

Article

# Complete Chloroplast Genome Characterization of *Oxalis Corniculata* and Its Comparison with Related Species from Family Oxalidaceae

Lubna <sup>1</sup>, Sajjad Asaf <sup>2,\*</sup> , Rahmatullah Jan <sup>3</sup> , Abdul Latif Khan <sup>2,\*</sup> and In-Jung Lee <sup>3,\*</sup>

<sup>1</sup> Department of Botany, Garden Campus, Abdul Wali Khan University, Mardan 23200, Pakistan; lubnabilal68@gmail.com

<sup>2</sup> Natural and Medical Sciences Research Center, University of Nizwa, Nizwa 616, Oman

<sup>3</sup> School of Applied Biosciences, Kyungpook National University, Daegu 41566, Korea; rehmatbot@yahoo.com

\* Correspondence: sajadasif2000@gmail.com (S.A.); latifepm78@yahoo.co.uk (A.L.K.); ijlee@knu.ac.kr (I.-J.L.)

Received: 27 May 2020; Accepted: 17 July 2020; Published: 23 July 2020



**Abstract:** *Oxalis corniculata* L. (family Oxalidaceae) is a small creeper wood sorrel plant that grows well in moist climates. Despite being medicinally important, little is known about the genomics of this species. Here, we determined the complete chloroplast genome sequence of *O. corniculata* for the first time and compared it with other members of family Oxalidaceae. The genome was 152,189 bp in size and comprised of a pair of 25,387 bp inverted repeats (IR) that separated a large 83,427 bp single copy region (LSC) and a small 16,990 bp single copy region (SSC). The chloroplast genome of *O. corniculata* contains 131 genes with 83 protein coding genes, 40 tRNA genes, and 8 rRNA genes. The analysis revealed 46 microsatellites, of which 6 were present in coding sequences (CDS) regions, 34 in the LSC, 8 in the SSC, and 2 in the single IR region. Twelve palindromic repeats, 30 forward repeats, and 32 tandem repeats were also detected. Chloroplast genome comparisons revealed an overall high degree of sequence similarity between *O. corniculata* and *O. drummondii* and some divergence in the intergenic spacers of related species in Oxalidaceae. Furthermore, the seven most divergent genes (*ccsA*, *clpP*, *rps8*, *rps15*, *rpl22*, *matK*, and *ycf1*) among genomes were observed. Phylogenomic characterization on the basis of 60 shared genes revealed that *O. corniculata* is closely related to *O. drummondii*. The complete *O. corniculata* genome sequenced in the present study is a valuable resource for investigating the population and evolutionary genetics of family Oxalidaceae and can be used to identify related species.

**Keywords:** Oxalidaceae; chloroplast genome comparison; inverted repeats; divergence; SSRs; phylogeny

## 1. Introduction

The largest genus of family Oxalidaceae is *Oxalis* L., which is distributed mostly in Southern Africa and South America, and comprises of more than 500 species. About one-half of the total species (>200 spp) growing in Southern Africa share a bulbous or tuberous in herbaceous taxa [1–3]. A huge morphological variation have been observed among approximately 250 species in South America, where this genus seems to have originated and diversified [2,4]. However, *Oxalis* sections Corniculatae DC. consist of creeping herbs, and many of the species grows in the temperate and humid areas of the Americas [5]. A cytogenetic study showed that Corniculatae tends to be categorized into two sub-groups: (1) a large number of diploid and polyploid species with a base chromosome number of  $x = 6$ , symmetrical karyotypes, and medium and small chromosomes size, and (2) a minor number of diploid species having  $x = 5$ , more asymmetrical karyotypes, and average to large chromosomes [3,6,7]. The taxonomy has been suffering with similarities in phenotypes across different species.

Among species from genus *Oxalis*, *Oxalis corniculata* L. is a small creeping herb that is adaptive to moist conditions with yellow flowers and cylindrical fruits [4]. The origin of *O. corniculata* L. has been unknown until now. *O. corniculata* was first described by Carl Linnaeus from the Mediterranean region and various others suggested that this region is the native range [4,8]. Currently, among vascular plant species, *O. corniculata* has the third largest distribution [9]. The early spread and global distribution of *O. corniculata* has contributed to the difficulty in identifying its native range [4]. It is highly persistent in the horticulture industry and commonly found in gardens and as a hitchhiker in plant pots in gardens and nurseries [10]. Additionally, *O. corniculata* had medicinal values to treat various infectious diseases [11]. It has been known as antifungal, antibacterial, anti-inflammatory, anthelmintic, antidiuretic, astringent, emmenagogue, depurative, febrifuge, lithontriptic, stomachic, and styptic [12]. It has sticky seeds, explosive capsules, persistent flowers, and a brief life cycle, which make it an effective colonizer and constant weed. *O. corniculata* is not a strong competitor; however, its abundance and enormous range increase its overall impact as a weed.

Chloroplasts (cp) are important plant organelles that carry out photosynthesis and the biosynthesis of amino acid, fatty acids, and pigments [13,14]. Similarly, cp genomes have proven to be a valuable resource for species identification, plant phylogenetics, population genetics, and genetic engineering. Its DNA is inherited maternally in not all but in the majority of angiosperm species [15]. As a result of its inheritance approach, cp DNA plays a significant role in population genetics and molecular evolution. Therefore, its DNA can not only be used for species discrimination but can also be used to solve many questions related to taxonomy [15,16]. Chloroplasts contain their own independent genomes and genetic systems, and DNA replication and transmission to daughter organelles result in a cytoplasmic inheritance of characteristics associated with primary events in photosynthesis [17,18]. The cp genome structure of angiosperms is circular—about 120–217 kb in size—and contain a quadripartite conformation [19,20]. It has small single copy (SSC) and large single copy (LSC) regions that are generally segregated by double copies of an inverted repeat region (IRA and IRb) [20]. In terms of gene order and content, the cp genome is usually conserved in various families of the angiosperm, such as Campanulaceae, Fabaceae, Geraniaceae, and Oleaceae [21–23]. Due to its conserved structure, small size, and recombination-free nature, the chloroplast genome is broadly used in plant phylogenetic analysis [24,25]. Similarly, the cp genome has a highly conserved structure that simplifies sequencing and primer designing, and chloroplast DNA can be used for the identification of plants as a barcode [26,27].

Molecular phylogenetic tools were broadly used to assess previous taxonomy and evolutionary processes. The number of chromosomes and contents of DNA (cytogenetic data) are usually associated with phylogenetic trees in order to understand the karyological differences involved in the group diversification [28,29]. Molecular phylogenomic studies using cp genome sequence data from the genes and slowly evolving inverted repeat regions have been applied to reveal the deep-level evolutionary relationships of plant taxa [30], producing robust phylogenies that are corroborated by sequence data from mitochondrial and nuclear genomes [31]. With the advancement of different genomic methods and tools, next-generation technologies have provided the rapid sequencing of different cp genomes from both angiosperms and gymnosperms in recent years, which enabled the confirmation of evolutionary relationships and detailed phylogenetic classifications to be conducted at the group, family, genus, and even species levels in plants [32,33]. Therefore, cp genome-scale data have increasingly been used to infer phylogenetic relationships at high taxonomic levels, and even at lower levels, great progress has been made [27,34–36].

The phylogenetic analysis of various *Oxalis* taxa has been reported previously by using nuclear Internal Transcribed Spacer (ITS) regions [37]; however, more detailed insight is still missing, keeping in view the complex taxonomy of *O. corniculata*. It is variable cytologically and genetically, but it is also phenotypically plastic [38,39]. Its taxonomy is complicated by the description of many subspecific taxa and other species now considered to be synonyms [5]. Furthermore, there has been taxonomic confusion of *O. corniculata* with closely related species, such as *O. stricta* and *O. dillenioides* [4]. Similarly,

there is little information available on their genetic structure, especially their chloroplast genomes or their detailed phylogenetic placement. Hence, the current study was performed with the aim to sequence and analyze the complete chloroplast genome of *O. corniculata* and compare it with related species from the family Oxalidaceae (*Oxalis drummondii*, *Averrhoa carambola* and *Cephalotus follicularis*). We also aimed to elucidate and compare the global pattern of structural variation in the cp genome of *O. corniculata* and related three species. In addition, we compared the IR region contraction and expansion, intron contents, regions of high sequence divergence, and phylogenomic of *O. corniculata* with related species cp genomes to reveal more insight regarding the comparative genome architecture.

