

RESEARCH

Open Access



Complete chloroplast genome of the desert date (*Balanites aegyptiaca* (L.) Del. comparative analysis, and phylogenetic relationships among the members of Zygophyllaceae

Widad S. AL-Juhani^{1*}, Samah A. Alharbi¹, Nora M. Al Aboud¹ and Ashwaq Y. Aljohani²

Abstract

Background: *Balanites aegyptiaca* (L.) Delile, commonly known as desert date, is a thorny evergreen tree belonging to the family Zygophyllaceae and subfamily Tribuloideae that is widespread in arid and semiarid regions. This plant is an important source of food and medicines and plays an important role in conservation strategies for restoring degraded desert ecosystems.

Results: In the present study, we sequenced the complete plastome of *B. aegyptiaca*. The chloroplast genome was 155,800 bp, with a typical four-region structure: a large single copy (LSC) region of 86,562 bp, a small single copy (SSC) region of 18,102 bp, and inverted repeat regions (IRa and IRb) of 25,568 bp each. The GC content was 35.5%. The chloroplast genome of *B. aegyptiaca* contains 107 genes, 75 of which coding proteins, 28 coding tRNA, and 4 coding rRNA. We did not observe a large loss in plastid genes or a reduction in the genome size in *B. aegyptiaca*, as found previously in some species belonging to the family Zygophyllaceae. However, we noticed a divergence in the location of certain genes at the IR-LSC and IR-SSC boundaries and loss of *ndh* genes relative to other species. Furthermore, the phylogenetic tree constructed from the complete chloroplast genome data broadly supported the taxonomic classification of *B. aegyptiaca* as belonging to the Zygophyllaceae family. The plastome of *B. aegyptiaca* was found to be rich in single sequence repeats (SSRs), with a total of 240 SSRs.

Conclusions: The genomic data available from this study could be useful for developing molecular markers to evaluate population structure, investigate genetic variation, and improve production programs for *B. aegyptiaca*. Furthermore, the current data will support future investigation of the evolution of the family Zygophyllaceae.

Keywords: Plastome, *Balanites aegyptiaca*, Zygophyllaceae, Phylogenetic relationship, Genome structure, Comparative analysis

Introduction

The Zygophyllaceae family includes approximately 22 genera and 230–240 species [1, 2]. *Balanites aegyptiaca* (L.) Delile, desert date or heglig, belonging to the Zygophyllaceae family, is native to dry and semiarid regions in Asia and Africa. It is a thorny evergreen shrub or small tree with multiple branches, grey bark [3, 4], compound leaves, hermaphrodite flowers that

*Correspondence: wsjuhani@uqu.edu.sa

¹ Department of Biology, Faculty of Applied Science, Umm Al-Qura University, Makkah 24381, Saudi Arabia

Full list of author information is available at the end of the article



are usually greenish-yellow, and brown fruit with a hardstone seed [4].

B. aegyptiaca has been the subject of various classic studies because of its importance as a source of food and medicines. In traditional medicine, the fruit of this plant is used to treat diabetes, asthma, epilepsy, and malaria [5]. The seeds of *B. aegyptiaca* fruit have a high concentration of oils (46.0–54.7%), especially unsaturated fatty acids (up to 75% of total fatty acids), and proteins (26.1–34.3%). The oil extracted from *B. aegyptiaca* can potentially be used as biodiesel, an alternative to chemical diesel [5]. Furthermore, *B. aegyptiaca* plays a key role in conservation strategies owing to its resistance to drought, and it has been used as a native plant in the restoration of a degraded ecosystem in Africa [4–6]. In the Great Green Wall (GGW) project, *B. aegyptiaca* was selected as one of the native plants that was convenient for restoring degraded Sahelian ecosystems. The GGW project aimed to plant a green belt of trees extending south of the Sahara Desert across 11 countries [5].

Despite the ability of plants in the Zygophyllaceae family to adapt to harsh conditions, the stability and natural diversity of *B. aegyptiaca* currently face threats from anthropogenic pressure (wood is used as a fuel), animal overgrazing, and environmental pressure due to the increasing occurrence of drought episodes [5]. Civil wars and consequent instability also represent a worrisome factor, as the migration of people negatively impacts certain areas, which become vulnerable to the depletion of natural resources [7]. In addition, *B. aegyptiaca* faces deterioration when large areas of virgin land are urbanised [8].

Historically, the genus *Balanites* has undergone numerous changes in name and taxonomic position. The species *B. aegyptiaca* was first described by Alpino in 1592 under the name Agihalid [9]. In 1753, Linnaeus described it as *Ximenia aegyptiaca*, while in 1813, Delile replaced the name Agihalid with *Balanites*, from a Greek word meaning “the fruit” [10]. Harms in 1904 [11] proposed keeping the name *Balanites*, and *Balanites aegyptiaca* was formally adopted at the Vienna Botanical Congress in 1905 [9]. Initial classifications in the genus *Balanites* were based on morphological characteristics and were vulnerable to conflict, which led to the movement of *Balanites* between plant families. Initially, Bentham in 1862 [12] placed it within the family Simaroubaceae, while Engler 1896 [13] moved it to the family Zygophyllaceae. Cronquist in 1968 [14] returned it again to Simaroubaceae. Hegnauer in 1973 [15] provided evidence that *Balanites* did not contain quassia-like alkaloids, which are the main chemical characteristic of the family Simaroubaceae. Hegnauer 1973, Scholz 1964 and Cronquist 1981 [15–17] supported the return of

Balanites to the Zygophyllaceae family again. Maksoud 1988 [18] mentioned that similarities between the flavonoids of *Balanites* and Zygophyllaceae did not support the treatment of *Balanites* as a separate family. Sheahan 1993 [19] studied 37 species in 19 genera within Zygophyllaceae. Based on anatomy and c4 activity in 27 species, the results of their study supported the separation of *Balanites* into an independent family named Balanitaceae. Boesewinkel 1994 [20] also supported the separation of *Balanites* in a special family based on distinguishing features of ovule and seed characters. Next, the molecular and anatomical characteristics of flowers and embryos, in addition to pollen characteristics, supported the classification of *Balanites* within the family Zygophyllaceae [21, 22]. However, the last comprehensive review performed by [23] supported the separation of *Balanites* and its placement in the family Balanitaceae.

While a few molecular phylogenetic studies have been conducted for the Zygophyllaceae family, no study has specifically addressed *B. aegyptiaca*. The molecular aspects of Zygophyllaceae were studied by [24] using the plastid gene *rbcl*, in combination with anatomical and morphological data from 20 Zygophyllaceae species, including *B. maughamii*. Based on morphological and phylogenetic data, the redistribution of Zygophyllaceae into five subfamilies, namely, Morkillioideae, Tribuloideae, Seetzenioideae, Larreioideae, and Zygophylloideae, was proposed [24]. Recently, [1] presented a phylogenetic tree for Zygophyllaceae based on the Bayesian analysis tree of combined sequence data (*rbcl*, *trnL-E*, and *ITS*) for these five subfamilies. The results were consistent with previous studies [21, 22, 24, 25], and the *Balanites* genus was affiliated with the Tribuloideae subfamily within the Zygophyllaceae family. On the other hand, [26] reported that *B. aegyptiaca* includes five varieties; however, this number was later reduced to two by [26]. In a study by [27], three ecotypes were reported within the *B. aegyptiaca* population in Egypt based on RAPD markers. This discrepancy in the number of varieties belonging to *Balanites aegyptiaca* may be due to variation in morphological characteristics affected by environmental conditions.

Chloroplasts, which have an essential role in photosynthesis, are organelles in plant cells that contain their own genome, the plastome [28]. Typically, chloroplast genome sizes range between 120 and 170 kilobase pairs (kb) [29]. A chloroplast genome presents a four-region structure comprising a large single copy (LSC), a small single copy (SSC), and two inverted repeats (IRa and IRb). During the evolutionary history of plant families, plastomes have been subjected to strong selective pressure [30]. Thus, chloroplast genomes include useful phylogenetic information that can be used to study

evolutionary relationships at different taxonomic levels and resolve difficult problems in plant phylogenetics. In addition, chloroplast genomes represent a database for identifying and developing efficient polymorphic molecular markers for studying genetic diversity and population structure, and for DNA barcoding, which is a tool for identification [31, 32].

Zygophyllaceae is an angiosperm family that utilizes the C4 pathway [33], which helps plants adapt to harsh, dry environments. Molecular phylogenetic analyses suggest that the evolutionary history of Zygophyllaceae is related to an arid period that began in the Oligocene in Asia and Africa [34, 35]. The most recent study of three chloroplast genomes of *Zygophyllum* species belonging to the Zygophyllaceae family reported a significant reduction in the genome size of these species [36]. In this study, we validated the hypothesis that dry environments will significantly reduce the genome size of *B. aegyptiaca* in the family Zygophyllaceae. The aims of the present study are to provide baseline molecular information on the *B. aegyptiaca* plastome (*B. aegyptiaca* is considered typical of plants in dry and semiarid environments), to verify the plastome structure of *B. aegyptiaca*, and to perform comparative analyses and investigate phylogenetic relationships and variations between *B. aegyptiaca* and related Zygophyllaceae species, based on an available complete plastome dataset in GenBank.

Results

Characteristics of the *B. aegyptiaca* chloroplast Genome

The cp genome of *B. aegyptiaca*, shown in Fig. 1, is a circular molecule with a length of 155,800bp. It has a four-region structure comprising a large single copy, a small single copy, and two inverted repeats. The LSC and SSC regions were 86,562bp and 18,102bp long, respectively, while IRa and IRb regions were 25,568bp each (Table 1). The length of the coding region is 75,890bp and represents 49% of the whole genome, while the noncoding region length is 79,910bp (51%). The percentage of AT in the whole genome was 64.5%, whereas the percentage of GC was 35.5%. The genome of *B. aegyptiaca* consists of A=31.8%, T(U)=32.7%, C=18.1% and G=17.4%, as shown in Table 1.

The total number of unique genes was 107, and the number of duplicated genes in the inverted region was 18. The LSC region was composed of 60 protein-coding genes and 21 tRNA genes, the SSC region contains 11 protein-coding genes and one tRNA, and 10 protein-coding genes and 22 tRNAs are located in the IR regions. Most of the protein-coding genes start with a methionine codon (AUG).

Sixteen of the 107 genes in *B. aegyptiaca* contain introns, 11 are protein coding genes, and five were tRNA

genes, as illustrated in Table 2. The *clpP* and *ycf3* genes present two introns, while the remaining genes present only one intron. Ten introns were included in the LSC region, one intron is included in the SSC region, and five introns were specifically located within the IRa and IRb regions.

Relative synonymous codon usage (RSCU)

The nucleotides of protein coding and tRNA genes were used to determine the codon usage bias of the plastome. The results obtained from the analysis of protein-coding genes and tRNA genes (78,624bp) of the *B. aegyptiaca* plastome are shown in Table S2. Genes are encoded by 22,415 codons. Leucine is the most frequent amino acid (11.1%), as shown in Fig. 2, whereas cysteine is the least frequent (1.2%). The RSCU values in Table S2 show that half (30) of the codons are >1, all with an A/T ending. It can be seen from the data that tryptophan and methionine with no codon usage bias have an RSCU value of 1.

