

## Article

# Complete Chloroplast Genome of *Abutilon fruticosum*: Genome Structure, Comparative and Phylogenetic Analysis

Dhafer A. Alzahrani

Department of Biological Sciences, Faculty of Sciences, King Abdulaziz University, P.O. Box 80203, Jeddah 21589, Saudi Arabia; dalzahrani@kau.edu.sa or dhaferalzahrani@hotmail.com; Tel.: +966-5555-6409

**Abstract:** *Abutilon fruticosum* is one of the endemic plants with high medicinal and economic value in Saudi Arabia and belongs to the family Malvaceae. However, the plastome sequence and phylogenetic position have not been reported until this study. In this research, the complete chloroplast genome of *A. fruticosum* was sequenced and assembled, and comparative and phylogenetic analyses within the Malvaceae family were conducted. The chloroplast genome (cp genome) has a circular and quadripartite structure with a total length of 160,357 bp and contains 114 unique genes (80 protein-coding genes, 30 tRNA genes and 4 rRNA genes). The repeat analyses indicate that all the types of repeats (palindromic, complement, forward and reverse) were present in the genome, with palindromic occurring more frequently. A total number of 212 microsatellites were identified in the plastome, of which the majority are mononucleotides. Comparative analyses with other species of Malvaceae indicate a high level of resemblance in gene content and structural organization and a significant level of variation in the position of genes in single copy and inverted repeat borders. The analyses also reveal variable hotspots in the genomes that can serve as barcodes and tools for inferring phylogenetic relationships in the family: the regions include *trnH-psbA*, *trnK-rps16*, *psbI-trnS*, *atpH-atpI*, *trnT-trnL*, *matK*, *ycf1* and *ndhH*. Phylogenetic analysis indicates that *A. fruticosum* is closely related to *Althaea officinalis*, which disagrees with the previous systematic position of the species. This study provides insights into the systematic position of *A. fruticosum* and valuable resources for further phylogenetic and evolutionary studies of the species and the Malvaceae family to resolve ambiguous issues within the taxa.

**Keywords:** *Abutilon*; Malvaceae; chloroplast genome; phylogenomics



**Citation:** Alzahrani, D.A. Complete Chloroplast Genome of *Abutilon fruticosum*: Genome Structure, Comparative and Phylogenetic Analysis. *Plants* **2021**, *10*, 270. <https://doi.org/10.3390/plants10020270>

Academic Editor: Jung Sung Kim  
Received: 24 November 2020  
Accepted: 24 January 2021  
Published: 30 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the author. 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

The genus *Abutilon* Mill. [1,2], whose members are widely distributed in tropical and subtropical regions [3], is considered as one of the largest genera of Malvaceae [4,5], with ca. 200 accepted species in all continents except Antarctica [3]. The systematic position of some of the taxa in the genus is still not clear [6]; hence, it is the most difficult genera in the Malvaceae with a need for critical systematic studies. The genus is distinguished from sister taxa by the presence of an endoglossum and dorsal dehiscence and a lack of an epicalyx [7]. Members of the genus received a large amount of attention due to their medicinal and economic value [8]. In addition, parts of the plant including the flower, bark, fruit and seeds are reported to contain some phytoconstituents that are responsible for their biological activity [9]. The plants contain no toxins; therefore, many researchers are focusing on them [10–12]. *Abutilon fruticosum* is reported to have medicinal values; all parts of the plant are used in the treatment of various ailments including ulcers, leprosy, inflammation of the bladder, piles, bronchitis, rheumatism and jaundice [10,13,14]. The fiber from the plant is used as a substitute of jute [8]. Despite its importance, the phylogenetic position of the genus is still not clear, and its complete chloroplast genome had not yet been reported until this study. The phylogenetic position of the species within the genus *Abutilon* and the family Malvaceae has not been reported. According to the available literature, there

has been no study that tried to address the phylogenetic position of the species at any taxonomic level where the species belong. As the plant is of medicinal and economic importance, there is a need to report its plastome sequence for conservation purposes, for authentication and to resolve its systematic position.

The family Malvaceae (commonly known as mallows), to which *Abutilon* belongs, is one of the largest families of angiosperm. It comprises ca. 4225 identified species distributed in 244 genera [15,16]. Until this study, only a few complete cp genomes of genera of the family Malvaceae had been known: *Gossypium*, *Abelmoschus*, *Hibiscus*, *Firmiana*, *Bombax*, *Theobroma*, *Craigia* and *Talipariti*.

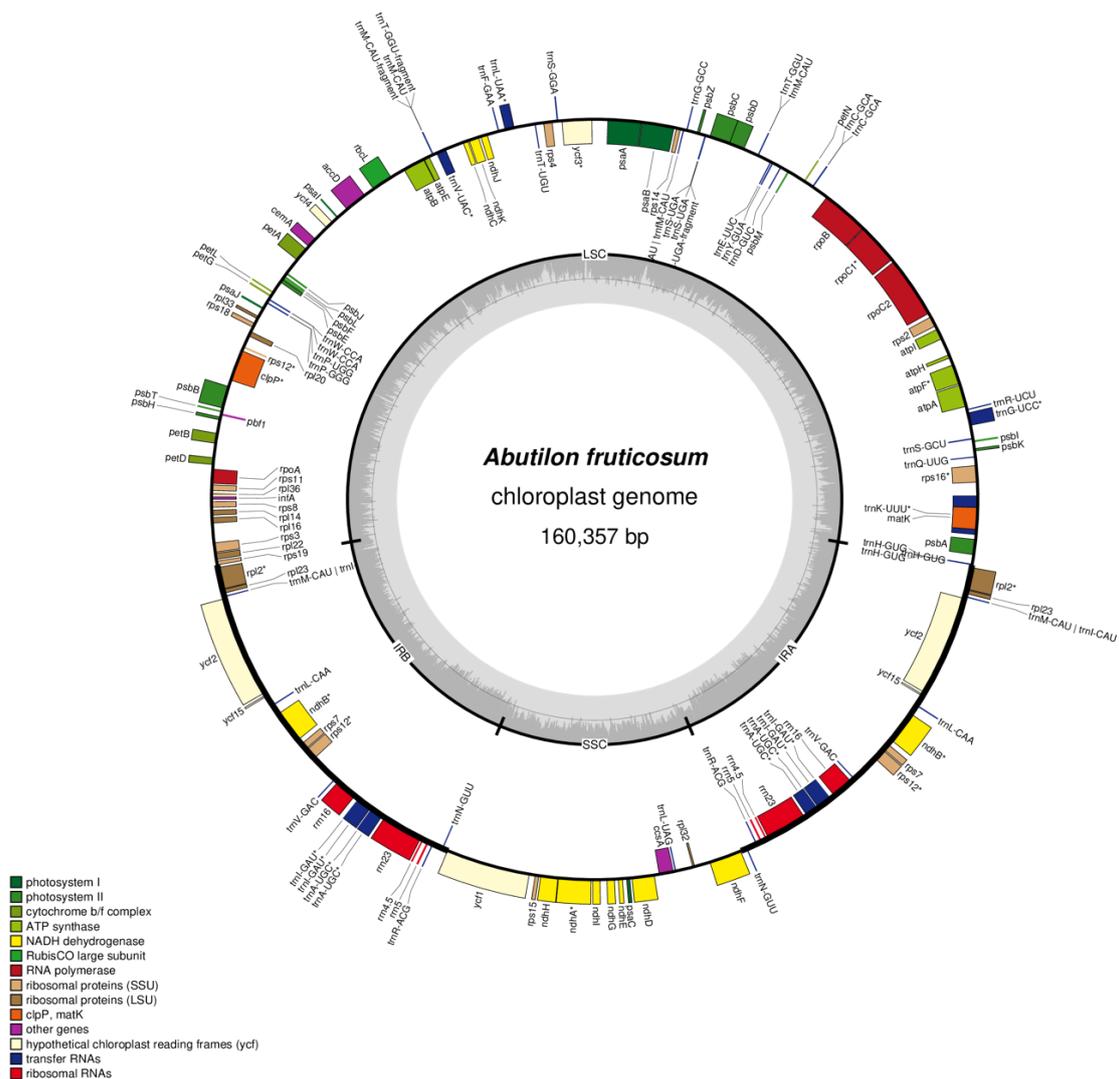
Genetic information is a reliable source of understanding evolutionary relationships among species at various taxonomic levels (categories, ranks). The genetic information in the chloroplast genome contains sufficient information for comparative analysis and studies of species diversification because of the presence of functional genes which have a vital role in plant cells [17]. The chloroplast organelle takes part in carbon fixation and photosynthesis in plants [18]. Among the three genomes present in plants, the chloroplast is the most conserved. In general, chloroplast DNA has a double-stranded, circular and, typically, four-segment structure which includes a large single copy, a small single copy and a pair of inverted repeats [19]. However, recently, Oldenburg and Bendich [20] demonstrated that some plants' ptDNA is of a linear structure. The cp genome is uniparentally inherited and also non-recombinant, and these characteristics make it highly conserved in structure and content [21]. However different kinds of mutations do occur [22], which, as a result, lead to sequence divergence among species and could be used to study evolutionary relationships in plants [23]. Despite the importance of the plastome in modern taxonomy, only chloroplast genomes of few genera in the whole Malvaceae family have been reported.

Here, the whole genome data of *Abutilon fruticosum* were obtained for the first time using Illumina sequencing technology, and the complete chloroplast genome was assembled using NOVOPlasty3.8.1. The features of the cp genomes were analyzed and compared with other Malvaceae species to provide resources for identification and evolutionary, phylogenetic and population genetics studies of the taxon.

## 2. Results and Discussion

### 2.1. Characteristics of *A. fruticosum* Chloroplast Genome

Studies have shown that the complete chloroplast genome of angiosperms is highly conserved in content and structural organization; however, contraction and expansion do occur [24,25]. The complete plastome sequence of *A. fruticosum* has a circular and quadripartite structure with a total length of 160,357 bp. The plastome has four distinct regions which are a small single copy (SSC), a large single copy (LSC) and a pair of inverted repeats (IRa and IRb) which separates the SSC and LSC (Figure 1; GeneBank MT772391). The gene coding region is 81,205 bp in length which constitutes 50.64% of the genome, and the remaining 69,517 bp is the non-coding region which includes introns and intergenic spacers (43.35%). The length of the SSC, LCS, IRa and IRb is 20,031, 89,034, 25,646 and 25,646 bp, respectively. The LSC and SSC regions possess a GC content of 34.76 % and 31.97 %, respectively, while the inverted repeats IRa and IRb have 42.9 % (Table 1). The percentage of GC in the inverted repeat regions is found to be higher than the large and small single regions.



**Figure 1.** Gene map of the *A. fruticosum* chloroplast genome. Genes outside the circles are transcribed in the counter-clockwise direction and those inside in the clockwise direction. Known functional genes are indicated by colored bar. The GC and AT contents are denoted by the dark gray and light gray colors in the inner circle, respectively. LSC indicates large single copy; SSC indicates small single copy; and IR indicates inverted repeat.

