RESEARCH ARTICLE

SABRAO Journal of Breeding and Genetics
58 (4) 1598-1608, 2026
http://doi.org/10.54910/sabrao2026.58.4.12
http://sabraojournal.org/
pISSN 1029-7073; eISSN 2224-8978

CHLOROPLAST SEQUENCE VARIATIONS AMONG THE *GYRINOPS VERSTEEGII* (GILG.) DOMKE GENOTYPES GENERATED THROUGH NANOPORE SEQUENCING

H. HARTATI[1*], R. PRATAMA[2], I.Z. SIREGAR[3], U.J. SIREGAR[3], N.S. HARTATI[4], S. RAHMAWATI[4], M. ARDIYANI[5], and M. MAJIIDU[6]

[1]Research Center for Applied Botany, National Research and Innovation Agency, Indonesia
[2]Department of Biochemistry, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, Indonesia
[3]Department of Silviculture, Faculty of Forestry and Environment, IPB University, Bogor, Indonesia
[4]Research Center for Genetic Engineering, National Research and Innovation Agency, Indonesia
[5]Research Center for Biosystematics and Evolution, National Research and Innovation Agency, Indonesia
[6]Molecular Science Research Group, Advanced Research Laboratory, IPB University, Bogor, Indonesia
[*]Corresponding author's email: hartati72lipi@gmail.com
Email addresses of co-authors: rahadian@apps.ipb.ac.id, siregar@apps.ipb.ac.id, ulfahjs@apps.ipb.ac.id, nsri001@brin.go.id, syam006@brin.go.id, marl004@brin.go.id, mmajiidu@apps.ipb.ac.id

SUMMARY

Gyrinops versteegii (Gilg.) Domke, an endangered agarwood-producing tree, faces threats from overexploitation and genetic erosion in Indonesia. This study sought to evaluate the intraspecific genetic diversity and molecular markers for conservation via sequencing and assembling of the chloroplast genomes of six *G. versteegii* genotypes from four regions using Oxford Nanopore Technology (ONT). The assemblies ranged from 32,442 to 138,782 bp, and the genotype CLD2 produced the largest and most complete chloroplast draft assembly among the six genotypes (138,782 bp; 126 genes). Comparative analysis of nine chloroplast genes revealed notable variations, particularly in the *atpA* (Pi = 0.09178). The highest genetic distance emerged between genotypes NTB2 (from Nusa Tenggara) and CLD1 (from West Java). The phylogenetic analysis suggested clustering of *G. versteegii* genotypes into three subclusters, separated from other species of the family Thymelaeaceae. Additionally, 235 chloroplast simple sequence repeats (cpSSRs) entailed identification, predominantly A/T motifs, including several novel loci not reported in the reference genomes. These cpSSRs provided valuable resources for species authentication, population monitoring, and conservation breeding. The results revealed intraspecific variation among *G. versteegii* genotypes and demonstrated the utility of ONT sequencing for developing molecular markers in tropical tree species. The genomic insights will support effective conservation strategies and sustainable management of agarwood-producing tree resources.

Communicating Editor: Prof. P.I. Prasanthi Perera

Manuscript received: November 04, 2025; Accepted: June 13, 2026.

Citation: Hartati H, Pratama R, Siregar IZ, Siregar UJ, Hartati NS, Rahmawati S, Ardiyani M, Majiidu M (2026). Chloroplast sequence variations among the *Gyrinops versteegii* (Gilg.) Domke genotypes generated through nanopore sequencing. *SABRAO J. Breed. Genet. 58 (4) 1598-1608. http://doi.org/10.54910/sabrao2026.58.4.12*.

Keywords: *G. versteegii*, chloroplast genomes, genetic distance, genetic diversity, conservation, intraspecific variation, microsatellites, nucleotide diversity

Key findings: This study explored the crucial intraspecific variations in *G. versteegii* genotypes as the endangered agarwood-producing tree, offering insights essential for developing effective conservation strategies for this endangered species.

INTRODUCTION

Agarwood, a high-value non-timber forest product, develops when trees of the family Thymelaeaceae acquire wounds from physical injury and infection via fungi, triggering the accumulation of the fragrant black resin in the wood of stems, branches, and roots. This black resin is well known for its antibacterial properties and extensive uses in perfumery, traditional medicines, and cultural practices (Lee *et al*., 2018; Chen *et al*., 2021). Among agarwood-producing species, the species *Gyrinops versteegii* (Gilg.) Domke is one of Indonesia's most important agarwood resources. This endemic and endangered species has a natural distribution in Papua, Sulawesi, Maluku, Lombok, Flores, and Sumba regions, Indonesia (Lee *et al*., 2018). Given its highest commercial and cultural values, the demand for agarwood has gained considerable enhancement worldwide, often driving unsustainable harvesting and exploitation. As a result, the wild germplasm of the species *G. versteegii* has declined, raising urgent concerns about its genetic resource management and conservation.

The current listing of *G. versteegii* falls under Appendix II of CITES (Convention on International Trade in Endangered Species) and has become an endemic species with increasing threats in its natural habitat. The overexploitation and the naturally low occurrence of agarwood formation in trees together reduce the regeneration potential and genetic diversity of this important species. In addressing these issues, effective conservation strategies, including habitat protection, sustainability, and regeneration programs, are urgently necessary (Irsyad *et al*., 2020).

