Bangladesh J. Plant Taxon. **31**(2): 205-223, 2024 (December) DOI: https://doi.org/10.3329/bjpt.v31i2.78749

COMPLETE CHLOROPLAST GENOME SEQUENCE OF A NOVEL *WITHANIA SOMNIFERA* (L.) DUNAL: COMPARATIVE GENOMICS AND PHYLOGENETIC INSIGHTS

NADIA MOHAMMAD ALSUHAIMI[1], MOHAMMAD AJMAL ALI[1,*], MONA SOLAIMAN ALWAHIBI[1], SHEIKH SUNZID AHMED[2], M. OLIUR RAHMAN[2*], SHANKAR KUMAR PANDEY[3], MOHAMED S ELSHIKH[1], SAYFI RASHED SAYFI ALSHALLALI[4], JOONGKU LEE[5] AND SOO-YONG KIM[6]

[1]*Department of Botany and Microbiology, College of Science, King Saud University, Riyadh-11451, Saudi Arabia*
[2]*Department of Botany, Faculty of Biological Sciences, University of Dhaka, Dhaka 1000, Bangladesh*
[3]*Department of Botany, SSV College Kahalgaon, Tilka Manjhi Bhagalpur University, Bhagalpur, Bihar, India*
[4]*Department of Pharmacognosy, College of Pharmacy, King Saud University, Riyadh 11451, Saudi Arabia*
[5]*Department of Environment and Forest Resources, Chungnam National University, Daejeon, Republic of Korea*
[6]*International Biological Material Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 34141, Republic of Korea*

Keywords: Chloroplast genome; Phylogeny; Solanaceae; Nucleotide diversity; Bioinformatics; *Withania somnifera* (L.) Dunal var. *abhaica* Nadia, A. Ali & M.S. Alwahibi, var. nov.

Abstract

This study introduces a novel variety of the highly esteemed medicinal plant *Withania somnifera* (L.) Dunal from the family Solanaceae. The new variety, *Withania somnifera* var. *abhaica* Nadia, A. Ali & M.S. Alwahibi, var. nov., is distributed at high altitudes in the Abha hills of Saudi Arabia. The distinct characteristics of the novel variety of *W. somnifera* include elliptic-elongated leaves that are thick and semi-succulent in nature, and a fruiting calyx with a bifurcated tip measuring approximately 0.5 mm in length, with each bifurcated tip being botuliform in shape. Using next-generation sequencing (NGS) techniques, we investigated the chloroplast genome of this variety. The complete chloroplast genome of *W. somnifera* var. *abhaica* from the Abha region, Saudi Arabia, spans 153,621 bp, with a GC content of 37.7%. It includes a large single-copy (LSC) region of 84,972 bp (GC 35.8%), a small single-copy (SSC) region of 18,400 bp (GC 31.7%), and two inverted repeats (IRs) of 50,249 bp (GC 43.2%). Annotation of the chloroplast genome identified 131 genes, comprising 86 protein-encoding genes (PCGs), eight ribosomal RNA genes, and 37 transfer RNA genes. Repeat analysis identified 38 simple sequence repeats (SSRs) and 50 longer repeat sequences in the plastome. A total of 66 RNA-editing sites were detected across 24 PCGs of the plastome. Comparative genomic studies including synteny analysis, supported and validated the assembled plastome. Nucleotide diversity analysis revealed *psbJ*, *psbA*, *ndhF*, and *ycf1* as the most hypervariable barcodes. Phylogenetic analyses suggested the monophyly of the genus *Withania*. Moreover, the newly sequenced chloroplast genome of *W. somnifera* var. *abhaica* was found to be distinct from the typical *W. somnifera.*

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

Introduction

The genus *Withania* Pauquy, a member of the Solanaceae family, includes 23 species widely distributed across North Africa, West Asia, and Southern Europe (Olmstead *et al*., 2008). In Saudi Arabia, this genus is represented by *Withania somnifera* (L.) Dunal, a highly esteemed medicinal plant, commonly known as "Indian Ginseng" (Paul *et al*., 2021). *W. somnifera* has been extensively utilized in traditional medicine for centuries, either alone or in combination with other herbs (Visweswari *et al*., 2013). It thrives in a diverse range of habitats, from desert plains to altitudes as high as 2,995 meters in the Abha region, including Jabal Sawda (Rahman *et al*., 2004). This herb exhibits a broad spectrum of biological activities due to its diverse phytochemical composition, including anti-inflammatory, antimicrobial, anti-tumor, neuroprotective, cardioprotective, and antidiabetic properties (Dar *et al*., 2015). Additionally, studies on *W. somnifera* have demonstrated its efficacy to lower reactive oxygen species, alter mitochondrial operations, control apoptosis, lessen inflammation, as well as improve endothelial function. Due to these pharmacological properties, *W. somnifera* holds significant potential as a therapeutic option for various clinical disorders, specifically those affecting the nervous system (Kulkarni and Dhir, 2008).

W. somnifera var. *abhaica* is characterized by reduced plant height, elliptic-elongated, thick and puberulous leaves that resemble semi-succulent morphological nature. Under a scanning electron microscope, the tip of the fruiting calyx appeared bifurcated, c. 0.5 mm in length, and each bifurcated tip being botuliform in shape (Fig. 1). These unique morphological characteristics, which differ from the typical features of *W. somnifera*, have sparked interest in conducting a chloroplast (Cp) genome-based systematic investigation. Unveiling the Cp genome of *W. somnifera* var. *abhaica* holds significant importance, even though the Cp genome of *W. somnifera* is already available (GenBank Accession MK142783) (Mehmood *et al*., 2020). While the plastome of *W. somnifera* provides a comprehensive reference for the species, studying the plastome of *W. somnifera* var. *abhaica* is essential to clarify taxonomic boundaries and confirm its classification as a distinct variety. Variations at the genomic level, particularly within the chloroplast DNA, can offer critical insights into the evolutionary relationships and genetic divergence between the variety and the typical species. These genomic differences can help resolve ambiguities in taxonomy, ensuring that *W. somnifera* var. *abhaica* is accurately classified and distinguished from other closely related taxa. Moreover, such studies can reveal unique genetic features of the variety, which may have implications for its ecological adaptation, medicinal properties, and conservation strategies (Dobrogojski *et al.*, 2020).

Chloroplasts (Cp) are essential cellular components of angiosperms involved in photosynthesis and the synthesis of important macromolecules, including amino acid and fatty acid. In molecular systematics research, plastome is a crucial element because of its distinctive characteristics and adaptive processes (Ahmed and Rahman, 2024). Typically, the Cp genome represents a quadripartite structure, comprising two inverted repeats separated by the large single-copy (LSC) and small single-copy (SSC) regions. In angiosperms, plastome sizes typically range from 107 kb to 218 kb (Wang *et al*., 2020). Diversity in gene content and organization can result from dynamic shifting in this genomic configuration, including contractions, expansions, and even unfolding (Ravi *et al*., 2008). In angiosperms, the plastome is typically inherited through the maternal line, whereas, in certain gymnosperms, it is inherited through the paternal line. The Cp genome encodes several groups of genes, including transfer RNAs (tRNAs), protein-coding genes (PCGs), and ribosomal RNAs (rRNAs), all of which are necessary for chloroplast function. Due to the inherent diversity found in Cp genomes, expressed through structural variants and polymorphisms, phylogenetic analysis provides new opportunities to resolve systematic

relationships, track the evolution of species, and investigate how species adapt to certain environments (Daniell *et al.*, 2016; Dobrogojski *et al.*, 2020).

Advancements in bioinformatics and next-generation sequencing (NGS) technologies have revolutionized genomics, significantly enriching the GenBank repository with plastomes (Jongsun *et al.*, 2020). Bioinformatic tools have streamlined the analysis and assembly of large-scale genomic data, while NGS has accelerated the sequencing process, making it faster, more cost-effective, and widely accessible. These advancements have resulted in a swift rise in the number of complete plastomes deposited in the GenBank database, providing a valuable resource for comparative genomics, phylogenetics, and evolutionary research. The availability of full plastome sequences has also elevated their role as a "superbarcode" for resolving phylogeny (Zhang *et al.*, 2019). Unlike traditional barcoding methods that rely on a few gene regions, the full plastome offers a comprehensive genetic blueprint, enabling more accurate and reliable phylogenetic analysis of *W. somnifera* var. *abhaica*. This enhanced resolution is particularly crucial for distinguishing closely related species, resolving complex evolutionary relationships, and improving the accuracy of species identification within the Solanaceae family (Olmstead *et al.*, 2008).

