Applied Biological Research 24(2): 156-167; (2022)
DOI: 10.5958/0974-4517.2022.00019.2

CHARACTERIZATION OF PLASTID GENOME OF A MEDICINAL PLANT *Capparis decidua* (Forsk) Edgew

Dhafer Alzahrani[1,*] and Enas Albokhari[1,2]

[1]*Department of Biological Sciences, Faculty of Sciences, King Abdulaziz University, Jeddah 21589 (Saudi Arabia)*
[2] *Department of Biological Sciences, Faculty of Applied Sciences, Umm Al-Qura University, Makkah 24351 (Saudi Arabia)*
*e-mail: *dalzahrani@kau.edu.sa, dhaferalzahrani@hotmail.com*

(Received 16 December, 2021; accepted 21 March, 2022)

ABSTRACT

***Capparis decidua* is one of the important medicinal plants belonging to the family Capparaceae. The complete plastid genome of *C. decidua*, which is 157,573 bp long, was examined. In this study, we analyzed the complete plastid genome of species using high-throughput sequencing technology and assembled using NOVOPlasty. The large single copy (LSC), small single copy (SSC) and inverted repeats (IRa and IRb) were 85950, 18621 and 26526 bp long. There were 136 genes in the genome: 80 protein-coding genes, 31 tRNA genes, and 4 rRNA genes. The repeat sequences findings showed that chloroplast genome contained a variety of repeats, with palindromic repeat occurring more frequently. A set of 279 SSR were mined from the *C. decidua* genome and most of them possess mononucleotide repeat motifs (A/T).**

Keywords: *Capparis decidua*, chloroplast genome, genes, plastid, SSR

INTRODUCTION

Capparis decidua (Forsk.) Edgew. is a member of genus Capparis, the biggest genus in plant family Capparaceae. The genus *Capparis* contains about 350 species that are found in tropics, and the Mediterranean region (Fici, 2001; Inocencio *et al*., 2006). *C. decidua* is a spinous perennial tree or shrub, densely branched, distributed in India, Saudi Arabia and Africa (Singh and Singh, 2011). *C. decidua* is a medicinal herb with most of its parts used to cure illnesses, such as toothache, arthritis, bronchitis, constipation, inflammation, intermittent fevers, cardiac issues, and swelling all of which have laxative, astringent, and vermifuge properties (Ghazanfar, 1994; Joseph and Jini, 2011; Singh *et al*., 2011). It is also used as a carminative, emmenagogue, alexipharmic, and appetizer; and is useful for treatment of rheumatism, hiccough, lumbago, blisters, boils, swelling, eruptions, as a poison antidote, in the treatment of hypercholesterolemia, and for anti-inflammatory, analgesic, anti-microbial, antiplaque and anti-hypertensive purposes (Chopra *et al*., 2006; Nadkarni, 2009; Verma *et al*., 2011). Cholera and dysentery are treated using the fruits and seeds; it is diuretic in nature and can treat diabetic symptoms. The seed oil is used to cure skin diseases (Singh and Singh, 2011). The plant contains chemical compounds like alkaloids, terpenoids, glycosides and certain fatty acids (Chahlia, 2009).

Recently, the chloroplast genome was used to solve taxonomic problems, along with the discrimination, classification and study of phylogenetic connections among the species (Reboud and Zeyl, 1994; McCauley *et al*., 2007; Liu *et al*., 2018). The cpDNA genome consists of quadripartite configuration, including a small single copy (SSC) and a large single copy (LSC); each is split with

a pair of inverted repeats (IR) with size in between 120 to 160 kb (Yang *et al*., 2010), which conserves the gene content and gene order to a high degree (Jansen *et al*., 2005; Wicke *et al*., 2011). There have been many molecular studies which help us understand the phylogenetic relationships and interspecific variations between *Capparis* species, such as Inocencio *et al*. (2005), Bhoyar *et al*. (2012), Nosrati *et al.* (2012), Aichi-Yousfi *et al.* (2016), and Tamboli *et al.* (2018). For the first time, a whole plastid genome of *Capparis decidua* was sequenced and analyzed in this study, providing crucial data for systematics and comprehension of the *Capparis* genus' variety and evolutionary links for future investigations.

MATERIALS AND METHODS

Plant material and DNA extraction

Young leaves of *C. decidua* were collected from Rabigh area of Saudi Arabia (22°48'31.3"N 39°02'08.9"E). A Qiagen genomic DNA extraction kit was used to obtain genomic DNA from young leaves as per the manufacturer's instructions. A total amount of 1.0 µg DNA was used as input material for DNA sample preparations. Sequencing library was prepared, and PCR done as per Yaradua *et al*. (2019) technique. The complete genome was sequenced using Illuminahiseq 2500 (Novogene-Hong Kong). The generated reads were assembled with NOVOPlasty (Dierckxsens *et al*., 2017) using Kmer (K-mer = 33) to build the entire plastid genome out of the entire genome sequence. *Brassica oleracea* var. *italica* (AY752711) was used as seed for the assembly of chloroplast genome, while entire plastome of *Brassica juncea* (NC 028272) was used as a reference for the assembly of cp genome. After that a single contig comprising the entire chloroplast genome sequence was created.

Gene annotation

The genes were annotated using DOGMA (Dual Organellar GenoMe Annotator, University of Texas, Austin, USA (Wyman *et al.*, 2004). The positions of the starting and end codons were adjusted with care. To find tRNA genes, a trnAscan-SE service (*http://lowelab.ucsc.edu/tRNAscan-SE/*) (Schattner *et al.*, 2005) was used. The circular chloroplast genome maps were created using OGDRAW (Organellar Genome DRAW) (Lohse *et al.*, 2007). The plastid genome sequences of *C. decidua* were deposited in the GenBank under the Accession No.: MT948186.

