Integrating DNA Barcoding and Traditional Taxonomy for the Identification of Dipterocarps in Remnant Lowland Forests of Sumatra

Carina Carneiro de Melo Moura [1] , **Fabian Brambach** [2] , **Kevin Jair Hernandez Bado** [1] ,
Konstantin V. Krutovsky [1,3,4,5,6] , **Holger Kreft** [2] , **Sri Sudarmiyati Tjitrosoedirdjo** [7] ,
Iskandar Z. Siregar [8] and **Oliver Gailing** [1,3,*]

[1] Department of Forest Genetics and Forest Tree Breeding, University of Göttingen, Büsgenweg 2,
37077 Göttingen, Germany; carinamoura@uni-goettingen.de (C.C.d.M.M.); kev.hdez.b@gmail.com (K.J.H.B.);
konstantin.krutovsky@forst.uni-goettingen.de (K.V.K.)

[2] Biodiversity, Macroecology and Biogeography, University of Göttingen, Büsgenweg 1, 37077 Göttingen,
Germany; fbrambach@uni-goettingen.de (F.B.); holger.kreft@forst.uni-goettingen.de (H.K.)

[3] Center for Integrated Breeding Research, University of Göttingen, Albrecht-Thaer-Weg 3,
37075 Göttingen, Germany

[4] Laboratory of Population Genetics, N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences,
3 Gubkin Str., Moscow 119333, Russian

[5] Laboratory of Forest Genomics, Genome Research and Education Center, Institute of Fundamental Biology
and Biotechnology, Siberian Federal University, 50a/2 Akademgorodok, Krasnoyarsk 660036, Russia

[6] Department of Ecosystem Science and Management, Texas A&M University, College Station,
TX 77843-2138, USA

[7] Southeast Asian Regional Center for Tropical Biology (SEAMEO BIOTROP), Jalan Raya Tajur Km. 6,
Bogor 16144, Indonesia; sritjitro@gmail.com

[8] Department of Silviculture, Faculty of Forestry, Bogor Agricultural University, Dramaga Campus,
Bogor 16680, Indonesia; izsiregar@gmail.com

* Correspondence: ogailin@gwdg.de; Tel.: +49-551-3933536

Abstract: DNA barcoding has been used as a universal tool for phylogenetic inferences and diversity assessments, especially in poorly studied species and regions. The aim of this study was to contrast morphological taxonomy and DNA barcoding, using the three frequently used markers *matK*, *rbcL*, and *trnL-F*, to assess the efficiency of DNA barcoding in the identification of dipterocarps in Sumatra, Indonesia. The chloroplast gene *matK* was the most polymorphic among these three markers with an average interspecific genetic distance of 0.020. The results of the molecular data were mostly in agreement with the morphological identification for the clades of *Anthoshorea*, *Hopea*, *Richetia*, *Parashorea*, and *Anisoptera*, nonetheless these markers were inefficient to resolve the relationships within the *Rubroshorea* group. The maximum likelihood and Bayesian inference phylogenies identified *Shorea* as a paraphyletic genus, *Anthoshorea* appeared as sister to *Hopea*, and *Richetia* was sister to *Parashorea*. A better discriminatory power among dipterocarp species provided by *matK* and observed in our study suggests that this marker has a higher evolutionary rate than the other two markers tested. However, a combination of several different barcoding markers is essential for reliable identification of the species at a lower taxonomic level.

Keywords: *matK*; *rbcL*; *trnL-F*; Dipterocarpoideae; tropical tree diversity; genetic distance; reference DNA library

1. Introduction

The Dipterocarpaceae family consists of approximately 680 species that are commonly placed in the two subfamilies, Monotoideae (~30 species in Africa and Madagascar, and the monotypic *Pseudomonotes* in Colombia) and Dipterocarpoideae, which hold the majority of species and are confined to the eastern tropics from India to New Guinea [1–6]. Members of the Dipterocarpoideae dominate the diverse rainforests of Sundaland, often with many co-occurring species [7]. Due to their high abundance, mechanical wood-properties, and tall stature, they provide some of the most valued and abundant sources of tropical hardwood, and at the same time are significant stores of aboveground carbon [8]. Hence, dipterocarps are of great ecological and economic importance [6], and understanding their diversity is necessary to advance our general knowledge of Southeast Asian rainforests. Despite being some of the most prevalent trees in the Asian tropics, dipterocarps are increasingly threatened by deforestation and land-use change [9]. For effective conservation of the threatened species and genetic diversity of the Dipterocarpaceae, it is important to understand their species-level taxonomy as well as their origin and the evolutionary processes that have led to the astonishing diversity of the family. To this end, molecular phylogenies are needed that provide basic knowledge on the evolutionary history and phylogenetic relationships of extant species [10]. Indeed, the dipterocarp family has already received great interest in traditional morphology-based taxonomy [11–13] and different phylogenetic studies based on DNA-markers ranging from Restriction Fragment Length Polymorphism (RFLP) of plastid regions [1], plastid DNA sequences [14,15], internal transcribed spacers [16], and nuclear genes [17] to genome-wide markers [3,18,19].

The exact taxonomic placement of the Dipterocarpaceae family within the angiosperms was disputed for a long time, including placements into the orders Theales and Malvales [20]. Based on recent phylogenetic and phylogenomic studies, the placement in Malvales is now universally accepted, but the relationships among the Dipterocarpaceae and the closely related families Cistaceae and Sarcolaenaceae are still insufficiently resolved [21]. The largest tribe of Dipterocarpaceae, Shoreeae, consists of the genera *Shorea*, *Hopea*, *Parashorea*, and the monotypic *Neobalanocarpus* [22]. Generic limits in the tribe are obscure and recent studies have shown that *Shorea* is in fact paraphyletic with *Hopea*, *Parashorea*, and *Neobalanocarpus* nested within it [3,23]. Nevertheless, based on phylogenomic data, well-defined clades within the tribe Shoreeae can be identified. They are consistent with most of the traditional genera and recognized subgenera of *Shorea*: *Anthoshorea* (white meranti), *Richetia* (yellow meranti), *Shorea* (balau/selangan batu), *Rubroshorea* (red meranti), and *Doona* (*Pentacme* has not been included in phylogenomic studies so far) [3,19].

Despite the comparatively large attention that this group has received due to its economic importance, identification of dipterocarps can be challenging [11–13]. This is due mainly to the large size of most dipterocarps and their characteristic periodic mass flowering and fruiting, which results in trees with reproductive structures absent for most of the time. Another complicating factor—especially in ecological studies—is that several closely related and morphologically similar species may co-occur. In this context, DNA barcoding can provide an independent source of information to delimit and identify species [24,25].

The *matK* and *rbcL* loci are considered standard plant DNA barcoding markers due to their universality, relatively high overall sequence quality, low cost, and high discriminatory power between angiosperms [26,27]. Specifically, *rbcL* has a higher PCR amplification success, but lower discriminatory power than *matK*. The use of both barcoding markers, *rbcL* together with *matK*, was proposed by the CBOL Plant Working Group of the Consortium for the Barcoding of Life [28] to have a higher combined discriminatory power [29]. Non-coding regions have also been implemented as barcoding marker [30]. Thus, joint use of coding and non-coding regions is an important step to implement a plant barcode database as a tool for accurate diversity assessments and to develop conservation strategies.