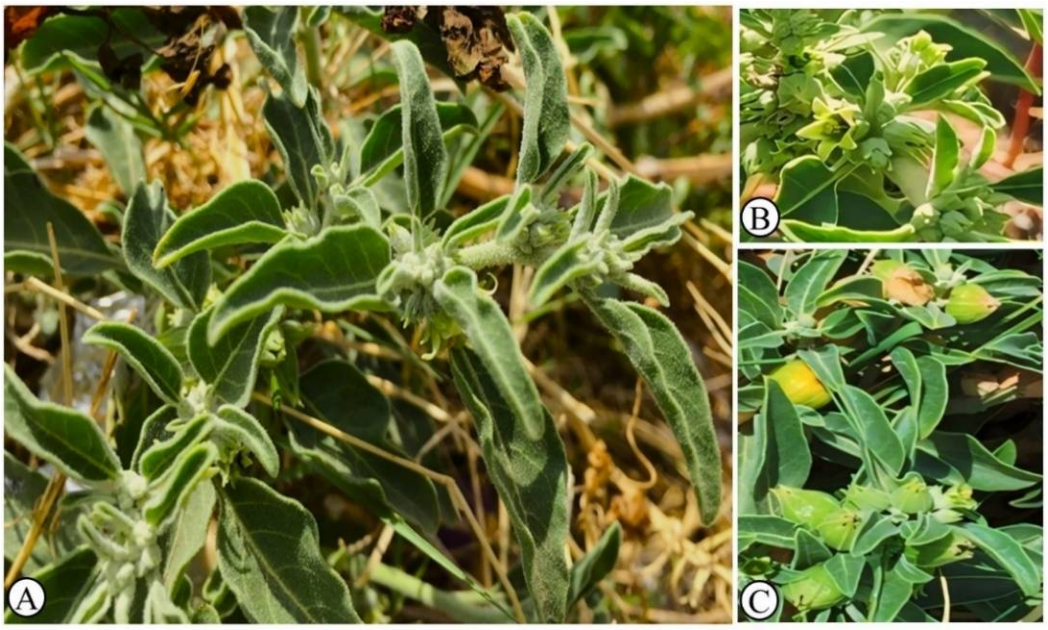

Fig. 1. Morphology of *Withania somnifera* var. *abhaica* collected from the hilly terrain of Abha region of Saudi Arabia. A. Habit, B. Flower, C. Fruit.

Given the intriguing morphological features observed in *W. somnifera* var. *abhaica* from the Abha region, a chloroplast genome-based approach is crucial for uncovering the genetic foundations and elucidating the phylogenetic relationships within this species. In this study, we aim to deepen the understanding of the complete chloroplast genome of this novel variety, focusing on the unique ecological context of the Abha hills in Saudi Arabia, by utilizing a comprehensive NGS-driven bioinformatics approach.

Materials and Methods

Specimen collection

The plant specimen was collected from the hilly regions of the Abha area in Saudi Arabia (coordinates: 18º17′44″N, 42º25′40″E; altitude: 2,491 m). The voucher specimen is preserved at the King Saud University Herbarium (KSUH) in Riyadh, Saudi Arabia, under the collection code Nadia, M.A. & Ali, M.A. 2021-1, ecotype (Abha hills, Saudi Arabia).

Genome sequencing

Total genomic DNA was isolated from silica gel-dried leaves utilizing the Qiagen DNA Extraction Kit. Paired-end reads of 151 bp were generated with a NextSeq 500 sequencer. The NGS (Next-generation sequencing) reads were assessed using FASTQC tool v.0.12.1 to evaluate Phred quality scores (Ahmed and Rahman, 2024). The raw sequencing data are publicly accessible on NCBI under the SRA accession ID SRR27753935.

Construction of the plastome and its annotation

The high-quality SRA data were configured into the Cp genome using Unipro UGENE v45.1 (Okonechnikov *et al.*, 2012). Annotation of the plastome was carried out employing CPGAVAS2 server and subsequently verified with CPGView (Shi *et al.*, 2019; Liu *et al.*, 2023). The manually curated annotation was used to construct the circular plastome diagram employing OGDraw server (Greiner *et al.*, 2019). The assembled plastome has been deposited in GenBank under the accession number OR166175.

Evaluation of longer repeats and SSRs

The REPuter server was used to identify longer repeat structures, while SSRs (simple sequence repeats) were identified utilizing the MISA-Web server (Kurtz *et al.*, 2001; Beier *et al.*, 2017). For the analysis of longer repeats in REPuter, all matching directions were considered. SSRs were analyzed using the default settings of the MISA-Web server.

Assessment of RNA editing sites and GC skewness

The plastome was examined for RNA editing sites using the PREPACT 3.0 server (Lenz *et al.*, 2018). The BLASTx module was utilized to identify forward editing sites (C→U), with *Nicotiana tabacum* L. (Solanaceae) as the reference database and an e-value threshold of 0.001. For GC content skewness analysis, the assembled plastome was uploaded in FASTA format to the Proksee server (Grant *et al.*, 2023). After initial processing, GC content and GC skew analysis options were applied to visualize the circular map.

Genome rearrangement and collinearity analysis

The assembled plastome of *W. somnifera* var. *abhaica* was subjected to comparative genomic analysis using the Mauve v.20150226 tool to identify gene order similarities with other taxa (Darling *et al.*, 2004). GenBank flat files of the relevant taxa were initially imported into the Java console to run the progressive Mauve module. The analysis employed the HOXD scoring matrix, with gap opening and gap extension penalties set to -400 and -30, respectively. These parameters were optimized to maximize alignment accuracy and synteny detection before initiating the final comparative analysis. For collinearity analysis, the plastome was analyzed using the Circoletto server (Darzentas, 2010).

Nucleotide diversity analysis

The nucleotide diversity analysis commenced with the alignment of chloroplast genome sequences utilizing the MAFFT online tool to ensure precise sequence alignment across the studied genomes (Katoh *et al.*, 2005). Subsequently, nucleotide diversity was assessed with

DnaSP v.5 tool. A sliding window approach was employed, with a window length of 600 base pairs and a step size of 200 base pairs, allowing for a detailed examination of nucleotide diversity across the genome (Librado and Rozas, 2009). The genomic coordinates of each window were then compared with the annotated gene regions of the chloroplast genome to identify and characterize patterns of nucleotide diversity.

Plastome-wide molecular phylogeny

To conduct a comprehensive plastome-wide molecular phylogenetic analysis within the Solanaceae family, 36 taxa, including *W. somnifera* var. *abhaica*, were chosen and retrieved from the NCBI GenBank database. *Mentha spicata* L. and *Phyla nodiflora* (L.) Greene were included as outgroup taxa to root the phylogenetic tree and provide context for evolutionary relationships. The sequences were compiled into a multi-FASTA file and subsequently aligned using the MAFFT server to ensure accurate sequence alignment across all taxa. The aligned sequences were then analyzed in MEGA v.11, where a maximum-likelihood (ML) tree was constructed. The Tamura 3-parameter model was employed as the nucleotide substitution model, which effectively captures the evolutionary dynamics within the plastomes (Ahmed and Rahman, 2024). The substitution rates were set to uniform to maintain a consistent rate of nucleotide changes across the sequences. A partial deletion approach was adopted for gap treatment, allowing for the exclusion of gaps that could potentially skew the results. To evaluate the robustness of the phylogenetic relationships, the ML tree was generated with 1,000 bootstrap replicates, providing a measure of confidence for each branch (Tamura *et al*., 2021).

Results and Discussion

Quality evaluation of NGS reads

The sequencing run using the NextSeq 500 platform generated approximately 5.6 GB of high-quality, adapter-removed reads. Analysis of these raw reads revealed a GC content of 43% and an AT content of 57%, indicating a balanced nucleotide composition that reflects the genomic characteristics of the sample. Quality control metrics further confirmed the reliability of the sequencing process, with 96.6% of the bases achieving a Q20 score, indicating a 99% base call accuracy. Additionally, 90.8% of the bases reached the Q30 threshold, corresponding to a 99.9% base call accuracy, underscoring the overall robustness and precision of the sequencing data obtained. The quality of the raw reads in this study was found to be consistent with a recently published plastome of *Tribulus macropterus* variety, where the Phred quality scores were 95.9% and 89.7% for the Q20 and Q30 indices, respectively (Albediwi *et al.*, 2024).

Genome structure and contents

The orbicular quadripartite plastome spanned a total length of 153,621 bp, comprising 84,972 bp in the LSC region, 18,400 bp in the SSC region, and 50,249 bp in the IRs regions (Fig. 2). The present investigation demonstrated a remarkable consistency with the plastome of *W. somnifera* as reported by Mehmood *et al.* (2020). The plastome of *W. somnifera* (MK142783) was characterized by a total length of 154,386 bp, with LSC, SSC, and IR regions measuring 85,688 bp, 18,464 bp, and 50,234 bp, respectively (Mehmood *et al.*, 2020). This strong correlation reinforces the validity of the plastome structure for *W. somnifera* var. *abhaica* constructed in this study.