Sequence analysis

MEGA 6.0 tool was used to examine the relative synonymous codon usage (RSCU) values, base structure, and codon usage. The PREP suite (Kurtz *et al*., 2001) find the RNA editing sites in the protein coding genes with default value of 0.8.

Mining of repetitive regions within the chloroplast genome

SSRs in chloroplast genome were mined with the help of a software tool MIcroSAtellite (MISA) (Thiel *et al*., 2003) using the following criteria: In each, 8, 5, 4, and 3 modules constantly repeat for mononucleotides, dinucleotides, trinucleotides, then tetra, penta, and hexa nucleotide SSR patterns. The REPuter tool (*https://bibiserv.cebitec.uni-bielefeld.de/reputer*) (Kurtz *et al*., 2001) was used with the default settings to detect the position of the lengthy repeats.

RESULTS AND DISCUSSION

Sequence characterization of Capparis decidua chloroplast genome

The chloroplast genome of *C. decidua* has a length of 157,573 bp. The plastome was found to have 4 unique regions: the large single copy (LSC) region of 85,950 bp length, the small single copy

(SSC) region of 18,621 bp length, and the inverted repeats (IRa and IRb) regions of 26,501 bp length, which separated the SSC and LSC (Fig. 1). The coding region in length was 77,822 bp (49.38%) of genome; the non-coding section, which included introns and intergenic spacers, was 70,230 bp long (44.56%). The quantity or proportion of GC observed in IR regions was greater than those in LSC and SSC regions. The GC content inside the IRa and IRb regions was 42.24 and 33.68% in LSC region, while 28.63% in SSC region (Table 1). The IRa and IRb areas had 42.24% GC, the LSC region contained 33.68% GC, as well as the SSC region contained 28.63% GC (Table 1).

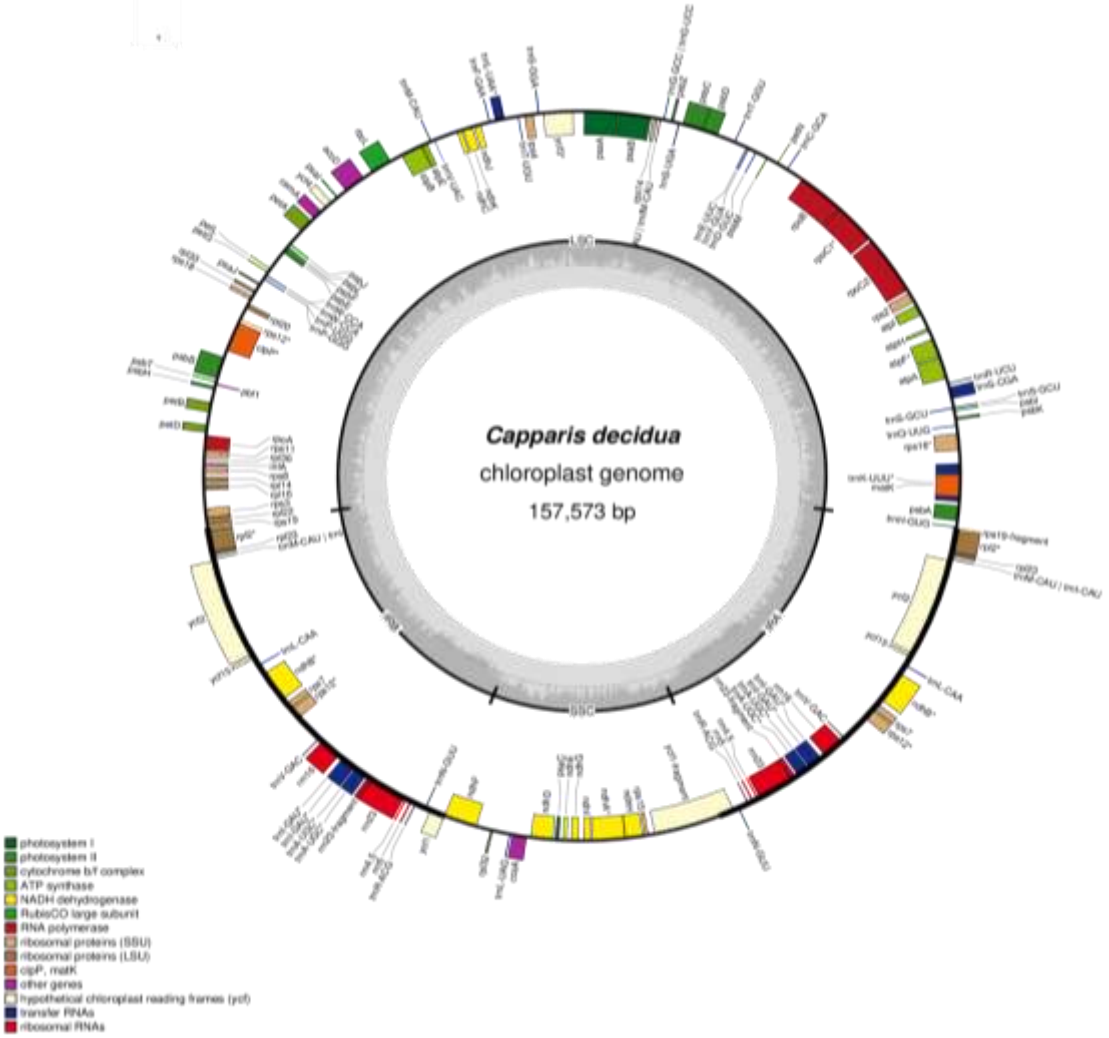

Fig. 1: Gene map of the *Capparis decidua* chloroplast genome. Genes shown outside the circles are transcribed in anticlockwise direction, and those inside are in clockwise direction. The coloured bars indicate functional genes. In inner circle, the dark gray area indicates GC, while the light grey colour indicates the AT contents. LSC indicates the large single copy, SSC indicates the small single copy, and IR indicates inverted repeats

