

UNVEILING THE COMPLETE CHLOROPLAST GENOME OF *TRIBULUS MACROPTERUS* VAR. *ARABICUS* (HOSNI) AL-HEMAID & J. THOMAS: GENOME STRUCTURE, COMPARATIVE ANALYSIS AND PHYLOGENY

ALBATOOL SULIMAN ALBEDIWI¹, MOHAMMAD AJMAL ALI^{1*}, MONA SOLAIMAN ALWAHIBI¹, SHEIKH SUNZID AHMED², M. OLIUR RAHMAN^{2*}, SOO-YONG KIM³, MOHAMED S. ELSHIKH¹ AND NADIA MOHAMMAD ALSUHAIMI¹

¹Department of Botany and Microbiology, College of Science, King Saud University, Riyadh-11451, Saudi Arabia

²Department of Botany, Faculty of Biological Sciences, University of Dhaka, Dhaka 1000, Bangladesh

³International Biological Material Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 34141, Republic of Korea

Keywords: Chloroplast genome assembly; *Tribulus macropterus* var. *arabicus*; Simple sequence repeats; Nucleotide diversity; Phylogenetics; Zygophyllaceae.

Abstract

The present investigation unveiled the first complete chloroplast (cp) genome of *Tribulus macropterus* var. *arabicus* (Hosni) Al-Hemaid & J. Thomas (Zygophyllaceae), a medicinal herb, indigenous to Saudi Arabia. The cp genome, comprising a length of 158,179 bp and a GC content of 35.8%, exhibited the typical circular quadripartite arrangement of flowering plants, including two inverted repeat regions (25842 bp) separated by a large single-copy (88873 bp) and a small single-copy region (17622 bp). Genome annotation unraveled 132 genes, comprising of 87 protein-coding genes, 37 tRNAs and eight rRNAs. A comparative plastomics approach demonstrated a very similar genome structure and gene organization in closely related taxa. Whole-genome alignment indicated that the inverted repeat regions exhibited greater conservation when compared to the single copy zones. Repeat analysis of the cp genome identified 80 simple sequence repeats, with the majority (64) being mononucleotides. Among the longer repeats, forward repeats were frequent (20) followed by palindromic and reverse repeats. The nucleotide diversity endeavor identified some hypervariable sites (*rpl22*, *cemA*, *trnL-UAG*) in the small and large single copies which would offer opportunities to design molecular markers for potential application in taxonomic identification. Phylogenetic analysis with *rbcL* barcode elucidated the distinct position of *T. macropterus* var. *arabicus* compared to *T. macropterus* within Zygophyllaceae and further validated the assembly. The findings of this investigation provide significant insights into biological fields, particularly enhancing the current understanding of the genetic and evolutionary variations within Zygophyllaceae.

Introduction

The family Zygophyllaceae *sensu lato* predominantly occurs in the arid and semi-arid regions within the tropics and subtropics (Sheahan and Chase, 2000). In Saudi Arabia, this family consists of six genera: *Balanites* Delile, *Fagonia* L., *Seetzenia* R. Br., *Tetraena* Maxim., *Tribulus* L., and *Zygophyllum* L. (Chaudhary, 2001). The genus *Tribulus* is distinguished by its buttercup-like yellow flowers featuring 5 separate corolla, 5 calyx, and 5 or 10 stamens. The fruits are typically

*Corresponding authors, Email: alimohammad@ksu.edu.sa; oliur.bot@du.ac.bd

hard and dry, often round-shaped and adorned with wings, tubercles, or spines, occasionally displaying a mixture of these features. The genus *Tribulus* encompasses approximately 30 species globally. Within Saudi Arabia, *Tribulus* includes a total of nine taxa such as *T. macropterus* var. *arabicus* (Hosni) Al-Hemaid & J. Thomas, *T. macropterus* var. *macropterus* Boiss., *T. macropterus* var. *mollis* (Ehr. ex Sch.) Al-Hemaid & J. Thomas, *T. pentandrus* var. *bimucronatus* (Viv.) Al-Hemaid & J. Thomas, *T. pentandrus* var. *pentandrus* Forssk., *T. terrestris* var. *inermis* Boiss., *T. terrestris* var. *parvispinus* (Presl.) Al-Hemaid & J. Thomas, *T. terrestris* var. *rajasthanensis* (Bhad. & Shar.) Al-Hemaid & J. Thomas, and *T. terrestris* var. *terrestris* L. (Chaudhary, 2001).

T. macropterus Boiss. is an annual herb, non-succulent and xerophytic, thriving within the scorching desert sand dunes of Rub' al-Khali region in south-central Arabia. This species is employed to address sexual dysfunction and cardiac ailments. Investigations into the constituents of *T. macropterus* revealed the existence of different alkaloids and flavonoids. Research into its antihyperglycemic and antihyperlipidemic properties has yielded encouraging results (Yonbawi *et al.*, 2021). As a variety of *T. macropterus*, *T. macropterus* var. *arabicus* presents an enticing avenue for delving into its chloroplast (cp) genome to unveil significant properties and insights linked to its taxonomic identification, cp genome engineering and biological conservation.

The cp genome or plastome, with its unique characteristics and evolutionary dynamics, holds significant importance in phylogenetics. Serving as the powerhouse of photosynthesis and facilitating the synthesis of essential biomolecules such as amino acids and fatty acids, the chloroplast plays a pivotal role in plant biology. The Cp genome is maternally inherited in angiosperms but paternally in certain gymnosperms. Typically ranging from 107 kb to 218 kb in size, the plastome contains a suite of genes crucial for chloroplast function, such as (rRNAs, tRNAs, and protein-coding genes (PCGs) (Daniell *et al.*, 2016). Structured with a quadripartite architecture comprising two inverted repeat regions separating the large single-copy and small single-copy regions, the plastome undergoes dynamic rearrangements, including contractions, expansions, and even linearization, leading to variations in gene content and organization. Exploiting this inherent diversity, cp genome polymorphism has emerged as an important device for phylogenetic inference, taxonomic resolution, and understanding species adaptation to specific environments. By leveraging cp genome sequencing data, it becomes possible to conduct species barcoding, population genetics studies, and conservation efforts for medicinal species. Previous studies have successfully utilized complete cp genomes to develop robust molecular markers for phylogenetic reconstruction and species identification (Ahmed *et al.*, 2013; Nguyen *et al.*, 2018). These markers offer reliability and authenticity, facilitating a deeper understanding of plant evolution and biodiversity. Thus, the cp genome stands as a valuable resource in unraveling the intricate tapestry of plant phylogenetics and evolution.

