THE USE OF BARCODING SEQUENCES FOR THE CONSTRUCTION OF PHYLOGENETIC RELATIONSHIPS IN THE DIPTEROCARPACEAE FAMILY

By

Kevin Jair Hernandez Bado

Master's Thesis at the Forest Genetics and Forest Tree Breeding Department

Faculty of Forest Science and Forest Ecology

Georg- August University Göttingen

Göttingen, 2018

Supervisor:	**Prof. Dr. Oliver Gailing**
Co-Supervisor:	**Prof. Dr. Konstantin V. Krutovsky**
Submitted:	**30.11.2018**

To My Beloved Family

TABLE OF CONTENTS

1. INTRODUCTION 1

1.1. OBJECTIVES 4

1.2. HYPOTHESES 4

2. MATERIALS AND METHODS 5

2.1. Study Area 5

2.2. Study Design and Specimen Collection 5

2.3. DNA Extraction 7

2.4. Polymerase Chain Reaction (PCR) Amplification and Sequencing 8

2.4.1. Leaf Samples 8

2.4.2. Honey and Pollen Samples 9

2.5. Data Analysis 10

2.6. Phylogenetic Analysis 11

2.6.1. Neighbor Joining 12

2.6.2. Maximum Parsimony 13

2.6.3. Maximum Likelihood 14

3. RESULTS 16

3.1. LEAF SAMPLES RESULTS 16

3.1.1. Neighbor Joining Phylogenetic Analyses 17

3.1.2. Maximum Parsimony Phylogenetic Analyses 33

3.1.3. Maximum Likelihood Phylogenetic Analyses 39

3.2. POLLEN AND HONEY SAMPLES RESULTS 47

4. DISCUSSION 54

4.1. Leaf Samples 54

4.2. Pollen and Honey Samples 56

5. CONCLUSIONS 58

6. REFERENCES 60

7. APPENDIXES 65

SUMMARY

Biodiversity has been one of the main hot topics in international and national levels since the Earth Summit in Rio de Janeiro, 1992. The increased awareness of the nature that surround us has steered the way human development was accepted, the beginning of a new Era has then led the path to preserve Biodiversity. Nevertheless, applying ‘eco-friendly’ processes to industrialized human activities has had a low to zero positive impact on nature, since the tendency is still towards biodiversity loss.

In order to preserve the variety of life conforming nature, it is necessary not only to create protected areas or establish guidelines to protect fauna and flora from negative human impacts, but also to understand the inherent relationships of biodiversity. Taxonomy is one of the main branches of sciences to acknowledge organisms’ identification, including the process of describing and classifying all living organisms. However, the amount of expertise required to such labor is amusingly gigantic due to the lack of taxonomic systems’ updates, which include taxonomic experts, morphological traits identification guidelines, genetic systems, among others.

Thankfully, the new and advancing technologies in computational processes has allowed the birth of new approaches for understanding evolutionary and genetic relationships of organisms, being molecular phylogeny one example. Additionally, the use of short sequences of DNA, or DNA barcodes, for promoting identification of any given species has also gained interest in recent years. It should be taken into account that this process does not seek to replace traditional taxonomic identification of species. Alternately, it rises as an aid to the taxonomic system by highlighting relationships among known taxa, atypical specimens or genetically divergent groups.

In that order of ideas, this research study aims to analyze the phylogenetic relationships of the *Dipterocarpaceae* family of trees by using DNA barcoding with two plastid regions (*rbcL* and *matK*), as well as evaluating the use of phylogenetic trees to assess phylogenetic diversity of pollen and honey.

The study sites were located in Jambi Province, Indonesia. In two regions, the ‘Bukit Duabelas landscape’ and the ‘Harapan landscape’, thirty-two subplots were established from which specimens of trees with diameter at breast high (DBH) ≥ 10 cm were identified and collected. Each species found was prepared for morphological identification at Indonesian Herbaria and a leaf tissue was collected from each sample and dried in silica-gel until DNA

extraction. As for the honey samples, *Apisdorsata* honey was collected from two bee colonies, Kampar colony and Kerinci colony. Additionally, two pollen traps were established for a period of one year, in a jungle rubber plantation and in a tropical rain forest.

The PCR amplification and sequencing was conducted using *rbcL* and *matK* markers for the samples identified as belonging to *Dipterocarpaceae* family. Honey and pollen samples were only amplified using rbcL marker.

Following, each sequence retrieved was aligned using the CodonCode Aligner software (CodonCode Corporation, https://www.codoncode.com/aligner/), taking into account the guidelines suggested by the CBOL Plant Working Group (2009). Multiple alignments were carried out with the consensus of all the aligned samples per genetic marker, i.e. *rbcL* and *matK*, using CodonCode Aligner's built-in alignment algorithms. Additionally, a concatenated alignment was created from the multiple alignments of *rbcL* and *matK* markers.

Results of the multiple alignments were exported to the MEGA-X software (Kumar et al., 2018) for the phylogenetic analyses with Neighbor Joining, Maximum Parsimony and Maximum Likelihood methods. A phylogenetic tree for each marker (*rbcL* and *matK*), and an additional concatenated tree, was generated per chosen method.

The topology of the constructed phylogenetic trees was generally similar across the different phylogenetic construction methodologies. Initially, the phylogenetic trees helped to discard five problematic samples that were clustering as a total outgroup from the *Dipterocarpaceae* set of samples.

The combination of the two genetic DNA barcode markers (*rbcL* and *matK*) proved to be superior in discrimination of taxa to species-level than the use of single-locus DNA barcodes. The *rbcL* marker is characterized by its universality and the *matK* marker is known to be a powerful discrimination tool in phylogenetic assessments.

Despite of having a low resolution, phylogenetic trees constructed of honey, pollen, and tree samples, and based on the *rbcL* marker, showed a modest discriminatory level to family and genus level, suggesting a benefit on phylogenetic diversity estimates.

ZUSAMMENFASSUNG

Seit der Konferenz der Vereinten Nationen über Umwelt und Entwicklung im Jahr 1992 ist Biodiversität eines der wichtigsten Themen im internationalen Kontext. Das höhere Bewusstsein für die uns umgebende Natur hat die Art der menschlichen Entwicklung geprägt und ein neues Zeitalter zum Schutz von Biodiversität eingeläutet. Dennoch hat die Anwendung von „umweltfreundlichen" Prozessen in industriellen menschlichen Aktivitäten nur einen geringen bis gar keinen positiven Effekt auf die Natur, da der Trend weiter in Richtung Biodiversitätsverlust geht.

Um die Vielfalt der Lebewesen zu erhalten sowie Flora und Fauna vor negativen menschlichen Einflüssen zu schützen, ist es nicht nur notwendig Schutzgebiete einzurichten oder Regelwerke aufzustellen, sondern auch zugehörigen Beziehungen der Biodiversität zu verstehen. Die Taxonomie ist eine der Hauptzweige der Wissenschaft zur Bestätigung der Identität eines Organismus', einschließlich des Prozesses der Beschreibung und Klassifizierung aller lebenden Organismen. Allerdings ist das Maß an Expertenwissen, das für diese Art von Arbeit notwendig ist, wegen des Mangels an Aktualisierungen von taxonomischen Systemen, unvergleichlich groß. Diese umfassen unter anderem taxonomische Experten und Leitfaden zur Identifizierung von morphologischen Eigenschaften.

Glücklicherweise erlauben neue und fortschrittliche Technologien eine immer höhere Rechenleistung und erlauben so die Geburt von neuen Herangehensweisen für das Verständnis von evolutionären und genetischen Zusammenhängen von Organismen. Hierbei ist die molekulare Phylogenie nur ein Beispiel. Darüber hinaus hat, in Hinblick auf die Identifizierung von Arten, der Gebrauch von kurzen Sequenzen der DNA, oder DNA Barcodes, in den letzten Jahren zunehmend Interesse erlangt. Dabei sollte jedoch beachtet werden, dass diese Prozesse nicht darauf abzielen die traditionelle taxonomische Identifizierung von Arten zu ersetzen. Sie bieten vielmehr eine nützliche Hilfe für das taxonomischen System indem sie Beziehungen zwischen bekannten Taxa, atypischen Exemplaren oder genetisch verschiedenen Gruppen, herausstellen.

In dieser Hinsicht hat die vorliegende Studie das Ziel die phylogenetischen Beziehungen von Bäumen der Familie der *Dipterocarpaceae* mit Hilfe von DNA Barcoding in zwei Plastidregionen (*rbcL* und *matK*) zu analysieren und den Nutzen von phylogenetischen Bäumen zur Beurteilung der phylogenetischer Vielfalt von Pollen und Honig zu evaluieren.

Die Untersuchungsstandorte befinden sich in der Provinz von Jambi in Indonesien. In zwei Regionen, der „Bukit Duabelas Landschaft“ und der „Harapan Landschaft“, wurden 32 Unterplots eingerichtet. Dort wurden Bäume mit einem BHD von ≥ 10 cm identifiziert und beprobt. Jede gefundene Art wurde für die morphologische Identifizierung im indonesischen Herbarium vorbereitet. Darüber hinaus wurde von jeder Probe Blattmaterial entnommen und bis zur DNA Extraktion in Silicagel getrocknet. Zur Gewinnung der Honigproben wurde *Apisdorsata*-Honig von zwei Bienenvölkern in Kampar und Kerinci, gesammelt. Zudem wurden zwei Pollenfallen für die Periode eines Jahres eingerichtet, eine in einer Urwald-Gummi-Plantage und eine im tropischen Regenwald.

Die PCR-Amplifizierung und Sequenzierung wurde mit Hilfe von *rbcL*- und *matK*-Markern für alle Proben, welche als zur Familie der *Dipterocarpaceae* gehören, durchgeführt. Bei den Honig- und Pollenproben wurden zur Amplifizierung lediglich *rbcL*-Marker eingesetzt.

Jede abgerufene Sequenz wurde mithilfe der Software „CodonCode Aligner“ (CodonCode Corporation, https://www.codoncode.com/aligner/) und unter Berücksichtigung der von der CBOL Plant Arbeitsgruppe (2009) vorgeschlagenen Leitfäden aligniert. Multiple Alignierungen wurden mit dem Konsensus aller alignierter Proben je genetischen Marker, d.h. *rbcL* und *matK*, mit den im CodonCode Aligner inbegriffenen Alignierungsalgorithmen, durchgeführt. Zudem wurde aus den multiplen Alignierungen von rbcL und matK eine zusammengeführte Alignierung erstellt.

Die Ergebnisse der multiplen Alignierungen wurden in die Software „MEGA-X“ (Kumar et al., 2018) exportiert um phylogenetische Untersuchungen mit Methoden der „Neighbor Joining“, „Maximum Parsimony“ und „Maximum Likelihood“ durchzuführen. Zusätzlichen zu einem phylogenetischen Baum je Marker (*rbcL* und *matK*) wurde ein zusammengeführter phylogenetischer Baum erstellt.

Die Strukturen der erstellten phylogenetischen Bäume war über die verschiedenen phylogenetischen Erstellungsmethoden generell ähnlich. Zunächst halfen die phylogenetischen Bäume dabei fünf problematische Proben, welche eine abgeschiedene Gruppe außerhalb der Proben der *Dipterocarpaceae* bildeten, zu verwerfen.

Die Kombination der beiden genetischen DNA-Barcode-Marker (*rbcL* und *matK*) zeigte sich gegenüber der Nutzung von Einzellokus-DNA-Barcodes bei der Unterscheidung von Taxa auf der Ebene von Arten deutlich überlegen. Der *rbcL* Marker zeichnet sich durch seine

Universalität aus, wohingegen der *matK* Marker als wirkungsvolles Unterscheidungsinstrument bei phylogenetischen Untersuchungen gilt.

Trotz der eher niedrigen Auflösung zeigten die phylogenetischen Bäume, welche für Honig, Pollen und Baumproben erstellt wurden und auf dem *rbcL*-Marker basieren, eine geringe Unterscheidung auf der Familien- und Gattungsebene. Dies deutet einen positiven Nutzen bei phylogenetischen Vielfaltsmaßen an.

1. INTRODUCTION

Since the 1992 Earth Summit in Rio de Janeiro, different political and non-governmental actors have been interested in nature as a whole, the amount of total species living on every ecosystem, the impacts caused by human development, ecosystems functions and dynamics and especially, how the biodiversity loss can impact on the environment and human activities (Cardinale et al., 2012). Nevertheless, the general perception on biodiversity is the tendency of increased rate loss, which led to global leaders, through the Convention on Biological Diversity (CBD) in 2002, to achieve by 2010 a significant reduction of biodiversity rate loss. Furthermore, according to Butchart et al. (2010), biodiversity has still continued to decline for at least the past four decades, with declines in population trends of vertebrates and extension of forests and coral reefs.

Three main goals were set within the framework of the CBD, to sustain the conservation of biological diversity, the sustainable use of its components and the wise and fair use of the benefits from the use of genetic resources (Secretariat of Convention on Biological Diversity, 2007). In order to fulfill the set of goals and hence, provide a sustainable biodiversity for future generations, it is important to asset and identify what the biodiversity is and how can be measured. Generalizing, biodiversity is defined as the variety of life, which includes variation among genes, species identification and functional traits and it can be measured from three perspectives, richness (number of unique life forms), evenness (equitability among life forms) and heterogeneity (dissimilarity among life forms) (Cardinale et al., 2012).

Therefore, it is crucial to at least acknowledge the richness of species belonging to every ecosystem because, as it follows in the Darwin Declaration, the success of the CBD lays on the expertise to surpass taxonomic impediment. This term refers to the noticeable gap of knowledge in the taxonomic system, including genetic systems, taxonomic experts, and the need to strengthen taxonomic infrastructure to manage, discover and understand the biodiversity's relationships (Environment Australia, 1998).

What is taxonomy and why is so important? The taxonomy science is the responsible of naming, describing and classifying all living organisms, including microorganism from all the world. Naming a species reflects its real biological difference, which is the potential of a group of organisms to interbreed and produce viable offspring that in turn can also interbreed themselves (Secretariat of Convention on Biological Diversity, 2007). By observing morphological, behavioral, genetic, biochemical, among others features' organisms, taxonomists sort the specimens into some classifications and check (reading descriptions of known species, comparing herbaria vouchers and/or museums) whether or not they already have names, giving a unique dual name in Latin format to any new specimen that has not been scientifically described before. The new species then is described including the observed features inherent from the organism, paths to distinguish from others or special characteristics. Thus, this process ensures that a species is properly referred to when talking about it regarding of common names or given names to organisms in different languages. Sometimes, the comparison process involves additional parameters to better differentiate the unique features of a species, such as dissections, environmental adaptations or even molecular analyses of DNA (Secretariat of Convention on Biological Diversity, 2007).

There have been estimates suggesting that all living organisms in Earth account to up to 100 million, including described and undescribed species. However, a recent estimate model suggests there are around 8.7 million species (±1.3 million species) from which only 1.4 million species have been morphological described after 250 years of taxonomic classifications (Mora et al., 2011; Guiry, 2012). Particularly, the clade of green plants or *Viridiplantae* is likely comprised of 500.000 species dominating terrestrial and aquatic environments. Despite of uncertainty in the relationships from this ecological and economically important group, the new advances and progress on molecular and paleobotanical research has given some understanding of the evolution of *Viridiplantae* to the current living species of plants (Gitzendanner, 2018).

One of those new advances is the current increase of DNA sequencing and computational technologies, giving new information for understanding evolutionary and genetic relationships (Hajibabaei, 2007). From this perspective, the idea to identify and promote the use of specific DNA sequence(s) for identification of any given species, as quickly as possible, was born and named DNA barcode (Naciri et al., 2012). Nevertheless, despite of being able to identify specimens to a species level, DNA barcoding must not be seen as a replacement of taxonomic analysis. Instead, barcoding can provide aid to the taxonomic framework by highlighting atypical specimens or genetically divergent groups. For example, when the barcode analysis is not able to classify certain sample to a known species or genera, more extensive taxonomic analysis is needed for this sample rather than describing it as new species with only the barcoding report. This process is advantageous since it can facilitate the task of identifying and describing new species with the conventional taxonomic approach (Hajibabaei, 2007).

Even though the *Dipterocarpaceae* family is of great biological and economic importance in vast forests areas from Southeast Asia, as well as being one of the most notorious trees in the tropics, it is constantly threatened by deforestation and land use changes (CIFOR, 1998). *Dipterocarpaceae* consists of approximately 695 species from two subfamilies, *Dipterocarpoideae* from tropical Asian forests with 470 species in 13 genera and *Monotoideae* from Africa with 40 species in three genera, including the monotypic *Pseudomonotes* genus (Ashton, 1982, according to Indrioko et al., 2006; Londoño et al., 1995; CIFOR, 1998; Christenhusz and Byng, 2016). Formerly, the monotypic *Pakaraimoideae* genus was included in *Monotoideae* subfamily, but recently moved to the *Cistaceae* family (APG, 2016).

The phylogenetic placement of this family within the angiosperms has been problematic for a long time, ranging from placements in the Theales order or in Malvales order (Dayanandan et al., 1999). Despite of keeping the family *Dipterocarpaceae* in the Malvales order, the Angiosperm Phylogeny Group (2016) has considered to combine *Cistaceae*, *Dipterocarpaceae* and *Sarcolaenaceae* families into one, but until more extensive and comprehensive studies of taxa have been concluded from this group of families, they abstain from making any additional changes on their phylogenetic arrangement.

Furthermore, the threatened genetic richness and diversity of the *Dipterocarpaceae* family can only be preserved by not only conserving and creating protected areas, but also understanding the evolutionary processes, relationships, origin and evolution of its inner and higher taxonomic levels (CIFOR, 1998). Since molecular phylogeny is able to provide the knowledge on evolutionary histories and relationships of any living species through DNA sequencing (Avise, 2006), it is possible to broaden the comprehension on the *Dipterocarpaceae* evolutionary processes for ensuring its conservation. Besides, this family has already been subject of great interest in different molecular phylogenetic studies, ranging from plastid DNA sequences to the use of six plastid markers for a provisional phylogeny framework (Heckenhauer et al., 2018). Hence, molecular phylogeny based on plant DNA barcode, with two genetic plastid region markers, is used in this study as a new angle to understand *Dipterocarpaceae* phylogenetic relationships.

1.1. OBJECTIVES

- The aim of this study is to analyze the phylogenetic relationships of *Dipterocarpaceae* samples collected in Jambi, Indonesia by using plant DNA barcoding with two genetic markers (*rbcL* and *matK*).
- To compare the constructed phylogenetic trees with concatenated genetic markers (*rbcL* and *matK*) against single genetic marker (*rbcL* or *matK*) phylogenetic trees, using different phylogenetic tree construction methodologies, i.e. Neighbor Joining, Maximum Parsimony and Maximum Likelihood.
- To interpret the relationships between pollen and honey samples with the *Dipterocarpaceae* collected samples, only using the *rbcL* genetic marker.

1.2. HYPOTHESES

- The use of plant DNA barcode with two concatenated genetic markers (*rbcL* and *matK*) provides usefulness on identifying phylogenetic relationships among taxa.
- Phylogenetic trees can be used to assess phylogenetic diversity of pollen and honey samples.

2. MATERIALS AND METHODS

2.1. Study Area

The study was carried out as part of the Collaborative Research Centre 990: Ecological and Socio-economic Functions of Tropical Lowland Rainforest Transformation Systems project (CRC990: EFForTS project, https://www.uni-goettingen.de/crc990), in Jambi Province, Indonesia. In two regions, the 'Bukit Duabelas landscape' and the 'Harapan landscape' (Figure 1), four different land use systems, lowland rain forest, jungle rubber, rubber monoculture plantation and oil palm monoculture were compared to evaluate the biodiversity and ecological functions of transformed rain forest systems. Within the CRC990 framework, in the Z02 project, a barcoding system was established for the study sites to support species identification of vascular plants (EFForTS project). The overall region is characterized by an average temperature of 26.7 °C and a mean annual precipitation of approximately 2235 mm (Drescher et al., 2016).

2.2. Study Design and Specimen Collection

In the two landscapes a total of four core plots were established per land use system, for a total of 32 sampling plots with a size of 50 m x 50 m. Within each plot, five subplots of 5 m x 5 m were selected randomly. In all core plots, specimen of trees with diameter at breast high (DBH) ≥ 10 cm were identified and collected from the whole core plot area. In the subplots, all vascular plant individuals (shrubs, lianas, seedlings and overall understory vegetation) were identified, sampled and measured in height (F. Amandita, 2015; Rembold et al., 2017).

For each species found, herbarium specimens of three individuals were collected, stored and prepared for later morphological identification at Indonesian Herbaria (Herbarium Bogoriensis and BIOTROP Herbarium). Additionally, a leaf tissue of approximately 2 cm^2 was collected from each sample and dried in silica-gel until DNA extraction. For DNA extraction and further analyses, the material was shipped to the Forest Genetics and Forest Tree Breeding Department, Faculty of Forest Sciences and Forest Ecology, Georg-August-Universität Göttingen, Germany.

Collected herbarium specimens were cross-referenced with the available specimens at Indonesian herbaria and identified by species, genus and/or family level by associated taxonomists. Morphological identification of samples was then compared to DNA Barcoding identification. By using DNA barcoding for species identification, the morphological identification made by the taxonomists can be corroborated, also giving insight for the cases where the species identification was not possible or insufficient.

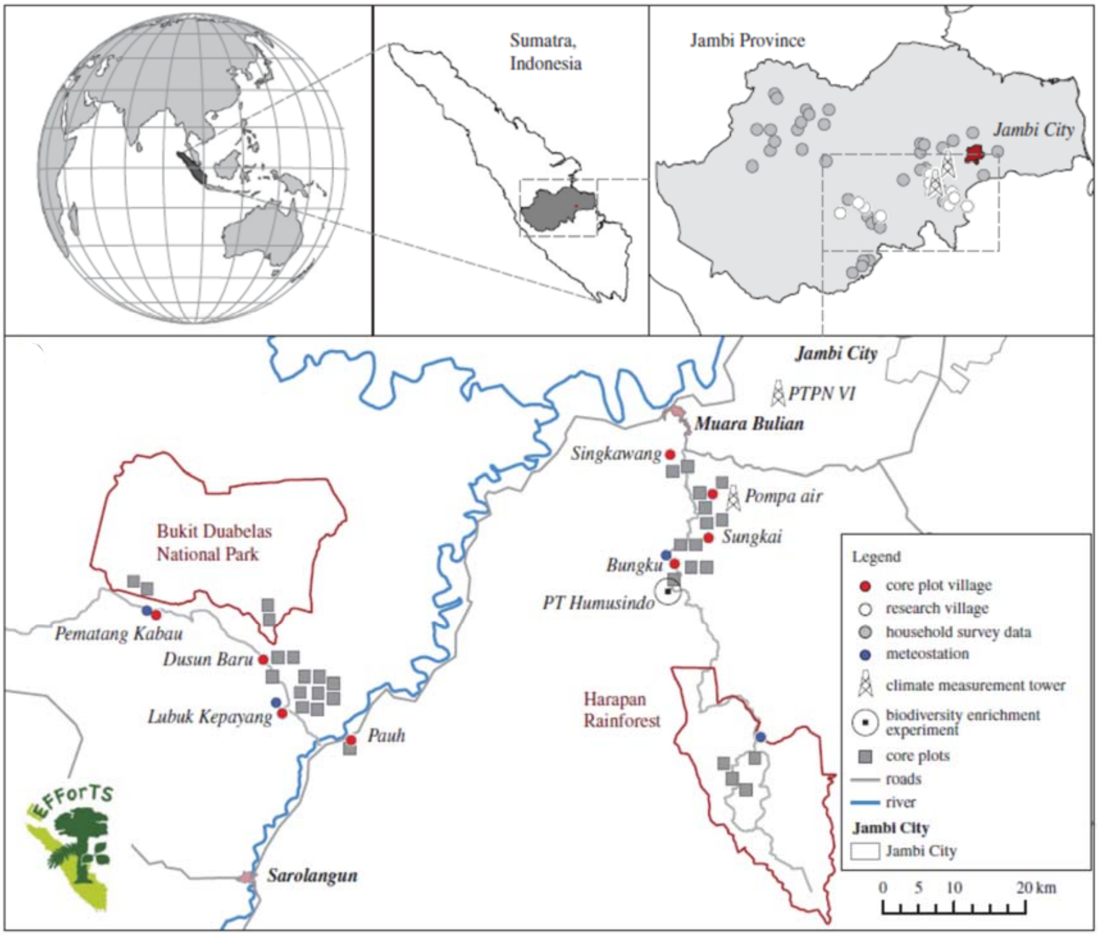


Image No. 1 Location of 32 plots in two study regions in Jambi Province, Sumatra, Indonesia, named the 'Bukit Duabelas landscape' and the 'Harapan landscape' after the respective national park in each region (Drescher et al. 2016).

In the analysis of honey samples, *Apisdorsata* honey was collected from two bee colonies; Kampar colony 1 (sample 8068) is located on a remnant forest surrounded by plantations of Eucalyptus, oil palm (*Elaeis guineensis*) and some Acacia, and Kerinci colony 1 (sample 8406) is located in a secondary forest partially surrounded by a pristine forest, agricultural land and few patches of oil palm plantations.

Additionally, pollen traps were established in two different plots over a period of one year, in a jungle rubber plantation (sample NA 20) and in a tropical rain forest (sample NA 30).

Pollen traps (Behling, Cohen, and Lara, 2001; Jantz, Homeier, and Behling, 2013), using 50 ml PVC test tubes with 3 ml glycerin and synthetic cotton covered with a fine mesh, were installed with a stick about 10 cm above the ground to collect the pollen rain from the surrounding vegetation.

Pollen was extracted from the pollen traps by centrifuging and sieving the samples (2 mm and 200 µm). Afterwards, 1 tablet of Lycopodium spores was added to each sample. A solution of 10% HCl was added to dissolve the tablet (Faegri and Iversen, 1989). The pollen residue obtained was kept in distilled water.

2.3. DNA Extraction

For the dried leaf tissue, DNA extraction was performed 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 1x TAE buffer solution, and 4 µl Roti-Safe dye. Later, the band patterns were visualized on a UV trans-illuminator. Following the manufactory's protocol for innuPREP Gel Extraction Kit protocol (analytikjena, Jena, Germany), DNA fragments for each sample were then isolated and purified from the electrophoresis agarose gel with a volume of 13 µl Elution Buffer (innuPREP Gel Extraction Kit).

Whereas in the case of the honey and pollen samples, DNA was isolated using a modified protocol of the DNeasy 96 Plant Mini Kit (Qiagen, Hilden, Germany) in line with modifications described by de Vere et al. (2017).

Pollen samples, from pollen traps, were washed three times with 200 µl of sterile water and centrifuged (Eppendorf Centrifuge 5424) at 20.000 rpm. The pellet was resuspended in 400 µl AP1 buffer (DNeasy 96 Plant Mini kit) and 80 µl proteinase K (1 mg/ml) and incubated at 65°C for one hour in a water bath. Two 3 mm tungsten carbide beads were added to each tube to disrupt pollen grains in a mixer millMM 300 (Retsch, Haan, Germany) at 30 Hz for 4 minutes.

For honey samples, 30 ml sterile water was added to 10 g honey and incubated overnight at 65 °C. Samples were then centrifuged (Eppendorf Centrifuge 5810R) for 30 min at 3.700 rpm. The supernatant was discarded and samples were lyophilized in a Christ Alpha 1-2 LD plus freeze dryer (Christ, Osterode, Germany). The resulting pellets were resuspended in

400 µl AP1 buffer (DNeasy 96 Plant Mini kit) and 80 µl proteinase K (1 mg/ml) and further processed as described for the pollen traps.

After extraction, the DNA was diluted 1:10 with ddH$_2$O and stored at -20°C for further processing.

2.4. Polymerase Chain Reaction (PCR) Amplification and Sequencing

2.4.1. Leaf Samples

For each extracted DNA sample, Polymerase Chain Reaction (PCR) was carried out using universal primers for the plastid regions *rbcL* and *matK* (Table 1). Nevertheless, the primers "MatK new F" and "MatK new R", designed in the Forest Genetics and Tree Breeding Department (Georg-August-Universität Göttingen) by Dr. Barbara Vornam, were used when the PCR amplification resulted in low success rates with the *matK* primers recommended by Ki-Joong Kim.

Table 1 Details *rbcL* and *matK* plastid regions.

Region	Primer name	Sequence Orientation (5´→ 3´)	Reference
matK	KIM1R_f	ACCCAGTCCATCTGGAAATCTTGGTTC	Ki-Joong Kim, unpublished
	KIM3F_r	CGTACAGTACTTTTGTGTTTACGAG	Ki-Joong Kim, unpublished
	MatK new F	GTTCAAACTCTTCGCTACTGG	Forest Genetics and Tree Breeding, Georg-August-Universität Göttingen
	MatK new R	GAGGATCCACTGTAATAATGAG	
rbcL	rbcLa_f	ATGTCACCACAAACAGAGACTAAAGC	Kress and Erickson, 2007
	rbcLa_r2	GAAACGGTCTCTCCAACGCAT	Fazekas et al., 2008

The use of a core barcode for plants with two plastid regions, *rbcL*+*matK*, was proposed by the Plant Working Group of the Consortium for the Barcoding of Life (CBOL, 2009) due to the easy recovery of high-quality sequences for *rbcL* and the discriminatory power of *matK* (CBOL Plant Working Group, 2009). Besides, the combination of *rbcL*+*matK* plastid regions showed an overall increased resolution at the species level than the use of multi locus marker barcodes, single markers or other dual markers (Hollingsworth et al., 2011a).

PCR was performed in a Peltier Thermal Cycler PTC-200 (MJ Research Inc.) 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_2O.

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 gel, excised from the gel and purified with the innuPREP Gel Extraction Kit protocol (analytikjena, Jena, Germany).

Sequencing reactions were done with the BrilliantDye v3.1 Terminator Cycle Sequencing Kit optimized for Dye Set Z (NIMAGEN, Nijmegen, 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. The total sequencing reaction mixture included 2 µl DNA template (5 – 10 ng), 4,5 µl ddH_2O, 0.5 µl BrilliantDye v3.1, 2 µl 5x Sequencing Buffer, 1 µl Forward/Reverse primer (5 pM/µl) for a total of 10 µl volume reaction.

The PCR sequencing program was set for an initial denaturation at 96 °C for 1 min, followed by 34 cycles of denaturation at 96 °C for 10 secs, annealing at 45 °C for 10 secs and elongation at 60 °C for 4 min. Sequencing results were detected using an ABI Prism Genetic Analyzer 3130xl with the Sequence Analysis v5.3.1 software (Applied Biosystems, Foster City, USA).

2.4.2. Honey and Pollen Samples

Using the same universal primer pair described in Table 1 for *rbcL*, the barcoding region *rbcL* was amplified. PCR reactions and program followed the same guidelines as the ones stated with the dried leaves.

PCR products were separated on 1.5% agarose gels, excised from the gel and purified with the innuPREP Gel Extraction Kit protocol (analytikjena, Jena, Germany). The purified PCR products were cloned into a pCR™4-TOPO® vector using a TOPO® TA Cloning® Kit

(Invitrogen, Carlsbad, USA). For each sample, *rbcL* sequences were obtained for at least five clones using colony PCR with M13 forward and reverse primers.

As for the leaf samples, sequencing reactions were done with the BrilliantDye v3.1 Terminator Cycle Sequencing Kit optimized for Dye Set Z (NIMAGEN, Nijmegen, Netherlands), purified following the manufacturer's protocol of DyeEx® 96 Kit (Qiagen, Hilden, Germany) and the results were obtained with an ABI Prism Genetic Analyzer 3130xl (Applied Biosystems, Foster City, USA).

2.5. Data Analysis

Each sequence retrieved from sequencing was visualized and aligned using the CodonCode Aligner software (CodonCode Corporation, https://www.codoncode.com/aligner/). Forward and reverse sequences were checked to have a length of more than 100 base pairs (bp) with a minimum average Quality Value (QV) of 30, segments at the beginning and end of each sequence, with more than 2 bp of <20 QV, were manually trimmed for each sample taking into account that post-trim lengths are >50% of the original length, and that >50% overlap of the forward and reverse sequences is in the assembled alignment (CBOL Plant Working Group, 2009). If necessary, mismatches between the aligned sequences were manually checked and edited, with the help of the traces visualization per sequence.

A multiple alignment was carried out with the consensus of all the aligned samples per marker, i.e. *rbcL* and *matK*, using CodonCode Aligner's built-in alignment algorithms. The resulting multiple alignments were then trimmed to have the same length across all samples and, when necessary, manually edited by deleting/adding gaps in the sequences for achieving better matches. BLAST searches (https://blast.ncbi.nlm.nih.gov/Blast.cgi) were performed to identify best matches of the samples in the National Center for Biotechnology Information (NCBI) GeneBank database (https://www.ncbi.nlm.nih.gov/) and Barcode of Life Data Systems (BOLDSYSTEMS) database (http://www.boldsystems.org/index.php/) to include them in the multiple alignment.

Furthermore, a concatenated alignment was created from the multiple alignments of *rbcL* and *matK* markers using the SequenceMatrix v1.8 software (Vaidya et al., 2011).

2.6. Phylogenetic Analysis

Currently, there are two major approaches for estimating phylogenetic trees, algorithmic and tree-searching. The algorithmic approach is fast and estimates a single tree from a dataset based on an algorithm. The tree-searching concept estimates several trees from which a final or best tree (or set of trees) is determined by evaluating each individual possible tree according to a criterion. However, this process could be impractical and exhaustive with increasing number of taxa. Therefore, a branch-addition algorithm is used for searching the best possible tree or set of trees (Hall, 2018).

It should also be taken into account that there are two more additional methodologies when referring to phylogenetic analysis estimations, distance methods and character-based methods.

Distance methods convert the multiple aligned sequences into a matrix of distances between the sequences. Branch lengths and order are computed based on this matrix. In distance methods, distances are understood as the fraction of sites that differ between two sequences. On the other hand, character-based methods compare each site within each column (character) from the whole multiple alignment directly, without converting it into something else (Hall, 2018).

Results of the multiple alignments were exported to the MEGA-X software (Kumar et al., 2018) for the phylogenetic analyses with Neighbor Joining, Maximum Parsimony and Maximum Likelihood methods. A phylogenetic tree for each marker (*rbcL* and *matK*), and an additional concatenated tree, was generated per chosen method.

In the construction of phylogenetic trees, it is common to use an outgroup for the dataset given. This outgroup taxon (or set of taxa) is closely related to the analyzed taxa, also referred as ingroup, in having a more ancient common ancestor with the ingroup than the most recent common ancestor of the ingroup. The outgroup is traditionally used as a point of reference for unrooted phylogenetic construction, giving a root for the ingroup (Felsenstein, 2004, according to Drummond and Bouckaert, 2015, p. 98).

The subfamily *Monotoideae*, of the *Dipterocarpaceae* family, was chosen as an outgroup and included in the multiple alignments prior the phylogenetic analysis for each phylogenetic method.

2.6.1. Neighbor Joining

The algorithmic method Neighbor Joining is a distance based method. The tree is reconstructed from a series of matrices, by reducing in size the original distance matrix at each step. Instead of constructing clusters, Neighbor Joining directly calculates distances to internal nodes (Hall, 2018). This method is like the minimum-evolution or maximum-parsimony method, since it produces a single tree under a minimum evolution principle. However, it is not guaranteed to produce a maximum parsimony tree in all the estimations with this method. One of the advantages of this method is the efficiency in obtaining the correct tree topology while providing the branch lengths (Saitou and Nei, 1987).

Different evolutionary models can be applied to phylogenetic construction methods, calculating branch lengths and thus, generating several tree topologies according to the selected model. Branch lengths indicate the amount of genetic variation among data and, some models take assumptions about this variation by creating substitutions of a nucleotide for another in different sites. The Jukes-Cantor model is one of the first developed models, it has only one parameter to calculate, the substitution rate, considering the probabilities of any nucleotide changing into another one as equal. The Kimura 2-parameter model takes into account the possibility that the occurrence rates for transversions and transitions mutations can be different. The Tamura-Nei model develops on the previous model by adding a correction for compositional bias and discriminating between transitional substitution rates among purines and transversional substitution rates among pyrimidines. The Maximum Composite Likelihood model is based on the sum of related log-likelihoods from the pairwise distances in the distance matrix, increasing the accuracy of calculating these pairwise distances and implementing a likelihood-based approach of the Tamura-Nei model (Hall, 2018). The selected evolutionary model for Neighbor Joining was Maximum Composite Likelihood, which is also the default preference for this method.

Phylogenetic trees with this method were calculated with 1000 bootstrap replications to test the reliability of the constructed trees. Transitions and transversions were included as substitutions to be considered by the chosen model. The option ´pattern among lineages´ refers to the assumption of homogeneity in the substitution patterns among lineages and it was set as homogeneous, which is the default selection, and ´rates among sites´ is used to include a rate variation among sites in the model, again the default option was selected, uniform rates. The treatment selected for the gaps and missing data was set as pairwise deletion, i.e. all sites with missing data are initially retained, excluding them as needed when the pairwise distance is estimated.

2.6.2. Maximum Parsimony

The tree-searching method Maximum Parsimony is a character-based method. This method is characterized by the principle of minimum evolution, meaning that the most likely tree is the one with the smallest number of nucleotide substitutions required to explain the evolutionary changes of the data (Saitou and Imanishi, 1989). A first assumption is established when using this method, all taxa sharing a common character have inherited that character from a common ancestor. However, some explanations are needed when conflicts occur with that assumption: Reversal which refers to a character that has changed but then reverted back to its original state; Convergence is when unrelated taxa have evolved the same character independently; or Parallelism, which is the case when different taxa may have similar features predisposing the development of a character in a certain way. All of these explanations are additional steps or hypotheses to explain the data, and can be referred to as homoplasies (Hall, 2018).

Maximum Parsimony choses the tree with the minimum number of evolutionary steps (including homoplasies) required to explain the aligned sequences. It should be noted that under this method, a common character is more likely to be inherited from a common ancestor than because of homoplasy explanations, i.e. Reversal, Convergence or Parallelism. Since this is a character-based method, each site in the alignment is a character and not all of these characters are useful for the tree construction. Some characters are the same in all taxa, providing no information to the tree construction and hence, ignored by the method. Characters that only vary in one taxon are also ignored (Hall, 2018).

Instead of selecting an evolutionary model to estimate phylogenetic trees, a Maximum Parsimony Search Method is chosen to implement Parsimony. After all, this method searches for the minimum number of steps, the most parsimonious tree, to explain the data and does not need to calculate distance matrices or genetic variation as in the Neighbor Joining's tree estimation.

There are four choices for search methods in the program, Subtree-Pruning-Regrafting (SPR), Tree-Bisection-Reconnection (TBR), Min-Mini Heuristic and Max-mini Branch-&-bound. The chosen Search Method was Subtree-Pruning-Regrafting (SPR), which is one of the fastest among the methods. The algorithm of this method reduces the number of tree topologies by detaching a subtree from the current best tree and regrafting it later onto another branch, creating a new topology with a new likelihood value. This process is constantly repeated with all the regrafting positions producing new topologies. If one of those new topologies has a better likelihood score than the previous trees, it becomes the new current best tree. Again, the overall process is repeated until no more likelihood score improvements are obtained. To test the reliability of the phylogenetic analysis with this method, 1000 bootstrap replications were set to run. Additionally, the gaps and missing data treatment selected was partial deletion with 95% site coverage cutoff, meaning that sites with a higher percentage of ambiguous sites than 95%, will be removed from the analysis. Default options were used for the initial number of trees for random addition, Maximum Parsimony Search level and Maximum number of trees to retain.

2.6.3. Maximum Likelihood

The tree-searching method Maximum Likelihood is a character-based method. The estimation on this method is based on a statistical inference which finds the evolutionary tree holding the highest probability of evolving the observed data. However, it should be taken into account that the likelihood of a tree is taken as a function of the tree itself and not the data, meaning that the likelihoods for different trees are not summed. Additionally, the likelihood of a tree does not immediately mean is the correct one (Felsenstein, 1981).

When this method is searching for the tree that makes the data most likely, it applies an explicit criterion for comparing different models of nucleotide substitution. In other words, the model tries to find the evolutionary tree that maximizes or explains the probability of observing the given data with an evolutionary model dictating the rates of nucleotide substitutions. The criterion is the probability of observing all of the data at all of the possible sites and is usually expressed as a log likelihood (due to computational easiness on handling those small numbers), and the sum of the log likelihoods for each of the sites is the total log likelihood of the tree. Likewise, the Maximum Likelihood method searches for the tree with the largest log likelihood, since the largest number explains the largest observed data (Hall, 2018).

Most of the options required to handle prior the Maximum Likelihood method are similar to the used in Neighbor Joining and Maximum Parsimony methods. The bootstrap test was selected to test the reliability of the constructed phylogenetic tree with 1000 replicates.

