

Characterization of the Complete Chloroplast Genome of *Blepharis ciliaris* (Acanthoideae, Acanthaceae)

Samaila S. Yaradua^{1, 2,*}, Dhafer A. Alzahrani¹, Abidina Abba¹ and Enas J. Albokhary¹

¹Department of Biology, King Abdulaziz University, Jeddah, Saudi Arabia; ²Centre for Biodiversity and Conservation, Department of Biology, Umaru Musa Yaradua University, Katsina, Nigeria

Received January 24, 2020; Revised February 17, 2020; Accepted: February 22, 2020

Abstract

The complete chloroplast genome of *Blepharis ciliaris*, medicinal and endangered plant in Saudi Arabia was sequenced and characterized. NOVOPlasty was used to assemble the complete chloroplast genome from the whole genome data. The cp genome of *B. ciliaris* is 149,717 bp in length with GC content of 38.5% and has a circular and quadripartite structure; the genome harbors one pair of inverted repeat (IRa and IRb 25,331bp each) separated by large single copy (LSC, 87,073 bp) and small single copy (SSC, 16,998 bp). There are 131 genes in the genome, which include 79 protein-coding genes, 30 tRNA and 4 rRNA; 113 are unique while the remaining 18 are duplicated in IR regions. The repeat analysis indicates that the genome contained all types of repeats with palindromic occurring more frequently; the analysis also identified a total number of 91 simple sequence repeats (SSR) of which the majority are mononucleotides A/T and are found in the intergenic spacer. This study reported the first cp genome of the genus *Blepharis* and provides resources for studying the genetic diversity of *B. ciliaris* as well as resolving phylogenetic relationship within the core Acanthaceae.

Keywords: Acanthaceae; *Blepharis ciliaris*; chloroplast genome; SSR

1. Introduction

Blepharis ciliaris (L.) B. L. Burtt. is a member of the tribe Acantheae (Acanthaceae). The species is known to be distributed in Saudi Arabia, East Africa, East Pakistan and Egypt (Kamal and Abdul, 1956). The plant is used as fodder for ruminant animals particularly camel and sheep. The plant seeds (powdered) are used as an antibacterial on boils, wounds and sores. The seed of the plant is being used to treat cough and has diuretic, aphrodisiac and antibacterial activity (Deshpande, 2006). In addition, the charcoal from the root is used to improve eye vision and treat sore eyes (Boulos, 1981; Tackholm, 1974). Despite the endangered nature and uses in traditional medicine of the species, the complete chloroplast genome of the species was not characterized until this study.

Comparison of complete chloroplast genome provides very informative information for the reconstruction of phylogeny and resolving evolutionary relationship issues at various taxonomic levels (Shaw *et al.*, 2007; Mardanov *et al.*, 2008; Moore *et al.*, 2010; Park *et al.*, 2017; Sun *et al.*, 2017). This is as a result of the conservative nature of the chloroplast genome (Wolfe *et al.*, 1987); this conservative nature is because the plastome evolves about half the rate of other genomes like the nuclear genome (Jansen *et al.*, 2005; Walker *et al.*, 2014). However, rearrangements in the sequence of chloroplast genome were reported by various plant chloroplast genome studies (Doyle *et al.*, 1996; Tangphatsornruang *et al.*, 2011;

Walker *et al.*, 2015; Sun *et al.*, 2017). These rearrangements occur as a result of contractions, expansions and inversions in the single copy regions (large single copy and small single copy) and the inverted repeats (Palmer *et al.*, 1987; Tangphatsornruang *et al.*, 2009). The rearrangements of the genes and inversion in the chloroplast genome are reported to be useful in phylogenetic analyses to solve taxonomic problems at various taxonomic levels because they do not occur often and estimation of their homology and inversion event polarity is simple (Johansson, 1999; Lee *et al.*, 2007; Jansen *et al.*, 2008; Yan *et al.*, 2017). With the importance of complete chloroplast genome in evolutionary studies, resolving phylogenetic relationship issues and the large number of genera and species in Acanthaceae, only complete chloroplast genome of 8 genera have been so far reported (*Andrographis paniculata* (Burm.f.) Nees, NC_022451; *Ruellia breedlovei* T. F. Daniel, KP300014; *Strobilanthes cusia* (Nees) O. Kuntze, MG874806 and four species *Echinacanthus* MF490441, MH045155, MH045156, and MH045157). Seven genera belong to cystolith clade, and only 1 genus belongs to non cystolith clade

In this study, we reported the characteristics of the complete chloroplast genome of *Blepharis ciliaris*, the second cp genome to be sequenced in Acantheae lineage. Moreover, its simple sequence repeats were reported to provide the tools for genetic diversity and identification of the species and lastly to resolve the status of Acantheae

* Corresponding author e-mail: dryaradua@gmail.com.

2. Materials and Methods

2.1. Plant material and DNA extraction

Plant material (vegetative and floral part) was collected through field survey of *B. ciliaris* in North Jeddah, Saudi Arabia. Plant identification and DNA extraction were done according to Samaila *et al.*, (2019).

2.2. Library construction, sequencing and assembly

A total amount of 1.0µg DNA was used as input material for the DNA sample preparations. Sequencing libraries were generated using NEBNext® DNA Library Prep Kit following the manufacturer's recommendations and indices were added to each sample (Samaila *et al.*, 2019). The raw reads were filtered to get the clean reads (5 Gb) using PRINSEQ lite v0.20.4 (Schmieder and Edwards, 2011) and were subjected to de novo assembly using NOVOPlasty2.7.2 (Dierckxsens *et al.*, 2016) with kmer=31 (Samaila *et al.*, 2019)

2.3. Gene annotation

The program DOGMA (Dual Organellar Genome Annotator, University of Texas at Austin, Austin, TX, USA) (Wyman, *et al.*, 2004) was used to annotate the genes in the assembled chloroplast genome. The positions of start and stop codon were adjusted manually. trnAscan-SE server (<http://lowelab.ucsc.edu/trnAscan-SE/>)(Schattner *et al.*, 2005) was used to verify the tRNA genes and finally, the plastome genome circular map was drawn using OGDRAW (Organellar Genome DRAW) (Lohse *et al.*, 2007). The sequence of the chloroplast genome of *B. ciliaris* was deposited in the GenBank database with the accession number (MK548576).

