

Article

The Complete Chloroplast Genome of a Medical Herb, *Myricaria squamosa* Desv. (Tamaricaceae), from Qinghai–Tibet Plateau in China

Likuan Liu ^{1,2,†}, Yifan Wang ^{1,†}, Haoyu Liu ¹, Tianyu Zhao ¹, Jianfen Ma ¹, Haotian Wang ¹, Kaidi Su ¹ and Jinping Li ^{1,2,*}

- ¹ Qinghai Provincial Key Laboratory of Medicinal Plant and Animal Resources of Qinghai–Tibet Plateau, School of Life Sciences, Qinghai Normal University, Xining 810008, China; 2025105@qhnu.edu.cn (L.L.); 15237886253@163.com (Y.W.)
- ² Plateau Science and Sustainability, Qinghai Normal University, Xining 810008, China
- * Correspondence: 2025030@qhnu.edu.cn; Tel.: +86-131-3907-7708
- † These authors contributed equally to this work.

Abstract

Background: *Myricaria squamosa* Desv., a perennial shrub of the Tamaricaceae family, holds significant ecological, medicinal and economic values on the Qinghai–Tibet Plateau. **Methods:** This study performed chloroplast whole-genome sequencing of *M. squamosa* and analyzed its genomic characteristics, codon usage bias, and phylogenetic relationships. **Results:** The *M. squamosa* chloroplast genome spans 154,689 bp, with a GC content of 36.3%. A total of 129 functional genes were annotated, including 84 protein-coding genes, 37 transport RNA (tRNA) genes, and eight ribosomal RNA (rRNA) genes. Structural analysis revealed that contraction and expansion of the inverted repeat (IR) boundaries in Tamaricaceae species primarily involve the *rps19* and *ycf1* genes. We identified 214 simple sequence repeat (SSR) loci. Comparative genomic analysis indicated higher variability in the large single-copy (LSC) and small single-copy (SSC) regions compared to the IR regions, and greater divergence in intergenic spacers than in protein-coding regions. All protein-coding sequences contain a total of 26,045 codons, with leucine (2750) being the most frequent, and a pronounced preference for adenine/thymine (A/T) endings. Phylogenetic reconstruction revealed that *M. squamosa* is most closely related to *Myricaria laxiflora*. **Conclusions:** This study provides valuable insights into the chloroplast genomic structure, codon usage patterns, and SSR distribution of *M. squamosa*, contributing to genomic evolution studies and taxonomic revision within the Tamaricaceae family.

Keywords: *Myricaria squamosa* Desv.; chloroplast genome; medical herb; Qinghai–Tibet Plateau



Academic Editor: Zhiqiang Wu

Received: 30 June 2026

Revised: 25 July 2026

Accepted: 27 July 2026

Published: 29 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Myricaria squamosa Desv. (Tamaricaceae), is a fundamental component in traditional Tibetan medicine, where it is locally known as “Umbu”. The taxonomic classification of the genus *Myricaria* has long been a subject of debate, particularly concerning the taxonomic status of *Myrtama elegans*. The authors of [1,2] carried out a comprehensive taxonomic revision of the genus *Myricaria germanica* in China based on traditional morphological traits. They concluded that *M. germanica* is not present in China and proposed its geographical replacement as the newly described species *Myricaria paniculata*. Furthermore, they argued that certain morphological features of *Myrtama elegans*, such as the partial fusion of

filaments at the base and the presence of white, long, soft hairs on the seed tips, are also observed in other species within the genus *Myricaria*, suggesting that *M. elegans* should not be classified under a separate genus. The morphological definition of species within the genus *Myricaria* has traditionally relied on type specimens and limited individual samples, which contributes to the ongoing challenges in achieving taxonomic consensus due to significant morphological differences and extensive interspecific hybridization [3]. Consequently, an increasing number of scholars worldwide are employing molecular biological techniques to address taxonomic uncertainties. Zhang et al. [4] collected samples representing 12 species from the genus *Myricaria*, as well as specimens from the genera *Tamarix* and *Reaumuria*. Amplification and sequencing of these samples were carried out using four molecular markers: ITS, *rps16*, *psbB-psbH*, and *trnL-trnF*. Phylogenetic analysis and molecular dating were conducted, employing the relaxed Bayesian molecular clock method implemented in BEAST. The results indicate that *Myricaria* comprises four distinct phylogenetic clades, each corresponding to a specific lineage within the genus. Molecular dating suggests that the initial diversification of *Myricaria* occurred during the early Miocene. Hu et al. [5] collected 25 samples from three species and 10 samples from other species within *Myricaria*, such as *M. laxiflora*. Phylogenetic tree analyses indicated that *Myricaria bracteata*, *M. squamosa*, and *M. paniculata* could not be clearly differentiated, forming two clustered monophyletic lineages collectively referred to as the *Squamosa* complex. Plastid phylogenetic analysis further revealed that this scaly *Myricaria* clade consists of two evolutionarily distinct units: one primarily distributed in the western Tarim Basin, and the other in the eastern Qinghai–Tibet Plateau. These findings contradict previous taxonomic classifications and suggest an urgent need for a re-evaluation and revision of the species boundaries within *Myricaria*. The desiccated shoots are extensively employed as medicinal remedies, exhibiting the therapeutic properties of dispersing exanthema, alleviating heat, and detoxifying. *Myricaria squamosa* exhibits a wide range of biological activities, including antibacterial, anti-inflammatory, analgesic, antioxidant, and anti-fatigue effects, as well as promoting cellular immunity, protecting against liver damage, and demonstrating anti-HIV properties [6].

The genus *Myricaria* comprises approximately 13 species, with a distribution centered on the Qinghai–Tibet Plateau and the northern temperate regions, and extending into Central Asia and Western Europe. In China, there are around 10 species and one variety [4,7]. They are predominantly found in mountainous areas, coastal regions, and sandy lands near lakes, at elevations ranging from 2400 to 4600 m, in provinces such as Qinghai, Tibet, Sichuan, Xinjiang, and Gansu. These species occupy critical ecological niches in wetlands, deserts, and riparian zones, where their robust root systems stabilize soils, mitigate erosion, and enhance water retention, underscoring their role in maintaining ecosystem resilience [8]. Chloroplasts possess their own genomes and serve as the primary sites of photosynthesis [9]. Their genomes exhibit high conservation in most higher plants, characterized by low mutation rates and gradual evolutionary dynamics, making them powerful tools for plant taxonomy and phylogenetics [10]. Advances in high-throughput sequencing have substantially reduced cost and turnaround time, enabling broad application of chloroplast genome data in species identification, phylogenetic reconstruction, and evolutionary studies. DNA barcoding, which uses one or more standardized short DNA regions to delimit taxonomic units, has gained accuracy, speed, and reliability through next-generation sequencing. Wang et al. [11] evaluated the efficacy of six candidate DNA barcode regions—ITS, ITS2, *psbA-trnH*, *matK*, *rbcL*, and *trnL-F*—for species-level identification within the genus *Stephania*. The findings indicated that multi-locus barcode combinations achieved higher identification success compared to individual barcodes. Using both best match and best proximity criteria, the ITS+*psbA-trnH* combination exhibited the strongest discriminatory power, achieving a species identification

rate of 93.93%. Shang et al. [12] identified four mutation hotspots (*accD-psaI*, *atpH-atpI*, *ndhC-trnV* (UAC), and *ndhF-rpl23*) within the chloroplast genomes of seven *Ajuga* species, which can serve as valuable markers for distinguishing among species within this genus. Furthermore, chloroplast genome-based phylogenetic analyses have clarified the placement of *Lysimachia* within Primulaceae [13], establishing a framework to investigate interactions between the chloroplast genome, other plastid genes, and the nuclear genome. These findings underscore how integrating multi-locus barcodes and phylogenomic signals can improve the resolution and trustworthiness of species identification, while enhancing the interpretability of evolutionary relationships [14].