The Hasegawa, Kishino and Yano (HKY) model was chosen as the nucleotide substitution model. The HKY model contemplates different rates of transitions and transversions from the four base nucleotides as well as unequal frequencies (Rambaut and Grassly, 1997). The considered rate of variation among sites (or rates among sites) for this model was the gamma distribution with 5 discrete gamma categories. Gaps and missing data treatment was selected as partial deletion with 95% site coverage cutoff, which was the same as in the Maximum Parsimony method analyses.

When using this method, sometimes it becomes impossible to evaluate all possible trees which leads to the employment of heuristic search methods. Generally, the starting point for finding the best Maximum Likelihood tree is to begin with the construction of a Neighbor Joining or Maximum Parsimony tree (Hall, 2018). There are two available options for heuristic methods, Subtree-Pruning-Regrafting (SPR) and Nearest-Neighbor-Interchange (NNI). The NNI heuristic search method improves the likelihood of a tree by specifying a neighbor relation between two unrooted trees, and then switch their subtrees in order to obtain a higher likelihood tree. Lastly, the initial tree for the Maximum Likelihood tree was kept in default preference, Neighbor Joining (NJ/BioNJ).

3. RESULTS

3.1. LEAF SAMPLES RESULTS

The initial number of samples gathered and correctly identified were 48 from the Dipterocarpaceae family. However, the number of samples with successful PCR amplifications and sequencing for *rbcL* and *matK* were 35, from which only 30 (Table 2) were in line within the accepted quality values.

Table 2 Identified samples after herbarium cross-referencing.

No.	Plot	Land Use System	Sample ID	Species name	Family
1	BF3	Rain Forest	721	*Shorea acuminata*	Dipterocarpaceae
2	BF3	Rain Forest	753	*Shorea acuminata*	Dipterocarpaceae
3	BF3	Rain Forest	754	*Shorea acuminata*	Dipterocarpaceae
4	BF4	Rain Forest	822	*Shorea acuminata*	Dipterocarpaceae
5	BF4	Rain Forest	842	*Shorea singkawang*	Dipterocarpaceae
6	BF4	Rain Forest	865	*Shorea singkawang*	Dipterocarpaceae
7	BF4	Rain Forest	869	*Shorea singkawang*	Dipterocarpaceae
8	BF4	Rain Forest	876	*Shorea singkawang*	Dipterocarpaceae
9	BF4	Rain Forest	885	*Shorea acuminata*	Dipterocarpaceae
10	BF4	Rain Forest	891	*Shorea ovalis*	Dipterocarpaceae
11	BF4	Rain Forest	896	*Shorea parvifolia*	Dipterocarpaceae
12	BF4	Rain Forest	898	*Shorea bracteolata*	Dipterocarpaceae
13	BF4	Rain Forest	901	*Parashorea cf. lucida*	Dipterocarpaceae
14	BF4	Rain Forest	957	*Shorea acuminata*	Dipterocarpaceae
15	BF4	Rain Forest	992	*Parashorea cf. lucida*	Dipterocarpaceae
16	BF2	Rain Forest	2940	*Parashorea lucida*	Dipterocarpaceae
17	BF2	Rain Forest	2941	*Parashorea lucida*	Dipterocarpaceae
18	HF1	Rain Forest	4121	*Shorea acuminata*	Dipterocarpaceae
19	HF1	Rain Forest	4225	*Shorea parvifolia*	Dipterocarpaceae
20	HF1	Rain Forest	4231	*Hopea ferruginea*	Dipterocarpaceae
21	HF1	Rain Forest	4285	*Shorea parvifolia*	Dipterocarpaceae
22	HF2	Rain Forest	4539	*Shorea ovalis*	Dipterocarpaceae
23	HF2	Rain Forest	4551	*Shorea parvifolia*	Dipterocarpaceae
24	HF2	Rain Forest	4565	*Hopea sangal*	Dipterocarpaceae
25	HF2	Rain Forest	4569	*Hopea beccariana*	Dipterocarpaceae
26	HF2	Rain Forest	4591	*Hopea ferruginea*	Dipterocarpaceae
27	HF3	Rain Forest	4807	*Shorea pauciflora*	Dipterocarpaceae
28	HF3	Rain Forest	4967	*Anisoptera costata*	Dipterocarpaceae
29	HF4	Rain Forest	5089	*Hopea ferruginea*	Dipterocarpaceae
30	HF4	Rain Forest	5168	*Dipterocarpaceae sp. 27*	Dipterocarpaceae

Herbarium specimens cross-referencing was carried out with the available specimens at Indonesian herbaria. Most samples were identified to a species level and all belonging to Dipterocarpaceae family. Nevertheless, in the time of processing and writing of this thesis, sample ID 5168 was not possible to identify beyond its initial characterization.

3.1.1. Neighbor Joining Phylogenetic Analyses

The initial tree was created with the sequences for the *matK* marker (Figure 1), the condensed tree with a cutoff higher than 50% shows the stronger relationships amongst samples and can be seen on Appendix 1 for further reference. One of the remarks on this image is that the lower cluster of samples showed no relationship between them and with the rest of the data. Moreover, the branch lengths of this cluster is larger than the ingroup branches, which can be interpreted as having a more evolutionary distance than the rest of samples and explaining its behavior as outgroup of the whole *Dipterocarpaceae* family.

To contrast this later cluster and compare the results of the *matK* marker, the second phylogenetic analysis was generated for the *rbcL* marker (Figure 2). The samples clustered amongst the rest of the *Dipterocarpaceae* samples, the sample ID 4121 (*Shorea acuminata*), despite of not having any resolved relationship with any other sample, the sample performed as an ingroup, meaning it is related to the family tree. In the case of the other problematic sample, sample ID 4591 (*Hopea ferruginea*) it is moderately related to the rest of *Hopea* samples. However, other samples presented an outgroup behavior as in Figure 1, samples ID 721 (*Shorea acuminata*), 753 (*Shorea acuminata*) and 754 (*Shorea acuminata*). This behavior is only present for these samples in Figure 2 and on Appendix 2, the most significant clades for this figure can be consulted.

Furthermore, the concatenated results are shown for comparison and determining if there is a change or more consistent result against the overall results from the single phylogenetic trees for each marker (Figure 3). Once more, the two problematic samples from the *matK* marker (Figure 1) are clustered together, possibly due to the problematic *matK* concatenated section, and behaving as an outgroup from the rest of samples. Additionally, problematic samples from the Figure 2 are showing a high number of evolutionary distances in contrast with the rest of samples. Despite of behaving as an ingroup, the cluster of samples from the *rbcL* marker are more distant than the rest of the taxa. A condensed tree for the concatenation can be found on Appendix 3.

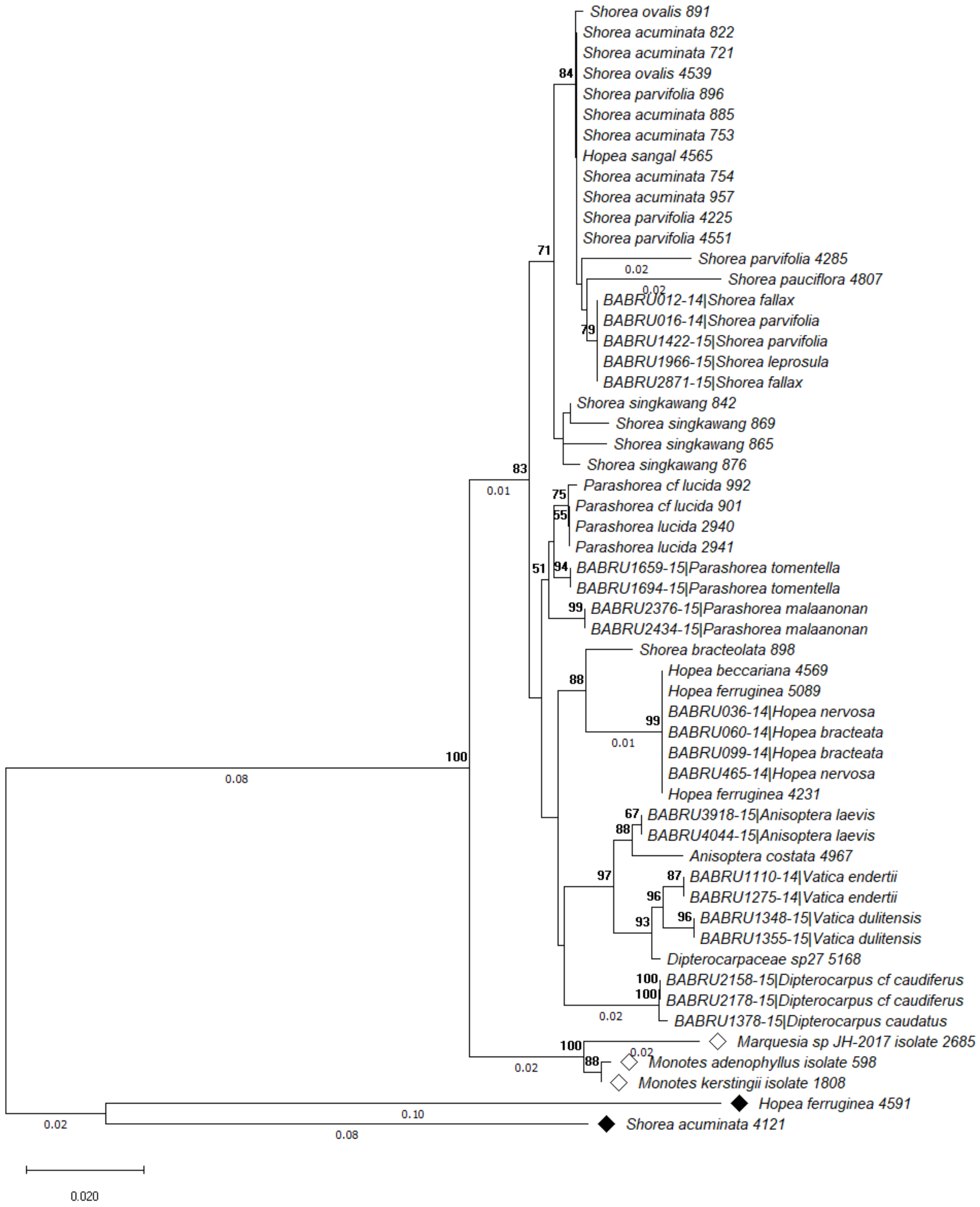

Shorea ovalis 891
Shorea acuminata 822
Shorea acuminata 721
84
Shorea ovalis 4539
Shorea parvifolia 896
Shorea acuminata 885
Shorea acuminata 753
Hopea sangal 4565
Shorea acuminata 754
Shorea acuminata 957
Shorea parvifolia 4225
Shorea parvifolia 4551
71
Shorea parvifolia 4285
0.02
Shorea pauciflora 4807
0.02
BABRU012-14|Shorea fallax
BABRU016-14|Shorea parvifolia
79
BABRU1422-15|Shorea parvifolia
BABRU1966-15|Shorea leprosula
BABRU2871-15|Shorea fallax
Shorea singkawang 842
Shorea singkawang 869
Shorea singkawang 865
Shorea singkawang 876
83
0.01
Parashorea cf lucida 992
75
Parashorea cf lucida 901
55
Parashorea lucida 2940
Parashorea lucida 2941
51
94
BABRU1659-15|Parashorea tomentella
BABRU1694-15|Parashorea tomentella
99
BABRU2376-15|Parashorea malaanonan
BABRU2434-15|Parashorea malaanonan
Shorea bracteolata 898
88
Hopea beccariana 4569
Hopea ferruginea 5089
BABRU036-14|Hopea nervosa
99
BABRU060-14|Hopea bracteata
0.01
BABRU099-14|Hopea bracteata
100
0.08
BABRU465-14|Hopea nervosa
Hopea ferruginea 4231
67
BABRU3918-15|Anisoptera laevis
88
BABRU4044-15|Anisoptera laevis
Anisoptera costata 4967
97
87
BABRU1110-14|Vatica endertii
96
BABRU1275-14|Vatica endertii
96
BABRU1348-15|Vatica dulitensis
93
BABRU1355-15|Vatica dulitensis
Dipterocarpaceae sp27 5168
100
BABRU2158-15|Dipterocarpus cf caudiferus
100
BABRU2178-15|Dipterocarpus cf caudiferus
0.02
BABRU1378-15|Dipterocarpus caudatus
100
◊ Marquesia sp JH-2017 isolate 2685
0.02
88
◊ Monotes adenophyllus isolate 598
0.02
◊ Monotes kerstingii isolate 1808
♦ Hopea ferruginea 4591
0.10
0.02
♦ Shorea acuminata 4121
0.08
0.020

Figure 1 The neighbor joining phylogenetic tree of trees' samples collected in the EFForTS project based on the *matK* marker and genetic distances computed using the Maximum Composite Likelihood method. The numbers at the tree nodes represent bootstrap values based on 1000 replicates. The number after the sample name refers the sample ID. ♦ - samples behaving as a total outgroup from the *Dipterocarpaceae* family; ◊ - the database samples chosen as the outgroup for the analysis**.**

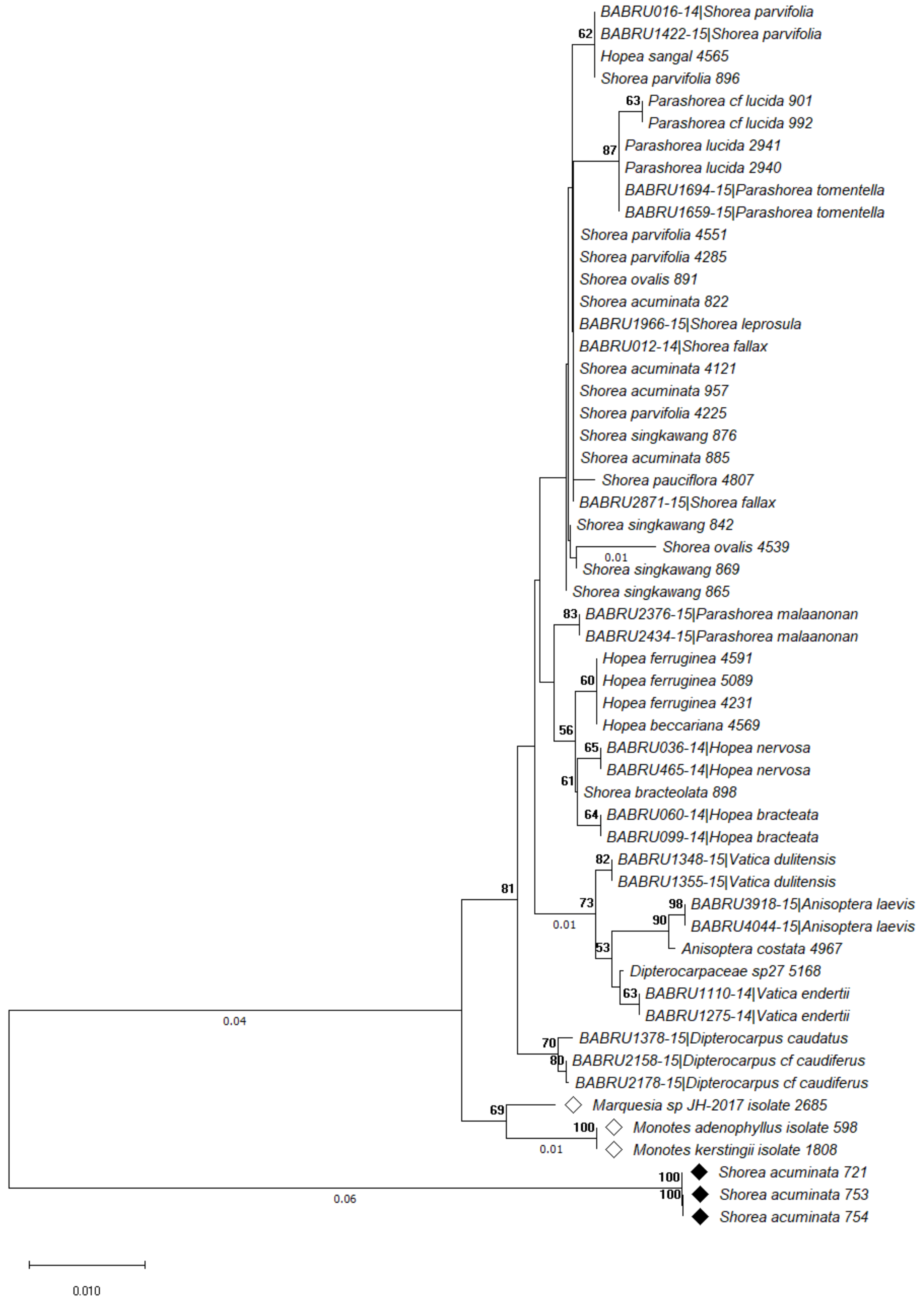

BABRU016-14|Shorea parvifolia
62
BABRU1422-15|Shorea parvifolia
Hopea sangal 4565
Shorea parvifolia 896
63
Parashorea cf lucida 901
Parashorea cf lucida 992
87
Parashorea lucida 2941
Parashorea lucida 2940
BABRU1694-15|Parashorea tomentella
BABRU1659-15|Parashorea tomentella
Shorea parvifolia 4551
Shorea parvifolia 4285
Shorea ovalis 891
Shorea acuminata 822
BABRU1966-15|Shorea leprosula
BABRU012-14|Shorea fallax
Shorea acuminata 4121
Shorea acuminata 957
Shorea parvifolia 4225
Shorea singkawang 876
Shorea acuminata 885
Shorea pauciflora 4807
BABRU2871-15|Shorea fallax
Shorea singkawang 842
Shorea ovalis 4539
0.01
Shorea singkawang 869
Shorea singkawang 865
83
BABRU2376-15|Parashorea malaanonan
BABRU2434-15|Parashorea malaanonan
Hopea ferruginea 4591
60
Hopea ferruginea 5089
Hopea ferruginea 4231
56
Hopea beccariana 4569
65
BABRU036-14|Hopea nervosa
BABRU465-14|Hopea nervosa
61
Shorea bracteolata 898
64
BABRU060-14|Hopea bracteata
BABRU099-14|Hopea bracteata
82
BABRU1348-15|Vatica dulitensis
BABRU1355-15|Vatica dulitensis
81
73
98
BABRU3918-15|Anisoptera laevis
0.01
90
BABRU4044-15|Anisoptera laevis
53
Anisoptera costata 4967
Dipterocarpaceae sp27 5168
63
BABRU1110-14|Vatica endertii
0.04
BABRU1275-14|Vatica endertii
70
BABRU1378-15|Dipterocarpus caudatus
80
BABRU2158-15|Dipterocarpus cf caudiferus
BABRU2178-15|Dipterocarpus cf caudiferus
69
◇ Marquesia sp JH-2017 isolate 2685
100
◇ Monotes adenophyllus isolate 598
0.01
◇ Monotes kerstingii isolate 1808
100
◆ Shorea acuminata 721
0.06
100
◆ Shorea acuminata 753
◆ Shorea acuminata 754
0.010

Figure 2 The neighbor joining phylogenetic tree of trees' samples collected in the EFForTS project based on the *rbcL* marker and genetic distances computed using the Maximum Composite Likelihood method. The numbers at the tree nodes represent bootstrap values based on 1000 replicates. The number after the sample name refers the sample ID. ♦ - samples behaving as a total outgroup from the *Dipterocarpaceae* family; ◊ - the database samples chosen as the outgroup for the analysis.

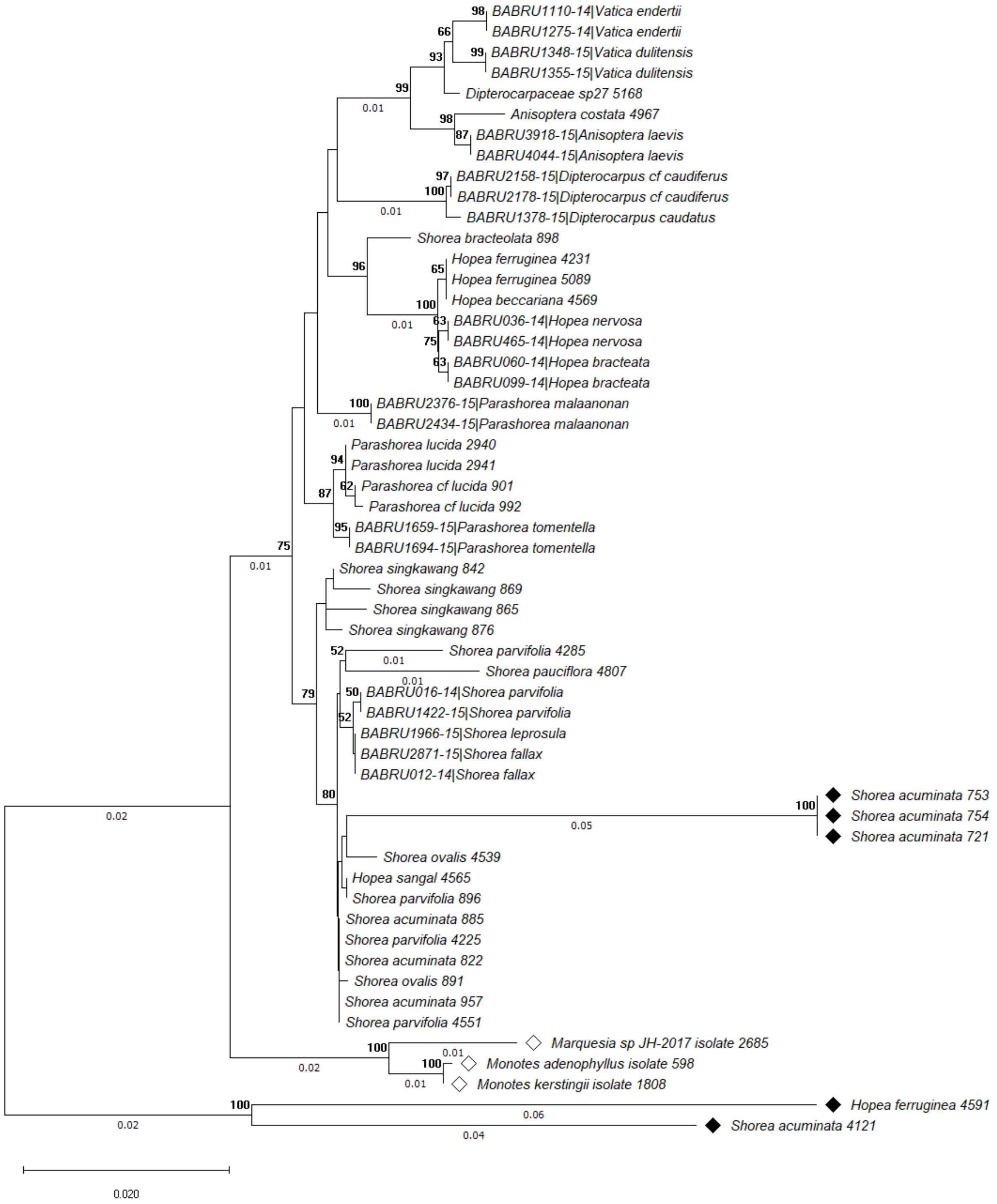
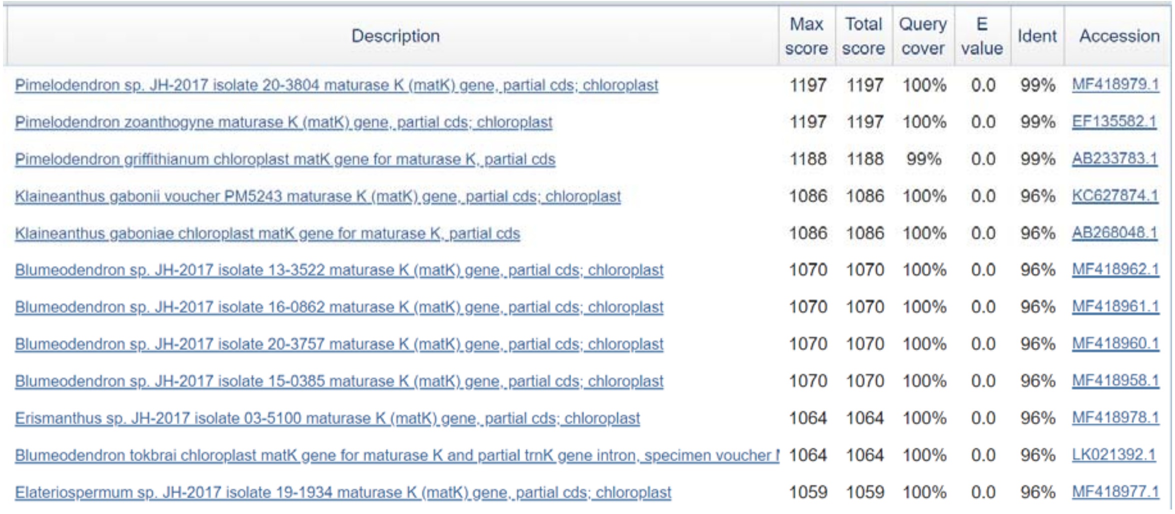

98 BABRU1110-14|Vatica endertii
BABRU1275-14|Vatica endertii
66
99 BABRU1348-15|Vatica dulitensis
93
BABRU1355-15|Vatica dulitensis
99
Dipterocarpaceae sp27 5168
0.01
98
Anisoptera costata 4967
87 BABRU3918-15|Anisoptera laevis
BABRU4044-15|Anisoptera laevis
97 BABRU2158-15|Dipterocarpus cf caudiferus
100
BABRU2178-15|Dipterocarpus cf caudiferus
0.01
BABRU1378-15|Dipterocarpus caudatus
Shorea bracteolata 898
96
Hopea ferruginea 4231
65
Hopea ferruginea 5089
100
Hopea beccariana 4569
0.01
63 BABRU036-14|Hopea nervosa
75
BABRU465-14|Hopea nervosa
63 BABRU060-14|Hopea bracteata
BABRU099-14|Hopea bracteata
100 BABRU2376-15|Parashorea malaanonan
0.01
BABRU2434-15|Parashorea malaanonan
Parashorea lucida 2940
94
Parashorea lucida 2941
62 Parashorea cf lucida 901
87
Parashorea cf lucida 992
95 BABRU1659-15|Parashorea tomentella
75
BABRU1694-15|Parashorea tomentella
0.01
Shorea singkawang 842
Shorea singkawang 869
Shorea singkawang 865
Shorea singkawang 876
52
Shorea parvifolia 4285
0.01
Shorea pauciflora 4807
0.01
79
50 BABRU016-14|Shorea parvifolia
52
BABRU1422-15|Shorea parvifolia
BABRU1966-15|Shorea leprosula
BABRU2871-15|Shorea fallax
BABRU012-14|Shorea fallax
80
100
◆ Shorea acuminata 753
0.02
0.05
◆ Shorea acuminata 754
◆ Shorea acuminata 721
Shorea ovalis 4539
Hopea sangal 4565
Shorea parvifolia 896
Shorea acuminata 885
Shorea parvifolia 4225
Shorea acuminata 822
Shorea ovalis 891
Shorea acuminata 957
Shorea parvifolia 4551
100
◇ Marquesia sp JH-2017 isolate 2685
0.01
0.02
100
◇ Monotes adenophyllus isolate 598
0.01
◇ Monotes kerstingii isolate 1808
100
◆ Hopea ferruginea 4591
0.06
0.02
◆ Shorea acuminata 4121
0.04
0.020

Figure 3 The neighbor joining phylogenetic tree of trees' samples collected in the EFForTS project based on the concatenated *rbcL* and *matK* markers, with genetic distances computed using the Maximum Composite Likelihood method. The numbers at the tree nodes represent bootstrap values based on 1000 replicates. The number after the sample name refers the sample ID. ♦ - samples clustered with problematic behavior; ◊ - the database samples chosen as the outgroup for the analysis.

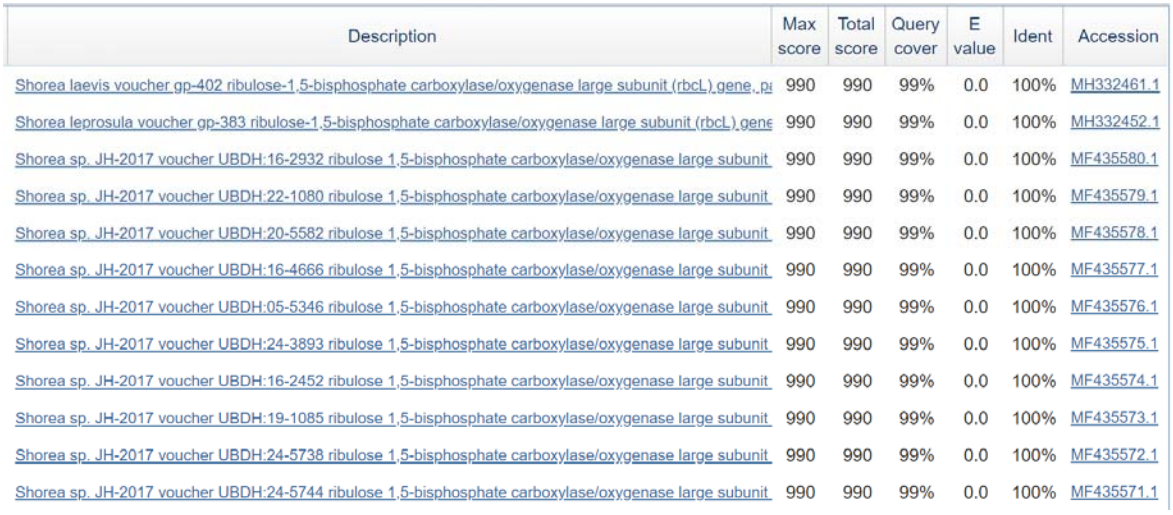
A BLAST search was performed for each of the five problematic samples in order to determine the causes of misplacement or not resolution of the samples. For comparison, the BLAST search was carried out for both markers, *rbcL* and *matK*, of all five problematic samples.



Description	Max score	Total score	Query cover	E value	Ident	Accession
Pimelodendron sp. JH-2017 isolate 20-3804 maturase K (matK) gene, partial cds; chloroplast	1197	1197	100%	0.0	99%	MF418979.1
Pimelodendron zoanthogyne maturase K (matK) gene, partial cds; chloroplast	1197	1197	100%	0.0	99%	EF135582.1
Pimelodendron griffithianum chloroplast matK gene for maturase K, partial cds	1188	1188	99%	0.0	99%	AB233783.1
Klaineanthus gabonii voucher PM5243 maturase K (matK) gene, partial cds; chloroplast	1086	1086	100%	0.0	96%	KC627874.1
Klaineanthus gaboniae chloroplast matK gene for maturase K, partial cds	1086	1086	100%	0.0	96%	AB268048.1
Blumeodendron sp. JH-2017 isolate 13-3522 maturase K (matK) gene, partial cds; chloroplast	1070	1070	100%	0.0	96%	MF418962.1
Blumeodendron sp. JH-2017 isolate 16-0862 maturase K (matK) gene, partial cds; chloroplast	1070	1070	100%	0.0	96%	MF418961.1
Blumeodendron sp. JH-2017 isolate 20-3757 maturase K (matK) gene, partial cds; chloroplast	1070	1070	100%	0.0	96%	MF418960.1
Blumeodendron sp. JH-2017 isolate 15-0385 maturase K (matK) gene, partial cds; chloroplast	1070	1070	100%	0.0	96%	MF418958.1
Erismanthus sp. JH-2017 isolate 03-5100 maturase K (matK) gene, partial cds; chloroplast	1064	1064	100%	0.0	96%	MF418978.1
Blumeodendron tokbrai chloroplast matK gene for maturase K and partial trnK gene intron, specimen voucher I	1064	1064	100%	0.0	96%	LK021392.1
Elateriospermum sp. JH-2017 isolate 19-1934 maturase K (matK) gene, partial cds; chloroplast	1059	1059	100%	0.0	96%	MF418977.1

Image 2 Sequences producing significant alignments for the sample ID 4121 (*Shorea acuminata*) with *matK* marker. Image taken from BLAST webpage: https://blast.ncbi.nlm.nih.gov/Blast.cgi

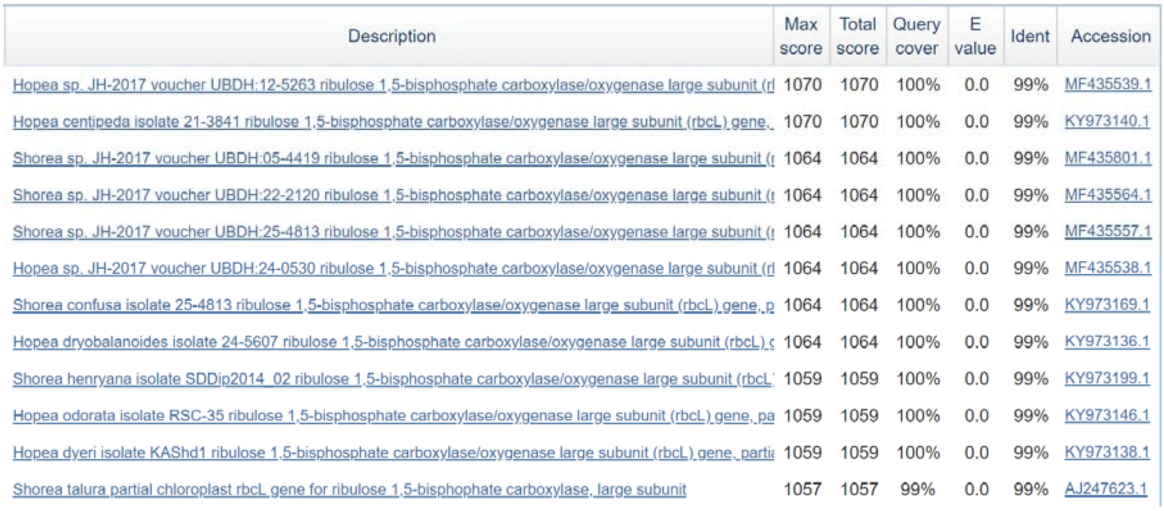
The sample ID 4121 (*Shorea acuminata*) with *matK* marker was the first one to be compared with the BLAST GeneBank database (Image 2). Moreover, all sample sequences belonged to a single family (*Euphorbiaceae*) which is different than the studied one, leading to the cause of misplacement and not resolution on the analyses. In contrast, the *rbcL* marker sequence of this sample showed a correct correlation with the BLAST database, giving a very high value on query cover and a perfect score of identity on the nucleotide level for the *Shorea* genus of *Dipterocarpaceae* family (Image 3).



Description	Max score	Total score	Query cover	E value	Ident	Accession
Shorea laevis voucher gp-402 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, pa	990	990	99%	0.0	100%	MH332461.1
Shorea leprosula voucher gp-383 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene	990	990	99%	0.0	100%	MH332452.1
Shorea sp. JH-2017 voucher UBDH:16-2932 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	990	990	99%	0.0	100%	MF435580.1
Shorea sp. JH-2017 voucher UBDH:22-1080 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	990	990	99%	0.0	100%	MF435579.1
Shorea sp. JH-2017 voucher UBDH:20-5582 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	990	990	99%	0.0	100%	MF435578.1
Shorea sp. JH-2017 voucher UBDH:16-4666 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	990	990	99%	0.0	100%	MF435577.1
Shorea sp. JH-2017 voucher UBDH:05-5346 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	990	990	99%	0.0	100%	MF435576.1
Shorea sp. JH-2017 voucher UBDH:24-3893 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	990	990	99%	0.0	100%	MF435575.1
Shorea sp. JH-2017 voucher UBDH:16-2452 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	990	990	99%	0.0	100%	MF435574.1
Shorea sp. JH-2017 voucher UBDH:19-1085 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	990	990	99%	0.0	100%	MF435573.1
Shorea sp. JH-2017 voucher UBDH:24-5738 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	990	990	99%	0.0	100%	MF435572.1
Shorea sp. JH-2017 voucher UBDH:24-5744 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	990	990	99%	0.0	100%	MF435571.1

Image 3 Sequences producing significant alignments for the sample ID 4121 (*Shorea acuminata*) with *rbcL* marker. Image taken from BLAST webpage: https://blast.ncbi.nlm.nih.gov/Blast.cgi

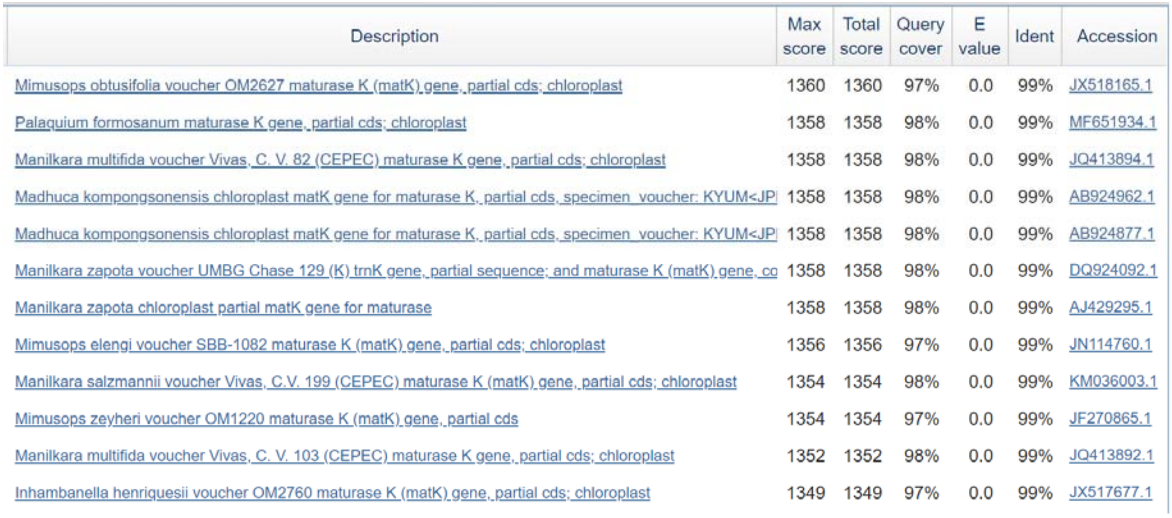
Following, sample ID 4591 (*Hopea ferruginea*) had a very good correlation with the BLAST GeneBank database for the *rbcL* marker. The sequences, with perfect cover and very good identity, were mostly related to *Shorea* and *Hopea* genera. Furthermore, the sample is clustered amongst other *Hopea* samples in Figure 2, confirming the statement the sequence is correctly labeled as a sample from the *Dipterocarpaceae* family (Image 4).



Description	Max score	Total score	Query cover	E value	Ident	Accession
Hopea sp. JH-2017 voucher UBDH:12-5263 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rl	1070	1070	100%	0.0	99%	MF435539.1
Hopea centipeda isolate 21-3841 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene,	1070	1070	100%	0.0	99%	KY973140.1
Shorea sp. JH-2017 voucher UBDH:05-4419 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (r	1064	1064	100%	0.0	99%	MF435801.1
Shorea sp. JH-2017 voucher UBDH:22-2120 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (r	1064	1064	100%	0.0	99%	MF435564.1
Shorea sp. JH-2017 voucher UBDH:25-4813 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (r	1064	1064	100%	0.0	99%	MF435557.1
Hopea sp. JH-2017 voucher UBDH:24-0530 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rl	1064	1064	100%	0.0	99%	MF435538.1
Shorea confusa isolate 25-4813 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, p	1064	1064	100%	0.0	99%	KY973169.1
Hopea dryobalanoides isolate 24-5607 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) g	1064	1064	100%	0.0	99%	KY973136.1
Shorea henryana isolate SDDip2014_02 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL)	1059	1059	100%	0.0	99%	KY973199.1
Hopea odorata isolate RSC-35 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, pa	1059	1059	100%	0.0	99%	KY973146.1
Hopea dyeri isolate KAShd1 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, parti	1059	1059	100%	0.0	99%	KY973138.1
Shorea talura partial chloroplast rbcL gene for ribulose 1,5-bisphophate carboxylase, large subunit	1057	1057	99%	0.0	99%	AJ247623.1

Image 4 Sequences producing significant alignments for the sample ID 4591 (*Hopea ferruginea*) with *rbcL* marker. Image taken from BLAST webpage: https://blast.ncbi.nlm.nih.gov/Blast.cgi

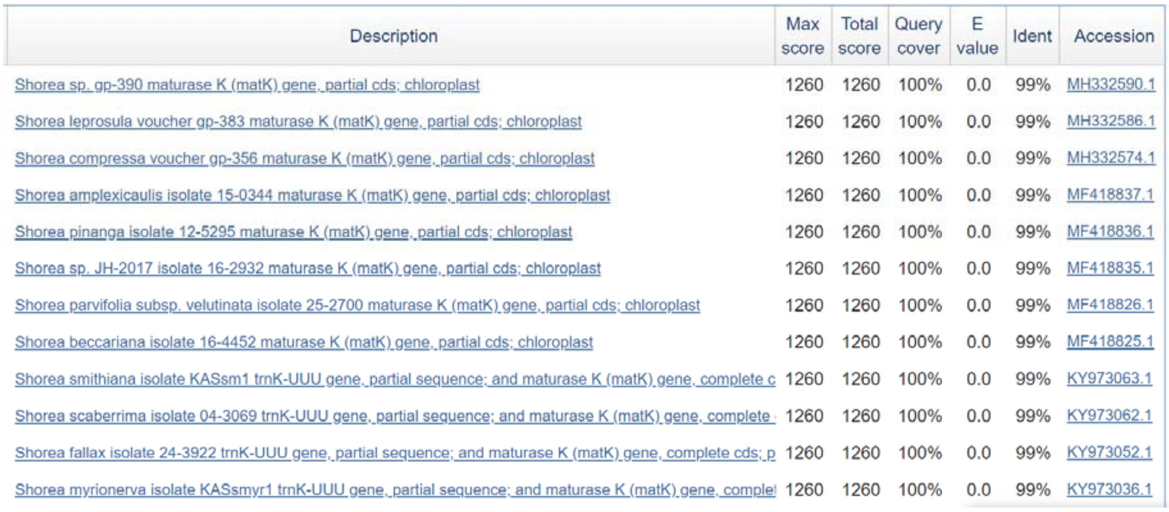
However, when cross-checking the sample ID 4591 (*Hopea ferruginea*) with the *matK* marker (Image 5), the resulting list of sequences from the database belonged to the *Sapotaceae* family. The query cover and identity have enough high values to realize why in Figure 1 the sample ID 4591 is not clustering with any other *Hopea* samples and behaves as a total outgroup from the *Dipterocarpaceae* data set.