2.4. Sequence Analysis

MEGA 6.0 was used to analyze the relative synonymous codon usage values (RSCU), base composition and codon usage. Possible RNA editing sites present in the protein coding genes of *B. ciliaris* cp genome were determined using PREP suite (Kurtz *et al.*, 2001) with 0.8 as the cutoff value.

2.5. Repeat analysis in *B. ciliaris* chloroplast genome

Simple sequence repeats (SSRs) were identified in the *B. ciliaris* chloroplast genome using the online software MIMoSAtellite (MISA) (Thiel *et al.*, 2003) with the following parameters: eight, five, four and three repeats units for mononucleotides, dinucleotides, trinucleotides and tetra, penta, hexa nucleotides SSR motifs respectively. For analysis of long repeats (palindromic, forward, reverse and complement) the program REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) (Kurtz *et al.*, 2001) with default parameters was used to identify the size and location of the repeats in *B. ciliaris* chloroplast genome

3. Results and Discussion

3.1. Characteristics of *B. ciliaris* chloroplast genome

The complete plastome sequence of *B. ciliaris* has a circular and quadripartite structure. The total length of the genome is 149, 717 bp (Samaila *et al.*, 2019). The plastome has four distinct regions which are Small Single Copy (SSC), Large Single Copy (LSC), and a pair of Inverted repeats (IRa and IRb) which separate the small single copy and large single copy (Figure 1). The region coding for genes is 87, 073bp in length which constitutes 58.15% of the genome, the remaining 62, 644 bp is the non-coding region which includes intron and intergenic spacer (41.85%). The length of the SSC, LCS, IRa and IRb is 16, 998 bp, 82, 057 bp, 25, 331 bp and 25, 331 bp, respectively. The LSC and SSC regions possessed GC content of 36.6 % and 32.4 %, respectively, while the inverted repeats IRa and IRb have 43.6 % (Table 1). The architecture of the *B. ciliaris* cp genome is like other reported Acanthoideae cp genomes (Chunming *et al.*, 2019; Yongbin and Erin, 2017). The plastome sequence has GC of 38.5% and AT content of 61.5%; this is consistent with the plastome data of *Strobilanthes cusia* (Chen *et al.*, 2018). The percentage of GC in the inverted repeat regions is found to be higher than the large single copy region and small single region.

Table 1. Base composition in the *B. ciliaris* chloroplast genome.

Region	T(U)	C	A	G	Total
cp genome	31.1	19.6	30.4	18.9	149717.0
LSC	32.4	18.8	31.0	17.8	82057.0
SSC	33.7	16.7	33.9	15.7	16998.0
IRA	28.3	22.6	28.1	21.0	25331.0
IRB	28.1	21.0	28.3	22.6	25331.0
1st Position	31	19.5	30.4	18.7	49906.0
2nd Position	31	19.6	30.0	19.1	49906.0
3rd Position	31	19.7	30.7	19.0	49905.0

The complete chloroplast genome of *B. ciliaris* contained a total of 131 genes; 113 genes out of the 131 are unique and are present in the single copy regions (large single copy containing 83 genes and the small single copy contained 13 genes); 18 genes are duplicated in the pair of the inverted repeat region which include 8 protein-coding genes, 4 rRNAs and 7 tRNAs. There are 79 protein-coding genes, 4 rRNAs and 30 tRNAs in the plastome (Table 2 and Figure 1). The structural organization and the number of genes in the plastome is consistent with that of the sequenced Acanthoideae cp genome (Chunming *et al.*, 2019; Yongbin and Erin, 2017). Irregular start codons like ACG, GTG and ATC were discovered in some of the annotated genes while majority of genes starts with normal start codon ATG. This phenomenon is also present in some genes in the sequenced chloroplast genome of angiosperms (Chen *et al.*, 2018; Park *et al.*, 2017; Li *et al.*, 2017).

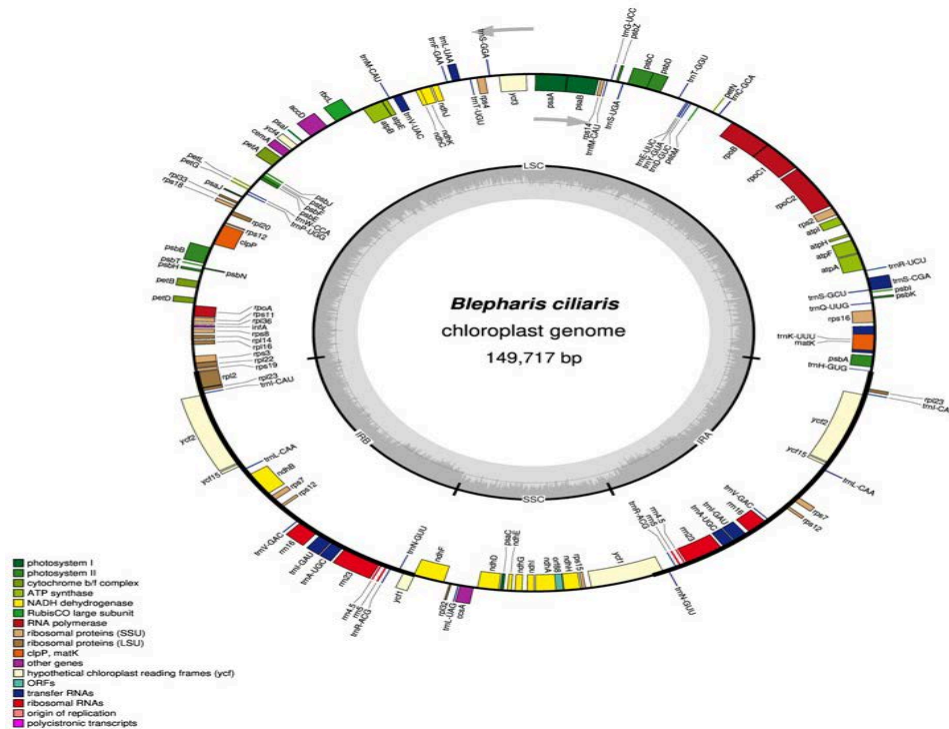


Figure 1. Gene map of the *B. ciliaris* chloroplast genome. Genes outside the circles are transcribed in counter-clockwise direction and those inside in clockwise direction. Known functional genes are indicated in the colored bar. The GC and AT contents are denoted by the dark grey and light grey color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