In this study, we sequenced and assembled the complete chloroplast genome of *M. squamosa*. The research foci of this study are as follows:

1. Characterize the complete chloroplast genome of *M. squamosa* through sequencing and assembly, elucidating its structural features, repeat sequences, and codon usage patterns.
2. Determine the evolutionary relationships and taxonomic position of *M. squamosa* by constructing a phylogenetic tree based on chloroplast genomic data.
3. Clarify phylogenetic demarcations within and across relevant genera by conducting comparative genomic analyses against related species.

Thus, we intend to furnish molecular genetic evidence that elucidates the evolutionary affinities of *Myricaria squamosa* chinensis and the phylogenetic demarcations within and across genera.

2. Materials and Methods

Myricaria plants were systematically collected from extensive regions across Qinghai Province and the northwestern part of Sichuan Province during the summers of 2022 and 2023.

2.1. Comprehensive DNA Extraction and Detection

After the complete grinding of the desiccated leaves, the total DNA was extracted utilizing the DP305 DNA extraction kit (TIANGEN BIOTECH, Beijing, China). For detailed operational procedures, kindly refer to the enclosed instructions provided with the kit. The concentration and purity of the extracted total DNA were determined using a micronucleic acid concentration tester (Thermo Fisher Scientific, Shanghai, China). The ratio of OD260 to OD280 for the highly pure DNA ranged between 1.7 and 1.9. A 2 µL aliquot of the DNA sample was mixed with 2 µL of 6× loading buffer, followed by agar gel electrophoresis at a voltage of 120 V.

2.2. The Assembly and Annotation of Chloroplast Genomes

The quality-assessed DNA samples were submitted to Novogene Bioinformatics Technology Co., Ltd. (Beijing, China) for second-generation sequencing and paired-end sequencing on the Illumina HiSeq X platform (Illumina, Inc., San Diego, CA, USA). Raw reads were filtered using fastp v0.11.7 with the following criteria: reads containing adapter sequences were removed; reads in which >50% of bases exhibited a Phred quality score ≤ 5 were discarded; and reads with an N content exceeding 10% were eliminated. The resulting high-quality clean reads were used for plastid genome reconstruction [15]. The clean reads were compared with the previously published chloroplast genome of the *Tamarix* family, and subsequently, the reads associated with the chloroplast genome were chosen. These selected reads were then utilized for contig assembly using SOAPdenovo2 (r240) software, followed by subsequent optimization of the assembly outcomes [16]. The completeness and structural accuracy of the assembly were validated by mapping all clean reads back to

the assembled genome; uniform coverage depth and concordant insert size distributions confirmed the absence of misassemblies and structural variants.

The complete chloroplast genome sequence of *M. squamosa* was annotated using Geneious Prime 2022.2.2 software, with manual adjustments made to the start codon and stop codon in the annotation results. Occasionally, the automatic annotation incorrectly assigned initiation codons (typically ATG) or termination codons (TAA, TAG, TGA) due to sequence noise or alignment biases, necessitating manual adjustments. For instance, the initial software annotation assigned non-standard start codons (e.g., GTG) to certain protein-coding genes, such as *rps16* and *atpF*. These were subsequently corrected to ATG based on manual adjustments informed by homology-based comparisons with genes from closely related Tamaricaceae species, including *M. laxiflora* and *Tamarix ramosissima*. Annotation fidelity was further assessed by BLASTP searches against the NCBI non-redundant protein database (e-value $< 1 \times 10^{-5}$) and by verifying conserved domain architectures. Finally, the gene annotation and sequence information were uploaded to the NCBI site with registration number NC_060684. Using the Chloroplot online tools (<https://irscope.shinyapps.io/Chloroplot/> (accessed on 24 April 2026)) [17], a chloroplast genome map was generated with a scale bar for *M. squamosa*.

2.3. The Analysis of IR Boundaries

In the study of chloroplast genomic structure evolution, the expansion and contraction of the IR region represent a key research focus. Variations in genome length among species are generally considered to be closely associated with changes in the IR region [18]. The GB files of nine species belonging to the *Tamarix* species were downloaded from the NCBI website, including *Tamarix taklamakanensis*, *T. ramosissima*, *Tamarix laxa*, *Tamarix karelinii*, *Tamarix arceuthoides*, *M. laxiflora*, *Myricaria prostrata*, *M. elegans* and *Myricaria Desv.* To analyze the contraction and expansion of IR boundary regions among these different species, we employed the online tool IRSCOPE available at <https://irscope.shinyapps.io/irapp/> (accessed on 24 April 2026) [19].

2.4. The Analysis of Repeated Sequences

SSRs, also known as microsatellites, are short tandem repeats characterized by a high degree of uniformity in repeat unit length and base composition, with only minor variations observed. SSR molecular markers developed based on chloroplast genomes display significant genetic polymorphism both within and across species, making them widely applicable in the fields of population genetics and phylogenetic research [18]. The analysis of an SSR locus in the chloroplast genome of *M. squamosa* was conducted using the MISA-web online tool (<https://webblast.ipk-gatersleben.de/misa/> (accessed on 24 April 2026)) [20]. The parameters were set as follows: the minimum number of repetitions for single nucleotides was set to ≥ 8 , the minimum number of repetitions for dinucleotides and trinucleotides was set to ≥ 4 , and the minimum number of repetitions for tetranucleotides up to six nucleotides was set to ≥ 3 .

2.5. Comparative Analysis of the Genome

The chloroplast genomes of the *Tamarix* species, including *Reaumuria songarica*, *Tamarix arceuthoides*, *Tamarix chinensis*, *Tamarix gracilis*, and *Tamarix karelinii*, were converted from GB format to gb_mvista_annotation format using scaly *M. squamosa* as a reference. The visualization analysis of these chloroplast genomes was conducted using mVISTA online tools (<https://genome.lbl.gov/vista/mvista/submit.shtml> (accessed on 24 April 2026)) [21].

2.6. The Analysis of Codon Bias

The software Geneious Primer 2022.2.2 was utilized for exporting a total of 84 protein-coding sequences from the chloroplast genome of *M. squamosa*, while the CodonW 1.4.2 software was employed to analyze the codon usage frequency and relative codon usage frequency within the same chloroplast genome [22,23].