The comparative analysis of the guanine-cytosine (GC) ratio and adenine-thymine (AT) ratio across different compartments of the plastome revealed distinct patterns (Table 1). The overall plastome exhibited an AT content of 62.26% and a GC content of 37.74%. The LSC region had the highest AT content of 64.20%, with a correspondingly lower GC content of 35.80%, reflecting

a higher proportion of adenine and thymine bases. The SSC region further amplified this trend, showing the highest AT content of 68.22% and the lowest GC content of 31.78%. In contrast, the inverted repeats (IRA and IRB) displayed a higher GC content, both around 43.2%, with corresponding AT contents of approximately 56.8%. These differences underscore the variability in nucleotide composition across the plastome.

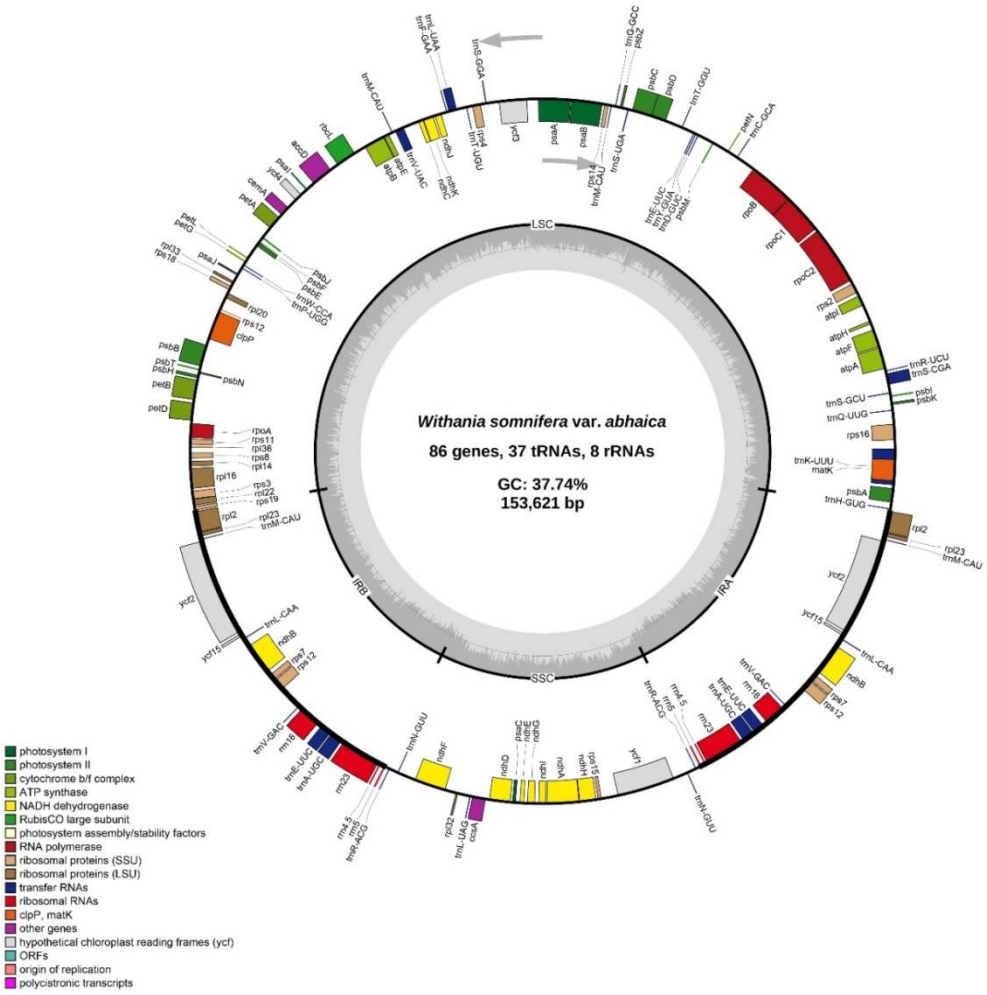

Fig. 2. Complete chloroplast genome of *Withania somnifera* var. *abhaica* representing gene orders and quadripartite junction sites.

An elevated AT ratio in the SSC and LSC regions, in contrast to the IRs indicates an evolutionary trend where these protein-coding gene-rich regions show a selection for AT-rich codons to potentially enhance gene expression and protein function (Qian *et al.*, 2013). This increased AT content is likely a consequence of higher recombination rates in these dynamic regions, driving nucleotide variability. In contrast, the IR regions, characterized by fewer recombination events, exhibit a lower AT content and greater structural stability, resulting in a more preserved nucleotide pattern. This contrast underscores how distinct evolutionary pressures

have shaped the structure and composition of the *Withania somnifera* var. *abhaica* plastome, reflecting both functional requirements and genomic stability (Saina *et al*., 2018).

Table 1. Proportion of nucleotides in the quadripartite sites of *Withania somnifera* var. *abhaica*.

Zones	A (%)	T (U) (%)	C (%)	G (%)	C + G (%)	A + T (%)
Plastome	30.74	31.52	19.19	18.55	37.74	62.26
LSC	31.43	32.77	18.32	17.48	35.80	64.20
SSC	33.86	34.36	16.65	15.14	31.78	68.22
IRA	28.42	28.35	22.43	20.80	43.23	56.77
IRB	28.43	28.37	20.77	22.44	43.21	56.79

Annotation of genes

The annotation of the *W. somnifera* var. *abhaica* Cp genome revealed a total of 131 genes, comprising 86 PCGs, 37 tRNA genes, and eight were rRNA genes (Fig. 2). Among the 19 genes associated with photosystems, 14 genes encode photosystem II, whereas 5 genes are involved with photosystem I. The genes encoding the small subunit of the ribosome (15 genes) outnumbered those encoding the large subunit (11 genes) (Table 2). Most PCGs were localized in the single-copy zones, while the inverted repeats predominantly contained RNA genes. The SSC region was notable for its high concentration of genes encoding subunits of NADH dehydrogenases. Furthermore, the LSC region featured two key DNA barcodes, *rbcL* and *matK*. In addition, the LSC harbored the *cemA* gene, which encodes the chloroplast envelope membrane protein. The high concentration of PCGs in the single-copy regions highlights their role in essential processes like photosynthesis and metabolism, contributing to functional diversity. The predominance of RNA genes in the IRs emphasizes their importance in maintaining genomic stability and supporting efficient protein synthesis. Additionally, the abundance of NADH dehydrogenase genes in the SSC region underscores its crucial role in energy metabolism and photosynthetic efficiency (Dobrogojski *et al*., 2020).

Table 2. Classification of the protein-coding genes present in *Withania somnifera* var. *abhaica* plastome.

Categories	Gene groups	Gene names
Genes for photosynthesis	Photosystem I	*psaA, psaB, psaC, psaI, psaJ*
	Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbM, psbN, psbT, psbZ*
	ATP synthase	*atpA, atpB, atpE, atpF, atpH, atpI*
	Cytochrome b/f complex	*petA, petB, petD, petG, petL, petN*
	Rubisco	*rbcL*
	NADH-dehydrogenase	*ndhA, ndhB*(×2), *ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
Self-replication	Small subunit of ribosome	*rps2, rps3, rps4, rps7*(×2), *rps8, rps11, rps12*(×3), *rps14, rps15, rps16, rps18, rps19*
	Large subunit of ribosome	*rpl2*(×2), *rpl14, rpl16, rpl20, rpl22, rpl23*(×2), *rpl32, rpl33, rpl36*
	DNA dependent RNA polymerase	*rpoA, rpoB, rpoC1, rpoC2*
Other genes	Maturase	*matK*
	Envelop membrane protein	*cemA*
	Acetyl-CoA-carboxylase	*accD*
	C-type cytochrome synthesis gene	*ccsA*
	Protease	*clpP*
Unknown	Conserved open reading frames	*ycf1, ycf2*(×2), *ycf3, ycf4, ycf15*(×2)

In a recent study, Mehmood *et al.* (2020) identified 86 PCGs, 37 tRNAs, and eight rRNAs in the plastome of *W. somnifera*. The gene content, organization, and localization in their investigation were congruent with the findings of the present study, showcasing the genetic similarity between *W. somnifera* and *W. somnifera* var. *abhaica.*

