

COMPARATIVE PLASTOME STRUCTURE REVEALS TAXONOMIC INCONSISTENCIES IN THE CURRENT SECTIONAL DELIMITATIONS IN *RUBROSHOREA*

Liu J^{1,2}, Ogary KM^{3,4}, Amos K^{1,2}, Malcom D⁵ & Wong SY^{3,*}

¹Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China

²Germplasm Bank of Wild Species & Yunnan Key Laboratory of Crop Wild Relatives Omics, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China

³Institute of Biodiversity and Environmental Conservation, Universiti Malaysia Sarawak, 94300 Samarahan, Sarawak, Malaysia

⁴Sarawak Forestry Corporation, Lot 218, KCLD, Jalan Tapang, Kota Sentosa, 93250 Kuching, Sarawak, Malaysia

⁵Forest Department Sarawak, Level 15, East Wing, Bangunan Baitul Makmur II, Medan Raya, Petra Jaya, 93050 Kuching, Sarawak, Malaysia

*sywong@unimas.my

Submitted December 2025; accepted February 2026

Rubroshorea (Meijer) P.S.Ashton & J.Heck. is a genus comprising 71 species with Borneo recognised as the center of species diversity for this genus, hosting 64 species, including 44 endemic species. In Borneo, several species of *Rubroshorea* and a species of *Shorea* are known for their non-timber value, via the nut oils, and are collectively locally known as Engkabang species. However, only a few Engkabang species have been included in previous phylogenetic studies, leaving their taxonomic placement at both generic and intrageneric levels insufficiently resolved understood. In this study, we conducted sequencing and assembly of the plastid genomes (plastomes) of 27 *Rubroshorea* accessions (representing 18 Bornean species) and nine *Shorea* accessions (representing seven species). The *Rubroshorea* plastid genome exhibits typical quadripartite structure and encodes 110–112 unique genes, comprising 77–78 protein-coding genes (PCGs), 29–30 tRNAs, and 4 rRNAs. Sliding-window analyses identified five highly variable regions, including *rps16-trnQ-UUG*, *ClpP1*, *rps15-ycf1*, *rpl32-trnL-UAG*, and *trnG-GCC-trnM-CAU* within *Rubroshorea* and a species of *Shorea*, representing promising molecular markers for future taxonomic and population genetic studies. Our phylogenomic analysis established a well-resolved and strongly supported phylogenetic framework for *Rubroshorea* but did not support the currently recognized sectional classification. By expanding plastid genomic data for Bornean endemics, this study provides new insights into the evolutionary relationships and taxonomy of *Rubroshorea* taxonomic clarity with practical conservation, securing the future of a genus integral to both biodiversity and human economies.

Keywords: Borneo, Dipterocarpaceae, endemics, *Shorea*

INTRODUCTION

The commercial use of Engkabang oil (known as Tengkawang in Indonesian Kalimantan) dates to the 1800s, when its high-quality fats were utilised for candle production and as lubricants in steam machinery and tram cars. Beyond such industrial applications, Engkabang oil serves as a Cocoa Butter Equivalent (CBE), offering a substitute for cocoa butter in food and cosmetics. Harvested from the nuts of *Shorea* species (colloquially termed Borneo Illipe nuts; Wong & Ogary 2026), Engkabang oils are primarily sourced from species currently placed in the genus *Rubroshorea* (Meijer) P.S.Ashton

& J.Heck (Ashton & Heckenbauer 2022). This revised genus framework, comprising 71 species with a center of diversity in Borneo, includes key oil-producing taxa such as *R. macrophylla* (de Vriese) P.S.Ashton & J.Heck., *R. mecistopteryx* (Ridl.) P.S.Ashton & J.Heck., *R. pinanga* (Scheff.) P.S.Ashton & J.Heck., *R. stenoptera* (Burck) P.S.Ashton & J.Heck., and *R. splendida* (de Vriese) P.S.Ashton & J.Heck. Among these, *R. stenoptera* is favoured for its large fruit size, while *R. macrophylla* and *R. pinanga* are widely harvested as a result of their broad distribution (Fambayun & Hut 2014,

Yulita 2016, Darmawan et al. 2020). Additional species from *Shorea* s.s.— including *S. seminis* (de Vriese) Slooten, *S. sumatrana* Sym. ex Desch, and the lesser-known *S. havilandii* Brandis — are also tapped for nut collection (Ashton 1982, Saridan et al. 2013). Both *Rubroshorea* and *Shorea* belong to the tribe Shoreeae together with *Parashorea*, *Hopea*, and the balance of the recently resurrected or new genera (Ashton & Heckenhauer 2022). Despite taxonomic reclassification, the fatty acid profiles of Engkabang species remain strikingly similar across genera (Anderson 1975), highlighting their shared biochemical utility.

Recent genomic efforts have focused on establishing the genetic architecture of *Rubroshorea*. Whole genome sequences for *R. leprosula* (Miq.) P.S.Ashton & J.Heck. (Ng et al. 2021) and *R. macrophylla* (Chung et al. 2025) have been published, alongside studies exploring phylogenomic relationships (Heckenhauer et al. 2017, 2018, 2019, Chew et al. 2023). However, taxonomic ambiguities persist: Ashton & Heckenhauer (2022) noted overlapping morphological traits among the sections within the genus *Rubroshorea*, complicating clear intersectional delimitation. This underscores the need for wider taxa sampling for plastid genome analysis to resolve phylogenetic uncertainties. Investigating the phylogeny of Engkabang species is critical to resolving taxonomic uncertainties and to safeguard Borneo's ecologically and economically vital forests (Ashton 1982). Engkabang underpin livelihoods and ecosystems, yet their evolutionary relationships remain poorly resolved owing to overlapping morphological traits and limited genomic data. Clarifying these relationships through plastid genome analysis will aid in identifying distinct lineages, ensuring accurate species delineation for sustainable harvesting and conservation.

The plastid genome, essential for plant growth, offers unique advantages for phylogenetic studies due to its uniparental inheritance, conserved structure, and moderate evolutionary rate (Gitzendanner et al. 2018). Recent advancements in sequencing technologies have enabled rapid assembly of plastid genomes, driving progress in reconstructing the evolutionary history of Dipterocarpaceae (Heckenhauer et al. 2017, 2019). While plastomes of *Rubroshorea*, *Shorea*, and *Richetia* have expanded our understanding of the family's phylogeny, Bornean *Rubroshorea* species remain

underrepresented (Heckenhauer et al. 2017, 2019, Chew et al. 2023). This gap limits insights into plastome evolution and phylogeographic patterns within the tribe Shoreeae.

It is crucial to conduct comparative analyses with more *Rubroshorea* plastid genomes to better identify the evolutionary pattern of the plastid genome within *Rubroshorea* and assess phylogenetic relationships. In this study, we assembled the complete plastomes for 27 *Rubroshorea* accessions (representing 19 species, including 15 Bornean endemics) and nine *Shorea* accessions (representing seven species, including *Shorea seminis*). Our objectives were to (1) examine the variation in plastome structure and features of *Rubroshorea* and *Shorea*, focusing on the Engkabang species, (2) identify the hypervariable loci across these *Rubroshorea* and *Shorea* plastomes, and (3) infer the phylogenomic relationships of the Engkabang species. This study will contribute to a better understanding of the evolution of Engkabang species based on plastome genomes.

MATERIALS AND METHODS

Taxon sampling, DNA extraction and genome sequencing

In this study, we collected a total of 14 accessions (representing seven Engkabang species of Bornean *Rubroshorea* section Pachycarpae (F.Heim) P.S.Ashton & J.Heck.), three accessions (representing one Engkabang species of *Shorea*, *S. seminis*), and one *Hopea* sp. from various localities in Sarawak, Malaysian Borneo. Voucher specimens were formally identified by the last author and deposited at Sarawak Forestry Herbarium (SAR). Silica- gel dried leaves were used for DNA extraction by the cetyltrimethylammonium bromide (CTAB) method. Paired-end (PE) sequencing libraries were built using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs, Ipswich, Massachusetts, USA) according to the manufacturer's instructions. Sequencing was done on an Illumina DNBSEQ-T7 (Illumina, San Diego, CA, United States) platform. Additionally, raw sequencing data (FASTQ files) for 25 samples were downloaded from GenBank (accessed on December 8, 2024), and complete plastid genomes were subsequently assembled following the same assembly pipeline used in this study, which belong to 13 *Rubroshorea*

