

Article

Analysis of Complete Chloroplast Genome: Structure, Phylogenetic Relationships of *Galega orientalis* and Evolutionary Inference of Galegeae

Junjie Feng ^{1,2}, Yi Xiong ¹ , Xiaoli Su ¹, Tianqi Liu ¹, Yanli Xiong ¹, Junming Zhao ¹, Xiong Lei ², Lijun Yan ², Wenlong Gou ^{2,*} and Xiao Ma ^{1,*} 

¹ College of Grassland Science and Technology, Sichuan Agricultural University, Chengdu 611130, China

² Sichuan Academy of Grassland Science, Chengdu 611130, China

* Correspondence: gw11975117@163.com (W.G.); maroar@126.com (X.M.)

Abstract: *Galega orientalis*, a leguminous herb in the Fabaceae family, is an ecologically and economically important species widely cultivated for its strong stress resistance and high protein content. However, genomic information of *Galega orientalis* has not been reported, which limiting its evolutionary analysis. The small genome size makes chloroplast relatively easy to obtain genomic sequence for phylogenetic studies and molecular marker development. Here, the chloroplast genome of *Galega orientalis* was sequenced and annotated. The results showed that the chloroplast genome of *G. orientalis* is 125,280 bp in length with GC content of 34.11%. A total of 107 genes were identified, including 74 protein-coding genes, 29 tRNAs and four rRNAs. One inverted repeat (IR) region was lost in the chloroplast genome of *G. orientalis*. In addition, five genes (*rpl22*, *ycf2*, *rps16*, *trnE-UUC* and *pbf1*) were lost compared with the chloroplast genome of its related species *G. officinalis*. A total of 84 long repeats and 68 simple sequence repeats were detected, which could be used as potential markers in the genetic studies of *G. orientalis* and related species. We found that the Ka/Ks values of three genes *petL*, *rpl20*, and *ycf4* were higher than one in the pairwise comparison of *G. officinalis* and other three Galegeae species (*Calophaca sinica*, *Caragana jubata*, *Caragana korshinskii*), which indicated those three genes were under positive selection. A comparative genomic analysis of 15 Galegeae species showed that most conserved non-coding sequence regions and two genic regions (*ycf1* and *clpP*) were highly divergent, which could be used as DNA barcodes for rapid and accurate species identification. Phylogenetic trees constructed based on the *ycf1* and *clpP* genes confirmed the evolutionary relationships among Galegeae species. In addition, among the 15 Galegeae species analyzed, *Galega orientalis* had a unique 30-bp intron in the *ycf1* gene and *Tibetia liangshanensis* lacked two introns in the *clpP* gene, which is contrary to existing conclusion that only *Glycyrrhiza* species in the IR lacking clade (IRLC) lack two introns. In conclusion, for the first time, the complete chloroplast genome of *G. orientalis* was determined and annotated, which could provide insights into the unsolved evolutionary relationships within the genus Galegeae.

Keywords: *Galega orientalis*; Galegeae; chloroplast genome; phylogenetic relationship



Citation: Feng, J.; Xiong, Y.; Su, X.; Liu, T.; Xiong, Y.; Zhao, J.; Lei, X.; Yan, L.; Gou, W.; Ma, X. Analysis of Complete Chloroplast Genome: Structure, Phylogenetic Relationships of *Galega orientalis* and Evolutionary Inference of Galegeae. *Genes* **2023**, *14*, 176. <https://doi.org/10.3390/genes14010176>

Academic Editor: Zhiqiang Wu

Received: 19 November 2022

Revised: 24 December 2022

Accepted: 5 January 2023

Published: 9 January 2023



Copyright: © 2023 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

As the main semi-autonomous genetic organelle in plant cells, chloroplasts contain their own independent DNA (cpDNA) and play an important role in carbon fixation, oxygen production and regulation of many metabolic pathways [1]. Chloroplast (cp) genome is typically a closed loop loaded structure, usually ranging in size from 115 to 180 kb, which consists of a large single-copy region, a small single-copy region and a pair of inverted repeats (IRs) regions in most angiosperms [2]. In comparison with the nuclear genome, cp genome has a small size, haploid nature, maternal inheritance, low rate of nucleotide substitution and highly conserved genome structure [3]. There are striking

variations in genome size, gene number, and GC content in different plant lineages' cp genomes, as well as signatures of selection, genomic rearrangement, intron gains and losses, and expansion/contraction of the IR [4]. Comparative analyses of the chloroplast genomes played a crucial role in understanding the evolution of the chloroplast genome [5–8].

Fabaceae, the third largest family of angiosperms after Orchidaceae and Asteraceae, contains many species with wide distribution [9,10]. Currently, Fabaceae is classified into six subfamilies, namely Caesarpinioideae, Certifidoideae, Detarioideae, Dialioideae, Duparquetioideae and Papilionoideae, according to a new classification system based on cp *matK* gene [11]. Among them, Papilionoideae is an excellent model for evolutionary analysis and investigation of cp genomic structure variation because of numerous rearrangements, gene deletions, intron deletions, and local hypervariable regions present within this taxon [12]. As a result, Papilionoideae has been divided into seven main evolutionary clades, including Cladrastis, Genistoids, Dalbergioids, Mirbelioids, Millettoids, Robinoids and the IR lacking clade (IRLC). Notably, IRLC is considered a useful unit for studying the evolutionary mechanism of Papilionoideae due to its high nucleotide mutation rate and obvious structural variation [13]. In the IRLC lineage, cp genomes were found rearranged in many species, including gene replication, loss, and sequence inversion, which is particularly common in *Trifolium* [14]. However, the cause of these rearrangements in the cp genome is unclear.

Galega orientalis is a member of the Galegeae tribe in Papilionoideae, which belongs to the genus *Galega* along with *Galega officinalis*. It is a perennial cool-season forage legume species mainly naturally distributed in grassland and subalpine meadows at an altitude of 330–1800 m around the Black Sea region extending from southeastern Europe through the Caucasus to western Asia [15,16]. It has higher crude protein content than common forage (*Medicago sativa*, *Trifolium pratense*) and excellent disease and pest resistance [17,18], good livestock palatability and low *galegine* content [19]. In addition, *G. orientalis* is a nectar plant useful for producing excellent quality of honey, and is also used to extract *dimethylbiguanide* to treat diabetes as a medicinal plant [20,21]. Therefore, because of its high ecological value and economic value, *G. orientalis* has been introduced and promoted in many countries [22].

However, the improvement of biomass trait by transgenic technique in *Galega* species has been reported [22]. Currently, only few reports are available on the chloroplast genome sequence of the genus *Galega*, which seriously hampers our understanding of molecular identification, evolution and phylogeny of Galegeae tribe [23,24]. Especially, as the model genus of Galegeae, *G. officinalis* is the only *Galega* species with an available complete chloroplast genome [25]. Therefore, sequencing other *Galega* species would greatly enrich the cp genome of this genus. Here, we sequenced the complete chloroplast genome of *G. orientalis*, and compared it with the chloroplast genome of *G. officinalis* as well as with the cp genomes of other six genera species in Galegeae: *Astragalus*, *Calophaca*, *Caragana*, *Oxytropis*, *Tibetia* and *Glycyrrhiza*. We also elucidated the evolutionary relationships within the genus *Galega* in Galegeae based on their complete cp genomes. Our results will provide important information and markers for phylogenetic and evolutionary studies in Galegeae.

2. Methods

2.1. Genomic DNA Extraction and Sequencing

The young leaves of *Galega orientalis* were collected from Hongyuan County, Sichuan, China. DNA was extracted from leaves immediately after drying with silica gel. Total genomic DNA was extracted from 500 mg of dried leaves using the TruSeq DNA Sample Preparation Kit (Vanzyme, China). In accordance with the Illumina manual (San Diego, CA, USA), sheared low-molecular-weight DNA fragments were used to construct paired-end libraries. Sequencing of the completed libraries was performed on an Illumina HiSeq 2500 using a PE150 sequencing strategy and 300 bp insert size. The data that support the findings of this study have been deposited into CNGB Sequence Archive (CNSA) [26] of China National GeneBank DataBase (CNGBdb) [27] with accession number CNP0003740.

2.2. Cp Genome Assembly and Annotation

The sequence containing more than 3 unidentified bases in the raw data of *Galega orientalis* was eliminated. Sequences with Q20 less than 60% were eliminated. Low quality sequences at 3' end, and sequences with length less than 60 bp were eliminated. Based on the reference genome of *Galega officinalis* (NC_051885.1), cp DNA was assembled. Briefly, sequences were assembled into Contigs using MetaSPAdes software [28], and non-target sequences were removed. The sequences of *G. orientalis* and *G. officinalis* were verified using NCBI BLAST, with the following parameters: E value less than 1×10^{-5} identity values of nearly 100%, and almost no gap. Then, CPGAVAS2 was used to annotate chloroplast genes [29]. The coding sequence (CDS) and rRNA of each organism were analyzed using BLAST Version 2.2.25, HMMER Version 3.1b2 and the NCBI genome database CP. ARAGORN V 1.2.38 [30] was used with a search server TRNAscan-SE [31] for predicting and confirming tRNA sequences. Finally, the consensus annotation was obtained using CPGAVAS2 and visualized with OGDRAW [32].