Capparis decidua had 136 genes in its plastome, of which 116 were distinct; 19 genes were replicated inside the pair of IR sections. The plastome contained 80 protein-coding genes, 31 tRNAs, and 4 rRNAs (Fig. 1, Table 2). There were 8 protein-coding genes, 7 tRNA genes, and 4 rRNA genes in IR area. The LSC region had 60 protein-coding genes and 23 tRNA genes, whereas the SSC zone contained 12 protein-coding genes but one tRNA gene. Similar to most flowering plants (angiosperms), all the protein-coding genes in chloroplast genome started with methionine (ATG codon), whereas some genes started with other codons, like ATC, GTG and ACG (Raman and Park, 2016; Park *et al.*, 2017; Li *et al.*, 2017).

Table 1: Base composition in the *C. decidua* chloroplast genome

Region	T or U (%)	C (%)	A (%)	G (%)	Total (bp)
cp genome	32.49	18.30	31.53	17.66	157,573
LSC	33.97	17.30	32.33	16.38	85,950
SSC	35.93	14.91	35.42	13.72	18,621
IRA	28.89	21.98	28.85	20.26	26,501
IRB	28.86	20.25	28.90	21.96	26,501
1st position	32.47	17.94	32.07	17.50	52,525
2nd position	32.24	18.71	31.10	17.92	52,524
3rd position	32.76	18.24	31.42	17.56	52,524

*A (Adenine), T (Thymine), G (Guanine), and C (Cytosine)

Introns were found in some protein-coding genes and tRNA genes in plastomes, as they were in all angiosperms (Raman and Park, 2016; Park *et al.*, 2017). There were 15 introns in 116 coding genes, 8 protein-coding genes, and 7 tRNAs (Table 3). The LSC area had 10 genes, the IR region had 4 genes, and the SSC region had only 1 gene. Two genes, *ycf3* and *clpP*, had 2 introns, while the remaining 13 genes had only 1 intron. In gene *trnK-UUU*, the biggest intron was 2,560 bp long.

Table 2: Gene contents in the plastomes of *Capparis decidua*

Category	Group of genes	Name of genes
RNA genes	Ribosomal RNA genes (rRNA)	*rrn5, rrn4.5, rrn16, rrn23*
	Transfer RNA genes (tRNA)	*trnH-GUG, trnK-UUU*$^{+}$*, trnQ-UUG, trnS-GCU, trnS-CGA, trnR-UCU, trnC-GCA, trnD-GUC, trnY-GUA trnE-UUC, trnT-GGU, trnS-UGA, trnG-GCC, trnfM-CAU, trnS-GGA, trnT-UGU, trnL-UAA*$^{+}$*, trnF-GAA, trnV-UAC*$^{+}$*, trnM-CAU, trnW-CCA, trnP-UGG, trnP-GGG, trnI-CAU, trnL-CAA*^{a}*, trnV-GAC*^{a}*, trnI-GAU*$^{+,a}$*, trnA-UGC*$^{+,a}$*, trnR-ACG*^{a}*, trnN-GUU*^{a}*, trnL-UAG*
Ribosomal proteins	Small subunit of ribosome	*rps2, rps3, rps4, rps7*^{a}*, rps8, rps11, rps12*^{a}*, rps14, rps15, rps16*$^{+}$*, rps18, rps19*
Transcription	Large subunit of ribosome	*rpl2*$^{+,a}$*, rpl14, rpl16, rpl20, rpl22, rpl23a, rpl32, rpl33, rpl36*
	DNA dependent RNA polymerase	*rpoA, rpoB, rpoC1*$^{+}$*, rpoC2*
Protein genes	Photosystem I	*psaA, psaB, psaC, psaI, psaJ, ycf3*$^{++}$
	Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
	Subunit of cytochrome	*petA, petB, petD, petG, petL, petN*
	Subunit of synthase	*atpA, atpB, atpE, atpF*$^{+}$*, atpH, atpI*
	Large subunit of rubisco	*RbcL*
	NADH dehydrogenase	*ndhA*$^{+}$*, ndhB*$^{+,a}$*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
	ATP dependent protease subunit P	*clpP*$^{++}$
	Chloroplast envelope membrane protein	*CemA*
Other genes	Maturase	*MatK*
	Subunit acetyl-CoA carboxylase	*AccD*
	C-type cytochrome synthesis	*CcsA*
	Translation initiation factor	*InfA*
	Hypothetical proteins	*ycf2*^{a}*, ycf4, ycf15*^{a}
	Component of TIC complex	*ycf1*^{a}

$^{+}$ Gene with one intron, $^{++}$ Gene with two introns and ^{a} Gene with copies

Table 3: Length of introns and exons in the *Capparis decidua* chloroplast genome

Gene	Location	Exon I (bp)	Entron I (bp)	Exon II (bp)	Entron II (bp)	Exon III (bp)
rps16	LSC	230	910	35		
atpF	LSC	407	747	143		
rpoC1	LSC	1610	781	446		
ycf3	LSC	149	794	227	714	125
clpP	LSC	224	604	296	902	68
rpI2	IR	431	679	392		
ndhB	IR	752	686	776		
ndhA	SSC	530	1107	551		
trnK-UUU	LSC	28	2560	36		
trnS-CGA	LSC	30	705	59		
trnL-UAA	LSC	34	539	49		
trnV-UAC	LSC	34	614	38		
trnE-UUC	LSC	31	949	39		
trnI-GAU	IR	34	944	41		
trnA-UGC	IR	34	803	37		