2.7. Phylogenetic Analysis

Complete chloroplast genome sequences of 16 species representing Tamaricaceae and Plumbaginaceae were retrieved in FASTA format from the NCBI database. The Tamaricaceae taxa comprised *Tamarix aphylla*, *Tamarix chinensis*, *Tamarix gracilis*, *T. ramosissima*, *T. laxa*, *Tamarix karelinii*, and *Tamarix arceuthoides*; *Myricaria wardii*, *Myricaria rosea*, *M. laxiflora*, *Myricaria prostrata*, and *M. elegans*; and *Reaumuria songarica* and *Reaumuria trigyna*. The outgroup consisted of *Plumbago zeylanica* and *Plumbago auriculata* (Plumbaginaceae). *M. squamosa* was included as the focal taxon.

Whole-chloroplast genome sequences were aligned using MAFFT v7 with the FFT-NS-2 algorithm under default parameters [24]. Phylogenetic trees were reconstructed using two independent methods: neighbor joining (NJ) and maximum likelihood (ML), both implemented in MEGA version 11 [25]. Branch support was assessed through bootstrap analysis with 1000 pseudoreplicates for both NJ and ML trees.

3. Results

3.1. Structural Features of the Chloroplast Genome

The chloroplast genome exhibited a total length of 154,689 base pairs (bps), featuring the characteristic four-segment circular double-stranded structure. Additionally, the GC content accounted for 36.3%. The LSC region has a length of 84,378 bp, the SSC region has a length of 18,303 bp, and the two inverted repeat (IR) regions have a combined length of 26,004 bp (Figure 1).

The chloroplast genome of *M. squamosa* harbors a total of 129 genes, comprising 84 protein-coding genes (PCG), 37 transfer RNA genes (tRNA), and eight ribosomal RNA genes (rRNAs). Among them, there were a total of 18 genes that exhibited copy gene characteristics (*ndhB*, *rps7*, *rps12*, *rpl2*, *rpl23*, *trnV-GAC*, *trnR-ACG*, *trnN-GUU*, *trnL-UAA*, *trnL-CAA*, *trnI-GAU*, *trnI-CAU*, *trnA-UGC*, *rrn4.5*, *rrn5*, *rrn16*, *rrn23*, *ycf2*), all of which were precisely located within the reverse repeat region. There are 13 genes (*ndhA*, *ndhB*, *atpF*, *petB*, *rpoC1*, *rps16*, *rpl16*, *trnV-UAC*, *trnL-UAA*, *trnK-UUU*, *trnI-GAU*, *trnG-UCC* and *trnA-UGC*) that each contain a single intron. Additionally, three genes (*rps12*, *clpP* and *ycf3*) possess two introns, as indicated in Table 1. Hu et al. [5] conducted a comparative analysis of the three species *M. bracteata*, *M. paniculata* and *M. squamosa*, which encoded a total of 88–90 PCGs and 8 rRNAs. Liu et al. [26] conducted an in-depth analysis of the complete chloroplast genome sequences of 12 *Tamarix* species, which includes 120 to 130 genes, encompassing 76 to 85 PCGs, 36 to 37 tRNAs, and eight rRNAs. This consistency confirms our gene count is within the typical range for the family.

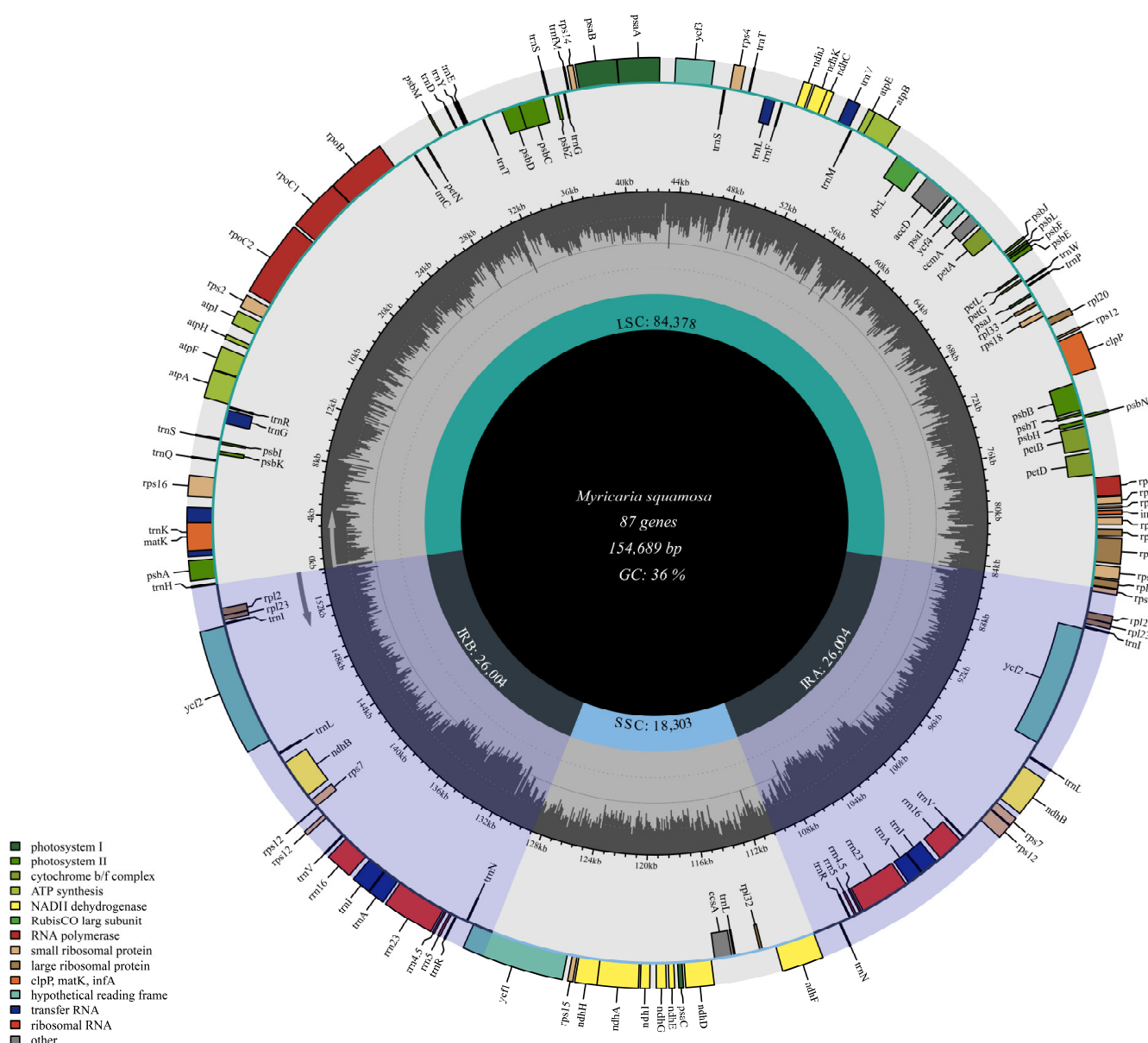


Figure 1. Gene map of *Myricaria squamosa* chloroplast genome. Genes are displayed on the interior and exterior of the circle. Different colors represent the functional categories of each gene. The black in the middle circle depicts the chloroplast structure, divided into four regions: the small single-copy region (SSC), large single-copy region (LSC), and two inverted repeats (IRa and IRb).

Table 1. List of genes of *Myricaria squamosa* chloroplast genome.