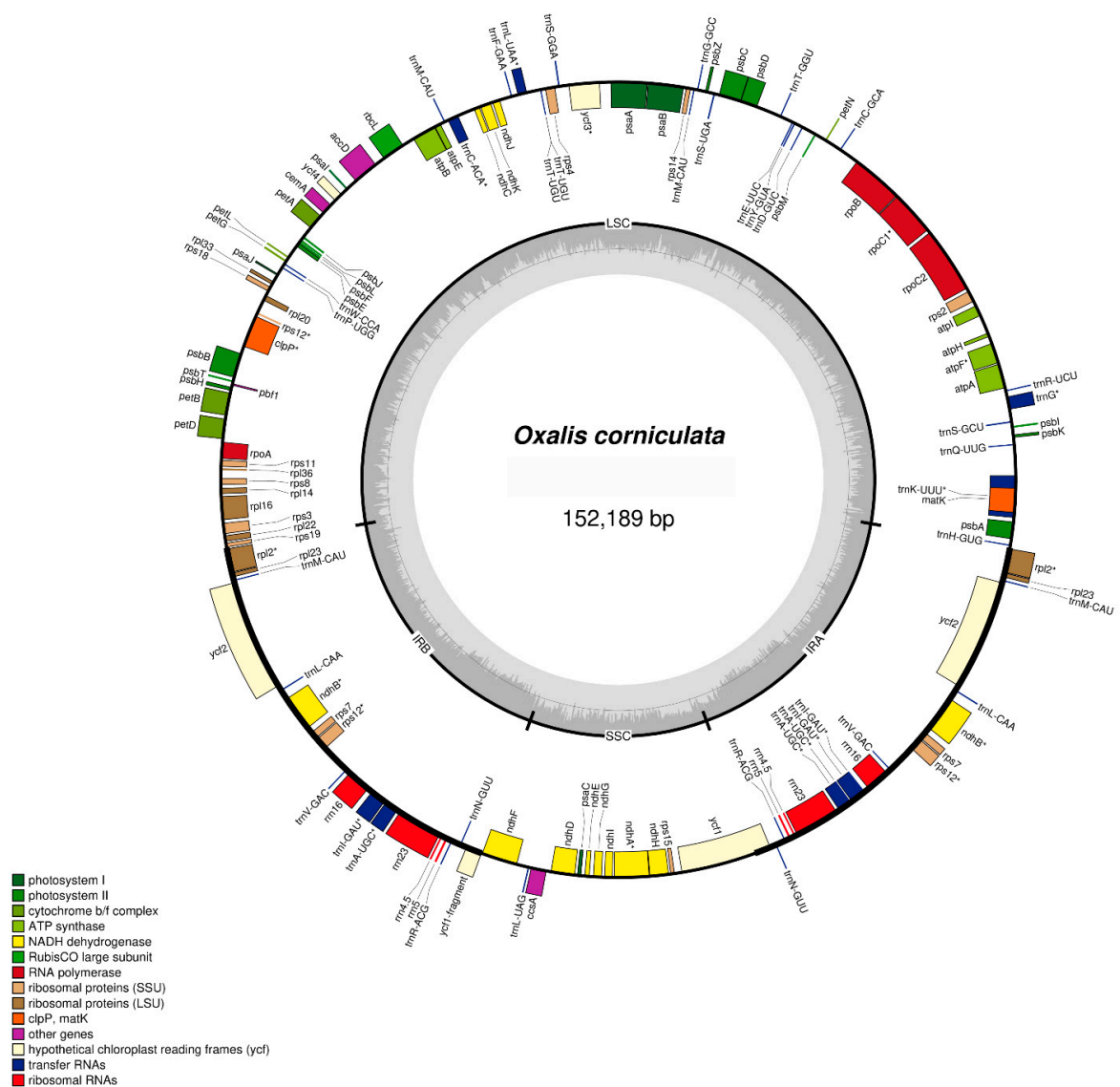
## 2. Results

### 2.1. Chloroplast Genome Structure of *O. corniculata*

The chloroplast genome of *O. corniculata* is 152,189 bp and displays a distinctive quadripartite structure with a pair of 25,387 bp inverted repeats (IRs) that separate 83,427 bp single copy regions and 16,990 bp single copy regions (Figure 1). The total GC content is 36.7% with uneven distribution across the whole genome. The GC contents of IRs are higher (42.6%) than the large single copy and small single copy regions (34.4% and 30.3%, respectively). Furthermore, the *O. corniculata* cp genome consists of 131 genes and among all these genes, 82 are protein-coding genes, 40 are tRNA, and 8 are rRNA (Figure 1; Table 1). The protein-coding genes present in the *O. corniculata* cp genome included nine genes for large ribosomal proteins (*rpl2*, 14, 16, 20, 22, 23, 32, 33, 36), 11 genes for small ribosomal proteins (*rps2*, 3, 4, 7, 8, 11, 12, 14, 15, 18, 19), five genes for photosystem I (*psaA*, B, C, I, J), 15 genes for photosystem II (*psbA*, B, C, D, E, F, H, I, J, K, L, M, T, Z) and six genes (*atpA*, B, E, F, H, I) for ATP synthesis and the electron transport chain (Figure 1, Table 2). Similarly, 17 genes contain introns (11 protein-coding genes and 6 tRNA genes), of which three comprised of two introns (*rps12*, *clpP*, *ycf3*), while the rest have a single intron (Table 3). The small ribosomal protein 12 gene (*rps12*) is trans-spliced with a single intron. Its 5' exon is located in the LSC region, while its 3' exon is located in the IRb region and duplicated in the IRa region (Figure 1). The largest intron was present in *trnK*-UUU (2558 bp), whilst *trnL*-UAA contains the smallest intron (492 bp) (Table 3). The protein-coding region accounts for 52%, while the tRNA and rRNA regions constitute 1.99% and 5.94%, respectively, in the cp genome. The length of the protein-coding region is about 79,239 bp while those of tRNA and rRNA are 3042 bp and 9048 bp, respectively. Similarly, the *rps16* gene, which is found in most angiosperm plastid genomes, is absent in *O. corniculata*. Furthermore, *infA* is also absent in the *O. corniculata* cp genome.

**Table 1.** Summary of complete chloroplast genome of *O. corniculata* and its comparison with related species from family Oxalidaceae.

	<i>A. carambola</i>	<i>C. follicularis</i>	<i>O. drummondii</i>	<i>O. corniculata</i>
Size (bp)	155,965	142,706	152,112	152,189
Overall GC contents	36.5	37.1	36.5	36.7
LSC size in bp	86,218	81,071	83,341	83,427
SSC size in bp	17,497	8739	16915	16,990
IR size in bp	25,626	25,949	25,429	25,387
Protein coding regions size in bp	78,528	66,672	77,826	79,239
tRNA size in bp	2790	2790	2790	3042
rRNA size in bp	9046	9050	9046	9048
Number of genes	131	122	127	131
Number of protein coding genes	83	71	82	83
Number of rRNA	8	8	8	8
Number of tRNA	37	37	37	40



**Figure 1.** Gene map of the *O. corniculata* cp genome. Thick lines in the red area indicate the extent of the inverted repeat regions (IRa and IRb; 25,387 bp), which separate the genome into small (SSC; 16,990 bp) and large (LSC; 83,427 bp) single copy regions. Genes drawn inside the circle are transcribed clockwise, and those outside are transcribed counter clockwise. Genes belonging to different functional groups are color-coded. The dark gray in the inner circle corresponds to the GC content, and the light gray corresponds to the AT content.

**Table 2.** Genes in the sequenced *O. corniculata* chloroplast genome.

Category	Group of Genes	Name of Genes
Self-replication	Large subunit of ribosomal proteins	<i>rpl2</i> *, <i>rpl14</i> , <i>rpl16</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> *, <i>rpl33</i> , <i>rpl36</i>
	Small subunit of ribosomal proteins	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> *, <i>rps8</i> , <i>rps11</i> , <i>rps12</i> *, <i>rps14</i> , <i>rps15</i> , <i>rps18</i> , <i>rps19</i>
	DNA dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>rpoC2</i>
	rRNA genes	<i>rrn4.5</i> , 5, 16, 23
	tRNA genes	<i>trnA</i> -UGC *, <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC, <i>trnF</i> -GAA, <i>trnM</i> -CAU, <i>trnG</i> -GCC *, <i>trnH</i> -GUG, <i>trnI</i> -GAU *, <i>trnK</i> -UUU, <i>trnL</i> -CAA *, <i>trnL</i> -UAA, <i>trnL</i> -UAG, <i>trnM</i> -CAU *, <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> -GGU, <i>trnT</i> -UGU *, <i>trnV</i> -GAC *, <i>trnW</i> -CCA, <i>trnY</i> -GUA
Photosynthesis	Photosystem I	<i>psaA</i> , B, C, I, J
	Photosystem II	<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>
	NadH oxidoreductase	<i>ndhA</i> , <i>ndhB</i> *, <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>
	Cytochrome b6/f complex	<i>petA</i> , <i>petB</i> , <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>
	ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i>
	Rubisco	<i>rbcL</i>
Other genes	Maturase	<i>matK</i>
	Protease	<i>clpP</i>
	Envelop membrane protein	<i>cemA</i>
	Subunit Acetyl- CoA-Carboxylate	<i>accD</i>
	c-type cytochrome synthesis gene	<i>ccsA</i>
Unknown	Conserved Open reading frames	<i>ycf1</i> *, 2 *, 3, 4,

\* Duplicated genes.

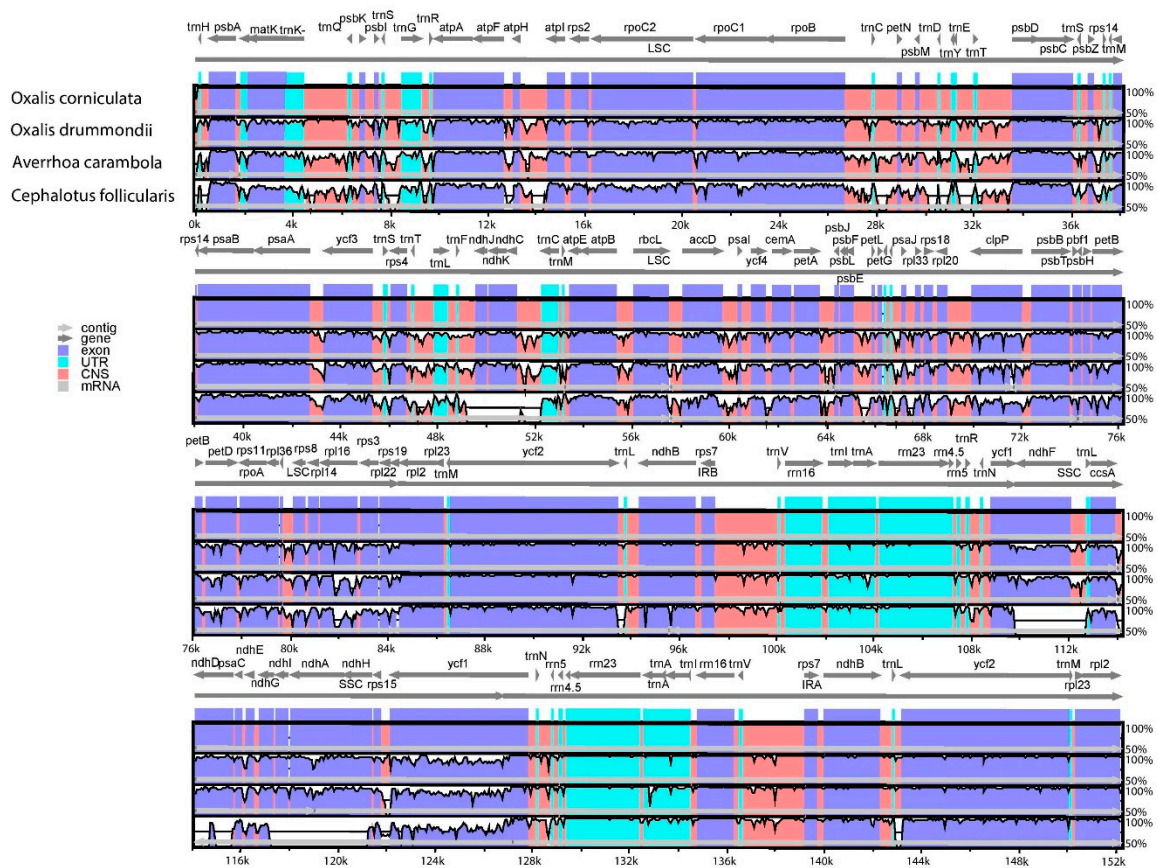
**Table 3.** The genes with introns in the four species chloroplast genome and the length of exons and introns.