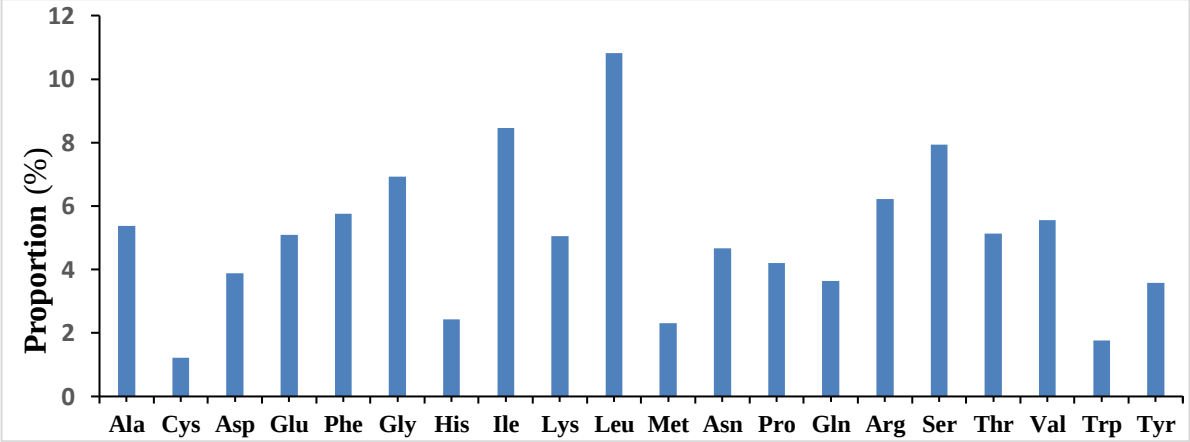

Fig. 2: Protein-coding sequence amino acid concentrations in the chloroplast genome of *C. decidua*

Because of gene bias, the codon use is a factor that impacts the evolution of plastid, and it varies among the species (Srivastava and Shanker, 2016; Li *et al*., 2017). The nucleotide sequence of protein-coding genes and tRNA genes used to calculate the frequency of codon usage in plastomes was 80,726 bp long. Table 4 shows the relative synonymous codon usage of genes in the genome. The results revealed that the genes in plastomes were encoded by 26,809 codons. Leucine was the most frequently occurring codons in the genome, with 2,884 (10.75%) in *C. decidua*, while cysteine was the least with 251 (0.93%) (Fig. 2). With the exception of two with an RSCU value of 1, all amino acids possessed codon use bias, as seen in Table 4 (tryptophan and methionine). The PREP suite tool was used to identify the RNA editing site within plastid. A first position of the first nucleotide was used in experiment. According to the findings, the bulk of codon location swaps occurred from amino acid serine to leucine (Table 5). The system identified 46 editing sites within the sequence, which were distributed across the 18 genes. The results showed higher editing in *ndhA* having 9, then *ndhB* with 8, *matK, rpoB*, while *rpoC1* with 3, and additional genes having 1 or 2 editing sites in the sequence were: *accD, atpF, clpP, ndhA, ndhF, ndhG, psbF, rpoA, rpoC2, rps2, rps14* and *rps16* (Table 5). Conversions of proline to serine have been reported, implying that the amino acids in RNA editing site have shifted from apolar to polar group. Other genes that lacked an RNA predicted site in their first codon were: *atpA, atpB, atpI, ccsA, petB, petD, petG, petL, psaB, psaI, psbB, psbL, rpl2, rpl20, rpl23, rps8* and *ycf3*.

Table 4: *Capparis decidua* chloroplast genome codon-anticodon recognition variations and codon usage

Codon	Amino acid	RSCU	tRNA	Codon	Amino acid	RSCU	tRNA
UUU	Phe	1.29	*trnF-GAA*	UAU	Tyr	1.60	*trnY-GUA*
UUC	Phe	0.71		UAC	Tyr	0.40	
UUA	Leu	1.91	*trnL-UAA*	UAA	Stop	1.20	
UUG	Leu	1.19	*trnL-CAA*	UAG	Stop	1.11	
CUU	Leu	1.25	*trnL-UAG*	CAU	His	1.55	*trnH-GUG*
CUC	Leu	0.45		CAC	His	0.45	
CUA	Leu	0.82		CAA	Gln	1.55	*trnQ-UUG*
CUG	Leu	0.38		CAG	Gln	0.45	
AUU	Ile	1.50	*trnI-GAU*	AAU	Asn	1.55	*trnN-GUU*
AUC	Ile	0.57		AAC	Asn	0.45	
AUA	Ile	0.93	*trnI-CAU*	AAA	Lys	1.49	*trnK-UUU*
AUG	Met	1.00	*trnM-CAU*	AAG	Lys	0.51	
GUU	Val	1.43	*trnV-GAC*	GAU	Asp	1.62	*trnD-GUC*
GUC	Val	0.55		GAC	Asp	0.38	
GUA	Val	1.40		GAA	Glu	1.49	*trnE-UUC*
GUG	Val	0.61	*trnV-UAC*	GAG	Glu	0.51	
UCU	Ser	1.69	*trnS-GGA*	UGU	Cys	1.50	*trnC-GCA*
UCC	Ser	0.93		UGC	Cys	0.50	
UCA	Ser	1.22		UGA	Stop	0.69	
UCG	Ser	0.55	*trnS-UGA*	UGG	Trp	1.00	*trnW-CCA*
CCU	Pro	1.64	*trnP-UGG*	CGU	Arg	1.28	*trnR-ACG*
CCC	Pro	0.71		CGC	Arg	0.39	*trnR-UCU*
CCA	Pro	1.11		CGA	Arg	1.38	
CCG	Pro	0.53		CGG	Arg	0.46	
ACU	Thr	1.68		AGA	Arg	1.81	
ACC	Thr	0.74		AGG	Arg	0.68	
ACA	Thr	1.18	*trnT-GGU*	AGU	Ser	1.20	*trnS-GCU*
ACG	Thr	0.41	*trnT-UGU*	AGC	Ser	0.41	
GCU	Ala	1.85	*trnA-UGC*	GGU	Gly	1.34	
GCC	Ala	0.57		GGC	Gly	0.37	
GCA	Ala	1.15		GGA	Gly	1.66	
GCG	Ala	0.43		GGG	Gly	0.64	*trnG-UCC*

