DNA Sequence-Based Identification and Molecular Phylogeny Within Subfamily Dipterocarpoideae (Dipterocarpaceae)

Dissertation

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Ph.D.)

at Forest Genetics and Forest Tree Breeding, Büsgen Institute

Faculty of Forest Sciences and Forest Ecology

Georg-August-Universität Göttingen

By

Essy Harnelly

(Born in Banda Aceh, Indonesia)

Göttingen, 2013

Supervisor	:	Prof. Dr. Reiner Finkeldey
Referee	:	Prof. Dr. Reiner Finkeldey
Co-referee	:	Prof. Dr. Holger Kreft
Date of Disputation	:	09.01.2013

To My Family

Acknowledgments

First of all, I would like to express my deepest gratitude to Prof. Dr. Reiner Finkeldey for accepting me as his PhD student, for his support, helpful advice and guidance throughout my study. I am very grateful that he gave me this valuable chance to join his highly motivated international working group. I would like to thank Prof. Dr. Holger Kreft and Prof. Dr. Raphl Mitlöhner, who agreed to be my co-referee and member of examination team.

I am grateful to Dr. Kathleen Prinz for her guidance, advice and support throughout my research as well as during the writing process. My deepest thankfulness goes to Dr. Sarah Seifert (*in memoriam*) for valuable discussion of my topic, summary translation and proof reading. I would also acknowledge Dr. Barbara Vornam for her guidance and numerous valuable discussions about my research topic. I would present my deep appreciation to Dr. Amarylis Vidalis, for her brilliant ideas to improve my understanding of my project.

My sincere thanks are to Prof. Dr. Elizabeth Gillet for various enlightening discussions not only about the statistical matter, but also my health issues. Many thanks are to Prof. Dr. Martin Ziehe for his support and recommendation to DAAD. I am also grateful to get acquainted with Alexandra Dolynska, who gave me enormous helps during my lab works. Without her help it would be difficult for me to finish my study on time. I would also like to acknowledge Gerold Dinkel for his help to solve computer problem. My thanks also go to Regina Berkeley, who always helps me with the administration matters during my study.

I would like to acknowledge former students: Dr. Nga Phi Nguyen, Dr. Yanti Rachmayanti, and Dr. Hani Siti Nuroniah for the data and discussion. Many thanks to my roommates: Dr. Alexandra Kuchma and Dr. Yazar Minn, for their kindness, help and support during my stay in Germany. I also would like to thank all my friend in the Institute; Randy Villarin, Konstantina Kameubun, Dr. Rajendra K.C., Dr. Chunxia Chung, Dr. Marius Ekue, Markus Müller, Fitri Yola Amandita, and Crusty Estoque, for knowledge sharing and discussion.

I need to acknowledge the Aceh government for granting me the scholarship, and Syiah Kuala University for allowing me to take study leave overseas. I would also like to acknowledge The German academic exchange service (DAAD) who gives me their best assistant during my study in Germany.

My sincerest thanks to all Acehnese student societies in Germany for their kind support, to my friends from Aceh in Göttingen for making me feel like home, especially to Maimun Rizal, Ikhlas Ali Amran , Rezky Syahrezal, Ichsan, Rita Khathir, Faradilla Fadlia and Febi Mutia for their help and support during my study. I also appreciate the Indonesian community for their support and warm friendship during my study in Göttingen. Many thanks to Dr. Siti Nurleily Marliana and Alex for proof reading this dissertation. I would like to deliver my sincere thanks to my best friends in Biology Department Syiah Kuala University: Betty Mauliya, Zumaidar, Yekki Yasmin, Fauziah and Leni Fitri, who always give support and caring during my study.

My deepest gratitude to my mother (*in memoriam*), who has a dream about going to Germany, which finally drove me to this country, to my mother in law for her support and prayer throughout my study, and to my brothers and sisters for their love and caring. My sincere thanks to my sister Enny Rohainy and her husband Prof. Dr. Herry Suhardiyanto for their love and support in my life. I would also like to acknowledge the support and warm discussion with my niece, Asadatun Abdullah.

Finally, I would like to express my profound gratitude to my husband, Eka Oktavianus, for his patient, love, and encouragement, as well as to my great children: Shofiya Assyifa, Adullah Abdul Aziz, and Najiya Assyifa, who showed me their brave, patient, and independent that made my study life easier.

Content

List of Figures

List of Tables

1 Introduction

1.1 Background of the study

Taxonomy has been in development for more than 250 years. However, only about 1.7 million species have been identified so far (Hawksworth, 1995). Approximately 86% of the species on earth still remain unknown, despite Carolus Linnaeus' introduction of the modern taxonomic system in 1758 (Lawton and May, 1995; Mora et al., 2011). The decrease of taxonomist is one of the impediment in limiting our understanding of species diversity in plants. Traditional taxonomic practices that are mainly based on morphological characters have a limited potential to enhance our understanding of species diversity in plants.

There are at least three problems in identifying plant species using traditional taxonomic practices: (a) limited literature and herbaria data access (Meyer and Paulay, 2005), (b) mis-identification of a taxon because of the resilience of a taxonomic trait, and (c) high time consumption. Moreover, the lack of trained taxonomists and parataxonomists, particularly in biodiversity hotspots in developing countries, is another problem that needs to be overcome.

Furthermore, the rapid decline of plant species as a result of many factors such as climate change and illegal logging are encouraging plant taxonomists to find fast and accurate methods of plant species identification before they become extinct (Finkeldey et al., 2010; Meyer and Paulay, 2005). The fast growth of DNA sequencing technologies in the past 20 years opens opportunities to resolve the mentioned problems.

Besides relying on morphological characters, plant taxonomists increasingly use molecular data. Much research has been done on the reliability of molecular data in supporting the plant identification system (Finkeldey et al., 2010; Nuroniah et al., 2010; Wesselink and Kuiper, 2008).

Therefore, in an attempt to find a fast and accurate method of plant identification, the main goal of this study is to develop a molecular identification key based on chloroplast regions using the subfamily Dipterocarpoideae as an object. More details about the project and the plant tribe used will be explained in separate chapters.

1.2 The Requirement of Molecular Taxonomic Tools

Molecular taxonomy is the classification of plant species using DNA data. These molecular methods provide broad taxonomic information for species delineation, which is available at the interspecific levels (Mayo et al., 2008). Generally, the comparison of species among lower or higher taxonomic levels by molecular systematic data requires a particular homologus region of the DNA sequence to be compared.

In 2003, Professor Paul Hebert from Guelph University proposed a quick, simple and economic tool for identifying biological diversity known as DNA barcoding (Hebert et al., 2003). This method involves comparing a short, standardized DNA region of an unknown species with that of a described species in a database. This method requires two components to obtain best results: a particular DNA sequence that has been named (Tautz et al., 2003) and high sequence variation of this DNA barcode. The mitochondrial region, *cox*I (cytochrome c and oxidase subunit 1) gene was at that time proposed as a standard DNA barcode. However, the implementation of this DNA barcode does not work as well in the identification of land plants as it does for animals. Thus, the *cox*I gene is only applied well in animal identification, moreover, the mitochondrial region in land plants has a high number of invariance, low rate of nucleotide substitution and non-conserved regions (Haider, 2011; Kress, 2005).

In an effort to determine the most reliable barcode region for land plants, chloroplast genes (cpDNA) were proposed. This is because cpDNA has quite conserved regions uniparental inheritance, is easy to isolate, and has stable genetic structure (Kress, 2005). Therefore, in 2009, The Plant Working Group in The Consortium for the Barcode of Life (CBOL) recommended the chloroplast regions *matK* and *rbc*L as a core barcode region for land plants because this sequences have high variation between species but low within species (Hollingsworth et al., 2011). Nevertheless, this barcode core for land plant is still debatable because both *mat*K and *rbc*L sometimes fail to work in some plants (Roy et al., 2010). Therefore, until now there still no a universal barcode available for land plant. However, cpDNA regions are still extensively used for plant molecular phylogenies at different taxonomy ranks. This is because the non-coding region (intergenic spacer) of cpDNA are usually quite variable to accomplish systematic studies at lower taxonomic levels (Haider, 2011; Shaw et al., 2005).

The discriminatory power proposed by DNA barcoding is based on sequences similarity and homology within species to conduct the identification. Sequence comparisons are facilitated by search tools such as Basic Local Alignment Tools (BLAST) and MEGABLAST to perform fast identification (Cowan and Fay, 2012). The user provides a query sequence before starting a BLAST search. The BLAST program will find regions of similarity between the query sequence against the sequence database in National Center for Biotechnology Information (NCBI) (Kerfeld and Scott, 2011). The higher the match query sequence to the reference sequence in NCBI the closer the sequence to that species. Unfortunately, these search tools cannot be used as taxon identification tools because they are unable to accurately differentiate between highly similar sequences (Little, 2011).

In addition, DNA barcoding needs the support of phylogenetic analyses. Normally, closely related sequences will indicate sister groups which indicate that these groups share a recent common ancestor (Soltis and Soltis, 2003). The phylogenetic tree will guide us to understand the genetic relationships of the organism as well as to figure out the evolutionary changes which happened during the time. Reliable DNA markers are very important as identification tools (Cawthorn et al., 2011).

Since DNA barcoding concepts are not well-established yet with regard to the definition of discriminatory regions especially for land plants, it is urgently needed to expand related concepts that can be used as tools to identify species. A promising concept is relying on the phylogenetic analysis which depends on DNA polymorphism among sequences so that it can be used as a discriminatory key to distinguish among species. This method can help to minimize misidentifications because it rests on comparative analyses of nucleotide differentiation as important characters to reveal similarities and differences among taxa. An identification key based on molecular sequence data can be a good alternative for identification purposes. The arrangement of the key will be based on the nucleotide polymorphism among sequences in monophyletic groups. The polymorphic nucleotides will be the character state to discriminate among species. This can be conducted because molecular sequence data and DNA molecular techniques are widely available now.

1.3 Molecular Identification Tools for Dipterocarps

The Dipterocarpaceae family plays a very important role as a source of timber in the tropical lowland rainforests of Southeast Asia. This family has three subfamilies, 17 genera and approximately 500 species that are spread across the tropical regions of Africa, Asia and South America (Ashton, 1982). Dipterocarpaceae *sensu lato* includes the following three subsubfamilies: Dipterocarpoideae in Asia, Pakaraimoideae in South America and Monotoideae in Africa and South America (Apanah, 1993).

The subfamily Dipterocarpoideae was selected as the subject of this study for the following reasons:

a. It has the highest number of species compared with the other subfamilies.
   Dipterocarpoideae has approximately 400 species, and is considered to have high biological diversity (Ashton, 1982).
b. Species belonging to the subfamily have good timber quality.
   It is well known that Dipterocarpoideae consists of many species with good timber quality, such as those in *Shorea*, the main genus in the subfamily.
c. Many species are threatened.
   As a consequence of the high demand of good timber, many Dipterocarpoideae species are endangered (IUCN, 2011)

To prevent a rapid decline of threatened forest species as a result of illegal logging, reliable and efficient tracing methods for forest tree species are urgently needed (Finkeldey et al., 2010). Although many countries have been using wood tags/wood labels to certify certain woods that can be cut down, many industrial wood processers fraudulently remove the labels. When the labels are removed, it is extremely difficult to distinguish the wood because of the high similarity in wood morphology and anatomy, but not necessarily in terms of DNA sequence variation. Therefore, DNA extraction protocols from woody tissue have been developed to apply molecular markers for wood certification (Rachmayanti et al., 2006).

Several molecular studies have attempted to develop tools for Dipterocarpaceae in the context of wood certification and timber forensic profiling. Rachmayanti et al. (2009) optimized DNA extraction protocols for Dipterocarp woods, and Nuroniah et al. (2009) developed a diagnostic marker for the identification of the tree species *Shorea leprosula* Miq. and *Shorea*

parvifolia Dyer, as well as the geographic origin of *Shorea leprosula* Miq using specific PCR (Polymerase Chain Reaction) markers/SCAR (Sequence Characterized Amplified Region) markers. Tnah et al. (2010) developed STR markers of *Neobalanocarpus hemii* for forensic DNA profiling.

In addition, molecular analyses have been conducted to clarify phylogenetic relationships among Dipterocarpaceae species (Kajita et al., 1998; Morton, 1999; Kamiya et al., 2005; Ishiyama et al., 2003; Yulita et al., 2005, Indrioko et al., 2006). The taxonomic treatment and phylogenetic arrangement of taxa is particularly controversial for the species-rich genera *Shorea* and *Hopea* (Dayanandan et al., 1999;Yulita et al., 2005).

Recent advances in molecular sequence technologies have enabled rapid and reliable authentication of Dipterocarp timber. A specific molecular database has been promoted for classifying *Shorea* species and the technique has been used for checking the legitimacy of timber and wood products (Tsumura et al., 2011). This database enables the identification of *Shorea* and its closely-related species among Dipterocarps using the FASTA software (http://f5002.ffpri-108.affrc.go.jp/shorea/). However, for effective certification programs, the development of a database should go along with the enhanced use of advanced molecular taxonomic identification tools in order to reliably discriminate as many species as possible.

1.4 Rationale of the study

The aims of this study are to evaluate the suitability of the *mat*K and *rbc*L regions in distinguishing Dipterocarpaceae species and to study the possibility of developing a molecular taxonomic identification key for Dipterocarpaceae based on phylogenetic analyses. This study also aims to investigate partial sequences of four chloroplast DNA regions in order to elucidate the phylogenetic relationships within subfamily Dipterocarpoideae.

2 Plant molecular systematics

Systematics refers to discovering, describing, naming, documenting and then classifying species based on phylogenetic analyses of evolutionary changes. Systematics plays a central role in the field of biology as the means of characterizing and identifying organisms (Schuh, 2000; Singh, 2004). One of the most important aspects of systematic and phylogenetic analyses is reconstructing the historical relationships of groups of biological organisms. A correctly inferred phylogeny may provide knowledge of species' relationships, which can then benefit studies in related fields as, for example, ecology and biogeography (Kreft and Jetz, 2010; Soltis and Soltis, 2003).

Plant molecular systematics can be defined as the use of genetic information, such as that obtained from nucleotides, to support taxonomic identification. In molecular systematic analyses, a hierarchy's arrangement is based on the homology of a DNA sequence from closely related species. The homology of DNA sequences needs to consider whether similar sequences share a common evolutionary history (Simpson, 2006). An advantage of using DNA sequences instead of morphological characters is related to the evolution of DNA sequences: DNA sequences maintain records of their ancient past as well as of their more recent history during evolution (Tautz et al., 2003).

The choice of a suitable DNA region is the most important consideration when inferring a phylogenetic relationship from molecular data (Soltis et al., 1998). The selection of a proper region is important since slowly evolving regions provide little information to the fully-resolved phylogeny, while quickly evolving regions lead to homoplasy as a result of multiple changes (Soltis and Soltis, 2000).

2.1 DNA Sequence Data

DNA sequencing is the process of determining the order of the nucleotide bases-A (adenine), G (guanine), C (cytosine) and T (thymine) present in a target DNA molecule. The process of DNA sequencing has been developing for over forty years. In the early-1970s, researchers used methods based on chromatography to obtain the first sequences. These techniques were followed by dye-based methods with automated analyses (Simpson, 2006).

The fast advance in DNA sequencing technology in the late-20th century has resulted in a tremendously high amount of DNA sequences, also leading to advances in the concepts with which species are identified and classified. DNA sequence analyses became useful tools in helping taxonomists to characterize the evolutionary relationships between lineages, and even identify the early stages of speciation (Brinegar, 2009). Comparing the homology of DNA among understudied taxa will provide the characters that can be used to infer the phylogenetic relationships among a large number of species (Simpson, 2006).

The use of DNA sequences for phylogenetic analysis of evolutionary processes at the molecular level requires information contained in nuclear and extranuclear genomes. The three basic types of sequence data generated from the genomes are nuclear DNA (nDNA), chloroplast DNA (cpDNA) and mitochondrial DNA (mtDNA).

2.1.1 Nuclear DNA

Nuclear DNA is generally used in evolutionary as well as phylogenetic studies. Nuclear DNA is transmitted from parent to offspring by nuclear division through sexual or asexual reproduction (Simpson, 2006). Since a nuclear genome is biparentally inherited, it is expected to provide more information than a chloroplast or mitochondrial genome on species identity, including hybridization. One of the more useful types of nuclear DNA sequences is the internal transcribed spacer region (ITS), which contains multiple DNA copies. The ITS region lies between 18S and 26S nuclear ribosomal DNA (nrDNA) (Fig. 2.1).

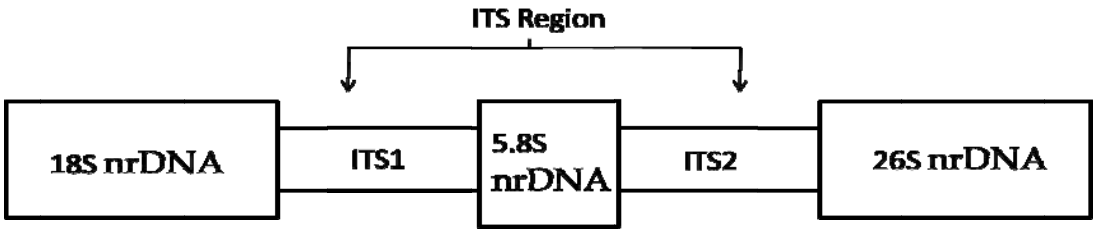


Fig 2.1 The internal transcribed spacer (ITSs) of nuclear ribosomal DNA scheme

Currently, ITS sequence data has been most valuable for inferring phylogenetic relationships at lower levels because this region is found in numerous copies in both animal and plant cells,

biparentally inherited, and has the potential to distinguish between closely-related species. This region is recommended for inferring the phylogenetic relationship among plants because of intra-genomic variability and high mutation rates (Kress, 2005). Hollingsworth et al. (2009) advocates the use of the ITS region for plant species that have limited variation in the plastid genome.