Category of Genes	Group of Genes	Names of Genes
Photosynthesis genes	ATP synthase genes	<i>atpA, atpB, atpE, atpF *, atpH, atpI</i>
	NADH dehydrogenase genes	<i>ndhA *, ndhB * (×2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	Cytochrome b/f complex genes	<i>petA, petB *, petG, petL, petN</i>
	Photosystem I genes	<i>psaA, psaB, psaC, psaI, psaJ</i>
	Photosystem II genes	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>

Table 1. Cont.

Category of Genes	Group of Genes	Names of Genes
Self-replication genes	RubisCO large subunit gene	<i>rbcL</i>
	RNA polymerase genes	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> *, <i>rpoC2</i>
	Ribosomal proteins (SSU) genes	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> (×2), <i>rps8</i> , <i>rps11</i> , <i>rps12</i> ** (×2), <i>rps14</i> , <i>rps15</i> , <i>rps16</i> *, <i>rps18</i> , <i>rps19</i>
	Ribosomal proteins (LSU) genes	<i>rpl2</i> (×2), <i>rpl14</i> , <i>rpl16</i> *, <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> (×2), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
	Transfer RNA genes	<i>trnY</i> -GUA, <i>trnV</i> -GAC (×2), <i>trnV</i> -UAC *, <i>trnW</i> -CCA, <i>trnT</i> -UGU, <i>trnT</i> -GGU, <i>trnS</i> -UGA, <i>trnS</i> -GGA, <i>trnS</i> -GCU, <i>trnR</i> -UCU, <i>trnR</i> -ACG (×2), <i>trnQ</i> -UUG, <i>trnP</i> -UGG, <i>trnN</i> -GUU (×2), <i>trnM</i> -CAU, <i>trnL</i> -UAG, <i>trnL</i> -UAA * (×2), <i>trnL</i> -CAA (×2), <i>trnK</i> -UUU *, <i>trnI</i> -GAU * (×2), <i>trnI</i> -CAU (×2), <i>trnH</i> -GUG, <i>trnG</i> -UCC *, <i>trnG</i> -GCC, <i>trnM</i> -CAU, <i>trnF</i> -GAA, <i>trnE</i> -UUC, <i>trnD</i> -GUC, <i>trnC</i> -GCA, <i>trnA</i> -UGC * (×2)
	Ribosomal RNA genes	<i>rrn4.5</i> (×2), <i>rrn5</i> (×2), <i>rrn16</i> (×2), <i>rrn23</i> (×2)
	Protease gene	<i>clpP</i> **
Other genes	Maturase K gene	<i>matK</i>
	Envelop membrane protein gene	<i>cemA</i>
	Subunit of acetyl-CoA gene	<i>accD</i>
	C-type cytochrome synthesis gene	<i>ccsA</i>
	Hypothetical chloroplast reading frames	<i>ycf1</i> , <i>ycf2</i> (×2), <i>ycf3</i> **, <i>ycf4</i>
	Translational initiation factor gene	<i>infA</i>

×2 indicates that the gene has two copies; * indicates that the gene contains one intron, and ** indicates that the gene contains two introns.

3.2. IR Boundary Analysis

By comparing the chloroplast genomes of nine *Tamarix* species in the IR boundary region, it was observed that the LSC/IRb boundaries of four *Myricaria* species, *T. laxa*, and *Tamarix karelinii* were located on the *rps19* gene. Conversely, the remaining three *Tamarix* species lacked the *rps19* gene. The boundaries between the inverted repeat B (IRb) and SSC regions of *T. taklamakanensis*, *T. laxa*, *Tamarix karelinii*, *Myricaria laxiflora*, and *Myricaria prostrata* sparsely flowering and creeping ramosarix were identified on the *ycf1* gene. Additionally, it was observed that a major portion of the *ycf1* gene is located within the SSC region. Furthermore, for other species, the IRb/SSC boundaries were found to be positioned 1–2 base pairs away from the *ndhF* gene. The *ycf1* genes of *T. ramosissima*, *Tamarix arceuthoides*, *M. laxiflora*, and *M. squamosa* Desv were all found across the SSC/IRa boundary, with a majority of them located within the SSC region. The boundaries between *T. laxa* and *Tamarix karelinii* species were situated between the *ndhF* gene and *ycf1* gene, while other species had boundaries 2–3 bp away from the *ndhF* gene. The IRa/LSC boundaries for all species were positioned at one end of the *trnH* gene or in close proximity to each other. Additionally, there was only a single base pair separating the *rps19* gene from the TRNH gene in both *T. laxa* and *Tamarix karelinii* (Figure 2).