Table S1 Detailed information for species and accessions included in this study

Species	Section	Locality and date	Collectors	GenBank accession no.
<i>Hopea dryobalanoides</i> SRR7067698	–	–	–	SAMN08094297
<i>Hopea</i> sp. DIP-263	–	Malaysia. Sarawak, Kapit, Taman Rekreasi Seabai, 1°56'45.6"N 112°54'16.8"E, 50 m asl. 28 Dec. 2021	Wong S. Y. & P. C. Boyce	PX508760
<i>Parashorea macrophylla</i> SRR7067730	–	–	–	SAMN08094316
<i>Richetia faguetiana</i> SRR7067717	–	–	–	SAMN08094349
<i>Rubroshorea acuta</i> SRR7067698	Rubroshorea	–	–	–
<i>Rubroshorea albida</i> SRR7067726	Rubella	–	–	SAMN08094297
<i>Rubroshorea beccariana</i> DIP-261	Pachycarpae	Malaysia. Sarawak, Kuching, Padawan, Bengoh Dam lake, hills along the lake, 1°14'54.9"N 110°10'12.5"E, 85 m asl. 8 Dec. 2021	Wong S. Y. & P. C. Boyce	PX508758
<i>Rubroshorea ferruginea</i> SRR7067720	Rubroshorea	–	–	SAMN08094354
<i>Rubroshorea</i> <i>inaequilateralis</i> SRR7067691	Brachyptera	–	–	SAMN08094359
<i>Rubroshorea johorensis</i> SRR7067690	Brachyptera	–	–	SAMN08094361
<i>Rubroshorea leprosula</i> SRR7067689	Rubroshorea	–	–	SAMN08094365
<i>Rubroshorea macrophylla</i> DIP-105	Pachycarpae	Malaysia. Sarawak, Kuching, Bau, Kampung Segong, Soli, Sungai Adis, 1°32'22.8"N 110°08'48.5"E, 65 m asl. Date unknown.	Wong S. Y. & P. C. Boyce	PX508751
<i>Rubroshorea macrophylla</i> DIP-189	Pachycarpae	Malaysia. Sarawak, Serian, Mongkos. Date unknown.	Wong S. Y. & P. C. Boyce	PX508754
<i>Rubroshorea macrophylla</i> DIP-265	Pachycarpae	Malaysia. Sarawak, Sarikei, Maradong, Sungai Mato b, 1.8683883, 111.9193822.2 Feb. 2022	Wong S. Y. & P. C. Boyce	PX508762
<i>Rubroshorea macrophylla</i> DIP-60	Pachycarpae	Malaysia. Sarawak, Kuching, Bau, Kampung Grogo, road to Kampung Apar, 50 m asl. Date unknown.	Wong S. Y. & P. C. Boyce	PX508764
<i>Rubroshorea macroptera</i> ssp. <i>macropterifolia</i> SRR7067687	Rubroshorea	–	–	SAMN08094368
<i>Rubroshorea mecistopteryx</i> DIP-243	Pachycarpae	Malaysia. Sarawak, Kuching, Bau, Krokong, Kpg Kaman. Date unknown.	Wong S. Y. & P. C. Boyce	PX508757
<i>Rubroshorea mecistopteryx</i> SRR7067708	Pachycarpae	–	–	SAMN08094401
<i>Rubroshorea ovalis</i> ssp. <i>sarawakensis</i> SRR7067706	Ovalis	–	–	SAMN08094382
<i>Rubroshorea pachyphylla</i> SRR7067704	Brachyptera	–	–	SAMN08094387
<i>Rubroshorea pinanga</i> DIP-241	Pachycarpae	Malaysia, Sarawak, Kuching, Bau, Kampung Segar, Jalan Kg Ngiramak – Skio, 1°25'23"N 110°14'15"E. Date unknown.	Wong S. Y. & P. C. Boyce	PX508756

Species	Section	Locality and date	Collectors	GenBank accession no.
<i>Rubroshorea quadrinervis</i> SRR7067705	Rubroshorea	–	–	SAMN08094393
<i>Rubroshorea rubra</i> SRR7067710	Rubroshorea	–	–	SAMN08094395
<i>Rubroshorea scaberrima</i> DIP-262	Pachycarpae	Malaysia. Sarawak, Kapit, Taman Rekreasi Seabai, 1°56'45.6"N 112°54'16.8"E, 50 m asl. 28 Dec. 2021	Wong S. Y. & P. C. Boyce	PX508759
<i>Rubroshorea scaberrima</i> DIP-264	Pachycarpae	Malaysia. Sarawak, Kapit, Kapit town, Sungai Seranau, 2°02'00.7"N 112°57'01.8"E, 66 m asl. 29 Dec. 2021	Wong S. Y. & P.C. Boyce	PX508761
<i>Rubroshorea scaberrima</i> DIP-266	Pachycarpae	Malaysia. Sarawak, Sarikei, Sungai Lepong, 1°57'12.9"N 111°30'34.9"E, 55 m asl. 3 Feb. 2022	Wong S. Y. & P. C. Boyce	PX508763
<i>Rubroshorea smithiana</i> SRR7067708	Brachyptera	–	–	SAMN08094401
<i>Rubroshorea splendida</i> DIP-81	Pachycarpae	Malaysia. Sarawak, Kuching, Bau, Kampung Grogo, road to Kampung Apar, 50 m asl. Date unknown.	Wong S. Y. & P. C. Boyce	PX508765
<i>Rubroshorea splendida</i> DIP-82	Pachycarpae	Malaysia. Sarawak. Kuching, Bau, Jalan Opar. Date unknown.	Wong S. Y. & P. C. Boyce	PX508766
<i>Rubroshorea stenoptera</i> DIP-181	Pachycarpae	Malaysia. Sarawak, Kuching, Bau, Kampung Grogo, road to Kampung Apar, 50 m asl. Date unknown.	Wong S. Y. & P. C. Boyce	PX508752
<i>Rubroshorea stenoptera</i> DIP-190	Pachycarpae	Malaysia. Sarawak, Serian, Mongkos. Date unknown.	Wong S. Y. & P. C. Boyce	PX508755
<i>Shorea atrinervosa</i> SRR7067712	–	–	–	SAMN08094334
<i>Shorea biawak</i> SRR7067713	–	–	–	SAMN08094340
<i>Shorea crassa</i> SRR7067714	–	–	–	
<i>Shorea exelliptica</i> SRR7067716	–	–	–	SAMN08094343
<i>Shorea oblongifolia</i> SRR70676	–	–	–	SRR70676.1
<i>Shorea seminis</i> DIP-185	–	Malaysia. Sarawak, Kuching, Bau, Krokong. Date unknown.	Wong S. Y. & P. C. Boyce	PX508753
<i>Shorea seminis</i> DIP-93	–	Malaysia. Sarawak, Sri Aman, Tempat Rekreasi Sungai Raya, Sungai Raya, 1°06'47.7"N 111°31'02.3"E, 86 m asl. Date unknown.	Wong S. Y. & P. C. Boyce	PX508767
<i>Shorea seminis</i> DIP-94	–	Malaysia. Sarawak, Kuching, Bau, Kampung Segong, Soli, Sungai Adis, 1°32'22.8"N 110°08'48.5"E, 65 m asl. Date unknown.	Wong S. Y. & P. C. Boyce	PX508768
<i>Dryobalanops lanceolata</i> SRR7067701	–	–	–	SAMN08094284
<i>Vatica mangachapoi</i>	–	–	–	MH716496

species, six *Shorea* species, one *Richetia* species (*Richetia faguetiana* (F.Heim) P.S.Ashton & J.Heck.), one *Parashorea* species (*Parashorea macrophylla* Wyatt-Sm. ex P.S.Ashton), one *Hopea* species (*Hopea dryobalanoides* Miq.), and two outgroups, *Dryobalanops lanceolata* Burck and *Vatica mangachapoi* Blanco for phylogenetic inferences. The detailed sampling information is summarised in Table S1.

Data assembly and annotation

The GetOrganelle v1.7.5.3 toolkit (Jin et al. 2020) was used for *de novo* assembly of the newly sequenced plastomes, employing a range of k-mer values (21, 45, 65, 85, 105, and 127). The complete plastome of *Rubroshorea rubra* (G.H.S.Wood ex P.S.Ashton) P.S.Ashton & J.Heck. (accession SRR7067710) was used as a reference for both assembly and annotation of *Rubroshorea* individuals. The resulting assemblies were visualized and filtered using Bandage v0.8 (Wick et al. 2015) to obtain complete circular plastome structures. Inverted repeat (IR) regions were identified using the “Find Repeats” function in Geneious v8.1 (Kearse et al. 2012), which automatically detects and flanks the IRs. Manual curation was subsequently performed to validate gene boundaries, including start and stop codons. The presence and structure of tRNA genes were further confirmed using tRNAscan-SE v1.21 (Schattner et al. 2005) with default settings.

Plastome structural variation

IR boundary shift and plastome gene content

Six newly sequenced representative plastomes of *Rubroshorea* species were aligned and compared using Geneious v8.1 to investigate structural variation in the IR regions. The positions of genes at the four junctions between the IR and single-copy (SC) regions—JLB (LSC/IRB), JSB (IRB/SSC), JSA (SSC/IRA), and JLA (IRA/LSC)—were examined to assess IR-SC boundary variation and visualized using IRscope (Amiryousefi et al. 2018). To detect potential gene inversions, the six *Rubroshorea* plastomes and a reference plastome (*Rubroshorea rubra*, SRR7067710) were aligned using the progressive Mauve algorithm (Wang et al. 2017) with default parameters, including automatic seed weight calculation (15) and a minimum Locally Collinear Block (LCB) score of 30,000. Additionally, we analysed structural

features of the 18 newly sequenced *Rubroshorea*, *Shorea*, and *Hopea* plastomes, including overall genome size, genomic components (LSC, SSC, and IR regions), and GC content.

