



DATA NOTE

Long-read sequencing of *Dipterocarpus littoralis* (Dipterocarpaceae), a critically endangered tree endemic to Nusakambangan Island, Indonesia: Resource for chloroplast genome, phylogenetic, and biosynthetic gene clusters studies

[version 1; peer review: 2 approved with reservations, 1 not approved]

Fifi Gus Dwiyantri ¹, Rahadian Pratama ², Dwi Wahyuni ³, Satria Ardhe Kautsar⁴, Fitri Indriani ⁵, Henti Hendalastuti Rachmat ⁶, Koichi Kamiya⁷, Iskandar Zulkarnaen Siregar ¹

¹Department of Silviculture, Faculty of Forestry and Environment, Institut Pertanian Bogor (IPB University), Bogor, West Java, 16680, Indonesia

²Department of Biochemistry, Faculty of Mathematics and Natural Sciences, Institut Pertanian Bogor (IPB University), Bogor, West Java, 16680, Indonesia

³State Agricultural Polytechnic Kupang, Kupang, East Nusa Tenggara, 85011, Indonesia

⁴Genome Institute of Singapore (GIS), Agency for Science, Technology and Research (A*STAR), 60 Biopolis Street, Genome, 138672, Singapore

⁵Research Center for Applied Botany, National Research and Innovation Agency (BRIN), Cibinong, West Java, 16911, Indonesia

⁶Research Center for Ecology, National Research and Innovation Agency (BRIN), Cibinong, West Java, 16911, Indonesia

⁷Graduate School of Agriculture, Ehime University, Matsuyama, Ehime, 790-8566, Japan

V1 First published: 04 May 2026, 15:660
<https://doi.org/10.12688/f1000research.179128.1>

Latest published: 04 May 2026, 15:660
<https://doi.org/10.12688/f1000research.179128.1>

Abstract

Abstract*

Background

Long-read sequencing of *Dipterocarpus littoralis* (Blume) Kurz. (Dipterocarpaceae) is essential for generating important genetic data to support biodiversity conservation, taxonomy resolution, and the sustainable management of this critically endangered species. However, this type of study remains largely unaddressed. Therefore, this study aims to generate genetic information for *D. littoralis*, including the characterization of chloroplast genome (cp genome) sequences, the construction of a phylogenetic tree, and the identification of the biosynthetic gene cluster (BGC).

Open Peer Review

Approval Status   

	1	2	3
version 1			
04 May 2026	view	view	view

1. **Nunzio D'Agostino**, University of Naples Federico II, Portici, Italy

2. **Abhisek Chakraborty** , Johannes Gutenberg University Mainz, Mainz, Germany

3. **Sheng-Kai Hsu** , Cornell University, Ithaca, USA

Any reports and responses or comments on the article can be found at the end of the article.

Methods

The genomic DNA extracted from *D. littoralis* leaf tissue was sequenced using the MinION Oxford Nanopore Technologies. The cp genome was reconstructed from an assembly with GetOrganelle, and the resulting genome was annotated with GeSeq. A maximum-likelihood (ML) phylogenetic tree was inferred using IQ-TREE2, while biosynthetic gene clusters were annotated and classified through BRAKER2 and PlantSMASH.

Conclusions

This study revealed that the length of the chloroplast genome was 153,775 base pairs (bp) with a GC content of 37%, and it consists of four subregions: a large single-copy (LSC) region of 85,079 bp, a small single-copy (SSC) region of 20,232 bp, and two inverted repeats (IR) regions (IRA 24,232 bp; IRB 24,232 bp). The cp genome of *D. littoralis* encodes 111 genes, including 78 protein-coding genes, 29 transfer RNA (tRNA) genes, and 4 ribosomal RNA (rRNA) genes. The phylogenetic tree derived from the cp genome indicated that *D. littoralis* formed a monophyletic group alongside other species within the same genus. Furthermore, a total of 77 genes coding for BGCs were identified, comprising 44 terpene clusters, 6 alkaloid clusters, 1 lignan cluster, 4 saccharide clusters, 1 Saccharide-Polyketide cluster, 1 saccharide-terpene cluster, and 20 putative clusters. The findings of this study may enhance molecular identification, clarify phylogenetic relationships, and support comparative genomics within the genus *Dipterocarpus*, as well as facilitate the discovery of potential natural products from *D. littoralis*.

Keywords

biosynthetic gene, chloroplast genome, *Dipterocarpus*, pelahlar, phylogeny

Corresponding author: Iskandar Zulkarnaen Siregar (siregar@apps.ipb.ac.id)

Author roles: **Dwiyanti FG:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation; **Pratama R:** Data Curation, Formal Analysis, Software, Visualization, Writing – Review & Editing; **Wahyuni D:** Data Curation, Writing – Review & Editing; **Kautsar SA:** Formal Analysis, Writing – Review & Editing; **Indriani F:** Validation, Writing – Review & Editing; **Rachmat HH:** Validation, Writing – Review & Editing; **Kamiya K:** Validation, Writing – Review & Editing; **Siregar IZ:** Funding Acquisition, Resources, Validation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The study was supported by The National Competitive Basic Research (Skema Penelitian Dasar Kompetitif Nasional) 2021-2022 with contract No. 1/E1/KP.PTNBH/2021 and No. 001/E5/PG.02.00PT/2022, entitled "Pilot Sequencing of 100 Native Forest Tree Genomes to Support Ecosystem Restoration" (Rintisan Sekuensing 100 Genom Pohon Hutan Asli untuk Mendukung Restorasi Ekosistem) awarded by the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia to the corresponding author. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2026 Dwiyanti FG *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Dwiyanti FG, Pratama R, Wahyuni D *et al.* **Long-read sequencing of *Dipterocarpus littoralis* (Dipterocarpaceae), a critically endangered tree endemic to Nusakambangan Island, Indonesia: Resource for chloroplast genome, phylogenetic, and biosynthetic gene clusters studies [version 1; peer review: 2 approved with reservations, 1 not approved]** F1000Research 2026, 15:660 <https://doi.org/10.12688/f1000research.179128.1>

First published: 04 May 2026, 15:660 <https://doi.org/10.12688/f1000research.179128.1>

Introduction

This section should include why the data were gathered or produced. Conservation for critically endangered endemic trees is essential to prevent immediate extinction, as these species often have extremely small, restricted, and vulnerable populations. Due to their limited geographic range and specialized, often damaged habitats, they are far more susceptible to extinction than species with broader distributions (Işik 2011). However, they are vital to ecosystem health, supporting biodiversity, regulating the climate, and purifying the air. Protecting them maintains ecological balance and preserves unique genetic resources for resilience. One of the key endemic trees in Indonesia is *Dipterocarpus littoralis* (Blume) Kurz., or locally known as the pelahlar or plahlar tree (Yulita and Partomihardjo 2011; Dwiyantri et al. 2014; Primajati et al. 2017; Widiastuti et al. 2022), which belongs to the Dipterocarpaceae family and is endemic to Nusakambangan Island, located in Cilacap Regency, Central Java Province, Indonesia, specifically in the West Nusakambangan Nature Reserve (Dwiyantri et al. 2014; Robiansyah and Davy 2015).

