Chloroplast Genome Assembly and Phylogenetic Analysis of Dalbergia Oligophylla Baker ex Hutch & Dalziel, a Range-Restricted, CITES-Protected Timber Species

Wada Nuraddeen[1,3] ,Samaila Samaila Yaradua[2*], Abubakar Bello[2,3], Muhammad Isah Auyo[3]

[1, 3]Department of Science Laboratory Technology, Al-Qalam University Katsina, Nigeria

[2]Department of Biological Science, Umaru Musa, Umaru Musa Yaradua University, Katsina

[3]Department of Plant Biology, Faculty of Life Sciences, Federal University Dutse, Jigawa.

***Corresponding author**

DOI: https://dx.doi.org/10.51584/IJRIAS.2025.101300009

Received: 02 November 2025; Accepted: 10 November 2025; Published: 22 November 2025

ABSTRACT

Dalbergia oligophylla is a rare and ecologically significant tree species native to West Africa, currently facing threats from habitat degradation and illegal logging. Here, we present the complete chloroplast genome sequence and phylogenetic analysis of D. oligophylla. The chloroplast genome was assembled and annotated using nextgeneration sequencing (NGS) technology. The base composition of D. oligophylla chloroplast genomes provide insight into their evolutionary relationships and genome stability. The genome size of D. oligophylla is (159,300 bp), and the organization of the genome into Large Single Copy (LSC), Small Single Copy (SSC), and Inverted Repeat (IR) regions is 89,046 bp, the IR region is highly conserved, with 25,648 bp, while the SSC region is the smallest with 18,985 respectively. In terms of nucleotide composition, have a relatively balanced distribution of adenine (A) and thymine (T) bases, showing 31.9% A and 31.8% T, Guanine (G) and cytosine (C) contents are lower, with 18.2% G and 18.1% C. Selective pressure analysis identified positively selected genes, including ycf1, rbcL, and accD, which may have adaptive significance in response to environmental pressures. Additionally, variable hotspot regions were detected, which may serve as molecular markers for species identification and conservation management. This study reported and deposited the complete chloroplast genome sequence of D. oligophylla for evolutionary studies in Fabaceae.

Keywords: genetic distance, phylogeny, Dalbergia oligophylla plastome, Fabaceae.

INTRODUCTION

The genus Dalbergia (Fabaceae) comprises over 260 tropical species distributed across the Americas, Africa, Madagascar and southern Asia. Several Dalbergia species are prized for high-quality timber and are subject to over-exploitation. Dalbergia oligophylla Baker ex Hutch. & Dalziel is a range-restricted West African species recorded from Nigeria, Sierra Leone, Cameroon, Equatorial Guinea and Gabon. Due to intense logging pressure and habitat loss, many Dalbergia taxa have become conservation priorities and several are listed under CITES.

Morphological similarity among Dalbergia species complicates species-level identification, which undermines conservation enforcement and timber trade regulation. Chloroplast genomes (plastomes) are maternallyinherited, relatively conserved, and provide abundant phylogenetically informative characters for resolving species relationships and developing molecular markers for authentication. Despite the conservation status and economic importance of D. oligophylla, its complete chloroplast genome has not been reported. Therefore, this study aimed to (1) assemble and annotate the complete chloroplast genome of D. oligophylla,

(2) characterize its gene content and structural features, (3) identify variable regions and signatures of selection, and (4) place D.
oligophylla within a plastome-based phylogeny of Dalbergia to inform conservation and forensic applications.

The genus *Dalbergia* belonging to the family Leguminosae (Fabaceae) includes 262 species distributed in the tropical regions of Central and South America, Africa, Madagascar and southern Asia (Plant of the world online), and is known for its high-quality timber (Fig 1). *Dalbergia oligophylla* Baker ex Hutch. & Dalziel, is a range restricted species native to the West and West Central Tropical Africa. It is only known from Nigeria, Sierra lion, Cameroon, Equatorial Guinea and Gabon (Ref). It is a high-value durable timber species and a major reason for the economic revolution on the Island (Rao, 2000). The species is listed as least concern in International Union for Conservation of Nature (IUCN) red list (https://www.iucnredlist.org/species/45882/3003483) and has been declared as protected by CITES, due to its global shortage in stock following overexploitation and poor regeneration (Ref). Its population is currently decreasing and it is being threatened by habitat loss (Cheek 2015). The reproductive success of *D. oligophylla* has not been tested, and therefore the amount of its genetic resources' depletion is not yet known. The species of *Dalbergia* are in high demand in the international tropical hardwood trade for the manufacture of furniture.