2.1.2 Mitochondrial DNA

Mitochondrial genomes are ubiquitous throughout the eucaryotic cell, encoding necessary proteins involved in energy production, as well as playing an important role in the development and reproduction of the plant (Stuessy, 2008).

Mitochondrial DNA is not recommended as a source of phylogenetic information by plant systematists, since this region is poorly conserved. One of the reasons why this genome is not often used in systematic studies in plants is because it is large in size (ranging between 200 and 300 kb), and it has widespread intra- and inter-molecular recombination. This genome is also not appropriate for most plant species because of a much slower rate of cytochrome c oxidase I gene evolution in vascular plants compared with animals (Kress, 2005).

2.1.3 Chloroplast DNA

The genomes of chloroplasts, which are responsible for photosynthesis, provide rich evolutionary and phylogenetic information. The chloroplast genome is most widely used as a source of information on the inference of the evolutionary patterns and processes of plants (Raubeson & Jansen, 2005), because this genome is thought to evolve slowly, with low mutation rates and maternal inheritance in most angiosperms, along with being a conserved region in structure and gene order.

A chloroplast DNA marker that is maternally inherited shows more conserved DNA patterns compared with a nuclear gene that is biparentally inherited. Chloroplast DNA replicates and divides independently of the nucleus. The chloroplast genome can be divided into three functional categories: exons, introns and intergenic spacers (Fig. 2.2), the latter two of which

do not encode proteins and are referred to as noncoding region (Shaw et al., 2005). Noncoding sequences such as introns and intergenic spacer regions of chloroplast DNA, became important tools in the phylogenetic analysis of a broad range of plant groups at a variety of taxonomic levels (Kelchner, 2000). These regions are supposed to evolve more rapidly than coding regions, enabling it to serve as a primary source of data for molecular systematic, phylogeographic, and population genetic studies of plants.

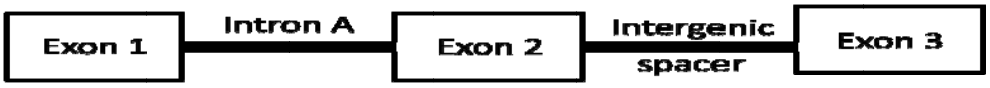


Fig 2.2 The scheme of chloroplast genome with three exons, an intron and an intergenic spacer

Noncoding chloroplast DNA sequences were first used to generate plant phylogenies when three noncoding regions were amplified by universal primers (Taberlet et al., 1991). In 1994, Gilley and Taberlet revealed that the noncoding chloroplast regions are more phylogenetically informative than the coding regions at lower taxonomic levels. This is because they can provide useful characters for phylogeny reconstruction based on a large number of mutations as shown for the genera *Gentiana* and *Euphorbia* (Gielly and Taberlet, 1994). There are many chloroplast regions used in phylogenetic studies. The following four chloroplast regions were frequently used in recent studies.

2.1.3.1 *trn*L intron region

The intron of the encoding chloroplast tRNA for Leucin (*trn*L_{UAA}), also known as the *trn*L intron, is presently widely used for reconstructing phylogenies between closely-related species and for identifying plant species (Zhou et al., 2008). This intron is present in almost all land plants and carophyte algae (Shaw et al., 2005). The location of the *trn*L intron is between the tandemly-arranged *trn*A gen, *trn*T_{UGU} and *trn*F_{GAA} in the large single-copy region of the chloroplast genome in land plants (Taberlet et al., 1991). The length of this region in land plant chloroplasts varies between ca 250 and 1400 bp (Shaw et al., 2005). The region is often jointly amplified with the intergenic spacer *trn*L-*trn*F (Shaw, 2007). There are several reasons why the *trn*L group I intron and intergenic spacer between *trn*L and *trn*F are

among the most widely-used noncoding DNA regions in plant systematics, namely they are easily amplified, the molecular structures are well known and present in nearly all plant taxa.

2.1.3.2 *pbs*C-*trn*S IGS region

*The psb*C gene is one of the plastid genes encoded for the subunit P680 protein (Photosystem II) that is important for photosynthesis. This gene lies within the large single-copy region (LSC) of the plastid genome. There is an intergenic spacer between the *psb*C and $trnS_{(UGA)}$ genes that is not known to be used for phylogenetic studies, although it has shown potential during preliminary screenings as alternative plastid genes of sufficient length and variation for use in molecular phylogenetic studies in some plants (Graham and Olmstead, 2000). Sequence data on the *psb*C-*trn*S region in *Abies alba* provides preliminary evidence of high intraspecific variation in the noncoding intergenic region compared with the highly conserved *psb*C gene sequence (Ziegenhagen and Fladung, 1997). This intergenic spacer region, combination with other gene regions succeeded to classify *Shorea* species (Tsumura et al., 2011).

2.1.3.3 *rbc*L region

The ribulose-1, 5-bisphosphate carboxylase large subunit (*rbc*L) is encoded for the RuBisCO enzyme, which is important for photosynthesis. The *rbc*L gene, the first gene to be sequenced in plants, exists as a single copy and contains no introns (Zurawski et al., 1981). Since it is one of the most conserved genes in the chloroplast genome, this gene has been widely used in molecular phylogenetic analysis. Because of this gene's conserved region, it is well known for its use as a tool to retrace the evolutionary history of plant groups that diverged a long time ago. Thus, even distantly-related plants will have sequences similar to each other.
The *rbc*L gene, along with a few other highly conserved genes, has assisted in answering questions about the origins of some of the major flowering plant groups.

Most plant phylogenetic studies suggest that the *rbc*L gene is best-used to reconstruct the evolutionary relationship until the generic level but not the species level (Soltis et al., 1998). Therefore, to increase the power of this gene for phylogenetic purposes, it should be combined with more variable region (Vijayan and Tsou, 2010).

2.1.3.4 *mat*K region

The chloroplast *matK* gene, which encodes a maturase enzyme, is one of the most-utilized genes in phylogenetic studies after *rbc*L because it evolves nearly two to three times faster than *rbc*L (Soltis et al., 1998). The *matK* sequence information data have been used successfully to resolve generic and even species-level relationships. The length of this region is about 1550 bp, located within the intron of the chloroplast gene *trn*K and embedded in the group II intron of the lysin gene *trn*K (Vijayan and Tsou, 2010).

The capability of this region as a marker for phylogeny construction is related to the observation that this gene evolves quickly and is abundant in the plant. The *mat*K gene is also frequently used for phylogenetics studies, because with the flanking noncoding intron parts, it is able to co-amplify the gene, so that the complete *trn*K intron is increasingly used. As a consequence, the utility of this region could be extended to the inter- and intra specific level (Muller and Borsch, 2005). The *matK* gene is considered to be one of the most informative loci for determining phylogenetic relationships (Hilu et al., 2003).

The *mat*K gene have been used to study the molecular phylogeny of Dipterocarpaceae in Southeast Asia (Kajita et al., 1998). Another research was done to infer the molecular phylogeny of the subfamily Dipterocarpoideae including 14 genera and 79 species (Gamage et al., 2006).

2.2 Molecular phylogenies of plants

A phylogenetic system classifies taxa based on the evolutionary relationship among them which are often illustrated in a phylogenetic tree (Wiley and Lieberman, 2011). A phylogenetic analysis that using DNA as an information is known as molecular phylogeny. Here, the DNA sequences are used as characters to construct the phylogenetic tree (Lemey et al., 2009).

The most important step before starting a molecular phylogenetic study is choosing the right DNA region and gene. There are several points that need to be considered when choosing the DNA region: (1) the gene should be universal for all species studied, and (2) variation among

the sequences should not be too high or too low (Shaw et al., 2005). If the gene evolves too slowly, there will be very little variation among sequences, whereas if it evolves too fast, it will be difficult to get a reliable alignment of the sequence and estimation of the evolutionary distance.

The advantage of using cpDNA in molecular evolutionary studies has been emphasized by systematists, not only as it facilitates straightforward PCR amplification, as a result of the high copy number, but also because of its uniparental inheritance, which produces unambiguous ancestor descendant relationships where the confounding effect of recombination is alleviated (Birky, 1995).

The evolutionary history among the DNA or protein sequences can be revealed by a phylogeny tree. The trees are built to represent the relationship of the sequences to their ancestor and show which sequence are most closely related (Lemey et al., 2009). Statistical methods are needed to determine the tree topology and calculate the branch lengths that best describe the phylogenetic relationships of the aligned sequences in a dataset.

Many different statistical methods can be used for reconstructing the phylogenetic tree. These methods differ from each other in their assumptions and algorithms of the character state. The most common computational methods applied include distance methods such as Unweighted Pair - Group Methods with Arithmetic Mean (UPGMA) and Neighbour Joining (NJ), and discrete data methods, such as Maximum Parsimony (MP) and Maximum Likelihood (ML) and Bayesian method (Hall, 2011; Lemey et al., 2009; Zhu et al., 2010). The principle of distance methods are to calculate all pairwise distances of the sequences as a distance matrix and group the most similar sequences together. Character-based methods use each character data in all steps of the analysis. In the maximum parsimony method, the observed input sequences are explained with a minimum number of substitutions. In this method, the likely tree is the one that requires the fewest number of changes. Maximum likelihood tries to infer an evolutionary tree by finding a tree which maximizes the probability for the observed data (Hall, 2011; Tamura et al., 2011).

2.3 DNA-based identification

Traditionally, taxa have been identified using morphological characters. Morphological characters have been used to identify species for centuries. It was only recently that botanists realized the limitations of taxonomic analyses based on morphological characters which are influenced by genetic and environmental factors (Tautz et al., 2003).

When no differentiating morphological characters are available, plant identification becomes increasingly challenging. Unclear morphological characters or specimens in poor condition, as well as the existence of cryptic taxa in which the species are reproductively isolated and morphologically similar can lead to misidentification (Hajibabaei et al., 2007; Zulkifli et al., 2012). With the increasing availability of molecular data, overcoming the limitations of morphological characters is much easier because DNA sequences will help to overcome some problems in plant systematics.

In principle, we can use DNA variation as a character to study systematics similar to how we use morphological characters. Even though molecular data have been widely used for species separation and identification throughout the past two decades (Mayo et al., 2008), this method is seen as a new concept requiring specific amplification of plant DNA to reveal enough variability to differentiate species (Ridgway et al., 2003).

2.4 DNA barcoding

DNA barcodes can be defined as short, standard DNA sequences that are used to identify species. This method allows the delimitation of an organism at any stage of development from a tiny tissue sample, whether it is fresh, broken or old. This method also helps to discover new species, which is particularly important for cryptogamic plants (Bell et al., 2012). This new molecular technique benefits from plant diversity surveys, especially those of closely-related species and species-rich genera lacking variation in morphological characters (Dick and Webb, 2012). DNA barcoding was first introduced by Paul Hebert when he succeeded in using a part of the mitochondrial region cytochrome oxidase subunit I (*cox*I) to discriminate animals (Hebert et al., 2003). This region, unfortunately, is not suitable for plants because of the slow rate of evolution of the plant mitochondrial genome (Chase et al., 2005).

To coordinate works on DNA barcoding in eukaryotes, the Consortium for the Barcode of Life (CBOL) was established within the secretariat of the National Museum of Natural History in Washington in 2004. CBOL includes organizations and researchers working in the framework of this approach. The region selected as a DNA barcode, as well as standards of its use, should be approved and ratified by the Consortium. CBOL consists of five working groups, namely, the Data Analysis Working Group, Database Working Group, DNA Working Group, Technology Development Working Group, and, most noteworthy in the context of this review, the Plant Working Group (PWG CBOL).

The principle of this method is to compare the DNA barcode region from a query sample with an available sample in a DNA barcoding database. For this reason, an established DNA barcoding database is critical. The Barcode of Life Database BOLD provides an integrated platform that supports all phases of the analytical pathway, from specimen collection to validation (Ratnasingham and Hebert, 2007).

Searching for DNA barcodes in plants has so far proven to be a challenging task. An appropriate DNA region is necessary for plant. In September 2009, the Consortium for Barcode of Life (CBOL) approved *rbc*L and *matK* as the core barcodes for land plants, because *rbc*L is easy to use, but has modest discriminatory power, while *matK* has higher discriminatory power, but lower universality. Peter Hollingsworth (2011), the chair of the Plant Working Group, explained that there are three important factors in DNA barcoding: standardization, minimalism and scalability. Thus, there should be one or more standard DNA regions that can apply to a large and diverse set of samples, and that enables them to be distinguished from one another.

2.5 Molecular taxonomic identification key

An identification key can be defined as a tool to simplify the specimen identification. A good structured key provides clarity and convinience for the user (Wiley and Lieberman, 2011). During the species identification process, an identification device such as an identification key is required. The identification key, used to narrow down the identity of a taxon, is simply a series of questions consisting of contrasting statements. Traditionally, identification keys are constructed using morphological characters, but for a molecular taxonomic identification key, a DNA sequence serves as an analog. Here, we can use each base position in the gene as a character, and use the specific base that occurs there (A,T,C or G) as a character state.

The construction of a dichotomous key starts with the first pair of leads deciding which base is true for the particular position, with the answer directing the user to a following question until the specimen is identified. There are always two possible bases in every site position. For a molecular identification key, the sites of the polymorphic base refer to a character while the polymorphic base in that position refers to the character state.

2.6 The Dipterocarpaceae family

The Dipterocarpaceae family plays an essential role as the main timber family in the tropical lowland rainforests of Southeast Asia. This family has approximately 470 species in 13 genera which are recognized as the Asian subfamily Dipterocarpoideae, 39 species in two African genera and a monotypic South American genus in the subfamily Monotoideae, and one species of one genus in the South American subfamily Pakaraimoideae (Ashton et al., 1984). Although the center of species diversity of this family is now located in Borneo and its surrounding regions, Ashton et al. (1984) suggest that subfamilies of Dipterocarpaceae originally invaded Asia by way of the Indian fragment of Gondwana (Fig. 2.3).

The name of the Dipterocarpaceae refers the family's characteristic fruit with two wings, which developed from persistent sepals (Ashton, 1982). The long sepals, in general, are considered to have evolved from ancestors that themselves did not have long sepals. This is seen in the family's relatives, none of which have long and persistent sepals (Suzuki & Ashton, 1996). However, in some emergent trees, the wings have become redundant with the

reduction of the larger fruits' propelling function, and species with more than two wings attached to their fruits are common a well.

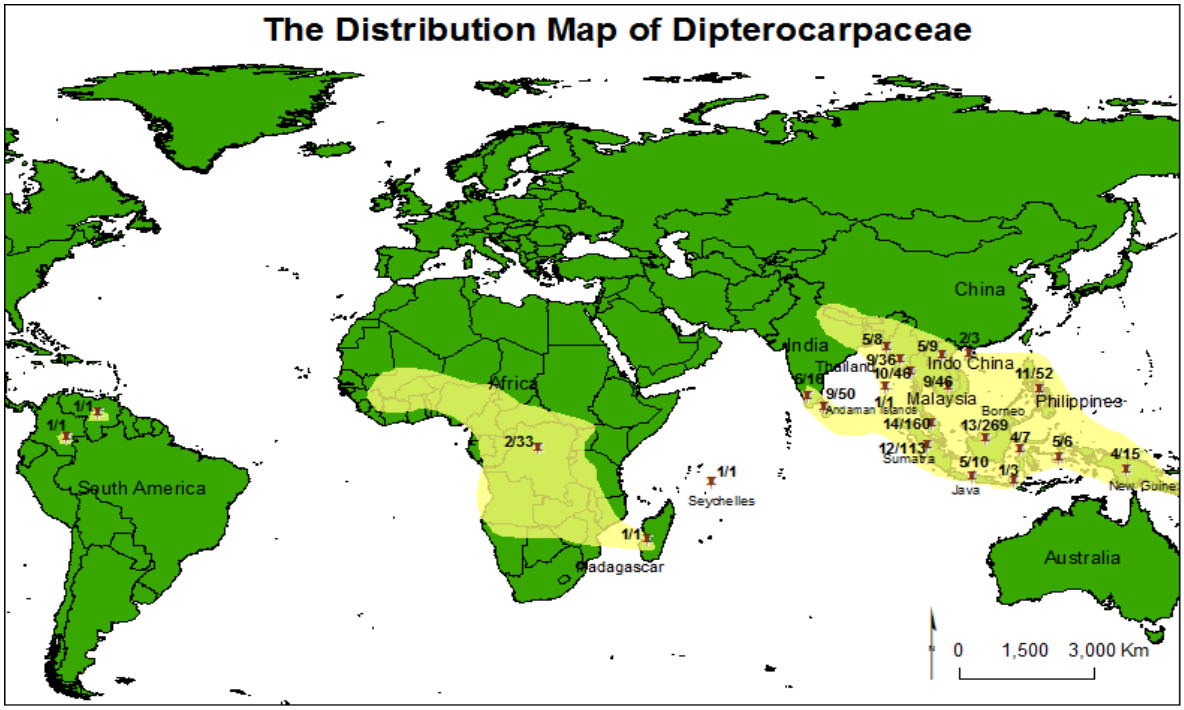


Fig 2.3 The Distribution map of Dipterocarpaceae in South America, Africa and Asia. The shaded areas indicate the extent of the family, labels indicate the numbers of genera/species (Symington, 1943)

Identification of Dipterocarpaceae is not an easy task, because some characteristics vary with a tree's age and habitat (Symington, 1943). Despite the difficulty created by this variability, Rath et al. (1998) reported that DNA polymorphisms are able to discriminate closely-related genotypes.