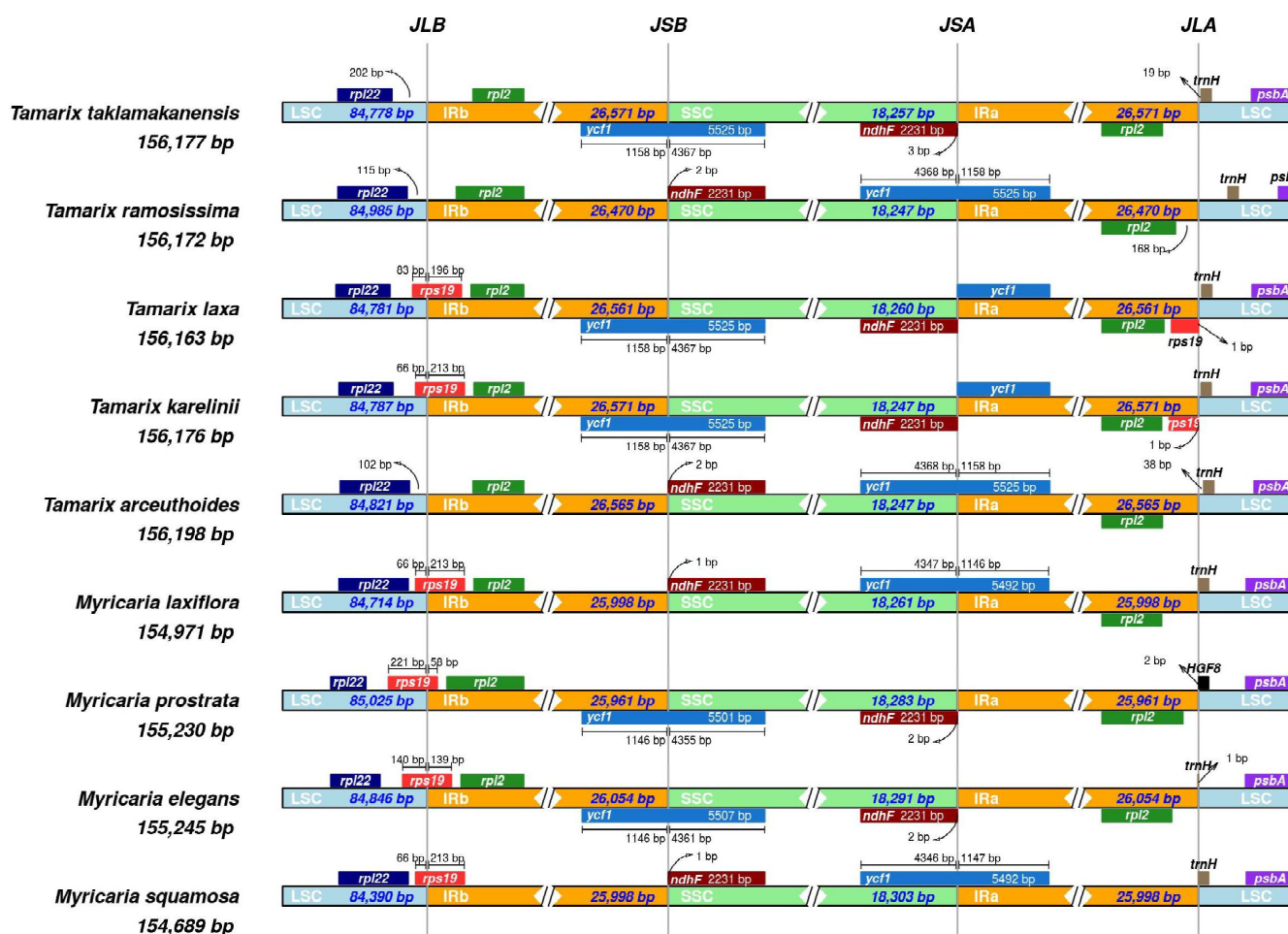


Figure 2. Schematic diagram of IR boundaries for 9 species in the Tamaricaceae family. Genes are represented by colored boxes, while arrows show the coordinate positions of each gene near the junctions. The gaps between the genes and boundaries are proportional to the distances in bps. Abbreviations denote the junction site of the plastid genome JLA (IRa/LSC), JLB (IRb/LSC), JSA (SSC/IRa), and JSB (IRb/SSC).

3.3. Repeat Sequence Analysis

The *M. squamosa* chloroplast genome contained a total of 214 SSR loci, comprising 151 mononucleotide repeats, 51 dinucleotide repeats, six trinucleotide repeats, and six tetranucleotide repeats. No pentanucleotide or hexanucleotide repeats were detected. The proportion of A/T in mononucleotide repeats was the highest, reaching 99.34%; in dinucleotide repeats, AT/AT accounted for the highest proportion, at 66.67%; similarly, in trinucleotide repeats, AAT/ATT had the highest proportion, at 66.67%; lastly, among tetranucleotide repeats, AATC/ATTG had the highest count and accounted for half of the total number (Table 2). The base composition of SSR loci in the chloroplast genome of *M. squamosa* is consistent with the high AT content (63.7%) observed across the entire genome.

Table 2. Type and number of SSR loci of *M. squamosa* chloroplast genome.

Type	Repeats	Number	Location of Repeats			Proportion
			LSC	SSC	IR	
mononucleotide						
A/T	8	64	42	12	10	99.34%
	9	42	22	8	12	
	10	24	19	5	0	
	11	8	7	1	0	
	12	2	2	0	0	
	13	5	5	0	0	
	14	3	3	0	0	
	16	1	1	0	0	
	17	1	1	0	0	
C/G	8	1	1	0	0	0.66%
dinucleotide						
AG/CT	4	16	4	0	12	33.33%
	5	1	1	0	0	
AT/TA	4	24	20	2	2	66.67%
	5	4	4	0	0	
	6	2	2	0	0	
	7	1	0	1	0	
	8	2	0	0	2	
	9	1	1	0	0	
trinucleotide						
AAG/CTT	4	2	0	0	2	33.33%
AAT/ATT	4	4	2	2	0	66.67%
tetranucleotide						
AAAG/CTTT	3	2	2	0	0	33.33%
AAAT/ATTT	3	1	1	0	0	16.67%
AATC/ATTG	3	3	0	1	2	50.00%

3.4. Comparative Genomic Analysis

Based on the comparison results of the six selected species, it was observed that the degree of variation in the LSC and SSC regions exceeded that of the IR region, while the gene spacer region exhibited a greater degree of variation than the protein-coding region. The gene spacer regions, including *rps16-psbK*, *ndhC-atpB*, *PBM-PSBD*, *ndhF-rpl32*, and *rpl32-ccsA*, exhibited significant variations.

The interspecific divergence between *Myricaria* and *Tamarix* was more pronounced in the *rps16-psbK* segment. The species difference between *Myricaria* and *Reaumuria* was more pronounced in *ndhC-atpB* and *rpl32-ccsA*. The *ycf1*, *accD*, and other genes exhibit simultaneous variations among different species (Figure 3).

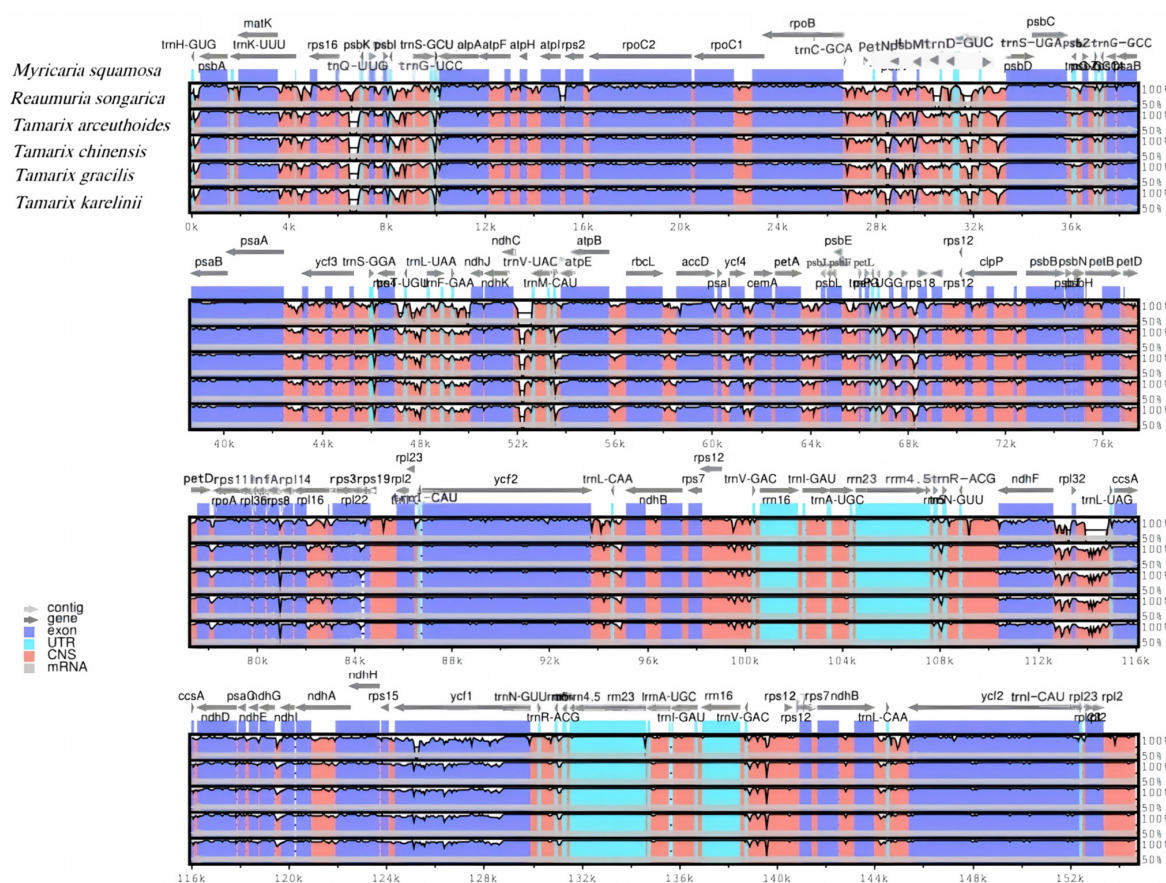


Figure 3. Comparative analysis of chloroplast genomes between *M. squamosa* and related species. The y-axis represents the percentage of interspecies consistency from 50% to 100%. Arrows represent genes and their directions. The dark blue area represents the protein-coding region, the cyan area represents rRNA or tRNA, and the orange area represents the gene spacer region.