2.3. Characterization of Repeat Sequences

The REPuter program [33] was used to identify forward repeats, reverse repeats, complementary sequences and palindromic sequences of at least 30 bases, with hamming distances of three and a 90% identity rate. Simple sequence repeats (SSRs) were detected using the microsatellite identification tool MISA [34] with the following parameters: the minimum SSR motifs number of mononucleotide, dinucleotide, trinucleotide repeats, other nucleotides were set to 10, 5, 4, and 3, respectively.

2.4. Ka/Ks Calculation, Divergent Hotspots Identification and Genome Structure Comparison

In the present study, the shared protein coding genes of 14 closely related species of *G. orientalis* in Galegeae (*G. officinalis*, *Astragalus mongholicus*, *A. galactites*, *A. laxmannii*, *Calophaca sinica*, *Caragana jubata*, *C. korshinskii*, *Oxytropis diversifolia*, *O. splendens*, *O. bicolor*, *Tibetia liangshanensis*, *Glycyrrhiza zaissanica*, *G. gobica* and *G. alaschanica*) were analyzed based on Ka/Ks value using KaKs Calculator v2.0 [35]. Furthermore, to evaluate the nucleotide polymorphism (Pi) of *G. orientalis* and *G. officinalis* cp genomes, the complete chloroplast genome sequence was aligned with MAFFT v7.402 [36] and then DnaSP v6.12 [37] software was used to perform sliding window analysis to calculate Pi values with the following parameters: window length of 800 bp and step size of 200 bp. mVISTA [38] software was used to compare the complete cp genomes of *G. orientalis* and 14 Galegeae species in the shuffle-LAGAN mode and with *G. orientalis* annotation as a reference.

2.5. Phylogenetic Analysis

Complete chloroplast genome sequences of 15 Galegeae species and another 16 legume species, including *Crotalaria albida* (NC_061361.1), *Crotalaria pallida* (NC_053562.1), *Piptanthus concolor* (NC_051876.1), *Piptanthus nepalensis* (NC_056139.1), *Ammopiptanthus mongolicus* (MK704436.1), *Ammopiptanthus nanus* (NC_034743.1), *Lotus corniculatus* (MT528596), *Carmichaelia australis* (MF597719), *Vicia sepium* (NC_039595), *Lupinus micranthus* (KU726828), *Clitoria ternatea* (MN709849.1), *Cicer arietinum* (NC_011163.1), *Medicago sativa* (KU321683), *Medicago hybrida* (NC_027153.1), *Onobrychis viciifolia* (MW007721.1) and *Trifolium alexandrinum* (MN857160) were used to construct a phylogenetic tree. Multiple sequence alignment of the 31 complete cp genome sequences was performed using MAFFT [34] with default parameters. Phylogenetic trees were constructed using the IQ-TREE (Maximum Likelihood Estimation) software applying *Pterocarpus santalinus* (MN251250.1) and *Arachis hypogaea* (NC_037358.1) as outgroups, with a bootstrap number of 1000 [39]. The ModelFinder module confirmed that the TVM+F+R7 was the best nucleotide substitution model.

3. Results

3.1. Chloroplast Genome Characteristics and Structure of *Galega orientalis*

In total, there were at least 16 million ReadSum (pair-end reads) in the original sequencing data of *Galega orientalis*, among which Q20 > 96% and Q30 > 90%. After removing adaptor sequences and low-quality sequences, 1,283,368 clean reads were remained. Both *G. orientalis* and *G. officinalis* cp genomes were found lacked IR region. The cp genome size of *G. orientalis* and *G. officinalis* was 125,280 and 125,086 bp, with an overall GC content of 34.11% and 34.18%, respectively (Figures 1 and 2). Based on chloroplast genome loop map, the *G. orientalis* chloroplast showed significant similarity with that of *G. officinalis* in gene content, order, and direction.

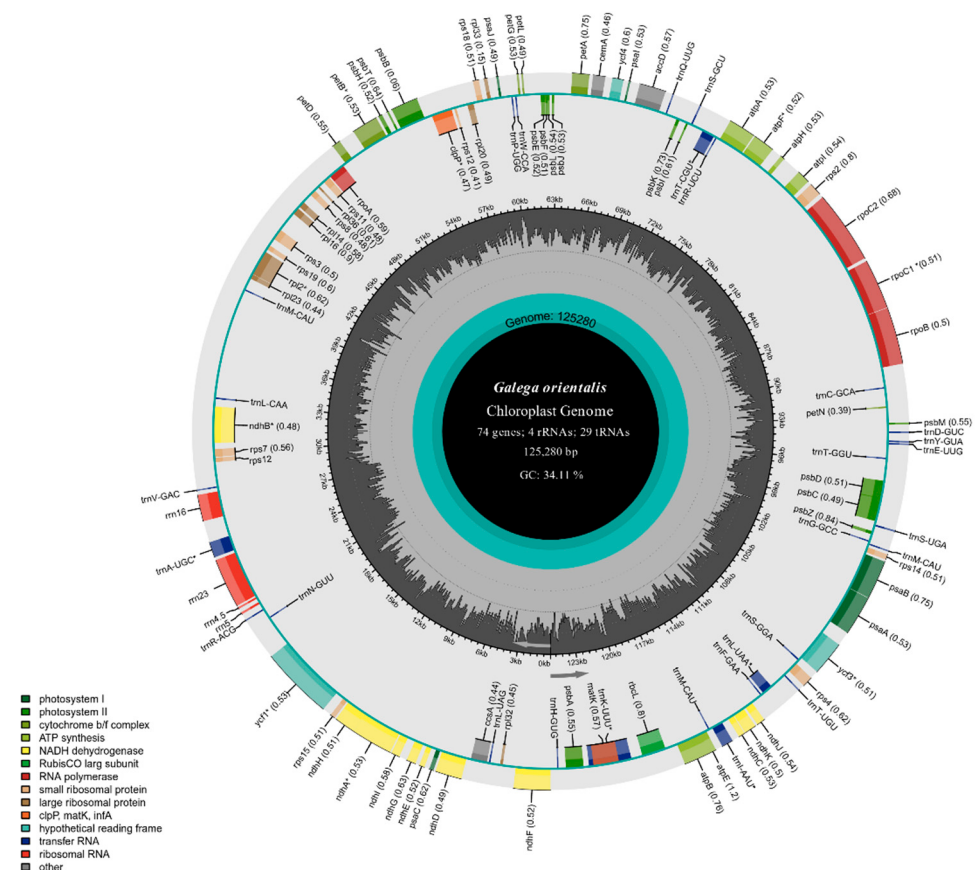


Figure 1. The chloroplast genome map of *Galega orientalis*. The genes drawn inside and outside of the circle are transcribed counterclockwise and clockwise, respectively. The dark gray inside the circle represents GC content, and the light gray corresponds to AT content. * Genes with introns.

In total, 107 genes were annotated in *G. orientalis*, including 74 protein-coding genes, 29 tRNA and four rRNA. Overall, 14 genes contained introns, of which nine encoded proteins, and five encoded tRNA. Of these, 13 genes had one intron, whereas *ycf3* had two. *ycf1* and *trnK-UUU* had the smallest intron (30 bp) and largest intron (2475 bp), respectively (Table 1). The main structural differences were as follows: there was an intron in the *ycf1* gene of *G. orientalis*, but not in *G. officinalis* (Table S1). Furthermore, five genes, including *rpl22*, *ycf2*, *rps16*, *trnE-UUC* and *pb1* were lost in *G. orientalis* compared with the chloroplast genome of *G. officinalis*.

