

## Article

# Comparative Analyses of Complete Chloroplast Genomes of *Microula sikkimensis* and Related Species of Boraginaceae

Yunqing Gao<sup>1,2,3,4,5,6</sup>, Zhenjiang Chen<sup>1,2,3,\*</sup>, Xiuzhang Li<sup>7</sup>, Kamran Malik<sup>1,2,3,4,5,6</sup>  and Chunjie Li<sup>1,2,3,4,5,6,\*</sup> 

- <sup>1</sup> State Key Laboratory of Herbage Improvement and Grassland Agro-Ecosystems, Lanzhou University, Lanzhou 730000, China; 220220902301@lzu.edu.cn (Y.G.); malik@lzu.edu.cn (K.M.)
  - <sup>2</sup> Key Laboratory of Grassland Livestock Industry Innovation, Ministry of Agriculture and Rural Affairs, Lanzhou University, Lanzhou 730000, China
  - <sup>3</sup> College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou 730000, China
  - <sup>4</sup> Engineering Research Center of Grassland Industry, Ministry of Education, Lanzhou University, Lanzhou 730000, China
  - <sup>5</sup> Gansu Tech Innovation Centre of Western China Grassland Industry, Lanzhou University, Lanzhou 730000, China
  - <sup>6</sup> Centre for Grassland Microbiome, Lanzhou University, Lanzhou 730000, China
  - <sup>7</sup> Qinghai Academy of Animal and Veterinary Science, Qinghai University, Xining 810016, China; xiuzhang11@163.com
- \* Correspondence: chenchenjiang@lzu.edu.cn (Z.C.); chunjie@lzu.edu.cn (C.L.)

**Abstract:** The present study provides a detailed analysis of the chloroplast genome of *Microula sikkimensis*. The genome consisted of a total of 149,428 bp and four distinct regions, including a large single-copy region (81,329 bp), a small single-copy region (17,261 bp), and an inverted repeat region (25,419 bp). The genome contained 112 genes, including 78 protein-coding genes, 30 tRNA genes, and 4 rRNA genes, and some exhibited duplication in the inverted repeat region. The chloroplast genome displayed different GC content across regions, with the inverted repeat region exhibiting the highest. Codon usage analysis and the identification of simple sequence repeats (SSRs) offer valuable genetic markers. Comparative analysis with other Boraginaceae species highlighted conservation and diversity in coding and noncoding regions. Phylogenetic analysis placed *M. sikkimensis* within the Boraginaceae family, revealing its distinct relationship with specific species.

**Keywords:** *Microula sikkimensis*; chloroplast genomes; comparative analysis; phylogenetic analysis; Boraginaceae



**Citation:** Gao, Y.; Chen, Z.; Li, X.; Malik, K.; Li, C. Comparative Analyses of Complete Chloroplast Genomes of *Microula sikkimensis* and Related Species of Boraginaceae. *Genes* **2024**, *15*, 226. <https://doi.org/10.3390/genes15020226>

Academic Editors: Lassaad Belbahri, Chao Shi and Shuo Wang

Received: 23 December 2023

Revised: 5 February 2024

Accepted: 7 February 2024

Published: 10 February 2024



**Copyright:** © 2024 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) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

*Microula sikkimensis*, a biennial herbaceous plant from the Boraginaceae family, thrives in the high-altitude grasslands, forests, shrublands, and secondary vegetation at elevations ranging from 2500 to 4000 m on the eastern edge of the Qinghai-Tibet Plateau [1]. *M. sikkimensis* is a plant resource rich in  $\gamma$ -linolenic acid, with promising development prospects. Abundant experiments have confirmed that *M. sikkimensis* oil significantly reduces the levels of total cholesterol (TC), triglycerides (TG), and serum malondialdehyde (MDA) in the liver and serum. It also increases the ratio of high-density lipoprotein cholesterol to total cholesterol (HDL-C/TC) in the serum. Moreover, it effectively reduces the deposition of cholesterol in peripheral tissue cells, thus preventing atherosclerosis and maintaining the integrity of the biomembrane structure. Its capacity to lower triglycerides in the liver and serum surpasses that of atorvastatin. *M. sikkimensis* oil can improve the high blood lipidemia by reducing blood viscosity, preventing thrombosis, and exhibiting unique solvent properties [1–3]. The harvested stalks of *M. sikkimensis* exhibit heightened palatability and are nutritionally dense, making them a crucial coarse fodder for supplementing livestock in the winter and spring seasons in high-altitude pastoral regions. The research of *M. sikkimensis* oil will provide enough supply of raw materials for the production of a

series of high-nutrition health foods, fortified dairy products, healthful edible oils, as well as specialized pharmaceuticals and novel cosmetics [4].

Chloroplasts are specialized energy converters unique to higher plants and certain algae. These organelles are essential for their role in carrying autonomous genetic information, necessary for cellular functions [5–7]. The plant chloroplast genome is generally a double-stranded circular molecule composed of four main regions: a large single copy (LSC) region, a small single copy (SSC) region, and two inverted repeat (IRA and IRB) regions, which are identical in sequence but opposite in orientation [8–11]. The relatively small (115–165 kb) and conservative nature of the chloroplast genome makes it a crucial tool for the exploration of genomic evolution and the analysis of phylogenetic relationships within angiosperms [12–14]. Additionally, the chloroplast genome finds widespread applications in diverse domains, including population genetics, molecular-assisted breeding, gene mapping, plant barcode sequence screening and gene diversity studies [15–19].

Currently, research findings on the chloroplast genome of *M. sikkimensis* remain unpublished. In this research, we presented the first report and analysis of the complete chloroplast genome sequence of *M. sikkimensis* by acquiring insights into the fundamental genome structure, simple sequence repeats (SSRs), and codon usage bias. Additionally, comparative genomic analysis and phylogenetic analysis of *M. sikkimensis* chloroplast genome was conducted in relation to other species within the Boraginaceae family. These research findings not only fill the existing gap in scientific research regarding the chloroplast genome of *M. sikkimensis* but also lay a foundational stone for future genetic and evolutionary studies.

## 2. Materials and Methods

### 2.1. Sample Collection and DNA Extraction

The *M. sikkimensis* plants were collected from Zaduo county, Qinghai province, China. The fresh leaves were frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . The cetyltrimethylammonium bromide (CTAB) method was used to extract total genomic DNA. The quality of DNA was measured by NanoDrop 2000 (Thermo Scientific, Wilmington, NC, USA) and agarose gel electrophoresis.

### 2.2. Genome Sequencing, Assembly and Annotation

Library construction was carried out by NexteraXT DNA Library Preparation Kit (Illumina, Shanghai, China), with 300 bp-short-insert fragments. Library quality was assessed using GeneRead DNA QuantiMIZE Assay Kit (Qiagen, Duesseldorf, Germany). Sequencing of qualified library was conducted utilizing Illumina NovaSeq 6000 (Wuhan Benagen Tech Solutions Co., Ltd., Wuhan, China). Initial raw sequencing output was subjected to quality control using SOAPnuke (Version: 2.1.0), with filtering parameters set to remove reads with more than 5% N base content, reads where the number of low-quality bases (quality score less than or equal to 5) reaches 50%, reads contaminated with adapters and duplicate sequences caused by PCR amplification. GetOrganelle software (version 1.7.7.0) [20] was utilized for assembling the chloroplast genome with the default settings. CPGAVAS software (v2) [21] was used to annotate the chloroplast genome while OGDRAW software (v1.1.1) [22] was used to visualize the circular chloroplast genome map of *M. sikkimensis*. The tRNA of the chloroplast genome was annotated using tRNAscanSE software (v2.0.11) [23]. The rRNA of the chloroplast genome was annotated using BLASTN software (v2.13.0) [24]. The annotation errors of each chloroplast genome were manually corrected using online tool CPGView [25] and Apollo software (v1.11.8) [26]. The fully annotated chloroplast genome was subsequently submitted to the GenBank database (Accession Number: OR866440).