**Table 1.** Base composition in the *A. fruticosum* chloroplast genome.

Region	T(U) (%)	C (%)	A (%)	G (%)	Total (bp)
cp genome	31.13	18.26	31.9	18.69	160,357
LSC	33.38	17.94	31.84	16.82	89,032
SSC	33.88	15.03	34.57	16.94	20,031
IRA	28.51	22.26	28.59	20.63	25,646
IRB	28.58	20.61	28.53	22.26	25,646
1st Position	31.24	18.34	31.81	18.59	53,453
2nd Position	31.07	18.25	32.02	18.64	53,452
3rd Position	31.09	18.18	31.86	18.85	53,452

The complete chloroplast genome of *A. fruticosum* contained a total of 133 genes, where 114 genes out of the 133 are unique and are present in the single copy regions; 18 genes are duplicated in the inverted repeat region which includes 7 protein-coding genes, 4 rRNAs and 7 tRNAs. There are 80 protein-coding genes, 4 rRNAs and 30 tRNAs in the plastome (Table 2 and Figure 1). The inverted repeat region contained seven protein-coding genes, seven tRNA and four rRNA, while in the single copy region, the LSC contained 62 protein-

coding genes and 22 tRNA genes; the rest of the 12 protein-coding genes and 1 tRNA are located within the SSC region. Almost all the protein-coding genes start with the ATG codon that codes for methionine, whereas some of the genes contained alternative start codons such as ATC, GTG and ACG; this is common in most chloroplast genomes of flowering plants (angiosperms) [26–28].

**Table 2.** Genes present in the chloroplast genome of *A. fruticosum*.

Category	Group of Genes	Name of Genes
RNA genes	Ribosomal RNA genes (rRNA)	<i>rrn5, rrn4.5, rrn16, rrn23</i>
	Transfer RNA genes (tRNA)	<i>trnH-GUG, trnK-UUU<sup>+</sup>, trnQ-UUG, trnS-GCU, trnG-UCC, trnR-UCU, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnM-CAU, trnT-GGU, trnS-UGA, trnG-GCC, trnI-M-CAU, trnS-GGA, trnT-UGU, trnL-UAA<sup>+</sup>, trnF-GAA, trnV-UAC<sup>+</sup>, trnW-CCA, trnP-UGG, trnP-GGG, trnL-CAA<sup>a</sup>, trnV-GAC<sup>a</sup>, trnI-GAU<sup>+,a</sup>, trnA-UGC<sup>+,a</sup>, trnR-ACG<sup>a</sup>, trnN-GUU<sup>a</sup>, trnL-UAG.</i>
Ribosomal proteins	Small subunit of ribosome	<i>rps2, rps3, rps4, rps7<sup>a</sup>, rps8, rps11, rps12<sup>a</sup>, rps14, rps15, rps16<sup>+</sup>, rps18, rps19</i>
Transcriptiongenes	Large subunit of ribosome	<i>rpl2<sup>+,a</sup>, rpl14, rpl16, rpl20, rpl22, rpl23<sup>a</sup>, rpl32, rpl33, rpl36.</i>
	DNA-dependent RNA polymerase	<i>rpoA, rpoB, rpoC1<sup>+</sup>, rpoC2</i>
Protein genes	Photosystem I	<i>psaA, psaB, psaC, psal, psaj, ycf3<sup>++</sup></i>
	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Subunit of cytochrome	<i>petA, petB, petD, petG, petL, petN</i>
	Subunit of synthase	<i>atpA, atpB, atpE, atpF<sup>+</sup>, atpH, atpI</i>
	Large subunit of rubisco	<i>rbcl</i>
	NADH dehydrogenase	<i>ndhA<sup>+</sup>, ndhB<sup>+,a</sup>, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	ATP-dependent protease subunit P	<i>clpP<sup>++</sup></i>
	Chloroplast envelope membrane protein	<i>cemA</i>
	Maturase	<i>matK</i>
Other genes	Subunit acetyl-coA carboxylase	<i>accD</i>
	C-type cytochrome synthesis	<i>ccsA</i>
	Translational initiation factor	<i>infA</i>
	Hypothetical proteins	<i>ycf2<sup>a</sup>, ycf4, ycf15<sup>a</sup></i>
	Component of TIC complex	<i>ycf1<sup>a</sup></i>

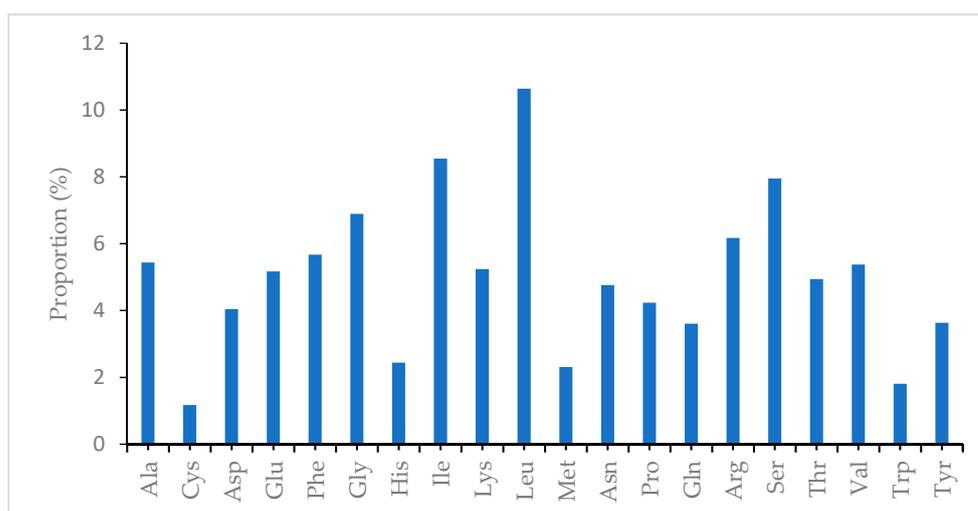
<sup>+</sup> Gene with one intron, <sup>++</sup> gene with two introns and <sup>a</sup> gene with copies.

The *A. fruticosum* chloroplast genome is found to contain an intron in some of the coding genes, such as in other chloroplast genomes of flowering plants [26,27]. Among the 114 coding genes in *A. fruticosum*, 17 contain introns (Table 3). Out of the 17 genes with an intron, 11 are protein-coding genes and six are tRNAs. The LSC region contains introns in 11 genes and the IR region contains introns in 5 genes, while the SSC region contains introns in only 1 gene. Two genes, *ycf3* and *clpP*, possess two introns and the other 15 genes have only one intron. *trnK-UUU* has the longest intron, while *accD* has the shortest intron (Table 3).

**Table 3.** The introns in the genes of the *A. fruticosum* plastome.

Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
<i>rps16</i>	LSC	224	889	35		
<i>atpF</i>	LSC	410	790	158		
<i>rpoC1</i>	LSC	1634	759	434		
<i>ycf3</i>	LSC	152	773	227	763	125
<i>ndhK</i>	LSC	716	7	155		
<i>accD</i>	LSC	44	6	1487		
<i>clpP</i>	LSC	227	640	293	914	70
<i>rp12</i>	IR	434	682	392		
<i>ycf15</i>	IR	131	24	125		
<i>ndhB</i>	IR	755	684	776		
<i>ndhA</i>	SSC	551	1124	539		
<i>trnK-UUU</i>	LSC	34	2574	36		
<i>trnG-UCC</i>	LSC	31	789	60		
<i>trnL-UAA</i>	LSC	36	560	49		
<i>trnV-UAC</i>	LSC	36	590	37		
<i>trnI-GAU</i>	IR	41	958	34		
<i>trnA-UGC</i>	IR	37	800	34		

Codon usage compares the frequencies of each codon that codes for a particular amino acid [29]. Codons are used in transmitting genetic information because they are the building blocks of proteins [30]. Codon usage is a factor shaping the evolution of chloroplast genomes because of bias in mutation [28], and it varies across different species [31]. The frequency of the codon present in the chloroplast genome was computed using the nucleotide sequence of protein-coding genes and tRNA genes 84,048 bp. The relative synonymous codon usage (RSCU) of the genes in the genome is presented in Table 4. The results show that genes in the plastome are encoded by 27,967 codons. Codons that code for leucine appear more frequently in the genome 2957 (10.57%) (Figure 2). Meanwhile, codons coding for cysteine are the least with 325 (1.16%) in the genome. Guanine and cytosine endings are found to be more frequent than their counterparts adenine and thymine; this is not the case in other plastome sequences [32–34]. The result of the analysis (Table 4) shows that codon usage bias is low in the chloroplast genome of *A. fruticosum*. The RSCU values of 30 codons were greater than 1 and all of them have an A/T ending, while for 31 codons, the values were less than 1 and are all of the G/C ending. Only two amino acids, tryptophan and methionine, have an RSCU value of 1 and therefore they are the only amino acids with no codon bias.

**Figure 2.** Amino acid frequencies in *A. fruticosum* chloroplast genome protein-coding sequences.

**Table 4.** Codon–anticodon recognition patterns and codon usage of the *A. fruticosum* chloroplast genome.

Codon	Amino Acid	RSCU	tRNA	Codon	Amino Acid	RSCU	tRNA
UUU	Phe	1.27	<i>trnF-GAA</i>	UAU	Tyr	1.57	<i>trnY-GUA</i>
UUC	Phe	0.73		UAC	Tyr	0.43	
UUA	Leu	1.85	<i>trnL-UAA</i>	UAA	Stop	1.35	
UUG	Leu	1.24	<i>trnL-CAA</i>	UAG	Stop	1.06	
CUU	Leu	1.22	<i>trnL-UAG</i>	CAU	His	1.49	<i>trnH-GUG</i>
CUC	Leu	0.46		CAC	His	0.51	
CUA	Leu	0.83		CAA	Gln	1.5	<i>trnQ-UUG</i>
CUG	Leu	0.41		CAG	Gln	0.5	
AUU	Ile	1.46	<i>trnI-GAU</i>	AAU	Asn	1.52	<i>trnN-GUU</i>
AUC	Ile	0.62		AAC	Asn	0.48	
AUA	Ile	0.93	<i>trnI-CAU</i>	AAA	Lys	1.48	<i>trnK-UUU</i>
AUG	Met	1	<i>trnM-CAU</i>	AAG	Lys	0.52	
GUU	Val	1.45	<i>trnV-GAC</i>	GAU	Asp	1.61	<i>trnD-GUC</i>
GUC	Val	0.53		GAC	Asp	0.39	
GUA	Val	1.46		GAA	Glu	1.45	<i>trnE-UUC</i>
GUG	Val	0.56	<i>trnV-UAC</i>	GAG	Glu	0.55	
UCU	Ser	1.65	<i>trnS-GGA</i>	UGU	Cys	1.5	<i>trnC-GCA</i>
UCC	Ser	0.98		UGC	Cys	0.5	
UCA	Ser	1.22		UGA	Stop	0.59	
UCG	Ser	0.57	<i>trnS-UGA</i>	UGG	Trp	1	<i>trnW-CCA</i>
CCU	Pro	1.47	<i>trnP-UGG</i>	CGU	Arg	1.2	<i>trnR-ACG</i>
CCC	Pro	0.77		CGC	Arg	0.49	<i>trnR-UCU</i>
CCA	Pro	1.15		CGA	Arg	1.35	
CCG	Pro	0.62		CGG	Arg	0.52	
ACU	Thr	1.57		AGA	Arg	1.71	
ACC	Thr	0.78		AGG	Arg	0.73	
ACA	Thr	1.16	<i>trnT-GGU</i>	AGU	Ser	1.14	<i>trnS-GCU</i>
ACG	Thr	0.49	<i>trnT-UGU</i>	AGC	Ser	0.43	
GCU	Ala	1.76	<i>trnA-UGC</i>	GGU	Gly	1.28	
GCC	Ala	0.7		GGC	Gly	0.46	
GCA	Ala	1.03		GGA	Gly	1.51	
GCG	Ala	0.52		GGG	Gly	0.75	<i>trnG-UCC</i>