A detailed DNA barcoding dataset of dipterocarps with comprehensive coverage across taxonomic groups and geographic areas is currently lacking [25]. So far, most efforts have concentrated on the Malay Peninsula and Borneo [3,16,24], while material from Sumatra has hardly been included in

analyses (but see [25]). At the same time, large-scale logging and subsequent deforestation in Sumatra over the last decades [31] have decimated dipterocarp populations on the island, and many of the formerly widespread species are now threatened with extinction [32]. Detailed assessments of dipterocarp diversity and composition in the remaining forests of Sumatra are therefore a requirement for effective conservation measures. Here, we contrast the traditional morphological taxonomy and the DNA barcoding approach for the identification of dipterocarp species in remnant lowland rainforests of Sumatra, Indonesia. We used three DNA barcoding markers—*rbcL*, *matK*, and *trnL-F* to (i) assess the dipterocarp identification using phylogenetic trees; and (ii) test the efficiency of these markers for the identification of dipterocarps.

2. Materials and Methods

2.1. Study Area and Specimen Collection

The study was a part of the Collaborative Research Centre 990: Ecological and Socio-economic Functions of Tropical Lowland Rainforest Transformation Systems (CRC990: EFForTS project, https://www.uni-goettingen.de/efforts) in Jambi Province, central Sumatra, Indonesia. The study region is characterized by an average annual temperature of 26.7 ± 0.2 °C and mean annual precipitation of approximately 2235 ± 381 mm [25,33,34]. Samples were collected in two areas, the 'Bukit Duabelas landscape' and the 'Harapan landscape', respectively, as part of plot based (0.25 ha, 50 × 50 m, four plots per landscape and land-use) inventories on well-drained soils in four land-use types: (1) logged-over primary rain forest, (2) jungle rubber agroforestry, (3) rubber plantations, and (4) oil palm plantations. In the 'Harapan landscape', we also collected samples from 12 riparian plots, four each in logged-over forest, rubber and oil palm plantations. Details about the sampling design can be found in Drescher et al. [33] and Paoletti et al. [35]. Over well-drained soils, dipterocarp species were abundant and diverse in all forest plots (c. 280 individuals in 13 species) but were mainly absent from the more intensely used land-use types (one species rarely present in rubber agroforestry, another in rubber monoculture) [34]. We collected a total of 80 herbarium specimens assigned initially to Dipterocarpaceae in the field.

2.2. Morphology-Based Species Identification

During plot-inventories, all species were pre-identified as morphospecies in the field. For each morphospecies, herbarium specimens of at least one individual were collected, stored and prepared for later morphological identification at Indonesian herbaria (Herbarium Bogoriense and BIOTROP Herbarium). The herbarium specimens were cross-referenced with the available specimens at the Indonesian herbaria and identified to species or morphospecies level by associated taxonomists. Subsequently, we checked and revised all identifications by comparing collected specimens and high-quality standardized photographs taken in the field to keys and descriptions in standard taxonomic literature [11–13,36] and the online repositories of herbarium specimens at BioPortal (http://bioportal.naturalis.nl) and JSTOR Global Plants (http://plants.jstor.org). During the identification, we focused on vegetative traits, as flowering or fruiting material was not available. Species were distinguished based on traits of the trunk (e.g., presence/form of buttresses and stilt-roots), bark (including the inner layers), twigs and stipules (size, color, indument), and leaves (petiole, size, venation, surface, indument). Representative specimens for all species of *Shorea* sect. *Rubroshorea* are shown in Figure 1. We also checked the identifications based on morphology against the placement of all specimens in our phylogenetic trees.
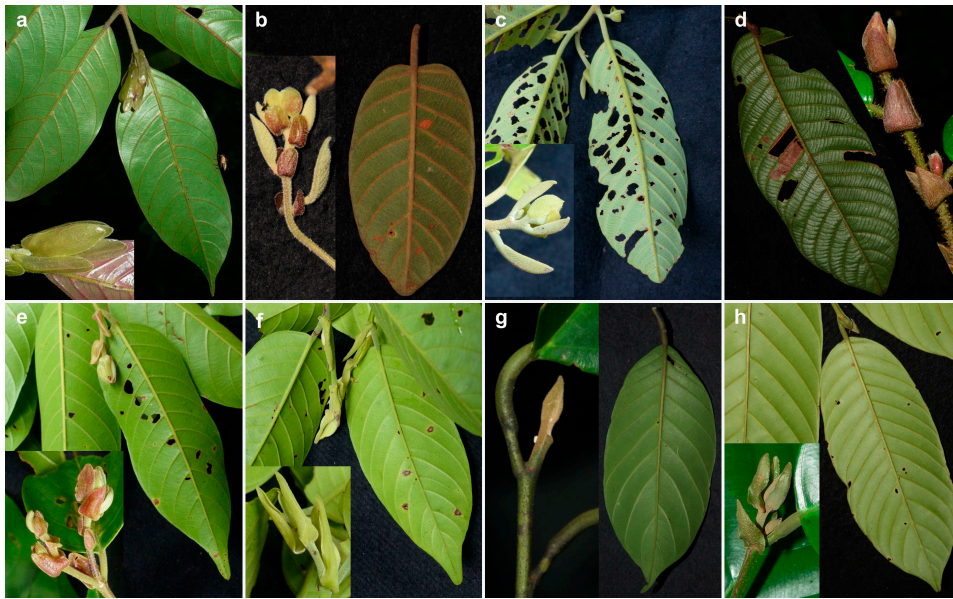
Figure 1. Morphological traits of all Sumatran species and subspecies of red meranti (*Shorea* sect. *Rubroshorea*) sampled for this study. Shown are branchlets, stipules, and leaves of (**a**) *Shorea acuminata* (from specimen Rembold KR0822), (**b**) *S. dasyphylla* (KR0546), (**c**) *S. leprosula* (KR5454), (**d**) *S. ovalis* (KR0891), (**e**) *S. parvifolia* subsp. *parvifolia* (KR5463), (**f**) *S. parvifolia* subsp. *velutinata* (KR5509), (**g**) *S. pauciflora* (KR4807), and (**h**) *S. singkawang* (KR0842). Not to scale, photographs by K. Rembold.

2.3. DNA Extraction, PCR Amplification, and Sequencing

Together with herbarium specimens, leaf tissues of approximately 2 cm^2 were collected from each sample and dried in silica-gel until DNA extraction. DNA was extracted from the dried leaf tissue following the manufacturer's protocol for the DNeasy 96 Plant Mini Kit (Qiagen, Hilden, Germany). The concentration of the extracted DNA was checked using 1% agarose gel electrophoresis with 1× TAE buffer solution, and 4 µL Roti-Safe dye. DNA fragments for each sample were then isolated and purified from the agarose gel with a volume of 13 µL Elution Buffer (innuPREP Gel Extraction Kit, Analytik Jena, Jena, Germany).

For each extracted DNA sample, polymerase chain reaction (PCR) was carried out using universal primers for the chloroplast DNA markers *rbcL*, *matK*, and *trnL-F* (Table 1). PCR was performed in a Peltier Thermal Cycler PTC-200 (MJ Research Inc., Waltham, MA, USA) with a total reaction mixture volume of 14 µL, which included a diluted 1 µL DNA sample, 1.5 µL PCR buffer (with 0.8 M Tris-HCl, 0.2 M (NH$_4$)2SO$_4$), 1.5 µL MgCl$_2$ (25 mM), 1 µL dNTPs (2.5 mM of each dNTP), 1 µL of forward primer, and 1 µL reverse primer (5 pM/µL each), 0.2 µL (5 U/µL) HOT FIREPol® Taq-Polymerase (Solis BioDyne, Tartu, Estonia), and 6.8 µL ddH$_2$O.