Gene	Location	Exon I (bp)				Intron 1 (bp)				Exon II (bp)				Intron II (bp)				Exon III (bp)			
		<i>O. c</i>	<i>O. d</i>	<i>A. c</i>	<i>C. f</i>	<i>O. c</i>	<i>O. d</i>	<i>A. c</i>	<i>C. f</i>	<i>O. c</i>	<i>O. d</i>	<i>A. c</i>	<i>C. f</i>	<i>O. c</i>	<i>O. d</i>	<i>A. c</i>	<i>C. f</i>	<i>O. c</i>	<i>O. d</i>	<i>A. c</i>	<i>C. f</i>
<i>atpF</i>	LSC	144	145	145	145	717	714	714	717	411	410	410	410								
<i>petB</i>	LSC	6	6	6	6	746	746	785	793	645	642	642	642								
<i>PetD</i>	LSC	8	8	7	7	746	778	707	721	645	475	473	473								
<i>rpl2</i> *	IR	391	391	391	391	658	661	659	671	434	434	434	525								
<i>rps16</i>	LSC	–	–	40	40	–	–	935	929	–	–	224	227								
<i>rpoC1</i>	LSC	430	432	453	459	727	729	751	759	1634	1629	1608	1617								
<i>rps12</i> *	IR/LSC			391		527				232		434						25			
<i>clpP</i>	LSC	71	71	69	69	808	812	833	833	289	289	291	291	603	612	568		228	228	228	
<i>ndhA</i>	SSC	557	557	559	–	1077	1067	1037	–	541	541	557	–								
<i>ndhB</i> *	IR	777	777	777	–	685	685	685	–	756	756	756	–								
<i>ycf3</i>	LSC	124	124	126	126	718	712	718	714	228	230	228	228	684	678	685	678	155	153	153	150
<i>trnA-UGC</i> *	IR	38	38	38	38	799	798	763	811	35	35	35	35								
<i>trnI-GAU</i> *	IR	37	37	37	37	925	924	932	946	35	35	35	35								
<i>trnL-UAA</i>	LSC	35	35	35	35	497	497	507	492	50	50	50	50								
<i>trnK-UUU</i>	LSC	37	34	37	35	2519	2515	2545	2558	35	37	35	37								
<i>trnG-GCC</i>	LSC	71	71	71	71	–	–	–	–	–	–	–	–								
<i>trnV-UAC</i>	LSC	–	–	35	35	–	–	617	618	–	–	39	39								

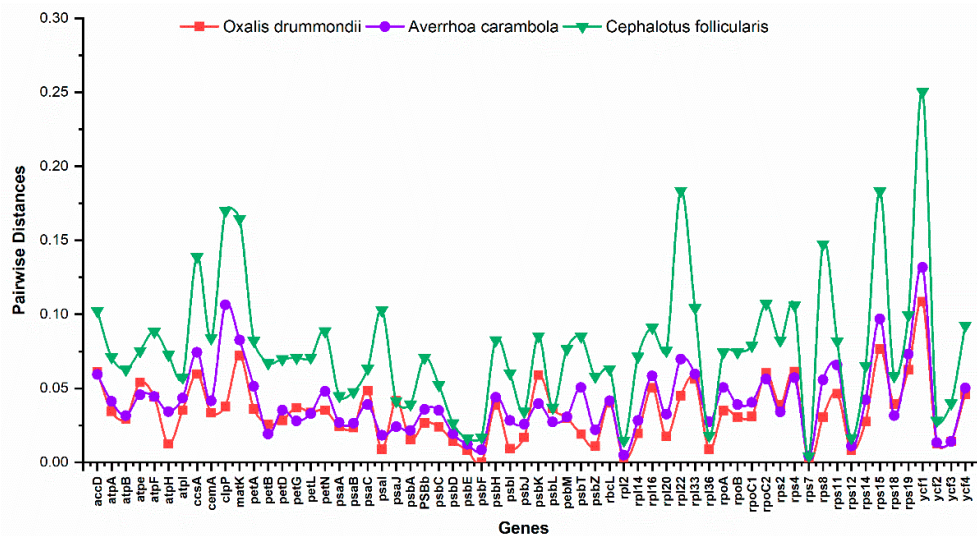
*O. c* = *Oxalis corniculata*, *O. d* = *Oxalis drummondii*, *A. c* = *Averrhoa carambola*, *C. f* = *Cephalotus follicularis*.

## 2.2. Comparative Analysis of *O. corniculata* Chloroplast Genome with Related Species

The *O. corniculata* chloroplast genome was compared with three already sequenced genomes from family Oxalidaceae i.e., *O. drummondii*, *A. carambola*, and *C. follicularis* (Table 1). Variations were observed in cp genomes where *C. follicularis* has the smallest cp genome, 142,706 bp, whilst the *A. carambola* cp genome was the largest, 155,965 bp, amongst analyzed species. We also compared the *O. corniculata* cp genome for pairwise sequence divergence (Table S1). Results showed that *O. corniculata* exhibited the lowest pairwise sequence divergence as compared to *O. drummondii* (0.044) and *A. carambola* (0.057; Table S1). Similarly, the whole cp genomes of *O. corniculata* were compared to identify sequence divergence via mVISTA (Figure 2). Results showed that the coding regions of all cp genomes are conserved compare to non-coding regions, whilst non-coding regions showed a higher divergence rate than the coding regions. The most divergence was observed in intergenic spaces. The *matK* gene exhibited a high degree of divergence in all genomes, but it was higher in *A. carambola* and *C. follicularis* as compared to others. Similarly, *rpoC1*, *rpoC2*, and *rpoB* exhibited more divergence in four cp genomes with the highest in *C. follicularis*. The region between *rpoB* and *psbD* genes showed a high degree of divergence in all species in the LSC. The *ndhK* and *rpl16* genes showed significant divergence compared to the *ycf2* gene with lesser values. The *ndhB* is less divergent in all species except *C. follicularis*, where it showed high divergence. Similarly, the *ndhF* gene is highly divergent in all cp genomes, while it is absent in the *C. follicularis* cp genome. In the SSC, the region between *ndhG* and *ndhH* is highly divergent and the cp genome of *C. follicularis* lack most of the NadH oxidoreductase genes. Furthermore, about 63 protein-coding gene sequences were compared to obtain the average pairwise sequence distance among these species. The results showed that a majority of the genes maintained low levels of average sequence divergence. A relatively lower sequence identity was observed between the chloroplast genomes of *O. corniculata* with related species, especially in the *ccsA*, *clpP*, *rps8*, *rps15*, *rpl22*, *matK*, and *ycf1* genes (Figure 3).



**Figure 2.** Visual alignment of plastid genomes of *O. corniculata* and three other members (*O. drummondii*, *A. carambola*, and *C. follicularis*) from the family Oxalidaceae. VISTA-based identity plot showing sequence identity among three species, using *O. corniculata* as a reference. The vertical scale indicates percent identity, ranging from 50% to 100%. The horizontal axis indicates the coordinates within the chloroplast genome. Arrows indicate the annotated genes and their transcription direction. The thick black lines show the inverted repeats (IRs).

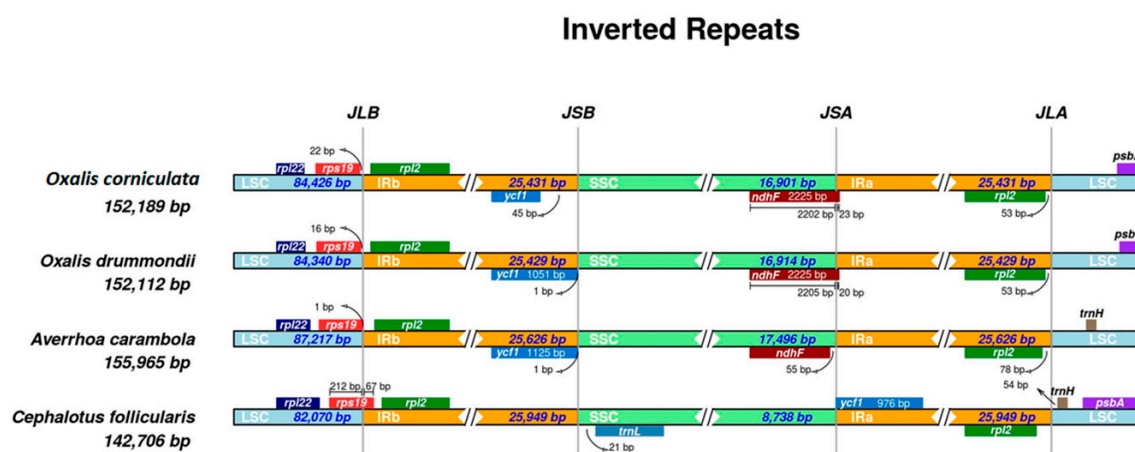


**Figure 3.** Pairwise sequence distance of 63 genes from of *O. corniculata* and three related species (*O. drummondii*, *A. carambola*, and *C. follicularis*) from family Oxalidaceae.