2.6.1 The subfamily Dipterocarpoideae

The Dipterocarpoideae is the most species-rich subfamily of the Dipterocarpaceae and the one with the highest diversity (Cao et al, 2006), with most of the species found in the genus *Shorea.* Classification of taxa within Dipterocarpoideae has been based on fruit, embryo and seedling characters, chromosome number and wood anatomy (Ashton, 1982; Maury & Curted, 1998). Based on the chromosome number, this subfamily is divided into two tribes, Dipterocarpeae and Shoreae (Ashton, 1982), with haploid chromosome numbers of 11 and

seven, respectively. Based on seed, embryo and seedling characters, (Maury et al., 1975) two main groups are recognized, one with imbricate fruit sepals and the other with valvate fruit sepals. The imbricate group includes two monophyletic genera, *Hopea* and *Shorea*, while the valvate clade includes *Dipterocarpus* and *Vatica*.

In conclusion, most taxonomists agree that the subfamily Dipterocarpoideae comprises two tribes, 13 genera, 17 sections and 12 sub sections. Separation of the tribe is based on the imbricate arrangement of fruit sepals and base chromosome number (Ashton, 1982). Tribe Dipterocarpeae consists of more than 150 species in eight genera (*Dipterocarpus*, *Upuna Cotylelobium*, *Stemonoporus*, *Anisoptera*, *Vatica*, *Vateria* and *Vateriopsis*) and four sections characterized by the valvate arrangement of the fruit sepals. The species rich tribe Shoreae comprises over 300 species (about 200 species of *Shorea* and over 100 species of *Hopea*) in five genera (*Hopea*, *Shorea*, *Neobalanocarpus*, *Parashorea* and *Dryobalanops*), 13 sections and 12 subsections The genus *Dryobalanops* has been considered an intermediate between the two groups and the type genus *Dipterocarpus* is recognized as a basal lineage of the subfamily Dipterocarpoideae (Maury and Curtet, 1998).

2.6.2 Tribe Shoreae

Shorea is the largest and economically most important genus in tribe Shoreae. Based on its wood's anatomy and how it is utilized, this genus can be classified into four sections: White Meranti, Yellow Meranti, Red Meranti and Balau (Selangan Batu), corresponding to the four sections *Anthoshorea*, *Richetioides*, *Rubroshorea* and *Shorea* (Symington, 1943). Compared with the other sections, Red Meranti's plywood is the most expensive.

The long – standing problem in placing *Hopea* and *Shorea* revealed that the taxa are not easily identified at the species level (Yulita et al., 2005). Ashton (1979) pointed out the difficulties in classifying *Hopea, Shorea* and *Neobalanocarpus* because of their morphological similarities. Several recent dipterocarp classification systems generally agree on the placement of *Hopea* and *Shorea* as two closely-related genera, although the placement of most species within these two genera is not clear (Whitmore, 1962; Meijer & Wood, 1964; Meijer & Wood, 1976; Maury-Lechon, 1979).

The differences in the placement and circumscription of *Hopea* and *Shorea* result mainly from the use of different diagnostic characters for the genera and infrageneric groups. For example, Symington (1943) used wood anatomy to divide *Shorea* into four main groups. Ashton (1982) largely followed this classification, but recognized some of Symington's groups at a lower taxonomic rank, thus dividing *Shorea* into 11 sections and giving greater importance to the characters of the fruit calyx, androecium and bark.

Both *Shorea* and *Hopea* have remarkable similarities and exhibit continuous morphological variations at the generic, infrageneric and specific levels. Yulita et al. (2005) assumed that the similar characters between these two genera have led to the recognition of intermediate 'forms' or taxa. Examples of such intermediates are *Parahopea*, *Parashorea* and *Richetia*. This in turn has created controversy as to whether *Hopea* and *Shorea* should be placed as a single genus or be classified into different genera.

2.6.3 Current research on Dipterocarpaceae

Recent studies have focused on using molecular phylogenetic analysis to order some Dipterocarpaceae species (Kajita et al., 1998; Morton et al., 1999 ; Kamiya et al., 2005; Ishiyama et al., 2003; Indrioko *et a*l., 2006; Yulita et al., 2005). The arrangement of taxa within the tribe of Dipterocarpaceae is easily identifiable, but taxonomists have long disagreed on the placement of different genera (Dayanandan et al., 1999; Yulita et al., 2005).

The phylogeny of Dipterocarpaceae has been assessed using several kinds of molecular methods, such as PCR-RFLP analysis of chloroplast genes (Tsumura et al., 1996; Indrioko et al., 2006), analyses of sequences of cpDNA regions (Kajita et al., 1998; Kamiya et al., 1998; Dayanandan et al., 1999), analyses internal transcribed spacer (ITS) regions (Yulita et al., 2005) and analysis of AFLPs (Cao et al., 2006).

Several molecular studies have been conducted for Dipterocarpaceae certification purposes and timber forensic profiling (Nuroniah et al., 2010; Rachmayanti et al., 2009). In 2010, Tnah et al. developed STR markers of *Neobalanocarpus hemii* for forensic DNA profiling. However, there is still insufficient information for robust molecular classification, and further sequence data, covering a greater range of species, are required (Tsumura et al., 2011).

3 Material and Methods

3.1 Material

3.1.1 Data mining for the selection of DNA regions

Dipterocarpaceae molecular sequence data were retrieved from the NCBI website (http://www.ncbi.nlm.nih.gov/Taxonomy/). Homology searches were done by applying the Basic Local Alignment System Tools (BLAST) for nucleotides (nBLAST) in the NCBI database using the MEGABLAST algorithm for highly similar sequences from the public database website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The BLAST program takes the query sequence and searches for the best hits of similar sequence against the entire database of the sequences maintained at NCBI (Mount, 2007b). Prior data mining was done for all types of DNA markers deposited in the NCBI database to have an overview of the most abundant and reliable data for a molecular taxonomic key of Dipterocarpaceae. Based on prior data mining, metadata were developed to summarize the information about the DNA regions and taxa deposited in the NCBI database that were most frequently used by scientists. Those DNA regions and taxa were also selected for the present study.

3.1.2 Additional sequence information from leaf collections

Additional sequences were obtained from leaf collections that are available at the Section Forest Genetics and Forest Tree Breeding, Georg-August–University Göttingen. As a result, it was possible to analyze an additional 76 species (Table 3.1). These additional sequences also included outgroup samples from *Monotoideae* (*Monotes kerstingii*), which originates from Benin, Africa.

Table 3.1 List of additional species from the collection of the Section Forest Genetics and Forest Tree Breeding, Göttingen University

No	Species	Origin country	No	Species	Origin country
1	*Shorea lepida*	Indonesia	41	*Parashorea globosa*	Indonesia
2	*Shorea latifolia*	Indonesia	42	*Shorea montigena*	Indonesia
3	*Shorea fallax*	Indonesia	43	*Shorea javanica*	Indonesia
4	*Shorea pinanga*	Indonesia	44	*Shorea andulensis*	Indonesia
5	*Hopea mengarawan*	Indonesia	45	*Shorea johorensis*	Indonesia
6	*Shorea platyclados*	Indonesia	46	*Shorea splendida*	Indonesia
7	*Shorea guiso*	Indonesia	47	*Hopea malibato*	Philippines
8	*Shorea palembanica*	Indonesia	48	*Hopea philippinensis*	Philippines
9	*Shorea stenoptera*	Indonesia	49	*Hopea plagata*	Philippines
10	*Hopea odorata*	Indonesia	50	*Parashorea malaanonan*	Philippines
11	*Shorea leprosula*	Indonesia	51	*Shorea almon*	Philippines
12	*Hopea dryobalanoides*	Indonesia	52	*Shorea astylosa*	Philippines
13	*Shorea macrophylla*	Indonesia	53	*Shorea contorta*	Philippines
14	*Shorea martiniana*	Indonesia	54	*Shorea negrosensis*	Philippines
15	*Shorea chrysophylla*	Indonesia	55	*Shorea squamata*	Philippines
16	*Shorea parvifolia*	Indonesia	56	*Shorea multiflora*	Indonesia
17	*Shorea acuminata*	Indonesia	57	*Shorea mecystopteryx*	Indonesia
18	*Shorea xantophylla*	Indonesia	58	*Shorea seminis*	Indonesia
19	*Shorea acuminatissima*	Indonesia	59	*Shorea selanica*	Indonesia
20	*Shorea andulensis*	Indonesia	60	*Shorea leptoclados*	Indonesia
21	*Hopea bancana*	Indonesia	61	*Shorea dasyphylla*	Indonesia
22	*Hopea sangal*	Indonesia	62	*Shorea blumuthensis*	Indonesia
23	*Shorea ovalis*	Indonesia	63	*Shorea compressa*	Indonesia
24	*Shorea virescens*	Indonesia	64	*Shorea polysperma*	Indonesia
25	*Shorea materialis*	Indonesia	65	*Shorea pauciflora*	Indonesia
26	*Shorea macroptera*	Indonesia	66	*Shorea atrynervosa*	Indonesia
27	*Shorea leprosula*	Indonesia	67	*Shorea singkawang*	Indonesia
28	*Shorea kuntsleri*	Indonesia	68	*Shorea hofeifolia*	Indonesia
29	*Shorea mujongensis*	Indonesia	69	*Shorea eminiens*	Indonesia
30	*Shorea laevis*	Indonesia	70	*Shorea beccariana*	Indonesia
31	*Shorea smithiana*	Indonesia	71	*Shorea brachteolata*	Indonesia
32	*Shorea teysmaniana*	Indonesia	72	*Shorea pauciflora*	Indonesia
33	*Shorea sandakanensis*	Indonesia	73	*Shorea ochracea*	Indonesia
34	*Hopea celebica*	Indonesia	74	*Shorea sumatrana*	Indonesia
35	*Hopea grifithii*	Indonesia	75	*Upuna borneensis*	Indonesia
36	*Hopea nigra*	Indonesia	76	*Monotes kerstingii*	Africa
37	*Shorea balangeran*	Indonesia			
38	*Shorea scaberrima*	Indonesia			
39	*Shorea faguetiana*	Indonesia			
40	*Parashorea lucida*	Indonesia			

3.2 Methods

3.2.1 Laboratory methods

3.2.1.1 DNA extraction

The total genomic DNA was extracted from about 40 mg of dried, healthy leaves using the DNeasy® 96 Plant Kit (Qiaqen, Hilden, Germany), following the manufacturers protocol. The concentration and quality of the extracted DNA were checked by 0.8-1% agarose gel electrophoresis with a Lambda DNA size marker (Roche) (Sambrook et al., 1989), visualized by UV illumination using a polaroid camera after ethidium bromide staining.

3.2.1.2 Polymerase Chain Reaction (PCR)

Parts of four chloroplast regions were amplified by PCR using previously described primers (Table 3.2). All primers were recommended by different sources. The two recommended plastid regions from the CBOL Plant Working Group (2009), *rbc*L and *mat*K, were included.

Table 3.2 Primers used in the present study

NO	Region	Name of primers	Sequence orientation (5′→ 3′)	Reference
1	*pbs*C-*trn*S IGS	cp6F	GGTCGTGACCAAGAAACCAC	Tsumura et.al., 2011
		cp6iR2	CCCAGAACAAAATGAGAGGT	
2	*trn*L intron	Cp2F	CGA AAT CGG TAG ACG CTA CG	Taberlet et al.,1991
		Cp2R	GGG GAT AGA GGG ACT TGA AC	
3	*mat*K	390f	CGATCTATTCATTCAATATTTC	Cuenoud et al. 2002
		990R	GGACAATGATCCAATCAAGGC	Dayananda et al., 2006
4	*rbc*L	rbcLa_f	ATGTCACCACAAACAGAGACTAAAGC	Kress and Erickson., 2007
		rbcLa_r2	GAAACGGTCTCTCCAACGCAT	Fazekas et al., 2008

The PCR was performed in a Peltier Thermal Cyler PTC-200 (MJ Research Inc.) with a volume of 15µl reaction mixture (Table 3.3). The PCR temperature profiles for the four chloroplast regions are shown in Table 3.4.

Table 3.3 Reaction mixture of PCR reagents

Reagents	Volume (15 µl)
PCR buffer	1,5 µl
$MgCl_2$	1,5 µl
Forward Primer (5 pmol/µl)	1 µl
Reverse Primer (5 pmol/µl)	1 µl
dNTPs	1 µl
Tag	0,2 µl
H_2O	6.8 µl
Template DNA (5-10 ng)	2 µl (5-10ng)

Table 3.4 Temperature profiles for PCR reactions

Step	Condition
Step 1:	Initial denaturation at 95^0 C for 15 minutes
Step 2:	35 cycles of
	Denaturation at 94^0 C for 1 minute
	Annealing at 50^0C for 1 minute
	Elongation at 72^0C for 1:30 minutes
Step 3:	Final extension at 72C for 20 minutes

To obtain purified DNA for sequencing, the DNA products were separated in agarose gels by electrophoresis. The DNA fragments in the agarose gel were sliced with a razor and then purified using the GENECLEAN® Kit (MP Biomedicals, Illkirch, France).

3.2.1.3 Direct DNA sequencing

The sequence data of the chloroplast DNA were obtained through direct sequencing. The sequencing reactions were performed using the ABI Prism^{TM} Big Dye ^{TM} Terminator Cycle Sequencing Ready Reaction Kit v1.1 (Applied Biosystems), based on the principles described by Sanger et al. (Sanger et al., 1977). Data were collected from capillary electrophoresis on an ABI Prism 3100® Genetic Analyzer with the Sequence Analysis Software v3.1 (Applied Biosystems). The sequencing was performed with forward and reverse primers in both directions. The sequencing reaction mixture is shown in Table 3.5, while the temperature profiles of the PCR for sequencing are shown in Table 3. 6.

Table 3.5 Reaction mix of PCR Sequencing reagents

Reagents	Volume (10 µl)
Big Dye	1µl
5X buffer	1,5 µl
Primer (F or R)	1 µl
H_2O	4.5 µl
Template DNA	2 µl (5-10ng)

Table 3.6 Temperature Profiles for Sequencing PCR reactions

Step	Condition
Step 1:	Initial denaturation at 96^0 C for 1 minutes
Step 2:	35 cycles of
	Denaturation at 96^0 C for 10 second
	Annealing at 45^0C for 10 second
	Elongation at 60^0C for 4 minutes
Step 3:	Final extension at 72C for 20 minutes

3.2.2 DNA sequence analysis

The sequences retrieved from the NCBI are a consensus sequence that has been assembled containing both forward and reverse strands. Meanwhile, for the sequences obtained from the laboratory, the CodonCode aligner version 3.7.1 (CodonCode Corporation) was used to edit and confirm the electropherograms of the sequences. The sequences data from the sequencer resulted in an Applied Biosystems (ABI) chromatogram file, which was then scored for quality assignments using the base calling program Phred (CodonCode Corporation). Phred reads DNA sequence chromatogram files and analyzes the peaks to call bases, assigning quality scores ("Phred scores") to each base call. Phred was also used for the assembly of consensus sequences for each sample from the replicate bidirectional sequence reads. The retrieved sequences can be found in the appendix 1 for the four chloroplast regions used in this research, *psb*C-*trn*S IGS, *trn*L intron, *mat*K and *rbc*L regions.

3.2.2.1 Multiple sequence alignment

The chloroplast sequences were aligned using the Clusthal W (Thompson et al., 1994) multiple sequence alignment program, found in Bioedit version 7.0.9. (Hall, 1999). The alignment results were corrected manually. The alignment data of those four regions then

transfered to DNASP v.5.10.01 software in order to get the information about sequence characteristics.

3.2.2.2 Phylogenetic analysis

Phylogenetic analyses of Dipterocarpaceae based on four chloroplast regions were carried out using parsimony and maximum likelihood analysis with MEGA 5 (Tamura et al., 2011). The tree topology was formed using MEGA 5 and the trees were rooted with an outgroup. Poor PCR product quality prevented the same chosen outgroup species from being used for all four chloroplast regions, possibly because the primers were not suitable for each outgroup species. For the *trn*L intron, we used *Monetes kerstingii* from Benin (subfamily Monotoideae) as an out-group, for *psb*C-*trn*S IGS *Upuna borneensis* (tribe Dipterocarpoideae), for *mat*K *Monotes madagascariensis* (subfamily Monotoideae) and for *rbc*L *Monotes kerstingii* (subfamily Monotoideae).

3.2.2.3 Taxonomic identification key based on phylogenetic tree

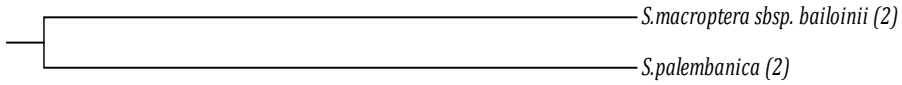
One of the aims of this study is to develop a molecular taxonomic identification key. A molecular identification key was developed based on the clades formed in the phylogeny analysis using maximum parsimony method. The tree is shown in Appendix 12. The plastid region *trn*L intron was chosen as a model for the key because this gene region has the most sequence data available in the NCBI database for the members of subfamily Dipterocarpoideae. The tool is similar to a dichotomous key that uses morphological characters, except that in this study DNA sequences from chloroplast regions were used instead of the morphological characters.

The arrangements of the characters were based on the topology of the phylogenetic trees that formed from the parsimony analysis. The cladogram produced by the phylogenetic analysis then classify based on the clades. According to (McLennan, 2010), a clade is a group of organisms that includes an ancestor and all descendents of that ancestor. Clades are nested within one another and they form a nested hierarchy within a phylogenetic tree. Since every clade share homologous sequences (Chao and Zhang, 2009), the species belong to one clade

should be closely related. However, even though they are closely related, there should be distinct characters that make them separated into different branches.

The following steps describe the construction of the dichotomous key based on the DNA sequences.

Step 1. The key was split based on the clades that were formed by the cladogram produced by the phylogenetic analysis.



Step 2. Using multiple alignments in Clusthal W (Thompson, 1994) the polymorphisms among the monophyletic groups of each clade were characterized.

Step 3. Polymorphic sites were summarized in a table (see table 3.7 as an example).