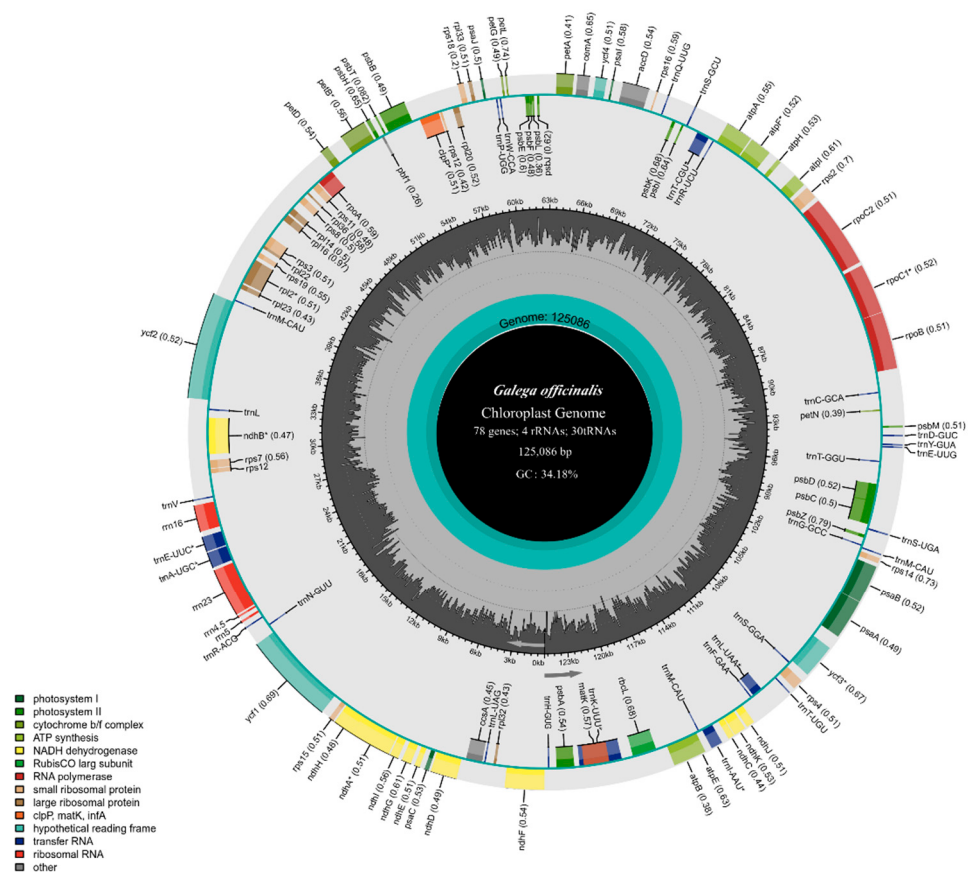


Figure 2. The chloroplast genome map of *Galega officinalis*. The genes drawn inside and outside of the circle are transcribed counterclockwise and clockwise, respectively. The dark gray inside the circle represents GC content, and the light gray corresponds to AT content. * Genes with introns.

Table 1. Lengths of introns and exons of the split genes in the *Galega orientalis* complete chloroplast genome.

Gene Name	Gene Location			Length (bp)				
	Strand	Start	End	Exon I	Intro I	Exon II	Intro II	Exon III
<i>ndhA</i>	—	9250	11,603	553	1262	539		
<i>ycf1</i>	—	13,671	17,798	1881	30	2217		
<i>trnA-UGC</i>	—	24,166	25,050	37	812	36		
<i>ndhB</i>	+	30,629	32,792	721	685	758		
<i>rpl2</i>	+	41,007	42,519	391	688	434		
<i>petB</i>	—	51,034	52,517	9	827	648		
<i>clpP</i>	+	55,465	56,683	360	631	228		
<i>trnT-CGU</i>	+	71,994	72,731	35	660	43		
<i>atpF</i>	—	74,721	75,956	145	681	410		
<i>rpoC1</i>	—	84,108	86,933	430	771	1625		
<i>ycf3</i>	—	106,655	108,601	124	719	232	721	151
<i>trnL-UAA</i>	+	111,044	111,688	35	540	50		
<i>trnI-AAU</i>	—	114,670	115,325	32	566	58		
<i>trnK-UUU</i>	+	120,448	122,994	37	2475	35		

Among the 107 chloroplast genes in *G. orientalis*, 11 encoded ribosomal proteins, 12 were associated with transcription, 33 were involved in self-replication, and 44 genes were associated with photosynthesis. These genes encoded photosystem I/II subunits, Rubisco subunits, ATP synthetase subunits, cytochrome b/f complexes, C-type cytochrome synthesis and NADH dehydrogenase subunits. In addition, four genes encoded maturase,

protease, chloroplast envelope membrane protein and Subunit of acetyl-CoA, and normal nucleotide sequences of the three structural genes (Table 2).

Table 2. Genes predicted in the chloroplast genome of *Galega orientalis*.

Category	Function	Name of Genes					
Ribosomal proteins (11)	Small subunit of ribosome (SSU)	<i>rps15</i>	<i>rps7</i>	<i>rps19</i>	<i>rps3</i>	<i>rps8</i>	<i>rps18</i>
Transcription (12)	Large subunit of ribosome (LSU)	<i>rps2</i>	<i>rps14</i>	<i>rps4</i>	<i>rps11</i>	<i>rps12</i>	
		<i>rpl32</i>	<i>rpl23</i>	<i>rpl2</i>	<i>rpl16</i>	<i>rpl14</i>	<i>rpl36</i>
		<i>rpl20</i>	<i>rpl33</i>				
Self-replication (33)	RNA polymerase subunits	<i>rpoA</i>	<i>rpoC1</i>	<i>rpoC2</i>	<i>rpoB</i>		
	Ribosomal RNA Genes	<i>rnr5</i>	<i>rnr4.5</i>	<i>rnr23</i>	<i>rnr16</i>		
	Transfer RNA genes	<i>trnL-UAG</i>	<i>trnN-GUU</i>	<i>trnR-ACG</i>	<i>trnA-UGC</i>	<i>trnV-GAC</i>	<i>trnL-CAA</i>
Photosynthesis related genes (44)		<i>trnM-CAU</i> ^a	<i>trnP-UGG</i>	<i>trnW-CCA</i>	<i>trnQ-UUG</i>	<i>trnS-GCU</i>	<i>trnT-CGU</i>
		<i>trnR-UCU</i>	<i>trnC-GCA</i>	<i>trnD-GUC</i>	<i>trnY-GUA</i>	<i>trnE-UUC</i>	<i>trnT-GGU</i>
		<i>trnS-UGA</i>	<i>trnG-GCC</i>	<i>trnM-CAU</i>	<i>trnS-GGA</i>	<i>trnT-UGU</i>	<i>trnL-UAA</i>
		<i>trnF-GAA</i>	<i>trnI-AAU</i>	<i>trnM-CAU</i>	<i>trnK-UUU</i>	<i>trnH-GUG</i>	
	Subunits of Photosystem I	<i>psaJ</i>	<i>psaI</i>	<i>psaB</i>	<i>psaA</i>	<i>psaC</i>	
	Subunits of Photosystem II	<i>psbH</i>	<i>psbT</i>	<i>psbB</i>	<i>psbK</i>	<i>psbI</i>	<i>psbC</i>
		<i>psbE</i>	<i>psbF</i>	<i>psbL</i>	<i>psbJ</i>	<i>psbM</i>	<i>psbZ</i>
		<i>psbA</i>	<i>psbD</i>				
	Large subunit of Rubisco	<i>rbcL</i>					
	Subunits of ATP synthase	<i>atpA</i>	<i>atpF</i>	<i>atpH</i>	<i>atpI</i>	<i>atpE</i>	<i>atpB</i>
	Cytochrome b/f complex	<i>petD</i>	<i>petB</i>	<i>petG</i>	<i>petL</i>	<i>petA</i>	<i>petN</i>
	C-type cytochrome synthesis gene	<i>ccsA</i>					
	Subunits of NADH dehydrogenase	<i>ndhF</i>	<i>ndhD</i>	<i>ndhE</i>	<i>ndhG</i>	<i>ndhI</i>	<i>ndhA</i>
		<i>ndhH</i>	<i>ndhB</i>	<i>ndhJ</i>	<i>ndhK</i>	<i>ndhC</i>	
		<i>matK</i>					
Other genes (7)	Maturase	<i>clpP</i>					
	Protease						
	Chloroplast envelope membrane protein	<i>cemA</i>					
	Subunit of acetyl-CoA	<i>accD</i>					
	Hypothetical open reading frames	<i>ycf1</i>	<i>ycf3</i>	<i>ycf4</i>			

Note: ^a duplicated gene.

3.2. SSRs and Repeat Sequences Analysis

Analysis of repeat sequences is important for gene rearrangement and phylogenetic construction. Therefore, we analyzed and compared the dispersed repeats in *G. orientalis* with those of *G. officinalis*. In total, 84 long repeats were found in the complete chloroplast genome of *G. orientalis*, including 52 forward repeats (30–170 bp), 26 palindromic repeats (30–72 bp), three reverse repeats (30–31 bp) and three complement repeats (30 bp). Among them, the maximum number of repeats (N = 170) was located in the *ycf1-trnN-GUU* region. In addition, 73.8% of the repeat sequence occurred in the intergenic spacer region (IGS), 14.3% in coding regions (*psaA*, *psaB*, *ndhF*, *ycf1*, *ycf4*, *psbA*, *rbcL* and *trnS—GCU*), 11.9% in the introns of *ycf1*, *ndhA*, *ndhB*, and *rpoC1* (Table S2-1). In *G. officinalis*, 59 long repeats were found, of which 64.4% were located in IGS, 27.1% in coding regions, and 8.5% in the introns (Table S2-2). *G. officinalis* had a low number of repeats, but the distribution of repeats was similar to that of *G. orientalis* (Figure 3A).