RNA editing sites

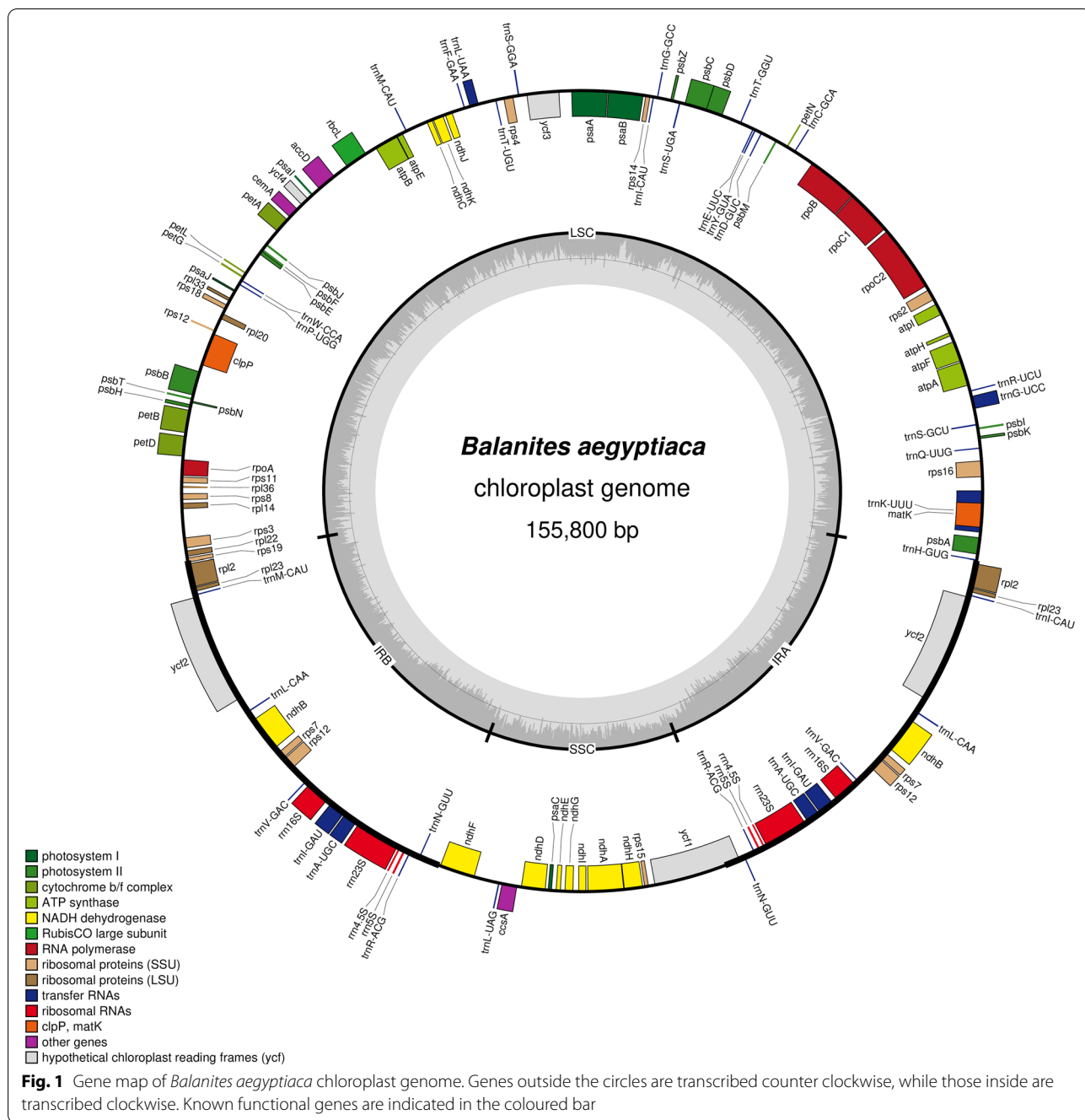
The PREP suite was used to predict the RNA editing sites in the *B. aegyptiaca* plastome, and the first codon position of the first nucleotide was used in the analysis.

The RNA editing sites are presented in Table S3. Overall, there were 41 editing sites in the genome distributed among protein-coding genes. Most of the codon position exchanges involved serine (S) and leucine (L) amino acids (S to L). The results obtained show the highest number of editing sites in the *ndhB* and *ndhD* genes (12 and 6 sites, respectively), followed by *rpoB* and *ndhF* genes with three sites. Moreover, the *accD*, *atpA*, *matK*, and *ndhA* genes have 2 sites, while 1 site is present for each of the remaining genes. The results of RNA editing show that certain genes, namely, *atpI*, *ccsA*, *petB*, *petD*, *petG*, *petL*, *psaB*, *psaI*, *psbB*, *psbE*, *psbF*, *rpl2*, *rpl23*, *rpoA*, *ycf3*, and *atpB*, do not possess a predicted site in the first codon of the first nucleotide.

Repeat Analysis

Long repeats

There are four types of repeats in the cp genome of *B. aegyptiaca*: palindromic (21), forward (12), reverse (15), and complement (1), as evident from Table S4. Overall, there were 49 repeats in the *B. aegyptiaca* plastome. The majority (75.5%) of repeats were found in the intergenic spacer (IGS) region. The sizes of most of the repeats range from 20 to 29bp (71.4%), followed by 10–19bp (14.3%), 30–39bp (10.2%), and 40–49bp (4.1%). The tRNAs include 4 repeats (8.2%), while the remaining 8 repeats (16.3%) were in the protein-coding genes *ndhC*, *ycf2*, *clpP*, *ycf1*, and *ndhA*. We noticed that the



protein-coding gene *ycf2* had the most repeat locations: 2 palindromic repeats and 2 forward repeats.

A comparison of the number of repeats in six Zygo-phyllaceae species (*Balanites aegyptiaca* (L.), *Guaiacum angustifolium* Engelm., *Larrea tridentata* (DC.) Coville, *Tetraena mongolica* Maxim., *Tribulus terrestris* L., and *Zygophyllum xanthoxylon* (Bunge) Maxim.) is provided in Fig. 3 (species description is available in Table S5); the genome of *L. tridentata* has the highest frequency

of palindromic repeats (26), while that of *Z. xanthoxylon* has the lowest (15). *L. tridentata* and *T. mongolica* have the same number of forward repeats (17), while the same number of reverse repeats occur in the cp genome of *T. mongolica* and *T. terrestris* (13). We can also see that complement repeats are the least abundant in the three genomes; *B. aegyptiaca*, *L. tridentata* and *G. angustifolium* have only one complement repeat, and complement repeats are absent in the other three species.

Table 1 Chloroplast genome features of *Balanites aegyptiaca*

Feature	<i>B. aegyptiaca</i>
Genome size (bp)	155,800
IRA (bp)	25,568
IRB (bp)	25,568
LSC (bp)	86,562
SSC (bp)	18,102
Total No of Unique Genes	107
rRNA	4
tRNA	28
Protein-Coding genes	75
A%	31.8
T (U)%	32.7
G%	17.4
C%	18.1
GC%	35.5

Simple sequence repeats (SSRs)

A total of 240 SSRs are present in the plastid genome of *B. aegyptiaca*, which is a larger number than in the five other species of Zygophyllaceae shown in Table 3. The majority of SSRs in the cp genome are mononucleotides (86.3%), mostly poly T (53.6%) and A (44%) (Figs. 4 and 5); poly C and G represent 1.5 and 1% of

repeats, respectively. The results obtained from the analysis of SSRs frequency in the genomes of six species of Zygophyllaceae are presented in Table 3 and Fig. 4. The dinucleotide AT/AT was found in the genome of all species, while the dinucleotide AG/CT was found in four species but was absent from *L. tridentata* and *G. angustifolium*. Furthermore, the existence of three trinucleotides (AAT/ATT, AAG/CTT, and AAC/GTT), as well as nine tetra-repeats (AAAC/GTTT, AAAG/CTTT, AAAT/ATTT, ACAT/ATGT, AATC/ATTG, AATC/ATTG, AATG/ATTC, ACCT/AGGT, and ACTG/AGTC), and three pentanucleotides (AATAT/ATATT, AATCG/ATTTCG, and AAA TAT/ATATTT) was observed only in *T. mongolica*, *Z. xanthoxylon*, and *L. tridentata* (Fig. 4).

A comparison between the frequency of SSRs in the plastomes of the six species is presented in Fig. 5; it is clear that mononucleotides are the most frequent in all genomes. *B. aegyptiaca* had the highest number of mononucleotides, trinucleotides, and tetranucleotides (207, 6, and 12, respectively). However, pentanucleotides were not present in the cp genomes of *B. aegyptiaca*, *G. angustifolium*, or *T. terrestris*. In addition, hexa-repeats were not present in any of the six species.

Overall, the IGS region included most of the SSRs repeats (83.3%), shown in Fig. 6, followed by the coding regions (16.7%).

Table 2 Genes with introns in the chloroplast genome of *Balanites aegyptiaca*

Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
<i>rpoC1</i>	LSC	432	798	1617		
<i>rpl2</i>	IRA	394	665	434		
		434	665	394		
<i>ndhA</i>	SSC	539	1165	553		
<i>ycf3</i>	LSC	153	793	230	733	124
<i>clpP</i>	LSC	247	634	294	836	71
<i>petB</i>	LSC	6	785	642		
<i>petD</i>	LSC	8	780	477		
<i>rps16</i>	LSC	228	757	40		
<i>atpF</i>	IRA	410	736	145		
<i>ndhB</i>	IRB	758	679	775		
		775	679	758		
<i>rps12</i>	LSC	114		26	538	232
		114		232	538	26
<i>trnL-UAA</i>	LSC	35	493	50		
<i>trnK-UUU</i>	LSC	35	2537	37		
<i>trnI-GAU</i>	IRB	32	956	40		
		40	956	32		
<i>trnA-UGC</i>	IRB	37	802	36		
		36	802	37		
<i>trnG-UCC</i>	LSC	32	715	60		

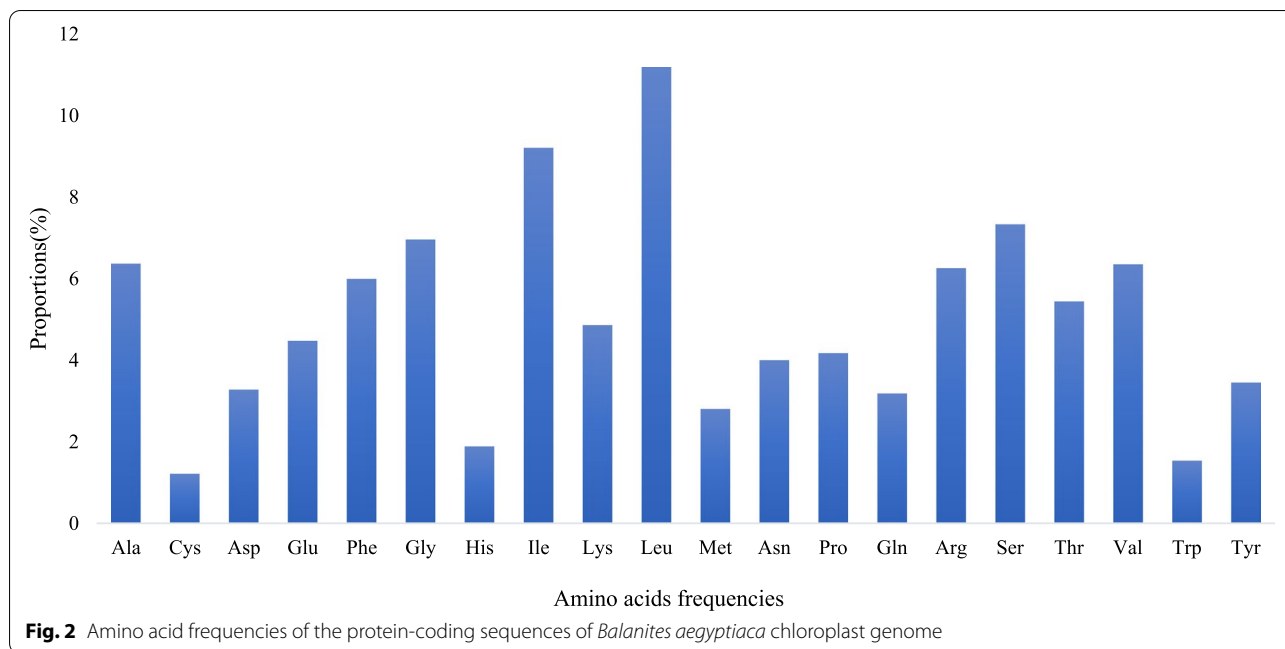


Fig. 2 Amino acid frequencies of the protein-coding sequences of *Balanites aegyptiaca* chloroplast genome

