Journal of Applied Botany and Food Quality 94, 148 - 161 (2021), DOI:10.5073/JABFQ.2021.094.018

[1]Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia
[2]Department of Biological Sciences, Faculty of Sciences, Federal University Lokoja, Kogi State, Nigeria
[3]Department of Biological Sciences, Umm Al-Qura University, Makkah, Saudi Arabia

Complete chloroplast genome sequencing of *Caralluma quadrangula* and comparative analysis of the Asclepiadoideae subfamily (Apocynaceae)

Abidina Abba[1, 2*], Dhafer A. Alzahrani[1], Samaila S. Yaradua[1], Enas J. Albokhari[1, 3]

(Submitted: March 18, 2021; Accepted: June 19, 2021)

Summary

Caralluma quadrangula is an herbaceous plant distributed in Saudi Arabia, India, Africa, and the Canary Islands. To assess the phylogenetic and evolutionary relationships among the Asclepiadoideae subfamily in the Apocynaceae family, we carried out a comparative chloroplast (cp) genome analysis of four species. The size of the *C. quadrangula* chloroplast genome is 161,456 bp, while the large single-copy (LSC) is 85,992 bp, the small single-copy (SSC) is 13,380 bp, and the two inverted repeats (IRs) are 31,042 bp each. The genome has a total of 129 genes, among which 19 are located at the inverted repeat region, 85 are protein coding genes, 36 are tRNA genes, and eight are rRNA genes. Visualization of single-copy and inverted repeat borders showed that expansion and contraction have occurred in the cp genomes of the compared species of Asclepiadoideae. The genome comparative analysis shows that the coding regions are more highly conserved than the non-coding region; thus, the variable hotspot can serve as a barcode for species in the subfamily. Phylogenetic analysis conducted based on cp sequences identified *Stapelia* as sister to *Caralluma*, in agreement with previous studies. The current study may facilitate the development of molecular markers to ease the identification and authentication of members of the genus *Caralluma* in the Apocynaceae.

Keywords: *Caralluma quadrangula;* Asclepiadoideae; Apocynaceae; Molecular phylogeny; Complete cp genome

Introduction

Caralluma quadrangula (Forssk) N. E.Br., is a medicinal succulent xerophyte plant with small, caduceus leaves. They are perennial and edible herbs (NAIK and KRISHNAMURTHY, 2012). Arabian folk medicine has recorded several ethno-medicinal uses of *C. quadrangula*; for example, to treat diabetes, cancer, snake venom, tuberculosis, scorpion bite, fever, skin disorders, inflammation, and so on (LEO et al., 2005). In many cultural communities, as it has no established side effects, the plant is commonly used as a hunger food. It is distributed throughout the Arabian Peninsula, India, South Africa, Europe, Sri Lanka, Iran, Africa, and the Canary Islands (MEVE and LIEDE, 2004).

Several species have been reported to contain pregnane glycosides, different esters, and megastigmane glycosides, confirming their medicinal properties (WAHEED et al., 2011). Antioxidant, anticancer, antidiabetic, anti-inflammatory, antimicrobial, antieczema, antimalarial, and antifungal properties of various *Caralluma* extracts have also been reported (WAHEED et al., 2011; Rauf et al., 2013).

Caralluma is a genus of 23 species in the subfamily Asclepiadoideae of Apocynaceae (formerly family Asclepiadaceae), known as the milkweeds (BENSUZAN, 2009). Apocynaceae are among the ten largest Angiosperm families with presently about 5,300 species in 378 genera of which more than 3,000 species in 180 genera belong to subfamily Asclepiadoideae (ENDRESS et al., 2018). Apocynaceae is one of the five families in the order Gentianales, in the Lamiid (Asterid I) clade of the Asterids (Magnoliopsida) (STEVENS, 2001 onwards).

The Asclepiadoideae subfamily is made up of five tribes: Fockeeae, Eustegieae, Marsdenieae, Ceropegieae, and Asclepiadeae (ENDRESS et al., 2018). In our study, however, we focus on the three major tribes: Ceropegieae (*C. quadrangula*), Marsdenieae *(Hoya pottsii* J. Traill, *Gymnema sylvestre* (Retz.) R. Br. ex Schult.), and Asclepiadeae (*Calotropis gigantea* (L.) R.Br. and *Cynanchum wilfordii* (Maxim.) Hemsley.

In Saudi Arabia, *C. quadrangula* is included in the region's endangered species list. For proper species management and conservation, characterizing the genetic profile of the species is important. To the best of our knowledge, this is the first study recording its complete chloroplast (cp) genome.

Although the cp genome is generally highly conserved, it has undergone serious evolutionary changes in most plants either of the same generic group or those plants belonging to a different genus in terms of intron loss, expansion and contraction of the inverted repeats, and inversion of the DNA strand (JANSEN et al., 2002; SLOAN et al., 2014). As such, cp genomic information might be very useful in molecular phylogenetic and evolutionary studies. This technique can also unmask the degree of differences in cp genomes and investigate cp nucleus gene transfer in the evolution of eukaryotic cells. In addition, a peculiar diversity of cp genomes and the absence of recombination and heteroplasmy have allowed scientists to assess the role of seed dispersal in gene flow evolution (MCCAULEY, 1995). Even though whole cp genome studies can provide more genetic data, in comparison to only one or few cpDNA fragments (JIN et al., 2011), several historic phylogenetic problems relating to different angiosperms have been successfully resolved at various taxonomic grades (WICKE et al., 2010; YANG et al., 2017). Further, cp genome provides many advantages in analyzing cp transformation which are superior to nuclear transformation, considering the enhanced trans-gene expression and a lack of pollen (DANEILL et al., 2016). Due to the cost-effectiveness of sequencing technology, many plant genomes, especially of land plants, have become available in the NCBI Genbank database, after the first published study (FERRARINI et al., 2013).

Our study involves comparative cp genome analysis of the four species detailed above, from the same subfamily but different tribes, in order to evaluate the variations in their cp genome architecture.

Sequencing technology has greatly increased the number of cp genomic sequences in recent years. Nevertheless, to date, a small number of plastid genomes of Apocynaceae species have been published and, thus, it is crucial to evaluate additional complete cp genomes in the Apocynaceae family, in order to further phylogenetic and genomic research studies.

The present study was conducted in order to assemble and characterize the cp genome of *C. quadrangula* by using Illumina sequencing,

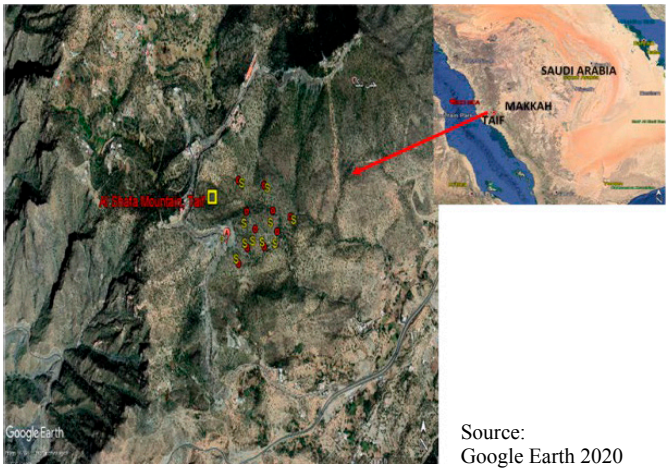
* Corresponding author

for the purposes of 1) evaluating common structural characteristics and variable regions, 2) exploring comparative evolution in the cp genomes (*C. gigantea*, *H. pottsii*, and *C. wilfordii*). We hope that our research can provide useful genetic resources for further investigation of the subfamily Asclepiadoideae.

Materials and methods

Plant sample collection and extraction of DNA

Caralluma quadrangula plant material was collected from Taif in the Al-Shafa mountains (23° 4.7'32" N. 40° 18.6'28'E), a town about 25 km from the holy city of Makkah (Fig. 1), and identified at the King Abdulaziz University herbarium in Jeddah. Fresh material of plants, both floral and vegetative parts, was collected from Al-Shafaa Hills Taif, Makkah Province, Saudi Arabia, and identified using herbarium specimens and morphological characters from literatures. The voucher specimen was stored at the King Abdulaziz University's herbarium in Jeddah, Saudi Arabia. The specimen's leaves were collected for genomic DNA extraction in which Qiagen DNeasy plant minikit (Qiagen, Korea) was used to extract genomic DNA according to the manufacturer's protocol. DNA was diluted to 30 ng/L and UV absorption values at A260 were used.



Source: Google Earth 2020

Fig. 1: Map of Shafa Mountains showing the sampling locations.

Library preparation, sequencing, and assembly

A quantity of 1.0 µg of leaf sample was measured for DNA preparation, to serve as an input. An NEBNext DNA Library Kit was used to generate the sequencing libraries, following the manufacturer's instructions, with indices included in each sample. The DNA was fragmented erratically by shearing, obtaining 350 bp size; polished strands of DNA were then ligated with NEBNext adapter for Illumina sequencing and tire indexed P7 and P5. This was followed by purification for PCR then analysis for the effective magnitude of the DNA by the Agilent 2100 kit and measured using real-time Polymerase chain reaction (RT-PCR). Raw reads obtained from the illumine sequences were filtered to obtain a 5 GB file with PRINSEQ lite v0.20.4 (Schmieder and Edwards, 2011). The FastQ files were assembled using NOVOPlasty2.7.2 (Dierckxsens et al., 2016), (Kmer = 32), in order to obtain a complete cp genome from the whole genome. The ndhax1-ndhAx2 intergenic spacer from *C. wilfordii* (KX352467K) served as a seed, while *Calotropis procera* (NC_041440) was the reference in which a contig was generated.

Gene annotations

For annotation of the *C. quadrangula* cp genome, the DOGMA program (University of Texas, Austin) (Wyman et al., 2004) was employed, with manual adjustments for stop and start codons applied. Annotation of tRNA genes was executed using the tRNAscan-SE program; while the cp circular structure was drawn by the use of OGDRAW (Organellar Genomes Draw) (Lohse et al., 2007). Sequences were deposited to Genbank with assigned accession number MT413385.

Analysis of the sequence and repeats

To analyze relative synonymous codon usage and base compositions, MEGA version 6.0 (Tamura et al., 2013) was used, while the PREP-suite program (Mower, 2009), was used in determining the RNA editing sites, using a cutoff value of 0.8. The identification of Simple sequence repeats of *C. quadrangula* was done using the Microsatellite (MISA) program (Theil et al., 2003). We carried out simple sequence repeats (SSR) analyses in the cp genomes of *C. quadrangula*, *H. pottsii*, *Asclepias nivea*, and *C. gigantea*, with the parameters of 8, 5, 4, and 3 repeat units for mono-nucleotide, di-nucleotide, tri-nucleotide, tetra-nucleotide, penta-nucleotide, and hexa-nucleotide motifs, respectively. The Reputer software (Kurtz et al., 2001) was used for the identification of tandem, complement, reverse, palindromic, and forward repeats in the genome, following the manual of >15 base pairs and a sequence identity of 90%.