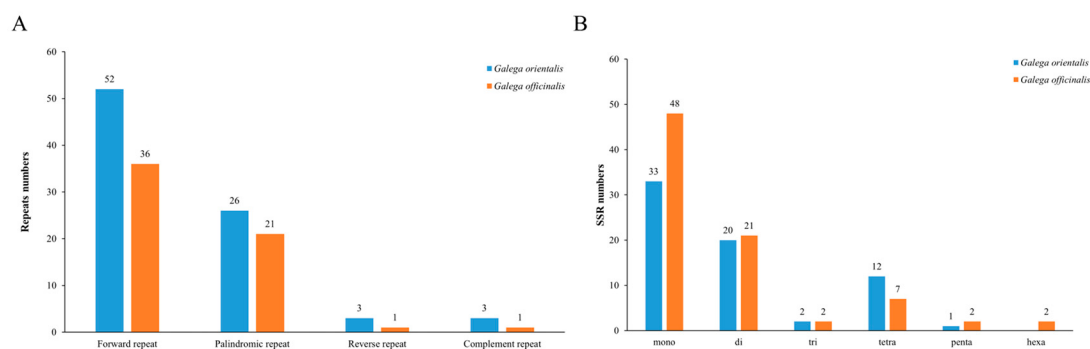


Figure 3. Analysis of repeat sequences in the chloroplast genome of *G. orientalis* and *G. officinalis* (A). SSR analysis of chloroplast genome of *G. orientalis* and *G. officinalis* (B).

SSRs are highly polymorphic and are valuable tools for genetic diversity studies and analysis of phylogenetic relationships among plant populations. A total of 68 and 82 SSRs (ranging in size from 10 to 24 bp and 10 to 36 bp) were detected in *G. orientalis* and *G. officinalis*, respectively. In *G. orientalis*, the 68 SSRs comprised 33 mono-repeats, 20 di-repeats, two tri-repeats, 12 tetra-repeats and one penta-repeat. The 82 SSRs in *G. officinalis* included 48 mono-repeats, 21 di-repeats, two tri-repeats, seven tetra-repeats, two penta-repeats and two hexa-repeats (Figure 3B). Overall, the frequency of mononucleotide repeats was higher in *G. officinalis*. In addition, mononucleotide repeats in *G. orientalis* were predominantly of the A/T motif (96.9%). All dinucleotide repeats were composed of the AT/TA motif. In total, 43 and 21 SSRs were distributed in IGS and CDS regions of *G. orientalis*, respectively (Figure 4A). However, no C/G motif was found in *G. officinalis* (Figure 4B), but CGTCCA and CTTATA hexanucleotide were found in *ycf1* coding gene and *clpP-Intron1*, respectively. Notably, the majority of SSRs were found in *ycf1* gene in both *G. orientalis* and *G. officinalis* (Table S3).

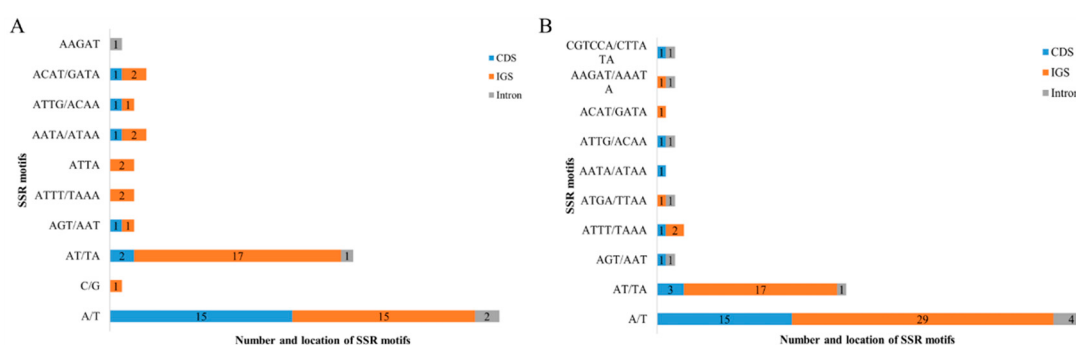


Figure 4. The types and distribution of SSRs along the chloroplast genomes of *G. orientalis* (A) and *G. officinalis* (B). Different locations, including CDS, IGS and intron regions, are represented as colored boxes.

3.3. Adaptive Evaluation Analysis

The Ka/Ks ratio was used to evaluate the degree of selection constraint on each gene and estimate the selective pressure of protein-coding genes. Ka/Ks > 1 indicates the gene is under positive selection, Ka/Ks = 1 indicates the gene is under neutral selection and Ka/Ks < 1 indicates purification selection.

In the present study, KaKs Calculator was used to calculate the non-synonymous (Ka) to synonymous (Ks) ratio (Ka/Ks) of 71 effectively shared protein-coding genes in *G. orientalis* and species belonging to six other genera (Table S4). The results showed that the Ka/Ks values between *G. orientalis* and Galegeae species ranged from 0 (*petN*) to 1.33757 (*petL*). Among them, the Ka/Ks value of *rpl20* gene between *G. orientalis* and *G. officinalis* was >1 (1.00749). However, the Ka/Ks values of *rpl20* among *G. orientalis* and other species of Galegeae were all lower than 0.5 (Table S4). The Ka/Ks values of *petL* and *ycf4* genes between *G. orientalis* and *Calophaca* were >1 (1.33757, 1.02613). *Caragana ycf4* gene ratio was also >1 (1.04846). These results indicated that *rpl20* gene was positively selected within *Galega*, but was under strong purification selection pressure in other Galegeae genera. *petL* and *ycf4* genes were positively selected between *G. orientalis* and *Calophaca*. The *ycf4* gene evolved under beneficial mutations between *G. orientalis* and *Caragana* (Figure 5).

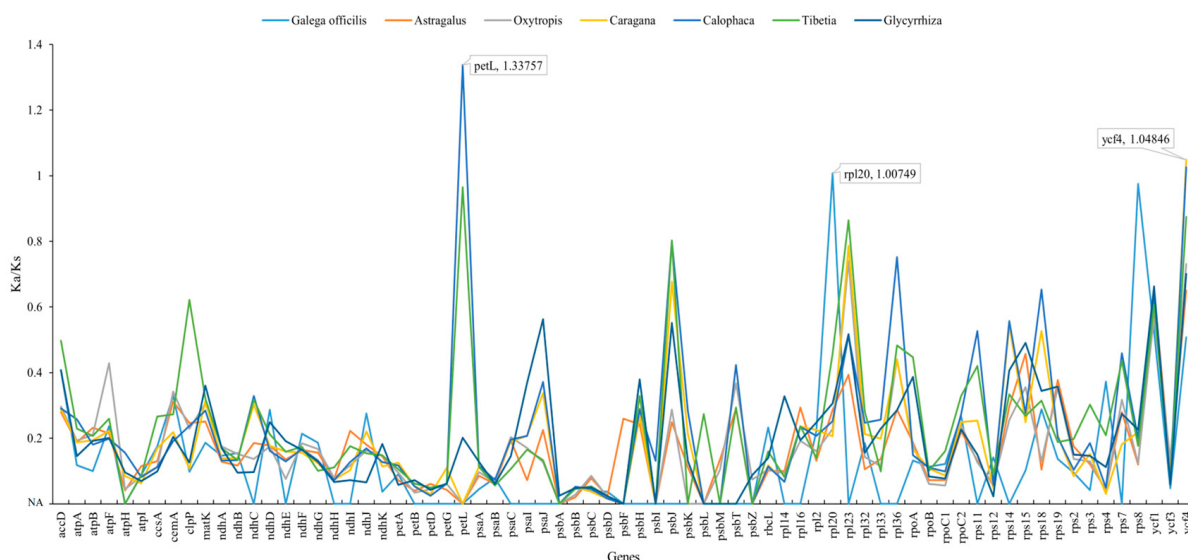


Figure 5. The Ka/Ks values of 71 shared genes between *G. orientalis* and other Galegeae species. The Ka/Ks values between *G. orientalis* and each genus are the average values of the Ka/Ks between *G. orientalis* and the species within the genus.

