

Article

Complete Chloroplast Genome Sequence of *Medicago falcata*: Comparative Analyses with Other Species of *Medicago*

Wei Duan ^{1,2}, Xueli Zhang ^{1,2}, Yuxiang Wang ^{1,2,*} and Qian Li ^{1,2,*}

¹ Key Laboratory of Ministry of Education of Grassland Resources and Ecology of Western Arid Region, Xinjiang Agricultural University, Urumqi 830052, China; dwcjatfx@163.com (W.D.); zzhangxl@163.com (X.Z.)

² Key Laboratory of Grassland Resources and Ecology of Xinjiang, College of Grassland Science, Xinjiang Agricultural University, Urumqi 830052, China

* Correspondence: wyx9868@163.com (Y.W.); liqian@xjau.edu.cn (Q.L.)

Abstract

Medicago falcata is one of the most important perennial forage legumes in the *Medicago* genus. In this study, we reported the complete chloroplast genome of two *M. falcata* ecotypes grown in different regions, and compared them with those of *Medicago truncatula* and *Medicago sativa*. We found that the *M. falcata* chloroplast genome lacks a typical quadripartite structure, containing 78 protein-coding genes, 30 tRNA genes, and four ribosomal RNA genes. They shared high conservation in size, genome structure, gene order, gene number and GC content with those of *M. truncatula* and *M. sativa*. High nucleotide diversity occurred in the coding gene regions of *rps16*, *rps3*, and *ycf4* genes. Meanwhile, mononucleotide repeats are the most abundant repeat type, followed by the di-, tri-, tetra-, and pentanucleotides, and forward repeats were more abundant than reverse and palindrome repeats for all these three *Medicago* species. Phylogenetic analyses using both coding sequences and complete chloroplast genomes revealed that *M. falcata* shares the closest phylogenetic relationship with *M. hybrida* and *M. sativa*. This study provided valuable information for further studies on the genetic relationship of the *Medicago* genus.

Keywords: chloroplast genome; legumes; *Medicago*; phylogeny; simple sequence repeats



Academic Editor: Matthew Hegarty

Received: 4 July 2025

Revised: 28 July 2025

Accepted: 29 July 2025

Published: 31 July 2025

Citation: Duan, W.; Zhang, X.; Wang, Y.; Li, Q. Complete Chloroplast Genome Sequence of *Medicago falcata*: Comparative Analyses with Other Species of *Medicago*. *Agronomy* **2025**, *15*, 1856. <https://doi.org/10.3390/agronomy15081856>

Copyright: © 2025 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

Legumes played central roles in the development of agriculture and civilization, and they account for approximately one-third of the world's primary crop production. In addition, legumes are important due to their ecologically vital role in biological nitrogen fixation [1]. The *Medicago* genus is one of the most important forage resources, which is cultivated worldwide [2]. In the *Medicago* genus, *M. truncatula* has been adopted as a model species for legumes [3]. *M. sativa* (alfalfa) is a highly productive and stress-tolerant forage crop valuable for livestock feeding, known as “the king of forage crops” [4,5]. *M. falcata* is mainly distributed in the north of China, Russia, Mongolia, and Europe [6], and grows in adverse environments, with great tolerance against abiotic stresses [7].

The inheritance of the chloroplast genome with conserved gene content and order made it a valuable asset for studies in plant phylogenetic and evolutionary relationships [8,9]. Chloroplast genomes of legumes have undergone considerable diversification in gene/intron content and gene order during phylogenetic evolution [1]. It has been reported that chloroplast genomes of some legume species have undergone rearrangements, including the loss of inverted repeats (IR) or genes (e.g., *rpl22* and *rps16*) [10,11], as

well as inversions of long fragments [12,13], which have been observed in genera such as *Glycyrrhiza* L., *Astragalus* L., *Medicago* L., *Pisum* L., and *Vicia faba* L. [14]. As for alfalfa, its chloroplast DNA was thought to be unrearranged, except for the deletion of one segment of the IR [10,15]. A recent study on the complete chloroplast genome of *Medicago sativa* L. cv. Qingda no. 1 further confirmed this structural feature: the chloroplast genome of this alfalfa cultivar is adapted to high-altitude environments [16]. These findings provide additional evidence for the conservation and diversity of chloroplast genome structure in *Medicago* species.

M. falcata is considered to be a wild species as well as a subspecies of *M. sativa* complex [4,17]. It was still difficult to clearly distinguish among *M. sativa*, *M. falcata*, and their hybrid *Medicago* $\times$ *varia* based on the molecular and morphological evidence [2]. Therefore, more chloroplast genomes of *M. falcata* will be valuable genetic resources for the study of the population genetics and evolutionary relationships of *Medicago* species. In this study, we chose two *M. falcata* ecotypes from different regions (e.g., Russia and Xinjiang, China) for complete chloroplast genome sequencing. Our detailed analyses on the chloroplast genomes of two new *M. falcata* ecotypes enrich and refine the chloroplast genome information of *M. falcata*. In addition, this study would be helpful to further understand the plastid evolution and phylogeny of the *Medicago* genus.

2. Materials and Methods

2.1. Plant Material, DNA Extraction, and Sequencing

The plants of two *Medicago falcata* ecotypes were used in this study: one was obtained from the Federal Research Center of Russia Vavilov Institute of Plant Genetic Resources, and the other was collected in Xinjiang, China. Both ecotypes were cultivated in the greenhouse of the College of Grassland Science, Xinjiang Agricultural University, Urumqi, China. Total genomic DNA was extracted from fresh leaves using the modified CTAB method [18]. Raw sequencing data were generated using the NovaSeq 6000 platform (Illumina, Inc., San Diego, CA, USA), yielding 3.4G of raw data for the chloroplast genomes of the two *M. falcata* ecotypes. The sequencing data and assembled chloroplast genome sequences have been deposited in the GenBank database, with the following accessions: MW271002 (genome sequence) and SRR15182922 (raw data) for the ecotype from Russia; and MW271003 (genome sequence) and SRR15182921 (raw data) for the ecotype collected from Xinjiang, China.

2.2. Chloroplast Genome Assembly and Annotation

The software GetOrganelle v1.5 [19] was used to assemble the chloroplast genome, with the chloroplast genome of another *M. falcata* ecotype (GenBank accession number: NC 032066.1) as a reference. Chloroplast genome annotation was performed using GeSeq [20] (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>, accessed on 12 May 2023). In order to ensure the prediction accuracy of the encoded protein and RNA genes, the program Hmmer was used to predict the protein coding sequences, ARAGORN v1.2.38 was used to predict the tRNA genes [21], and the final annotation results were manually corrected. According to the annotation, circular diagram of the chloroplast genomes of *M. falcata* was subsequently drawn using OGDRAW v1.3.1 [22].

2.3. Genome Structure Analysis of Chloroplast Sequence

Custom Perl v5.34.0 and Python v 3.10.0 scripts were used to process the chloroplast genome annotation files of five *Medicago* samples, and to calculate the basic data of the chloroplast genome structure, including the number of chloroplast genes, the total length of the chloroplast genome (bp), GC content, protein-coding gene number, CDS (Coding

DNA Sequence) number, rRNA number and proportion, tRNA number and proportion, IR number, and the classification of chloroplast genes in the *Medicago* plant subgenus.

2.4. Analysis of the Chloroplast Genome Consistency

The Python script was used to process the annotation files of the five *Medicago* plant samples, and the alignment of the complete chloroplast genomes of the five *Medicago* samples was sequenced using the online mVISTA program. (<http://genome.lbl.gov/vista/mvista/submit.shtml>, accessed on 6 August 2023) [23]. Then we selected Shuffle-LAGAN as the parameter for sequence comparison [24]. The sequence identities of the chloroplast genomes of all *Medicago* plant samples were analyzed, with the chloroplast genome of *M. falcata* (NC 032066.1) as a reference sequence.