Description	Max score	Total score	Query cover	E value	Ident	Accession
Mimusops obtusifolia voucher OM2627 maturase K (matK) gene, partial cds; chloroplast	1360	1360	97%	0.0	99%	JX518165.1
Palaquium formosanum maturase K gene, partial cds; chloroplast	1358	1358	98%	0.0	99%	MF651934.1
Manilkara multifida voucher Vivas, C. V. 82 (CEPEC) maturase K gene, partial cds; chloroplast	1358	1358	98%	0.0	99%	JQ413894.1
Madhuca kompongsonensis chloroplast matK gene for maturase K, partial cds, specimen_voucher: KYUM<JP	1358	1358	98%	0.0	99%	AB924962.1
Madhuca kompongsonensis chloroplast matK gene for maturase K, partial cds, specimen_voucher: KYUM<JP	1358	1358	98%	0.0	99%	AB924877.1
Manilkara zapota voucher UMBG Chase 129 (K) trnK gene, partial sequence; and maturase K (matK) gene, co	1358	1358	98%	0.0	99%	DQ924092.1
Manilkara zapota chloroplast partial matK gene for maturase	1358	1358	98%	0.0	99%	AJ429295.1
Mimusops elengi voucher SBB-1082 maturase K (matK) gene, partial cds; chloroplast	1356	1356	97%	0.0	99%	JN114760.1
Manilkara salzmannii voucher Vivas, C.V. 199 (CEPEC) maturase K (matK) gene, partial cds; chloroplast	1354	1354	98%	0.0	99%	KM036003.1
Mimusops zeyheri voucher OM1220 maturase K (matK) gene, partial cds	1354	1354	97%	0.0	99%	JF270865.1
Manilkara multifida voucher Vivas, C. V. 103 (CEPEC) maturase K gene, partial cds; chloroplast	1352	1352	98%	0.0	99%	JQ413892.1
Inhambanella henriquesii voucher OM2760 maturase K (matK) gene, partial cds; chloroplast	1349	1349	97%	0.0	99%	JX517677.1

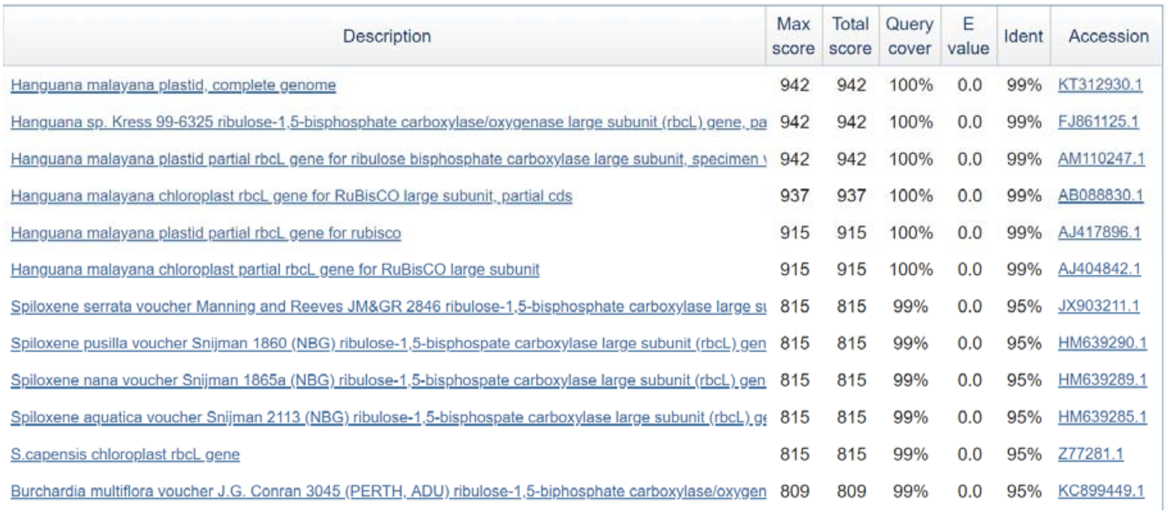
Image 5 Sequences producing significant alignments for the sample ID 4591 (*Hopea ferruginea*) with *matK* marker. Image taken from BLAST webpage: https://blast.ncbi.nlm.nih.gov/Blast.cgi

The previous samples had an outgroup behavior for Figure 1 and Figure 3, and with the help of BLAST searches it is more clear the reason of this outcome. Three more samples had a similar behavior in Figure 2 and a considerable large evolutionary distance in Figure 3. The first of those three samples to have a BLAST search was sample ID 721 (*Shorea acuminata*) with *matK* marker (Image 6). All resulting sequences had a perfect query cover and an identity close to be perfect for *Shorea* genus, belonging to *Dipterocarpaceae* family, which is in correct relation to the sample's label as *Shorea acuminata*. Contrarily, the BLAST search of this sample ID 721 with *rbcL* marker (Image 7) resulted in some sequences belonging to the *Hanguanaceae* family, with perfect query cover and 99% identity on the nucleotide level, and others belonging to the *Hypoxidaceae* family, with an almost perfect query cover and 95% identity. Both resulting families are characterized for being herbs and can be found on South East Asia.



Description	Max score	Total score	Query cover	E value	Ident	Accession
Shorea sp. gp-390 maturase K (matK) gene, partial cds; chloroplast	1260	1260	100%	0.0	99%	MH332590.1
Shorea leprosula voucher gp-383 maturase K (matK) gene, partial cds; chloroplast	1260	1260	100%	0.0	99%	MH332586.1
Shorea compressa voucher gp-356 maturase K (matK) gene, partial cds; chloroplast	1260	1260	100%	0.0	99%	MH332574.1
Shorea amplexicaulis isolate 15-0344 maturase K (matK) gene, partial cds; chloroplast	1260	1260	100%	0.0	99%	MF418837.1
Shorea pinanga isolate 12-5295 maturase K (matK) gene, partial cds; chloroplast	1260	1260	100%	0.0	99%	MF418836.1
Shorea sp. JH-2017 isolate 16-2932 maturase K (matK) gene, partial cds; chloroplast	1260	1260	100%	0.0	99%	MF418835.1
Shorea parvifolia subsp. velutinata isolate 25-2700 maturase K (matK) gene, partial cds; chloroplast	1260	1260	100%	0.0	99%	MF418826.1
Shorea beccariana isolate 16-4452 maturase K (matK) gene, partial cds; chloroplast	1260	1260	100%	0.0	99%	MF418825.1
Shorea smithiana isolate KASsm1 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complete c	1260	1260	100%	0.0	99%	KY973063.1
Shorea scaberrima isolate 04-3069 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complete	1260	1260	100%	0.0	99%	KY973062.1
Shorea fallax isolate 24-3922 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complete cds; p	1260	1260	100%	0.0	99%	KY973052.1
Shorea myrionerva isolate KASsmyr1 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complet	1260	1260	100%	0.0	99%	KY973036.1

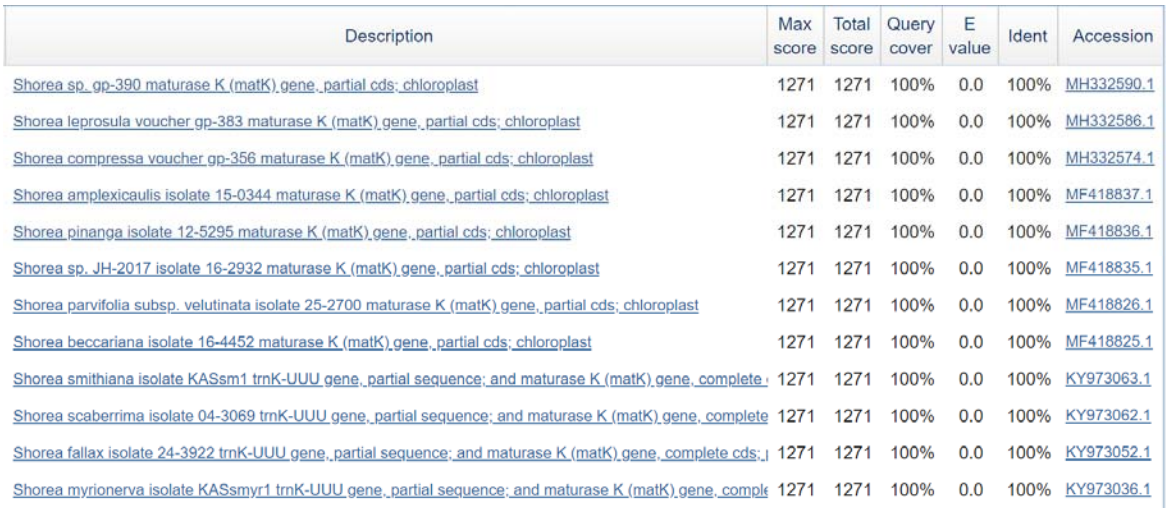
Image 6 Sequences producing significant alignments for the sample ID 721 (*Shorea acuminata*) with *matK* marker. Image taken from BLAST webpage: https://blast.ncbi.nlm.nih.gov/Blast.cgi



Description	Max score	Total score	Query cover	E value	Ident	Accession
Hanguana malayana plastid, complete genome	942	942	100%	0.0	99%	KT312930.1
Hanguana sp. Kress 99-6325 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, pa	942	942	100%	0.0	99%	FJ861125.1
Hanguana malayana plastid partial rbcL gene for ribulose bisphosphate carboxylase large subunit, specimen	942	942	100%	0.0	99%	AM110247.1
Hanguana malayana chloroplast rbcL gene for RuBisCO large subunit, partial cds	937	937	100%	0.0	99%	AB088830.1
Hanguana malayana plastid partial rbcL gene for rubisco	915	915	100%	0.0	99%	AJ417896.1
Hanguana malayana chloroplast partial rbcL gene for RuBisCO large subunit	915	915	100%	0.0	99%	AJ404842.1
Spiloxene serrata voucher Manning and Reeves JM&GR 2846 ribulose-1,5-bisphosphate carboxylase large su	815	815	99%	0.0	95%	JX903211.1
Spiloxene pusilla voucher Snijman 1860 (NBG) ribulose-1,5-bisphospate carboxylase large subunit (rbcL) gen	815	815	99%	0.0	95%	HM639290.1
Spiloxene nana voucher Snijman 1865a (NBG) ribulose-1,5-bisphospate carboxylase large subunit (rbcL) gen	815	815	99%	0.0	95%	HM639289.1
Spiloxene aquatica voucher Snijman 2113 (NBG) ribulose-1,5-bisphospate carboxylase large subunit (rbcL) ge	815	815	99%	0.0	95%	HM639285.1
S.capensis chloroplast rbcL gene	815	815	99%	0.0	95%	Z77281.1
Burchardia multiflora voucher J.G. Conran 3045 (PERTH, ADU) ribulose-1,5-biphosphate carboxylase/oxygen	809	809	99%	0.0	95%	KC899449.1

Image 7 Sequences producing significant alignments for the sample ID 721 (*Shorea acuminata*) with *rbcL* marker. Image taken from BLAST webpage: https://blast.ncbi.nlm.nih.gov/Blast.cgi

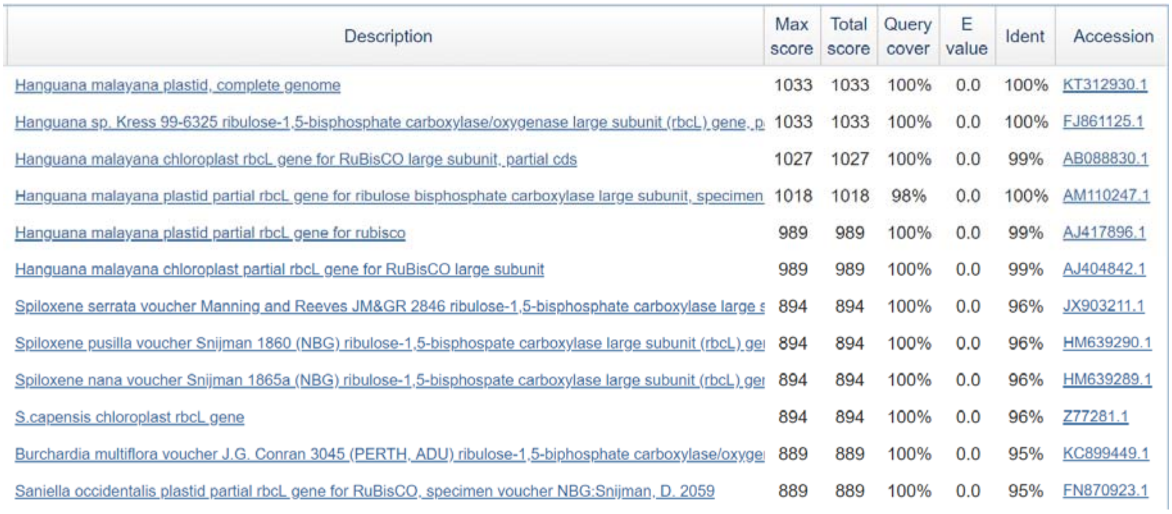
Next, sample ID 753 (*Shorea acuminata*) with *matK* marker was compared with the BLAST GeneBank database (Image 8). The resulting samples were almost the same as for the case of sample ID 721 with *matK* marker, displaying a perfect query cover and perfect identity on the nucleotide level from *Shorea* genus of *Dipterocarpaceae* family.



Description	Max score	Total score	Query cover	E value	Ident	Accession
Shorea sp. gp-390 maturase K (matK) gene, partial cds; chloroplast	1271	1271	100%	0.0	100%	MH332590.1
Shorea leprosula voucher gp-383 maturase K (matK) gene, partial cds; chloroplast	1271	1271	100%	0.0	100%	MH332586.1
Shorea compressa voucher gp-356 maturase K (matK) gene, partial cds; chloroplast	1271	1271	100%	0.0	100%	MH332574.1
Shorea amplexicaulis isolate 15-0344 maturase K (matK) gene, partial cds; chloroplast	1271	1271	100%	0.0	100%	MF418837.1
Shorea pinanga isolate 12-5295 maturase K (matK) gene, partial cds; chloroplast	1271	1271	100%	0.0	100%	MF418836.1
Shorea sp. JH-2017 isolate 16-2932 maturase K (matK) gene, partial cds; chloroplast	1271	1271	100%	0.0	100%	MF418835.1
Shorea parvifolia subsp. velutinata isolate 25-2700 maturase K (matK) gene, partial cds; chloroplast	1271	1271	100%	0.0	100%	MF418826.1
Shorea beccariana isolate 16-4452 maturase K (matK) gene, partial cds; chloroplast	1271	1271	100%	0.0	100%	MF418825.1
Shorea smithiana isolate KASsm1 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complete	1271	1271	100%	0.0	100%	KY973063.1
Shorea scaberrima isolate 04-3069 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complete	1271	1271	100%	0.0	100%	KY973062.1
Shorea fallax isolate 24-3922 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complete cds;	1271	1271	100%	0.0	100%	KY973052.1
Shorea myrionerva isolate KASsmyr1 trnK-UUU gene, partial sequence; and maturase K (matK) gene, comple	1271	1271	100%	0.0	100%	KY973036.1

Image 8 Sequences producing significant alignments for the sample ID 753 (*Shorea acuminata*) with *matK* marker. Image taken from BLAST webpage: https://blast.ncbi.nlm.nih.gov/Blast.cgi

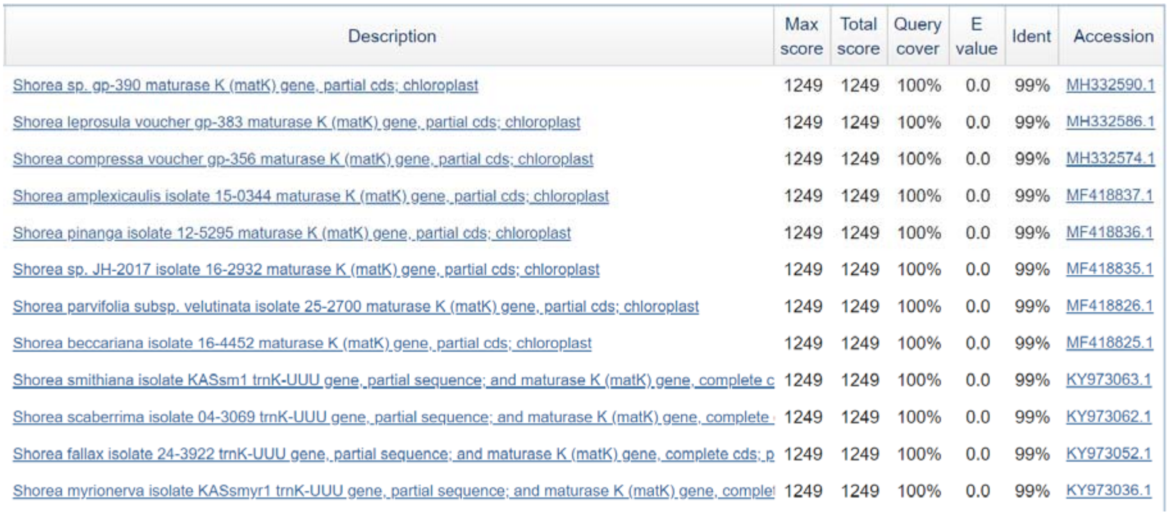
On the other hand, the BLAST search for sample ID 753 (*Shorea acuminata*) with *rbcL* marker presented a similar result from the BLAST search of sample ID 721 with *rbcL* marker (Image 9). Some of the resulting sequences showed perfect values for query cover and identity from *Hanguanaceae* family and others from the *Hypoxidaceae* family with also perfect values on query cover and values of 95% - 96% on identity.



Description	Max score	Total score	Query cover	E value	Ident	Accession
Hanguana malayana plastid, complete genome	1033	1033	100%	0.0	100%	KT312930.1
Hanguana sp. Kress 99-6325 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, p	1033	1033	100%	0.0	100%	FJ861125.1
Hanguana malayana chloroplast rbcL gene for RuBisCO large subunit, partial cds	1027	1027	100%	0.0	99%	AB088830.1
Hanguana malayana plastid partial rbcL gene for ribulose bisphosphate carboxylase large subunit, specimen	1018	1018	98%	0.0	100%	AM110247.1
Hanguana malayana plastid partial rbcL gene for rubisco	989	989	100%	0.0	99%	AJ417896.1
Hanguana malayana chloroplast partial rbcL gene for RuBisCO large subunit	989	989	100%	0.0	99%	AJ404842.1
Spiloxene serrata voucher Manning and Reeves JM&GR 2846 ribulose-1,5-bisphosphate carboxylase large s	894	894	100%	0.0	96%	JX903211.1
Spiloxene pusilla voucher Snijman 1860 (NBG) ribulose-1,5-bisphospate carboxylase large subunit (rbcL) gei	894	894	100%	0.0	96%	HM639290.1
Spiloxene nana voucher Snijman 1865a (NBG) ribulose-1,5-bisphospate carboxylase large subunit (rbcL) gei	894	894	100%	0.0	96%	HM639289.1
S.capensis chloroplast rbcL gene	894	894	100%	0.0	96%	Z77281.1
Burchardia multiflora voucher J.G. Conran 3045 (PERTH, ADU) ribulose-1,5-biphosphate carboxylase/oxyge	889	889	100%	0.0	95%	KC899449.1
Saniella occidentalis plastid partial rbcL gene for RuBisCO, specimen voucher NBG:Snijman, D. 2059	889	889	100%	0.0	95%	FN870923.1

Image 9 Sequences producing significant alignments for the sample ID 753 (*Shorea acuminata*) with *rbcL* marker. Image taken from BLAST webpage: https://blast.ncbi.nlm.nih.gov/Blast.cgi

Finally, the last sample having a BLAST search was sample ID 754 (*Shorea acuminata*) with *matK* marker (Image10) and *rbcL* marker (Image 11). One more time, *matK* results are correlated to the sample's label with a perfect query cover and a 99% identity for *Shorea* genus of *Dipterocarpaceae* family, as the previous two samples (ID 721 and 753).



Description	Max score	Total score	Query cover	E value	Ident	Accession
Shorea sp. gp-390 maturase K (matK) gene, partial cds; chloroplast	1249	1249	100%	0.0	99%	MH332590.1
Shorea leprosula voucher gp-383 maturase K (matK) gene, partial cds; chloroplast	1249	1249	100%	0.0	99%	MH332586.1
Shorea compressa voucher gp-356 maturase K (matK) gene, partial cds; chloroplast	1249	1249	100%	0.0	99%	MH332574.1
Shorea amplexicaulis isolate 15-0344 maturase K (matK) gene, partial cds; chloroplast	1249	1249	100%	0.0	99%	MF418837.1
Shorea pinanga isolate 12-5295 maturase K (matK) gene, partial cds; chloroplast	1249	1249	100%	0.0	99%	MF418836.1
Shorea sp. JH-2017 isolate 16-2932 maturase K (matK) gene, partial cds; chloroplast	1249	1249	100%	0.0	99%	MF418835.1
Shorea parvifolia subsp. velutinata isolate 25-2700 maturase K (matK) gene, partial cds; chloroplast	1249	1249	100%	0.0	99%	MF418826.1
Shorea beccariana isolate 16-4452 maturase K (matK) gene, partial cds; chloroplast	1249	1249	100%	0.0	99%	MF418825.1
Shorea smithiana isolate KASsm1 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complete c	1249	1249	100%	0.0	99%	KY973063.1
Shorea scaberrima isolate 04-3069 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complete	1249	1249	100%	0.0	99%	KY973062.1
Shorea fallax isolate 24-3922 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complete cds; p	1249	1249	100%	0.0	99%	KY973052.1
Shorea myrionerva isolate KASsmyr1 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complet	1249	1249	100%	0.0	99%	KY973036.1

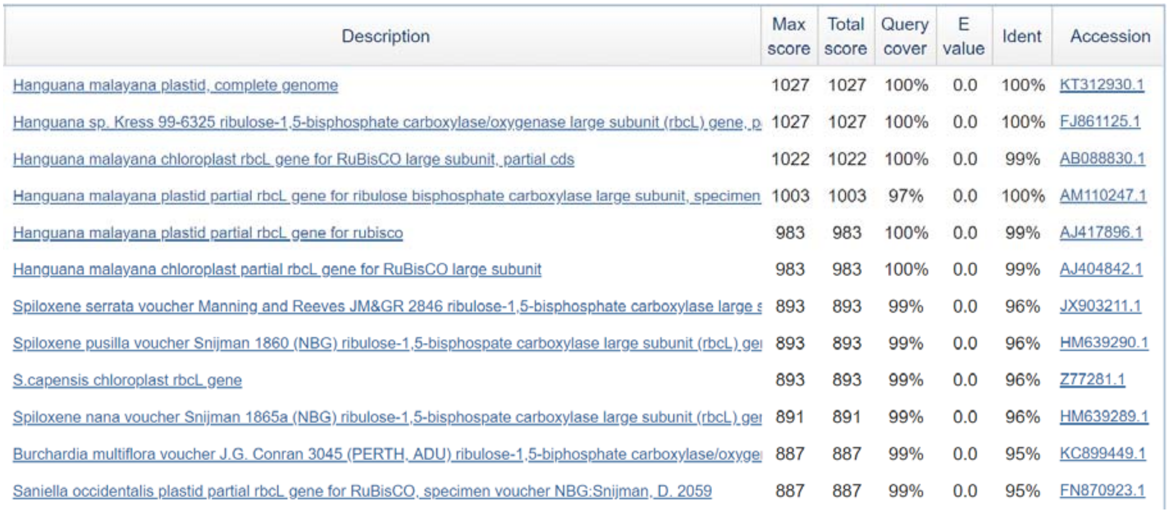
Image 10 Sequences producing significant alignments for the sample ID 754 (*Shorea acuminata*) with *matK* marker. Image taken from BLAST webpage: https://blast.ncbi.nlm.nih.gov/Blast.cgi

Furthermore, results from the *rbcL* marker (Image 11) were also similar to the previous results of samples ID 721 and 753. These results can explain why the three samples are clustering together and behaving as a total outgroup from the *Dipterocarpaceae* data set in Figure No. 2 since, according to BLAST database search results, they are not part of this family.

Summarizing, sample ID 4121 (*Shorea acuminata*) and sample ID 4591 (*Hopea ferruginea*) presented a mislabeled treatment with *matK* marker, which was evident in Figure 1 for their behavior as total outgroup from the data set. Second, samples ID samples ID 721 (*Shorea acuminata*), 753 (*Shorea acuminata*) and 754 (*Shorea acuminata*) also had a mislabel treatment with only the *rbcL* marker, explaining their outgroup behavior in Figure 2. In the concatenated tree (Figure 3), and due to this situation, all of the samples in question manifested different behaviors against to what it was expected. First two samples (4121 and

4591) still acted as a total outgroup and, the three latter (721, 753 and 754) were in the ingroup section of the tree but with a noticeable large evolutionary distance in comparison with the rest of the ingroup samples.

The reasons behind the mislabeled of the samples could be traced back from voucher comparison errors to human error in the Laboratory management of samples.



Description	Max score	Total score	Query cover	E value	Ident	Accession
Hanguana malayana plastid, complete genome	1027	1027	100%	0.0	100%	KT312930.1
Hanguana sp. Kress 99-6325 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, p	1027	1027	100%	0.0	100%	FJ861125.1
Hanguana malayana chloroplast rbcL gene for RuBisCO large subunit, partial cds	1022	1022	100%	0.0	99%	AB088830.1
Hanguana malayana plastid partial rbcL gene for ribulose bisphosphate carboxylase large subunit, specimen	1003	1003	97%	0.0	100%	AM110247.1
Hanguana malayana plastid partial rbcL gene for rubisco	983	983	100%	0.0	99%	AJ417896.1
Hanguana malayana chloroplast partial rbcL gene for RuBisCO large subunit	983	983	100%	0.0	99%	AJ404842.1
Spiloxene serrata voucher Manning and Reeves JM&GR 2846 ribulose-1,5-bisphosphate carboxylase large s	893	893	99%	0.0	96%	JX903211.1
Spiloxene pusilla voucher Snijman 1860 (NBG) ribulose-1,5-bisphospate carboxylase large subunit (rbcL) ge	893	893	99%	0.0	96%	HM639290.1
S.capensis chloroplast rbcL gene	893	893	99%	0.0	96%	Z77281.1
Spiloxene nana voucher Snijman 1865a (NBG) ribulose-1,5-bisphospate carboxylase large subunit (rbcL) ge	891	891	99%	0.0	96%	HM639289.1
Burchardia multiflora voucher J.G. Conran 3045 (PERTH, ADU) ribulose-1,5-biphosphate carboxylase/oxyge	887	887	99%	0.0	95%	KC899449.1
Saniella occidentalis plastid partial rbcL gene for RuBisCO, specimen voucher NBG:Snijman, D. 2059	887	887	99%	0.0	95%	FN870923.1

Image 11 Sequences producing significant alignments for the sample ID 754 (*Shorea acuminata*) with *rbcL* marker. Image taken from BLAST webpage: https://blast.ncbi.nlm.nih.gov/Blast.cgi

Having no guarantee on the correct taxonomic identification of these problematic samples because of the BLAST results showed different taxonomic families than *Dipterocarpaceae*, and since both markers are being used to construct concatenated phylogenetic trees, a conservative approach is advisable. The five samples with both *rbcL* and *matK* markers, sample ID 4121, sample ID 4591, sample ID 721, sample ID 753 and sample ID 754, will stay excluded for the rest of analyses regardless whether one of its markers was correctly labeled.

Once more, phylogenetic trees were constructed with the Neighbor Joining method, excluding the problematic samples in order to have a more consistent analyses with the given data set.

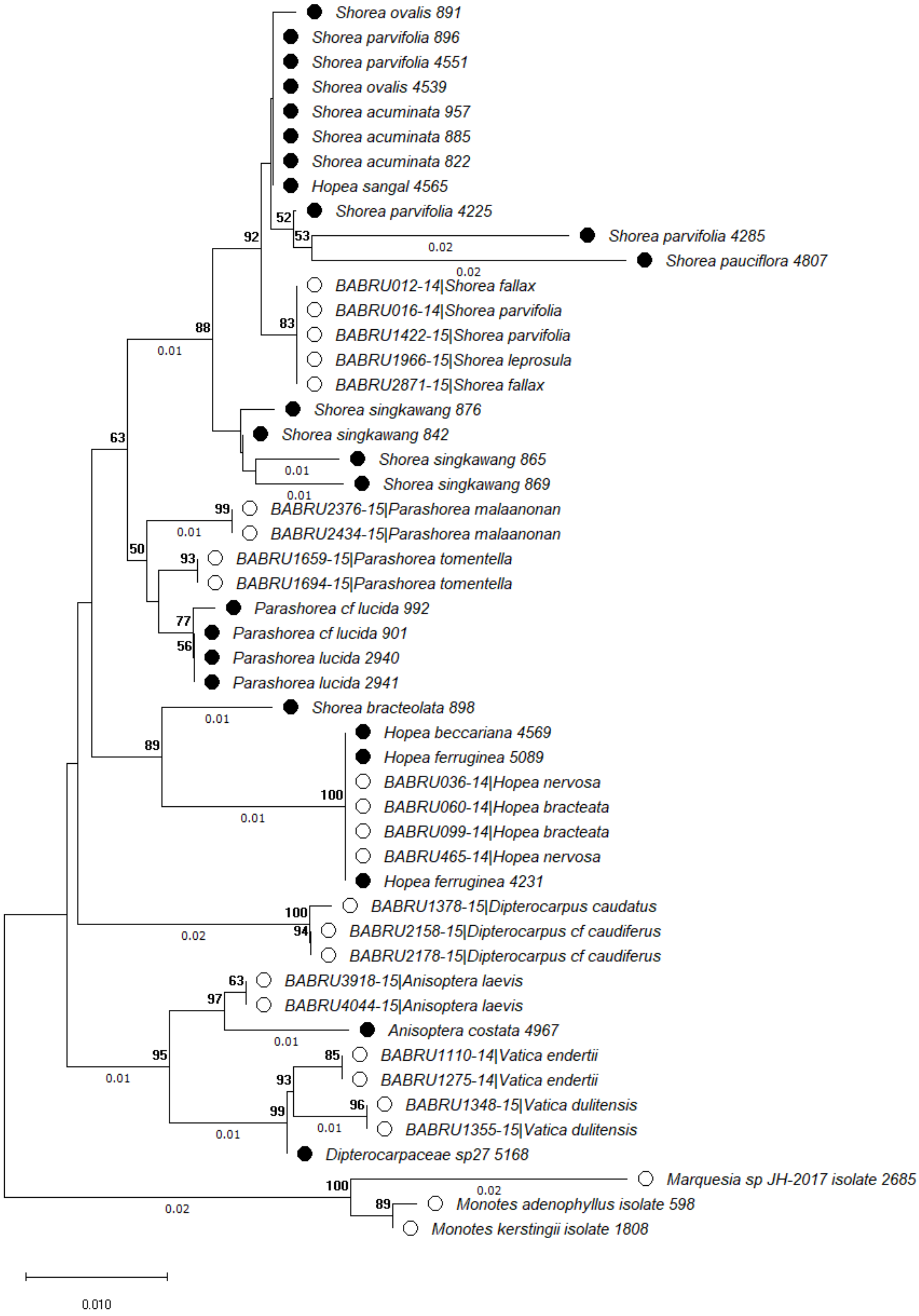

Shorea ovalis 891
Shorea parvifolia 896
Shorea parvifolia 4551
Shorea ovalis 4539
Shorea acuminata 957
Shorea acuminata 885
Shorea acuminata 822
Hopea sangal 4565
52
Shorea parvifolia 4225
92
53
Shorea parvifolia 4285
0.02
Shorea pauciflora 4807
0.02
BABRU012-14|Shorea fallax
BABRU016-14|Shorea parvifolia
83
BABRU1422-15|Shorea parvifolia
88
0.01
BABRU1966-15|Shorea leprosula
BABRU2871-15|Shorea fallax
Shorea singkawang 876
63
Shorea singkawang 842
Shorea singkawang 865
0.01
Shorea singkawang 869
0.01
99
BABRU2376-15|Parashorea malaanonan
0.01
BABRU2434-15|Parashorea malaanonan
50
93
BABRU1659-15|Parashorea tomentella
BABRU1694-15|Parashorea tomentella
Parashorea cf lucida 992
77
Parashorea cf lucida 901
56
Parashorea lucida 2940
Parashorea lucida 2941
Shorea bracteolata 898
0.01
89
Hopea beccariana 4569
Hopea ferruginea 5089
BABRU036-14|Hopea nervosa
100
BABRU060-14|Hopea bracteata
0.01
BABRU099-14|Hopea bracteata
BABRU465-14|Hopea nervosa
Hopea ferruginea 4231
100
BABRU1378-15|Dipterocarpus caudatus
94
BABRU2158-15|Dipterocarpus cf caudiferus
0.02
BABRU2178-15|Dipterocarpus cf caudiferus
63
BABRU3918-15|Anisoptera laevis
97
BABRU4044-15|Anisoptera laevis
Anisoptera costata 4967
0.01
95
85
BABRU1110-14|Vatica endertii
0.01
93
BABRU1275-14|Vatica endertii
96
BABRU1348-15|Vatica dulitensis
99
0.01
BABRU1355-15|Vatica dulitensis
0.01
Dipterocarpaceae sp27 5168
100
Marquesia sp JH-2017 isolate 2685
0.02
89
Monotes adenophyllus isolate 598
0.02
Monotes kerstingii isolate 1808
0.010

Figure 4 The neighbor joining phylogenetic tree of trees' samples collected in the EFForTS project based on the *matK* marker and genetic distances computed using the Maximum Composite Likelihood method. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.

The topology of the new constructed tree (Figure 4) is similar to the one including the problematic samples (Figure 1), without the outgroup behavior those samples had. Nevertheless, the *Monotoideae* group can now be easily recognized as the outgroup in the lower part of the figure with 100% bootstrap value. On the contrary, at the top of the figure a group of clustered *Shorea* samples can be observed with some of the plot samples, with most of them belonging to the *Shorea* genus.

Sample ID 992 (*Parashorea cf. lucida*), sample ID 901 (*Parashorea cf. lucida*), sample ID 2940 (*Parashorea lucida*) and sample ID 2941 (*Parashorea lucida*) are clustered with great relationships amongst them. However, it is noticeable that those samples have a weak relationship with the downloaded samples of *Parashorea* genus.

Most of the samples identified as belonging to *Hopea* genus, sample ID 4231 (*Hopea ferruginea*), sample ID 4569 (*Hopea beccariana*) and sample ID 5089 (*Hopea ferruginea*), are strongly related to the downloaded *Hopea* samples. Additionally, sample ID 4967 (*Anisoptera costata*) had a strong relationship with the downloaded *Anisoptera* samples and, sample ID 5168 (*Dipterocarpaceae sp. 27*) presented a good relationship with the rest of the downloaded *Vatica* samples, which can infer the sample belongs to a species from this particular genus. For reference, a condensed tree showing the clades with significant values of more than 50% is shown in Appendix 4.

A second phylogenetic tree was created with the exclusion of the five problematic samples, in this case, with the *rbcL* marker (Figure 5). The topology is also similar to the shown in Figure 2 and the outgroup is correctly placed at the bottom.

Sample ID 901 (*Parashorea cf. lucida*), 992 (*Parashorea cf. lucida*), 2940 (*Parashorea lucida*) and 2941 (*Parashorea lucida*) are associated together with great relationships amongst them and to the downloaded *Parashorea* samples, as it occurred in Figure 4. The *Hopea* samples clustered together, sample ID 4231 (*Hopea ferruginea*), sample ID 4569 (*Hopea beccariana*) and sample ID 5089 (*Hopea ferruginea*). However, they are weakly related to the downloaded *Hopea* samples, contrary as in the case of Figure 4.

Just like in Figure 4, sample ID 4967 (*Anisoptera costata*) clustered again with downloaded *Anisoptera* samples with strong support values, but sample ID 5168 (*Dipterocarpaceae sp. 27*) showed a weaker but clear relationship with the downloaded *Vatica* samples.

Most of the *Shorea* samples and collected samples showed no good relationships and hence, are not clustered amongst them with significant values (Appendix 5).