The tree grows in lowland tropical forest, reaching 50 cm in diameter and 150 cm in height, sometimes with high buttresses (Hamidi and Robiansyah 2018). The bark tends to peel, especially in old trees, pale grey to grey, resinous (Hamidi and Robiansyah 2018). Furthermore, the species is heavily harvested for local timber needs and is sometimes tapped for resin (dammar) used for fire or to produce glue for boats (Hamidi and Robiansyah 2018). Since 1998, the species has been classified as critically endangered (CR) by the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (Hamidi and Robiansyah 2018) due to illegal logging and is increasingly threatened due to the expansion of the invasive *Arenga obtusifolia*, which inhibits the regeneration of *D. littoralis* seedlings (Robiansyah and Davy 2015).

Dwiyantri et al. (2014) reported the presence of 18 adult trees on the island, while Robiansyah and Davy (2015) found 11 adult trees. Furthermore, the genetic diversity of this species, assessed using microsatellite markers, was low, with an expected heterozygosity of 0.476. This suggests that reductions in population size have been occurring for an extended period (Dwiyantri et al. 2014). Therefore, urgent efforts are needed to prevent this species' extinction. Several studies have been conducted to support conservation strategies for *D. littoralis*, including assessments of population status and habitat preferences (Robiansyah and Davy 2015; Primajati et al. 2017), and of genetic diversity (Yulita and Partomihardjo 2011; Dwiyantri et al. 2014). However, investigations into the chloroplast genome (cp genome), molecular phylogeny, and biosynthetic gene cluster (BGC) have not yet been addressed. Studying the chloroplast genome and molecular phylogeny is crucial, as these provide essential tools for biodiversity conservation and taxonomic resolution, helping to prioritize and implement protective measures. For instance, high-throughput sequencing of the entire chloroplast genome can facilitate the identification of unique lineages, map population structures, and develop DNA barcodes to accurately identify endangered species (Yin et al. 2025; Ma et al. 2025). Additionally, biosynthetic gene clusters (BGCs) are conserved genomic regions that encode specialized metabolites, which play a critical role in species adaptation, defense, and ecological interactions (Cawood and Ton 2025). Therefore, this study aims to characterize the chloroplast genome (cp genome) sequences, construct a phylogenetic tree, and identify the biosynthetic gene cluster (BGC) of *D. littoralis*.

Materials and methods

Plant material, DNA extraction, and sequencing

Fresh leaf samples were collected from a single *Dipterocarpus littoralis* sapling (Figure 1A) planted in 2022 at the area of Faculty of Forestry and Environment, Institut Pertanian Bogor (IPB University), located in Bogor Regency, West Java Province, Indonesia (6°33'21.775" S, 106°43'48.45781" E). The sapling was originally sourced from the West Nusakambangan Nature Reserve in Cilacap Regency, Central Java Province, Indonesia. *Dipterocarpus littoralis* exhibits unique morphological characteristics that distinguish it from other species in its genus. Notably, it features red velvet and hairy stipules (Figure 1B), along with elliptical to ovate leaves of 16-25 cm x 10-18 cm (Newman et al. 1998) (Figure 1C). The leaves are spirally arranged and clustered around the apices of the twigs (Hamidi and Robiansyah 2018). A herbarium specimen was collected and deposited in the Forest Genetics and Molecular Forestry Laboratory, the Department of Silviculture, Faculty of Forestry and Environment, at IPB University in Bogor Regency, West Java Province, Indonesia, under voucher number LGH.0369. The species name was verified by Fifi Gus Dwiyantri through comparison with the *Dipterocarpus littoralis* specimen number PL.VI-C.75 available at the Herbarium Center of the Center for Sustainable Forest Development, Ministry of Forestry of the Republic of Indonesia, Bogor City, West Java Province, Indonesia. This verification was also supported by prior research by Dwiyantri et al. (2014).

The laboratory work, which includes DNA extraction and sequencing, was conducted at the Laboratory of Forest Genetics and Molecular Forestry within the Department of Silviculture at the Faculty of Forestry and Environment, Institut Pertanian Bogor (IPB University) in Bogor Regency, West Java Province, Indonesia. A total of 100 mg of fresh leaf material was used for genomic DNA extraction and isolation using the modified Cetyltrimethylammonium Bromide (CTAB) protocol established by Doyle and Doyle (1990). The leaf sample was ground in a mortar and pestle with 1000

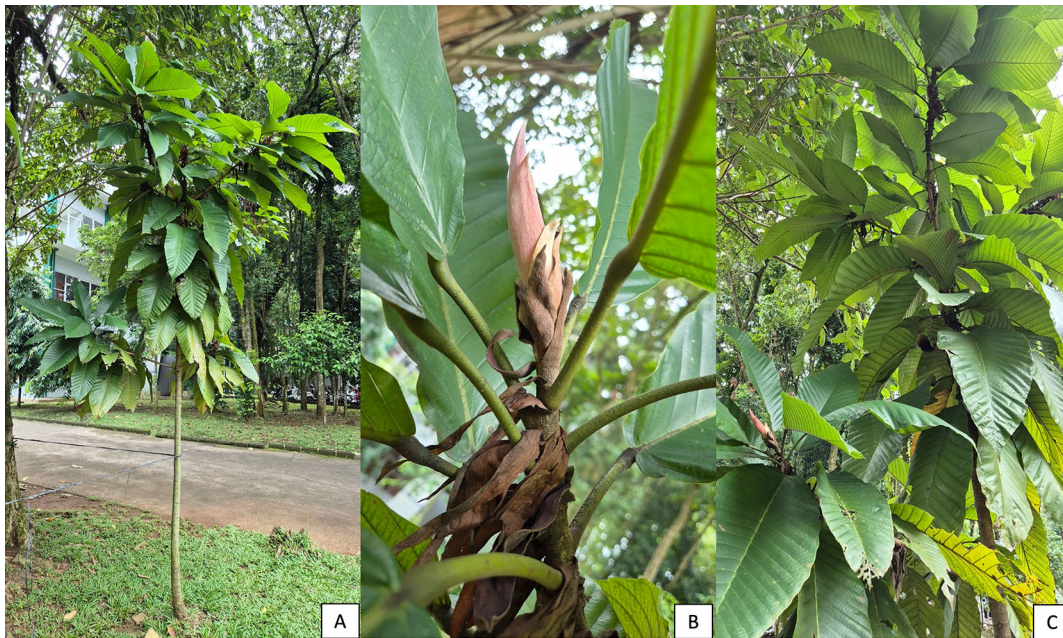


Figure 1. The *Dipterocarpus littoralis* sapling planted in the Faculty of Forestry and Environment of IPB University (A) has unique morphological characteristics, such as red velvet and hairy stipules (B), and elliptical to ovate leaves (C). Pictures were taken by the author Fifi Gus Dwiyantri on 2026-02-12.