Genome comparison and characterizations of substitution rates (Ka/Ks)

The mVISTA (Mayor et al., 2000) program was employed to compare the four cp genomes of Asclepiadoideae species namely, *C. quadrangula*, *H. pottsii*, *G. sylvestre*, and *C. wilfordii* using the annotation of *C. quadrangula* as reference with the Shuffle-LAGAN method (Frazer et al., 2004). Border junctions of SSC, LSC, and IRs were all compared between the four species. DNAsp v5.10.01 (Librado and Rozas, 2009) was used for the analysis of the Ka/Ks ratio between the sequence of *C. quadrangular* and sequences of *H. pottsii*, *G. sylvestre*, and *C. wilfordii*, all of the Asclepiadoideae subfamily (Apocynaceae). The protein encoding genes were individually aligned using the GeniousV8.1.3 program and then translated to protein sequences.

Phylogenetic analysis

The complete cp genomes of *C. quadrangula* and 13 other species (9 from the Asclepiadoideae subfamily) and 2 out group species of the Rauvolfioideae subfamily were downloaded from Genbank and aligned with the MAFFT program v.7 with (L-INS-i) option selected (Katoh and Standley, 2003), aligned sequences were further analyzed with Maximum Parsimony PAUP ver. 4.0b 10 (Felsenstein, 1978), using heuristic searches with 1,000 replicate tree bisection-reconnections, branch swapping, and random taxon addition; further, MulTrees was kept on and tree saving of 100 as a peak value for all replicates. Missing characters were considered as a gap, while support was determined using 1,000 replicates for the non-parametric bootstrap method. Bayesian Inference (BI) analyses were carried out using MrBayes v.3.2.6 (Ronquist and Huelsenbeck, 2003), running four Markov chains (one cold and three heated) simultaneously for 50 million generations. Sampling frequency was set to 1000 and all other parameters were set as default. Model selection for the BI analyses was conducted with MrModeltest v.3.06 (Posada and Crandall, 1998) under the Akaike information criterion. Mr-Model-test selected the general time reversible (GTR) model of DNA substitutions with gamma-distributed rate variation across invariant sites.

Results

Our results include details regarding genes with introns, SSR identification, codon usage, RNA editing sites, comparative genomics with Mvista, Ka/Ks ratio, IR border junction, and the Phylogenetic relationship of Asclepiadoideae.

Genome characterization of *C. quadrangula*

The *C. quadrangula* cp genome is161, 456 bp in length and comprises four sections, namely, the small single-copy (SSC) region, the large single-copy (LSC) region, and a pair of inverted repeats. The coding region has a length of 94,775 bp, thus making up 58.7% of the genome. Annotation results of the genome show that there are 129 genes, of which 19 are duplicative genes, 85 are protein-coding genes, 36 are *tRNA* genes, and eight are *rRNA* genes (Fig. 2 and Tab. 1). The genome has an overall GC content of 38%, while the inverted repeat regions have high GC content of 44.36%, followed by the LSC region (31.6%).

Tab. 2 shows the comparative genome features of *C. quadrangula*, *H. pottsii*, *C. wilfordii*, and *C. gigantea*. The genome sizes indicate that *C. wilfordii* (161,180 bp) has the smallest, followed by *C. quadrangula* (161,456 bp), *H. pottsii* (161,565 bp), and *C. gigantea* (165,928 bp) is the largest. Among the four genomes, *C. gigantea* has the largest LSC region (94,372 bp); while the smallest were *C. quadrangula* (85,992 bp) and C. *wilfordii* (91,976 bp).

The largest SSC region was in *C. gigantea* (22,444 bp); similarly, *H. pottsii* has the largest protein coding, tRNA, and rRNA region sizes (75,455, 3170, and 9118 bp, respectively), while the smallest belong to *C. wilfordii*, *C. quadrangula*, and *C. gigantea* (74,371, 3050, and 9054 bp each); we also found that the genomes with the highest number of genes belonged to *H. pottsii* and *C. gigantea* (133 genes each), followed by *C. wilfordii* (131 genes);while the smallest genome belongs to *C. quadrangula* (129 genes). All the genomes have the same number of tRNA genes (36 each); while *C. quadrangula* and *H. pottsii* have higher number of *rRNA* genes (8 each), followed by *C. gigantea* and *C. wilfordii* (4 each) (Tab. 4).

Genes with introns

Introns are an important factor in determining gene expression. They are not only in charge of gene structure and functions, but also intracellular gene expression as well as transformation effectiveness (Li et al., 2013; Xu et al., 2002). There are some genes in the *C. quadrangula* cp genome which hold one or two introns in a number of protein-coding and *tRNA* genes, as seen in the majority of angiosperms (Raman and Park, 2016). There are 18 genes with introns out of the129 genes in its genome (Tab. 3); of which, 6 are tRNA genes and the remaining 12 are protein-coding genes. The genes with the longest introns are *trn*KUU (2480), *acc*D (1194), and *ndh*A (1104). The only genes with two introns are *atp*F, *acc*D, and *ycf3*; *trn*KUU

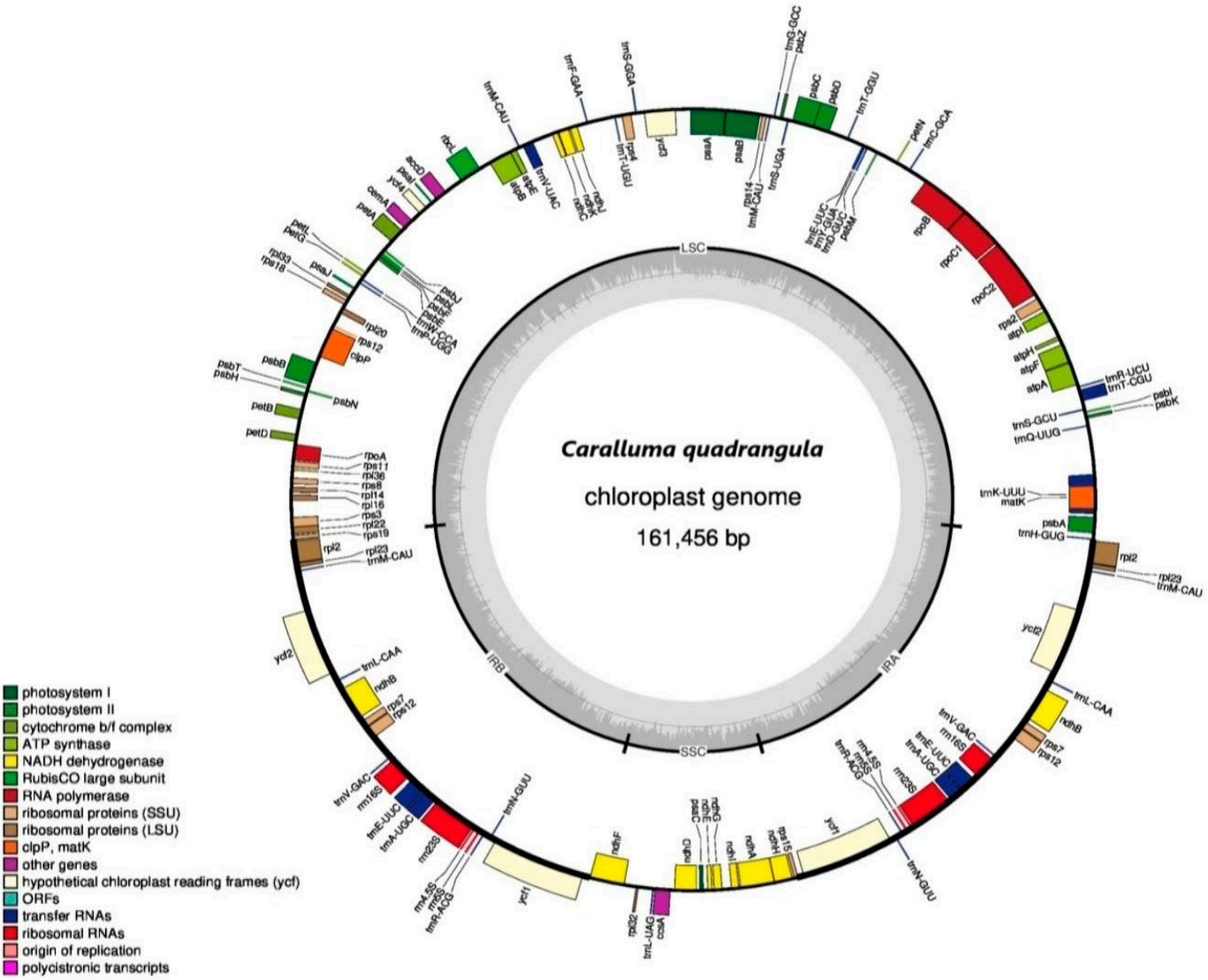


Fig. 2: Cp genome map of *Caralluma quadrangula*: Genes on the outside are pulling towards the counterclockwise direction, while the inner genes are pulling in the clockwise direction. Functional groups are represented with colors. GC content represented by dark portion (slight grey), while AT content is shown in light gray. The LSC, SSC, and IRs are all represented in the map.

Tab. 1: *Caralluma quadrangular* cp genome base compositions.

	T (U) (%)	C (U) (%)	A (U) (%)	G (U) (%)	Total (bp)
cp genome	31.2915	19.19532	30.87405	18.63913	161.456
LSC	33.53284	16.2523	34.80663	15.40823	85.992
SSC	33.80058	16.75449	33.89778	15.54714	13.380
IRA	27.24265	20.22416	30.34451	22.18868	31.042
IRB	27.24265	20.22416	30.34451	22.18868	31.042
1st location	31.3681	18.90782	31.15443	18.56965	53.819
2nd location	31.5632	19.24042	30.52639	18.66999	53.819
3rd location	30.94318	19.43773	30.94132	18.67777	53.818

Tab. 2: Summary of the complete cp genome features of *C. quadrangula*, *H. pottsii*, *C. wilfordii*, and *C. gigantea*.

	C. quadrangula	*H. pottsii*	*C. wilfordii*	*C. gigantea*
Genome size (bp)	161.456	161.565	161.180	165.928
Overall GC content (%)	38	38	38	37.5
LSC (bp)	85.992	92.532	91.976	94.372
SSC (bp)	13.380	19.719	19.930	22.444
IR (bp)	31.042	24.657	24.637	74.383
Protein coding region (bp)	74.505	75.455	74.371	74.383
tRNA (bp)	3050	3212	3170	3168
rRNA (bp)	9062	9118	9064	9054
Total number of genes	129	133	131	133
Number of Protein coding genes	85	89	91	89
Number of genes in SSC	12	13	13	12
Number of genes in LSC	100	102	101	106
Number of genes in IR	17	18	17	15
No. of tRNA genes	36	36	36	36
No. of rRNA genes	8	8	4	8

Tab. 3: Genes with introns in the sequence of *Caralluma quadrangula* cp genome.

Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
rps16	LSC	225	856	40		
atp F	LSC	410	706	143	751	51
rpoC1	LSC	1609	751	450		
ycf3	LSC	154	799	225	743	125
clpP	LSC	383	682	293		
rpl2	IR	390	640	433		
ndhB	IR	755	685	776		
ndhA	SSC	544	1104	552		
trnK-UU	LSC	35	2480	36		
trnG-UCC	LSC	22	672	36		
trnL-UAA	LSC	36	490	49		
trnV-UAC	LSC	36	582	37		
trnM-GAU	IR	41	944	37		
trnA-UGC	IR	37	832	34		
accD	LSC	655	1194	194	628	08
rpoC2	IR	433	640	390		
rps12	IR	09	545	235		
ycf1	SSC	552	953	538		

(2480) was the largest intron, as it contains the *mat*K gene, as stated previously (Saina et al., 2016; Kim and Lee, 2016; Shen et al., 2017). The *rps12* gene feature has a 5' end trans-spliced gene in the LSC region and 3' ends in the IR regions. The *ycf3* gene is very critical for stable photosystem I accumulation (Boudreau et al., 2016; Naver et al., 2001). The c*lp*P gene is part of the gene responsible for protease synthesis (Ramundo and Rochaix, 2014). In addition, introns are valuable resources and promote the creation of new plant varieties with high disease and insect pest resistance.

Codon usage analysis

Single-nucleotide substitution in protein-coding sequences results in

the formation of synonymous or non-synonymous codons this occurs in all 20 amino acids; except for Methionine and Tryptophan, where they are encoded by at least two codon usage biases. Synonymous codon usage bias is the consequence of genetic events overtime, such as drift, mutation, and natural selection during evolutionary process (Akashi, 2001; Wang et al., 2014). Previous studies have shown that synonymous codon usage bias is heterogeneous in the organellar genomes of land plants and plays a major role in plant evolution (Pal, 2006). About 27,342 codons were identified, based on the protein coding genes (except those with stop codons). All genes had the ATG canonical starting codon, with the exception of *ndh*, which began with ACG. The three most frequent amino acids were leucine (5265; 19.2%), isoleucine (2215; 8.55%), and serine (2063; 7.97%), compared with the least amino acid of cysteine 1098 (4.02%) (Tab. 5). For amino acids coded by several codons, codon use at synonymous third-place sites was directed towards A and U (Alzahrani et al., 2020b; Abba et al., 2020), a similar bias was found in the genome of *C*. *quadrangula* cp. Of the 28 favored codons (Relative synonymous codon Usage, RSCU > 1), 9 finished in A or U. Of the less frequently used 37 Codons (RSCU < 1), however, 13 terminated in a G or C. Moreover, Methionine and Tryptophan do not have codon bias; this is because their RSCU value is one (Abba et al., 2021). The codon usage analysis in Tab. 5 indicates mostly high variability in the possible coding of the amino acids leucine and arginine (six codons each); meanwhile, valine, serine, proline, threonine, alanine, and glycine each had four codons. The least of the amino acids only had two codons each. The high number of codons in leucine is similar to that observed in other Angiosperms, as reported previously (Park et al., 2018).

RNA editing sites

In this study, we recorded RNA editing in the cp genome of *C*. *quadrangula* for the first time. The 20 cp gene analysis using the PREP suite software (Tab. 6) revealed that 45 RNA editing sites were authentic in 14 genes, all editing with C-to-U sort. U-to-A is the largest editing setting reported in this species. Most of the editing took place in the first and second codon bases, which resulted in changes of amino acids. Changes in codons resulted in hydrophobicity and expansion in protein regions; particularly *ndh*A, *ndh*B, *ndh*D, *ndh*G, *ndh*I, and *ndh*K. For this purpose, amino acid modifications by RNA editing of genera indicate that genes carrying these editing sites can potentially use DNA sequences to determine evolutionary ties (Alzahrani et al., 2020a; Brooks et al., 2002). The *ndh*B and *ndh*D genes are the largest in number of RNA editing (7), followed by *psaI* (6), *atp*I (5), *mat*K, *atp*F,*ndh*F, and *ndh*A (3); *acc*D and *rpo*B (2); and *ndh*G, *pet*B, *rpl2*, and *rpl20* (1). Conversion from Proline to serine (P-S) in *acc*D and *atp*I genes, in which a polar amino acid group can be transformed into a polar type, in an unusual phenomenon, which has been confirmed previously (Xu et al., 2015). Some genes, however, did not have editing sites in the genome, such as *ccs*A, *ndh*C, *ndh*E, *pet*L, *pet*G, and *clp*P.

Amino acid frequency

Amino acid composition analysis is a very important tool in the study of evolution. It can be used to understand the composition of proteins in ancestral genomes, based on the conserved residues in evolutionary descendent species, along with knowledge about the relative probability of conservation of various amino acids (Brooks et al., 2002). It has also been reported that the amino acid composition of a protein reflects its historical changes, when determining its genetic code (Brooks et al., 2002). The amino acid frequencies in the genome of *C*. *quadrangular* (Fig. 2) feature a high proportion of leucine (10.66%), followed by Serine (8.3%), and arginine and phenylalanine (both 7.8%). The least present amino acids in the genome were methionine (1.78%) and tryptophan, cysteine, tyrosine, valine, alanine, and asparagine, which ranged from 2.11% to 3.78%.

Tab. 4: Genes present in the *Caralluma quadrangula* cp genome.

Category	Class of genes	Gene identity
RNA genes	ribosomal RNA genes (rRNA)	*rrn5*, *rrn4.5*, *rrn16*, *rrn23*
	Transfer RNA (tRNA)	*trnH-GUG*, *trnK-UUU*$^{+}$, *trnQ-UUG*, *trnS-GCU*, *trnV-GAC*^{a}*trnS-CGA*$^{+}$, *trnR-UCU*, *trnC-GC; trnD-GUC*, *trnY-GUA*, *trnE-UUC*, *trnT-GGU*, *trnS-UGA*, *trnfM-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}, *trnA-UGC*$^{+,a}$, *trnR-ACG*^{a}, *trnN-GUU*^{a}, *trnL-UAG*
Ribosome proteins	Small sub-unit of ribosome	*rps2*, *rps3*, *rps4*, *rps7*^{a}, *rps8*, *rps11*, *rps12*^{a}, *rps14*, *rps15*, *rps16*$^{+}$, *rps18*, *rps19*
Transcription genes	Large sub-unit of ribosome	*rpl2*$^{+,a}$, *rpl14*, *rpl16*, *rpl20*, *rpl22*, *rpl23*^{a}, *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*, *psbTpsbZ*
	Sub-unit of cytochrome	*petA*, *petB*, *petD*, *petG*, *petL*, *petN*
	Sub-unit of synthase	*atpA*, *atpB*, *atpE*, *atpF*$^{+}$, *atpH*, *atpI*
	Large sub-unit of rubisco	*rbcL*
	NADH dehydrogenase	*ndhA*$^{+}$, *ndhB*$^{+a}$, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH; ndhI; ndhJ; ndhK*
Other genes	ATP-dependent protease subunit P	*clpP*$^{++}$
	Cp envelope membrane protein	*cemA*
	Maturase	*matK*
	Sub-unit acetyl-CoA carboxylase	*accD*
	C-type cytochrome synthesis	*ccsA*
	Hypothetical proteins	*ycf2*^{a}, *ycf4*
	Component of TIC complex	*ycf1*^{a}

Tab. 5: Codon-anti codon recognition and codon usage in the cp genome of *Caralluma quadrangula*.

Codon	Amino Acid	Count	RSCU	tRNA	Codon	Amino Acid	Count	RSCU	tRNA
UUU	Phe	2318	1.18	*trnF-GAA*	UAU	Tyr	1443	1.35	*trnY-GUA*
UUC	Phe	1606	0.82		UAC	Tyr	688	0.65	
UUA	Leu	1141	1.3	*trnL-UAA*	UAA	Stop	1290	1.22	
UUG	Leu	1129	1.29	*trnL-CAA*	UAG	Stop	876	0.83	
CUU	Leu	1067	1.22	*trnL-UAG*	CAU	His	878	1.39	*trnH-GUG*
CUC	Leu	675	0.77		CAC	His	386	0.61	
CUA	Leu	800	0.91		CAA	Gln	1077	1.4	*trnQ-UUG*
CUG	Leu	453	0.52		CAG	Gln	459	0.6	
AUU	Ile	1838	1.23	*trnI-GAU*	AAU	Asn	1805	1.4	*trnG-GUU*
AUC	Ile	1094	0.73		AAC	Asn	770	0.6	
AUA	Ile	1560	1.04	*trnI-CAU*	AAA	Lys	2302	1.37	*trnK-UUU*
AUG	Met	847	1	*trnM-CAU*	AAG	Lys	1053	0.63	
GUU	Val	853	1.39	*trnV-GAC*	GAU	Asp	1034	1.4	*trnD-GUC*
GUC	Val	453	0.74		GAC	Asp	440	0.6	
GUG	Val	735	1.2		GAA	Glu	1361	1.41	*trnE-UUC*
GUA	Val	418	0.68	*trnV-UAC*	GAG	Glu	566	0.59	
UCU	Ser	1104	1.37	*trnS-GGA*	UGU	Cys	653	1.19	*trnC-GCA*
UCC	Ser	912	1.13		UGC	Cys	445	0.81	
UCG	Ser	900	1.12		UGA	Stop	1016	0.96	
UCA	Ser	660	0.82	*trnS-UGA*	UGG	Trp	701	1	*trnW-CCA*
CCU	Pro	709	1.1	*trnP-UGG*	CGU	Arg	426	0.71	*trnR-ACG*
CCC	Pro	699	1.08		CGC	Arg	290	0.48	*trnR-UCU*
CCA	Pro	709	1.1		CGA	Arg	632	1.05	
CCG	Pro	465	0.72		CGG	Arg	451	0.75	
ACU	Thr	713	1.17		AGA	Arg	715	0.89	
ACC	Thr	635	1.04		AGG	Arg	538	0.67	
ACG	Thr	718	1.18	*trnT-GGU*	AGU	Ser	1144	1.9	*trnS-GCU*
ACA	Thr	372	0.61	*trnT-UGU*	AGC	Ser	662	1.1	
GCU	Ala	543	1.23	*trnA-UGC*	GGU	Gly	554	0.93	*trnG-GCC*
GCC	Ala	427	0.97		GGC	Gly	403	0.68	
GCA	Ala	472	1.07		GGA	Gly	795	1.34	
GCG	Ala	317	0.72		GGG	Gly	623	1.05	*trnG-UCC*

Microsatellite analysis

The results show that mono-nucleotides (165-168) had the highest number in all the four genomes. Di- and tetra-nucleotides (16 each) were the second most frequent in *H. posttsii*, while hexa-nucleotides (8) and tri-nucleotides (5) were the least present SSRs. In *Asclepias nivea*, tetra-nucleotides (23) were the second highest, followed by di-nucleotides (15), tri-nucleotides (6), and penta-nucleotides (3). In *C. gigantea*, the SSRs were in the order of di-nucleotides (16), tetra-nucleotides (15), hexa-nucleotides (8), and penta-nucleotides (3). For *C. quadrangula*, tri-nucleotides (9) were followed by di- and tetra-nucleotides (7 each) (Fig. 3, Fig. 4).