Repeat analyses

The repeat analyses were performed using three approaches: (1) REPuter and the parameters for the dispersed repeat searches was set at a minimum size of 30 bp, Hamming distance of 3, and sequence identity of $\geq 90\%$, (2) Tandem Repeats Finder (TRF v 4.09; Benson 1999) web-interface was used with default settings, and (3) simple sequence repeats (SSRs) were detected using MISA (Beier et al. 2017) with preset parameters followed by the core PERL function MISA.pl. The parameters were set to 10 repeat units (≥ 10) for mononucleotide SSRs, six repeat units (≥ 6) for dinucleotide, five repeat units (≥ 5) for trinucleotide, four repeat units (≥ 4) for tetranucleotide, and three repeat units (≥ 3) for pentanucleotide and hexanucleotide.

Sequence divergence analyses

Comparison of the six representative plastomes of *Rubroshorea* and *Shorea* species was performed using LAGAN mode in mVISTA (Brudno et al. 2003, Frazer et al. 2004) with *Rubroshorea rubra* (accession SRR7067710) as a reference genome. The sequence divergent sites were identified using complete plastomes (CP) aligned in Geneious. The analysis included the CDs (protein-coding genes [cpPCGs], tRNAs, and rRNAs) and noncoding (NCDs) regions. The individual gene alignment was performed in Geneious, and gaps were treated as “missing data” during the analysis. The variable parsimony uninformative sites (PUIS) and parsimony-informative sites (PIS) were estimated in PAUP*4.0a (Swofford 2004).

Phylogenetic analyses

Maximum Likelihood (ML) and Bayesian Inference (BI) were used to reconstruct phylogenetic relationships of the Engkabang species based on 37 ingroup individuals of tribe Shoreeae and two outgroup taxa. ML analysis was performed in RAxML v.8.2.12 (Stamatakis 2014) based on the best sequence substitution model, determined by Akaike information criterion (AIC; Akaike 1973) in jModelTest2 v.2.1.7 (Darriba et al. 2012), with 1,000 bootstrap

replicates. The BI was performed using MrBayes (Ronquist & Huelsenbeck 2003) with Markov Chain Monte Carlo (MCMC) analysed for 10,000,000 generations, and the tree sampling frequency was set at every 1000th generations. The Bayesian posterior probability (PP) was calculated with two MCMC chains. The MCMC convergence was determined based on effective sample size ($ESS \geq 200$) for each parameter per run in TRACER (Rambaut & Drummond 2007). The first 25% were discarded as burn-in, and the PP was estimated. The resulting trees (ML and BI) were visualized and edited in Figtree v.1.4.2 (Rambaut 2014) and the web server of Interactive tree of life (iTOL, v.4) (Letunic & Bork 2019).

RESULTS

Structural features and genome composition of plastid genomes

The complete plastid genomes exhibited the typical quadripartite circular structure, consisting of a large single-copy (LSC) region, a small single-copy (SSC) region, and two inverted repeat (IR)

regions (Figure 1). The length of the genomes ranged from 150,394 bp (*Rubroshorea splendida*, DIP-81) to 152,552 bp (*Shorea seminis*, DIP-94), while the LSC region length ranged from 83,424 bp (*Hopea* sp.) to 84,245 bp (*Shorea seminis*, DIP-94). The shortest SSC region was 19,661 bp, found in *Rubroshorea splendida* (DIP-81), while the longest SSC region was 20,211 bp in *Shorea seminis* (DIP-94). The IR region length ranged from 23,621 bp (*Hopea* sp.) to 24,048 bp (*Shorea seminis*, DIP-94). The GC content of the plastid genome is 37% (Table 1).

The annotated *Rubroshorea* and *Shorea* plastid genomes encoded 110–112 unique genes, comprising 77–78 protein-coding genes (PCGs), 29–30 tRNAs, and 4 rRNAs (Table 1). A total of 94 genes were present as single copies, including *atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI*, *ccsA*, *cemA*, *clpP1*, *matK*, *ndhA*, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*, *pafI*, *pafII*, *pbfI*, *petA*, *petB*, *petD*, *petG*, *petL*, *petN*, *psaA*, *psaB*, and *psaC*. Nineteen genes were duplicated within the inverted repeat (IR) regions, including *ndhB*, *rpl2*, *rpl23*, *rps7*, *rps12*, all four rRNA genes (*rrn4.5*, *rrn5*, *rrn16*, *rrn23*), seven tRNA genes (*trnA-UGC*, *trnI-CAU*,

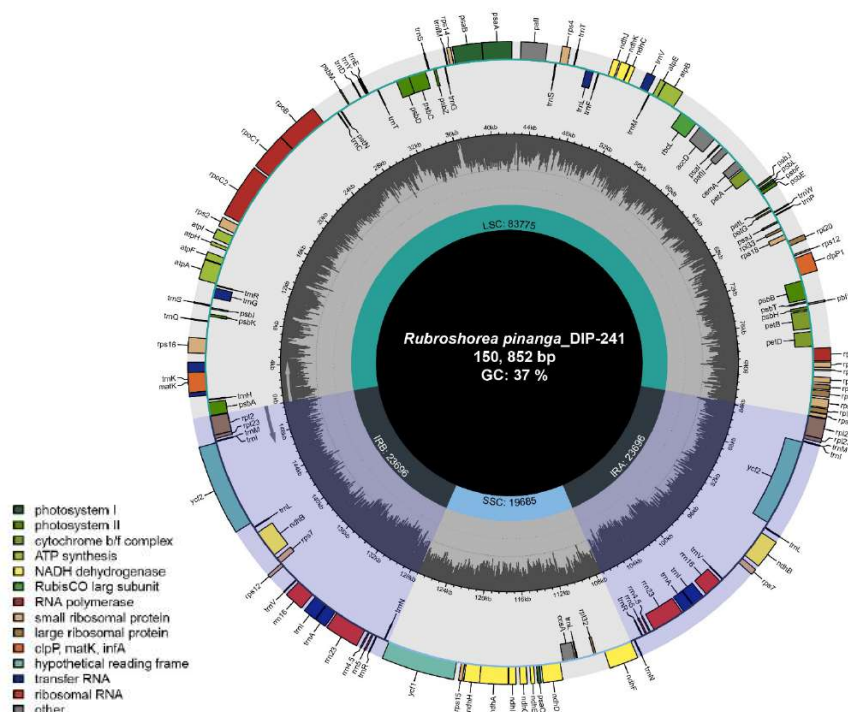


Figure 1 Structural map of the typical complete plastome of a representative *Rubroshorea* species (*Rubroshorea pinanga*_DIP-241). Genes inside the circle are transcribed clockwise, and genes outside are transcribed counterclockwise. The functional groups are colour-coded. The plastome exhibits a typical quadripartite structure with conserved gene order.

Table 1 Comparative plastome features of *Rubroshorea* and *Shorea* species. Summary statistics include total plastome size (bp), GC content (%), and lengths of the large single-copy (LSC), small single-copy (SSC), protein-coding gene (PCG), and inverted repeat (IR) regions

Species	Genome size (bp)	LSC (bp)	SSC (bp)	IR (bp)	GC content (%)	PCG	rRNAs	tRNAs
<i>Rubroshorea beccariana</i> DIP-261	150686	83639	19806	23622	37	78	4	30
<i>Rubroshorea macrophylla</i> DIP-60	150599	83547	19801	23625	37	77	4	29
<i>Rubroshorea macrophylla</i> DIP-105	150720	83643	19817	23630	37	77	4	29
<i>Rubroshorea macrophylla</i> DIP-189	150702	83633	19807	23631	37	77	4	29
<i>Rubroshorea macrophylla</i> DIP-265	150563	83499	19802	23631	37	77	4	29
<i>Rubroshorea meciscopteryx</i> DIP-243	150755	83670	19820	23632	37	78	4	30
<i>Rubroshorea pinanga</i> DIP-241	150852	83775	19685	23696	37	78	4	30
<i>Rubroshorea scaberrima</i> DIP-262	150772	83655	19815	23626	37	77	4	29
<i>Rubroshorea scaberrima</i> DIP-264	150590	83510	19818	23631	37	77	4	29
<i>Rubroshorea scaberrima</i> DIP-266	150712	83645	19805	23631	37	77	4	29
<i>Rubroshorea splendida</i> DIP-81	150394	83475	19661	23630	37	78	4	30
<i>Rubroshorea splendida</i> DIP-82	150569	83522	19803	23622	37	77	4	29
<i>Rubroshorea stenoptera</i> DIP-181	150612	83566	19797	23624	37	77	4	29
<i>Rubroshorea stenoptera</i> DIP-190	150720	83643	19817	23630	37	77	4	29
<i>Shorea seminis</i> DIP-185	152436	84234	20146	24028	37	77	4	29
<i>Shorea seminis</i> DIP-93	152436	84234	20146	24028	37	77	4	29
<i>Shorea seminis</i> DIP-94	152552	84245	20211	24048	37	77	4	29
<i>Hopea</i> sp. DIP-263	150488	83424	19825	23621	37	77	4	29

trnI-GAU, *trnL*-CAA, *trnN*-GUU, *trnR*-ACG, *trnV*-GAC), and *ycf2* (Table S2).