Deciphering the complete cp genome of *T. macropterus* var. *arabicus*, native to the Saudi Arabian desert holds immense significance for both scientific understanding and practical applications. The unique environmental conditions of the desert, characterized by extreme temperatures, water scarcity, and high levels of salinity, have driven the evolution of specialized adaptations in this variety. This taxon demonstrates remarkable adaptability to harsh environmental conditions, including heat stress (Mandaville, 1986). Studying the chloroplast genome of various desert plants can uncover the genetic mechanisms underlying their resilience and adaptation strategies, offering valuable insights into plant stress tolerance and survival in arid environments (Eshel *et al.*, 2021). Moreover, understanding the genetic diversity and evolutionary history of desert flora can aid in conservation efforts, guiding the preservation of native species and their habitats amidst ongoing environmental challenges and human interventions. Furthermore, the identification of novel genetic resources within the cp genome may hold promise for biotechnological applications, such as the development of drought-resistant crops or the

production of bioactive compounds with pharmaceutical potential (Al-Juhani *et al.*, 2022; Alshegaihi, 2024). Thus, unraveling the chloroplast genome of Saudi Arabian desert plants, particularly *T. macropterus* var. *arabicus*, will enriches the current knowledge of desert ecosystems and offer practical avenues for sustainable agriculture and ecosystem management in arid regions.

Over the recent decades, the extensive utilization of Next-Generation Sequencing (NGS) has revolutionized research on desert plants (Shi *et al.*, 2013; Dickinson *et al.*, 2018; Eshel *et al.*, 2021; Zeng *et al.*, 2021). Despite these advancements, a notable gap persists in understanding the genetic mechanisms underpinning the environmental adaptability of *T. macropterus* var. *arabicus*. Furthermore, uncertainties linger regarding the taxonomic classification of this variety. Consequently, there is a pressing need to embark on a comprehensive sequencing effort targeting the cp genome of *T. macropterus* var. *arabicus*. Therefore, in this investigation, we aim to elucidate the genetic makeup of the cp genome using a comparative phylogenomic framework to resolve the taxonomic status and enhance the evolutionary understanding within Zygophyllaceae.

Materials and Methods

Plant sample collection, DNA isolation and sequencing

Tribulus macropterus var. *arabicus*, thriving in extremely hot desert conditions, was collected [Voucher: MAA 130 (KSUH)], Riyadh, Saudi Arabia (24°23'07.6"N 46°53'37.2"E). Upon collection, it was promptly transported to the laboratory and stored under controlled conditions at 4°C to preserve its integrity. Subsequently, the collected samples were desiccated in silica gel, and stored at -80° C until further use for *de novo* genome sequencing. The vouchers were deposited at the King Saud University Herbarium (KSUH). Identification of the taxon was confirmed with the aid of the Flora of Saudi Arabia (Chaudhary, 2001). Total genomic DNA was isolated from the collected leaf sample using the Qiagen DNA extraction Kit (QIAGEN Inc., Crawley, West Sussex, UK), followed by generation of the paired-end reads of 150 bp using a Novaseq 6000 sequencer (Illumina, San Diego, CA) at Macrogen, South Korea. The raw reads have been submitted to NCBI and are publicly accessible with the SRA accession ID “SRR29254160”. The corresponding BioProject and BioSample accessions are PRJNA1119014 and SAMN41632360, respectively.

Plastome assembly and annotation

The raw reads underwent filtration first employing the fastqc to procure high-quality clean data by eliminating adapter sequences and low-quality reads with a Q-value ≤ 20 . Subsequently, Unipro Ugene v.45.1 was employed to assemble the high-quality reads (Okonechnikov *et al.*, 2012). The assembled cp genome was annotated employing the GeSeq server (Tillich *et al.*, 2017). The annotated GenBank file of the plastome served as the basis for constructing a circular gene map using the Chloroplot server (Zheng *et al.*, 2020). The plastome assembled in this study was submitted to NCBI GenBank with the accession ID “OR750460”.

Repeat structures and codon usage analysis

The presence of repeat elements within the cp genomes of *T. macropterus* var. *arabicus* was assessed employing two distinct servers. The MICroSATellite identification tool (Beier *et al.*, 2017) was utilized for detecting SSRs (Simple sequence repeats), while the REPuter program (Kurtz *et al.*, 2001) was applied to detect longer repeat sequences. The RSCU (Relative synonymous codon usage) calculations were performed employing MEGA v.11 software (Tamura *et al.*, 2021).

Inverted repeat (IR) expansion and contraction analysis

Quadripartite junction sites and the genes located on the junction sites were analyzed utilizing the IRscope server (Amiryousefi *et al.*, 2018). The manually curated GenBank file of *T.*

macropterus var. *arabicus* was uploaded to the server while for other taxa, GenBank accessions (MK341055.1, NC_066813.1) were used for pursuing. Following the generation of the plot on the server, it was retrieved to assess the expansion and contraction of the inverted repeat regions.

Genome divergence evaluation

Two taxa were considered along with *T. macropterus* var. *arabicus*, such as *T. terrestris* L. and *Balanites aegyptiaca* (L.) Delile. with GenBank accession numbers MK341055.1 and NC_066813.1, respectively. The sequences were retrieved from NCBI and aligned using the mVISTA server. Shuffle-LAGAN mode was enabled before running the analysis in mVISTA (Frazer *et al.*, 2004).

Nucleotide diversity via sliding window technique

Initially, the cp genome sequences were aligned employing the MAFFT server (Katoh *et al.*, 2002). Afterwards, nucleotide variation was estimated using DnaSP v.5 software (Librado and Rozas, 2009). The parameters for window length and step size were set to 600 base pairs and 200 base pairs, respectively. Comparative analysis of genomic coordinates for each window was conducted against gene annotations of the cp genome to elucidate the characteristic features of nucleotide diversity indices.