Long repeats analysis

The analysis of long repeat types in the four genomes (*C. quadrangula*, *A. niveae*, *C. gigantea*, and *C. wilfordii*) indicated that *C. quadrangula* has a high number of palindromic and complement repeats (25 each), followed by forward (3), with no reverse type identified in the sequence of *A. nivea* also had high palindromic (17), forward (13), complement (11), and reverse (8) repeats. Similarly, *C. gigantea* had high palindromic (19) repeats, followed by forward (15), reverse (4), and complement (3) repeats. In *C. wilfordii*, a high number of forward repeat (33) sequences were recorded, followed by palindromic (13), while reverse and complement repeats were not present in the sequence (Fig. 5).

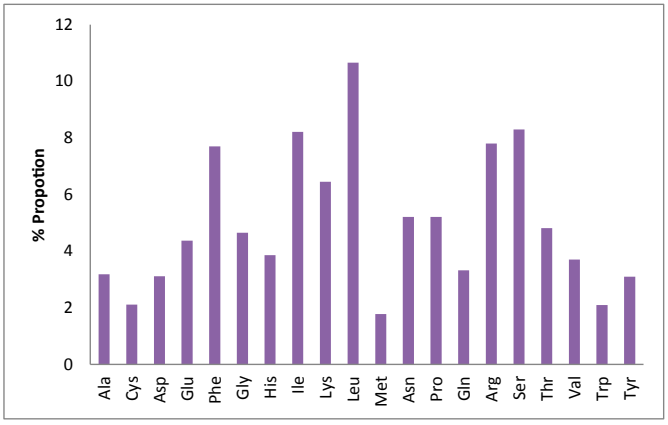


Fig. 3: Amino acids frequencies in *Caralluma quadrangula* plastome.

Comparative cp genome analysis of *C. quadrangula* with related species

The *C. quadrangula* plastome was compared with three genomes from the Asclepiadoideae subfamily (*C. wilfordii, A. nivea*, and *C. gigantea*). The smallest genome, in terms of size, was *C. wilfordii* (161,180 bp), followed by *C. quadrangula* (161,456 bp), then *H. pott-*

Tab. 6: Number of predicted RNA editing sites in the cp genome of *Caralluma quadrangula*.

Genes	Nucleotide Position	Amino Acid Position	Codon conversion	Amino Acid Conversion	Score
accD	248	83	TCA=>TTA	S=>L	1
	607	203	CCG=>TCG	P=>S	0.86
atpF	80	27	TCA=>TTA	S=>L	1
	914	305	TCA=>TTA	S=>L	1
atpI	1511	504	ACG=>ATG	T=>M	0.86
	2287	763	CGG=>TGG	R=>W	1
	2815	939	CCA=>TCA	P=>S	1
	3035	1012	CCT=>CTT	P=>L	0.86
	3734	1245	TCA=>TTA	S=>L	0.86
matK	533	178	TCC=>TTC	S=>F	0.86
	1195	399	CGG=>TGG	R=>W	1
	1520	507	CCG=>CTG	P=>L	1
ndhA	341	114	TCA=>TTA	S=>L	1
	566	189	TCA=>TTA	S=>L	1
	1079	360	TCC=>TTC	S=>F	1
ndhB	467	156	CCA=>TCA	P=>L	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	TCA=>TTA	S=>L	1
	1481	494	CCA=>CTA	P=>L	1
ndhD	2	1	ACG=>ATG	T=>M	1
	47	16	TCC=>TTC	S=>F	0.8
	599	200	TCA=>TTA	S=>L	1
	674	225	TCG=>TTG	S=>L	1
	878	293	TCA=>TTA	S=>L	1
	1298	433	TCA=>TTA	S=>L	0.8
	1310	437	TCA=>TTA	S=>L	0.8
ndhF	290	97	TCA=>TTA	S=>L	1
	365	122	CCG=>CTG	P=>L	1
	2221	750	CTT=>TTT	L=>F	1
ndhG	314	105	ACA=>ATA	T=>A	0.8
petB	611	204	CCA=>CTA	P=>L	1
PsaB	473	158	TCA=>TTA	S=>L	0.86
	551	184	TCA=>TTA	S=>L	1
	566	189	TCG=>TTG	S=>L	1
	1360	454	CTT=>TTT	L=>F	1
	2000	667	TCT=>TTT	S=>F	1
	2426	809	TCA=>TTA	S=>F	0.86

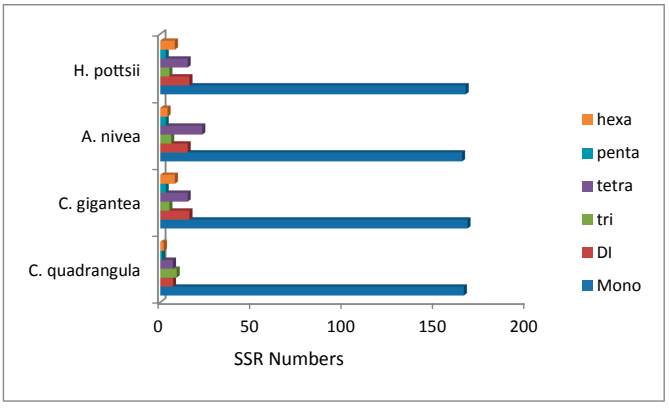


Fig. 4: SSR types in *C. quadrangula*, *H. pottsii*, *A. nivea*, and *C. gigantea* cp genomes.

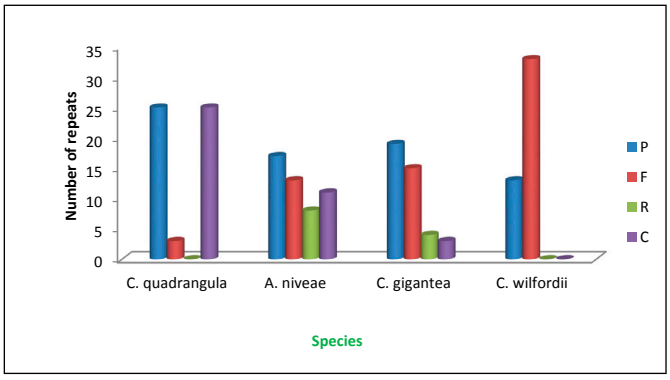


Fig. 5: Long repeats in four plastome genomes of Asclepiadoideae, where (P) indicates palindromic, (F) forward, (R) reverse, and (C) complement repeats

sii (161,565 bp), while *C. gigantea* was the largest genome (165,928 bp). The genome of *C. quadrangula* was conserved, as shown in Fig. 6. Generally, coding regions are more conserved than genome non-coding regions; at the same time, the non-coding region is generally much more diverse, compared to the coding region. The regions of *psbA* and *matK* were more diverse in *C. quadrangula* than in *H. pottsii*, *A. nivea*, and *C. gigantea*. Further, *rpoB*, *rbcL*, *rpoc1*, and *rpoc2* were highly divergent in all of the genomes. Meanwhile, *psaA*, *ndhK*, and *accD* also showed higher degrees of divergence, while *petB* was less divergent across the genomes. A highly divergent region was observed between *petB* and *petD*.

A study of the expansion of the Inverted repeats in the genomes of *C. quadrangula*, *H. pottsii*, *C. wilfordii*, and *C. gigantea* was carried out, inorder to compare the cp genome border junctions; see (Fig. 7). Variations between three genes (*trnH-Gug*, *rps19* and *ycf1*) were indicated, on the basis of their positions: *trnH-Gug* was located at the LSC-IRa border region of C. *quadrangula*, *H. pottsii*, *A. niveae*, and *C. gigantea*, varying in size (7, 63, 1, and 64 bp, respectively). Further, *rps19* was located at the LSC region in *H. pottsii*, *C. wilfordii*, and *C. gigantea;* genomes while, in *C. quadrangula*, it extended into the IRb region. This disparity was due to contraction and expansion of the genome. Meanwhile, *ycf1*was located at the extension SSC-IRa border region in *H. pottsii*, *C. wilfordii*, and *C. gigantea*, thereby creating pseudo genes between the regions; while, in *C. quadrangula*, it was located at the SSC region.

Protein-coding genes divergence (Ka/Ks)

In the study of molecular evolution, the Ka/Ks ratio is used to explain the mechanism of DNA sequence evolution, for the reconstruction of phylogenies, and for the identification of protein-coding genes. It can be used as tool for estimation of the selective pressure of gene evolution, with a Ka/Ks ratio of > 1 denoting positive selection and a Ka/Ks ratio of < 1 indicating negative selection; a value closer to 1 indicates neutral mutation (Abba et al., 2021). Ka/Ks ratios were evaluated for protein-coding genes, in order to estimate the level of divergence and selective pressure within the genes. Positive selection genes (Ka/Ks > 1) were the majority in the result (Fig. 8). Sequence divergence and purifying selection in the protein-coding genes were calculated using the synonymous (Ka) and non-synonymous (Ks) substituted rates and the Ka/Ks ratio. The results suggest low sequence divergence in several genes (i.e., Ks < 0.1). The Ka/Ks analysis demonstrated that most protein-coding genes were already under selective pressure, although a few genes (i.e., *atpI*, *ndhA*, *petB*, *psbJ*, and *rps16)* had values close to 1(neutral mutation); similar results have been reported previously (Yaradua et al., 2019).