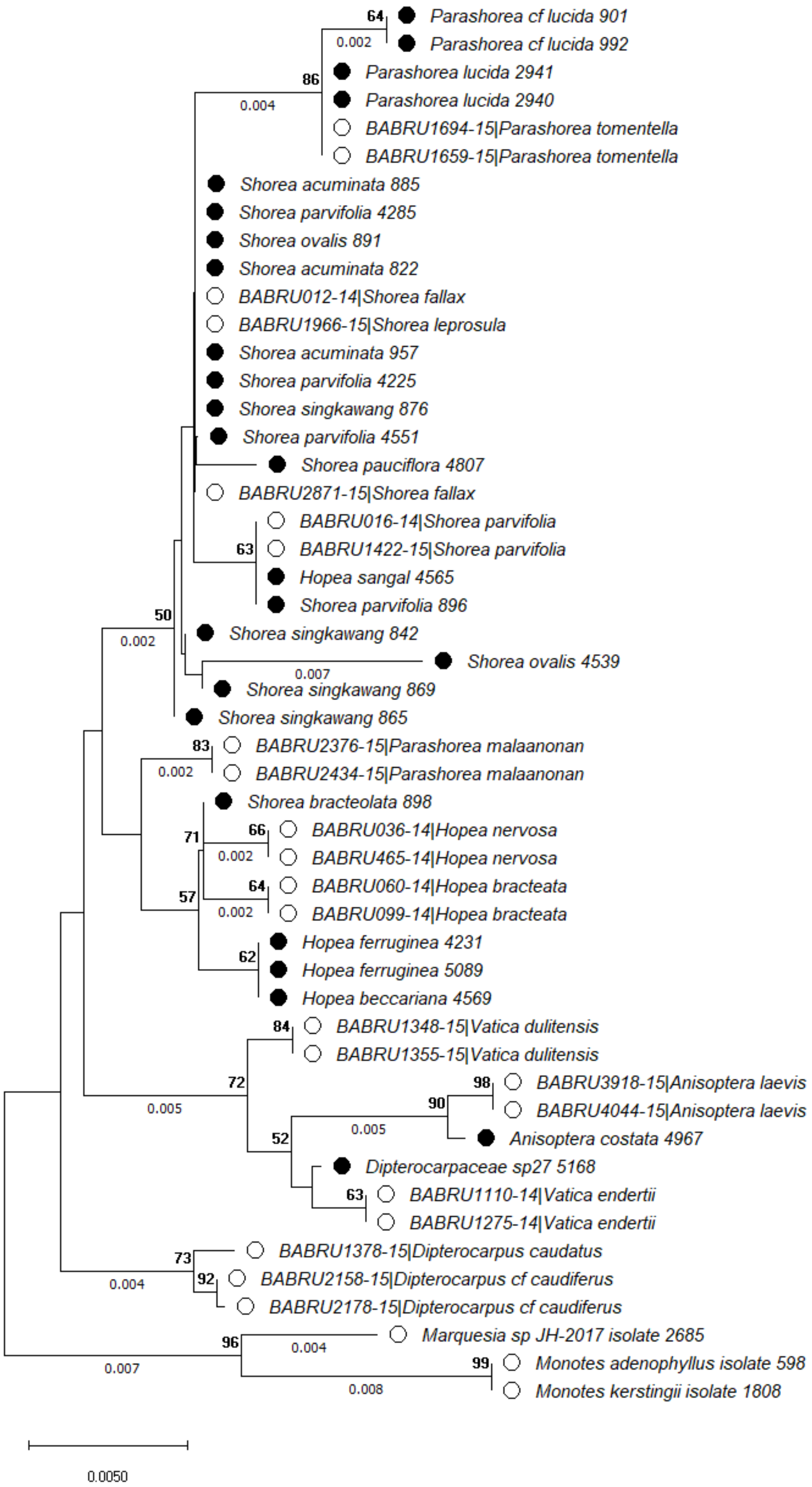
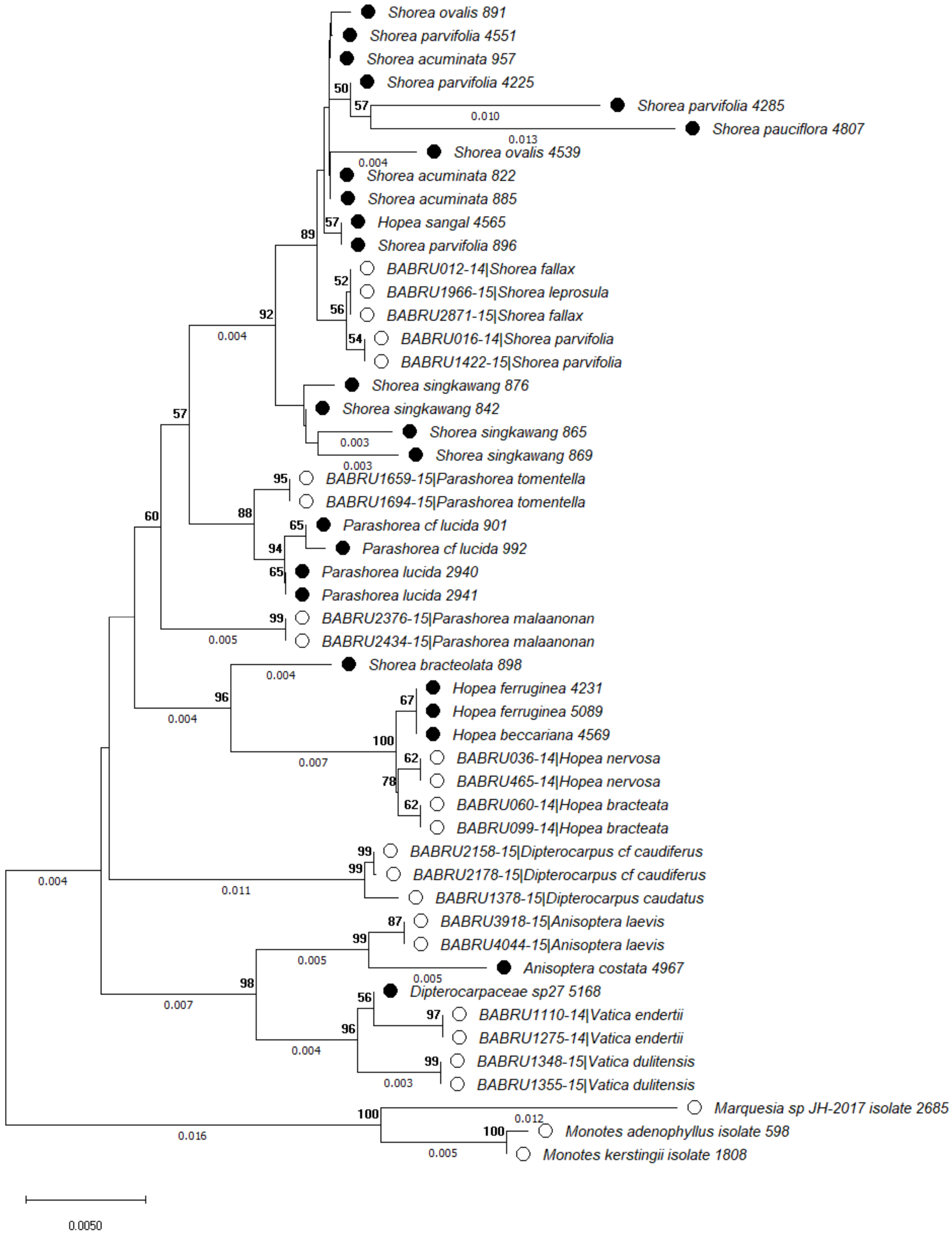

64
Parashorea cf lucida 901
0.002
Parashorea cf lucida 992
86
Parashorea lucida 2941
0.004
Parashorea lucida 2940
BABRU1694-15|Parashorea tomentella
BABRU1659-15|Parashorea tomentella
Shorea acuminata 885
Shorea parvifolia 4285
Shorea ovalis 891
Shorea acuminata 822
BABRU012-14|Shorea fallax
BABRU1966-15|Shorea leprosula
Shorea acuminata 957
Shorea parvifolia 4225
Shorea singkawang 876
Shorea parvifolia 4551
Shorea pauciflora 4807
BABRU2871-15|Shorea fallax
BABRU016-14|Shorea parvifolia
63
BABRU1422-15|Shorea parvifolia
Hopea sangal 4565
Shorea parvifolia 896
50
0.002
Shorea singkawang 842
Shorea ovalis 4539
0.007
Shorea singkawang 869
Shorea singkawang 865
83
BABRU2376-15|Parashorea malaanonan
0.002
BABRU2434-15|Parashorea malaanonan
Shorea bracteolata 898
66
BABRU036-14|Hopea nervosa
71
0.002
BABRU465-14|Hopea nervosa
64
BABRU060-14|Hopea bracteata
57
0.002
BABRU099-14|Hopea bracteata
Hopea ferruginea 4231
62
Hopea ferruginea 5089
Hopea beccariana 4569
84
BABRU1348-15|Vatica dulitensis
BABRU1355-15|Vatica dulitensis
72
98
BABRU3918-15|Anisoptera laevis
0.005
90
BABRU4044-15|Anisoptera laevis
52
0.005
Anisoptera costata 4967
Dipterocarpaceae sp27 5168
63
BABRU1110-14|Vatica endertii
BABRU1275-14|Vatica endertii
73
BABRU1378-15|Dipterocarpus caudatus
0.004
92
BABRU2158-15|Dipterocarpus cf caudiferus
BABRU2178-15|Dipterocarpus cf caudiferus
96
0.004
Marquesia sp JH-2017 isolate 2685
0.007
99
Monotes adenophyllus isolate 598
0.008
Monotes kerstingii isolate 1808
0.0050

Figure 5 The neighbor joining phylogenetic tree of trees' samples collected in the EFForTS project based on the *rbcL* marker and genetic distances computed using the Maximum Composite Likelihood method. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.



● Shorea ovalis 891
● Shorea parvifolia 4551
● Shorea acuminata 957
50
● Shorea parvifolia 4225
57
0.010
● Shorea parvifolia 4285
0.013
● Shorea pauciflora 4807
0.004
● Shorea ovalis 4539
● Shorea acuminata 822
● Shorea acuminata 885
57
● Hopea sangal 4565
89
● Shorea parvifolia 896
○ BABRU012-14|Shorea fallax
52
○ BABRU1966-15|Shorea leprosula
92
56
○ BABRU2871-15|Shorea fallax
0.004
54
○ BABRU016-14|Shorea parvifolia
○ BABRU1422-15|Shorea parvifolia
● Shorea singkawang 876
57
● Shorea singkawang 842
● Shorea singkawang 865
0.003
● Shorea singkawang 869
0.003
95
○ BABRU1659-15|Parashorea tomentella
○ BABRU1694-15|Parashorea tomentella
60
88
65
● Parashorea cf lucida 901
94
● Parashorea cf lucida 992
65
● Parashorea lucida 2940
● Parashorea lucida 2941
99
○ BABRU2376-15|Parashorea malaanonan
0.005
○ BABRU2434-15|Parashorea malaanonan
● Shorea bracteolata 898
0.004
96
67
● Hopea ferruginea 4231
● Hopea ferruginea 5089
0.004
● Hopea beccariana 4569
100
0.007
62
○ BABRU036-14|Hopea nervosa
78
○ BABRU465-14|Hopea nervosa
62
○ BABRU060-14|Hopea bracteata
○ BABRU099-14|Hopea bracteata
99
○ BABRU2158-15|Dipterocarpus cf caudiferus
0.004
99
○ BABRU2178-15|Dipterocarpus cf caudiferus
0.011
○ BABRU1378-15|Dipterocarpus caudatus
87
○ BABRU3918-15|Anisoptera laevis
99
○ BABRU4044-15|Anisoptera laevis
0.005
● Anisoptera costata 4967
0.005
98
56
● Dipterocarpaceae sp27 5168
0.007
97
○ BABRU1110-14|Vatica endertii
96
○ BABRU1275-14|Vatica endertii
0.004
99
○ BABRU1348-15|Vatica dulitensis
0.003
○ BABRU1355-15|Vatica dulitensis
100
○ Marquesia sp JH-2017 isolate 2685
0.012
0.016
100
○ Monotes adenophyllus isolate 598
0.005
○ Monotes kerstingii isolate 1808
0.0050

Figure 6 The neighbor joining phylogenetic tree of trees' samples collected in the EFForTS project based on the concatenated *rbcL* and *matK* markers, with genetic distances computed using the Maximum Composite Likelihood method. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.

Furthermore, another tree was constructed to better understand these relationships with the concatenation of *rbcL* and *matK* markers (Figure 6).

The first clade at the top of the figure contains samples from *Shorea* genus. Some of the downloaded *Shorea* samples demonstrated a fairly good relationship with the samples labeled as belonging to the *Shorea* genus, sample ID 891 (*Shorea ovalis*), sample ID 4551 (*Shorea parvifolia*), sample ID 957 (*Shorea acuminata*), sample ID 822 (*Shorea acuminata*), sample ID 4539 (*Shorea ovalis*), sample ID 885 (*Shorea acuminata*) and, in a subcluster, sample ID 4225 (*Shorea parvifolia*), sample ID 4285 (*Shorea parvifolia*) and sample ID 4807 (*Shorea pauciflora*). In the following subcluster, sample ID 4565 (*Hopea sangal*) and sample ID 896 (*Shorea parvifolia*) presented a weak relationship between them, and both of them belong to the same group (*Shoreae*) from the *Dipterocarpoideae* subfamily. Lastly, the final subcluster of this clade also belongs to the *Shorea* genus with sample ID 876 (*Shorea singkawang*), sample ID 842 (*Shorea singkawang*), sample ID 865 (*Shorea singkawang*) and sample ID 869 (*Shorea singkawang*), which showed a good relationship to the above mentioned samples and overall, to some of the downloaded *Shorea* genus samples.

Another noticeable remark, the *Parashorea tomentella* downloaded samples clustered, with a high probability, with the sample ID 901 (*Parashorea cf. lucida*), sample ID 992 (*Parashorea cf. lucida*), sample ID 2940 (*Parashorea lucida*) and sample ID 2941 (*Parashorea lucida*), affirming a strong relationship with the *Parashorea* genus.

One of the most interesting clades is the one containing the *Hopea* samples. *Hopea bracteata* and *Hopea nervosa* samples, related in almost all bootstrap iterations with sample ID 4231 (*Hopea ferruginea*), sample ID 4569 (*Hopea beccariana*) and sample ID 5089 (*Hopea ferruginea*). Thus, confirming that these relationships are closely partnered from the identified taxonomic level to the DNA barcoding level and phylogenetic analyses.

As final remarks, sample ID 4967 (*Anisoptera costata*) continued to have a solid relationship with the *Anisoptera* genus and, sample ID 5168 (*Dipterocarpaceae sp. 27*) proved a close relationship to all downloaded *Vatica* samples, which can infer the belonging of the sample to this genus.

The condensed tree for this figure, showing all significant clades with bootstrap values higher than 50%, can be found on Appendix 6.

3.1.2. Maximum Parsimony Phylogenetic Analyses

The Maximum Parsimony method was used for contrasting the Neighbor Joining method results and analyzing the performance of the new tree topologies. Only the set of samples excluding the problematic samples was used for generating the most parsimonious phylogenetic trees. The first phylogenetic tree was constructed with the sequences for the *matK* marker (Figure 7). For a condensed tree of this figure, showing only the clades with a significant value of relationships, i.e. more than 50%, the Appendix 7 can be consulted.

Interestingly, the first clade shows a similar topology configuration as seen in Figure 4 with samples ID 822 (*Shorea acuminata*), 4551 (*Shorea parvifolia*), 891 (*Shorea ovalis*), 957 (*Shorea acuminata*), 4539 (*Shorea ovalis*), 885 (*Shorea acuminata*), 4565 (*Hopea sangal*), 896 (*Shorea parvifolia*), showing a fairly good relationship amongst themselves and other *Shorea* samples, 4225 (*Shorea parvifolia*), 4285 (*Shorea parvifolia*) and 4807 (*Shorea pauciflora*), which clustered in a subclade with acceptable bootstrap values. Some downloaded *Shorea* samples also clustered together, presenting a high correlation value to the above samples, as occurred in Figure 4.

Moreover, samples ID 876 (*Shorea singkawang*), 869 (*Shorea singkawang*) and 842 (*Shorea singkawang*) and 865 (*Shorea singkawang*), formed a cluster with a good connection to the first clade, despite of having low association amongst them. This relationship indicates that the *Shorea singkawang* samples may hold a common ancestor to the *Shorea* samples from the first clade.

As appeared in Figure 4, samples ID 901 (*Parashorea cf. lucida*), 992 (*Parashorea cf. lucida*), 2940 (*Parashorea lucida*) and 2941 (*Parashorea lucida*) clustered all together with an acceptable support value. Besides, the cluster of samples manifested a stronger relationship to the downloaded samples of *Parashorea* genus than in Figure 4.

Sample ID 4967 (*Anisoptera costata*) is reiteratively clustered to the downloaded *Anisoptera* samples with great support values and, sample ID 5168 (*Dipterocarpaceae sp. 27*) is again present amongst the downloaded *Vatica* samples with high support values as in Figure 4.

In the case of the *Hopea* samples, sample ID 4231 (*Hopea ferruginea*), sample ID 5089 (*Hopea ferruginea*) and sample ID 4569 (*Hopea beccariana*) did not group with any particular downloaded species. Nevertheless, all *Hopea* samples are present in a complete cluster amongst the rest of all the downloaded *Hopea* samples with perfect support value.

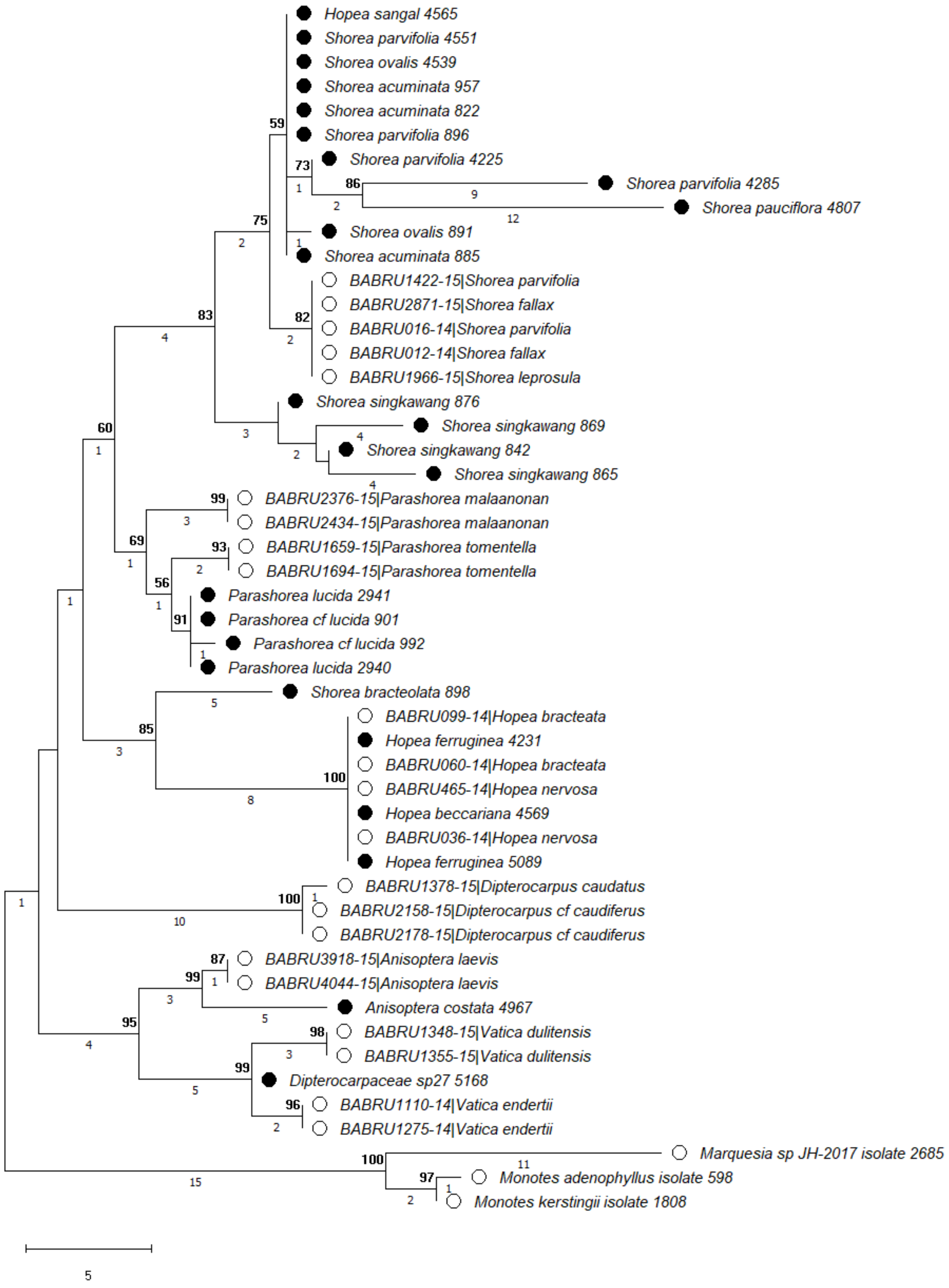

Hopea sangal 4565
Shorea parvifolia 4551
Shorea ovalis 4539
Shorea acuminata 957
Shorea acuminata 822
Shorea parvifolia 896
Shorea parvifolia 4225
Shorea parvifolia 4285
Shorea pauciflora 4807
Shorea ovalis 891
Shorea acuminata 885
BABRU1422-15|Shorea parvifolia
BABRU2871-15|Shorea fallax
BABRU016-14|Shorea parvifolia
BABRU012-14|Shorea fallax
BABRU1966-15|Shorea leprosula
Shorea singkawang 876
Shorea singkawang 869
Shorea singkawang 842
Shorea singkawang 865
BABRU2376-15|Parashorea malaanonan
BABRU2434-15|Parashorea malaanonan
BABRU1659-15|Parashorea tomentella
BABRU1694-15|Parashorea tomentella
Parashorea lucida 2941
Parashorea cf lucida 901
Parashorea cf lucida 992
Parashorea lucida 2940
Shorea bracteolata 898
BABRU099-14|Hopea bracteata
Hopea ferruginea 4231
BABRU060-14|Hopea bracteata
BABRU465-14|Hopea nervosa
Hopea beccariana 4569
BABRU036-14|Hopea nervosa
Hopea ferruginea 5089
BABRU1378-15|Dipterocarpus caudatus
BABRU2158-15|Dipterocarpus cf caudiferus
BABRU2178-15|Dipterocarpus cf caudiferus
BABRU3918-15|Anisoptera laevis
BABRU4044-15|Anisoptera laevis
Anisoptera costata 4967
BABRU1348-15|Vatica dulitensis
BABRU1355-15|Vatica dulitensis
Dipterocarpaceae sp27 5168
BABRU1110-14|Vatica endertii
BABRU1275-14|Vatica endertii
Marquesia sp JH-2017 isolate 2685
Monotes adenophyllus isolate 598
Monotes kerstingii isolate 1808
59
73
86
75
83
82
60
99
69
93
56
91
85
100
100
87
99
95
98
99
96
100
97
5

Figure 7 The maximum parsimony phylogenetic tree of trees' samples collected in the EFForTS project based on the *matK* marker and obtained by using the Subtree-Pruning-Regrafting (SPR) algorithm. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.

The most significant clades for Figure 7 can be consulted on Appendix 7.

Following, the second phylogenetic tree was created for the *rbcL* marker without the problematic samples (Figure 8). The topologies from this new constructed tree and the Neighbor Joining method tree for *rbcL* (Figure 5) are similar but with important differences, such as the arrangement of some clades and content of some subclusters.

The *Shorea* samples clustering in Figure 4 and Figure 7, with the downloaded *Shorea* samples, are not present as a strong clade in this figure. On the contrary, there are no good relationships above support values higher than 50% with the exception of the subclade consisting of sample ID 896 (*Shorea parvifolia*), 4565 (*Hopea sangal*) and two downloaded samples of *Shorea parvifolia*, which is also present in Figure 5.

Only the *Parashorea tomentella* downloaded samples clustered acceptably enough to samples ID 901 (*Parashorea cf. lucida*), sample ID 992 (*Parashorea cf. lucida*), sample ID 2940 (*Parashorea lucida*) and sample ID 2941 (*Parashorea lucida*). This also happened in Figure 5 with similar relationships amongst the samples.

The samples ID 4231 (*Hopea ferruginea*), 4569 (*Hopea beccariana*) and 5089 (*Hopea ferruginea*) clustered in a separate clade without directly relating to any of the downloaded *Hopea* samples. However, they showed a weak relationship (lower than 50%) with the rest of downloaded *Hopea* samples and sample ID 898 (*Shorea bracteolata*).

Again, sample ID 4967 (*Anisoptera costata*) grouped with the downloaded *Anisoptera* genus samples having a strong support value. Also, sample ID 5168 (*Dipterocarpaceae sp. 27*) showed a noticeable, yet weak, relationship with the downloaded *Vatica* samples. The overall performance relationship amongst sample 5168 and the downloaded *Vatica* genus samples was weak.

Appendix 8 can be checked for reviewing only the clades with 50%, or more, support values from this figure.

Finally, the concatenation of *rbcL* and *matK* markers was carried out for the construction of a more complete phylogenetic tree, giving a different angle of previous seen perspectives and without problematic samples (Figure 9).

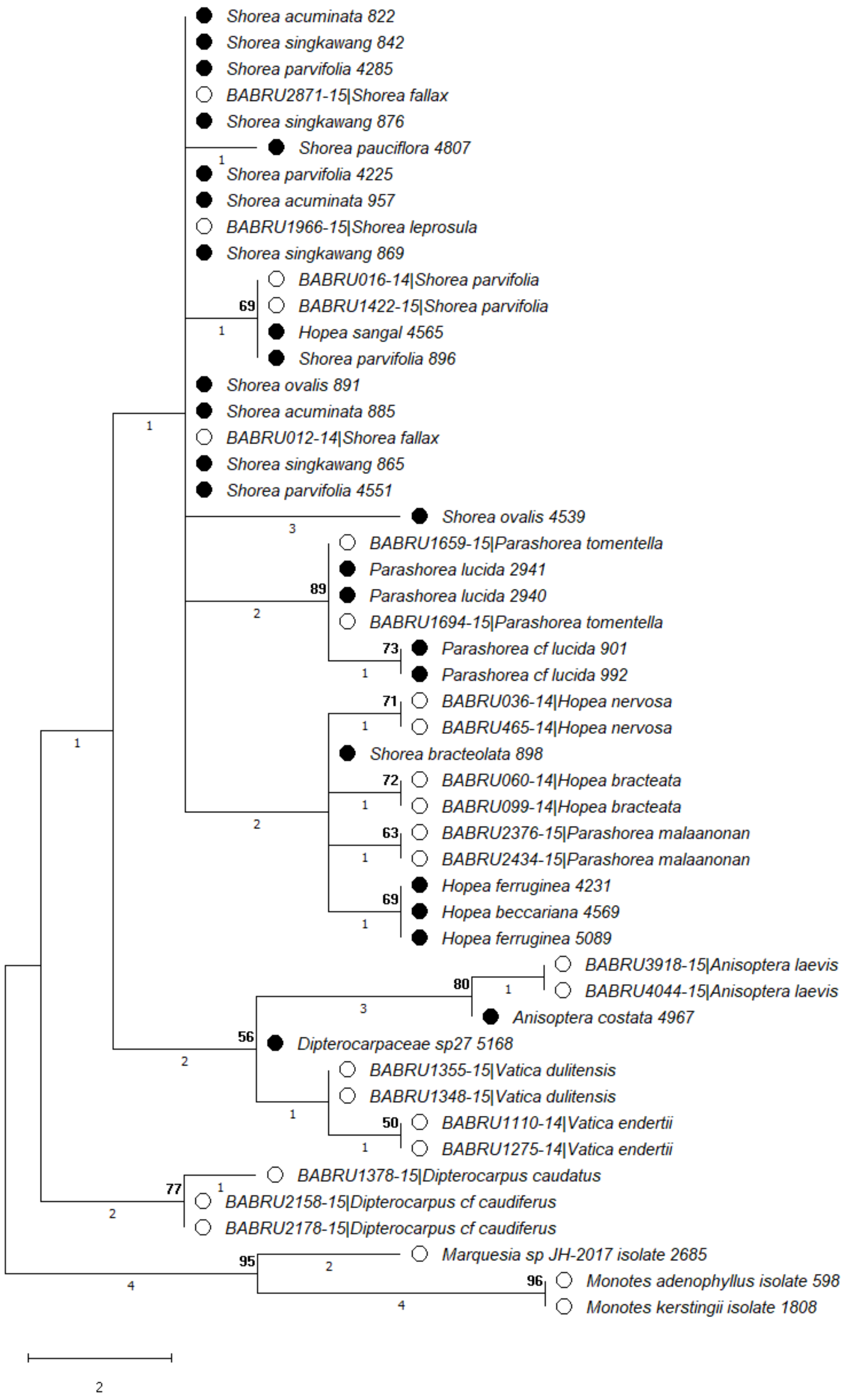

Shorea acuminata 822
Shorea singkawang 842
Shorea parvifolia 4285
BABRU2871-15|Shorea fallax
Shorea singkawang 876
Shorea pauciflora 4807
1
Shorea parvifolia 4225
Shorea acuminata 957
BABRU1966-15|Shorea leprosula
Shorea singkawang 869
BABRU016-14|Shorea parvifolia
69
BABRU1422-15|Shorea parvifolia
1
Hopea sangal 4565
Shorea parvifolia 896
Shorea ovalis 891
Shorea acuminata 885
1
BABRU012-14|Shorea fallax
Shorea singkawang 865
Shorea parvifolia 4551
Shorea ovalis 4539
3
BABRU1659-15|Parashorea tomentella
Parashorea lucida 2941
89
Parashorea lucida 2940
2
BABRU1694-15|Parashorea tomentella
73
Parashorea cf lucida 901
1
Parashorea cf lucida 992
71
BABRU036-14|Hopea nervosa
1
BABRU465-14|Hopea nervosa
1
Shorea bracteolata 898
72
BABRU060-14|Hopea bracteata
1
BABRU099-14|Hopea bracteata
2
63
BABRU2376-15|Parashorea malaanonan
1
BABRU2434-15|Parashorea malaanonan
Hopea ferruginea 4231
69
Hopea beccariana 4569
1
Hopea ferruginea 5089
BABRU3918-15|Anisoptera laevis
80
1
BABRU4044-15|Anisoptera laevis
3
Anisoptera costata 4967
56
Dipterocarpaceae sp27 5168
2
BABRU1355-15|Vatica dulitensis
BABRU1348-15|Vatica dulitensis
1
50
BABRU1110-14|Vatica endertii
1
BABRU1275-14|Vatica endertii
BABRU1378-15|Dipterocarpus caudatus
77
1
BABRU2158-15|Dipterocarpus cf caudiferus
2
BABRU2178-15|Dipterocarpus cf caudiferus
95
2
Marquesia sp JH-2017 isolate 2685
96
Monotes adenophyllus isolate 598
4
4
Monotes kerstingii isolate 1808
2

Figure 8 The maximum parsimony phylogenetic tree of trees' samples collected in the EFForTS project based on the *rbcL* marker and obtained by using the Subtree-Pruning-Regrafting (SPR) algorithm. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.

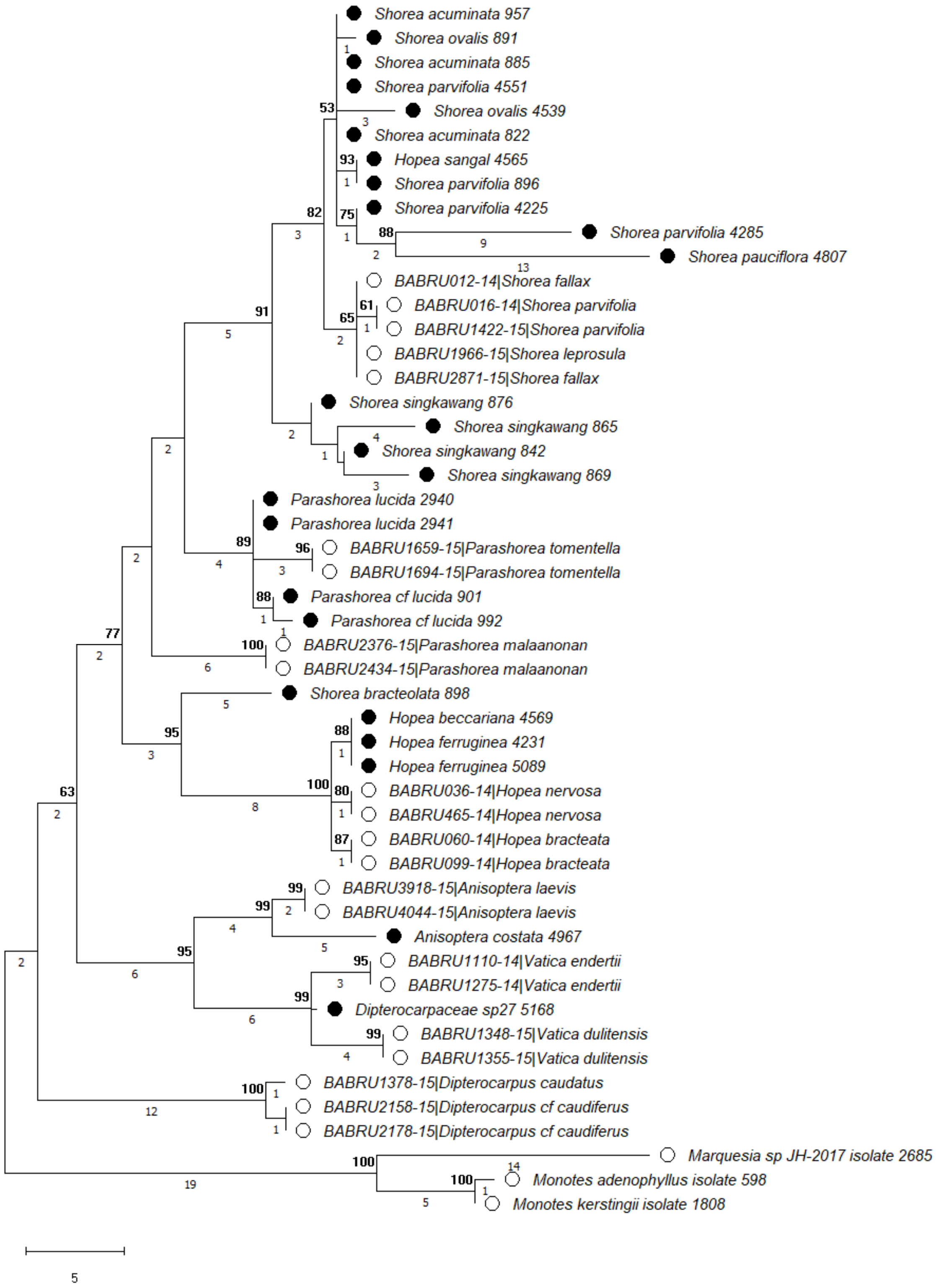

Shorea acuminata 957
Shorea ovalis 891
1
Shorea acuminata 885
Shorea parvifolia 4551
53
Shorea ovalis 4539
3
Shorea acuminata 822
93
Hopea sangal 4565
1
Shorea parvifolia 896
82
75
Shorea parvifolia 4225
3
88
Shorea parvifolia 4285
1
9
2
Shorea pauciflora 4807
13
BABRU012-14|Shorea fallax
61
BABRU016-14|Shorea parvifolia
91
65
1
BABRU1422-15|Shorea parvifolia
5
2
BABRU1966-15|Shorea leprosula
BABRU2871-15|Shorea fallax
Shorea singkawang 876
2
Shorea singkawang 865
4
2
1
Shorea singkawang 842
Shorea singkawang 869
3
Parashorea lucida 2940
Parashorea lucida 2941
89
96
BABRU1659-15|Parashorea tomentella
2
4
3
BABRU1694-15|Parashorea tomentella
88
Parashorea cf lucida 901
1
1
Parashorea cf lucida 992
77
100
BABRU2376-15|Parashorea malaanonan
2
6
BABRU2434-15|Parashorea malaanonan
Shorea bracteolata 898
5
Hopea beccariana 4569
88
95
Hopea ferruginea 4231
3
1
Hopea ferruginea 5089
100
80
BABRU036-14|Hopea nervosa
63
8
1
BABRU465-14|Hopea nervosa
2
87
BABRU060-14|Hopea bracteata
1
BABRU099-14|Hopea bracteata
99
BABRU3918-15|Anisoptera laevis
99
2
BABRU4044-15|Anisoptera laevis
4
Anisoptera costata 4967
95
5
2
6
95
BABRU1110-14|Vatica endertii
3
BABRU1275-14|Vatica endertii
99
Dipterocarpaceae sp27 5168
6
99
BABRU1348-15|Vatica dulitensis
4
BABRU1355-15|Vatica dulitensis
100
1
BABRU1378-15|Dipterocarpus caudatus
12
BABRU2158-15|Dipterocarpus cf caudiferus
1
BABRU2178-15|Dipterocarpus cf caudiferus
Marquesia sp JH-2017 isolate 2685
100
14
100
Monotes adenophyllus isolate 598
19
1
5
Monotes kerstingii isolate 1808
5

Figure 9 The maximum parsimony phylogenetic tree of trees' samples collected in the EFForTS project based on the concatenated *rbcL* and *matK* markers, and obtained by using the Subtree-Pruning-Regrafting (SPR) algorithm. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.

Once more, the top is characterized for a big clade consisting of various clusters consisting some downloaded *Shorea* samples and analyzed samples. One of the similarities of this figure from Figure 6 is the appearance of various subclades. The first of them was present in most of the iterations, consisting of sample ID 4565 (*Hopea sangal*) and 896 (*Shorea parviflia*). Next subclade is a good relationship amongst samples ID 4225 (*Shorea parviflia*), 4285 (*Shorea parviflia*) and 4807 (*Shorea pauciflora*). Following subclade is just a fairly good bonding amongst some downloaded *Shorea* samples. The last subclade is a weak relationship amongst samples ID 876 (*Shorea singkawang*), 865 (*Shorea singkawang*), 842 (*Shorea singkawang*) and 869 (*Shorea singkawang*). The rest of the samples, 891 (*Shorea ovalis*), 822 (*Shorea acuminata*), 957 (*Shorea acuminata*), 4539 (*Shorea ovalis*), 885 (*Shorea acuminata*), and 4551 (*Shorea parvifolia*), were also present within the first clade, which despite of not having any specific relationship amongst themselves, the support value showed a good overall relationship to *Shorea* genus samples

Beginning to look as a constant in the figures, *Parashorea tomentella* samples manifested a strong relationship with samples ID 2940 (*Parashorea lucida*), 2941 (*Parashorea lucida*) and a subclade of 901 (*Parashorea cf. lucida*) and 992 (*Parashorea cf. lucida*). Having encountered this relationship in both concatenated *rbcL* and *matK* figures (Figure 6 and Figure 9) from two different methods, it is a certainty that samples are strongly correlating to this particular species which is notably endemic to east Borneo.

Most of *Hopea* samples, 4569 (*Hopea beccariana*), 4231 (*Hopea ferruginea*) and 5089 (*Hopea ferruginea*), clustered in one subclade with strong support value. Additionally, the subclade is highly related to all other downloaded *Hopea* samples, validating its belonging to this genus.

Final remarks of this figure include the sample ID 4967 (*Anisoptera costata*) showing a strong relationship to *Anisoptera laevis* downloaded samples, and sample ID 5168 (*Dipterocarpaceae sp. 27*) having a strong placement amongst the *Vatica* downloaded samples, as in Figure 6 for both cases.

In Appendix 9, the most significant clades with bootstrap values of more than 50% are shown.

3.1.3. Maximum Likelihood Phylogenetic Analyses

Starting to look as a pattern, the display and arrangement of the tree's topology (Figure 10) is similar to the previous *matK* marker figures with Neighbor Joining and Maximum Parsimony, Figure 4 and Figure 7, respectively.

Shorea samples are forming a big clade at the top of the figure with samples ID 4551 (*Shorea parvifolia*), 896 (*Shorea parvifolia*), 4539 (*Shorea ovalis*), 957 (*Shorea acuminata*), 885 (*Shorea acuminata*), 822 (*Shorea acuminata*), 4565 (*Hopea sangal*), 891 (*Shorea ovalis*) having a low support value amongst them. However, they also presented a fairly acceptable relationship to some downloaded *Shorea* samples and an overall great relationship with subclade of samples ID 876 (*Shorea singkawang*), 842 (*Shorea singkawang*), 865 (*Shorea singkawang*) and 869 (*Shorea singkawang*). Another interesting fact is the constantly appearing of the subclade with samples ID 4225 (*Shorea parvifolia*), 4285 (*Shorea parvifolia*) and 4807 (*Shorea pauciflora*), within the first clade of *Shorea* samples in all of the figures.

Following the same shape from previous figures, *Parashorea* samples, ID 901 (*Parashorea cf. lucida*), 992 (*Parashorea cf. lucida*), 2940 (*Parashorea lucida*) and 2941 (*Parashorea lucida*) are weakly associated to the rest of downloaded *Parashorea* samples even though they are forming a cluster.

As for the *Hopea* samples, ID 4569 (*Hopea beccariana*), 4231 (*Hopea ferruginea*) and 5089 (*Hopea ferruginea*), they all are firmly correlated to the downloaded *Hopea* samples, just as it has been observed in previous *matK* figures, being the Figure 7 the most similar to Figure 10.

One more time, both samples, ID 4967 (*Anisoptera costata*) and 5168 (*Dipterocarpaceae sp. 27*), are strongly associated with high support values to the downloaded *Anisopterea* and *Vatica* samples, respectively.

On Appendix 10 a condensed tree with collapsed low support branches (less than 50%) can be consulted.