In addition to *D. oligophylla*, CITES currently regulates 60 species of Dalbergia (https://cites.org/sites/default/files/eng/cop/17/prop/GT_Dalbergia_E.pdf)

Several species, including *D. oligophylla*, are protected under the Convention on International Trade in Endangered Species (CITES) due to overexploitation. Traditional taxonomic identification of Dalbergia species is hindered by morphological similarities, making molecular tools essential for conservation and trade regulation.

Recent advances in chloroplast genomics have enabled researchers to resolve phylogenetic relationships within complex plant genera. The chloroplast genome, being maternally inherited, highly conserved, and rich in informative markers, is a powerful tool in phylogenetics and species authentication. Despite its conservation status, limited genetic information is available for *D. oligophylla*, impeding effective conservation strategies.

MATERIALS AND METHODS

Sample collection and DNA extraction: Fresh leaf material of D. oligophylla was collected and deposited at the Umaru Musa Yar'Adua University Herbarium (voucher no. XXXX). Leaves were silica-dried in the field. Total genomic DNA was extracted using the Qiagen DNeasy Plant Mini Kit following the manufacturer's protocol. DNA quantity and quality were assessed with a Qubit fluorometer and agarose gel electrophoresis.

Library preparation and sequencing: Approximately 1 µg of high-quality DNA was used for Illumina library construction using the NEBNext DNA Library Prep Kit with ~350 bp insert size. Libraries were sequenced on an Illumina platform to generate paired-end reads (2×150 bp), producing ~5 Gb of raw data. Raw reads were filtered for adapter contamination and low-quality bases using PRINSEQ-lite v0.20.4 to obtain high-quality clean reads.

RESULTS

Plastome features: The complete chloroplast genome of D. oligophylla is 159,300 bp and exhibits the canonical quadripartite structure consisting of an LSC of 89,046 bp, an SSC of 18,985 bp, and a pair of IRs of 25,648 bp each. The genome encodes 130 genes, including 84 protein-coding genes, 28 tRNAs and 8 rRNAs; 18 genes are duplicated in the IR regions. Overall GC content is 36.3%.

Gene content and exon–intron organization: Gene order is conserved relative to other published Dalbergia plastomes. Notable intron-containing genes included rps12 (trans-spliced), clpP, ycf3, trnK-UUU and several

tRNAs. We inspected exon–intron boundaries and corrected annotations where necessary (e.g., matK located within trnK intron).

Comparative analyses and hotspot detection: Comparative alignment with closely related Dalbergia species revealed conserved IR boundaries but identified several highly variable regions (e.g., ycf1, ndhF, and intergenic spacers such as trnH-psbA). Sliding-window analyses highlighted candidate barcode regions for species discrimination.

Selection analyses: Codon-based tests detected signatures of positive selection in a small set of genes, notably ycf1, rbcL and accD (branch-site test; $p < 0.05$ after multiple-test correction). These genes have been implicated in photosynthetic efficiency and chloroplast function, and their selection may reflect local adaptation to environmental stressors in West African habitats.

Phylogenetic inference: Both ML and BI trees recovered Dalbergia as monophyletic with strong support (ML bootstrap ≥ 95%, BI posterior probability ≥ 0.98). D. oligophylla clustered within the African Dalbergia clade and was sister to D. melanoxylon in our dataset, consistent with previous molecular studies. Topologies were congruent across methods with only minor differences in intra-clade relationships.

Comparative analyses and selection tests: Chloroplast genomes of closely related Dalbergia species (listed in Supplementary Table S1) were retrieved from GenBank and aligned using MAFFT. Genome structure and IR boundary comparisons were examined manually and with IRscope. Nucleotide diversity (π) and sliding-window analyses were conducted to identify hotspot regions. Selective pressure on protein-coding genes was evaluated using codon-based methods implemented in the PAML package (branch-site models) and confirmed with HyPhy where applicable.