The IR boundaries of the *Rubroshorea* and *Shorea* plastomes were largely conserved, with only minor variations in gene positioning and overlap lengths, and no evidence of major IR expansion or contraction events (Figures 2 and 3). At the LSC/IRb junction (JLB), the *rps19* gene extended four bp into LSC in all species, while *rpl2* was located in IRb, with LSC lengths ranging from 83,190 to 84,234 bp. At the IRb/SSC junction (JSB), *ndhF* in the IRb overlapped by 22–65 bp, while *ycf1* extended from the SSC into IRb by 2,235–2,259 bp. At the SSC/IRa junction (JSA), *ycf1* overlapped into SSC by 64–251 bp, followed by *trnN* within IRa. At the IRa/LSC junction (JLA), *rpl2* was found in IRa with *psbA* in the LSC extended by one bp in all species.

Repeat sequence analysis

The plastomes of six representative *Rubroshorea* and *Shorea* species harboured 91 to 110 SSRs, 110 to 120 long repeats, and 63 to 105 tandem repeats (Figure 4 and Table 2). Mononucleotide repeats

were the most common type, ranging from 58 to 79. Notably, no hexanucleotide repeats were observed in any of the plastomes (Figure 4A and 4D; Table 2). The mononucleotide repeats were the most abundant, with notable variation in their abundance among the species examined (Figure 4A)

All forward (F), reverse (R), Complement (C), and palindrome (P) repeat types were identified among the long repeats, with forward repeats being the dominant type (70 occurrences in *Shorea seminis*) (Figure 4B). Tandem repeats were detected 63 times in *Rubroshorea meciscopteryx* to 105 times in *Shorea seminis*.

Sequence divergence analysis

By using the annotation of *Rubroshorea rubra* as a reference, the mVISTA program was used to align the chloroplast genomes and visualise the pattern of sequence identity among closely related species with *Rubroshorea beccariana*, *Rubroshorea macrophylla*, *Rubroshorea meciscopteryx*, *Rubroshorea pinanga*, *Rubroshorea splendida*, and *Shorea seminis* (Figure 5). Overall, coding regions were more conserved than non-coding regions, with most

Table S2 Annotated gene information of the plastomes of *Rubroshorea* and *Shorea* species

Gene Function Classification	Group of Genes	Gene Names (CDS/gene entries consolidated)
Self-replication	Ribosomal RNA genes	<i>rrn4.5, rrn5, rrn16, rrn23</i>
	Transfer RNA genes	<i>trnA-UGC, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnJ-M-CAU, trnG-GCC, trnG-UCC, trnH-GUG, trnI-CAU, trnI-GAU, trnK-UUU, trnL-CAA, trnL-UAA, trnL-UAG, trnM-CAU, trnN-GUU, trnP-UGG, trnQ-UUG, trnR-ACG, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC, trnV-UAC, trnW-CCA, trnY-GUA</i>
	RNA polymerase	<i>rpoA, rpoB, rpoC1, rpoC2</i>
	Small ribosomal subunits	<i>rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps15, rps16, rps18, rps19</i>
	Large ribosomal subunits	<i>rpl2, rpl14, rpl16, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36</i>
Genes for photosynthesis	Photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>
	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbT, psbZ</i>
	RubisCO large subunit	<i>rbcL</i>
	Cytochrome *b/f* complex	<i>petA, petB, petD, petG, petL, petN</i>
	ATP synthase subunits	<i>atpA, atpB, atpE, atpF, atpH, atpI</i>
Other genes	NADH dehydrogenase	<i>ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	Protease	<i>clpP1</i>
	Maturase	<i>matK</i>
	Envelope membrane protein	<i>cemA</i>
	Acetyl-CoA carboxylase	<i>accD</i>
	C-type cytochrome synthesis	<i>ccsA</i>
	Other functions	<i>ycf1, ycf2, pafI, pafII, pbf1</i>

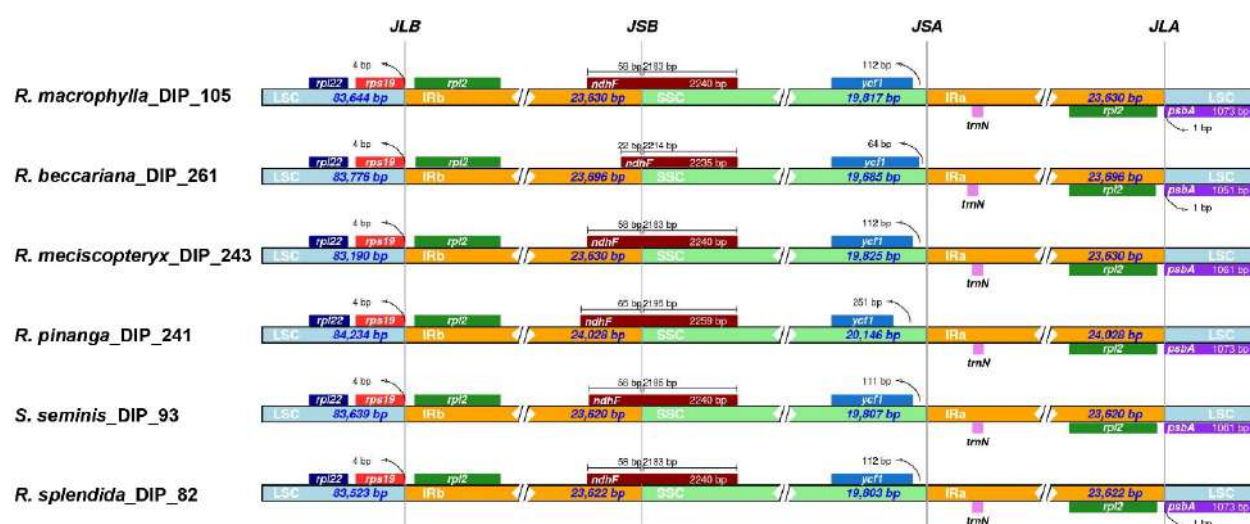


Figure 2 Comparison of LSC, SSC, and IR borders among the six representative plastomes of *Rubroshorea* and *Shorea* showing the inter- and intra-specific variation at the borders. The genes around the edges are shown above or below the main line. Abbreviations: LSC- Large single copy, SSC- Small single copy, and IR (a & b) – Inverted repeat a &b. The genes above the line transcribe on a negative strand, whereas the gene below transcribes on a positive strand

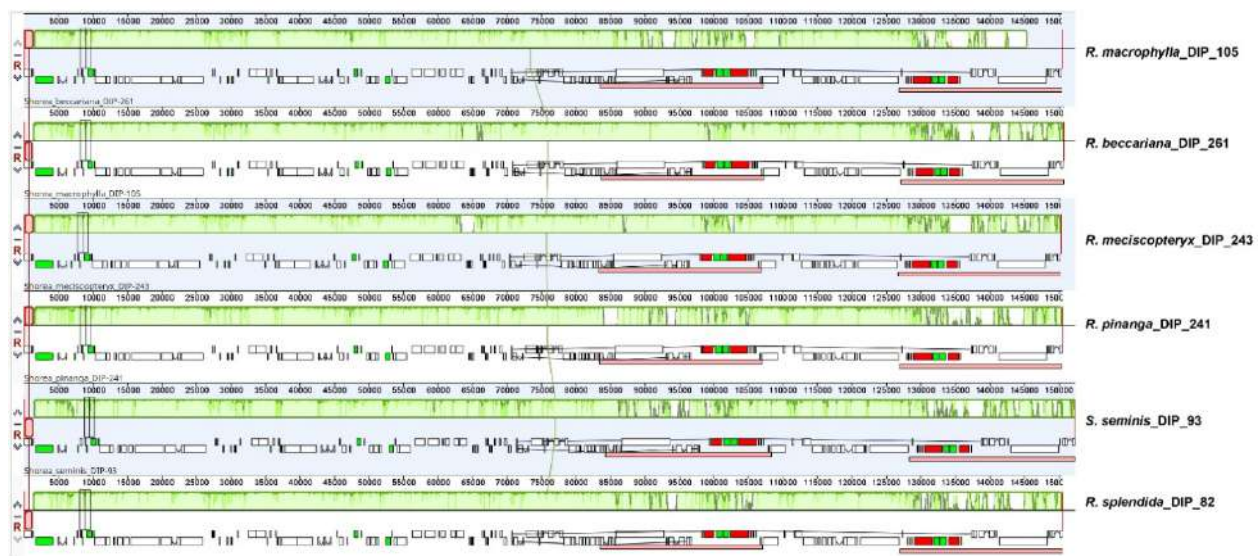


Figure 3 Comparative alignment among the six representative plastomes of *Rubroshorea* and *Shorea* plastomes using Mauve. The green colored blocks indicate locally collinear regions (LCBs), with lines connecting homologous sequences across species