μL of 10% CTAB buffer (Cat. No. MB101-100G, HiMedia Laboratories). To eliminate proteins and RNA, 40 μL of 26% Polyvinylpyrrolidone (Cat. No. MB102-100G, HiMedia Laboratories), 5 μL of mercaptoethanol (Cat. No. 8.05740.0250, Merck), and 200 μL of NaCl (Cat. No. 1.06404.1000, Merck) were added. The sample was then homogenized using a vortex MX-5 (Biologix) and subsequently incubated in a water bath (VWR Scientific) at 65°C for 30 minutes. The partitioning of lipids and cellular debris into the organic phase was completed by adding 500 μL of chloroform (Cat. No. 1.02445.2500, Merck)-isoamyl alcohol (Cat. No. 1.00979.1000, Merck) with a ratio of 24:1 and 10 μL of phenol, followed by centrifugation at 13,000 rpm for 3 minutes using Centrifuge (MPW Med. Instruments) to obtain purified DNA. The purified DNA was precipitated by adding an isopropanol (Cat. No. 1.09634.2500, Merck) solution at a 1:1 volume ratio to the supernatant, along with NaCl at a 1:4 volume ratio. The samples were then stored at -20°C for approximately 2 hours to allow DNA to precipitate. Subsequently, the DNA was washed with 500 μL of 70% ethanol (Cat. No. 1.00983.2500, Merck) and dried in a desiccator (Normax). Finally, 50 μL of TE buffer (Cat. No. 93283-500ml, Merck) was added to the genomic DNA before the samples were stored at -20°C.

The quality of the genomic DNA was assessed via agarose gel electrophoresis utilizing a UV Transilluminator TFX-20 LM (Vilber Lourmat). Quantitative measurement of genomic DNA was performed using a Qubit 1.0 fluorometer (Invitrogen-Thermo Fisher Scientific) with the Qubit dsDNA BR assay kit (Cat. No. Q32850, Thermo Fisher Scientific). The high-molecular-weight genomic DNA extracted from *D. littoralis* was subsequently employed for library preparation to facilitate long-read sequencing. DNA libraries were prepared according to the Nanopore protocol using a ligation sequencing kit (Cat. No. SQK-LSK110, Oxford Nanopore Technologies), version GDE_9108_v110_revN_10Nov2020 (Oxford Nanopore Technologies). Long-read sequencing was then performed on a MinION Mk1B R9.4.1 flow cell (Cat. No. FLO-MIN106D, Oxford Nanopore Technologies), connected to MinKNOW software (Oxford Nanopore Technologies) to control the device, monitor flow cells, and perform real-time basecalling. All data analysis was performed at the Forest Genetics and Molecular Forestry Laboratory within the Department of Silviculture, Faculty of Forestry and Environment, IPB University, Bogor Regency, West Java Province, Indonesia.

Chloroplast genome assembly and annotation

The raw Fastq data were analyzed using NanoPlot v1.46.1 (De Coster et al. 2018) to calculate and visualize read distributions. De novo assembly of the organelle genome was performed using Flye v2.9.6 (Kolmogorov et al. 2019) to construct the plastome. The chloroplast genome was reconstructed from an assembly using GetOrganelle (Jin et al. 2020). The resulting chloroplast genomes were then annotated using GeSeq (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>) (Tillich et al. 2017). The fully annotated circular genome was visualized using OrganellarGenomeDRAW (OGDRAW) v1.3.1 accessible through the MPI-MP Chlorobox platform (Greiner et al. 2019).

Phylogenetic tree construction

A total of 18 chloroplast genomes of taxa closely related to *Dipterocarpus littoralis* (from the Dipterocarpaceae family) were downloaded from GenBank (the National Center for Biotechnology Information/NCBI) and aligned with the obtained plastomes. A complete list of the accessions used is given in [Table 1](#). *Gyrinops versteegii* (AP018453.1) was used as the outgroup. The alignment was performed using MAFFT v7.526 ([Katoh and Standley 2013](#)) using default parameters. A maximum-likelihood (ML) phylogenetic tree was inferred using the IQ-TREE2 ([Trifinopoulos et al., 2016](#)) with 1,000 bootstrap replicates. The phylogenetic tree was visualized using iTOL ([Letunic and Bork 2024](#)).

Biosynthetic gene cluster

All DNA sequence datasets from the assembly were annotated using BRAKER2 (without RNA sequencing) ([Brůna et al., 2021](#)) and then assigned to putative proteins and protein regions using GeneMark-ES. The predicted proteins are aligned with the NCBI protein database using DIAMOND to identify the target proteins. These proteins are subsequently aligned to the seed region using Spaln (a space-efficient spliced alignment tool). Through these steps, genome sequences, protein predictions, and the corresponding sequence regions were obtained. Afterward, the exon boundaries were predicted by identifying introns, start codons, and stop codons. The sequences are classified as high or low confidence based on the predictions' confidence levels. High confidence sequences can be processed further with AUGUST2, while those with low confidence are reviewed and reordered to enhance gene predictions using the EP+ gene tag. Upon completing all phases, the results from BRAKER2 yield DNA sequences with predictable genes, which are then clustered using the PlantSMASH platform (plantismash.secondarymetabolites.org) ([Kautsar et al., 2017](#)).

Results

Long-read sequencing of *Dipterocarpus littoralis* using the MinION ONT has produced a dataset of 40,219,030 reads, resulting in a total of 48.4 gigabase pairs (Gbp) of raw data. The mean read length was 1,204.4 base pairs (bp), and the mean raw data read quality score was 11.8. Following the filtering process, all reads have met the quality assessment criteria and retain the same quality score. The complete chloroplast genome of *D. littoralis* has been effectively assembled from long-read sequencing data, enabling the resolution of structural regions within the genome. The chloroplast genome

Table 1. Detailed information on the chloroplast genome used for the phylogenetic analysis in Figure 3.

