Bangladesh J. Plant Taxon. **32**(2): 139-160, 2025 (December) DOI: https://doi.org/10.3329/bjpt.v32i2.86664

THE FIRST CHLOROPLAST AND NUCLEAR GENOME ASSEMBLIES OF *PRASOXYLON EXCELSUM* (MELIACEAE) REVEAL PHYLOGENETICS, EVOLUTIONARY AND FUNCTIONAL INSIGHTS

M. OLIUR RAHMAN[1], SHEIKH SUNZID AHMED[1*], MOHAMMAD AJMAL ALI[2] AND JOONGKU LEE[3]

[1] *Department of Botany, Faculty of Biological Sciences, University of Dhaka, Dhaka 1000, Bangladesh*
[2]*Department of Botany and Microbiology, College of Science, King Saud University, Riyadh-11451, Saudi Arabia*
[3]*Department of Environment and Forest Resources, College of Agricultural Life Science, Chungnam National University, Daejeon, South Korea*

Keywords: Molecular phylogenetics; *Prasoxylon excelsum*; Cp genome; Molecular dating; Nuclear genome; BUSCO; KEGG pathway.

Abstract

Prasoxylon excelsum (Spreng.) Mabb., a medicinal tree of the family Meliaceae, is traditionally used to treat diarrhoea, dysentery, and food poisoning. Despite its ethnobotanical significance, no genomic resources have been available to date. This study presents the first *de novo* assembly and characterization of both the chloroplast (Cp) and nuclear genomes of *P. excelsum*. The complete Cp genome spans 160,299 bp and exhibits a typical quadripartite structure, comprising a large single-copy (LSC) region of 87,832 bp, a small single-copy (SSC) region of 18,360 bp, and a pair of inverted repeats (IRs) of 27,054 bp each. The plastome encodes 81 protein-coding genes, 37 transfer RNAs, and eight ribosomal RNAs. A total of 98 simple sequence repeats (SSRs) were identified, predominantly mononucleotide motifs (68). Nucleotide diversity analysis revealed *atpB* and *ycf1* as the most variable regions, suggesting their potential as DNA barcodes. Phylogenetic reconstruction using the maximum-likelihood method confirmed the placement of *P. excelsum* within the subfamily Melioideae. Molecular dating analysis indicated that *P. excelsum* diverged during the late Burdigalian age of the Miocene epoch, approximately 16.21 million years ago. The nuclear genome assembly totaled 676.45 Mb with an N50 of 3,686 bp, showing 87.8% BUSCO completeness. Gene prediction identified 44,465 genes, with KEGG analysis indicating enrichment in metabolic pathways. These findings provide a valuable foundation for evolutionary, phylogenomic, and functional studies in *Prasoxylon* and related Meliaceae species, aiding future research in conservation genomics, phytochemistry, and molecular breeding.

Introduction

Prasoxylon excelsum (Spreng.) Mabb., belonging to Meliaceae, is a medicinally and ecologically significant tree distributed across Bangladesh, China, Myanmar, Malaysia, Sri Lanka, and India (Govaerts *et al.*, 2021). Members of the family Meliaceae are known for their diverse secondary metabolites and ethnobotanical importance, including applications in traditional medicine, timber, and agroforestry (Oyedeji-Amusa *et al.*, 2021). *P. excelsum* typically grows up to 13 m in height and the branchlets are brown to reddish-brown and glabrous, while leaves are paripinnate with seven to nine elliptic to oblong leaflets. The tree produces small white flowers in

*Corresponding author, E-mail: sunzid79@gmail.com

axillary thyrses and bears globose to pyriform fruits containing seeds with bright red testa. Flowering takes place from September to November, followed by fruit maturation between April and June (Brach and Song, 2006). *P. excelsum* has notable indigenous medicinal importance, with tender shoots and leaves traditionally consumed to treat diarrhoea, dysentery, and food poisoning (Lalmuanpuii *et al.*, 2024). Recent phytochemical investigations have revealed the presence of bioactive triterpenoids in its stem bark, exhibiting cytotoxic activity against MCF-7 breast cancer cells (Meilanie *et al.*, 2022; Zainuddin *et al.*, 2020). Despite these ethnomedicinal insights, *P. excelsum* remains poorly studied at the genomic level. Comprehensive chloroplast (Cp) and nuclear genome assemblies are therefore essential to provide foundational resources for understanding its biosynthetic pathways, evolutionary relationships, and potential biotechnological applications (Clarke *et al.*, 2011).

The chloroplast genome serves as a powerful tool for resolving phylogenetic relationships and elucidating evolutionary history due to its conserved structure, uniparental inheritance, and moderate mutation rate (Ahmed and Rahman, 2024). In Meliaceae, where morphological diversity can obscure species boundaries, complete plastome sequencing of *Prasoxylon excelsum* enables a comprehensive analysis of all protein-coding genes, transfer RNAs (tRNAs), and ribosomal RNAs (rRNAs), thereby improving species delimitation and clarifying its position within the family (Nie *et al.*, 2025). Beyond phylogenetics, the nuclear genome offers complementary insights, particularly for medicinally important plants. Characterizing the nuclear genome of *P. excelsum* enables the identification of genes involved in key biological processes, metabolic pathways, and secondary metabolite biosynthesis through gene ontology (GO) analyses (Romdhane *et al.*, 2025). Integrating chloroplast and nuclear genomic data provides a holistic view of the species' genetic architecture, facilitating evolutionary studies, functional genomics, and conservation strategies. Together, these genome resources establish a foundation for linking phylogenomic relationships with the molecular basis of medicinal traits, paving the way for future research into bioactive compound biosynthesis and supporting the systematic classification of *P. excelsum* within Meliaceae (Chakravartty and Neelapu, 2023).

The rapid growth of next-generation sequencing (NGS) technologies has revolutionized the exploration of plant genomes, enabling the accurate assembly and analysis of both plastid and nuclear genomes. By leveraging publicly available datasets, it is possible to generate complete plastome assemblies and draft nuclear genomes without the need for fresh biological material, thereby reducing technical and financial constraints (Samji *et al.*, 2023; Ramadan *et al.*, 2023). This approach not only facilitates phylogenetic and evolutionary studies using plastid genomes but also supports the investigation of nuclear genomic features, including genes involved in key biological processes, metabolic pathways, and secondary metabolite biosynthesis. Utilizing existing sequence data enhances reproducibility and transparency, as raw reads remain accessible for validation and further analysis. In addition, integrating plastid and nuclear genomic information opens new avenues for comparative genomics, functional annotation, and evolutionary research, even in species that are rare or understudied (Chakravartty and Neelapu, 2023, 2024).

To date, no comprehensive genomic study has been conducted for *P. excelsum*, leaving significant gaps in our understanding of its plastome-wide phylogenetic position, evolutionary history, and the genetic basis of its medicinal properties. In this study, we present the first complete Cp genome together with a draft nuclear genome for *P. excelsum*, integrating phylogenetic reconstruction, molecular dating, and functional gene annotation analyses. These genomic resources are expected to clarify its systematic placement within Meliaceae, provide insights into lineage diversification, and identify genes associated with key biological processes and secondary metabolite biosynthesis.

Materials and Methods

Retrieval of NGS reads

Raw Illumina sequencing data for *Prasoxylon excelsum* (accession SRR7122084) were obtained from the NCBI Sequence Read Archive (SRA). The dataset, generated using the BGISEQ-500 platform, comprised approximately 325.91 million paired-end reads with an average length of 100 bp and an insert size of 200 bp, totaling around 65.2 Gb of sequencing output (43.5 Gb of clean data). Prior to assembly, read quality and base composition were assessed using FastQC to ensure high-confidence input for subsequent analyses (Brandine and Smith, 2021).

Cp genome: Assembly and annotation

The complete Cp genome was assembled from Illumina short-read data using GetOrganelle v1.7.7.0 (Jin *et al*., 2020), which efficiently assembles circular organellar genomes. Assembly coverage and sequence integrity were evaluated in UGENE v.52.1 (Okonechnikov *et al*., 2012). Gene annotation was conducted through the CPGAVAS2 (Shi *et al*., 2019) web platform and further validated using the CPGView server (Liu *et al*., 2023) to ensure annotation accuracy. A circular genome map was subsequently generated with Chloroplot (Zheng *et al*., 2020). The finalized and annotated plastome has been deposited in the GenBank database under the accession number PQ790014.1.

SSR and IR evaluation

Microsatellite motifs within the *P. excelsum* plastome were identified using MISA-Web to characterize the distribution and composition of Simple Sequence Repeats (SSRs) (Beier *et al*., 2017). To investigate structural variation across the inverted repeat (IR) boundaries, the assembled plastome was analyzed through IRscope (Amiryousefi *et al*., 2018). The annotated GenBank file of *P. excelsum* was compared with plastomes of closely related Meliaceae species to visualize and interpret junction dynamics among the large single-copy (LSC), small single-copy (SSC), and IR regions.

Comparative plastomics

Genome-wide synteny and structural variations within the *P. excelsum* plastome were assessed through multiple sequence alignment (MSA) using Mauve v.20150226 (Darling *et al*., 2004). Collinearity patterns among homologous regions were further visualized using the Circoletto platform under default configurations (Darzentas, 2010), providing insights into conserved and rearranged genomic segments. In addition, GC-skewness distribution across the plastome was analyzed via the Proksee server (Grant *et al*., 2023) to examine compositional asymmetry and strand bias.

Nucleotide diversity

Complete Cp genomes of *P. excelsum* and its related taxa were subjected to MSA using MAFFT to ensure accurate positional homology across genomes (Katoh and Standley, 2013). The aligned sequences were subsequently analyzed in DnaSP v.5 (Librado and Rozas, 2009) to quantify nucleotide diversity (π) and identify highly variable regions within the plastome, offering insights into mutational hotspots and potential molecular markers.