RNA editing is a set of processes including the insertion, deletion and modification of nucleotides that alters the DNA-encoded sequence [35], which is a way to create transcript and protein diversity [36]. Some chloroplast RNA editing sites are preserved in plants [37]. The program PREP suite was used to predict the RNA editing sites in the chloroplast genome of *A. fruticosum*. The first nucleotide of the codon was used in all the analyses. The result of the analysis shows that most of the conversions in the codons are from serine to leucine (Table 5).

Generally, 50 editing sites in the genome were revealed which were distributed within 22 protein-coding genes. The gene *ndhB* has the highest number of editing sites with 12 sites, and this is consistent with previous studies [38–40]. One gene with eight editing sites is *ndhD* and other genes with a high number of editing sites are *ndhF* and *rpoB* having four and *matK* with three editing sites. The genes *accD*, *atpA*, *ndhA*, *ndhG*, *rpoA*, *rpoC1*, *rpoC2* and *rps2* have two editing sites.

The following genes: *atpF*, *atpI*, *ccsA*, *clpP*, *petB*, *psbF*, *rpl20*, *rps8* and *rps14*, with one editing site, have the lowest number of editing sites. Conversions of proline to serine were observed, which involve the change of amino acids in the RNA editing site from a nonpolar to a polar group. Genes such as *atpB*, *petD*, *petG*, *petL*, *psaB*, *psaI*, *psbB*, *psbE*, *psbL*, *rpl2*, *rpl23*, *rps16* and *ycf3* do not possess predicted RNA sites in their first codon.

**Table 5.** Predicted RNA editing sites in the *A. fruticosum* chloroplast genome.

Gene	Nucleotide Position	Amino Acid Position	Codon	Amino Acid	Score
<i>accD</i>	854	285	TCG => TTG	S => L	0.8
	1463	488	CCT => CTT	P => L	1
<i>atpA</i>	914	305	TCA => TTA	S => L	1
	1148	383	TCA => TTAA	S => L	1
<i>atpF</i>	92	31	CCA => CTA	P => L	0.86
<i>atpI</i>	629	210	TCA => TTA	S => L	1
<i>ccsA</i>	662	221	ACT => ATT	T => I	0.86
<i>clpP</i>	559	187	CAT => TAT	H => Y	1
<i>matK</i>	457	153	CAT => TAT	H => Y	1
	634	212	CAT => TAT	H => Y	1
	1237	413	CAC => TAC	H => Y	1
<i>ndhA</i>	341	114	TCG => TTG	S => L	1
	566	189	TCA => TTA	S => L	1
<i>ndhB</i>	149	50	TCA => TTA	S => L	1
	467	156	CCA => CTA	P => L	1
	542	181	ACG => ATG	T => M	1
	586	196	CAT => TAT	H => Y	1
	611	204	TCA => TTA	S => L	0.8
	737	246	CCA => CTA	P => L	1
	746	249	TCT => TTT	S => F	1
	830	277	TCAG => TTG	S => L	1
	836	279	TCA => TTA	S => L	1
	1255	419	CAT => TAT	H => Y	1
	1291	431	CTC => CTA	L => F	1
	1481	494	CCA => CTA	P => L	1
	<i>ndhD</i>	2	1	ACG => ATG	T => M
26		9	ACA => ATA	T => I	1
47		16	TCT => TTT	S => F	0.8
383		128	TCA => TTA	S => L	1
568		190	CCT => TCT	P => S	1
674		225	TCG => TTG	S => L	1
878		293	TCA => TTA	S => L	1
1298		433	TCA => TTA	S => L	0.8
290		97	TCA => TTA	S => L	1
1549		517	CTT => TTT	L => F	1
<i>ndhF</i>	1826	609	ACA => ATA	T => I	0.8
	1892	631	GCG => GTG	A => V	0.8
	166	56	CAT => TAT	H => Y	0.8
	314	105	ACA => ATA	T => I	0.8
<i>petB</i>	425	142	CGG => TGG	R => W	1
<i>psbF</i>	77	26	TCT => TTT	S => F	1
<i>rpl20</i>	308	103	TCA => TTA	S => L	0.86
<i>rpoA</i>	329	110	GCC => GTC	A => V	0.86
	830	277	TCA => TTA	S => L	1
<i>rpoB</i>	338	113	TCT => TTT	S => F	1
	551	184	TCA => TTA	S => L	1
	566	189	TCG => TTG	S => L	1
<i>rpoC1</i>	2426	809	TCA => TTA	S => L	0.86
	41	14	TCA => TTA	S => L	1
	1273	425	CCG => TCG	P => S	0.86
<i>rpoC2</i>	2296	766	CGG => TGG	R => W	1
	3188	1063	CCC => CTC	P => L	0.86
<i>rps2</i>	248	83	TCG => TTG	S => L	1
	325	109	CCC => TCC	P => S	1
<i>rps8</i>	217	73	CAT => TAT	H => Y	1
<i>rps14</i>	149	50	TCA => TTA	S => L	1

## 2.2. Repeat Analyses

### 2.2.1. Long Repeats

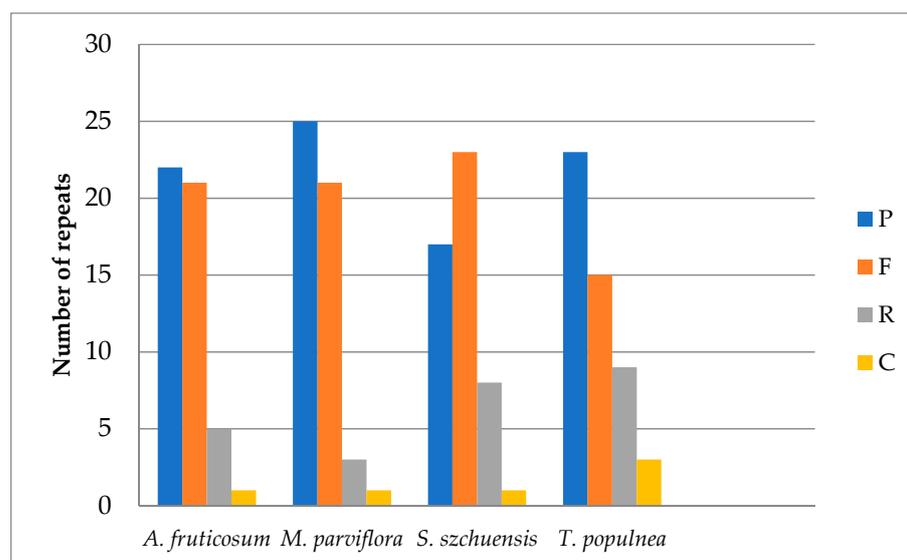
The program REPuter was used to identify long repeat sequences present in the *A. fruticosum* chloroplast genome. It was discovered from the results that all four types of long repeats (palindromic, forward, reverse and complement) were present in the plastome of *A. fruticosum* (Table 6). The analysis showed 22 palindromic repeats, 21 forward repeats, 5 reverse repeats and 1 complement repeat (Table 6). In total, there were 49 long repeats in the chloroplast genome of *A. fruticosum*. The majority of the repeats were between 20 and 29 bp (87.75%) in size, followed by 30–39 bp (8.16%) and 50–59 bp (4.08%) long repeats. In the first location, the intergenic spacer harbored 61.22% of the repeats. The tRNA contained four repeats (8.16%), and eight repeats (16.32%) were located in the protein-coding genes. The length of repeated sequences in the *A. fruticosum* chloroplast genome ranged from 10 to 69 bp, analogously to the other angiosperm plants [41–43]. I compared the frequency of repeats among four Malvaceae cp genomes and found that all the types of repeats (palindromic, forward, reverse and complement) were present in all genomes (Figure 3). *Malva parviflora* has the highest number of palindromic repeats (25), while *Sida szechuensis* has the lowest with 17. *A. fruticosum* and *M. parviflora* have the same number of forward repeats—21 for each of them. *T. populnea* has the highest number of reverse repeats (9), while *M. parviflora* has the lowest (3). Complement repeats were found to be the least numerous types of repeat across the genome in *A. fruticosum*, in *S. szechuensis* and in *M. parviflora*, occurring once. In the plastome of *T. populnea*, there were three complement repeats.