### 2.3. Comparative Genome Analysis

The complete chloroplast genomes of *M. sikkimensis* and other four species were compared using the MVISTA program [27] with the shuffle-LAGAN model, with *M. sikkimensis*

as the reference. The IRSCOPE program [28] was applied to analyze the LSC, IR and LSC boundary locations in five Boraginaceae species complete chloroplast genomes.

The complete chloroplast genomes of the five Boraginaceae plants were multiple aligned by MAFFT (v7) [29], and nucleotide variations (Pi) was calculated by DnaSP (v6) [30] with the following parameters: window length, 600; and step size, 200. The protein-coding sequences of genomes were extracted using Phylosuite software (v1.1.16) [31]. The codon preference of protein-coding genes in chloroplast genome was analyzed and the RSCU value was calculated using Mega software (v7.0) [32].

#### 2.4. Repeat Sequence Analysis

MISA (v2.1) (<https://webblast.ipk-gatersleben.de/misa/>) (accessed on 4 May 2023) [33], TRF (v4.09) (<https://tandem.bu.edu/trf/trf.unix.help.html>) (accessed on 4 May 2023) [34] and REPuter web server (<https://bibiserv.cebitec.uni-bielefeld.de/reputer/>) (accessed on 4 May 2023) [35] were used to identify the repeat sequences including microsatellite sequence repeats, tandem repeats and scattered repeats. The results were visualized using Excel (2021) software (v2207).

#### 2.5. Phylogenetic Analysis

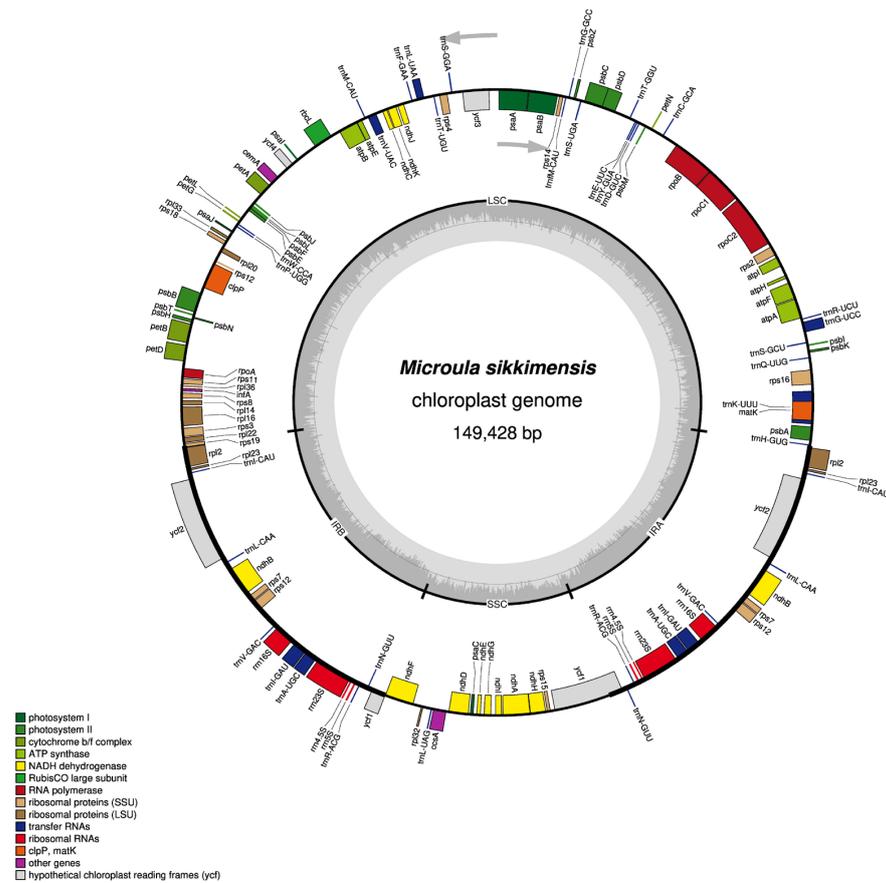
The complete chloroplast genomes of 30 species were downloaded from the National Center for Biotechnology Information (NCBI). *Isodon serra* and *Forsythia suspensa* were chosen as the outgroups. The detailed list of all the species along with their respective accession numbers for the chloroplast genomes available in the NCBI database can be found in Table S1. All the sequences were aligned using MAFFT with the default parameters. Ambiguously aligned fragments were filtered by GBLOCKS 0.91b [36], and the parameter settings were as follows: minimum number of sequences for a conserved position, 20; minimum number of sequences for a flank position, 20; maximum number of contiguous nonconserved positions, 6; minimum length of a block, 11; and allowed gap positions, 0. Based on the result of alignment, a phylogenetic tree was constructed using IQ-TREE 2 [37] with 5000 bootstraps. Bayesian inference (BI) phylogenies were conducted through MR-BAYES 3.2.0 [38], employing the GTR+I+G model across eight parallel runs for a total of 2,000,000 generations. We discarded the initial 25% of the sampled data as burn-in. For visualization, the tree files were uploaded to the online tool iTOL (v6) [39].

### 3. Results

#### 3.1. Chloroplast Genome Assembly and Genome Features

In *M. sikkimensis*, the chloroplast genome spans a total of 149,428 bp, characterized by a conventional quadripartite circular structure, which is divided into four specific regions: the LSC region (81,329 bp), the SSC region (17,261 bp), and two identical IR regions (25,419 bp of each), which separated SSC and LSC (Figure 1).

In the chloroplast genome of *M. sikkimensis*, gene annotation efforts revealed the presence of 112 genes in total, including 78 protein-coding genes, 30 transfer RNA (tRNA) genes, and 4 ribosomal RNA (rRNA) genes. The protein-coding genes could be grouped into 15 distinct gene families, including 16 photosystem II genes, 12 ribosome small genes, 11 NADH dehydrogenase genes, 9 ribosome major genes, 6 cytochrome b/f complex genes, 6 ATP synthase genes, 5 photosystem I genes, 4 DNA-dependent RNA polymerase genes, 3 conserved open reading frame genes, 1 1, 5-diphosphate ribulose carboxylase/oxygenase large subunit gene, 1 mature enzyme gene, 1 membrane protein gene, 1 protease gene, 1 C-type cytochrome synthesis gene, 1 translation initiation factor (Table 1). Among the 112 genes, 7 protein-coding genes, 7 tRNA genes, and 4 rRNA genes were duplicated in IR regions (Figure 1).



**Figure 1.** Chloroplast genome map of *M. sikkimensis*. Genes inside the circle are transcribed clockwise, while those outside the circle are transcribed anticlockwise. The large single copy (LSC) region, inverted repeat (IRA, IRB) regions and small single copy (SSC) region are shown in the figure. The darker gray in the inner circle shows the GC content, while the lighter gray shows the AT content. Genes with different functions are represented by different colors.