Molecular phylogenetics and dating

Phylogenetic inference was conducted in MEGA v.11 (Tamura *et al*., 2021) using the Maximum Likelihood (ML) approach to reconstruct evolutionary relationships among *P. excelsum* and related Meliaceae taxa. Divergence time estimation was subsequently performed with the RelTime-ML module, integrating fossil-calibrated nodes obtained from the TimeTree database (Kumar *et al*., 2017).

Nuclear genome assembly

The nuclear genome was assembled *de novo* using ABySS v.2.3.10 with a k-mer length of 41 bp to achieve optimal scaffold continuity. Reads with base quality below Q3 were trimmed, and low-quality positions were masked as 'N'. Assembly parameters included a minimum scaffold length of 200 bp and bubble sequence identity of 0.9. Scaffolds were built using a minimum of 10 read pairs and refined with 15–20 supporting pairs, permitting a distance estimation error of 6 bp (Simpson *et al.*, 2009).

Quality assessment of the nuclear genome

The completeness and structural quality of the assembled nuclear genome were evaluated using QUAST v.5.3.0 and BUSCO v.5.0.0 (Gurevich *et al.*, 2013; Seppey *et al*., 2019). In QUAST, scaffolds shorter than 500 bp were excluded, and a minimum local misassembly size of 200 bp was applied to identify potential assembly errors. BUSCO analysis was performed using the Viridiplantae and Embryophyta lineage datasets, with AUGUSTUS employed for gene prediction and an E-value cutoff of 0.001 for BLAST searches.

Functional annotation of the nuclear genome

The assembled scaffolds from ABySS were first subjected to gene prediction using AUGUSTUS (Stanke *et al.*, 2006), with *Arabidopsis thaliana* set as the model organism. This produced predicted protein sequences in FASTA format, which were used for downstream functional analysis. To identify homologous proteins, a local BLAST protein database was created using the complete *Arabidopsis thaliana* proteome (TAIR10 release) with the NCBI makeblastdb tool. Predicted protein sequences from AUGUSTUS were queried against this database using BLASTp (Camacho *et al.*, 2009) with default parameters to detect significant sequence similarity. The resulting list of candidate genes with significant BLAST hits was compiled and subjected to functional enrichment analysis using ShinyGO v.0.85 (Ge *et al*., 2020).

Results and Discussion

Assembled and annotated plastome

The complete chloroplast genome of *P. excelsum* was 160,299 bp in length and exhibited the typical quadripartite structure, comprising a large single-copy (LSC) region, a small single-copy (SSC) region, and two inverted repeat (IR) regions (Fig. 1). The overall nucleotide composition of the plastome was A+T biased, with 62.43% A+T and 37.57% G+C content. Analysis of individual regions revealed distinct nucleotide composition patterns (Table 1).

Table 1. Composition of the nucleotides in the plastome of *P. excelsum*.

Compartments	A+T (%)	C+G (%)	T (U) (%)	G (%)	A (%)	C (%)
IRA	57.25	42.75	28.63	22.08	28.62	20.67
IRB	57.25	42.75	28.62	20.67	28.63	22.08
LSC	64.45	35.56	32.98	17.20	31.47	18.36
SSC	68.07	31.93	34.20	16.66	33.87	15.27
Plastome	62.43	37.57	31.65	18.54	30.78	19.02

Both IRa and IRb showed identical composition, with 57.25% A+T and 42.75% G+C, reflecting the relatively higher GC content typical of inverted repeats. The LSC region contained 64.45% A+T, whereas the SSC region exhibited the highest A+T content at 68.07%, indicating strong AT richness in the single-copy regions (Table 1).

The sequencing coverage depth of the *P. excelsum* chloroplast genome was analyzed to assess the uniformity and reliability of the assembly. The maximum coverage was 50,925X at position 69,581 bp, while the minimum coverage was 2,793X at the terminal position of the plastome (160,299 bp). The average coverage depth across the entire genome was 9,673.05X (Fig. 2). These results indicate that the chloroplast genome assembly was highly reliable, with consistently high sequencing depth across most regions. The slight variation in coverage, with the lowest depth observed at the genome terminus, is typical of plastome assemblies due to mapping biases at the edges of circular genomes. The findings were congruent with previous studies (Ahmed and Rahman, 2024; 2025).

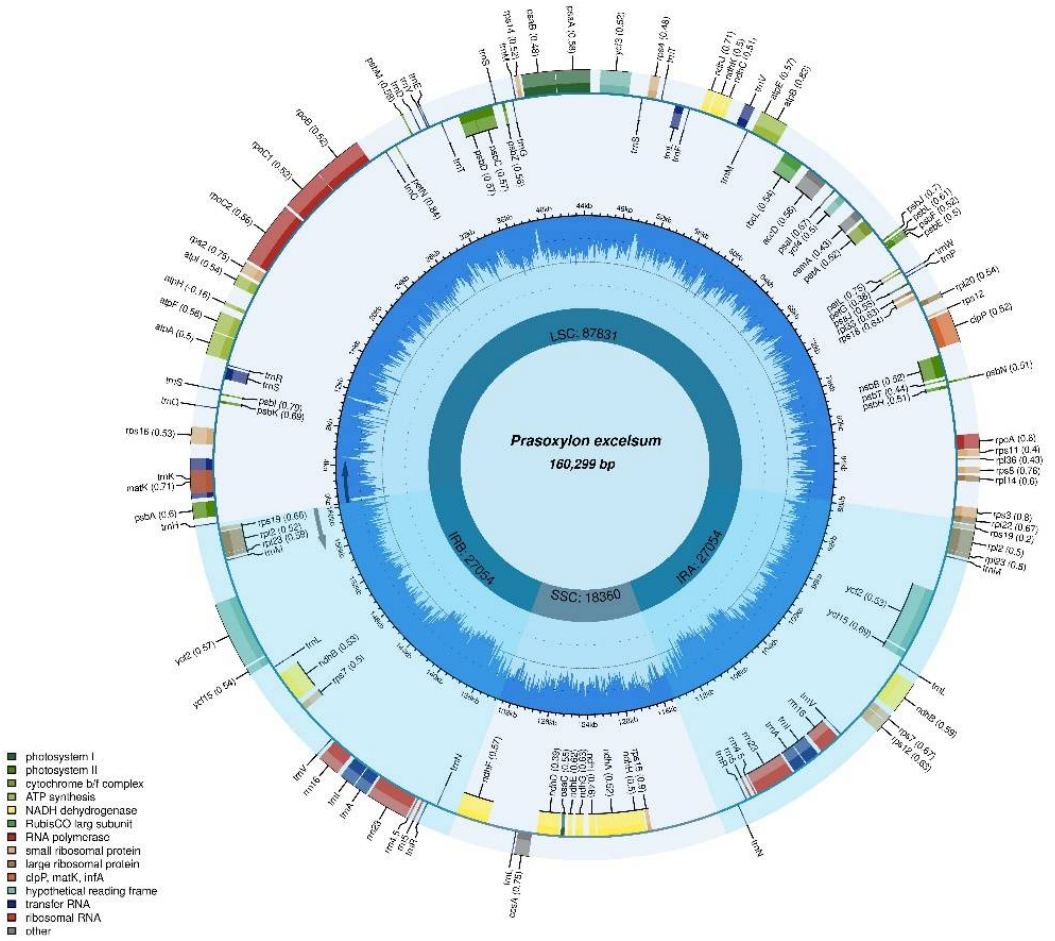

Fig. 1. Map of the assembled and annotated Cp genome of *P. excelsum*.

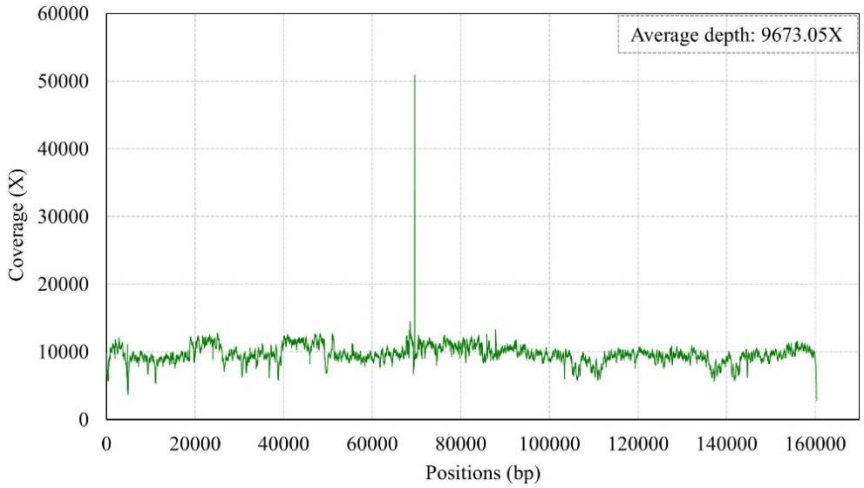

Fig. 2. Coverage depth analysis of the plastome of *P. excelsum*.

The chloroplast genome of *P. excelsum* contained 126 genes, including 81 protein-coding genes (PCGs), 37 tRNAs, and eight rRNAs. The PCGs were primarily associated with photosynthesis (*psaA*, *psbA*, *petA*, *atpA*, *rbcL* etc.), transcription and translation (*rpoA*, *rpl2*, *rps19* etc.), and energy metabolism (*ndhA*, *ndhB*, *ndhC* etc.), as well as other functions such as maturase (*matK*), protease (*clpP*), c-type cytochrome synthesis (*ccsA*), and envelope membrane protein (*cemA*). Several conserved open reading frames (*ycf* genes) were also identified (Table 2).

Table 2. Protein-coding genes present in the plastome of *P. excelsum*.