Phylogenetic analysis

We used the complete cp genomes to reconstitute a phylogenetic tree and to establish the phylogenetic relationships, as well as tribal positions, of the fourteen Apocynaceae species. Phylogenetic analysis was performed using Bayesian inference (BI) and maximum pro-

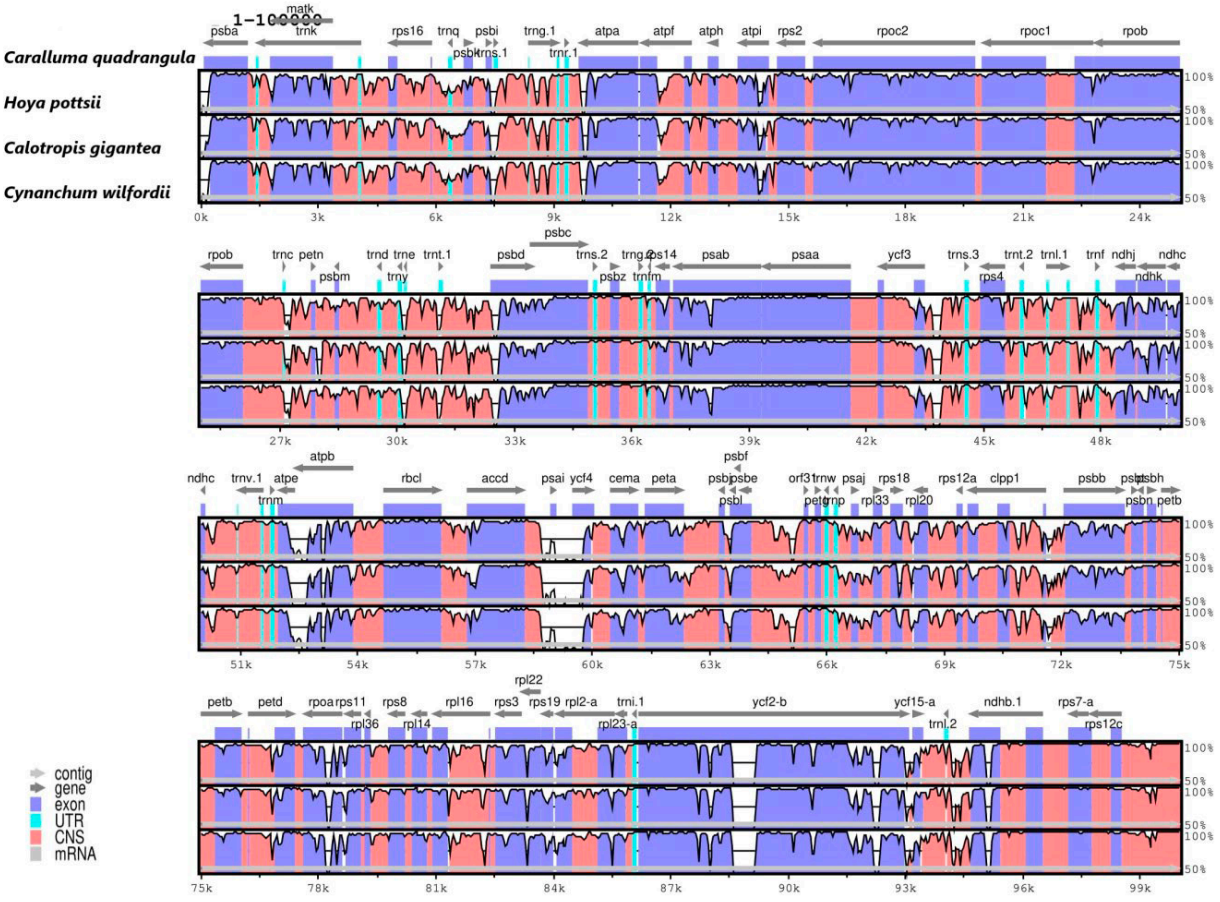


Fig: 6: Comparative analysis of cp genomes for *C. quadrangula*, *H.pottsii*, *C. wilfordii*, and *C. gigantea*, computed with VISTA program, in which the sequence of *C. quadrangula* was used as a reference. The black-grayish arrows on top of the alignment show the orientations of genes and Inverted repeat positions. Purple, pink, and gray color bars, as well as white peaks, denote exons. Similarly, untranslated Region (UTRs), mRNA, conserved non-coding Sequence (CNSs), and the genome differences are all represented in (Fig. 6); a 70% identity value was used for the plot, while a 50-100% percentage scale was used to measure the Y-axis.

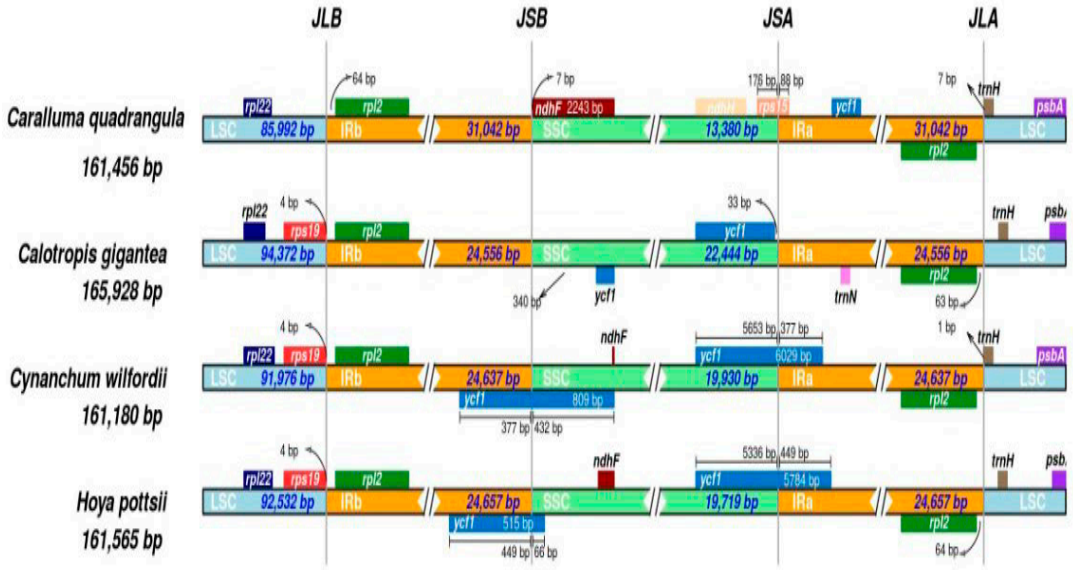


Fig. 7: Cp genome border comparison of LSC, SSC, and IR regions in the four species *C. quadrangula, C. gigantea, C. wilfordii, and H, pottsii*. Gene features are shown, as well as the distance between respective genes and the border points in *H. pottsii* and *C. wilfordii*. IRb and SSC borders extend into *ycf1* while, in *C. quadrangula*, it does not extend. The number of base pairs in the border of *trnH* varied in *C. quadrangula* (6bp) while, in the genomes of *C. gigantea*, *H. pottsii*, and *C. wilfordii*, one base pair was observed.

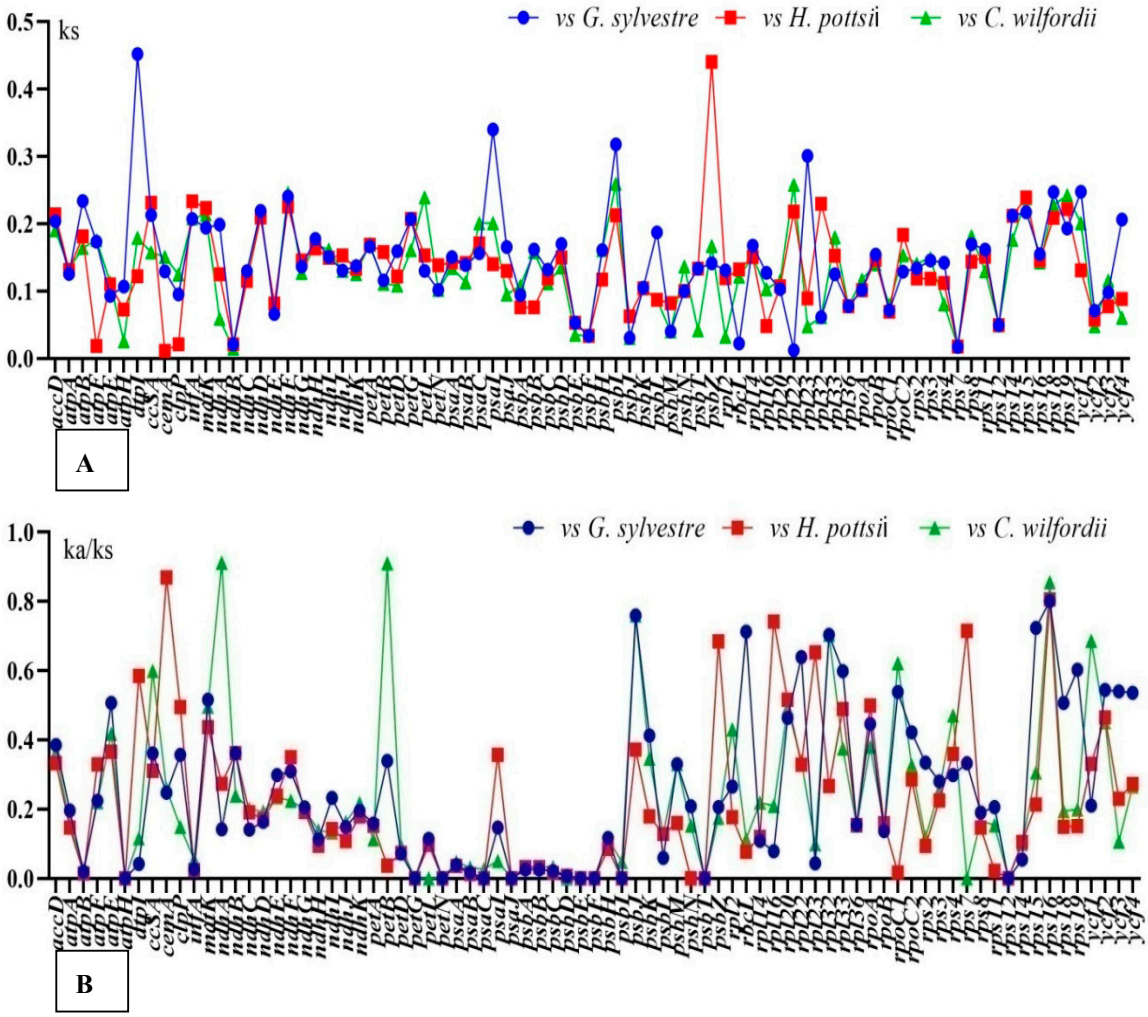


Fig. 8: Ka/Ks ratio of *C. quadrangula* against *G. sylvestre, H. pottsii*, and *C. wilfordii* genomes.

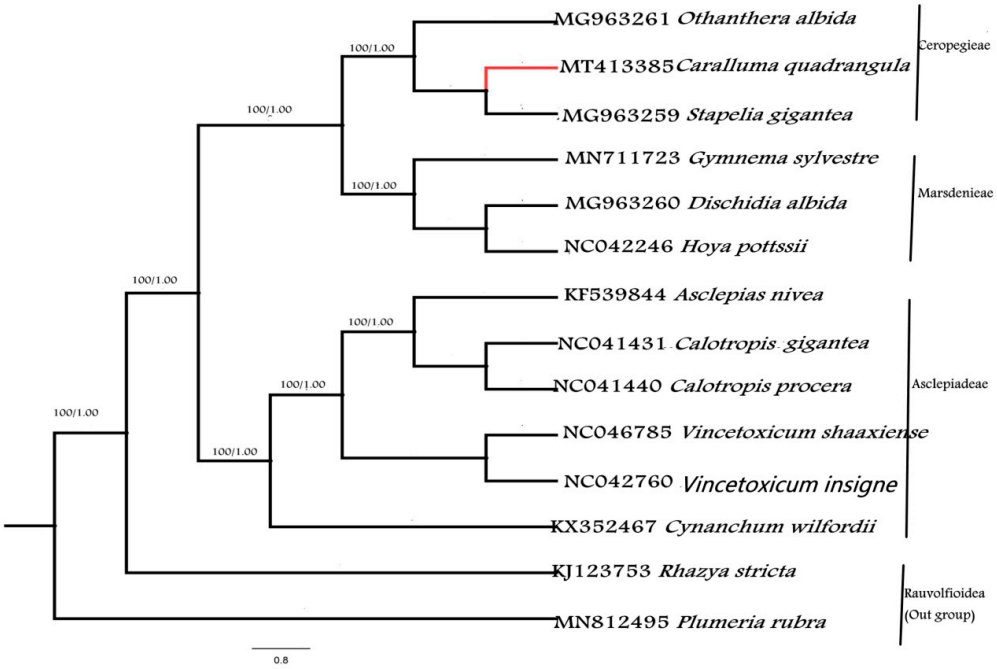


Fig. 9: Phylogenetic tree reconstruction of 14 taxa based on the complete chloroplast genome using Maximum Parsimony (MP) and Bayesian Inference (BI) methods, showing relationship within the fourteen species of Apocynaceae. The numbers in the branch nodes represent bootstrap percentage (BP)/ posterior probability (PP).

bability with an out-group of *Plumeria rubra* and *Rhazya stricta*. All 12 species clustered with strong support in one clade, split into three subclades. Subclade 1, which is monophyletic, comprises *O. albida*, *S. gigantea*, and *C. quadrangula* (Ceropegieae); this was a sister subclade to class 2, comprising *H. pottsii*, *D. albida* and *G. sylvestre* (Marsdenieae). Class 3 was the main clade, including *C. procera*, *C. gigantea*, *V. shaanxiense*, *V. insigne*, and *C. wilfordii* (Asclepiadeae) (Fig. 9).