No.	Taxon	GenBank accession number	Current taxon (Ashton and Heckenbauer, 2022)**
1.	NC_058774.1	<i>Vatica rassak</i>	-
2.	NC_071231.1	<i>Vatica bantamensis</i>	-
3.	NC_058773.1	<i>Vatica xishuangbannaensis</i>	-
4.	NC_054172.1	<i>Vatica odorata</i>	-
5.	NC_051531.1	<i>Vatica guangxiensis</i>	-
6.	NC_041485.1	<i>Vatica mangachapoi</i>	-
7.	NC_081465.1	<i>Dipterocarpus littoralis</i>	-
8.	NC_058777.1	<i>Dipterocarpus alatus</i>	-
9.	NC_067812.1	<i>Dipterocarpus retusus</i>	-
10.	NC_046842.1	<i>Dipterocarpus turbinatus</i>	-
11.	NC_065503.1	<i>Dipterocarpus hasseltii</i>	-
12.	NC_041191.1	<i>Neobalanocarpus heimii</i>	-
13.	NC_044642.1	<i>Hopea hainanensis</i>	-
14.	NC_052744.1	<i>Hopea reticulata</i>	-
15.	NC_053766.1	<i>Hopea chinensis</i>	-
16.	NC_057187.1	<i>Dryobalanops aromatica</i>	-
17.	NC_046579.1	<i>Parashorea chinensis</i>	-
18.	NC_064148.1	<i>Shorea macrophylla</i>	<i>Rubroshorea macrophylla</i>
19.	AP018453.1	<i>Gyrinops versteegii</i> *	-

*The plastome of *Gyrinops versteegii*, not a member of the Dipterocarpaceae family, was used as an outgroup.

**Ashton PS, Heckenbauer J. 2022. Tribe Shoreae (Dipterocarpaceae subfamily Dipterocarpoideae) Finally Dissected. Kew Bulletin. 77:885-903. doi:10.1007/S12225-022-10057-W.

exhibits a typical quadripartite structure (Figure 2a) with a total length of 153,775 bp. The genome comprises a Small Single-copy Region (SSC: 20,232 bp) and a Large Single-copy Region (LSC: 85,079 bp), which are separated by two inverted repeat regions: Inverted Repeat A (IRA: 24,232 bp) and Inverted Repeat B (IRB: 24,232 bp) (Figure 2a). The GC content of the *D. littoralis* sequence is 37 %. This finding exceeds the chloroplast genome size of *Dipterocarpus turbinatus*, as reported by Ci et al. (2019) using short-read sequencing on the Illumina HiSeq 2000 platform, which reported a size of 152,279 bp. However, it is lower than that of *Dipterocarpus retusus*, which was generated via short-read sequencing on the Illumina NovaSeq platform by Tao et al. (2024), at 154,303 bp. The findings support the conclusion that the plastome size in the genus *Dipterocarpus* is relatively conserved (typically in the 152–154 kb range).

The *D. littoralis* chloroplast genome (cp genome) contains 111 genes, including 78 protein-coding genes, 29 transfer RNA (tRNA) genes, and 4 ribosomal RNA (rRNA) genes (Table 2). These genes are categorized into four functional groups: self-replicating genes, photosynthetic genes, genes with other functions, and genes of unknown function (Table 2). This composition is similar to that of *D. retusus* and *D. turbinatus*, whose chloroplast genomes encode 128 genes, including 84 protein-coding genes, 36 tRNA genes, and 8 rRNA genes (Ci et al., 2019; Tao et al., 2024).

The phylogenetic tree of *D. littoralis* based on the complete chloroplast genome showed that the studied *D. littoralis* was in the same clade as *D. littoralis* (NC_081465.1) with a bootstrap value of 100% (Figure 3). Additionally, the studied *D. littoralis* formed a monophyletic group with other species within the same genus, specifically *D. alatus*