In germplasm, the loss of intraspecific variation is one of the most crucial threats to species persistence. Genetic variation underpins capacity, enabling the population to respond to environmental variations (Henry, 2022). Reduced genetic diversity increases vulnerability and limits the efficiency of restoration and breeding programs. Consequently, for conservation strategies, ex-situ collections, and trade regulations, understanding the extent and structure of genetic variation in the species *G. versteegii* is essential. Genomic tools also provide powerful approaches for characterizing genotypic variations, identifying molecular markers, and supporting both conservation and breeding initiatives for endangered species (Henry, 2022).

Chloroplast (cp) genome sequencing has emerged as a valuable strategy and plays a vital role in resolving genetic diversity and evolutionary relationships in plants. In angiosperms, the cp genome is a relatively conserved circular molecule that contains 100–130 genes involved in managing photosynthesis and other key metabolic processes. Its quadripartite structure comprises a large single-copy (LSC) region, a small single-copy (SSC) region, and two inverted repeats (IRs) (Wang *et al*., 2016, 2018). Despite its conserved nature, the cp genome has numerous microsatellites (SSRs), single nucleotide polymorphisms (SNPs), and indels, which serve as informative markers in the germplasm at interspecific and intraspecific levels (Cao *et al*., 2018). The maternal inheritance of cp genomes also provides critical insights into plant phylogeography, population differentiation, and evolutionary history (Lee *et al*., 2022; Sun *et al*., 2023). Compared with nuclear genomes, chloroplast genomes are smaller, more conserved, and easier to assemble, making them suitable for comparative genomics and marker

development in non-model species (Wang *et al*., 2016; Wang *et al*., 2018). Their uniparental inheritance and limited recombination also provide valuable information for phylogeographic and evolutionary studies (Lee *et al*., 2022; Sun *et al*., 2023). In *Gyrinops versteegii*, where publishing nuclear genome research has yet to succeed, chloroplast genome sequencing offers a practical approach for assessing genetic variation and supporting conservation efforts.

Despite the potential of chloroplast genomes for genetic resource assessment, information on chloroplast genome variation within *G. versteegii* remains limited. Understanding such variation is important for identifying informative markers, evaluating genetic diversity, and supporting conservation planning in this endangered species. Recent advances in long-read sequencing (LRS) technologies, such as Oxford Nanopore Technology (ONT), have transformed the genome sequencing efforts, particularly in tropical and non-model species. The ONT enables real-time sequencing of single DNA molecules without amplification and facilitates recovery of the complete cp genomes (Wang *et al*., 2018; Siregar *et al*., 2021). Previous studies have reported chloroplast genome assemblies of *G. versteegii* based on Illumina sequencing, including an assembly from a single genotype (Hartati *et al*., 2024). However, comparative chloroplast genome analyses across multiple genotypes representing different geographic origins have not progressed. Additionally, the potential of Oxford Nanopore Technology (ONT) for recovering chloroplast genomic resources and identifying chloroplast simple sequence repeat (cpSSR) markers in this endangered species remains largely unexplored. Therefore, this study represents the first comparative ONT-based analysis of the chloroplast genome variation across multiple *G. versteegii* genotypes from different regions of Indonesia and provides new genomic resources for genetic monitoring and conservation. However, the chloroplast genome variation across geographically distinct *G. versteegii* genotypes remains poorly understood, with no comparative ONT-based study having been conducted for this species.

The following study aimed to a) assemble chloroplast genomes from six *G. versteegii* genotypes collected across four Indonesian regions using ONT sequencing, b) assess intraspecific genetic variation through comparative analysis of chloroplast genes, c) study phylogenetic relationships within the family Thymelaeaceae, and d) identify chloroplast microsatellites as potential molecular markers. By integrating the genome assembly, nucleotide diversity analysis, and SSR discovery, this research work provides new insights into the genetic resources of the species *G. versteegii* and supports conservation strategies for this endangered agarwood tree species.

MATERIALS AND METHODS

Plant material and DNA extraction

Six *G. versteegii* genotypes collected from different regions of Indonesia included Bangka Island (voucher specimen FR01 Bangka) and Bogor Botanical Garden (voucher specimen KRB VIIG.232, KRB VIIB.117A). The others came from West Nusa Tenggara (NTB2; originating from West Nusa Tenggara and sampled from an *ex situ* individual maintained at KST Soekarno, BRIN, Cibinong, West Java [−6.49565, 106.84592]) and Cilodong, West Java (CLD1: −6.43732, 106.84438; CLD2: −6.43743, 106.84428). Genomic DNA from three genotypes, CLD1, CLD2, and NTB2, underwent extraction from fresh leaves using the Plant Genomic DNA Mini Kit (Geneaid). Meanwhile, DNA from three other genotypes, KRB VIIB.117A, KRB VIIG.232, and FR01 Bangka dried leaf specimens, incurred isolation using the CTAB method (Doyle and Doyle, 1990). The DNA quality and concentration assessment employed three complementary methods: gel electrophoresis, NanoPhotometer NP80 (IMPLEN), and Qubit dsDNA BR Assay (Life Technologies). Only DNA samples that

met quality requirements for ONT library preparation were successful for sequencing use. Thus, despite differences in sample preservation and extraction methods, all samples passed the same quality-control procedures before sequencing.