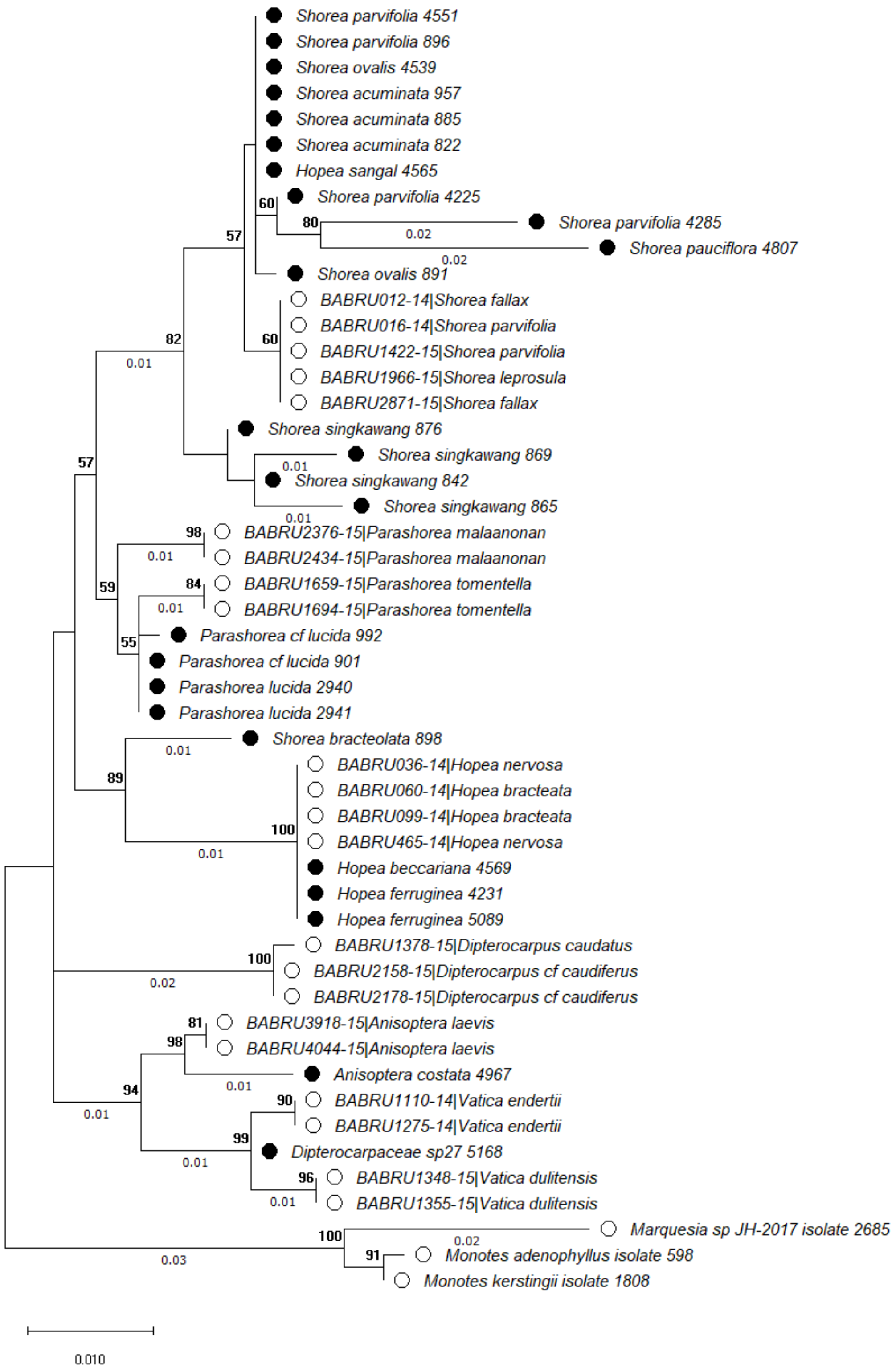

Shorea parvifolia 4551
Shorea parvifolia 896
Shorea ovalis 4539
Shorea acuminata 957
Shorea acuminata 885
Shorea acuminata 822
Hopea sangal 4565
60
Shorea parvifolia 4225
80
Shorea parvifolia 4285
0.02
57
Shorea pauciflora 4807
0.02
Shorea ovalis 891
BABRU012-14|Shorea fallax
BABRU016-14|Shorea parvifolia
82
60
BABRU1422-15|Shorea parvifolia
0.01
BABRU1966-15|Shorea leprosula
BABRU2871-15|Shorea fallax
Shorea singkawang 876
Shorea singkawang 869
57
0.01
Shorea singkawang 842
Shorea singkawang 865
0.01
98
BABRU2376-15|Parashorea malaanonan
0.01
BABRU2434-15|Parashorea malaanonan
84
BABRU1659-15|Parashorea tomentella
59
0.01
BABRU1694-15|Parashorea tomentella
Parashorea cf lucida 992
55
Parashorea cf lucida 901
Parashorea lucida 2940
Parashorea lucida 2941
Shorea bracteolata 898
0.01
BABRU036-14|Hopea nervosa
89
BABRU060-14|Hopea bracteata
BABRU099-14|Hopea bracteata
100
BABRU465-14|Hopea nervosa
0.01
Hopea beccariana 4569
Hopea ferruginea 4231
Hopea ferruginea 5089
BABRU1378-15|Dipterocarpus caudatus
100
BABRU2158-15|Dipterocarpus cf caudiferus
0.02
BABRU2178-15|Dipterocarpus cf caudiferus
81
BABRU3918-15|Anisoptera laevis
98
BABRU4044-15|Anisoptera laevis
Anisoptera costata 4967
94
0.01
90
BABRU1110-14|Vatica endertii
0.01
BABRU1275-14|Vatica endertii
99
Dipterocarpaceae sp27 5168
0.01
96
BABRU1348-15|Vatica dulitensis
0.01
BABRU1355-15|Vatica dulitensis
100
Marquesia sp JH-2017 isolate 2685
0.02
91
Monotes adenophyllus isolate 598
0.03
Monotes kerstingii isolate 1808
0.010

Figure 10 The maximum likelihood phylogenetic tree of trees' samples collected in the EFForTS project based on the *matK* marker and the Hasegawa-Kishino-Yano model. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. The tree log likelihood is (-1689.38). The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.

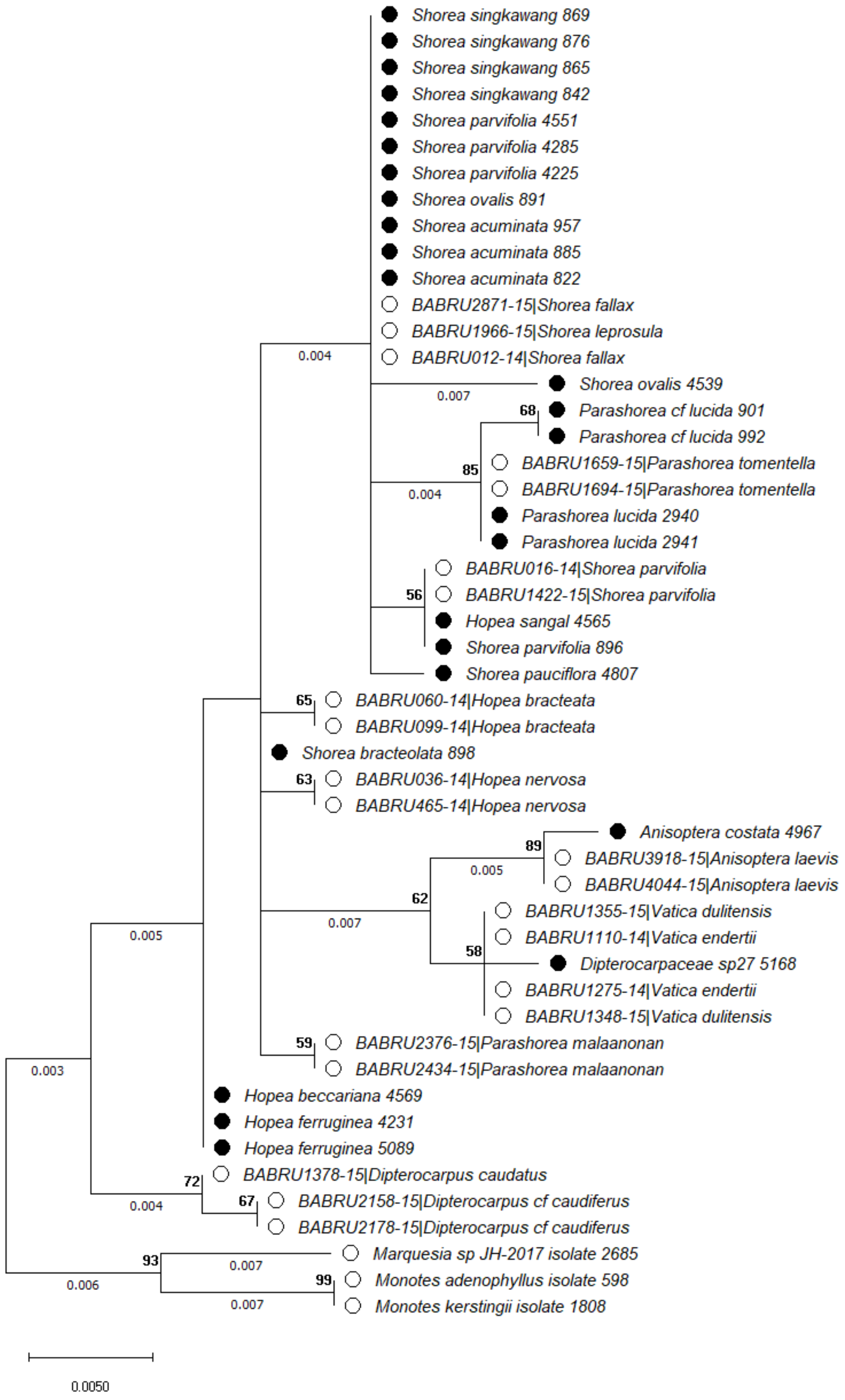

● Shorea singkawang 869
● Shorea singkawang 876
● Shorea singkawang 865
● Shorea singkawang 842
● Shorea parvifolia 4551
● Shorea parvifolia 4285
● Shorea parvifolia 4225
● Shorea ovalis 891
● Shorea acuminata 957
● Shorea acuminata 885
● Shorea acuminata 822
○ BABRU2871-15|Shorea fallax
○ BABRU1966-15|Shorea leprosula
0.004
○ BABRU012-14|Shorea fallax
● Shorea ovalis 4539
0.007
68
● Parashorea cf lucida 901
● Parashorea cf lucida 992
85
○ BABRU1659-15|Parashorea tomentella
0.004
○ BABRU1694-15|Parashorea tomentella
● Parashorea lucida 2940
● Parashorea lucida 2941
○ BABRU016-14|Shorea parvifolia
56
○ BABRU1422-15|Shorea parvifolia
● Hopea sangal 4565
● Shorea parvifolia 896
● Shorea pauciflora 4807
65
○ BABRU060-14|Hopea bracteata
○ BABRU099-14|Hopea bracteata
● Shorea bracteolata 898
63
○ BABRU036-14|Hopea nervosa
○ BABRU465-14|Hopea nervosa
● Anisoptera costata 4967
89
○ BABRU3918-15|Anisoptera laevis
0.005
○ BABRU4044-15|Anisoptera laevis
62
○ BABRU1355-15|Vatica dulitensis
0.005
0.007
○ BABRU1110-14|Vatica endertii
58
● Dipterocarpaceae sp27 5168
○ BABRU1275-14|Vatica endertii
○ BABRU1348-15|Vatica dulitensis
59
○ BABRU2376-15|Parashorea malaanonan
0.003
○ BABRU2434-15|Parashorea malaanonan
● Hopea beccariana 4569
● Hopea ferruginea 4231
● Hopea ferruginea 5089
72
○ BABRU1378-15|Dipterocarpus caudatus
0.004
67
○ BABRU2158-15|Dipterocarpus cf caudiferus
○ BABRU2178-15|Dipterocarpus cf caudiferus
93
○ Marquesia sp JH-2017 isolate 2685
0.007
0.006
99
○ Monotes adenophyllus isolate 598
0.007
○ Monotes kerstingii isolate 1808
0.0050

Figure 11 The maximum likelihood phylogenetic tree of trees' samples collected in the EFForTS project based on the *rbcL* marker and the Hasegawa-Kishino-Yano model. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. The tree log likelihood is (-943.45). The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.

The phylogenetic tree based on the *rbcL* marker (Figure 11) manifested a similar topology to the previous *rbcL* marker Figure 5 and Figure 8. All *Shorea* samples had no resolution of their relationships to each other, except for the subcluster of sample ID 4565 (*Hopea sangal*), 896 (*Shorea parvifolia*) and two downloaded database samples of *Shorea parvifolia*, which is present in all the *rbcL* phylogenetic tree figures.

Parashorea samples, sample ID 901 (*Parashorea cf. lucida*), 992 (*Parashorea cf. lucida*), 2940 (*Parashorea lucida*) and 2941 (*Parashorea lucida*), also behaved as usual, having a high support value with *Parashorea tomentella* downloaded samples.

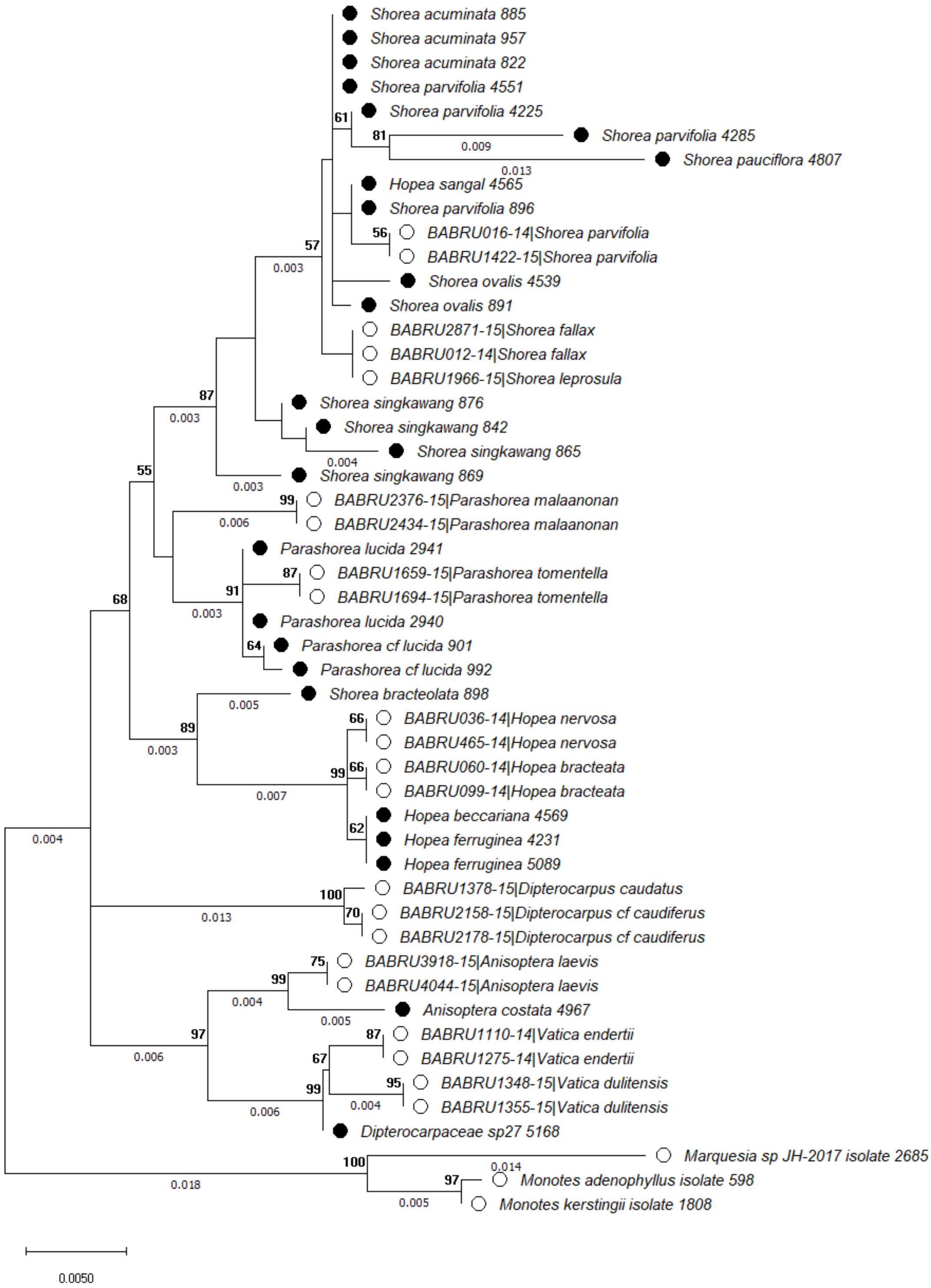
Nevertheless, one of the differences from previous results is the behavior of *Hopea* samples, ID 4569 (*Hopea beccariana*), 4231 (*Hopea ferruginea*) and 5089 (*Hopea ferruginea*), which did not group to any particular cluster, nor even to the downloaded *Hopea* samples.

As for the case of sample ID 4967 (*Anisoptera costata*), the relationship shown to the *Anisoptera laevis* samples was once again very strong. In addition, sample ID 5168 (*Dipterocarpaceae sp. 27*) was vaguely to acceptably related to the downloaded *Vatica* samples.

The condensed tree for Figure 11, with collapsed branches of lower support values, can be seen on Appendix 11 to highlight the most significant clades.

The third maximum likelihood phylogenetic tree with the concatenation of both markers, *rbcL* and *matK*, was created (Figure 12). It is important to note how the topology of the concatenated trees from Neighbor Joining method and Maximum Parsimony method (Figure 6 and Figure 9) are strikingly similar to the maximum likelihood concatenated tree (Figure 11).

Similar features can be spotted; most *Shorea* samples are present in a big clade having low support value; samples ID 4225 (*Shorea parvifolia*), 4285 (*Shorea parvifolia*) and 4807 (*Shorea pauciflora*) are constantly appearing as a subclade with acceptable to high bootstrap values; the next recurring subclade consists of samples ID 4565 (*Hopea sangal*) and 896 (*Shorea parvifolia*), clustering with low (Figure 6 and Figure 11) to high (Figure 9) support values, but in the case of Figure 11 the subclade also presented a weak relation to two downloaded *Shorea parvifolia* samples; the weak subcluster of *Shorea singkawang* samples (ID 876, 842, 869) and their strong association to the *Shorea* samples' clade.



● Shorea acuminata 885
● Shorea acuminata 957
● Shorea acuminata 822
● Shorea parvifolia 4551
61
● Shorea parvifolia 4225
81
● Shorea parvifolia 4285
0.009
0.013
● Shorea pauciflora 4807
● Hopea sangal 4565
● Shorea parvifolia 896
56
○ BABRU016-14|Shorea parvifolia
○ BABRU1422-15|Shorea parvifolia
57
0.003
● Shorea ovalis 4539
● Shorea ovalis 891
○ BABRU2871-15|Shorea fallax
○ BABRU012-14|Shorea fallax
○ BABRU1966-15|Shorea leprosula
87
0.003
● Shorea singkawang 876
● Shorea singkawang 842
0.004
● Shorea singkawang 865
55
0.003
● Shorea singkawang 869
99
○ BABRU2376-15|Parashorea malaanonan
0.006
○ BABRU2434-15|Parashorea malaanonan
● Parashorea lucida 2941
87
○ BABRU1659-15|Parashorea tomentella
68
91
○ BABRU1694-15|Parashorea tomentella
0.003
● Parashorea lucida 2940
64
● Parashorea cf lucida 901
● Parashorea cf lucida 992
● Shorea bracteolata 898
0.005
66
○ BABRU036-14|Hopea nervosa
○ BABRU465-14|Hopea nervosa
89
0.003
99
66
○ BABRU060-14|Hopea bracteata
○ BABRU099-14|Hopea bracteata
0.007
● Hopea beccariana 4569
62
● Hopea ferruginea 4231
0.004
● Hopea ferruginea 5089
100
○ BABRU1378-15|Dipterocarpus caudatus
70
○ BABRU2158-15|Dipterocarpus cf caudiferus
0.013
○ BABRU2178-15|Dipterocarpus cf caudiferus
75
○ BABRU3918-15|Anisoptera laevis
99
○ BABRU4044-15|Anisoptera laevis
0.004
● Anisoptera costata 4967
0.005
97
87
○ BABRU1110-14|Vatica endertii
0.006
67
○ BABRU1275-14|Vatica endertii
99
95
○ BABRU1348-15|Vatica dulitensis
0.006
0.004
○ BABRU1355-15|Vatica dulitensis
● Dipterocarpaceae sp27 5168
100
○ Marquesia sp JH-2017 isolate 2685
0.014
0.018
97
○ Monotes adenophyllus isolate 598
0.005
○ Monotes kerstingii isolate 1808
0.0050

Figure 12 The maximum likelihood phylogenetic tree of trees' samples collected in the EFForTS project based on the concatenated *rbcL* and *matK* markers, and the Hasegawa-Kishino-Yano model. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. The tree log likelihood is (-2708.98). The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.

Parashorea samples, ID 2941 (*Parashorea lucida*), 2940 (*Parashorea lucida*), 901 (*Parashorea cf. lucida*) and 992 (*Parashorea cf. lucida*), clustered strongly to *Parashorea tomentella* downloaded samples, the same way it has occurred in Figure 6 and Figure 9 with similar support values.

In this figure, the *Hopea* samples, ID 4569 (*Hopea beccariana*), 4231 (*Hopea ferruginea*) and 5089 (*Hopea ferruginea*), grouped together in a subcluster within a clade containing the rest of *Hopea* samples. This also was present in the neighbor joining phylogenetic tree (Figure 6) and maximum parsimony phylogenetic tree (Figure 9) with the same configuration pattern.

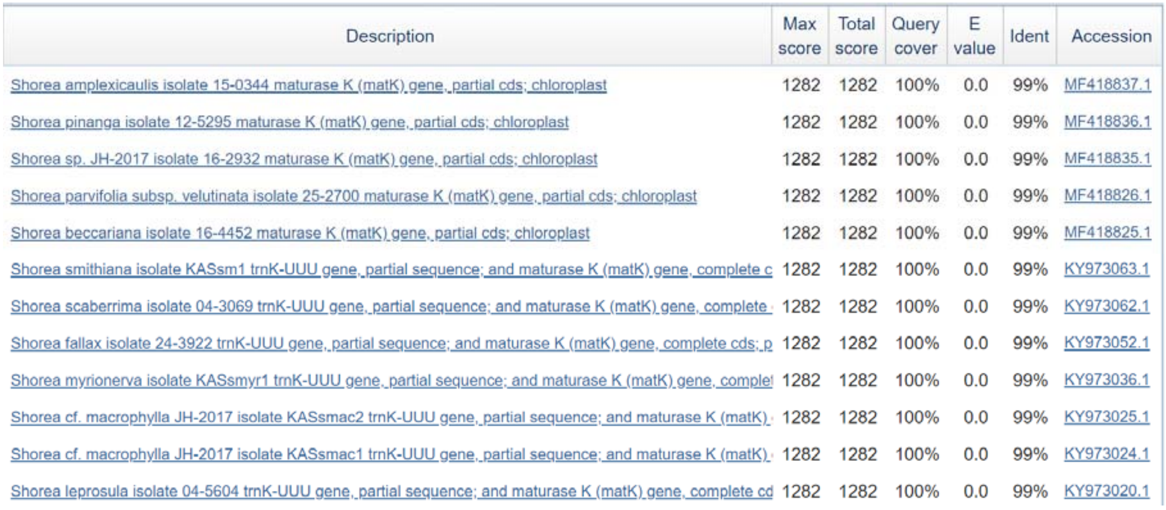
Sample ID 4967 (Anisoptera costata) and simple ID 5168 (Dipterocarpaceae sp. 27) also manifested the same configuration pattern as seen in previous concatenated figures, strongly associating to *Anisoptera* genus database samples and *Vatica* genus database samples, respectively.

Most of the constructed phylogenetic trees exhibited a similar configuration on their topologies and clades arrangement, especially the concatenated phylogenetic trees. Despite of being constructed with three different phylogenetic methods, the results were much alike consistent and showing interesting relationships among the analyzed data.

Nonetheless, two samples in particular presented an attractive pattern amongst the constructed phylogenetic trees. Sample ID 4565 labeled as *Hopea sangal* clustered in almost all trees with the sample ID 896 (*Shorea parvifolia*) and, in some cases, with two more database samples belonging to *Shorea parvifolia* species. After doing a BLAST search for both markers, it is clear the reason of these associations (Image 12 and Image 13).

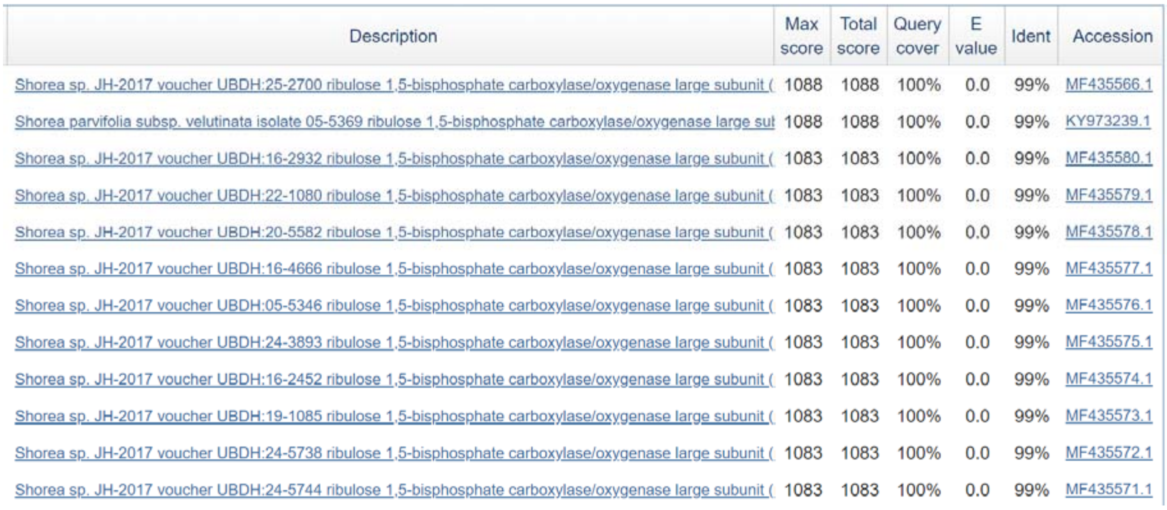
In both BLAST searches the results had a full query cover and a 99% identity on the nucleotide level for *Shorea* genus species, one of them being *Shorea parvifolia*. Since *Shorea* genus and *Hopea* genus belong to the same Imbricate *Shoreae* group, it could be possible a mislabel or misidentification error for sample 4565.

Another attractive behavior is the shown by sample ID 5168 (*Dipterocarpaceae sp. 27*), which at the time of writing this study, has not been identify to a species level. However, a BLAST search was performed with both markers to at least have a notion on the genus from which this sample belongs to (Image 14 and Image 15).



Description	Max score	Total score	Query cover	E value	Ident	Accession
Shorea amplexicaulis isolate 15-0344 maturase K (matK) gene, partial cds; chloroplast	1282	1282	100%	0.0	99%	MF418837.1
Shorea pinanga isolate 12-5295 maturase K (matK) gene, partial cds; chloroplast	1282	1282	100%	0.0	99%	MF418836.1
Shorea sp. JH-2017 isolate 16-2932 maturase K (matK) gene, partial cds; chloroplast	1282	1282	100%	0.0	99%	MF418835.1
Shorea parvifolia subsp. velutinata isolate 25-2700 maturase K (matK) gene, partial cds; chloroplast	1282	1282	100%	0.0	99%	MF418826.1
Shorea beccariana isolate 16-4452 maturase K (matK) gene, partial cds; chloroplast	1282	1282	100%	0.0	99%	MF418825.1
Shorea smithiana isolate KASsm1 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complete c	1282	1282	100%	0.0	99%	KY973063.1
Shorea scaberrima isolate 04-3069 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complete	1282	1282	100%	0.0	99%	KY973062.1
Shorea fallax isolate 24-3922 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complete cds; p	1282	1282	100%	0.0	99%	KY973052.1
Shorea myrionerva isolate KASsmyr1 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complet	1282	1282	100%	0.0	99%	KY973036.1
Shorea cf. macrophylla JH-2017 isolate KASsmac2 trnK-UUU gene, partial sequence; and maturase K (matK)	1282	1282	100%	0.0	99%	KY973025.1
Shorea cf. macrophylla JH-2017 isolate KASsmac1 trnK-UUU gene, partial sequence; and maturase K (matK)	1282	1282	100%	0.0	99%	KY973024.1
Shorea leprosula isolate 04-5604 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complete cd	1282	1282	100%	0.0	99%	KY973020.1

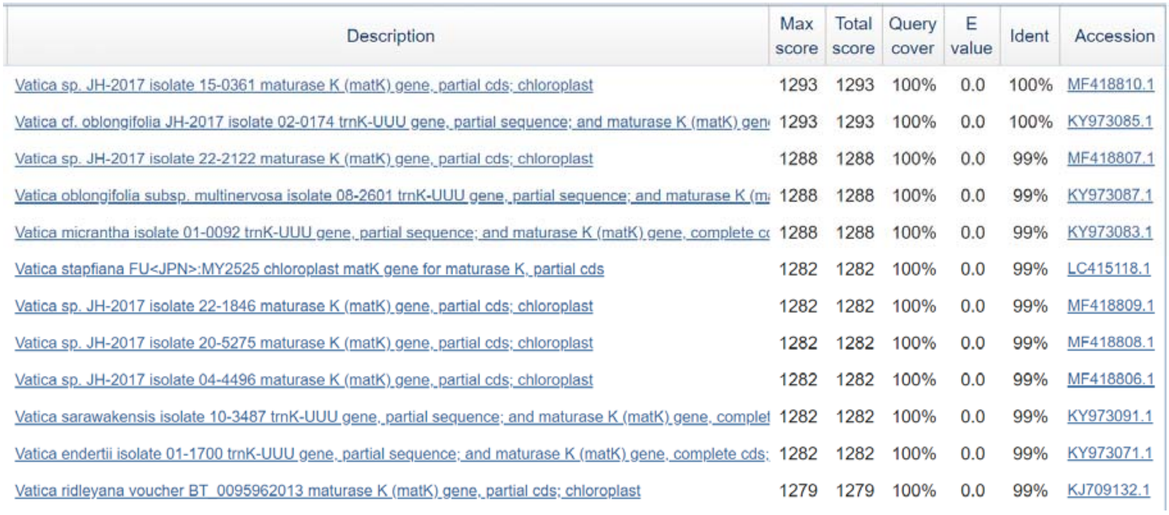
Image 12 Sequences producing significant alignments for the sample ID 4565 (*Hopea sangal*) with *matK* marker. Image taken from BLAST webpage: https://blast.ncbi.nlm.nih.gov/Blast.cgi



Description	Max score	Total score	Query cover	E value	Ident	Accession
Shorea sp. JH-2017 voucher UBDH:25-2700 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (	1088	1088	100%	0.0	99%	MF435566.1
Shorea parvifolia subsp. velutinata isolate 05-5369 ribulose 1,5-bisphosphate carboxylase/oxygenase large sul	1088	1088	100%	0.0	99%	KY973239.1
Shorea sp. JH-2017 voucher UBDH:16-2932 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (	1083	1083	100%	0.0	99%	MF435580.1
Shorea sp. JH-2017 voucher UBDH:22-1080 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (	1083	1083	100%	0.0	99%	MF435579.1
Shorea sp. JH-2017 voucher UBDH:20-5582 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (	1083	1083	100%	0.0	99%	MF435578.1
Shorea sp. JH-2017 voucher UBDH:16-4666 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (	1083	1083	100%	0.0	99%	MF435577.1
Shorea sp. JH-2017 voucher UBDH:05-5346 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (	1083	1083	100%	0.0	99%	MF435576.1
Shorea sp. JH-2017 voucher UBDH:24-3893 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (	1083	1083	100%	0.0	99%	MF435575.1
Shorea sp. JH-2017 voucher UBDH:16-2452 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (	1083	1083	100%	0.0	99%	MF435574.1
Shorea sp. JH-2017 voucher UBDH:19-1085 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (	1083	1083	100%	0.0	99%	MF435573.1
Shorea sp. JH-2017 voucher UBDH:24-5738 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (	1083	1083	100%	0.0	99%	MF435572.1
Shorea sp. JH-2017 voucher UBDH:24-5744 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (	1083	1083	100%	0.0	99%	MF435571.1

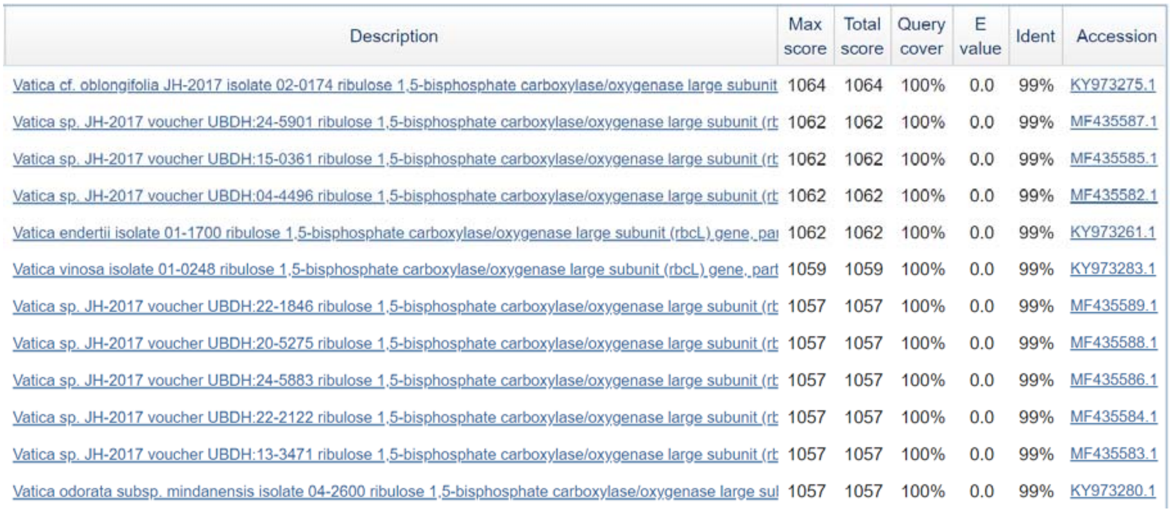
Image 13 Sequences producing significant alignments for the sample ID 4565 (*Hopea sangal*) with *rbcL* marker. Image taken from BLAST webpage: https://blast.ncbi.nlm.nih.gov/Blast.cgi

Interestingly, in both BLAST searches the results for the sample ID 5168 (*Dipterocarpaceae sp. 27*) fully belonged to *Vatica* genus. The query cover was perfect for every sequence and the identity on the nucleotide level ranged from 99% to 100%. Furthermore, these results can explain the clustering behavior of this sample to the downloaded database *Vatica* samples in all constructed phylogenetic trees.



Description	Max score	Total score	Query cover	E value	Ident	Accession
Vatica sp. JH-2017 isolate 15-0361 maturase K (matK) gene, partial cds; chloroplast	1293	1293	100%	0.0	100%	MF418810.1
Vatica cf. oblongifolia JH-2017 isolate 02-0174 trnK-UUU gene, partial sequence; and maturase K (matK) gen	1293	1293	100%	0.0	100%	KY973085.1
Vatica sp. JH-2017 isolate 22-2122 maturase K (matK) gene, partial cds; chloroplast	1288	1288	100%	0.0	99%	MF418807.1
Vatica oblongifolia subsp. multinervosa isolate 08-2601 trnK-UUU gene, partial sequence; and maturase K (m	1288	1288	100%	0.0	99%	KY973087.1
Vatica micrantha isolate 01-0092 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complete c	1288	1288	100%	0.0	99%	KY973083.1
Vatica stapfiana FU<JPN>:MY2525 chloroplast matK gene for maturase K, partial cds	1282	1282	100%	0.0	99%	LC415118.1
Vatica sp. JH-2017 isolate 22-1846 maturase K (matK) gene, partial cds; chloroplast	1282	1282	100%	0.0	99%	MF418809.1
Vatica sp. JH-2017 isolate 20-5275 maturase K (matK) gene, partial cds; chloroplast	1282	1282	100%	0.0	99%	MF418808.1
Vatica sp. JH-2017 isolate 04-4496 maturase K (matK) gene, partial cds; chloroplast	1282	1282	100%	0.0	99%	MF418806.1
Vatica sarawakensis isolate 10-3487 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complet	1282	1282	100%	0.0	99%	KY973091.1
Vatica endertii isolate 01-1700 trnK-UUU gene, partial sequence; and maturase K (matK) gene, complete cds;	1282	1282	100%	0.0	99%	KY973071.1
Vatica ridleyana voucher BT_0095962013 maturase K (matK) gene, partial cds; chloroplast	1279	1279	100%	0.0	99%	KJ709132.1

Image 14 Sequences producing significant alignments for the sample ID 5168 (*Dipterocarpaceae sp. 27*) with *matK* marker. Image taken from BLAST webpage: https://blast.ncbi.nlm.nih.gov/Blast.cgi



Description	Max score	Total score	Query cover	E value	Ident	Accession
Vatica cf. oblongifolia JH-2017 isolate 02-0174 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	1064	1064	100%	0.0	99%	KY973275.1
Vatica sp. JH-2017 voucher UBDH:24-5901 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rt	1062	1062	100%	0.0	99%	MF435587.1
Vatica sp. JH-2017 voucher UBDH:15-0361 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rt	1062	1062	100%	0.0	99%	MF435585.1
Vatica sp. JH-2017 voucher UBDH:04-4496 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rt	1062	1062	100%	0.0	99%	MF435582.1
Vatica endertii isolate 01-1700 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, pa	1062	1062	100%	0.0	99%	KY973261.1
Vatica vinosa isolate 01-0248 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, part	1059	1059	100%	0.0	99%	KY973283.1
Vatica sp. JH-2017 voucher UBDH:22-1846 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rt	1057	1057	100%	0.0	99%	MF435589.1
Vatica sp. JH-2017 voucher UBDH:20-5275 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rt	1057	1057	100%	0.0	99%	MF435588.1
Vatica sp. JH-2017 voucher UBDH:24-5883 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rt	1057	1057	100%	0.0	99%	MF435586.1
Vatica sp. JH-2017 voucher UBDH:22-2122 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rt	1057	1057	100%	0.0	99%	MF435584.1
Vatica sp. JH-2017 voucher UBDH:13-3471 ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rt	1057	1057	100%	0.0	99%	MF435583.1
Vatica odorata subsp. mindanensis isolate 04-2600 ribulose 1,5-bisphosphate carboxylase/oxygenase large sul	1057	1057	100%	0.0	99%	KY973280.1

Image 15 Sequences producing significant alignments for the sample ID 5168 (*Dipterocarpaceae sp. 27*) with *rbcL* marker. Image taken from BLAST webpage: https://blast.ncbi.nlm.nih.gov/Blast.cgi

3.2. POLLEN AND HONEY SAMPLES RESULTS

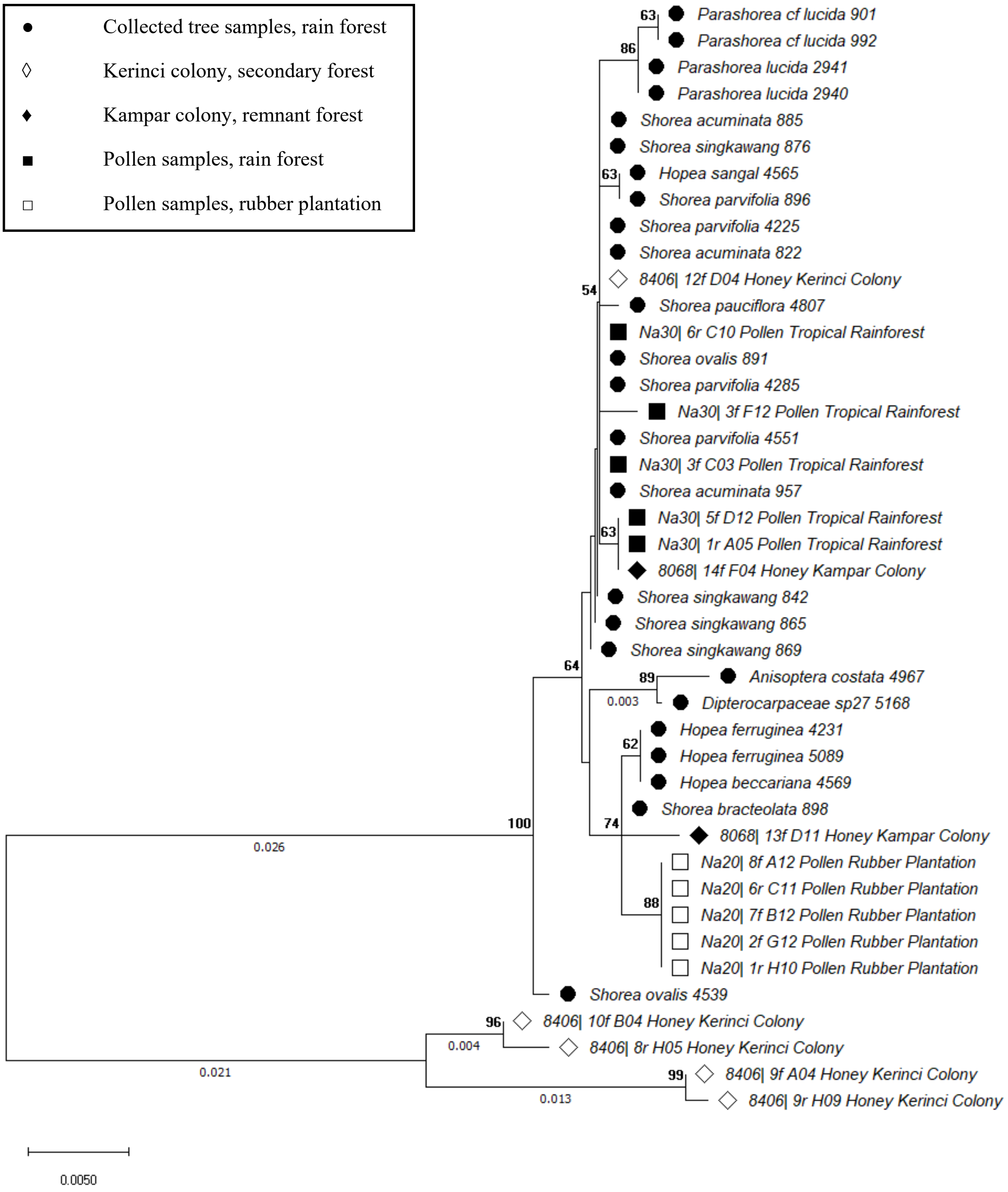
The collected tree samples in the EFForTS project were used in a phylogenetic tree to examine the relationships with the collected honey and pollen samples (Figure 13), since both samples were gathered in the same region. The initial created phylogenetic tree was obtained by the Neighbor Joining method with the same configuration used with the leaf samples. Only the *rbcL* marker sequences were used since the honey and pollen samples were PCR amplified and sequenced with this marker.