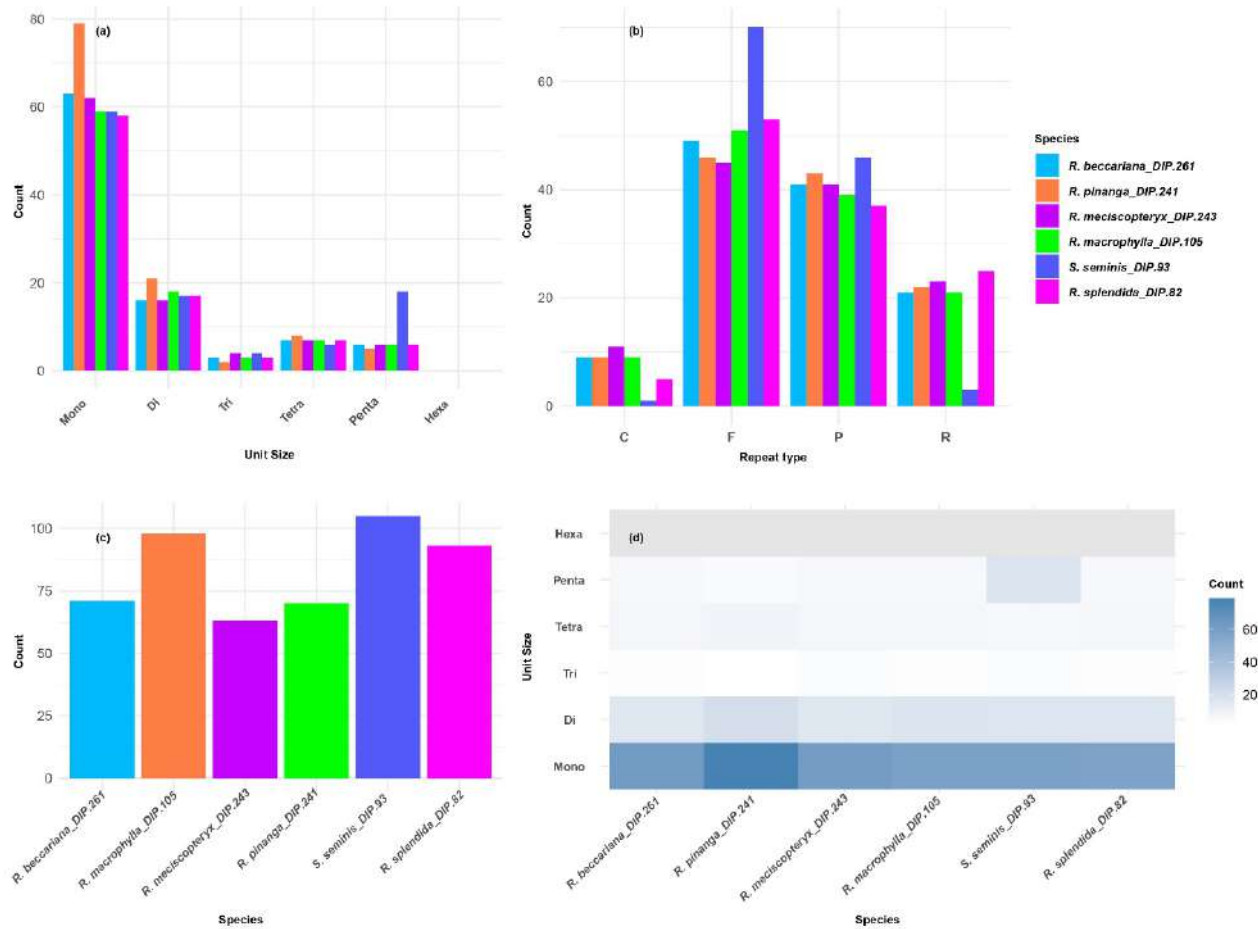
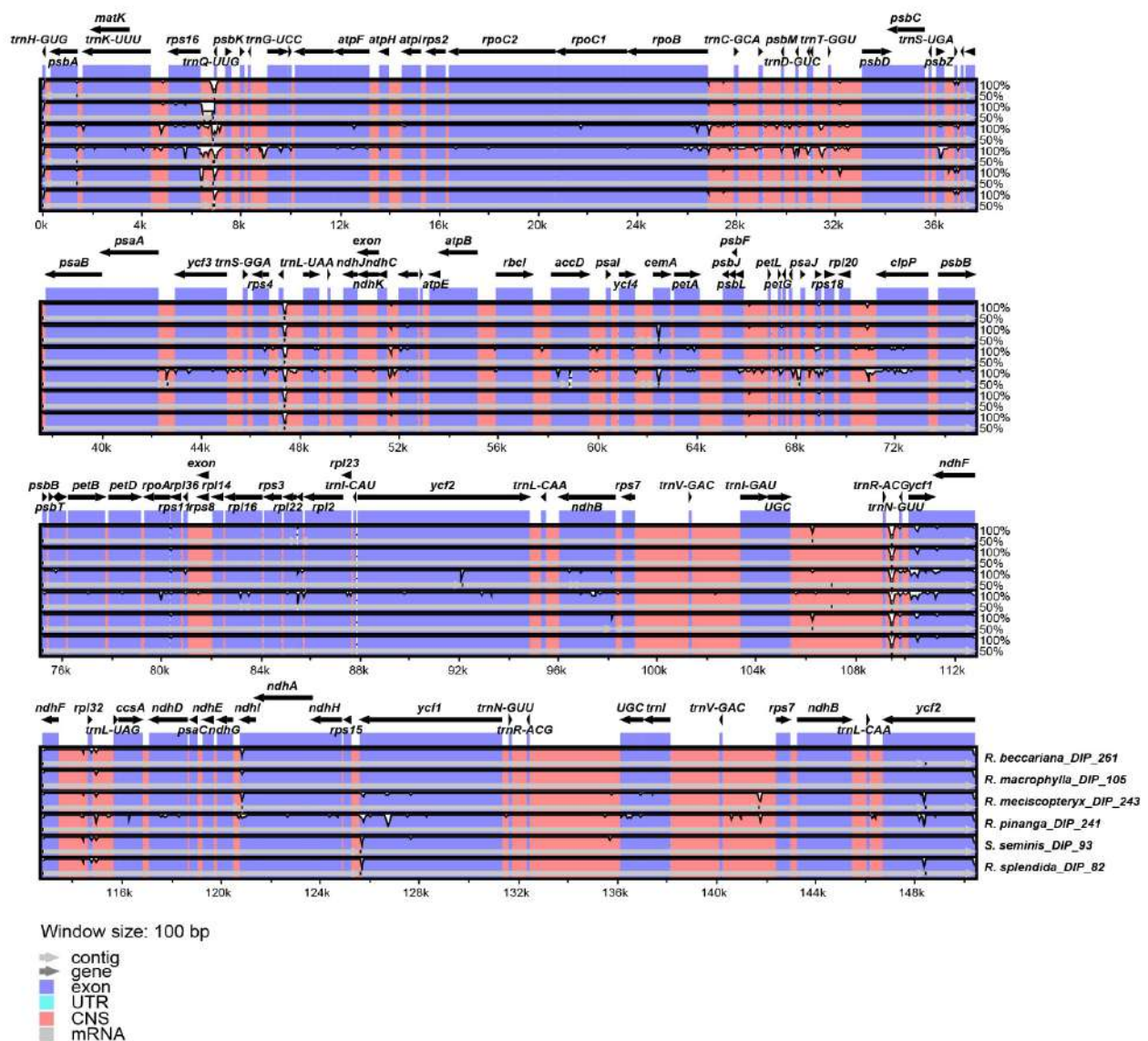


Figure 4 Repeat analysis across six representative plastomes of *Rubroshorea* and *Shorea* species. (a) SSR type distribution (: mono-, di-, tri-, tetra-, penta- and hexanucleotide motifs). (b) Frequency of four dispersed repeat types—Forward (F), Reverse (R), Complement (C), and Palindromic (P). (c) Tandem repeats across the six representative species of *Rubroshorea* and *Shorea*. (d) Heatmap of SSR counts of the six representative species of *Rubroshorea* and *Shorea*

Table 2 The distribution of repeat sequences among the six representative plastomes of *Rubroshorea* and *Shorea* species including simple sequence repeats (SSRs), long repeats, and tandem repeats

Species	SSR						Long repeats				Tandem repeats
	Mono	Di	Tri	Tetra	Penta	Hexa	C	P	R	F	
<i>Rubroshorea beccariana</i> DIP-261	63	16	3	7	6	-	9	41	21	49	71
<i>Rubroshorea pinanga</i> DIP-241	79	21	2	8	5	-	9	43	22	46	70
<i>Rubroshorea macrophylla</i> DIP-105	59	18	3	7	6	-	9	39	21	51	98
<i>Rubroshorea meciscopteryx</i> DIP-243	62	16	4	7	6	-	11	41	23	45	63
<i>Rubroshorea splendida</i> DIP-82	58	17	3	7	6	-	5	37	25	53	93
<i>Shorea seminis</i> DIP-93	59	17	4	6	18	-	1	46	3	70	105

**Figure 5** The visual comparison of the six representative plastomes of *Rubroshorea* and *Shorea* species using mVISTA with the *Rubroshorea rubra*_SRR7067710 as the reference genome. The y-axis represents the similarity identity within 50-100%. The arrows indicate the direction of gene transcription, with the blue representing coding regions and the pink representing non-coding regions

variations were detected in non-coding sequences (Figure 5). The exon and UTR sequences were nearly identical across the six taxa, reflecting their conserved structure and gene order.