Library preparation, sequencing, and assembly

The library preparation, sequencing process, and bioinformatic analyses proceeded by following the protocol of Siregar *et al*. (2021). Total genomic DNA used was without prior chloroplast enrichment or chloroplast-specific read extraction. The raw reads filtering used NanoFilt v1.28.2 (Q ≥ 7; length ≥ 500 bp), quality-checking was with NanoPlot (De-Coster *et al.,* 2018), and further processing was with Filtlong v0.2.0 to retain the high-quality long reads (Wick, 2018). Upon performing the de novo assembly, the study used Flye v2.8.3 (Kolmogorov *et al.,* 2020), with an estimated genome size of 700 Mb. Assembly statistics came from using QUAST v5.0.2 (Mikheenko *et al.,* 2018) with a comparison against *Gyrinops walla* (NC_060363.1) and *G. versteegii* (NC_065047.1) references (Chen *et al.,* 2021; Lee *et al*., 2022). Carrying out the cp genome annotation and visualization used GeSeq (Tillich *et al.,* 2017), followed by gene completeness assessment with BUSCO using *G. versteegii* (NC_065047.1) as a reference. The assemblies served as draft chloroplast genomes for comparative analysis of gene content and sequence variation, without additional polishing. Read depth had no separate estimation for chloroplast-specific reads due to the whole-genome assembly approach. Afterward, depositing the assembled sequences in the DNA Data Bank of Japan (DDBJ) under BioProject PRJDB12000 had DDBJ accession numbers corresponding to assembled cp sequences: FR01 Bangka (DRR314202), KRB VIIG.232 (DRR314206), KRB VIIB.117A (DRR314205), NTB2 (DRR314207), CLD1 (DRR314203), and CLD2 (DRR314204).

Genetic variation and phylogenetic analysis

The use of full-length chloroplast-derived genes obtained from six *G. versteegii* genotypes sought to assess intra- and interspecific variations before further performing phylogenetic analyses. Homologous cp sequences from the related species of the family Thymelaeaceae came from NCBI GenBank. Using *Eucalyptus grandis* (NC_014570.1) served as the outgroup, while *Gonystylus affinis* (NC_052860.1), *Daphne tangutica* (NC_042950.1), *G. walla* (NC_060363.1), *Aquilaria malaccensis* (NC_041117.1), and *Aquilaria sinensis* (MN147870.1) were the included sister taxa. Multiple sequence alignment, as performed, utilized MAFFT v7 (Rozewicki *et al*., 2019). Genes used for nucleotide diversity and phylogenetic analyses entailed selection based on their recovery from the assembled chloroplast genomes and their reported informativeness in agarwood-producing species. The study included nine chloroplast genes (*matK*, *atpA*, *rpoC1*, *rpoC2*, *cemA*, *rps16*, *rpl33*, *petA,* and a*ccD*) because they were among the best-recovered loci across the genotypes, having been previously used for comparative and phylogenetic analyses in *Aquilaria* and related taxa (Hishamuddin *et al*., 2020). Maximum likelihood (ML) phylogenetic trees succeeded in their generation in MEGA7 (Kumar *et al.,* 2016) using the Tamura-3 parameter model with 1,000 bootstrap replicates (Hartati *et al*., 2024). Estimation of the nucleotide diversity for intra- and interspecific comparisons engaged DnaSP v6 (Rozas *et al.*, 2017).

Identification of chloroplast microsatellites

Chloroplast simple sequence repeats (cpSSRs) identification used MISA (http://pgrc.ipk-gatersleben.de/misa/) (Thiel *et al*., 2003), with thresholds of ≥10 repeats for mononucleotides, ≥6 for dinucleotides, and ≥5 for tri- to

hexanucleotides (Kuntal *et al*. 2012). Detected SSRs received categorization by motif type and repeat length.

RESULTS AND DISCUSSION

Chloroplast genome assembly and annotation

Draft chloroplast genome assemblies succeeded in their generation from six genotypes of the *G. versteegii* using Oxford Nanopore sequencing. Assembly quality varied substantially in contiguity, size, and completeness (Table 1), resulting in assemblies that ranged from partial chloroplast genome fragments to a near-complete draft assembly. Nevertheless, a substantial proportion of chloroplast genes attained recovery across the genotypes. Among the six genotypes, CLD2 produced the largest and most complete draft chloroplast assembly (138,782 bp; 126 genes; GC 38.7%; DRR314204), while other genotypes showed shorter assemblies of varying lengths as follows: CLD1, 54,807 bp (DRR314203); KRB VIIB.117A, 97,859 bp (DRR314205); KRB VIIG.232, 101,014 bp (DRR314206); NTB2, 54,778 bp (DRR314207); and Bangka, 32,442 bp (DRR314202). Consequently, subsequent comparative analyses continued using the recovered chloroplast regions and underwent interpretations considering differences in assembly completeness. Our previous Illumina-based assembly of the genotype CLD1 produced a complete cp genome of 174,814 bp (Hartati *et al*., 2024). The size of the CLD2 assembly approaches the expected chloroplast genome length reported for *G. versteegii* (~174.8 kb, encoding 99 protein-coding genes, 42 tRNAs, and 8 rRNAs); although, it remains shorter than the complete chloroplast genomes previously published using Illumina sequencing (Hartati *et al*., 2024; Lee *et al*., 2022), indicating that this present assembly should only be a draft chloroplast genome.

Genotype CLD2 annotation had a total of 126 genes. By comparing with the reference genome of *Aquilaria* and *G. versteegii* CLD1 (Hartati *et al*. 2024), nearly all core genes' identification was successful in the partial assemblies. However, a few genes did not assemble completely. Among photosynthesis-related genes, non-recovery was with clpP1, while in the self-replication category, the non-recovered genes included *pbf1, rps8, rps11, rps22, rpl12, rpl14, rpl16, rpl32, rpl36, trnL-GAU, trnL-UAG, rpoA,* and *psaC* (Table 2). The absence of these genes likely reflects incomplete assembly coverage rather than true gene loss. Previous studies have reported that the same loci are present in the complete genomes of *A. sinensis* and *G. versteegii* (Chen *et al*., 2021; Hartati *et al*., 2024).