Table 2. Genes present in the chloroplast genome of *B. ciliaris*

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⁺, trnQ-UUG, trnS-GCU, trnS-CGA⁺, trnR-UCU, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnT-GGU, trnS-UGA, trnM-CAU, trnG-GCC, trnS-GGA, trnL-UAA⁺, trnT-UGU, trnF-GAA, trnV-UAC⁺, trnM-CAU, trnW-CCA, trnP-UGG, trnI-CAU^a, trnL-CAA^a, trnV-GAC^a, trnI-GAU^{+,a}, trnA-UGC^{+,a}, trnR-ACG^a, trnN-GUU^a, trnL-UAG,</i>	
Ribosomal proteins	Small subunit of ribosome	<i>rps2, rps3, rps4, rps7^a, rps8, rps11, rps12^a, rps14, rps15, rps16⁺, rps18, rps19</i>	
Transcription	Large subunit of ribosome	<i>rpl2^{+,a}, rpl14, rpl16, rpl20, rpl22, rpl23^a, rpl32, rpl33, rpl36</i>	
	DNA dependent RNA polymerase	<i>rpoA, rpoB, rpoCI⁺, rpoC2</i>	
Protein genes	Photosystem I	<i>psaA, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	
	Photosystem II	<i>psaA, 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⁺, atpH, atpI</i>	
	Large subunit of rubisco	<i>rbcL</i>	
	NADH dehydrogenase	<i>ndhA⁺, ndhB^{+,a}, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	
	ATP dependent protease subunit P	<i>clpP⁺⁺</i>	
	Chloroplast envelope membrane protein	<i>cemA</i>	
	Other genes	Maturase	<i>matK</i>
		Subunit acetyl-coA carboxylase	<i>accD</i>
C-type cytochrome synthesis		<i>ccsA</i>	
Hypothetical proteins		<i>yef2^a, yef4, yef15^a</i>	
Component of TIC complex	<i>yef1^a</i>		

⁺ Gene with one intron, ⁺⁺ Gene with two intron and ^a Gene with copies

Some of the annotated genes in the chloroplast genome of *B. ciliaris* contained intron, like the sequenced cp genomes of Lamiales (Li *et al.*, 2017, Josphat *et al.*, 2018). Among the 113 coding genes, 14 contained the intron (Table 3). Out of the 14 genes with intron, 8 are protein coding genes and 6 are tRNAs. The large single copy region contained 12 genes while the other four are situated in the inverted repeat region which includes *ndhB*, *trnA-UGC*, *trnI-GAU* and *rpl2*. Two genes, *clpP* and *ycf3* have 2 introns and the other 12 genes have only 1 intron; this has been reported in cp genome of *S. cusia* (Chen *et al.*, 2018). *trnK-UUU* has the longest intron while *trnL-UAA* has the shortest (Table 3).

Table 3: Genes with introns in the *B. ciliaris* chloroplast genome and length of introns and exons

Gene	Location	Exon		Intron		Exon III (bp)
		I (bp)	Intron I (bp)	II (bp)	Intron II (bp)	
<i>rps16</i>	LSC	34	902	225		
<i>atp F</i>	LSC	143	658	470		
<i>rpoC1</i>	LSC	434	783	1628		
<i>ycf3</i>	LSC	128	716	227	716	152
<i>clpP</i>	LSC	68	731	290	557	275
<i>rpl2</i>	IR	392	669	434		
<i>ndhB</i>	IR	776	685	755		
<i>ndhA</i>	SSC	551	905	539		
<i>trnK-UU</i>	LSC	36	2458	37		
<i>trnS-CGA</i>	LSC	57	1059	37		
<i>trnL-UAA</i>	LSC	36	489	49		
<i>trnV-UAC</i>	LSC	37	578	36		
<i>trnI-GAU</i>	IR	41	943	34		
<i>trnA-UGC</i>	IR	37	810	34		

The frequency of the codon usage present in the chloroplast genome was computed using the nucleotide sequence of protein-coding genes and tRNA genes 87,

073bp; the relative synonymous codon usage of the genes in the genome is presented in (Table 4). The results showed the genes in the plastome are encoded by 25, 233 codons. Codons that code for the amino acids leucine appear more frequently in the genome 2, 762 (11.0%) (Figure 2), like that of *Justicia flava* (Samaila *et al.*, 2019), whereas codons coding for Tryptophan are the least 446 (1.90%) in the genome. Guanine and Cytosine ending are found to be more frequent than their counterpart Adenine and Thymine; this is not the case in other plastome sequences (Zhou *et al.*, 2017; Jiang *et al.*, 2017; Zhou *et al.*, 2018). The result of the analysis (Table 4) showed that codon usage bias is low in the chloroplast genome of *B. ciliaris*. The RSCU values of 30 codons were greater than 1 and all of them have A/T ending while for 31 codons were less than 1 and are all of G/C ending. Only two amino acids Tryptophan and Methionine have RSCU value of 1; therefore, they are the only amino acids with no codon bias.

The program PREP suite was used to predict the RNA editing site in the chloroplast genomes of *B. ciliaris*. The first codon of the first nucleotide was used in all the analysis. The result of the analysis shows that most of the conversions in the codon positions are from the amino acid serine to leucine (Table 5). In all, the program revealed 50 editing sites in the genome and they are distributed within 20 protein-coding genes. The gene *ndhB* has the highest number of editing sites with (9 sites); this is consisted with previous researches (Wang *et al.*, 2017; Kumbhar *et al.*, 2018; Park *et al.*, 2018). Other genes with highest number of editing sites in the genome are *rpoB* and *rpoC2* having 6 and 5 editing sites, respectively. The following genes: *accD*, *atpF*, *atpI*, *ndhG*, *rpl2*, *rpl20*, *rpoA* and *rps2* have the lowest number of editing site with 1 editing site. Conversions of proline to serine were observed, which involves the changing of the amino acids in the RNA editing site from apolar to polar group. Genes such as *atpB*, *ccsA*, *clpP*, *ndhC*, *ndhE*, *ndhG*, *petD*, *petG* and *petL*, among others, do not possess RNA predicting site in their first codon of the nucleotide.