**Table 6.** Repeat sequences present in the *A. fruticosum* chloroplast genome.

SN	Repeat Size	Repeat Position 1	Repeat Type	Repeat Location	Repeat Position 2	Repeat Location 2	E-Value
1	58	33,156	F	IGS	33,213	IGS	$8.71 \times 10^{-26}$
2	50	0	P	IGS	89,084	IGS	$5.71 \times 10^{-21}$
3	36	44,187	F	IGS	44,205	IGS	$1.53 \times 10^{-12}$
4	36	103,838	P	IGS	123,909	<i>ndhA</i> -Intron	$1.53 \times 10^{-12}$
5	36	123,909	F	<i>ndhA</i> -Intron	145,617	IGS	$1.53 \times 10^{-12}$
6	30	64,915	F	IGS	64,930	IGS	$6.27 \times 10^{-9}$
7	29	8535	P	<i>trnS-GCU</i>	47,062	IGS	$2.51 \times 10^{-8}$
8	26	6039	P	IGS	6039	IGS	$1.61 \times 10^{-6}$
9	26	10,662	P	IGS	10,662	IGS	$1.61 \times 10^{-6}$
10	26	96,437	F	<i>ycf2</i>	96,455	<i>ycf2</i>	$1.61 \times 10^{-6}$
11	26	96,437	P	<i>ycf2</i>	153,010	<i>ycf2</i>	$1.61 \times 10^{-6}$
12	26	96,455	P	<i>ycf2</i>	153,028	<i>ycf2</i>	$1.61 \times 10^{-6}$
13	26	153,010	F	<i>ycf2</i>	153,028	<i>ycf2</i>	$1.61 \times 10^{-6}$
14	25	53,987	P	IGS	58,381	IGS	$6.42 \times 10^{-6}$
15	25	54,230	R	IGS	54,230	IGS	$6.42 \times 10^{-6}$
16	24	19,010	F	<i>rpoC2</i>	19,034	<i>rpoC2</i>	$2.57 \times 10^{-6}$
17	24	40,578	F	<i>psaB</i>	42,802	<i>psaA</i>	$2.57 \times 10^{-5}$
18	24	48,652	F	IGS	48,672	IGS	$2.57 \times 10^{-5}$
19	23	48,902	R	IGS	48,902	IGS	$1.03 \times 10^{-4}$
20	23	88,769	F	<i>rps19</i>	88,792	<i>rps19</i>	$1.03 \times 10^{-4}$
21	23	112,986	F	IGS	113,018	IGS	$1.03 \times 10^{-4}$
22	23	112,986	P	IGS	136,450	IGS	$1.03 \times 10^{-4}$
23	23	113,018	P	IGS	136,482	IGS	$1.03 \times 10^{-4}$
24	23	136,450	F	IGS	136,482	IGS	$1.03 \times 10^{-4}$
25	22	10,362	P	IGS	10,388	IGS	$4.11 \times 10^{-4}$
26	22	38,111	P	IGS	38,111	IGS	$4.11 \times 10^{-4}$
27	22	113,764	F	IGS	113,785	IGS	$4.11 \times 10^{-4}$
28	22	113,764	P	IGS	135,684	IGS	$4.11 \times 10^{-4}$
29	22	113,785	P	IGS	135,705	IGS	$4.11 \times 10^{-4}$

Table 6. Cont.

SN	Repeat Size	Repeat Position 1	Repeat Type	Repeat Location	Repeat Position 2	Repeat Location 2	E-Value
30	22	135,684	F	IGS	135,705	IGS	$4.11 \times 10^{-4}$
31	21	8540	F	<i>trnS-GCU</i>	37,036	<i>trnS-UGA</i>	$1.64 \times 10^{-3}$
32	21	9201	F	IGS	9220	IGS	$1.64 \times 10^{-3}$
33	21	10,136	F	IGS	10,157	IGS	$1.64 \times 10^{-3}$
34	21	10,491	R	<i>trnR-UCU</i>	10,491	<i>trnR-UCU</i>	$1.64 \times 10^{-3}$
35	21	37,036	P	<i>trnS-UGA</i>	47,065	IGS	$1.64 \times 10^{-3}$
36	21	43,822	R	IGS	43,822	IGS	$1.64 \times 10^{-3}$
37	21	78,782	P	<i>psbN</i>	78,810	<i>psbN</i>	$1.64 \times 10^{-3}$
38	21	96,428	F	<i>ycf2</i>	96,482	<i>ycf2</i>	$1.64 \times 10^{-3}$
39	21	96,428	P	<i>ycf2</i>	152,988	<i>ycf2</i>	$1.64 \times 10^{-3}$
40	21	96,482	P	<i>ycf2</i>	153,042	<i>ycf2</i>	$1.64 \times 10^{-3}$
41	21	152,988	F	<i>ycf2</i>	153,042	<i>ycf2</i>	$1.64 \times 10^{-3}$
42	20	402	P	IGS	402	IGS	$6.58 \times 10^{-3}$
43	20	5288	F	IGS	5307	IGS	$6.58 \times 10^{-3}$
44	20	10,128	C	IGS	82,502	IGS	$6.58 \times 10^{-3}$
45	20	14,151	P	IGS	54,042	IGS	$6.58 \times 10^{-3}$
46	20	50,997	P	<i>trnF-GAA</i>	55,404	IGS	$6.58 \times 10^{-3}$
47	20	53,309	R	<i>ndhC</i>	53,309	<i>ndhC</i>	$6.58 \times 10^{-3}$
48	20	55,405	P	IGS	108,890	<i>ycf2</i>	$6.58 \times 10^{-3}$
49	20	55,405	F	IGS	140,581	<i>ycf2</i>	$6.58 \times 10^{-3}$



**Figure 3.** Number of different repeats in chloroplast genomes of five species of Malvaceae. P = palindromic, F = forward, R = reverse and C = complement.

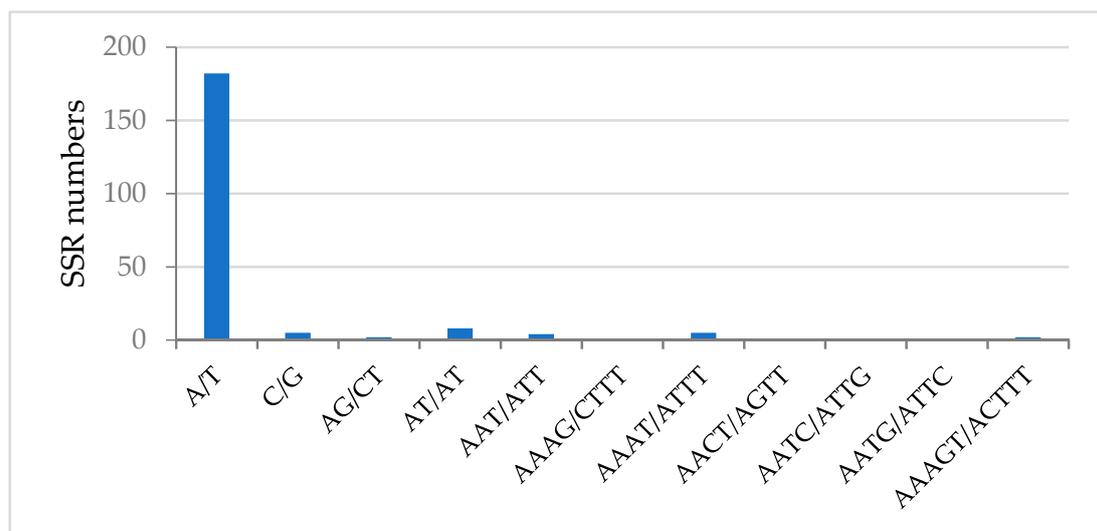
### 2.2.2. Simple Sequence Repeats (SSRs)

There were short repeats of nucleotide series (1–6 bp) that were dispersed through the whole genome called microsatellites (SSRs). These short repeats in the plastid genome were passed from a single parent. As a result, they are used as molecular indicators in developmental studies such as genetic diversity and also contribute to the recognition of species [44–46]. A total of 212 microsatellites were found in the chloroplast genome of *A. fruticosum* in this study (Table 7). The majority of SSRs in the cp genome are mononucleotides (88.88), where poly-A (polyadenine) and poly-T (polythymine) are dominant (Figure 4). Poly-A constituted 45.06%, whereas poly-T constituted 41.97%. This is consistent with previous studies [47]. Among the dinucleotide repeats, only AG/CT and AT/AT were

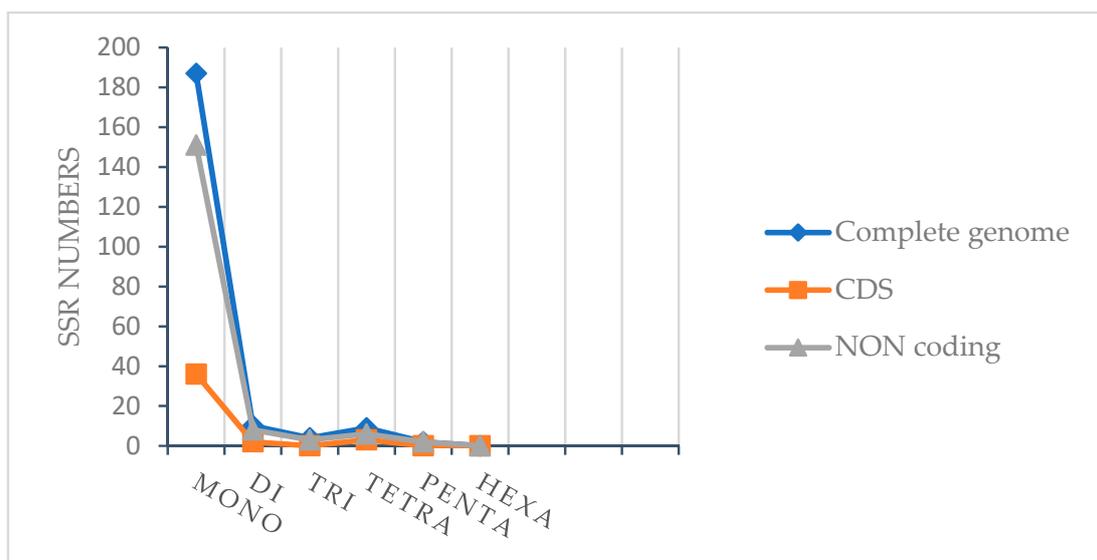
found in the cp genome. Taking into account the complementarity of series, only one trinucleotide (AAT/ATT), five tetranucleotides (AAAG/CTTT, AAAT/ATTT, AACT/AGTT, AATC/ATTG and AATG/ATTC) and only one pentanucleotide (AAAGT/ACTTT) were present in the cp genome (Figure 4). The intergenic/non-coding regions harbored most of the microsatellites (75.92%) (Figure 5).