3.4. Sequence Divergence Analysis

The highly variable Cp genome is not only useful for identifying closely related species, but also can provide abundant phylogenetic information for evolutionary research. We compared the cp genomes of *G. orientalis* and 14 other Galegeae species. The results showed that the cp genome size in Galegeae species ranged from 122,210 bp (*Oxytropis diversifolia*) to 129,331 bp (*Caragana korshinskii*) and GC content ranged from 33.94% (*Astragalus galactites*) to 34.68% (*Tibetia liangshanensis*). *G. orientalis* differed significantly from other species in the length of the CDS and non-coding regions (Table S5). We used mVISTA to compare the cp genomes of 14 species of Galegeae with *G. orientalis* as a reference. In general, *G. orientalis* and *G. officinalis* showed highly sequence similarity in the coding regions, and only *ycf1* and *rpl23* genes were found to be significantly different and highly variable. In contrast, the sequence variation was significantly higher in conserved non-coding sequence regions (CNS) than in other regions. In addition to *ndhD-ccsA*, *ndhE-psaC*, *petB-petD*, *trnS-GGA-ycf3*, *trnK-UUU-matK* and *psaJ-rpl33*, almost all CNS regions showed variation. In addition, we found a high degree of sequence variation in the *accD*, *rpoC2* and *ycf4* genes. In the UTR region of the *rrn23* gene, the degree of regional variation in the CNS was similar between *G. orientalis* and *G. officinalis* but significantly different among 14 species. Overall, the degree of sequence variation in conserved non-coding sequence regions was significantly higher than that in the coding regions (Figure 6).

We also calculated *p*-values between *G. orientalis* and *G. officinalis* to further clarify the extent of nucleotide variation within these species (Figure S1). In general, the *p*-values of all genes in *Galega* were low, and only the *clpP* gene exceeded 1, indicating high nucleotide variation. The *clpP* gene could provide a wealth of information about the phylogeny of Galegeae species.

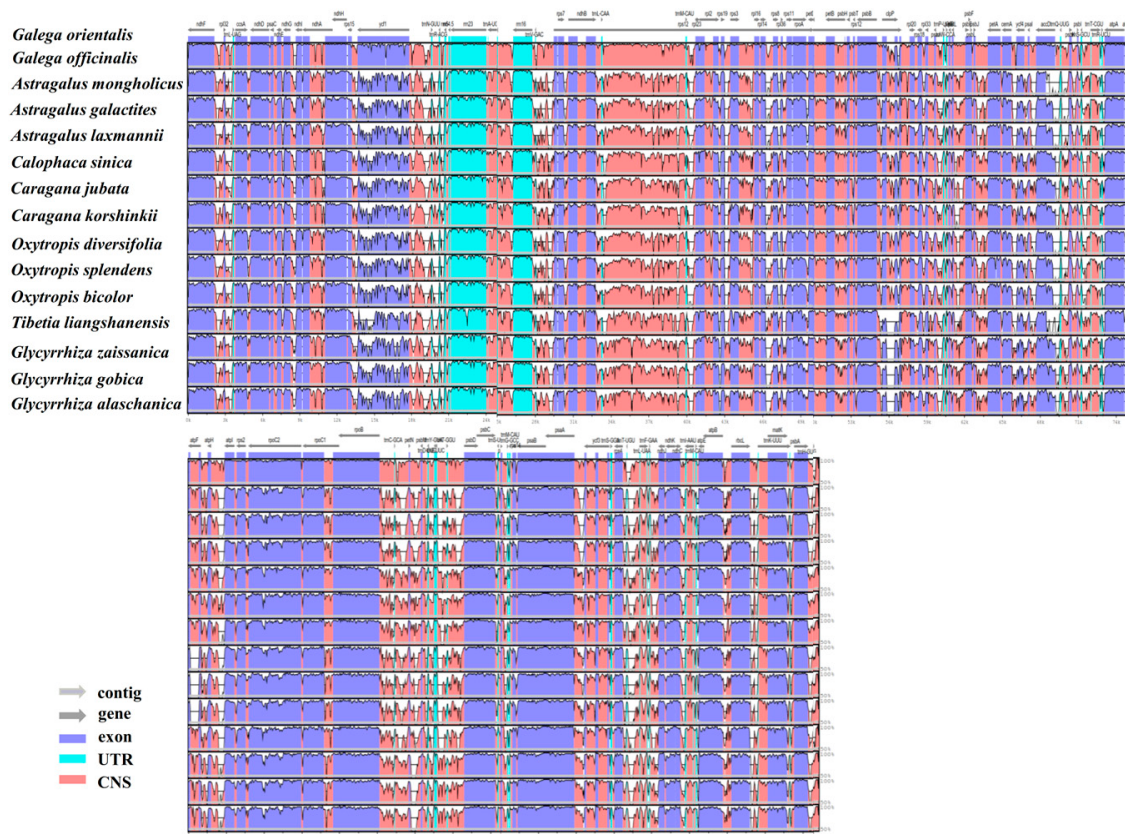


Figure 6. The comparison of 15 Galegeae species cp genomes by using mVISTA. *G. orientalis* as a reference, genomic regions are color-coded to indicate protein-coding regions, exons, UTRs, and CNS.

3.5. Phylogenetic Relationships Based on Chloroplast Genomes

To identify the phylogenetic position of *G. orientalis* in the Papilionoideae subfamily, we used 31 complete chloroplast genomes of Papilionoideae subfamily species (15 Galegeae species) for phylogenetic analysis (Figure 7A), with *Pterocarpus santalinus* and *Arachis hypogaea* as outgroups. The results showed the seven genera of Galegeae were differentiated, with bootstrap support values of most nodes >95. *G. orientalis* and *G. officinalis* were clustered into one clade, and were more closely related to *Lotus corniculatus*.

In addition, we selected two genes (*ycf1* and *clpP*) with strong sequence differences and high Pi values to construct phylogenetic trees of 15 Galegeae species by comparing chloroplast genomes of *G. orientalis* and *G. officinalis*. The results showed that the phylogenetic topology of Galegeae species based on *ycf1* (Figure 7B) and *clpP* (Figure 7C) was almost the same as the phylogenetic tree constructed using the complete chloroplast genome. Seven Galegeae genera were clearly discriminated based on *ycf1* and *clpP* genes. *Tibetia liangshanensis* and *Galega* species (*G. orientalis* and *G. officinalis*) were grouped together in one cluster (Figure 7B). However, as shown in Figure 7C, *T. liangshanensis* was clustered in a separate clade. Furthermore, all three evolutionary trees placed *Calophaca sinica* in the same group as *Caragana* (*C. jubata* and *C. korshinskii*), suggesting that they are more closely related.