Table 3.7 Table for polymorphic sites of the species and their nucleotides.

Species	Polymorphic site			
	244	246	275	276
S. macrotera sbsp.bailonii	-	G	A	C
S. palembanica	A	A	C	T

Step 4. The key was constructed based on the polymorphic sites and bases.

1 a. site 244 is (-) …….. 2a

b. if (A) 2b

2. a. site 246 is (G) ……..…….. 3a

b. if (A)... .....3b

3. a. site 275 is (A) ……..……...…. 4a

b. if (C) .....…4b

4. a. site 276 is (T) …….……..... *S. palembanica*

b. If (C) .....*S. macroptera sbsp bailonii*

3.2.2.4 Barcode analysis

We used two barcode regions, *matK* and *rbcL,* which were adopted from the Consortium for the barcode of life (CBOL plant working group 2009) to assess the suitability of these two gene regions to discriminate Dipterocapaceae species. To support the barcode analysis, we performed the phylogenetic analysis based on distance algorithm methods. The query sequences from the laboratory (marked with X) were combined with the sequences retreived from the NCBI database and analyzed using the K2P distance NJ method with MEGA 5 (Kimura, 1980 ; Tamura et al., 2011). The neighbor joining method, which is embedded in MEGA 5, was the chosen as method to construct the phylogenetic trees for the barcode analysis, with the following settings: Kimura's 2 parameter was the chosen model/method. Beside the neighbor joining analysis, we also executed the nBLAST identification from the NCBI website. The purpose of this analysis was to evaluate the reliability of the nBLAST tool as a taxonomic identification method using sequence data because this tool was lately used worldwide as a routine and quick identification system (Kool et al., 2012; Mount, 2007a; Pons, 2006). The known samples from the own laboratory analyses were used as queries in the nBLAST.

4 Results

4.1 DNA sequence characteristics

The numbers of successfully sequenced Dipterocarp samples were not the same for the studied regions, because not all of the four primers used work well with Dipterocarpaceae. Of the four primers sequenced, the *trn*L intron and *rbc*L had a 94% sequencing success rate, while the *mat*K region and *psb*C-*trn*S IGS had success rates of 70% and 76%, respectively (Table 4.1).

The highest numbers of taxa available from the NCBI (Appendix 1) belong to the *psb*C-*trn*S IGS region (210 sequences). Conversely, there are very few *rbc*L region sequences from Dipterocarp species deposited in the NCBI (5 sequences). Of all four primers, the *trn*L intron has the highest combined number of species (145, both deposited in the NCBI and the Forest Genetics and Forest Tree Breeding Institute laboratory) from different genera and tribes, whereas *rbc*L has the lowest number (67 species). The *mat*K gene and *psb*C-*trn*S IGS had 116 and 117 species, respectively (Table 4.1).

Table 4.1 The sequence information of four chloroplast regions

Parameter	*rbc*L	*mat*K	*psb*C-*trn*S IGS	*trn*L intron
Number of Sequences from NCBI	5	109	210	191
Number of additional sequences	71	53	56	143
Sequencing success	93%	70%	75%	93%
Number of species	67	116	117	145
Aligned length	647 bp	635 bp	1136 bp	537 bp

The length of the obtained sequences varied, but the final alignment lengths ranged from 537 bp for the *trn*L intron to 1136 bp for the *psb*C-*trn*S IGS (Table 4.1).

4.1.1 *psb*C- *trn*S IGS region

The amplification and sequencing results using the primer of this region was only moderately successful; only 57 species from 76 leaf samples (75%) from the additional data samples were successfully sequenced. Combining these results with the available data in the NCBI database, which totaled 210 sequences, resulted in a total of 118 species restricted to tribe

Shoreae. The final lengths of the sequences after being aligned and manually edited were 1136 bp (Table 4.1).

4.1.2 *trn*L intron region

Amplification using this region was mostly successful for the additional leaf samples. From 76 samples, we were able to amplify and sequence 71 samples (93%). The individual sequences' length was around 570 bp. The combination of 191 DNA sequences from the NCBI - 71 sequences from the leaf sample collection and 72 from Rachmayanti (2009), from whom the samples were obtained personally, and Nguyen (2009) yielded the highest number of sequences among the chloroplast regions that were used in this study; a total of 334 *trn*L intron sequences representing 145 species from subfamily Dipterocarpoideae. The final length of the refined sequences that will be used for further analysis was 537 bp (Table 4.1).

Table 4.2 Sequence characteristics of four chloroplast regions

Parameter	***rbc*L**	***mat*K**	***psb*C-*trn*S-IGS**	***trn*L intron**
Number of nucleotides	647 bp	635 bp	1136 bp	537 bp
Number of variable sites	47	309	117	112
Number of informative characters	45	234	110	103
Number of haplotypes	27	81	70	61
Haplotype diversity (Hd)	0.875	0.950	0.825	0.850
G+C content	0.431	0.329	0.433	0.320

4.1.3 *mat*K region

This region produced the lowest, albeit moderately, successful sequencing results; `70% of the samples (53 of 76) were successfully sequenced. The final lengths after alignment and manual refinement were 635 bp. The total *mat*K sequences comprised both tribes Shoreae and Dipterocarpeae. Among the four chloroplast regions, this gene region gave the highest

number of informative characters (234) as well as number of variable sites (309) (Table 4.2). The number of haplotypes in this region was also the highest (81).

4.1.4 *rbc*L region

The amplification of the *rbc*L region was mostly successful for Dipterocarpaceae species, particularly those in Shoreae. From a total 76 species sequenced, only 5 species could not be amplified using this region (93% success rate). Using the *rbc*L region to obtain both successful PCR products and sequencing results was easy. Combining these results with the 5 *rbc*L sequences downloaded from the NCBI resulted in a total of 69 different species. The final lengths of the sequences after alignment and manual correction were 647 bp. The *rbc*L gene region has 47 variable sites and 45 informative characters (Table 4.2).

4.2 Molecular phylogeny based on four chloroplast regions

The analyses of the four chloroplast regions using three statistical methods (maximum parsimony, maximum likelihood and neighbor joining) yielded a total of 12 phylogenetic trees. Using *U. borneensis*, *M. madagascariensis* and *M. kerstingii* as an outgoup, the common topologies of the trees showed similar, though not exactly identical patterns. As the three phylogenetic analysis methods resulted in similar patterns, only the most interesting result will be described.

4.2.1 *psbC- trn*S IGS region

The maximum parsimony tree of the *psb*C-*trn*S IGS is shown in Fig. 4.1. Using *U. borneensis* as an outgroup; this gene was able to resolve tribe Shoreae into two clades with a strong bootstrap value of 98%. Some of those clades formed subclades with paraphyletic groups based on the section. This clade comprises a mix of some sections of the *Shorea* group, sister with *Parashorea*, and formed a sister subclade with *Hopea* with a strong bootstrap value (100%). The *Shorea* group of the subclade comprises section *Brachyptera* (S. *almon, S. platyclados, S. pachycarpae, S. kuntsleri, S. scaberrima, S. pauciflora, S.*

johorensis, S. andulensis, S. smithiana, S. pubistylla, S. bullata), section *Mutica* (*S. curtisii, S. macroptera* subsp. *sandakanensis, S. macroptera* subsp. *macropterifolia, S. parvifolia, S. ovata, S. ferruginea, S. quadrinervis, S. teysmaniana, S. rugosa, S. platycarpa, S. acuta, S. macroptera, S. rubra, S. slootenii, S. leprosula, S. dasyphylla, S. argentifolia*), one member of section *Ovalis* (*S. ovalis*), section *Pachycarpae* (*S. amplexicaulis, S. pilosa, S. splendida, S. beccariana, S. mecystopteryx, S. stenoptera, S. macrophylla*) and one member of section *Rubella* (*S. albida*). This subclade branch also comprises three *Hopea* lineages from section *Dryobalanoides* (*H. grifithii, H. nigra*) and section *Hopea* (*H. celebica*) (Fig. 4.1a). Another clade formed a monophyletic group of section *Anthoshorea* that excluded *S. obscura*, which belongs to section *Shorea.* (Fig. 4.1b). Fig. 4.1c shows lineage from section *Richetioides.* Fig. 4.2.1d shows a paraphyletic clade that mostly dominated with section *Shorea* (*S. biawak, S. maxwelliana, S. laevis, S. falciferoides, S. havilandii, S. foxworthyi, S. guiso, S. seminis, S. superba, S. crassa, S. materialis, S. domatiosa, S. inappendiculata, S. atrinervosa*), section *Neohopea* (*S. isoptera*), section *Richetioides* (*S.blumuthensis and S. polysperma*). The results in Fig. 4.1e show the paraphyletic group that comprises a *Hopea* group from section Dryobalanoides (*H. dryopbalanoides, H. mengarawan*) and section *Hopea* (*H. bancana, H. odorata, H. sangal*), sister subclade with *N. hemii,* which is clustered together with *S. astylosa*.

The phylogenetic tree based on maximum likelihood methods of this region (Appendix 4) did not show a great ability to resolve tribe Shoreae. This method resolve one big paraphyletic clade from other member of tribe Dipterocarpeae with high bootstrap support (97%).

This paraphyletic clade comprises species from *Shorea*, *Hopea*, *Neobalanocarpus* and *Parashorea* genera in one group that is separated from other members of the *Dipterocarpeae* tribe, specifically *Anisoptera laevis, Cotylelobium lanceolatum, Vatica bella* and *Vatica oblongifolia*, that formed a sister clade with *Upuna borneensis* as a single outgroup.

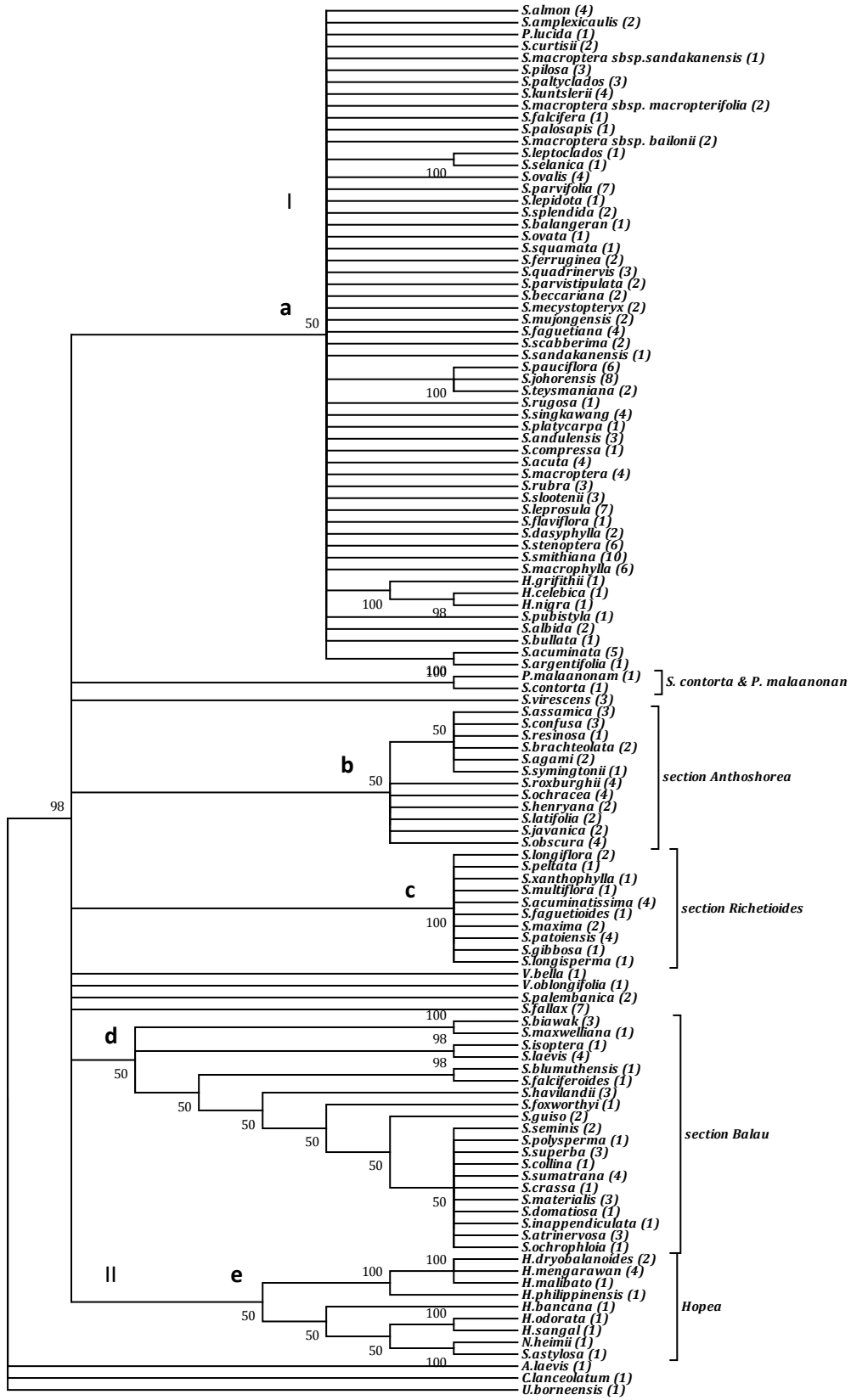


Fig 4.1 The tree of the *psb*C-*trn*S region using maximum parsimony method. (a) Is the paraphyletic clade consist the member of *Shorea*, *Hopea* and *Parashorea*. (b) Is a subclade section *Anthoshorea* that excluded *S. obscura* (c) is lineage from section *Richetioides*. (d) is section Shorea that excluded *S. isoptera* from section *Neohopea*, and *S. blumuthensis* and *S. polysperma* from section *Richetioides*. (E) is paraphyletic group that comprises a *Hopea* sister subclade with *N. hemii* and *S. astylosa*.

The neighbor joining tree of this region that utilized *U. borneensis* as an outgroup, could separate *Cotylelobium* as a sister subclade. This method was also able to separate tribe Shoreae from members of tribe Dipterocarpeae (specifically, *V. bella* and *V. oblongifolia*) with high bootstrap value (100%). Compared with the ML tree, this method was able to form a monophyletic group for subclade *Richetioides* and *Anthoshorea,* with bootstrap values of 62 % and 80 %, respectively. This tree was able to resolve C. *lanceolatum* from other species with a bootstrap value of 61%, and *A. laevis* from other species with a bootstrap value of 67%. This method formed a paralyphyletic clade that comprised several subclades from sections *Shorea,* Richetioides and *Anthoshorea.* Those subclades formed polytomies with the Red Meranti group of *Shorea* (Appendix 5).

4.2.2 *trn*L intron region

The topology of the tree construction using maximum parsimony, maximum likelihood and neighbor joining, were similar for the *trn*L intron region (Fig 4.2 and Appendices 2 and 3). The three methods were able to separate the Dipterocarpeae group (X=11) from the *Shorea*e group (X=7) with bootstrap values between 87% and 99%. However, the cladogram shows that the trees yielded many polytomies.

Using maximum parsimony (Fig. 4.2), the *trn*L intron gene was not able to resolve the Dipterocarpoideae very well. The tree topology showed that the gene could not resolve most of the members of Dipterocarpoideae, with low bootstrap support (12%) (data not shown). This gene could only resolve the *Dipterocarpus* group (tribe Dipterocarpeae) from the members of tribe Shoreae. Additionally, this gene could not resolve other genera from *Dipterocarpeae*, specifically *Vatica, Anisoptera* and *Upuna.*

The results showed many polytomies in one clade, but the subclade showed that the *trn*L intron gene was able to resolve the *Dipterocarpus* group from other sister branches, with a high bootstrap value (99%). A monophyletic group was also formed for members of

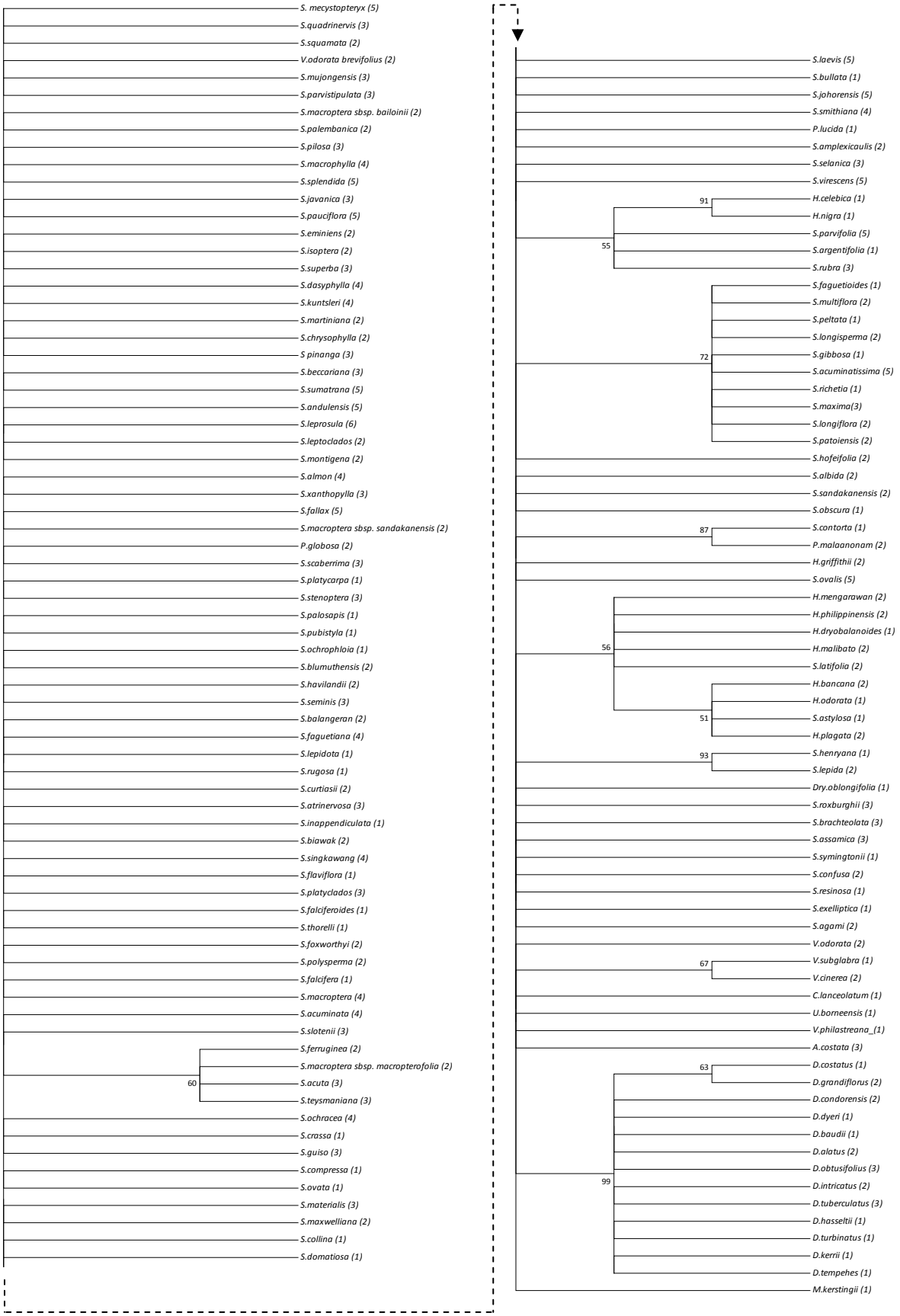


Fig 4.2 The tree of *trn*L intron using maximum parsimony method. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The number in parentheses means number of species tested.