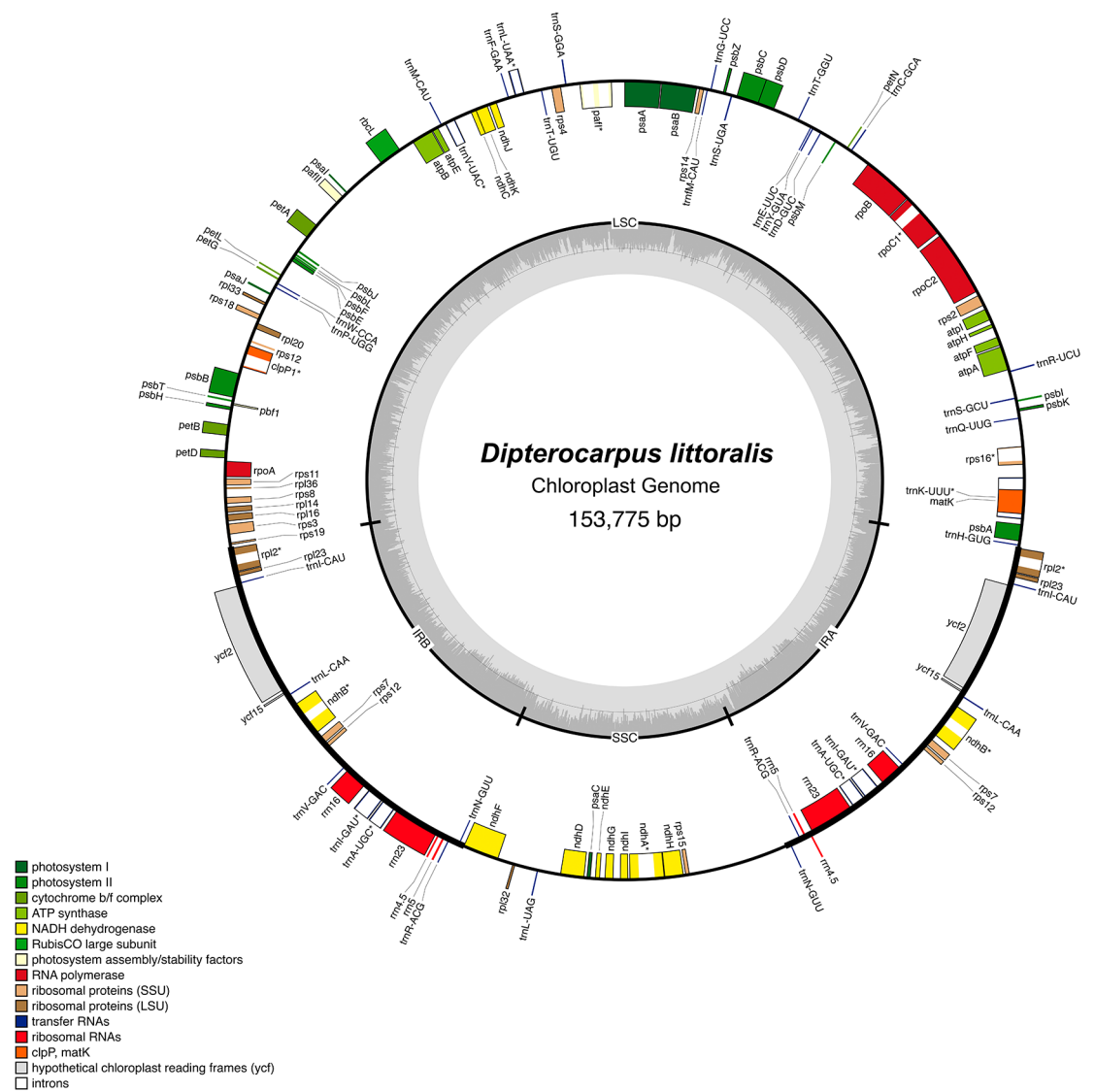


Table 2. List of genes in the chloroplast genomes of *Dipterocarpus littoralis*

Functional category	Gene group	Gene name
Self-replication	Ribosomal RNA	<i>rrn16^d</i> , <i>rrn23^d</i> , <i>rrn4.5^d</i> , <i>rrn5^d</i>
	Transfer RNA	<i>trnA-UGC^{d*}</i> , <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnFM-CAU</i> , <i>trnG-UCC</i> , <i>trnH-GUG</i> , <i>trnI-CAU^d</i> , <i>trnI-GAU^{d*}</i> , <i>trnK-UUU[*]</i> , <i>trnL-CAA^d</i> , <i>trnL-UAA[*]</i> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnN-GUU^d</i> , <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG^d</i> , <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> , <i>trnV-GAC^d</i> , <i>trnV-UAC[*]</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i>
	Large subunit ribosomal proteins (LSU)	<i>rpl2^{d*}</i> , <i>rpl14</i> , <i>rpl16^{d*}</i> , <i>rpl20</i> , <i>rpl23^d</i> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
	Small subunit ribosomal proteins (SSU)	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7^d</i> , <i>rps8</i> , <i>rps11</i> , <i>rps12^{d*}</i> , <i>rps14</i> , <i>rps15</i> , <i>rps16^{d*}</i> , <i>rps18</i> , <i>rps19</i>
	DNA-dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1[*]</i> , <i>rpoC2</i>
Photosynthesis	Subunits of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i>
	Subunits of NADH-dehydrogenase	<i>ndhA[*]</i> , <i>ndhB^{d*}</i> , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
	Subunits of photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i> , <i>pbf1</i>
	Subunits of photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbT</i> , <i>psbZ</i>
	Subunits of cytochrome b/f complex	<i>petA</i> , <i>petB</i> , <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>
	Subunit rubisco	<i>rbcl</i>
Other function	Subunit of acetyl-CoA-carboxylase	<i>accD</i>
	C-type cytochrome synthesis gene	<i>ccsA</i>
	Protease	<i>clpP1[*]</i> , <i>pafI^{**}</i> , <i>pafII</i>
	Maturase	<i>matK</i>
	Envelope membrane protein	<i>cemA</i>
	Translation initiation factor	<i>infA</i>
Unknown function	Conserved open reading frames	<i>ycf1^d</i> , <i>ycf2^d</i>

^dGene duplication in inverted repeat (IR);^{*}Single intron;^{**}Double intron; *rps12* is trans-spliced in Large Single Copy (LSC) and IR.

(NC_058777.1), *D. retusus* (NC_067812.1), *D. turbinatus* (NC_046842), and *D. hasseltii* (NC_065503.1), also showing a bootstrap value of 100% (Figure 3). This pattern aligns with the findings of Tao et al. (2024), which reported a monophyletic clade for the phylogenetic tree of *Dipterocarpus*, indicating a specialized evolutionary lineage. The present study contributes valuable insights for future studies on the molecular identification and evolutionary dynamics of *D. littoralis* and related species, enhancing understanding of the phylogeny of *Dipterocarpus* and the family Dipterocarpaceae.

In this study, 8,399 contigs encoding biosynthetic gene clusters (BGCs) of *D. littoralis* and 77 genes coding for BGCs were identified, comprising 44 terpene clusters, 6 alkaloid clusters, 1 lignan cluster, 4 saccharide clusters, 1 Saccharide-Polyketide cluster, 1 saccharide-terpene cluster, and 20 putative clusters (Table 3), indicating a high potential for

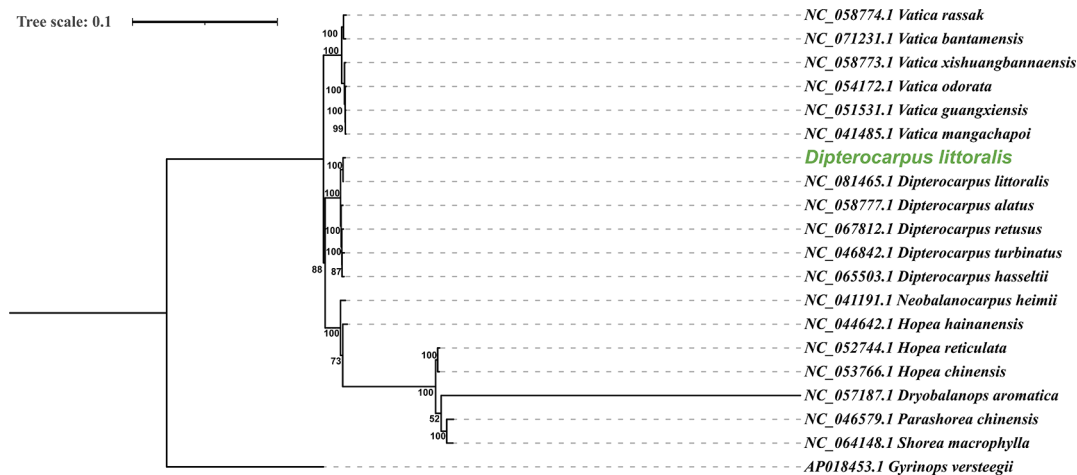


Figure 3. Phylogenetic relationships of *Dipterocarpus littoralis*.

Table 3. The Biosynthetic gene clusters (BGC) of *D. littoralis* in several contigs.

BGC	Total clusters	Cluster size (Kbp)	Sub-clusters
Terpene	44	506,27	Terpene_synth, Terpene_synth_C, p450, SQHop_cyclase_C, SQHop_cyclase_N
Alkaloid	6	111,68	Amino_oxidase, Bet_v_1, AMP-binding, Str_synth, Cu_amine_oxid, Epimerase, p450, Bet_v_1, Lipxygenase, AMP-binding, Str_synth
Lignan	1	10,47	Dirigent, Peptidase_S10
Saccharide	4	62,33	Glycos_transf_1, SQS_PSY, UDPGT_2, p450
Saccharide-Polyketide	1	26,6	Chal_sti_synt_C, Chal_sti_synt_N, UDPGT_2
Saccharide-Terpene	1	35,19	Terpene_synth, Terpene_synth_C, UDPGT_2
putative	20	557,31	AMP-binding, FA_desaturase_2, 2OG-FeII_Oxy, Cellulose_synt, DIOX_N, Methyltransf_11, Methyltransf_2, COesterase, Transferase, p450, Lipxygenase, Aminotran_1_2, Prenyltransf, SE, polyprenyl_synt.

secondary metabolism. At 44 BGC terpene, *D. littoralis* was found to contain gene-producing protein domain including Terpene synthase (involved in sesquiterpene and monoterpene pathways) and SQHop cyclase (involved in triterpenoid pathways) (Yang et al., 2018). Six *D. littoralis* alkaloid BGCs contain gene-producing protein domains, namely, Copper amine oxidase (involved in the biosynthesis of pyrrolizidine/PA alkaloids), Pictet-Spengler enzymes, and Strictosidine synthase-like (Bet v1), both of which are involved in the biosynthesis of monoterpene indole alkaloids (MIA) (Maresh et al., 2008). This information enables an overview of *D. littoralis* by identifying the Pictet-Spengler enzyme and the Strictosidine synthase-like (Bet v1) gene.