**Table 1.** List of genes annotated in the chloroplast genomes of *M. sikkimensis*.

Group of Genes	Name of Genes	Number of Genes
NADH-dehydrogenase	<i>ndhA, ndhB</i> (×2), <i>ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	11
photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>	5
photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ, ycf3</i>	16
cytochrome b/f complex	<i>petA, petB, petD, petG, petL, petN</i>	6
ATP synthase	<i>atpA, atpB, atpE, atpF, atpH, atpI</i>	6
Large subunit of rubisco	<i>rbcL</i>	1
Small subunit of ribosome	<i>rps2, rps3, rps4, rps7</i> (×2), <i>rps8, rps11, rps12</i> (×2), <i>rps14, rps15, rps16, rps18, rps19</i>	12
Large subunit of ribosome	<i>rpl2</i> (×2), <i>rpl14, rpl16, rpl20, rpl22, rpl23</i> (×2), <i>rpl32, rpl33, rpl36</i>	9
DNA dependent RNA polymerase	<i>rpoA, rpoB, rpoC1, rpoC2</i>	4
rRNA genes	<i>rrn4.5S</i> (×2), <i>rrn5S</i> (×2), <i>rrn16S</i> (×2), <i>rrn23S</i> (×2)	4
tRNA genes	30 tRNAs	30
Maturase	<i>matK</i>	1
Envelope membrane protein	<i>cemA</i>	1
Protease	<i>clpP</i>	1
c-type cytochrom synthesis gene	<i>ccsA</i>	1
Translation al initiation factor	<i>infA</i>	1
Genes of unknown functions Open Reading	<i>ycf1</i> (×2), <i>ycf2</i> (×2), <i>ycf4</i>	3

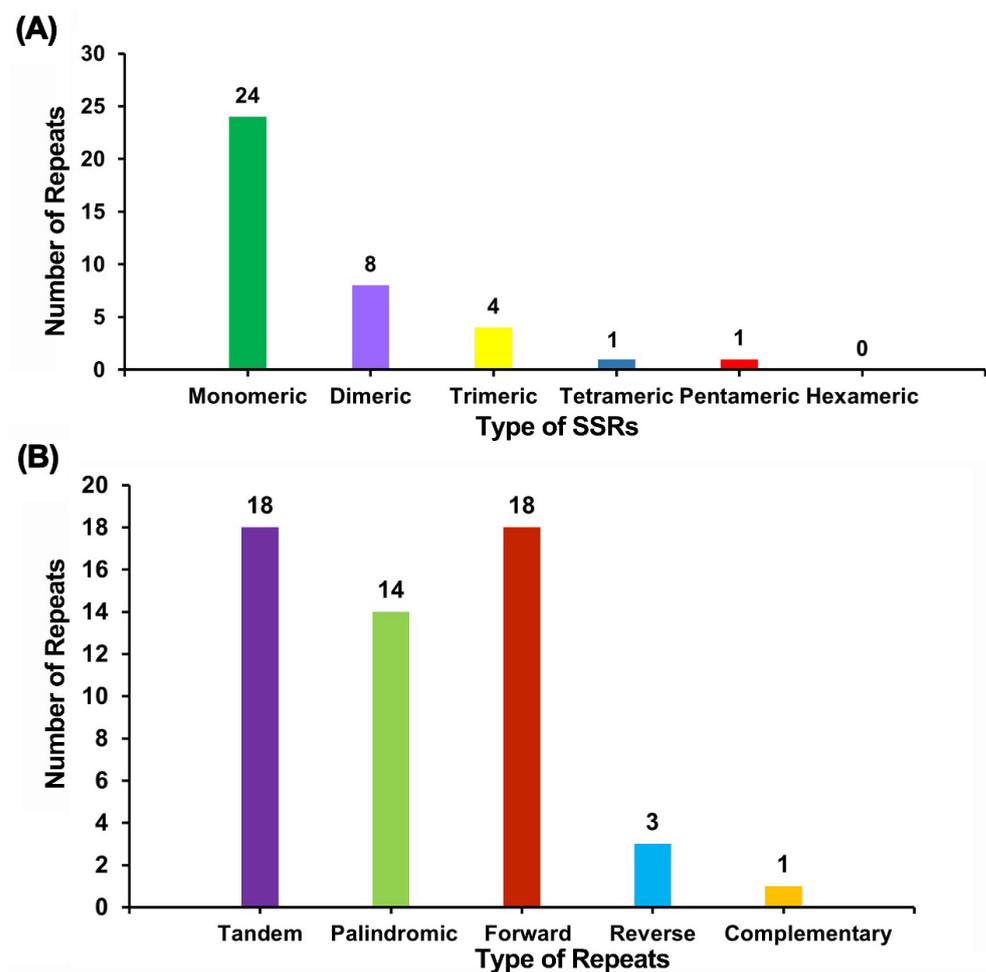
The chloroplast genome of *M. sikkimensis* exhibited a GC content of 37.51%. Notably, the GC content varies considerably across the different regions within the chloroplast



### 3.3. Repeat Sequences and SSR Analysis

In the comprehensive analysis of the *M. sikkimensis* chloroplast genome, 38 SSRs were detected, distributed as follows: 26 located in the LSC region, 11 located in the SSC region, and 1 located in the IR region. Mononucleotide repeats consisted of 10–12 repetitions, dinucleotide repeats had 5–8 repetitions, trinucleotide repeats had 4 repetitions, while tetra- and pentanucleotide repeats had 3 repetitions. The analysis revealed a predominance of SSRs characterized by A/T motifs over those with G/C motifs. Specifically, A/T mononucleotide repeats were identified as the most prevalent type ( $n = 22$ ), with AT dinucleotide-containing repeats emerging as the second most common ( $n = 8$ ) (Table S4).

Tandem repeats, often referred to as satellite DNA, consist of core sequences ranging from 7 to 200 bases in length, which are repetitively aligned in the sequence. Such repeats exist widely in eukaryotic and prokaryotes genomes. Within the chloroplast genome of *M. sikkimensis*, a total of 18 tandem repeats have been identified, each of which exhibited a similarity of over 74% and lengths ranging from 8 to 24 bp (Figure 3).