2.5. Analysis of Simple and Complex Repeats

The Tandem Repeats Finder program [25] and the online REPuter tool (<https://bibiserv.cebitec.uni-bielefeld.de/bibi/Administration>Contact.html>, accessed on 12 December 2023) [26] were used to predict repetitive sequences and dispersed repetitive sequences. The local MISA program was used to predict the simple repeat sequence (SSR) [27], and the maximal number of mononucleotide repeats, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, pentanucleotide repeats, and hexanucleotide repeats were set to 10, 6, 4, 3, 3, and 3, respectively.

2.6. Analysis of Nucleotide Polymorphism of the Chloroplast Genome

The commonly shared CDS sequences of five *Medicago* chloroplasts were aligned with the muscle v3.8 program. DNAsp v6 was used to analyze the nucleotide polymorphism and calculate the nucleotide diversity.

2.7. Phylogenetic Analysis

A total of 37 chloroplast genome sequences were used for phylogenetic analysis, including 11 sequences of the *Medicago* genus and 25 other leguminous species, with the chloroplast genome of *Arabidopsis thaliana* L. as the outgroup. The phylogenetic trees were constructed based on two datasets: the CDS regions of 78 protein-coding genes and the complete chloroplast genome sequences of the 36 leguminous species, and the outgroup *Arabidopsis thaliana* L.

The sequences of CDS and the complete chloroplast genome sequences were retrieved using custom Python scripts, and homologous sequences from different species were aligned via a Python-based pipeline incorporating the clustalW2 program. Individual alignment outputs were concatenated sequentially. Phylogenetic trees were constructed from the concatenated CDS dataset and complete chloroplast genome sequences using both maximum likelihood (ML) and Bayesian inference (BI) methods. For ML analysis, RAxML-NG v1.0.1 was used with the GTR+G4+F+I substitution model, which was selected by ModelFinder integrated into Phylosuite v1.2.2 based on the corrected Akaike Information Criterion (AICc), with 1000 bootstrap replicates. For BI analysis, the MrBayes 3.2.6 program was run within Phylosuite v1.2.2 (citation) using the GTR+G4+F+I model, as determined by the Bayesian Information Criterion (BIC). The Bayesian inference tree was constructed using MrBayes 3.2.7 [28]. The Markov chain Monte Carlo (MCMC) algorithm was executed for 1,000,000 generations, with trees sampled every 1000 generations until convergence (average standard deviation of split frequencies <0.01). The initial 20% of trees were discarded as burn-in, and the remaining trees were used to generate a 50% majority-rule consensus tree.

3. Results

3.1. Characterization of the Chloroplast Genomes of Two *M. falcata* Ecotypes

The lengths of the complete chloroplast genomes of these two *M. falcata* ecotypes are 125,657 bp and 125,479 bp, respectively (Figure 1, Table 1). By comparison, the chloroplast genome of *M. falcata* strain 1210, *M. sativa*, and *M. truncatula* are 124,430, 125,330, and 124,033 bp in length, respectively (Table 1). In detail, their chloroplast genome sequences contained 112 unique genes, including 78 protein-coding genes, 30 tRNA genes and four rRNA genes. Except for *M. truncatula*, all the other samples had the same numbers of protein-coding genes (78) and rRNA genes (4). It is worth noting that *M. falcata* (NC 032066.1) from Inner Mongolia in China lacks two tRNA, while the other two ecotypes of *M. falcata* have 30 tRNAs (Table 2).

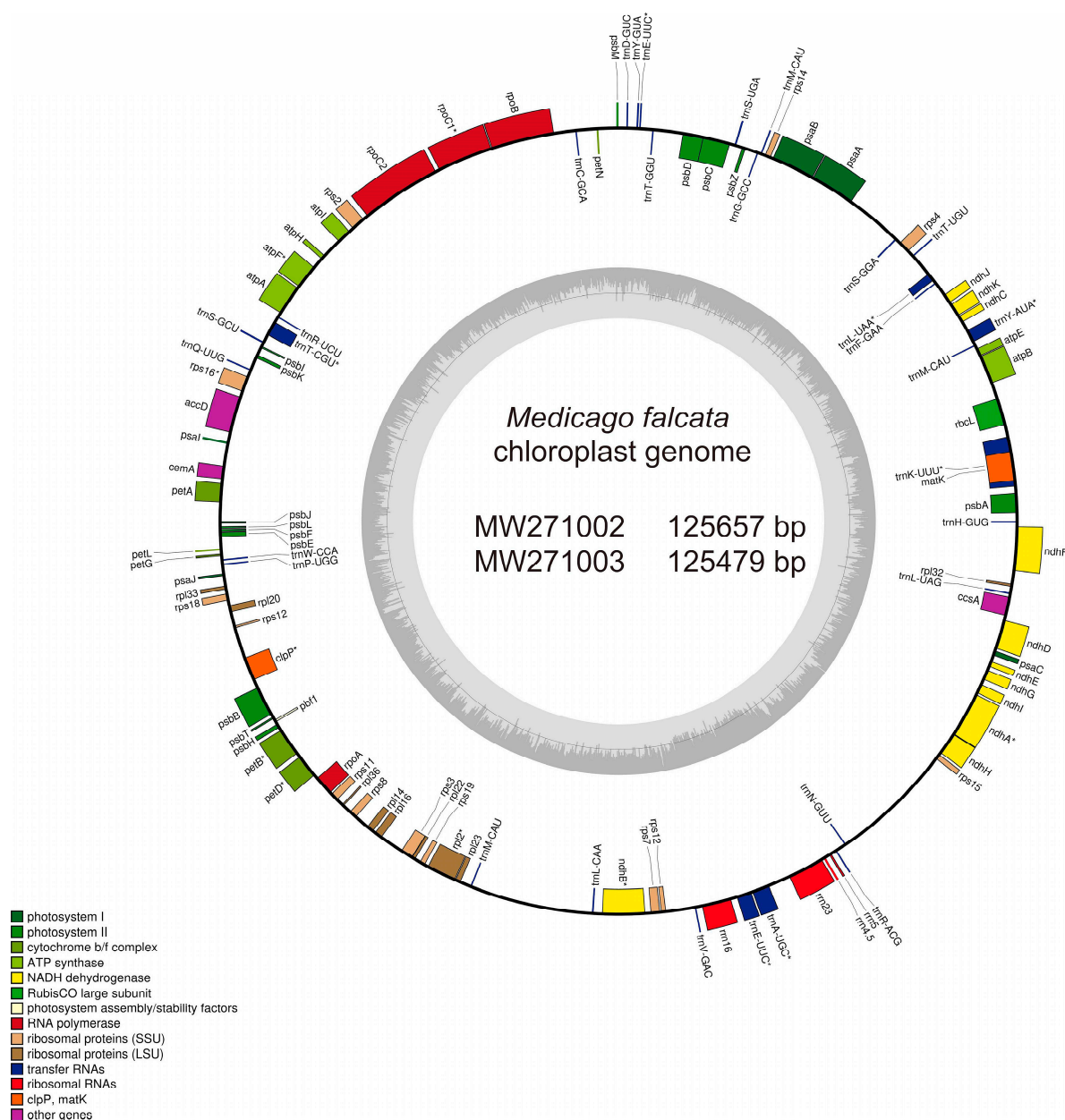


Figure 1. Visualization of *Medicago falcata* chloroplast gene map with annotations. The inner circle is for GC content. Genes are color-coded based on function as per the legend. Genes on the inside of the outer circle are minus (−) strand and genes on the outside of the outer circle are plus (+) strand. Genes belonging to different functional groups are color-coded.