Richetioides (*S. faguetioides, S. multiflora, S. longisperma, S. peltata, S. gibbosa, S. acuminatissima, S. richetia, S. maxima, S. longiflora, S. patoiensis*).

This region could not resolve the *Hopea* genera as a single monophyletic group. The *Hopea* genera were still nested with some of *Shorea* species (*S. astylosa*, and *S. latifolia*) with a bootstrap value of 56%. Meanwhile other member of *Hopea* (*H. celebica* and *H. nigra*) formed sister branch with member of *Shorea* (*S. parvifolia*, *S. argentifolia*, *S. rubra*). (Fig. 4.2).

When using the maximum likelihood method for the *trn*L intron gene, two groups were formed (Appendix 2). One group consisted of members of the genus *Dipterocarpus,* with high bootstrap support (99%), and the other was a mixed group of tribe Shoreae. However, other members of the *Dipterocarpeae* tribe (*Cotylelobium*, *Anisoptera*, *Upuna* and *Vatica*) formed a subclade in this group with high bootstrap support (92%). This method was also able to resolve some sections of tribe Shoreae in the subclade; for example, section *Richetioides* with a bootstrap value of 91%, and section *Anthoshorea* as well as parts of section *Mutica*, with bootstrap values of 62% and 52%, respectively. Some *Hopea* genera formed a group but were still nested with one species of section *Shorea* (*S. astylosa*), while others formed polytomies among the *Shorea* species. The *trn*L intron gene could not group two species of *Parashorea* into a single group. *P. malaanonan* was grouped together with *S. contorta,* while *P. lucida was* nested with other polytomies of *Shorea*.

The neighbor joining method showed similar patterns as the maximum likelihood method (Appendix 3). This method was also unable to resolve members of the Dipterocarpoideae, although some of the genera were grouped within a single subclade. The subclade of *Dipterocarpus* showed one group with high bootstrap support (99%). Other *Dipterocarpeae* members formed another group, also with a high bootstrap value (88%).

A subclade of *Richetioides* formed one monophyletic group, with a high bootstrap value (92%), while a part of section *Anthoshorea*, as well as part of section *Mutica*, which formed a monophyletic group with a bootstrap value of 64% for *Anthoshorea* and 62 % for *Mutica*. The *Hopea* genera did not form one monophyletic group, instead remaining nested with some *Shorea* species. In this tree, some *Hopea* genera formed a paraphyletic group. There were three sister branches among a subclade of *Hopea*. One branch belonged to *S. guiso*, one to genus *Hopea* from section *Hopea* (*H. plagata*, *H. bancana, H. odorata*), which is nested with *S. astylosa*, and another branch belonged to *H. mengarawan* from section *Hopea* and *H. dryobalanoides* from section *Dryobalanops*, itself nested with *S. latifolia*. The neighbor

joining method also produced the same results as the maximum likelihood method for the genus *Parashorea*, the subclade showing that *P. malaanonan* was grouped together with *S. contorta*, while *P. lucida* was nested with other polytomies of *Shorea*.

4.2.3 *mat*K region

The tree topology using *M. madagascariensis* as an outgroup showed that all three statistical methods (MP, ML and NJ) were unable to resolve Diptercarpoidae well. However, all three statistical methods succeeded in using the *mat*K gene region to resolve section *Doona* of Shorea group. The section *Doona* (*S. affinis*, *S. zeylanica*, *S. cordifolia*, *S. gardneri*, *S. worthingtonii*, *S. trapezifolia*, *S. congestiflora*, *S. disticha*, *S. megistophylla*) maintained a stable monophyletic group with strong bootstrap support (< 91), while other sections of both tribes formed paraphyletic groups (Fig. 4.3 Appendix , 6 and 7). The topology of the trees did not show a significant difference. All the trees were able to resolve *Dipterocarpus* as one monophyletic group which formed a sister subclade with the other subclades.

For the maximum parsimony tree, the first subclade, shown in Fig 4.3, was a paraphyletic group, with only moderate bootstrap support (52%). This subclade comprised numerous members of tribe Shoreae (sections *Brachyptarae*, *Mutica*, *Richetioides*, *Ovalis* and *Pachycarpae*).

The second subclade belongs to two members of *Parashorea* (*P. chinensis* and *P.chinensis* var. *kwangsiensis*) in one group supported with a moderately high bootstrap value (63%). This subclades formed a sister branch with other *Parashorea* members; *P.lucida* which is nested in the *Shorea* sub clade, and formed a sister branch with *P.malaanonan*. The tree topology obtained using MP and NJ methods were similar.

The *Neobalanocarpus* genus formed its own sister branch with the *Hopea* (*H. sangal*, *H. bancana*, *H. helferi*, *H. jucunda, H. subalata*, *H. discolor*, *H. nervosa*, *H. latifolia*, *H. jucunda* subsp. *modesta*, *H. malibato, H. mengarawan*, *H. dryobalanoides*, *H. philippinensis*, *H. odorata*, *H. wightiana*, *H. plagata*) and *Shorea* (*S. brachteolata*, *S. virescens*, *S. lepida*,

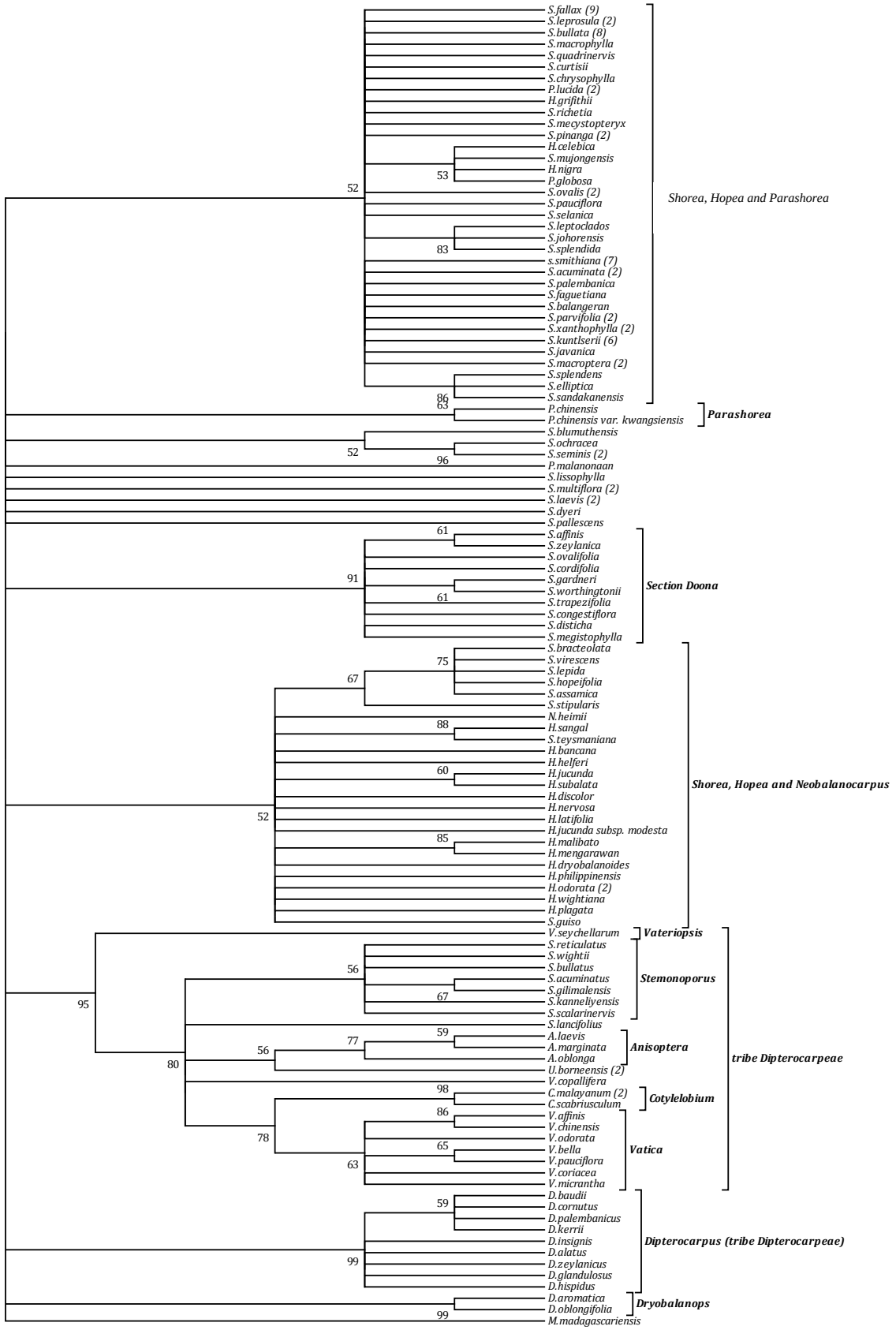


Fig 4.3 The tree of *mat*K region using the maximum parsimony method. The percentage of bootstrap support shown next to the branch. Tree type 1 out of 136 most parsimonious trees shown. consistency index is (0.650246), the retention index is (0.924708), and the composite index is 0.755496. The number in parentheses means number of species tested

S. hopeifolia, *S. assamica*, *S. stipularis*, *S. guiso, S. teysmaniana*) groups, with a bootstrap value of 52%.

Using maximum parsimony (Fig 4.3), a subclade of tribe *Dipterocarpeae member* (excluded *Dipterocarpus*) showed a paraphyletic group. In this clade, *Stemonoporus* species allied with member of *Vatica*, *Anisoptera*, *Upuna* and *Cotylelobium* with strong bootstrap support (95%).

The *Dipterocarpus* genus was resolved into a monophyletic group, comprised *D. baudii*, *D. cornutus*, *D. palembanicus*, *D. kerii*, *D. insignis*, *D. alatus*, *D. zeylanicus* and *D. glandulosus*, supported by a bootstrap value of 99%.

The *Dryobalanops* genus was formed a monophyletic group comprised *D. aromatica* and *D. oblongifolia*supported with a bootstrap value of 99%.

The maximum likelihood method resulted in a tree similar to that of the maximum parsimony method (Appendix 6). This tree showed that several subclades composed of paraphyletic groups were formed, excluding section *Doona,* which formed a monophyletic group, with a high bootstrap value (93%). *Neobalanocarpus hemii,* whose place in the family is still debated, is placed on the sister branch with the *Hopea* group in this tree with a bootstrap value of 66%. *Shorea guiso* was nested with the *Hopea* group.

This maximum likelihood method also formed a paraphyletic group in one subclade of tribe Dipterocarpeae. In this subclade*, Vatica seychellarum* formed a sister branch with another subclade supported with a high bootstrap value (97%). The subclade consisted of members of the genus *Stemonoporus* (*Stemonoporus acuminatus, Stemonoporus gilimalensis, Stemonoporus wightii, Stemonoporus scalarinervis, Stemonoporus reticulatus, Stemonoporus kanneliyensis, Stemonoporus bullatus*), *U. borneensis,* which formed a sister branch with members of *Anisoptera* (*A. laevis*, *A. marginata*, *A. oblongata*), and *C. malayanum* and *C. scabriusculum,* which formed a sister branch with some *Vatica* members (*V. afinis*, *V. pauciflora*, *V. bella, V. odorata*, *V. coriacea*, *V. micrantha*).

An analysis based on the neighbor joining method of the *mat*K region showed a separation of the subclades, concurring with the results of maximum parsimony and maximum likelihood. A subclade formed a paraphyletic group with a bootstrap value of 53%. The first branch of this subclade belonged to *H. sangal* and *S. teysmaniana* supported with a high bootstrap value (96%); the second to the *Shorea* group (*S. assamica*, *S.virescens*, *S. brachteolata*,

S.hopeifolia, *S. stipularis*), supported by a value of 75%, and the third branch belonged to *Neobalanocarpus,* which was sister branch with the *Hopea* group (Appendix 7).

The *Dryobalanops* group was separated into a monophyletic group; *D. aromatica* and *D. oblongifolia* in one subclade with a high bootstrap value (99%) and formed sister branch with other subclades.

The member of genus *Dipterocarpus* showed a monophyletic group subclade comprising *D. insignis*, *D. zeylanicus*, *D. hispidus*, *D. glandulosus*, *D. alatus*, *D. cornutus*, *D. palembanicus*, *D. baudi* and *D. kerii*, with a high bootstrap support (99%). Other member of the *Dipterocarpeae* groups (*Vatica*, *Anisoptera*, *Upuna*, *Cotylelobium*) showed a similar topology as the one produced by the maximum likelihood method.

4.2.4 *rbc*L region

Using maximum parsimony and *M. kerstingii* as an outgroup; this gene region was able to resolve tribe Dipterocarpeae from tribe Shoreae, with a bootstrap value of 64%. However, the resolutions within the member of the tribe were not clear (Fig. 4.4). The members of the *Shorea* genus were still allied with *Hopea* and *Parashorea* genera. Our result showed that *P*arashorea *malaanonan* was grouped together with *S. contorta,* with a high bootstrap value (98%).

The evolutionary history was inferred using the maximum likelihood method. This method could resolve the *Dipterocarpeae* tribe from tribe Shoreae. The *rbc*L region was able to separate some members of *Shorea*, *Hopea*, and *Parashorea,* as well as members of *Dipterocarpeae* (*U. borneensis*, *A. marginata*, *V. machapagoi*) from other members of tribe Shoreae (*Shorea*, *Hopea* and *Parashorea*), with a low bootstrap value of 51% (Appendix 8). Using neighbor joining to infer the evolutionary relationship of Diperocarpoideae members produced the same results as the maximum parsimony and maximum likelihood methods. The neighbor joining method was also unable to resolve members of tribe Dipterocarpeae from members of tribe Shoreae (Appendix 9).

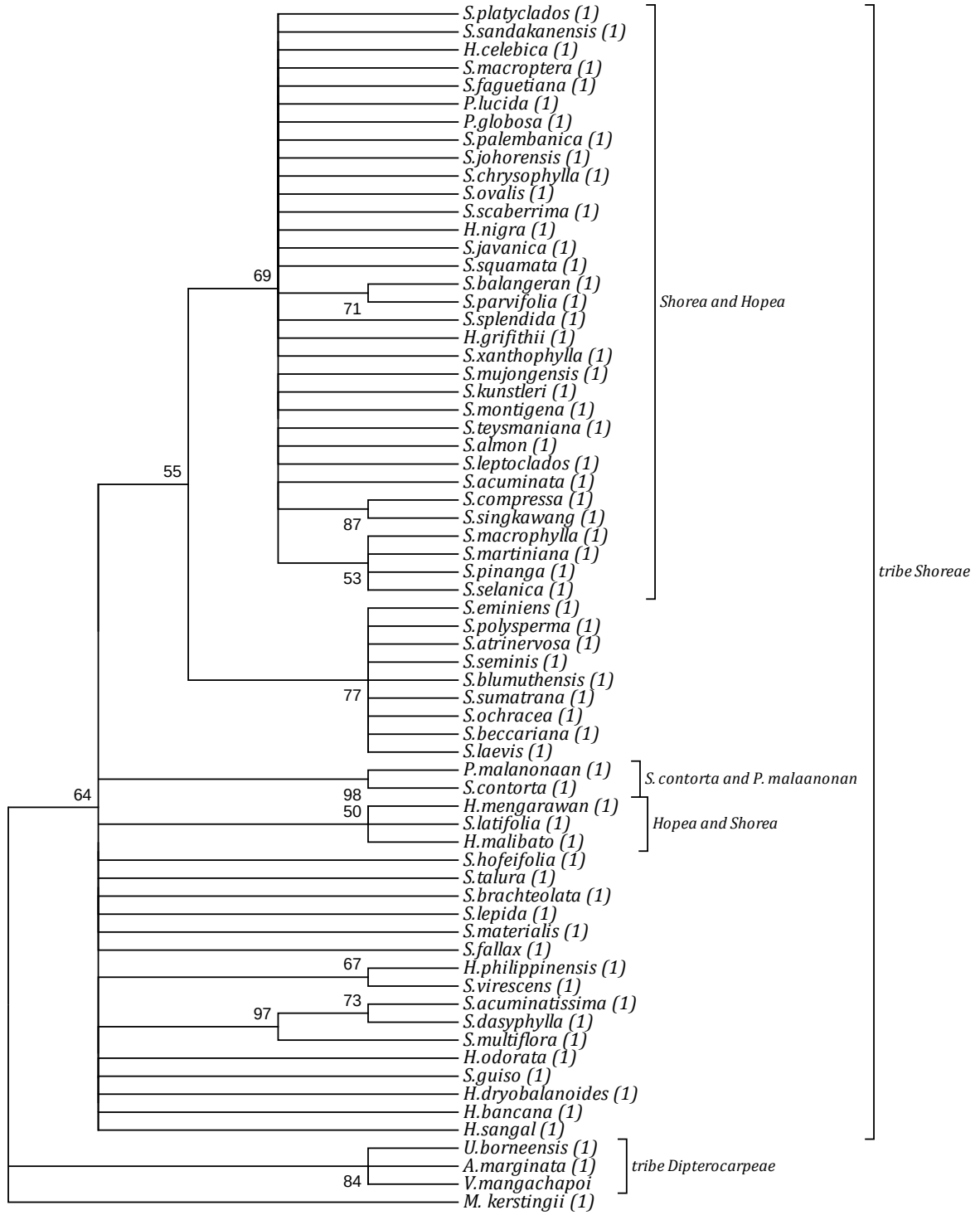


Fig 4.4 The tree of *rbc*L gene using the maximum parsimony method. The percentage of bootstrap support shown next to the branch. Tree type 1 out of 428 most parsimonious trees (length = 147) is shown. The number in parentheses mean number of species tested.