**Table 7.** Simple sequence repeats in the chloroplast genome of *A. fruticosum*.

Repeat	Length (bp)	Number	Start Position
A	8	40	4686; 5481; 5865; 6926; 7897; 12,271; 13,473; 13,702; 15,318; 19,663; 22,147; 23,354; 30,009; 30,421; 31,419; 49,433; 50,206; 51,576; 54,400; 69,630; 74,881; 76,238; 82,296; 84,962; 85,726; 115,568; 117,338; 118,072; 120,375; 120,736; 128,085; 130,563; 131,446; 131,580; 135,564; 135,895; 140,843; 141,163; 145,401; 148,617
	9	12	4151; 8835; 16,318; 23,595; 67,970; 80,923; 87,139; 94,543; 117,759; 119,376; 125,042; 125,285
	10	11	5166; 7193; 28,188; 37,725; 48,400; 74,710; 75,051; 76,083; 115,690; 128,759; 131,209
	11	5	52,345; 82,469; 116,213; 117,460; 118,643
	12	3	320; 2221; 65,189
	14	1	14,115
	15	1	145,189
C	8	1	27,432
	10	1	14,953
	11	1	30,215
T	8	28	116; 4444; 9679; 28,812; 32,051; 36,848; 37,359; 58,437; 61,319; 63,814; 67,383; 67,842; 68,540; 71,227; 79,674; 81,661; 88,975; 100,748; 103,964; 108,202; 108,522; 113,470; 113,801; 115,834; 126,161; 127,095; 130,350; 131,860
	9	13	2643; 6292; 13,241; 20,064; 31,032; 43,959; 51,110; 54,805; 71,700; 82,737; 87,273; 123,499; 154,821
	10	15	5982; 9102; 10,369; 10,642; 12,750; 14,567; 17,590; 30,779; 34,153; 49,053; 58,034; 64,821; 66,148; 81,282; 84,685
	11	6	19,520; 21,952; 29,839; 70,361; 75,959; 85,189
	12	4	27,254; 54,171; 57,834; 63,361
	15	1	104,169
AT	5	3	20,902; 53,741; 63,067
	6	1	28,039
TA	5	2	10,115; 33,119
TC	5	1	65,402
CT	5	1	17,231
AAT	4	1	13,837
TTA	4	1	160,252
AATA	3	1	13,014
AGAA	3	1	115,404
ATCA	3	1	126,521
GAAT	3	1	118,455
TAGT	3	1	62,483
TTTA	3	1	72,098
ACTTT	3	1	139,519
TAAAG	3	1	109,838

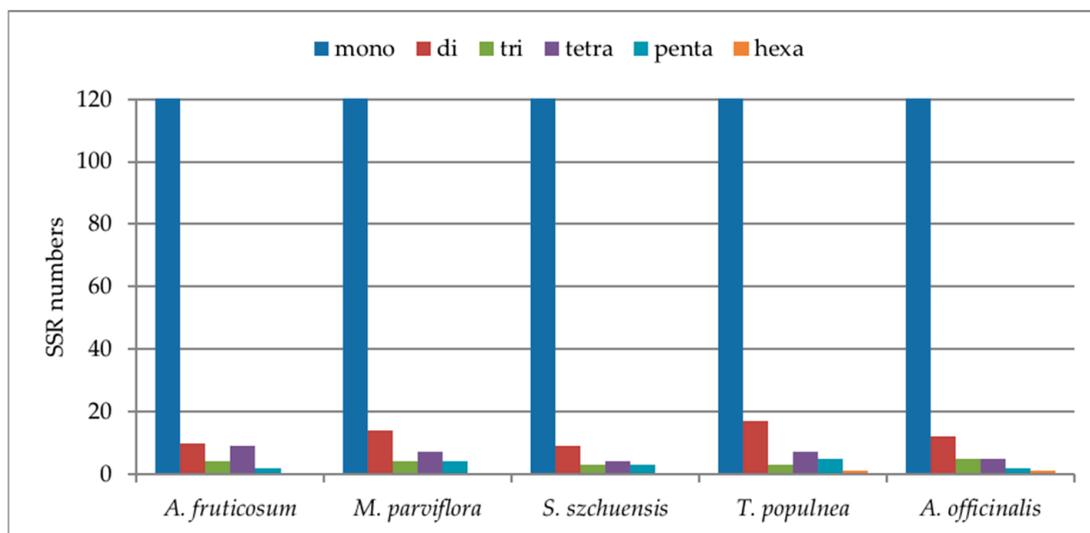


**Figure 4.** Frequency of different simple sequence repeat (SSR) motifs in different repeat types in the *A. fruticosum* chloroplast genome.



**Figure 5.** Number of SSR types in the complete cp genome, and protein-coding and non-coding sequences in *A. fruticosum*.

The rate of occurrence of SSRs among the plastomes of the five species of Malvaceae was compared (Figure 6); the comparison indicates a high number of mononucleotides across all the plastomes. *E. attenuatus* and *A. paniculata* had the highest number of mononucleotides with 107 and 104, respectively. Pentanucleotides were not found in the plastome of *B. prionitis*, *E. attenuatus*, *A. knappiae*, *B. ciliaris* and *R. breedlovei*, while hexanucleotides were only present in *B. prionitis*, *R. breedlovei* and *A. knappiae*.

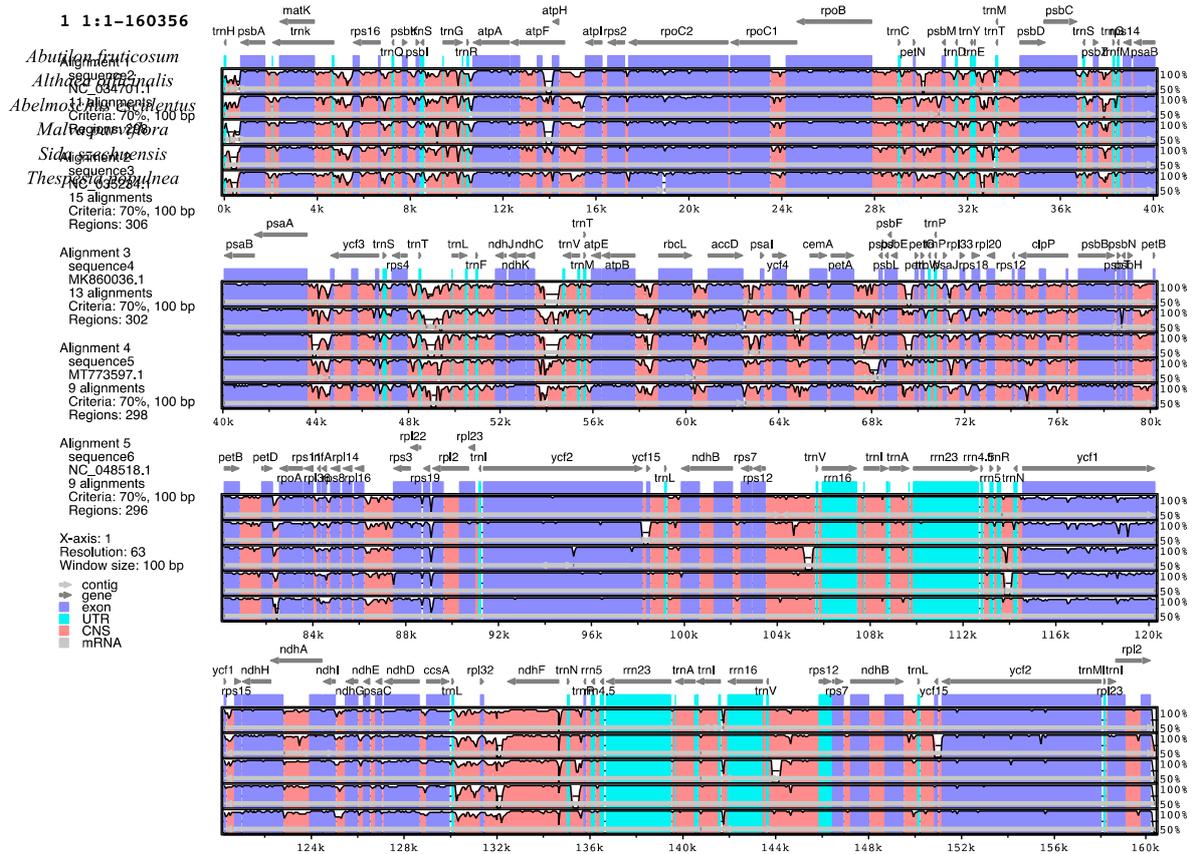


**Figure 6.** Number of different SSR types in the chloroplast genome of five Malvaceae.

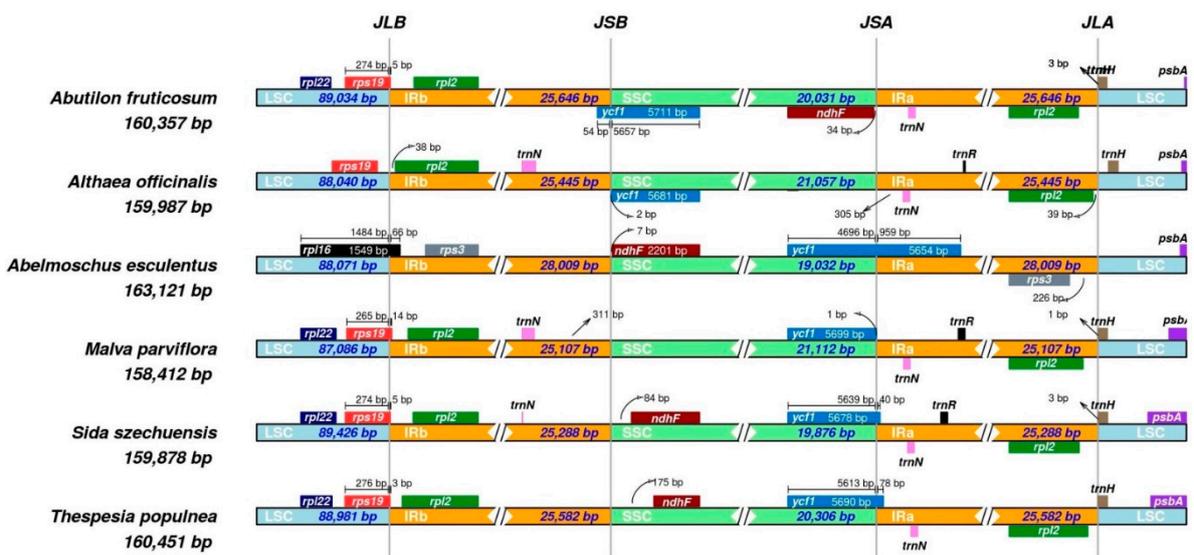
### 2.3. Comparative Analysis of Plastomes of Malvaceae Species

To examine the degree of divergence in the chloroplast genome of the six species of Malvaceae, comparative analysis was conducted using the mVISTA program to align the sequences using the annotation of *A. fruticosum* as a reference. The alignment showed that the genomes are highly conserved with some degree of variation. The coding regions are more conserved than the non-coding regions and the inverted repeat regions are more conserved than the single copy regions (Figure 7). This was reported in the chloroplast genomes of some genera in previous studies [47,48]. The most divergent non-coding regions among the six cp genomes are *trnH-psbA*, *trnK-rps16*, *psbI-trnS*, *atpH-atpI*, *trnT-trnL*, *ndhC-trnV*, *accD-psaI*, *petA-psbJ*, *atpB-rbcL*, *rps12* and *trnL-rpl32*. A slightly lower level of variability was observed in the following genes: *matK*, *ycf1*, *ndhH*, *ycf2* and *accD*. These regions can be used as a source of potential barcodes for identification/authentication of Malvaceae species as well as resources for inferring phylogenetic relationships of the family.