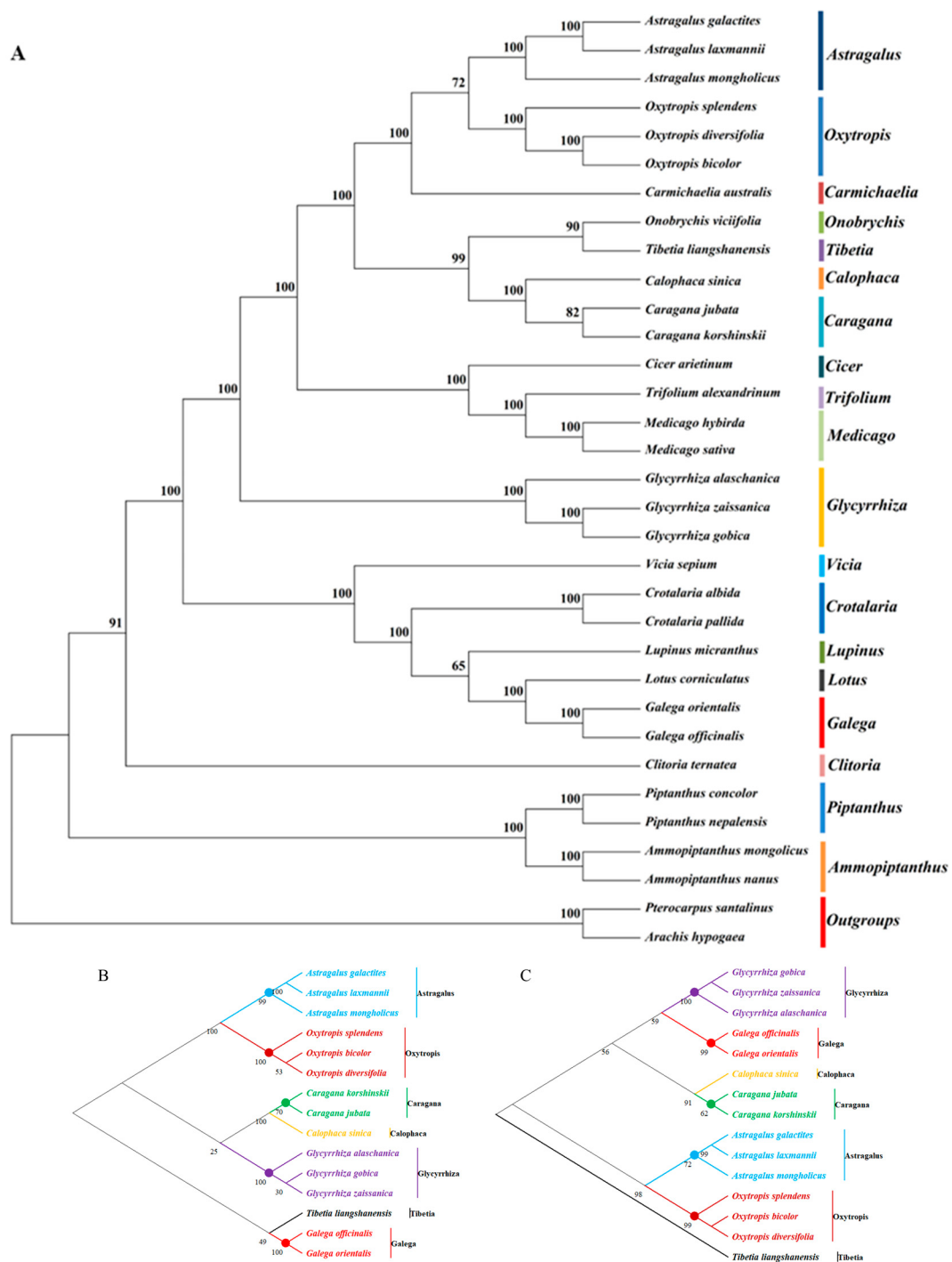


Figure 7. (A) Phylogenetic tree of the Leguminosae species (Galegeae) based on 31 complete chloroplast genomes and constructed using the *ycf1* gene (B) and *clpP* gene (C), with the *Galega* species indicated in red.

4. Discussion

4.1. General Features of *G. orientalis*cp Genome

In the present study, we sequenced and assembled the complete chloroplast genome of *G. orientalis*, which was in a length of 125,280 bp and structurally similar to other Papilionoideae species without an IR region [40]. The chloroplast genomes of *G. orientalis* and *G. officinalis* are highly conserved and similar in gene composition, structure and GC content, showing close species relationships within *Galega*. Regarding genetic differences, one intron in the *ycf1* gene was found deleted in the chloroplast genome of *G. officinalis* but

present in *G. orientalis*. The *ycf1* gene length of *G. orientalis* is 4128 bp with one 30-bp intron and two exons: exon1 and exon2 with length of 1881 and 2217 bp, respectively. There is no intron in *ycf1* gene in other known Galegeae chloroplast genomes. The reason for the existence of an intron in *ycf1* gene in Papilionoideae cp genome has not been reported. In land plants, introns are crucial for promoting gene transcription [41]. We found a short intron in *ycf1* gene of *G. orientalis*, indicating that the structural variation of intron in *ycf1* gene of *G. orientalis* can be used as a molecular marker to provide important information at low taxonomic level in phylogenetic studies of leguminous species [42]. The *clpP* gene also showed loss of introns. Studies have shown that *clpP* genes of most species in the IRLC lineage of Papilionoideae have lost one intron, and only *Glycyrrhiza glabra* has lost two introns. In the present study, the loss of one intron in the *clpP* gene of *G. orientalis* and *G. officinalis*, and two introns lost in the *clpP* gene of *Glycyrrhiza zaisanica*, *Glycyrrhiza gobica* and *Glycyrrhiza alaschanica*, indicating that introns lost is a common phenomenon could occur in *Glycyrrhiza* species [43,44]. In addition, we also found the loss of two introns in the *clpP* gene of *Tibetia liangshanensis* (Table S1).

Furthermore, we identified 74 protein-coding genes in the chloroplast genome of *Galega orientalis*. Compared with *G. officinalis*, *G. orientalis* cp genome was found lacked five unique genes, namely *trnE-UUC*, *rps16*, *rpl22*, *ycf2* and *pbf1*. This may be due to the nuclear translocation of these genes during the evolution of *G. orientalis*. A similar result was also revealed in the cp genome of *Cicer arietinum* with deletion of the *rpl22* gene [44]. Among them, *pbf1* gene is also absent in cp genome of other Galegeae species, but present in cp genome of *G. officinalis*. The reason may be that the deletion of IR region has led to gene rearrangement. Numerous studies have shown that IRLC lineage often results in gene deletion due to gene rearrangement, such as loss of *accD/rpl2* genes in some species of *Trifolium* [14], loss of *rpl23* gene in *Lathyrus* and *Vicia* [45], absence of *atpF* gene in *Cistanche* [46] and loss of *psbN* gene in *Haloxylon* [47]. The lack of IR region, size variation of IR region and many tandemly repeated sequences could cause chloroplast rearrangements, genetic structural variation or deletions.

Compared with angiosperms with intact chloroplast tetrad structure, *Galega* has fewer protein-coding genes, which may be caused by the deletion of IR region. Five of these genes (*accD*, *ycf1*, *ycf2*, *rpl23*, *infA*) were found only in certain species like *Trifolium* and *Medicago* [48]. *InfA*, which is considered to be the most transferable gene in the cp genome, and only exists in 24 angiosperms lineages, was absent in the cp genome of any known Galegeae species [49].

4.2. SSRs and Tandem Repeat Analysis

In the present study, we found high similarity in the types and number of tandem repeats between *G. orientalis* and *G. officinalis*. In addition, forward repeat was the most common type and most repeats were distributed in the IGS region, suggesting their potential as genetic markers for phylogenetic studies [50]. As the chloroplast genome is characterized by maternal inheritance and possess low frequency of genetic recombination, cp simple repeats (SSRs) can be used as effective molecular markers for genetic diversity research, species identification and phylogenetic analysis [51]. We observed abundant mononucleotide repeats in *G. orientalis* and *G. officinalis* chloroplast genomes with A/T motif repeats being dominant. An extremely strong A/T bias in SSR loci has also been observed in other legumes such as *Vigna radiate* [52] and *Arachis hypogaea* [53], which may result in base composition bias. At present, the potential role of cpSSRs in ecology and plant evolution has not been explored. Therefore, the cpSSRs detected in the two species could be used to assess the genetic relationships between different species and detect the polymorphisms in the Galegeae species at the population level.

4.3. Sequence Variation Analysis and Hotspots

Chloroplast genome is useful for the study of species evolution and classification. The CNS showed more obvious hypervariable regions than the other regions when *G. orientalis*

was compared with 14 other Galegeae species. In the protein-coding region (CDS), only *ycf1* and *rpl23* genes showed strong differences. At the same time, *ycf1* gene had a higher Ka/Ks value, suggesting that it is under less purifying selection pressure. The accelerated evolution rate of *petL* gene was observed in *Galega*. The *ycf4* gene has undergone adaptive evolution between *G. orientalis* and *Calophaca/Caragana*. It is also found in *Lathyrus*, *Pisum* and *Vavilovia* [54,55]. However, high sequence variation was found in the CNS regions except *ndhD-ccsA*, *ndhE-psaC*, *petB-petD*, *trnS-GGA-ycf3*, *trnK-UUU-matK* and *psaJ-rpl33*. These hypervariable genomic regions were considered to be useful markers for elucidating the phylogenetic relationships among the Galegeae species. The *ycf1* gene also showed potential as a genetic marker for *Quercus bawanglingensis* [56]. *atpB-rbcL* was used to assess the germplasm differentiation of *P. lunatus* [57], *trnW-CCA-petG* and *psbN-psbH* were also used for phylogenetic and domestication studies of *Astragalus* species [58]. The polymorphism of chloroplast genome is useful for evolutionary analysis of *G. orientalis*. Furthermore, we found that the *clpP* gene had the highest Pi value in the chloroplast genomes of *G. orientalis* and *G. officinalis* and could be used for phylogenetic analysis and population genetic study of *Galega* species.

4.4. Phylogenetic Relationships

Based on complete chloroplast genome sequences of 31 legume species, we determined the position of the genus in the Papilionoideae using phylogenetic analysis, and the topological structure was relatively consistent. *G. orientalis* and *G. officinalis* are closely related, clustered into one group, and were grouped in the same clade with *Lotus corniculatus*. These results indicate that besides the *Galega* family species, the *Lotus* also has a close evolutionary relationship with the *G. orientalis*. However, in Galegeae, the species of the *Tibetia* did not belong to the same clade as other genera, which was also confirmed in the phylogenetic tree constructed based on *clpP* gene. This may be due to the absence of IR region in the ancestral species of the Galegeae species, resulting in a series of alternations (including high rearrangements and duplications) during evolution, leading to differences among congeners [58].