Exon-intron distribution and cis-trans splicing

The exon-intron distribution in the plastome of *W. somnifera* var. *abhaica* revealed a diverse and intricate gene structure across different regions (Table 3). In the LSC region, several genes, such as *trnK-UUU*, *rps16*, and *atpF*, have two exons separated by a single intron. The genes *ycf3* and *clpP* exhibit more complex structures, each containing three exons and two introns. Notably, the *ycf3* gene stands out with its relatively long intron II of 745 bp. The IRA and IRB regions display a mirrored arrangement of genes, such as *rpl2* and *ndhB*, both containing two exons separated by a substantial intron. The SSC region includes the *ndhA* gene, which also features two exons and a notably a long intron of 1159 bp. This distribution highlights the complex organization of the plastome, with multiple genes featuring introns, contributing to the regulatory mechanisms of gene expression and the potential for alternative splicing in the plastid genome. The pattern of exon-intron distribution observed in this study aligns closely with that of the plastome of *Capparis decidua* (Forsk) Edgew (Alzahrani and Albokhari, 2022). Similar to our findings in *W. somnifera* var. *abhaica, C. decidua* also exhibited complex structures in the *ycf3* and *clpP* genes within the LSC region, with both genes containing three exons separated by two introns. This resemblance corroborated the accuracy of the exon-intron prediction in the plastome of *W. somnifera* var. *abhaica.*

Table 3. Exons-introns distribution in the plastome of *Withania somnifera* var. *abhaica.*

Location	Genes	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
IRB	*trnA*-UGC	37	811	36		
IRA	*ndhB*	775	679	758		
SSC	*ndhA*	553	1159	539		
IRB	*rpl2*	391	666	434		
IRB	*trnE*-UUC	32	723	40		
IRB	*ndhB*	775	679	758		
IRA	*trnE*-UUC	32	723	40		
IRA	*trnA*-UGC	37	811	36		
LSC	*rpl16*	9	1028	396		
IRA	*rpl2*	391	666	434		
LSC	*rpoC1*	453	737	1614		
LSC	*ycf3*	124	727	232	745	151
LSC	*trnK*-UUU	37	2477	36		
LSC	*petB*	6	746	642		
LSC	*petD*	8	745	475		
LSC	*trnL*-UAA	35	492	50		
LSC	*trnV*-UAC	36	552	56		
LSC	*rps16*	40	855	227		
LSC	*trnS*-CGA	31	674	60		
LSC	*atpF*	145	700	410		
LSC	*clpP*	71	790	294	632	244

The plastome of *W. somnifera* var. *abhaica* unveiled several cis-splicing genes, such as *rps16*, *atpF*, *rpoC1*, *ycf3*, *clpP*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, and *ndhA* (Fig. 3). These genes feature diverse intron-exon structures, with some containing multiple introns and others having a single intron, exhibiting significant variability in intron lengths. This diversity in intron structures and lengths suggests a range of splicing requirements and potential impacts on gene expression and chloroplast function, reflecting adaptations to specific functional needs and environmental conditions (Dobrogojski *et al.*, 2020). The *rps12* is a trans-spliced gene with exons located in separate regions of the genome (Fig. 4). Exon 1 is located in the LSC, while Exons 2 and 3 are situated in the IRs. These exons were transcribed separately and then spliced together to form a mature mRNA, which facilitates the correct assembly of the *rps12* coding sequence. This trans-splicing mechanism is essential for the gene's proper expression, ensuring that exons from distinct genomic regions are combined to produce a functional protein. The plastome of *Mandragora caulescens* C. B. Clarke (tribe Solaneae) revealed a similar structural organization of cis- and trans-spliced genes, supporting the findings of the present investigation (Ma *et al.*, 2024).

Fig. 3. *W. somnifera* var. *abhaica* plastome showing genes responsible for cis-splicing.

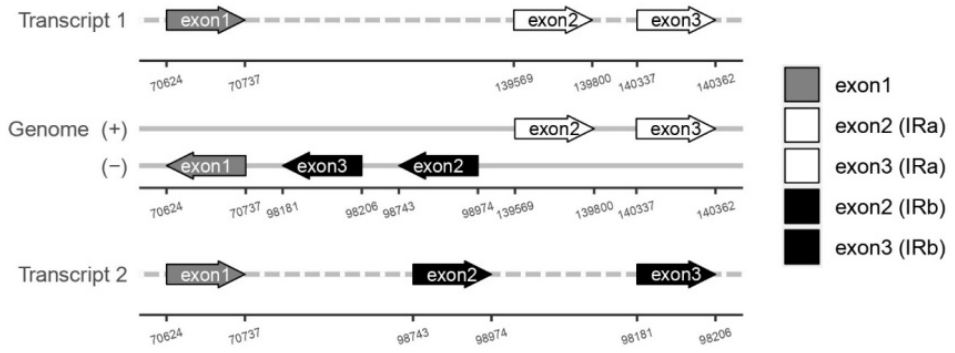

Fig. 4. *W. somnifera* var. *abhaica* plastome showing *rps12* responsible for trans-splicing.

Longer repeats and SSRs

In the analysis of longer repeats, *W. somnifera* var. *abhaica* exhibited 15 forward, eight reverse, 26 palindromic, and one complement repeats (Fig. 5A). Compared to other species, *W. somnifera* var. *abhaica* had fewer forward and reverse repeats than *W. coagulans*, *W. riebeckii*, and *Dunalia obovata*. However, *W. somnifera* var. *abhaica* had more palindromic repeats than *W. riebeckii* and *W. coagulans*, and fewer complement repeats compared to most species, except for *D. obovata*, which had none. *Physalis peruviana* had one longer repeat, similar to *W. somnifera* var. *abhaica*.

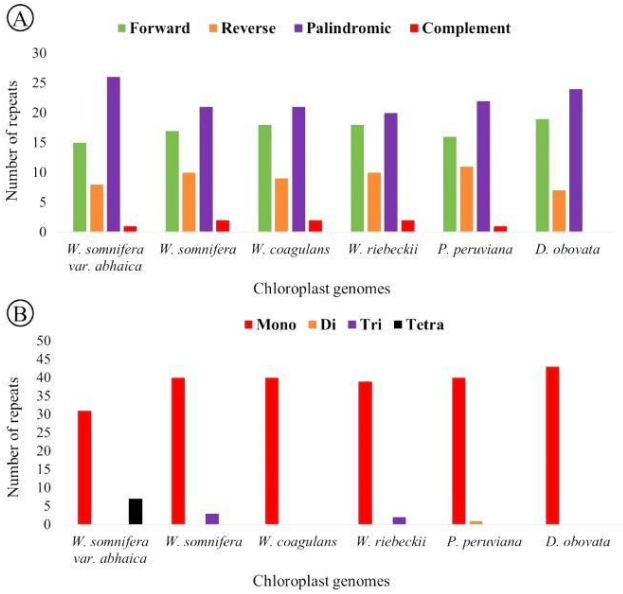

Fig. 5. Comparative overview of the repeat structures present in the plastome of *W. somnifera* var. *abhaica* and allied genera. A. Longer repeats, B. Simple sequence repeats.

The evaluation of SSR profile revealed *W. somnifera* var. *abhaica* had 31 mononucleotide repeats, zero dinucleotide and trinucleotide repeats, and seven tetranucleotide repeats (Fig. 5B). In comparison, *W. somnifera* and *W. coagulans* had more mononucleotide repeats, with 40 each, and *D. obovata* had the highest number of 43. Unlike *W. somnifera* var. *abhaica*, which had no dinucleotide or trinucleotide repeats, *P. peruviana* showed one dinucleotide repeat, and *W.*

riebeckii had two trinucleotide repeats. *W. somnifera* var. *abhaica* was unique in having seven tetranucleotide repeats, a characteristic feature not found in the other species studied. These findings suggest that *W. somnifera* var. *abhaica* possesses a unique profile compared to the other analyzed species. SSRs in the plastome exhibit lower mutation rates compared to nuclear SSRs, enhancing their stability and reliability in phylogenetic studies (Albediwi *et al.*, 2024).

RNA-editing sites and GC skewness

The RNA-editing analysis of the *W. somnifera* var. *abhaica* cp genome revealed a varied distribution of editing sites across different regions, with the LSC region comprising 40% of the sites, the IR region 21%, and the SSC region 39% (Fig. 6). A total of 66 RNA-editing sites were identified in the plastome across various compartments. The SSC region exhibited the highest concentration of RNA-editing sites, with the *ycf1* gene containing 12 sites, followed by the *ndhD* gene with 7 sites. In contrast, the LSC region, covering the largest portion of the genome, had a wider range of genes with editing sites, though the *rpoB* gene had the highest count at 5 sites. The IR region showed significant editing activity in the *ycf2* gene with 5 sites. Overall, the SSC region, despite having slightly fewer editing sites than the LSC, showed a higher frequency in specific genes, such as *ycf1* and *ndhD* compared to the more evenly distributed editing sites among various genes in the LSC and IR regions.