Generally, angiosperms retain the structure and size of the chloroplast genome [46]; however, due to evolutionary events such as an expansion and contraction in the genome, slight variations in the size and location of the boundaries of inverted repeats and single copy regions do occur [49,50]. I compared IR–LSC and IR–SSC boundaries of six cp genomes of Malvaceae (*Abutilon fruticosum*, *Althaea officinalis*, *Abelmoschus esculentus*, *Malva parviflora*, *Sida szechuensis*, *Thespesia populnea*) (Figure 8). The length of the six cp genomes ranged from 158,412 (*M. parviflora*) to 163,121 bp (*A. esculentus*). The genes *rps19*, *rpl2* and *trnH* were located at the junctions LSC–IR and SSC–IR of the compared cp genomes with the exception of *A. esculentus*. The cp genome of *A. fruticosum* is different from the other cp genomes by having the *ndhF* gene in the reverse strand and in the junction of SSC and IRa. The *ycf1* gene is located on the SSC–IRa border in the *A. esculentus*, *M. Parviflora*, *S. szechuensis* and *T. populnea* cp genomes and extends into IRa with 959 bp in *A. esculentus*. The cp genome of *A. esculentus* is unique by having the *rpl16* gene on the LSC–IRb border and the *rps3* gene on the IRa–LSC border. The cp genome of *S. szechuensis* has the smallest IR region, at 25,288 bp, while *A. esculentus* has the longest, at 28,009 bp. The *ndhF* gene is found on the IRb–SSC border of *A. esculentus*, *S. szechuensis* and *T. populnea*.



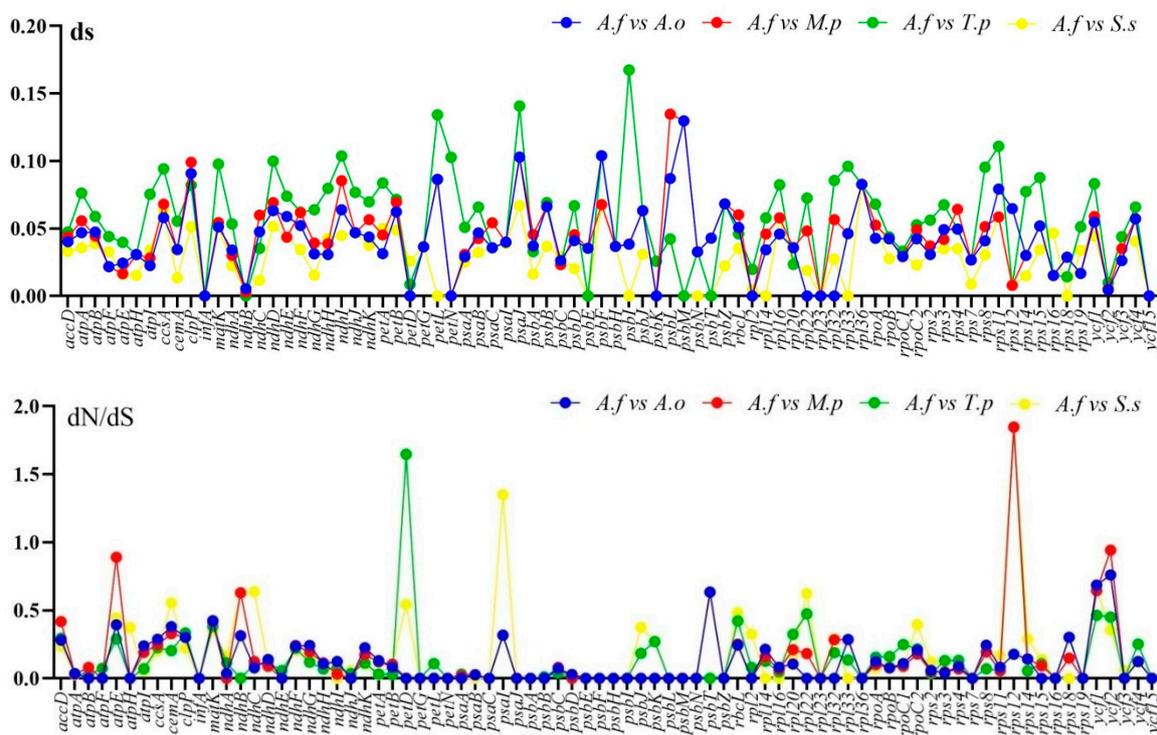
**Figure 7.** Sequence alignment of six chloroplast genomes of representatives of the Malvaceae family performed with mVISTA using annotation of *A. fruticosum* as a reference. The top arrow shows transcription direction, blue color indicates protein-coding sequences, pink color shows conserved non-coding sequences (CNS) and light green indicates tRNAs and rRNAs. The x-axis represents the coordinates in the cp genome, while the y-axis represents percentage identity within 50–100%.



**Figure 8.** Comparison of the borders of the IR, SSC and LSC regions among six chloroplast genomes of Malvaceae (JLB: junction of LSC and IRB; JSB: junction of SSC and IRB; JSA: junction SSC and IRA; JLA: junction LSC and IRA).

#### 2.4. Divergence of Protein-Coding Gene Sequences

The rates of synonymous (dS) and nonsynonymous (dN) substitutions and the dN/dS ratio were calculated using DNAsp among the plastome of six species of Malvaceae to detect whether the 80 shared protein-coding genes were under selective pressure. The results show that the dN/dS ratio is less than 1 in almost all of the paired genes except *petD* of *A. fruticosum* vs. *T. populnea*, *psaI* of *A. fruticosum* vs. *S. szechuensis* and *rps12* of *A. fruticosum* vs. *T. populnea*, *A. fruticosum* vs. *S. szechuensis* and *A. fruticosum* vs. *T. populnea* (Figure 9). This indicates that the majority of the genes were under negative selection, and only three of them underwent positive selection. The synonymous (dS) values range from 0.01 to 0.16 in all the genes (Figure 9). Some of the genes including *infA*, *petG*, *petN*, *psaJ*, *psbA*, *psbZ*, *psbF*, *psbH*, *psbI*, *psbL* and *rps7* showed that no nonsynonymous changes occur in the plastome of the paired species of Malvaceae.

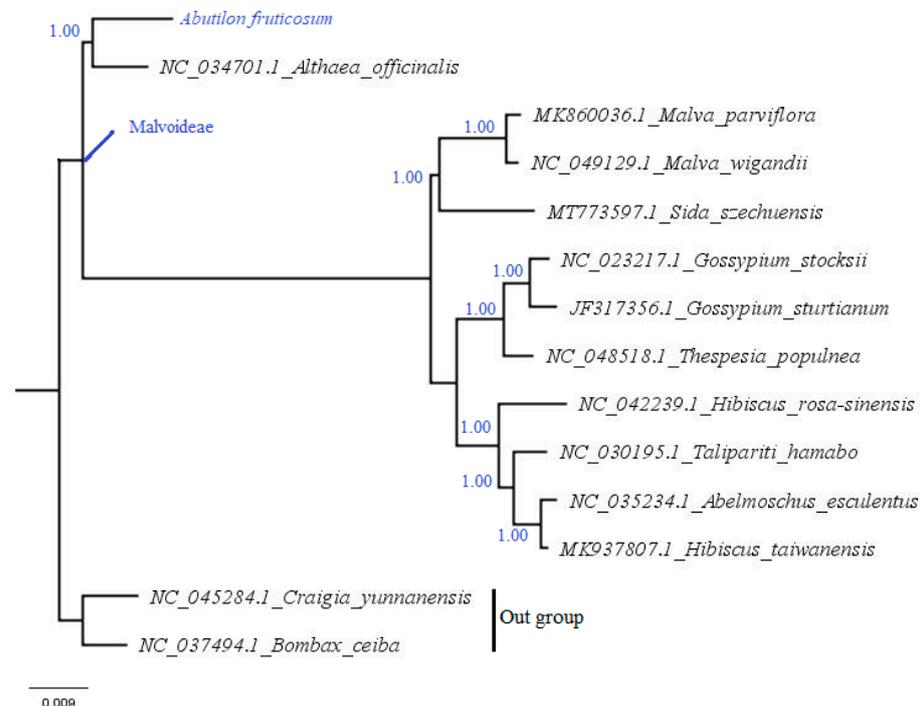


**Figure 9.** The synonymous (dS) and nonsynonymous (dN)/dS ratio values of 80 protein-coding genes from five Malvaceae cp genomes (*A.f.*: *A. fruticosum*; *A.o.*: *A. officinalis*; *M.P.*: *M. parviflora*; *T.p.*: *T. populnea*; *S.s.*: *S. szechuensis*).

#### 2.5. Phylogenetic Analysis

A complete chloroplast genome is a good resource for inferring evolutionary and phylogenetic relationships [51–53]. Many researchers have used plastome sequences to resolve phylogenetic relationships at various taxonomic levels [54,55]. To understand the evolutionary relationship of Malvoideae, Malvaceae and the phylogenetic position of *A. fruticosum* in the family, the complete plastome sequences of 10 species belonging to Malvoideae were downloaded from the GenBank database. In addition, two species, *C. yunnanensis* (Tilioideae, Malvaceae) and *Bombax ceiba* (Bombacoideae, Malvaceae), used as an outgroup, were also downloaded from GenBank. The downloaded cp genomes and the plastome of *A. fruticosum* were aligned using MAFFT. The phylogenetic tree was constructed using the Bayesian inference approach. The results reveal (Figure 10) that the species belonging to the subfamily Malvoideae are in one clade (monophyletic) with highly strong support, with a posterior probability (PP) value of (1.00). This is congruent with previous studies using molecular and morphological data [56–58]. The tree showed four distinct clades: a first clade containing *Abutilon* and *Altheae* and a second clade including

Malvea species and being sister to a large clade containing two clades (Hibisceae and Gossypieae). A similar tree was obtained in a previous study using ITS [59] with slight variation. The species *A. fruticosum* is closely related and sister to *A. officinalis*. This result is incongruent with the earlier systematic position of *A. fruticosum* and *S. szechuensis*. Previous studies [60] reported that two species are sister taxa. In a recent classification, the subfamily Malvoideae [61] was divided into four tribes, namely, Malveae, Hibisceae, Gossypieae and Kydieae. Traditionally, *Abutilon* was placed in Malveae together with *Malva* and *Sida* by various researchers [62,63]. Later, Hutchinson [64] restructured the traditional classification using morphological data, particularly the ovule positions and their number. He proposed an introduction of such tribes as Abutileae (comprising two subtribes Abutilinae and Sidinae), Malveae, Malopeae and Hibisceae. Traditionally, Bentham, Hooker and Schumann classified *Abutilon* (tribe Malveae, subtribe Abutilinae), *Malva* (tribe Malveae, subtribe Eumalvinae), *Sida* (tribe Malveae, subtribe Sidinae) and *Altheae* (tribe Malveae, subtribe Eumalvinae); Hutchinson, later revised *Abutilon* (tribe Abutileae, subtribe Abutilinae), *Malva* and *Altheae* (tribe Malveae, subtribe Malvinae). Here, my results disagree with all the previous tribal positions of the genera. The tree showed that *Abutilon* is closely related to *Altheae* (with strong support) and *Sida*, which was reported as a sister to *Abutilon*, is in a different clade. Additionally, *Malva* and *Altheae* are also in different clades but were included in the same subtribe by previous classification. Based on the result in this study, I proposed the exclusion of *Altheae* from the tribe Malvae and its placement in Abututileae. Comparative analysis in this study (Figures 6 and 7) also showed high similarity between cp genomes of *Abutilon* and *Altheae*. More sequenced chloroplast genomes of the representatives of the subfamily Malvoideae and phylogenetic analysis based on them would still be useful to establish the final systematic position of the genera within it.