The PCR program consisted of an initial denaturation at 95 °C for 15 min, followed by 35 cycles of denaturation at 94 °C for 1 min, annealing at 50 °C for 1 min, elongation at 72 °C for 1.5 min and a final extension at 72 °C for 20 min. PCR products were separated and visualized on 1% agarose gels, excised from the gel and purified with the innuPREP Gel Extraction Kit protocol (Analytik Jena, Jena, Germany).

Sequencing reactions were done with the BrilliantDye v3.1 Terminator Cycle Sequencing Kit optimized for Dye Set Z (NIMAGEN, Nijmegen, The Netherlands), and purified following the manufacturer's protocol of DyeEx® 96 Kit (Qiagen, Hilden, Germany). The same primers used for amplification were also used for sequencing (Table 1). The total sequencing reaction mixture included 2 µL DNA template (5–10 ng), 4.5 µL ddH$_2$O, 0.5 µL BrilliantDye v3.1, 2 µL 5× Sequencing Buffer, 1 µL Forward/Reverse primer (5 pM/µL). Nucleotide sequences were analyzed using an ABI Prism Genetic Analyzer 3130xl with the Sequence Analysis v5.3.1 software (Applied Biosystems, Foster City, CA, USA).

Table 1. List of primers used in this study.

Barcode Region	Name of Primer	Primer Sequence (5′ → 3′)	Reference
matK	3F_KIM_f	CGTACAGTACTTTTGTGTTTACGAG	[28]
	1R_KIM_r	ACCCAGTCCATCTGGAAATCTTGGTTC	[28]
	390_f	CGATCTATTCATTCAATATTTC	[28,37]
	990_r	GGACAATGATCCAATCAAGGC	[14]
rbcL	rbcLa_f	ATGTCACCACAAACAGAGACTAAAGC	[38,39]
	rbcLajf634R_r	GAAACGGTCTCTCCAACGCAT	[40]
trnL-F	B49317_f	CGAAATCGGTAGACGCTACG	[30]
	B49873_f	GGTTCAAGTCCCTCTATCCC	[30]
	A50272_r	ATI'TGAACTGGTGACACGAG	[30]

2.4. Nucleotide Sequence Data Analysis

Both forward and reverse nucleotide sequence were visualized and aligned using the CodonCode Aligner software (https://www.codoncode.com/aligner). Sequences were manually checked; sequencing errors, if any, were corrected, consensus sequences were generated and then used for multiple sequence alignments. BLAST searches were performed for consensus sequences to identify best matches in the National Center for Biotechnology Information (NCBI) GenBank and Barcode of Life Data Systems (BOLD) [41] databases. Additionally, sequences from BOLD were included in the phylogenetic reconstruction (accession numbers are presented in the figures). All amplified sequences of the Dipterocarpaceae family obtained in this study were uploaded to the NCBI Genbank database, accession numbers MN444889-MN445045.

2.5. Genetic Distance and Phylogenetic Analysis

The nucleotide divergence between sequences was estimated using the Kimura-2-parameter genetic distance for each barcode markers *matK*, *rbcL* and *trnL-F* and for the combined markers (*matK* + *rbcL* and *matK* + *rbcL* + *trnL-F*) using their concatenated sequences. A uniform distribution was set as rate variation among sites. The overall mean genetic distance, as well as intraspecific and interspecific genetic distances were calculated for each species identified by traditional taxonomical features.

Phylogenetic trees were generated for each marker separately and based on the three markers combined (total length of alignment = 2204 bp; *matK* = 614 bp; *rbcL* = 603; *trnL-F* = 987 bp; see Table 2) using maximum likelihood (ML) methods in MEGA-X software [42] and Bayesian inference in BEAST and BEAUti 1.8.0 [43] by choosing the Hasegawa, Kishino and Yano (HKY) model as a nucleotide substitution model for nucleotide sites, "Yule process" option (Yule model of branching) for trees and "strict model" for molecular clock that assumes homogeneous rates among branches [44]. The HKY model considers different rates of transitions and transversions as well as unequal frequencies [45]. The considered rate of variation among sites for this model was the gamma distribution with five discrete gamma categories. Stationarity and convergence of runs were checked using Tracer 1.5 [46]. The maximum clade credibility tree was generated from trees produced by BEAST using TreeAnnotator 1.8.0 [43]. ML trees were calculated with 1000 bootstrap replications using the HKY model. The initial tree for the ML tree was kept as default preference, and the nearest-neighbor-interchange (NNI) heuristic method was used to search for the final ML tree. Gaps and missing data treatment were selected as partial deletion with 95% site coverage cutoff.

Table 2. Sequence information and characteristics of the *rbcL*, *matK* and *trnL-F* loci.

Parameter	*matK*	*rbcL*	*trnL-F*
Number of samples used for amplification and sequencing	78	78	78
Number of obtained sequences	66	66	42
Sequencing success rate, %	81.0	83.5	54.0
Variable sites (proportion), %	23.1	21.3	18.0
Parsimony-informative sites (proportion), %	11.1	7.8	-
CG content mean (range), %	32.3	43.7	29.0
Length of the alignment, bp	614	603	987

3. Results

3.1. Taxonomic Resolution of DNA Barcoding Markers

The sequencing success rate of the Dipterocarpaceae family for *matK*, *rbcL*, and *trnL-F* markers was 81%, 83.7%, and 54%, respectively (Table 2). The results of the BLAST performed using NCBI and BOLD platforms allowed us to correct the taxonomic identification for a significant number of specimens. Thirty-seven percent of the species, 5% of genera and 4% of all families were reassigned after comparison with the barcoding dataset and based on the subsequent new morphological identification performed using phylogenies as support (see Table S1).

The *matK* marker was efficient to identify samples at species level for the specimens belonging to the groups *Anthoshorea*, *Hopea*, *Richetia*, and *Parashorea*. However, this marker proved inefficient to resolve the relationships within the *Rubroshorea* clade.