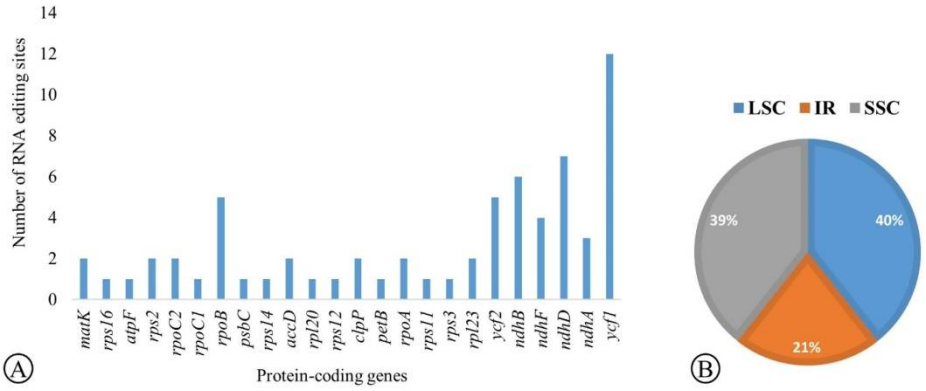

Fig. 6. Distribution of RNA editing loci in the plastome of *W. somnifera* var. *abhaica.* A. Distribution across PCGs, B. Distribution across compartments.

A high concentration of RNA-editing sites in the SSC region, as observed in our findings, has also been reported in the chloroplast genome of *Capparis decidua* (Alzahrani and Albokhari, 2022), where 46 RNA-editing sites were detected across 18 genes, including nine sites within the *ndhD* gene. This pattern of RNA-editing is further supported by findings in the plastome of *Solanum dulcamara* L., which exhibited a similar tendency, reinforcing the significance of our results (Amiryousefi *et al.*, 2018). Analyzing RNA-editing sites in the *W. somnifera* var. *abhaica* plastome is crucial for uncovering the complex regulatory mechanisms that govern gene expression and photosynthesis in this medicinal plant. Variations in RNA-editing efficiency across different genes, or even at different loci within the same gene, suggest intricate layers of control that can impact protein structure and function. These modifications have the potential to regulate key physiological processes, including photosynthesis, which is vital for the survival and metabolic activities of this variety (Amiryousefi *et al.*, 2018).

Skewness analysis revealed a highly similar pattern of GC content and GC skew across all species examined (Fig. 7). A positive GC skew indicated that guanine (G) was more abundant than cytosine (C) in the analyzed genome region, whereas a negative GC skew suggested the opposite. The consistent patterns in GC skew and content across these species suggest a conserved DNA structure and stability, implying that the mechanisms governing guanine and cytosine distribution have been preserved within the Solanaceae family (Wang *et al.*, 2023). These metrics provide insights into DNA density, as regions with high GC content are more stable and denser due to the stronger hydrogen bonding between G and C pairs. The uniformity observed between *W. somnifera* var. *abhaica* and other closely related taxa further supports the evolutionary stability of these species and reinforces the accuracy of the assembled plastome of *W. somnifera* var. *abhaica*.

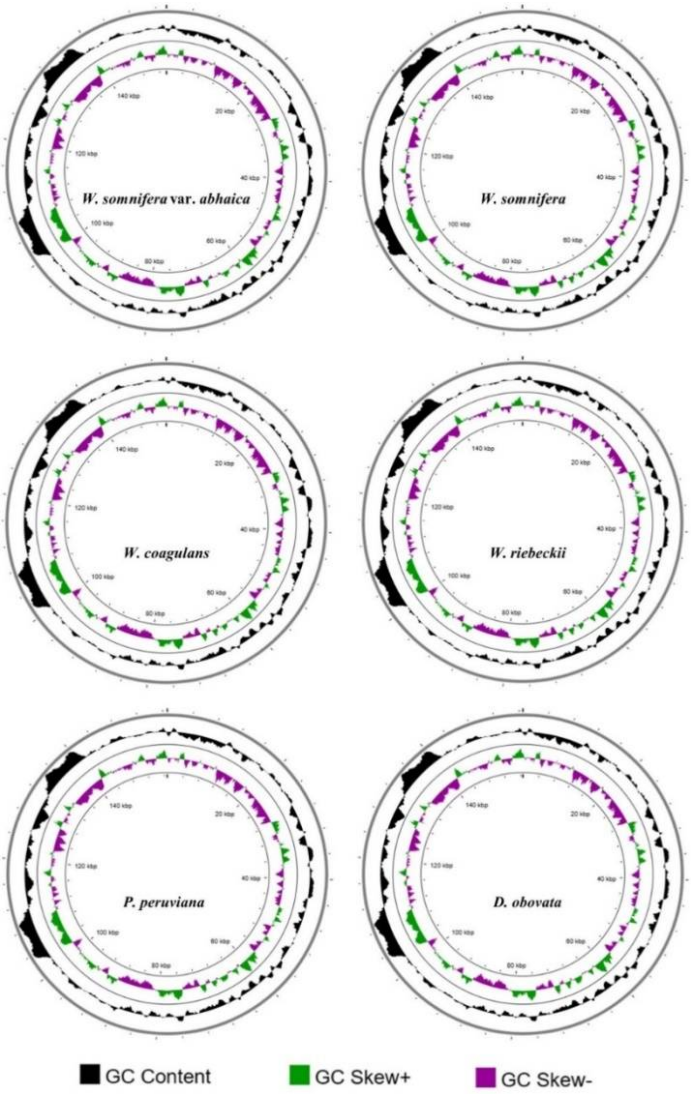

Fig. 7. GC content and skewness analysis of the *W. somnifera* var. *abhaica* along with closely related species.

Comparative genomic assessments

Plastome-wide alignment revealed locally collinear blocks (LCBs) with high similarities (Fig. 8). Gene orders and arrangements were represented by multi-colored mini blocks: white for PCGs, black for tRNAs, green for intron-containing tRNAs, and red for rRNAs. The high similarity of *W. somnifera* var. *abhaica* with other closely related taxa supports the accuracy of its plastome assembly and annotation. The results of the Mauve whole-genome alignment are congruent with findings from other similar studies (Henriquez *et al.*, 2020; Munyao *et al.*, 2020). The collinearity analysis unveiled a high synteny of *W. somnifera* var. *abhaica* with other closely related taxa (Fig. 9). No significant rearrangements were detected among the taxa studied, highlighting their structural similarity and integrity. Sequence identity was notably similar among *W. somnifera*, *W. coagulans*, and *W. riebeckii*. Unlike other *Withania* species, *P. peruviana* and *D. obovata* exhibited syntenic blocks (red) with over 75% sequence similarity (Fig. 9). This high synteny in *P. peruviana* and *D. obovata* suggests that, despite belonging to different genera, these species share significant conserved genomic regions. This conservation could be due to evolutionary constraints or functional necessities that have maintained these sequences (Ding *et al.*, 2022). Table 4 presents a detailed comparative account of the plastomes of various taxa examined in the present study, underscoring differences in chloroplast genome length, GC content, and gene count.