### 2.3. Expansion and Contraction of IR Regions

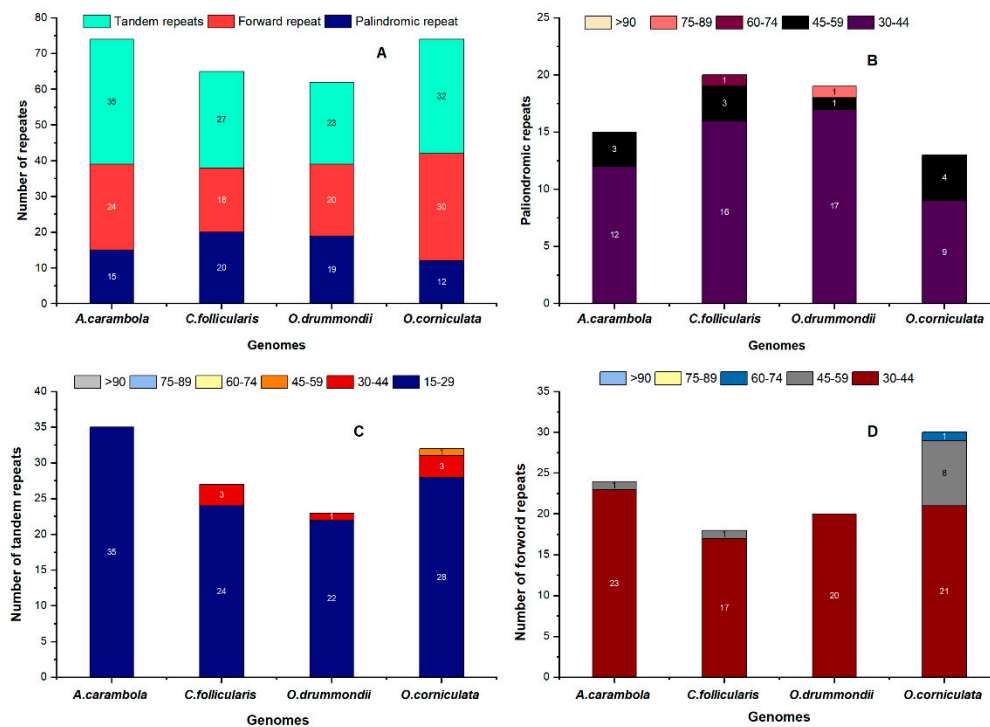
A detailed comparison was performed of the four junctions ( $J^{LA}$ ,  $J^{LB}$ ,  $J^{SA}$ , and  $J^{SB}$ ) amongst IRs (IRa and IRb) and the LSC/SSC regions for the *O. corniculata*, *O. drummondii*, *A. carambola*, and *C. follicularis* (Figure 4). We carefully analyzed and compared the exact IR border positions and the adjacent genes among these cp genomes. The results revealed that at the LSC/IRb ( $J^{LB}$ ) border *rps19* gene is present 22 bp away from the  $J^{LB}$  junction and located in the LSC region in *O. corniculata*. However, in *O. drummondii*, the *rps19* gene is located 16 bp away, while in *A. carambola*, it is 1 bp away from the  $J^{LB}$  junction. On the other hand, in *C. follicularis*, *rps19* is located in the  $J^{LB}$  junction and extended 67 bp in the IRb region. The *ycf1* gene is located 45 bp away from IRb/SSC ( $J^{SB}$ ) in *O. corniculata*, while in *O. drummondii* and *A. carambola*, the *ycf1* gene is located 1 bp away in the IRb region. However, in *C. follicularis*, the *ycf1* gene is absent in the IRb region and located 976 bp apart from the  $J^{SA}$  junction and located in the IRa region. Similarly, the *ndhF* gene is located in the ( $J^{SA}$ ) border and extended 23 bp and 20 bp in the IRa region in both *O. corniculata* and *O. drummondii* genomes, respectively, revealing the expansion of the IR region. In *A. carambola*, the *ndhF* gene is located 55 bp away from the  $J^{SA}$  junction in the SSC region. However, the *ndhF* gene is completely absent in the *C. follicularis* cp genome. Similarly, the *rpl2* gene is located 53 bp away from the IRa/LSC ( $J^{LA}$ ) junction in both *O. corniculata* and *O. drummondii*, while in *A. carambola* and *C. follicularis*, it is 78 bp and 124 bp away from the  $J^{LA}$  junction, respectively (Figure 4).



**Figure 4.** Distance between adjacent genes and junctions of the small single-copy (SSC), large single-copy (LSC), and two inverted repeat (IR) regions of *O. corniculata* with related species cp genomes. Boxes above and below the main line indicate the adjacent border genes. The figure is not to scale regarding sequence length, and it only shows relative changes at or near the IR/SC borders.

### 2.4. Repeat Sequence Analysis

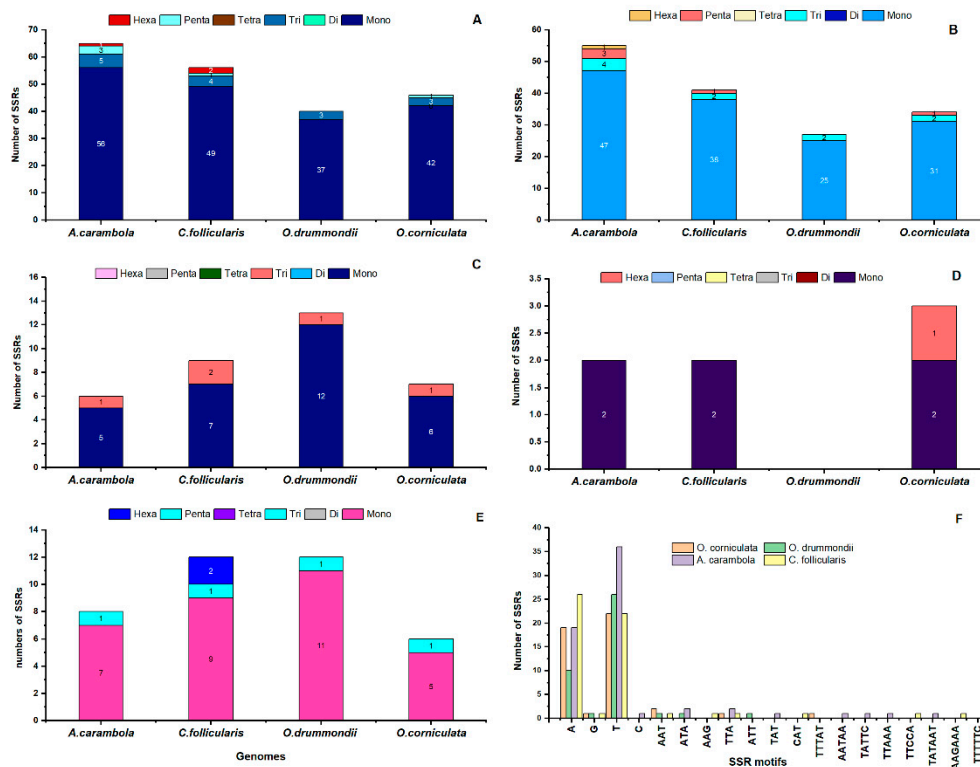
We investigated the repeat sequences of the *O. corniculata* chloroplast genome with the related species. The results revealed that *O. corniculata* contains 12 palindromic, 30 forward, and 32 tandem repeats, *O. drummondii* contains 19 palindromic, 20 forward, and 23 tandem repeats, *A. carambola* contains 15 palindromic, 18 forward, and 27 tandem repeats, and *C. follicularis* contain 20 palindromic, 30 forward, and 32 tandem repeats (Figure 5A). In *O. corniculata*, out of these repeats, the sizes of 9 palindromic repeats were 30–44 bp, while the sizes of 4 repeats were 45–59 bp. Likewise, the size of 28 and 3 tandem repeats were 15–29 bp and 30–44 bp, respectively, whereas the size of 21 forward repeats was found to be about 30–44 bp (Figure 5B–D). Amongst all these cp genomes 74 repeats (the highest) were detected in both *A. carambola* and *O. corniculata*. In all types of repeats, tandem repeats are the highest in number in all the cp genomes, followed by forward and palindromic repeats.



**Figure 5.** Analysis and graphical representation of repeated sequences in the four Oxalidaceae cp genomes. (A) Totals numbers of three repeat types; (B) Number of palindromic repeats by length; (C) Number of tandem repeats by length; (D) Number of forward repeats by length.