**Figure 10.** Phylogenetic tree reconstruction of 14 taxa based on the complete chloroplast genomes using the Bayesian inference (BI) method showing relationships within the 14 species of Malvaceae. The numbers in the branch nodes represent posterior probability (PP).

### 3. Materials and Methods

#### 3.1. Plant Material and DNA Extraction

Leaf material of *Abutilon fruticosum* was collected during field research in Jeddah, Saudi Arabia. Total genomic DNA was extracted from the samples using the Qiagen genomic DNA extraction kit according to the manufacturer's protocol.

#### 3.2. Library Construction, Sequencing and Assembly of the Chloroplast Genome

A total amount of 1.0 µg DNA was used as an input material for the DNA sample preparations. Sequencing libraries were generated using the NEBNext DNA Library Prep Kit for Illumina following the manufacturer's recommendations. The genomic DNA was randomly fragmented into 350 bp long sequences. The raw reads were filtered to get the clean reads (5 Gb) using PRINSEQlite v0.20.4 [65] and were subjected to de novo assembly using NOVOPlasty2.7.2 [66] with kmer (K-mer= 31–33) to assemble the complete chloroplast genome from the whole genome sequence. One contig containing the complete chloroplast genome sequence was generated. The chloroplast genome sequence of *A. fruticosum* has been submitted to GenBank (accession number: MT772391)

#### 3.3. Gene Annotation

Genes were annotated using DOGMA (Dual Organellar GenoMe Annotator, University of Texas at Austin, Austin, TX, USA) [67]. The positions of start and stop codons were adjusted manually. tRNA genes were identified by the trnAscan-SE server (<http://lowelab.ucsc.edu/trnAscan-SE/>) [68]. The circular chloroplast genome maps were drawn using OGDRAW (Organellar Genome DRAW) [69].

#### 3.4. Sequence Analysis

The relative synonymous codon usage values (RSCU), base composition and codon usage were computed using MEGA 6.0. Possible RNA editing sites present in the protein-coding genes of the cp genome of Malvaceae species were determined using PREP suite [35] with 0.8 as the cutoff value.

#### 3.5. Repeat Analysis

Simple sequence repeats (SSRs) were identified in the *Abutilon fruticosum* chloroplast genome using the online software MicroSAteLLite (MISA) [70] with the following parameters: eight, five, four and three repeat units for mononucleotides, dinucleotides, trinucleotides and tetra-, penta- and hexanucleotide SSR motifs, respectively. For analysis of long repeats (palindromic, forward, reverse and complement), the program REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) [71] with default parameters was used to identify the size and location of the repeats in the genome.

#### 3.6. Genome Comparison

The complete chloroplast genomes of six species of Malvaceae were compared with the program mVISTA [72] using the annotation of *A. fruticosum* as a reference in the Shuffle-LAGAN mode [73]. The border regions between the large single copy (LSC) and inverted repeat (IR) and small single copy (SSC) and inverted repeat (IR) junctions were compared using an IR scope.

#### 3.7. Characterization of Substitution Rate

DNAsp v5.10.01 [74] was used to analyze synonymous (dS) and nonsynonymous (dN) substitution rates and the dN/dS ratio to detect the genes that are under selection pressure; the chloroplast genome of *A. fruticosum* was compared with the cp genome of *M. parviflora*, *S. szchuensis*, *T. populnea* and *A. officinalis*.

### 3.8. Phylogenetic Analysis

The complete chloroplast genomes of eleven Malvoideae and two species, *Craigia yunnanensis* (Tilioideae) and *Bombax ceiba* (Bombacoideae), were downloaded from GenBank. The downloaded sequences were aligned with the sequenced cp genome of *A. fruticosum* using MAFFT v.7 [75]. The data were analyzed with the Bayesian inference approach using MrBayes version 3.2.6 [76]. jModelTest version 3.7 [77] was used to select the suitable model.

**Funding:** This research was funded by the DEANSHIP OF SCIENTIFIC RESEARCH (DSR), KING ABDULAZIZ UNIVERSITY, JEDDAH, grant number DF-295-130-1441" and "The APC was funded by D. A. A.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data that support the findings of this study are openly available in GenBank of NCBI at <https://www.ncbi.nlm.nih.gov>, reference number (*A. fruticosum*, MT772391).

**Conflicts of Interest:** The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Miller, P. *The Gardeners Dictionary* 3:23; John and James Rivington: London, UK, 1731. [CrossRef]
2. Sweet, R. *Hortus Britannicus*, 1st ed.; J. Ridgway: London, UK, 1826.
3. Sivarajan, V.V.; Pradeep, A.K. *Malvaceae of Southern Peninsular India: A Taxonomic Monograph*; Daya Publ. House: Delhi, India, 1996; pp. 201–204.
4. Kearney, T.H. A tentative key to the South American species of *Abutilon*. *Leafl. West. Bot.* **1958**, *8*, 201–216.
5. Fryxell, P.A. The American genera of Malvaceae-II. *Brittonia* **1997**, *49*, 204–269. [CrossRef]
6. Fryxell, P.A. An *Abutilon* nomenclator (Malvaceae). *Lundellia* **2002**, *5*, 79–118. [CrossRef]
7. Esteves, G.L.; Krapovickas, A. New species of *Abutilon* (Malvaceae) from Sao Paulo State, Brazil. *Kew Bull.* **2002**, *57*, 479–482. [CrossRef]
8. Husain, S.A.; Baquar, S.R. Biosystematic studies in the genus *Abutilon* from Pakistan. *Phyton* **1974**, *15*, 219–234.
9. Patel, M.K.; Rajput, A.P. Therapeutic significance of *Abutilon indicum*: An overview. *Am. J. Pharm. Tech. Res.* **2013**, *4*, 20–35.
10. Khadabadi, S.S.; Bhajipale, N.S. A review on some important medicinal plants of *Abutilon* spp. *Res. J. Pharm. Biol. Chem. Sci.* **2010**, *1*, 718–729.
11. Pingale, S.S.; Virkar, P.S. Evaluation of acute toxicity for *Abutilon indicum*. *Pharm. Lett.* **2011**, *3*, 37–42.
12. Bano, I.; Deora, G.S. Studies on micro morphological taxonomic variations in *Abutilon* species of Indian Thar Desert. *IOSR J. Pharm. Biol. Sci.* **2017**, *12*, 60–68.
13. Baquar, S.R. *Medicinal and Poisonous Plants of Pakistan*; Printas: Karachi, Pakistan, 1989.
14. Ramar, K.; Ayyadurai, V. The present investigation deals with in-vitro Callus induction and plant regeneration of *Abutilon indicum* (L.). *J. Pharmacogn. Phytochem.* **2015**, *3*, 248–251.
15. Christenhusz, M.J.M.; Byng, J.W. The number of known plants species in the world and its annual increase. *Phytotaxa* **2016**, *261*, 201–217. [CrossRef]
16. Xu, Z.; Deng, M. *Malvaceae, Identification and Control of Common Weeds*, 1st ed.; Springer: Dordrecht, The Netherlands, 2017; pp. 717–735.
17. Grevich, J.J.; Daniell, H. Chloroplast Genetic Engineering: Recent Advances and Future Perspectives. *Crit. Rev. Plant Sci.* **2005**, *24*, 83–107. [CrossRef]
18. Neuhaus, H.; Emes, M. Nonphotosynthetic metabolism in plastids. *Ann. Rev. Plant Biol.* **2000**, *51*, 111–140. [CrossRef] [PubMed]
19. Wicke, S.; Schneeweiss, G.M.; Depamphilis, C.W.; Kai, F.M.; 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]
20. Oldenburg, D.J.; Bendich, A.J. The linear plastid chromosomes of maize: Terminal sequences, structures and implications for DNA replication. *Curr. Genet.* **2016**, *62*, 431–442. [CrossRef]
21. 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]
22. Guisinger, M.M.; Chumley, T.W.; Kuehl, J.V.; Boore, J.L.; Jansen, R.K. Implications of the plastid genome sequence of *Typha* (Typhaceae, poales) for understanding genome evolution in poaceae. *J. Mol. Evol.* **2010**, *70*, 149–166. [CrossRef]