Sliding window analysis revealed the nucleotide polymorphisms in *Rubroshorea* and *Shorea* plastid genomes with variation distribution across the LSC, IR and SC region. Similarly, the non-coding regions exhibited more pronounced variation than coding regions (Figure 6). The nucleotide polymorphism (π) ranged from 0.01773 to 0.04211. Relatively high π values (> 0.011) were observed in *rps16-trnQ-UUG*, *ClpP1*, *rps15-ycf1*, *rpl32-trnL-UAG*, and *trnG-GCC-trnM-CAU*. *rps16-trnQ-UUG* had the highest percentage of variable and parsimony informative sites (Table 3).

Phylogenetic analysis

The phylogenetic trees constructed using ML and BI methods displayed highly congruent topologies, with robust bootstrap support and high posterior probabilities for most nodes (Figure 7). The findings revealed that tribe Shoreeae could be delineated into two well-supported clades, arranged from the first diverging lineage to the terminal positions in the phylogenetic tree, as follows: (1) a clade comprising *Shorea* species, Clade II; and (2) a clade comprising *Rubroshorea* species, Clade I. Three accessions of *Shorea seminis* were robustly supported (BS = 100, PP = 1.00) and well embedded within the *Shorea* clade (BS = 100, PP = 1.00). *Rubroshorea* is separated into four groups:

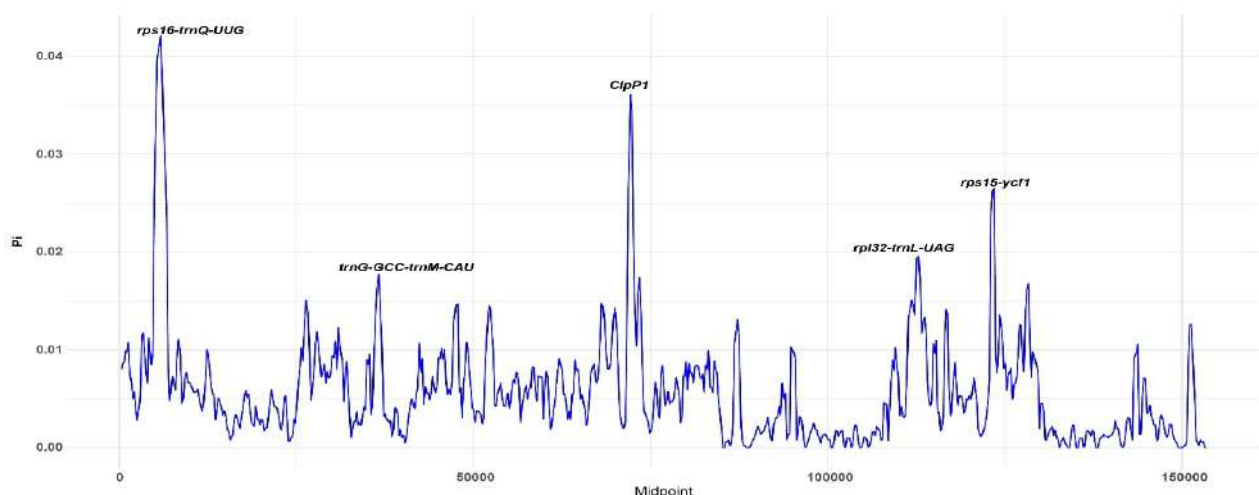


Figure 6 Sliding-window analysis of nucleotide diversity (π) across *Rubroshorea* and *Shorea* plastomes with window length: 600 bp and; step size: 200 bp. Peaks indicate hypervariable regions, while troughs denote conserved regions. The top five most variable regions (labeled) correspond to those ranked in Table 3 (highest to lowest π). X-axis = plastome midpoint positions; Y-axis = π values

Table 3 The sequence variation in the top five hypervariable regions of *Rubroshorea* and *Shorea* plastomes including nucleotide position, nucleotide diversity, number of total characters, PUIS (Parsimony uninformative sites), PIS (Parsimony informative sites) and the total number of variable sites

Nucleotide position	Nucleotide diversity (π)	Region	Total characters	Constant sites	PUIS	PIS	Variables
5889	0.04211	<i>rps16-trnQ-UUG</i>	1955	1747	126	82	208
72174	0.03606	<i>ClpP1</i>	638	562	23	53	76
123468	0.02651	<i>rps15-ycf1</i>	405	364	9	32	41
112569	0.01937	<i>rpl32-trnL-UAG</i>	1381	1298	44	39	83
36677	0.01773	<i>trnG-GCC-trnM-CAU</i>	146	143	1	2	3

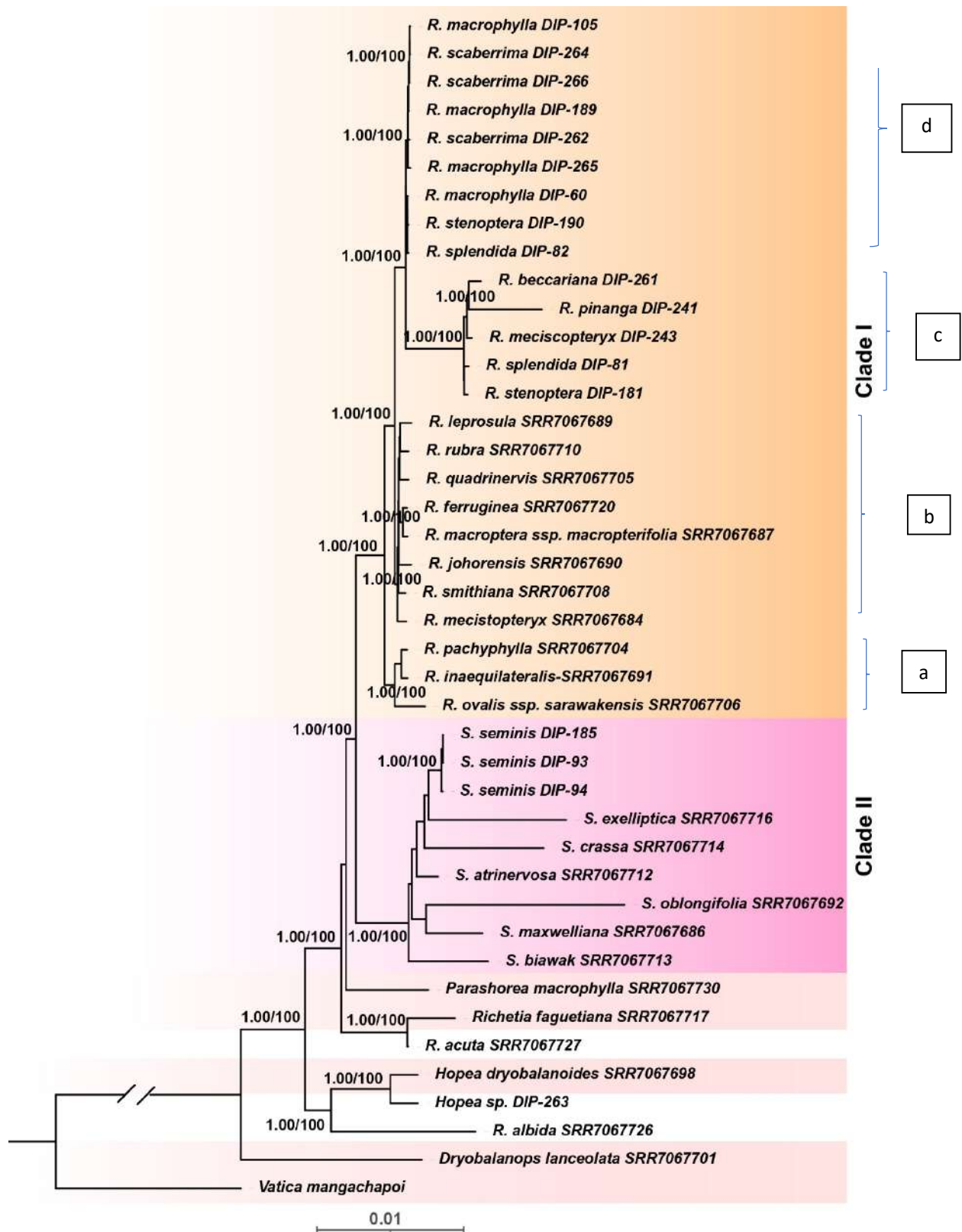


Figure 7 Phylogenetic relationships of *Rubroshorea* and *Shorea* species inferred from whole plastome sequences. The tree was reconstructed using maximum likelihood (ML) and Bayesian inference (BI) based on 36 *Rubroshorea*, *Shorea* individuals, and two outgroup taxa. Support values are indicated at nodes (Bayesian posterior probabilities [PP]; left / ML bootstrap [BS]; right); only values with maximum support are shown. Clade I is highlighted in light orange, Clade II in light pink, and the outgroup taxa in light red. Unhighlighted taxa of *Rubroshorea* represent species that did not group within either major clade

(a) *Rubroshorea* species from section Brachyptera and section Ovalis (BS = 100, PP = 1.00); (b) *Rubroshorea* species from section Rubroshorea, section Pachycarpae, and section Brachyptera (BS = 100, PP = 1.00); (c) *Rubroshorea* species from section Pachycarpae (BS = 100, PP = 1.00); and (d) *Rubroshorea macrophylla* (from central Sarawak and one accession from Mongkos, West Sarawak) and *Rubroshorea scaberrima* formed a strong clade (BS = 100, PP = 1.00), which together were a sister to *Rubroshorea macrophylla* (Bau, West Sarawak), *Rubroshorea stenoptera* (Mongkos, West Sarawak) and *Rubroshorea splendida* (Opar, West Sarawak).