A total of four *D. littoralis* saccharide BGCs contain gene-producing protein domains, namely glycosyltransferases (GTs), which are carbohydrate-active enzymes (CAZy) and are the primary catalysts in the biosynthesis and modification of plant cell walls (Hao et al., 2012). Although *D. littoralis* has only one cluster of BGC lignans, this cluster acts as a catalyst for bimolecular coupling reactions that result in regiochemical and stereochemical control (Pickel and Schaller, 2013). Hybrid BGCs that identify saccharides joined to other clusters were also identified in this study, namely, one saccharide-polyketide BGC cluster and one saccharide-terpene BGC cluster (Table 3). The sub-clusters in the saccharide-polyketide BGCs were Stilbene synthase domains and UDPGT2, and the sub-clusters in the saccharide-terpene BGCs were Terpene synthase and UDPGT2. Both BGCs identified the unique characteristics of the secondary metabolites in *D. littoralis*. In addition to these clusters, *D. littoralis* also has 20 putative clusters with sub-clusters, such as CoA-ligase, Fatty acid desaturase, Dioxygenase, Cellulose synthase-like, cytochrome 450, Dioxygenase, Methyltransferase,

COesterase, BAHD acyltransferase, Methyltransferase, Polyphenyl synthetase, Oxidoreductase, Lipoxigenase, Amino-transferase, Squalene epoxidase, Aminotransferase, and Prenyltransferase.

Dataset validation

Not applicable.

Ethical considerations

Not applicable.

Data availability

DNA Data Bank of Japan (DDBJ) BioProject: Forest Tree Species Genome. BioProject Accession number PRJDB20739; <https://ddbj.nig.ac.jp/search/entry/bioproject/PRJDB20739> (Dwiyanti and Rahadian 2025).

DNA Data Bank of Japan (DDBJ) BioSample: *Dipterocarpus littoralis*. The BioSample Accession number SAMD01813819; <https://ddbj.nig.ac.jp/search/entry/biosample/SAMD01813819> (Dwiyanti 2026).

Acknowledgements

The authors gratefully acknowledged Central Java Natural Resources Conservation Agency (BKSDA Jawa Tengah) for providing permission to access the genetic resources of *Dipterocarpus littoralis* seedlings (No. SK/94/K.21/TU/KSA.2/07/2021) and permission to transport the seedlings from West Nusakambangan Nature Reserve, Cilacap Regency, Central Java Province to the Faculty of Forestry and Environment, IPB University, Bogor Regency, West Java Province.

References

- Ashton PS, Heckenhauer J: **Tribe Shoreae (Dipterocarpaceae subfamily Dipterocarpoideae) finally dissected.** *Kew Bull.* 2022; **77**(4): 885–903.
[Publisher Full Text](#)
- Brüna T, Hoff KJ, Lomsadze A, *et al.*: **BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database.** *NAR Genomics and Bioinformatics.* 2021; **3**(1): 108.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cawood GL, Ton J: **Decoding resilience: ecology, regulation, and evolution of biosynthetic gene clusters.** *Trend in Plant Science.* 2025; **30**(2): 185–198.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ci X, Peng J, Shi C, *et al.*: **The complete chloroplast genome of *Dipterocarpus turbinatus* Gaertn. F.** *Mitochondrial DNA Part B Resources.* 2019; **4**(2): 3636–3637.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Doyle JJ, Doyle JL: **Isolation of Plant DNA from Fresh Tissue.** *Focus (Madison).* 1990; **12**(1): 13–15.
- Dwiyanti FG: *Dipterocarpus littoralis*. DNA Data Bank of Japan (DDBJ) BioSample; 2026.
[Reference Source](#)
- De Coster W, D'Hert S, Schultz DT, *et al.*: **NanoPack: visualizing and processing long-read sequencing data.** *Bioinformatics.* 2018; **34**(15): 2666–2669.
[Publisher Full Text](#)
- Dwiyanti FG, Harada K, Siregar IZ, *et al.*: **Population genetics of the critically endangered species *Dipterocarpus littoralis* Blume (Dipterocarpaceae) endemic on Nusakambangan Island, Indonesia.** *Biotropia.* 2014; **21**(1): 1–12.
[Publisher Full Text](#)
- Dwiyanti FG, Pratama R: *Forest Tree Species Genome*. DNA Data Bank of Japan (DDBJ) BioProject; 2025.
[Reference Source](#)
- Greiner S, Lehwark P, Bock R: **OrganelleGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes.** *Nucleic Acids Res.* 2019; **47**(W1): W59–W64.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hamidi A, Robiansyah I: ***Dipterocarpus littoralis*. The IUCN Red List of Threatened Species 2018: e.T33376A125628315.** 2018. Accessed on 08 March 2026.
[Publisher Full Text](#)
- Hao DC, Chen SL, Xiao PG, *et al.*: **Application of high-throughput sequencing in medicinal plant transcriptome studies: Medicinal plant transcriptome studies.** *Drug Dev. Res.* 2012; **73**(8): 487–498.
[Publisher Full Text](#)
- İşik K: **Rare and endemic species: why are they prone to extinction?** *Turk. J. Bot.* 2011; **35**(4): 411–417.
[Publisher Full Text](#)
- Jin JJ, Bin YW, Yang JB, *et al.*: **GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes.** *Genome Biol.* 2020; **21**(1): 231–241.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Katoh K, Standley DM: **MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability.** *Mol. Biol. Evol.* 2013; **30**(4): 772–780.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kautsar SA, Duran HGS, Blin K, *et al.*: **plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters.** *Nucleic Acids Res.* 2017; **45**(1): W55–W63.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kolmogorov M, Yuan J, Lin Y, *et al.*: **Assembly of long, error-prone reads using repeat graphs.** *Nat Biotechnol.* 2019; **37**: 540–546.
[Publisher Full Text](#)
- Letunic I, Bork P: **Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool.** *Nucleic Acids Res.* 2024; **52**(W1): W78–W82.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ma X, Ye Y, Lin R, *et al.*: **Complete chloroplast genome characterization, and phylogenetic analyses of the rare and endangered plant *Platy crater arguta*.** *Biology.* 2025; **14**: 1726.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Maresh JJ, Giddings LA, Friedrich A, *et al.*: **Strictosidine synthase: mechanism of a Pictet-Spengler catalyzing enzyme.** *J. Am. Chem. Soc.* 2008; **130**(2): 710–723.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Newman MF, Burgess PF, Whitmore TC: *Manuals of Dipterocarps for Foresters: Java to New Guinea*. Edinburgh, UK: Royal Botanic Garden Edinburgh and CIFOR; 1998; 107.
- Pickel B, Schaller A: **Dirigent proteins: molecular characteristics and potential biotechnological applications.** *Appl. Microbiol. Biotechnol.* 2013; **97**(19): 8427–8438.
[PubMed Abstract](#) | [Publisher Full Text](#)