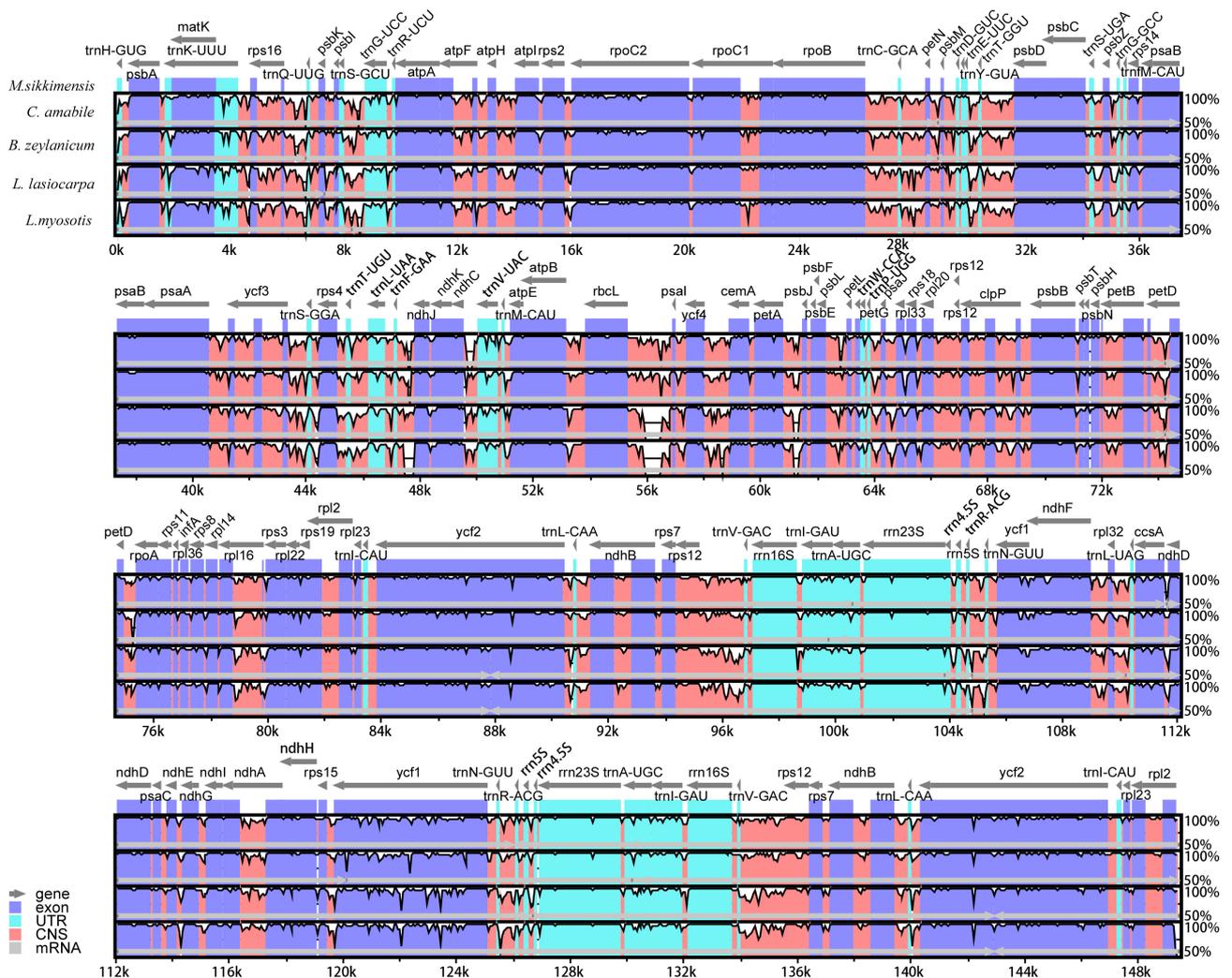


**Figure 3.** Repeat sequence and SSR analysis of *M. sikkimensis* chloroplast genome. **(A)** The horizontal coordinate represents the type of SSRs, the vertical coordinate represents the number of repeats, the green represents monomer SSRs, the purple represents dimer SSRs, the yellow represents trimer SSRs, the blue represents tetramer SSRs, and the red represents pentamer SSRs. No hexamer SSRs were detected in the chloroplast genome. **(B)** The horizontal coordinate indicates the type of repeat sequence, the vertical coordinate indicates the number of repeat segments, purple indicates tandem repeats, green indicates palindromic repeats, red indicates forward repeats, blue indicates reverse repeats, and yellow indicates complementary repeats.

The analysis of dispersed repeat sequences within the *M. sikkimensis* chloroplast genome, conducted by REPuter, identified 36 pairs of repeats each with a minimum length of 30 base pairs. Among these pairs, 14 were identified as palindromic repeats, 18 as forward repeats, 3 as reverse repeats, and 1 as complementary repeat. Notably, the maximum length recorded for palindromic repeats was 46 bp, and for forward repeats, it extended to 52 bp (Figure 3).

### 3.4. Comparative Genome Analysis

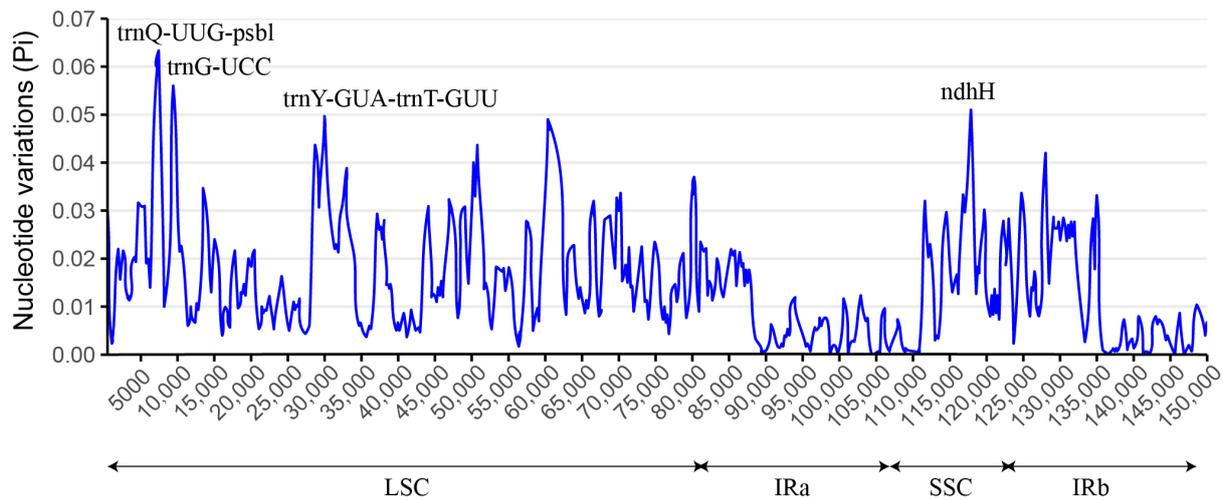
Comparative analysis of the chloroplast genomes from five Boraginaceae species was conducted using the mVISTA tool, with the *M. sikkimensis* genome serving as the reference sequence. As a result, these five genomes were basically identical in coding regions, whereas were more diverse in noncoding regions. Significant variation was observed in the intergenic spacers of the chloroplast genomes, notably within regions such as *matK-rps16*, *rps16-trnQ-UUG*, *trnS-GCU*, *trnF-GAA-ndhJ*, *rbcl-psaI*, *ycf4-cemA*, and *petA-psbJ* in the LSC region, along with *ccsA-ndhD* and *rps15-ycf1* in the SSC region (Figure 4). These regions may serve as potential molecular markers for the identification of Boraginaceae species.



**Figure 4.** Sequence alignment of five Boraginaceae genomes in MVISTA. The grey arrows above the alignment indicate the genes transcription directions. The Y-axis represents identity, ranging from 50% to 100%.

To evaluate the extent of genetic diversity among chloroplast genomes from five Boraginaceae species, alignments were generated, and nucleotide variability ( $\Pi$ ) was

calculated using the DnaSP software (v6). The findings, illustrated in Figure 5, indicated that Pi values varied from 0 to 0.062, with a mean of approximately 0.013. Moreover, analysis pointed out four loci with high Pi values ( $\geq 0.05$ ) as mutational hotspots within these species, including 1 protein coding gene (*ndhH*), 1 tRNA coding gene (*trnG-UCC*) and 2 intergenic regions (*trnQ-UUG-psbI* and *trnY-GUA-trnT-GUU*). These identified regions could serve as potential molecular markers for Boraginaceae species. Additionally, the region with the minimum Pi value was found within the IR region, indicating their significant conservation across the Boraginaceae family.



**Figure 5.** Nucleotide diversity (Pi) analysis for chloroplast genomes from the Boraginaceae plants. (window length: 600 bp; step size: 200 bp).