Name of genes	Group of genes
rpoA, rpoB, rpoC1, rpoC2	DNA dependent RNA polymerase
petA, petG, petL, petN	Subunits of cytochrome b/f complex
accD	Subunit of Acetyl-CoA-carboxylase
cemA	Envelop membrane protein
rpl14, rpl2 (×2), *rpl20, rpl22, rpl23* (×2), *rpl33, rpl36*	Large subunit of ribosome
ycf15 (×2), *ycf2* (×2), *ycf4*	Conserved open reading frames
rps11, rps12, rps14, rps15, rps16, rps18, rps19 (×2), *rps2, rps3, rps4, rps7* (×2), *rps8*	Small subunit of ribosome
psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ, ycf3	Subunits of photosystem II
rbcL	Subunit of rubisco
ccsA	c-type cytochrom synthesis gene
psaA, psaB, psaC, psaI, psaJ	Subunits of photosystem I
matK	Maturase
clpP	Protease
ndhA, ndhB (×2), *ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*	Subunits of NADH-dehydrogenase
atpA, atpB, atpE, atpF, atpH, atpI	Subunits of ATP synthase

SSR and IR evaluation

A total of 98 SSRs were identified in the chloroplast genome of *P. excelsum*, comparable to those observed in *Aphanamixis polystachya* (98), *Aglaia odorata* (104), and *Lansium domesticum* (83). Among the SSRs in *P. excelsum*, mononucleotide repeats were the most abundant (68), followed by tetranucleotide (9), dinucleotide (12), trinucleotide (7), and pentanucleotide repeats (2) (Fig. 3). These results indicate that mononucleotide repeats dominate the plastomes of Meliaceae species, whereas other repeat types occur at lower frequencies. The overall SSR distribution was similar among the four species analyzed, suggesting a conserved pattern of simple sequence repeats in their chloroplast genomes (Nie *et al*., 2025). Comparative analysis of the IR regions in *P. excelsum* and related members of the subfamily Melioideae revealed both expansion and contraction events (Fig. 4). Within the subfamily, the LSC regions ranged from 87,062 to 88,173 bp, while the SSC regions varied from 18,228 to 18,709 bp, and the IRs ranged from 26,958 to 27,089 bp. The *rpl22* gene was consistently located at the LSC/IRB junction, and the *trnH* gene was positioned near the LSC/IRA border across all species. The IR regions of *P. excelsum* were expanded relative to *Aphanamixis polystachya*, *Lansium domesticum*, *Leplaea laurentii*, *Heynea velutina*, *Sandoricum koetjape*, *Melia azedarach*, *Azadirachta excelsa*, and *Azadirachta indica*. In contrast, IR contraction was observed when compared with *Aglaia odorata*, *Turraeanthus africanus*, *Leplaea thompsonii*, and *Cipadessa cinerascens* (Fig. 4).

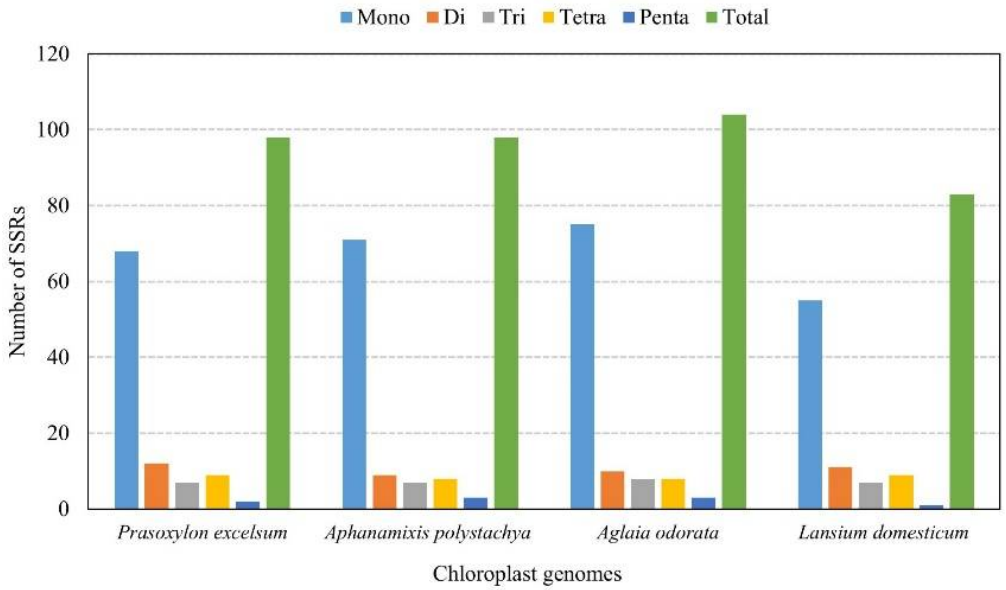

Fig. 3. Evaluation of SSRs in *P. excelsum* plastome and other closely related members.

Comparative plastomics

Comparative analysis of the *P. excelsum* Cp genome revealed high structural similarity with the plastomes of *A. polystachya*, *A. odorata*, and *L. domesticum* (Fig. 5). Mauve alignment and local collinear block (LCB) analysis showed that the plastomes were largely collinear, with minimal structural rearrangements. The multi-colored blocks in the alignment represented conserved genomic regions, where red blocks indicated rRNA genes, black blocks corresponded to tRNA genes, green blocks indicated intron-containing tRNAs, and white blocks represented

PCGs. These results suggest that the overall gene order and structure of *P. excelsum* plastome are highly conserved within the subfamily Melioideae. Synteny analysis revealed a high degree of conservation in gene order and arrangement among closely related species, including *L. domesticum*, *A. odorata*, and *A. polystachya* (Fig. 6).

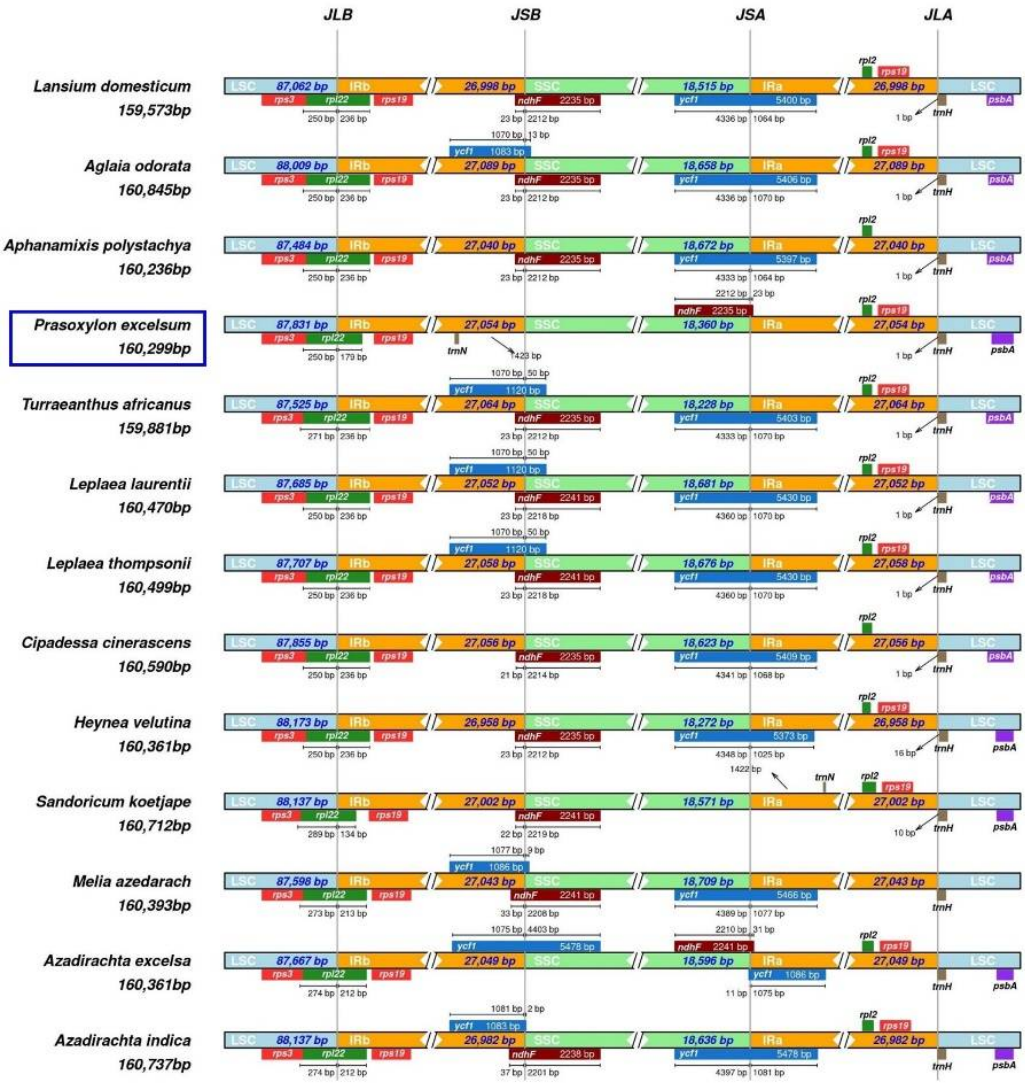

Fig. 4. Evaluation of expansion and contraction of IRs within the subfamily Melioideae.

No large-scale genome rearrangements, inversions, or translocations were detected, indicating that the overall plastome architecture is highly conserved among these taxa. The analysis revealed extensive collinear blocks across the plastomes, reflecting strong conservation of both coding and non-coding regions. GC-skewness analysis of the *P. excelsum* plastome unraveled a balanced distribution of guanine (G) and cytosine (C) content across the genome (Fig. 7).