The overall genetic distance estimated for the *matK* sequences was 0.020, for *rbcL* 0.017, for *trnL-F* 0.026, for both *matK* and *rbcL* 0.019, and 0.021 for the three barcodes together. Figure 2 shows the boxplots of the genetic distances, revealing a clear difference between the intraspecific and interspecific genetic distances for each barcode marker and for the combined dataset (*matK* + *rbcL* + *trnL–F* and *matK* + *rbcL*) for all clades except *Rubroshorea*. Low differences were observed between pairwise intra- and interspecific genetic distances within section *Rubroshorea*, as the barcode makers used in this study were unsuccessful to distinguish the section *Rubroshorea* at species level (Figure 2).
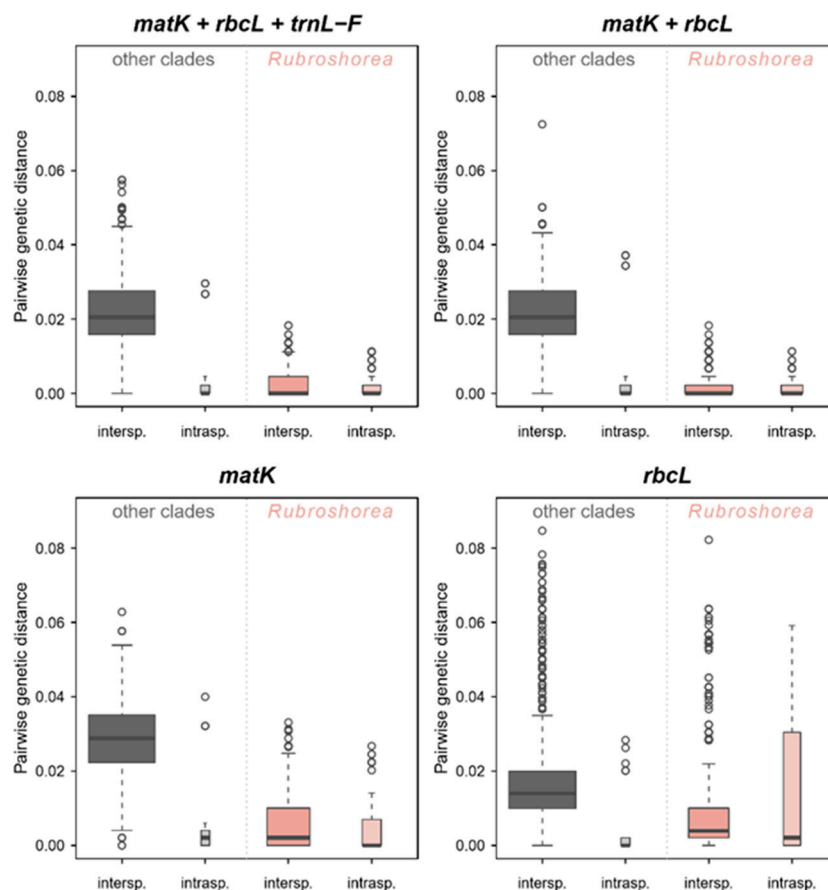

Figure 2. Boxplot representation of intraspecific (intrasp.) and interspecific (intersp.) pairwise genetic distances for the Dipterocarpaceae family based on each traditional barcode marker and the combined dataset: *matK* + *rbcL* + *trnL-F*; *rbcL* + *matK*; *matK*; and *rbcL*. *Rubroshorea* was plotted separately since the barcode makers used in this study failed to differentiate this section at species level. The width of each boxplots is proportional to the square-roots of the number of observations.

3.2. Species Assignment Using Phylogenetic Trees

All markers were efficient to distinguish the taxa at the family level, and the combination of the DNA barcodes was more efficient to allow the taxonomic identification of the Dipterocarpaceae at lower taxonomic levels (Table 3). In all phylogenetic trees (Figures 3 and 4, and Figures S1–S4), the phylogenetic relationships within *Rubroshorea* remained unresolved.

Table 3. Number of taxa genetically resolved for Dipterocarpaceae using individual barcode markers and the combined dataset. Subgenera of *Shorea* s.l. (*Anthoshorea*, *Doona*, *Richetia*, *Shorea* s.s., *Rubroshorea*) are counted as distinct genera.

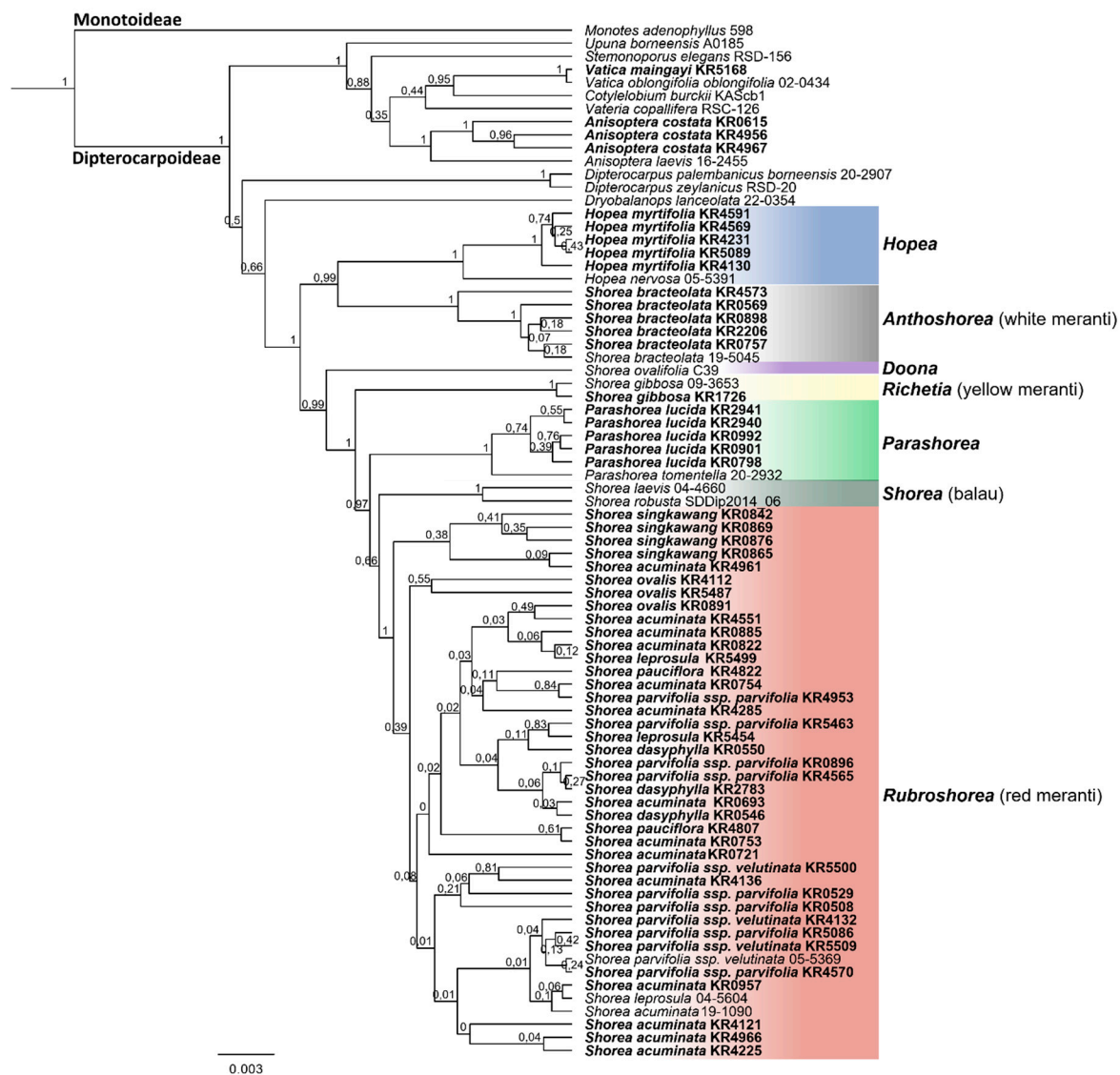| Barcode Regions	Species	Genera	Subfamily	Family
matK	17	12	2	1
rbcL	12	11	2	1
trnL-F	18	14	2	1
matK + rbcL	20	16	2	1
matK + rbcL + trnL-F	21	16	2	1