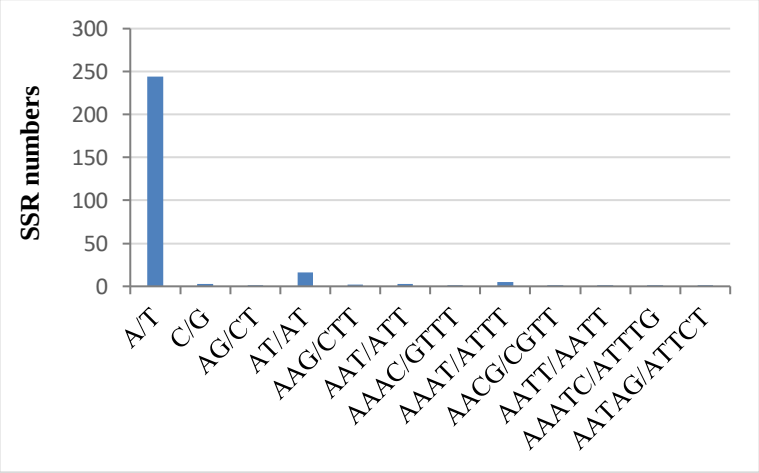

Fig. 3: The frequency of specific SSR motifs in *C. decidua* plastome

Using default settings, the tool REPuter was used to discover lengthy repeat sequences in the plastid. As a result of this, all the four kinds of repeat (palindromic, forward, reverse, and complement) were found in plastome (Table 6). The chloroplast genome had 49 repeats, comprising of 24 palindromic repeats, 16 forward repeats, 7 reverse repeats, and 2 complement repeats (Table 6; Fig. 3). The bulk of repeats (83.67%) were 20-29 bp in length,

Table 5: Predicted RNA editing site in the *Capparis spinosa* chloroplast genome

Gene	Nucleotide position	Amino acid position	Codon	Amino acid	Score
accD	415	139	CCT => TCT	P => S	1.00
	803	268	TCG => TTG	S => L	0.80
atpF	92	31	TCA => TTA	P => L	0.86
clpP	533	178	GCC => GTC	A => V	1.00
	562	188	CAT => TAT	H => Y	1.00
matK	706	236	CAT => TAT	H => Y	1.00
	1250	417	TCA => TTA	S => L	0.86
	1309	437	CAC => TAC	H => Y	1.00
ndhA	341	114	TCA => TTA	S => L	1.00
ndhB	149	50	TCA => TTA	S => L	1.00
	467	156	CCA => CTA	P => L	1.00
	586	196	CAT => TAT	H => Y	1.00
	611	204	TCA => TTA	S => L	0.80
	746	249	TCT => TTT	S => F	1.00
	840	277	TCA => TTA	S => L	1.00
	1255	419	CAT => TAT	H => Y	1.00
	1481	494	CCA => CTA	P => L	1.00
ndhD	20	7	ACG =>ATG	T => M	1.00
	65	22	TCT => TTT	S => F	0.80
	692	231	TCA => TTA	S => L	1.00
	896	299	TCA => TTA	S => L	1.00
	905	302	CCT => CTT	P => L	1.00
	1202	401	GCT => GTT	A => V	0.80
	1316	439	TCA => TTA	S => L	0.80
	1328	443	TCA => TTA	S => L	0.80
	1423	475	CTT => TTT	L => F	0.80
ndhF	205	69	CAT => TAT	H => Y	0.80
	586	196	CTT => TTT	L => F	0.80
ndhG	166	56	CAT => TAT	H => Y	0.80
	314	105	ACA => ATA	T => I	0.80
psbE	214	72	CCT => TCT	P => S	1.00
psbF	77	26	TCT => TTT	S => F	1.00
rpoA	515	172	ACA => ATA	T => I	1.00
	887	296	TCG => TTG	S => L	1.00
rpoB	551	184	TCA => TTA	S => L	1.00
	566	189	TCA => TTA	S => L	1.00
	2432	811	TCA => TTA	S => L	0.86
rpoC1	41	14	TCA => TTA	S => L	1.00
	1717	573	CAC => TAC	H => Y	1.00
	1946	649	ACT => ATT	T => I	0.86
rpoC2	2287	763	CGG => TGG	R => W	1.00
	2336	779	GCC => GTC	A => V	0.86
rps2	248	83	TCA => TTA	S => L	1.00
rps14	80	27	TCA => TTA	S => L	1.00
	149	50	TCA => TTA	S => L	1.00
rps16	176	59	TCA => TTA	S => L	0.83

accompanied by 10-19 bp (6.12%), 30-39 bp, 40-49 bp (4.08% each), and 50-59 bp (2.04%). The intergenic spacer region held 67.34% of repetitions in the first site. Seven repeats (14.28%) were found in tRNA, while ten repeats (20.4%) were found in protein-coding genes (Table 6). The short