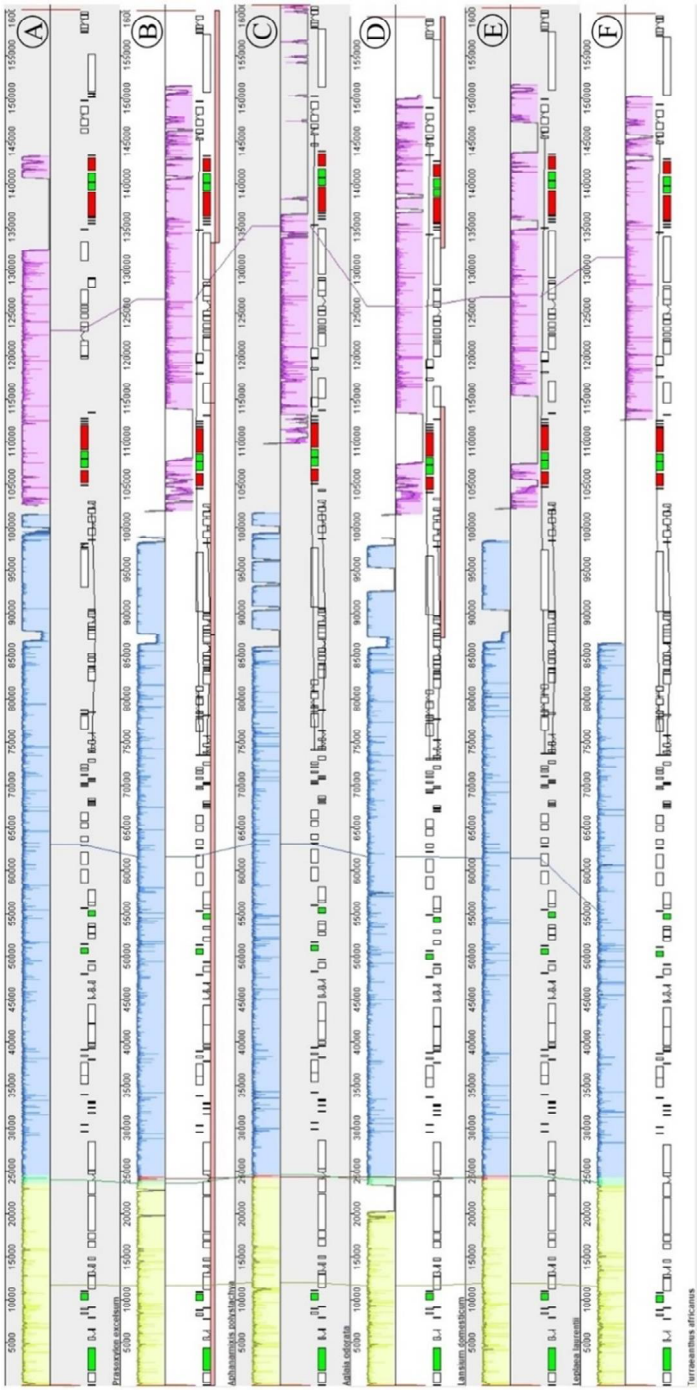

Fig. 5. Locally Collinear Block analysis of the Cp genomes of *P. excelsum* and closely related members illustrating conserved genomic regions. A. *P. excelsum*, B. *Aphanamixis polystachya*, C. *Aglaia odorata*, D. *Lansium domesticum*, E. *Leplaea laurentii*, F. *Turraeanthus africanus.*

The overall GC skew pattern, characterized by alternating positive and negative regions, was highly similar to those observed in closely related taxa, indicating that the bias in nucleotide composition is conserved within the subfamily Melioideae. Regions with positive GC skew corresponded to areas where guanine content exceeded cytosine, whereas negative skew regions indicated higher cytosine content relative to guanine. The similarity in these patterns across the studied species suggests that the replication and transcriptional dynamics of the chloroplast genome are largely conserved. These findings are consistent with previous observations in angiosperm plastomes, where GC-skew patterns often indicate replication origin and terminus regions, as well as conserved structural and functional features of the genome (Alsuhaimi *et al.*, 2024).

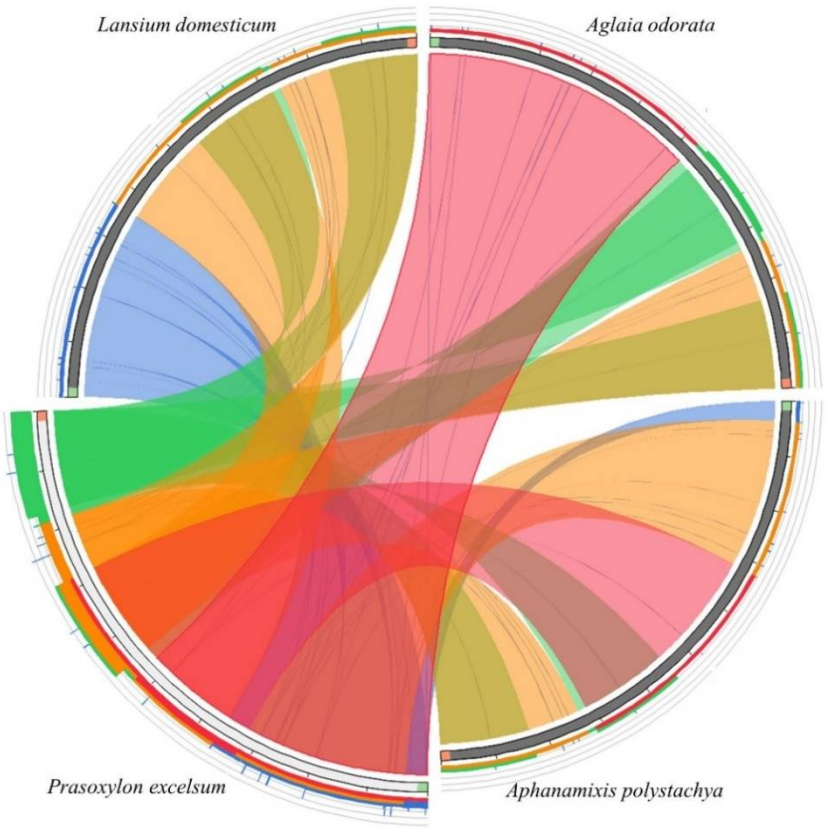

Fig. 6. Synteny analysis of *P. excelsum* plastome with some closely related members of Melioideae.

Nucleotide diversity

Nucleotide diversity (π) analysis across the plastomes of Melioideae subfamily revealed an average π value of 0.0061, indicating a generally conserved genomic composition within the subfamily. The highest nucleotide diversity was detected in the *atpB* gene (π = 0.05389) located in the LSC region, followed by *ycf1* (π = 0.03556) in the SSC region (Fig. 8). The LSC and SSC regions exhibited greater sequence variation than the IR regions, while both IRs displayed nearly identical diversity profiles, reflecting their conserved and duplicated nature.

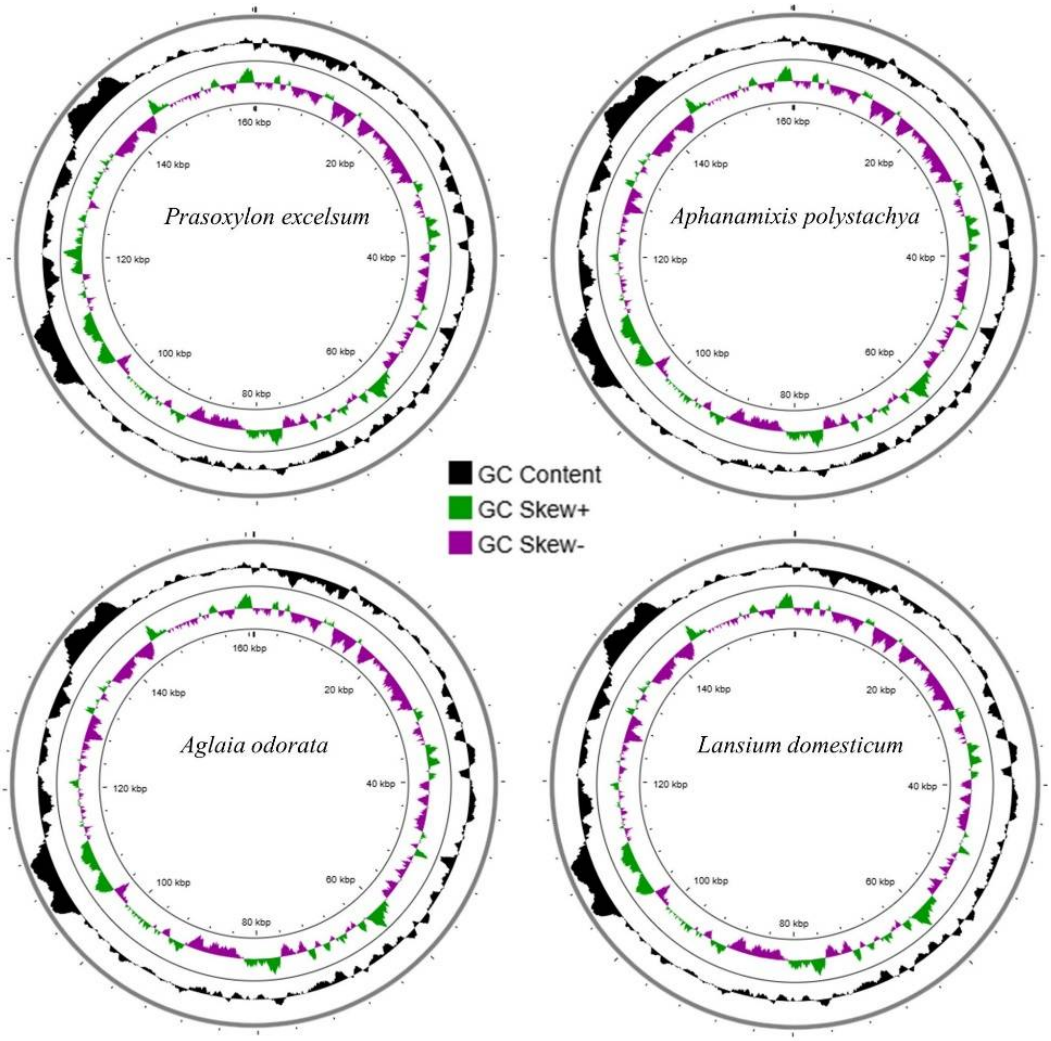

Fig. 7. GC contents distribution and skewness analysis in *P. excelsum* and allied members within Melioideae.