Molecular phylogenetic analysis

For the molecular phylogenetic analyses, *rbcL* gene sequences of 10 taxa including outgroups, were retrieved from the NCBI GenBank (Table 1). *Morkillia mexicana* (DC.) Rose & Painter and *Sericodes greggii* A. Gray were used as outgroup taxa in this phylogenetic endeavor. The *rbcL* gene sequence of *T. macropterus* var. *arabicus* was identified from the GenBank annotation file and included in the data set. The sequence alignment was conducted using ClustalX v.1.81 (Thompson *et al.*, 1994), followed by the Maximum Parsimony (MP) analysis employing MEGA v.11 (Tamura *et al.*, 2021).



Fig. 1. Habit of *Tribulus macropterus* var. *arabicus*. A. Flowering twig, B. Fruiting twig.

Results and Discussion

Genome assembly and annotation

Total genomic DNA was isolated from fresh leaf materials (Fig. 1) and proceeded further for Illumina sequencing, which yielded approximately 5.7 GB of clean data, accounting for a total of 5,732,394,578 base pairs (bp) from a total of 37,962,878 raw reads. For the raw reads, GC content and AT content were 40% and 60%, respectively. Quality control analysis revealed 95.9% and 89.7% scores for the Q20 and Q30 parameters, respectively.

Table 1. Taxon and their *rbcL* gene sequences used in the phylogenetic analyses to infer taxonomic status of *Tribulus macropterus* var. *arabicus*.

No.	Taxon	GenBank accession
<i>Ingroup</i>		
1	<i>Balanites maughamii</i> Sprague	Y15016
2	<i>Kallstroemia maxima</i> (L.) Hook. & Arn.	Y15020
3	<i>Kelleronia revoilii</i> (Franch.) Chiov.	Y15021
4	<i>Neoluederitzia sericeocarpa</i> Schinz	Y15023
5	<i>Sisyndite sparteae</i> E. Mey. ex Sond.	Y15026
6	<i>Tribulus macropterus</i> Boiss.	Y15028
7	<i>Tribulopsis pentandra</i> R. Br.	AJ133860
8	<i>Tribulus terrestris</i> L.	MN525779
<i>Outgroup</i>		
9	<i>Morkillia mexicana</i> (DC.) Rose & Painter	AJ133857
10	<i>Sericodes greggii</i> A. Gray	AJ133859

Following the assembly of raw reads, the length of the cp genome was found to be 158,179 bp, which exhibited the typical quadripartite structure found in angiosperms, comprising a large single-copy (LSC) region (88873 bp), a small single-copy (SSC) region (17622 bp) and two inverted repeat (IR) regions (25842 bp each) (Fig. 2). In the assembled plastome, the GC content was found to be 35.80% and the base frequency was 31.67% (A), 32.54% (T), 18.26% (C) and 17.54% (G). In the LSC and SSC zones, the GC content was lower than that of the IR zones. Conversely, the AT content was higher in the LSC and SSC regions compared to the IR region (Table 2). The observed variation in GC content across the plastome of *T. macropterus* var. *arabicus* can be attributed to a combination of factors, including gene density, structural variation, and recombination dynamics. The higher GC content observed in the LSC and SSC regions than in the IR region, is probably attributable to the higher density of protein-coding genes in these regions.

Genes typically exhibit a higher GC content due to selection pressures favoring GC-rich codons for plastid-encoded proteins (Qian *et al.*, 2013). Furthermore, the SSC and LSC regions experience more frequent recombination events, leading to greater variability in nucleotide composition, including higher GC content. In contrast, the IR regions, characterized by their conserved sequence and structure, undergo fewer recombination events and thus maintain a more stable nucleotide composition (Saina *et al.*, 2018).

The plastome contained a comprehensive set of 132 genes, comprising 37 tRNAs (transfer RNA), 8 rRNAs (ribosomal RNA), and 87 protein-coding genes (PCGs) (Fig. 2). Within the PCGs, 44 were identified as photosynthesis-related, with 19 specifically linked to photosystem I and II functions. Among the different categories of genes, the highest number falls under protein genes (Fig. 3A). The functional emphasis on transcription and translation highlighted 76

genes, primarily comprising tRNAs. Among these, 26 genes were dedicated to ribosomal components, encompassing both small (15) and large (11) subunits. Gene duplication was notably prominent within tRNAs, while equal duplications were observed across all the rRNAs (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*) (Fig. 3B). In the inverted repeats region, tRNAs and rRNAs prevailed, while the SSC region exhibited a prevalence of NADH dehydrogenases. Two prominent DNA barcodes, such as *rbcL* and *matK*, were identified in the LSC region. In addition, the LSC also harbored the gene *cemaA* for encoding the cp envelope membrane protein. Photosystem I and II assembly factors (*pafI* and *pafII*) shared their positions in the LSC where *pafI* displayed a counter-clockwise direction and *pafII* showcased a clockwise direction for translation. Gene organization and genome structure were consistent with closely related species, *i.e.*, *T. terrestris* that reported the same number of tRNAs and rRNAs with a very similar number of PCGs (Yan *et al.*, 2019).