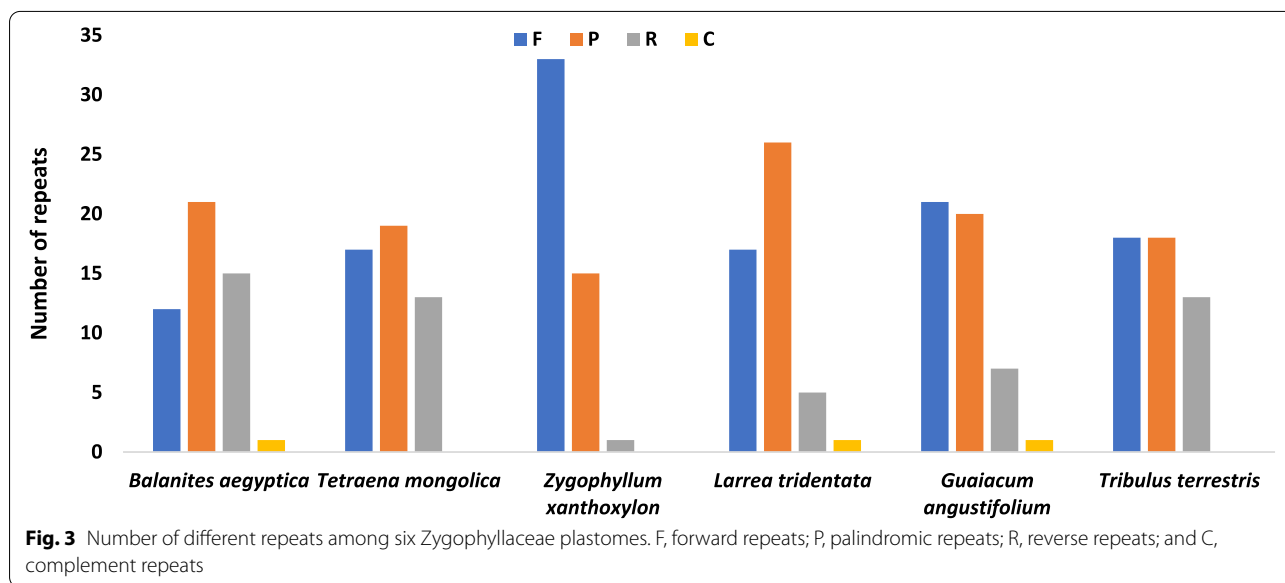


Fig. 3 Number of different repeats among six Zygothylaceae plastomes. F, forward repeats; P, palindromic repeats; R, reverse repeats; and C, complement repeats

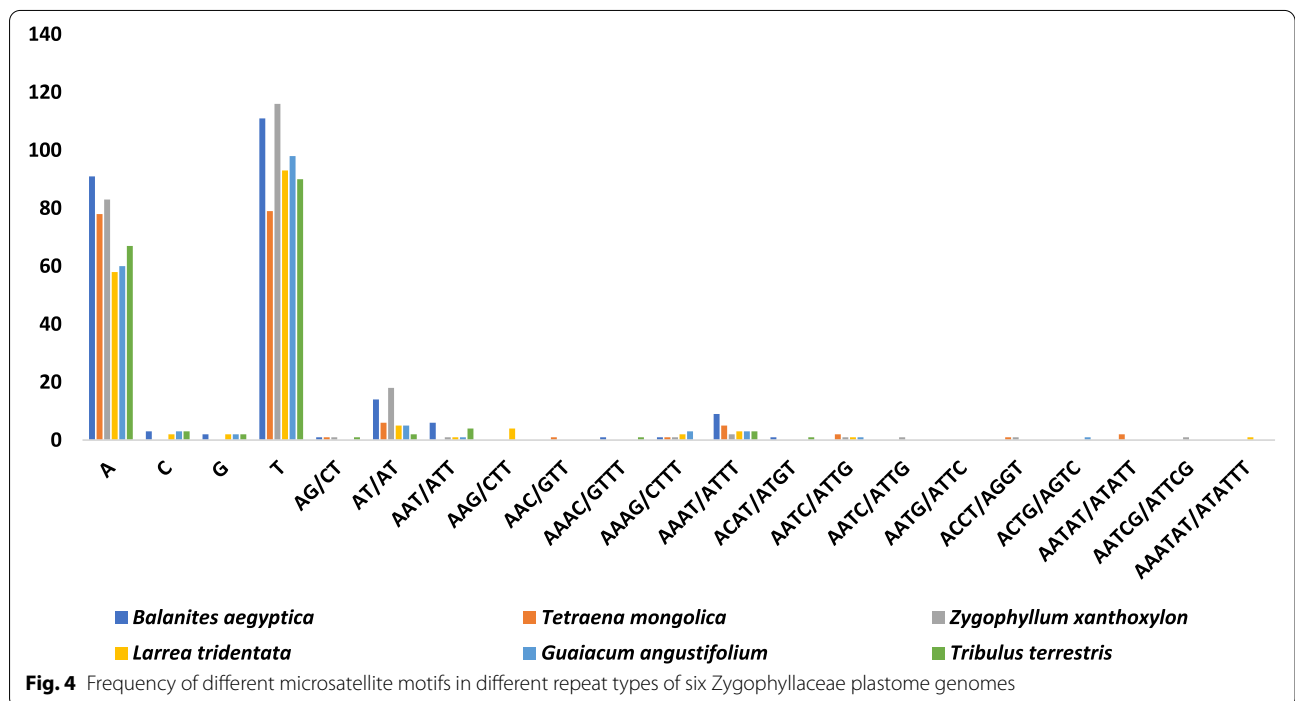
Sequence divergence

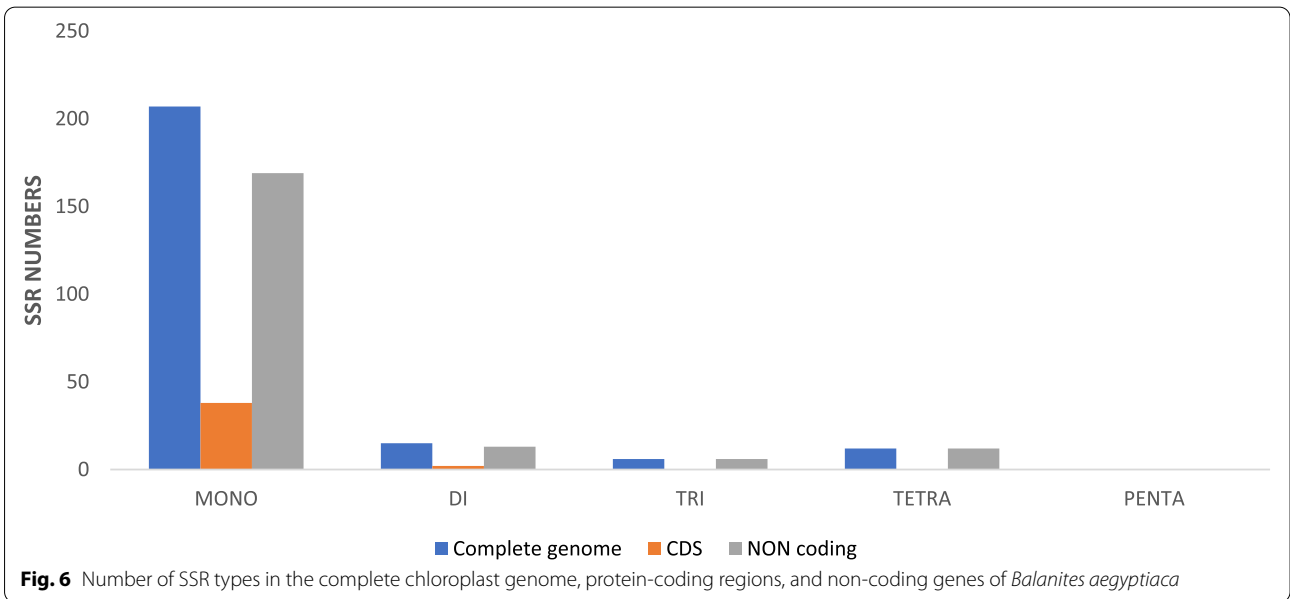
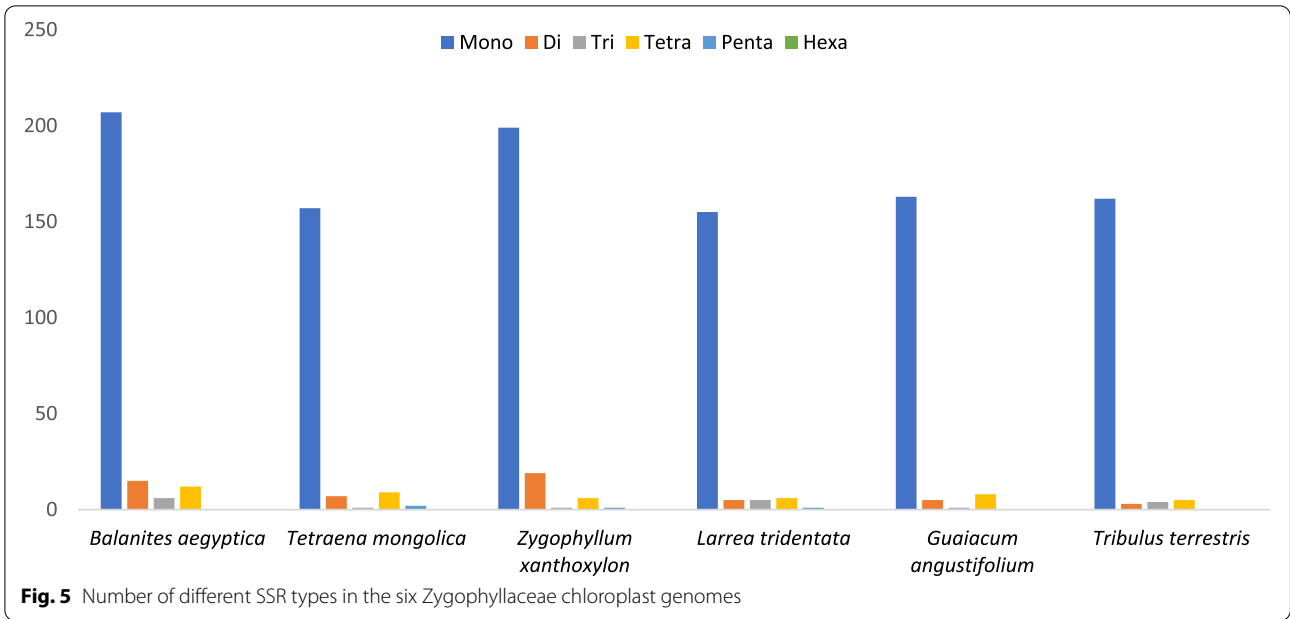
To investigate the degree of genome sequence divergence, the program mVISTA was used to align the cp genome sequence of *B. aegyptiaca* with five Zygothylaceae chloroplast genomes available in GenBank: *T. mongolica*, *Z. xanthoxylon*, *L. tridentata*, *G. angustifolium*, and *T. terrestris*. The alignment showed that genes were less conserved in the genomes of *T. mongolica* and *Z. xanthoxylon*, especially the *ndh* genes *rps16* and *ycf2*. In general, the protein-coding genes were more conserved than the noncoding regions (Fig. 7). The

noncoding regions presented high divergence in the following genes: *psabA-trnK-UUUU*, *trnK-UUUU-rps16*, *psbK-psbI*, *trnG-UCC-trnR-UCU*, *atpF-atpH*, *atpH-atpI*, *rps2-rpoC2*, *rpoC2-rpoC1*, *rpoC1-rpoB*, *rpoB-trnC-GCA*, *psbD-trnS-UGA*, *psaA-ycf3*, *ycf3-trnS-GGA*, *trnS-GGA-rps4*, *trnT-UGU-trnL-UAA*, *ndhC-trnM-CAU*, *ntrnM-CAU-atpE*, *atpB-rbcL*, *rbcL-accD*, *cemA-petA*, *petA-psbJ*, *psbJ-psbE*, *psbE-petL*, *psaJ-rpl33*, *rps18-rpl20*, *rps12-psbB*, *psbB-psbT*, *psbH-petB*, *petB-petD*, *petD-rpoA*, *rps8-rpl14*, *rpl14-rps3*, *rpl2-rpl23*, *ycf2-trnL-CAA*, *trnL-CAA-ndhB*, *ndhF-trnL-UAG*, *psaC-ndhE*, *trnN-GUU-trnR-ACG*,

Table 3 cpSSRs detected in six Zygophyllaceae chloroplast genomes

SSR type	Repeat unit	<i>Balanites aegyptiaca</i>	<i>Tetraena mongolica</i>	<i>Zygophyllum xanthoxylon</i>	<i>Larrea tridentata</i>	<i>Guaiacum angustifolium</i>	<i>Tribulus terrestris</i>
Mono	A	91	78	83	58	60	67
	C	3	0	0	2	3	3
	G	2	0	0	2	2	2
	T	111	79	116	93	98	90
Di	AG/CT	1	1	1	0	0	1
	AT/AT	14	6	18	5	5	2
Tri	AAT/ATT	6	0	1	1	1	4
	AAG/CTT	0	0	0	4	0	0
	AAC/GTT	0	1	0	0	0	0
Tetra	AAAC/GTTT	1	0	0	0	0	1
	AAAG/CTTT	1	1	1	2	3	0
	AAAT/ATTT	9	5	2	3	3	3
	ACAT/ATGT	1	0	0	0	0	1
	AATC/ATTG	0	2	1	1	1	0
	AATC/ATTG	0	0	1	0	0	0
	AATG/ATTC	0	0	0	0	0	0
	ACCT/AGGT	0	1	1	0	0	0
	ACTG/AGTC	0	0	0	0	1	0
Penta	AATAT/ATATT	0	2	0	0	0	0
	AATCG/ATTCG	0	0	1	0	0	0
	AAATAT/ATATTT	0	0	0	1	0	0
Total		240	176	226	172	177	174





(See figure on next page.)