5. Conclusions

In this study, the complete chloroplast genome of *Galega orientalis* was sequenced and annotated, and compared with the cp genome of other Galegeae species. The results showed that the length and structure of the chloroplast genome of *G. orientalis* were similar to those of *G. officinalis*, but the length of the non-coding region varied significantly. mVISTA analysis results showed that the mutation rate of the non-coding region was higher in *G. orientalis* than other Galegeae species, and the coding region was more conserved. Some hypervariable regions were identified, which could be used to distinguish species. In addition, we identified two hotspot genes (*ycf1* and *clpP*) that contained rich information for species identification and phylogenetic reconstruction of Galegeae, as well as provide insights into the evolution of Galegeae species.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes14010176/s1>, Table S1: Intron, exon statistics for *Galega orientalis* and other Galegeae species; Table S2: Repeat Sequences of *Galega orientalis* and *Galega officinalis*; Table S3: SSRs motifs of *Galega orientalis* and *Galega officinalis*; Table S4: Ka/Ks analysis between *Galega orientalis* and Galegeae species; Table S5: Structure of genome of Galegeae species; Figure S1: Calculation of Pi values between *Galega. orientalis* and *Galega. officinalis*.

Author Contributions: Data curation, J.F., Y.X. (Yi Xiong) and X.S.; Formal analysis, Y.X. (Yi Xiong) and T.L.; Funding acquisition, W.G. and X.M.; Investigation, J.F., T.L., Y.X. (Yanli Xiong) and J.Z.; Methodology, J.F. and X.L.; Project administration, W.G. and X.M.; Resources, L.Y.; Visualization, J.F.; Writing—original draft, J.F.; Writing—review & editing, J.F., X.S. and L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Sichuan Provincial Regional Innovation Integration and demonstration of soil fertility improvement technology for mowing grassland in southern hills (NK2022181006), Sichuan Provincial Key Science and Technology Project (2022YFQ0076, 2023SYSX0014).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy. The complete chloroplast genome sequence of *G. orientalis* was deposited at CNGBdb (accession number CNP0003740).

Acknowledgments: We acknowledge Sichuan Academy of Grassland Science and Genesky Biotechnologies Inc., Shanghai, 201315 for its technical support.

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

References

1. Ruhlman, T.A.; Jansen, R. Plastid Genomes of Flowering Plants: Essential Principles. *Methods Mol. Biol.* **2021**, *2317*, 3–47. [PubMed]
2. Tang, C.; Chen, X.; Deng, Y.; Geng, L.; Ma, J.; Wei, X. Complete chloroplast genomes of *Sorbus sensu stricto* (Rosaceae): Comparative analyses and phylogenetic relationships. *BMC Plant Biol.* **2022**, *22*, 1–21. [CrossRef] [PubMed]
3. Qian, R.; Ye, Y.; Hu, Q.; Ma, X.; Zheng, J. Complete Chloroplast Genome of *Gladiolus gandavensis* (Gladiolus) and Genetic Evolutionary Analysis. *Genes* **2022**, *13*, 1599. [CrossRef] [PubMed]
4. Li, J.; Li, J.; Ma, Y.; Kou, L.; Wei, J.; Wang, W. The complete mitochondrial genome of okra (*Abelmoschus esculentus*): Using nanopore long reads to investigate gene transfer from chloroplast genomes and rearrangements of mitochondrial DNA molecules. *BMC Genom.* **2022**, *23*, 481. [CrossRef] [PubMed]
5. Brouat, C.; Gielly, L.; McKey, D. Phylogenetic relationships in the genus *Leonardoxa* (Leguminosae: Caesalpinioideae) inferred from chloroplast trnL intron and trnL-trnF intergenic spacer sequences. *Am. J. Bot.* **2001**, *88*, 143–149. [CrossRef]
6. Ernst, K.; Michael, W. Phylogenetic Relationships in the Papilionoideae (Family Leguminosae) Based on Nucleotide Sequences of cpDNA (rbc L) and ncDNA (ITS 1 and 2). *Mol. Phylogenet. Evol.* **1997**, *8*, 65–88.
7. Vincent, M.; Anne, B. Phylogeny reconstruction in the Caesalpinieae grade (Leguminosae) based on duplicated copies of the sucrose synthase gene and plastid markers. *Mol. Phylogenet. Evol.* **2012**, *65*, 149–162.
8. Li, Y.; Zhang, L.; Wang, T.; Zhang, C.; Wang, R.; Zhang, D.; Xie, Y.; Zhou, N.; Wang, W.; Zhang, H.; et al. The complete chloroplast genome sequences of three lilies: Genome structure, comparative genomic and phylogenetic analyses. *J. Plant Res.* **2022**, *135*, 723–737. [CrossRef]
9. Li, C.; Liu, Y.; Lin, F.; Zheng, Y.; Huang, P. Characterization of the complete chloroplast genome sequences of six *Dalbergia* species and its comparative analysis in the subfamily of Papilionoideae (Fabaceae). *PeerJ* **2022**, *10*, e13570. [CrossRef]
10. Yamuna, S.; De-Long, G.; Wen-Qiang, W.; Liang, Z.; Sheng-Quan, X. The Complete Chloroplast Genomes of Two *Lespedeza* Species: Insights into Codon Usage Bias, RNA Editing Sites, and Phylogenetic Relationships in Desmodieae (Fabaceae: Papilionoideae). *Plants* **2019**, *9*, 51.
11. Azani, N.; Bruneau, A.; Wojciechowski, M.F.; Zarre, S. Miocene climate change as a driving force for multiple origins of annual species in *Astragalus* (Fabaceae, Papilionoideae). *Mol. Phylogenet. Evol.* **2019**, *137*, 210–221. [CrossRef]
12. Schwarz, E.N.; Ruhlman, T.A.; Sabir, J.S.M.; Hajrah, N.H.; Alharbi, N.S.; Al-Malki, A.L.; Bailey, C.D.; Jansen, R.K. Plastid genome sequences of legumes reveal parallel inversions and multiple losses of rps16 in papilionoids. *J. Syst. Evol.* **2015**, *53*, 458–468. [CrossRef]
13. Zha, X.; Wang, X.; Li, J.; Gao, F.; Zhou, Y. Complete chloroplast genome of *Sophora alopecuroides* (Papilionoideae): Molecular structures, comparative genome analysis and phylogenetic analysis. *J. Genet.* **2020**, *99*, 13. [CrossRef]
14. Yanli, X.; Yi, X.; Jun, H.; Qingqing, Y.; Junming, Z.; Xiong, L.; Zhixiao, D.; Jian, Y.; Yan, P.; Xinquan, Z.; et al. The Complete Chloroplast Genome of Two Important Annual Clover Species, *Trifolium alexandrinum* and *T. resupinatum*: Genome Structure, Comparative Analyses and Phylogenetic Relationships with Relatives in Leguminosae. *Plants* **2020**, *9*, 478.
15. Zolotarev, V.N.; Klimenko, I.A.; Kosolapov, V.M.; Korovina, V.L.; Antonov, A.A. Marker-Trait Association for Breeding Fodder Galega (*Galega orientalis* Lam). *Russ. Agric. Sci.* **2022**, *48*, 270–275. [CrossRef]
16. Lang, J.; Zikmundová, B.; Hájek, J.; Barták, M.; Váczi, P. The Effects of Foliar Application of Phenoxy and Imidazoline Family Herbicides on the Limitation of Primary Photosynthetic Processes in *Galega orientalis* L. *Agronomy* **2021**, *12*, 96. [CrossRef]
17. Balakina, A.A.; Nefedieva, E.E.; Khramova, V.N.; Gorlov, I.F.; Slozhenkina, M.I.; Poletayev, R.N. The structure of the seed coat of valuable forage plant *Galega orientalis* Lam. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *677*, 032077. [CrossRef]
18. Povilaitis, V.; Šlepetienė, A.; Šlepetys, J.; Lazauskas, S.; Tilvikienė, V.; Amalevičiūtė, K.; Feizienė, D.; Feiza, V.; Liaudanskienė, I.; Cesevičienė, J.; et al. The productivity and energy potential of alfalfa, fodder galega and maize plants under the conditions of the nemoral zone. *Acta Agric. Scand. Sect. B—Soil Plant Sci.* **2016**, *66*, 259–266. [CrossRef]

