021) Molecular evolution and SSRs analysis based on the Chloroplast genome of *Callitropsis funebris*. *Ecol Evol* 11:4786–4802
52. Jayaswall K, Sharma H, Jayaswal D et al (2023) Development of Chloroplast derived SSR markers for genus *Allium* and their characterization in the allies for genetic improvement of alliums. *South Afr J Bot* 162:304–313
53. Kumar B, Kumar U, Yadav HK (2015) Identification of EST-SSRs and molecular diversity analysis in *Mentha Piperita*. *Crop J* 3:335–342

54. Fukui Y, Saito M, Nakamura N et al (2022) Classification of Southeast Asian mints (*Mentha* spp.) based on simple sequence repeat markers. *Breed Sci* 72:181–187
55. Wang R-J, Cheng C-L, Chang C-C et al (2008) Dynamics and evolution of the inverted repeat-large single copy junctions in the Chloroplast genomes of monocots. *BMC Evol Biol* 8:1–14
56. Kress WJ, Wurdack KJ, Zimmer EA et al (2005) Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci* 102:8369–8374
57. Ruhlman TA, Jansen RK (2014) The plastid genomes of flowering plants. *Chloroplast Biotechnol Methods Protoc* 3–38
58. Kim K-J, Lee H-L (2004) Complete Chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res* 11:247–261
59. Khan A, Asaf S, Khan AL et al (2019) Complete Chloroplast genomes of medicinally important *Teucrium* species and comparative analyses with related species from Lamiaceae. *PeerJ* 7:e7260. <https://doi.org/10.7717/peerj.7260>
60. Du Q, Yang H, Zeng J et al (2022) Comparative genomics and phylogenetic analysis of the Chloroplast genomes in three medicinal *Salvia* species for bioexploration. *Int J Mol Sci* 23:12080
61. He L, Qian J, Li X et al (2017) Complete Chloroplast genome of medicinal plant *Lonicera japonica*: genome rearrangement, intron gain and loss, and implications for phylogenetic studies. <https://doi.org/10.3390/molecules22020249>. *Molecules* 22:
62. Hu J, Zhao M, Hou Z, Shang J (2020) The complete Chloroplast genome sequence of *Salvia miltiorrhiza*, a medicinal plant for preventing and treating vascular dementia. *Mitochondrial DNA Part B* 5:2460–2462
63. Zhu A, Guo W, Gupta S et al (2016) Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol* 209:1747–1756
64. Raubeson LA, Peery R, Chumley TW et al (2007) Comparative Chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 8:1–27
65. Dong W, Xu C, Li C et al (2015) *ycf1*, the most promising plastid DNA barcode of land plants. *Sci Rep* 5:8348
66. Chumley TW, Palmer JD, Mower JP et al (2006) The complete Chloroplast genome sequence of *pelargonium* × *hortorum*: organization and evolution of the largest and most highly rearranged Chloroplast genome of land plants. *Mol Biol Evol* 23:2175–2190
67. Bobur K, Jorabek T, Sarviniso M, Ziyoviddin Y (2025) First complete Chloroplast genome of *Thymus seravschanicus*: insights into genome structure and phylogenetic relationships within Lamiaceae. *J Asia-Pacific Biodivers* 18:185–189
68. Fu G, Liu Y, Caraballo-Ortiz MA et al (2022) Characterization of the complete Chloroplast genome of the Dragonhead herb, *Draacocephalum heterophyllum* (Lamiaceae), and comparative analyses with related species. *Diversity* 14:110
69. Moghaddam M, Wojciechowski MF, Kazempour-Osaloo S (2023) Characterization and comparative analysis of the complete plastid genomes of four *Astragalus* species. *PLoS ONE* 18:e0286083
70. Liu Y, Zheng C, Su X et al (2024) Comparative analysis and characterization of the Chloroplast genome of *Krascheninnikovia ceratoides* (Amarathaceae): a xerophytic semi-shrub exhibiting drought resistance and high-quality traits. *BMC Genomic Data* 25:10
71. Li L, Wu Q, Zhai J et al (2024) Comparative Chloroplast genomics of 24 species shed light on the genome evolution and phylogeny of subtribe *Coelogyneae* (Orchidaceae). *BMC Plant Biol* 24:31
72. He L, Xu S, Cheng X et al (2024) Chloroplast genomes in seven *Lagerstroemia* species provide new insights into molecular evolution of photosynthesis genes. *Front Genet* 15:1378403

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.