The dynamics of the IR region, through its expansion and contraction, played a crucial role in plant evolutionary processes. These changes could induce structural modifications within the chloroplast genome, potentially influencing the expression and functionality of genes within the chloroplast genes [10,42,46]. The comparative analysis was performed on the boundaries of the chloroplast genome regions across five species: *M. sikkimensis*, *Bothriospermum zeylanicum*, *Trigonotis zhuokejiensis*, *Trigonotis tibetica*, and *Cynoglossum amabile*. The results revealed variations in the chloroplast genome sizes among the five examined species, with lengths extending from 148,193 bp (*T. tibetica*) to 152,532 bp (*C. amabile*). Specifically, the LSC regions displayed lengths from 80,767 bp (*T. tibetica*) to 83,692 bp (*B. zeylanicum*). The lengths of SSC regions were observed to range from 17,181 bp (*B. zeylanicum*) to 17,366 bp (*C. amabile*). In addition, the IR regions exhibited lengths from 25,088 bp (*T. tibetica*) to 25,632 bp (*C. amabile*). Across all examined species, the boundaries of the LSC-IRb and SSC-IRa regions were identified within the *rps19* and *ycf1* genes, respectively. Notably, the junction between the SSC and IRb regions occurred within the *ycf1* gene across most species. However, an exception was observed in *B. zeylanicum*, where the *ycf1* gene was absent near the IRb-SSC junction. Instead, the *ndhF* gene, located entirely within the SSC region, was found to be next to the IRb/SSC boundary, with only a 1 bp distance. In *T. zhuokejiensis* and *T. tibetica* chloroplast genomes, the boundary between the IRb and SSC regions was situated within the *ndhF* gene, which had a 2 bp insertion in IRb. Additionally, genes such as *rpl22*, *rpl2*, and *trnH* were also identified at the boundaries of LSC/IR and SSC/IR across the five species examined (Figure 6).

### Inverted Repeats

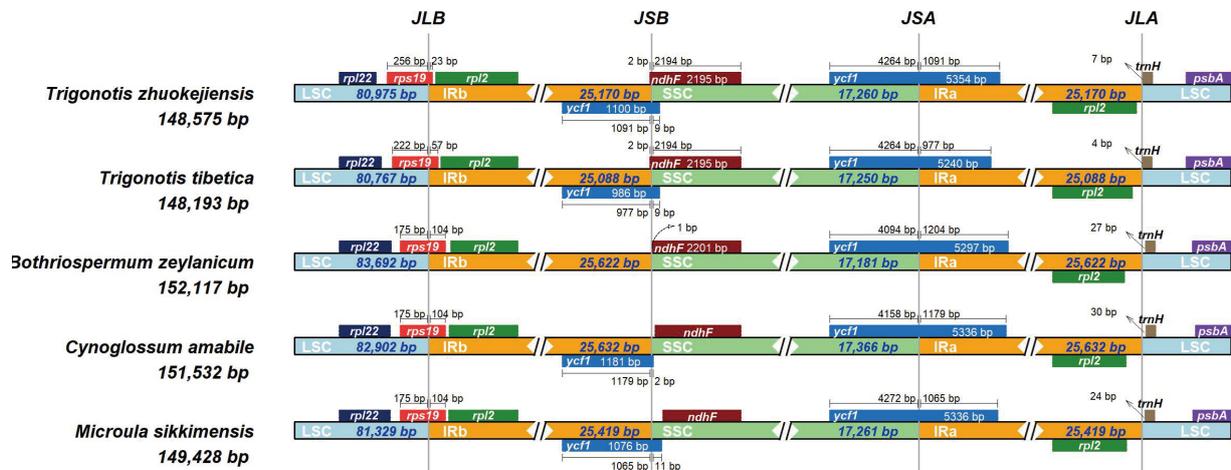


Figure 6. Comparison of the junction positions between the LSC, SSC and IR regions among the chloroplast genomes of five species.

### 3.5. Phylogenetic Analysis

Phylogenetic analyses employing ML and BI approaches were conducted on the chloroplast genome sequences of 30 Boraginaceae species, utilizing *I. serra* and *F. suspensa* as the outgroups. As shown in Figure 7, the generated phylogenetic trees demonstrated high support for most clades. This comprehensive analysis led to the categorization of the examined Boraginaceae species into two distinct subfamilies: Cynoglossoideae and Boraginoideae.

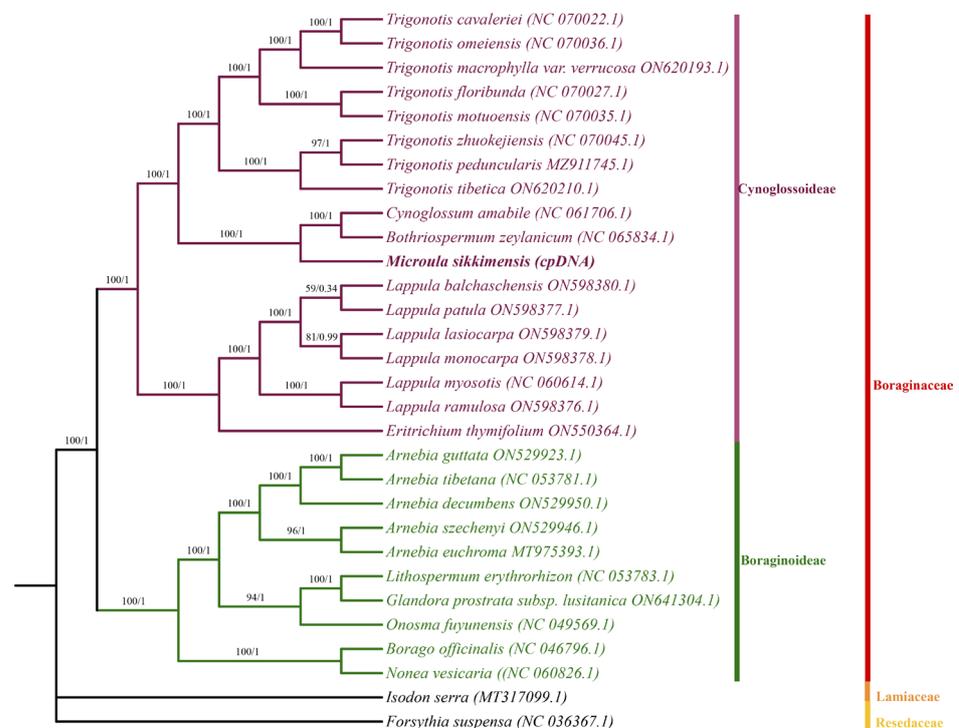


Figure 7. Phylogenetic tree reconstructed based on the complete chloroplast genome sequences from 30 species using ML method. The numerical annotations above the branches represented the ML bootstrap support values/BI probability support values.

*M. sikkimensis* displayed a sister relationship with *C. amabile* and *B. zeylanicum*, but failed to cluster into the same branch, indicating significant distinctions in their chloroplast genomes, albeit with high similarity. This study had provided crucial genetic data that were essential for elucidating the phylogenetic relationships within the Boraginaceae family.