### 2.5. Simple Sequence Repeat (SSR) Analysis

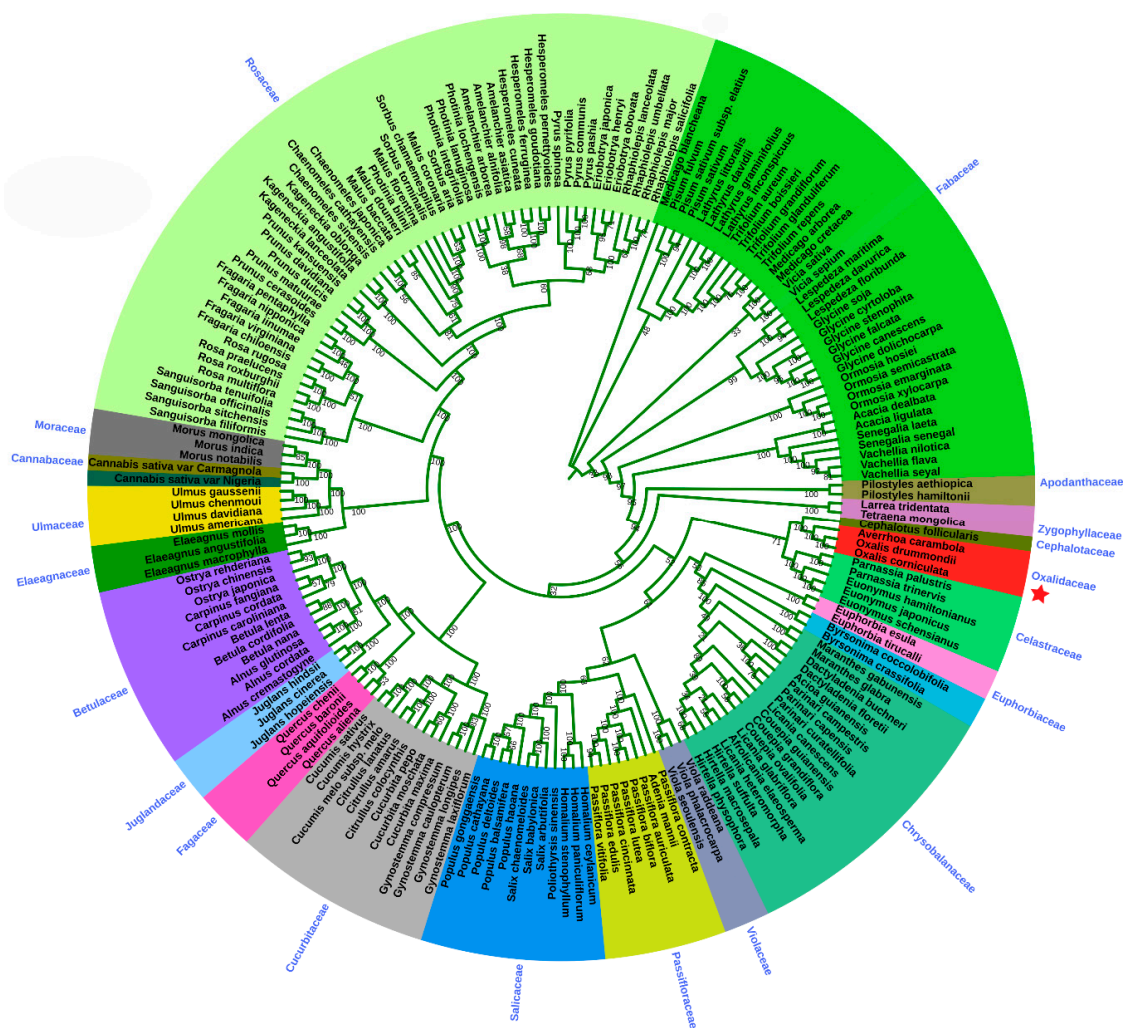
In simple sequence repeats (SSRs) analyses, a total of 46 SSRs were detected in the *O. corniculata* genome; among them, 42 are mononucleotide repeats, 3 are trinucleotide repeats, and 1 is a pentanucleotides repeat (Figure 6). There are no dinucleotides, tetranucleotides, and hexanucleotides in the *O. corniculata* genome. In *O. corniculata*, 13% SSRs are present in the CDS region, 73.9% is present in the LSC region, 15.2% is present in the SSC region, and 2.1% is present in the IR region (Figure 6B–E). Similarly, the highest numbers of SSRs in the other three species are located in the intergenic regions; i.e., *O. drummondii* (70%), *A. carambola* (87.6%), and *C. follicularis* (78.5%) followed by the LSC region—that is, 67.5%, 84.6%, and 73.2%, respectively (Figure 6B–E). On the other hand, in the cp genome of *O. drummondii*, a total of 40 SSRs were found, of which 37 are mono and 3 are trinucleotide, while di, tetra, Penta and hexanucleotides were not detected. In *A. carambola* and *C. follicularis*, 56 and 49 are mononucleotide repeats (Figure 6F). *A. carambola* have 5 trinucleotides, 3 penta, and 1 hexanucleotides repeat, while di and tetranucleotides repeats were missing. Similarly, in *C. follicularis*, 4 tri, 1 penta, and 2 hexanucleotides repeats were found, while di and tetranucleotides repeats were absent in this cp genome. Among the four cp genomes, *A. carambola* has a high number of SSRs; i.e., 56.



**Figure 6.** Analysis and graphical representation of simple sequence repeats (SSRs). Analysis of simple sequence repeats in the four Oxalidaceae chloroplast genomes (A), SSR numbers detected in the four species' LSC regions (B), SSR numbers detected in the four species' SSC regions (C), SSR numbers detected in the four species' IR regions (D), Frequency of identified SSRs in the coding sequences (CDS) region (E) and frequency of identified SSR motifs in different repeat class types (F).

## 2.6. Phylogenetic Analysis

For the phylogenetic analysis of *O. corniculata*, we have downloaded about 191 genomes from the 20 families mentioned in the Materials and Methods section. We inferred the phylogenetic position of *O. corniculata* on the basis of 60 shared genes among these genomes. The study revealed that *O. corniculata* forms a single clade with *O. drummondii* and *A. carambola* in the family Oxalidaceae (Figure 7). These results also showed that *O. corniculata* is closer to *O. drummondii* than *A. carambola*, which is a different genus. Furthermore, the phylogenetic tree also inferred that the Oxalidaceae family is close to Cephalotaceae and Celastraceae with high bootstrap support (100%), followed by Zygophyllaceae and Euphorbiaceae. The phylogenetic trees in this study also exhibited that Rosaceae is highly interlinked to Moraceae. Similarly, the phylogenetic trees also indicate the close relationship of Fabaceae with Apodanthaceae.



**Figure 7.** Phylogenetic trees of *O. corniculata* with related species. The 60 shared genes dataset was analyzed by using maximum likelihood (ML). Numbers above the branches represent bootstrap values in the ML. The red star represents the position of *O. corniculata* (MN998500).

### 3. Discussion

In case of genetic and evolutionary relationship assessments among plant species, chloroplast DNA sequences have been extensively used [40–42]. The complete chloroplast genome sequences provided sufficient information to reconstruct both current and prehistoric diversifications [43]. The powerful and flexible nature of Next Generation Sequencing (NGS) has permeated many areas of study, enabling the development of a broad range of applications that have transformed study designs capable of unlocking information of the genome, transcriptome, and epigenome of any organism [44]. In the current study, we have sequenced the complete genome of *O. corniculata* chloroplast for the first time. The results revealed that the chloroplast genome size of *O. corniculata* is in line with the chloroplast of those flowering plants, which ranges from 125,373 bp to 176,045 bp in *Cuscuta exaltata* and *Vaccinium macrocarpon*, respectively [45,46]. The CG content of *O. corniculata* is 36.7% (Table 1), which is slightly lower than *C. follicularis* and *Paeonia obovata* (38.43%) [47]. The GC content in the IR region is higher (42.6%) than that of the LSC and SSC regions. As a result of the presence of the rich GC nucleotide, higher GC content was present in the IR region of rRNA genes such as *rrn5*, *rrn4.5*, *rrn23*, and *rrn16*, which is consistent with what has been investigated in other cp genomes [48–50].

In most angiosperms, it is believed that the gene(s) of the chloroplast genome and their organization are extremely conserved [51]. In correlation, we detected 131 genes in the cp genome of *O. corniculata*

while other studies also show that many angiosperms have retained these genes [52,53]. With the increasing number of chloroplast genome sequences, the diverse organization of the chloroplast genome is becoming more evident, as demonstrated by genome rearrangement and gene losses in the chloroplast genomes of Oxalidaceae. For example, the *rps16* gene, which is found in most angiosperm plastid genomes, has been lost in *O. corniculata*. Similar results have been reported in various cp genomes previously [54]. Furthermore, in *O. corniculata*, the cp genome the *infA* gene was lost, as reported previously by various researchers, the *infA* gene has been independently lost multiple times from angiosperms and especially in most Rosids [32,51]. Moreover, we found 11 protein-coding genes and 6 tRNAs genes containing introns in the *O. corniculata* cp genome. Among them, three genes, *clpP*, *rps12*, and *ycf3* have two introns, while the others have one intron. Similar results were also reported previously in the *Manihot esculenta* chloroplast genome [32] and *Oresitrophe* chloroplast genome [55]. In this study, genes *ccsA*, *clpP*, *rps8*, *rps15*, *rpl22*, *matK*, and *ycf1* were found to have high evolution rates among the four cp genomes (Figure 3), which agreed with earlier reports of Cuenoud et al. [41]. Similar results of these genes were reported previously among 17 vascular plants and *Panax* species [56].

In the terrestrial plants, the cp genome is very conserved structurally, and the large inverted repeats (IRs) junction is not essential to the function of the cp genome [57]. It is believed that IRs are the most conserved region due to which the rate of natural nucleotide substitution in IRs is lesser as compared to single copy regions, and the variation in IR/LSC and IR/SSC boundaries is the key reason for the size variation among the cp genomes of different groups. The variation in size among four genomes was exhibited by the slight expansion of the IRb (JLB border) in *C. follicularis* compared to *O. corniculata* (Figure 4). These results are in agreement with previous work where IRs are one of the efficient tools for conformational reorganizations within the plastids genomes and are regularly subjected to expansion, contraction, or even complete loss [20]. Similarly, previous results showed that contractions and expansions of the IR regions triggered the diversification of size among the cp genomes [58].