Most of the honey samples from the Kerinci colony behaved as an outgroup from the rest of the data set, clustering at the bottom of the figure and inferring there is not a direct association with the *Dipterocarpaceae* family. Moreover, after doing a BLAST search to the samples forming this cluster the results suggested they belonged to the *Asteraceae* family of plants, which is a large family of mostly herbaceous flowering plants.

A new neighbor joining phylogenetic tree was constructed with some database *Asteraceae* samples of the most known species from the same region where the tree, honey and pollens samples were collected (Figure 14).

Clearly, the majority of the honey samples from the Kerinci colony (8406) are weakly related to the *Asteraceae* samples, especially, forming a subcluster to the *Sphagneticola trilobata*, *Synedrella nodiflora* and *Clibadium alatum* database samples. On the other hand, pollen samples from the rubber plantation (NA 20) can be traced back and acceptably related to the *Hopea* samples, ID 4231 (*Hopea ferruginea*), 5089 (*Hopea ferruginea*) and 4569 (*Hopea beccariana*), as well as weakly related to sample ID 898 (*Shorea bracteolata*) and a honey sample from Kampar colony (8068).

Moreover, a relationship amongst pollen from tropical rain forest (NA 30) and honey from Kampar colony (8068) can be observed with an acceptable support value of 64% in a subcluster in the first clade containing *Shorea* samples. The honey in Kampar colony (8068) could have an origin in the tropical rainforest pollen samples (NA 30). The rest of the honey and pollen samples did not group to any particular tree sample species but they all were placed amongst the *Shorea* genus samples.



● Collected tree samples, rain forest
◊ Kerinci colony, secondary forest
♦ Kampar colony, remnant forest
■ Pollen samples, rain forest
□ Pollen samples, rubber plantation
63
86
Parashorea cf lucida 901
Parashorea cf lucida 992
Parashorea lucida 2941
Parashorea lucida 2940
Shorea acuminata 885
Shorea singkawang 876
63
Hopea sangal 4565
Shorea parvifolia 896
Shorea parvifolia 4225
Shorea acuminata 822
8406| 12f D04 Honey Kerinci Colony
54
Shorea pauciflora 4807
Na30| 6r C10 Pollen Tropical Rainforest
Shorea ovalis 891
Shorea parvifolia 4285
Na30| 3f F12 Pollen Tropical Rainforest
Shorea parvifolia 4551
Na30| 3f C03 Pollen Tropical Rainforest
Shorea acuminata 957
Na30| 5f D12 Pollen Tropical Rainforest
63
Na30| 1r A05 Pollen Tropical Rainforest
8068| 14f F04 Honey Kampar Colony
Shorea singkawang 842
Shorea singkawang 865
Shorea singkawang 869
64
89
Anisoptera costata 4967
0.003
Dipterocarpaceae sp27 5168
Hopea ferruginea 4231
62
Hopea ferruginea 5089
Hopea beccariana 4569
Shorea bracteolata 898
100
74
8068| 13f D11 Honey Kampar Colony
0.026
Na20| 8f A12 Pollen Rubber Plantation
Na20| 6r C11 Pollen Rubber Plantation
88
Na20| 7f B12 Pollen Rubber Plantation
Na20| 2f G12 Pollen Rubber Plantation
Na20| 1r H10 Pollen Rubber Plantation
Shorea ovalis 4539
96
8406| 10f B04 Honey Kerinci Colony
0.004
8406| 8r H05 Honey Kerinci Colony
0.021
99
8406| 9f A04 Honey Kerinci Colony
0.013
8406| 9r H09 Honey Kerinci Colony
0.0050

Figure 13 The neighbor joining phylogenetic tree of honey, pollen and tree samples collected in the EFForTS project based on the *rbcL* marker, with genetic distances computed using the Maximum Composite Likelihood method. The numbers at the tree nodes represent bootstrap values based on 1000 replicates. ● - the plot tree samples located in a rain forest; ◊ - honey samples from Kerinci colony, located in a secondary forest partly surrounded by a pristine forest; ♦ - honey samples from Kampar colony, located in a remnant forest surrounded by plantations of Eucalyptus, oil palm (*Elaeis guineensis*) and some Acacia; ■ – pollen samples from pollen traps located in a rain forest; □ – pollen samples from pollen traps located in a jungle rubber plantation.

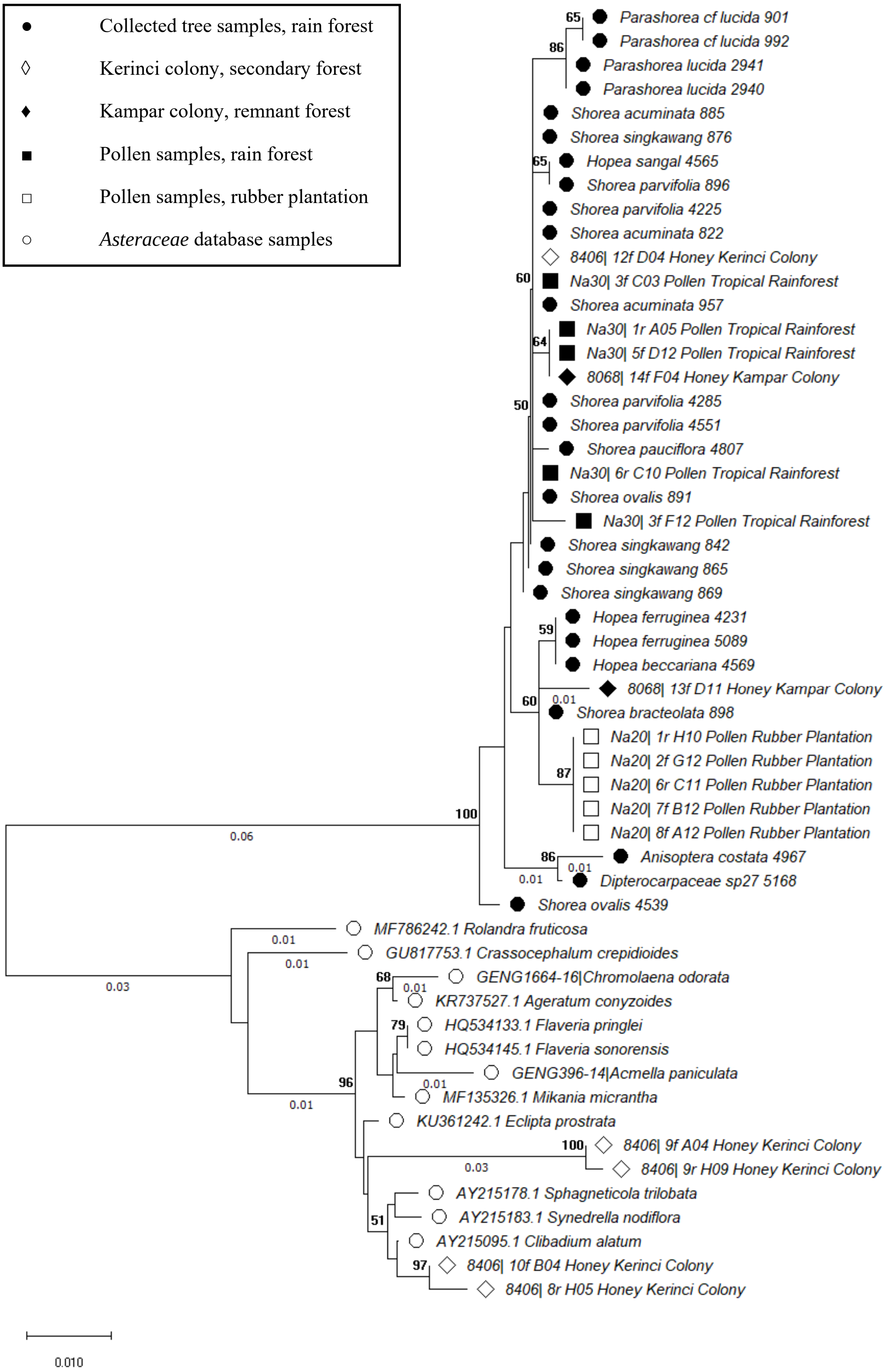

● Collected tree samples, rain forest
◊ Kerinci colony, secondary forest
♦ Kampar colony, remnant forest
■ Pollen samples, rain forest
□ Pollen samples, rubber plantation
○ *Asteraceae* database samples
65
86
Parashorea cf lucida 901
Parashorea cf lucida 992
Parashorea lucida 2941
Parashorea lucida 2940
Shorea acuminata 885
Shorea singkawang 876
65
Hopea sangal 4565
Shorea parvifolia 896
Shorea parvifolia 4225
Shorea acuminata 822
8406| 12f D04 Honey Kerinci Colony
60
Na30| 3f C03 Pollen Tropical Rainforest
Shorea acuminata 957
64
Na30| 1r A05 Pollen Tropical Rainforest
Na30| 5f D12 Pollen Tropical Rainforest
8068| 14f F04 Honey Kampar Colony
50
Shorea parvifolia 4285
Shorea parvifolia 4551
Shorea pauciflora 4807
Na30| 6r C10 Pollen Tropical Rainforest
Shorea ovalis 891
Na30| 3f F12 Pollen Tropical Rainforest
Shorea singkawang 842
Shorea singkawang 865
Shorea singkawang 869
59
Hopea ferruginea 4231
Hopea ferruginea 5089
Hopea beccariana 4569
8068| 13f D11 Honey Kampar Colony
60
0.01
Shorea bracteolata 898
87
Na20| 1r H10 Pollen Rubber Plantation
Na20| 2f G12 Pollen Rubber Plantation
Na20| 6r C11 Pollen Rubber Plantation
Na20| 7f B12 Pollen Rubber Plantation
Na20| 8f A12 Pollen Rubber Plantation
100
0.06
86
0.01
Anisoptera costata 4967
0.01
Dipterocarpaceae sp27 5168
Shorea ovalis 4539
0.01
MF786242.1 Rolandra fruticosa
0.01
GU817753.1 Crassocephalum crepidioides
0.03
68
0.01
GENG1664-16|Chromolaena odorata
KR737527.1 Ageratum conyzoides
79
HQ534133.1 Flaveria pringlei
HQ534145.1 Flaveria sonorensis
GENG396-14|Acmella paniculata
96
0.01
MF135326.1 Mikania micrantha
0.01
KU361242.1 Eclipta prostrata
100
8406| 9f A04 Honey Kerinci Colony
0.03
8406| 9r H09 Honey Kerinci Colony
AY215178.1 Sphagneticola trilobata
51
AY215183.1 Synedrella nodiflora
AY215095.1 Clibadium alatum
97
8406| 10f B04 Honey Kerinci Colony
8406| 8r H05 Honey Kerinci Colony
0.010

Figure 14 The neighbor joining phylogenetic tree of honey, pollen and tree samples collected in the EFForTS project and additional *Asteraceae* database samples, based on the *rbcL* marker with genetic distances computed using the Maximum Composite Likelihood method. The numbers at the tree nodes represent bootstrap values based on 1000

replicates. ● - the plot tree samples located in a rain forest; ◊ - honey samples from Kerinci colony, located in a secondary forest partly surrounded by a pristine forest; ♦ - honey samples from Kampar colony, located in a remnant forest surrounded by plantations of Eucalyptus, oil palm (*Elaeis guineensis*) and some Acacia; ■ - pollen samples from pollen traps located in a rain forest; □ - pollen samples from pollen traps located in a jungle rubber plantation; ○ - database samples from BOLDSYSTEMS.

The third phylogenetic tree was created using Maximum Parsimony method (Figure 15) to interpret the previous results with a different angle. The topology is similar to the observed in Figure 14, with one main different feature, the *Hopea* samples, ID 4231 (*Hopea ferruginea*), 5089 (*Hopea ferruginea*) and 4569 (*Hopea beccariana*) did not group to any sample. In contrast, the pollen samples from rubber plantation (NA 20) were once more associated to the sample ID 898 (*Shorea bracteolata*) and a honey sample from Kampar colony (8068), with lower than 50% support values.

The subcluster containing two samples of pollen from tropical rain forest (NA 30) and a honey sample from Kampar colony (8068) had a stronger relationship than the previous stated in Figure 14.

In the case of the lower clade of *Asteraceae* database samples, the honey samples from Kerinci colony (8406) samples once more showed a weak relationship to the *Sphagneticola trilobata*, *Synedrella nodiflora* and *Clibadium alatum* database samples. Additionally, it is noticeable that two of the honey samples from Kerinci colony (8406) showed a larger evolutionary distance against the rest of samples, as it also occurred in Figure 14, which could mean the honey mixture comes from more different species of *Asteraceae*.

The rest of samples, three pollen samples from rain forest (NA 30) and one honey sample from Kerinci colony (8406), were placed amongst the first clade of *Shorea* samples without associating to any of the database samples, just as in Figure 14. The placement amongst the *Shorea* clade could mean the samples belong to a different *Shorea* species not present in the sampled trees.

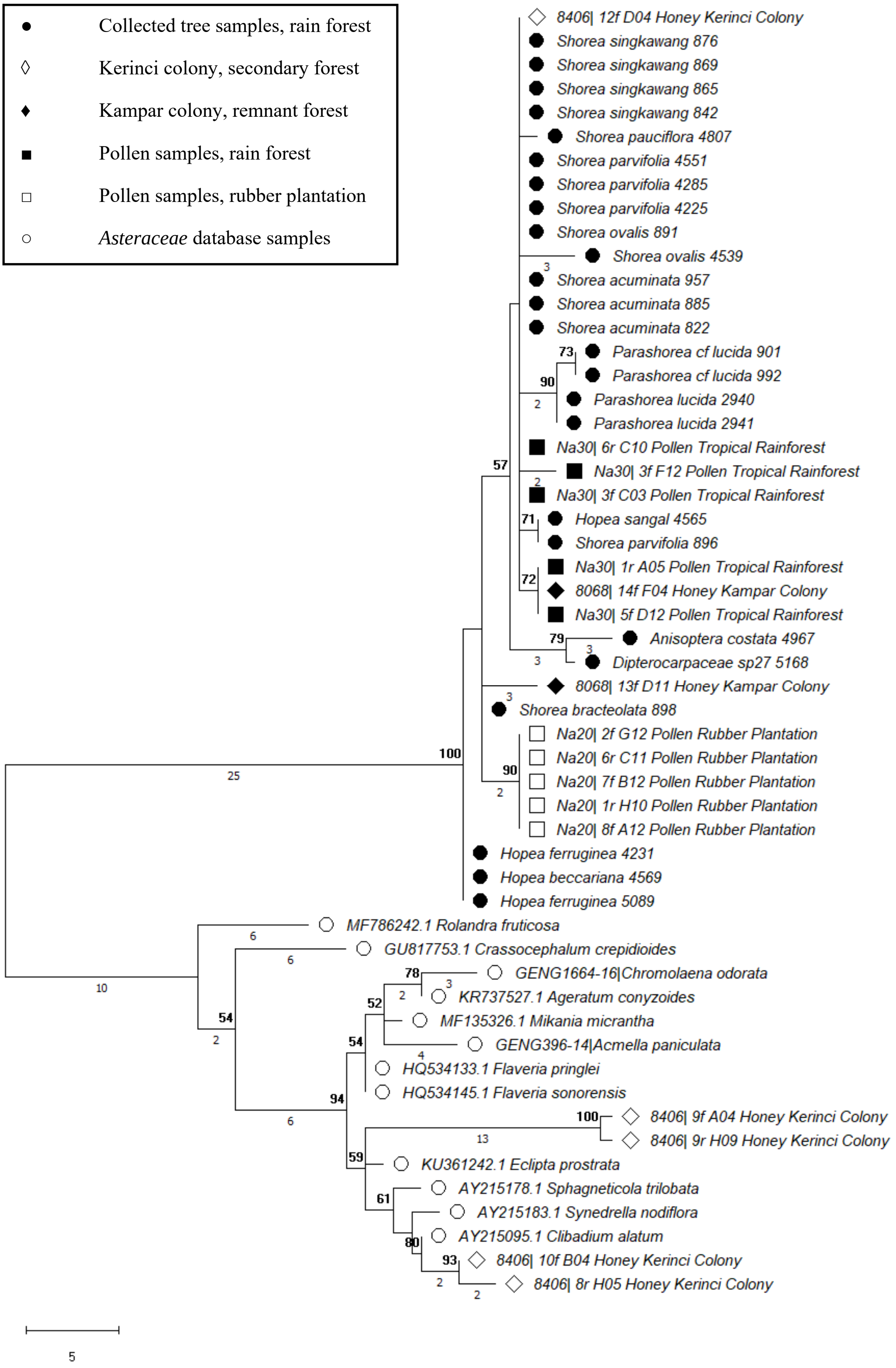

● Collected tree samples, rain forest
◊ Kerinci colony, secondary forest
♦ Kampar colony, remnant forest
■ Pollen samples, rain forest
□ Pollen samples, rubber plantation
○ *Asteraceae* database samples
8406| 12f D04 Honey Kerinci Colony
Shorea singkawang 876
Shorea singkawang 869
Shorea singkawang 865
Shorea singkawang 842
Shorea pauciflora 4807
Shorea parvifolia 4551
Shorea parvifolia 4285
Shorea parvifolia 4225
Shorea ovalis 891
Shorea ovalis 4539
3
Shorea acuminata 957
Shorea acuminata 885
Shorea acuminata 822
73
Parashorea cf lucida 901
90
Parashorea cf lucida 992
2
Parashorea lucida 2940
Parashorea lucida 2941
Na30| 6r C10 Pollen Tropical Rainforest
57
Na30| 3f F12 Pollen Tropical Rainforest
2
Na30| 3f C03 Pollen Tropical Rainforest
71
Hopea sangal 4565
Shorea parvifolia 896
Na30| 1r A05 Pollen Tropical Rainforest
72
8068| 14f F04 Honey Kampar Colony
Na30| 5f D12 Pollen Tropical Rainforest
79
Anisoptera costata 4967
3
3
Dipterocarpaceae sp27 5168
8068| 13f D11 Honey Kampar Colony
3
Shorea bracteolata 898
Na20| 2f G12 Pollen Rubber Plantation
100
Na20| 6r C11 Pollen Rubber Plantation
25
90
Na20| 7f B12 Pollen Rubber Plantation
2
Na20| 1r H10 Pollen Rubber Plantation
Na20| 8f A12 Pollen Rubber Plantation
Hopea ferruginea 4231
Hopea beccariana 4569
Hopea ferruginea 5089
MF786242.1 Rolandra fruticosa
6
GU817753.1 Crassocephalum crepidioides
6
78
GENG1664-16|Chromolaena odorata
10
3
2
KR737527.1 Ageratum conyzoides
52
54
MF135326.1 Mikania micrantha
2
54
GENG396-14|Acmella paniculata
4
HQ534133.1 Flaveria pringlei
94
HQ534145.1 Flaveria sonorensis
6
100
8406| 9f A04 Honey Kerinci Colony
13
8406| 9r H09 Honey Kerinci Colony
59
KU361242.1 Eclipta prostrata
AY215178.1 Sphagneticola trilobata
61
AY215183.1 Synedrella nodiflora
80
AY215095.1 Clibadium alatum
93
8406| 10f B04 Honey Kerinci Colony
2
8406| 8r H05 Honey Kerinci Colony
2
5

Figure 15 The maximum parsimony phylogenetic tree of honey, pollen and tree samples collected in the EFForTS project and additional *Asteraceae* database samples, based on the *rbcL* marker and obtained by using the Subtree-Pruning-Regrafting (SPR) algorithm. The numbers at the tree nodes represent bootstrap values based on 1000

replicates. ● - the plot tree samples located in a rain forest; ◊ - honey samples from Kerinci colony, located in a secondary forest partly surrounded by a pristine forest; ♦ - honey samples from Kampar colony, located in a remnant forest surrounded by plantations of Eucalyptus, oil palm (*Elaeis guineensis*) and some Acacia; ■ - pollen samples from pollen traps located in a rain forest; □ - pollen samples from pollen traps located in a jungle rubber plantation; ○ - database samples from BOLDSYSTEMS.

Lastly, another phylogenetic tree was created using the Maximum Likelihood method (Figure 16). The topology and arrangement of the clades were very similar to the maximum parsimony phylogenetic tree (Figure 15).

The same pollen samples from tropical rain forest (NA 30) and the one honey sample from Kerinci colony (8406) were once more amongst the first clade of *Shorea* samples, with no direct association to a particular tree species.

With the same support values seen in Figure 14, a subcluster consisting of two pollen samples from rain forest (NA 30) and a honey sample from Kampar colony (8068), was present among the *Shorea* clade. Having this subcluster present in all figures could indicate the confirmation of a correlation between the honey sample in Kampar colony and the honey sample collected in the rain forest, as well as their linkage to the *Shorea* tree species.

Another present constant is the cluster of pollen samples from the jungle rubber plantation (NA 20), all of the samples are strongly related to each other and without visible relation to sample ID 898 (*Shorea bracteolata*) and a honey sample from Kampar colony (8068).

The larger part of pollen samples from the Kerinci colony (8406) clustered with the *Asteraceae* database samples, just like in previous figures. The samples in question are forming a low support subcluster with *Sphagneticola trilobata*, *Synedrella nodiflora* and *Clibadium alatum* database samples.

All of the results were consistent among the different used phylogenetic construction methods, which demonstrates to a certain point the fidelity of the observed relationships among the samples.

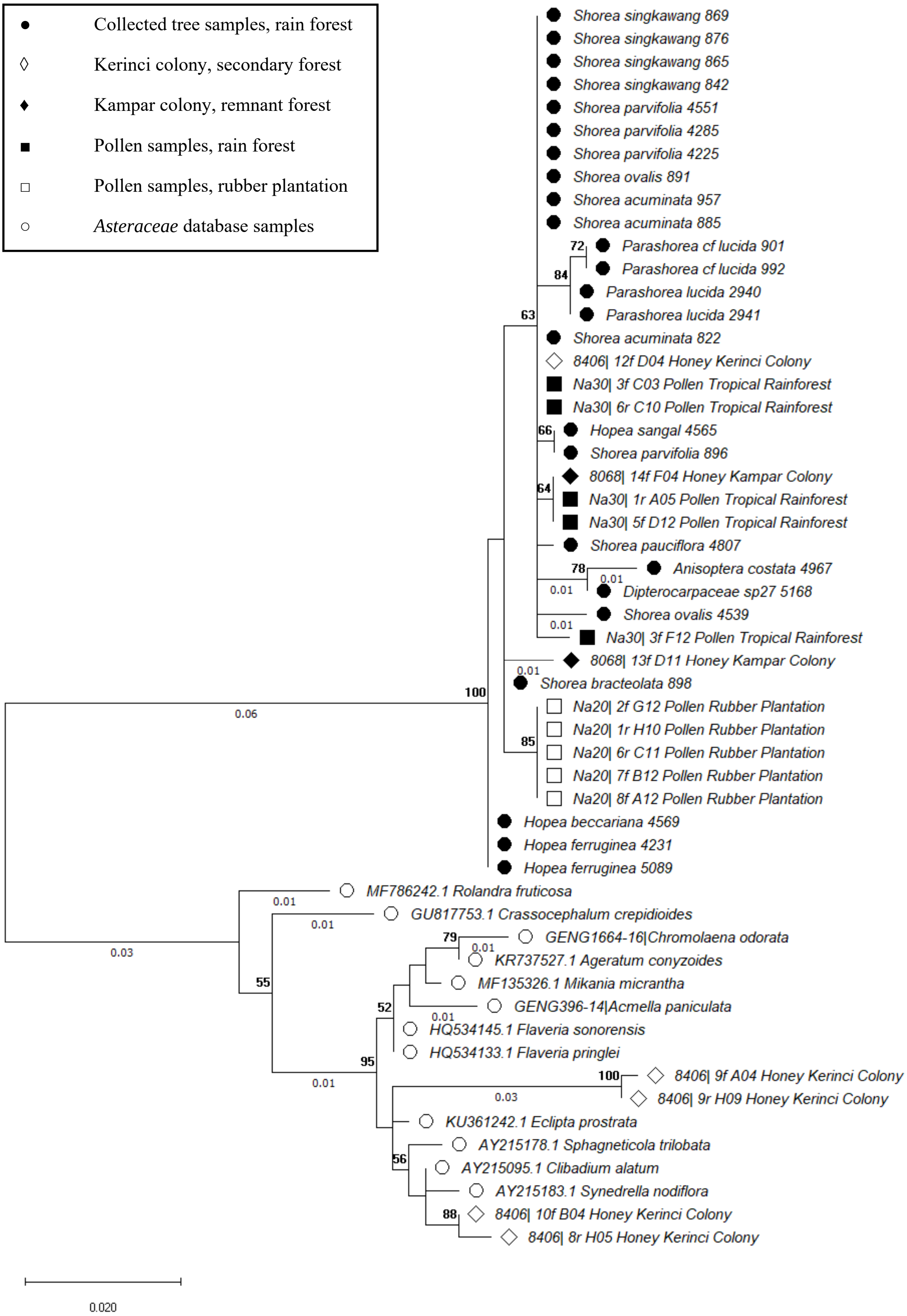

● Collected tree samples, rain forest
◊ Kerinci colony, secondary forest
♦ Kampar colony, remnant forest
■ Pollen samples, rain forest
□ Pollen samples, rubber plantation
○ *Asteraceae* database samples
Shorea singkawang 869
Shorea singkawang 876
Shorea singkawang 865
Shorea singkawang 842
Shorea parvifolia 4551
Shorea parvifolia 4285
Shorea parvifolia 4225
Shorea ovalis 891
Shorea acuminata 957
Shorea acuminata 885
72
Parashorea cf lucida 901
Parashorea cf lucida 992
84
Parashorea lucida 2940
Parashorea lucida 2941
63
Shorea acuminata 822
8406| 12f D04 Honey Kerinci Colony
Na30| 3f C03 Pollen Tropical Rainforest
Na30| 6r C10 Pollen Tropical Rainforest
66
Hopea sangal 4565
Shorea parvifolia 896
8068| 14f F04 Honey Kampar Colony
64
Na30| 1r A05 Pollen Tropical Rainforest
Na30| 5f D12 Pollen Tropical Rainforest
Shorea pauciflora 4807
78
Anisoptera costata 4967
0.01
0.01
Dipterocarpaceae sp27 5168
Shorea ovalis 4539
0.01
Na30| 3f F12 Pollen Tropical Rainforest
8068| 13f D11 Honey Kampar Colony
0.01
100
Shorea bracteolata 898
0.06
Na20| 2f G12 Pollen Rubber Plantation
Na20| 1r H10 Pollen Rubber Plantation
85
Na20| 6r C11 Pollen Rubber Plantation
Na20| 7f B12 Pollen Rubber Plantation
Na20| 8f A12 Pollen Rubber Plantation
Hopea beccariana 4569
Hopea ferruginea 4231
Hopea ferruginea 5089
MF786242.1 Rolandra fruticosa
0.01
GU817753.1 Crassocephalum crepidioides
0.01
0.03
79
GENG1664-16|Chromolaena odorata
0.01
KR737527.1 Ageratum conyzoides
55
MF135326.1 Mikania micrantha
52
GENG396-14|Acmella paniculata
0.01
HQ534145.1 Flaveria sonorensis
95
HQ534133.1 Flaveria pringlei
0.01
100
8406| 9f A04 Honey Kerinci Colony
0.03
8406| 9r H09 Honey Kerinci Colony
KU361242.1 Eclipta prostrata
AY215178.1 Sphagneticola trilobata
56
AY215095.1 Clibadium alatum
AY215183.1 Synedrella nodiflora
88
8406| 10f B04 Honey Kerinci Colony
8406| 8r H05 Honey Kerinci Colony
0.020

Figure 16 The maximum likelihood phylogenetic tree of honey, pollen and tree samples collected in the EFForTS project and additional *Asteraceae* database samples, based on the *rbcL* marker and the Hasegawa-Kishino-Yano model. The numbers at the tree

nodes represent bootstrap values based on 1000 replicates. The tree log likelihood is (-1368.42). ● - the plot tree samples located in a rain forest; ◊ - honey samples from Kerinci colony, located in a secondary forest partly surrounded by a pristine forest; ♦ - honey samples from Kampar colony, located in a remnant forest surrounded by plantations of Eucalyptus, oil palm (*Elaeis guineensis*) and some Acacia; ■ - pollen samples from pollen traps located in a rain forest; □ - pollen samples from pollen traps located in a jungle rubber plantation; ○ - database samples from BOLDSYSTEMS.

4. DISCUSSION

4.1. Leaf Samples

One of the allegedly advantages of DNA barcode, its usefulness for organism diversity studies, has been proposed and demonstrated to be feasible through the use of the short DNA sequence *CO1*, which has led to the premise of practical, standardized, species-level identification in all animal groups for biodiversity assessments. However, the use of the *CO1* genetic marker for land plants barcoding proved to be not successful as in the case of animals because of its low performance (low substitution rates of mitochondrial DNA) in species-level discrimination for flowering plants (Hebert et al., 2003; Kress et al., 2005; CBOL, 2009).

After extensive research, by comparing 7 candidate loci in more than 900 samples representing 445 angiosperm specimen approximately and checking universality, discrimination, sequence quality and coverage criteria, the Consortium for the Barcode of Life (CBOL, 2009) chose two locus DNA barcode markers, i.e. *rbcL* and *matK.*

Despite of Hajibabaei et al. (2007) suggesting the DNA barcoding does not have enough discrimination to create reliable phylogenetic trees for resolving evolutionary relationships among taxa, the present study states the contrary with the two locus DNA genetic markers, *rbcL* and *matK*, applied to the *Dipterocarpaceae* family of trees.

The *matK* genetic marker has a high evolutionary rate, which gives a powerful discriminatory utility among angiosperm species (CBOL, 2009; Li, 2014). Most of the phylogenetic trees constructed based on the *matK* region had an impressive resolution to species-level, giving a broad view of the relationships among *Dipterocarpaceae* species. Besides, in the initial phylogenetic tree based on *matK* matker (Figure 1) the discriminatory power inferred two possibly mislabel or misidentified samples, sample ID 4121 (*Shorea acuminata*) and sample ID 4591 (*Hopea ferruginea*), due to their outgroup behavior. Kajita et al. (1998), used *matK*, *trnL-trnF* IGS plastid regions as DNA barcode markers for constructing phylogenetic trees of samples belonging to *Dipterocarpaceae* family and in their findings, the majority of relationships among genera were resolved.

Nevertheless, the *matK* marker has been reported to have a lower universality, meaning that it is difficult to PCR amplify using the currently known primers (CBOL, 2009; Hollingsworth et al., 2011a). Given that precept, the primers designed by the Forest Genetics and Tree Breeding Department of the Georg-August-Universität Göttingen, were used in this study with higher recovery rates than the *matK* primers recommended by Ki-Joong Kim. In addition, Li et al. (2014) indicated, on their review about single-locus DNA barcodes, a variable rate of discrimination across different taxonomic groups for the *matK* marker, ranging from 49% to 90%.

Contrarily, the *rbcL* marker provides a high universality, which translates in easily recovered PCR amplifications, high-quality bidirectional sequencing and alignment in most land plants. However, *rbcL* marker does not possess the best discriminatory power, having the lowest divergence among plastid genes in flowering plants. Thus, it can perform as a good DNA barcoding region only at a genus or family levels as a single-locus DNA barcode (CBOL, 2009; Hollingsworth et al., 2011a; Li et al., 2014; Techen et al., 2014). The latter can be checked with the observable topology of the constructed phylogenetic trees based on the *rbcL* marker, with some of the relationships not being able to be resolved at species-level. Particularly in Figure 2, the discriminatory power of *rbcL* marker was strong enough to differentiate the three problematic samples, ID 721 (*Shorea acuminata*), 753 (*Shorea acuminata*) and 754 (*Shorea acuminata*), not clustering to the rest of the *Dipterocarpaceae* family, which gave an indication of their belonging to a different family.

Dayanandan et al. (1999) constructed phylogenetic trees based on the *rbcL* DNA barcoding marker and most of the resolved relationships were close to the known phylogeny of *Dipterocarpaceae* family at the time of their research. However, some relationships were not clearly resolved, some *Hopea* genus samples clustered to other *Shorea* samples, as well as unresolved placements of *Dipterocarpus* and *Dryobalanops* genus.

Using the proposed two locus DNA barcode, *rbcL* and *matK*, has proven to be a powerful tool in phylogenetic analyses since it combines the two strong features of both markers. The universality of the *rbcL* marker, which despite of not meeting desired attributes for barcoding itself alone, it is able to perform accurate identifications in combination to other plastid markers; and the discriminatory power of *matK* marker, which has been considered as the plant analogue to *CO1* animal barcode (CBOL, 2009; Hollingsworth et al., 2011a; Li et al., 2014; Kress et al., 2007). Furthermore, the constructed phylogenetic trees based on the concatenation of both plastid regions, *rbcL+matK*, showed a clear resolution on most of the relationships at a genus and species level for *Dipterocarpaceae* family, and even highlighting the possible misidentification of sample ID 4565 labeled as *Hopea sangal* and closely relating to sample ID 896 *Shorea parvifolia*, as well as the strong relationship of sample ID 5168 (*Dipterocarpaceae sp. 27*) to *Vatica* genus samples. Thus, confirming that the core DNA barcode for land plants (*rbcL+matK*) can provide the necessary tools for analysis of relationships among taxa and ultimately, to provide aid at the taxonomic systems.

A recent study (Heckenhauer et al., 2017) has even provided a tentative framework phylogeny for the whole *Dipterocarpaceae* family using three plastid regions (*rbcL*, *trnK-matK-trnK*, *trnT-trnL-trnF*) which included *rbcL* and *matK* markers. Despite of some relationships still remaining unresolved in the *Shoreeae* group, the results showed a wide range of relationships among different *Dipterocarpaceae* genus and other families from which *Dipterocarpaceae* is related (*Pakaraimaea, Cistaceae, Sarcolaenaceae*).

4.2. Pollen and Honey Samples

The most common used approach for identifying the origin of bee's honey is called melissopalynology and, in principle, is a micromorpological analysis of pollen. However, this process can be time consuming, laborious counting procedure and challenging to interpret the results since some plants can be difficult to distinguish (Bruni et al., 2015; Hawkings et al., 2015).

DNA barcoding raises as an applied molecular tool for analyzing the composition of honey and pollen grains. Thus, improving the rates of plant species identification on the honey mixture as it was stated by Bruni et al. (2015), using *rbcL* and *trnH-psbA* markers. Hawkings et al. (2015) used only *rbcL* marker based on its universality characteristic for the possibility to broaden the target taxonomic groups, and found that applying DNA barcoding on honey and pollen grains has the potential to provide identification of floral composition of honey.

As for the case of the present research, the universality of *rbcL* proved to be efficient on acquiring the sequences for the collected pollen and honey samples as it was previously stated by other studies. However, the low discrimination power did not fully resolve to a species level the possible relationships and connections of honey and pollen to the collected tree samples in the constructed phylogenetic trees, but it was enough to interpret that some honey samples (Kerinci colony) did not have a family level relationship to the *Dipterocarpaceae* family. Through the identified common wayside plants of Jambi Province by Katja et al. (2017), it was possible to select the known species of *Asteraceae* family, and download their counterpart from BOLDSYSTEMS database, to include them in the phylogenetic trees for honey and pollen samples. It is possible to confirm that using DNA barcoding markers, even with the lowest discriminatory power, can give an insight on phylogenetic diversity of pollen and honey, especially when cross-referencing the surrounding flora.

5. CONCLUSIONS

There is no doubt the use of DNA barcodes is a promising tool for organisms' identification, especially for land plants. However, this approach cannot be taken as a replacement for taxonomic identification, rather a complementary aid to maximize the time consuming and difficult labor under the classical taxonomic systems.

Constructing phylogenetic trees based on single locus DNA barcodes was not reliable enough to interpret phylogenetic relationships from a family to a species level, even though in some cases it did manifested the necessary discrimination power to realize unexpected outgroups or mislabeled samples. On the contrary, when using the suggested concatenation of two plastid region (*rbcL* and *matK*) by the CBOL (2009) for DNA barcoding analyses, the display of the phylogenetic trees was smoother and the relationships were clear to interpret from all levels. The discriminatory power of *matK* marker and the universality of *rbcL* marker provide a reliable core barcode for land plants, allowing the assessment of phylogenetic relationships.

The topology of the constructed phylogenetic trees, disregarding the used DNA barcode marker, was mainly consistent trough the applied construction methods, Neighbor Joining, Maximum Parsimony and Maximum Likelihood, which could indicate the consistency of the collected data and their relationships. Additionally, the phylogenetic trees based on the concatenated genetic markers (*rbcL* and *matK*) had virtually the same arrangement, topology and relationships. One of the most remarking features was the constant clustering of *Parashorea lucida* samples to another *Parashorea* genus sample (from the database); *Hopea ferruginea* and *Hopea beccariana* samples forming a cluster with database *Hopea* samples; and *Anisoptera costata* sample showing strong relationships to *Anisoptera laevis* database samples; which in turn proves the species level discrimination power of the phylogenetic trees.

However, not all the times the resolution and level of discrimination is narrowed to species-level with the suggested core barcode (*rbcL* and *matK*). There is still the need to increase the discrimination power to all land plants, since their combined discrimination is typically lower than *CO1* in animals (Hollingsworth et al., 2011a). Thus, it is necessary to keep improving and refining the DNA barcode for land plants, modifying primers or adding specific plastid regions to the core barcode as it was suggested by Hollingsworth (2011b), or even developing new plastid regions that fit the universality, discrimination and quality coverage criteria recommended by the CBOL (2009).

On the other hand, constructed phylogenetic trees based on the *rbcL* marker for honey and pollen samples proved the ability to explain phylogenetic diversity when, despite of the low discriminatory power of *rbcL* marker, a clear distinction of two possible origin of honey samples was observed. A second clade of honey samples from Kerinci colony was related to *Asteraceae* family, instead of the first clade consisting of pollen and honey samples correlated to *Dipterocarpaceae* family. Nevertheless, extensive research and the inclusion of additional markers for phylogenetic assessments of honey and pollen samples is required to keep helping taxonomic labor and providing useful identification on floral composition of honey.

6. REFERENCES

Amandita F. (2015). DNA Barcoding of Flowering Plants in Jambi, Indonesia. Ph.D. Thesis. Georg- August University Göttingen.

Ashton P.S. (1982) Dipterocarpaceae. Flora malesiana. Series I. Spermatophyta. 9: 237 - 552.

Avise J. (2006). Evolutionary Pathways in Nature a Phylogenetic Approach. New York, USA. Cambridge University Press.

Behling, H., Cohen, M. & Lara, R. (2001). Studies on Holocene mangrove ecosystem dynamics of the Bragança Peninsula in north-eastern Pará, Brazil. Palaeogeography, Palaeoclimatology, Palaeoecology 167: 225 - 242.

Bruni, I. et al. (2015). A DNA barcoding approach to identify plant species in multiflower honey. Food Chemistry 170: 308 - 315.

Butchart et al. (2010). Global Biodiversity: Indicators of Recent Declines. SCIENCE 328: 1164 -1168.

Cardinale B., Duffy E., Gonzalez A., Hooper D., Perrings C., Venail P., Narwani A., Mace G., Tilman D., Wardle D., Kinzig A., Daily G., Loreau M., Grace J., Larigauderie A., Srivastava D., Naeem S. (2012). Biodiversity loss and its impact on humanity. NATURE 486: 59 - 87.

Center for International Forestry Research (CIFOR). (1998). A review of dipterocarps: taxonomy, ecology and silviculture.

Christenhusz, M. J. & Byng, J. W. (2016). The number of known plants species in the world and its annual increase. Phytotaxa 261 (3): 201–217.