Fig. 7 Whole chloroplast genome alignments for six Zygophyllaceae species via the mVISTA program, using the annotation of *Balanites aegyptiaca* as reference. The x-axis represents the coordinates in the cp genome, while the y-axis indicates percentage identity from 50 to 100%. The top grey arrows indicate the position and direction of each gene. Pink indicates non-coding sequences (NCS), blue indicates protein-coding genes, and light green indicates tRNAs and rRNAs

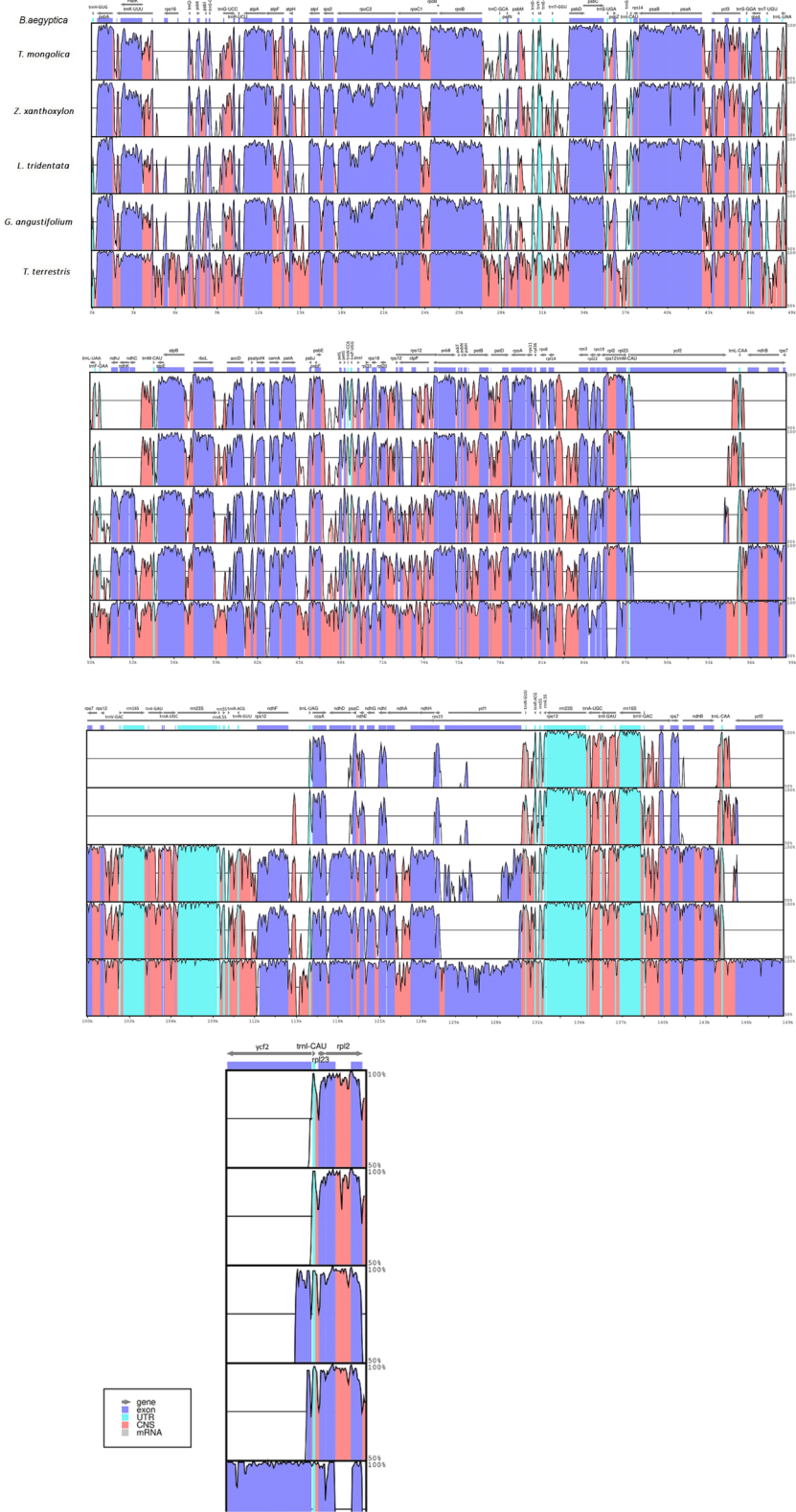
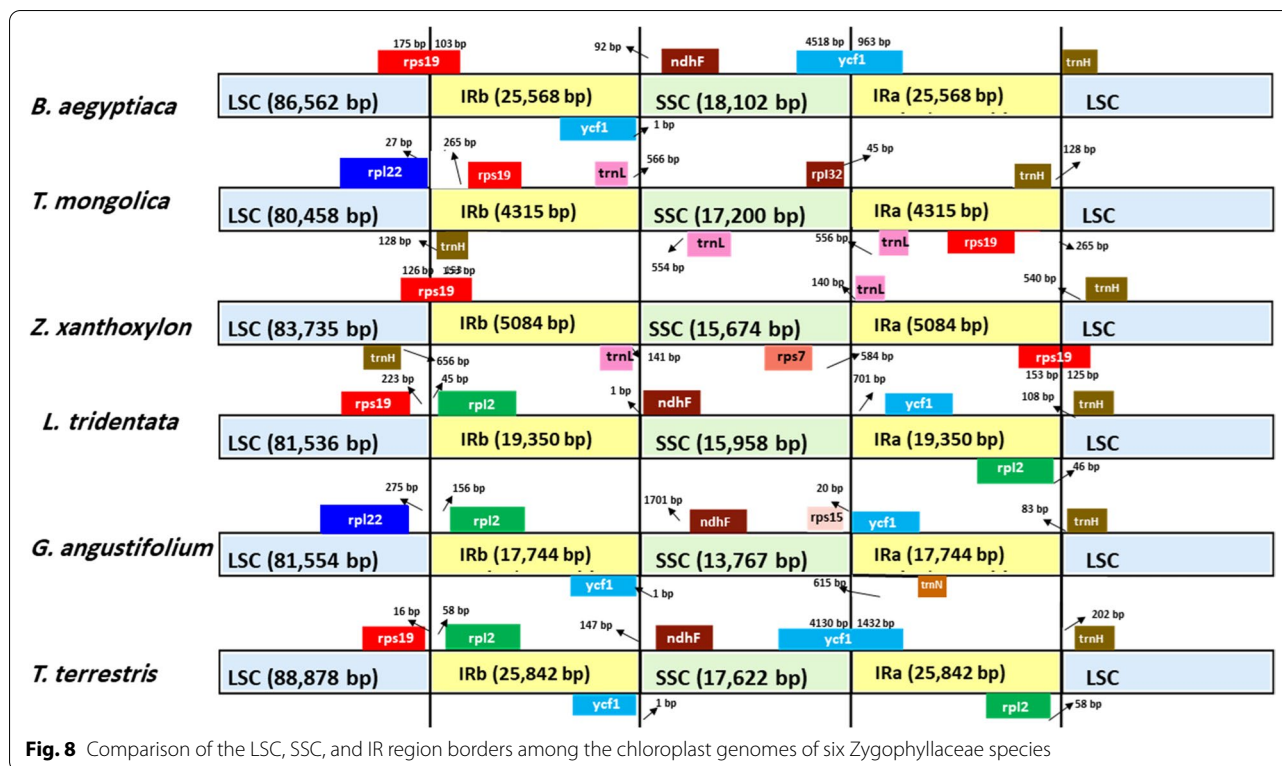


Fig. 7 (See legend on previous page.)



trnA-UGC-trnI-GAU, *trnI-GAU-rrn16S*, *rrn16S-trnV-GAC*, *trnL-CAA-ycf2*, *trnI-CAU-rpl23*, and *rpl23-rpl2*. However, the following protein-coding genes showed divergence in fewer regions: *matK*, *rpoC2*, *psaA*, *accD*, *cemA*, *clpP*, *rpl23*, *ndhF*, *ccsA*, and *rpl23*.

Boundary between LSC/SSC and IR regions

The comparisons between IR-SC boundaries for the six species of Zygophyllaceae are shown in Fig. 8. In general, the variation in length of the two LSC/SSC regions was lower than that of the IRa/IRb regions. The shortest LSC and SSC regions were 80,458 bp and 13,767 bp, respectively, in *T. mongolica* and *G. angustifolium*, while the longest LSC and SSC regions were 88,878 bp and 18,102 bp, in *T. terrestris* and *B. aegyptiaca*, respectively. The IRa/IRb regions of *B. aegyptiaca* and *T. terrestris* were the longest (25,568/25,842), and contraction was noticed in the IR region in both *Z. monogolica* (4315 bp) and *Z. xanthoxylon* (5084 bp) compared to the other species (Table S6).

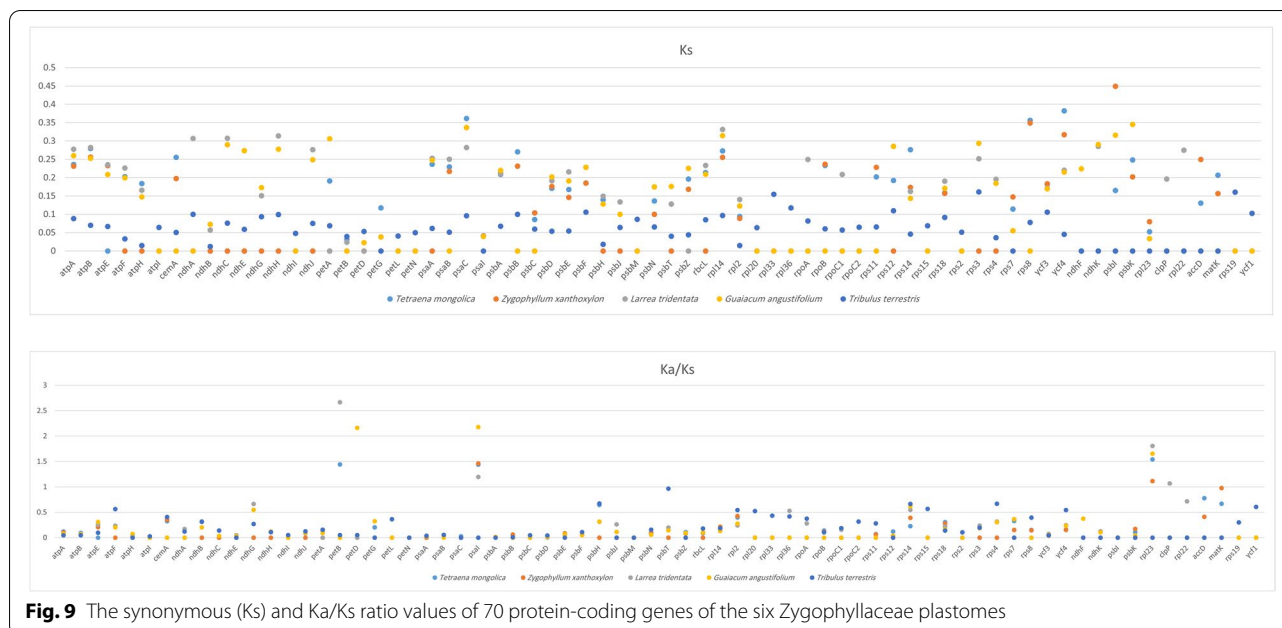
The *ycf1* gene was located within the SSC/IR border in *B. aegyptiaca*, *T. terrestris*, and *G. angustifolium* but was located within the IRa region in *L. tridentata* (701 bp). The pseudogene *ycf1* was present in the IRb/SSC border in *B. aegyptiaca*, *G. angustifolium* and *T. terrestris*. In contrast, the *ycf1* gene was not present in *Z. xanthoxylon* or *T. mongolica*; instead, the *rpl32* gene was located

in the SSC region, close to the SSC/IRa border (45 bp) in *T. mongolica*, and the *rps7* gene was present in the SSC region in *Z. xanthoxylon*. Both *T. mongolica* and *Z. xanthoxylon* possessed two copies of the *trnL* gene (in the IRa and IRb regions).