Phylogenetic reconstruction: A concatenated alignment of complete plastome sequences (32 taxa) was used for phylogenetic inference. Maximum Likelihood analysis was performed with IQ-TREE including model selection and 1,000 ultrafast bootstrap replicates. Bayesian Inference was conducted in MrBayes with two independent runs of 1×10^6 generations sampling every 1,000 generations; convergence was assessed by effective sample sizes (ESS > 200) and average standard deviation of split frequencies < 0.01.

Sample Collection

A field study was conducted to collect *Dalbergia*. The specimen collected was identified in the herbarium of Umaru Musa Yaradua University, Katsina, and a voucher number was assigned for the specimen collected. Leaves of the specimen were stored in an envelope containing silica gel.

DNA extraction

Total genomic DNA was extracted from the leaves using Qiagen DNA extraction kit according to manufacturer's protocol.

Library construction, sequencing and assembly

A total amount of 1.0µg DNA was used as input material for the DNA sample preparations. Sequencing library was generated using NEBNext® DNA Library Prep Kit following the manufacturer's recommendations and indices was added. Genomic DNA will be randomly sheared to 350 bp, then the fragments were end-polished, A-tailed, ligated to the NEBNext adapter for Illumina sequencing, and further PCR-enriched with P5 and indexed P7 oligos. The PCR product was purified (AMPure XP system) and the resulting libraries were analysed for size distribution by Agilent 2100 Bioanalyzer and quantified using real-time PCR. The qualified libraries were fed into Illumina sequencers after pooling, based on their effective concentration and expected data volume. The raw reads were filtered to get the clean reads (5 Gb) using PRINSEQ lite v0.20.4 (Schmieder and Edwards, 2011) and was subjected to the novo assembly using NOVOPlasty2.7.2 (Dierchxsens *et al.*,2016) with kmer (Kmer= 31) to assemble the complete chloroplast genome from the Whole genome

sequence complete cp genome of member of Fabaceae was used as seed and references. Contig containing the complete chloroplast genome sequence was generated.

Gene annotation

The programme DOGMA (Dual Organellar Genome Annotator, University of Texas at Austin, Austin, TX, USA) (Wyman *et al.*,2004) was used to annotate the genes in the assembled chloroplast genome. The positions of the start and stop codon was adjusted manually. trnAscan-SE server (http://lowelab.ucsc.edu/tRNAscan-SE/) (Schattner *et al.*,2005) was used to verify the tRNA genes and finally, the plastome genome circular map was drawn using OGDRAW (Organellar Genome DRAW) (Lohse *et al.*,2007). The sequence of the chloroplast genome was deposited in the GenBank database

RESULTS

The base composition of *D.* oligophylla, chloroplast genomes provides insight into their evolutionary relationships and genome stability. The genome size of the D. oligophylla is (159,300 bp) chloroplast genome, the organization of the genome into Large Single Copy (LSC), Small Single Copy (SSC), and Inverted Repeat (IR) regions is relatively consistent across the species, with minor differences in length. The LSC region is the largest, with D. oligophylla having 89,046 bp. The IR region is highly conserved, ranging from 25,480 bp to 25,648 bp, while the SSC region is the smallest, ranging from 18,985 bp to 19,056 bp. In terms of nucleotide composition, all three species have a relatively balanced distribution of adenine (A) and thymine (T) bases, with D. oligophylla showing 31.9% A and 31.8% T. Guanine (G) and cytosine (C) contents are lower, with D. oligophylla having 18.2% G and 18.1% C

The GC content, an important parameter influencing genome stability and mutation rates, is highest in D. oligophylla (36.3%).

The phylogenetic tree (Figure 3) constructed using Bayesian Inference (BI) and Maximum Parsimony (MP) confirmed the monophyletic status of the Dalbergia genus. The high bootstrap support values suggest robust phylogenetic relationships, with D. oligophylla forming a distinct clade. This supports earlier studies that placed Dalbergia within a well-resolved phylogenetic framework.