Figure 3. Bayesian inference tree based on the concatenated sequences of the *matK*, *rbcL*, and *trnL-F* markers. The numbers at the tree nodes represent the posterior probability. Tips display species IDs, samples collected for this study are depicted in bold face (see Table S1 for details), major clades of Shoreeae are color-highlighted.
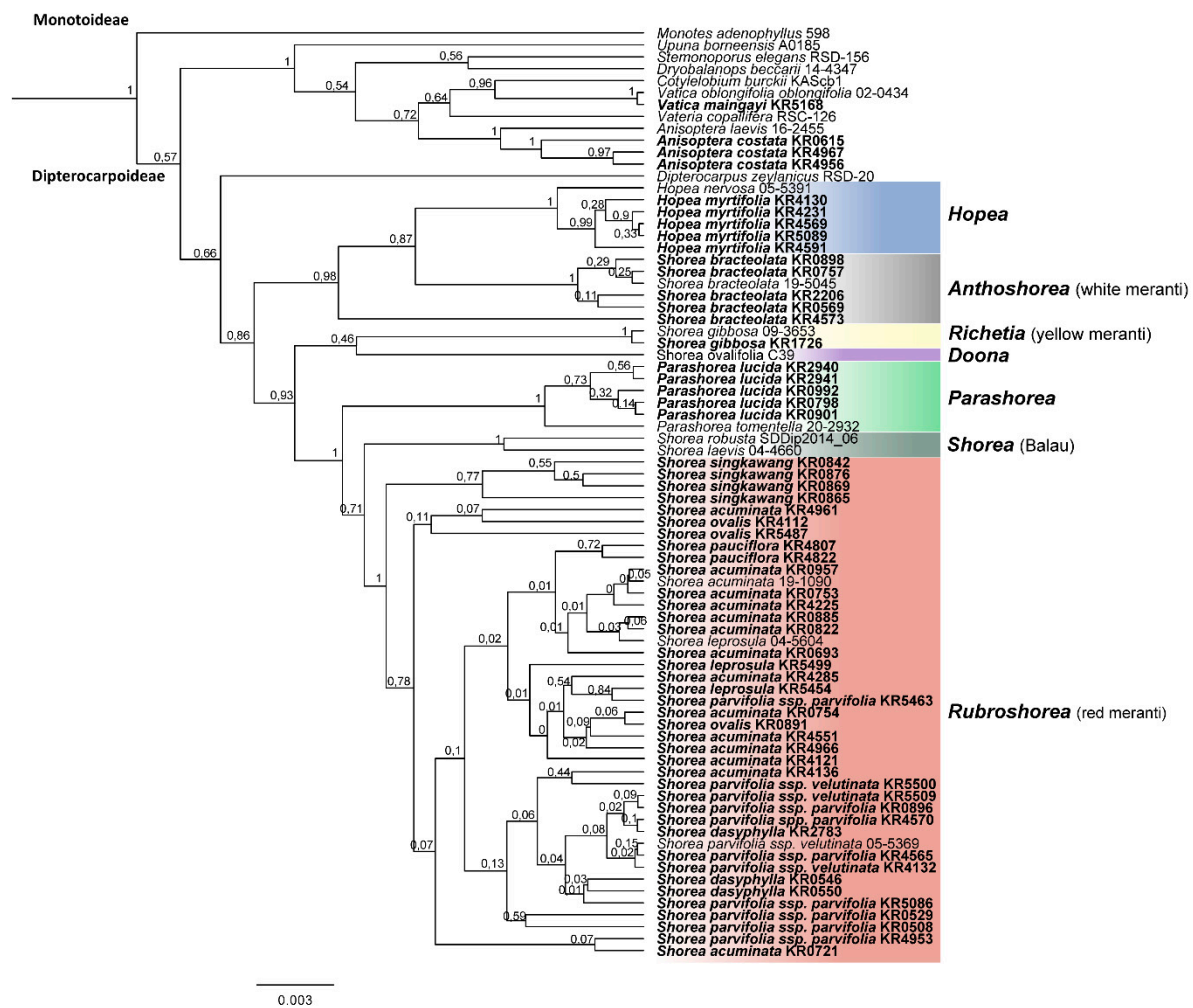
Figure 4. Bayesian inference tree based on the concatenated sequences of the *matK* and *rbcL* markers. The numbers at the tree nodes represent the posterior probability. Tips display species IDs, samples collected for this study are depicted in bold (see Table S1 for details), major clades of Shoreeae are color-highlighted.

The topology of the BI and ML trees mirrored each other, the main clades presented good support (bootstrap/posterior probabilities > 0.7/70%) and were consistent for each marker and for the concatenated sequences. *Monotes* was set as outgroup in the phylogenetic analysis (Figures 3 and 4, and Figures S1–S4).

Overall, the phylogenetic tree based on the sequences of the three concatenated markers showed stronger node support and better resolution of the relationships between species of the Dipterocarpaceae family than the individual markers (Table 3). Dipterocarpoideae was resolved as monophyletic lineage with strong support (PP = 1). *Vatica* and *Anisoptera* were retrieved with strong support (PP = 0.95 and 1, respectively, Figure 3 and Figure S1) and *Dryobalanops* was sister to Shoreeae with moderate support (PP = 0.66). Paraphyletic *Shorea* (including *Hopea* and *Parashorea*) was divided into seven major lineages with high posterior probabilities (0.97 to 1.0, Figure 3 and Figure S1): *S. bracteolata* (*Shorea* subgenus *Anthoshorea*) and *Hopea* appeared as sister to the remaining clades. *Shorea* subgenus *Doona*, subgenus *Richetia*, *Parashorea*, and subgenus *Shorea* (Balau lineage) were then successively sister to the large subgenus *Rubroshorea*. The latter comprised a monophyletic group, but resolution within the clade was low (Figure 3 and Figure S1). The combined tree and the individual markers failed to resolve the taxonomic relationship within *Rubroshorea*, but were efficient to assess a precise taxonomic

identification at species level for the following taxa: *Anisoptera costata*, *Vatica maingayi*, *Hopea myrtifolia*, *Shorea bracteolata*, *Shorea gibbosa*, and *Parashorea lucida*.

The resolution of the two-marker tree based on *rbcL* and *matK* was equivalent to the tree using the additional intergenic spacer *trnL-F*, and grouped all main lineages in monophyletic clades, however with lower bootstrap support for the Shoreeae clade (PP = 0.86). Nevertheless, low posterior probability was found supporting the lineages Balau, *Parashorea,* and *Richetia* (Figure 4). The combination of *matK* and *rbcL* was efficient to identify the following taxa at species level: *A. costata*, *V. oblongifolia*, *H. myrtifolia*, *S. bracteolata*, *P. lucida*, *S. gibbosa*, and *S. singkawang*.

Overall, the phylogenetic tree based on the *matK* sequences (Figure S2) displayed a similar topology in comparison with the three-marker tree with reference to the position of most main lineages in Shoreeae (*Hopea*, *Anthoshorea*, *Doona*, *Parashorea,* and *Richetia*) but the phylogenetic relationships of *Rubroshorea* and Balau lineages remained unresolved based on this single marker. Still, *matK* alone was not efficient to correctly place the species within the genus *Hopea*: *H. nervosa* and the specimen *H. myrtifolia* KR4130 showed low differentiation. A similarly dubious position was observed for the sample *S. bracteolata* KR4573, which clustered with low support in the same clade as *Parashorea* (Figure S2).

A lack of resolution was observed in the phylogenetic analysis based on the *rbcL* marker regarding the topology of the main subfamilies (Monotoideae and Dipterocarpoideae) and main sections (Figure S3), while the phylogenetic relationships of these lineages were clarified in the phylogenies based on *matK* (Figure S2) and the concatenated markers (Figures 3 and 4 and Figures S1). Species of genera *Anisoptera* (PP > 0.8), *Vatica* (PP = 1.0) and *H. myrtifolia* (PP = 1.0) clustered with high support at species level (Figure S3). In contrast with the *matK* tree (Figure S2), the genus *Parashorea* fell into the *Rubroshorea* clade with low node support (PP = 0.39) in the *rbcL* tree (Figure S3).