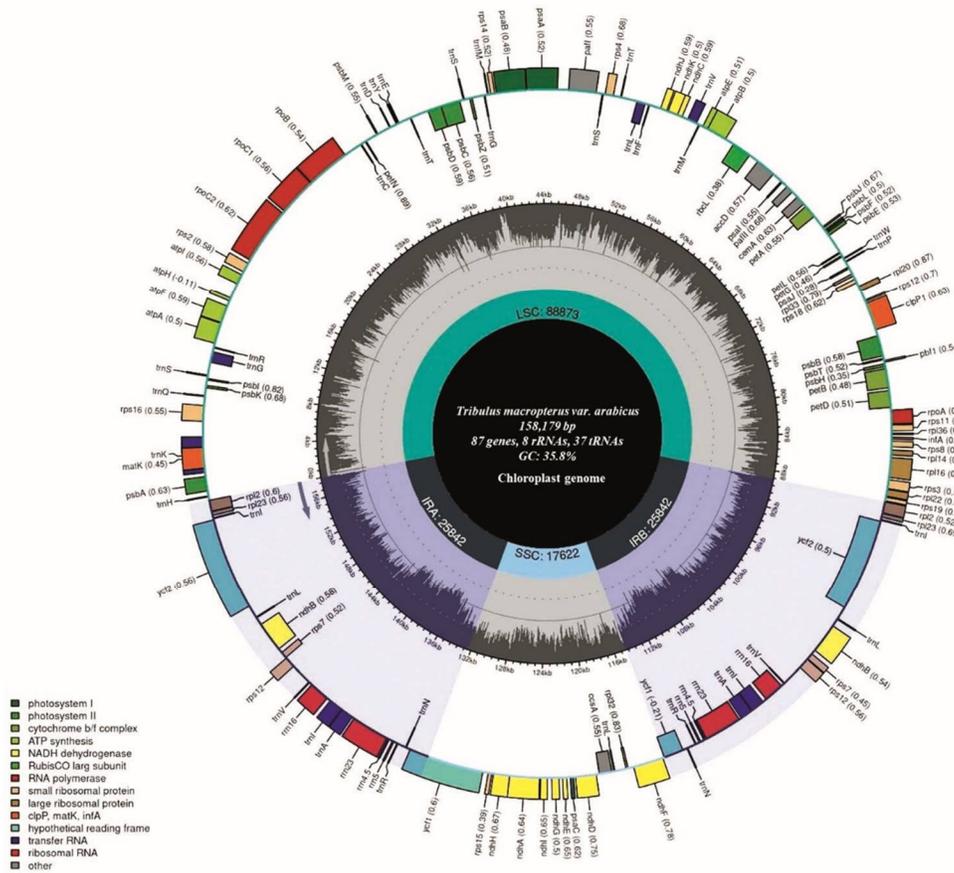


Fig. 2. Genome map of the complete chloroplast genome of *Tribulus macropteris* var. *arabicus*.

Repeat structures and codon usage analysis

A total of 80 SSRs (Simple Sequence Repeats) were identified in the plastome of *T. macropteris* var. *arabicus*. Among the six different types of repeats, only hexanucleotide repeats were absent (Fig. 4A). Comparative analysis also revealed a very similar organization of SSRs in the plastomes of *T. terrestris* and *B. aegyptiaca*, depicting frequent occurrence of mononucleotides

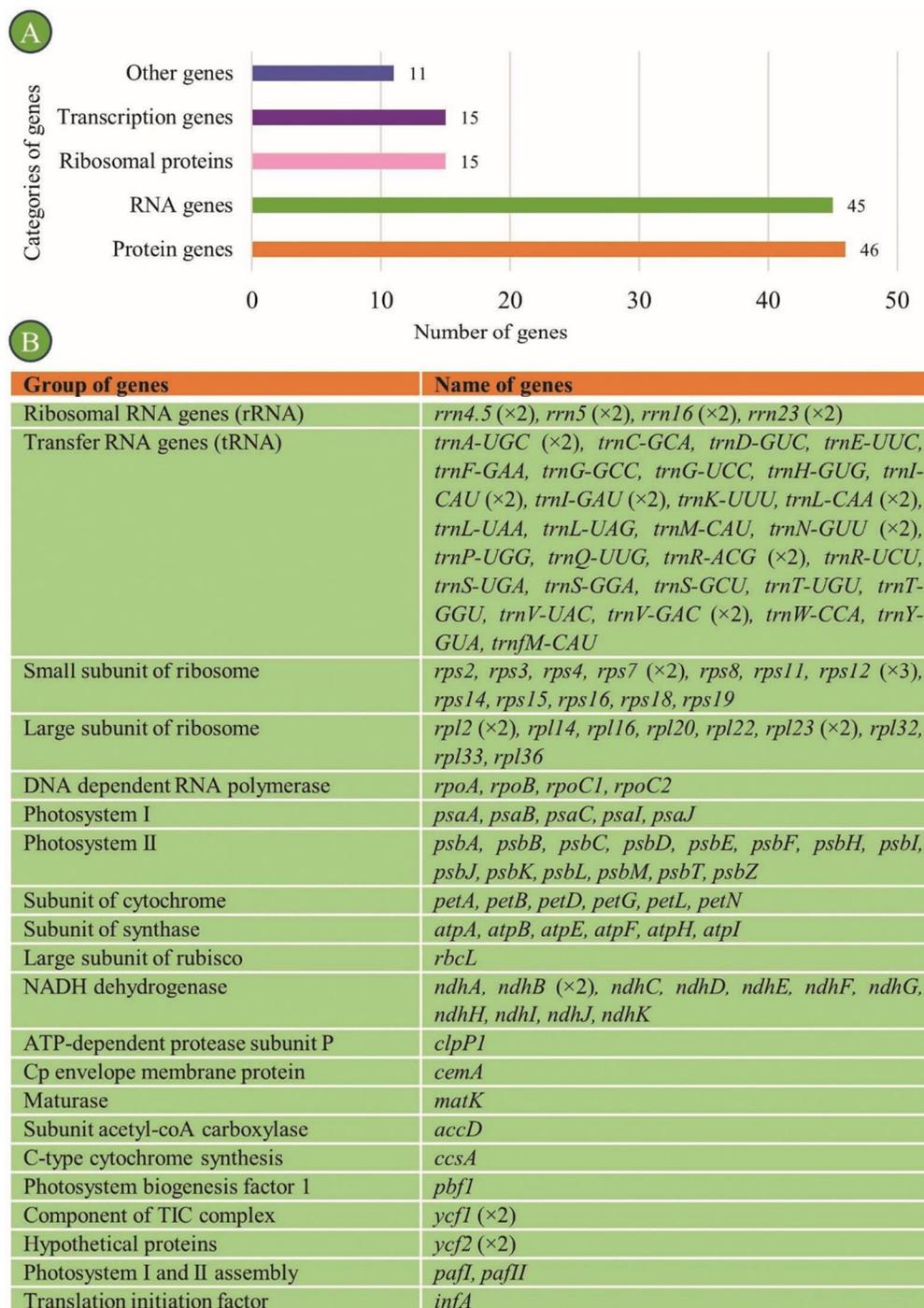


Fig. 3. Gene contents in the complete chloroplast genome of *T. macropterus* var. *arabicus*. A. Categories and number of genes, B Functional groups and names of genes.

than other types of SSRs, and our findings align with other studies (Zhang *et al.*, 2021). Evaluation of longer repeats revealed a total of 49 sequences classified into three types such as forward, reverse, and palindromic repeats. No complement repeats were observed in the plastome of *T. macropterus* var. *arabicus* (Fig. 4B). The plastome of *T. terrestris* also demonstrated zero occurrence of complement repeats but in the case of *B. aegyptiaca*, a small percentage of complement repeats was observed. Abundance of SSRs in the *T. macropterus* var. *arabicus* cp genome offers several advantages, including high polymorphism rates, codominant inheritance, and Mendelian segregation, making them potential markers for population genetics, phylogenetic studies, and molecular breeding. Furthermore, SSRs in the cp genome exhibit lower mutation rates compared to nuclear SSRs, enhancing their stability and reliability in evolutionary analyses (Nguyen *et al.*, 2021).