4.2.5 Combination dataset of *psb*C-*trn*S IGS, *trn*L intron, *matK* and *rbc*L

There were only 40 species with the same gene regions available for the phylogenetic analysis; the combined total length of all four gene regions was 2098 bp. The combined data comprises species from *Shorea*, *Parashorea* and *Hopea*. The strict consensus tree for the maximum likelihood, maximum parsimony and neighbor joining methods are shown in Fig. 4.5. All trees showed congruent patterns, only the neighbor joining tree showed a slight difference with regard to the number of *Shorea* members including in the second clade. The trees separated the species into two paraphyletic clades containing a mixture of *Shorea*, *Hopea* and *Parashorea*. The first clade was dominated by *Shorea* genera, with three genera belonging to *Hopea* (*H. celebica*, *H. nigra* and *H. grifithii*) and two to *Parashorea* (*P. lucida* and *P. malaanonan*). The second clade was dominated by members of *Hopea* (*H.odorata, H.bancana, H.philippinensis, H. dryobalanoides, H. malibato, H, mengarawan*), *with* Shorea making up the rest (*S. brachteolata, S. virescens, S. fallax, S. guiso, S. acuminatissima, S.multiflora*).

The topology patterns of the subclades were stable for all three statistical methods, with the exception of a small number of clades that differed slightly (Fig.4.5). There were 11 subclades, of which 10 had the same group pattern, for all three methods; they are labeled by Roman numbers (I-XI)

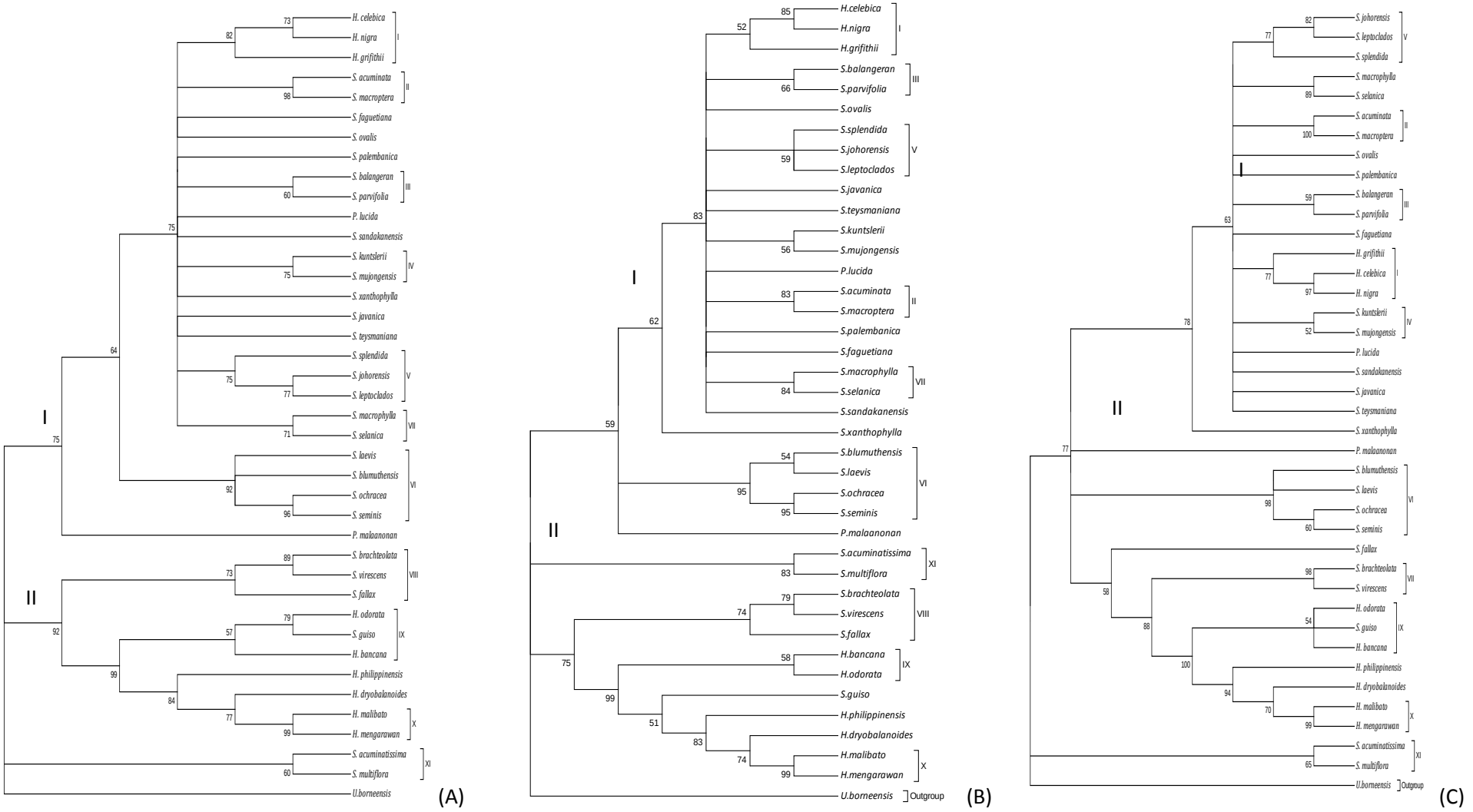


Fig 4.5 The combination dataset trees using different statistical method. Maximum likelihood (A), maximum parsimony (B) neighbor joining (C) trees from the combination dataset (2098 bp). Numbers above the nodes denote bootstrap support based on 1000 bootstrap replicates.

4.3 The barcode analysis for *mat*K and *rbc*L region

Both DNA regions were successfully utilized in obtaining Dipterocarpaceae data sequences. Within the Dipterocarpoideae tribe, 119 and 67 sequences were available for the *mat*K and *rbc*L regions, respectively. The neighbor joining trees for the different regions revealed both regions' abilities to distinguish tribe Dipterocarpeae with high bootstrap support. The *mat*K region showed a potential discriminatory power to distinguish genera and species within tribe Shoreae (Appendix 10), while *rbc*L could not resolve the genera of the tribe (Appendix 11). In our study, most of the *mat*K sequences analyzed in our laboratory (marked with X) allied with the corresponding sequences from the same species available in the NCBI database (appendix 10).

The combination of *mat*K and *rbc*L was not able to resolve the *Shorea*, *Hopea* and *Parashorea* genera (Fig. 4.6). Using *U. borneensis* as an outgroup, the gene region combination was able to separate the 40 species into two paraphyletic clades with strong bootstrap support (98%). The first clade was dominated by *Shorea* members, whereas the second clade was dominated by *Hopea* members.

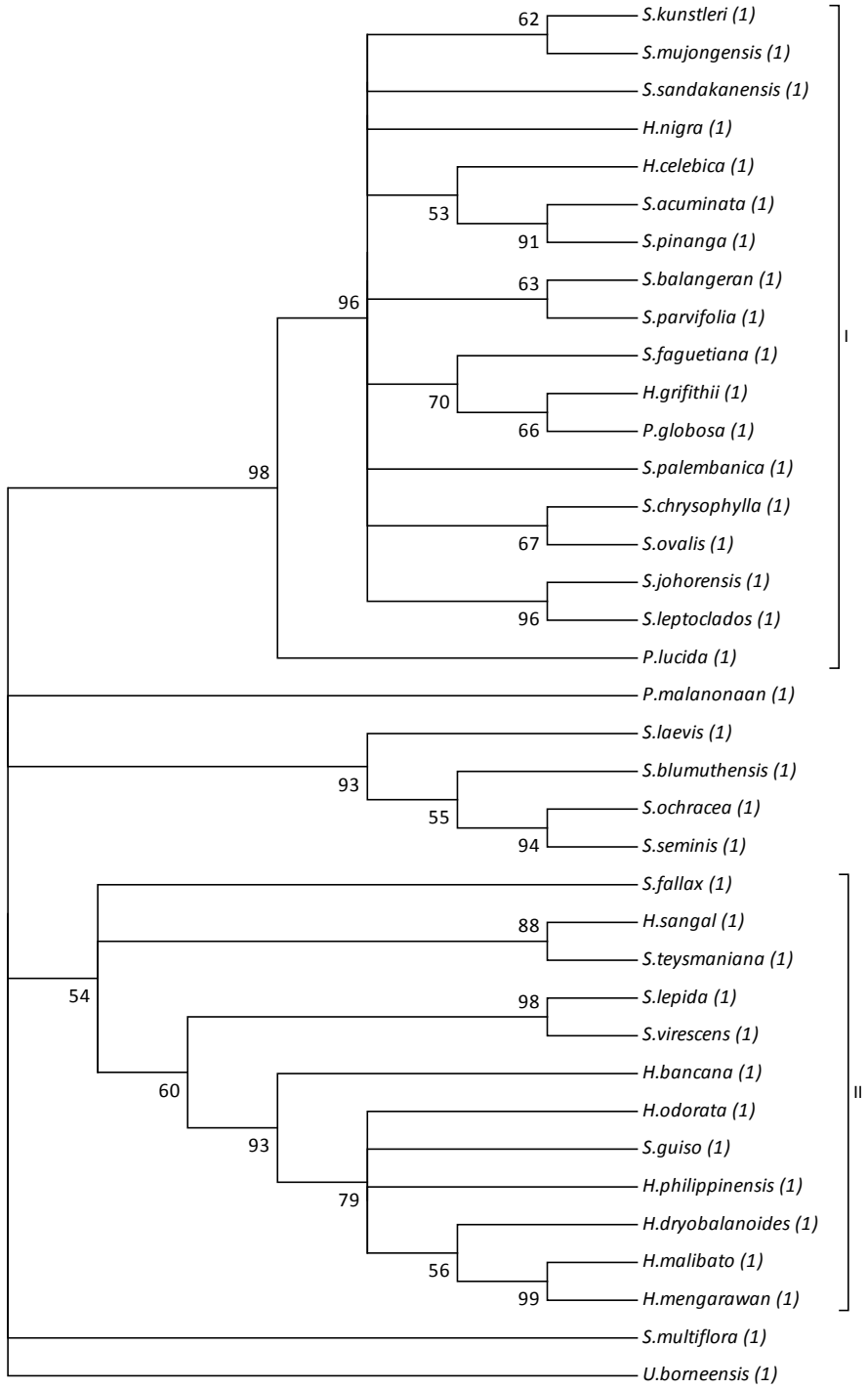


Fig 4.6 The neighbor joining tree of combined dataset *mat*K and *rbc*L (1282 nucleotides) using Kimura2 Parameter distance method. The number next to the branch is the bootstrap support test (1000 replicates). The tree was analyzed using MEGA 5.

The results of the nBLAST identification showed that of the 44 query sequences, 37 (84%) were successfully assigned to the correct genus. Regarding the individual species, however, nBLAST was rarely successful in finding the best match. This is likely due to the lower number of the corresponding species in the NCBI. In addition, almost all of the tested

sequences showed a low E-value (0.0). Only the best hits for *H. sanggal* showed an E-value higher than 0.0 (Table 4.3).

Some of the query sequences (seven out of 44) were matched with highly similar species from different genera. For instance, *H. celebica, H. nigra, H. grifithii* and *P. globosa* were all matched with with *S. smithiana* with 100% coverage and similarity. In addition, *P. lucida* was matched with *S. smithiana* with 100% coverage and 99 % similarity, respectively. *S. guiso* was matched with *H. wightiana* with 100% coverage and 99% similarity, and *P. malaanonan* was matched with *S. palescence* with 100% coverage and 99% similarity (Table 4.3).

Some of the best Megablast hits led to the same species showing unique results; for example, *S. ovalis,* S. *leprosula* and *S. acuminata* revealed a higher similarity to S. *smithiana* than to themselves, despite the coverage being 100% and similarity 100% or 99%. However, these species appeared at ranks lower than *S. smithiana* in the significant alignments'table (nBLAST result table). *Shorea smithiana* itself was matched with itself, although with a maximum identity of only 96%.

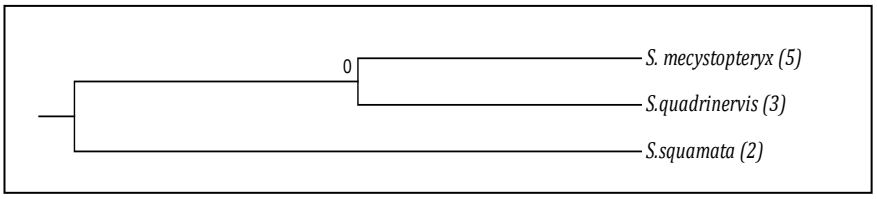
Table 4.3 The best match hits of *mat*K sequence samples from laboratory samples using nBlast and Megablast for highly similar sequences.

No	Investigated species	Best hit at NCBI database	Coverage (%)	Similarity (%)	E-value
1	*H. bancana*	*H. discolor*	100	99	0.0
2	*H. celebica*	*S. smithiana*	100	100	0.0
3	*H. dryobalanoides*	*H. discolor*	99	99	0.0
4	*H. grifithii*	*S. smithiana*	100	100	0.0
5	*H. malibato*	*H. latifolia*	100	99	0.0
6	*H. nigra*	*H. smithiana*	100	100	0.0
7	*H. odorata*	*H. wightiana*	100	99	0.0
8	*H. philippinensis*	*H. discolor*	100	99	0.0
9	*H. plagata*	*H. wightiana*	100	99	0.0
10	*H. sangal*	*H. jucunda subsp. modesta*	100	84	3e-167
11	*P. globosa*	*S. smithiana*	100	100	0.0
12	*P. lucida*	*S. smithiana*	100	99	0.0
13	*P. malaanonan*	*S. smithiana*	100	99	0.0
14	*S. acuminata*	*S. smithiana/S. acuminata*	100	100	0.0
15	*S. andulensis*	*S. smithiana*	100	100	0.0
16	*S. balangeran*	*S. smithiana*	99	99	0.0
17	*S. blumuthensis*	*S. palescens*	100	99	0.0
18	*S. chrysophylla*	*S. smithiana*	100	99	0.0
19	*S. faguetiana*	*S. smithiana*	100	100	0.0
20	*S. fallax*	*S. brachteolata*	99	99	0.0
21	*S. guiso*	*H. wightiana*	100	99	0.0
22	*S. hopeifolia*	*S. brachteolata*	100	99	0.0
23	*S. javanica*	*S. smithiana*	100	98	0.0
24	*S. johorensis*	*S. smithiana*	100	99	0.0
25	*S. laevis*	*S. pallescens*	100	99	0.0
26	*S. lepida*	*S. brachteolata*	100	99	0.0
27	*S. leprosula*	*S. smithiana/S.leprosula*	100	99	0.0
28	*S. leptoclados*	*S. smithiana*	100	99	0.0
29	*S. mecystopryx*	*S. smithiana*	100	99	0.0
30	*S. mujongensis*	*S. kuntslerii*	100	99	0.0
31	*S. multiflora*	*S. xanthophylla*	100	99	0.0
32	*S. ochracea*	*S. seminis*	100	99	0.0
33	*S. ovalis*	*S. smithiana/S.ovalis*	100	99	0.0
34	*S. palembanica*	*S. smithiana*	100	99	0.0
35	*S. pauciflora*	*S. smithiana*	100	99	0.0
36	*S. pinanga*	*S. smithiana*	100	100	0.0
37	*S. sandakanensis*	*S. splendens*	100	100	0.0
38	*S. selanica*	*S. smithiana*	100	99	0.0
39	*S. seminis*	*S. seminis*	100	99	0.0
40	*S. smithiana*	*S. smithiana*	100	96	0.0
41	*S. splendida*	*S. smithiana*	100	99	0.0
42	*S. teysmaniana*	*S. pinanga*	100	87	0.0
43	*S. virescens*	*S. brachteolata*	100	100	0.0
44	*S. xanthophylla*	*S. xanthophylla*	100	99	0.0

4.4 Sequence-based identification key using *trn*L intron as a model

In the maximum parsimony tree, using *M. kerstingii* as an outgroup and 145 species as an ingroup, 29 subclades were formed. Every subclade had a identical sequence.

Clade 1.