For the selected species, we focused on the closeness of the evolutionary relationship of the species of Asclepiadoideae reflecting their tribes (Asclepiadeae, Ceropegieae, and Marsdenieae) in the subfamily.

Discussion

Cp DNA sequences have been frequently utilized in the study of genetic and evolutionary relationship analyses for plant species (Ohnishi and Matsuoka, 1996; Cuénoud et al., 2002). Total cp genome sequences provide ample information to reproduce both current and ancient diversification records (Cui et al., 2014). The solid, flexible invention of Next Generation Sequencing (NGS) has affected many areas of science, leading to the creation of a number of algorithms, which have changed research designs, making researchers able to detect the information of nearly every organism's genome, transcriptome, and epigenome (McPherson, 2013). To the best of our knowledge, we sequenced the complete cp genome of *C. quadrangula* for the first time in present study. Our findings show that the *C. quadrangula* cp genome size aligns with the cp of flowering plants, ranging from 154,841 bp to 164,213 bp in the *Rhazya stricta* and *Pergularia tomentosa* genomes respectively (Park et al., 2014; Abba et al., 2020). The GC content of *C. quadrangula, H. pottsi*, *C. wilfordii*, was 38% (Tab. 1 and 2), slightly higher than that of *C. gigantea* 37.5% (Dong et al., 2013).

The distribution and percentage rate of the AT and GC contents of the genome showed that LSC AT (62.7%) and GC (37.3%), SSC AT (66.3%) and GC (33.7%), IRa and IRb AT (51.8%) and (48.2%) for each. The overall AT content was 62.2%. Moreover, the GC content of the single-copy region was less than GC content of the IR, LSC, and SSC regions.

In the IR region, the GC content was higher (42.6%) than in the LSC and SSC regions. Due to the presence of GC-rich nucleotides in rRNA genes (e.g. *rrn5*, *rrn4.5*, *rrn23* and *rrn16*) in the IR region, the GC content was greater, consistent with data previously reported in other cp genomes (Park et al., 2014; Yang et al., 2013).

It is generally assumed that the gene(s) of the cp genome, as well as their organization, are strongly conserved in the majority of angiosperms (Jansen et al., 2008). A total of 129 genes (Fig. 2) were found in the *C. quadrangula* cp genome, where additional studies have also shown that many angiosperms have retained identical genes (Jansen et al., 2007; Park et al., 2014; Liu et al., 2018). Due to the large amount of cp genome sequences, gene alteration and gene loss in cp genomes have become more and more evident in the general organization of the cp genome; for example, an*ycf15* gene discovered in most of angiosperm plastid genomes was lost in *C. quadrangula*. Similar findings have also been observed in many cp genomes (Wu et al., 2015). In the *C. quadrangula* cp genome, here was a loss of *trnI-UGC*, as previously reported by different researchers, in line with other angiosperms; particularly, in most Rosids (Goulding et al., 1996; Leclerq et al., 2007). Similarly, 12 protein coding genes and 6 tRNA genes with introns were found in the *C. quadrangula* cp genome (Tab. 3). Among the genes with introns, only *atpF*, *ycf3*, and *accD* had two introns, while the rest had only one intron. Similar findings have been documented in both the *Pergularia tomentosa* cp genome (Abba et al., 2020) and the cp genome of *Razhya stricta* (Park et al., 2014).

Genes such as *clpP*, *rps15*, *rpl22*, and *ycf1* have been reported to have high rates of evolution in cp genomes, according to (Cuénoud et al., 2002). Similar results have recently been published among 17 plant species and species of *Panax* (Kim et al., 2016).

The genes with the longest introns were *trnKUU* (2480), *accD* (1194), and *ndhA* (1104). The Large single-copy (LSC) region had the highest number of genes with introns, 10 genes (*rps16*, *atpF*, *rpoC1*, *ycf3*,

clpP, *trnk-UUU*, *trnG-UCC*, *trnL-UAA*, *trnV-UAC*, and *accD)*, followed by the inverted repeat (IR) region, with 6 genes (*rpl2*, *ndhB*, *trnM-GAU*, *trnA-UGC*, *rpoC2*, and *rps12*), then the small single-copy (SSC) region, with 2 genes (*ndhA* and *ycf1*) (Tab. 2). Meanwhile, *atpF*, *accD*, and *ycf3* are the only genes with two introns. The largest intron was in *trnKUU* (2480), as it contains the *matK* gene, as previously reported (Saina et al., 2016; Kim et al., 2016; Shen et al., 2017).The *rps12* gene is a 5' end trans-spliced gene in the LSC region with a duplicated 3' end in the IR areas; *ycf3* is very critical for stable photo-system I accumulation (Boudreu et al., 2016; Naver et al., 2001); and *clpP* also forms part of the protease synthesis gene (Ramundo et al., 2014). Introns are useful instruments, which aid in the development of new plant varieties with high resistance to diseases and insect pests.

Codons were represented by 27,342 nucleotides, with the highest codon rate identified for Leucine 5265 (19.2%), while the lowest was for Cysteine 1098 (4.02%). Similar findings have been reported by Raman et al. (2017). The number and location of genes in the cp genome of *C*. *quadrangula* were same as previously published cp genomes of Angiosperms (Alzahraini et al., 2020a; Qian et al., 2013). The IR region comprised six tRNAs, four rRNAs, and seven protein-coding genes. The LSC region enclosed 63 protein genes with 23 tRNA genes, whereas the remaining 13 protein genes and the tRNA pair were located in the SSC region. Most protein-coding genes began with ATG (for Methionine), and some with GTG, ATC, and ACG; this is commonly found throughout Angiosperms (Raman and Park, 2017; Kuang et al., 2011).

In this research, for the first time, we described RNA editing in the cp of *C*. *quadrangula*. Analysis of 20 cp genes using the PREP suite software (Tab. 6) revealed 45 RNA editing sites in 14 genes, with the editing all of C-to-U type. U-to-A is the highest editing context observed in this species. Most of the editing happened at the first and second codon bases, resulting in changes in amino acids. Codon modifications resulted in increased hydrophobicity and the expansion of regions in a number of proteins, particularly *ndhA*, *ndhB*, *ndhD*, *ndhG*, *NdhI*, and *ndhK*. For this reason, changes in amino acids from closely related organisms through RNA edits indicate the potential use of DNA sequences of genes carrying these editing sites for testing evolutionary relationships (Park et al., 2018; Xu et al., 2015; Alzahrani et al., 2020a). The *ndhB* and *ndhD* genes hadthe largest number of RNA edits (7 each), followed by *psaI* (6); *atpI* (5); *matK*, *atpF*, *ndhF*, and *ndhA*(3 each); *accD*, and *rpoB* (2 each); and *ndhG*, *petB*, *rpl2*, and *rpl20*(1 each). Conversion of proline to serine (P-S) was present in the *accD*, and *atpI* genes, where the apolar amino acid group could potentially transition to polar type; this is a very unusual occurrence, which has previously been recorded (Xu et al., 2015). However, certain genes did not have RNA editing sites in the genome, such as *ccsA*, *ndhC*, *ndhE*, *petL*, *petG*, and *clpP*. The amino acid composition of proteins reflects its historical changes in deciding its genetic code (Brooks et al., 2002). Amino acid frequencies in the *C*. *quadrangula* genome (Fig. 3) displayed a high amount of leucine (10.66%), accompanied by serine (8.3%) and arginine and phenylalanine (both 7.8%). The least present amino acids in the genome were methionine (1.78%); as well as tryptophan, cysteine, tyrosine, valine, alanine, and asparagine (2.11-3.78%).

Simple sequence repeats (SSRs) have a higher mutation rate than the other neutral DNA regions, due to scraped strand mispairing. Because of the haploid and non-parental ancestry features of cpSSRs, the population structure is commonly determined in genetic studies by molecular markers (Asaf et al., 2017; Khan et al., 2019). Herein, we examined proper SSRs amongst four species: *C*. *quadrangula*, *A*. *nivea*, *H*. *pottsii*, and *C*. *gigantea* (Fig. 4). The largest proportion of SSRs was discovered in *H*. *pottsii*, accompanied by *C*. *quadrangula*. Mononucleotide SSRs were the most frequent SSRs in all four species; where mononucleotide repeats of A or T were the most abundant SSR in *C*. *quadrangular* (Fig. 4), consistent with earlier observations that cpSSRs generally comprise A or T repeats and infrequently repeats G or C (Jansen et al., 2011). In our sequenced cp genome, different recurrences (Palindromic, Forward, Reverse, and Complement) showed heterogeneity in the number of repeats; similar to the species previously studied (Brooks et al., 2002). Palindromic repeats, in all forms, were more numerous than forward, complement, and reverse repeats, as observed in the four cp genomes (Fig. 5), consistent with previous studies of *Barleria prionitis* (Alzahrani et al., 2020b) and *Pergularia tomentosa* (Abba et al., 2020).

The analysis of long repeat types in the genomes of *C*. *quadrangula*, *A*. *niveae*, *C*. *gigantea* and *C*. *wilfordii*, as shown in (Fig. 5), indicated that *C*. *quadrangula* had more palindromic repeats (25 each), followed by forward (3), with no reverse form found in the sequence. *A*. *nivea* also had high palindromic (17) repeats, as well as forward (13), complement (11) and reverse (8) repeats. Likewise, *C*. *gigantea* also had high palindromic (19) repeats, followed by forward (15), reverse (4), and complement (3) repeats. In *C*. *wilfordii*, a high number of forward repeats (33) were observed, followed by palindromic (13) repeats, while reverse and complement repeats were not present in the sequence.