The *ndhF* gene was in the IRb/SSC border of *L. tridentata* (1 bp), whereas gene *ndhF* was in the SSC (92 bp) of IRb/SSC border in *B. aegyptiaca* and *T. terrestris* (147 bp), while *G. angustifolium* having *ndhF* gene in (1701 bp) of the IRb/SSC border. However, the *ndhF* gene was missing from the SSC region in both *Z. xanthoxylon* and *T. mongolica* as well.

The *trnH* gene showed variation in its location in the IR/LSC border. *B. aegyptiaca* had the *trnH* gene in the IRa/LSC border, while it was located in the LSC region and far from the IRb/LSC border in *Z. xanthoxylon* (540 bp), *L. tridentata* (108 bp), *G. angustifolium* (83 bp), and *T. terrestris* (202 bp). *T. mongolica* was different; the *trnH* gene was located in the IRa region far from the IRa/LSC border (128 bp). *T. mongolica* and *Z. xanthoxylon* had another copy of the *trnH* gene; it was present in the IRb region and far from the IRb/SSC border (128 bp) in *T. mongolica*, and it was in the LSC region far from the IRb/SSC border (656 bp) in *Z. xanthoxylon*.

Variations in the location of the *rps19* gene in the IR/LSC border also occurred, with it being located at different sites in the cp genomes (Fig. 8). The *rps19* gene



spanned the border of LSC/IR in *B. aegyptiaca* and *Z. xanthoxylon*. *Z. xanthoxylon* duplicated the *rps19* gene in the IRa region. *rps19* was located in the LSC region, distant to the LSC/IR border in *L. tridentata* (223 bp) and *T. terrestris* (16 bp). Both species possessed duplicated *rpl2* genes in the IRa and IRb regions. *T. mongolica* had two copies of the *rps19* gene located in the inverted repeat region (IRa and IRb), whereas the *rpl22* gene was in the LSC region. The *rpl22* and *rpl2* genes were located in the LSC and IRa regions in *G. angustifolium*, respectively.

Characterisation of substitution rate

The value of synonymous (Ks) and nonsynonymous (Ka) substitutions and the Ka/Ks ratio were calculated among the 70 protein-coding genes that represent the common genes in the species that were compared within Zygophyllaceae (*T. mongolica*, *Z. xanthoxylon*, *L. tridentata*, *G. angustifolium*, and *T. terrestris*). Several genes were under positive selection with Ka/Ks values > 1, as shown in Fig. 9: *atpE*, *ndhG*, *petB*, *petD*, *psaI*, *psbH*, *psbT*, *rpl2*, *rps14*, *rps4*, *rps7*, *ycf4*, *rpl23*, and *matK*. In addition, most Ks values were < 1 in all genes (Fig. 9), except for *atpA*, *atpB*, *psaC*, *rpl14*, and *ycf4*.

Phylogenetic Analysis

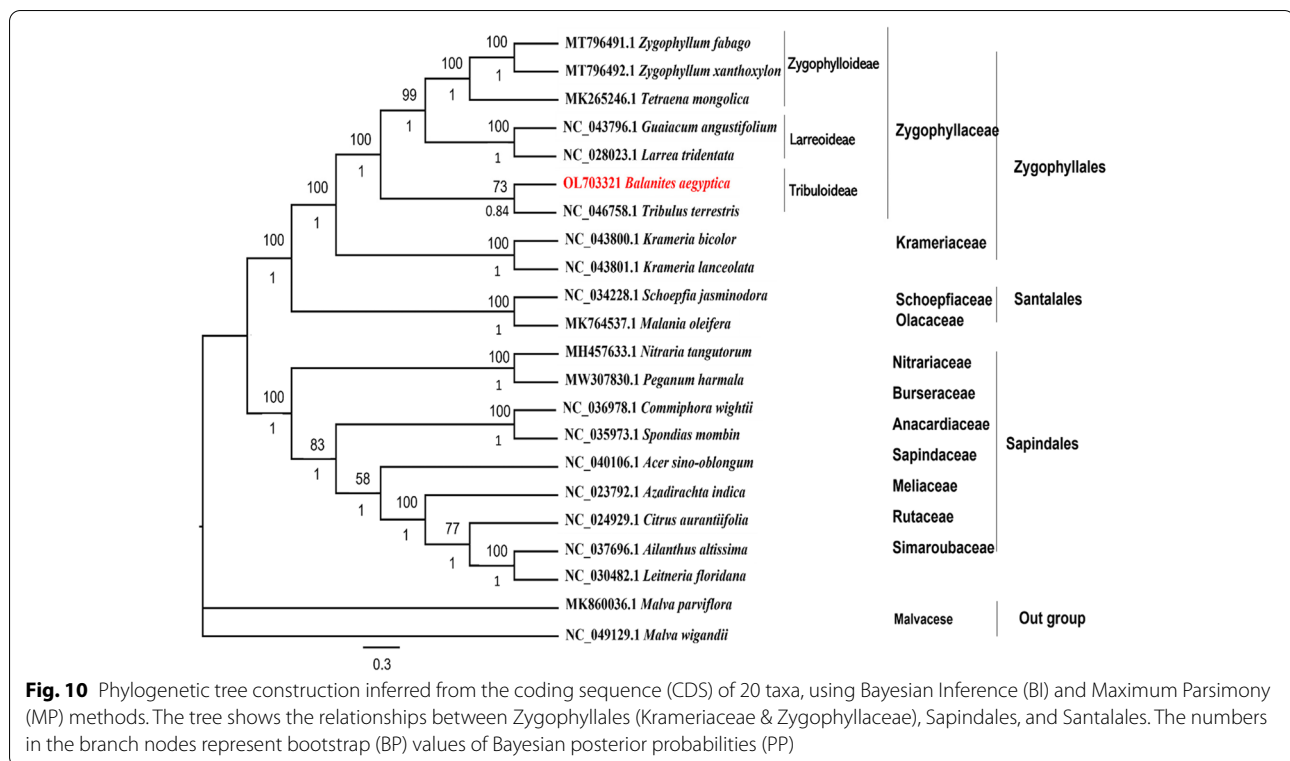
Phylogenetic relationships were assessed among all currently available cp genome sequences in GenBank for the order Zygophyllales (six species from the family Zygophyllaceae), two species from Krameriaceae, and 11 plastome sequences from the orders Sapindales and

Santalales, where *B. aegyptiaca* was classified formerly. Phylogenetic analysis was performed via MP and BI analyses. The two phylogenetic trees were topologically similar, with the majority of nodes having 100% bootstrap (BP) values and 1.00 Bayesian posterior probabilities (PP) (Fig. 10). The monophyly of the Zygophyllales order is strongly supported (PP = 1.00; BS = 100%) by this study.

Zygophyllales was divided into two highly supported clades (PP = 1.00; BS = 100%): one clade contained species from the Krameriaceae family, and the other comprised Zygophyllaceae species. Zygophyllaceae was further divided into three separate clades representing three subfamilies: Tribuloideae (PP = 0.84; BS = 73%), Larreoideae (PP = 1.00; BS = 100%), and Zygophylloideae (PP = 1.00; BS = 100%). In this study, *B. aegyptiaca* was nested in the Zygophyllaceae family and formed a moderately supported clade with *T. terrestris*, representing the Tribuloideae subfamily.

Discussion

This study presents the first chloroplast genome from *B. aegyptiaca*, a plant within the Tribuloideae (Zygophyllaceae) subfamily. The length of the cp genome in *B. aegyptiaca* (155 kb) was similar to that seen in the cp genome of *Tribulus terrestris* (158 kb), which is also within the Tribuloideae subfamily. Previous studies have [36] indicated a great reduction in plastid genome size in three species of the subfamily Zygophylloideae (Zygophyllaceae), namely, *T. mongolica*, *Z. xanthoxylon*, and *Z. fabago*. The sizes of these genomes ranged between 104 and 106 kb. In the current study, a large decrease in



the size of the *B. aegyptiaca* genome, which has mentioned in regard to some Zygophyllaceae species, was not observed [36]. Size, structure, gene content, and organization are usually conserved in the chloroplast genomes of angiosperms [30]. However, the chloroplast genomes in some species are smaller than those in most other angiosperms. The cp genome in Saguaro (*Carnegiea giganteais*) considered to the smallest known plastid genome in autotrophic angiosperms having size (113 kb), where lost whole of IR region and *ndh* genes [37]. As well in *Astragalus membranaceus* of the Fabaceae whose chloroplast genome size is 124 kb [38]. In addition, to decrease in genome size is a common feature of parasitic plants and is caused by their conversion from an autotrophic to parasitic strategy. This decrease in genome size is accompanied by several other changes, such as pseudogenization, gene loss, structural rearrangement, and size reduction [39, 40]. For example, the plastid genome size in parasitic species of the Loranthaceae family ranges from 116 to 139 kb and the reductions in the plastid sizes of some holoparasites may be even greater. Indeed, the root parasite species in the family Cynomoriaceae has a total plastid genome length of 45,519 bp [41].

The organization, size, and structure of the *B. aegyptiaca* chloroplast genome is similar to those of other angiosperms; the size of the LSC region is 86 kb, the SSC region is 18 kb, and the two IR regions are 25 kb. The

LSC region in angiosperms ranges from 80 to 90 kb, SSC regions are approximately 16–27 kb, and the sizes of the two IRs range from 20 to 28 kb [36].

A typical angiosperm chloroplast genome consists of 113 genes, including 79 protein coding genes, 30 tRNA genes and four rRNA genes [30]. The *B. aegyptiaca* chloroplast genome had a similar number of genes (107 genes), including 75 protein-coding genes, 28 tRNA genes and 4 rRNA genes. This was similar to other species in the *Zygophyllum* genus of the family Zygophyllaceae [36], which contained 107 genes, including 75 protein-coding genes, 33 tRNA genes and 4 rRNA genes.

We found that a GC content of 35% in the *B. aegyptiaca* cp genome was akin to GC content in *Zygophyllum* spp. [36], which ranged between 33 and 36%. Codon usage bias plays an important role in chloroplast genome evolution and occurs as a result of natural selection and mutations [42, 43]. Codons encoding the amino acid leucine were the most frequently observed in the plastome of *B. aegyptiaca*, whereas those encoding cysteine were the least frequently observed. This finding has also been reported in the chloroplast genomes of *Populus* species [44]. While isoleucine is the most encoded amino acid among *Zygophyllum* spp. (Zygophyllaceae), methionine is a less abundant amino acid [36].

Introns play a significant role in gene expression and regulation [45, 46]. The *clpP* and *ycf3* genes in the *B.*

aegyptiaca plastome had two introns, while the remaining genes contained only one intron; these findings are consistent with previously published reports [47, 48]. The chloroplast genomes of three species of *Zygophyllum* also contained two introns in the *ycf3* gene [36].

RNA editing includes the processes of inserting, deleting, or modifying nucleotides, which leads to changes in the DNA coding sequence during RNA transcription [49]; in turn, this allows the creation of different protein transcripts [50]. Only C-U RNA editing has been described in plastomes [51]. The results obtained from the present study show that in the *B. aegyptiaca* plastome, most of the transformation in the codons was from the amino acid exchange of serine to leucine and that the *ndhB* gene had the highest number of editing sites, which could be attributed to high preservation of RNA editing [52, 53].

Prior studies have noted the importance of repeat sequences in the cp genome, which play a significant role in genomic recombination [54, 55]. In the *B. aegyptiaca* plastome, most repeats are in the *ycf2* gene, which is consistent with data obtained from other angiosperms [56, 57].

Several reports [58–61] have shown the importance of chloroplast SSRs (cpSSRs) as reliable molecular markers to discriminate between specimens at lower taxonomic levels and in studying population structure. In the *B. aegyptiaca* cp genome, there are 240 SSRs repeats, the largest number compared to five other species of *Zygophyllaceae*. Mononucleotide A/T SSRs were the most frequent, and most SSRs were marked in noncoding regions (83.3%). Thus, we recommended used plastid genome data in this study for developing cpSSR loci and for studying the levels of genetic variation in the *B. aegyptiaca* population.