Table 1. Summary of the chloroplast genome assembly data for *Medicago*.

Name	Length (bp)	Gene Number	Protein-Coding Gene Number	Protein Coding Gene (%)	rRNA_Gene Number	rRNA (%)	tRNA Gene Number	tRNA (%)	GC Content (%)	IR Length/bp
<i>Medicago falcata</i> MW 271002	125,657	112	78	69.64	4	3.57	30	26.79	33.85	N/A
<i>Medicago falcata</i> MW 271003	125,479	112	78	69.64	4	3.57	30	26.79	33.84	N/A
<i>Medicago falcata</i> NC 032066.1	124,430	110	78	70.91	4	3.64	28	25.45	33.96	N/A
<i>Medicago truncatula</i> NC 003119.6	124,033	111	77	69.37	4	3.6	30	27.03	33.97	N/A
<i>Medicago sativa</i> NC 042841.1	125,330	112	78	69.64	4	3.57	30	26.79	33.87	N/A

Table 2. Gene content and functional classification of the *Medicago* chloroplast genomes.

Gene Category	Gene Group	Gene Names
Other genes	Envelope membrane protein (1)	<i>cemA</i>
	Maturase (1)	<i>matK</i>
	Protease (1)	<i>ClpP</i> ^a
	Subunit of acetyl-CoA carboxylase (1)	<i>accD</i>
	c-type cytochrome synthesis gene (1)	<i>ccsA</i>

Table 2. Cont.

Gene Category	Gene Group	Gene Names
Photosynthesis	Others (3)	<i>pbf1</i> , <i>ycf3</i> ^b , <i>ycf4</i>
	Subunits of ATP synthase (6)	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> ^a , <i>atpH</i> , <i>atpI</i>
	Subunits of NADH dehydrogenase (11)	<i>NdhA</i> ^a , <i>ndhB</i> ^a , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
	Subunits of cytochrome (6)	<i>petA</i> , <i>petB</i> ^a , <i>petD</i> ^a , <i>petG</i> , <i>petL</i> , <i>petN</i>
	Subunits of photosystem II (14)	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbT</i> , <i>psbZ</i>
	Subunits of photosystem (5)	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>
	Subunits of rubisco (1)	<i>rbcL</i>
Self-replication	DNA dependent RNA polymerase (4)	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> ^a , <i>rpoC2</i>
	Large subunits of ribosome (9)	<i>rpl14</i> , <i>rpl16</i> , <i>rpl2</i> ^a , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
	Small subunits of ribosome (12)	<i>rps11</i> , <i>rps12</i> ^a , <i>rps14</i> , <i>rps15</i> , <i>rps16</i> ^a , <i>rps18</i> , <i>rps19</i> , <i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> , <i>rps8</i>
	rRNA genes (4)	<i>rrn16</i> , <i>rrn23</i> , <i>rrn4.5</i> , <i>rrn5</i>
	tRNA genes (30)	<i>trnA</i> -UGC, <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC (×2), <i>trnF</i> -GAA, <i>trnG</i> -GCC, <i>trnH</i> -GUG, <i>trnK</i> -UUU, <i>trnL</i> -CAA, <i>trnL</i> -UAA, <i>trnL</i> -UAG, <i>trnM</i> -CAU (×3), <i>trnN</i> -GUU, <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnR</i> -ACG, <i>trnR</i> -UCU, <i>trnS</i> -GCU, <i>trnS</i> -GGA, <i>trnS</i> -UGA, <i>trnT</i> -CGU, <i>trnT</i> -GGU, <i>trnT</i> -UGU, <i>trnV</i> -GAC, <i>trnW</i> -CCA, <i>trnY</i> -AUA, <i>trnY</i> -GUA
Unknown function	Conserved open reading frames (2)	<i>ycf1</i> , <i>ycf2</i>

^a represents gene containing one intron, and ^b represents gene containing two introns, ×2 shows two copies, ×3 shows three copies.

3.2. Comparative Analysis of the Chloroplast Genome of Three *Medicago* Species

The mVISTA-based identity plot indicated conservation in DNA sequence and gene synteny across the whole chloroplast genome, and revealed the areas with increased genetic variation (Figure 2). The gene number, order, and orientation were found to be highly conserved. However, their CDS regions showed distinct variation (Figure 2). Further analysis of nucleotide polymorphisms (nucleotide diversity) revealed that 87 out of 108 genes exhibited variations among these five *Medicago* species, while no variations were detected in the remaining 21 genes. In this study, we found that all rRNA and tRNA genes are highly conserved (Figure 3). Compared with tRNA genes and rRNA genes, protein-coding genes had relatively higher nucleotide diversity (Figure 3, Table S1), and the highest nucleotide diversity was 0.486 for *rps18*, while the lowest value was 0.198 for *psbE*. The analyses of nucleotide diversity revealed the variation size of the chloroplast genome in different *Medicago* species, and regions with high nucleotide diversity (e.g., *rps18*, *rps3*, and *ycf3*) may be developed as potential molecular markers for population genetics.