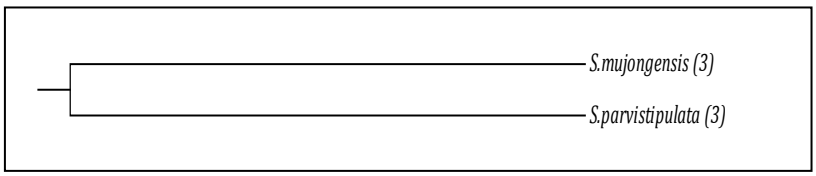
Species	Position of polymorphic site and its characters	
	52	212
S. squamata	-	G
S. mecystopteryx	A	G
S. quadrinervis	A	(R) A

1. A. Site 52 is (-) ……....2
   B. If (A) .....…….3
2. Site 212 is (G) ……...*S. squamata*
3. A. Site 212 is (G) ……..*S. mecystopteryx*
   B. if (A) .....……...……....*S. quadrinervis*

Clade 2

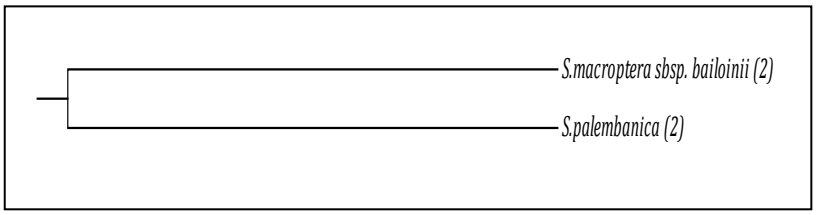
Only comprised one member of *Vatica* (*V. odorata)*. This species formed a polytomy with the other sequences.

Clade 3.



There was no polymorphism between *S. mujongensis* and *S. parvistipulata*. These two species showed identical sequences in the *trn*L intron gene region, therefore these two species cannot be distinguished using the *trn*L intron.

Clade 4.



Species	The polymorphic site and its nucleotide			
	244	246	275	276
S. macroptera sbsp. bailonii	-	G	A	C
S. palembanica	A	R	M	Y

1 a. site 244 is (-) …….....…….... 2a

b. if (A) 2b

2. a. site 246 is (G) …….....…….... 3a

b. if (A/R) 3b

3. a. site 275 is (A) .....……...…….... 4a

b. if (C/M) .....…….4b

4. a. site 276 is (T/Y) ……..…….... *S. palembanica*

b. If (C) .....*S. macroptera* subsp. *bailonii*

Not all the members of the clade could be identified using this key. If the compared species had identical sequences, this gene region was unable to identify them. Our results showed that there were many identical sequences included in one subclade, even though they belonged to different species. This was the case in the 3rd subclade (*S. mujongensis* and *S. parvistipulata*), 6th subclade (*S. javanica* and *S. pauciflora*), 11th subclade (*S. andulensis* and *S. leprosula*), 14th subclade (*S. paltycarpa* and *S. stenoptera*), 15th subclade (*S. palosapis* and *S. pubystila*), 16th subclade (*S. blumuthensis* and *S. havilandii)*, 17thsubclade (*S. balangeran*, *S. lepidota*, *S. rugosa*, *S. curtisii*), 18th subclade (*S. atrinervosa* and *S. inappendiculata*), 19th subclade (*S. flaviflora and S. platyclados)*,20th subclade (*S. falciferoides* and *S. polysperma*), 21st subclade (*S. falcifera* and *S. macroptera)*, 24th subclade (*S. colina* and *S. domatiosa)*, 25th subclade (*P. lucida*, *S. amplexicaulis*, *S. johorensis*, *S. smithiana*), 26th subclade (*S. argentifolia* and *S. rubra*), 27th subsubclade (a) (*S. patoiensis*, *S. Richetia* and *S. gibbosa, S. longiflora, S. maxima, S. faguetioides*), 27th subsubclade (b) (*S. albida* and *S. sandakanensis*), 27th subsubclade (c) (*S. latifolia*, *H. malibato*, *H. mengarawan*, *H. philippinensis*), 27th

subsubclade (d) (*S. henryana*, *S. lepida*) (*S. agami*, *S. confusa* and *S. symingtonii*). 29th (*D. baudii*, *D. Condorensis; D. grandiflorus, D. kerii, D. tempehes, D. Turbinatus; D. costatus, D. Haseltii*). The rest of the identification key is shown Appendix 12.

5 Discussion

5.1 The phylogeny of Dipterocarpoideae

Dipterocarpaceae species are known to be difficult to identify, especially when there is no flower available because of their infrequent flowering periods. Additionally, the species are also difficult to differentiate based on morphological characters because many species resemble each other (Symington, 1974).

The data presented here comprise a different number of sequences from a varying number of species from four DNA regions, depending on the sequences' availability in the NCBI and tissue samples in the laboratory. This means that the four regions are not directly comparable, but they do allow examination of the relationship between the groups of taxa in each DNA region.

Since the topology of the three statistical analyses (maximum parsimony, maximum likelihood and neighbor joining) for each region was generally congruent, with only small differences in bootstrap support, the discussion mostly refers to the maximum parsimony method.

5.1.1 Combination of four chloroplast DNA regions and overview of single regions

The combination dataset of four chloroplast regions was unable to clearly separate the members of *Shorea*, *Hopea* and *Parashorea*, reflecting the difficulties of using chloroplast DNA to classify the genera into one monophyletic group for each genus. The three statistical methods also formed a stable group within the tree but most of the grouping was paraphyletic. The combined trees produced in this study were representative of the ability of some of the chloroplast genes to resolve the phylogenetic relationship among *Shorea*, *Parashorea* and *Hopea* genera. This result is in accordance with previous results (Dayanandan et al., 1999; Kajita et al., 1998; Rath et al., 1998; Yulita et al., 2005). Thus, neither morphological nor molecular studies have been able to separate *Shorea*, *Hopea* and the putative genus *Parashorea* into three monophyletic groups.

Using *M. kerstingii* as an outgroup, our results showed that the *trn*L intron gene region is unlikely to reflect the evolutionary relationships of subfamily Dipterocarpoideae's members.

There are three assumptions for why a polytomy was formed in this tree: (1) recent speciation has occurred, as revealed by this gene region, (2) we don't have enough data to fully resolve the species and (3) there has been hybridization between closely related species. Since the reported data comprised GeneBank data and samples deposited in our laboratory, the sources of sampled species were heterogeneous. The species possibly showed polytomies as a result of internodes occurring in a short period of evolutionary time. The phylogenetic "bushes" in this tree might be because the *trn*L intron does not contain phylogenetically relevant information to resolve the tree, forming a bifurcating pattern (Humphries and Winker, 2010). The results of a previous study by Taberlet et al.(2007) agrees with this result.

The *psb*C-*trn*S IGS has shown the best delineation of the genus *Shorea* based on a classification by Symington (1943) and Ashton (1982). Since this region was mostly based on data for *Shorea*, *Hopea* and *Parashorea*, we could not assess the ability of this region to infer the complete phylogeny within tribe Dipterocarpeae. However, the parsimony tree of this region showed that *Anisoptera* and *Cotylelobium* genera formed an outgroup with *U. borneensis*, while the members of genus *Vatica* formed a sister branch with other subclades of tribe Shoreae.

The *mat*K gene region in this study was able to resolve tribe Dipterocarpeae and tribe Shoreae. Using *M. madagascariensis* as an outgroup, this gene was able to resolve two subclades of tribe Dipterocarpeae; one subclade was a monophyletic group of genus *Dipterocarpus*, and the other a paraphyletic group of other members of tribe Dipterocarpeae. This gene region also succeeded in placing section *Doona* of genus *Shorea* into a monophyletic group. Even though this gene region was reported by previous research to have more power to resolve phylogenetic relationships on the intra and inter species levels, the *mat*K gene did not show an ability to resolve the placement of *Shorea*, *Hopea* and *Parashorea,* with the exception of section *Doona* of the *Shorea* group.

Analysis of the *rbc*L data revealed that this gene region could not resolve the Dipterocarpaceae group above the generic level. This gene could separate tribes Dipterocarpeae and Shoreae effectively, but the separation within tribe Shoreae is still unclear using this gene. This result agreed with a previous result by Dayanandan et al. (1999), who succeeded in studying the affinity of the Dipterocarpaceae family to the Sarcolaenaceae family and allied with Malvales, but their study could not separate the tribes of Dipterocarpaceae.

The most important photosynthetic enzyme is encoded by the *rbc*L gene (Zurawski et al., 1981); this gene is extensively used as the first DNA sequenced from plants in plant phylogenetics studies. According to Vijayan and Shou (2010), the *rbc*L gene is the best characterized gene sequence among the plastid genes. However, most phylogenetic studies suggest that this gene is best suited to reconstruct the relationship down to the generic level but not the species level.

5.1.2 Phylogeny within tribe Dipterocarpeae

We could retrieve sequence data of tribe Dipterocarpeae for all of the studied DNA regions, but only the *trn*L intron and *mat*K gene regions provided extensive data compared with the *psb*C-*trn*S IGS and *rbc*L gene regions. The *trn*L intron sequence data of tribe Dipterocarpeae (*Anisoptera*, *Cotylelobium*, *Dipterocarpus* and *Vatica*) genera in our study were the same as the data used by Nguyen (2009).

Based on our result, the *trn*L intron region for the three methods showed the same pattern; all trees showed a low resolution and formed a polytomic clade. However, some subclades formed monophyletic groups. The *Dipterocarpus* group for all three methods showed a stable pattern that was distinct within the subclade. All *Dipterocarpus* species were grouped together; their affinity was supported by a high bootstrap value (99%). Nguyen's (2009) results also showed that this genus' members also formed a monophyletic group using nuclear genes (ITS1 and ITS2).

Because *Dipterocarpus* was only available for the *trn*L intron and *mat*K gene regions in our dataset, we were only able to observe this genus for those two DNA regions. This genus always formed a unified and distinct group that was separated from the other genera.

The generic relationships of Dipterocarpeae members using the *trn*L intron and *mat*K gene regions and revealed by the three statistical methods was mostly in accordance with previous results (Dayanandan et al., 1999; Gamage et al., 2006; Nguyen, 2009), which also found that *Dipterocarpus* always formed a monophyletic group, separated from the other members of tribe Dipterocarpeae and with high bootstrap support, indicating that this genus diverged earlier than other members of Dipterocarpeae. There were several indels found in all members of *Dipterocarpus* when aligned with another Dipterocarpeae member. These indels

might be important characters that resolved this group into one monophyletic group (data not shown).

Dipterocarpus may represent the basal lineage of Dipterocarpoideae (Meijer, 1979), and the family's name was taken based on this genus, probably because this genus was a primitive group among Dipterocarpaceae members (Maury - Lechon, 1979). This genus is well defined in the Dipterocarpaceae family, both in terms of morphological characters and molecular analyses. Through its morphological characters, *Dipterocarpus* can be identified by the large yellow anthers of its flower (2.5–8 cm across), with long appendages and columnar styles that are enclosed in large pink and white petals. There are generally two wing-like fruits. All *Dipterocarpus* species produce an oleo-resin called minyak keruing (Ashton, 1988; Symington, 1943).

The three statistical methods using the *trn*L intron showed that the members of Dipterocarpeae, excluding *Dipterocarpus,* showed an affiliation with members of tribe Shoreae. This gene region was unable to trace the evolutionary relationship among the studied taxa. The three statistical methods did not indicate that the *trn*L intron gene is a suitable region for studying the evolutionary relationship of dipterocarps. The low ability of the *trn*L intron region to resolve Dipterocarpaceae members was because of the lower intraspecific variation compared with the other noncoding regions of the chloroplast DNA (Shaw et al., 2005). Despite the fact that this region was easy to amplify, it doesn't represent the best choice either to delimit species or study the phylogenetic relationship among closely-related species (Taberlet et al., 2007a)

On the other hand, when using the *mat*K gene region, all the other members of tribe Dipterocarpeae, namely *Cotylelobium, Upuna, Anisoptera, Vatica, Vateriopsis* and *Stemonoporus* formed a paraphylethic group and sister clade with *Dipterocarpus.* This result revealed that the *mat*K gene region was better than the *trn*L intron in distinguishing the members of tribe Dipterocarpeae. The *mat*K gene region could provide a better depiction of the evolutionary relationships within *Dipterocarpeae*.

The results of our analyses agreed with those of Gamage et al. (2006), as well as the results of Nguyen (2009), who found that *Vateriopsis seychellarum* diverged earlier and formed a sister group with the other members of tribe Dipterocarpeae, excluding the *Dipterocarpus* branch. This species is more morphologically resemblant to *Dipterocarpus* than the other members of the Dipterocarpaceae family, however, in that it has a micropyle formed by both the inner and outer integument (Oginuma et al., 1999)

Our results also agreed with the results of Gamage et al. (2006) and Nguyen (2009), who placed the members of genus *Stemonoporus* in one monophyletic subclade. *Stemonoporus* is a well-known endemic genus in Sri Lanka. In addition, this genus diverged from the other members of the Dipterocarpeae tribe based on its morphological characters, specifically its peculiar anther with apical dehiscence and apical leaf traces, which separates from the central vascular cylinder well before the node (Ashton, 1982; Gamage et al., 2006; Kostermans, 1981).

On our *mat*K tree, genus *Cotylelobium* grouped together and formed a sister branch with a monophyletic group of genus *Vatica*. This was similar to the neighbor joining tree using the *psb*C-*trn*S IGS, which showed that *C. lanceolatum* was a root of the *Vatica* group. This result agreed with previous studies in placing *Cotylelobium* in a separate branch from *Vatica* (Cao et al., 2006; Dayanandan et al., 1999; Gamage et al., 2006; Nguyen, 2009). However, this result was contrary to Kosterman (1981), who placed *Cotylelobium* in a group with genus *Vatica* section *Sunaptea*. Our results supported the results of Indrioko et al. (2006), where genus *Cotylelobium* diverged earlier than all the members of tribe Dipterocarpeae but *Dipterocarpus*.

Our *mat*K tree results were similar to previous results by Parameswaran and Gottwald (1979), in placing *U. borneensis* in a sister branch with genus *Anisoptera* (*A. laevis* and *A. marginata*). Based on the morphological characters, genus *Upuna* was similar to *Anisoptera* and *Vatica* in medium-large solitary and partial multiple pores (120–150 μm), diffuse resin canals, thick-walled fibre and lack of SiO_2 (Parameswaran & Gottwald, 1979).

5.1.3 Phylogeny within tribe Shoreae

Our phylogenetic analyses showed similar topologies for all the trees with regard to the separation of tribe Dipterocarpacae and tribe Shoreae. Tribe Shoreae encompasses *Shorea, Hopea, Parashorea* and *Neobalanocarpus.* In this tribe, *Shorea* comprises 196 tree species found in lowland tropical forests in Southeast Asia. The placement of *Shorea* in our results is also in agreement with the classification proposed by Symington (1943), as well as the classification by Ashton (1982) (Fig 5.1). Symington classified *Shorea* based on the wood color (*Balau*, *Yellow Meranti, Red Meranti* and *White Meranti*) and treated *Pentacme* as a separate genus. Meanwhile, Ashton (1982) classified *Shorea* based on morphological characters, specifically the fruit calix, androecium and bark, separating the genus into 11

sections and treating *Doona* and *Pentacme* as two sections within the group. The sections of Ashton's classification were similar to those in Symington's classification. Sections *Doona*, *Pentacme* and *Anthoshorea* correspond to White Meranti, sections *Shorea*, *Pentacme* and *Neohopea* correspond to Balau and section *Richetioides* corresponds to Yellow Meranti, while Red Meranti belongs to sections *Ovalis*, *Rubella*, *Brachyptera*, *Pachycarpae*, and *Mutica*.

5.1.3.1 Placement of genus *Shorea*

The placement of *Shorea* species in our tree was revealed best using the *psb*C-*trn*S IGS gene region. The maximum parsimony tree of the *psb*C-*trn*S IGS was in accordance with the classifications of previous taxonomists (Ashton, 1982; Symington, 1974). The first clade was paraphyletic because some sections of the Red *Meranti* group of *Shorea* (sections *Brachyptera*, *Mutica*, *Ovalis*, *Pachycarpae* and *Rubella*) mixed with some members of section *Richetioides* (Fig. 4.1.a). Meanwhile, other subclades formed a monophyletic group based on the section, which corresponded to wood color.

None of the four studied DNA region trees succeeded in placing the Red Meranti group of *Shorea* into a monophyletic group, likely because the group is well known to have numerous species among other groups. Red Meranti species are mainly distributed in Sumatra, west Borneo and throughout the Malay Peninsula. The specific characteristics of this group are large, stoutly-buttressed trees, and red, pink, reddish-brown or orange-brown inner bark (Symington,1943). Yulita et. al., (2005), using *trn*L-*trn*F and ITS regions, could not resolve *Shorea* into a monophyletic group separated from genus *Hopea*, and suggested that the *Hopea* group may have originated from *Shorea*.