The phylogenetic tree based on the intergenic spacer *trnL-F* was efficient to resolve the relationship only of the species *H. myrtifolia*, *S. bracteolata*, *S. gibbosa,* and *A. costata,* and presented clear distinction at subfamily level. However, the overall topology of the tree displayed low resolution concerning the position of the main lineages of the Dipterocarpaceae family (Figure S4).

4. Discussion

4.1. Applicability of DNA Barcoding

We used a dataset of dipterocarp samples from Sumatra to explore the utility of DNA barcoding for species identification in this poorly sampled tropical region and for groups where traditional morphology-based species identification is challenging. Our results show that the applicability of barcoding depends on the chosen markers and the analyzed clades.

The *matK* marker has a high evolutionary rate, which gives a high discriminatory power among angiosperm species [28,47,48]. The phylogenetic trees reconstructed in this study using *matK* had a reasonable resolution to the species-level, giving a broad view of the relationships among Dipterocarpaceae species. Nevertheless, *matK* has been reported to have a lower universality, meaning that it is difficult to amplify specimens from evolutionary distant clades if they are arranged in a high-throughput format with the currently established PCR primers [28,29,48]. The currently established PCR primers for *matK* showed a high rate of recovery within family Dipterocarpaceae, which helped to avoid this problem. In their review about single-locus DNA barcodes, Li et al. [48] indicated that the discrimination rate of *matK* ranges from 49% to 90% across different taxonomic groups.

Contrarily, the *rbcL* marker provides a high universality in terms of steady PCR amplification, high-quality bidirectional sequencing, and reliable nucleotide sequence alignment in most land plants. However, *rbcL* does not have sufficient discriminatory power due the relatively low divergence of this locus in flowering plants observed also in the current study; *rbcL* alone was inefficient to access the main lineages of the Dipterocarpaceae [28,29].

Similar low resolution of the phylogenetic relationships among the main lineages of the Dipterocarpaceae family was observed for the intergenic spacer *trnL-F*. However, the combined use of *rbcL* and *matK* has proven to be a powerful tool in phylogenetic analyses by combining the two strong features of both markers (high levels of polymorphism in *matK* and the universality of *rbcL*) [24,28,29,40] and the applicability of both barcode markers is confirmed by our study for most clades (Figure 3, Figure 4 and Figure S1). The two-marker tree (*matK* + *rbcL*) displayed an equivalent topology to the phylogenetic tree based on the three combined markers (*matK, rbcL,* and *trnL-F*) and it was efficient to access the main lineages with an equal level of resolution (Figure 4). The phylogenetic tree based on the three markers was superior to the two-marker system (*matK* + *rbcL*) only by displaying higher support of the nodes, allowing a better interpretation of the evolutionary history of the group.

4.2. Phylogenetic Relationships of Sumatran Dipterocarpaceae

The phylogenetic analysis based on two (*matK* and *rbcL*) and three (*matK, rbcL,* and *trnL-F*) DNA barcoding markers confirmed the monophyly of the subfamily Dipterocarpoideae and assembled its main lineages in agreement with previous studies using denser taxon sampling and more markers [23] or phylogenomic methods [3]: Dipterocarpoideae is composed of two major clades. The first clade contains all genera of tribe Dipterocarpeae except the toponymous *Dipterocarpus* – i.e., *Upuna, Stemonoporus, Anisoptera* (three specimens of one species sampled in our study), *Vatica* (one specimen sampled), *Cotylelobium, Vateria,* and *Vateriopsis* (not sampled here) – and corresponds to clade IV of Heckenhauer et al. [23] (Figures 3 and 4). In the second clade, *Dipterocarpus* and *Dryobalanops* are successively sister to the tribe Shoreeae, which contains the bulk of our specimens. The largest genus of Shoreeae, *Shorea,* has been shown to be paraphyletic with respect to the smaller monophyletic genera *Hopea, Parashorea, Neobalanocarpus,* and *Pentacme* (the latter two not sampled here) based on plastid [1,14,20,23,49] and nuclear markers [17], a combination of both [16], and RADseq [3]. However, when *Shorea* is split into subgenera, monophyletic groups can be retrieved in this tribe [3,14,23,40], and these groups are supported by morphological characters [12,22,36]. The topology of Shoreeae from our study differs somewhat from that of previous studies [3,14,17], mainly in the placement of *Doona* and *Richetia* (Figures 3 and 4), but overall finds the same monophyletic groups: *Hopea, Parashorea,* and the *Shorea* subgenera *Anthoshorea* (white meranti), *Doona, Richetia* (yellow meranti), *Shorea* (balau), and *Rubroshorea* (red meranti). The paraphyletism of *Shorea* calls for a redefinition of generic boundaries as previously suggested [3,14] in Shoreeae, either including all species of Shoreeae in an expanded *Shorea* s.l. or by raising the mentioned subgenera to generic rank in line with the classifications based on general morphology [12,22,36].

The phylogenetic relationships of the relatively young (evolutionary age ca. 15 Ma [23]) and species rich (68 species [13]) *Rubroshorea* could so far only be reliably resolved by using genomic data, possibly indicating incomplete lineage sorting, which would be consistent with a recent and ongoing diversification of the group and/or adaptive introgression. Most of the species of *Rubroshorea,* for which we had several samples available, were retrieved as polyphyletic, especially in the two-markers tree (*matK* + *rbcL*), but also in the three-markers analysis (Figures 3 and 4). This could indicate mis-identification of species based on morphology. Correct species identification in dipterocarps is hampered by the fact that often, only vegetative material from the specimens is available, as was the case in our study. However, decades of work by dedicated field and herbarium taxonomists have produced vast literature [11–13,36] for identification based on vegetative traits of trunk, bark, twigs, stipules, and leaves (Figure 1). Taking into account these traits, species of red meranti (*Rubroshorea*) in Sumatra can be distinguished fairly easily. An exception is the distinction between the two subspecies of *Shorea parvifolia* (Figure 1e–f), which remains challenging as traits are variable depending on the life stage of the trees and because forms that are intermediate between the subspecies can occur. Notably, even morphologically clearly distinct species such as *Shorea ovalis* (Figure 1d) appeared in different clades, so even if we failed in identifying all specimens correctly, a strong mismatch between morphology and the barcoding results for the *Rubroshorea* clade remains. The prevalent low support

values in the clade indicate that the markers used in our study do not provide sufficient resolution for species-level identification of red meranti (*Rubroshorea*) taxa. For better results, the inclusion of additional markers or phylogenomic approaches would be desirable. In addition, reference databases are often geographically strongly biased, which may hamper the reliability of phylogenetic trees. With our study, we provide DNA barcodes for 13 species and two subspecies of Dipterocarpaceae from the under-sampled island of Sumatra.

Although DNA barcoding does not always have enough discriminatory power to retrieve species phylogenetically, the present study supports the applicability of the markers *rbcL*, *matK*, and *trnL-F* for placing dipterocarp specimens in highly-supported major clades corresponding to taxonomic groups at the level of genus or subgenus. In addition, in all clades, except *Rubroshorea*, multiple specimens per species were resolved in highly supported monophyletic clades, often clearly distinguished from other congeneric species (Figure 3). These results thus confirm that (with the exception of *Rubroshorea* species) dipterocarp species in Sumatra can reliably be separated using DNA barcoding.