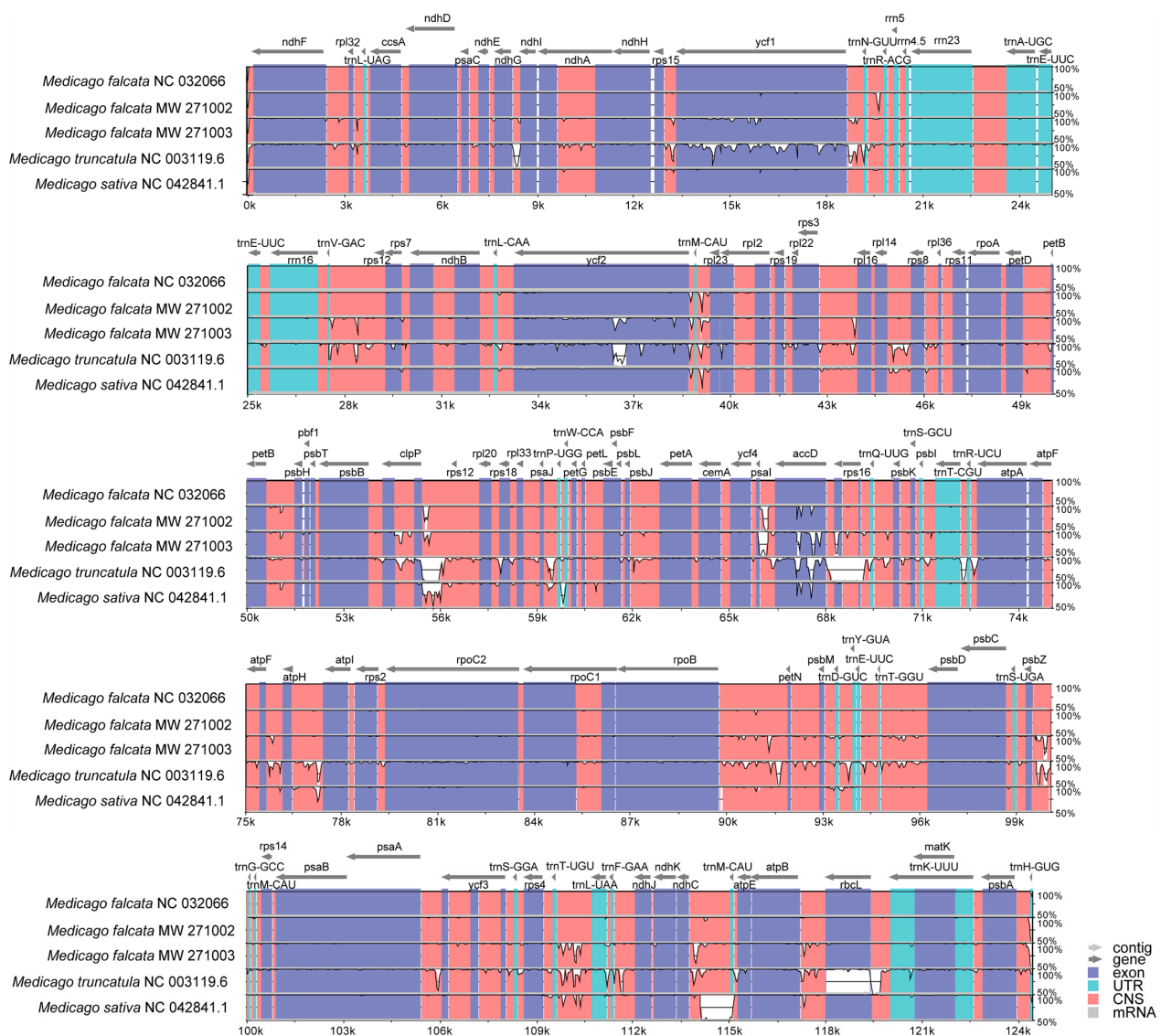


Figure 2. Comparison of chloroplast genomes of *Medicago* species using the mVISTA program. A cut-off of 70% identity was used for the plots. The Y-scale axis represents the percent identity between 50% and 100%. Gray arrows above the alignment indicate gene positions and their orientation. UTR: untranslated region; CNS: conserved noncoding sequences.

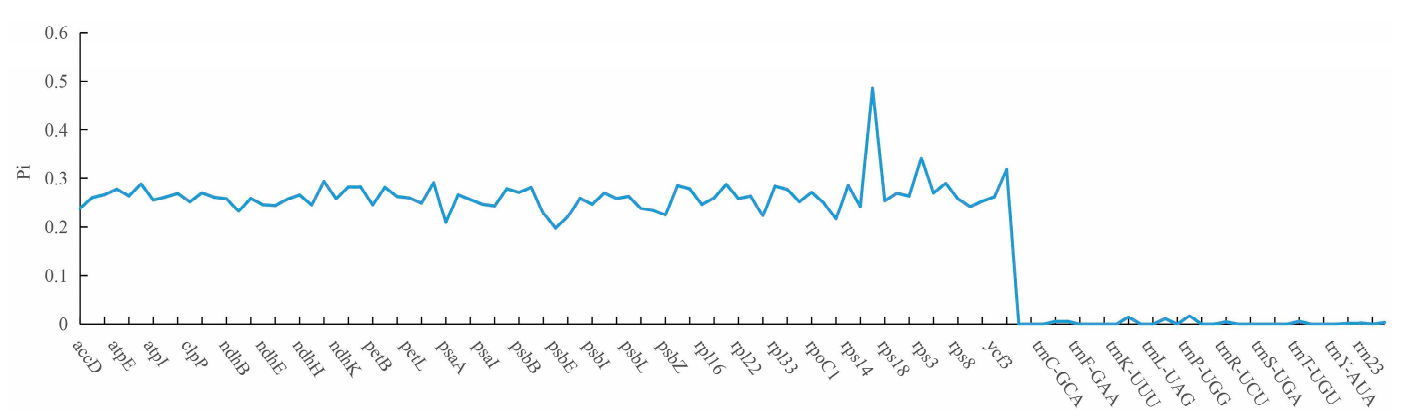


Figure 3. Nucleotide diversity values between five *Medicago* species determined using whole chloroplast genomes. The x-axis represents chloroplast genome genes, and the y-axis represents nucleotide diversity. Detailed Pi values were shown in Table S1.

3.3. Features of the cpDNA Repeats of Medicago

Complete chloroplast genomes of all five *Medicago* species contained mono-, di-, tri-, tetra-, penta- and hexanucleotide SSRs. The most abundant repeats are mononucleotide repeats, accounting for more than 71.73% of the total SSRs, followed by the di-, tri-, tetra- and pentanucleotides (Figure 4A). Considering a complementary sequence, 16 classified repeat types were found in all five *Medicago* species. The most abundant repeat type was A/T, which were 67, 75, 76, 76, and 79 in *M. falcata* NC 032066.1, *M. falcata* MW 271002, *M. falcata* MW 271003, *M. truncatula* NC 003119.6, and *M. sativa*, respectively (Figure 4B).

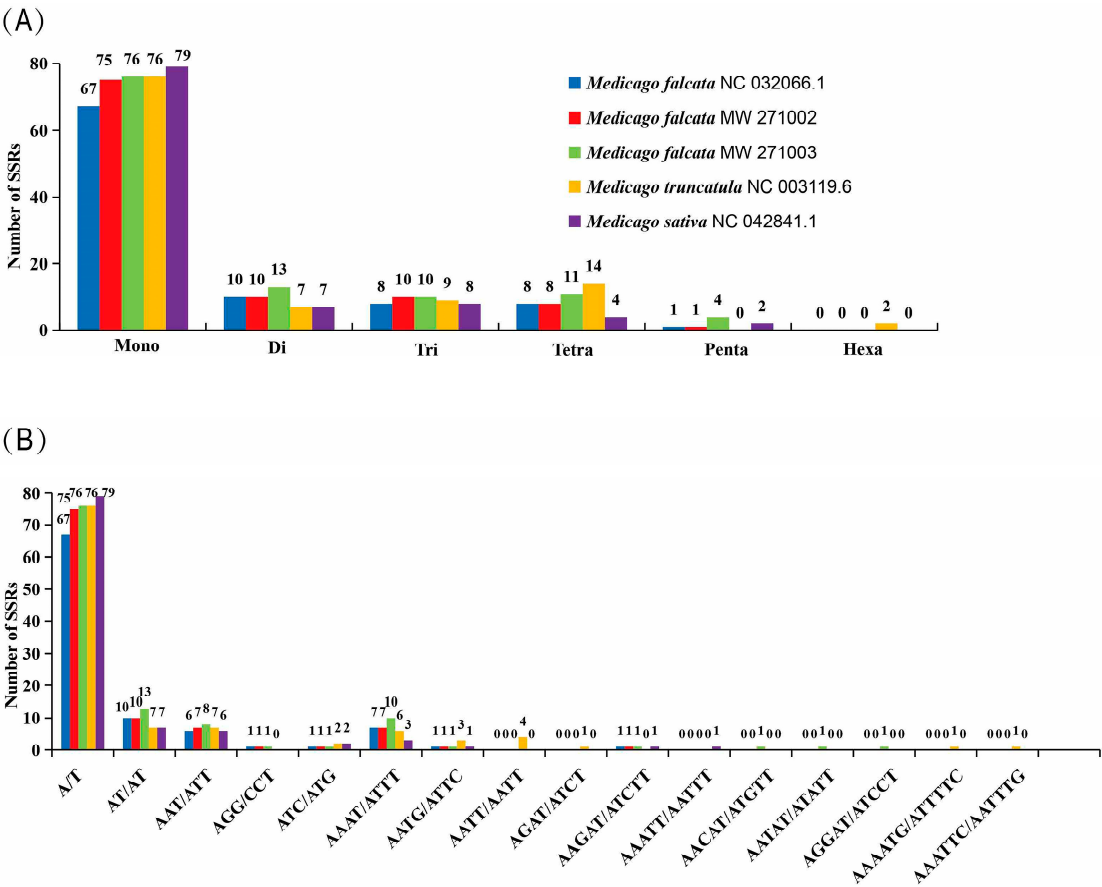


Figure 4. Type and amount of simple sequence repeats in the chloroplast genome of *Medicago*. (A) Distribution of simple sequence repeat (SSR) types across five *Medicago* species. (B) Frequency of specific SSR motifs in the five *Medicago* species.