#### 4. Discussion

This study had for the first time successfully sequenced the complete chloroplast genome of *M. sikkimensis*, presenting its genetic information. The analysis demonstrated that the *M. sikkimensis* chloroplast genome spanned a length of 149,428 bp, and it featured the circular quadripartite typical of chloroplast genomes, consisting of the LSC, SSC, and a pair of IR regions. Gene annotation indicated that the genome contains 112 genes, which include 78 protein-coding genes, 30 tRNA genes, and 4 rRNA genes, which is consistent with the genomic structure and gene quantity observed in the chloroplasts of other species within the Boraginaceae family [47–51]. Introns were crucial for the regulatory mechanisms that control gene expression. They influenced the synthesis and function of proteins through splicing regulation, thus significantly impacting the development, growth and environment adaptation of organisms [52]. In the *M. sikkimensis* chloroplast genome, our analysis identified 12 genes containing introns, 10 of these genes possess a single intron, while 2 contain a pair of introns. GC content was one of the important characteristics of the nucleic acid sequence composition, exhibiting variation across different species. This variability influenced the distribution, adaptability to environmental conditions, and lifestyles of species [53]. The chloroplast genome of *M. sikkimensis* had an overall GC content of 37.51%. Within this genome, the GC content varied considerably across different regions, reflecting the diversity in structural and functional demands of these regions. The highest GC content was found in IR regions at 43.13%, suggesting a possible role in enhancing the stability of the genome structure or in specific functional aspects related to gene expression.

SSRs, as a crucial category of molecular genetic markers, have found extensive applications across various domains of biological research. They served as important tools applied in genetic relationships, population structures, and evolutionary processes among species [54–56]. In plant populations, SSRs could serve as highly effective markers, facilitating the examination of genetic diversity within closely related taxa. Here, we found 38 SSRs in the chloroplast genome in *M. sikkimensis*, including the following five types of SSRs: mononucleotide, dinucleotide, trinucleotide, tetranucleotide, and pentanucleotide. A/T mononucleotide repeats were the most frequent SSR, followed by AT dinucleotide repeats, that were consistent with the previous reports [57,58]. These SSR markers provided insights for genetic diversity studies and conservation strategies in the Boraginaceae.

Research on codon usage bias could help explain the intricacies of gene expression and the mechanisms of translation. The preference for certain synonymous codons was associated with the number of introns and exhibited variability across different exons. This variability in codon usage patterns is distinct among various species. Furthermore, evidence suggests that DNA methylation plays a significant role in influencing synonymous codon usage bias [59]. Additionally, the GC content in codons is recognized as a key factor in the development of codon usage bias [60]. Within the analysis, it was observed that out of the 30 codons RSCU values above 1, 29 concluded with either A or U. Conversely, of the 32 codons showing RSCU values below 1, 29 concluded with G or C. These findings suggest a codon preference in *M. sikkimensis* for ending with A or U at the third base, a pattern also observed in species such as *Fagopyrum dibotrys* and *Salix wilsonii* [61,62]. This preference for A or T in the codon composition could stem from evolutionary pressures and genetic alterations. Within the plant chloroplast genome, it was observed that preferred synonymous codons predominantly terminated with A or U. This trend might be attributed to a higher content of A and T bases, resulting in an obvious bias for A or T ending codons [63].

By comparing the chloroplast genomes from various species, we could enhance our understanding of the evolutionary relationships and genomic structural disparities among these species [64,65]. Comparative analysis of the chloroplast genomes across five Boraginaceae species revealed notable similarities within their coding sequences alongside pronounced differences in their non-coding regions. Such variability in the non-coding regions presents their potential for use as molecular markers in future Boraginaceae investigations, providing crucial insights for species classification and phylogenetic analyses. Additionally, phylogenetic analysis revealed that *M. sikkimensis* had a sister relationship with *C. amabile* and *B. zeylanicum*. However, they failed to cluster into the same branch, indicating that they were highly similar in chloroplast genomes, but there were still discernible differences. It could be inferred that certain regions within the chloroplast genomes of these three species have experienced divergent evolutionary paths, resulting in slight but observable genetic variations.

## 5. Conclusions

This study presented a comprehensive analysis of the chloroplast genome of *M. sikkimensis*, shedding light on its genetic characteristics and evolutionary context within the Boraginaceae family.

The chloroplast genome, spanning over 149,428 bp, exhibited the characteristic circular quadripartite structure common to higher plants. The gene annotation identified a total of 112 genes, consistent with the structure observed in related Boraginaceae species. The identification of SSRs, with a predominance of A/T mononucleotide repeats followed by AT dinucleotide repeats, provided valuable genetic markers for exploring genetic diversity within closely related plant populations. Comparative analysis with other species chloroplast genomes revealed both conservation and variability, particularly in non-coding regions. These highly variable regions served as crucial molecular markers for species classification and phylogenetic research within the Boraginaceae family. Furthermore, phylogenetic analysis placed *M. sikkimensis* in a distinct relationship with *C. amabile* and *B. zeylanicum*, indicating a high degree of similarity in their chloroplast genomes. However, discernible differences existed, suggesting subtle genomic distinctions that endorse further investigation.

Overall, this study not only provided a comprehensive understanding of the chloroplast genome of *M. sikkimensis* but also established a valuable genetic resource for future research in phylogenetics, molecular marker development, and conservation strategies among the Boraginaceae family. These findings could contribute to the broader field of plant genomics and facilitate a deeper appreciation of the evolutionary dynamics within this plant family.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes15020226/s1>, Table S1 All information of species and the accession numbers of their chloroplast genomes in NCBI; Table S2 The intron-containing genes in the chloroplast genomes of *M. sikkimensis*; Table S3 Sequences of protein-coding genes of *M. sikkimensis* chloroplast genome; Table S4 repeat sequences in the chloroplast genome of *M. sikkimensis*.

**Author Contributions:** C.L. conceived and designed the experiments. Z.C., Y.G. and X.L. sampled plant specimens and conducted experiments. Y.G. performed data analyses. Y.G. wrote the manuscript. Z.C. and K.M. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Major Science and Technology Project of Qinghai Province (2021-SF-A4); Chinese Academy of Sciences–People’s Government of Qinghai Province on Sanjiangyuan National Park (LHZX-2022-01); Gansu Province Grassland Monitoring and Evaluation Technology Support Project of Gansu Province Forestry and Grassland Administration (2021794).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are openly available in NCBI (GenBank accession number: OR866440).