Table 6: Predicted RNA editing site in the *Capparis decidua* chloroplast genome

S. No.	Repeat size	Repeat position 1	Repeat type	Repeat location 1	Repeat position 2	Repeat location 2	E-value
1	52	29924	P	IGS	29924	IGS	0 3.44E-22
2	44	75878	P	IGS	75878	IGS	0 2.26E-17
3	41	39721	F	*psaB*	41945	*psaA*	0 1.44E-15
4	36	48507	P	*trnL-UAA* Intron	48507	*trnL-UAA* Intron	0 1.48E-12
5	30	8233	P	IGS	45863	*trnS-GGA*	0 6.06E-09
6	28	128062	P	*ycf1*	128062	*ycf1*	0 9.69E-08
7	27	361	P	IGS	411	IGS	0 3.88E-07
8	27	10085	P	IGS	119862	IGS	0 3.88E-07
9	27	10132	F	IGS	10158	IGS	0 3.88E-07
10	26	9798	P	IGS	9830	IGS	0 1.55E-06
11	26	27540	F	IGS	27566	IGS	0 1.55E-06
12	24	31721	P	IGS	31721	IGS	0 2.48E-05
13	23	12648	R	*atpF*-Intron	12648	*atpF*-Intron	0 9.92E-05
14	23	96259	F	IGS	96281	IGS	0 9.92E-05
15	23	96259	P	IGS	147211	IGS	0 9.92E-05
16	23	96281	P	IGS	147233	IGS	0 9.92E-05
17	23	128550	F	*ycf1*	129262	*ycf1*	0 9.92E-05
18	23	147211	F	IGS	147233	IGS	0 9.92E-05
19	22	28213	P	IGS	28213	IGS	0 3.97E-04
20	22	39698	F	*psaB*	41922	*psaA*	0 3.97E-04
21	22	51707	R	IGS	51707	IGS	0 3.97E-04
22	22	61914	P	IGS	61914	IGS	0 3.97E-04
23	22	112024	P	*ycf1*	112024	*ycf1*	0 3.97E-04
24	22	112024	F	*ycf1*	131469	*ycf1*	0 3.97E-04
25	22	112449	F	*ycf1-ndhF*	112468	*ndhF*	0 3.97E-04
26	22	131469	P	*ycf1*	131469	*ycf1*	0 3.97E-04
27	21	8239	F	*psbI*	36415	*trnS-UGA*	0 1.59E-03
28	21	36415	P	*trnS-UGA*	45866	*trnS-GGA*	0 1.59E-03
29	21	37691	F	*trnfM-CAU*	68004	*trnP-UGG-trnP-GGG*	0 1.59E-03
30	21	46956	F	IGS	120878	IGS	0 1.59E-03
31	21	67437	R	IGS	67437	IGS	0 1.59E-03
32	21	115026	R	IGS	115026	IGS	0 1.59E-03
33	20	187	F	IGS	7995	IGS	0 6.35E-03
34	20	397	P	IGS	65284	IGS	0 6.35E-03
35	20	5103	C	IGS	73216	*clpP*-Intron	0 6.35E-03
36	20	6746	R	IGS	6746	IGS	0 6.35E-03
37	20	6925	P	IGS	78414	IGS	0 6.35E-03
38	20	14066	F	*rrn16*	89841	*ycf2*	0 6.35E-03
39	20	14066	P	*rrn16*	153654	*ycf2*	0 6.35E-03
40	20	27533	P	IGS	27533	IGS	0 6.35E-03
41	20	32448	P	IGS	37322	IGS	0 6.35E-03
42	20	36483	P	*trnS-UGA*	45805	IGS	0 6.35E-03
43	20	51232	R	*ndhC*	51232	*ndhC*	0 6.35E-03
44	20	71745	P	*clpP*-Intron	101941	IGS	0 6.35E-03
45	20	71745	F	*clpP*-Intron	141554	IGS	0 6.35E-03
46	20	115996	R	IGS	115996	IGS	0 6.35E-03
47	19	9706	F	IGS	37517	*trnG-UCC*	0 2.54E-02
48	19	12648	P	*atpF*-Intron	61873	IGS	0 2.54E-02
49	19	12652	C	*atpF*-Intron	61873	IGS	0 2.54E-02

Table 7: The various SSRs in *Capparis decidua* chloroplast genome

Repeat	Length (bp)	Number	Start position
A	8	31	3,569; 4,590; 6,033; 6,957; 13,264; 13,764; 14,836; 15,040; 19,145; 22,810; 30,590; 31,217; 36,714; 37,675; 48,129; 49,244; 52,023; 57,841; 60,068; 67,313; 72,564; 84,513; 98,946; 113,152; 115,687; 116,767; 118,612; 120,358; 126,397; 130,940; 138,470
	9	17	3,996; 13,575; 31,923; 33,138; 33,339; 43,315; 48,304; 67,939; 69,865; 70,611; 71,757; 91,500; 116,322; 118,102; 119,858; 133,246; 144,341
	10	9	28,317; 47,842; 66,780; 66,780; 83,681; 101,131; 112,428; 141,566; 157,438
	11	5	9,778; 16,777; 30,837; 72,177; 110,107
	12	1	123,420
C	8	2	37,076; 63,301
	9	1	49,470
T	7	1	10,342
	8	28	1,701; 3,044; 4,200; 8,676; 12,805; 14,631; 26,646; 29,484; 32,165; 32,959; 43,068; 46,825; 48,799; 53,403; 61,256; 64,763; 66,274; 72,876; 75,665; 78,777; 83,342; 105,039; 115,840; 123,946; 126,153; 127,551; 127,849; 144,563
	9	26	2,376; 4,744; 7,996; 17,106; 19,537; 27,966; 48,515; 49,585; 51,040; 52,466; 60,682; 68,126; 77,234; 81,329; 81,648; 84,127; 84,388; 86,032; 99,167; 110,262; 115,282; 116,548; 123,618; 125,986; 129,821; 152,008
	10	14	3,864; 5,755; 45,324; 47,536; 55,399; 67,591; 101,941; 114,741; 121,718; 126,953; 128,560; 129,272; 131,079; 142,376
	11	7	15,954; 19,002; 28,507; 58,888; 68,650; 72,721; 133,399
	12	3	185; 23,526; 30,192
	13	5	55,792; 58,137; 61,697; 70,431; 79,762
	14	2	81,105; 82,804
	15	1	12,653
AT	5	3	14,255; 20,375; 43,506
	6	4	32,457; 37,323; 46,958; 120,880
	7	1	27,537
TA	5	2	10,082; 128,072
	6	1	64,456
TC	5	1	62,081
AAT	4	1	66,393
GAA	4	1	6,644
TAT	4	1	51,713
TTC	4	1	130,389
AAAT	3	1	4,931
AAAC	3	1	28,687
GTTC	3	1	44,461
TTCTA	3	1	64,216