Four distinct types of long dispersed repeats (LDRs) were identified: forward (F), palindromic (P), reverse (R), and complementary (C) repeats. Among these repeats, forward repeats were the most abundant, with counts ranging from 47 to 182, followed by palindromic repeats (16 to 54) and reverse repeats (2 to 3) (Figure 5). Additionally, two complementary repeats were identified in *Medicago falcata* MW271002, which were absent in the other four accessions (Figure 5).

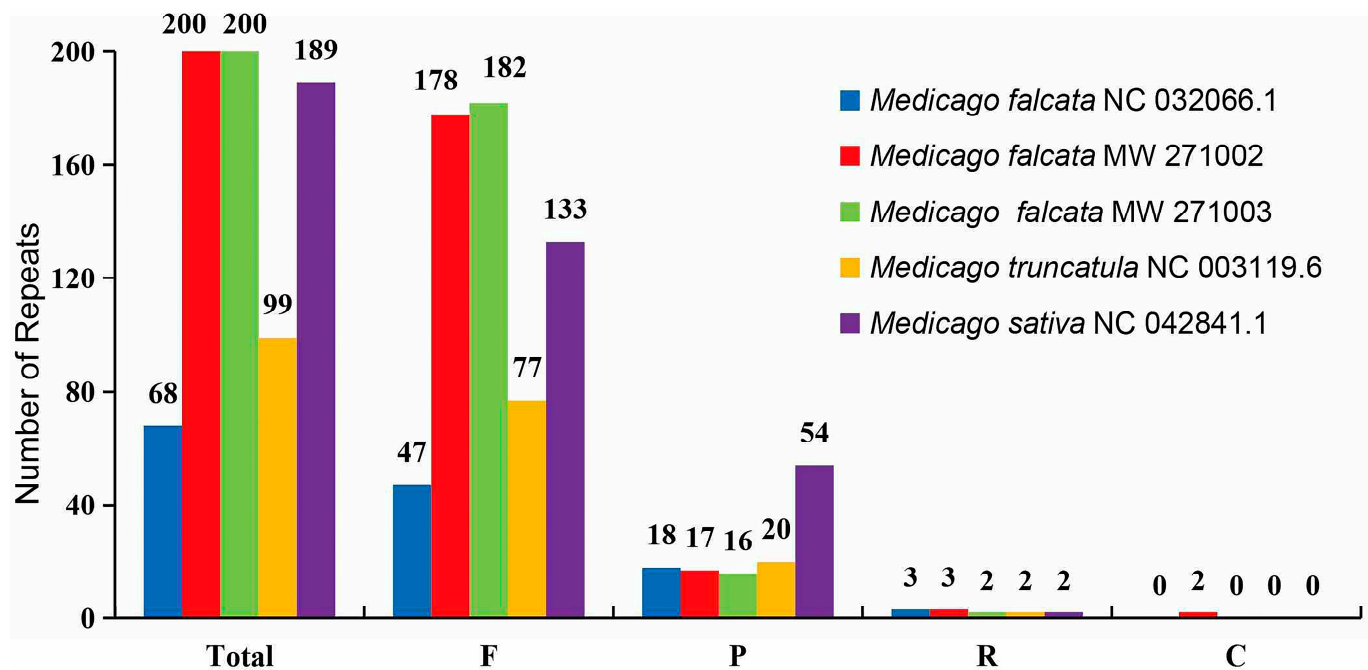


Figure 5. Analyses of long dispersed repeats (LDRs) in *Medicago* chloroplast genomes. The frequency of LDRs classified by the length and type of repeat: Total: total numbers of all repeats, F: forward repeats, P: palindromic repeats, R: reverse repeats, C: complementary repeats.

3.4. Phylogenetic Relationship Between *Medicago* and Related Species

Within the cluster of the *Medicago* genus, a slight difference was found for the relationship between the trees clustering with the CDS or clustering with the complete chloroplast genome sequences (Figure 6). The phylogenetic tree constructed with the full length is more accurate than the phylogenetic tree clustering with the CDS regions. The phylogenetic analysis with the complete chloroplast genome sequences showed that *M. falcata* MW271002 and *M. falcata* MW271003 were both clustered with *M. falcata*, and they were close to *M. sativa* and *M. hybrida* (Figure 6B), which was supported by a high bootstrap value (>98%). This result is consistent with previous phylogenetic analyses [4,16]. In both phylogenetic trees, *M. falcata* MW271003 was closely related to *M. hybrida* and *M. sativa*, suggesting that *M. falcata* and *M. sativa* may have evolved from *M. hybrida* during their evolution.

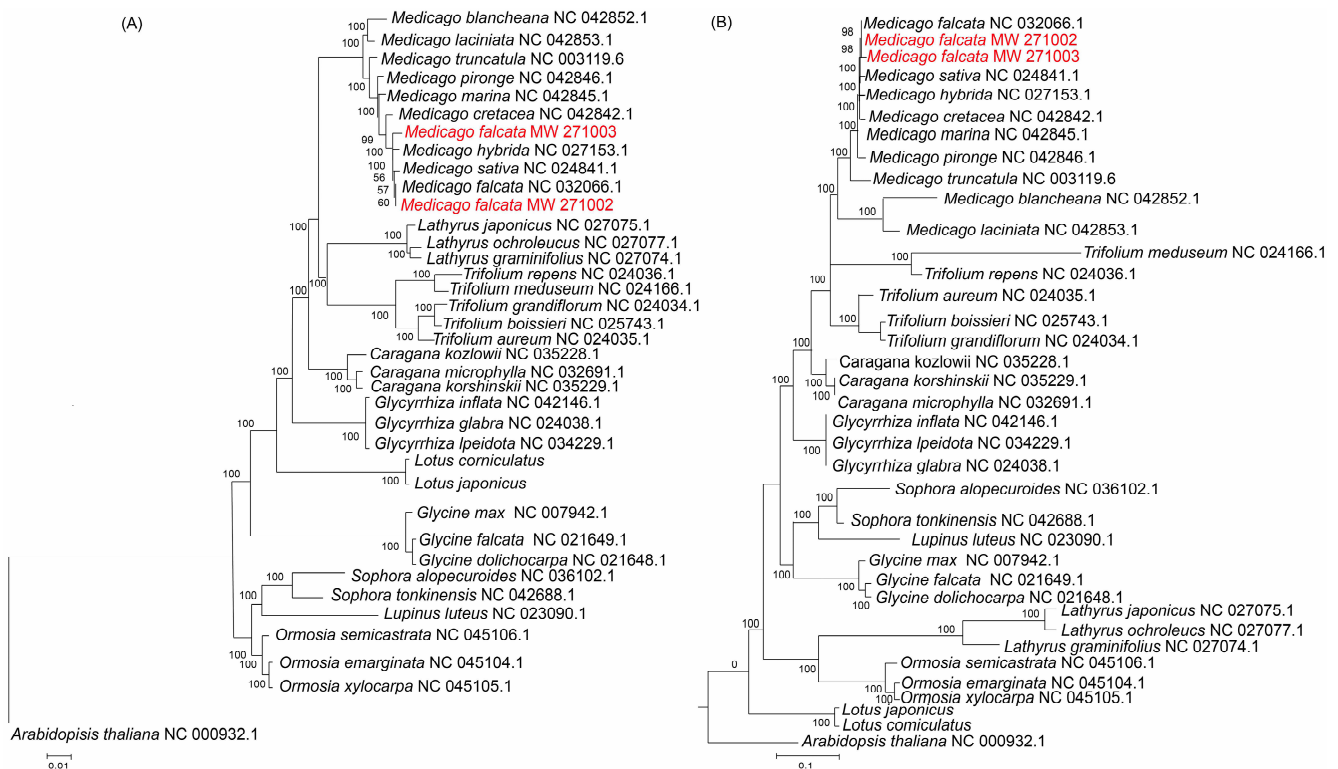


Figure 6. Phylogenetic trees constructed using the maximum likelihood (ML) method. ML tree based on the CDS sequences of protein-coding genes (A), or based on the complete chloroplast genomes (B).