3.5. Codon Usage Bias Analysis

After analyzing the protein-coding sequence in the *M. squamosa* chloroplast genome, it was observed that out of the 26,045 codons examined, leucine coding accounted for the highest proportion, with 2750 (10.56%), followed by isoleucine coding, with 2261 (8.68%). Conversely, cysteine codons were found to be least abundant, at 301 (1.16%) and TAA emerged as the preferred terminator codon, constituting approximately 58.8% of the total count. The relative synonymous codon usage analysis revealed that, out of the 30 codons with RSCU > 1, 13 and 15 had A or T as their terminal nucleotides, respectively. This observation indicates that the chloroplast genome of *M. squamosa* exhibits a preference for codons ending in A or T (Table 3).

Table 3. Usage of CDS sequence codons in chloroplast genome of *M. squamosa*.

Amino Acid	Codon	Number	RSCU
Phenylalanine Phe	TTT	1023	1.34
	TTC	499	0.66
Leucine Leu	TTA	887	1.94
	TTG	540	1.18
	CTT	594	1.30
	CTC	186	0.41
	CTA	381	0.83
	CTG	162	0.35

Table 3. *Cont.*

Amino Acid	Codon	Number	RSCU
Serine Ser	TCT	603	1.80
	TCC	302	0.90
	TCA	421	1.25
	TCG	171	0.51
	AGT	400	1.19
	AGC	116	0.35
Proline Pro	CCT	421	1.60
	CCC	186	0.71
	CCA	307	1.17
	CCG	138	0.52
Isoleucine Ile	ATT	1136	1.51
	ATC	414	0.55
	ATA	711	0.94
Methionine Met	ATG	623	1.00
Valine Val	GTT	518	1.51
	GTC	152	0.44
	GTA	525	1.53
	GTG	177	0.52
Threonine Thr	ACT	549	1.67
	ACC	223	0.68
	ACA	410	1.24
	ACG	136	0.41
Alanine Ala	GCT	609	1.78
	GCC	214	0.63
	GCA	428	1.25
	GCG	115	0.34
Tyrosine Tyr	TAT	794	1.64
	TAC	175	0.36
terminator TER	TAA	50	1.76
	TAG	19	0.67
	TGA	16	0.56
Cysteine Cys	TGT	238	1.58
	TGC	63	0.42
Tryptophan Trp	TGG	454	1.00
Histidine His	CAT	475	1.54
	CAC	143	0.46
Glutamine Gln	CAA	719	1.56
	CAG	200	0.44
Arginine Arg	CGT	367	1.39
	CGC	87	0.33
	CGA	365	1.38
	CGG	113	0.43
	AGA	465	1.76
	AGG	185	0.70
Asparagine Asn	AAT	1007	1.57
	AAC	278	0.43
Lysine Lys	AAA	1082	1.54
	AAG	326	0.46

Table 3. Cont.

Amino Acid	Codon	Number	RSCU
Aspartate Asp	GAT	853	1.57
	GAC	231	0.43
Glutamate Glu	GAA	1046	1.54
	GAG	317	0.46
Glycine Gly	GGT	566	1.33
	GGC	150	0.35
	GGA	728	1.72
	GGG	253	0.60

3.6. Phylogenetic Relationships Analysis

The phylogenetic trees reconstructed using neighbor joining (NJ) and maximum likelihood (ML) methods yielded identical topologies (Figure 4). Bootstrap support values were calculated from 1000 replicates to assess branch reliability. The two *Plumbago* species (*Plumbaginaceae*), designated as the outgroup, formed a distinct and strongly supported clade, confirming their distant relationship to *Tamaricaceae*. Within *Tamaricaceae*, the two *Reaumuria* species were resolved as a separate lineage, consistent with current taxonomic treatments. However, several topological features deviated from traditional generic delimitations. *Myricaria prostrata* and *M. elegans* grouped with five *Tamarix* species (*Tamarix karelinii*, *Tamarix chinensis*, *Tamarix gracilis*, *T. laxa*, and *Tamarix aphylla*) in one cluster. Moreover, *T. ramosissima* and *Tamarix arceuthoides* were placed with the remaining four *Myricaria* species in a second cluster, indicating that neither *Tamarix* nor *Myricaria* was recovered as monophyletic under the current sampling. Among all analyzed taxa, *M. laxiflora* was identified as the closest relative of *M. squamosa* with robust bootstrap support. These findings highlight phylogenetic incongruences that challenge the existing genus-level classification and underscore the need for an interpretable phylogenetic framework that can reliably resolve intergeneric boundaries within *Tamaricaceae*.

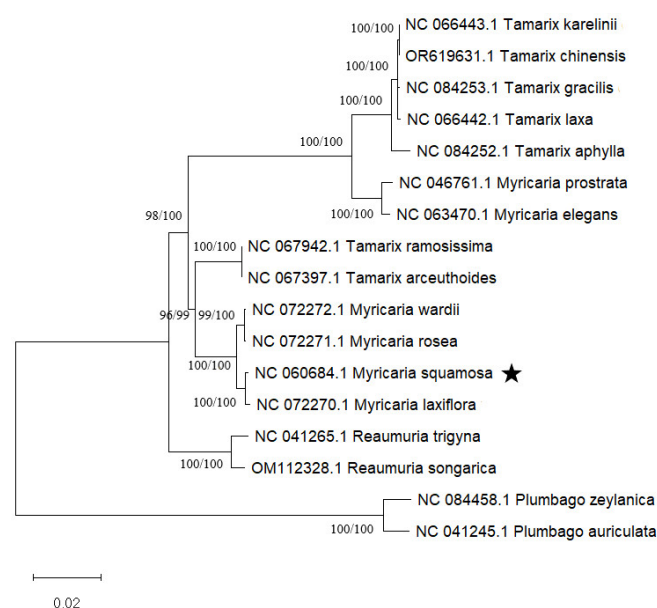


Figure 4. Chloroplast genome phylogenetic tree of 17 species, including *M. squamosa*, based on NJ and ML methods. Bootstrap values are shown near each node. ClustalW, integrated in MEGA-11 software, was utilized to perform sequence comparison, and phylogenetic trees were constructed using two methods: NJ and ML.

4. Discussion

The complete chloroplast genome of *M. squamosa* (154,689 bp, GC content 36.3%) exhibits the conserved quadripartite structure and gene content characteristics of Tamaricaceae [26], comprising 129 functional genes (84 protein-coding, 37 tRNA, and eight rRNA), 18 of which are duplicated in the inverted repeat regions. Thirteen genes contain one intron; *rps12*, *clpP*, and *ycf3* contain two introns, a pattern conserved across Tamaricaceae. Retention of these introns suggests possible roles in RNA processing [27].