The cp genome is generally very well preserved in terrestrial plants and, so, the large inverted repeat (IR) junctions are not necessary for the cp genome function (Jansen et al., 2011). It is assumed that the IRs are the most conserved regions, as the rate of natural nucleotide substitution in the IRs is lower than in the Single-copy regions. The difference between the IR/LSC and IR/SSC borders is the main reason for the variations in size between cp genomes of divergent groups. The small expansion of the IRb (JLB) border in *C*. *quadrangula* suggested a difference in size between the four genomes, compared to *C*. *wilfordii* (Fig. 7). These findings are consistent with previous work, in which IRs have been used as successful tools for assessing conformational reorganizations within the plastid genomes and are frequently subject to expansion, contraction, or even complete loss (Yaradua et al., 2019).

The sizes of the inverted repeats were 31,042 bp for *C*. *quadrangula*, 24,556 bp for *C*. *gigantea*, 24,637 bp for *C*. *wilfordii*, and 24,657 bp for *H*. *pottsii*. The IR borders of the four sequences were compared, in which three types of junctions were recognized, based on the*rps19*, *trn*H, and *ycf1-ndhF* positioning. In the first border, there was a similar orientation of the SSC, LSC, IRa, and IRb in*C*. *wilfordii* and *H*. *pottsii*. In the *C*. *quadrangula* sequence, there was a clear variation in the orientations of the junctions. In the second position, a uniform Border junction was observed in the four sequences while, at the third position, *ycf1* in *C*. *quadrangula* and *C*. *gigantean* was unique in its position due to its appearance at the forward strand while, in *C*. *wilfordii* and *H*. *pottsii*, it appeared in both forward and reverse strands.

Cp genomes have been reported to be much conserved, although variation between species has been reported (Abba et al., 2021). Previous systemic phylogenies of Angiosperms, based on the ITS region (nuclear) and plastid fragment sequences, have shown significant variation from the phylogenetic tree (Kim et al., 2009).

To assess the phylogenetic relationship and tribal status of the 14 Apocynaceae species, we used the full cp genomes of the eight species to reconstruct a phylogenetic tree. The phylogenetic analyses were performed using Bayesian inference (BI) and Maximum likelihood, with *Plumeria rubra* and *Rhazya stricta* as outgroup. The resulting tree was congruent, with strong support (PP = 1.0 and MP = 100) for all relationships and in agreement with previously published phylogenies (e.g., Fishbein et al., 2018; Surveswaran et al., 2014). All 12 species were clustered into one clade with strong support, and further divided into three major sub-clades. Sub-clade 1, which was monophyletic, included *O*. *albida*, *S*. *gigantea*, and *C*. *quadrangular* (Ceropegieae), and was a sister to clade 2, containing *H*. *pottsii*, *D*.

albida, and *G. sylvestre* (Marsdenieae). Clade 3 was the largest clade, including *C. procera, C. gigantea, V. shaanxiense*, and *V. insigne* (Asclepiadeae). The species were selected to comprise members of the three main tribes (Asclepiadeae, Ceropegieae, and Marsdenieae) of the subfamily Asclepiadoideae of Apocynaceae.

Conclusion

The results presented in this research elucidate the structure of the chloroplast genome and its variability in Asclepiadoideae and may help to identify molecular markers for the identification and authentication in the subfamily. The availability of this comparative cp genome sequence may serve as a reference for the advancement of genomics studies in *C. quadrangula*, thus helping researchers to investigate genetic issues specific to this species.

Author contributions: A.A. proposed the research, performed the experiments, and drafted the manuscript; S.S.Y., D.A.A., and A.A. collected data, while A.A. and E.J.A. analyzed the data; D.A.A. supervised the project. All the authors edited the manuscript.

Acknowledgements
The project was financially supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, with grant No. D1441-260-130. Therefore, the authors appreciate the gesture.

Conflicts of interest
No potential conflict of interest was reported by the author .

Data availability
Sequences were deposited to the NCBI Genbank database https://www.ncbi.nlm.nih.gov with the accession number MT413385.

References

Abba, A., Alzahrani, D., Yaradua, S., Albokhari, E., 2021: Comparative chloroplast genome analysis and evolutionary relationships in some species of Asclepiadeae, Apocynaceae. Leg. Res. 44(2). DOI: 10.18805/LR-565

Abba, A., Alzahrani, D., Yaradua, S., Albokhari, E.B., 2020: Complete plastome genome of *Pergularia tomentosa* L. (Asclepiadoideae, Apocynaceae). Mitochondrial DNA B Res. 5(1), 566-567. DOI: 10.1080/23802359.2019.1710291

Akashi, H., 2011: Gene expression and molecular evolution. Curr. Opin. Genet. Dev. 11, 660-666. DOI: 10.1016/s0959-437x(00)00250-1

Alzahrani, D., Albokhari, E., Yaradua, S., Abba, A., 2020: Complete plastome genome of *Dipterygium glaucum*, Dipterygieae, Cleomaceae. Mitochondrial DNA Part B, 5, 1872-1873. DOI: 10.1080/23802359.2020.1750978

Alzahrani, D.A., Yaradua, S.S., Albokhari, E.J., Abba, A., 2020b: Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. BMC Genom. 21, 393. DOI: 10.1186/s12864-020-06798-2

Aruna, V., Kiranmai, C., Karripusamy, S., 2009: Micropropagation of three varieties of *Caralluma adscendens* via nodal explants. J. Plant Biochem. Biotech. 18, 121-123. DOI: 10.1007/BF03263309

Asaf, S., Khan, A.L., Khan, M.A., Waqas, M., Kang, S.M., Yun, B.W., Lee, I.J., 2017: Chloroplast genomes of *Arabidopsis halleri* ssp. gemmifera and *Arabidopsis lyrata* ssp. *petraea*: Structures and comparative analysis. Sci. Rep. 7, 7556. DOI: 10.1038/s41598-017-07891-5

Bensuzan, K., 2009: Taxonomy and conservation status of Moroccan stapeliads (Apocynaceae-Asclepiadoideae-Ceropegieae-Stapeliinae). Bull. Inst. Sci. 31, 67-77.

Boudreau, E., Takahashi, Y., Lemieux, C., Turmel, M., Rochaix, J.D., 1997: The chloroplast ycf3 and ycf4 open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. The EMBO J. 16(20), 6095-6104. DOI: 10.1093/emboj/16.20.6095

Brooks, D.J., Fresco, J.R., Lesk, A.M., Singh, M., 2002: Evolution of amino acid frequencies in proteins over deep time: inferred order of introduction of amino acids into the genetic code. Mol. Bio. Evol. 19, 1645-1655. DOI: 10.1093/oxfordjournals.molbev.a003988

Cuénoud, P., Savolainen, V., Chatrou, L.W., Powell, M., Grayer, R.J., Chase, M.W., 2002: Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences. Ame. J. Bot. 89, 132-144. DOI: 10.3732/ajb.89.1.132

Cui, Y., Qin, S., Jiang, P., 2014: Chloroplast transformation of *Platymonas* (Tetraselmis) *subcordiformis* with the bar gene as selectable marker. PLoS ONE 9, e98607. DOI: 10.1371/journal.pone.0098607

Daniell, H., Lin, C., Yu, M., Chang, W., 2016: Chloroplast genomes: diversity, evolution, and applications in genetic engineering, Genome Biol. 17, 134. DOI: 10.1186/s13059-016-1004-2

Dierckxsens, N., Mardulyn, P., Smits, G., 2017: NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucl. Acids Res. 45, e18-e18. DOI: 10.1093/nar/gkw955

Dong, W., Xu, C., Cheng, T., Zhou, S., 2013: Complete chloroplast genome of *Sedum sarmentosum* and chloroplast genome evolution in Saxifragales. PloS one 8, e77965. DOI: 10.1371/journal.pone.0077965

Ebert, D., Peakall, R., 2009: Cp simple sequence repeats (cpSSRs): technical resources and Recommendations for expanding cpSSR discovery and applications to a wide array of plant species. Mol. Ecol. Res. 9, 673-690. DOI: 10.1111/j.1755-0998.2008.02319.x

Echt, C.S., DeVerno, L.L., Anzidei, M., Vendramin, G.G., 1998: Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. Mol. Eco. 7, 307-316. DOI: 10.1046/j.1365-294X.1998.00350.x

Endress M.E., Meve, U., Middleton, D.J., Liede-Schumann, S., 2018: *Apocynaceae*. In: Flowering Plants. Eudicots, 207-411. Springer, Cham.

Evans, W., Evans, E., Trease, G., 2012: *Trease and Evans Pharmacognosy*, 15th edn. London.

Felsenstein, J., 1978: Cases in which parsimony or compatibility methods will be positively misleading, Syste. Bio. 27, 401-410. DOI: 10.2307/2412923

Ferrarini, M., Moretto, M., Ward, J., 2013: An evaluation of the PacBio R S platform for sequencing and de novoAssembly of a chloroplast genome. BMC Genom. 14, 670. DOI: 10.1186/1471-2164-14-670

Frazer, K., Pachter, L., Poliakov, A., Rubin, E., Dubchak, I., 2004: VISTA: Computational tools for comparative genomics, Nucl. Acids Res. 32, 273-279. DOI: 10.1093/nar/gkh458

Google Maps, 2020: Google Earth, Al-shafa mountains, Taif Saudi Arabia. Google Maps [online] Available through: [Accessed 21 August 2020].

Goulding, S.E., Wolfe, K.H., Olmstead, R.G., Morden, C.W., 1996: Ebb and flow of the chloroplast inverted repeat. Mol. Genet. 252, 195-206. DOI: 10.1007/BF02173220

Jansen, R., Raubeson, L., Boore, J., 2002: Methods for obtaining and analyzing whole chloroplast Genome Sequences, Methods in Enzymology. Mol. Evol. 395, 348-384. DOI: 10.1016/S0076-6879(05)95020-9

Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Depamphilis, C.W., Leebens-Mack, J., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., 2007: Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc. Natl. Acad. Sci USA 104, 19369-19374. DOI: 10.1073/pnas.0709121104

Jansen, R.K., Saski, C., Lee, S.B., Hansen, A.K., Daniell, H., 2011: Complete plastid genome sequences of three rosids (*Castanea, Prunus, Theobroma*): evidence for at least two independent transfers of rpl22 to the nucleus. Mol. Bio. Evol. 28, 835-847. DOI: 10.1093/molbev/msq261

Jansen, R.K., Wojciechowski, M.F., Sanniyasi, E., Lee, S.B., 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. Phyl. Evo. 48, 1204-1217. DOI: 10.1016/j.ympev.2008.06.013

Jin, D., Choi, I., Choi, B., 2019: Plastid genome evolution in tribe Desmodieae (Fabaceae: Papilionoideae). PLoS ONE 14, e0218743. DOI: 10.1371/journal.pone.0218743

Katoh, K., Standley, D., 2003: MAFFT multiple sequence alignment software Version7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772-780. DOI: 10.1093/molbev/mst010

Khan, A., Asaf, S., Khan, A.L., Khan, A., Al-Harrasi, A., Al-Sudairy, O., AbdulKareem, N.M., Al-Saady, N., Al-Rawahi, A., 2019: Complete chloroplast genomes of medicinally important *Teucrium* species and comparative analyses with related species from Lamiaceae. PeerJ, 7, e7260. DOI: 10.7717/peerj.7260