Genotype CLD2 spanned with all four regions, suggesting structural conservation typical of angiosperm cp genomes (Figure 1A). The results revealed the LSC extended from *trnH–rps19*, IRB from *rps19–ndhF*, SSC from *ndhF–rpl32*, and IRA from *trnL-UAG–rpl2*. The overall gene order was consistent with reference species *G. versteegii* cp genomes (NC_065047.1) (Lee *et al*., 2022) and genotype CLD1 sequence in our earlier studies (Hartati *et al*., 2024). According to chloroplasts in angiosperms, the LSC contained more genes than the SSC and GC contents of 30%–40% (Wang *et al*., 2016; Zhang *et al*., 2019; Chen *et al*., 2021). Similar quadripartite structural stability has also been evident in tropical woody taxa, such as *Diospyros celebica*, confirming that ONT sequencing produced the reliable assemblies even in the non-model forest species (Siregar *et al*., 2021).

The presented study represents the first ONT-based cp genome assemblies for multiple *G. versteegii* genotypes, providing a new genomic resource for comparative analysis and future conservation. Despite variations in assembly completeness, the recovered chloroplast sequences enabled comparative analyses and the identification of informative cpSSR markers. This is especially important since agarwood production still relies heavily on wild harvesting in Indonesia. As a result, such data are crucial for monitoring genetic diversity and supporting conservation of this CITES-listed endangered species.

Table 1. Statistics of raw, filtered, and assembled reads of six *G. versteegii* genotypes.

Information	Genotypes					
	FR01 Bangka	CLD1	CLD2	KRB VIIB.117A	KRB VIIG.232	NTB2
Raw reads						
Number of raw reads	150,958	373,195	276,853	787,412	279,608	342,262
Mean read length	651.8	3,392.1	3,732.7	926.2	1,092.1	4,011.3
Mean read quality	12.3	12.4	12.4	12.3	12.3	12.4
Read length N50 (bp)	1,003	10,574	11,057	1,494	2,027	11,800
Total base (bp)	98,387,442	1,265,918,527	1,033,399,890	729,286,993	305,396,123	1,372,927,446
Filtered reads						
Number of filtered reads	55,909	237,964	188,908	418,166	148,768	246,603
Mean read length	1,265.3	5,136.7	5,306.0	1,454.4	1,766.1	5,430
Mean read quality	12.1	12.2	12.3	12.2	12.1	12.3
Read length N50 (bp)	1,475	10,977	12,694	1,803	2,461	12,166
Total base (bp)	70,740,753	1,222,348,714	1,002,344,578	608,192,262	262,745,389	1,339,049,827
Assembled reads						
Number of contig	133	3,090	1953	247	160	3833
Largest contig	26,148	219,270	198,786	97,859	101,014	260,780
Total length	728,287	76,564,006	49,054,637	1,269,936	947,501	104,172,910
N50	7,958	42,447	47,677	10,995	13,737	50,436
GC (%)	38.95	38.47	38.68	40.77	39.74	38.02
Percentage of cp genome fraction	29.89%	84.656%	75.948%	81.380%	73.963%	75.853%

Table 2. Chloroplast genes identified in CLD2 genotypes.

Category and gene functions	*G. versteegii* CLD2
Gene for photosynthesis	
Subunits of photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbT, psbZ*
Subunits of Cytochrome b/f complex	*petA, petB, petD, petG, petL, petN*
Subunit of ATP synthase	*atpA, atpB, atpE, atpF, atpH, atpI*
Subunit of rubisco	*rbcL*
Maturase	*matK*
Self-replicating system	
Ribosomal RNA genes	*rrn4.5, rrn4.5/ rrn4.5S, rrn5S, rrn5, rrn5S/rrn5, rrn16/rrn16S (2), rrn23/ rrn23S*
Transfer RNA genes	*trnA-UGC, trnC-GCA, trnD-GUC*, *trnE-UUC, trnF-GAA, trnG-GCC*|*trnG-UCC, trnH-GUG, trnI-CAU, trnI-GAU, trnK-UUU, trnL-UAA, trnfM-CAU, trnM-CAU, trnN-GUU, trnP-UGG, trnQ-UUG, trnR-ACG, trnS-GCU, trnS-UGA, trnS-GGA, trnT-GGU, trnV-UAC, trnW-CCA, trnY-GUA*
Small subunit of ribosome	*rps2, rps3, rps4, rps7, rps12, rps14, rps15, rps16, rps18*
Large subunit of ribosome	*rpl2, rpl14, rpl16, rpl20, rpl22, rpl23, rpl33*
Subunit of NADH dehydrogenase	*ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
DNA dependent RNA polymerase	*rpoB, rpoC1, rpoC2*
Subunit of photosystem I	*psaA, psaB, psaI, psaJ*
Other genes	
Photosystem assembly/RNA polymerase associated factor	*paf1*, pafII*
Photosystem biogenesis factor 1	*pbf1*
Subunit of acetyl-CoA carboxylase	*accD*
C-type cytochrome synthesis gene	*ccsA*
Envelope membrane protein	cemA
Genes of unknown function	
Conserved open reading frames	*ycf1, ycf2, ycf15*