23. Yang, J.B.; Tang, M.; Li, H.T.; Zhang, Z.R.; Li, D.Z. Complete chloroplast genome of the genus *Cymbidium*: Lights into the species identification, phylogenetic implications and population genetic analysis. *BMC Evol. Biol.* **2013**, *13*. [[CrossRef](#)]
24. Chen, H.; Shao, J.; Zhang, H.; Jiang, M.; Huang, L.; Zhang, Z.; Yang, D.; He, M.; Ronaghi, M.; Luo, X.; et al. Sequencing and analysis of *Strobilanthes cusia* (Nees) Kuntze chloroplast Genome revealed the rare simultaneous contraction and expansion of the inverted repeat region in Angiosperm. *Front. Plant Sci.* **2018**, *9*, 324. [[CrossRef](#)]
25. Chang, C.C.; Lin, H.C.; Lin, I.P.; Chow, T.Y.; Chen, H.H.; Chen, W.H.; Cheng, C.H.; Lin, C.Y.; Liu, S.M.; Chang, C.C.; et al. The chloroplast genome of *Phalaenopsis Aphrodite* (Orchidaceae): Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol. Biol. Evol.* **2006**, *23*, 279–291. [[CrossRef](#)]
26. Raman, G.; Park, S. The complete chloroplast genome sequence of *Ampelopsis*: Gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. *Front. Plant Sci.* **2016**, *341*, 7. [[CrossRef](#)] [[PubMed](#)]
27. Park, I.; Kim, W.J.; Yeo, S.-M.; Choi, G.; Kang, Y.-M.; Piao, R.; Moon, B.C. The complete chloroplast genome sequences of *Fritillaria ussuriensis* maxim. In addition, *Fritillaria cirrhosa* D. Don, and comparative analysis with other *Fritillaria* species. *Molecules.* **2017**, *282*, 22.
28. Li, B.; Lin, F.; Huang, P.; Guo, W.; Zheng, Y. Complete chloroplast genome sequence of *Decaisnea insignis*: Genome organization, genomic resources and comparative analysis. *Sci. Rep.* **2017**, *7*, 10073. [[CrossRef](#)] [[PubMed](#)]
29. Campbell, W.H.; Gowri, G. Codon usage in higher plants, green algae, and cyanobacteria. *Plant Physiol.* **1990**, *92*, 1–11. [[CrossRef](#)] [[PubMed](#)]
30. Liu, Q.; Dou, S.; Ji, Z.; Xue, Q. Synonymous codon usage and gene function are strongly related in *Oryza sativa*. *Biosystems* **2005**, *80*, 123–131. [[CrossRef](#)]
31. Srivastava, D.; Shanker, A. Identification of simple sequence repeats in chloroplast genomes of Magnoliids through bioinformatics approach. *Interdiscip. Sci. Comput. Life Sci.* **2016**, *8*, 327–336. [[CrossRef](#)]
32. Zhou, J.; Chen, X.; Cui, Y.; Sun, W.; Li, Y.; Wang, Y. Molecular structure and phylogenetic analyses of complete chloroplast genomes of two *Aristolochia* medicinal species. *Int. J. Mol. Sci.* **2017**, *18*, 1839. [[CrossRef](#)]
33. Jiang, D.; Zhao, Z.; Zhang, T.; Zhong, W.; Liu, C.; Yuan, Q.; Huang, L. The chloroplast genome sequence of *Scutellaria baicalensis* provides insight into intraspecific and interspecific chloroplast genome diversity in *Scutellaria*. *Gene* **2017**, *8*, 227. [[CrossRef](#)]
34. Zhou, J.; Cui, Y.; Chen, X.; Li, Y.; Xu, Z.; Duan, B.; Li, Y.; Song, J.; Yao, H. Complete chloroplast genomes of *Papaver rhoeas* and *Papaver orientale*: Molecular structures, comparative analysis, and phylogenetic analysis. *Molecules* **2018**, *23*, 437. [[CrossRef](#)]
35. Mower, J.P. The PREP suite: Predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res.* **2009**, *37*, 253–259. [[CrossRef](#)]
36. Bundschuh, R.; Altmüller, J.; Becker, C.; Nurnberg, P.; Gott, J.M. Complete characterization of the edited transcriptome of the mitochondrion of *Physarum polycephalum* using deep sequencing of RNA. *Nucleic Acids Res.* **2011**, *39*, 6044–6055. [[CrossRef](#)] [[PubMed](#)]
37. Zeng, W.H.; Liao, S.C.; Chang, C.C. Identification of RNA editing sites in chloroplast transcripts of *Phalaenopsis Aphrodite* and comparative analysis with those of other seed plants. *Plant Cell Physiol.* **2007**, *48*, 362–368. [[CrossRef](#)]
38. Wang, W.; Yu, H.; Wang, J.; Lei, W.; Gao, J.; Qiu, X.; Wang, J. The complete chloroplast genome sequences of the medicinal plant *Forsythia suspense* (oleaceae). *Int. J. Mol. Sci.* **2017**, *18*, 2288. [[CrossRef](#)] [[PubMed](#)]
39. Kumbhar, F.; Nie, X.; Xing, G.; Zhao, X.; Lin, Y.; Wang, S.; Weining, S. Identification and characterisation of RNA editing sites in chloroplast transcripts of einkorn wheat (*Triticum monococcum*). *Ann. Appl. Biol.* **2018**, *172*, 197–207. [[CrossRef](#)]
40. Park, M.; Park, H.; Lee, H.; Lee, B.-H.; Lee, J. The complete plastome sequence of an antarctic Bryophyte *Sanionia uncinata* (Hedw.) loeske. *Int. J. Mol. Sci.* **2018**, *19*, 709. [[CrossRef](#)] [[PubMed](#)]
41. Li, Y.G.; Xu, W.Q.; Zou, W.T.; Jiang, D.Y.; Liu, X.H. Complete chloroplast genome sequences of two endangered *Phoebe* (Lauraceae) species. *Bot. Stud.* **2017**, *58*, 37. [[CrossRef](#)]
42. Greiner, S.; Wang, X.; Rauwolf, U.; Silber, M.V.; Mayer, K.; Meurer, J.; Haberer, G.; Herrmann, R.G. The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. sequence evaluation and plastome evolution. *Nucleic Acids Res.* **2008**, *36*, 2366–2378. [[CrossRef](#)]
43. Song, Y.; Wang, S.; Ding, Y.; Xu, J.; Li, M.F.; Zhu, S.; Chen, N. Chloroplast Genomic Resource of Paris for Species Discrimination. *Sci. Rep.* **2017**, *7*, 3427. [[CrossRef](#)]
44. Bryan, G.J.; McNicol, J.W.; Meyer, R.C.; Ramsay, G.; De Jong, W.S. Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. *Theor. Appl. Genet.* **1999**, *99*, 859–867. [[CrossRef](#)]
45. Provan, J. Novel chloroplast microsatellites reveal cytoplasmic variation in *Arabidopsis thaliana*. *Mol. Ecol.* **2000**, *9*, 2183–2185. [[CrossRef](#)]
46. Ebert, D.; Peakall, R. Chloroplast simple sequence repeats (cpSSRs): Technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. *Mol. Ecol. Resour.* **2009**, *9*, 673–690. [[CrossRef](#)] [[PubMed](#)]
47. Jacinta, N.M.; Xiang, D.; Jia-Xin, Y.; Elijah, M.M.; Vincent, O.W.; Millicent, A.O.; Josphat, K.S.; Paulm, M.M.; Guang-Wan, H. Complete chloroplast genome of *Chlorophytum comosum* and *Chlorophytum gallabatense*: Geome structures, comparative and phylogenetic analysis. *Plants* **2020**, *9*, 296.
48. Dhafer, A.A.; Samaila, S.Y.; Enas, J.A.; Abidina, A. Complete chloroplast genome sequence of *Barleria prionitis*, Comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. *BMC Genom.* **2020**. [[CrossRef](#)]

49. Philippe, H.; Delsuc, F.; Brinkmann, H.; Lartillot, N. Phylogenomics, Annual Review of Ecology. *Evol. Syst.* **2005**, *36*, 541–562. [[CrossRef](#)]
50. Raubeson, L.A.; Peery, R.; Chumley, T.W.; Dziubek, C.; Fourcade, H.M.; Boorem, J.L.; Jansen, R.K. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms *Nupharadvena* and *Ranunculus macranthus*. *BMC Genom.* **2007**, *8*, 174–201. [[CrossRef](#)] [[PubMed](#)]
51. Borsch, T.; Quandt, D. Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. *Plant Syst. Evol.* **2009**, *282*, 169–199. [[CrossRef](#)]
52. Dong, W.P.; Liu, J.; Yu, J.; Wang, L.; Zhou, S.L. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* **2012**, *7*, e35071. [[CrossRef](#)]
53. Tong, W.; Kim, T.S.; Park, Y.J. Rice chloroplast genome variation architecture and phylogenetic dissection in diverse *Oryza* species assessed by whole-genome resequencing. *Rice* **2016**, *9*, 57. [[CrossRef](#)]
54. Dong, W.-P.; Liu, H.; Xu, C.; Zuo, Y.; Chen, Z.; Zhou, S. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: A case study on ginsengs. *BMC Genet.* **2014**, *15*, 138. [[CrossRef](#)]
55. Du, Y.-P.; Bi, Y.; Yang, F.-P.; Zhang, M.-F.; Chen, X.-Q.; Xue, J.; Zhang, X.-H. Complete chloroplast genome sequences of *Lilium*: Insights into evolutionary dynamics and phylogenetic analyses. *Sci. Rep.* **2017**, *7*, 5751. [[CrossRef](#)]
56. Judd, W.S.; Manchester, S.R. Circumscription of Malvaceae (Malvales) as determined by a preliminary cladistic analysis of morphological, anatomical, palynological, and chemical characters. *Brittonia* **1997**, *49*, 384–405. [[CrossRef](#)]
57. Alverson, W.S.; Whitlock, B.A.; Nyffeler, R.; Bayer, C.; Baum, D.A. Phylogeny of the core Malvales: Evidence from *ndhF* sequence data. *Am. J. Bot.* **1999**, *86*, 1474–1486. [[CrossRef](#)] [[PubMed](#)]
58. Bayer, C.; Fay, M.F.; Bruijn, A.Y.D.; Savolainen, V.; Morton, C.M.; Kubitzki, K.; Alverson, W.S.; Chase, M.W. Support for an expanded family concept of Malvaceae within a recircumscribed order Malvales: A combined analysis of plastid *atpB* and *rbcL* DNA sequences. *Bot. J. Linn. Soc.* **1999**, *129*, 267–303. [[CrossRef](#)]
59. Jennifer, A.T.; Javier, F.A.; Steven, J.W.; John, C.L.; Tracey, A.B.D.; Beryl, B.S. Phylogenetic relationships within the tribe Malveae (Malvaceae, subfamily Malvoideae) as inferred from *ITS* sequence data. *Am. J. Bot.* **2005**, *92*, 602.
60. La Duke, J.C.; Doebley, J. A chloroplast DNA based phylogeny of the Malvaceae. *Syst. Bot.* **1995**, *20*, 259–271. [[CrossRef](#)]
61. Bayer, C.; Kubitzki, K. Malvaceae. In *Flowering Plants, Dicotyledons: Malvales, Capparales, and Nonbetalain Caryophyllales*; Kubitzki, K., Bayer, C., Eds.; Springer: Berlin, Germany, 2003; pp. 225–311.
62. Bentham, G.; Hooker, J.D. Malvaceae. In *Genera Plantarum*; Bentham, G., Hooker, J.D., Eds.; Reeve & Co.: London, UK, 1862; pp. 195–213.
63. Schumann, K. Malvaceae. In *Die natuerlichen Pflanzenfamilien*; Wilhelm Engelmann: Leipzig, Germany, 1890; pp. 30–53.
64. Hutchinson, J. *The Genera of Flowering Plants*; Clarendon Press: Oxford, UK, 1967.
65. Schmieder, R.; Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **2011**, *27*, 863–864. [[CrossRef](#)]
66. Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **2016**, *45*, e18.
67. Wyman, S.K.; Jansen, R.K.; Boore, J.L. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **2004**, *20*, 3252–3255. [[CrossRef](#)]
68. Schattner, P.; Brooks, A.N.; Lowe, T.M. The tRNA scan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **2005**, *33*, 686–689. [[CrossRef](#)]
69. 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. Genet.* **2007**, *52*, 267–274. [[CrossRef](#)]
70. Thiel, T.; Michalek, W.; Varshney, R.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422. [[CrossRef](#)] [[PubMed](#)]
71. 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](#)] [[PubMed](#)]
72. Mayor, C.; Brudno, M.; Schwartz, J.R.; Poliakov, A.; Rubin, E.M.; Frazer, K.A.; Pachter, L.S.; Dubchak, I. VISTA: Visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **2000**, *16*, 1046–1047. [[CrossRef](#)] [[PubMed](#)]
73. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, 273–279. [[CrossRef](#)] [[PubMed](#)]
74. Librado, P.; Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451–1452. [[CrossRef](#)] [[PubMed](#)]
75. 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](#)]
76. Fredrik, R.; Maxim, T.; Paul, V.M.; Daniel, L.A.; Aaron, D.; Sebastian, H.; Bret, L.; Liang, L.; Mar, A.S.; John, P.H. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst. Biol.* **2012**, *61*, 539–542.
77. Posada, D. jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* **2008**, *25*, 1253–1259. [[CrossRef](#)]