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

## References

1. Han, F.; Cheng, D.Z.; Shi, S.B.; Ran, F.; Li, Y.K.; Bao, S.K.; Ren, F.; Shi, L.N.; Han, Q. The research process of a high—Quality wild resource—*Microula sikkimensis*. *Chin. Wild Plant Resour.* **2007**, *26*, 5–9. (In Chinese) [[CrossRef](#)]
2. Zheng, S.Z.; Yang, H.P.; Meng, J.C.; Ma, X.M.; Shen, X.W. Studies on the constituents from the seeds of *M. sikkimensis*. *J. Northwest Norm. Univ. (Nat. Sci.)* **2003**, *39*, 54–57. (In Chinese) [[CrossRef](#)]
3. Wu, L.P. Exploitation and research of *Microula sikkimensis* resources. *China Oils Fats* **1994**, *19*, 41–42. (In Chinese)
4. Luo, G.R.; Deng, Y.C. Experiment on feeding *Microula sikkimensis* straw to Tibetan sheep. *J. Grass Forage Sci.* **2001**, *2*, 56–57. (In Chinese) [[CrossRef](#)]
5. Szabò, I.; Spetea, C. Impact of the ion transportome of chloroplasts on the optimization of photosynthesis. *J. Exp. Bot.* **2017**, *68*, 3115–3128. [[CrossRef](#)] [[PubMed](#)]
6. Mullineaux, P.M.; Exposito-Rodriguez, M.; Laissue, P.P.; Smirnov, N. ROS-dependent signalling pathways in plants and algae exposed to high light: Comparisons with other eukaryotes. *Free Radic. Biol. Med.* **2018**, *122*, 52–64. [[CrossRef](#)] [[PubMed](#)]
7. Pollari, M.; Ruotsalainen, V.; Rantamäki, S.; Tyystjärvi, E.; Tyystjärvi, T. Simultaneous inactivation of sigma factors B and D interferes with light acclimation of the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* **2009**, *191*, 3992–4001. [[CrossRef](#)] [[PubMed](#)]
8. Yang, X.; Xie, D.F.; Chen, J.P.; Zhou, S.D.; Yu, Y.; He, X.J. Comparative Analysis of the Complete Chloroplast Genomes in *Allium* Subgenus *Cyathophora* (Amaryllidaceae): Phylogenetic Relationship and Adaptive Evolution. *BioMed Res. Int.* **2020**, *2020*, 1732586. [[CrossRef](#)] [[PubMed](#)]
9. Yang, J.B.; Tang, M.; Li, H.T.; Zhang, Z.R.; Li, D.Z. Complete chloroplast genome of the genus *Cymbidium*: Lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evol. Biol.* **2013**, *13*, 84. [[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](#)]
11. Kim, G.B.; Lim, C.E.; Kim, J.S.; Kim, K.; Lee, J.H.; Yu, H.J.; Mun, J.H. Comparative chloroplast genome analysis of *Artemisia* (Asteraceae) in East Asia: Insights into evolutionary divergence and phylogenomic implications. *BMC Genom.* **2020**, *21*, 415. [[CrossRef](#)]
12. Olmstead, R.G.; Palmer, J.D. Chloroplast DNA systematics: A review of methods and data analysis. *Am. J. Bot.* **1994**, *81*, 1205–1224. [[CrossRef](#)]
13. Wang, A.; Wu, H.; Zhu, X.; Lin, J. Species Identification of *Conyza bonariensis* Assisted by Chloroplast Genome Sequencing. *Front. Genet.* **2018**, *9*, 374. [[CrossRef](#)]
14. Kelchner, S. The Evolution of Non-Coding Chloroplast DNA and Its Application in Plant Systematics. *Ann. Mo. Bot. Gard.* **2000**, *87*, 482–498. [[CrossRef](#)]
15. Tang, J.; Xia, H.; Cao, M.; Zhang, X.; Zeng, W.; Hu, S.; Tong, W.; Wang, J.; Wang, J.; Yu, J.; et al. A comparison of rice chloroplast genomes. *Plant Physiol.* **2004**, *135*, 412–420. [[CrossRef](#)] [[PubMed](#)]
16. Li, J.; Tang, J.; Zeng, S.; Han, F.; Yuan, J.; Yu, J. Comparative plastid genomics of four *Pilea* (Urticaceae) species: Insight into interspecific plastid genome diversity in *Pilea*. *BMC Plant Biol.* **2021**, *21*, 25. [[CrossRef](#)] [[PubMed](#)]
17. Sasaki, C.; Lee, S.B.; Daniell, H.; Wood, T.C.; Tomkins, J.; Kim, H.G.; Jansen, R.K. Complete chloroplast genome sequence of *Gycine max* and comparative analyses with other legume genomes. *Plant Mol. Biol.* **2005**, *59*, 309–322. [[CrossRef](#)] [[PubMed](#)]
18. Du, Q.; Li, J.; Wang, L.; Chen, H.; Jiang, M.; Chen, Z.; Jiang, C.; Gao, H.; Wang, B.; Liu, C. Complete chloroplast genomes of two medicinal *Swertia* species: The comparative evolutionary analysis of *Swertia* genus in the Gentianaceae family. *Planta* **2022**, *256*, 73. [[CrossRef](#)] [[PubMed](#)]
19. Daniell, H.; Lee, S.-B.; Grevich, J.; Sasaki, C.; Quesada-Vargas, T.; Guda, C.; Tomkins, J.; Jansen, R.K. Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. *Theor. Appl. Genet.* **2006**, *112*, 1503–1518. [[CrossRef](#)] [[PubMed](#)]
20. Jin, J.J.; Yu, W.B.; Yang, J.B.; Song, Y.; dePamphilis, C.W.; Yi, T.S.; Li, D.Z. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **2020**, *21*, 241. [[CrossRef](#)]
21. Shi, L.; Chen, H.; Jiang, M.; Wang, L.; Wu, X.; Huang, L.; Liu, C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* **2019**, *47*, W65–W73. [[CrossRef](#)] [[PubMed](#)]
22. Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **2019**, *47*, W59–W64. [[CrossRef](#)] [[PubMed](#)]
23. Lowe, T.M.; Eddy, S.R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **1997**, *25*, 955–964. [[CrossRef](#)]
24. Chen, Y.; Ye, W.; Zhang, Y.; Xu, Y. High speed BLASTN: An accelerated MegaBLAST search tool. *Nucleic Acids Res.* **2015**, *43*, 7762–7768. [[CrossRef](#)] [[PubMed](#)]