Consortium for the Barcoding of Life (CBOL) Plant Working Group. (2009). A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America (PNAS) 106: 12794 - 12797.

Dayanandan, S., Ashton, P. S., Williams, S. M. & Primack, R. B. (1999). Phylogeny of the tropical tree family Dipterocarpaceae based on nucleotide sequences of the chloroplast RBCL gene. American Journal of Botany 86: 1182 - 1190.

De Vere, N. et al. (2017). Using DNA metabarcoding to investigate honey bee foraging reveals limited flower use despite high floral availability. Scientific Reports 7:42838.

Drescher, J. et al. (2016). Ecological and socio-economic functions across tropical land use systems after rainforest conversion. Philosophical Transactions of the Royal Society B: Biological Sciences 371: 20150275.

Drummond, A. J. & Bouckaert R. R. (2015) Bayesian Evolutionary Analysis with BEAST. Cambridge, UK. Cambridge University Press.

Environment Australia (1998). The Darwin Declaration. Australian Biological Resources Study

Faegri, K., & Iversen J. (1989). Textbook of Pollen Analysis 4th ed. Chichester, UK. John Wiley & Sons Ltd.

Fazekas, A. J. et al. (2008). Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. PLoS ONE 3(7): e2802.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. Journal of Molecular Evolution 17: 368 - 376.

Felsenstein, J. (2004). Inferring phylogenies. Sunderland, USA. Sinauer Associates, Inc.

Gitzendanner, M. A., Soltis, P. S., Wong, G. K.-S., Ruhfel, B. R. & Soltis, D. E. (2018). Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. American Journal of Botany 105: 291–301.

Guiry, M. D. (2012). How Many Species of Algae Are There? Journal of Phycology 48: 1057–1063.

Hajibabaei, M., Singer, G. A., Hebert, P. D. & Hickey, D. A. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends in Genetics 23: 167–172.

Hajibabaei, M., Singer, G. A., Hebert, P. D. & Hickey, D. A. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends in Genetics 23: 167 - 172.

Hall, B. G. (2018). Phylogenetic trees made easy: A how-to manual. New York, USA. Oxford University Press.

Hawkins, J. et al. (2015). Using DNA Metabarcoding to Identify the Floral Composition of Honey: A New Tool for Investigating Honey Bee Foraging Preferences. PLoS ONE 10(8): e0134735.

Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society B 270: 313 - 321.

Heckenhauer, J. 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, P. S., Salim, K. A. & Paun, O. (2018). Phylogenomics resolves evolutionary relationships and provides insights into floral evolution in the tribe Shoreeae (Dipterocarpaceae). Molecular Phylogenetics and Evolution 127: 1 - 13.

Hollingsworth, P. M. (2011b). Refining the DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America (PNAS) 108: 19451–19452.

Hollingsworth, P. M., Graham, S. W. & Little, D. P. (2011a). Choosing and Using a Plant DNA Barcode. PLoS ONE 6(5): e19254.

Indrioko, S., Gailing, O. & Finkeldey, R. (2006). Molecular phylogeny of Dipterocarpaceae in Indonesia based on chloroplast DNA. Plant Systematics and Evolution 261: 99–115.

Jantz, N., Homeier, J. & Behling, H. (2013). Representativeness of tree diversity in the modern pollen rain of Andean montane forests. Journal of Vegetation Science 25: 481 - 490.

Kajita, T. et al. (1998). Molecular Phylogeny of Dipetrocarpaceae in Southeast Asia Based on Nucleotide Sequences of *matK*, *trnL* Intron, and *trnL-trnF* Intergenic Spacer Region in Chloroplast DNA. Molecular Phylogenetics and Evolution 10: 202 - 209.

Kress, W. J. & Erickson, D. L. (2007). A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. PLoS ONE 2(6): e508.

Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A. & Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. Proceedings of the National Academy of Sciences of the United States of America (PNAS) 102: 8369 - 8374.

Kumar S., Stecher G., Li M., Knyaz C., & Tamura K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35: 1547 - 1549.

Li, X. et al. (2014). Plant DNA barcoding: from gene to genome. Biological Reviews 90: 157 - 166.

Londono, A. C., Alvarez, E., Forero, E. & Morton, C. M. (1995). A New Genus and Species of Dipterocarpaceae from the Neotropics. I. Introduction, Taxonomy, Ecology, and Distribution. Brittonia 47: 225 - 236.

Mora C., Tittensor D., Adl S., Simpson A., Worm B. (2011). How Many Species Are There on Earth and in the Ocean? PLoS Biology 9(8): e1001127.

Naciri, Y., Caetano, S. & Salamin, N. (2012). Plant DNA barcodes and the influence of gene flow. Molecular Ecology Resources 12: 575–580.

Rambaut, A. & Grass, N. C. (1997). Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. CABIOS 13: 235 - 238.

Rembold, K., Mangopo, H., Tjitrosoedirdjo, S. S. & Kreft, H. (2017). Plant diversity, forest dependency, and alien plant invasions in tropical agricultural landscapes. Biological Conservation 213: 234 - 242.

Rembold, K., Sri Tjitrosoedirdjo, S. S., Kreft, H. (2017). Common wayside plants of Jambi Province (Sumatra, Indonesia). Version 2. Biodiversity, Macroecology & Biogeography, Faculty of Forest Sciences and Forest Ecology of the University of Goettingen, Germany.

Saitou N. & Imanishi T. (1989). Relative efficiencies of the Fitch-Margoliash, Maximum Parsimony, Maximum-Likelihood, Minimum-Evolution, and Neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. Molecular Biology and Evolution 6(5): 514 - 525.

Saitou N. & Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406 - 425.

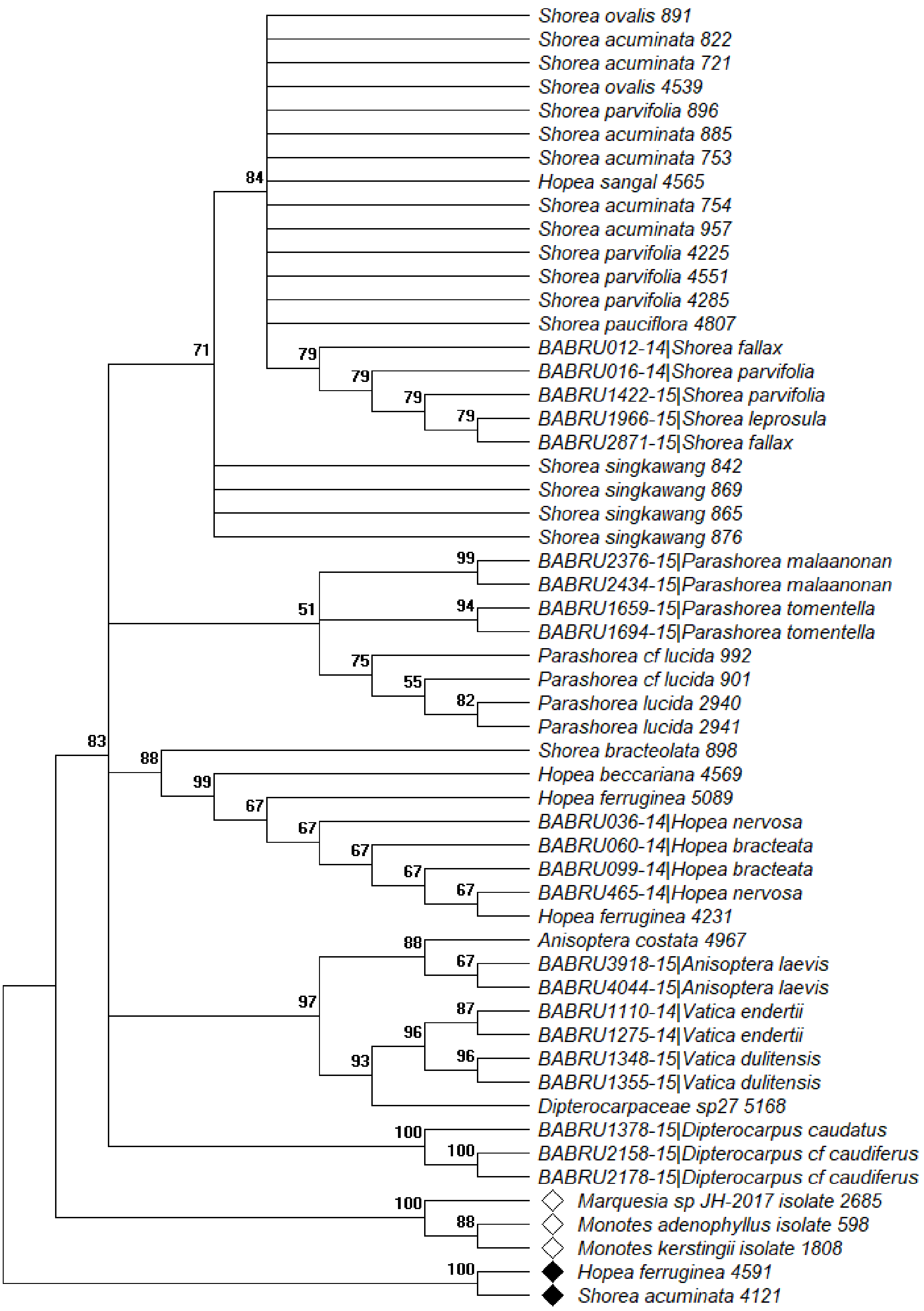
Secretariat of Convention on Biological Diversity (2007). Guide to The Global Taxonomy Initiative. CBD Technical Series No. 30.

Techen, N., Parveen, I., Pan, Z. & Khan, I. A. (2014). DNA barcoding of medicinal plant material for identification. Current Opinion in Biotechnology 25: 103 - 110.

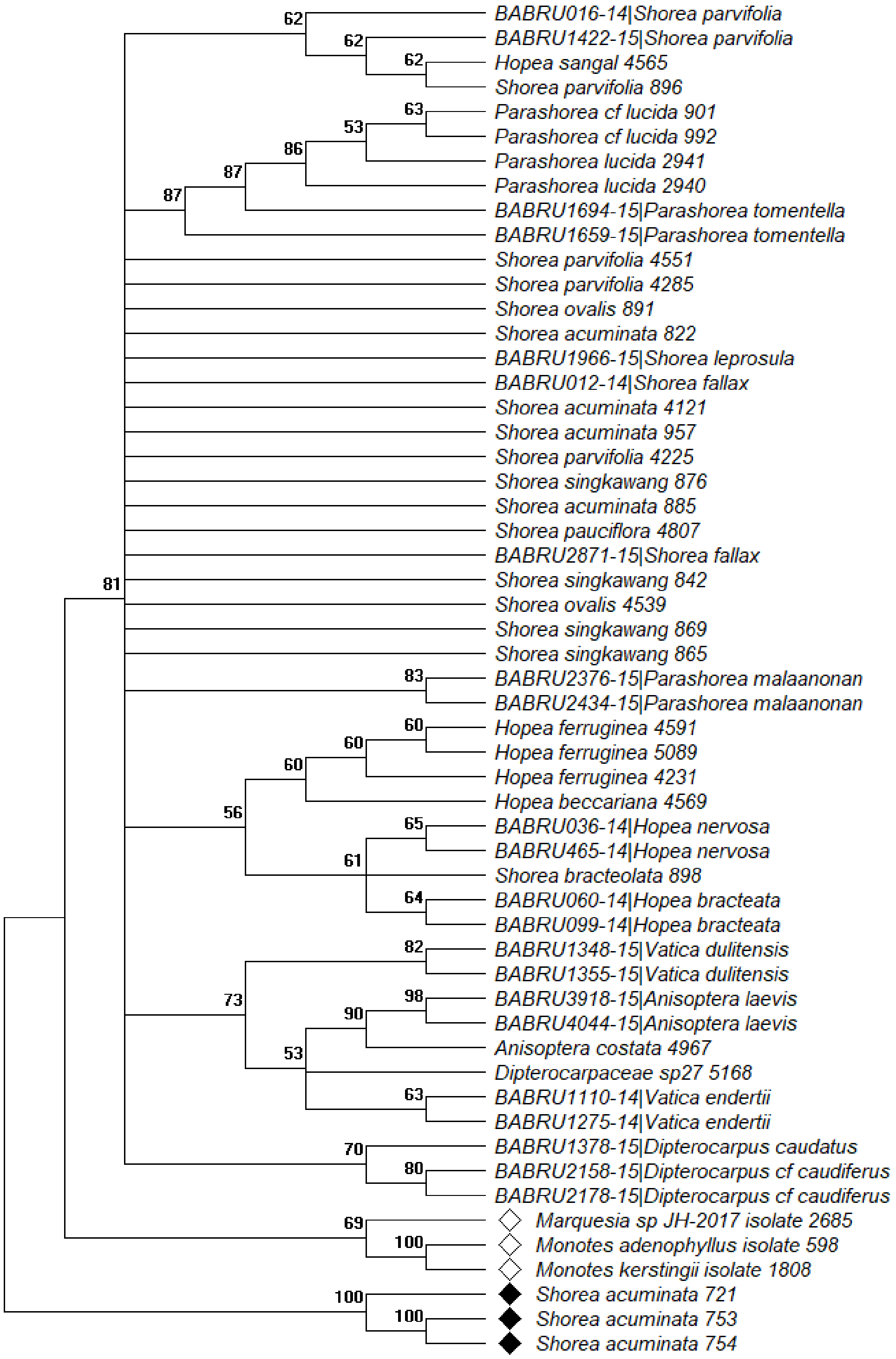
The Angiosperm Phylogeny Group (APG) IV. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Botanical Journal of the Linnean Society 181: 1 - 20.

Vaidya, G., Lohman, D. J. & Meier, R. (2011). SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27: 171 - 180.

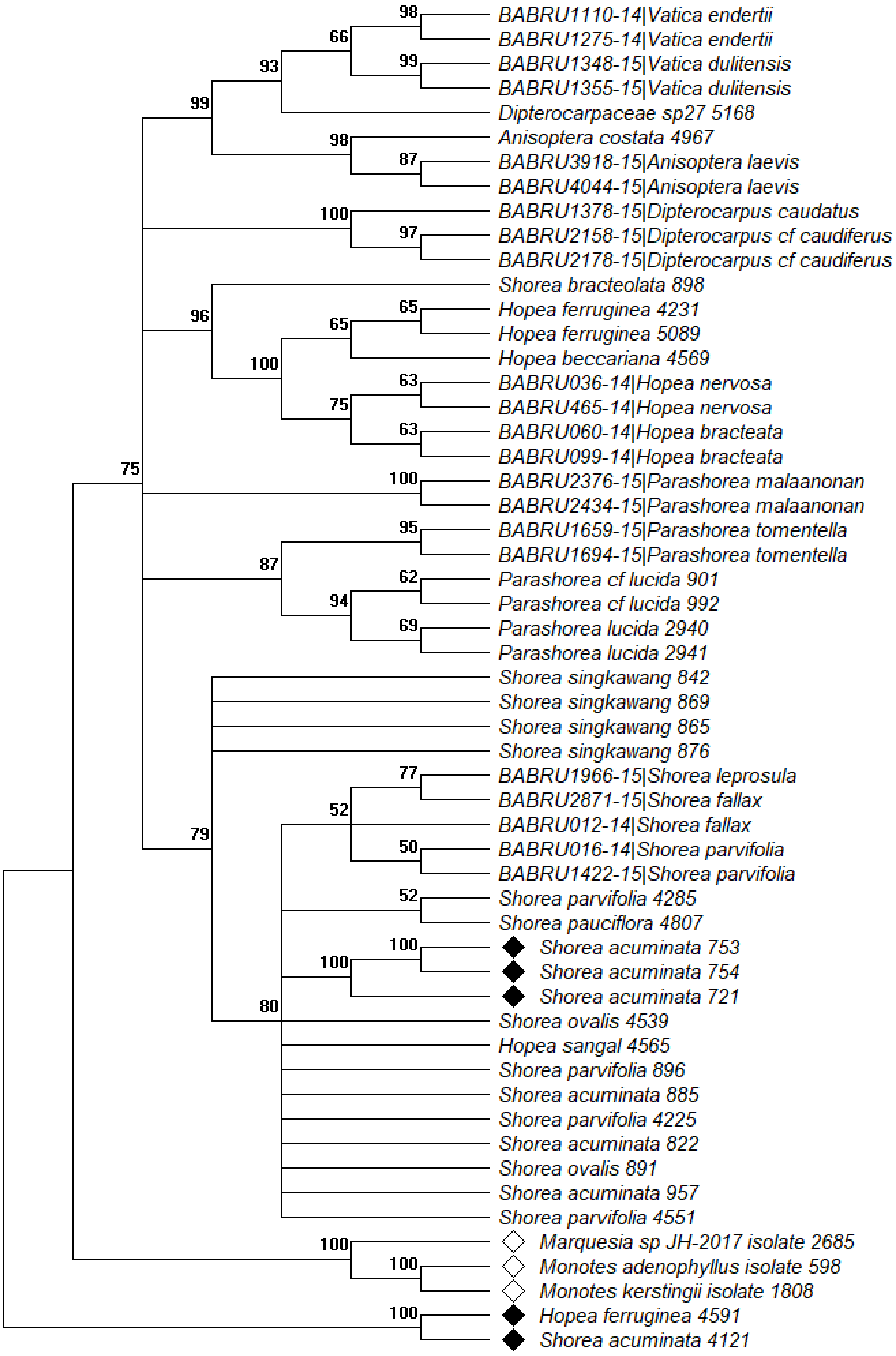
7. APPENDIXES


Shorea ovalis 891
Shorea acuminata 822
Shorea acuminata 721
Shorea ovalis 4539
Shorea parvifolia 896
Shorea acuminata 885
Shorea acuminata 753
84
Hopea sangal 4565
Shorea acuminata 754
Shorea acuminata 957
Shorea parvifolia 4225
Shorea parvifolia 4551
Shorea parvifolia 4285
Shorea pauciflora 4807
71
79
BABRU012-14|Shorea fallax
79
BABRU016-14|Shorea parvifolia
79
BABRU1422-15|Shorea parvifolia
79
BABRU1966-15|Shorea leprosula
BABRU2871-15|Shorea fallax
Shorea singkawang 842
Shorea singkawang 869
Shorea singkawang 865
Shorea singkawang 876
99
BABRU2376-15|Parashorea malaanonan
BABRU2434-15|Parashorea malaanonan
51
94
BABRU1659-15|Parashorea tomentella
BABRU1694-15|Parashorea tomentella
75
Parashorea cf lucida 992
55
Parashorea cf lucida 901
82
Parashorea lucida 2940
Parashorea lucida 2941
83
Shorea bracteolata 898
88
99
Hopea beccariana 4569
67
Hopea ferruginea 5089
67
BABRU036-14|Hopea nervosa
67
BABRU060-14|Hopea bracteata
67
BABRU099-14|Hopea bracteata
67
BABRU465-14|Hopea nervosa
Hopea ferruginea 4231
88
Anisoptera costata 4967
67
BABRU3918-15|Anisoptera laevis
BABRU4044-15|Anisoptera laevis
97
87
BABRU1110-14|Vatica endertii
BABRU1275-14|Vatica endertii
96
96
BABRU1348-15|Vatica dulitensis
93
BABRU1355-15|Vatica dulitensis
Dipterocarpaceae sp27 5168
100
BABRU1378-15|Dipterocarpus caudatus
100
BABRU2158-15|Dipterocarpus cf caudiferus
BABRU2178-15|Dipterocarpus cf caudiferus
100
◇ Marquesia sp JH-2017 isolate 2685
88
◇ Monotes adenophyllus isolate 598
◇ Monotes kerstingii isolate 1808
100
◆ Hopea ferruginea 4591
◆ Shorea acuminata 4121

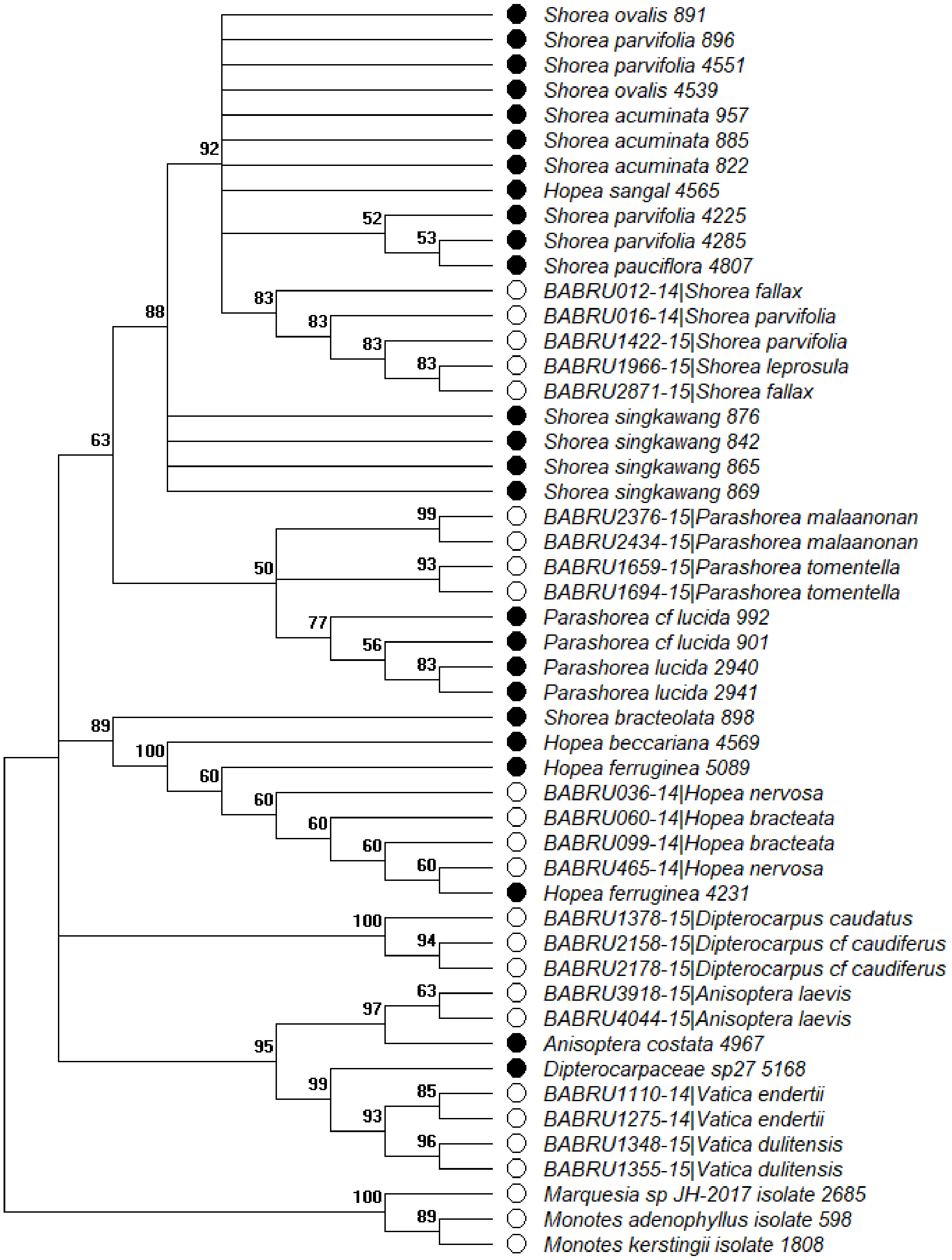
Appendix 1 The neighbor joining phylogenetic tree of trees' samples collected in the EFForTS project based on the *matK* marker and genetic distances computed using the Maximum Composite Likelihood method. The numbers at the tree nodes represent bootstrap values based on 1000 replicates. Condensed tree, branches with support values lower than 50% have been collapsed. The number after the sample name refers the sample ID. ♦ - samples behaving as a total outgroup from the *Dipterocarpaceae* family; ◊ - the database samples chosen as the outgroup for the analysis.


62
62
62
BABRU016-14|Shorea parvifolia
BABRU1422-15|Shorea parvifolia
Hopea sangal 4565
Shorea parvifolia 896
63
53
86
87
87
Parashorea cf lucida 901
Parashorea cf lucida 992
Parashorea lucida 2941
Parashorea lucida 2940
BABRU1694-15|Parashorea tomentella
BABRU1659-15|Parashorea tomentella
Shorea parvifolia 4551
Shorea parvifolia 4285
Shorea ovalis 891
Shorea acuminata 822
BABRU1966-15|Shorea leprosula
BABRU012-14|Shorea fallax
Shorea acuminata 4121
Shorea acuminata 957
Shorea parvifolia 4225
Shorea singkawang 876
Shorea acuminata 885
Shorea pauciflora 4807
BABRU2871-15|Shorea fallax
81
Shorea singkawang 842
Shorea ovalis 4539
Shorea singkawang 869
Shorea singkawang 865
83
BABRU2376-15|Parashorea malaanonan
BABRU2434-15|Parashorea malaanonan
60
60
60
56
Hopea ferruginea 4591
Hopea ferruginea 5089
Hopea ferruginea 4231
Hopea beccariana 4569
65
61
BABRU036-14|Hopea nervosa
BABRU465-14|Hopea nervosa
Shorea bracteolata 898
64
BABRU060-14|Hopea bracteata
BABRU099-14|Hopea bracteata
82
BABRU1348-15|Vatica dulitensis
BABRU1355-15|Vatica dulitensis
73
98
90
BABRU3918-15|Anisoptera laevis
BABRU4044-15|Anisoptera laevis
53
Anisoptera costata 4967
Dipterocarpaceae sp27 5168
63
BABRU1110-14|Vatica endertii
BABRU1275-14|Vatica endertii
70
BABRU1378-15|Dipterocarpus caudatus
80
BABRU2158-15|Dipterocarpus cf caudiferus
BABRU2178-15|Dipterocarpus cf caudiferus
69
◇ Marquesia sp JH-2017 isolate 2685
100
◇ Monotes adenophyllus isolate 598
◇ Monotes kerstingii isolate 1808
100
◆ Shorea acuminata 721
100
◆ Shorea acuminata 753
◆ Shorea acuminata 754

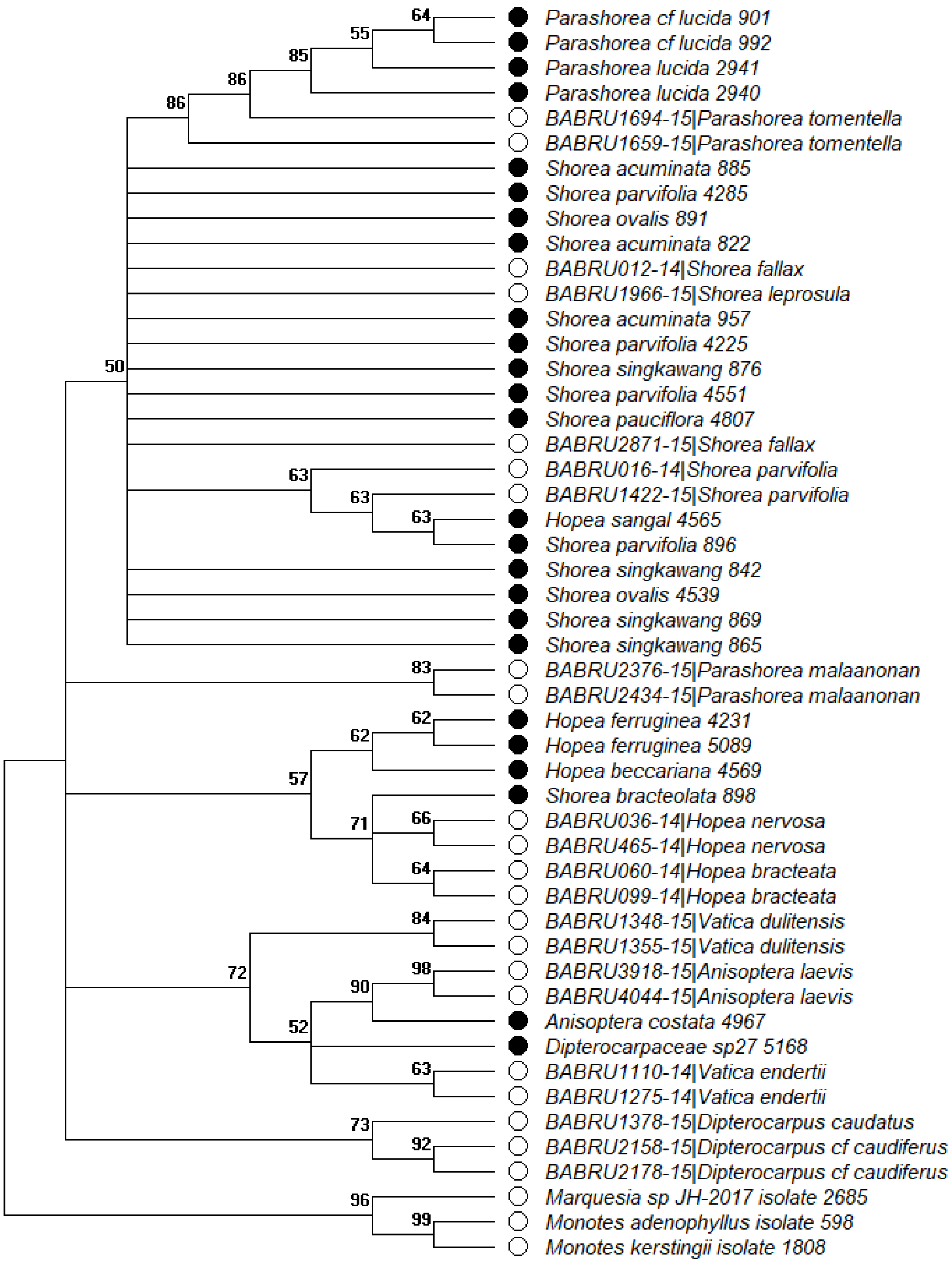
Appendix 2 The neighbor joining phylogenetic tree of trees' samples collected in the EFForTS project based on the *rbcL* marker and genetic distances computed using the Maximum Composite Likelihood method. The numbers at the tree nodes represent bootstrap values based on 1000 replicates. Condensed tree, branches with support values lower than 50% have been collapsed. The number after the sample name refers the sample ID. ♦ - samples behaving as a total outgroup from the *Dipterocarpaceae* family; ◊ - the database samples chosen as the outgroup for the analysis.


98
BABRU1110-14|Vatica endertii
66
BABRU1275-14|Vatica endertii
93
99
BABRU1348-15|Vatica dulitensis
BABRU1355-15|Vatica dulitensis
99
Dipterocarpaceae sp27 5168
98
Anisoptera costata 4967
87
BABRU3918-15|Anisoptera laevis
BABRU4044-15|Anisoptera laevis
100
BABRU1378-15|Dipterocarpus caudatus
97
BABRU2158-15|Dipterocarpus cf caudiferus
BABRU2178-15|Dipterocarpus cf caudiferus
Shorea bracteolata 898
96
65
Hopea ferruginea 4231
65
Hopea ferruginea 5089
100
Hopea beccariana 4569
63
BABRU036-14|Hopea nervosa
75
BABRU465-14|Hopea nervosa
63
BABRU060-14|Hopea bracteata
BABRU099-14|Hopea bracteata
75
100
BABRU2376-15|Parashorea malaanonan
BABRU2434-15|Parashorea malaanonan
95
BABRU1659-15|Parashorea tomentella
87
BABRU1694-15|Parashorea tomentella
62
Parashorea cf lucida 901
94
Parashorea cf lucida 992
69
Parashorea lucida 2940
Parashorea lucida 2941
Shorea singkawang 842
Shorea singkawang 869
Shorea singkawang 865
Shorea singkawang 876
77
BABRU1966-15|Shorea leprosula
52
BABRU2871-15|Shorea fallax
79
BABRU012-14|Shorea fallax
50
BABRU016-14|Shorea parvifolia
BABRU1422-15|Shorea parvifolia
52
Shorea parvifolia 4285
Shorea pauciflora 4807
100
♦ Shorea acuminata 753
100
♦ Shorea acuminata 754
♦ Shorea acuminata 721
80
Shorea ovalis 4539
Hopea sangal 4565
Shorea parvifolia 896
Shorea acuminata 885
Shorea parvifolia 4225
Shorea acuminata 822
Shorea ovalis 891
Shorea acuminata 957
Shorea parvifolia 4551
100
◊ Marquesia sp JH-2017 isolate 2685
100
◊ Monotes adenophyllus isolate 598
◊ Monotes kerstingii isolate 1808
100
♦ Hopea ferruginea 4591
♦ Shorea acuminata 4121

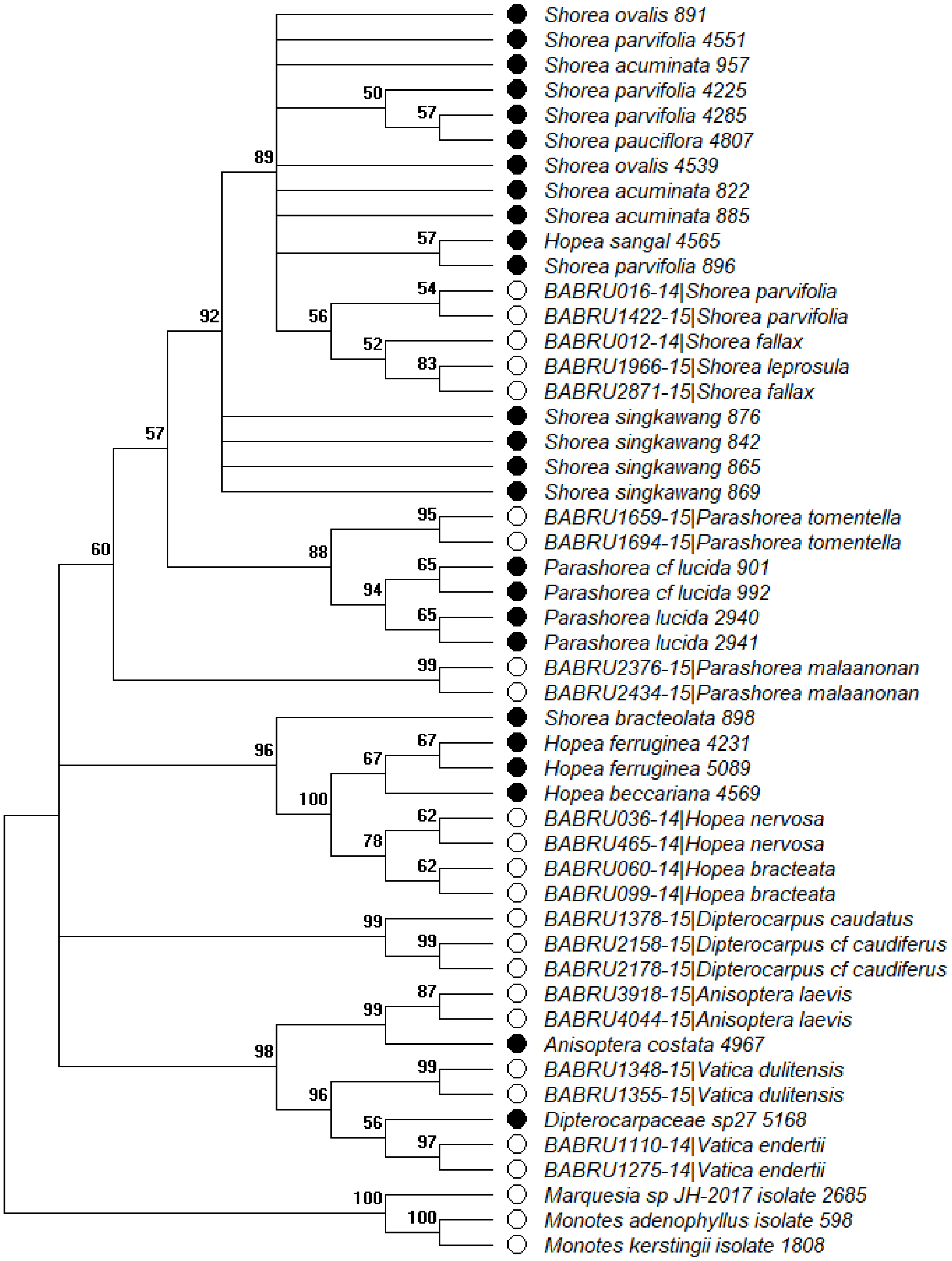
Appendix 3 The neighbor joining phylogenetic tree of trees' samples collected in the EFForTS project based on the concatenated *rbcL* and *matK* markers, with genetic distances computed using the Maximum Composite Likelihood method. The numbers at the tree nodes represent bootstrap values based on 1000 replicates. Condensed tree, branches with support values lower than 50% have been collapsed. The number after the sample name refers the sample ID. ♦ - samples clustered with problematic behavior; ◊ - the database samples chosen as the outgroup for the analysis.


● Shorea ovalis 891
● Shorea parvifolia 896
● Shorea parvifolia 4551
● Shorea ovalis 4539
● Shorea acuminata 957
● Shorea acuminata 885
● Shorea acuminata 822
● Hopea sangal 4565
● Shorea parvifolia 4225
● Shorea parvifolia 4285
● Shorea pauciflora 4807
○ BABRU012-14|Shorea fallax
○ BABRU016-14|Shorea parvifolia
○ BABRU1422-15|Shorea parvifolia
○ BABRU1966-15|Shorea leprosula
○ BABRU2871-15|Shorea fallax
● Shorea singkawang 876
● Shorea singkawang 842
● Shorea singkawang 865
● Shorea singkawang 869
○ BABRU2376-15|Parashorea malaanonan
○ BABRU2434-15|Parashorea malaanonan
○ BABRU1659-15|Parashorea tomentella
○ BABRU1694-15|Parashorea tomentella
● Parashorea cf lucida 992
● Parashorea cf lucida 901
● Parashorea lucida 2940
● Parashorea lucida 2941
● Shorea bracteolata 898
● Hopea beccariana 4569
● Hopea ferruginea 5089
○ BABRU036-14|Hopea nervosa
○ BABRU060-14|Hopea bracteata
○ BABRU099-14|Hopea bracteata
○ BABRU465-14|Hopea nervosa
● Hopea ferruginea 4231
○ BABRU1378-15|Dipterocarpus caudatus
○ BABRU2158-15|Dipterocarpus cf caudiferus
○ BABRU2178-15|Dipterocarpus cf caudiferus
○ BABRU3918-15|Anisoptera laevis
○ BABRU4044-15|Anisoptera laevis
● Anisoptera costata 4967
● Dipterocarpaceae sp27 5168
○ BABRU1110-14|Vatica endertii
○ BABRU1275-14|Vatica endertii
○ BABRU1348-15|Vatica dulitensis
○ BABRU1355-15|Vatica dulitensis
○ Marquesia sp JH-2017 isolate 2685
○ Monotes adenophyllus isolate 598
○ Monotes kerstingii isolate 1808
92
52
53
83
88
83
83
83
63
99
50
93
77
56
83
89
100
60
60
60
60
60
100
94
63
97
95
99
85
93
96
100
89

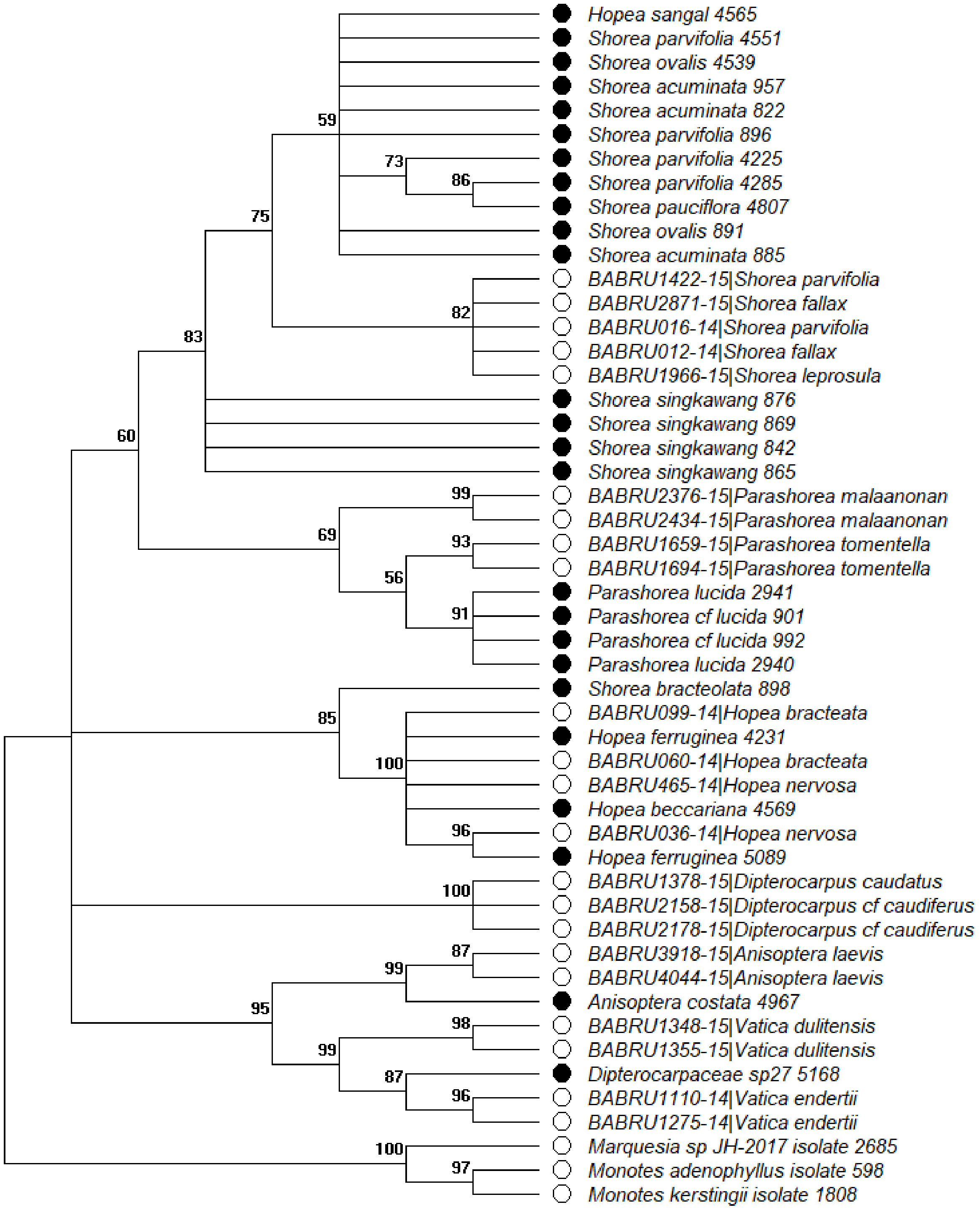
Appendix 4 The neighbor joining phylogenetic tree of trees' samples collected in the EFForTS project based on the *matK* marker and genetic distances computed using the Maximum Composite Likelihood method. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. Condensed tree, branches with support values lower than 50% have been collapsed. The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.