Primajati M, Hamidi A, Robiansyah I, *et al.*: **Predicting suitable habitat of the threatened and endangered tree *Dipterocarpus littoralis* in West Nusakambangan Nature Reserve, Indonesia using maximum entropy modelling.** *Proceeding of the International Conference on Tropical Conservation and Utilization. Bogor, 18-20 May 2027.* Center for Plant Conservation Botanic Gardens, Indonesian Institute of Sciences; 2017.

Robiansyah I, Davy AJ: **Population status and habitat preferences of critically endangered *Dipterocarpus littoralis* in West Nusakambangan, Indonesia.** *Makara Journal of Science.* 2015; **19**(4): 150–160.

[Publisher Full Text](#)

Tao YT, Yuan M, Li QQ, *et al.*: **Complete chloroplast genome of endangered species *Dipterocarpus retusus* Blume (Dipterocarpaceae) and its phylogenetic implications.** *Mitochondrial DNA Part B Resources.* 2024; **9**(9): 1162–1165.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Tillich M, Lehwark P, Pellizzer T, *et al.*: **GeSeq - Versatile and accurate annotation of organelle genomes.** *Nucleic Acids Res.* 2017; **45**(W1): W6–W11.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Trifinopoulos J, Nguyen LT, von Haeseler A, *et al.*: **W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis.** *Nucleic Acids Res.*

2016; **44**: W232–W235.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Widiatuti T, Sulistyati AN, Darwoto, *et al.*: **Development of cilacap batik utilizing natural dyes with main motifs of marine life and *Dipterocarpus littoralis* (pelahlar).** *IOP Conf. Ser.: Earth Environ. Sci.* 2022; **1114**: 012031.

[Publisher Full Text](#)

Yang X, Ding W, Yue Y, *et al.*: **Cloning and expression analysis of three critical triterpenoid pathway genes in *Osmanthus fragrans*.** *Electron. J. Biotechnol.* 2018; **36**: 1–8.

[Publisher Full Text](#)

Yin D, Li X, Xiao Z, *et al.*: **Chloroplast genome features and phylogeny of two nationally protected medicinal plants, *Euchresta tubulosa* and *Euchresta japonica*: Molecular resources for identification and conservation.** *Genes (Basel).* 2025; **16**(11): 1286.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Yulita KS, Partomihardjo T: **Population Genetic Diversity of Pelahlar (*Dipterocarpus littoralis* (Bl.) Kurz) in Nusakambangan Island Based on Enhanced Random Amplified Polymorphic DNA.** *Berita Biologi.* 2011; **10**(4): 541–548.

[Reference Source](#)

Open Peer Review

Current Peer Review Status: ? ? X

Version 1

Reviewer Report 22 July 2026

<https://doi.org/10.5256/f1000research.197606.r498950>

© 2026 Hsu S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Sheng-Kai Hsu 

Cornell University, Ithaca, NY, USA

This data note presents a useful conservation genomics resource, and the chloroplast assembly and phylogenetic placement appear sound. However, the biosynthetic gene cluster (BGC) analysis rests on an unstated and unexplained nuclear genome assembly step, and no sequencing depth, coverage, or assembly quality metrics are reported for either genome compartment. Both issues affect reproducibility and the interpretability of the reported results and should be resolved.

My specific comments are as follow:

1. Missing sequencing and assembly quality metrics

No coverage, depth, or completeness statistics are reported for either genome compartment: Total raw output (48.4 Gbp, 40.2M reads) is given without a genome size estimate, so coverage depth cannot be assessed for either the plastome or the implied nuclear assembly.

The statement that "following filtering process, all reads have met the quality assessment criteria and retain the same quality score" is unclear as written. Please state the actual filtering thresholds applied (e.g., minimum length, minimum Q score) and report pre- and post-filtering read/base counts.

No BUSCO (or equivalent) completeness score is reported for the chloroplast assembly or the nuclear contig set, if present. Given that plastome size and gene content are explicitly compared against related species, a completeness metric would substantially strengthen confidence in those comparisons.

2. Undisclosed nuclear genome assembly

The manuscript never states that a nuclear genome was assembled, yet the BGC analysis clearly requires one.

In the Methods, Flye is introduced solely "to construct the plastome," implying an organelle-only assembly target. However, GetOrganelle is introduced subsequently for chloroplast genome reconstruction. Was Flye used for "genome" assembly?

The Biosynthetic gene cluster subsection then opens with "All DNA sequence datasets from the assembly," without specifying which assembly, and proceeds directly into BRAKER2/GeneMark-based gene prediction.

The Results subsequently report "8,399 contigs encoding biosynthetic gene clusters." This is a nuclear scale contig set with no antecedent in the Methods. Given that 77 BGCs (including 44

terpene clusters) cannot plausibly reside on a 153,775 bp plastome, these must derive from a separate nuclear or whole-genome assembly that is never described. No assembler, assembly parameters, genome size estimate, contig N50/total length, or completeness metric (e.g., BUSCO) is provided for this nuclear dataset as suggested above. As written, the BGC pipeline cannot be reproduced or independently assessed.

I recommend the authors add an explicit "Nuclear genome assembly" subsection specifying the assembler and parameters used, basic assembly statistics, and a completeness estimate, and that they release the 8,399 contigs (or the assembly they derive from).

3. Lack of clarity in genome annotation pipeline

The authors mixed in genome annotation pipeline in the section of Biosynthetic gene cluster classification. I recommend the authors add another explicit "genome annotation" subsection explaining how the nuclear genome, if present, was annotated.

The pipeline as it is currently not really clear. The standard annotation workflow is approximately: Genome → GeneMark-ES seed predictions → translation into seed proteins → DIAMOND search against the reference protein database → Spaln alignment of selected homologous proteins to seed genomic regions → ProtHint protein hints → GeneMark-EP+ prediction → AUGUSTUS training and prediction → final BRAKER2 gene models

And then, one would use the final gene models as input for plantiSMASH BGC identification.

Please clarify and provide the exact versions and appropriate citations for tools and databases used in the pipeline.