Although the plastid genome is conserved in angiosperm plants as previously reported [30], several studies have reported variations in size and boundaries among IR/LSC and IR/SSC regions and variation in gene location [62, 63]. In the present study, comparisons between the IR-LSC and IR-SSC boundaries in the six complete *Zygophyllaceae* chloroplast genomes showed clear variation in the inverted repeat region in *Zygophyllaceae* chloroplast genomes and significant contraction in the IR region in the *Z. xanthoxylon* and *T. mongolica* genomes (synonym of *Zygophyllum mongolicum*). Our results are consistent with recent results published by [36], who mention that the most conspicuous change occurs in two IRs reduced by approximately 16–24 kb in size in the plastid genomes of *T. mongolica*, *Z. xanthoxylon*, and *Z. fabago*.

There is variation in the border of the IR-SC region among the six species of *Zygophyllaceae* (in this study

based on the position of the *rps19*, *rpl2* and *trnH* genes) at the LSC/IRa junction, which could occur as a result of contraction and expansion in the inverted repeat region. A similar observation of variation in the location of the *trnH*, *rpl2*, and *rps19* genes at the border was reported for species in the subfamily *Acanthoideae* by [64]. Furthermore, two copies of the *trnL* gene appeared in the IRs of both *T. mongolica* and *Z. xanthoxylon*. This could be related to IR border changes due to either gene loss, such as *ndhF*, or expansion/reduction. The *trnL* genes at the IR borders of *Tetraena mongolica* and *Zygophyllum xanthoxylon* are not the same, *trnL-CAA* and *trnL-UAG* respectively. The *trnL-CAA* is usually located in the IR regions and *trnL-UAG* in the SSC region. We noticed that the *ycf1* gene was located in the IRb/SSC border in *B. aegyptiaca*, *G. angustifolium* and *T. terrestris*, and studies [65] have indicated that the *ycf1* gene in the IRb region is a pseudogene in the angiosperm chloroplast genome.

In the current comparison, we observed that all the *ndh* genes that were usually located in the SSC and IR regions were lost in *T. mongolica* and *Z. xanthoxylon*. This was similar to the mVISTA results, which also showed a loss of *ndh* genes (specifically in the sequences from *T. mongolica* and *Z. xanthoxylon*). A similar loss of the *ndh* group (11 genes) from the chloroplast genomes of *T. mongolica*, *Z. xanthoxylon*, and *Z. fabago* was also reported by [36], and it was reported that rRNA genes, which are usually present in IRs, were in the SSC region. In our study, although we used plastome sequences from GenBank with accession numbers that were different than those used by [36] for the species *T. mongolica* and *Z. xanthoxylon*, we observed a similar loss of all *ndh* genes and the presence of ribosomal proteins codes (*rpl32* & *rps7*) genes in the SSC region in *Zygophyllum* spp., which could be a result of contraction of the IR regions. The *ndh* genes represent a complex group consisting of approximately 30 subunits, and 11 out of 30 subunits are used in encoding the NADH dehydrogenase complex in plant plastids and are involved in photosynthesis [66]. The partial or complete loss of genes associated with photosynthesis (*ndh*) has been reported in some species of the *Cactaceae* [37], *Pinaceae* [67], and *Orchidaceae* families [68]. It is also a common phenomenon in hemiparasites and holoparasites of the *Santalales* and *Orobanchaceae* [69–74].

On the other hand, in a study by [36] on three species of the subfamily *Zygophylloideae*, the authors indicated a significant loss in chloroplast genes (11 genes), and it was not certain if this loss was because of adaptation by plants to living in arid and semiarid environments or as a result of gene transmission to the nuclear genome. In our study, we present the cp genome of *B. aegyptiaca*

(subfamily Tribuloideae and family Zygophyllaceae), which is adapted to arid and semiarid lands. However, we did not observe a large loss in plastid genes and all 11 *ndh* genes were present in the *Balanites* plastome. Another possible explanation could be that gene loss observed in *Zygophyllum* spp. may be related to the evolutionary history of the subfamily (Zygophylloideae); this point is interesting, and we recommend future study of different species in the family Zygophyllaceae. We also recommend that future studies include species from subfamilies Morkillioideae, Seetzenioideae and Larreioideae in Zygophyllaceae.

The Ka/Ks value is usually used for evaluating sequence variations in different species or taxonomical species with unknown evolutionary status and to detect substitution, selection and beneficial mutations in genes under selective pressure [75]. The values of synonymous (Ks) and nonsynonymous (Ka) substitutions and the Ka/Ks ratio showed that 14 protein-coding genes (*atpF*, *ndhG*, *petB*, *petD*, *psaI*, *psbH*, *psbT*, *rpl2*, *rps14*, *rps4*, *rps7*, *ycf4*, *rpl23*, and *matK*) were under positive selection and may have a faster evolutionary rate [57]. Most of these genes play a role in maintaining the efficiency of photosynthesis. In the current study, the *ycf4* gene was under positive selection. *ycf4* is located in the thylakoid membrane and involved in the assembly of the photosystem I complex [76], it is possible that *ycf4* has high substitution rates in arid plant species. Further research is required to investigate the use of these regions in detecting phylogenetic relationships among Zygophyllaceae species.

The plastome consists of many highly efficient genes that may resolve phylogenetic questions at different levels of angiosperm taxonomy [77, 78]. In this study, we found that *B. aegyptiaca* was distantly related to the family Simaroubaceae. This provides additional evidence to confirm the current position of *B. aegyptiaca* in the family Zygophyllaceae, as suggested by previous studies [21, 22, 25]. The chloroplast phylogenetic tree showed a strong relationship among Zygophyllaceae species. *B. aegyptiaca* and *T. terrestris* formed a group representing the Tribuloideae subfamily, *L. tridentata* and *G. angustifolium* (within the same branch) represent the Larreioideae subfamily, while *Z. mongolica*, *Z. fabago*, and *Z. xanthoxylon* were part of the same group and represent the Zygophylloideae subfamily. These findings broadly support the previous results of [1, 21].

Materials and methods

Sample collection and DNA extraction

Fresh leaves of *B. aegyptiaca* were collected from the Wadi Fatima (Al-Jamoum) Makkah district (21° 38' 49.6" N, 39° 41' 49.3" E) in Saudi Arabia. The plants were identified by Dr. Widad S. Al-Juhani, assistant professor

of taxonomy and supervisor of the herbarium in the Biology Department of Umm AlQura University, based on herbarium specimens and morphologies in relevant literature. A sample specimen was prepared and deposited in the herbarium of Umm Al-Qura University, Makkah, with accession number UQU072021. Samples of fresh leaves were dried in silica gel for DNA extraction. DNA was extracted from the silica gel-dried leaves of *B. aegyptiaca* using the CTAB Plant DNA extraction protocol [79].

Library construction and De novo Genome sequencing

Library construction and sequencing using Illumina sequencing and read length 151bp paired-ends were carried out by Macrogen (<https://dna.macrogen.com/>, Seoul, South Korea), with a final yield raw data of 3.5 Gb.

Genome assembly and annotation

The FastQC tool was used to check raw read quality and remove adaptors. A Phred score above 30 was used. Clean reads were processed for genome assembly using NOVOPlasty 4.3.0 version [80] with kmer (K-mer = 33). The contig N50 value was high, and the plastome was assembled using the whole genomic sequence of *B. aegyptiaca*. *Tribulus terrestris* (NC_046758.1) was used as a reference in the assembly. Single contigs containing the plastome were generated. Gene prediction and annotation of the *B. aegyptiaca* chloroplast genome were carried out using the GeSeq tool [81], with default parameters and percent identity cut-off for protein-coding genes and RNAs set at ≥ 60 and ≤ 85 , respectively. tRNA genes were identified with trnAscan-SE version 2.0 [82]. The annotated (gb) format sequence files were used to draw the circular chloroplast genome maps with the OGDRAW tool (Organellar Genome DRAW), version 1.3.1 [83]. The sequence of the chloroplast genome of *B. aegyptiaca* was deposited in the GenBank database with accession number (OL703321).

Sequence Analysis

The relative synonymous codon usage (RSCU) values, base composition, and codon usage were analysed using MEGA software [84] version 11.0. Potential RNA editing sites present in the protein-coding genes were predicted by the PREP suite [85] with a cut-off value of 0.8.

Repeat Analysis in the chloroplast Genome

The online software MIcroSatellite (MISA) v2.1 [86] was used to identify simple sequence repeats (SSRs) in the chloroplast genome of *B. aegyptiaca* and five other species from the Zygophyllaceae family, namely, *Guaia-cum angustifolium*, *Larrea tridentata*, *Tetraena mongolica*, *Tribulus terrestris*, and *Zygophyllum xanthoxylon*.

Parameters eight, five, four and three repeats units were used: eight repeats for mononucleotide, five for dinucleotides, four for trinucleotides, three for each tetranucleotides, pentanucleotides, and hexanucleotides respectively.

In addition, REPuter [85] software was used with default settings to detect the size and location of long palindromic, forward, reverse, and complementary repeats in the *B. aegyptiaca* cp genome and the genomes of five species from Zygophyllaceae.

Sequence divergence and boundary

Comparison of the genome of *B. aegyptiaca* with five chloroplast genome sequences from Zygophyllaceae (*G. angustifolium*, *L. tridentata*, *T. mongolica*, *T. terrestris*, and *Z. xanthoxylon*; GenBank accession numbers are shown in Table 4) was performed using the mVISTA program [87]; the annotation of *B. aegyptiaca* was used as a reference in the Shuffle-LAGAN mode. Furthermore, comparisons between the borders of the IR, SSC, and LSC regions were generated using IRSCOPE [88].

Characterisation of the substitution rate

The methods for estimating nonsynonymous and synonymous substitution rates (Ka and Ks), selection, and

beneficial mutations among protein-coding sequences followed. Nonsynonymous (Ka) substitution, synonymous (Ks) substitution, and Ka/Ks ratios were calculated to detect variable mutation rates across chloroplast genome sequence, that contain an important information related to evolutionary history in *B. aegyptiaca* compared with those in the five aforementioned Zygophyllaceae species. We employed Ka/Ks Calculator version 2.0 [75] with default parameters, and the Nei and Gojobori substitution model was used.

Phylogenetic Analysis

Phylogenetic analysis was conducted based on the cp genome sequences of the members of the Zygophyllales order, including cp genome sequences available in GenBank of species from the families Zygophyllaceae and Krameriaceae.

Because an older taxonomy placed *Balanites aegyptiaca* in the orders Sapindales and Santalales, the current phylogenetic analysis included sequesters of the cp genome of some species belonging to these orders. Information on species names, families, and GenBank accessions is available in Table 4. Two species from the Malvaceae family (*Malva parviflora* and *Malva wigandii*) were used as an outgroup.

Table 4 Accession numbers of chloroplast genome analysed in the study

Orders	Family	Accession Number	Organism		
Zygophyllales	Zygophyllaceae	OL703321	<i>Balanites aegyptiaca</i> (L.) Delile		
		MK265246.1	<i>Tetraena mongolica</i> Maxim. synonym of <i>Zygophyllum mongolicum</i> (Maxim.) Christenh. & Byng		
		MT796492.1	<i>Zygophyllum xanthoxylon</i> (Bunge) Maxim.		
		NC_028023.1	<i>Larrea tridentata</i> (DC.) Coville		
		NC_043796.1	<i>Guaiaacum angustifolium</i> Engelm.		
		NC_046758.1	<i>Tribulus terrestris</i> L.		
		NC_043800.1	<i>Krameria bicolor</i> S.Watson		
		NC_043801.1	<i>Krameria lanceolata</i> Torr.		
		Santalales	Olacaceae	MK764537.1	<i>Malania oleifera</i> Chun & S.K.Lee
			Schoepfiaceae	NC_034228.1	<i>Schoepfia jasminodora</i> Siebold & Zucc.
Sapindales	Anacardiaceae	NC_035973.1	<i>Spondias mombin</i> L.		
	Burseraceae	NC_036978.1	<i>Commiphora wightii</i> (Arn.) Bhandar		
	Meliaceae	NC_023792.1	<i>Azadirachta indica</i> A.Juss.		
	Nitrariaceae	MH457633.1	<i>Nitraria tangutorum</i> Bobrov		
		MW307830.1	<i>Peganum harmala</i> L.		
	Rutaceae	NC_024929.1	<i>Citrus aurantiifolia</i> (Christm.) Swingle		
	Sapindaceae	NC_040106.1	<i>Acer sino-oblongum</i> F.P.Metcalf		
	Simaroubaceae	NC_030482.1	<i>Leitneria floridana</i> Chapm.		
NC_037696.1		<i>Ailanthus altissima</i> (Mill.) Swingle			
Outgroup	Malvaceae	MK860036.1	<i>Malva parviflora</i> L.		
		NC_049129.1	<i>Malva wigandii</i> (Alef) M.F.Ray		

All the common genes from the organisms were retrieved and the coding sequence CDS were joined in order. The sequences were further aligned using MAFFT version 7.475 [89]. Then phylogenetic trees were reconstructed based on the maximum parsimony (MP) method using 1000 bootstrap values and Mega software [84] version 11.0. The MP search method consisted of subtree-pruning-regrafting, with the number of initial trees (random addition) set at 10 and the number of threads set at 5.