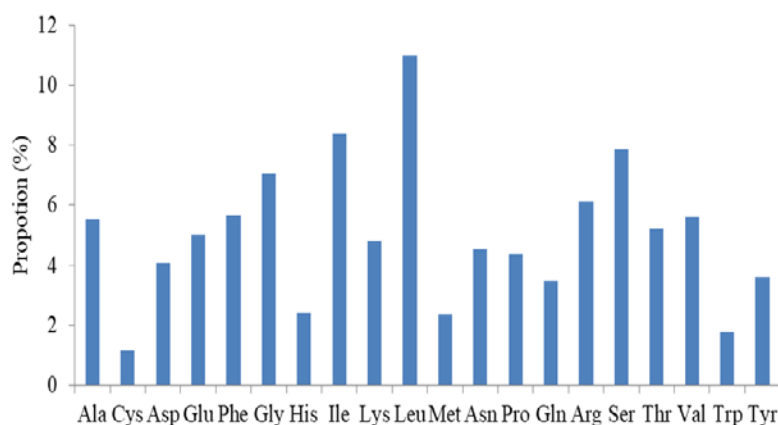


Figure 2: Amino acids frequencies in *B. ciliaris* chloroplast genome protein coding sequences

Table 4: Codon – anticodon recognition patterns and codon usage of the *B. ciliaris* chloroplast genome

Codon	Amino Acid	RSCU	tRNA	Codon	Amino Acid	RSCU	tRNA
UUU	Phe	1.29	<i>tmF-GAA</i>	UAU	Tyr	1.61	<i>tmY-GUA</i>
UUC	Phe	0.71		UAC	Tyr	0.39	
UUA	Leu	1.82	<i>tmL-UAA</i>	UAA	Stop	1.55	
UUG	Leu	1.23	<i>tmL-CAA</i>	UAG	Stop	0.74	
CUU	Leu	1.26	<i>tmL-UAG</i>	CAU	His	1.57	<i>tmH-GUG</i>
CUC	Leu	0.43		CAC	His	0.43	
CUA	Leu	0.86		CAA	Gln	1.53	<i>tmQ-UUG</i>
CUG	Leu	0.39		CAG	Gln	0.47	
AUU	Ile	1.48	<i>tmI-GAU</i>	AAU	Asn	1.51	<i>tmG-GUU</i>
AUC	Ile	0.67		AAC	Asn	0.49	
AUA	Ile	0.85	<i>tmI-CAU</i>	AAA	Lys	1.43	<i>tmK-UUU</i>
AUG	Met	1	<i>tmM-CAU</i>	AAG	Lys	0.57	
GUU	Val	1.49	<i>tmV-GAC</i>	GAU	Asp	1.62	<i>tmD-GUC</i>
GUC	Val	0.52		GAC	Asp	0.38	
GUG	Val	1.43		GAA	Glu	1.45	<i>tmE-UUC</i>
GUA	Val	0.56	<i>tmV-UAC</i>	GAG	Glu	0.55	
UCU	Ser	1.72	<i>tmS-GGA</i>	UGU	Cys	1.53	<i>tmC-GCA</i>
UCC	Ser	0.98		UGC	Cys	0.47	
UCG	Ser	1.17		UGA	Stop	0.71	
UCA	Ser	0.61	<i>tmS-UGA</i>	UGG	Trp	1	<i>tmW-CCA</i>
CCU	Pro	1.45	<i>tmP-UGG</i>	CGU	Arg	1.25	<i>tmR-ACG</i>
CCC	Pro	0.83		CGC	Arg	0.43	<i>tmR-UCU</i>
CCA	Pro	1.11		CGA	Arg	1.35	
CCG	Pro	0.61		CGG	Arg	0.51	
ACU	Thr	1.61		AGA	Arg	1.17	
ACC	Thr	0.76		AGG	Arg	0.35	
ACG	Thr	1.16	<i>tmT-GGU</i>	AGU	Ser	1.75	<i>tmS-GCU</i>
ACA	Thr	0.47	<i>tmT-UGU</i>	AGC	Ser	0.71	
GCU	Ala	1.77	<i>tmA-UGC</i>	GGU	Gly	1.26	<i>tmG-GCC</i>
GCC	Ala	0.65		GGC	Gly	0.42	
GCA	Ala	1.13		GGA	Gly	1.61	
GCG	Ala	0.45		GGG	Gly	0.71	<i>tmG-UCC</i>

Table 5. Predicted RNA editing sites in the *B. ciliaris* chloroplast genome.

gene	Nucleotide Position	Amino Acid Position	Codon Conversion	Amino Acid Conversion	Score
<i>accD</i>	230	77	TCT => TTT	S => F	1
<i>atpA</i>	791	264	CCC => CTC	P => L	1
	914	305	TCA => TTA	S => L	1
<i>atpF</i>	92	31	CCA => CTA	P => L	0.86
<i>atpI</i>	620	207	TCA => TTA	S => L	1
<i>ccsA</i>	71	24	ACT => ATT	T => I	1
	377	126	ACC => ATC	T => I	0.86
<i>matK</i>	70	24	CTT => TTT	L => F	1
	620	207	TCT => TTT	S => F	0.86
	640	214	CAT => TAT	H => Y	1
<i>ndhA</i>	1249	417	CAT => TAT	H => Y	1
	341	114	TCA => TTA	S => L	1
	875	292	TCT => TTT	S => F	1
	1073	358	TCC => TTC	S => F	1