The higher variability of *atpB* and *ycf1* suggests that these loci evolve more rapidly and may serve as hypervariable markers useful for species identification and phylogenetic inference within *Meliaceae*. Similar trends of elevated diversity in single-copy regions and conservation in IRs have been reported in other angiosperm plastomes, attributed to copy-correction and structural stability of IR sequences (Alsuhaimi *et al*., 2024; Ahmed and Rahman, 2025). Therefore, *atpB* and *ycf1* can be proposed as potential DNA barcodes for distinguishing closely related *Melioideae* taxa and improving resolution in evolutionary and taxonomic studies.

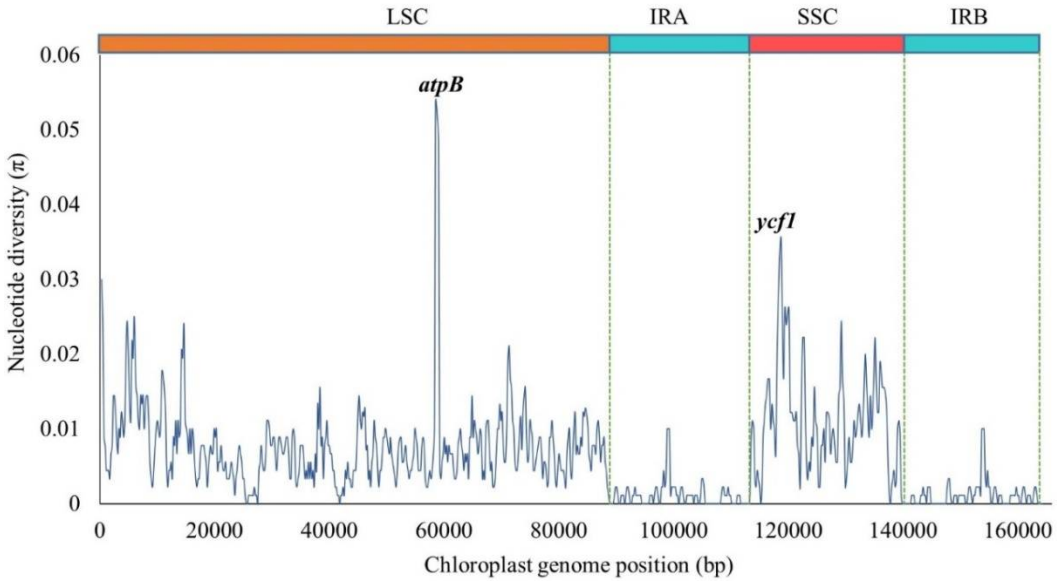

Fig. 8. Nucleotide diversity analysis illustrating hypervariable DNA barcodes within Melioideae.

Molecular phylogenetics and dating

Molecular phylogenetic analysis based on complete Cp genome sequences strongly supported the systematic placement of *P. excelsum* within the subfamily Melioideae of the family Meliaceae (Fig. 9). In the ML tree, *P. excelsum* formed a well-supported clade with members of the tribe Aglaieae, particularly showing a close affinity to *A. polystachya*, *A. odorata* and *L. domesticum*, indicating a shared evolutionary lineage within the subfamily. The ML topology further confirmed the monophyletic origin of the two recognized subfamilies, Cedreloideae and Melioideae, each forming distinct and well-resolved clades. Both clades received high bootstrap support values (~100%), reflecting strong confidence in the inferred relationships. The clear separation between the two subfamilies corresponds well with their previously recognized morphological and molecular distinctions, thereby reinforcing the current taxonomic classification within Meliaceae (Nei *et al*., 2025).

Molecular dating analysis revealed divergence time within Meliaceae based on calibration nodes generated from the TimeTree server (Fig. 10). The TimeTree results exhibited a pairwise divergence time of 18.5 MYA (million years ago) between *Toona ciliata* and *Khaya senegalensis* and of 68 MYA between *Xylocarpus moluccensis* and *Azadirachta indica.* The use of calibration nodes during molecular dating analysis was supported by similar previous studies (Ahmed and Rahman, 2024; 2025). The dating analysis estimated the divergence of the family *Meliaceae* at approximately 58.84 MYA, corresponding to the Paleocene–Eocene transition of the Cenozoic era (Fig. 11). The two major subfamilies, *Cedreloideae* and *Melioideae*, were inferred to have diverged around 52.32 MYA and 23.31 MYA, respectively, indicating an early radiation of Meliaceae followed by subsequent diversification within the subfamilies. The lineage leading to *P. excelsum* was estimated to have originated around 16.21 MYA, during the late Burdigalian age of the Miocene epoch, suggesting that speciation likely occurred during a period of significant climatic and ecological transformation in tropical regions.

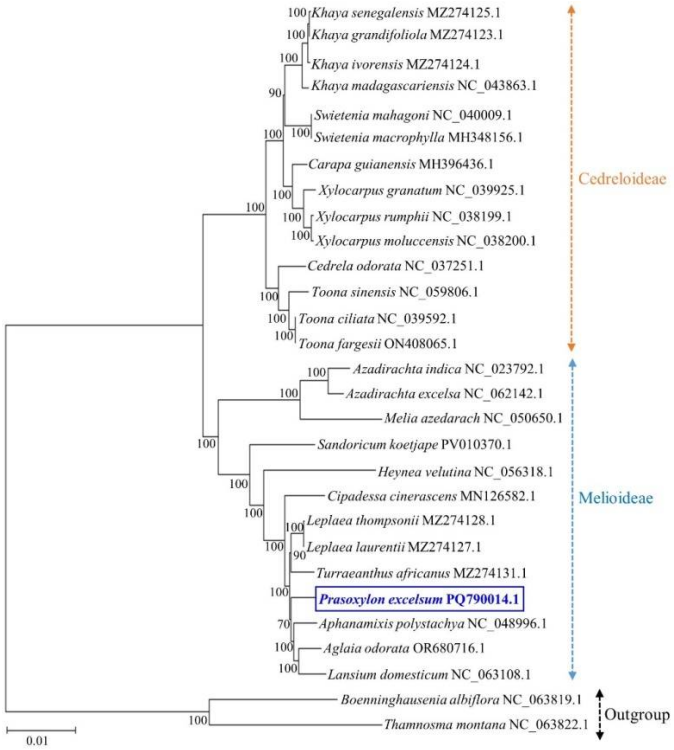

Fig. 9. Maximum-likelihood tree with 1000 bootstrap replicates showing phylogenetic relationships of *P. excelsum* within Meliaceae.

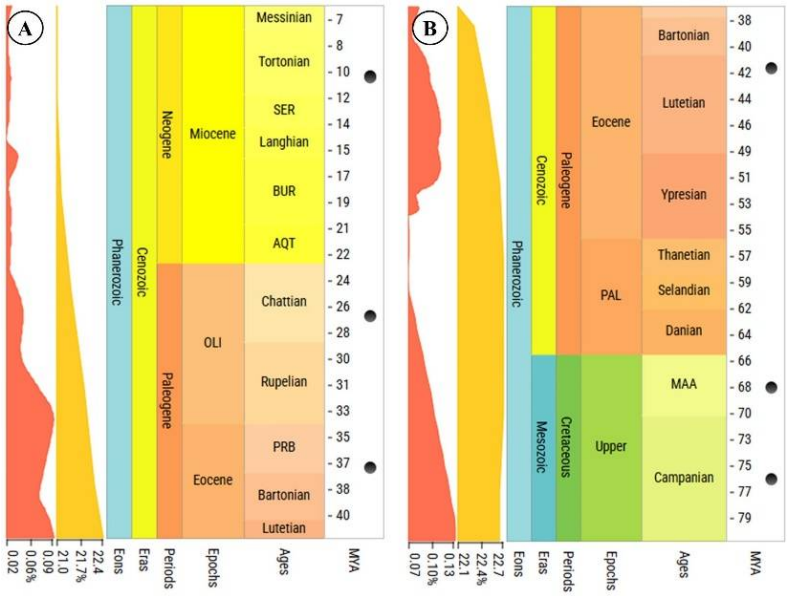

Fig. 10. Calibration node analysis using the TimeTree server. A. Pairwise divergence time estimation between *T. ciliata* and *K. senegalensis*, B. Pairwise divergence time estimation between *X. moluccensis* and *A. indica*.