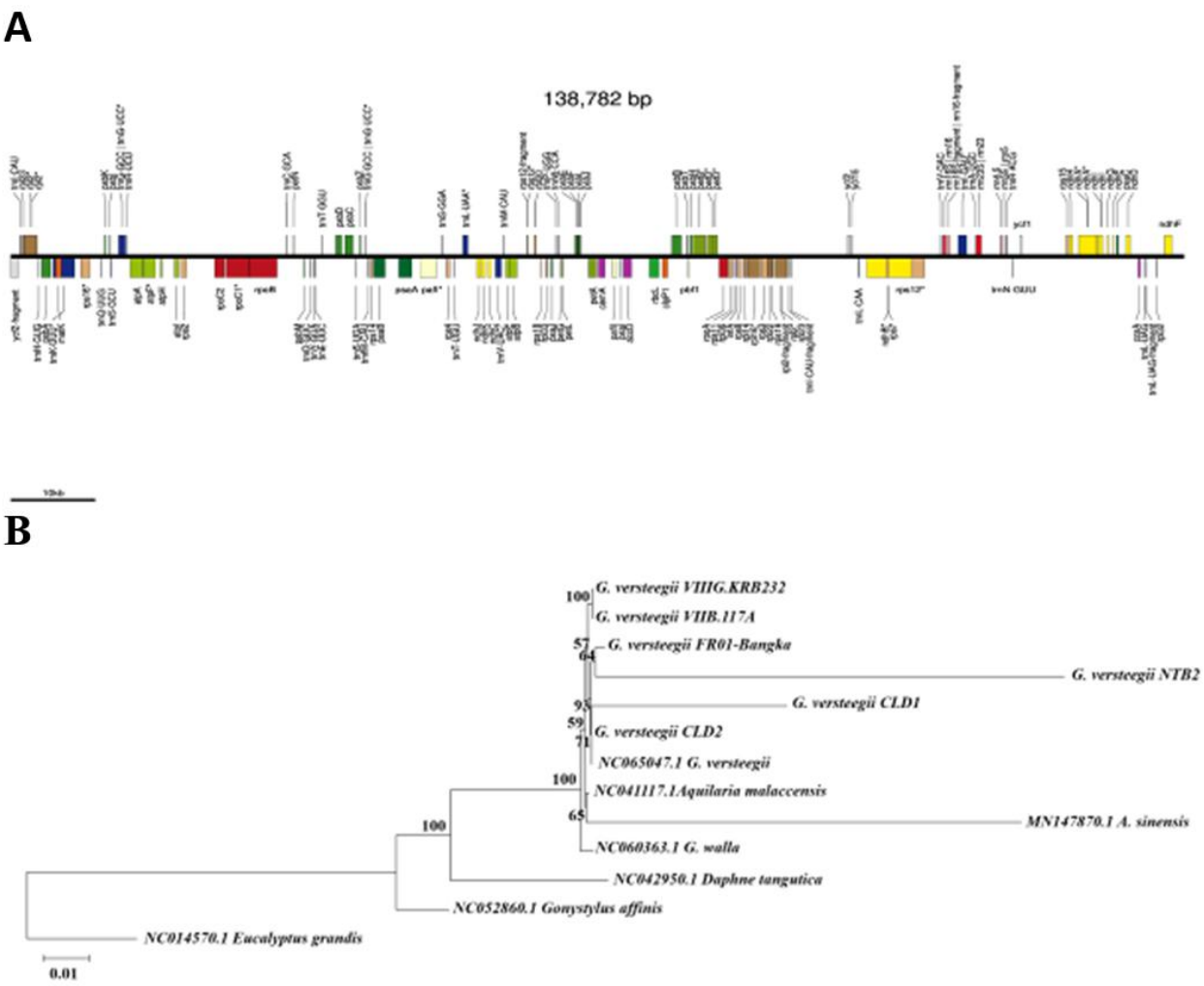

Figure 1. Comparative features of the *Gyrinops versteegii* cp genomes. A) Linear structure of partially assembled chloroplast genome of *G. versteegii* CLD2 (138,782 bp) generated with Flye and visualized with OGDRAW. B) Maximum likelihood phylogenetic tree constructed using nine chloroplast genes obtained from this study and from the NCBI GenBank. Using *E. grandis* served as an outgroup, while using *G. affinis* and *D. tangutica* was as a sister group. Bootstrap values are shown on the branches.

Intra- and interspecific variations and phylogenetic relationships

Nine chloroplast genes, i.e., *matK, atpA, rpoC1, rpoC2, cemA, rps16, rpl33, petA,* and *accD,* incurred comparison across the six *G. versteegii* genotypes. Of these full-length genes, obtaining *rpl33* and *rps16* resulted in all genotypes, while the genes *atpA, rpoC1, cemA*, and *accD* emerged in the subsets of genotype samples. The remaining three genes (*rpoC2, matK*, and *petA*) entailed partial assembly. Reference genome assessment and comparison revealed that these genes exhibited the expected gene sizes. Finally, the divergent regions (*matK, rps16, rpoC1, rpoC2, petA*, and *cemA*) succeeded in their selection for further analysis based on their reported informativeness in *Aquilaria* (Hishamuddin *et al*., 2020).