<i>ndhB</i>	149	50	TCA => TTA	S => L	1	
	467	156	TCA => TTA	S => L	1	
	586	196	CAT => TAT	H => Y	1	
	737	246	CCA => CTA	P => L	1	
	746	249	TCT => TTT	S => F	1	
	830	277	TCA => TTA	S => L	1	
	836	279	TCA => TTA	S => L	1	
	1292	431	TCC => TTC	S => F	1	
	1481	494	CCA => CTA	P => L	1	
	<i>ndhD</i>	20	7	ACG => ATG	T => M	1
563		188	GCT => GTT	A => V	1	
896		299	TCA => TTA	S => L	1	
<i>ndhF</i>	671	224	CCA => CTA	P => L	1	
	1535	512	ACC => ATC	T => I	1	
<i>ndhG</i>	314	105	ACA => ATA	T => I	0.8	
<i>petB</i>	424	142	CGG => TGG	R => W	1	
	617	206	CCA => CTA	P => L	1	
<i>rpl2</i>	596	199	GCG => GTG	A => V	0.86	
<i>rpl20</i>	308	103	TCA => TTA	S => L	0.86	
<i>rpoA</i>	887	296	TCG => TTG	S => L	1	
<i>rpoB</i>	338	113	TCT => TTT	S => F	1	
	473	158	TCA => TTA	S => L	0.86	
	551	184	TCA => TTA	S => L	1	
	566	189	TCG => TTG	S => L	1	
	2000	667	TCT => TTT	S => F	1	
	2426	809	TCA => TTA	S => L	0.86	
	<i>rpoC2</i>	1726	576	CTC => TTC	L => F	0.86
		1909	637	CCA => TCA	P => S	1
2248		750	CCC => TCC	P => S	1	
3127		1043	CGT => TGT	R => C	0.8	
3731		1244	TCA => TTA	S => L	0.86	
<i>rps2</i>	248	83	TCA => TTA	S => L	1	
<i>rps14</i>	80	27	TCA => TTA	S => L	1	
	149	50	CCA => CTA	P => L	1	
<i>ycf3</i>	374	125	ACT => ATT	T => I	1	
	388	130	CCT => TCT	P => S	0.86	

3.2. Repeat Analysis

3.2.1. Long repeats

The program REPuter was used to identify long repeat sequences present *B. ciliaris* chloroplast genome using default settings; from the result it was discovered that all the four types of repeats (palindromic, forward, reverse and complement) were present in the plastome of *B. ciliaris* (Table 6). The analysis showed 22 palindromic repeats, 17 forward repeats, 7 reverse repeats and 3 complement repeats (Table 6). In total, there are 49

repeats in the chloroplast genome of *B. ciliaris*. The majority of the repeats size are between 20-29 bp (61.22%), followed by 10-19bp (28.57%), whereas 30-39 bp and 40-49 bp are the least with (6.12% and 4.08%), respectively. In the first location, the intergenic spacer harbored 65.30% of the repeats; this has also been reported in cp genome of *Fagopyron dibotrys* (Xumei *et al.*, 2018). The tRNA contained 4 repeats (8.16%) and 11 repeats (22.44%) are located in the protein coding genes. Within the protein coding genes only *rpoC2*, *ndhC*, *psaB*, *ycf2* and *ycf1* contained the longest repeats.

Table 6: Repeat sequences present in the *B. ciliaris* chloroplast genome.

S/N	Repeat Size	Repeat Position 1	Repeat Type	Repeat Location 1	Repeat Position 2	Repeat Location 2	E-Value
1	41	96141	F	IGS	116548	IGS	1.30E-15
2	41	116548	P	IGS	135532	IGS	1.30E-15
3	32	35433	P	IGS	35433	IGS	3.42E-10
4	30	7483	P	IGS- <i>trnS-GCU</i>	44096	IGS- <i>trnS-GGA</i>	5.47E-09
5	30	113424	P	IGS	113424	IGS	5.47E-09
6	26	85621	P	<i>ycf2</i>	85621	<i>ycf2</i>	1.40E-06
7	26	85621	F	<i>ycf2</i>	146067	<i>ycf2</i>	1.40E-06
8	26	146067	P	<i>ycf2</i>	146067	<i>ycf2</i>	1.40E-06
9	23	42629	F	<i>ycf3</i> Intron	96143	IGS	8.96E-05
10	23	42629	F	IGS	116550	<i>ndhA</i> Intron	8.96E-05
11	23	42629	P	IGS	135548	IGS	8.96E-05
12	23	59041	F	IGS	59062	IGS	8.96E-05
13	22	9018	F	<i>trnG-GCC</i>	35559	<i>trnG-UCC</i>	3.58E-04
14	22	9075	P	IGS	9127	IGS	3.58E-04
15	22	89166	F	<i>ycf2</i>	89184	<i>ycf2</i>	3.58E-04
16	22	89166	P	<i>ycf2</i>	142508	<i>ycf2</i>	3.58E-04
17	22	89184	P	<i>ycf2</i>	142526	<i>ycf2</i>	3.58E-04
18	22	91574	P	IGS	91600	IGS	3.58E-04
19	22	91574	F	IGS	140092	IGS	3.58E-04
20	22	91600	F	IGS	140118	IGS	3.58E-04
21	22	140092	P	IGS	140118	IGS	3.58E-04
22	22	142508	F	<i>ycf2</i>	142526	<i>ycf2</i>	3.58E-04
23	21	7489	F	<i>trnS-GCU</i>	34623	<i>trnS-UGA</i>	1.43E-03
24	21	34623	P	<i>trnS-UGA</i>	44099	<i>trnS-GGA</i>	1.43E-03
25	21	35438	R	IGS	35438	IGS	1.43E-03
26	21	35438	C	IGS	35439	IGS	1.43E-03
27	21	35439	R	IGS	35439	IGS	1.43E-03
28	21	35765	F	<i>trnM-CAU</i>	64866	<i>trnP-UGG</i>	1.43E-03
29	21	111719	P	IGS	111759	IGS	1.43E-03
30	20	35438	P	IGS	35438	IGS	5.73E-03
31	20	35438	F	IGS	35440	IGS	5.73E-03
32	20	35440	P	IGS	35440	IGS	5.73E-03
33	20	49223	R	<i>ndhC</i>	49223	<i>ndhC</i>	5.73E-03
34	20	72300	P	IGS	72324	IGS	5.73E-03
35	20	121066	P	<i>ycf1</i>	121066	<i>ycf1</i>	5.73E-03
36	19	26706	F	IGS	36178	<i>rps14</i>	2.29E-02
37	19	35438	R	IGS	35438	IGS	2.29E-02
38	19	35438	C	IGS	35441	IGS	2.29E-02
39	19	35441	R	IGS	35441	IGS	2.29E-02
40	19	59093	R	IGS	59093	IGS	2.29E-02
41	19	68921	R	IGS	68921	IGS	2.29E-02
42	19	109803	F	IGS	109881	IGS	2.29E-02
43	18	7894	C	IGS	35739	IGS	9.17E-02
44	18	8131	P	IGS	8131	IGS	9.17E-02
45	18	15894	P	<i>rpoC2</i>	59092	IGS	9.17E-02
46	18	35438	P	IGS	35438	IGS	9.17E-02
47	18	35438	F	IGS	35442	IGS	9.17E-02
48	18	35442	P	IGS	35442	IGS	9.17E-02
49	18	37221	F	<i>psaB</i>	39436	<i>psaA</i>	9.17E-02