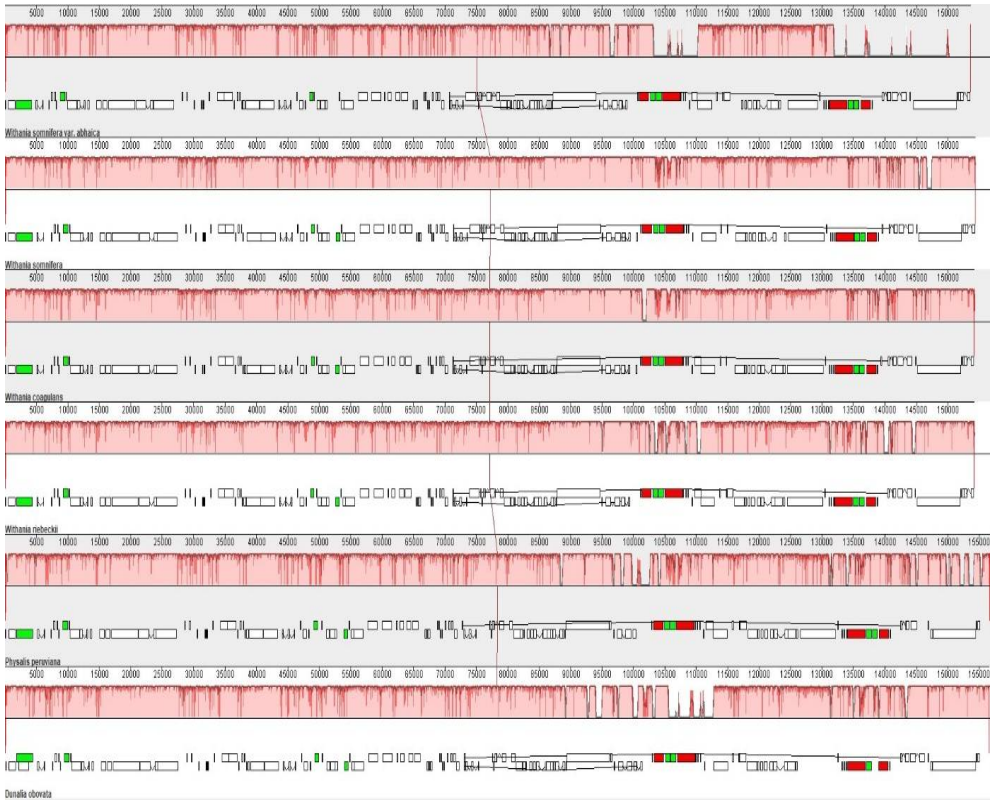

Fig. 8. Mauve progressive alignment of the complete plastome of *W. somnifera* var. *abhaica* showing resemblances with other closely related taxa within Solanaceae.

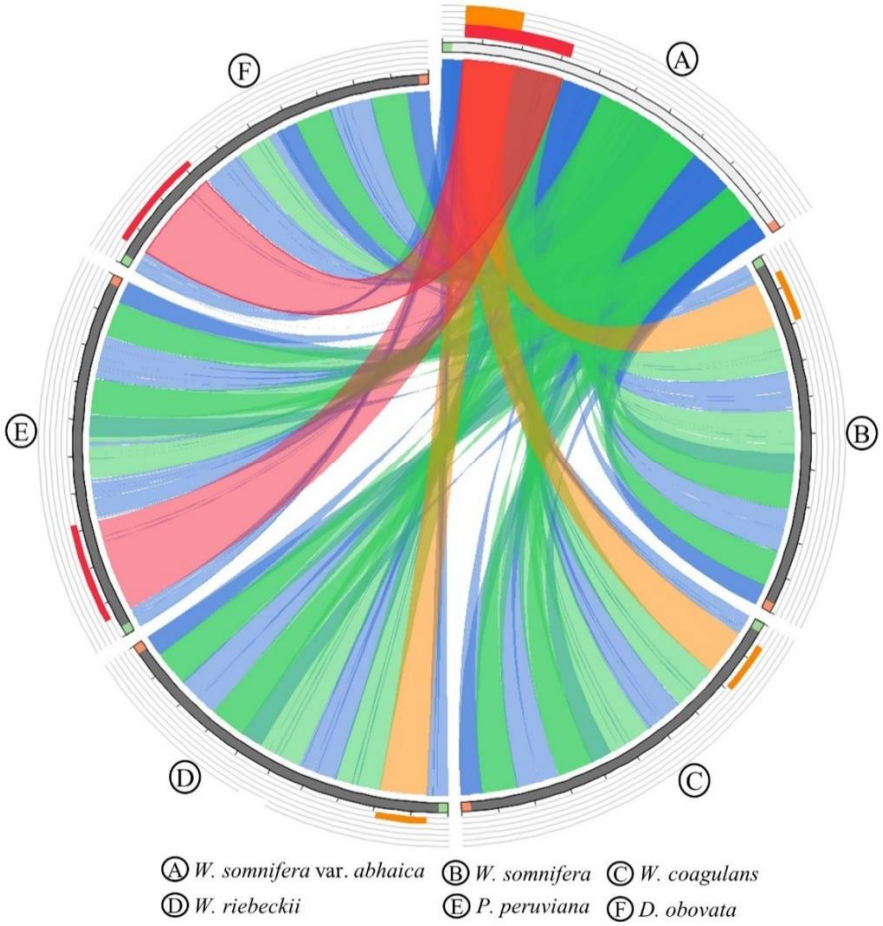

Fig. 9. Synteny analysis of *W. somnifera* var. *abhaica* with other closely related taxa within Solanaceae.

Nucleotide diversity assessments

The nucleotide diversity analysis of the *W. somnifera* var. *abhaica* plastome identified several hypervariable sites (Fig. 10). The mean Pi value across all genomic positions was 0.0048. The most hypervariable site was found in the *psbJ* gene (Pi = 0.04844), followed by *psbA* (Pi = 0.04222), both located in the LSC region. In the SSC region, the *ndhF* gene showed the highest variability (Pi = 0.3389), followed by *ycf1* (Pi = 0.3044). Nucleotide diversity was lower in the IRs compared to the LSC and SSC, reflecting the more conserved nature of the IR regions. Our results align with earlier findings for *Chlorophytum comosum*, *C. gallabatense*, and *Tribulus macropterus* var. *arabicus* (Munyao *et al.*, 2020; Albediwi *et al.*, 2024). Identifying hypervariable genes in the plastome of *W. somnifera* var. *abhaica* is important for developing genetic markers or barcodes. Due to their high variability, these genes can serve as accurate genetic identifiers for distinguishing closely related species or even different subspecies within the same species. These hypervariable barcodes enable precise identification and classification of *Withania* species. Moreover, these markers can enhance the understanding of unusual mutations within a lineage and help to elucidate evolutionary relationships (Breen *et al.*, 2009).

Table 4. Comparison of plastome features of the taxa analyzed in the current investigation.

Taxa	GenBank ID	Total length	GC content (%)	PCGs	tRNAs	rRNAs	Total genes
Withania somnifera var. *abhaica* Nadia, A. Ali & M.S. Alwahibi, var. nov.	OR166175.1	153,621	37.74	86	37	8	131
W. somnifera (L.) Dunal	MK142783.1	154,386	37.67	84	39	8	131
W. frutescens (L.) Pauquy	ON153173.1	153,771	37.73	89	44	8	141
Discopodium penninervium Hochst.	OR400640.1	155,033	37.52	93	38	8	139
Nothocestrum latifolium A. Gray	OR400642.1	155,669	37.53	93	38	8	139
Physalis peruviana L.	NC_026570.1	156,706	37.54	91	37	8	136
P. cordata Houst. *ex* Mill.	NC_072167.1	157,000	37.51	92	38	8	138
Dunalia obovata (Ruiz & Pav.) Dammer	NC_026563.1	156,559	37.69	88	36	8	132
Capsicum baccatum L.	NC_072696.1	157,475	37.64	87	37	8	132
C. lycianthoides Bitter	NC_026551.1	156,583	37.76	87	36	8	131
Jaltomata sinuosa (Miers) Mione	NC_062863.1	156,163	37.91	89	36	8	133
J. bicolor (Ruiz & Pav.) Mione	NC_062862.1	155,459	38.03	89	36	8	133
Solanum corneliomulleri J.F. Macbr.	NC_062080.1	155,544	37.85	93	37	8	138
S. huaylasense Peralta	NC_062081.1	155,571	37.83	93	37	8	138
S. americanum Mill.	NC_062693.1	155,266	37.95	91	37	8	136
S. scabrum Mill.	MT621038.1	155,552	37.90	91	37	8	136
S. villosum Mill.	MT621039.1	155,529	37.89	91	37	8	136
S. nigrum L.	MT621037.1	155,446	37.90	91	37	8	136
Brugmansia arborea (L.) Sweet	NC_081500.1	155,939	37.83	86	37	8	131
Datura stramonium L.	MT610897.1	155,884	37.86	86	37	8	131
D. metel L.	OK040953.1	155,934	37.86	86	38	8	132
Nicandra physalodes (L.) Gaertn.	MN165114.1	156,729	37.78	86	38	8	132
Mandragora caulescens C.B. Clarke	NC_086882.1	154,810	37.98	94	40	8	142
Atropa bella-donna L.	NC_004561.1	156,687	37.56	85	37	8	130
Lycium ferocissimum Miers	MN866909.1	155,894	37.85	86	37	8	131
L. chinense Mill.	MN102357.1	155,736	37.84	89	37	8	134
L. ruthenicum Murray	MT955897.1	154,911	37.91	89	37	8	134
L. qingshuiheense Jiang & Li	NC_084119.1	154,945	37.92	87	37	8	132
Nicotiana tomentosiformis Goodsp.	NC_007602.1	155,745	37.79	82	37	8	127
N. tabacum L.	NC_001879.2	155,943	37.85	84	37	8	129
N. sylvestris Speg.	NC_007500.1	155,941	37.85	81	38	8	127
N. attenuata Torr. *ex* Watson	MG182422.1	155,914	37.86	90	37	8	135
N. undulata Ruiz & Pav.	NC_016068.1	155,863	37.88	89	37	8	134
Petunia exserta Stehmann	MT644125.1	156,597	37.81	88	37	8	133
Mentha spicata L.	OM617844.1	152,048	37.85	88	37	8	133
Phyla nodiflora (L.) Greene	OQ673174.1	154,341	39.19	87	34	8	129