The optimal evolutionary model was the GTR+I+G model, as calculated by ModelFinder using [90] Akaike's information criterion (AIC) [91]. Bayesian inference analysis (BI) was conducted using MrBayes v. 3.2.6 [92, 93] in CIPRES Science Gateway 3.3 [94]. BI analysis included two separate runs; each of four Markov chain Monte Carlo chains was run for 10 million generations with sampling every 10,000 generations. Trees from the first 25% of the sampled generations were discarded as burn-in. The convergence of the runs was tested by using the effective sample size (ESS), calculated with Tracer v1.7.1 [95], with ESS values greater than 200 for all parameters considered good evidence. The majority rule (> 50%) consensus tree of BI was visualised using FigTree 1.4.3 [96].

Conclusions

The aim of the present research was to provide the complete chloroplast genome of *B. aegyptiaca*, a plant in subfamily Tribuloideae and family Zygophyllaceae, which has medical and nutritional importance and plays a key role in ecosystem conservation in arid lands. We compared the cp genomes of available genera and species in the Zygophyllaceae family to assess the systematic relationships within the family and between related families as well as the genome conservation state.

This study confirmed the taxonomic status of the species *B. aegyptiaca* as a member of the Zygophyllaceae family. This study did not record loss in the chloroplast genome of *B. aegyptiaca*, as mentioned for species in subfamily Zygophylloideae. However, plastomes vary mainly at the SC/IR boundary among members of Zygophyllaceae, and there is clear genome reduction and gene loss in some species of Zygophyllaceae. Thus, we recommended further study to investigate changes that could have occurred in the structure of the genome during the evolutionary history of the family. It is necessary that future studies also include samples from subfamilies Morkillioideae, Seetzenioideae and Larreoideae in Zygophyllaceae.

Abbreviations

Cp: Chloroplast; LSC: Large single copy region; SSC: Small single copy region; IR: Inverted repeat; RSCU: Relative synonymous codon usage; SSR: Simple sequence repeats; IGS: Intergenic spacer; CNS: Conserved non coding

sequence; cpSSRs: Chloroplast simple sequence repeats; RAPD: Random Amplified Polymorphic DNA; MP: Maximum Parsimony; BP: Bootstrap percentage; BI: Bayesian Inference; PP: Posterior probability; Ks: Synonymous; Ka: Non-synonymous; CTAB: Cetrimonium bromide; PCR: Polymerase chain reaction.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08850-9>.

Additional file 1.

Acknowledgments

Not applicable.

Authors' contributions

WAJ and SAA collected the data, designed and performed the experiment, WAJ, SAA, NMA, and AYA analyzed the data, investigation and drafted the manuscript, WAJ supervised the project, all the authors edited and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Availability of data and materials

The data presented in this study are available in this article and Supplementary Material. The complete chloroplast genome sequence of *Balanites aegyptiaca* was deposited in GenBank at <https://www.ncbi.nlm.nih.gov>, (accession numbers: OL703321).

Declarations

Ethics approval and consent to participate

The experiment was conducted with relevant institutional, national, and international guidelines and legislation. Permission was obtained to collect the samples.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interest.

Author details

¹Department of Biology, Faculty of Applied Science, Umm Al-Qura University, Makkah 24381, Saudi Arabia. ²Research Laboratories Centre, Faculty of Applied Science, Umm Al-Qura University, Makkah 24381, Saudi Arabia.

Received: 9 January 2022 Accepted: 18 August 2022

Published online: 31 August 2022

References

- Godoy-Bürki AC, Acosta JM, Aagesen L. Phylogenetic relationships within the New World subfamily Larreoideae (Zygophyllaceae) confirm polyphyly of the disjunct genus *Bulnesia*. *Syst Biodivers*. 2018;16(5):453–68 <https://www.tandfonline.com/doi/abs/10.1080/14772000.2018.1451406>.
- Al-Thobaiti SA, Abu ZI. Medicinal properties of desert date plants (*Balanites aegyptiaca*)-an overview. *Glob J Pharmacol*. 2018;12(1):01–12.
- Orwa C, Mutua A, Kindt R, Jamnadass R, Anthony S. Agroforestry database: a tree reference and selection guide; version 4.0. World Agroforestry Centre, Kenya; 2009.
- Chothani DL, Vaghasiya H. A review on *Balanites aegyptiaca* Del (desert date): phytochemical constituents, traditional uses, and pharmacological activity. *Pharmacogn Rev*. 2011;5(9):55–62. <https://doi.org/10.4103/0973-7847.79100>.
- Abdelaziz SM, Medraoui L, Alami M, Pakhrou O, Makkaoui M, Mohamed AO, et al. Inter simple sequence repeat markers to assess genetic diversity of the

- desert date (*Balanites aegyptiaca* Del.) for Sahelian ecosystem restoration. *Sci Rep.* 2020;10(1):1–8 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7486294/>.
6. Gardette J-L, Baba M. FTIR and DSC studies of the thermal and photochemical stability of *Balanites aegyptiaca* oil (Toogga oil). *Chem Phys Lipids.* 2013;170(171):1–7. <https://doi.org/10.1016/j.chemphyslip.2013.02.008>.
 7. Abdoulaye B, Béchir AB, Mapongmetsem PM. Morphological variability of *Balanites aegyptiaca* (L.) Del. In the region of Ouaddai in Chad Int. *J. Biol. Chem. Sci.* 2016;10(4):1733–46 <http://ajol.info/index.php/ijbcs>.
 8. Thomas J. Plant Diversity in Saudi Arabia. <http://www.plantdiversityofsaudiArabia.info/index.htm>. 2011. Accessed 10 Sept 2021.
 9. Sands MJS. The desert date and its relatives: a revision of the genus *Balanites*. *Kew Bull.* 2001;56(1):1–128 <http://www.jstor.org/stable/4119431>.
 10. Hall JB, Walker DH. *Balanites aegyptiaca*: a monograph. Bangor: School of Agricultural and Forest Sciences, University of Wales; 1991.
 11. Harms HAT. Vorschlag zur Erganzung der "Lois de la nomenclature botanique de 1867". *Notizbl Konigl Bot Gart Berlin-Dahlem, App.* 1904;13:1–37.
 12. Bentham G, Hooker JD. *Simarubaceae*. In: *Genera Plantarum*. London; Lovell Reeve & Co., London, 1862, 1(1).
 13. Engler HGA. *Zygophyllaceae*. In: Engler HGA, Prantl K, Die Natfirliehen Pflanzenfamilien, Leipzig; W Engelmann 1896: , 3(4).
 14. Cronquist A. Evolution and classification of flowering plants. London & Edinburgh. 1968.
 15. Hegnauer R. *Chemotaxonomy der pflanzen*. Basel und Stuttgart: Birkhuser Verlag; 1973. p. 6.
 16. Scholz H. *Zygophyllaceae*. In: Melchior, H, 12th edition (ed). A. Vol. 2. Berlin: Engler's Syllabus der Pfannenfamilien; 1964: 251–252c.
 17. Cronquist AJ. An integrated system of classification of flowering plants. New York, Columbia university press, 1981.
 18. Maksoud SA, El-Hadidi MN. The flavonoids of *Balanites aegyptiaca* from Egypt. *Plant Syst Evol.* 1988;160:153–8.
 19. Sheahan MC, Culter DF. Contribution of vegetation anatomy to systematics of the Zygophyllaceae r.br. *Bot J Linn Soc.* 1993;113:227–62.
 20. Boesewinkel FD. Ovule and seed characters of *Balanites aegyptiaca* and the classification of the Linales-Geraniales-Polygalales assembly. *Acta Bot Neerl.* 1994;43(1):15–25. <https://doi.org/10.1111/j.1438-8677.1994.tb00730.x>.
 21. Sheahan MC, Chase MW. Phylogenetic relationships within Zygophyllaceae based on DNA sequences of three plastid regions, with special emphasis on Zygophylloideae. *Syst Bot.* 2000;25(2):371–84. <https://doi.org/10.2307/2666648>.
 22. Singh KK, Samanta AK, Kundu SS, Sharma D. Evaluation of certain feed resources for carbohydrate and protein fractions and in situ digestion characteristics. *Indian J Anim Sci.* 2002;72(9):794–7.
 23. Sands MJS. Flora of tropical East Africa: Balanitaceae. In: Beentje HJ, editor. and Ghazanfar S.a. (subed.). *Royal Botanic Gardens, Kew: Flora of tropical East Africa*; 2013. p. 1–17.
 24. Sheahan MC, Chase MW. A phylogenetic analysis of Zygophyllaceae R. Br. Based on morphological, anatomical and rbc L DNA sequence data. *Bot J Linn Soc.* 1996;122(4):279–300. <https://doi.org/10.1111/j.1095-8339.1996.tb02077.x>.
 25. Boulos L. *Flora of Egypt, checklist*. Cairo: Al-Hadara Publishing; 2000.
 26. Sands MJS. Balanitaceae, in *Flora of Ethiopia and Eritrea* S. Edwards, M. Tadesse, and I. Hedberg, editors. National Herbarium, Addis Ababa; Department of Systematic Botany: Uppsala University; 1989.
 27. Amer WM, Soliman MM, Sheded MM. Biosystematics studies for *Balanites aegyptiaca* (Balanitaceae) populations in Egypt. *Flora Mediterr.* 2002;12:353–67.
 28. Palmer JD. Comparative organization of chloroplast genomes. *Annu Rev Genet.* 1985;19:325–54. <https://doi.org/10.1146/annurev.ge.19.120185.001545>.
 29. Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am J Bot.* 2007;94(3):275–88. <https://doi.org/10.3732/ajb.94.3.275> PMID 21636401.
 30. Wicke S, Schneeweiss GM, Depamphilis CW, Muller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol.* 2011;76:273–97. <https://doi.org/10.1007/s11103-011-9762-4>.
 31. Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *Am J Bot.* 2012;99:349–64.
 32. Liu HJ, Ding CH, He J, Cheng J, Pei LY, Xie L. Complete chloroplast genomes of *Archiclematis*, *Naravelia* and *Clematis* (Ranunculaceae), and their phylogenetic implications. *Phytotaxa.* 2018;343:214–26.
 33. Sage RF, Christin P-A, Edwards EJ. The C4 plant lineages of planet earth. *J Exp Bot.* 2011;62(9):3155–69 <https://academic.oup.com/jxb/article/62/9/3155/474202>.
 34. Bellstedt DU, Galley C, Pirie MD, Linder HP. The migration of the palaeotropical arid flora: Zygophylloideae as an example. *Syst Bot.* 2012;37:951–9.
 35. Wu SD, Lin L, Li HL, Yu SX, Zhang LJ, Wang W. Evolution of Asian interior arid-zone biota: evidence from the diversification of Asian *Zygophyllum* (Zygophyllaceae). *PLoS One.* 2015;10(9):e0138697. <https://doi.org/10.1371/journal.pone.0138697>.
 36. Wang X, Dorjee T, Chen Y, Gao F, Zhou Y. The complete chloroplast genome sequencing analysis revealed an unusual IRs reduction in three species of subfamily Zygophylloideae. *Plos One.* 2022;17(2):e0263253. <https://doi.org/10.1371/journal.pone.0263253>.
 37. Sanderson MJ, Copetti D, Burquez A, Bustamante E, Charboneau JLM, Eguiarre LE, et al. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): loss of the ndh gene suite and inverted repeat. *Am J Bot.* 2015;102(7):1115–27. <https://doi.org/10.3732/ajb.1500184>.
 38. Lei WJ, Ni DP, Wang YJ, Shao JJ, Wang XC, Yang D, et al. Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of *Astragalus membranaceus*. *Sci Rep.* 2016;6:21669. <https://doi.org/10.1038/srep21669> PMID: 26899134.
 39. Wicke S, Muller KF, Claude WD, Quandt D, Bellot S, Schneeweiss GM. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proc Natl Acad Sci U S A.* 2016;113:9045–50. <https://doi.org/10.1073/pnas.1607576113>.
 40. Wicke S, Naumann J. Molecular evolution of plastid genomes in parasitic flowering plants. *Adv Bot Res.* 2018;85:315–47. <https://doi.org/10.1016/bs.abr.2017.11.014>.
 41. Bellot S, Cusimano N, Luo S, Sun G, Zarre S, Groger A, et al. Assembled plastid and mitochondrial genomes, as well as nuclear genes, place the parasite family Cynomoriaceae in the Saxifragales. *Genome Biol. Evol.* 2016;8(7):2214–30. <https://doi.org/10.1093/gbe/evw147>.
 42. Staden R, McLachlan AD. Codon preference and its use in identifying protein coding regions in long DNA sequences. *Nucleic Acids Res.* 1982;10:141–56. <https://doi.org/10.1093/nar/10.1.141> PMID: 7063399.
 43. Li B, Lin F, Huang P, Guo W, Zheng Y. Complete chloroplast genome sequence of *Decaisnea insignis*: Genome organization, genomic resources and comparative analysis. *Sci Rep.* 2017;7(1):1–10. <https://doi.org/10.1038/s41598-017-10409-8>.
 44. Zong D, Zhang Y, Zou X, Li D, Duan A, He C. Characterization of the complete chloroplast genomes of five *Populus* species from the western Sichuan plateau, Southwest China: comparative and phylogenetic analyses. *PeerJ.* 2019;7:e6386. <https://doi.org/10.7717/peerj.6386>.
 45. Mattick JS, Gagen MJ. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol Biol Evol.* 2001;18:1611–30. <https://doi.org/10.1093/oxfordjournals.molbev.a003951> PMID: 11504843.
 46. Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, et al. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* 2007, 35(3):e14. doi: <https://doi.org/10.1093/nar/gkl938>. Epub 2006 Dec 14. PMID: 17169982; PMCID: PMC1807943.
 47. Raman G, Park S. The complete chloroplast genome sequence of *Ampelopsis*: gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. *Front Plant Sci.* 2016;7:341. <https://doi.org/10.3389/fpls.2016.00341>.
 48. Alzahrani D, Albokhari E, Yaradua S, Abba A. Complete chloroplast genome sequences of *Dipterygium glaucum* and *Cleome chrysantha* and other *Cleomaceae* species, comparative analysis and phylogenetic relationships. *Saudi J Biol Sci.* 2021;28(4):2476–90. <https://doi.org/10.1016/j.sjbs.2021.01.049>.
 49. Mower JP. The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res.* 2009;37:253–9. <https://doi.org/10.1093/nar/gkp337>.
 50. Bundschuh R, Altmuller J, Becker C, Nurnberg P, Gott JM. Complete characterization of the edited transcriptome of the mitochondrion of *Physarum polycephalum* using deep sequencing of RNA. *Nucleic Acids Res.* 2011;39:6044–55.