3.2.2. Simple sequence repeats (SSRs)

There are short repeats of nucleotide series (1-6 bp) that are dispersed all through genome called microsatellites (SSRs). These short repeats in plastid genome are passed from a single parent. As a result, they are used as molecular indicators in developmental studies such as genetic heterogeneity, and also contribute in recognition of species (Bryan *et al.*, 1999; Provan, 2000; Ebert and Peakall, 2009). A total of 91 microsatellites were found in the chloroplast genome of *B. ciliaris* in this

study (Table 7). Majority of SSRs in the cp genome are mononucleotide (83.51%) of which the majority are poly T and A (Figure 4). Poly T (polythymine) constituted 40.65%, whereas poly A (polyadenine) 38.46%; this is consistent with previous studies (Saina *et al.*, 2018; Zhou *et al.*, 2016). Only two poly C (polycytosine) and poly G (polyguanine) were present in the genome, each with 2.17%. Among the dinucleotide only AT/AT is found in the genome. Reflecting series complementary, two trinucleotide AAG/CTT, and AAT/ATT, five tetraAAAT/ATTT, AATC/ATTG, AATT/AATT,

ACAG/CTGT were present in the cp genome. Penta and hexa nucleotides were not discovered in the genome (Figure 4). The intergenic spacer region harbored most of the microsatellite (68.13%) than the coding region (31.86%) (Figure 5). Most, but not all the repeats (70.40%) were detected in the LSC region, and the SSC region incorporates the least number of repeats in the genome.

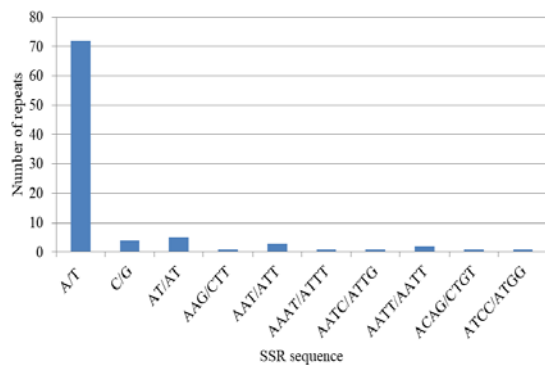


Figure 4: Frequency of different SSR motifs in different repeat types in *B. ciliaris* chloroplast genome

Table 7. Simple sequence repeats in the chloroplast genome of *B. ciliaris*

Repeat Motif	Length	start	End	Region	Annotation
(T)9	9	291	300	LSC	IGS
(AATT)3	3	5403	5414	LSC	IGS
(T)9	9	7230	7238	LSC	IGS
(T)8	8	7902	7909	LSC	IGS
(TA)5	5	7932	7941	LSC	IGS
(T)8	8	8404	8411	LSC	IGS
(A)8	8	9269	9276	LSC	IGS
(GTCT)3	3	10457	10468	LSC	IGS
(T)9	9	11716	11724	LSC	IGS
(T)9	9	12402	12410	LSC	IGS
(T)8	8	13743	13750	LSC	IGS
(T)9	9	14752	14760	LSC	IGS
(TA)5	5	14780	14789	LSC	IGS
(A)9	9	15600	15608	LSC	IGS
(T)9	9	15898	15906	LSC	<i>rpoC2</i>
(T)10	10	17802	17811	LSC	<i>rpoC2</i>
(A)8	8	17944	17951	LSC	<i>rpoC2</i>
(A)8	8	21615	21622	LSC	<i>rpoC1</i>
(T)8	8	25538	25545	LSC	<i>rpoB</i>
(A)8	8	30385	30392	LSC	IGS
(TTC)4	4	34253	34264	LSC	<i>psbC</i>
(T)11	11	34467	34477	LSC	IGS
(A)12	12	35239	35258	LSC	IGS
(A)9	9	35308	35316	LSC	IGS
(TA)11	11	35439	35460	LSC	IGS
(A)8	8	35747	35754	LSC	IGS
(T)8	8	42201	42208	LSC	IGS
(A)9	9	43289	43297	LSC	IGS
(A)8	8	43865	43872	LSC	IGS
(T)9	9	45059	45067	LSC	IGS
(A)9	9	46316	46324	LSC	IGS
(A)11	11	46494	46504	LSC	IGS
(T)8	8	46937	46944	LSC	IGS
(A)8	8	49698	49705	LSC	IGS