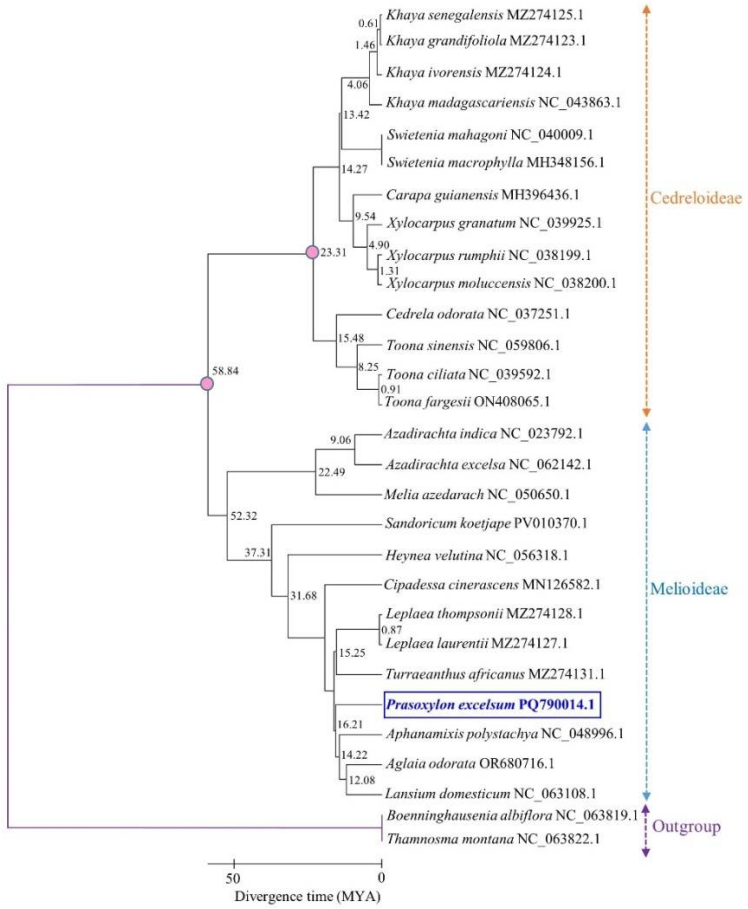

Fig. 11. Molecular dating analysis of *P. excelsum* within Meliaceae elucidating divergence time according to the geological time-scale.

The use of the complete plastome in divergence time estimation provided a robust temporal framework with improved phylogenetic resolution compared to traditional analyses based on single or a few chloroplast genes. Complete chloroplast genomes, often regarded as "superbarcodes," offer extensive phylogenetic signal due to their large number of informative sites, conserved structure, and low rates of recombination. This genome-scale approach minimizes stochastic error and enhances the precision of molecular clock estimates, thereby enabling more reliable reconstruction of evolutionary timelines within *Meliaceae* (Nei *et al.*, 2025; Jiang *et al.*, 2023).

Nuclear genome assembly and quality assessments

The draft nuclear genome of *P. excelsum* assembled using the ABySS assembler yielded a total length of 676.45 Mb with a GC content of 33.49% (Table 3). The assembly comprised 310,705 scaffolds, with the largest scaffold measuring 70,178 bp. The N50 and N90 values were 3,686 bp and 835 bp, respectively, while the auN value was 5,969.10 bp, indicating a relatively fragmented yet comprehensive draft assembly.

Table 3. Assembly statistics of the *P. excelsum* draft nuclear genome based on QUAST analysis.

Parameter	Value
Assembled length (total)	676,454,850 bp
GC content (%)	33.49
N50 (bp)	3,686
N90 (bp)	835
auN (bp)	5,969.10
L50	45,423
L90	205,434
Number of N's per 100 kbp	1,023.38
Scaffolds number (≥ 0 bp)	10,724,571
Scaffolds number (≥ 1,000 bp)	176,473
Total length (≥ 0 bp)	1,459,076,014 bp
Total length (≥ 1,000 bp)	582,365,969 bp
Scaffold number (total)	310,705
Largest scaffold (bp)	70,178

The L50 and L90 values were 45,423 and 205,434, respectively, suggesting that half of the assembled genome was contained within 45,423 of the largest scaffolds. The number of ambiguous bases was moderate, with 1,023.38 N's per 100 kbp, reflecting acceptable assembly quality for a complex plant nuclear genome. Scaffolds ≥ 1 kb in length accounted for 582.36 Mb, representing the majority of the assembled sequence (Table 3). BUSCO-based assessment demonstrated that the draft nuclear genome of *P. excelsum* is highly complete, recovering 87.8% (Viridiplantae) and 87.2% (Embryophyta) complete BUSCOs (Table 4, Fig. 12). These values indicate a well-assembled and gene-rich genome, comparable to or exceeding the completeness levels reported for several other plant genomes. For instance, *Chenopodium pallidicaule* (Ali *et al*., 2024) showed 269 complete BUSCOs (63.3%) out of 425 using the same Viridiplantae dataset, reflecting a moderately fragmented assembly relative to *P. excelsum.*

Table 4. BUSCO assessment of the draft nuclear genome of *P. excelsum*.

BUSCO Category	Description	Viridiplantae		Embryophyta	
		Number	Percentage	Number	Percentage
Complete (C)	Total complete BUSCOs (S + D)	373	87.8	1406	87.2
Single-copy (S)	Complete and single-copy BUSCOs	308	72.5	1150	71.3
Duplicated (D)	Complete and duplicated BUSCOs	65	15.3	256	15.9
Fragmented (F)	Partial or fragmented BUSCOs	43	10.1	129	8.0
Missing (M)	BUSCOs not detected in the assembly	9	2.1	79	4.8
Total BUSCO groups searched (n)	Reference BUSCOs from lineage dataset	425	100	1614	100

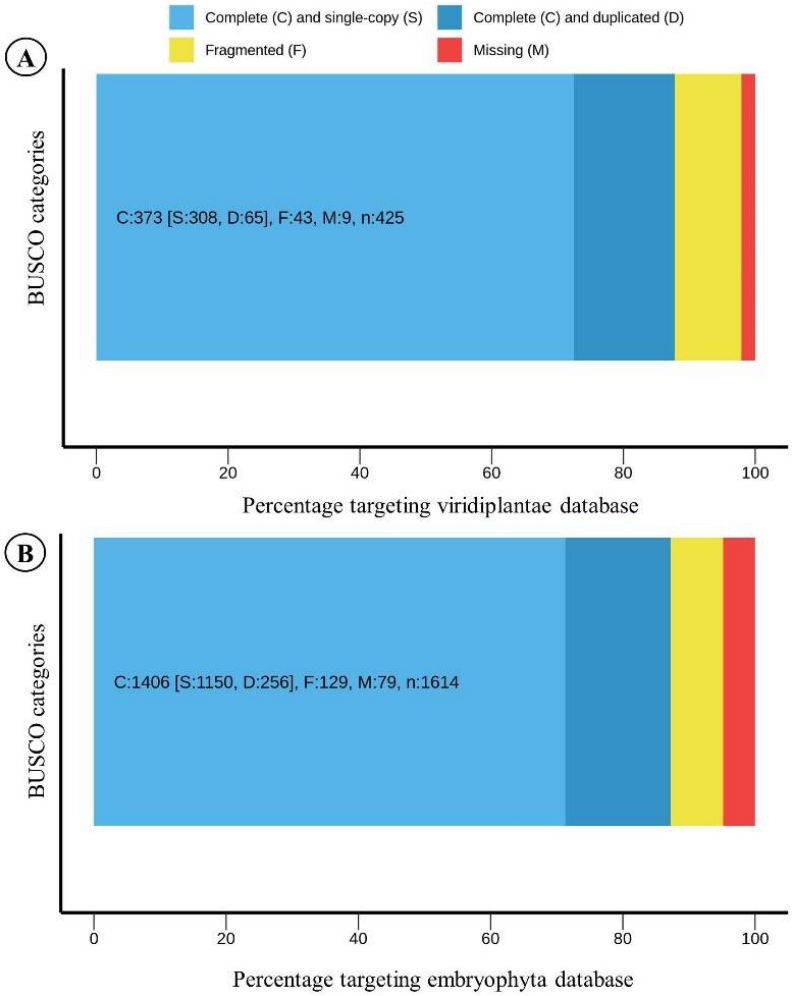

Fig. 12. BUSCO-based evaluation of completeness of the nuclear genome of *P. excelsum*. A. Comparison with Viridiplantae database, B. Comparison with Embryophyta database.

Similarly, *Phoenix roebelenii* (Chakravartty and Neelapu, 2023) achieved 84.2% completeness and 79.9% single-copy BUSCOs under the Embryophyta dataset, slightly lower than that of *P. excelsum*, indicating a comparable yet marginally less complete nuclear assembly. In *P. excelsum*, the relatively high proportion of single-copy BUSCOs (~72%) and low missing fraction (2-5%) demonstrate that the majority of conserved orthologous genes are represented, underscoring the assembly's quality despite its fragmented structure (N50: 3,686 bp). The moderate level of duplication (15-16%) likely reflects either allelic redundancy typical of heterozygous plant genomes or remnants of ancient polyploidy events, which are common within *Meliaceae* (Guimaraes and Forni-Martins, 2022).

Functional annotation of the nuclear genome

Gene prediction was conducted using AUGUSTUS, which identified a total of 44,465 putative PCGs in the *P. excelsum* nuclear genome. Functional annotation of the predicted genes

was performed through ShinyGO, where 24,498 genes were successfully mapped to orthologs in *Arabidopsis thaliana*, enabling downstream Gene Ontology (GO) and KEGG pathway enrichment analyses.

GO enrichment analysis revealed 18 highly enriched pathways of biological processes where most of the genes were associated with gene regulation and macromolecular activity (Fig. 13A). The most significantly enriched Biological Process (BP) categories included nucleobase-containing compound metabolism (4637 genes, FDR = 7.3E−175), gene expression (4636 genes, FDR = 3.3E−170), and nucleic acid metabolic processes (4212 genes, FDR = 1.7E−160). Other highly represented pathways encompassed protein metabolic processes, RNA metabolic and biosynthetic processes, and protein modification processes, all showing strong enrichment (FDR < 1E−120).

In the Cellular Component (CC) category, the majority of predicted genes were associated with key intracellular and membrane-bound structures (Fig. 13B). The most significantly enriched components included the cytosol (2754 genes, FDR = 2.0E−106), plasma membrane (2769 genes, FDR = 2.1E−102), and cell periphery (3188 genes, FDR = 4.9E−115). High enrichment was also observed for the protein-containing complex (3097 genes, FDR = 2.7E−104), plastid (2851 genes, FDR = 1.6E−85), and chloroplast (2350 genes, FDR = 2.6E−66), indicating the prominence of photosynthetically active organelles. Additional enriched components included the nuclear lumen, endomembrane system, Golgi apparatus, and organelle membrane, highlighting the functional diversity and compartmentalized organization of the *P. excelsum* cellular machinery.