Molecular phylogenetics

A plastome-wide molecular phylogeny within the Solanaceae family was reconstructed that supported the systematic position of *W. somnifera* var. *abhaica* as a novel variety of *W. somnifera* (Fig. 11). The maximum-likelihood (ML) tree showed a strong bootstrap support across most of the clades and subclades. The family Solanaceae consists of seven subfamilies, such as Cestroideae, Goetzeoideae, Nicotianoideae, Petunioideae, Schizanthoideae, Schwenckioideae, and Solanoideae (Olmstead *et al.*, 2008). However, Cp genomes in the NCBI GenBank database were

available for only three subfamilies, *viz*. Solanoideae, Nicotianoideae, and Petunioideae. The reconstructed ML tree provided a well-resolved phylogeny for these three subfamilies, all of which exhibited a monophyletic origin (Fig. 11). *W. somnifera* var. *abhaica* was found to be closely related to other *W. somnifera* accessions and grouped with other members of the tribe Physaleae within the Solanoideae subfamily. Tribe Physaleae displayed a monophyletic origin with 100% bootstrap support and showed a close relationship with the member taxa of the tribe Capsiceae. The tribe Solaneae, represented by eight taxa, also demonstrated a monophyletic nature with 100% bootstrap support. Similarly, tribe Datureae exhibited strong bootstrap support with its two representative genera, *Brugmansia* and *Datura*. The tribes Nicandreae and Mandragoreae, each represented by one species, formed a cluster. The remaining two tribes, Hyoscyameae and Lycieae grouped together with robust bootstrap support. The subfamily Nicotianoideae was represented solely by the tribe Nicotianeae, which depicted a well-resolved monophyletic nature. *Petunia exserta* was the only representative of the subfamily Petunioideae, occupying a distinct position in the ML tree.

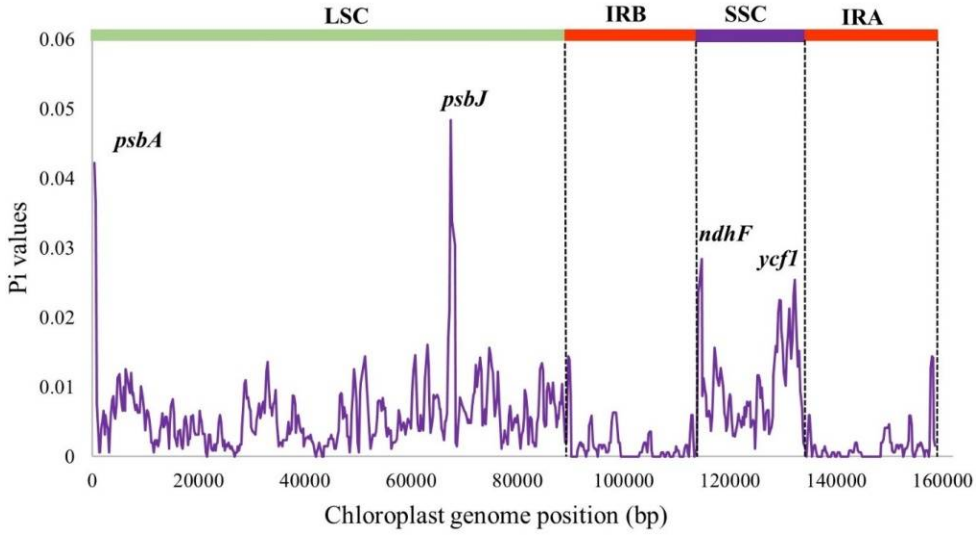

Fig. 10. Nucleotide diversity of *W. somnifera* var. *abhaica* Cp genome elucidating hypervariable barcodes across LSC, SSC and IRs compartments.

The accurate systematic positioning of the *W. somnifera* var. *abhaica* plastome justifies its assembly. Mehmood *et al.* (2020) constructed ML-based phylogenetic tree with 23 taxa of Solanaceae to validate the assembly of *W. somnifera* (MK142783). In their analysis, the tribe Physaleae exhibited a close affinity with Capsiceae tribe, while the tribe Hyoscyameae clustered with the tribe Lycieae. These tribal relationships are further supported by our current investigation (Fig. 11). Mehmood *et al.* (2020) included two species of Nicotianoideae, *viz. N. sylvestris* and *N. tabacum* in their ML tree. In the present study, we have included five species of *Nicotiana*, where *N. tabacum* clustered with *N. sylvestris*, and these consistent findings reinforce the well-resolved phylogeny (Fig. 11).

Given the distinctive characteristics of this novel variety of *W. somnifera*, including reduced plant height, elliptic-elongated, thick and puberulous leaves resembling semi-succulent morphological nature, a bifurcated fruiting calyx tip, c. 0.5 mm in length, with each bifurcation being botuliform in shape, and robust phylogenetic evidence supporting its uniqueness, we herein

propose that the collected ecotype accession be recognized as a new variety: *Withania somnifera* (L.) Dunal var. *abhaica* Nadia, A. Ali & M.S. Alwahibi, *var. nov.* This new variety is named after its '*Type'* locality.

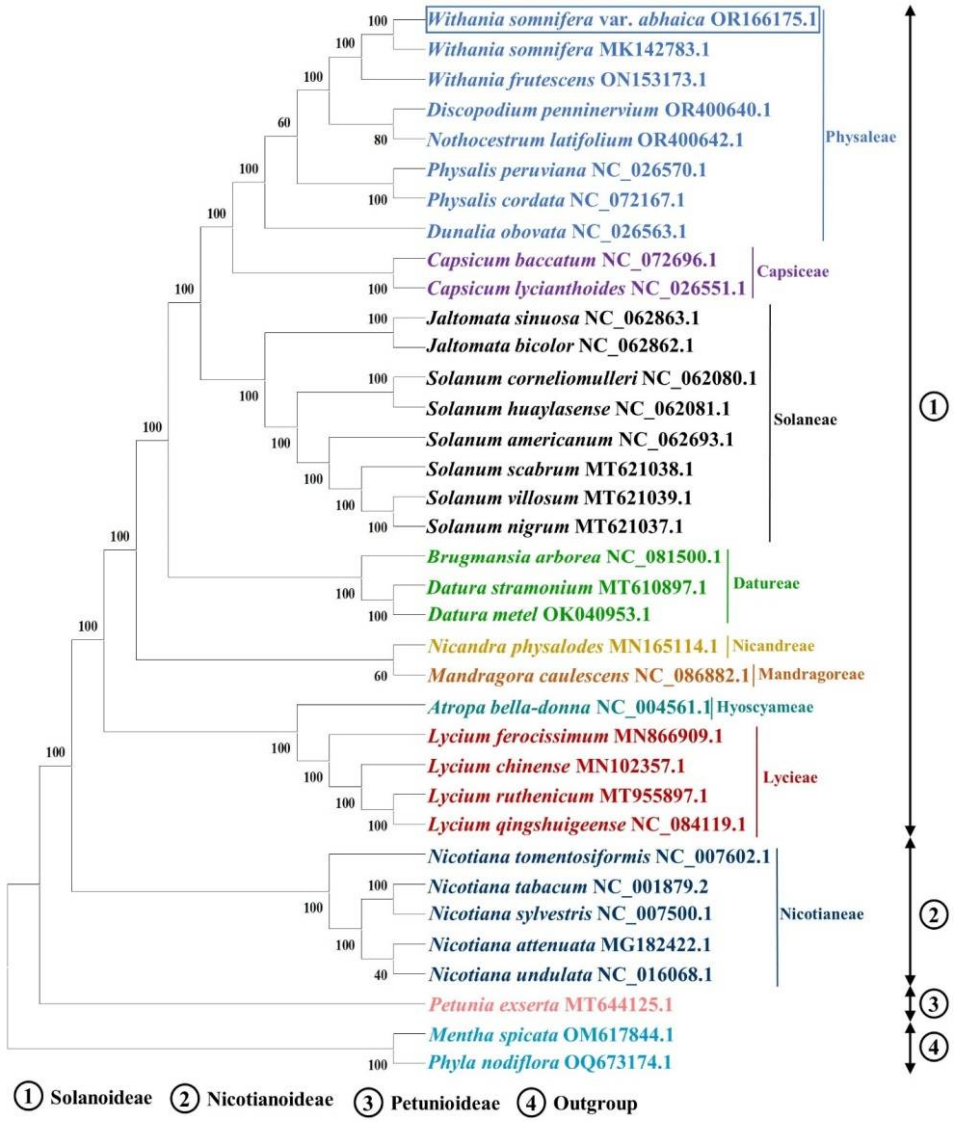

Fig. 11. Maximum-likelihood (ML) tree representing plastome-wide phylogenetic relationships of *W. somnifera* var. *abhaica.*

The complete chloroplast genome presented in this study will contribute valuable new data to the GenBank repository, providing an essential resource for comprehensive molecular phylogenetic and dating analyses. The identified hypervariable barcodes will advance DNA barcoding efforts, offering precise tools for the taxonomic identification of the medicinally significant taxon. Additionally, this study lays the foundation for deeper insights into the evolutionary processes and genetic diversity within the Solanaceae family.