The study of different repeats (palindromic, forward, and tandem) in our sequenced cp genome showed variation in the number of repeats, which is similar to other species previously studied [59]. In all types of repeats, tandem repeats were found more than palindromic and forward repeats in four cp genomes; these results are consistent with previous reports of *Teucrium* and *Commiphora* species [60,61], as well as *S. miltiorrhiza* [62]. Similarly, simple sequence repeats (SSRs) usually have a higher rate of mutation compared with other neutral regions of DNA due to slipped strand mispairing. In genetic studies, due to the haploid and nonparental inheritance nature of cp SSRs, they are commonly used for the assessment of population structure as molecular markers [63,64]. In this study, we comparatively studied the ideal SSRs among the four species *O. corniculata*, *O. drummondii*, *A. carambola*, and *C. follicularis* (Figure 6). The largest number of SSRs was found in *A. carambola*, followed by *C. follicularis*. Mononucleotide repeats were found to be the most common type of SSR in all four species; the A or T mononucleotide repeats are most abundant SSRs in *O. corniculata* (Figure 6), which is congruent to the previous result that SSRs in the chloroplast genome are commonly composed of A or T repeats and rarely G or C repeats [62,65].

Recently, cp genomes information has provided a large amount of data for improving phylogenetic resolution. Chloroplast genome sequences have been widely used for the reconstruction of phylogenetic relationships among plant lineages [66–69]. The phylogenetic evaluation of plant species might not be easy to resolve evolutionary relationships, specifically at taxonomic levels while using a small number of loci [70,71]. Previous phylogenetic studies based on the complete cp genomes and shared genes have been used to explain problematic phylogenetic relations among nearly associated species [34,68] and to increase our concept related to evolutionary relations of angiosperms [72,73]. Phylogenetic relationships of *O. corniculata* were inferred by using 60 shared genes datasets by using the ML method. The results showed that *O. corniculata* form a single clade with *O. drummondii*. Similarly, the phylogenetic tree also inferred that Oxalidaceae family is close to Cephalotaceae and Celastraceae with high bootstrap support (100%), followed by Zygophyllaceae and Euphorbiaceae (Figure 7).

Further cp genomes from the family Oxalidaceae should be explored to determine the phylogenetic position of *O. corniculata* within the section Corniculatae.

## 4. Materials and Methods

### 4.1. Chloroplast DNA Extraction, Sequencing, and Assembly

Young and immature leaves of *O. corniculata* were ground into fine powder in liquid nitrogen, and pure DNA was isolated through a DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA). The resultant chloroplast DNA, by using an Illumina HiSeq-2000 platform (San Diego, CA, USA) at Macrogen (Seoul, Korea) was sequenced. A total of 43,453,336 raw reads were generated for *O. corniculata*, and CLC Genomics Workbench v7.0 (CLC Bio, Aarhus, Denmark) was used to trim and filter reads for the de novo genome assembly. Trimmomatic 0.36 was used for filtering the reads and trailing and leading nucleotide with a Phred score of <20 or when the Phred score dropped below 20 on implementing a 4-bp sliding-window approach. Similarly, reads of <50 bp were discarded after quality filtering and adaptor trimming. The first assembly was formed using SPADIS v3.9.0, with an additional switchover to SOAP denovo v2.04. The resulting contigs were compared against the chloroplast genomes of *O. drummondii* using BLASTN with an E-value cut-off of  $1 \times 10^{-5}$ . The uncertain regions in these genomes, such as IR junction's region, were chosen from the already published genome mentioned above to adjust the sequence length using the iteration method and by employing the Geneious v11.1.2 software [74]. The chloroplast genome sequence of *O. corniculata* has been submitted to GenBank (accession number: MN998500).

### 4.2. Genome Annotation

The Dual Organellar Genome Annotator (DOGMA) [75] was used to annotate the cp genomes of the sequenced species and through BLASTX, the number and position of ribosomal RNAs, transfer RNAs, and other coding genes present in chloroplast genomes were identified and analyzed, while BLASTN tRNAscan-SE version 1.21 was used for tRNA annotation [76] software. Furthermore, Geneious (v11.0) and tRNAscan-SE [76] were used for manual adjustment to compare with the reference genomes reported previously. Similarly, the start and stop codon and intron boundaries were also manually adjusted and compared with the reference chloroplast genome already published. Additionally, by using Organellar Genome DRAW (OGDRAW) [77], the structural characteristics of chloroplast genomes of *O. corniculata* were demonstrated. Beside this, to determine the relative synonymous codon usage and deviations in synonymous codon usage by avoiding the effect of amino acid composition, MEGA6 software [78] was used.

### 4.3. Characterization of Repetitive Sequences and SSR

REPuter software [79] was used to determine the repetitive sequences (palindromic, reverse and direct repeats) within these four cp genomes (*O. corniculata*, *O. drummondii*, *A. carambola*, and *C. follicularis*). Subsequent settings were used for repeat identification through REPuter: (1) a minimum repeat size of 30 bp, (2)  $\geq 90\%$  sequence identity, and (3) a Hamming distance of 1. Tandem Repeats Finder version 4.07 b was used to find tandem repeats by using default settings [80]. The MicroSatellite (MISA) identification tool was used for the microsatellite analysis of *O. corniculata* and another three species' (*O. drummondii*, *A. carambola*, and *C. follicularis*) cp genomes [81]. The parameters such as unit\_size and min\_repeats were defined as follows: 1–10, 2–8, 3–4, 4–4, 5–3, and 6–3; the smallest distance between two SSRs was set to 100 bp. The following conditions were set for parametric significance: 10 or more repeats of one base, 6 or more repeats of two bases, 5 or more repeats of three bases, 5 or more repeats of four bases, 4 or more repeats of five bases, and 4 or more repeats of six bases.

#### 4.4. Sequence Divergence and Phylogenetic Analysis

In the *O. corniculata* chloroplast genome, the average pairwise sequence divergence with three related species (*O. drummondii*, *A. carambola*, and *C. follicularis*) from the family Oxalidaceae was determined. After a comparison of gene order and multiple sequence alignment, comparative sequence analysis was used to recognize missing and unclear gene annotations. For whole genome alignment, MAFFT version 7.222 [82], with default parameters were used, and pairwise sequence divergence was calculated by the use of the selected Kimura's two-parameter (K2P) model [83]. MEGA 6 software [78] was used to evaluate the relative synonymous codon usage by avoiding the effect of amino acid composition. Finally, the divergence of the new *O. corniculata* cp genomes from related species of family Oxalidaceae was determined using mVISTA [84] in Shuffle-LAGAN mode and by employing the genome of new *O. corniculata* as a reference. To resolve the phylogenetic position of *O. corniculata* within the family Oxalidaceae and to check the relationship of 20 families (Fabaceae, Apodanthaceae, Zygophyllaceae, Cephalotaceae, Oxalidaceae, Celastraceae, Euphorbiaceae, Malpighiaceae, Chrysobalanaceae, Violaceae, Passifloraceae, Salicaceae, Cucurbitaceae, Fagaceae, Juglandaceae, Betulaceae, Elaeagnaceae, Ulmaceae, Cannabaceae, Moraceae, Rosaceae) in monophyletic clade rosids, about 60 shared genes from 191 cp genomes were downloaded from the National Center for Biotechnology Information (NCBI) database. For the alignment of 60 shared genes, MAFFT version 7.222 [82] with default parameters was used. The maximum likelihood (ML) method was adopted to infer the phylogenetic trees with MEGA 6 [78], and parameters were adjusted with a BIONJ tree with 1000 bootstrap replicates using the Kimura two-parameter model with gamma-distributed rate heterogeneity and invariant sites.

## 5. Conclusions

The current findings reveal detailed understandings of the complete cp genome of *O. corniculata* for the first time through sequencing on Illumina HiSeq-2000 platform. The gene order and gene structure of *O. corniculata* was found to be similar with three related species from the family Oxalidaceae. Through detailed bioinformatic analysis and comparative assessments, we retrieved essential genetic features such as repetitive sequences, SSRs, codon usage, IR contraction and expansion, sequence divergence, and phylogenomic placement. Whole cp genome comparisons revealed an overall high degree of sequence similarity between *O. corniculata* and *O. drummondii* and some divergence in the intergenic spacers of other species. No major structural rearrangement in these four cp genomes was observed. Phylogenomic analyses of the complete plastid genomes revealed that *O. corniculata* is closely related to *O. drummondii*. A current plastome genomic dataset and the detailed analysis of *O. corniculata* and related species and their comparative analysis provide a powerful genetic resource for the future molecular phylogeny, evolution, population genetics, and biological functions of genus *Oxalis*.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2223-7747/9/8/928/s1>, Table S1. Average pairwise sequence distance of *O. corniculata* with related species cp genomes.

**Author Contributions:** L., S.A. performed experiments; A.L.K., S.A. and R.J. wrote the original draft and bioinformatics analysis; A.L.K., I.-J.L. supervision and arranging resources. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the Agenda Program (Project No. PJ015026022020) Rural Development Administration, Republic of Korea.

**Conflicts of Interest:** The authors have declared that no competing interests exist.

**Availability of Data and Materials:** All data generated or analyzed during this study are included in this published article.