In the Molecular Function (MF) category, *P. excelsum* genes were predominantly enriched in functions associated with binding and catalytic activities (Fig. 13C). The most significantly overrepresented terms included nucleic acid binding (4421 genes, FDR = 8.2E−146), heterocyclic compound binding (3256 genes, FDR = 1.1E−158), and anion binding (3270 genes, FDR = 5.3E−161), reflecting extensive molecular interactions essential for transcriptional and metabolic regulation. Strong enrichment was also observed for metal ion binding (3776 genes, FDR = 3.3E−166), transferase activity (4087 genes, FDR = 2.9E−163), and hydrolase activity (3296 genes, FDR = 1.2E−128), highlighting the genome's enzymatic versatility. Additionally, numerous genes were involved in ATP binding, nucleotide binding, and carbohydrate derivative binding, suggesting active participation in energy metabolism and signal transduction. The enrichment of transcription regulator activity (1805 genes, FDR = 5.2E−81) further indicates a well-developed regulatory network governing gene expression. Altogether, these results underscore the functional diversity of the *P. excelsum* nuclear genome, emphasizing its dynamic roles in catalysis, binding, and molecular regulation.

Functional classification of *P. excelsum* genes based on KEGG pathway analysis revealed significant enrichment across a wide spectrum of metabolic and signaling pathways (Table 5). The most represented categories were metabolic pathways (2343 genes, FDR = 2.5E−107) and biosynthesis of secondary metabolites (1313 genes, FDR = 2.5E−60), reflecting the species' diverse metabolic capacity and potential for producing bioactive compounds. Strong enrichment was also observed in plant hormone signal transduction (417 genes, FDR = 1.1E−16), underscoring the genome's regulatory complexity in growth and stress responses (Table 5, Fig. 14). Additionally, key primary metabolic processes such as biosynthesis of amino acids (240 genes, FDR = 3.1E−09), carbon metabolism (267 genes, FDR = 1.9E−09), and biosynthesis of cofactors (235 genes, FDR = 4.6E−09) were prominently represented, indicating robust core metabolic activity. Furthermore, genes involved in protein processing in the endoplasmic reticulum (218 genes, FDR = 2.7E−08) and ubiquitin-mediated proteolysis (160 genes, FDR = 1.3E−05) suggest active protein quality control and turnover mechanisms.

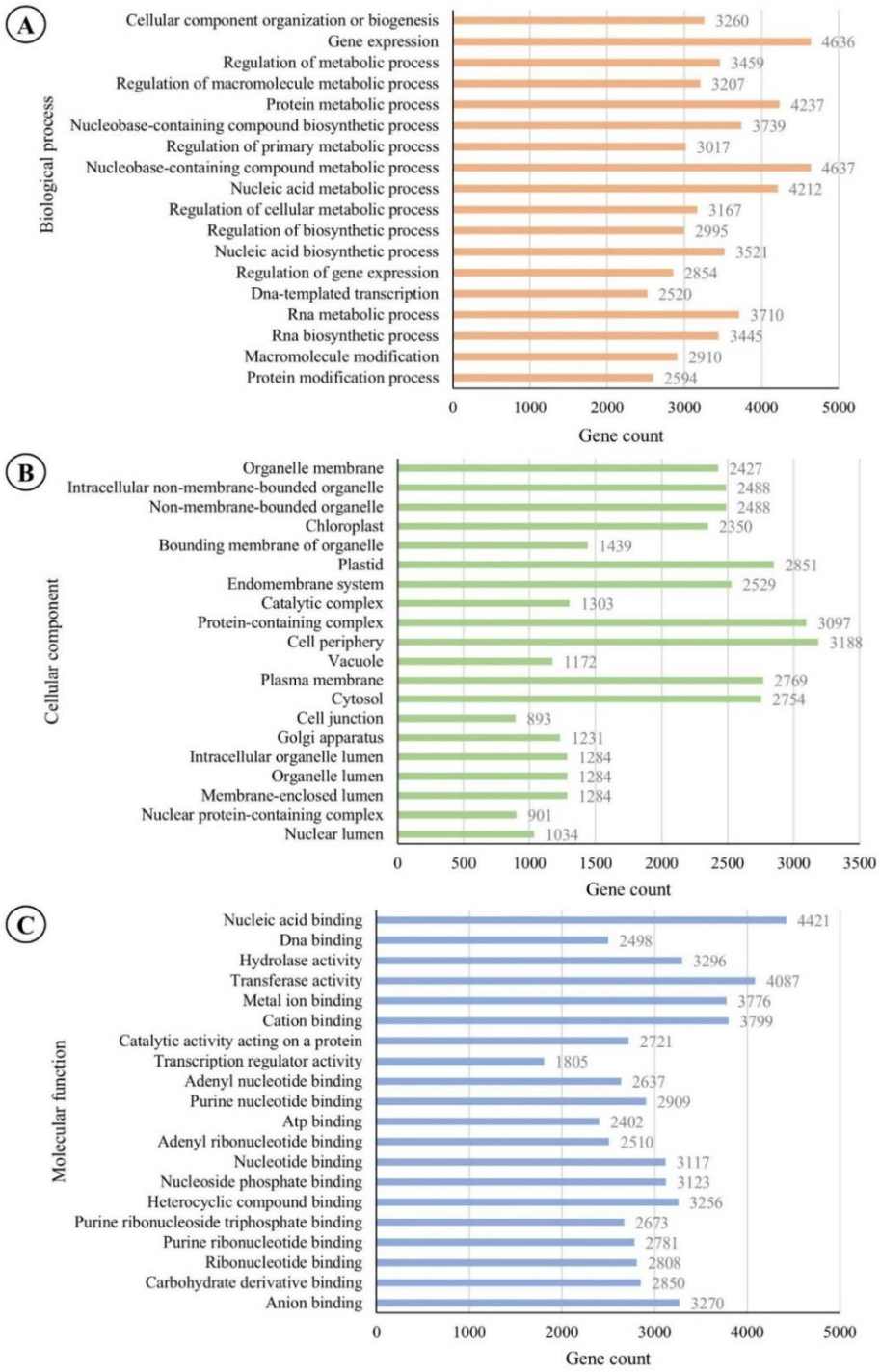

Fig. 13. Gene ontology-based enrichment analysis showing the distribution of nuclear genes of *P. excelsum* in various pathways. A. Biological process, B. Cellular component, C. Molecular function.

Table 5. Enriched KEGG pathways in the *P. excelsum* draft nuclear genome based on functional annotation of protein-coding genes.

KEGG Pathways	Genes	Enrichment FDR	Pathway genes
Biosynthesis of secondary metabolites	1313	2.50E-60	1318
Metabolic pathways	2343	2.50E-107	2354
Plant hormone signal transduction	417	1.10E-16	420
Biosynthesis of amino acids	240	3.10E-09	242
Biosynthesis of cofactors	235	4.60E-09	237
Protein processing in endoplasmic reticulum	218	2.70E-08	220
Plant-pathogen interaction	214	3.70E-08	216
Carbon metabolism	267	1.90E-09	270
Ubiquitin mediated proteolysis	160	1.30E-05	162
Endocytosis	152	2.60E-05	154
MAPK signaling pathway-plant	142	6.90E-05	144
Phenylpropanoid biosynthesis	125	3.60E-04	127
Cysteine and methionine metabolism	123	4.20E-04	125
Starch and sucrose metabolism	168	1.30E-04	172
Ribosome	314	2.50E-05	327

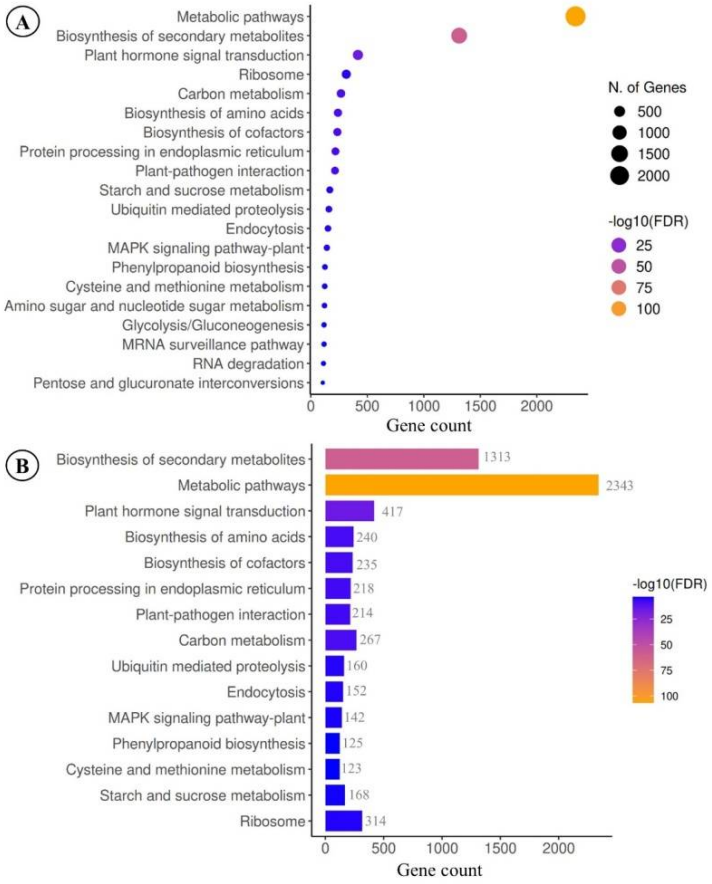

Fig. 14. Distribution of *P. excelsum* nuclear genes across the top significantly enriched KEGG pathways.