4.3. Combined Identification Using Morphology and DNA Barcodes

The use of DNA barcoding allows to recognize taxonomic misidentification of samples and facilitate the identification of phylogenetic species. In this study, DNA barcoding proved to be a useful tool to enhance the accuracy of the taxonomic identification of the taxa belonging to the Dipterocarpaceae family. Especially in the context of plot-based sampling, as applied in the EFForTS project, a large number of collected specimens usually contain juvenile or otherwise vegetative material. Placement of these specimens in the correct major taxonomic groups (families, genera) based on morphology alone is often a great challenge and a time-consuming pursuit. DNA barcoding can greatly facilitate this process. However, it does not replace the traditional taxonomic identification, at least for taxa with high diversification rates or lack of reliable reference data, as shown for the *Rubroshorea* clade. Rather, DNA barcoding and traditional taxonomic approaches complement each other for inventories of diversity. Once a comprehensive and standardized reference database is established for understudied regions such as Sumatra, Indonesia, we advocate that the DNA barcoding method can accelerate taxonomic inventories and species discovery with great precision and be applicable for analysis of phylogenetic diversity.

5. Conclusions

The joint use of DNA barcoding markers *rbcL* and *matK* is a reliable tool for identification of land plants from Sumatra, Indonesia and the addition of further markers, such as the *trnL-F* marker, provide better node support of the lineages. Remarkably, the phylogenetic tree based on the DNA barcoding markers employed in this study recovered a topology largely consistent with recent studies based on phylogenomic data [3,19].

In summary, we highlight the applicability of the DNA barcoding as a reliable tool for species inventories and evolutionary studies in tropical areas. Nevertheless, this method cannot be taken as a replacement for taxonomic identification, but should rather be seen as a complementary tool to support the classical taxonomy.

Author Contributions: Conceptualization, all authors; Methodology, C.C.d.M.M., F.B., K.J.H.B., O.G., H.K. and K.V.K.; Formal Analysis, C.C.d.M.M., F.B. and K.J.H.B.; Taxonomic identification of specimens, F.B.; Data Curation, C.C.d.M.M. and F.B.; Writing–Original Draft Preparation, C.C.d.M.M. and F.B.; Writing–Review and Editing, C.C.d.M.M., F.B., O.G., H.K. and K.V.K.; C.C.d.M.M., F.B. and K.J.H.B contributed equally to this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Indrioko, S.; Gailing, O.; Finkeldey, R. Molecular phylogeny of Dipterocarpaceae in Indonesia based on chloroplast DNA. *Plant Syst. Evol.* **2006**, *261*, 99–115. [CrossRef]

2. Christenhusz, M.J.M.; Byng, J.W. The number of known plants species in the world and its annual increase. *Phytotaxa* **2016**, *261*, 201–217. [CrossRef]

3. Heckenhauer, J.; Samuel, R.; Ashton, P.S.; Abu Salim, K.; Paun, O. Phylogenomics resolves evolutionary relationships and provides insights into floral evolution in the tribe Shoreeae (Dipterocarpaceae). *Mol. Phylogenetics Evol.* **2018**, *127*, 1–13. [CrossRef]

4. Ashton, P.S. Dipterocarpaceae. In *Flowering Plants—Dicotyledons: Malvales, Capparales and Non-Betalain Caryophyllales*; Springer: Berlin/Heidelberg, Germany, 2003; Volume 5, pp. 182–197.

5. Londono, A.C.; Alvarez, E.; Forero, E.; Morton, C.M. A new genus and species of Dipterocarpaceae from the Neotropics. Introduction, Taxonomy, Ecology, and Distribution. *Brittonia* **1995**, *47*, 225. [CrossRef]

6. Brearley, F.Q.; Banin, L.F.; Saner, P. The ecology of the Asian dipterocarps. *Plant Ecol. Divers* **2016**, *9*, 429–436. [CrossRef]

7. Slik, J.W.F.; Poulsen, A.D.; Ashton, P.S.; Cannon, C.H.; Eichhorn, K.A.O.; Kartawinata, K.; Lanniari, I.; Nagamasu, H.; Nakagawa, M.; van Nieuwstadt, M.G.L.; et al. A floristic analysis of the lowland dipterocarp forests of Borneo. *J. Biogeogr.* **2003**, *30*, 1517–1531. [CrossRef]

8. Surianegara, I.C.; Lemmens, R.H.M.J. Timber trees: Major commercial timbers. In *Plant Resources of South-East Asia*; van der Maesen, L.J.G., Somaatmadja, S., Eds.; ESCAP CGPRT Centre: Bogor, Indonesia, 1990.

9. Ang, C.C.; O'Brien, M.J.; Ng, K.K.S.; Lee, P.C.; Hector, A.; Schmid, B.; Shimizu, K.K. Genetic diversity of two tropical tree species of the Dipterocarpaceae following logging and restoration in Borneo: high genetic diversity in plots with high species diversity. *Plant Ecol. Divers* **2016**, *9*, 459–469. [CrossRef]

10. Avise, J. *Evolutionary Pathways in Nature a Phylogenetic Approach*; Cambridge University Press: New York, NY, USA, 2006.

11. Foxworthy, F.W. *Dipterocarpaceae of the Malay Peninsula*; Malayan Forest Records; Federated Malay States Government, Forest Department: Kepong, Malaysia, 1932; p. 10.

12. Symington, C.F.; Ashton, P.S.; Appanah, S. *Foresters' manual of dipterocarps*; Forest Research Institute Malaysia (FRIM): Kuala Lumpur, Malaysia, 2004.

13. Ashton, P.S. Dipterocarpaceae. In *Flora Malesiana I*; Martinus Nijhoff Publishers: Hague, The Netherlands, 1982; Volume 9, pp. 237–552.

14. Gamage, D.T.; de Silva, M.P.; Inomata, N.; Yamazaki, T.; Szmidt, A.E. Comprehensive molecular phylogeny of the subfamily Dipterocarpoideae (Dipterocarpaceae) based on chloroplast DNA sequences. *Genes Genet. Syst.* **2006**, *81*, 1–12. [CrossRef]

15. Harnelly, E.; Thomy, Z.; Fathiya, N. Phylogenetic analysis of Dipterocarpaceae in Ketambe Research Station, Gunung Leuser National Park (Sumatra, Indonesia) based on *rbcL* and *matK* genes. *Biodiversitas J. Biol. Divers* **2018**, *19*, 1074–1080. [CrossRef]

16. Yulita, K.S.; Bayer, R.J.; West, J.G. Molecular phylogenetic study of *Hopea* and *Shorea* (Dipterocarpaceae): Evidence from the *trnL–trnF* and internal transcribed spacer regions. *Plant Species Biol.* **2005**, *20*, 167–182. [CrossRef]

17. Kamiya, K.; Harada, K.; Tachida, H.; Ashton, P.S. Phylogeny of *PgiC* gene in *Shorea* and its closely related genera (Dipterocarpaceae), the dominant trees in Southeast Asian tropical rain forests. *Am. J. Bot.* **2005**, *92*, 775–788. [CrossRef]

18. Cvetković, T.; Hinsinger, D.D.; Strijk, J.S. Exploring evolution and diversity of Chinese Dipterocarpaceae using next-generation sequencing. *Sci. Rep.* **2019**, *9*, 11639. [CrossRef]

19. Heckenhauer, J.; Paun, O.; Chase, M.W.; Ashton, P.S.; Kamariah, A.S.; Samuel, R. Molecular phylogenomics of the tribe Shoreeae (Dipterocarpaceae) using whole plastid genomes. *Ann. Bot.* **2019**, *123*, 857–865. [CrossRef]