repeats of nucleotide series are called microsatellites (SSRs) (Schlötterer, 2000), which were found dispensed all through the genome. The chloroplast genome of *C. decidua* contained 279 microsatellites (Table 7). In cp genome of species, the majority of SSR were mononucleotide (88.53%), of which the majority had poly A and T (Fig. 3); Poly A (polyadenine) constituted 37.99%, whereas poly T (polythymine) made up 49.46%. Only three poly C (polycytosine) were

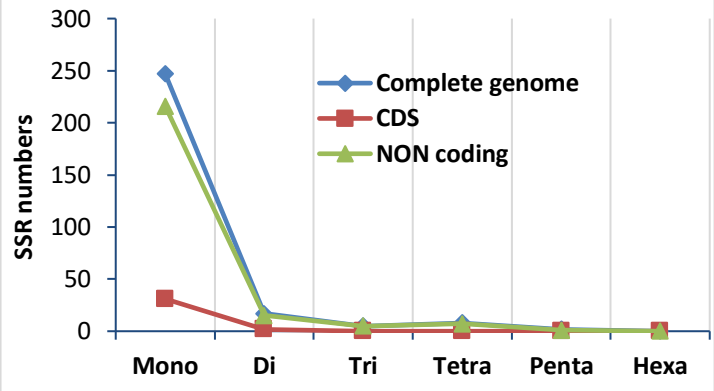

Fig. 4: Number of SSR types in the complete genome, protein-coding regions, and non-coding genes in *Capparis decidua*

present in the genome (1.07%), whereas poly G (polyguanine) was entirely absent in *C. decidua* genome. The sequence of *C. decidua* had 2 dinucleotides (AG/CT and AT/AT), 2 trinucleotides (AAT/ATT and AAG/CTT), 4 tetranucleotides (AAAC/GTTT, AAAT/ATTT, AACG/CGTT, and AATT/AATT), and 2 pentanucleotides (AAATC/ ATTTG and AATAG/ ATTCT); no hexanucleotide (Fig. 3). The inter-genic spacer region (87.45%) had higher representation of micro-satellite than the coding area (11.82%) (Fig. 4).

Conclusion: The current study used the Illumina HiSeq 2500 platform to sequence the complete chloroplast genome of *Capparis decidua* of Capparaceae which provided valuable plastid genomic resource for this medicinally important plant. The cp genome was annotated and identified the base composition, codon usage, RNA editing site, SSRs and long repeat analysis.

Availability of data and material: The data that support the findings of this study are openly available in GenBank of NCBI at *https://www.ncbi.nlm.nih.gov*, refer-ence number (*C. decidua* MT948186).

Code availability: Not applicable

Funding: Not applicable

Conflicts of interest/Competing interests: The authors declare that they have no conflict of interest.

Authors' contributions: Conceptualization and data collection by Dhafer Alzahrani (DA) and Enas Albokhari (EA); formal analysis of data and manuscript writing (review and editing) by EA; manuscript writing (original draft preparation) by DA. All the authors have read and agreed to the published version of the manuscript.

REFERENCES

Aichi-Yousfi, H., Bahri, B.A., Medini, M., Rouz, S., Rejeb M.N. and Ghrabi-Gammar, Z. 2016. Genetic diversity and population structure of six species of *Capparis* in Tunisia using AFLP markers. *Comptes Rendus Biologies,* **339**: 442-453.

Bhoyar, M.S., Mishra, G.P., Naik, P.K., Murkute, A.A. and Srivastava, R.B. 2012. Genetic variability studies among natural populations of *Capparis spinosa* from cold arid desert of trans-Himalayas using DNA markers. *National Academy of Sciences Letters,* **35**: 505-515.

Chahlia, N. 2009. Effect of *Capparis decidua* on hypolipidemic activity in rats. *Journal of Medicinal Plants Research,* **3**(6): 481-484.

Chopra, R.N., Nayar S.L. and Chopra, I.C. 2006. *Glossary of Indian Medicinal Plants.* National Institute of Science Communication and Information Resources, New Delhi. India.

Dierckxsens, N., Mardulyn, P. and Smits, G. 2017. NOVOPlasty: *De novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Research,* **45**(4): 18. [*doi:10.1093/nar/gkw955*].

Fici, S. 2001. Intraspecific variation and evolutionary trends in *Capparis spinosa* L. (Capparaceae). *Plant Systematics and Evolution*, **228**: 123-141.

Ghazanfar, S.A. 1994. *Handbook of Arabian Medicinal Plants*. CRC Press, Boca Raton, Florida, USA.