4. Discussion

4.1. Conservation of *Medicago* cpDNA

Chloroplast genomes are highly conserved in angiosperms with respect to gene content and order [29]. The highly conserved structure of the chloroplast genome is a potential source for the phylogenetic reconstruction of species relationships among legume plants [30]. The number, type, and order of genes were found to be very similar among the chloroplast genome sequences of these five *Medicago* samples [1,15,31]. In comparison, the complete chloroplast genomes of *M. falcata* MW271002 (125,657 bp) and *M. falcata* MW271003 (125,479 bp) are 327 bp and 149 bp longer than that of *M. sativa* (125,330 bp), respectively, and 1624 bp and 1446 bp longer than that of *M. truncatula* (124,033 bp), respectively. Notably, the lack of a typical quadripartite structure in *M. falcata* chloroplast genomes aligns with findings in other *Medicago* studies. For example, a study on *M. sativa* L. cv. Qingdao no. 1 reported that its chloroplast genome has a length of 125,637 bp, lacks a distinct quadripartite structure, and contains 111 genes (77 protein-coding genes, 30 tRNA genes, and four rRNA genes) with a GC content of 38.33% [16]. While the gene composition (numbers of protein-coding genes, tRNA genes, and rRNA genes) in this study is generally consistent with that of *M. falcata* in our research (78 protein-coding genes, 30 tRNA genes, and four rRNA genes), there is a notable difference in GC content between the two, with the GC content of *M. falcata* in our study ranging from 33.84% to 33.85%. Another study on two *Medicago* species also confirmed the conservation of the genome structure: the chloroplast genomes of *M. sativa* and *M. falcata* were 125,095 bp and 125,810 bp in length, respectively, and both contained 109 unique genes [31]. This further supports the conservation of gene composition and structural framework within the genus, while the difference in GC content may reflect adaptive divergence or genetic background variations among species.

It is worth noting that the genome of *M. falcata* (NC 032066.1) is 900 nucleotides shorter than that of *M. sativa* L. *M. falcata* (NC 032066.1) lacks two tRNA genes (e.g., *trnC*-GCA

and *trnY-AUA*) compared to other species (Table 1), while *M. truncatula* is deficient in the protein-coding gene *rps16* (Table 1). *M. falcata* has a distinctive chloroplast structure that does not exhibit a typical quadripartite organization (Figure 1, Table 1), which differs from the chloroplast genomes of most typical land plants that contain two copies of the inverted repeat (IR) region [32]. In addition, a lack of IR can cause extensive gene rearrangement. This phenomenon mainly occurs in the legume tribes, including subclover, broad bean, pea, and alfalfa [10,15,33]. The *infA* gene was found in most angiosperm chloroplast genomes, including representatives of the early branching lineages [32], but it was not present in the *M. falcata* or *M. sativa* chloroplast genome, which may be due to the presence of one IR. These results support the hypothesis that the presence of the large inverted repeat stabilizes the chloroplast genome against major structural rearrangements. The GC content of complete chloroplast genomes in angiosperms is relatively consistent, ranging from 36.7% to 37.0% [34,35]. However, in our study, the GC content of these *Medicago* species is approximately 33.8%, and this relatively low GC content may be attributed to the composition and number of pseudogenes [36].

4.2. Simple and Complex Repeats Analysis

Large and complex repeat sequences are likely to play crucial roles in driving plastid genome rearrangements and sequence divergence events [37,38]. Variations in the distribution patterns of these repetitive elements are linked to structural rearrangements of complete chloroplast genomes and nucleotide substitution processes, making them potential candidates for developing genetic markers in phylogenetic research [38]. Our findings align with previous reports indicating that simple sequence repeats (SSRs) within complete chloroplast genomes are predominantly composed of polyadenine (poly A) or polythymine (poly T) repeats, with tandem guanine (G) or cytosine (C) repeats being relatively rare [35,39]. The prevalence of A/T-rich SSRs in land plant chloroplast genomes has been documented in prior studies, and our observations are consistent with such patterns reported in other species [39–42]. In the analyzed *Medicago* samples, mononucleotide repeats were the most abundant SSR type, representing over 71.73% of total SSRs, followed by dinucleotide, trinucleotide, tetranucleotide, and pentanucleotide repeats (Figure 4A). This distribution pattern is comparable to that observed in *Lilium* L. [35]. A similar trend was reported in *M. sativa* L. cv. Qingda no. 1, where mononucleotide repeats were dominant among its 114 identified SSR loci [16]. Additionally, our results showed that tetranucleotide repeats were more frequent than pentanucleotide repeats, which is consistent with findings in *Quercus* L. [43]. Hexanucleotide repeats were scarce across the five complete *Medicago* chloroplast genomes, a pattern that mirrors observations in *Lilium* L. and *Allium* L. [35,36]. Forward repeats were more abundant than reverse and palindrome repeats (Figure 4B) [44], a trend also supported by the analysis of two other *Medicago* species [31]. These new resources will potentially be useful for population studies in the genus *Medicago*.

4.3. Phylogenetic

Phylogenetic analyses based on complete plastid genome sequences have provided valuable insights into relationships among and within plant genera. As recorded in the flora of China, *M. falcata* is not only considered to be a wild species, but also a subspecies of *M. sativa* [4,17], which is consistent with our phylogenetic analyses (Figure 6). The phylogenetic tree constructed with its full length is more accurate than the phylogenetic tree clustering with the CDS regions. The phylogenetic analysis with the complete chloroplast genome sequences showed that *M. falcata* MW271002 and *M. falcata* MW271003 were both clustered with *M. falcata*, and they were close to *M. sativa* and *M. hybrida* (Figure 6B), which

was supported by a high bootstrap value (>98%). This result is consistent with previous phylogenetic analyses [4,16,31].

Phylogenetic analyses based on complete plastid genome sequences have provided valuable insights into relationships among and within plant genera. As recorded in the flora of China, *M. falcata* is not only considered to be a wild species, but also a subspecies of *M. sativa* [4,17], which is consistent with our phylogenetic analyses (Figure 6). The phylogenetic tree constructed with its full length is more accurate than the phylogenetic tree clustering with the CDS regions. The phylogenetic analysis with the complete chloroplast genome sequences showed that *M. falcata* MW271002 and *M. falcata* MW271003 were both clustered with *M. falcata*, and they were close to *M. sativa* and *M. hybrida* (Figure 6B), which was supported by a high bootstrap value (>98%). This result is consistent with previous phylogenetic analyses [4,16,31].