DISCUSSION

Plastome characteristics and comparison

The plastomes of 18 species were successfully sequenced and assembled, revealing a conserved quadripartite structure (Figure 1), consistent with previously published plastomes of tribe Shoreeae (Heckenhauer et al. 2019, Chew et al. 2023). The size of the plastid genome ranges from 150,394 bp in *Rubroshorea splendida* to 152,552 bp in *Shorea seminis*, which is relatively small. The genome size of *Rubroshorea macrophylla* (ranging from 150,563 bp to 150,720 bp) varies from the previously published genome at 150,778 bp (Chew et al. 2023). There were 110–112 unique genes, comprising 77–78 PCGs, 29–30 tRNAs, and 4 rRNAs. Interestingly, all accessions of *R. macrophylla* are reported to include 77 PCGs and 29 tRNA genes, differing from 78 PCGs and 30 tRNA genes reported by Chew et al. (2023). Of these, 94 genes occurred as single copies, while 19 were duplicated within the inverted repeat (IR) regions. The duplicated genes included PCGs, all four rRNA genes and tRNAs (Table 1). The GC content averages 37% which is consistent with previous studies (Heckenhauer et al. 2019, Chew et al. 2023).

In angiosperms, the contraction and expansion of IRs are ubiquitous phenomena that have been recognized as primary drivers of plastome size variation (Weng et al. 2017). A comparative investigation of JLB, JSB, JSA, and JLA boundaries in the plastomes of six species from the tribe Shoreeae revealed that, although the overall boundary architecture was largely conserved, subtle variations existed (Figure 2). In all six species, the *rpl22* and *rps19* genes are completely covered by LSC regions.

The *rpl2* gene are found fully embedded in IR region for all species. However, Chew et al. (2023) reported that *rpl2* gene of *Rubroshorea macrophylla*, *Rubroshorea pachyphylla* (Ridl. ex Symington) P.S.Ashton & J.Heck., *Doona zeylanica* Thwaites, and *Parashorea macrophylla* Wyatt-Sm. ex P.S.Ashton is expanded into the LSC region. The *ndhF* gene expanded into IR and SSC regions for all species while Chew et al. (2023) reported similar patterns for several species but the *ndhF* gene of *Rubroshorea leprosula* (Miq.) P.S.Ashton & J.Heck., *Rubroshorea pachyphylla*, *Doona zeylanica*, and *Parashorea macrophylla* are found completely in SSC region. The *ycf1* gene of all the observed species are all completely located in the SSC region, however the location of *ycf1* gene to the boundary varies in all the species, ranging from 64 to 252bp, much smaller range than the species reported in Chew et al. (2023).

Repeat sequences and hypervariable regions analyses

Repetitive sequences played a crucial role in species identification, molecular marker development, and phylogenetic studies (Agrama et al. 2003, Doorduyn et al. 2011, Yang et al. 2011). Simple sequence repeats (SSRs), short tandem repeats consisting of 1–6 bp units, were widely dispersed throughout plastomes and served as valuable molecular markers for genetic and phylogenetic analyses (Tong et al. 2022). In this study, between 91 to 110 SSRs were identified in the six species analysed (Figure 4a; Table 2) which is much less than 262 SSRs reported in Chew et al. (2023). Mononucleotide repeats dominated the SSR category. 110 to 120 long repeats and 63 to 105 tandem repeats were detected in all the six species included. The long repeats sequences comprise all four forms. This result differed from Chew et al. (2023) which reported only 43 long repeats for *Rubroshorea macrophylla* as compared to 120 long repeats reported in this study. Forward repeats were the most prevalent, with most measuring 45–70 bp in length. Tandem repeats in *Shorea seminis* were higher than *Rubroshorea* species.

The mVISTA analysis of six Shoreeae species revealed higher levels of sequence divergence in the LSC and SSC regions relative to the IRs, with intergenic regions exhibiting greater variability compared to coding regions (Table 3). Highly variable regions serve as valuable resources for

developing molecular markers and investigating phylogenetic relationships. Such markers are fundamental for species identification and genetic diversity assessment, which ultimately inform evolutionary studies and conservation strategies. In the present study, the nucleotide diversity analysis of coding and intergenic sequences identified five hypervariable sequences: *rps16-trnQ-UUG*, *ClpP1*, *rps15-ycf1*, *rpl32-trnL-UAG*, and *trnG-GCC-trnM-CAU* which differs to the seven mutational hotspots discovered (*ycf2*, *psbI-trnG*, *ndhA-ndhI*, *psbZ-trnG*, *rps12-ndhB*, *rpl33-rpoA*, and *rpoA-psbB*) in *Rubroshorea macrophylla* (Chew et al. 2023). These regions approximate species-level nucleotide substitutions, allowing researchers to investigate and develop molecular markers that are critical for plant taxonomy and identification.

By using the annotation of *Rubroshorea rubra* as a reference, the mVISTA program was used to align the chloroplast genomes and visualise the pattern of sequence identity among the six species included. Similar to Chew et al. (2023), coding regions were more conserved than non-coding regions, with most variations were detected in non-coding sequences. In addition, as shown in other taxa's chloroplast genome research, the LSC and SSC regions diverged more than the IR regions (Lin et al. 2025). In the LSC, the majority of exons are conserved in intergenic spacers with a high amount of divergence (> 50%, as determined by the white space seen across alignments), similar to Chew et al. (2023).

Phylogenetic analyses

Systematics of Dipterocarpaceae has been widely discussed, with two published phylogenomic studies (Heckenhauer et al. 2019, Chew et al. 2022). The topologies of the phylogenetic trees obtained in this study, although with limited sampling, have the primary intention to investigate the Bornean *Rubroshorea* species and are like those revealed in previous studies (Heckenhauer et al., 2017, 2018, 2019) but with stronger support for the main clades of *Shorea* and *Rubroshorea*. The core Bornean *Rubroshorea* species are divided into two geographic regions: central Sarawak and West Sarawak. Plastomes are typically maternally inherited (via seeds in plants) and do not recombine. With seed dispersal limited within geographically isolated populations each will accumulate distinct plastid lineages, leading to geographically

structured clades (Bock 2007). The river basins of Borneo play a crucial role in the ecology and distribution of *Rubroshorea* species. The Dipterocarpaceae are often cited to have seed dispersal primarily by wind, with water acting as a secondary mechanism (Ashton 1982). However, Engkabang seeds are heavy (Krishnapillay & Tompsett 1998) dropping near the parent tree, and if near to flowing water the fruits floating. Dispersal is therefore constrained by river basins, with seeds carried by streams and rivers to new germination sites within, but unable to travel outside, of the basin. The Rejang River basin limits the dispersal of many endemic species, including Engkabang, creating water-mitigated phytogeographic separations. Intersectional relationships in *Rubroshorea* lack resolution due to low genetic variation, necessitating deeper taxonomic work to clarify lineages. Building on the reclassification of *Shorea* into seven genera (Ashton & Heckenhauer 2022), further intra- and inter sectional studies are critical. Notably, *Rubroshorea acuta* (subsection *Auriculatae*) aligns with *Richetia*, while *R. albida* (section *Rubella*) appears as sister to *Hopea*. As in Heckenhauer et al. (2019), *Hopea* and *Parashorea* nest within *Shorea* sensu Ashton & Heck., highlighting unresolved phylogeny. This study underscores the need for expanded research to clarify relationships among species of Shoreaeae.

CONCLUSIONS

This research involved a detailed investigation of the complete plastomes of 36 *Rubroshorea* and *Shorea* species, elucidating their structural characteristics, genetic diversity, and phylogenetic relationships. The plastomes of these species displayed the typical quadripartite structure, with conserved gene composition. The comparative analysis revealed higher variation in intergenic regions than in coding sequences. Notably, five hypervariable regions (*ndhF-rpl32*, *rps16-trnQ-UUG*, *ClpP1*, *rps15-ycf1*, *rpl32-trnL-UAG*, and *trnG-GCC-trnM-CAU*) were identified, providing potential molecular markers for phylogenetic studies. Based on complete plastome sequences, phylogenomic reconstruction indicated that *Shorea* is still not monophyletic. These results not only expand genomic resources for *Rubroshorea*, specifically in species known for their valuable nut oil, but also offer critical insights into species delimitation, evolutionary biology, and

conservation strategies, thereby establishing a solid foundation for future investigations of this taxonomically significant lineage.