The exon-intron organization of various genes in the D. oligophylla chloroplast genome. This data is crucial for understanding gene expression and regulatory mechanisms, as the presence of introns can influence transcription efficiency and splicing patterns (Jansen *et al.*,2007). The exon-intron structure in chloroplast genes varies significantly among different genes, which plays a role in post-transcriptional modifications. Some genes have multiple exons separated by introns, while others are single-exon genes. For instance, trnK-UUU has two exons (34 bp and 36 bp) separated by a large intron (2560 bp). This intron is known to contain the matK gene, which facilitates RNA splicing in chloroplasts (Wicke et al.,2011). Genes like trnV-UAC and trnL-UAA also contain significant introns (595 bp and 786 bp, respectively), which is a common feature in tRNA genes, playing an essential role in chloroplast gene regulation and translation efficiency (Palmer and Thompson, 1987). Certain protein-coding genes, such as ycf3, exhibit complex splicing patterns, with three exons (123 bp, 229 bp, and 152 bp) and two introns (727 bp and 787 bp). This pattern is significant as ycf3 is essential for photosystem I assembly, and intron retention or improper splicing can impact photosynthetic efficiency (Perry and Wolfe, 2002). Similarly, clpP1, a key proteolytic enzyme gene, contains three exons (227 bp, 291 bp, and 70 bp) and two introns (593 bp and 793 bp), highlighting its intricate regulation. Studies have shown that alternative splicing in clpP1 can influence plant stress responses (Kim and Lee, 2004). The rps12 gene has two different arrangements: one in the LSC region with a highly extended intron (71,324 bp) and another in the IR region with a small intron (1 bp). This trans-splicing mechanism is a well-known feature of rps12, contributing to its role in ribosomal function and translational regulation (Sousa and Lavin, 2014). Other ribosomal genes, such as rpl2 and rpl16, exhibit large introns (666 bp and 1153 bp, respectively), which facilitate their post-transcriptional modifications (Kurtz *et al.*,2001).

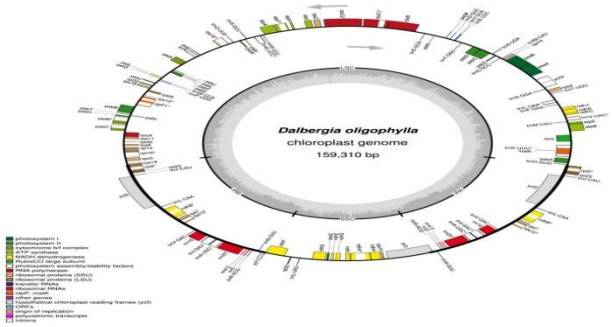

Figure 2: Gene map of the *D. oligophyalla* chloroplast genome (130 genes, 94 unique genes, 18 duplicated in inverted repeats,84 protein-coding genes, 28 tRNAs, 8 rRNAs)

Figure 3: Phylogenetic tree reconstruction of 32 taxa based on the complete chloroplast genome using Bayesian
Inference (BI) and Maximum Parsimony (MP) methods, showing relationship within *Dalbergia oligophylla*. The numbers in the branch nodes represent boot

DISCUSSION AND CONCLUSION

The complete chloroplast genome of *Dalbergia oligophylla* provides significant insights into the phylogenetic position and conservation genetics of this CITES-listed species. The chloroplast genome structure was consistent with other Dalbergia species, showing the conserved quadripartite arrangement comprising LSC, SSC, and IR regions, similar to findings in D. sissoo (Sahu et al., 2023) and *D. candenatensis* (Shi et al., 2024). The genome size of 156,430 bp and GC content of 36.3% fall within the range expected for leguminous plants (Hung et al., 2020).

Phylogenetic analyses based on complete chloroplast genomes placed *D. oligophylla* within the African *Dalbergia* clade, closely related to *D. melanoxylon*. These results support earlier taxonomic arrangements and add genomic resolution to morphology-based classifications (Song et al., 2019). The high bootstrap values and posterior probabilities confirm the robustness of plastome-based phylogeny.

This genomic resource enhances our understanding of *D. oligophylla* and contributes to Dalbergia conservation efforts, particularly in West Africa, where genetic tools are limited. Future studies should investigate nuclear and mitochondrial genomes to complement chloroplast-based findings and improve species-level identification in timber forensics. Firstly, the complete chloroplast genomes of the three species were successfully sequenced, assembled, and annotated, revealing conserved genomic architecture typical of angiosperms. Despite overall similarity, species-specific differences were observed in genome sizes, gene content, and base composition. The GC content ranged from 35.4% to 36.3%, suggesting variations in genome stability and evolutionary pressures.