Inverted repeat boundary analysis across nine Tamaricaceae species revealed substantial structural variation with potential phylogenetic utility. The LSC/IRb boundary resides within *rps19* in four *Myricaria* species, *T. laxa*, and *Tamarix karelinii*, whereas *rps19* is absent from this junction in three other *Tamarix* species. The IRb/SSC boundary lies within *ycf1* in *T. taklamakanensis*, *Tamarix laxa*, *Tamarix karelinii*, and two *Myricaria* species, but is positioned 1–2 bp from *ndhF* in the remaining taxa. These boundary shifts are not randomly distributed, but appear to coincide with major phylogenetic splits, suggesting they may serve as molecular markers for species delimitation [28]. However, the evolutionary mechanisms driving IR expansion and contraction—whether mediated by recombination, replication slippage, or selection—remain to be determined. Functional characterization of these boundary dynamics, particularly their impacts on gene expression or genome stability, would strengthen the interpretability of these markers beyond their descriptive value.

A total of 214 SSRs were identified. Mononucleotide repeats predominate (70.6%), with a strong A/T bias (99.34%), reflecting the genome's AT-rich composition (63.7%). No pentanucleotide or hexanucleotide repeats were found. Among dinucleotide repeats, AT/AT is most abundant (66.67%), followed by AG/CT (33.33%), similar to *Tamarix chinensis*, but differing from *Reaumuria songarica*. These SSRs are candidate markers for population genetic studies. However, empirical validation is required before they can be used as ready-to-use markers [29].

Comparative analysis confirmed that the LSC and SSC regions exhibit substantially higher sequence divergence than the IR regions, and intergenic spacers are more variable than coding sequences, consistent with established patterns across angiosperms [30]. Five hypervariable regions were identified: *rps16-psbK*, *ndhC-atpB*, *psbM-psbD*, *ndhF-rpl32*, and *rpl32-ccsA*. Of these, *rps16-psbK* showed the greatest divergence between *Myricaria* and *Tamarix*, whereas *ndhC-atpB* and *rpl32-ccsA* were most variable between *Myricaria* and *Reaumuria*. The elevated variations in *ycf1* and *accD* further support their established utility as phylogenetic markers at lower taxonomic levels [30]. These regions collectively provide a robust toolkit for species-level discrimination within Tamaricaceae. However, functional annotation of variation in *accD*, which encodes a subunit of acetyl-CoA carboxylase, is lacking; linking sequence divergence in this gene to possible adaptive significance would transform it from a purely diagnostic marker to a functionally informative locus.

Leucine is the most frequent amino acid (10.56%) and cysteine the least (1.16%). Among 30 codons with RSCU > 1, 13 end with A and 15 with T, reflecting the AT-rich genome composition, which is consistent with observations made in other plants [31]. The stop codon TAA is preferred (58.8%). Distinguishing mutational bias from translational selection would require tRNA or expression data [32].

The identical Phylogenetic analyses by NJ and ML analyses reveal that neither *Myricaria* nor *Tamarix* is monophyletic, with *Myricaria prostrata* and *M. elegans* embedded among *Tamarix* species and *T. ramosissima* and *Tamarix arceuthoides* nested within *Myricaria*. These results corroborate previously reported taxonomic uncertainties in Tamaricaceae and demonstrate that the morphological characters traditionally used to delimit these genera are homoplastic, underscoring the need for a molecularly grounded revision [5,33]. Beyond

cataloging incongruence, the observed intermingling demands mechanistic explanation. Two population-level processes—chloroplast capture via hybridization and incomplete lineage sorting of ancestral polymorphisms—are both plausible given the weak reproductive barriers and sympatric distributions in this family, yet they cannot be distinguished using chloroplast data alone [34]. More fundamentally, these discordant relationships may also arise from lineage-specific mutation dynamics of the organellar genome itself. As articulated in the rethinking of mutation hypotheses for plant organellar DNA, nucleotide composition biases, strand-specific substitution asymmetries, and IR-mediated gene conversion can generate homoplastic molecular patterns that mislead phylogenetic inference if not explicitly modeled [35]. The pronounced AT bias and variable IR boundaries observed in Tamaricaceae chloroplast genomes are consistent with such mutation-driven convergence, suggesting that some topological conflicts may reflect heterogeneous mutational pressures, rather than true genealogical history. This interpretation shifts the focus from simply reporting discordance to evaluating the evolutionary processes that shape organellar sequence evolution.

Several methodological considerations delimit the immediate applicability of our results. The plastid genome, as a single, uniparentally inherited locus, cannot resolve complex evolutionary scenarios involving hybridization or introgression. Furthermore, *in silico* marker discovery, while efficient, generates hypotheses that require empirical validation before translation into conservation or population genetic tools.

Author Contributions: L.L. and Y.W. designed and conceived the study. L.L.: Writing—original draft, Writing—review and editing, Resources, Supervision, Methodology, Investigation, Conceptualization; Y.W.: Writing—original draft, Writing—review and editing, Methodology, Formal analysis, Data curation, Visualization; H.L.: Writing—review and editing, Methodology; T.Z.: Writing—review and editing, Methodology; J.M.: Writing—review and editing, Methodology; H.W.: Writing—review and editing, Methodology; K.S.: Writing—review and editing; J.L.: Writing—review and editing, Resources, Supervision, Funding acquisition, Methodology, Conceptualization. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Applied Basic Research Project of Qinghai Province, grant number 2023-ZJ-749.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All relevant data generated or analyzed during this study are included in this published article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NCBI	National Center for Biotechnology Information
RSCU	Relative synonymous codon usage
LSC	Large Single-Copy
NJ	Neighbor Joining
ML	Maximum Likelihood
SSC	Small Single-Copy
SSR	Simple Sequence Repeat