Kim, K., Lee, H., 2016: Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng Nees*) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res. 11, 247-261. DOI: 10.1093/dnares/11.4.247

Kim, Y.K., Park, C.W., Kim, K.J., 2009: Complete chloroplast DNA sequence from a Korean endemic genus, *Megaleranthis saniculifolia*, and its evolutionary implications. Mol. Cells, 27, 365. DOI: 10.1007/s10059-009-0047-6

Kuang, D.Y., Wu, H., Wang, Y.L., Gao, L.M., Zhang, S.Z., Lu, L., 2011: Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. Genome, 54, 663-673. DOI: 10.1139/g11-026

Kurtz, S., Choudhuri, J., Ohlebusch, E., 2001: REPuter: The Manifold applications of repeat analysis on a genomic scale. Nucl. Acids Res. 29, 4633-4642. DOI: 10.1093/nar/29.22.4633

Leclercq, S., Rivals, E., Jarne, P., 2007: Detecting microsatellites within genomes: significant variation among algorithms. BMC Bioinform. 8, 125. DOI: 10.1186/1471-2105-8-125

Leo, M., DeTommasi, N., Sanogo, R., 2005: New pregnane glycosides from *Caralluma dalzielii*. Steroids 70, 573-585. DOI: 10.1016/j.steroids.2005.03.013

Li, B., Lin, F., Huang, P., 2017: Complete chloroplast genome sequence of *Decaisnea insignis*: genome organization, genomic resources and comparative analysis. Sci. Rep. 7, 10073. DOI: 10.1038/s41598-017-10409-8

Li, Y., Ren, B., Ding, L., Shen, Q., Peng, S., Guo, S., 2013: Does chloroplast size influence photosynthetic nitrogen use efficiency? PloS one 8, e62036. DOI: 10.1371/journal.pone.0062036

Librado, P., Rozas, J., 2009: DnaSPv5: software for comprehensive analysis of DNA polymorphism. Bioinformatics 25, 1451-1452. DOI: 10.1093/bioinformatics/btp187

Liu, L., Wang, Y., He, P., Li, P., Lee, J., Soltis, D.E., Fu, C., 2018: Chloroplast genome analyses and genomic resource development for epilithic sister genera *Oresitrophe* and *Mukdenia* (Saxifragaceae), using genome skimming data. BMC Genom. 19, 235. DOI: 10.1186/s12864-018-4633-x

Lohse, M., Drechsel, O., Bock, R., 2007: OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Current Genet. 52, 267-274. DOI: 10.1007/s00294-007-0161-y

Mayor, C., Brudno, M., Schwartz, J., 2000: VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics 16, 1046-1047. DOI: 10.1093/bioinformatics/16.11.1046

McCauley, D., 1995: The use of chloroplast DNA polymorphism in studies of gene flow in plants. Trends Ecol. Evol. 10, 198-202. DOI: 10.1016/S0169-5347(00)89052-7

McPherson, H., Van der Merwe, M., Delaney, S.K., Edwards, M.A., Henry, R.J., McIntosh, E., Rymer, P.D., Milner, M.L., Siow, J., Rossetto, M., 2013: Capturing chloroplast variation for molecular ecology studies: a simple next generation sequencing approach applied to a rainforest tree. BMC Ecol. 13, 8. DOI: 10.1186/1472-6785-13-8

Meve, U., Liede, S., 2004: Subtribal división of Ceropegieae (Apocynaceae-Asclepiadoideae). Taxon 53, 61-72. DOI: 10.2307/4135489

Meve, U., Liede-Schumann, S., 2007: Ceropegia (Apocynaceae, Ceropegieae, and Stapeliinae): Paraphyletic but still taxonomically sound. Ann. Missouri Bot. Gard. 94, 392-406.

Mower, J.P., 2009: The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucl. Acids. Res. 37, 253-259. DOI: 10.1093/nar/gkp337

Naik, R., Krishnamurthy, Y., 2012: Xerophyte *Caralluma stalagmifera* var. Longipetala (Asclepiadaceae): A new record to the flora of Karnataka, India. J. Threat. Taxa. 4, 2656-2659. DOI: 10.11609/JoTT.o2898.2656-9

Ohnishi, O., Matsuoka, Y., 1996: Search for the wild ancestor of buckwheat II. Taxonomy of *Fagopyrum* (Polygonaceae) species based on morphology, isozymes and cpDNA variability. Genes Genet. Syst. 71, 383-390.

Pal, C., Papp, B., Hurst, L., 2006: An integrated view of protein evolution. Rev. Genet. 7, 337-348. DOI: 10.1038/nrg1838

Park, M., Park, H., Lee, H., Lee, B.H., Lee, J., 2018: The complete plastome sequence of an Antarctic bryophyte *Sanionia uncinata* (Hedw.) Loeske. Int. J. Mol. Sci. 19, 709. DOI: 10.3390/ijms19030709

Park, S., Ruhlman, T.A., Sabir, J.S., Mutwakil, M.H., Baeshen, M.N., Sabir, M.J., Baeshen, N.A., Jansen, R.K., 2014: Complete sequences of organelle genomes from the medicinal plant *Rhazya stricta* (Apocynaceae) and contrasting patterns of mitochondrial genome evolution across Asterids. BMC Genom. 15, 405. DOI: 10.1186/1471-2164-15-405

Parks, M., Cronn, R., Liston, A., 2009: Increasing phylogenetic resolution at low taxonomic levels using Massively parallel sequencing of chloroplast genomes. BMC Biol. 7, 84. DOI: 10.1186/1741-7007-7-84

Qian, J., Song, J., Gao, H., Zhu, Y., Xu, J., Pang, X., Yao, H., Sun, C., Li, C., Liu, J., Xu, H., 2013: The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. PloS one 8, e57607. DOI: 10.1371/journal.pone.0057607

Raman, G., Park, S., 2016: The complete chloroplast genome sequence of *Ampelopsis*: gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. Front. Plant Sci. 7, 341. DOI: 10.3389/fpls.2016.00341

Ramundo, S., Rochaix, J., 2014: Chloroplast unfolded protein response, a new plastid stress signalingpathway. Plant Sign. Behav. 9, 972874. DOI: 10.4161/15592316

Rapini, A., 2002: Taxonomy "under construction": advances in the systematic of Apocynaceae, with emphasis on the Brazilian Asclepiadoideae. Rodriguésia 63, 075-088. DOI: 10.1590/S2175-78602012000100007

Rauf, A., Rehman, W., Muhammad, N., 2013: Phytochemical, phytotoxic and Antioxidant Profile of *Caralluma tuberculata* N.E. Brown. Wudpecker. J. Pharm. Pharmacol. 2, 21-25.

Ronquist, F., Teslenko, R., vander Mark, P., 2012: MrBayes 3.2: efficient Bayesian Phylogenetic inference and model choice across a large model space. Sys. Bio. 61, 539-542.

Saina, J.K., Li, Z.Z., Gichira, A.W., Liao, Y.Y., 2018: The complete chloroplast genome sequence of tree of heaven (*Ailanthus altissima* (Mill.) (Sapindales: Simaroubaceae), an important pantropical tree. Int. J. Mol. Sci. 19, 929. DOI: 10.3390/ijms19040929

Schmieder, R., Edwards, R., 2011: Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863-864. DOI: 10.1093/bioinformatics/btr026

Shen, X., Wu, M., Liao, B., 2017: Complete chloroplast genome sequence and phylogenetic analysis of the medicinal plant *Artemisia annua*. Molecules 22, 1330-1343. DOI: 10.3390/molecules22081330

Sloan, D., Triant, D., Forrester, N., 2014: A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe *Sileneae* (Caryophyllaceae). Mol. Phy. Evo. 72, 82-89. DOI: 10.1016/j.ympev.2013.12.004

Stevens, P., 2001: Angiosperm Phylogeny Website. Version 14, July 2017 [and more or less continuously updated since]. http://www.mobot.org/MOBOT/research/APweb/

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013: MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol.

Biol. Evol. 12, 2725-2729. DOI: 10.1093/molbev/mst197

Thiel, T., Michalek, W., Varshney, R., 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. DOI: 10.1007/s00122-002-1031-0

Waheed, A., Barker, J., Barton, S., 2011: Novel acylated steroidal glycosides from *Caralluma tuberculata* induce caspase dependent apoptosis in cancer cells. J. Ethno. Pharm. 137, 1189-1196. DOI: 10.1016/j.jep.2011.07.049

Wang, Z., San Lucas, F.A., Qiu, P., Liu, Y., 2014: Improving the sensitivity of sample clustering by leveraging gene co-expression networks in variable selection. BMC Bioinform. 15, 153. DOI: 10.1186/1471-2105-15-153

Wicke, S., Schneeweiss, G., Müller, K., Quandt, D., 2011: The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. Plant Mol. Biol. 76, 273-297.

Wu, C., Chaw, S., 2015: Evolutionary stasis in cycad plastomes and the first case of plastome GC-biased gene conversion. Gen. Biol. Evol. 7, 2000-2009.

Wyman, S.K., Jansen, R.K., Boore, J.L., 2004: Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20, 3252-3255. DOI: 10.1093/bioinformatics/bth352

Xu, C., Härtel, H., Wada, H., Hagio, M., Yu, B., Eakin, C., Benning, C., 2002: The pgp1 mutant locus of *Arabidopsis* encodes a phosphatidyl glycerol phosphate synthase with impaired activity. Plant Physiol. 129, 594-604. DOI: 10.1104/pp.002725

Xu, J.H., Liu, Q., Hu, W., Wang, T., Xue, Q., Messing, J., 2015: Dynamics of chloroplast genomes in green plants. Genomics, 106, 221-231. DOI: 10.1016/j.ygeno.2015.07.004

Yang, J., Yue, M., Niu, C., 2017: Comparative analysis of the complete chloroplast genome of four endangered herbals of *Notopterygium*. Genes 8, 124. DOI: 10.3390/genes8040124

Yang, M., Zhang, X., Liu, G., 2010: The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). PLoS ONE 5, e12762. DOI: 10.1371/journal.pone.0012762

Yaradua, S.S., Alzahrani, D.A., Albokhary, E.J., Abba, A., Bello, A., 2019: Complete chloroplast genome sequence of *Justicia flava*: Genome comparative analysis and phylogenetic relationships among Acanthaceae. BioMed Res. Int. 4, 4370258. DOI: 10.1155/2019/4370258

ORCID
Abidina Abba https://orcid.org/0000-0002-1886-2034
Dhafer A. Alzahrani https://orcid.org/0000-0002-4978-2731
Samaila S. Yaradua https://orcid.org/0000-0001-6164-4537
Enas J. Albokhari https://orcid.org/0000-0001-7124-9864

Address of the corresponding author:
Abidina Abba: Department of Biological Sciences, Faculty of Sciences, King Abdulaziz University, P.O. Box 80203, Jeddah 21589, Saudi Arabia
E-mail: abdin2007@gmail.com