Table 2. Composition of nucleotides and the percentage of AT and GC content across the four regions of *T. macropterus* var. *arabicus* cp genome.

Region	A (%)	T (U) (%)	C (%)	G (%)	C + G (%)	A + T (%)
cp Genome	31.67	32.54	18.26	17.54	35.8	64.2
LSC	32.63	34.19	17.07	16.11	33.18	66.82
SSC	35.39	35.33	15.45	13.84	29.29	70.71
IRA	28.58	28.92	22.20	20.29	42.49	57.51
IRB	28.92	28.58	20.29	22.20	42.49	57.51

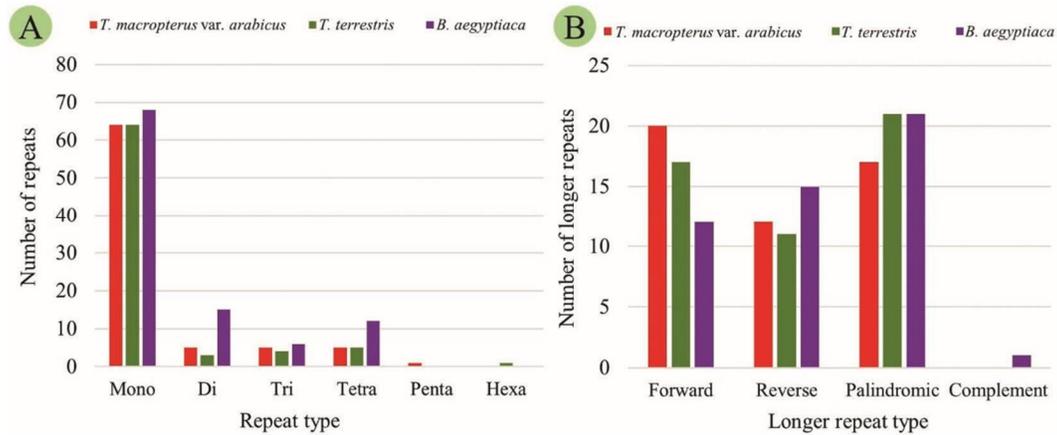


Fig. 4. Comparative analysis of repeat structures across various cp genomes. A. Simple sequence repeats, B. Longer repeats.

The codon usage analysis unveiled the utilization of 64 distinct codons encoding 20 unique amino acids, with a total codon frequency of 52,726. Alanine (GCG) exhibited the lowest codon frequency (217) whereas Phenylalanine demonstrated the highest frequency (2473). RSCU values ranged from 0.55 to 1.49 for different codons, showcasing varying degrees of usage bias. RSCU was recorded highest (6.06) for Leucine, followed by Arginine (6.00) (Fig. 5). Interestingly, 32 codons displayed usage frequencies exceeding the expected equilibrium ($RSCU > 1$), while 30 codons showcased usage bias ($RSCU < 1$). Of significance, AUG (Methionine) and UGG (Tryptophan) demonstrated unbiased usage, both with an RSCU value of 1. In the analysis of the entire cp genome of *Sophora tonkinensis*, Leucine emerged with the highest RSCU score. Except

for Methionine and Tryptophan, all other amino acids showed two to six codons (Wei *et al.*, 2020). This finding of codon usage is further supported by the present investigation.

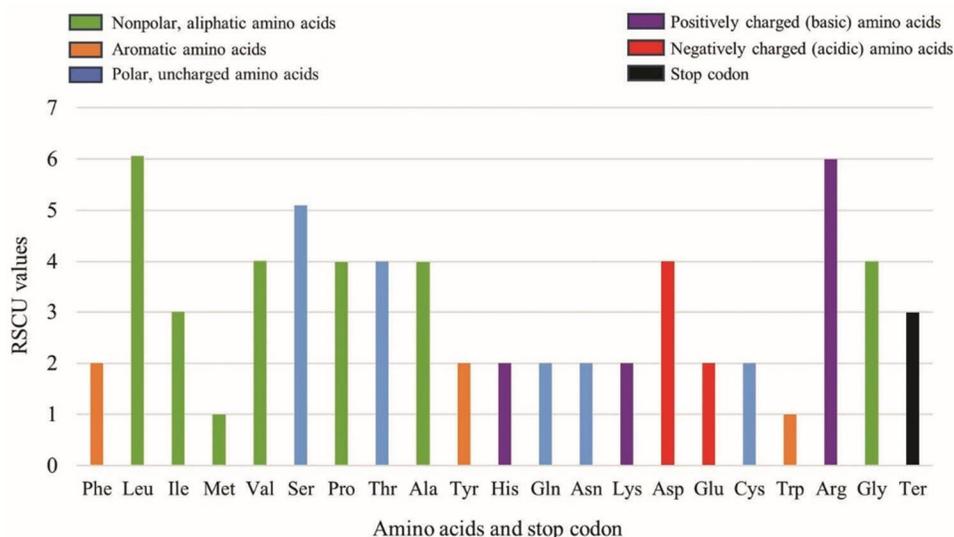


Fig. 5. Relative synonymous codon usage analysis of various amino acids of the cp genome of *T. macropterus* var. *arabicus*.