(TAT)4	4	51250	51261	LSC	IGS
(ATGG)4	4	51301	51316	LSC	IGS
(T)10	10	53435	53444	LSC	<i>AtpB</i>
Repeat Motif	Length	start	End	Region	Annotation
(ATA)4	4	53489	53500	LSC	IGS
(AAT)4	4	54070	54081	LSC	IGS
(G)8	8	57533	57540	LSC	<i>accD</i>
(ATCA)3	3	58090	58101	LSC	IGS
(TA)6	6	58156	58167	LSC	IGS
(A)9	9	59099	59107	LSC	IGS
(T)8	8	59334	59341	LSC	<i>ycf4</i>
(T)8	8	59773	59780	LSC	IGS
(A)8	8	61221	61228	LSC	<i>petA</i>
(A)8	8	62298	62305	LSC	IGS
(A)8	8	62907	62914	LSC	<i>psbF</i>
(A)8	8	66648	66655	LSC	IGS
(A)8	8	67389	67396	LSC	IGS
(A)8	8	68737	68744	LSC	IGS
(A)9	9	68866	68874	LSC	IGS
(A)12	12	69359	69370	LSC	IGS
(T)8	8	71197	71204	LSC	<i>PsbB</i>
(T)8	8	72022	72029	LSC	IGS
(A)10	10	72115	72124	LSC	IGS
(T)8	8	73151	73158	LSC	IGS
(T)10	10	76123	76132	LSC	<i>rpoA</i>
(T)8	8	78010	78017	LSC	IGS
(T)8	8	81973	81980	LSC	<i>rps19</i>
(T)9	9	82005	82013	LSC	<i>rps19</i>
(A)9	9	87411	87419	IRb	<i>ycf2</i>
(A)8	8	90040	90047	IRb	<i>ycf2</i>
(A)9	9	94536	94544	IRb	IGS
(T)8	8	96519	96526	IRb	IGS
(T)8	8	100252	100259	IRb	IGS
(G)8	8	101469	101476	IRb	IGS
(A)9	9	107656	107664	SSC	<i>ndhF</i>
(A)8	8	108920	108927	SSC	<i>ndhF</i>
(AATA)3	3	112096	112107	SSC	<i>ndhD</i>
(T)8	8	112487	112494	SSC	<i>ndhD</i>
(A)8	8	113461	113468	SSC	IGS
(A)8	8	116785	116792	SSC	IGS
(T)9	9	119309	119317	SSC	<i>rps15</i>
(A)8	8	119521	119528	SSC	<i>rps15</i>
Repeat Motif	Length	start	End	Region	Annotation
(T)9	9	121498	121506	SSC	<i>ycf1</i>
(AATT)3	3	121516	121527	SSC	<i>ycf1</i>
(T)8	8	121529	121536	SSC	<i>ycf1</i>
(T)11	11	121956	121966	SSC	<i>ycf1</i>
(T)8	8	121979	121986	SSC	<i>ycf1</i>
(A)9	9	122030	122038	SSC	<i>ycf1</i>
(T)8	8	122497	122504	SSC	<i>ycf1</i>
(T)9	9	122563	122571	SSC	<i>ycf1</i>
(A)8	8	123209	123216	SSC	<i>ycf1</i>
(C)8	8	130239	130246	IRa	IGS
(A)8	8	131456	131463	IRa	IGS
(A)8	8	135189	135196	IRa	IGS
(T)9	9	137171	137179	IRa	IGS
(T)8	8	141668	141675	IRa	<i>ycf2</i>
(T)9	9	144296	144304	IRa	<i>ycf2</i>

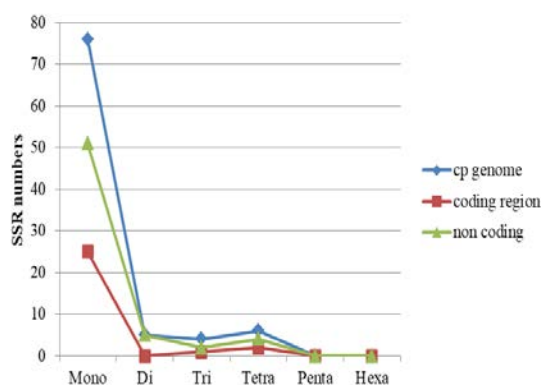


Figure 5: Number of SSR types in complete genome, protein coding regions and Non coding genes.

Acknowledgement

This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant No. (D-468-130-1441). The authors, therefore, gratefully acknowledge DSR technical and financial support.

References

- Boulos L. 1981. Medicinal Plants of North Africa. Reference Publications, Inc., Alonac.
- Bryan GJ, McNicol JW, Meyer RC, Ramsay G and De Jong WS. 1999. Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. *Theor. Appl. Genet.*, **99**:859–867.
- Chen H, Junie S, Hui Z, Mei J, Linfang H, Zhao Z, Dan Y, Molly H, Mostafa R, Xi L, Botao S, Wuwei W and Chang L. 2018. 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.*, **9**, 324.
- Chunming G, Yunfei D and Jun W. 2019. The Complete Chloroplast Genomes of *Echinacanthus* Species (Acanthaceae): Phylogenetic Relationships, Adaptive Evolution, and Screening of Molecular Markers. *Front. Plant Sci.*, **9**, 1989
- Deshpande DJ. 2006. A handbook of medicinal herbs. *Agrobios*, Jodhpur. 342-343.
- Dierckxsens N, Mardulyn P and Smits G. 2016. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.*, **45**, e18.
- Doyle JJ, Doyle JL, Ballenger JA and Palmer JD. 1996. The distribution and phylogenetic significance of a 50-kb chloroplast DNA inversion in the flowering plant family Leguminosae. *Mol. Phylogenet. Evol.*, **5**, 429–438.
- Ebert D and Peakall R. 2009. 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.*, **9**:673–690.
- Jansen RK, Raubeson LA, Boore JL, Chumley TW, Haberle RC and Wyman SK. 2005. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods Enzymol.*, **395**, 348–384.
- Jansen RK, Wojciechowski MF, Sanniyasi E, Lee SB and Daniell H. 2008. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae). *Mol. Phylogenet. Evol.*, **48**, 1204–1217.
- Jiang D, Zhao Z, Zhang T, Zhong W, Liu C, Yuan Q and Huang L. 2017. The chloroplast genome sequence of *Scutellaria baicalensis* provides insight into intraspecific

and interspecific chloroplast genome diversity in *Scutellaria*. *Genes*, **8**, 227.

Johansson JT. 1999. There large inversions in the chloroplast genomes and one loss of the chloroplast *genes16* suggest an early evolutionary split in the genus *Adonis* (Ranunculaceae). *Plant Syst. Evol.*, **218**, 133–143.

Kamal AM and Abdul G. 1956. *Blepharis ciliaris* (L.) B.L Burt in *Notes Roy. Bot. Gard. Edinburg.* **22**, 94.

Kumbhar F, Nie X, Xing G, Zhao X, Lin Y, Wang S and Weining S. 2018. Identification and characterisation of rna editing sites in chloroplast transcripts of einkorn wheat (*Triticum monococcum*). *Ann. Appl. Biol.*, **172**, 197–207.

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J and Giegerich R. 2001. Reputer: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.*, **29**, 4633–4642.