References

1. Gaskin, J.F. Tamaricaceae. In *Flowering Plants Dicotyledons*; Springer: Berlin/Heidelberg, Germany, 2003. [CrossRef]
2. Zhang, Y. Cluster analysis on pollen morphology of the Tamaricaceae from China. *Acta Botan. Boreali-Occident. Sin.* **2004**, *24*, 1702–1707. [CrossRef]
3. Wei, X.; Cui, X.; Yuan, F.; Zhou, K.; Zhou, L.; Zhao, C.; Guo, S.; Shang, C.; Zhang, Z. Species delimitation in the *Populus laurifolia* complex (Salicaceae) based on phylogenetic and morphometric evidence. *Front. Plant Sci.* **2025**, *16*, 1518122. [CrossRef]
4. Zhang, M.-L.; Meng, H.-H.; Zhang, H.-X.; Vyacheslav, B.V.; Sanderson, S.C. Himalayan Origin and Evolution of Myricaria (Tamaricaceae) in the Neogene. *PLoS ONE* **2014**, *9*, e97582. [CrossRef] [PubMed]
5. Hu, H.; Wang, Q.; Hao, G.; Zhou, R.; Luo, D.; Cao, K.; Yan, Z.; Wang, X. Insights into the phylogenetic relationships and species boundaries of the *Myricaria squamosa* complex (Tamaricaceae) based on the complete chloroplast genome. *PeerJ* **2023**, *11*, e16642. [CrossRef]
6. Zhang, M.G.; De, J.; Chen, D.F. Research advances on chemical compounds, pharmacological activities and quality standards of medical plants from Myricaria. *J. Pharm. Res.* **2023**, *42*, 73–81. [CrossRef]
7. Zhang, Y.M.; Pan, B.R.; Yin, L.G. The research history of the family Tamaricaceae. *Acta Bot. Boreali-Occident. Sin.* **2001**, *4*, 796–804.
8. Liu, H.; Chen, W.G.; Li, Y.; Fan, X. The complete chloroplast genome sequence of *Myricaria wardii* Marquand 1929 (Tamaricaceae): A shrub species endemic to the Tibet Plateau. *Mitochondrial DNA B Resour.* **2024**, *9*, 367–370. [CrossRef] [PubMed]
9. Hernández-Verdeja, T. Regulation of chloroplast biogenesis and differentiation. *J. Exp. Bot.* **2026**, *77*, 1568–1581. [CrossRef] [PubMed]
10. Daniell, H.; Lin, C.-S.; Yu, M.; Chang, W.-J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **2016**, *17*, 134. [CrossRef] [PubMed]
11. Xieli Wang, J.X.; Zhang, Y.; Xie, H.; Wang, Y.; Weng, W.; Kang, Y.; Huang, J. DNA barcodes for the identification of *Stephania* (Menispermaceae) species. *Mol. Biol. Rep.* **2020**, *47*, 2197–2203. [CrossRef] [PubMed]
12. Shang, M.; Wang, J.; Dai, G.; Zheng, J.; Liao, B.; Wang, J.; Duan, B. Comparative analysis of chloroplast genome and new insights into phylogenetic relationships of *Ajuga* and common adulterants. *Front. Plant Sci.* **2023**, *14*, 1251829. [CrossRef] [PubMed]
13. Lee, Y.; Yun, N.; Kang, J.; Choi, S.; Paik, J.H. The complete chloroplast genome of the medicinal plant *Lysimachia mauritiana* (Lamarck, 1792). *Mitochondrial DNA B Resour.* **2022**, *7*, 554–555. [CrossRef] [PubMed]
14. Sloan, D.B.; Warren, J.M.; Williams, A.M.; Wu, Z.; Abdel-Ghany, S.E.; Chicco, A.J.; Havird, J.C. Cytonuclear integration and co-evolution. *Nat. Rev. Genet.* **2018**, *19*, 635–648. [CrossRef] [PubMed]
15. Chen, S. fastp 1.0: An ultra-fast all-round tool for FASTQ data quality control and preprocessing. *iMeta* **2025**, *4*, e70078. [CrossRef] [PubMed]
16. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y.; et al. SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience* **2012**, *1*, 2047–217X-1-18. [CrossRef] [PubMed]
17. Zheng, S.; Poczai, P.; Hyvönen, J.; Tang, J.; Amiryousefi, A. Chloroplot: An Online Program for the Versatile Plotting of Organelle Genomes. *Front. Genet.* **2020**, *11*, 576124. [CrossRef] [PubMed]
18. Sun, Y.Q. Comparative Analysis of Chloroplast Genomes and Study Onpotential Suitable Habitats of *Glehnia littoralis*. Master's Thesis, Nanjing Forestry University, Nanjing, China, 2024. Available online: <https://link.cnki.net/doi/10.27242/d.cnki.gnjlu.2024.000610> (accessed on 25 July 2026).
19. Amiryousefi, A.; Hyvönen, J.; Poczai, P.; Hancock, J. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [CrossRef] [PubMed]
20. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M.; Valencia, A. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [CrossRef] [PubMed]
21. Frazer, K.A.; Lior, P.; Alexander, P. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, 273–279. [CrossRef] [PubMed]
22. Liang, H.H.; Fu, H.Y.; Li, Z.P. Analysis on Codon Usage Bias of Chloroplast Genome from *Chlorella*. *Mol. Plant Breed.* **2020**, *18*, 5665–5673. [CrossRef]
23. Tang, D.; Wei, F.; Cai, Z.; Wei, Y.; Khan, A.; Miao, J.; Wei, K.-H. Analysis of codon usage bias and evolution in the chloroplast genome of *Mesona chinensis* Benth. *Dev. Genes Evol.* **2020**, *231*, 1–9. [CrossRef] [PubMed]
24. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **2019**, *20*, 1160–1166. [CrossRef] [PubMed]
25. Tamura, K.; Stecher, G.; Kumar, S.; Battistuzzi, F.U. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef] [PubMed]
26. Liu, Y.; Ding, K.; Liang, L.; Zhang, Z.; Chen, K.; Li, H. Comparative study on chloroplast genome of *Tamarix* species. *Ecol. Evol.* **2024**, *14*, e70353. [CrossRef] [PubMed]
27. Porras-Tobias, A.L.; Caldera, A.; Castro-Piedras, I. Intron Retention: A Reemerging Paradigm in RNA Biology and Post-Transcriptional Gene Regulation. *Genes* **2025**, *16*, 986. [CrossRef] [PubMed]

28. Li, J.H.; Zheng, Q.Q.; Guan, R.R.; Liu, Y.P.; Su, X.; Zhang, Q.F. Structure Characterization and Phylogeny of Chloroplast Genomes in Tamaricaceae. *Bull. Bot. Res.* **2025**, *45*, 873–887.
29. Zhang, L.; Zhang, E.; Wei, Y. Phylogenetic analysis and divergence time estimation of *Lycium* species in China based on the chloroplast genomes. *BMC Genom.* **2024**, *5*, 569.
30. Diani Gohar, S.; Soorni, A. Plastome diversity and phylogenomic analysis of *Satureja* (Lamiaceae): Uncovering evolutionary patterns and diagnostic markers. *Planta* **2025**, *262*, 149. [[CrossRef](#)] [[PubMed](#)]
31. Parvathy, S.T.; Udayasuriyan, V.; Bhadana, V. Codon usage bias. *Mol. Biol. Rep.* **2022**, *49*, 539–565. [[CrossRef](#)] [[PubMed](#)]
32. Hu, X.; Li, Y.; Meng, F. Analysis of chloroplast genome characteristics and codon usage bias in 14 species of Annonaceae. *Funct. Integr. Genom.* **2024**, *24*, 109. [[CrossRef](#)]
33. Villar, J.L.; Alonso, M.Á.; Juan, A.; Gaskin, J.F.; Crespo, M.B. Out of the Middle East: New phylogenetic insights in the genus *Tamarix* (Tamaricaceae). *J. Syst. Evol.* **2019**, *57*, 488–507. [[CrossRef](#)]
34. Sheidai, M.; Koohdar, F. Evidence for ancient introgression and gene flow in the genus *Tamarix* L. (Tamicaceae): A computational approach. *Genet. Resour. Crop Evol.* **2023**, *70*, 1653–1661. [[CrossRef](#)]
35. Wang, J.; Zou, Y.; Mower, J.P.; Reeve, W.; Wu, Z. Rethinking the mutation hypotheses of plant organellar DNA. *Genom. Commun.* **2024**, *1*, e003. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.