19. Żarczyński, P.J.; Sienkiewicz, S.; Wierzbowska, J.; Krzebietke, S.J. Fodder Galega—A Versatile Plant. *Agronomy* **2021**, *11*, 1797. [\[CrossRef\]](#)
20. Darbyshire, S.J.; Francis, A.; Bromfield, E.S.P.; Mechanda, S. The Biology of Canadian Weeds: 158. *Galega officinalis* L. *Can. J. Plant Sci.* **2021**, *102*, 160–185. [\[CrossRef\]](#)
21. Eryashev, A.P.; Timoshkin, O.A. The Effect of Plant Protection Products and Albite on the Productivity and Quality of Fodder Galega (*Galega Orientalis*). *Int. J. Eng. Adv. Technol.* **2019**, *9*, 3286–3292. [\[CrossRef\]](#)
22. Shymanska, O.V.; Vergun, O.M.; Rakhmetov, D.B.; Brindza, J. Antiradical activity of plant extracts of *Galega officinalis* L. and *G. orientalis* Lam. *Plant Introd.* **2018**, *78*, 12–19.
23. Jun, L.; Hongwen, G.; Jishan, J.; Nikolay, D.; Vladimir, C.; Zan, W.; Xuemin, W. Overexpression of the *Galega orientalis* gibberellin receptor improves biomass production in transgenic tobacco. *Plant Physiol. Biochem.* **2013**, *73*, 1–6.
24. Bożena, C.; Barbara, M. The structure of fungal population from *Galega orientalis* root system formed as the result of fertilization. *Acta Mycol.* **2014**, *35*, 311–321.
25. Tian, C.; Li, X.; Wu, Z.; Li, Z.; Hou, X.; Li, F.Y. Characterization and Comparative Analysis of Complete Chloroplast Genomes of Three Species from the Genus *Astragalus* (Leguminosae). *Front. Genet.* **2021**, *12*, 705482. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Guo, X.; Chen, F.; Gao, F.; Li, L.; Liu, K.; You, L.; Hua, C.; Yang, F.; Liu, W.; Peng, C.; et al. CNSA: A data repository for archiving omics data. *Database J. Biol. Databases Curation* **2020**, *2020*, baaa055. [\[CrossRef\]](#)
27. Chen, F.Z.; You, L.J.; Yang, F.; Wang, L.N.; Guo, X.Q.; Gao, F.; Hua, C.; Tan, C.; Fang, L.; Shan, R.Q.; et al. CNGBdb: China National GeneBank DataBase. *Yi Chuan Hered.* **2020**, *42*, 799–809.
28. Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P.A. metaSPAdes: A new versatile metagenomic assembler. *Genome Res.* **2017**, *27*, 824–834. [\[CrossRef\]](#)
29. 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. [\[CrossRef\]](#)
30. Laslett, D.; Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **2004**, *32*, 11–16. [\[CrossRef\]](#)
31. Lowe, T.M.; Chan, P. tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **2016**, *44*, W54–W57. [\[CrossRef\]](#)
32. 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. [\[CrossRef\]](#)
33. 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. [\[CrossRef\]](#)
34. Thiel, T.; Michalek, W.; Varshney, R.K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *TAG Theor. Appl. Genet. Theor. Angew. Genet.* **2003**, *106*, 411–422. [\[CrossRef\]](#)
35. Kane, K.W. KaKs_Calculator: Calculating Ka and Ks Through Model Selection and Model Averaging. *Genom. Proteom. Bioinform.* **2006**, *4*, 259–263.
36. Katoh, K.; Standley, D. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [\[CrossRef\]](#)
37. 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. [\[CrossRef\]](#)
38. 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. [\[CrossRef\]](#)
39. Nguyen, L.-T.; Schmidt, H.; von Haeseler, A.; Minh, B. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [\[CrossRef\]](#)
40. Charboneau, J.L.M.; Cronn, R.C.; Liston, A.; Wojciechowski, M.F.; Sanderson, M.J. Plastome structural evolution and homoplastic inversions in Neo-Astragalus (Fabaceae). *Genome Biol. Evol.* **2021**, *13*, evab215. [\[CrossRef\]](#)
41. Moghaddam, M.; Ohta, A.; Shimizu, M.; Terauchi, R.; Kazempour-Osaloo, S. The complete chloroplast genome of *Onobrychis gaubae* (Fabaceae-Papilionoideae): Comparative analysis with related IR-lacking clade species. *BMC Plant Biol.* **2022**, *22*, 75. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Zhang, D.; Li, K.; Gao, J.; Liu, Y.; Gao, L.-Z. The Complete Plastid Genome Sequence of the Wild Rice *Zizania latifolia* and Comparative Chloroplast Genomics of the Rice Tribe Oryzaeae, Poaceae. *Front. Ecol. Evol.* **2016**, *4*, 88. [\[CrossRef\]](#)
43. Sabir, J.; Schwarz, E.; Ellison, N.; Zhang, J.; Baeshen, N.A.; Mutwakil, M.; Jansen, R.; Ruhlman, T. Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. *Plant Biotechnol. J.* **2014**, *12*, 743–754. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Robert, K.J.; Martin, F.W.; Elumalai, S.; Seung-Bum, L.; Henry, D. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of rps12 and clpP intron losses among legumes (Leguminosae). *Mol. Phylogenet. Evol.* **2008**, *48*, 1204–1217.
45. Li, C.; Zhao, Y.; Xu, Z.; Yang, G.; Peng, J.; Peng, X. Initial Characterization of the Chloroplast Genome of *Vicia sepium*, an Important Wild Resource Plant, and Related Inferences About Its Evolution. *Front. Genet.* **2020**, *11*, 73. [\[CrossRef\]](#)

46. Li, X.; Zhang, T.C.; Qiao, Q.; Ren, Z.; Zhao, J.; Yonezawa, T.; Hasegawa, M.; Crabbe, M.J.; Li, J.; Zhong, Y. Complete chloroplast genome sequence of holoparasite *Cistanche deserticola* (Orobanchaceae) reveals gene loss and horizontal gene transfer from its host *Haloxylon ammodendron* (Chenopodiaceae). *PLoS ONE* **2013**, *8*, e58747. [[CrossRef](#)]
47. Dong, W.; Xu, C.; Li, D.; Jin, X.; Li, R.; Lu, Q.; Suo, Z. Comparative analysis of the complete chloroplast genome sequences in psammophytic *Haloxylon* species (Amaranthaceae). *PeerJ* **2016**, *4*, e2699. [[CrossRef](#)]
48. Millen, R.S.; Olmstead, R.G.; Adams, K.L.; Palmer, J.D.; Lao, N.T.; Heggie, L.; Kavanagh, T.A.; Hibberd, J.M.; Gray, J.C.; Morden, C.W.; et al. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* **2001**, *13*, 645–658. [[CrossRef](#)]
49. Xiaoli, T.; Lichao, M.; Zhengshe, Z.; Wenxian, L.; Zhipeng, L. Characterization of the complete chloroplast genome of alfalfa (*Medicago sativa*) (Leguminosae). *Gene Rep.* **2017**, *6*, 67–73.
50. Vello, J. Phylogenetic relationships among sections *Vicia*, *Sepium* and *Lathyroides* of *Vicia* subgenus *Vicia*: Isozyme evidence. *Biochem. Syst. Ecol.* **2015**, *62*, 186–193.
51. Gregory, L.W.; Hanna, E.D.; Alenda, B.; Lavanya, C.; Lisa, E.W. A Review of the Prevalence, Utility, and Caveats of Using Chloroplast Simple Sequence Repeats for Studies of Plant Biology. *Appl. Plant Sci.* **2014**, *2*, 1400059.
52. Yang, M.T.; Chen, S.L.; Lin, C.Y.; Chen, Y.M. Chilling stress suppresses chloroplast development and nuclear gene expression in leaves of mung bean seedlings. *Planta* **2005**, *221*, 374–385. [[CrossRef](#)]
53. Prabhudas, S.K.; Prayaga, S.; Madasamy, P.; Natarajan, P. Shallow Whole Genome Sequencing for the Assembly of Complete Chloroplast Genome Sequence of *Arachis hypogaea* L. *Front. Plant Sci.* **2016**, *7*, 1106. [[CrossRef](#)]
54. Magee, A.M.; Aspinall, S.; Rice, D.W.; Cusack, B.P.; Semon, M.; Perry, A.S.; Stefanovic, S.; Milbourne, D.; Barth, S.; Palmer, J.D.; et al. Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Res.* **2010**, *20*, 1700–1710. [[CrossRef](#)]
55. Moghaddam, M.; Kazempour-Osaloo, S. Extensive survey of the *ycf4* plastid gene throughout the IRLC legumes: Robust evidence of its locus and lineage specific accelerated rate of evolution, pseudogenization and gene loss in the tribe Fabeae. *PLoS ONE* **2020**, *15*, e0229846. [[CrossRef](#)]
56. 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](#)]
57. Tian, S.; Lu, P.; Zhang, Z.; Wu, J.Q.; Zhang, H.; Shen, H. Chloroplast genome sequence of Chongming lima bean (*Phaseolus lunatus* L.) and comparative analyses with other legume chloroplast genomes. *BMC Genom.* **2021**, *22*, 194. [[CrossRef](#)]
58. Lei, W.; Ni, D.; Wang, Y.; Shao, J.; Wang, X.; Yang, D.; Wang, J.; Chen, H.; Liu, C. Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of *Astragalus membranaceus*. *Sci. Rep.* **2016**, *6*, 21669. [[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.