25. Liu, S.; Ni, Y.; Li, J.; Zhang, X.; Yang, H.; Chen, H.; Liu, C. CPGView: A package for visualizing detailed chloroplast genome structures. *Mol. Ecol. Resour.* **2023**, *23*, 694–704. [[CrossRef](#)] [[PubMed](#)]
26. Lewis, S.E.; Searle, S.M.; Harris, N.; Gibson, M.; Lyer, V.; Richter, J.; Wiel, C.; Bayraktaroglu, L.; Birney, E.; Crosby, M.A.; et al. Apollo: A sequence annotation editor. *Genome Biol.* **2002**, *3*, 1–14. [[CrossRef](#)]
27. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, W273–W279. [[CrossRef](#)]
28. Amiryousefi, A.; Hyvönen, J.; Poczai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [[CrossRef](#)]
29. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]
30. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sánchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [[CrossRef](#)]
31. Zhang, D.; Gao, F.; Jakovlić, I.; Zou, H.; Zhang, J.; Li, W.X.; Wang, G.T. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* **2020**, *20*, 348–355. [[CrossRef](#)] [[PubMed](#)]
32. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)] [[PubMed](#)]
33. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [[CrossRef](#)] [[PubMed](#)]
34. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)] [[PubMed](#)]
35. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [[CrossRef](#)]
36. Talavera, G.; Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **2007**, *56*, 564–577. [[CrossRef](#)]
37. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)]
38. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [[CrossRef](#)]
39. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [[CrossRef](#)]
40. Yi, D.K.; Kim, K.J. Two complete chloroplast genome sequences of genus *Paulownia* (Paulowniaceae): *Paulownia coreana* and *P. tomentosa*. *Mitochondrial DNA Part B Resour.* **2016**, *1*, 627–629. [[CrossRef](#)] [[PubMed](#)]
41. Zheng, L.P.; Li, L.J. Characterization of the complete chloroplast genome of *Centranthera grandiflora* Benth. (Orobanchaceae), an important species of medicinal herb. *Mitochondrial DNA Part B Resour.* **2021**, *6*, 1784–1785. [[CrossRef](#)]
42. Shen, X.; Wu, M.; Liao, B.; Liu, Z.; Bai, R.; Xiao, S.; Li, X.; Zhang, B.; Xu, J.; Chen, S. Complete Chloroplast Genome Sequence and Phylogenetic Analysis of the Medicinal Plant *Artemisia annua*. *Molecules* **2017**, *22*, 1330. [[CrossRef](#)]
43. Zhang, F.; Chen, H.; Zhou, Y.; Li, N.; Chong, X.; Li, Y.; Lu, X.; Wang, C. The complete chloroplast genome sequence and phylogenetic analysis of *Ilex 'Beryl'*, a hybrid of *Ilex cornuta* × *Ilex latifolia* (Aquifoliaceae). *Mitochondrial DNA Part B* **2021**, *6*, 227–228. [[CrossRef](#)] [[PubMed](#)]
44. Zeng, S.; Zhou, T.; Han, K.; Yang, Y.; Zhao, J.; Liu, Z.L. The Complete Chloroplast Genome Sequences of Six *Rehmannia* Species. *Genes* **2017**, *8*, 103. [[CrossRef](#)]
45. Nie, L.; Cui, Y.; Wu, L.; Zhou, J.; Xu, Z.; Li, Y.; Li, X.; Wang, Y.; Yao, H. Gene Losses and Variations in Chloroplast Genome of Parasitic Plant *Macrosolen* and Phylogenetic Relationships within Santalales. *Int. J. Mol. Sci.* **2019**, *20*, 5812. [[CrossRef](#)] [[PubMed](#)]
46. Xue, S.; Shi, T.; Luo, W.; Ni, X.; Iqbal, S.; Ni, Z.; Huang, X.; Yao, D.; Shen, Z.; Gao, Z. Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. *Hortic. Res.* **2019**, *6*, 89. [[CrossRef](#)] [[PubMed](#)]
47. Sun, J.; Wang, S.; Wang, Y.; Wang, R.; Liu, K.; Li, E.; Qiao, P.; Shi, L.; Dong, W.; Huang, L.; et al. Phylogenomics and Genetic Diversity of *Arnebia Radix* and Its *Allies* (*Arnebia*, Boraginaceae) in China. *Front. Plant Sci.* **2022**, *13*, 920826. [[CrossRef](#)]
48. Zhao, F.; Peng, H. The complete chloroplast genome of *Caryopteris incana* (Lamiaceae) and phylogenetic analysis. *Mitochondrial DNA Part B* **2020**, *5*, 1399–1400. [[CrossRef](#)]
49. Duan, H.C.; Zheng, X.H.; Li, Y.Y.; Li, S.M.; Ye, L.; Jing, H.Z.; Dong, Q. The complete chloroplast genome of *Fraxinus malacophylla* (Oleaceae, Oleoideae). *Mitochondrial DNA Part B Resour.* **2020**, *5*, 3588–3589. [[CrossRef](#)]
50. He, Y.; Xu, X.; Liu, Q. The complete chloroplast genome of *Onosma fuyunensis* Y. He & Q.R. Liu and its phylogenetic analysis. *Mitochondrial DNA Part B Resour.* **2021**, *6*, 3142–3143. [[CrossRef](#)]
51. Wu, J.H.; Li, H.M.; Lei, J.M.; Liang, Z.R. The complete chloroplast genome sequence of *Trigonotis peduncularis* (Boraginaceae). *Mitochondrial DNA Part B Resour.* **2022**, *7*, 456–457. [[CrossRef](#)]
52. Wu, L.; Fan, P.; Zhou, J.; Li, Y.; Xu, Z.; Lin, Y.; Wang, Y.; Song, J.; Yao, H. Gene Losses and Homology of the Chloroplast Genomes of *Taxillus* and *Phacellaria* Species. *Genes* **2023**, *14*, 943. [[CrossRef](#)]

53. Mann, S.; Chen, Y.-P.P. Bacterial genomic G+C composition-eliciting environmental adaptation. *Genomics* **2010**, *95*, 7–15. [[CrossRef](#)] [[PubMed](#)]
54. Ebert, D.; Peakall, R. Chloroplast simple sequence repeats (cpSSRs): Technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. *Mol. Ecol. Resour.* **2009**, *9*, 673–690. [[CrossRef](#)] [[PubMed](#)]
55. George, B.; Bhatt, B.S.; Awasthi, M.; George, B.; Singh, A.K. Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. *Curr. Genet.* **2015**, *61*, 665–677. [[CrossRef](#)] [[PubMed](#)]
56. Du, X.; Zeng, T.; Feng, Q.; Hu, L.; Luo, X.; Weng, Q.; He, J.; Zhu, B. The complete chloroplast genome sequence of yellow mustard (*Sinapis alba* L.) and its phylogenetic relationship to other Brassicaceae species. *Gene* **2020**, *731*, 144340. [[CrossRef](#)] [[PubMed](#)]
57. Mwanzia, V.M.; Nzei, J.M.; Yan, D.Y.; Kamau, P.W.; Chen, J.M.; Li, Z.Z. The complete chloroplast genomes of two species in threatened monocot genus *Caldesia* in China. *Genetica* **2019**, *147*, 381–390. [[CrossRef](#)] [[PubMed](#)]
58. Dong, S.; Ying, Z.; Yu, S.; Wang, Q.; Liao, G.; Ge, Y.; Cheng, R. Complete chloroplast genome of *Stephania tetrandra* (Menispermaceae) from Zhejiang Province: Insights into molecular structures, comparative genome analysis, mutational hotspots and phylogenetic relationships. *BMC Genom.* **2021**, *22*, 880. [[CrossRef](#)]
59. Qin, Z.; Cai, Z.; Xia, G.; Wang, M. Synonymous codon usage bias is correlative to intron number and shows disequilibrium among exons in plants. *BMC Genom.* **2013**, *14*, 56. [[CrossRef](#)] [[PubMed](#)]
60. Zhang, Z.; Dai, W.; Wang, Y.; Lu, C.; Fan, H. Analysis of synonymous codon usage patterns in torque teno sus virus 1 (TTSuV1). *Arch. Virol.* **2013**, *158*, 145–154. [[CrossRef](#)]
61. Wang, X.; Zhou, T.; Bai, G.; Zhao, Y. Complete chloroplast genome sequence of *Fagopyrum dibotrys*: Genome features, comparative analysis and phylogenetic relationships. *Sci. Rep.* **2018**, *8*, 12379. [[CrossRef](#)] [[PubMed](#)]
62. Chen, Y.; Hu, N.; Wu, H. Analyzing and Characterizing the Chloroplast Genome of *Salix wilsonii*. *BioMed Res. Int.* **2019**, *2019*, 5190425. [[CrossRef](#)] [[PubMed](#)]
63. Zhang, H.; Huang, T.; Zhou, Q.; Sheng, Q.; Zhu, Z. Complete Chloroplast Genomes and Phylogenetic Relationships of *Bougainvillea spectabilis* and *Bougainvillea glabra* (Nyctaginaceae). *Int. J. Mol. Sci.* **2023**, *24*, 13044. [[CrossRef](#)] [[PubMed](#)]
64. Jiao, L.; Lu, Y.; He, T.; Li, J.; Yin, Y. A strategy for developing high-resolution DNA barcodes for species discrimination of wood specimens using the complete chloroplast genome of three *Pterocarpus* species. *Planta* **2019**, *250*, 95–104. [[CrossRef](#)]
65. Bi, Y.; Zhang, M.F.; Xue, J.; Dong, R.; Du, Y.P.; Zhang, X.H. Chloroplast genomic resources for phylogeny and DNA barcoding: A case study on *Fritillaria*. *Sci. Rep.* **2018**, *8*, 1184. [[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.