51. Smith DR. Unparalleled variation in RNA editing among Selaginella Plastomes. *Plant Physiol.* 2020;182(1):12–4. <https://doi.org/10.1104/pp.19.00904>.
52. Wang W, Yu H, Wang J, Lei W, Gao J, Qiu X, et al. The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (Oleaceae). *Int J Mol Sci.* 2017;18(11):2288. <https://doi.org/10.3390/ijms18112288>.
53. Park M, Park H, Lee H, Lee B, Lee J. The complete plastome sequence of an Antarctic bryophyte *Sanionia uncinata* (Hedw.) Loeske. *Int J Mol Sci.* 2018;19:709. <https://doi.org/10.3390/ijms19030709>.
54. Asano T, Tsudzuki T, Takahashi S, Shimada H, Ki K. Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. *DNA Res.* 2004;11(2):93–9. <https://doi.org/10.1093/dnares/11.2.93>.
55. Weng ML, Ruhlman TA, Jansen RK. Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. *New Phytol.* 2017;214(2):842–51. <https://doi.org/10.1111/nph.14375>.
56. Yang Y, Zhou T, Duan D, Yang J, Feng L, Zhao G. Comparative Analysis of the complete chloroplast genomes of five *Quercus* species. *Front Plant Sci.* 2016;7(959) <https://www.frontiersin.org/article/10.3389/fpls.2016.00959>.
57. Zhou T, Ruhsam M, Wang J, Zhu H, Li W, Zhang X, et al. The complete chloroplast Genome of *Euphrasia regelii*, Pseudogenization of *ndh* genes and the Phylogenetic Relationships within Orobanchaceae. *Front Genet.* 2019;10(444):1–15 <https://www.frontiersin.org/article/10.3389/fgene.2019.00444>.
58. Provan J, Powell W, Hollingsworth PM. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol Evol.* 2001;16(3):142–7. [https://doi.org/10.1016/s0169-5347\(00\)02097-8](https://doi.org/10.1016/s0169-5347(00)02097-8).
59. Yang A-H, Zhang J-J, Yao X-H, Huang H-W. Chloroplast microsatellite markers in *Liriodendron tulipifera* (Magnoliaceae) and cross-species amplification in *L. chinense*. *Am J Bot.* 2011;98(5):e123–6 <https://bsapubs.onlinelibrary.wiley.com/doi/abs/10.3732/ajb.1000532>.
60. Xue J, Wang S, Zhou S-L. Polymorphic chloroplast microsatellite loci in *Nelumbo* (Nelumbonaceae). *Am J Bot.* 2012;99(6):e240–4 <https://bsapubs.onlinelibrary.wiley.com/doi/abs/10.3732/ajb.1100547>.
61. Hu Y, Woeste KE, Zhao P. Completion of the chloroplast genomes of five Chinese Juglans and their contribution to chloroplast phylogeny. *Front Plant Sci.* 2016;7(1955) <https://www.frontiersin.org/article/10.3389/fpls.2016.01955>.
62. Ruhsam M, Clark A, Finger A, Wulff AS, Mill RR, Thomas PI, et al. Hidden in plain view: cryptic diversity in the emblematic *Araucaria* of New Caledonia. *Am J Bot.* 2016;103(5):888–98. <https://doi.org/10.3732/ajb.1500487>.
63. Wang RJ, Cheng CL, Chang CC, Wu C-L, Su T-M, Chaw S-M. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol Biol.* 2008;8(1):3650. <https://doi.org/10.1186/1471-2148-8-36>.
64. Alzahrani DA, Yaradua SS, Albokhari EJ, Abba A. Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. *BMC Genomics.* 2020;21:393. <https://doi.org/10.1186/s12864-020-06798-2>.
65. Yao X, Tan Y-H, Liu Y-Y, Song Y, Yang J-B, Corlett RT. Chloroplast genome structure in *Ilex* (Aquifoliaceae). *Sci Rep.* 2016;6:28559.
66. Rauberson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics.* 2007;8(1):174. <https://doi.org/10.1186/1471-2164-8-174>.
67. Braukmann TWA, Kuzmina M, Stefanović S. Loss of all plastid *ndh* genes in Gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny. *Curr Genet.* 2009;55:323–37.
68. Dong WL, Wang RN, Zhang NY, Fan WB, Fang MF, Li Z-H. Molecular evolution of chloroplast genomes of orchid species: insights into Phylogenetic relationship and adaptive evolution. *Int J Mol Sci.* 2018;19:716. <https://doi.org/10.3390/ijms19030716> PMID: 29498674.
69. Friedrich T, Steinmüller K, Weiss H. The proton-pumping respiratory complex I of bacteria and mitochondria and its homologue in chloroplasts. *FEBS Lett.* 1995;367:107–11. [https://doi.org/10.1016/0014-5793\(95\)00548-n](https://doi.org/10.1016/0014-5793(95)00548-n).
70. Petersen G, Cuenca A, Seberg O. Plastome evolution in Hemiparasitic mistletoes. *Genome Biol Evol.* 2015;7:2520–32. <https://doi.org/10.1093/gbe/evw165>.
71. Shin HW, Lee NS. Understanding plastome evolution in Hemiparasitic Santalales: complete chloroplast genomes of three species, *Dendrotrophe varians*, *Helixanthera parasitica*, and *Macrosolen cochinchinensis*. *PLoS One.* 2018;13. <https://doi.org/10.1371/journal.pone.0200293>.
72. Zhu ZX, Wang JH, Cai YC, Zhao KK, Moore MJ, Wang HF. Complete plastome sequence of *Erythralpalum scandens* (Erythralpalaceae), an edible and medicinally important liana in China. *Mitochondr DNA Part B.* 2018;3:139–40. <https://doi.org/10.1080/23802359.2017.1413435>.
73. Su HJ, Hu JM. The complete chloroplast genome of hemiparasitic flowering plant *Schoepfia jasminodora*. *Mitochondr DNA Part B.* 2016;1:767–9. <https://doi.org/10.1080/23802359.2016.1238753>.
74. Fraile DC, Chaluvadi SR, Vaughn JN, Goatney G, Bennetzen JL. Gene loss and genome rearrangement in the plastids of five Hemiparasites in the family Orobanchaceae. *BMC Plant Biol.* 2018;18:30. <https://doi.org/10.1186/s12870-018-1249-x>.
75. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *GPB.* 2010;8(1):77–80. [https://doi.org/10.1016/S1672-0229\(10\)60008-3](https://doi.org/10.1016/S1672-0229(10)60008-3).
76. Boudreau E, Takahashi Y, Lemieux C, Turmel M, Rochaix JD. The chloroplast *ycf3* and *ycf4* open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. *EMBO J.* 1997;16(20):6095–6104. doi:<https://doi.org/10.1093/emboj/16.20.6095>. PMID 1326293. PMID 9321389.
77. Yaradua SS, Alzahrani DA, Albokhari EJ, Abba A, Bello A. Complete Chloroplast Genome Sequence of *Justicia flava*: Genome Comparative Analysis and Phylogenetic Relationships among Acanthaceae. *Biomed Res. Int.* 2019; 17 pages, Article ID 4370258. <https://doi.org/10.1155/2019/4370258>.
78. Dong W, Xu C, Li W, Xie X, Lu Y, Liu Y, et al. Phylogenetic resolution in *Juglans* based on complete chloroplast genomes and nuclear DNA sequences. *Front Plant Sci.* 2017;8:1148. <https://doi.org/10.3389/fpls.2017.01148>.
79. Doyle J, Doyle J. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 1987;19(1):11–5.
80. Dierckxens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 2016. <https://doi.org/10.1093/nar/gkw955>.
81. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 2017;45(W1):W6–W11 <https://chlorobox.mpimp-golm.mpg.de/geseq.html>.
82. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. In: Kollmar M, editor. *Gene prediction: methods and protocols.* New York: Springer; 2019. p. 1–14.
83. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organelle genomes. *Nucleic Acids Res.* 2019;47(W1):W59–64. <https://doi.org/10.1093/nar/gkz238>.
84. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33(7):1870–4. <https://doi.org/10.1093/molbev/msw054>.
85. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 2001;29:4633–42. <https://doi.org/10.1093/nar/29.22.4633>.
86. Thiel T, Michalek W, Varshney R. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet.* 2003;106:411–22.
87. Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, et al. VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics.* 2000;16:1046. <https://doi.org/10.1093/bioinformatics/16.11.1046>.
88. Amiryousefi A, Hyvönen J, Poczar P. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics.* 2018;34(17):3030–1. <https://doi.org/10.1093/bioinformatics/bty220>.
89. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* 2022;50(W1):W276–9. <https://doi.org/10.1093/nar/gkac240> Online ahead of print.
90. Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler AV, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14(6):587–9. <https://doi.org/10.1038/nmeth.4285> <https://doi.org/10.1038/nmeth.4285>.

91. Akaike H. A new look at the statistical model identification. *IEEE*. 1974;19(6):716–23. <https://doi.org/10.1109/TAC.1974.1100705>.
92. Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*. 2001;17:754–5.
93. Ronquist F, Huelsenbeck JP. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19:1572–4. <https://doi.org/10.1093/bioinformatics/btg180>.
94. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES science gateway for inference of large phylogenetic trees. In: 2010 gateway computing environments workshop (GCE): IEEE; 2010.
95. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian Phylogenetics using tracer 1.7. *Syst Biol*. 2018;67(5):901–4. <https://doi.org/10.1093/sysbio/syy032>.
96. Rambaut A, Drummond A. FigTree v. 1.4.0; 2012.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