20. Dayanandan, S.; Ashton, P.S.; Williams, S.M.; Primack, R.B. Phylogeny of the tropical tree family Dipterocarpaceae based on nucleotide sequences of the chloroplast *rbcL* gene. *Am. J. Bot.* **1999**, *86*, 1182–1190. [CrossRef]

21. The Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **2016**, *181*, 1–20. [CrossRef]

22. Maury, G. Diptérocarpacées: du Fruit à la Plantule. Ph.D. Thesis, Université Paul Sabatier, Tolouse, France, 1978.

23. Heckenhauer, J.; Samuel, R.; Ashton, P.S.; Turner, B.; Barfuss, M.H.J.; Jang, T.-S.; Temsch, E.M.; Mccann, J.; Salim, K.A.; Attanayake, A.M.A.S.; et al. Phylogenetic analyses of plastid DNA suggest a different interpretation of morphological evolution than those used as the basis for previous classifications of Dipterocarpaceae (Malvales). *Bot. J. Linn. Soc.* **2017**, *185*, 1–26. [CrossRef]

24. Heckenhauer, J.; Salim, K.A.; Chase, M.W.; Dexter, K.G.; Pennington, R.T.; Tan, S.; Kaye, M.E.; Samuel, R. Plant DNA barcodes and assessment of phylogenetic community structure of a tropical mixed dipterocarp forest in Brunei Darussalam (Borneo). *PLoS ONE* **2017**, *12*, e0185861. [CrossRef]

25. Amandita, F.Y.; Rembold, K.; Vornam, B.; Rahayu, S.; Siregar, I.Z.; Kreft, H.; Finkeldey, R. DNA barcoding of flowering plants in Sumatra, Indonesia. *Ecol. Evol.* **2019**, *9*, 1858–1868. [CrossRef]

26. Hollingsworth, M.L.; Andra Clark, A.; Forrest, L.L.; Richardson, J.; Pennington, R.T.; Long, D.G.; Cowan, R.; Chase, M.W.; Gaudeul, M.; Hollingsworth, P.M. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants: DNA barcoding. *Mol. Ecol. Resour.* **2009**, *9*, 439–457. [CrossRef]

27. Webb, C.O.; Slik, J.W.F.; Triono, T. Biodiversity inventory and informatics in Southeast Asia. *Biodivers. Conserv.* **2010**, *19*, 955–972. [CrossRef]

28. CBOL Plant Working Group. A DNA barcode for land plants. *PNAS* **2009**, *106*, 12794–12797. [CrossRef]

29. Hollingsworth, P.M.; Graham, S.W.; Little, D.P. Choosing and using a plant DNA barcode. *PLoS ONE* **2011**, *6*, e19254. [CrossRef]

30. Taberlet, P.; Gielly, L.; Pautou, G.; Bouvet, J. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* **1991**, *17*, 1105–1109. [CrossRef]

31. Laumonier, Y.; Uryu, Y.; Stüwe, M.; Budiman, A.; Setiabudi, B.; Hadian, O. Eco-floristic sectors and deforestation threats in Sumatra: identifying new conservation area network priorities for ecosystem-based land use planning. *Biodivers. Conserv.* **2010**, *19*, 1153–1174. [CrossRef]

32. IUCN. *IUCN Red List Categories and Criteria: Version 3.1*, 2nd ed.; IUCN: Gland, Switzerland; Cambridge, UK, 2012.

33. Drescher, J.; Rembold, K.; Allen, K.; Beckschäfer, P.; Buchori, D.; Clough, Y.; Faust, H.; Fauzi, A.M.; Gunawan, D.; Hertel, D.; et al. Ecological and socio-economic functions across tropical land-use systems after rainforest conversion. *Philos. Trans. R. Soc. B* **2016**, *371*, 20150275. [CrossRef]

34. Rembold, K.; Mangopo, H.; Tjitrosoedirdjo, S.S.; Kreft, H. Plant diversity, forest dependency, and alien plant invasions in tropical agricultural landscapes. *Biol. Conserv.* **2017**, *213*, 234–242. [CrossRef]

35. Paoletti, A.; Darras, K.; Jayanto, H.; Grass, I.; Kusrini, M.; Tscharntke, T. Amphibian and reptile communities of upland and riparian sites across Indonesian oil palm, rubber and forest. *Glob. Ecol. Conserv.* **2018**, *16*, e00492. [CrossRef]

36. Symington, C.F. *Foresters' Manual of Dipterocarps*; Malayan Forest Records; Penerbit Universiti Malaya: Kuala Lumpur, Malaysia, 1943; Volume 16.

37. Schmitz-Linneweber, C.; Maier, R.M.; Alcaraz, J.-P.; Cottet, A.; Herrmann, R.G.; Mache, R. The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. *Plant Mol. Biol.* **2001**, *45*, 307–315. [CrossRef]

38. Levin, R.A.; Wagner, W.L.; Hoch, P.C.; Nepokroeff, M.; Pires, J.C.; Zimmer, E.A.; Sytsma, K.J. Family-level relationships of Onagraceae based on chloroplast *rbcL* and *ndhF* data. *Am. J. Bot.* **2003**, *90*, 107–115. [CrossRef]

39. Soltis, P.S.; Soltis, D.E.; Smiley, C.J. An *rbcL* sequence from a Miocene *Taxodium* (bald cypress). *PNAS* **1992**, *89*, 449–451. [CrossRef]

40. Fazekas, A.J.; Burgess, K.S.; Kesanakurti, P.R.; Graham, S.W.; Newmaster, S.G.; Husband, B.C.; Percy, D.M.; Hajibabaei, M.; Barrett, S.C.H. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE* **2008**, *3*, e2802. [CrossRef]

41. Ratnasingham, S.; Hebert, P.D.N. BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Mol. Ecol. Notes* **2007**, *7*, 355–364. [CrossRef]

42. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]

43. Drummond, A.J.; Suchard, M.A.; Xie, D.; Rambaut, A. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **2012**, *29*, 1969–1973. [CrossRef]

44. Ferreira, M.A.R.; Suchard, M.A. Bayesian analysis of elapsed times in continuous-time Markov chains. *Can. J. Stat.* **2008**, *36*, 355–368. [CrossRef]

45. Rambaut, A.; Grass, N.C. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Bioinformatics* **1997**, *13*, 235–238. [CrossRef]

46. Rambaut, A.; Drummond, A.J. *Tracer*, version 1.5; Institute of Evolutionary Biology, University of Edinburgh: Edinburgh, UK, 2014.

47. Hajibabaei, M.; Singer, G.A.C.; Hebert, P.D.N.; Hickey, D.A. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet.* **2007**, *23*, 167–172. [CrossRef]

48. Li, X.; Yang, Y.; Henry, R.J.; Rossetto, M.; Wang, Y.; Chen, S. Plant DNA barcoding: from gene to genome: Plant identification using DNA barcodes. *Biol. Rev.* **2015**, *90*, 157–166. [CrossRef]

49. Tsumura, Y.; Kado, T.; Yoshida, K.; Abe, H.; Ohtani, M.; Taguchi, Y.; Fukue, Y.; Tani, N.; Ueno, S.; Yoshimura, K.; et al. Molecular database for classifying *Shorea* species (Dipterocarpaceae) and techniques for checking the legitimacy of timber and wood products. *J. Plant. Res.* **2011**, *124*, 35–48. [CrossRef]