Lee HL, Jansen RK, Chumley TW and Kim KJ. 2007. Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Mol. Biol. Evol.*, **24**, 1161–1180.

Li B, Lin F, Huang P, Guo W and Zheng Y. 2017. Complete chloroplast genome sequence of *Decaisnea insignis*: Genome organization, genomic resources and comparative analysis. *Sci. Rep.*, **7**, 10073.

Lohse M, Drechsel O and Bock R. 2007. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.*, **52**, 267–274.

Mardanov AV, Ravin NV, Kuznetsov BB, Samigullin TH, Antonov AS and Kolganova TV. 2008. Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. *J. Mol. Evol.*, **66**, 555–564.

Moore MJ, Soltis PS, Bell CD, Burleigh JG and Soltis DE. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4623–4628.

Palmer JD, Nugent JM and Herbon LA. 1987. Unusual structure of *Geranium* chloroplast DNA: a triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 769–773.

Park I, Kim WJ, Yang S, Yeo SM, Li H and Moon BC. 2017. The complete chloroplast genome sequence of *Aconitum coreanum* and *Aconitum carmichaelii* and comparative analysis with other *Aconitum* species. *PLoS ONE*, **12**, e0184257.

Park I, Kim WJ, Yeo SM, Choi G, Kang YM, Piao R and Moon BC. 2017. The complete chloroplast genome sequences of *Fritillaria ussuriensis* maxim. In addition, *Fritillaria cirrhosa* D. Don, and comparative analysis with other *Fritillaria* species. *Molecules*, **22**, 982.

Park M, Park H, Lee H, Lee BH and Lee J. 2018. The complete plastome sequence of an antarctic bryophyte *Sanionia uncinata* (Hedw.) loeske. *Int. J. Mol. Sci.*, **19**, 709.

Provan J. 2000. Novel chloroplast microsatellites reveal cytoplasmic variation in *Arabidopsis thaliana*. *Mol. Ecol.*, **9**:2183–2185.

Saina JK, Gichira AW, Li ZZ, Hu GW, Wang QF and Liao K. 2018. The complete chloroplast genome sequence of *Dodonaea viscosa*: Comparative and phylogenetic analyses. *Genetica*, **146**, 101–113.

Samaila SY, Dhafer AA, Enas JA, Abidina A and Abubakar B. 2019. Complete chloroplast genome of *Justicia flava*: Comparative Analysis and Phylogenetic Relationships among Acanthaceae. *Biomed Res. Int.*, 2019.

Samaila SY, Dhafer AA, Enas JA, Abidina A and Abubakar B. 2019. The Complete chloroplast genome of *Blepharis ciliaris* (Acanthaceae) *Mitochondrial DNA B.*, **4**, 3729–3730.

Schattner P, Brooks AN and Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.*, **33**, W686–W689.

- Schmieder R and Edwards R. 2011 Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, **27**, 863–864.
- Shaw J, Lickey EB, Schilling EE and Small RL. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am. J. Bot.*, **94**, 275–288.
- Sun M, Li J, Li D and Shi L. 2017. Complete chloroplast genome sequence of the medical fern *Drynaria roosii* and its phylogenetic analysis. *Mitochondrial DNA Part B.*, **2**, 7–8.
- Tackholm V. 1974. Student's flora of Egypt, 2nd edn. Cairo University, Cairo. 501 – 502.
- Tangphatsornruang S, Sangsrakru D, Chanprasert J, Uthapaisanwong P, Yoocha T and Jomchai N. 2009. The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. *DNA Res.*, **17**, 11–22.
- Tangphatsornruang S, Uthapaisanwong P, Sangsrakru D, Chanprasert J, Yoocha T and Jomchai N. 2011. Characterization of the complete chloroplast genome of *Hevea brasiliensis* reveals genome rearrangement, RNA editing sites and phylogenetic relationships. *Gene*, **475**, 104–112.
- Thiel T, Michalek W, Varshney R and Graner A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.*, **106**, 411–422.
- Walker JF, Jansen RK, Zanis MJ and Emery NC. 2015. Sources of inversion variation in the small single copy (SSC) region of chloroplast genomes. *Amer. J. Bot.*, **102**, 1–2.
- Walker JF, Zanis MJ and Emery NC. 2014. Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia burkei* (Madieae, Asteraceae). *Amer. J. Bot.*, **101**, 722–729.
- Wang W, Yu H, Wang J, Lei W, Gao J, Qiu X and Wang J. 2017. The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (oleaceae). *Int. J. Mol. Sci.*, **18**, 2288.
- Wolfe KH, Li WH and Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 9054–9058.
- Wyman SK, Jansen RK and Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, **20**, 3252–3255.
- Xumei W, Tao Z, Guoqing B and Yuemei Z. 2018 Complete chloroplast genome sequence of *Fagopyrum dibotrys*: genome features, comparative analysis and phylogenetic relationships. *Scientific Reports.*, **8**: 12379.
- Yan M, Moore MJ, Meng A, Yao X and Wang H. 2017. The first complete plastome sequence of the basal asterid family Styracaceae (Ericales) reveals a large inversion. *Plant Syst. Evol.*, **303**, 61–70.
- Yongbin Z and Erin AT. 2017. The draft genome of *Ruellia speciosa* (Beautiful Wild Petunia: Acanthaceae. *DNA Res.*, **24**: 179–192.
- Zhou J, Chen X, Cui Y, Sun W, Li Y, Wang Y, Song J and Yao H. 2017. Molecular structure and phylogenetic analyses of complete chloroplast genomes of two *Aristolochia* medicinal species. *Int. J. Mol. Sci.*, **18**, 1839.
- Zhou J, Cui Y, Chen X, Li Y, Xu Z, Duan B, Li Y, Song J and Yao H. 2018. Complete chloroplast genomes of *Papaver rhoeas* and *Papaver orientale*: Molecular structures, comparative analysis, and phylogenetic analysis. *Molecules*, **23**, 437.
- Zhou T, Chen C, Wei Y, Chang Y, Bai G, Li Z, Kanwal N and Zhao G. 2016. Comparative transcriptome and chloroplast genome analyses of two related *Dipteronia* species. *Front. Plant Sci.*, **7**, 1512.