Acknowledgements

The authors extend their appreciation to the Researchers Supporting Project number (RSP2025R306), King Saud University, Riyadh, Saudi Arabia. This research was also funded by the Korean Research Institute of Bioscience and Biotechnology (KRIBB) Initiative Program of the Republic of Korea (KGM4582423).

References

Ahmed, S.S. and Rahman, M.O. 2024. Deciphering the complete chloroplast genome sequence of *Meconopsis torquata* Prain: Insights into genome structure, comparative analysis and phylogenetic relationship. Heliyon **10**(2024): e36204.

Albediwi, A.S., Ali, M.A., Alwahibi, M.S., Ahmed, S.S., Rahman, M.O., Kim, S.Y., Elshikh, M.S. and Alsuhaimi, N.M. 2024. Unveiling the complete chloroplast genome of *Tribulus macropterus* var. *arabicus* (Hosni) Al-Hemaid & J. Thomas: Genome structure, comparative analysis and phylogeny. Bangladesh J. Plant Taxon. **31**(1): 1–14.

Alzahrani, D. and Albokhari, E. 2022. Characterization of plastid genome of a medicinal plant *Capparis decidua* (Forsk) Edgew. App. Biol. Res. **24**(2): 156–167.

Amiryousefi, A., Hyvönen, J. and Poczai, P. 2018. The chloroplast genome sequence of bittersweet (*Solanum dulcamara*): Plastid genome structure evolution in Solanaceae. PLoS One **13**(4): e0196069.

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.

Breen, A.L., Glenn, E., Yeager, A. and Olson, M.S. 2009. Nucleotide diversity among natural populations of a North American poplar (*Populus balsamifera*, Salicaceae). New Phytol. **182**(3): 763–773.

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.

Dar, N.J., Hamid, A. and Ahmad, M. 2015. Pharmacologic overview of *Withania somnifera*, the Indian Ginseng. Cell. Mol. Life Sci. **72**: 4445–4460.

Darling, A.C., Mau, B., Blattner, F.R. and Perna, N.T. 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Res. **14**(7): 1394–1403.

Darzentas, N. 2010. Circoletto: Visualizing sequence similarity with Circos. Bioinform. **26**(20): p2620.

Ding, S.X., Li, J.C., Hu, K., Huang, Z.J. and Lu, R.S. 2022. Morphological characteristics and comparative chloroplast genome analyses between red and white flower phenotypes of *Pyracantha fortuneana* (Maxim.) Li (Rosaceae), with implications for taxonomy and phylogeny. Genes **13**(12): 2404.

Dobrogojski, J., Adamiec, M. and Luciński, R. 2020. The chloroplast genome: A review. Acta Physiol. Plant. **42**(6): 98.

Grant, J.R., Enns, E., Marinier, E., Mandal, A., Herman, E.K., Chen, C.Y., Graham, M., Domselaar, G.V. and Stothard, P. 2023. Proksee: in-depth characterization and visualization of bacterial genomes. Nucleic Acids Res. **51**(W1): W484–W492.

Greiner, S., Lehwark, P. and Bock, R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: Expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. **47**(W1): W59–W64.

Henriquez, C.L., Abdullah, Ahmed, I., Carlsen, M.M., Zuluaga, A., Croat, T.B. and McKain, M.R. 2020. Molecular evolution of chloroplast genomes in Monsteroideae (Araceae). Planta **251** (2020): 72.

Jongsun, P., Hong, X.I. and Sang-Hun, O.H. 2020. Comparative chloroplast genomics and phylogenetic analysis of the *Viburnum dilatatum* complex (Adoxaceae) in Korea. Korean J. Plant Taxon. **50**(1): 8–16.

Katoh, K., Kuma, K.I., Toh, H. and Miyata, T. 2005. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. **33**(2): 511–518.

Kulkarni, S.K. and Dhir, A. 2008. *Withania somnifera*: An Indian ginseng. Prog. Neuropsychopharmacol. Biol. Psychiatry **32**(5): 1093–1105.

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.

Lenz, H., Hein, A. and Knoop, V. 2018. Plant organelle RNA editing and its specificity factors: Enhancements of analyses and new database features in PREPACT 3.0. BMC Bioinform. **19**: 1–18.

Librado, P. and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25**(11): 1451–1452.

Liu, S., Ni, Y., Li, J., Zhang, X., Yang, H., Chen, H. and Liu, C. 2023. CPGView: A package for visualizing detailed chloroplast genome structures. Mol. Ecol. Resour. **23**(3): 694–704.

Ma, H., Zhang, E., An, Y., Wei, Y. and Zhang, L. 2024. Characterization of the complete chloroplast genome of the rare medicinal plant: *Mandragora caulescens* (Solanaceae). Mitochondrial DNA B Resour. **9**(6): 812–817.

Mehmood, F., Shahzadi, I., Ahmed, I., Waheed, M.T. and Mirza, B. 2020. Characterization of *Withania somnifera* chloroplast genome and its comparison with other selected species of Solanaceae. Genomics **112**(2): 1522–1530.

Munyao, 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.

Okonechnikov, K., Golosova, O., Fursov, M. and Ugene Team. 2012. Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics **28**(8): 1166–1167.

Olmstead, R.G., Bohs, L., Migid, H.A., Santiago-Valentin, E., Garcia, V.F. and Collier, S.M. 2008. A molecular phylogeny of the Solanaceae. Taxon **57**(4): 1159–1181.

Paul, S., Chakraborty, S., Anand, U., Dey, S., Nandy, S., Ghorai, M., Saha, S.C., Patil, M.T., Kandimalla, R., Proćków, J. and Dey, A. 2021. *Withania somnifera* (L.) Dunal (Ashwagandha): A comprehensive review on ethnopharmacology, pharmacotherapeutics, biomedicinal and toxicological aspects. Biomed. Pharmacother. **143**: 112175.

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.

Rahman, M.A., Mossa, J.S., Al-Said, M.S. and Al-Yahya, M.A. 2004. Medicinal plant diversity in the flora of Saudi Arabia 1: A report on seven plant families. Fitoterapia 75(2): 149–161.

Ravi, V., Khurana, J.P., Tyagi, A.K. and Khurana, P.J.P.S. 2008. An update on chloroplast genomes. Plant Syst. Evol. **271**: 101–122.

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.

Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L. and Liu, C. 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. **47**(W1): W65–W73.

Tamura, K., Stecher, G. and Kumar, S. 2021. MEGA11: Molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. **38**(7): 3022–3027.

Visweswari, G., Christopher, R. and Rajendra, W. 2013. Phytochemical screening of active secondary metabolites present in *Withania somnifera* root: Role in traditional medicine. Int. J. Pharm. Sci. Res. **4**(7): 2770–2776.

Wang, L., Liang, J., Shang, Q., Sa, W. and Wang, L. 2020. The complete plastome of *Sorbaria kirilowii*: Genome structure, comparative analysis, and phylogenetic implications. Mol. Biol. Rep. **47**: 9677–9687.

Wang, X., Zhang, R., Wang, D., Yang, C., Zhang, Y., Sui, M., Quan, J., Sun, Y., You, C. and Shen, X. 2023. Molecular structure and variation characteristics of the plastomes from six *Malus baccata* (L.) Borkh. individuals and comparative genomic analysis with other *Malus* species. Biomolecules **13**(6): 962.

Zhang, Z., Zhang, Y., Song, M., Guan, Y. and Ma, X. 2019. Species identification of *Dracaena* using the complete chloroplast genome as a super-barcode. Front. Pharmacol. **10**: 1441.

(Manuscript received on 7 February 2024; revised on 27 November 2024)