Genetic variation assessment ensued through nucleotide differences and p-distance values. Interspecific distances ranged from 0.000 to 0.244, with the largest divergence observed between the species *E. grandis* and *A. sinensis*. Within species *G. versteegii*, the

highest intraspecific distance (0.135) appeared between the genotypes CLD1 and NTB2, while the lowest distance (0.000) recorded was between the genotype CLD2 and the *G. versteegii* reference (NC_065047.1), as well as between KRB VIIB.117A and KRB VIIG.232. Pairwise comparison revealed substantial variations (2.5%–13.5%) by comparing the genotype NTB2 with all other genotypes. The results disclosed substantial intraspecific variation, which could serve as a valuable genetic resource for future conservation. The results were consistent with a previous report that intraspecific variation had a general expectation to be <3% (Hebert *et al*., 2003). However, greater thresholds of 0%–5% (Lasso*,* 2008) and up to 12.16% to 18.1% (Ma *et al*., 2022) were also outcomes reported, suggesting that no universal barcoding gap exists. Moreover, in chloroplast genomes, the intraspecific variation was typically low due to reduced mutation rates and uniparental inheritance (Chang *et al*., 2021). The observed intraspecific differences indicate detectable chloroplast sequence variation among sampled genotypes.

Nucleotide diversity (Pi) reached further estimation for each cp gene using MAFFT-based alignments. Interspecifically, the genes *atpA* and *rpoC1* (in the LSC region) showed the highest diversity (Pi = 0.0736 and 0.0940, respectively), followed by the genes *petA* (0.0449) and *accD* (0.0315). However, other genes, including *matK*, were below 0.0300 and showed relatively low variation. Ardiyani *et al*. (2019) also reported that *matK* expressed insufficient phylogenetic resolution for distinguishing the family Thymelaeaceae species. Hishamuddin *et al.* (2020) also explored that genes *atpA* and *rpoC1* were evident among the most variable loci in *Aquilaria* cp genomes. Intraspecific comparisons within *G. versteegii* genotypes detailed that only the gene *atpA* exhibited notable variation (Pi = 0.09178), whereas other genes were highly conserved, such as *rpl33* = 0.0015, *rpoC1* = 0.0003, *accD* = 0.0010, and *cemA*, *matK, petA*, and *rpoC2* = 0.0000. The highest variability in the gene *atpA* likely explains the greater genetic distance of the genotypes CLD1 and NTB2. Similar elevated divergence in genes *atpA* and *accD* has shown links to local structural rearrangements and minor expansions of the LSC region in other woody taxa, including *A. sinensis* and *Cinnamomum camphora* (Xiao-Ming *et al*., 2017; Lee *et al*., 2022; Sun *et al*., 2023). Collectively, the results highlighted *atpA* as a potential high-resolution marker for intraspecific differentiation in the species *G. versteegii*.

Phylogenetic analysis grouped the *Gyrinops* and *Aquilaria* into distinct clades, separate from *D. tangutica* (Figure 1B). The species *G. walla* formed a separate branch from *G. versteegii* genotypes, with a moderate bootstrap value, consistent with previous cp phylogenies of the family Thymelaeaceae (Lee *et al*., 2018; Chen *et al*., 2021). Within *G. versteegii*, the recovery of three inferred subgroups comprised a) CLD1–CLD2, b) NTB2, and c) KRB VIIB.117A–KRB VIIG.232, suggesting potential chloroplast differentiation among the sampled genotypes. Although the tribe *Aquilariae* was monophyletic, moderate bootstrap support (59%) suggested limited resolution with the nine loci. A comparable low bootstrap support value for species resolution has had reports in *matK*-based phylogenies of *Aquilaria* and *Gyrinops* (Ardiyani *et al*., 2019). Nevertheless, the results revealed chloroplast sequence variation among *G. versteegii* genotypes, consistent with earlier findings from both nuclear and cp datasets (Irsyad *et al*., 2020; Hartati *et al*., 2024). Genetic diversity is fundamental to resilience against environmental variations and provides the basis for conservation breeding (Henry, 2022). The outcomes further highlighted the utility of ONT sequencing for detecting chloroplast sequence variation, thereby providing valuable tools for phylogenetics, barcoding, and conservation genomics of endangered agarwood species such as *Gyrinops* and *Aquilaria*.

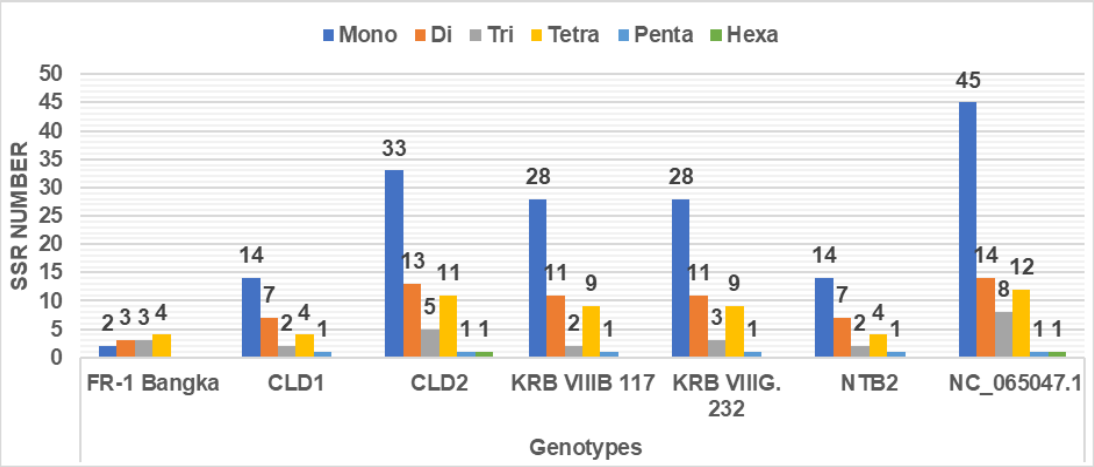

Figure 2. Microsatellite or cpSSRs in *G. versteegii* chloroplast genomes.