Supplementary Materials: The supplementary material for this article can be found in the attached additional files.

ACKNOWLEDGEMENTS

Fieldwork in Sarawak was carried out under Research Permit No. SFC.810-4/6/1(2023)-145 from Sarawak Forestry Corporation and SBC-2021-RDP-38-WSY from Sarawak Biodiversity Centre. This paper forms part of the results obtained from a Ph.D. study by the second author. Funding by the Sarawak Research and Development Council through Special Grant Scheme No. RDCRG02/THM02/2020/_1 is acknowledged. Jie Liu was supported by the CAS Light of West China Program and the Key Research Program of Frontier Sciences, CAS (ZDBS-LY-7001).

REFERENCES

- AGRAMA H & TUINSTRAL M. 2003. Phylogenetic diversity and relationships among Sorghum accessions using SSRs and RAPDs. *African Journal of Biotechnology* 2: 334–340.
- AKAIKE H. 1973. Information theory and an extension of the maximum likelihood principle. In Petrov BN & Csáki F (eds.) Second International Symposium on Information Theory, Akademiai Kiado.
- AMIRYUSEFI A, HYVÖNEN J & POCHAI P. 2018. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34: 3030–3031. <https://doi.org/10.1093/bioinformatics/bty220>
- ANDERSON JAR. 1975. Illipe Nuts (*Shorea* sp.) as potential agricultural crops." Pp. 217–230 in Williams JT et al. (eds.) *South East Asian Plant Genetic Resources*. PROSEA Bogor.
- ASHTON PS. 1982. Dipterocarpaceae. Flora Malesiana series I-Spermatophyta flowering plants 9: 237–552. Mentinun Nijhoff Publications, the Netherlands.
- ASHTON PS & HECKENHAUER J. 2022. Tribe Shoreae (Dipterocarpaceae subfamily Dipterocarpoideae) finally dissected. *Kew Bulletin* 77: 885–903.
- BEIER S, THIEL T, MÜNCH T ET AL. 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33: 2583–2585.
- BENSON G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 27: 573–580. <https://doi.org/10.1093/nar/27.2.573>
- BOCK R. 2007. Structure, function, and inheritance of plastid genomes. Pp. 29–63 in Bock R (ed.) *Cell and molecular biology of plastids*. Springer, London.
- BRUDNO M, MALDE S, POLIAKOV A ET AL. 2003. Global alignment: finding rearrangements during alignment. *Bioinformatics* 19: i54–i62. <https://doi.org/10.1093/bioinformatics/btg1005>
- CHEW IYY, CHUNG HH, LIM LWK ET AL. 2023. Complete chloroplast genome data of *Shorea macrophylla* (Engkabang): structural Features, comparative and phylogenetic analysis. *Data in Brief* 47: 109029.
- CHUNG HH, SOH AAL, LAU MML ET AL. 2025. The first engkabang jantung (*Rubroshorea macrophylla*) genome survey data. *Data in Brief* 58: 111248.
- DARMAWAN MA, MUHAMMAD BZ, HARAHAP AFP ET AL. 2020. Reduction of the acidity and peroxide numbers of Tengawang Butter (*Shorea stenoptera*) using thermal and acid activated bentonites." *Heliyon* 6: e05742.
- DARRIBA D, TABOADA GL, DOALLO R & POSADA D. 2012. jModelTest 2: more models, new heuristics and parallel computing." *Nature Methods* 9: 772.
- DOORDUIN L, GRAVENDEEL B, LAMMERS Y ET AL. 2011. The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. *DNA Research* 18: 93–105.
- FAMBAYUN RA & HUT S. 2014. Budidaya Tengawang untuk kayu pertukangan, bahan makanan dan kerajinan. PT Penerbit IPB Press, Bogor.
- FRAZER KA, PACHTER L, POLIAKOV A ET AL. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Research* 32: W273–W279. <https://doi.org/10.1093/nar/gkh458>
- GITZENDANNER MA, SOLTIS PS, YI TS, LI DZ & SOLTIS DE. 2018. Plastome phylogenetics: 30 Years of inferences into plant evolution. *Advances in Botanical Research* 85: 293–313.
- HECKENHAUER J, PAUN O, CHASE MW ET AL. 2019. Molecular phylogenomics of the Tribe Shoreae (Dipterocarpaceae) using whole plastid genomes. *Annals of Botany* 123: 857–865.
- HECKENHAUER J, SAMUEL R, ASHTON PS ET AL. 2017. Phylogenetic analyses of plastid DNA suggest a different interpretation of morphological evolution than those used as the basis for previous classifications of Dipterocarpaceae (Malvales). *Botanical Journal of the Linnean Society* 185: 1–26.
- HECKENHAUER J, SAMUEL R, ASHTON PS ET AL. 2018. Phylogenomics resolves evolutionary relationships and provides insights into floral evolution in the Tribe Shoreae (Dipterocarpaceae). *Molecular Phylogenetics and Evolution* 127: 1–13.
- JIN JJ, YU WB, YANG WB ET AL. 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>
- KEARSE M, MOIR R, WILSON A ET AL. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- KRISHNAPILLAY B. & TOMPSETT PB. 1998. Seed Handling. Pp. 73–88 in Appanah S & Turnbull JM (eds.) *A Review of Dipterocarps: Taxonomy ecology and silviculture*. Center for International Forestry Research, Bogor.

- LETUNIC I & BORK P. 2019. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* 47: W256–W259. <https://doi.org/10.1093/nar/gkz239>
- LIN Q, HU S, & WU Z ET AL. 2025. Comparative chloroplast genomics provides insights into the phylogenetic relationships and evolutionary history for *Actinidia* species.” *Scientific Reports* 15: 13291.
- NG KS, KOBAYASHI MJ, FAWCETT JA ET AL. 2021. The genome of *Shorea leprosula* (Dipterocarpaceae) highlights the ecological relevance of drought in aseasonal tropical rainforests. *Communications Biology* 4: 1166.
- RAMBAUT A. 2014. FigTree v1.4. Molecular evolution, phylogenetics, and epidemiology. <http://tree.bio.ed.ac.uk/software/figtree/>
- RAMBAUT A & DRUMMOND AJ. 2007. Tracer version 1.5. Molecular evolution, phylogenetics and epidemiology. <http://beast.bio.ed.ac.uk/Tracer>
- RONQUIST F & HUELSENBECK JP. 2003. MrBayes 3: Bayesian Phylogenetic Inference Under Mixed Models. *Bioinformatics* 19: 1572–1574. <https://doi.org/10.1093/bioinformatics/btg180>
- SARIDAN A, FERNANDES A & NOOR M. 2013. Sebaran dan potensi pohon tengkawang di hutan penelitian labanan, Kalimantan Timur. *Jurnal Penelitian Dipterokarpa* 7: 101–108.
- SCHATTNER P, BROOKS AN & LOWE TM. 2005. The tRNAscan-SE, Snoscan and SnoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research* 33: W686–W689. <https://doi.org/10.1093/nar/gki366>
- STAMATAKIS A. 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.” *Bioinformatics* 30: 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- SWOFFORD DL. 2004. PAUP: Phylogenetic analysis using parsimony (and other methods) (version 4.0a) [computer software]. Sinauer Associates.
- TONG R, GUI C & ZHANG Y ET AL. 2022. Phylogenomics, plastome structure and species identification in *Mahonia* (Berberidaceae). *BMC Genomics* 23: 766.
- WANG CL, DING MQ, ZOU CY ET AL. 2017. Comparative analysis of four buckwheat species based on morphology and complete chloroplast genome sequences. *Scientific Reports* 7: 6514. <https://doi.org/10.1038/s41598-017-06961-5>
- WENG ML, RUHLMAN TA & JANSEN RK. 2017. Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. *New Phytologist* 214: 842–851.
- WICK RR, SCHULTZ MB, ZOBEL J ET AL. 2015. Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 31: 3350–3352. <https://doi.org/10.1093/bioinformatics/btv383>
- WONG SY & OGARY KM. 2026. Multiplexed inter-simple sequence repeat genotyping by sequencing (Mig-Seq) of nut producing *Rubroshorea* and *Shorea* Species. *Journal of Tropical Forest Science* 38 (1). Accepted for publication.
- YANG AH, ZHANG JJ, YAO XH & HUANG HW. 2011. Chloroplast microsatellite markers in *Liriodendron tulipifera* (Magnoliaceae) and cross-species amplification in *L. chinense*. *American Journal of Botany* 98: e123–e126.
- YULITA KS. 2016. Phenetic and phylogenetic analyses of Tengkawang (*Shorea* spp., Dipterocarpaceae) based on morphological and molecular data. *Bulletin Kebun Raya* 19: 47–56.