Notably, although phylogenetic analyses based on plastid genomes indicate a close genetic relationship between *M. falcata* and *M. hybrida*, the two species exhibit marked morphological differences. *M. falcata* typically displays yellow flowers, linear-lanceolate leaflets, and a prostrate growth habit, which are adaptive traits to cold and arid environments [6,7]. In contrast, *M. hybrida* (recognized as a natural hybrid of *M. falcata* and *M. sativa*) displays intermediate characteristics: including pale purple to white flowers, broader leaflets, and an erect stature, reflecting its adaptation to more diverse habitats [4,17].

This paradox of high plastid genome similarity versus distinct morphology can be explained by two key factors. First, plastid genomes are maternally inherited and evolve slowly, retaining high sequence conservation even when nuclear genomes diverge [8,31]. Morphological traits are more likely governed by nuclear genes (e.g., those controlling floral pigmentation or growth regulation), which are not captured by plastome analyses. Second, *M. hybrida* may have accumulated nuclear genetic variations through introgression between *M. falcata* and *M. sativa*, leading to phenotypic novelty while retaining the maternal plastid signature [2,17]. Similarly, phylogenetic studies on *Medicago sativa* L. cv. Qingda no. 1 and the space-mutated *Medicago sativa* cv. Hangmu no. 1 also indicated close affiliations with *M. sativa* and *M. falcata* [16,45], reinforcing the close evolutionary relationship within this complex.

As reported previously, even when grown in the same area, the phenotypic traits of *M. falcata* also show considerable differences between individuals [2]. These variations between individuals even within the same population may be related to the characteristics of cross-pollination of *M. falcata* and its ability to adapt to adverse environments. Therefore, the characterization of multiple complete chloroplast genomes provided the opportunity for comparison and investigation with the current *M. falcata* chloroplast genome. In the genus *Medicago*, species with a relatively close phylogenetic relationship clustered together could be explained by frequent genetic exchange and gene introgression among species.

5. Conclusions

In this study, we determined the complete chloroplast genome sequences of two *M. falcata* samples from Russia and Xinjiang. The results revealed that gene orientation, genome structure, size, gene number, gene order, and GC content are conserved among the five *Medicago* accessions, including *M. falcata* NC 032066.1, *M. falcata* MW271002, *M. falcata* MW271003, *M. truncatula* NC 003119.6, and *M. sativa* NC 042841.1. However, there are slight differences in the number of protein-coding genes and tRNA genes. In the comparative analysis of sequence differences, the protein-coding gene similarity was low, and there was large variation among CDSs. Further analysis of nucleotide polymorphisms showed that 87 out of 108 surveyed regions exhibit variations among the five *Medicago* species, with the nucleotide diversity of other coding genes being relatively high (exceeding

0.2). In our study, we observed that all rRNA and tRNA genes are highly conserved. The most abundant are mononucleotide repeats, followed by di-, tri-, tetra-, and penta- repeats. Forward repeats were more abundant than reverse and palindrome repeats. The two phylogenetic tree analysis results are slightly different. The phylogenetic tree made from the full length is more accurate than the CDS phylogenetic tree clustering. These results offer valuable information regarding the identification of *Medicago* species and will benefit further investigations of these species.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy15081856/s1>, Table S1: Detailed nucleotide diversity values between five *Medicago* species determined using whole chloroplast genomes.

Author Contributions: Conceptualization: Y.W. and Q.L.; methodology: W.D. and X.Z.; formal analysis: W.D. and X.Z.; writing-original draft preparation: W.D.; writing-review and editing: Y.W. and Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Biological Breeding-National Science and Technology Major Project (2022ZD0401104), National Natural Science Foundation of China (32460349), the Key Research and Development Program of the Xinjiang Uygur Autonomous Region (2023B02031), Postdoctoral Mobile Station for Crop Science and Autonomous Region Talent Development Fund “Tianchi Talent” Introduction Plan Project.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding authors upon reasonable request.

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

References

- Jansen, R.K.; Cai, Z.; Raubeson, L.A.; Daniell, H.; Depamphilis, C.W.; Leebens-Mack, J.; Müller, K.F.; Guisinger-Bellian, M.; Haberle, R.C.; Hansen, A.K.; et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19369–19374. [CrossRef] [PubMed]
- Chen, J.; Wu, G.; Shrestha, N.; Wu, S.; Guo, W.; Yin, M.; Li, A.; Liu, J.; Ren, G. Phylogeny and Species Delimitation of Chinese *Medicago* (Leguminosae) and Its Relatives Based on Molecular and Morphological Evidence. *Front. Plant Sci.* **2021**, *11*, 619799. [CrossRef]
- Benedito, V.A.; Torres-Jerez, I.; Murray, J.D.; Andriankaja, A.; Allen, S.; Kakar, K.; Wandrey, M.; Verdier, J.; Zuber, H.; Ott, T.; et al. A gene expression atlas of the model legume *Medicago truncatula*. *Plant J.* **2008**, *55*, 504–513. [CrossRef]
- Small, E.; Jomphe, M. A synopsis of the genus *Medicago* (Leguminosae). *Can. J. Bot.* **1989**, *67*, 3260–3294. [CrossRef]
- Liu, Z.; Chen, T.; Ma, L.; Zhao, Z.; Zhao, P.X.; Nan, Z.; Wang, Y. Global transcriptome sequencing using the Illumina platform and the development of EST-SSR markers in autotetraploid alfalfa. *PLoS ONE* **2013**, *8*, e83549. [CrossRef]
- Shi, H.; He, S.; He, X.; Lu, S.; Guo, Z. An eukaryotic elongation factor 2 from *Medicago falcata* (MfEF2) confers cold tolerance. *BMC Plant Biol.* **2019**, *19*, 218. [CrossRef] [PubMed]
- Zhang, L.L.; Zhao, M.G.; Tian, Q.Y.; Zhang, W.H. Comparative studies on tolerance of *Medicago truncatula* and *Medicago falcata* to freezing. *Planta* **2011**, *234*, 445–457. [CrossRef]
- Birky, C.W., Jr. The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. *Annu. Rev. Genet.* **2001**, *35*, 125–148. [CrossRef]
- Wu, Z.Q.; Ge, S. The phylogeny of the BEP clade in grasses revisited: Evidence from the whole-genome sequences of chloroplasts. *Mol. Phylogenet. Evol.* **2012**, *62*, 573–578. [CrossRef]
- Palmer, J.D.; Osorio, B.; Aldrich, J.; Thompson, W.F. Chloroplast DNA evolution among legumes: Loss of a large inverted repeat occurred prior to other sequence rearrangements. *Curr. Genet.* **1987**, *11*, 275–286. [CrossRef]
- Saski, C.; Lee, S.B.; Daniell, H.; Wood, T.C.; Tomkins, J.; Kim, H.G.; Jansen, R.K. Complete Chloroplast Genome Sequence of *Glycine max* and Comparative Analyses with other Legume Genomes. *Plant Mol. Biol.* **2005**, *59*, 309–322. [CrossRef]
- Palmer, J.D.; Thompson, W.F. Rearrangements in the chloroplast genomes of mung bean and pea. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 5533–5537. [CrossRef]
- Bruneau, A.; Palmer, D.J.D. A Chloroplast DNA Inversion as a Subtribal Character in the *Phaseoleae* (Leguminosae). *Syst. Bot.* **1990**, *15*, 378–386. [CrossRef]