64
55
85
86
86
50
● Parashorea cf lucida 901
● Parashorea cf lucida 992
● Parashorea lucida 2941
● Parashorea lucida 2940
○ BABRU1694-15|Parashorea tomentella
○ BABRU1659-15|Parashorea tomentella
● Shorea acuminata 885
● Shorea parvifolia 4285
● Shorea ovalis 891
● Shorea acuminata 822
○ BABRU012-14|Shorea fallax
○ BABRU1966-15|Shorea leprosula
● Shorea acuminata 957
● Shorea parvifolia 4225
● Shorea singkawang 876
● Shorea parvifolia 4551
● Shorea pauciflora 4807
○ BABRU2871-15|Shorea fallax
63
63
63
○ BABRU016-14|Shorea parvifolia
○ BABRU1422-15|Shorea parvifolia
● Hopea sangal 4565
● Shorea parvifolia 896
● Shorea singkawang 842
● Shorea ovalis 4539
● Shorea singkawang 869
● Shorea singkawang 865
83
○ BABRU2376-15|Parashorea malaanonan
○ BABRU2434-15|Parashorea malaanonan
62
62
57
● Hopea ferruginea 4231
● Hopea ferruginea 5089
● Hopea beccariana 4569
● Shorea bracteolata 898
71
66
○ BABRU036-14|Hopea nervosa
○ BABRU465-14|Hopea nervosa
64
○ BABRU060-14|Hopea bracteata
○ BABRU099-14|Hopea bracteata
84
○ BABRU1348-15|Vatica dulitensis
○ BABRU1355-15|Vatica dulitensis
72
98
90
○ BABRU3918-15|Anisoptera laevis
○ BABRU4044-15|Anisoptera laevis
52
● Anisoptera costata 4967
● Dipterocarpaceae sp27 5168
63
○ BABRU1110-14|Vatica endertii
○ BABRU1275-14|Vatica endertii
73
○ BABRU1378-15|Dipterocarpus caudatus
92
○ BABRU2158-15|Dipterocarpus cf caudiferus
○ BABRU2178-15|Dipterocarpus cf caudiferus
96
○ Marquesia sp JH-2017 isolate 2685
99
○ Monotes adenophyllus isolate 598
○ Monotes kerstingii isolate 1808

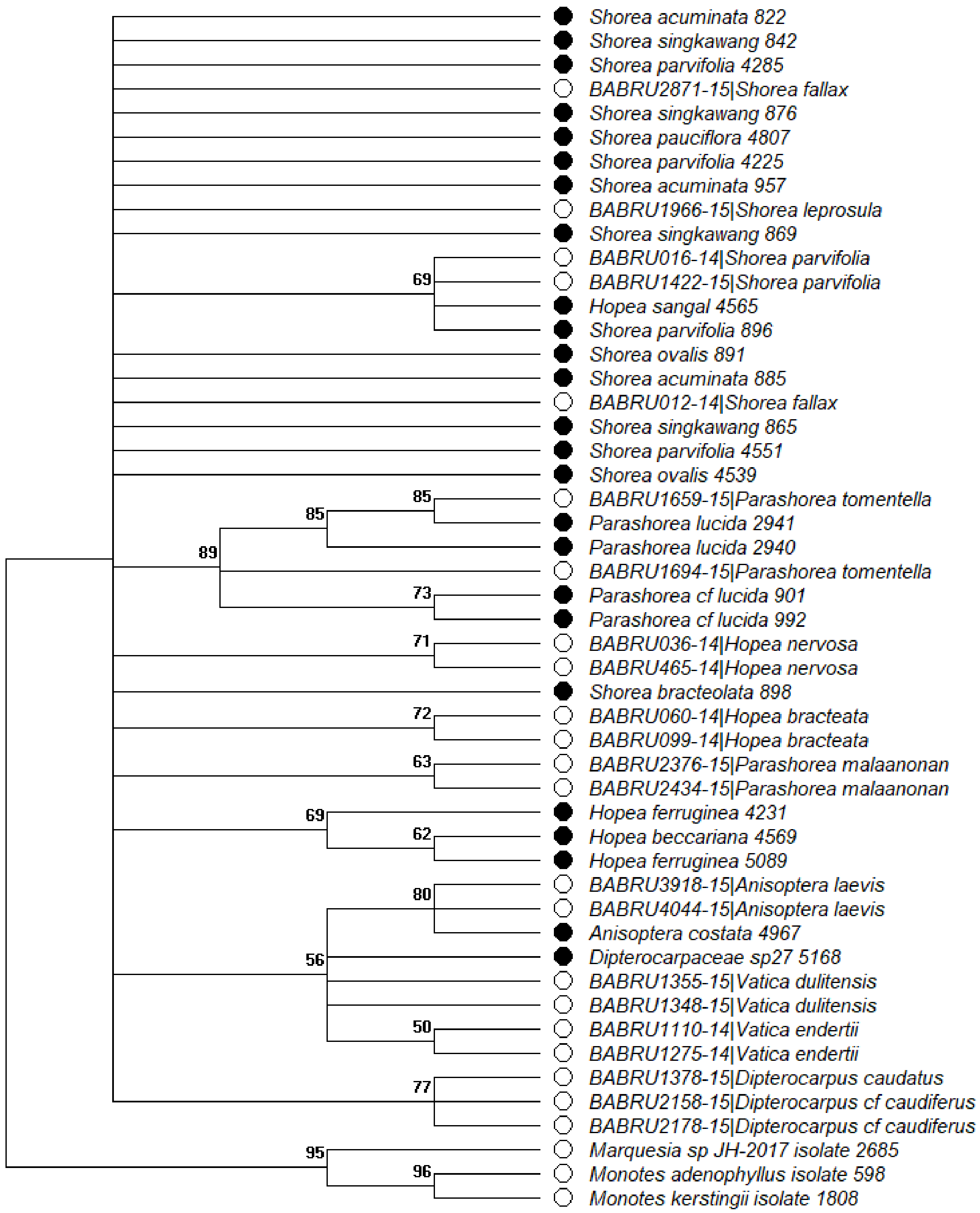
Appendix 5 The neighbor joining phylogenetic tree of trees' samples collected in the EFForTS project based on the *rbcL* marker and genetic distances computed using the Maximum Composite Likelihood method. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. Condensed tree, branches with support values lower than 50% have been collapsed. The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.


● Shorea ovalis 891
● Shorea parvifolia 4551
● Shorea acuminata 957
● Shorea parvifolia 4225
● Shorea parvifolia 4285
● Shorea pauciflora 4807
● Shorea ovalis 4539
● Shorea acuminata 822
● Shorea acuminata 885
● Hopea sangal 4565
● Shorea parvifolia 896
○ BABRU016-14|Shorea parvifolia
○ BABRU1422-15|Shorea parvifolia
○ BABRU012-14|Shorea fallax
○ BABRU1966-15|Shorea leprosula
○ BABRU2871-15|Shorea fallax
● Shorea singkawang 876
● Shorea singkawang 842
● Shorea singkawang 865
● Shorea singkawang 869
○ BABRU1659-15|Parashorea tomentella
○ BABRU1694-15|Parashorea tomentella
● Parashorea cf lucida 901
● Parashorea cf lucida 992
● Parashorea lucida 2940
● Parashorea lucida 2941
○ BABRU2376-15|Parashorea malaanonan
○ BABRU2434-15|Parashorea malaanonan
● Shorea bracteolata 898
● Hopea ferruginea 4231
● Hopea ferruginea 5089
● Hopea beccariana 4569
○ BABRU036-14|Hopea nervosa
○ BABRU465-14|Hopea nervosa
○ BABRU060-14|Hopea bracteata
○ BABRU099-14|Hopea bracteata
○ BABRU1378-15|Dipterocarpus caudatus
○ BABRU2158-15|Dipterocarpus cf caudiferus
○ BABRU2178-15|Dipterocarpus cf caudiferus
○ BABRU3918-15|Anisoptera laevis
○ BABRU4044-15|Anisoptera laevis
● Anisoptera costata 4967
○ BABRU1348-15|Vatica dulitensis
○ BABRU1355-15|Vatica dulitensis
● Dipterocarpaceae sp27 5168
○ BABRU1110-14|Vatica endertii
○ BABRU1275-14|Vatica endertii
○ Marquesia sp JH-2017 isolate 2685
○ Monotes adenophyllus isolate 598
○ Monotes kerstingii isolate 1808
50
57
89
57
54
92
56
52
83
57
95
60
88
65
94
65
99
67
96
67
100
62
78
62
99
99
87
99
98
99
96
56
97
100
100

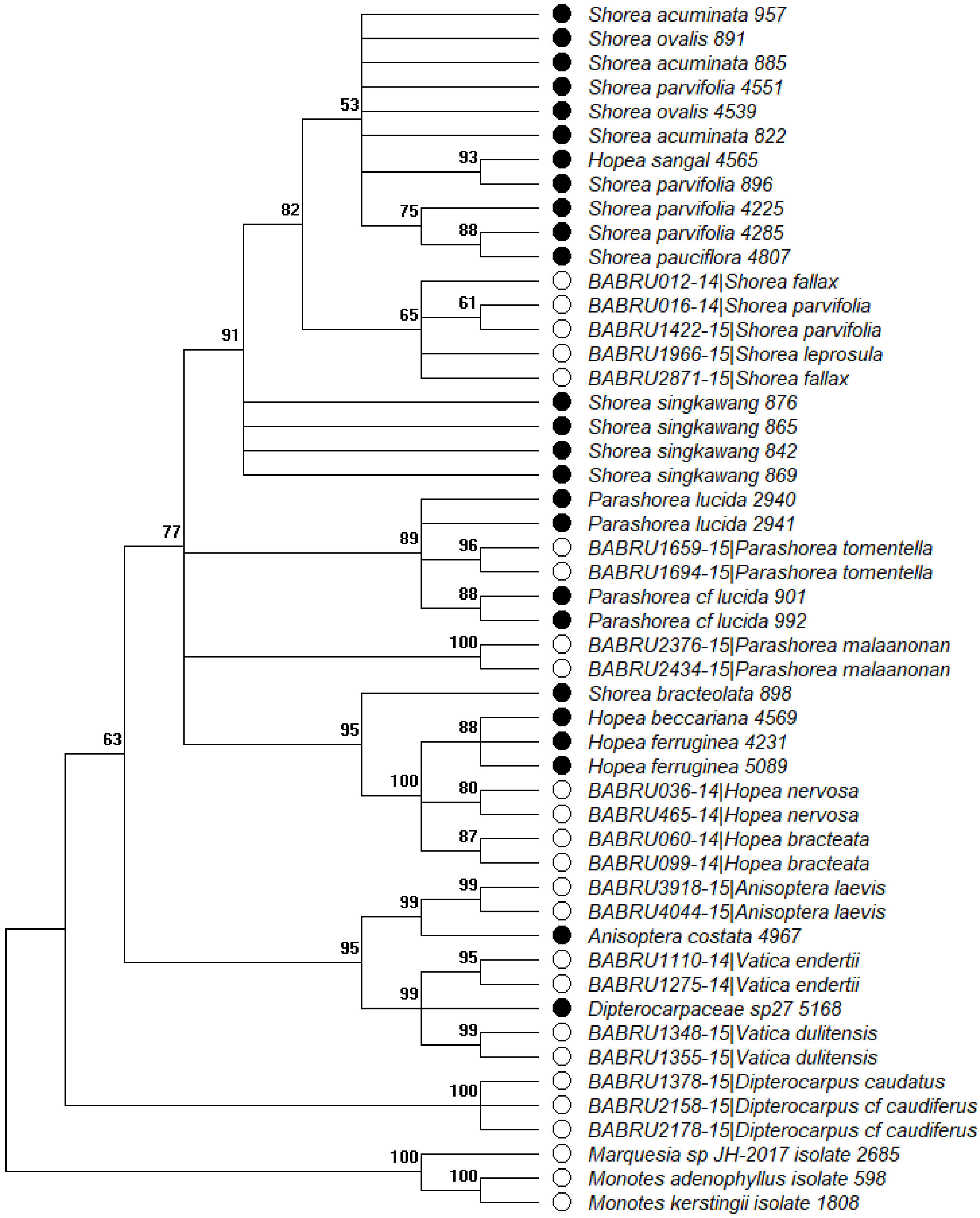
Appendix 6 The neighbor joining phylogenetic tree of trees' samples collected in the EFForTS project based on the concatenated *rbcL* and *matK* markers, with genetic distances computed using the Maximum Composite Likelihood method. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. Condensed tree, branches with support values lower than 50% have been collapsed. The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.


Hopea sangal 4565
Shorea parvifolia 4551
Shorea ovalis 4539
Shorea acuminata 957
Shorea acuminata 822
Shorea parvifolia 896
Shorea parvifolia 4225
Shorea parvifolia 4285
Shorea pauciflora 4807
Shorea ovalis 891
Shorea acuminata 885
BABRU1422-15|Shorea parvifolia
BABRU2871-15|Shorea fallax
BABRU016-14|Shorea parvifolia
BABRU012-14|Shorea fallax
BABRU1966-15|Shorea leprosula
Shorea singkawang 876
Shorea singkawang 869
Shorea singkawang 842
Shorea singkawang 865
BABRU2376-15|Parashorea malaanonan
BABRU2434-15|Parashorea malaanonan
BABRU1659-15|Parashorea tomentella
BABRU1694-15|Parashorea tomentella
Parashorea lucida 2941
Parashorea cf lucida 901
Parashorea cf lucida 992
Parashorea lucida 2940
Shorea bracteolata 898
BABRU099-14|Hopea bracteata
Hopea ferruginea 4231
BABRU060-14|Hopea bracteata
BABRU465-14|Hopea nervosa
Hopea beccariana 4569
BABRU036-14|Hopea nervosa
Hopea ferruginea 5089
BABRU1378-15|Dipterocarpus caudatus
BABRU2158-15|Dipterocarpus cf caudiferus
BABRU2178-15|Dipterocarpus cf caudiferus
BABRU3918-15|Anisoptera laevis
BABRU4044-15|Anisoptera laevis
Anisoptera costata 4967
BABRU1348-15|Vatica dulitensis
BABRU1355-15|Vatica dulitensis
Dipterocarpaceae sp27 5168
BABRU1110-14|Vatica endertii
BABRU1275-14|Vatica endertii
Marquesia sp JH-2017 isolate 2685
Monotes adenophyllus isolate 598
Monotes kerstingii isolate 1808
59
73
86
75
82
83
60
99
69
93
56
91
85
100
96
100
87
99
95
98
99
87
96
100
97

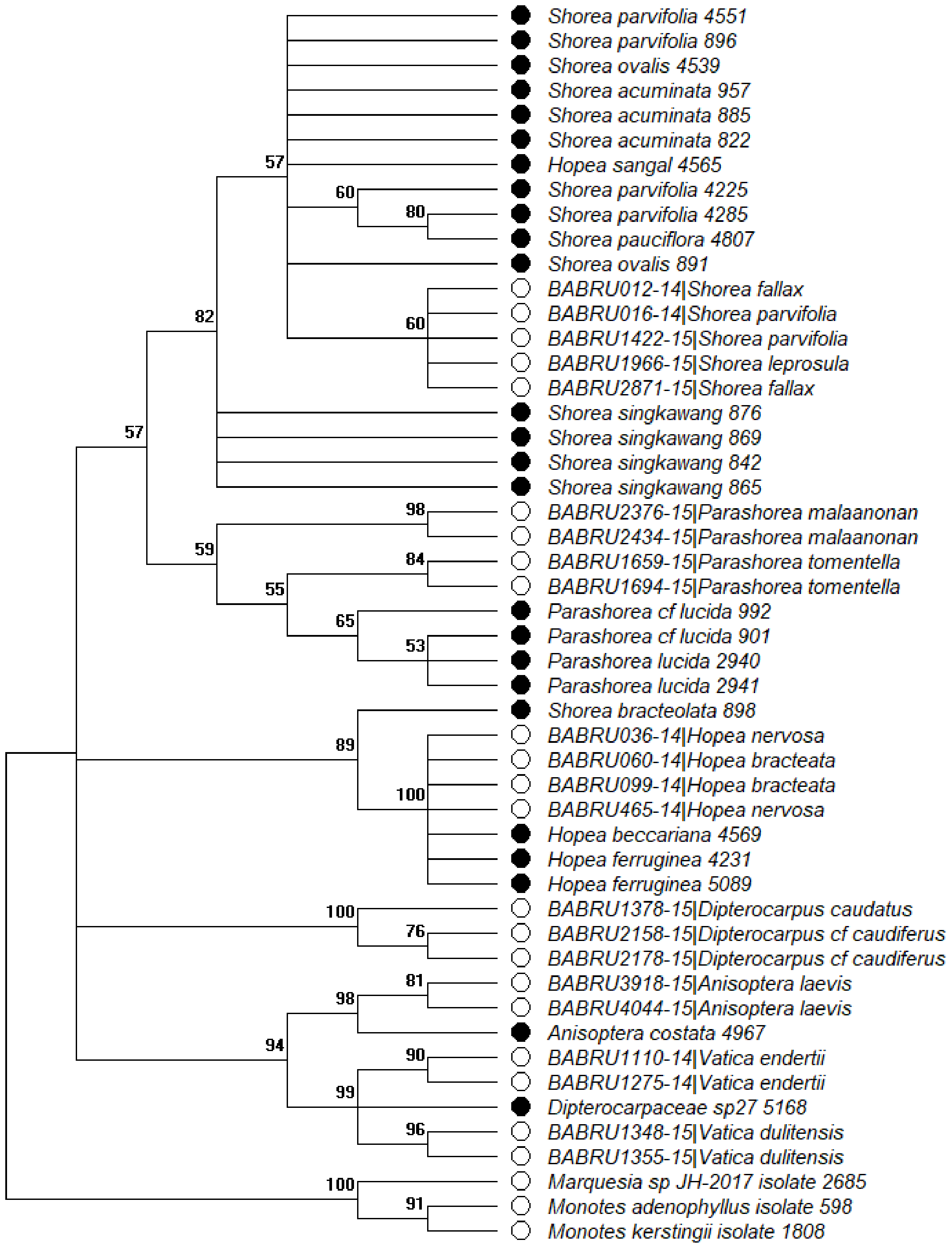
Appendix 7 The maximum parsimony phylogenetic tree of trees' samples collected in the EFForTS project based on the *matK* marker and obtained by using the Subtree-Pruning-Regrafting (SPR) algorithm. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. Condensed tree, branches with support values lower than 50% have been collapsed. The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.


● Shorea acuminata 822
● Shorea singkawang 842
● Shorea parvifolia 4285
○ BABRU2871-15|Shorea fallax
● Shorea singkawang 876
● Shorea pauciflora 4807
● Shorea parvifolia 4225
● Shorea acuminata 957
○ BABRU1966-15|Shorea leprosula
● Shorea singkawang 869
69
○ BABRU016-14|Shorea parvifolia
○ BABRU1422-15|Shorea parvifolia
● Hopea sangal 4565
● Shorea parvifolia 896
● Shorea ovalis 891
● Shorea acuminata 885
○ BABRU012-14|Shorea fallax
● Shorea singkawang 865
● Shorea parvifolia 4551
● Shorea ovalis 4539
85
85
89
○ BABRU1659-15|Parashorea tomentella
● Parashorea lucida 2941
● Parashorea lucida 2940
○ BABRU1694-15|Parashorea tomentella
73
● Parashorea cf lucida 901
● Parashorea cf lucida 992
71
○ BABRU036-14|Hopea nervosa
○ BABRU465-14|Hopea nervosa
● Shorea bracteolata 898
72
○ BABRU060-14|Hopea bracteata
○ BABRU099-14|Hopea bracteata
63
○ BABRU2376-15|Parashorea malaanonan
○ BABRU2434-15|Parashorea malaanonan
69
● Hopea ferruginea 4231
62
● Hopea beccariana 4569
● Hopea ferruginea 5089
80
○ BABRU3918-15|Anisoptera laevis
○ BABRU4044-15|Anisoptera laevis
● Anisoptera costata 4967
56
● Dipterocarpaceae sp27 5168
○ BABRU1355-15|Vatica dulitensis
○ BABRU1348-15|Vatica dulitensis
50
○ BABRU1110-14|Vatica endertii
○ BABRU1275-14|Vatica endertii
77
○ BABRU1378-15|Dipterocarpus caudatus
○ BABRU2158-15|Dipterocarpus cf caudiferus
○ BABRU2178-15|Dipterocarpus cf caudiferus
95
○ Marquesia sp JH-2017 isolate 2685
96
○ Monotes adenophyllus isolate 598
○ Monotes kerstingii isolate 1808

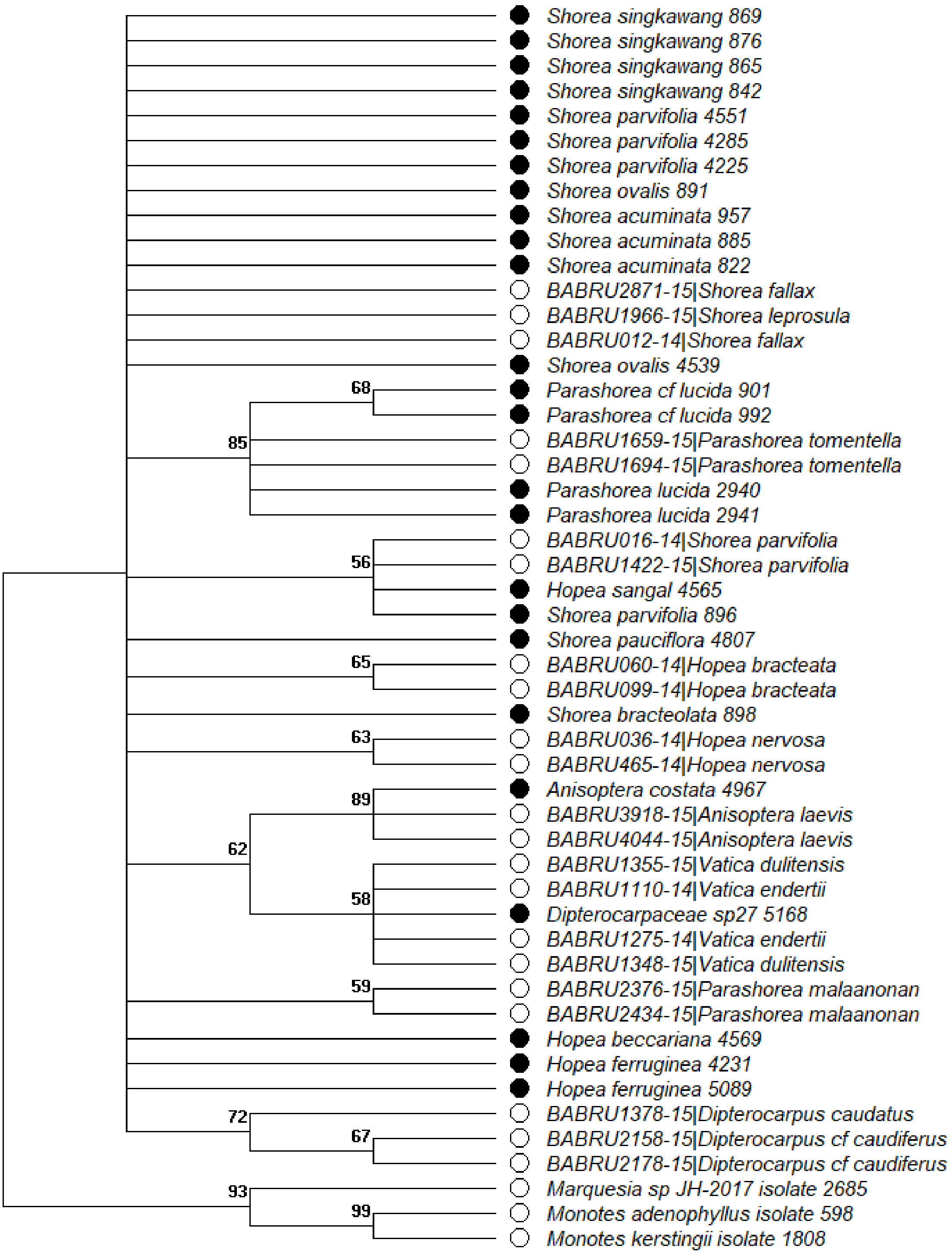
Appendix 8 The maximum parsimony phylogenetic tree of trees' samples collected in the EFForTS project based on the *rbcL* marker and obtained by using the Subtree-Pruning-Regrafting (SPR) algorithm. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. Condensed tree, branches with support values lower than 50% have been collapsed. The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.


● Shorea acuminata 957
● Shorea ovalis 891
● Shorea acuminata 885
● Shorea parvifolia 4551
● Shorea ovalis 4539
● Shorea acuminata 822
● Hopea sangal 4565
● Shorea parvifolia 896
● Shorea parvifolia 4225
● Shorea parvifolia 4285
● Shorea pauciflora 4807
○ BABRU012-14|Shorea fallax
○ BABRU016-14|Shorea parvifolia
○ BABRU1422-15|Shorea parvifolia
○ BABRU1966-15|Shorea leprosula
○ BABRU2871-15|Shorea fallax
● Shorea singkawang 876
● Shorea singkawang 865
● Shorea singkawang 842
● Shorea singkawang 869
● Parashorea lucida 2940
● Parashorea lucida 2941
○ BABRU1659-15|Parashorea tomentella
○ BABRU1694-15|Parashorea tomentella
● Parashorea cf lucida 901
● Parashorea cf lucida 992
○ BABRU2376-15|Parashorea malaanonan
○ BABRU2434-15|Parashorea malaanonan
● Shorea bracteolata 898
● Hopea beccariana 4569
● Hopea ferruginea 4231
● Hopea ferruginea 5089
○ BABRU036-14|Hopea nervosa
○ BABRU465-14|Hopea nervosa
○ BABRU060-14|Hopea bracteata
○ BABRU099-14|Hopea bracteata
○ BABRU3918-15|Anisoptera laevis
○ BABRU4044-15|Anisoptera laevis
● Anisoptera costata 4967
○ BABRU1110-14|Vatica endertii
○ BABRU1275-14|Vatica endertii
● Dipterocarpaceae sp27 5168
○ BABRU1348-15|Vatica dulitensis
○ BABRU1355-15|Vatica dulitensis
○ BABRU1378-15|Dipterocarpus caudatus
○ BABRU2158-15|Dipterocarpus cf caudiferus
○ BABRU2178-15|Dipterocarpus cf caudiferus
○ Marquesia sp JH-2017 isolate 2685
○ Monotes adenophyllus isolate 598
○ Monotes kerstingii isolate 1808
53
93
82
75
88
61
65
91
77
89
96
88
100
95
88
100
80
87
63
99
99
95
95
99
99
100
100
100

Appendix 9 The maximum parsimony phylogenetic tree of trees' samples collected in the EFForTS project based on the concatenated *rbcL* and *matK* markers, and obtained by using the Subtree-Pruning-Regrafting (SPR) algorithm. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. Condensed tree, branches with support values lower than 50% have been collapsed. The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.

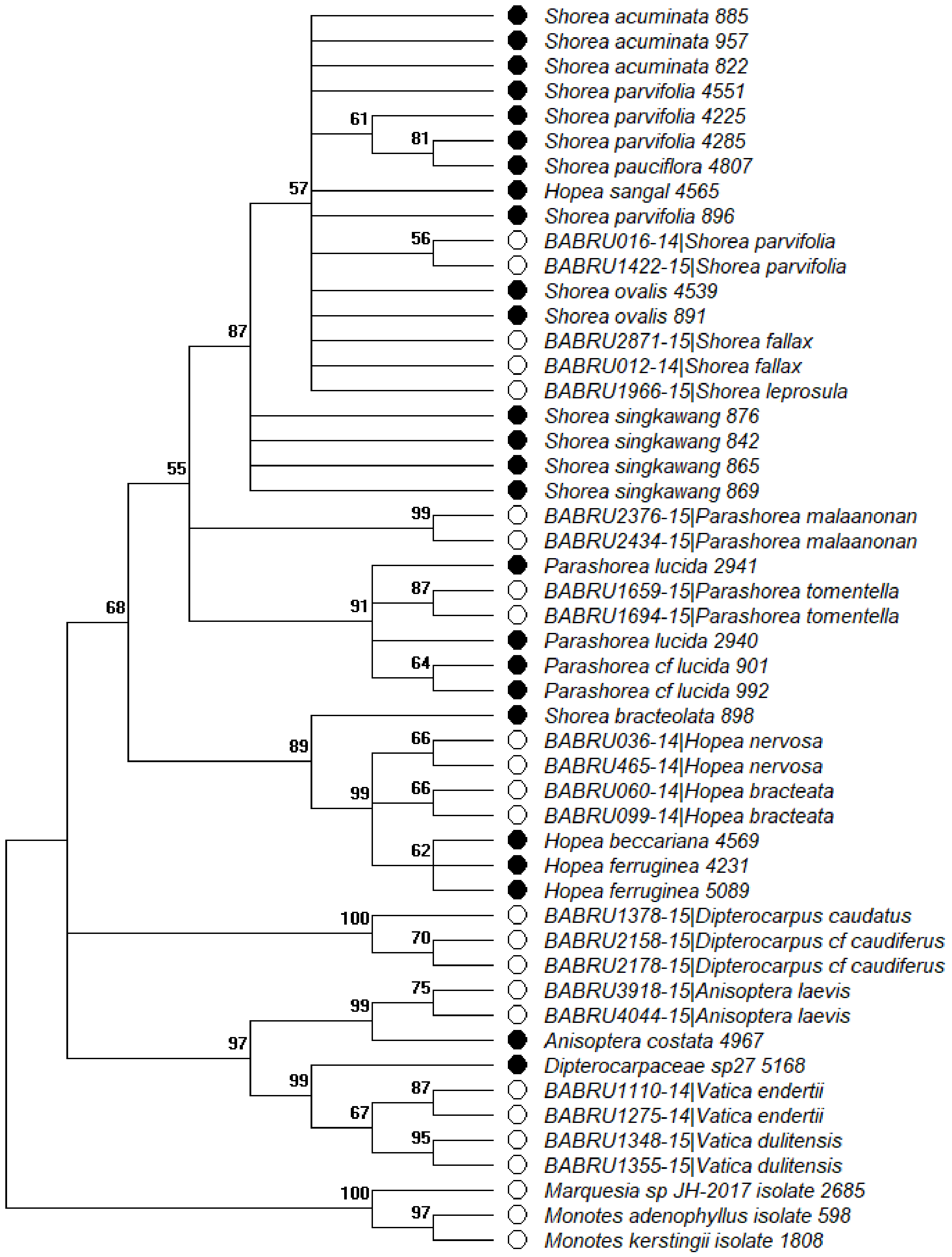

● Shorea parvifolia 4551
● Shorea parvifolia 896
● Shorea ovalis 4539
● Shorea acuminata 957
● Shorea acuminata 885
● Shorea acuminata 822
57
● Hopea sangal 4565
60
● Shorea parvifolia 4225
80
● Shorea parvifolia 4285
● Shorea pauciflora 4807
● Shorea ovalis 891
○ BABRU012-14|Shorea fallax
82
○ BABRU016-14|Shorea parvifolia
60
○ BABRU1422-15|Shorea parvifolia
○ BABRU1966-15|Shorea leprosula
○ BABRU2871-15|Shorea fallax
● Shorea singkawang 876
57
● Shorea singkawang 869
● Shorea singkawang 842
● Shorea singkawang 865
98
○ BABRU2376-15|Parashorea malaanonan
○ BABRU2434-15|Parashorea malaanonan
59
84
○ BABRU1659-15|Parashorea tomentella
55
○ BABRU1694-15|Parashorea tomentella
● Parashorea cf lucida 992
65
53
● Parashorea cf lucida 901
● Parashorea lucida 2940
● Parashorea lucida 2941
● Shorea bracteolata 898
○ BABRU036-14|Hopea nervosa
89
○ BABRU060-14|Hopea bracteata
○ BABRU099-14|Hopea bracteata
100
○ BABRU465-14|Hopea nervosa
● Hopea beccariana 4569
● Hopea ferruginea 4231
● Hopea ferruginea 5089
100
○ BABRU1378-15|Dipterocarpus caudatus
76
○ BABRU2158-15|Dipterocarpus cf caudiferus
○ BABRU2178-15|Dipterocarpus cf caudiferus
81
○ BABRU3918-15|Anisoptera laevis
98
○ BABRU4044-15|Anisoptera laevis
● Anisoptera costata 4967
94
90
○ BABRU1110-14|Vatica endertii
○ BABRU1275-14|Vatica endertii
99
● Dipterocarpaceae sp27 5168
96
○ BABRU1348-15|Vatica dulitensis
○ BABRU1355-15|Vatica dulitensis
100
○ Marquesia sp JH-2017 isolate 2685
91
○ Monotes adenophyllus isolate 598
○ Monotes kerstingii isolate 1808

Appendix 10 The maximum likelihood phylogenetic tree of trees' samples collected in the EFForTS project based on the *matK* marker and the Hasegawa-Kishino-Yano model. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. Condensed tree, branches with support values lower than 50% have been collapsed. The tree log likelihood is (-1689.38). The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.



● Shorea singkawang 869
● Shorea singkawang 876
● Shorea singkawang 865
● Shorea singkawang 842
● Shorea parvifolia 4551
● Shorea parvifolia 4285
● Shorea parvifolia 4225
● Shorea ovalis 891
● Shorea acuminata 957
● Shorea acuminata 885
● Shorea acuminata 822
○ BABRU2871-15|Shorea fallax
○ BABRU1966-15|Shorea leprosula
○ BABRU012-14|Shorea fallax
● Shorea ovalis 4539
68
● Parashorea cf lucida 901
● Parashorea cf lucida 992
85
○ BABRU1659-15|Parashorea tomentella
○ BABRU1694-15|Parashorea tomentella
● Parashorea lucida 2940
● Parashorea lucida 2941
○ BABRU016-14|Shorea parvifolia
56
○ BABRU1422-15|Shorea parvifolia
● Hopea sangal 4565
● Shorea parvifolia 896
● Shorea pauciflora 4807
65
○ BABRU060-14|Hopea bracteata
○ BABRU099-14|Hopea bracteata
● Shorea bracteolata 898
63
○ BABRU036-14|Hopea nervosa
○ BABRU465-14|Hopea nervosa
● Anisoptera costata 4967
89
○ BABRU3918-15|Anisoptera laevis
62
○ BABRU4044-15|Anisoptera laevis
○ BABRU1355-15|Vatica dulitensis
○ BABRU1110-14|Vatica endertii
58
● Dipterocarpaceae sp27 5168
○ BABRU1275-14|Vatica endertii
○ BABRU1348-15|Vatica dulitensis
59
○ BABRU2376-15|Parashorea malaanonan
○ BABRU2434-15|Parashorea malaanonan
● Hopea beccariana 4569
● Hopea ferruginea 4231
● Hopea ferruginea 5089
72
○ BABRU1378-15|Dipterocarpus caudatus
67
○ BABRU2158-15|Dipterocarpus cf caudiferus
○ BABRU2178-15|Dipterocarpus cf caudiferus
93
○ Marquesia sp JH-2017 isolate 2685
99
○ Monotes adenophyllus isolate 598
○ Monotes kerstingii isolate 1808

Appendix 11 The maximum likelihood phylogenetic tree of trees' samples collected in the EFForTS project based on the *rbcL* marker and the Hasegawa-Kishino-Yano model. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. Condensed tree, branches with support values lower than 50% have been collapsed. The tree log likelihood is (-943.45). The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.


● Shorea acuminata 885
● Shorea acuminata 957
● Shorea acuminata 822
● Shorea parvifolia 4551
61
● Shorea parvifolia 4225
81
● Shorea parvifolia 4285
● Shorea pauciflora 4807
57
● Hopea sangal 4565
● Shorea parvifolia 896
56
○ BABRU016-14|Shorea parvifolia
○ BABRU1422-15|Shorea parvifolia
● Shorea ovalis 4539
● Shorea ovalis 891
87
○ BABRU2871-15|Shorea fallax
○ BABRU012-14|Shorea fallax
○ BABRU1966-15|Shorea leprosula
● Shorea singkawang 876
● Shorea singkawang 842
55
● Shorea singkawang 865
● Shorea singkawang 869
99
○ BABRU2376-15|Parashorea malaanonan
○ BABRU2434-15|Parashorea malaanonan
● Parashorea lucida 2941
87
○ BABRU1659-15|Parashorea tomentella
68
91
○ BABRU1694-15|Parashorea tomentella
● Parashorea lucida 2940
64
● Parashorea cf lucida 901
● Parashorea cf lucida 992
● Shorea bracteolata 898
66
89
○ BABRU036-14|Hopea nervosa
○ BABRU465-14|Hopea nervosa
99
66
○ BABRU060-14|Hopea bracteata
○ BABRU099-14|Hopea bracteata
62
● Hopea beccariana 4569
● Hopea ferruginea 4231
● Hopea ferruginea 5089
100
○ BABRU1378-15|Dipterocarpus caudatus
70
○ BABRU2158-15|Dipterocarpus cf caudiferus
○ BABRU2178-15|Dipterocarpus cf caudiferus
75
○ BABRU3918-15|Anisoptera laevis
99
○ BABRU4044-15|Anisoptera laevis
97
● Anisoptera costata 4967
● Dipterocarpaceae sp27 5168
99
87
○ BABRU1110-14|Vatica endertii
67
○ BABRU1275-14|Vatica endertii
95
○ BABRU1348-15|Vatica dulitensis
○ BABRU1355-15|Vatica dulitensis
100
○ Marquesia sp JH-2017 isolate 2685
97
○ Monotes adenophyllus isolate 598
○ Monotes kerstingii isolate 1808

Appendix 12 The maximum likelihood phylogenetic tree of trees' samples collected in the EFForTS project based on the concatenated *rbcL* and *matK* markers, and the Hasegawa-Kishino-Yano model. Five problematic samples have been excluded (samples ID 721, 753, 754, 4121, 4591). The numbers at the tree nodes represent bootstrap values based on 1000 replicates. Condensed tree, branches with support values lower than 50% have been collapsed. The tree log likelihood is (-2708.98). The number after the sample name refers the sample ID. ○ - the database samples from BOLDSYSTEMS; ● - the plot samples.

Declaration of originality

I, Kevin Jair Hernandez Bado, hereby declare that I am the one and only author of this master thesis in all of the development processes, and entitled as “The Use of Barcoding Sequences for The Construction of Phylogenetic Relationships in The Dipterocarpaceae Family”. All data sources and references has been properly cited. Additionally, this project has not been submitted or shared in any form to another party outside the Forest Genetics and Forest Tree Breeding Department, Faculty of Forest Science and Forest Ecology at the Georg- August University Göttingen.

Signed: ________________________________