## References

1. Denton, M.F. *Monograph of Oxalis, Section Ionoxalis (Oxalidaceae) in North America*; Michigan State University: East Lansing, MI, USA, 1973.
2. Lourteig, A. *Oxalis, L. subg é nero Monoxalis (Small) Lourteig, Oxalis y Trifidus Lourteig*; Herbarium Bradeanum: Rio de Janeiro, Brazil, 2000.
3. De Azkue, D. Chromosome diversity of South American *Oxalis* (Oxalidaceae). *Bot. J. Linn. Soc.* **2000**, *132*, 143–152. [[CrossRef](#)]
4. Groom, Q.J.; Van der Straeten, J.; Hoste, I. The origin of *Oxalis corniculata* L. *PeerJ* **2019**, *7*, e6384. [[CrossRef](#)]
5. Lourteig, A. Oxalidaceae extra-austroamericanae. II. *Oxalis*, L. Sectio Corniculatae DC. *Phytologia* **1979**, *42*, 57–198.
6. Marks, G. Chromosome numbers in the genus *Oxalis*. *New Phytol.* **1956**, *55*, 120–129. [[CrossRef](#)]
7. Naranjo, C. Estudios citotaxonomicos y evolutivos en especies herbaceas sudamericanas de *Oxalis* (Oxalidaceae) I. *Bol. Soc. Argent. Bot* **1982**, *20*, 183–200.
8. Groom, Q.; Hoste, I.; Janssens, S. Observación confirmada de *Oxalis dillenii* en España. *Collect. Bot.* **2017**, *36*, 4. [[CrossRef](#)]
9. Pyšek, P.; Pergl, J.; Essl, F.; Lenzner, B.; Dawson, W.; Kreft, H.; Weigelt, P.; Winter, M.; Kartesz, J.; Nishino, M.; et al. Naturalized alien flora of the world. *Preslia* **2017**, *89*, 203–274. [[CrossRef](#)]
10. Marble, S.C.; Andrew, K.K.; Gitta, H.; Drew, M.; Annette, C. Efficacy and estimated annual cost of common weed control methods in landscape planting beds. *HortTechnology* **2017**, *27*, 199–211. [[CrossRef](#)]
11. Rehman, A.; Rehman, A.; Ahmad, I. Antibacterial, antifungal, and insecticidal potentials of *Oxalis corniculata* and its isolated compounds. *Int. J. Anal. Chem.* **2015**, *2015*, 1–5. [[CrossRef](#)]
12. Duke, J.A.; Ayensu, E.S. *Medicinal Plants of China*; Reference Publications: Algonac, MI, USA, 1985; Volume 2.
13. Neuhaus, H.; Emes, M. Nonphotosynthetic metabolism in plastids. *Ann. Rev. Plant Biol.* **2000**, *51*, 111–140. [[CrossRef](#)]
14. Rodríguez-Ezpeleta, N.; Brinkmann, H.; Burey, S.C.; Roure, B.; Burger, G.; Löffelhardt, W.; Bohnert, H.J.; Philippe, H.; Lang, B.F. Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. *Curr. Biol.* **2005**, *15*, 1325–1330. [[CrossRef](#)]
15. McCauley, D.E.; Sundby, A.K.; Bailey, M.F.; Welch, M.E. Inheritance of chloroplast DNA is not strictly maternal in *Silene vulgaris* (Caryophyllaceae): Evidence from experimental crosses and natural populations. *Am. J. Bot.* **2007**, *94*, 1333–1337. [[CrossRef](#)]
16. Reboud, X.; Zeyl, C. Organelle inheritance in plants. *Heredity* **1994**, *72*, 132–140. [[CrossRef](#)]
17. Allen, J.F. Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10231–10238. [[CrossRef](#)]
18. Olmstead, R.G.; Palmer, J.D. Chloroplast DNA systematics: A review of methods and data analysis. *Am. J. Bot.* **1994**, *81*, 1205–1224. [[CrossRef](#)]
19. Chumley, T.W.; Palmer, J.D.; Mower, J.P.; Fourcade, H.M.; Calie, P.J.; Boore, J.L.; Jansen, R.K. The complete chloroplast genome sequence of *Pelargonium × hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol. Biol. Evol.* **2006**, *23*, 2175–2190. [[CrossRef](#)]
20. Wicke, S.; Schneeweiss, G.M.; de Pamphilis, C.W.; Müller, K.F.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant. Mol. Biol.* **2011**, *76*, 273–297. [[CrossRef](#)]
21. Lee, H.L.; Jansen, R.K.; Chumley, T.W.; Kim, K.J. Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Mol. Biol. Evol.* **2007**, *24*, 1161–1180. [[CrossRef](#)]
22. Greiner, S.; Wang, X.; Herrmann, R.G.; Rauwolf, U.; Mayer, K.; Haberer, G.; Meurer, J. The complete nucleotide sequences of the 5 genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: II. A microevolutionary view using bioinformatics and formal genetic data. *Mol. Biol. Evol.* **2008**, *25*, 2019–2030. [[CrossRef](#)]
23. Frailey, D.C.; Chaluvadi, S.R.; Vaughn, J.N.; Coatney, C.G.; Bennetzen, J.L. Gene loss and genome rearrangement in the plastids of five Hemiparasites in the family *Orobanchaceae*. *BMC Plant. Biol.* **2018**, *18*, 30. [[CrossRef](#)]



24. Barrett, C.F.; Freudenstein, J.V.; Li, J.; Mayfield-Jones, D.R.; Perez, L.; Pires, J.C.; Santos, C. Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Mol. Biol. Evol.* **2014**, *31*, 3095–3112. [[CrossRef](#)] [[PubMed](#)]
25. Fan, W.-B.; Wu, Y.; Yang, J.; Shahzad, K.; Li, Z.-H. Comparative Chloroplast Genomics of Dipsacales Species: Insights Into Sequence Variation, Adaptive Evolution, and Phylogenetic Relationships. *Front. Plant Sci.* **2018**, *9*. [[CrossRef](#)] [[PubMed](#)]
26. Shaw, J.; Lickey, E.B.; Beck, J.T.; Farmer, S.B.; Liu, W.; Miller, J.; Siripun, K.C.; Winder, C.T.; Schilling, E.E.; Small, R.L. The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* **2005**, *92*, 142–166. [[CrossRef](#)] [[PubMed](#)]
27. Shaw, J.; Shafer, H.L.; Leonard, O.R.; Kovach, M.J.; Schorr, M.; Morris, A.B. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. *Am. J. Bot.* **2014**, *101*, 1987–2004. [[CrossRef](#)]
28. Koehler, S.; Cabral, J.S.; Whitten, W.M.; Williams, N.H.; Singer, R.B.; Neubig, K.M.; Guerra, M.; Souza, A.P.; Amaral, M.D.C.E. Molecular phylogeny of the neotropical genus *Christensonella* (Orchidaceae, Maxillariinae): Species delimitation and insights into chromosome evolution. *Ann. Bot.* **2008**, *102*, 491–507. [[CrossRef](#)]
29. Stuessy, T.F.; Blösch, C.; Villaseñor, J.L.; Rebernick, C.A.; Weiss-Schneeweiss, H. Phylogenetic analyses of DNA sequences with chromosomal and morphological data confirm and refine sectional and series classification within *Melampodium* (Asteraceae, Millerieae). *Taxon* **2011**, *60*, 436–449. [[CrossRef](#)]
30. Ruhfel, B.R.; Gitzendanner, M.A.; Soltis, P.S.; Soltis, D.E.; Burleigh, J.G. From algae to angiosperms—inferring the phylogeny of green plants (*Viridiplantae*) from 360 plastid genomes. *BMC Evol. Biol.* **2014**, *14*, 23. [[CrossRef](#)]
31. Soltis, D.E.; Smith, S.A.; Cellinese, N.; Wurdack, K.J.; Tank, D.C.; Brockington, S.F.; Refulio-Rodriguez, N.F.; Walker, J.B.; Moore, M.J.; Carlsward, B.S.; et al. Angiosperm phylogeny: 17 genes, 640 taxa. *Am. J. Bot.* **2011**, *98*, 704–730. [[CrossRef](#)]
32. 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](#)]
33. Parks, M.; Cronn, R.; Liston, A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* **2009**, *7*, 84. [[CrossRef](#)]
34. Carbonell-Caballero, J.; Alonso, R.; Ibañez, V.; Terol, J.; Talon, M.; Dopazo, J. A Phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus citrus. *Mol. Biol. Evol.* **2015**, *32*, 2015–2035. [[CrossRef](#)] [[PubMed](#)]
35. Moore, M.J.; Bell, C.D.; Soltis, P.S.; Soltis, D.E. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19363–19368. [[CrossRef](#)] [[PubMed](#)]
36. Asaf, S.; Khan, A.L.; Khan, A.; Al-Harrasi, A. Unraveling the chloroplast genomes of two prosopis species to identify its genomic information, comparative analyses and phylogenetic relationship. *Int. J. Mol. Sci.* **2020**, *21*, 3280. [[CrossRef](#)] [[PubMed](#)]
37. Oberlander, K.C.; Dreyer, L.L.; Bellstedt, D.U. Molecular phylogenetics and origins of southern African *Oxalis*. *TAXON* **2011**, *60*, 1667–1677. [[CrossRef](#)]
38. PM, M. Cytology of *Oxalidaceae*. *Cytologia* **1958**, *23*, 200–210.
39. Vaio, M.; Gardner, A.; Emshwiller, E.; Guerra, M. Molecular phylogeny and chromosome evolution among the creeping herbaceous *Oxalis* species of sections *Corniculatae* and *Ripariae* (*Oxalidaceae*). *Mol. Phyl. Evol.* **2013**, *68*, 199–211. [[CrossRef](#)]
40. Ohnishi, O.; Matsuoka, Y. Search for the wild ancestor of buckwheat II. Taxonomy of *Fagopyrum* (Polygonaceae) species based on morphology, isozymes and cpDNA variability. *Genes Genet. Syst.* **1996**, *71*, 383–390. [[CrossRef](#)]
41. Cuénoud, P.; Savolainen, V.; Chatrou, L.W.; Powell, M.; Grayer, R.J.; Chase, M.W. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences. *Am. J. Bot.* **2002**, *89*, 132–144. [[CrossRef](#)]
42. Yamane, K.; Yasui, Y.; Ohnishi, O. Intraspecific cpDNA variations of diploid and tetraploid perennial buckwheat, *Fagopyrum cymosum* (Polygonaceae). *Am. J. Bot.* **2003**, *90*, 339–346. [[CrossRef](#)]