14. Wojciechowski, M.F.; Lavin, M.; Sanderson, M.J. A phylogeny of legumes (*Leguminosae*) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *Am. J. Bot.* **2004**, *91*, 1846–1862. [\[CrossRef\]](#)
15. Tao, X.; Ma, L.; Zhang, Z.; Liu, W.; Liu, Z. Characterization of the complete chloroplast genome of alfalfa (*Medicago sativa*) (*Leguminosae*). *Gene Rep.* **2017**, *6*, 67–73. [\[CrossRef\]](#)
16. Ren, Y.L.; Ma, Y.J.; Li, X.; Li, X.A.; Yang, G.Z.; Li, P. Complete chloroplast genome sequence and characteristics analysis of Qingda no.1 alfalfa (*Medicago sativa* L. cv. Qingda no.1). *Czech J. Genet. Plant Breed.* **2023**, *59*, 160–168. [\[CrossRef\]](#)
17. Quiros, C.F.; Bauchan, G.R. The genus *Medicago* and the origin of the *Medicago sativa* complex. *Agronomy* **1988**, 93–124. [\[CrossRef\]](#)
18. Allen, G.C.; Flores-Vergara, M.A.; Krasynanski, S.; Kumar, S.; Thompson, W.F. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* **2006**, *1*, 2320–2325. [\[CrossRef\]](#)
19. 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. [\[CrossRef\]](#)
20. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. *GeSeq*—Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. [\[CrossRef\]](#)
21. Laslett, D.; Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **2004**, *32*, 11–16. [\[CrossRef\]](#)
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. [\[CrossRef\]](#)
23. 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. [\[CrossRef\]](#)
24. Brudno, M.; Do, C.B.; Cooper, G.M.; Kim, M.F.; Davydov, E.; NISC Comparative Sequencing Program; Green, E.D.; Sidow, A.; Batzoglou, S. LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **2003**, *13*, 721–731. [\[CrossRef\]](#)
25. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Kurtz, S.; Schleiermacher, C. REPuter: Fast computation of maximal repeats in complete genomes. *Bioinformatics* **1999**, *15*, 426–427. [\[CrossRef\]](#)
27. Beier, S.; Himmelbach, A.; Colmsee, C.; Zhang, X.Q.; Barrero, R.A.; Zhang, Q.; Li, L.; Bayer, M.; Bolser, D.; Taudien, S.; et al. Construction of a map-based reference genome sequence for barley, *Hordeum vulgare* L. *Sci. Data* **2017**, *4*, 170044. [\[CrossRef\]](#)
28. Ronquist, F.; Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **2003**, *19*, 1572–1574. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Sugiura, M. The chloroplast genome. *Plant Mol. Biol.* **1992**, *19*, 149–168. [\[CrossRef\]](#)
30. Raveendar, S.; Na, Y.W.; Lee, J.R.; Shim, D.; Ma, K.H.; Lee, S.Y.; Chung, J.W. The complete chloroplast genome of *Capsicum annuum* var. *glabriusculum* using Illumina sequencing. *Molecules* **2015**, *20*, 13080–13088. [\[CrossRef\]](#)
31. Yan, W.; Shi, W.; Liu, L.; Ma, Y.; Chen, L.; Wang, Z.; Hou, X. Complete sequencing of the chloroplast genomes of two *Medicago* species. *Mitochondrial DNA Part B Resour.* **2017**, *2*, 302–303. [\[CrossRef\]](#)
32. Raubeson, L.A.; Peery, R.; Chumley, T.W.; Dziubek, C.; Fourcade, H.M.; Boore, J.L.; Jansen, R.K. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genom.* **2007**, *8*, 174. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Strauss, S.H.; Palmer, J.D.; Howe, G.T.; Doerksen, A.H. Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 3898–3902. [\[CrossRef\]](#)
34. Kim, S.; Park, J.Y.; Yang, T. Comparative analysis of the complete chloroplast genome sequences of a normal male-fertile cytoplasm and two different cytoplasms conferring cytoplasmic male sterility in onion (*Allium cepa* L.). *J. Hortic. Sci. Biotechnol.* **2015**, *90*, 459–468. [\[CrossRef\]](#)
35. Du, Y.P.; Bi, Y.; Yang, F.P.; Zhang, M.F.; Chen, X.Q.; Xue, J.; Zhang, X.H. Complete chloroplast genome sequences of *Lilium*: Insights into evolutionary dynamics and phylogenetic analyses. *Sci. Rep.* **2017**, *7*, 5751. [\[CrossRef\]](#)
36. Huo, Y.; Gao, L.; Liu, B.; Yang, Y.; Kong, S.; Sun, Y.; Yang, Y.; Wu, X. Complete chloroplast genome sequences of four *Allium* species: Comparative and phylogenetic analyses. *Sci. Rep.* **2019**, *9*, 12250. [\[CrossRef\]](#)
37. Timme, R.E.; Kuehl, J.V.; Boore, J.L.; Jansen, R.K. A comparative analysis of the *Lactuca* and *Helianthus* (*Asteraceae*) plastid genomes: Identification of divergent regions and categorization of shared repeats. *Am. J. Bot.* **2007**, *94*, 302–312. [\[CrossRef\]](#)
38. Weng, M.L.; Blazier, J.C.; Govindu, M.; Jansen, R.K. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol. Biol. Evol.* **2014**, *31*, 645–659. [\[CrossRef\]](#)
39. Wang, L.; Guo, Z.H. The complete chloroplast genome of *Tamarix ramosissima* and comparative analysis of *Tamaricaceae* species. *Biol. Plant.* **2021**, *65*, 237–245. [\[CrossRef\]](#)

40. Kaila, T.; Chaduvla, P.K.; Rawal, H.C.; Saxena, S.; Tyagi, A.; Mithra, S.V.A.; Solanke, A.U.; Kalia, P.; Sharma, T.R.; Singh, N.K.; et al. Chloroplast Genome Sequence of Clusterbean (*Cyamopsis tetragonoloba* L.): Genome Structure and Comparative Analysis. *Genes* **2017**, *8*, 212. [\[CrossRef\]](#)
41. Li, X.; Tan, W.; Sun, J.; Du, J.; Zheng, C.; Tian, X.; Zheng, M.; Xiang, B.; Wang, Y. Author Correction: Comparison of Four Complete Chloroplast Genomes of Medicinal and Ornamental *Meconopsis* Species: Genome Organization and Species Discrimination. *Sci. Rep.* **2019**, *9*, 15163. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Somaratne, Y.; Guan, D.L.; Wang, W.Q.; Zhao, L.; Xu, S.Q. Complete chloroplast genome sequence of *Xanthium sibiricum* provides useful DNA barcodes for future species identification and phylogeny. *Plant Syst. Evol.* **2019**, *305*, 949–960. [\[CrossRef\]](#)
43. Yang, Y.; Zhou, T.; Duan, D.; Yang, J.; Feng, L.; Zhao, G. Comparative Analysis of the Complete Chloroplast Genomes of Five *Quercus* Species. *Front. Plant Sci.* **2016**, *7*, 959. [\[CrossRef\]](#)
44. Jung, J.; Do, H.D.K.; Hyun, J.; Kim, C.; Kim, J.H. Comparative analysis and implications of the chloroplast genomes of three thistles (*Carduus* L., Asteraceae). *PeerJ* **2021**, *9*, e10687. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Yang, C.; Wu, X.Y.; Guo, X.; Bao, P.J.; Chu, M.; Zhou, X.L.; Liang, Z.Y.; Ding, X.Z.; Yan, P. The complete chloroplast genome of *Medicago sativa* cv. Hangmu No.1, a plant of space mutation breeding. *Mitochondrial DNA Part B Resour.* **2019**, *4*, 603–604. [\[CrossRef\]](#)

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.