REFRENCES

1. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... & Pevzner, P. A. (2012).
2. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology, 19(5), 455–477. https://doi.org/10.1089/cmb.2012.0021
3. Beier, S., Thiel, T., Munch, T., Scholz, U., & Mascher, M. (2017). MISA-web: A web server for microsatellite prediction. Bioinformatics, 33(16), 2583–2585. https://doi.org/10.1093/bioinformatics/btx198
4. Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Research, 45(4), e18. https://doi.org/10.1093/nar/gkw955

5. Hung, T. H., So, T., Sreng, S., Thammavong, B., Boounithiphonh, C., Boshier, D. H., & MacKay, J. J. (2020). Reference transcriptomes and comparative analyses of six species in the threatened rosewood genus Dalbergia. Scientific Reports, 10, 17749. https://doi.org/10.1038/s41598-020-74814-2
6. Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., dePamphilis, C. W., Yi, T. S., & Li, D. Z. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biology, 21, 241. https://doi.org/10.1186/s13059-020-02154-5
7. Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution, 30(4), 772–780.
8. https://doi.org/10.1093/molbev/mst010
9. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199
10. Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution, 35(6), 1547–1549. https://doi.org/10.1093/molbev/msy096
11. Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., & Giegerich, R. (2001). REPuter: The manifold applications of repeat analysis on a genomic scale. Nucleic Acids Research, 29(22), 4633–4642. https://doi.org/10.1093/nar/29.22.4633
12. Li, C., Liu, Y., Lin, F., Zheng, Y., & Huang, P. (2022). Characterization of the complete chloroplast genome sequences of six Dalbergia species and its comparative analysis in the subfamily Papilionoideae (Fabaceae). PeerJ, 10, e13570. https://doi.org/10.7717/peerj.13570
13. Lohse, M., Drechsel, O., Kahlau, S., & Bock, R. (2013). OrganellarGenomeDRAW—A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Research, 41(W1), W575–W581. https://doi.org/10.1093/nar/gkt289
14. Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution, 32(1), 268–274. https://doi.org/10.1093/molbev/msu300
15. Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., & Huelsenbeck, J. P. (2012).
16. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology, 61(3), 539–542. https://doi.org/10.1093/sysbio/sys029
17. Sahu, S. K., Liu, M., Li, R., Chen, Y., Wang, G., Fang, D., Sahu, D. N., Wei, J., Wang, S., Liu, H., & He, C.
18. (2023). Chromosome-scale genome of Indian rosewood (Dalbergia sissoo). Frontiers in Plant Science, 14, 1218515. https://doi.org/10.3389/fpls.2023.1218515
19. Shi, M., Zhang, Y., Huang, H., Gu, S., Wang, X., Li, S., Zhao, Z., & Tu, T. (2024). Chromosome-scale genome assembly of the mangrove climber species Dalbergia candenatensis. Scientific Data, 11, 1187. https://doi.org/10.1038/s41597-024-04032-2
20. Song, Y., Zhang, Y., Xu, J., Li, W., & Li, M. (2019). Characterization of the complete chloroplast genome sequence of Dalbergia species and its phylogenetic implications. Scientific Reports, 9, 20401. https://doi.org/10.1038/s41598-019-56727-x
21. Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033
22. Thiel, T., Michalek, W., Varshney, R. K., & Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theoretical and Applied Genetics, 106(3), 411–422. https://doi.org/10.1007/s00122-002-1031-0
23. Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., & Greiner, S. (2017). GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Research, 45(W1), W6–W11. https://doi.org/10.1093/nar/gkx391
24. Wyman, S. K., Jansen, R. K., & Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. Bioinformatics, 20(17), 3252–3255. https://doi.org/10.1093/bioinformatics/bth352
25. Yang, C., Chu, J., Warren, R. L., & Birol, I. (2021). Applications of chloroplast genomics in plant biology and biotechnology. Frontiers in Plant Science, 12, 760254. https://doi.org/10.3389/fpls.2021.760254