Microsatellite identification

Chloroplast simple sequence repeats (cpSSR) identified totaled 235 across six *G. versteegii* genotypes, ranging from 12 to 64 loci (Figure 2). The majority of motifs were A/T mononucleotides (50.6%), followed by di- and tetra-nucleotides, while penta- and hexa-nucleotide motifs were rare observations. The distribution was slightly lower than that observed in *Aquilaria* species, where A/T mononucleotide repeats represent up to 97.8% of total cpSSRs (Hishamuddin *et al*., 2022). Notably, several unique repeat motifs (such as AG, ATT, AAGG, and TAAAA) gained detection in the species *G. versteegii* genotypes, however not evident in the published *Aquilaria,* thereby expanding the marker set for *G. versteegii*. Similar cases of novel cpSSR emergence have successful reports in the species *Dioscorea polystachya* and *Cinnamomum camphora*, suggesting lineage-specific microsatellite evolution within long A/T tracts (Xiao-Ming *et al*., 2017; Cao *et al*., 2018).

Moreover, longer mononucleotide repeats (up to 16 bases) were notable, exceeding the typical 8–15 repeats as previously reported in cp genomes (Ozyigit *et al*., 2015). Such longer repeats often occurred to be associated with higher mutation rates and served as potential mutational hotspots that promote local sequence instability (Fan and Chu, 2007). Overall, in species *G. versteegii*, the cpSSR distribution was broadly similar to *Aquilaria*, and the presence of longer and unique motifs revealed a more dynamic cp genome (Hishamuddin *et al*., 2022). Thus, these markers proved promising for population genetics monitoring, species authentication, and phylogeographic studies in *Gyrinops* and related agarwood-producing taxa.

Implications for breeding and conservation

This study provides the first ONT-based cp genome assemblies across multiple *G. versteegii* populations from different geographic origins in Indonesia and reveals detectable chloroplast sequence variation among genotypes. The newly identified variable loci (*atpA* and *rpoC1*) and 235 chloroplast microsatellites offer practical tools for genetic monitoring, species authentication, and enforcement of CITES trade regulations. Furthermore, these markers could play a vital role in conservation breeding and ex-situ collection by ensuring the preservation of divergent genotypes, thereby strengthening adaptive potential. By integrating chloroplast genome sequencing, phylogenetic analysis, and SSR discovery, the promising study provides genomic resources that can support conservation planning and sustainable agarwood management. The identified

chloroplast marker, including variable genes and cpSSRs, will be useful for assessing genetic diversity, monitoring population structure, tracing the geographic origin of planting materials, and guiding germplasm conservation strategies in *G. versteegii*.

CONCLUSIONS

The innovative study provides the first Oxford Nanopore-based draft chloroplast genome assemblies for multiple genotypes of the species *G. versteegii*, and the genotype CLD2 produced the most complete assembly. Comparative analysis of nine chloroplast genes revealed notable intraspecific variations, particularly within the gene *atpA,* which may contribute to genetic differentiation among genotypes. Phylogenetic reconstruction grouped the species *G. versteegii* genotypes into distinct subclusters, suggesting the species' complex genetic structure within the species. The 235 chloroplast microsatellite loci identified included several novel motifs, offering valuable tools for species authentication, genetic monitoring, and marker development. The results highlighted the Nanopore sequencing's utility for resolving fine-scale genetic variation, reinforced the need to conserve the population, and provided practical resources for sustainable management of agarwood-producing trees. Future studies should expand sampling across the species' natural distribution range and validate the identified cpSSR markers for population genetics and conservation applications.

ACKNOWLEDGMENTS

This work received support from the Rumah Program Organisasi Riset Hayati dan Lingkungan, BRIN (FY 2020; 9/III/HK/2022-RP1WB2-006), with sample use permitted by the Research Center for Applied Botany-BRIN, the Research Center for Biosystematics and Evolution, and the Plant Conservation and Botanic Gardens-BRIN under the CITES project. We sincerely thank Ayub Muswir for providing several *Gyrinops* samples.

REFERENCES

Ardiyani M, Wardani W, Sulistyaningsih LD, Susila (2019). DNA barcoding of Indonesian agarwood using markers from chloroplast and nuclear genomes. In: Proceedings of the National Seminar on Conservation and Utilization of Plants and Wildlife, Bogor, Indonesia.

Cao J, Jiang D, Zhao Z, Yuan S, Zhang Y, Zhang T, Zhong W, Yuan Q, Huang L (2018). Development of chloroplast genomic resources in Chinese Yam (*Dioscorea polystachya*). *Biomed Res. Int*. 2018:6293847.

Chang H, Zhang L, Xie H, Liu J, Xi Z, Xu X (2021). The conservation of chloroplast genome structure and improved resolution of infrafamilial relationships of Crassulaceae. *Front. Plant Sci.* 12: 631884.

Chen J, Lee SY, Munugoda KD, Mohamed R, Subasinghe SMCUP, Liao W (2021). The complete chloroplast genome of Walla Patta, *Gyrinops walla* (Thymelaeaceae), an agarwood-producing tree species from Sri Lanka. *Mitochondrial DNA B Resour*. 6(6): 1699–1701.