Inverted repeat (IR) expansion and contraction analysis

The cp genome of *T. macropterus* var. *arabicus* was compared with two closely related taxa, *T. terrestris* and *B. aegyptiaca*, revealing remarkably similar genome structure and gene organization across the LSC, SSC and IR regions. Junction site evaluation depicted the variation of the LSC, SSC and IR region within a narrow range (Fig. 6). The LSC ranged from 86,562 to 88,864 bp, while the SSC varied from 17,622 to 18,102 bp. IRb and IRa displayed similar results and justified the assembly of the *T. macropterus* var. *arabicus* plastome. Among the three studied taxa, the highest similarity was observed between the two *Tribulus* species compared to *B. aegyptiaca*. The *rps19* gene was found closely associated with the LSC/IRb border both in *T. terrestris* and *T. macropterus* var. *arabicus*. In *T. terrestris*, this gene originated from the IRb and expanded to the LSC, with an expansion of 294 bp. However, in *T. macropterus* var. *arabicus*, the *rps19* gene began 16 bp away from the LSC/IRb border within the LSC region. The *rpl2* gene, when translated counter-clockwise, was marked in the IRb zone in all three species, though in *B. aegyptiaca*, it was slightly deviated (175 bp) from the LSC/IRb border. However, in the clockwise direction, *rpl2* was observed in the IRa zone across all the three species. The gene *ndhF* was exclusively located in the SSC of *B. aegyptiaca*. The position of the *ycf1* was closely parallel in two *Tribulus* species, where it extended from IRb to SSC in a clockwise direction through the IRb/SSC border. This gene maintained a counter-clockwise expansion from IRa to SSC through the SSC/IRa border. The expansion volume matched exactly when moving in a clockwise and counter-clockwise fashion, as marked by 34 bp and 4130 bp, respectively. The positions of *psbA* and *trnH* genes were alike in the two *Tribulus* taxa. *trnH* was positioned in the IRa/LSC border in *B. aegyptiaca* (Fig. 6). The expansion and contraction of IR were found to be congruent with the findings of previous studies (Zhang *et al.*, 2021; Nguyen *et al.*, 2021). The presence of conserved IR boundaries suggests closer evolutionary relationships between the two *Tribulus* taxa, while

sequences (CNS) displayed the majority of variations. The features of LSC, SSC, and IR following whole genome alignment were concordant with other studies (Zhang *et al.*, 2021; Nguyen *et al.*, 2021).

Nucleotide diversity via sliding window technique

The sliding window analysis elucidated variability in the nucleotides and identified some hypervariable sites (Fig. 8). The highest nucleotide diversity (π) was recorded for the gene *rpl22* (0.17333), located in the LSC region. The second-highest peak was recorded in the LSC zone for the gene *cemA* with a π value of 0.17000. In the SSC, the gene *trnL-UAG* was identified as the most hypervariable site having the highest π value. The nucleotide diversity pattern was almost identical in the two inverted repeat regions. Among the two single copies, the LSC displayed a higher number of hypervariable sites than the SSC. In the IR region, π values were less than 0.05, and that justified the conserved nature of the IR over LSC and SSC. The inverted repeat zone within Asparagales chloroplast genomes and the Cinchonoideae subfamily showcased a notably lower nucleotide diversity ($\pi < 0.05$) (Munyao *et al.*, 2020; Castro *et al.*, 2023). The present investigation reinforces and extends these previously reported observations.

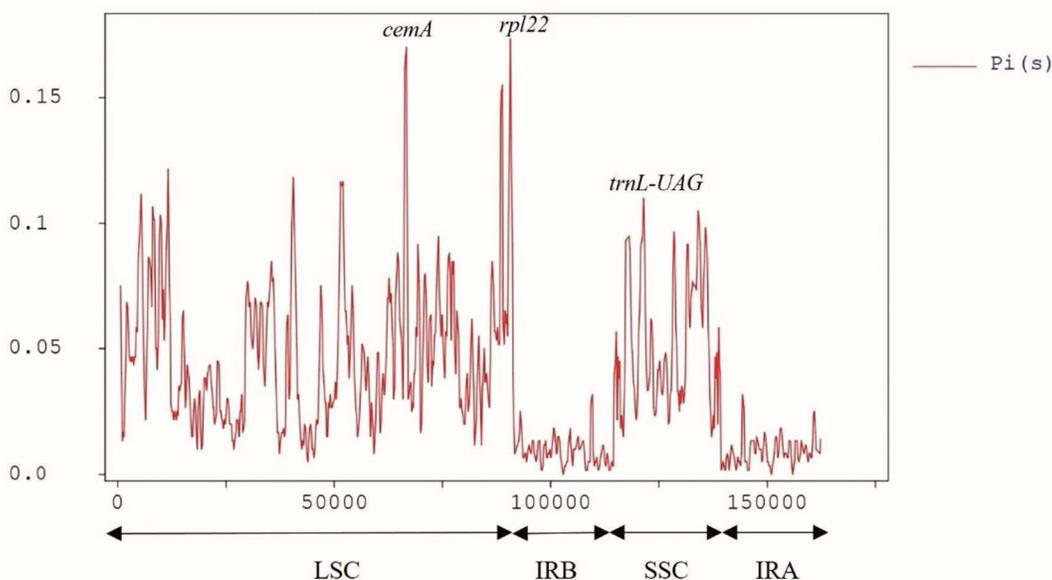


Fig. 8. Nucleotide diversity across the plastome of *T. macropterus* var. *arabicus* and its close relatives. The length and step of the sliding window were 600 bp and 200 bp, respectively.

Molecular phylogenetic analysis

Molecular phylogenetic analysis revealed clear segregation pattern of the member taxa within Zygophyllaceae. The *rbcL* gene sequences retrieved from GenBank at NCBI were used in the phylogenetic analysis. A total of 7 parsimonious trees were generated with 500 bootstrap replicates and the best tree (tree 1) was selected for interpretation and analysis. The MP tree supported the position of *T. macropterus* var. *arabicus* as a distinct taxon with good bootstrap support within Zygophyllaceae (Fig. 9). The tree was well rooted with the outgroup taxa, *Morkillia mexicana* and