43. Cui, Y.; Qin, S.; Jiang, P. Chloroplast transformation of *Platymonas* (*Tetraselmis*) *subcordiformis* with the bar gene as selectable marker. *PLoS ONE* **2014**, *9*, e98607. [[CrossRef](#)]
44. McPherson, H.; Van der Merwe, M.; Delaney, S.K.; Edwards, M.A.; Henry, R.J.; McIntosh, E.; Rymer, P.D.; Milner, M.L.; Siow, J.; Rossetto, M. Capturing chloroplast variation for molecular ecology studies: A simple next generation sequencing approach applied to a rainforest tree. *BMC Ecol.* **2013**, *13*, 8. [[CrossRef](#)] [[PubMed](#)]
45. McNeal, J.R.; Kuehl, J.V.; Boore, J.L.; de Pamphilis, C.W. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biol.* **2007**, *7*, 57. [[CrossRef](#)] [[PubMed](#)]
46. Fajardo, D.; Senalik, D.; Ames, M.; Zhu, H.; Steffan, S.A.; Harbut, R.; Polashock, J.; Vorsa, N.; Gillespie, E.; Kron, K. Complete plastid genome sequence of *Vaccinium macrocarpon*: Structure, gene content, and rearrangements revealed by next generation sequencing. *Tree Genet. Genomes* **2013**, *9*, 489–498. [[CrossRef](#)]
47. Dong, W.; Xu, C.; Cheng, T.; Zhou, S. Complete chloroplast genome of *Sedum sarmentosum* and chloroplast genome evolution in *Saxifragales*. *PLoS ONE* **2013**, *8*, e77965. [[CrossRef](#)] [[PubMed](#)]
48. Curci, P.L.; De Paola, D.; Danzi, D.; Vendramin, G.G.; Sonnante, G. Complete chloroplast genome of the multifunctional crop globe artichoke and comparison with other *Asteraceae*. *PLoS ONE* **2015**, *10*, e0120589. [[CrossRef](#)]
49. Yang, J.-B.; Yang, S.-X.; Li, H.-T.; Yang, J.; Li, D.-Z. Comparative chloroplast genomes of *Camellia* species. *PLoS ONE* **2013**, *8*, e73053. [[CrossRef](#)]
50. Asaf, S.; Khan, A.L.; Khan, A.; Khan, G.; Lee, I.-J.; Al-Harrasi, A. Expanded inverted repeat region with large scale inversion in the first complete plastid genome sequence of *Plantago ovata*. *Sci. Rep.* **2020**, *10*, 1–16. [[CrossRef](#)]
51. Jansen, R.K.; Saski, C.; Lee, S.-B.; Hansen, A.K.; Daniell, H. Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): Evidence for at least two independent transfers of *rpl22* to the nucleus. *Mol. Biol. Evol.* **2011**, *28*, 835–847. [[CrossRef](#)]
52. Nie, X.; Lv, S.; Zhang, Y.; Du, X.; Wang, L.; Biradar, S.S.; Tan, X.; Wan, F.; Weining, S. Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PLoS ONE* **2012**, *7*, e36869. [[CrossRef](#)]
53. Kim, K.-J.; Lee, H.-L. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* **2004**, *11*, 247–261. [[CrossRef](#)]
54. Jansen, R.K.; Wojciechowski, M.F.; Sanniyasi, E.; Lee, S.-B.; Daniell, H. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (*Leguminosae*). *Mol. Phylogenetics Evol.* **2008**, *48*, 1204–1217. [[CrossRef](#)] [[PubMed](#)]
55. Liu, L.; Wang, Y.; He, P.; Li, P.; Lee, J.; Soltis, D.E.; Fu, C. Chloroplast genome analyses and genomic resource development for epilithic sister genera *Orestitrophe* and *Mukdenia* (*Saxifragaceae*), using genome skimming data. *BMC Genom.* **2018**, *19*, 235. [[CrossRef](#)] [[PubMed](#)]
56. Wu, C.-S.; Chaw, S.-M. Evolutionary stasis in cycad plastomes and the first case of plastome GC-biased gene conversion. *Genome Biol. Evol.* **2015**, *7*, 2000–2009. [[CrossRef](#)] [[PubMed](#)]
57. Goulding, S.E.; Wolfe, K.; Olmstead, R.; Morden, C. Ebb and flow of the chloroplast inverted repeat. *Mol. Gen. Genet.* **1996**, *252*, 195–206. [[CrossRef](#)] [[PubMed](#)]
58. Kim, Y.-K.; Park, C.-W.; Kim, K.-J. Complete chloroplast DNA sequence from a Korean endemic genus, *Megaleranthis saniculifolia*, and its evolutionary implications. *Mol. Cells* **2009**, *27*, 365. [[CrossRef](#)]
59. Asaf, S.; Khan, A.L.; Khan, M.A.; Waqas, M.; Kang, S.-M.; Yun, B.-W.; Lee, I.-J. Chloroplast genomes of *Arabidopsis halleri* ssp. *gemmifera* and *Arabidopsis lyrata* ssp. *petraea*: Structures and comparative analysis. *Sci. Rep.* **2017**, *7*, 1–15. [[CrossRef](#)]
60. Khan, A.; Asaf, S.; Khan, A.L.; Al-Harrasi, A.; Al-Sudairy, O.; AbdulKareem, N.M.; Khan, A.; Shehzad, T.; Alsaady, N.; Al-Lawati, A.; et al. First complete chloroplast genomics and comparative phylogenetic analysis of *Commiphora gileadensis* and *C. foliacea*: Myrrh producing trees. *PLoS ONE* **2019**, *14*, e0208511. [[CrossRef](#)]
61. Khan, A.; Asaf, S.; Khan, A.L.; Khan, A.; Al-Harrasi, A.; Al-Sudairy, O.; AbdulKareem, N.M.; Al-Saady, N.; Al-Rawahi, A. Complete chloroplast genomes of medicinally important *Teucrium* species and comparative analyses with related species from *Lamiaceae*. *PeerJ* **2019**, *7*, e7260. [[CrossRef](#)]

62. Qian, J.; Song, J.; Gao, H.; Zhu, Y.; Xu, J.; Pang, X.; Yao, H.; Sun, C.; Li, X.E.; Li, C.; et al. The Complete Chloroplast Genome Sequence of the Medicinal Plant *Salvia miltiorrhiza*. *PLoS ONE* **2013**, *8*, e57607. [[CrossRef](#)]
63. Leclercq, S.; Rivals, E.; Jarne, P. Detecting microsatellites within genomes: Significant variation among algorithms. *BMC Bioinf.* **2007**, *8*, 125. [[CrossRef](#)]
64. Echt, C.S.; Deverno, L.L.; Anzidei, M.; Vendramin, G.G. Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. *Mol. Ecol.* **1998**, *7*, 307–316. [[CrossRef](#)]
65. Kuang, D.-Y.; Wu, H.; Wang, Y.-L.; Gao, L.-M.; Zhang, S.-Z.; Lu, L. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. *Genome* **2011**, *54*, 663–673. [[CrossRef](#)] [[PubMed](#)]
66. Burke, S.V.; Lin, C.-S.; Wysocki, W.P.; Clark, L.G.; Duvall, M.R. Phylogenomics and Plastome Evolution of Tropical Forest Grasses (Leptaspis, Streptochaeta: Poaceae). *Front. Plant Sci.* **2016**, *7*, 1993. [[CrossRef](#)] [[PubMed](#)]
67. Dong, W.; Xu, C.; Li, W.; Xie, X.; Lu, Y.; Liu, Y.; Jin, X.; Suo, Z. Phylogenetic resolution in Juglans based on complete chloroplast genomes and nuclear DNA sequences. *Front. Plant Sci.* **2017**, *8*. [[CrossRef](#)]
68. 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*, 1–10. [[CrossRef](#)]
69. Sun, L.; Fang, L.; Zhang, Z.; Chang, X.; Penny, D.; Zhong, B. Chloroplast phylogenomic inference of green algae relationships. *Sci. Rep.* **2016**, *6*, 20528. [[CrossRef](#)]
70. Majure, L.C.; Puente, R.; Griffith, M.P.; Judd, W.S.; Soltis, P.S.; Soltis, D.E. Phylogeny of *Opuntia* s.s. (Cactaceae): Clade delineation, geographic origins, and reticulate evolution. *Am. J. Bot.* **2012**, *99*, 847–864. [[CrossRef](#)]
71. Hilu, K.W.; Alice, L.A. A phylogeny of Chloridoideae (Poaceae) based on matK sequences. *Syst. Bot.* **2001**, *26*, 386–405.
72. Goremykin, V.V.; Nikiforova, S.V.; Biggs, P.J.; Zhong, B.; Delange, P.; Martin, W.; Woetzel, S.; Atherton, R.A.; McLenachan, P.A.; Lockhart, P.J. The evolutionary root of flowering plants. *Syst. Biol.* **2013**, *62*, 50–61. [[CrossRef](#)]
73. Luo, Y.; Ma, P.-F.; Li, H.-T.; Yang, J.-B.; Wang, H.; Li, D.-Z. Plastid Phylogenomic Analyses Resolve Tofieldiaceae as the Root of the Early Diverging Monocot Order Alismatales. *Genome Biol. Evol.* **2016**, *8*, 932–945. [[CrossRef](#)]
74. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [[CrossRef](#)] [[PubMed](#)]
75. Wyman, S.K.; Jansen, R.K.; Boore, J.L. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **2004**, *20*, 3252–3255. [[CrossRef](#)] [[PubMed](#)]
76. Schattner, P.; Brooks, A.N.; Lowe, T.M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **2005**, *33*, W686–W689. [[CrossRef](#)] [[PubMed](#)]
77. Lohse, M.; Drechsel, O.; Bock, R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Gen.* **2007**, *52*, 267–274. [[CrossRef](#)] [[PubMed](#)]
78. Kumar, S.; Nei, M.; Dudley, J.; Tamura, K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinf.* **2008**, *9*, 299–306. [[CrossRef](#)]
79. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucl. Acids Res.* **2001**, *29*, 4633–4642. [[CrossRef](#)]
80. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucl. Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)]
81. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [[CrossRef](#)]
82. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]

83. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **1980**, *16*, 111–120. [[CrossRef](#)]
84. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucl. Acids Res.* **2004**, *32*, W273–W279. [[CrossRef](#)] [[PubMed](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).