De-Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C (2018). Nanopack: Visualizing and processing long-read sequencing data. *Bioinformatics* 34(15): 2666–2669.

Doyle JJ, Doyle JL (1990). Isolation of plant DNA from fresh tissue. *Focus* 12(1): 13–15.

Fan H, Chu JY (2007). A brief review of short tandem repeat mutation. *Genom. Proteom. Bioinform.* 5(1): 7–14.

Hartati H, Cartealy IC, Supatmi S, Rahmawati S, Hartati NS, Siregar UJ, Siregar IZ (2024). A comparative study of the organellar genome of *Gyrinops versteegii* and *Aquilaria malaccensis*. *J. Manajemen Hutan Tropika*. 30(3): 326–335.

Hebert PD, Cywinska A, Ball SL, deWaard JR (2003). Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270: 313–321.

Henry RJ (2022). Progress in plant genome sequence. *Appl. Biosci.* 1: 113–128.

Hishamuddin MS, Lee SY, Ng WL, Ramlee SI, Lamasudin DU, Mohamed R (2020). Comparison of eight complete chloroplast genomes of the endangered *Aquilaria* tree species (Thymelaeaceae) and their phylogenetic relationships. *Sci. Rep.* 10: 13034.

Hishamuddin MS, Samsuddin AS, Mohamed R (2022). Mining and analysis of chloroplast

simple sequence repeat (SSRs) from eight species of *Aquilaria*. *Turkish J. Bot.* 46(1): 26–36.

Irsyad AF, Rindyastuti R, Yuslistyarini T, Darmayanti AS, Daryono BS (2020). Genetic variation of agarwood producing tree (*Gyrinops versteegii*) from Pongkor, Manggarai District, Flores Island Indonesia using ISSR molecular markers. *Biodiversitas* 21(2): 485–491.

Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, Pevzner PA (2020). MetaFlye: Scalable long-read metagenome assembly using repeat graphs. *Nat. Methods* 17: 1103–1110.

Kumar S, Stecher G, Tamura K (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33(7): 1870–1874.

Kuntal H, Sharma V, Daniell H (2012). Microsatellite analysis in organelle genomes of Chlorophyta. *Bioinformation* 8(6): 255–259.

Lasso E (2008). The importance of setting the right genetic distance threshold for identification of clones using amplified fragment length polymorphism: A case study with five species in the tropical plant genus *Piper. Mol. Ecol. Resour.* 8(1): 74–82.

Lee SY, Turjaman M, Chaveerach A, Subasinghe SMCUP, Fan Q, Liao W (2022). Phylogenetic relationships of *Aquilaria* and *Gyrinops* (Thymelaeaceae) revisited: Evidence from complete plastid genomes. *Bot. J. Linn. Soc.* 200(3): 344–359.

Lee SY, Turjaman M, Mohamed R (2018). Phylogenetic relatedness of several agarwood-producing taxa (Thymelaeaceae) from Indonesia. *Trop. Life Sci. Res.* 29(2): 13–28.

Ma Z, Ren J, Zhang R (2022). Identifying the genetic distance threshold for Entiminae (Coleoptera: Curculionidae) species delimitation via COI barcodes. *Insects* 13(3): 261.

Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34(13): 142–150.

Ozyigit II, Dogan I, Filiz E (2015). In silico analysis of simple sequence repeats (SSRs) in chloroplast genomes of 'Glycine' species. *Plant Omics* 8(1): 24–29.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol*. 34(12): 3299–3302.

Rozewicki J, Li S, Amada KM, Standley DM, Katoh K (2019). MAFFT-DASH: Integrated protein sequence and structural alignment. *Nucleic Acids Res*. 47(1): 5–10.

Siregar IZ, Dwiyanti FG, Pratama R, Matra DD, Majiidu M (2021). Generating long-read sequences using Oxford Nanopore Technology from *Diospyros celebica* genomic DNA. *BMC Res. Notes.* 14: 75.

Sun J, Wang Y, Qiao P, Zhang L, Li E, Dong W, Zhao Y, Huang L (2023). *Pueraria montana* population structure and genetic diversity based on chloroplast genome data. *Plants* 12(12): 2231.

Thiel T, Michalek W, Varshney RK, Graner A (2003). Exploiting EST databases for the development and characterization of gene-derived SSR markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 106: 411–422.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S (2017). GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45: 6–11.

Wang W, Schalamun M, Morales-Suarez A, Kainer D, Schwessinger B, Lanfear R (2018). Assembly of chloroplast genomes with long- and short-read data: A comparison of approaches using *Eucalyptus pauciflora* as a test case. *BMC Genomics* 19: 977.

Wang Y, Zhan DF, Jia X (2016). Complete chloroplast genome sequence of *Aquilaria sinensis* (Lour.) Gilg and evolution analysis within the Malvales order. *Front. Plant Sci.* 7: 280.

Wick R (2018). Filtlong: A tool for filtering long reads. GitHub repository. Available at: GitHub repository. https://github.com/rrwick/Filtlong (Accessed Oct 19, 2023).

Xiao-Ming Z, Junrui W, Li F, Sha L, Hongbo P, Lan Q, Jing L, Yan S, Weihua Q, Lifang Z, Yunlian C, Qingwen Y (2017). Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. *Sci. Rep*. 7: 1555.

Zhang YH, Huang Y, Li ZM, Zhang SD (2019). Characterization of the complete chloroplast genome of the vulnerable agarwood tree, *Aquilaria yunnanensis* (Thymelaeaceae). *Conserv. Genet. Resour.* 11: 161–164.