"AUGUST2" is most likely a typo of "AUGUSTUS"

4. Ethical considerations

This is currently listed as "Not applicable," yet the Acknowledgements describe a formal collection/transport permit from BKSDA Jawa Tengah for a CITES/IUCN-listed critically endangered species. In my opinion, this permit information should be moved into a proper Ethical considerations statement with the permit-issuing authority and number cited.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of methods and materials provided to allow replication by others?

No

Are the datasets clearly presented in a useable and accessible format?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Phylogenomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 22 July 2026

<https://doi.org/10.5256/f1000research.197606.r498943>

© 2026 Chakraborty A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Abhisek Chakraborty

Johannes Gutenberg University Mainz, Mainz, Germany

Dwiyanti et al. performed long-read sequencing of the critically endangered tree *Dipterocarpus littoralis* and provided a genomic resource including its chloroplast genome, phylogeny, and identification of the Biosynthetic Gene Clusters (BGCs). Although the resources will be useful for future research, I have the following comments for improving the manuscript quality:

1. The methods are not clearly described. First, how did the authors perform the raw data filtering? It is also unclear which software was finally used for the chloroplast genome assembly – Flye or GetOrganelle? Also, why is the mention of two chloroplast genome assemblies in the sentence “The resulting chloroplast genomes were then annotated using GeSeq”?
2. Similarly, the BGCs identification methodology is also misleading. For example, the authors should provide detailed methods for the nuclear genome assembly and annotation.
3. All softwares used in the analyses should be cited. For example, tools such as GeneMark-ES, DIAMOND, Spaln, etc. were not cited. “AUGUST2” will be “AUGUSTUS 2”.
4. The first line of the Introduction - “This section should include why the data were gathered or produced” - should be removed.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of methods and materials provided to allow replication by others?

No

Are the datasets clearly presented in a useable and accessible format?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics, Evolutionary Genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 20 May 2026

<https://doi.org/10.5256/f1000research.197606.r486250>

© 2026 D'Agostino N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Nunzio D'Agostino

University of Naples Federico II, Portici, Italy

The manuscript by Dwiyantri et al. describes the long-read sequencing of *Dipterocarpus littoralis*, a critically endangered tropical tree endemic to Nusakambangan Island, Indonesia, to support conservation efforts and clarify its phylogenetic relationships. The study identified 77 biosynthetic gene clusters (BGCs) in the nuclear genome, including terpene, alkaloid, lignan, saccharide, and hybrid clusters associated with secondary metabolite production, while also characterizing the chloroplast genome and confirming the phylogenetic placement of the species within the genus *Dipterocarpus*.

The manuscript contains several inaccuracies and requires improvement in English language and style. The methods applied are standard. I recommend minor revisions, although the overall quality of the manuscript should be strengthened.

The discussion points would also benefit from deeper critical analysis, as several points are currently rather generic. The authors should aim to provide more insightful interpretations while avoiding broad or unsupported statements.

I would also have appreciated the inclusion of line numbers, as they would have facilitated the review process.

Abstract/Background

- It should read "to support germplasm conservation".
- Please remove "However... Therefore" and start the sentence directly with "This study aims...".
- It should be "biosynthetic gene clusters (BGCs)".

Abstract/Methods

- Please remove "*D. littoralis*" from the first line. It should read "from leaf tissue".
- It should read "was reconstructed using GetOrganelle".

Abstract/Conclusions

- Please start the paragraph with "The length...".
- There is no need to report the length of IRA and IRB twice; stating it once is sufficient.
- It should read "the phylogenetic tree indicated that...".

It should also be clearly specified that the BGCs were searched within the nuclear genome, as this is not explicitly stated in the manuscript.

The sentence "as well as facilitate the discovery of potential natural products from *D. littoralis*" appears somewhat overstated. The authors should better clarify how the presented results would concretely facilitate such discoveries, or alternatively moderate the claim.

Introduction

- The sentence *"This section should include why the data were gathered or produced."* appears to be an editorial placeholder and should be removed from the manuscript.
- "Purifying the air" would be more appropriate than "purifying the air". In any case the statement regarding ecosystem services (e.g., supporting biodiversity, regulating climate, and purifying the air) appears overly general, as these functions are common to most forest tree species and are not specifically linked to *D. littoralis*. The authors may consider replacing this sentence with information more directly relevant to the ecological, evolutionary, or conservation significance of the species.
- The acronym should consistently be written as **BGCs**, since it refers to "biosynthetic gene clusters" in the plural form.
- Once the acronym **BGCs** has been defined, the full term followed by the acronym in parentheses does not need to be repeated throughout the manuscript.
- The statement: *"Studying the chloroplast genome and molecular phylogeny is crucial, as these provide essential tools for biodiversity conservation and taxonomic resolution, helping to prioritize and implement protective measures."* appears somewhat overstated with respect to biodiversity conservation. Chloroplast genome analysis and molecular phylogeny are primarily valuable tools for taxonomic resolution and evolutionary inference. The sentence could therefore be moderated by removing the direct reference to biodiversity conservation.

Methods

- The sentence should be revised to: *"DNA extraction and sequencing were conducted..."*
- The use of the plural form in *"The samples were then stored"* is unclear, as the manuscript appears to describe a single biological sample. Please clarify.
- The sentence should read: *"The cp genome was assembled using GetOrganelle."*
- Similarly, the sentence *"The resulting chloroplast genomes were annotated..."* should be revised for consistency, since only one chloroplast genome appears to have been assembled and analyzed.

Results and Discussion

- The authors should specify whether and how chloroplast-derived reads were filtered or extracted from the original FASTQ dataset prior to cp genome assembly.
- Page 7: remove the redundant expression "(cp genome)".
- The statement: *"The present study contributes valuable insights for future studies on the molecular identification and evolutionary dynamics of D. littoralis and related species, enhancing understanding of the phylogeny of Dipterocarpus and the family Dipterocarpaceae."* would benefit from a clearer indication of the types of future studies being referred to (e.g., phylogenomics, conservation genetics, comparative genomics, DNA barcoding, or population-level evolutionary analyses). Adding this information would improve the clarity and broader relevance of the statement.
- In addition, it should be clearly specified that the identified BGCs were detected in the nuclear genome, as this point is currently not sufficiently emphasized in the manuscript.

Data availability and reproducibility

- The authors are encouraged to submit the assembled chloroplast genome to the NCBI Organelle Genome Resources database.
- Furthermore, in accordance with FAIR data principles and to ensure the reproducibility of

the study, the 8,399 nuclear contigs used for the identification of the biosynthetic gene clusters (BGCs) should also be made publicly available.

- The manuscript currently lacks a sufficiently detailed description of how these nuclear contigs were generated, including the assembly strategy, software tools employed, and the parameters used during the analysis. Providing these methodological details would substantially improve the transparency, reproducibility, and overall robustness of the study.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of methods and materials provided to allow replication by others?

Partly

Are the datasets clearly presented in a useable and accessible format?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics and genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research