Sericodes greggii. In the MP tree, the consistency index was found to be 0.984, the retention index was 0.891, and the composite index was 0.798 across all sites, and 0.693 for parsimony-informative sites. The final dataset comprised a total of 1379 positions, with 41 being parsimoniously informative. The genus *Tribulus* with its three member taxa - *T. terrestris*, *T. macropterus*, and *T. macropterus* var. *arabicus* exhibited a clear monophyletic nature (Bootstrap support 96%) within Zygophyllaceae. The cladding of *T. macropterus* var. *arabicus* with *T. macropterus* (Bootstrap support 50%) further validated the accurate assembly of this variety and depicted that the variety is distinct. The MP tree also illustrated that the genus *Tribulus* is more closely related to *Kelleronia* than to other taxa present in the MP tree (Fig. 9). The *rbcL* gene holds significant importance in the phylogenetics of *Tribulus* due to its widespread presence across Zygophyllaceae. The *rbcL* with its relatively conserved sequence regions coupled with sufficiently variable regions, can act as an ideal molecular marker for evolutionary studies. It encodes a key enzyme involved in photosynthesis, thereby reflecting the evolutionary history of photosynthetic organisms. Additionally, the abundance of *rbcL* sequences in public databases enables extensive comparative analyses, contributing to the current understanding of the evolutionary processes within Zygophyllaceae (Albert *et al.*, 1994). Several other studies have utilized this efficient barcode to delineate phylogenetic relationships within Zygophyllaceae, which justified the selection of *rbcL* to reconstruct the phylogeny of *Tribulus* in the present study (Sheahan and Chase, 2000; Bellstedt *et al.*, 2008; Alzahrani and Albokhari, 2017).

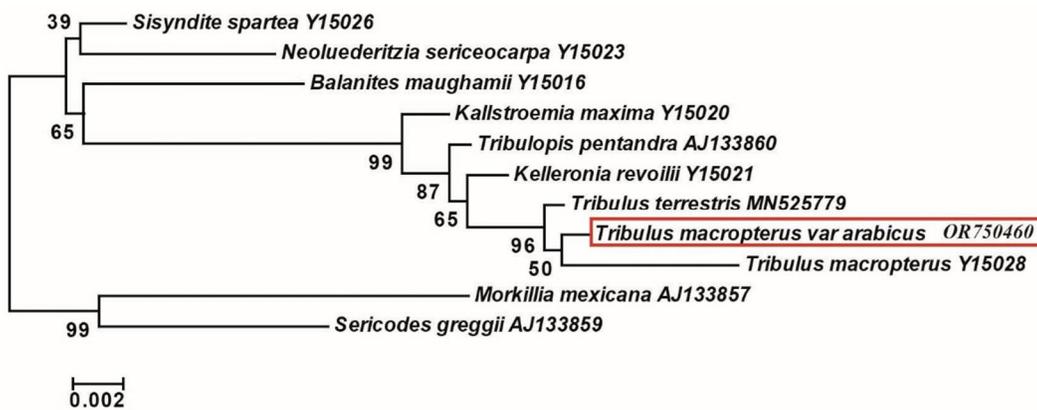


Fig. 9. Maximum-Parsimony (MP) tree with 500 bootstrap replicates showing phylogenetic relationships of *T. macropterus* var. *arabicus* based on *rbcL* gene.

The gene sequences of the genus *Tribulus* are inadequately represented in the GenBank database at NCBI. The genus *Tribulus* is highly polymorphic and showcases many intermediate forms. The complete Cp genome of *T. macropterus* var. *arabicus* unveiled in the present investigation will enrich the genomic information in GenBank and provide opportunities to conduct large-scale phylogenetic analyses in the future. Additionally, the Cp genome will facilitate DNA barcoding studies for accurate taxonomic identification of this medicinally important taxon and provide a foundation to enhance the current understanding of genetic and evolutionary variation within Zygophyllaceae.

Acknowledgements

The authors extend their appreciation to the Researchers Supporting Project (RSP2024R306), King Saud University, Riyadh, Saudi Arabia. This study was also supported by the KRIBB Initiative Program of the Republic of Korea (KGM4582423).

References

- Ahmed, I., Matthews, P.J., Biggs, P.J., Naeem, M., McLenachan, P.A. and Lockhart, P.J. 2013. Identification of chloroplast genome loci suitable for high-resolution phylogeographic studies of *Colocasia esculenta* (L.) Schott (Araceae) and closely related taxa. *Mol. Ecol. Resour.* **13**(5): 929-937.
- Albert, V.A., Backlund, A., Bremer, K., Chase, M.W., Manhart, J.R., Mishler, B.D. and Nixon, K.C. 1994. Functional constraints and *rbcL* evidence for land plant phylogeny. *Ann. Mo. Bot. Gard.* **81**(3): 534-567.
- Al-Juhani, W.S., Alharbi, S.A., Al Aboud, N.M. and Aljohani, A.Y. 2022. Complete chloroplast genome of the desert date (*Balanites aegyptiaca* (L.) Del. comparative analysis, and phylogenetic relationships among the members of Zygophyllaceae. *BMC Genomics* **23**(1): 1-19.
- Alshegaihi, R.M. 2024. The complete chloroplast genome of the halophyte flowering plant *Suaeda monoica* from Jeddah, Saudi Arabia. *Mol. Biol. Rep.* **51**(1): 60.
- Alzahrani, D.A. and Albokhari, E.J. 2017. Molecular phylogeny of Saudi Arabian *Tetraena* Maxim. and *Zygophyllum* L. (Zygophyllaceae) based on plastid DNA sequences. *Bangladesh J. Plant Taxon.* **24**(2): 155-164.
- Amiryousefi, A., Hyvönen, J. and Poczai, P. 2018. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinform.* **34**(17): 3030-3031.
- Beier, S., Thiel, T., Münch, T., Scholz, U. and Mascher, M. 2017. MISA-web: A web server for microsatellite prediction. *Bioinform.* **33**(16): 2583-2585.
- Bellstedt, D.U., Van Zyl, L., Marais, E.M., Bytebier, B., De Villiers, C.A., Makwarela, A.M. and Dreyer, L.L. 2008. Phylogenetic relationships, character evolution and biogeography of southern African members of *Zygophyllum* (Zygophyllaceae) based on three plastid regions. *Mol. Phylogenet. Evol.* **47**(3): 932-949.
- Castro, A.A., Nunes, R., Carvalho, L.R., Targueta, C.P., Braga-Ferreira, R.D., Melo-Ximenes, A.A., Corvalán, L.C.J., Bertoni, B.W., Pereira, A.M.S. and Telles, M.P.D. 2023. Chloroplast genome characterization of *Uncaria guianensis* and *Uncaria tomentosa* and evolutive dynamics of the Cinchonoideae subfamily. *Sci. Rep.* **13**(1), 8390.
- Chaudhary, S. 2001. Flora of the Kingdom of Saudi Arabia (Vol. 2). National Agriculture and Water Research Center Kingdom of Saudi Arabia, Riyadh.
- Daniell, H., Lin, C.S., Yu, M. and Chang, W.J. 2016. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **17**: 1-29.
- Dickinson, P.J., Kumar, M., Martinho, C., Yoo, S.J., Lan, H., Artavanis, G., Charoensawan, V., Schöttler, M.A., Bock, R., Jaeger, K.E. and Wigge, P.A. 2018. Chloroplast signaling gates thermotolerance in *Arabidopsis*. *Cell Rep.* **22**(7): 1657-1665.
- Eshel, G., Araus, V., Undurraga, S., Soto, D.C., Moraga, C., Montecinos, A., Moyano, T., Maldonado, J., Diaz, F.P., Varala, K., Nelson, C.W., Contreras-Lopez, O., Pal-Gabor, H., Kraiser, T., Carrasco-Puga, G., Nilo-Poyanco, R., Zegar, C.M., Orellana, A., Montecino, M., Maass, A., Allende, M.L., De Salle, R., Stevenson, D.W., Gonzalez, M., Latorre, C., Coruzzi, G.M. and Gutiérrez, R.A. 2021. Plant ecological genomics at the limits of life in the Atacama Desert. *Proc. Natl. Acad. Sci.* **118**(46): e2101177118.
- Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M. and Dubchak, I. 2004. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **32**(suppl.2): W273-W279.
- Katoh, K., Misawa, K., Kuma, K.I. and Miyata, T. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**(14): 3059-3066.
- Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J. and Giegerich, R. 2001. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **29**(22): 4633-4642.
- Librado, P. and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinform.* **25**(11): 1451-1452.