Pathways related to plant–pathogen interaction (214 genes, FDR = 3.7E−08) and MAPK signaling (142 genes, FDR = 6.9E−05) highlight the species' genetic potential for defense and environmental adaptation. Secondary metabolic routes, including phenylpropanoid biosynthesis (125 genes, FDR = 3.6E−04), together with starch and sucrose metabolism (168 genes, FDR = 1.3E−04), further emphasize the biochemical versatility of *P. excelsum*. Collectively, these findings illustrate a well-balanced genomic landscape integrating primary metabolism, secondary compound biosynthesis, and complex regulatory signaling networks, reflecting the species' adaptive and functional diversity.

Gene Ontology (GO)-based annotation provides a structured framework to classify genes according to their biological roles, molecular activities, and cellular localization, thereby enabling a functional overview of the nuclear genome. In *P. excelsum*, GO analysis revealed the predominance of genes associated with fundamental biological processes such as gene expression regulation, nucleic acid metabolism, and protein modification, underscoring the species' robust cellular and metabolic regulation. These annotations facilitate the interpretation of genomic complexity by linking predicted genes to well-characterized pathways and biological systems. Moreover, GO categorization enhances cross-species comparisons, aiding in identification of conserved and lineage-specific functional signatures within Meliaceae. Similar GO-based genomic annotations have been instrumental in elucidating gene function and adaptive evolution in other plant species such as *Chenopodium pallidicaule* and *Polygonum cuspidatum* (Ali *et al.*, 2024; Zhang *et al.*, 2019). Thus, functional categorization through gene ontology not only validates gene prediction outcomes but also provides a foundational resource for future investigations into the molecular basis of stress tolerance, secondary metabolite biosynthesis, and evolutionary adaptation in *P. excelsum*.

The current investigation reports the first comprehensive *de novo* assembly and annotation of both the chloroplast and nuclear genomes of *Prasoxylon excelsum*, a medicinally important species of the Meliaceae. The complete chloroplast genome (GenBank accession: PQ790014.1) revealed a conserved quadripartite structure and provided valuable insights into gene content, sequence variation, and phylogenetic placement within the subfamily Melioideae. The nuclear genome assembly further enriched the genomic landscape of *P. excelsum*, revealing genes associated with metabolic and regulatory pathways. Collectively, these genomic resources offer a robust framework for future research on the evolution, systematics, and functional genomics of *Prasoxylon* and related taxa. Moreover, the dataset lays a foundation basis for molecular breeding and conservation strategies aimed at preserving this ethnomedicinally significant species.

Acknowledgment

The authors extend their appreciation to Ongoing Funding Research Program (ORF-2025-306), King Saud University, Riyadh, Saudi Arabia, for funding this work.

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**(16): e36204.

Ahmed, S.S. and Rahman, M.O. 2025. Complete chloroplast genome of *Fraxinus griffithii* CB Clarke (Oleaceae): Insights into genome structure and molecular phylogenetics. Bangladesh J. Plant Taxon. **32**(1): 27–44.

Ali, M.A., Mahato, R. and Lee, J. 2024. Genomic analysis made easy (Game V1): An automated software for plant genome assembly and annotation from illumina sequencing. Bangladesh J. Plant Taxon. **31**(2): 225–238.

Alsuhaimi, N.M., Ali, M.A., Alwahibi, M.S., Ahmed, S.S., Rahman, M.O., Pandey, S.K., Elshikh, M.S., Alshallali, S.R.S., Lee, J. and Kim, S.Y. 2024. Complete chloroplast genome sequence of a novel *Withania somnifera* (L.) Dunal: Comparative genomics and phylogenetic insights. Bangladesh J. Plant Taxon. **31**(2): 205–223.

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.

Brach, A.R. and Song, H. 2006. eFloras: New directions for online floras exemplified by the Flora of China Project. Taxon **55**(1): 188–192.

Brandine, G.D.S. and Smith, A.D. 2021. Falco: High-speed FastQC emulation for quality control of sequencing data. F1000Res. **8**: 1874.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. 2009. BLAST+: Architecture and applications. BMC Bioinform. **10**(1): 421.

Chakravartty, N. and Neelapu, N.R. 2023. The *de novo* genome assembly (nuclear, chloroplast and mitochondria) of ornamental plant pygmy date palm *Phoenix roebelenii*. J. Appl. Biol. Biotechnol. **11**(3): 113–122.

Chakravartty, N. and Neelapu, N.R.R. 2024. The genome assembly of lemon grass to identify the genes in *de novo*. J. Appl. Biol. Biotechnol. **12**(2): 100–149.

Clarke, J.L., Daniell, H. and Nugent, J.M. 2011. Chloroplast biotechnology, genomics and evolution: Current status, challenges and future directions. Plant Mol. Biol. **76**(3): 207–209.

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.

Ge, S.X., Jung, D. and Yao R. 2020. ShinyGO: A graphical gene-set enrichment tool for animals and plants. Bioinform. **36**(8): 2628–2629.

Govaerts, R., Lughadha, E.N., Black, N., Turner, R. and Paton, A. 2021. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. Sci. Data **8**(1): 215.

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.

Guimaraes, R. and Forni-Martins, E.R. 2022. Chromosome numbers and their evolutionary meaning in the Sapindales order: An overview. Brazilian J. Bot. **45**(1): 77–91.

Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinform. **29**(8): 1072–1075.

Jiang, Y., Zhu, C., Wang, S., Wang, F. and Sun, Z. 2023. Identification of three cultivated varieties of *Scutellaria baicalensis* using the complete chloroplast genome as a super-barcode. Sci. Rep. **13**(1): 5602.

Jin, J.J., Yu, W.B., Yang, J.B., Song, Y., DePamphilis, C.W., Yi, T.S. and Li, D.Z. 2020. GetOrganelle: A fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. Genome Biol. **21**(241): 1–31.

Katoh, K. and Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. **30**(4): 772–780.

Kumar, S., Stecher, G., Suleski, M. and Hedges, S.B. 2017. TimeTree: A resource for timelines, timetrees, and divergence times. Mol. Biol. Evol. **34**(7): 1812–1819.

Lalmuanpuii, R., Zodinpuii, B., Bohia, B., Zothanpuia, Lalbiaknunga, J. and Singh, P.K. 2024. Wild edible vegetables of ethnic communities of Mizoram (Northeast India): An ethnobotanical study in thrust of marketing potential. J. Ethnobiol. Ethnomed. **20**(1): 58.

Librado, P. and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinform. **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.

Meilanie, S.R., Mayanti, T., Nurlelasari, N., Huspa, D.H.P., Maharani, R., Zainuddin, A., Darwati, D., Julaeha, E., Supratman, U. and Anshori, J.A. 2022. Triterpenoids from stem bark of *Dysoxylum excelsum* and their cytotoxic activity against MCF-7 breast cancer cells. Indonesian J. Chem. **22**(4): 1107–1115.

Nie, Z., Ma, J., Wang, C., Tang, M., Jia, T., Liao, G. and Zhang, L. 2025. Comparative analysis of chloroplast genomes on Meliaceae species: Insights into the evolution and species identification. Front. Plant Sci. **16**: 1536313.

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

Oyedeji-Amusa, M.O., Sadgrove, N.J. and Wyk, B.E.V. 2021. The ethnobotany and chemistry of south african Meliaceae: A review. Plants **10**(9): 1796.

Ramadan, A.M., Mohammed, T., Al-Ghamdi, K.M., Alghamdi, A.J. and Atef, A. 2023. The first report describes features of the chloroplast genome of *Withania frutescens*. Saudi J. Biol. Sci. **30**(3): 103600.

Romdhane, W.B., Saad, R.B., Guiderdoni, E., Ali, A.A.M., Tarroum, M., Al-Doss, A. and Hassairi, A. 2025. De novo, high-quality assembly and annotation of the halophyte grass *Aeluropus littoralis* draft genome and identification of A20/AN1 zinc finger protein family. BMC Plant Biol. **25**(1): 556.

Samji, A., Eashwarlal, K., Shanmugavel, S., Kumar, S. and Warrier, R.R. 2023. Chloroplast genome skimming of a potential agroforestry species *Melia dubia* Cav and its comparative phylogenetic analysis with major Meliaceae members. 3 Biotech **13**(1): 30.

Seppey, M., Manni, M. and Zdobnov, E.M. 2019. BUSCO: Assessing genome assembly and annotation completeness. *In*: Kollmar, M. (Ed.), Gene prediction: Methods and protocols. Springer New York, pp. 227–245.

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.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J. and Birol, I. 2009. ABySS: A parallel assembler for short read sequence data. Genome Res. **19**(6): 1117–1123.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S. and Morgenstern, B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res. **34**: W435–W439.

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

Zainuddin, A., Meilanie, S.R., Darwati, D., Kurniawan, K., Nurlelasari, N., Herlina, T., Saputra, A.R., Anshori, J.A. and Mayanti, T. 2020. Cytotoxic triterpenoids from the stem barks of *Dysoxylum arborescens* and *Dysoxylum excelsum* against MCF-7 breast cancer cell. Sains Malays. **49**(5): 989–994.

Zhang, Y., Zheng, L., Zheng, Y., Zhou, C., Huang, P., Xiao, X., Zhao, Y., Hao, X., Hu, Z., Chen, Q., Li, H., Wang, X., Fukushima, K., Wang, G. and Li, C. 2019. Assembly and annotation of a draft genome of the medicinal plant *Polygonum cuspidatum*. Front. Plant Sci. **10**: 1274.

Zheng, S., Poczai, P., Hyvönen, J., Tang, J. and Amiryousefi, A. 2020. Chloroplot: An online program for the versatile plotting of organelle genomes. Front. Genet. **11**: 576124.

(Manuscript received on 03 August 2025; revised on 05 December 2025)