Inocencio, C., Cowan, R.S., Alcaraz, F., Rivera, D. and Fay, M.F. 2005. AFLP fingerprinting in *Capparis* subgenus *Capparis* related to the commercial sources of capers. *Genetic Resources and Crop Evolution*, **52**: 137-144.

Inocencio, C., Rivera, D., Concepción Obón, M., Alcaraz, F. and Barreña, J.A. 2006. A systematic revision of *Capparis* section *Capparis* (Capparaceae) 1, 2. *Annals of the Missouri Botanical Garden*, **93**: 122-149.

Joseph, B. and Jini, D.A. 2011. A medicinal potency of *Capparis decidua* - A harsh terrain plant. *Research Journal of Phytochemistry*, **5**: 1-13.

Jansen, R.K., Raubeson, L.A., Boore, J.L., dePamphilis, C.W., Chumley, T.W., Haberle, R.C., Wyman, S.K., Alverson, A.J., Peery, R., Herman, S.J., Fourcade, H.M., Kuehl, J.V., McNeal, J.R., Leebens-Mack, J. and Cui, L. 2005. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods in Enzymology*, **395**: 348-384.

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

Li, B., Lin, F.R., Huang, P., Guo, W.Y. and Zheng, Y.Q. 2017. Complete chloroplast Genome sequence of *Decaisneainsignis*: Genome organization, Genomic resources and comparative analysis. *Scientific Reports*, 7: 10073 [*doi:10.1038/s41598-017-10409-8*].

Liu, H.J., Ding, C.H., He, J., Cheng, J., Pei, L.Y. and Xie, L. 2018. Complete chloroplast genomes of *Archiclematis*, *Naravelia* and *Clematis* (Ranunculaceae), and their phylogenetic implications. *Phytotaxa*, **343**: 214-226.

Lohse, M., Drechsel, O. and Bock, R. 2007. Organellar Genome DRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics*, **52**: 267-274.

McCauley, D.E., Sundby, A.K., Bailey, M.F. and Welch, M.E. 2007. Inheritance of chloroplast DNA is not strictly maternal in *Silene vulgaris* (Caryophyllaceae): evidence from experimental crosses and natural populations. *American Journal of Botany*, **94**: 1333-1337.

Nadkarni, K.M. 2009. *Indian Materia Medica* (3rd edn.). Bombay Popular Prakashan, Mumbai, India.

Nosrati, H., Hosseinpour-Feizi, M.A., Mazinani, M. and Haghighi, A.R. 2012. Effect of population size on genetic variation levels in *Capparis spinosa* (Capparaceae) detected by RAPDs. *Eurasian Journal of Medical and Biological Sciences*, **6**: 70-75.

Park, I., Kim, W.J., Yeo, S.-M., Choi, G., Kang, Y.-M., Piao, R. and Moon, B.C. 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**: 282. [*doi:10.3390/molecules22060982*].

Raman, G. and Park, S. 2016. The complete chloroplast genome sequence of *Ampelopsis*: Gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. *Frontiers in Plant Science*, **7**: 341. [*doi: 10.3389/fpls.2016.00341*].

Reboud, X. and Zeyl, C. 1994. Organelle inheritance in plants. Heredity, **72**: 132-140.

Schattner, P., Brooks, A.N. and Lowe, T.M. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research*, **33**: 686-689.

Schlötterer, C. 2000. Evolutionary dynamics of microsatellite DNA. *Chromosoma*, **109**: 365-371.

Singh, D. and Singh, R.K. 2011. Kair (*Capparis decidua*): A potential ethnobotanical weather predictor and livelihood security shrub of the arid zone of Rajasthan and Gujrat. *Indian Journal of Traditional Knowledge*, **10**: 146-155.

Singh, P., Mishra, G., Sangeet, S., Srivastava, K.K., Jha, K. and Khosa, R.L. 2011. Traditional uses, phytochemistry and pharmacological properties of *Capparis decidua*: An overview. *Der Pharmacia Lettre*, **3**: 71-82.

Srivastava, D. and Shanker, A. 2016. Identification of simple sequence repeats in chloroplast genomes of Magnoliids through bioinformatics approach. *Interdisciplinary Sciences: Computational Life Sciences*, **8**: 327-336.

Tamboli, A.S., Yadav, P.B., Gothe, A.A., Yadav, S.R. and Govindwar, S.P. 2018. Molecular phylogeny and genetic diversity of genus *Capparis* (Capparaceae) based on plastid DNA sequences and ISSR markers. *Plant Systematics and Evolution*, **304**: 205-217.

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.). *Theoretical and Applied Genetics*, **106**: 411-422.

Verma, P.D., Dangar, R.D., Shah, K.N., Gandhi, D.M. and Suhagia, B.N. 2011. Pharmacognostical potential of *Capparis decidua* Edgew. *Journal of Applied Pharmaceutical Science*, **1**(10): 6-11.

Wicke, S., Schneeweiss, G.M., dePamphilis, C.W., Muller, K.F. and Quandt, D. 2011. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Molecular Biology*, **76**: 273-297.

Wyman, S.K., Jansen, R.K. and Boore, J.L. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, **20**: 3252-3255.

Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., Yun, Q., Zhao, D., Al-Mssallem, I.S. and Yu, J. 2010. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS One*, **5**(9): [*e12762. https://doi.org/10.1371/journal.pone.0012762*].

Yaradua, S., Alzahrani, D., Albokhary, E. and Abba, A. 2019. Complete chloroplast genome sequence of *Justicia flava*: Genome comparative analysis and phylogenetic relationships among Acanthaceae. *BioMed Research International*, **21**: 1-17. [*https://doi.org/10.1155/2019/4370258*].