- Mandaville, J.P. 1986. Plant life in the Rub' al-Khali (the Empty Quarter), south-central Arabia. *Proc. Biol. Sci.* **89**: 147-157.
- Munayao, J.N., Dong, X., Yang, J.X., Mbandi, E.M., Wanga, V.O., Oulo, M.A., Saina, J.K., Musili, P.M. and Hu, G.W. 2020. Complete chloroplast genomes of *Chlorophytum comosum* and *Chlorophytum gallabatense*: Genome structures, comparative and phylogenetic analysis. *Plants* **9**(3): 296.
- Nguyen, H.Q., Nguyen, T.N.L., Doan, T.N., Nguyen, T.T.N., Pham, M.H., Le, T.L., Sy, D.T., Chu, H.H. and Chu, H.M. 2021. Complete chloroplast genome of novel *Adinandra megaphylla* Hu species: Molecular structure, comparative and phylogenetic analysis. *Sci. Rep.* **11**(1): 11731.
- Nguyen, V.B., Linh Giang, V.N., Waminal, N.E., Park, H.S., Kim, N.H., Jang, W., Lee, J. and Yang, T.J. 2018. Comprehensive comparative analysis of chloroplast genomes from seven *Panax* species and development of an authentication system based on species unique single nucleotide polymorphism markers. *J. Ginseng Res.* **44**(1): 135-144.
- Okonechnikov, K., Golosova, O., Fursov, M. and Ugene Team. 2012. Unipro UGENE: A unified bioinformatics toolkit. *Bioinform.* **28**(8): 1166-1167.
- Qian, J., Song, J., Gao, H., Zhu, Y., Xu, J., Pang, X., Yao, H., Sun, C., Li, X., Li, C., Liu, J., Xu, H. and Chen, S. 2013. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS ONE* **8**(2): e57607.
- Saina, J.K., Gichira, A.W., Li, Z.Z., Hu, G.W., Wang, Q.F. and Liao, K. 2018. The complete chloroplast genome sequence of *Dodonaea viscosa*: Comparative and phylogenetic analyses. *Genetica* **146**: 101-113.
- Sheahan, M.C. and Chase, M.W. 2000. Phylogenetic relationships within Zygophyllaceae based on DNA sequences of three plastid regions, with special emphasis on Zygophylloideae. *Syst. Bot.* **25**(2): 371-384.
- Shi, Y., Yan, X., Zhao, P., Yin, H., Zhao, X., Xiao, H., Li, X., Chen, G. and Ma, X.F. 2013. Transcriptomic analysis of a tertiary relict plant, extreme xerophyte *Reaumuria soongorica* to identify genes related to drought adaptation. *PLoS ONE* **8**(5): e63993.
- Tamura, K., Stecher, G. and Kumar, S. 2021. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**(7): 3022-3027.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**(22): 4673-4680.
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R. and Greiner, S. 2017. GeSeq-Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **45**(W1): W6-W11.
- Wei, F., Tang, D., Wei, K., Qin, F., Li, L., Lin, Y., Zhu, Y., Khan, A., Kashif, M.H. and Miao, J. 2020. The complete chloroplast genome sequence of the medicinal plant *Sophora tonkinensis*. *Sci. Rep.* **10**(1): 12473.
- Yan, J., Zhang, N. and Duan, Y. 2019. The complete chloroplast genome sequence of *Tribulus terrestris*, an important traditional Chinese medicine. *Mitochondrial DNA B Resour.* **4**(2): 3108-3109.
- Yonbawi, A.R., Abdallah, H.M., Alkhilaiwi, F.A., Koshak, A.E. and Heard, C.M. 2021. Anti-proliferative, cytotoxic and antioxidant properties of the methanolic extracts of five Saudi Arabian flora with folkloric medicinal use: *Aizoon canariense*, *Citrullus colocynthis*, *Maerua crassifolia*, *Rhazya stricta* and *Tribulus macropterus*. *Plants* **10**(10): 2073.
- Zeng, C., Jia, T., Gu, T., Su, J. and Hu, X. 2021. Progress in research on the mechanisms underlying chloroplast-involved heat tolerance in plants. *Genes* **12**(9): 1343.
- Zhang, L., Wang, S., Su, C., Harris, A.J., Zhao, L., Su, N., Wang, J.R., Duan, L. and Chang, Z.Y. 2021. Comparative chloroplast genomics and phylogenetic analysis of *Zygophyllum* (Zygophyllaceae) of China. *Front. Plant Sci.* **12**: 723622.
- Zheng, S., Poczai, P., Hyvönen, J., Tang, J. and Amiryousefi, A. 2020. Chloroplast: An online program for the versatile plotting of organelle genomes. *Front. Genet.* **11**: 576124.