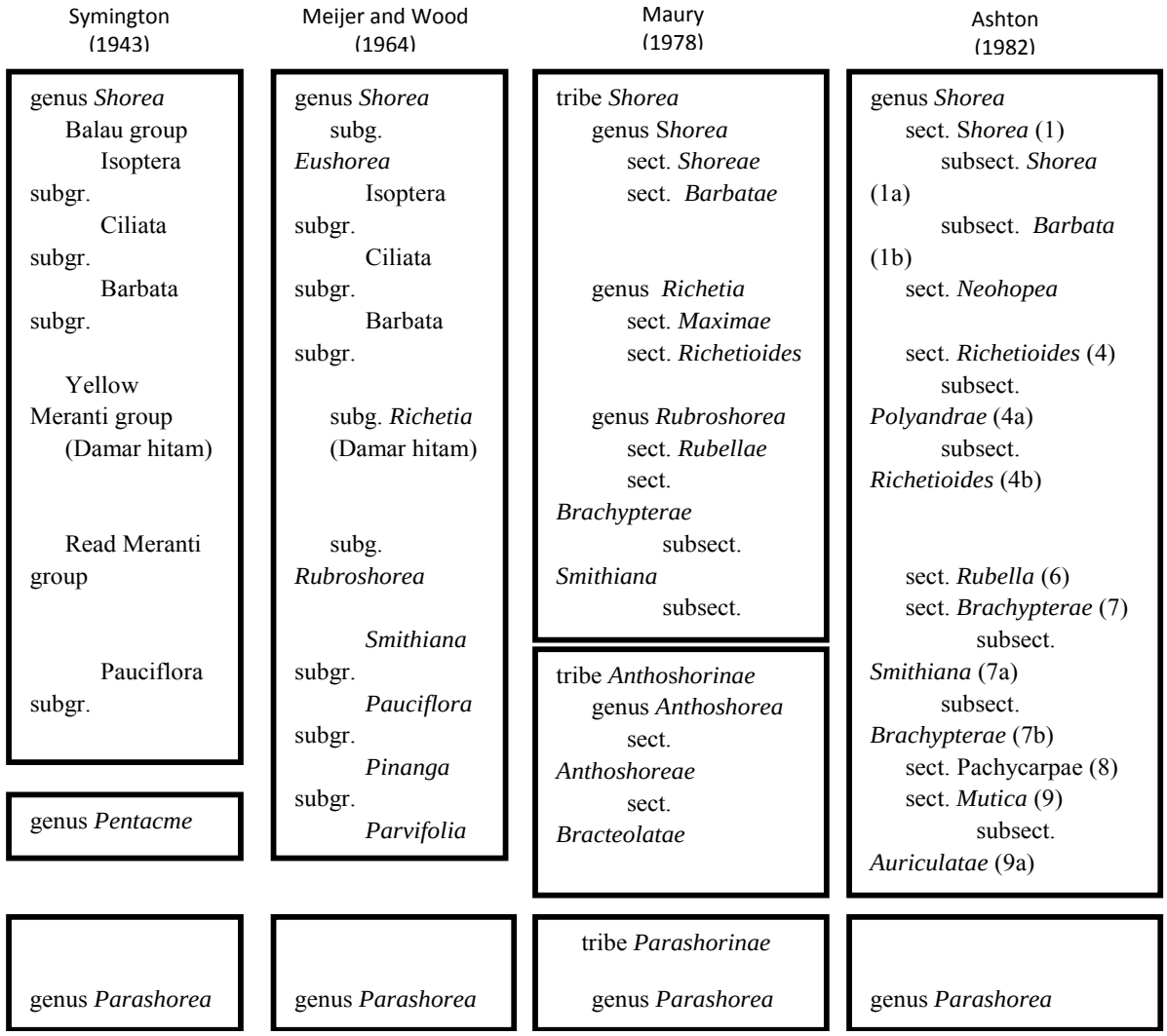


Fig 5.1 Comparison of classification of Shorea and its closely-related genera adopted from Kamiya et al. (2005)

Our results are similar to the results of Tsumura et al. (2011), who used the combination of four chloroplast DNA regions (*trn*L gene, *trn*L-*trn*F IGS, *trn*H-*psb*A-*trn*K and *psb*C-*trn*S IGS) and succeeded in separating *Shorea* in a similar manner as phylogenies based on the wood color, with the exception of the White Meranti group, which formed an affiliation with the members of genus *Hopea*. In the *psb*C-*trn*S IGS tree, using maximum parsimony analysis, the subclade of *Anthoshorea* was resolved as a monophyletic group, supported by a moderate bootstrap value (50%). According to Symington (1943), based on the production of yellow pale dammar resin, the White Meranti is similar to damar mata kucing, which is produced by some *Hopea* species.

The *psb*C-*trn*S IGS succeeded in resolving section *Richetioides*, which belongs to the Yellow Meranti group of *Shorea* as a monophyletic subclade (C in Fig. 4.1), with strong bootstrap support (100%). The monophyly of this section agreed with the results of Kamiya et al.

(2005), who also found that section *Richetioides* formed a monophyletic group *when* using the *Pgi*C gene region (Kamiya et al., 2005). Yulita et al., (2005) also reported similar results when using the *trn*L-*trn*F IGS region. In addition, besides the *psb*C-*trn*S IGS, the *trn*L intron tree using maximum parsimony analysis also showed a monophyletic group of section *Richetioides,* corresponding to the Yellow Meranti group of *Shorea* (*S. faguetioides*, *S. multiflora*, *S.peltata*, *S. longisperma*, *S. gibbosa*, *S. acuminatissima*, *S. richetia*, *S. maxima*, *S. longiflora*, *S. patoiensis*). The members of section *Richetioides* are known in the market as Meranti Damar Hitam. This section is known to have characters that are dissimilar to those of other sections, including subequal calyx lobes, anthers, wood and bark anatomy and dark brown or black dammar exudation (Ashton, 1982; Symington, 1943).

The subclade of section *Shorea* using the *psb*C-*trn*S IGS and the maximum parsimony method (D in Fig. 4.1) agreed with the results of Yulita et al. (2005), who used the *trn*L-*trn*F gene regions, as well as the ITS gene regions, and found that section *Shorea* (*Balau* group) did not form a monophyletic group, but a paraphyletic one, because of the inclusion of *S. isoptera* from section *Neohopea* and some sections of the *Shorea* genus (*Brachypterae*, *Mutica* and *Anthoshorea*). However, in our result, the subclade of section *Shorea* was not allied with the members of section *Brachypterae* or other sections from the Red Meranti group of *Shorea,* instead allying with the members of section *Richetioides* (*S. blumuthensis* and *S. polysperma*).

Using the three statistical methods with the *mat*K gene region resulted in the successful formation of section *Doona* (*S. affinis*, *S. zeylanica*, *S. cordifolia*, *S. gardneri*, *S. worthingtonii*, *S. trapezifolia*, *S. congestiflora*, *S. disticha*, *S. megistophylla*) into a single monophyletic group, with a high bootstrap value (91%). Among all the sections of the *Shorea* genus, section *Doona* is one of the easiest to characterize. This section consists of ten species, most of them endemic to Sri Lanka. Morphological studies, using *Doona*'s distinct characters, recommend that the section might be grouped in an own, separate genus (Maury - Lechon, 1979). However, Ashton (1982) suggested that it should be classified as a section of the *Shorea* genus.

5.1.3.2 Placement of genus *Parashorea*

The phylogenetic tree of the *psb*C-*trn*S IGS using maximum parsimoy showed that *Parashorea malaanonan* formed a sister subclade with all the other subclades of *Shorea* and *Hopea*, and was grouped with *S. contorta,* with a strong bootstrap value (100%). One member of *Parashorea* (*P. lucida*) in this tree allied with the Red Meranti group in the first clade (Fig. 4.1.a). The close relationship between *P. malaanonan* and *S. contorta* was also evident in the *rbc*L gene analysis. Based on our phylogenetic analyses of the *rbc*L gene region, the three statistical methods showed that *P. malaanonan* and *S. contorta* always grouped together with high bootstrap support (>97%). Meanwhile, other members of *Parashorea* (*P. lucida* and *P. globosa*) formed a sister branch with other *Shorea* members.

According to Parameswaran and Gottwald (1979), genus *Parashorea* is closely-related to section *Pentacme* because of their similarities in wood anatomy, solitary and multiple vessels, apotracheal and paratracheal parenchymes, calcium oxalate crystals in rays and resin canals in tangential rows. Some *Parashorea* members are also morphologically similar to the members of Red Meranti of Shorea (Symington, 1943), while Dayanandan et.al (1999) suggested that this genus is close to the section *Anthoshorea* and *Richetioides* are therefore also believed to be closely-related to genus *Shorea.* According to Ashton (1982), *S. contorta* is member of section *Pentacme*. *Parashorea malaanonan* and *Shorea contorta* are grouped together; both species have a wide distribution throughout the Philippines (Ashton, 2004), while *P. lucida* and *P. globosa*, which are nested with the Red Meranti group, are found only on the islands of Sumatra and Borneo.

The *mat*K gene using maximum parsimony analysis revealed that the placement of *Parashorea chinensis* was unlike that in the results of Li et al. (2004). *Parashorea chinensis* and *Parashorea chinensis* var. *kwangsiensis*, using the *mat*K gene, *trn*L-*trn*F IGS and *trn*L intron, were affiliated with *Parashorea lucida* and sistered with *S. macroptera* and *S. ovalis* (Li et al., 2004). According to Symington (1943), *Parashorea* resembles the members of the Red Meranti group of *Shorea* in leaf characteristics, such as glausescence, particularly in young leaves, needle-like leaves and older seedlings with white subpeltate leaves on the undersurface.

There are several of *Parashorea* generic characters that don't belong to *P. chinensis*. The nut is ovoid (not cylindrical), and the usually prominent pale lenticels are obscured by the tomentum. Additionally, the leaves are not folded, nor are seedling leaves peltate or silvery

on the underside, even though the flower is similar of those of the other species of *Parashorea* (Li et al., 2004).

5.1.3.3 Placement *of* genera *Hopea* and *Neobalanocarpus*

Our parsimony tree of the *psb*C-*trn*S IGS region showed that the genus *Hopea* in the second clade formed a paraphyletic subclade, because of the inclusion of *S. astylosa* and *N. hemii*. None of the chloroplast regions in this study could separate *Hopea* and *Shorea* into monophyletic groups. The affiliations of *Shorea* and *Hopea* genera in all of the phylogenetic trees were similar to those produced by Yulita (2005), who utilized a different chloroplast region (*trn*L-*trn*F IGS). Both Kamiya (2005), who used the *Pgi*C gene, and Yulita (2005), who used the ITS gene, found a monophyletic group of *Hopea*, with the exception of *H. celebica* in Yulita's analysis, which was nested with *Shorea* group. These results suggest that nuclear genes are more effective than chloroplast genes when classifying the *Hopea* group. According to Ashton (1988) morphological characteristics of *Hopea* can distinguish it from *Shorea*, such as the number of long fruit calyxes and height of their members.

Our *psb*C-*trn*S IGS tree showed that *Neobalanocarpus hemii* formed a sister branch with the *Hopea* group, agreeing with the classical taxonomic work of Ashton (Ashton, 1982) and previous results from Gamage et al. (2006), Tsumura et al. (1996) and Yulita et al. (2005), but contrary to a previous result from Kamiya (Kamiya et al., 2005), who had used part of the *Pgi*C gene and found that *Neobalanocarpus* was nested in section *Anthoshorea* of the *Shorea* group. Kamiya et al. (2005) assumed that the origin of *Neobalanocarpus* is the result of hybridization between *Hopea* and White Meranti of *Shorea*, with the former as the maternal progenitor and the latter as paternal. However, our maximum parsimony tree of the *mat*K gene showed a paraphyletic subclade in which *N. hemii* was nested together with *Hopea*, and *Shorea* section *Anthoshorea* (*S. brachteolata* and *S. virescens*), *Richetioides* (*S. hopeifolia*), and section *Balau* (*S. guiso*). These results cannot fully support the assumtion of Kamiya et al. (2005), whose hypothesis could be acceptable if the paternal progenitor was not restricted to the White Meranti group of *Shorea*, but the whole genus in general.

According to the argumentation of Parameswaran and Gottwald (Parameswaran & Gottwald, 1979) and Ashton (1982) which is based on its morphological characters, *Neobalanocarpus* is more closely-related to the *Hopea* group than the *Shorea* group, owing to the similarities in inflourescence features, embryo structure and germination modes, as well as leaves and wood

anatomy. Asthon (Ashton, 1982) has strongly suggested that *N. heimii* is closely-related to the genus *Hopea*, and his suggestion has been supported by subsequent research by Kajita et al., (1998).

5.1.3.4 Placement of genus *Dryobalanops*

Our parsimony tree of the *mat*K and *trn*L intron gene regions showed different result in the placement of the position of genus *Dryobalanops*. The *mat*K gene showed an affinity of this genus to genus *Dipterocarpus* and formed sister subclade with the outgroup (*Monotes madagascariensis*). This result is in agreement with the previous result of Indrioko (2005) using chloroplast microsatelite analyses and Yulita et al. (2005) using the *trn*L-*trn*F region that genus *Dryobalanops* forms a sister taxon with the *Shorea*, *Hopea* and *Parashorea* genera. Meanwhile, the three methods of the *trn*L intron statistical analysis showed that this genus nested with the *Shorea* group. The maximum likelihood tree of the *trn*L intron showed that *Dryobalanops oblongifolia* formed a sister branch with *S. seminis*. This result was similar with the result of ITS analysis of Yulita et.al (2005). In her result using that nuclear gene region, the genus *Dryobalanops* was placed within the genus *Shorea* clade.

According to Indrioko (2005) the genus *Dryobalanops* is morphologically similar to tribe Dipterocarpeae; he suggested that this genus is a basal of the Shoreae tribe. However, according to Symington (1943), the genus *Dryobalanops* is similar with genus *Dipterocarpus*, section Balau of the *Shorea* group and some species of the *Hopea* genus in having a scaly bark. Our result using the *mat*K gene agree to place genus *Dryobalanops* close to genus *Dipterocarpus*, while our *trn*L intron tree showed a relationship of this genus to the member of section Balau (*S. seminis*) which can support the assumption of Indrioko (2005) that this genus is the basal lineage of tribe Shoreae.

5.2 Utilizing two DNA barcode regions (*mat*K and *rbc*L) for dipterocarps

Using the *rbc*L gene and the maximum parsimony method, the tribes Shoreae and Dipterocarpeae were successfully separated. However, the ability of this gene to distinguish the members of Shoreae was low (Appendix 11) because the *rbc*L gene does not have sufficient variation at the species level. Although the *rbc*L gene has been used extensively in familial level phylogenetic studies (Gielly and Taberlet, 1994), it is reported to evolve slowly (Soltis et al., 1998). Our result agreed with previous research of the *Dioscorea* genus by Sun et al. (2012), in finding that the *rbc*L gene was not capable of discriminating species but genera and above.

Thus, to be used as a DNA barcode, it appears that the *rbc*L gene region cannot work alone. This gene region should be combined with gene regions to improve its discriminatory power.

Amplification of the *mat*K gene using the universal *mat*K primer proposed by Kim (Hollingsworth, 2011) was difficult. Additionally, compared with the *rbc*L gene, the sequencing results were of low quality. These results were similar to previous results regarding different taxa (de Vere et al., 2012; Hollingsworth et al., 2011; Yu et al., 2011).

Our results showed that using nBLAST in concert with the *mat*K gene can lead to an identification of the correct genus (84%). However, nBLAST's results can be misleading on the species level, possibly because the sequences deposited in NCBI that corresponded to our samples were limited. Additionally, the *mat*K region was conserved which was showen by the of E-values for almost all the species. The lower the E-value, the more similar the query sequence to the hit sequences in the database (Madden, 2002). Even though according to Olmstead and Palmer (1994) *mat*K is the most variable coding region among cpDNA, our results showed that this region is very conserved for Dipterocarp species. BOLD is thus required because it will provide a more reliable database than the NCBI for DNA barcoding. Little and Stevenson (2007) have suggested that using a reference database in which virtually all haplotypes in all species are represented will provide the most reliable identification.

The neighbor joining trees revealed that some of the *mat*K sequences analyzed in our laboratory allied with the corresponding sequences from the same species available in the NCBI database. The ability of most of the sequences from the laboratory analyses to group together based on the same genus revealed that this gene region is able to discriminate sequences above specific level. However, the neighbor joining tree of this region showed

many polytomies, indicating that this gene region is not able to effectively trace the evolutionary relationships of species.

The combination of the two DNA regions (*mat*K and *rbc*L) was not able to distinguish the *Shorea*, *Hopea* and *Parashorea* genera as one monophyletic group (Fig. 4.6). This combination of regions was able to separate the 40 species into two paraphyletic clades with strong bootstrap support (98%). The first clade was dominated by *Shorea* members, whereas the second clade was dominated by *Hopea* members. The difficulties of resolving *Shorea*, *Hopea* and *Parashorea* are because all three generas' members are closely-related species (Kress et al., 2009; Yulita et al., 2005). The difficulties in using these two regions as barcode regions for closely-related species were also revealed in a previous research (Zhang et al., 2012), who analyzed *Lysimachia* L. (Myrcinaceae family), and found the impossibility of using the *rbc*L and *mat*K gene regions together as barcode regions to distinguish closely-related species in Myrcinaceae. For dipterocarps, even though the *mat*K gene showed a moderate discriminatory power, many polytomies were formed in the resulting tree, suggesting its low ability to reveal the phylogeny of dipterocarps.

5.3 Sequence-based identification key

The presented results show that we were unable to produce a reliable identification key for identifying members of the Dipterocarpaceae family. The weakness of this key is that many identical sequences belong to different species; that is, multiple occasions were encountered in which one haplotype would belong to different species. The investigated gene regions were not suitable for distinguishing species, particularly closely-related ones. Other gene regions may be more suitable for dipterocarp species identification. Further research on the possibility of developing a molecular taxonomic identification key based on phylogenetic analyses is needed.

6 Conclusion and Outlook

The Dipterocarpaceae family dominates the lowland forests of Southeast Asia. It is divided into three subfamilies: Dipterocarpoideae, Pakaraimoideae and Monotoideae. Most of the genera in this family belong to species that produce valuable timber. Subfamily Dipterocarpoideae is the largest group and, based on the basic chromosome number, divided into two tribes, Shoreae (X=7) and Dipterocarpeae (x=11), with genus *Shorea* containing the highest number of species.

The evolutionary relationship between the members of subfamily Dipterocarpoideae was inferred using four chloroplast regions: *trn*L intron, *psb*C-*trn*S IGS, *mat*K and *rbc*L. This study also aims to evaluate DNA-based identification using DNA barcoding and a molecular taxonomic identification key.

The phylogenetic analysis using the four chloroplast regions and three statistical methods (maximum parsimony, maximum likelihood and neighbor joining) resulted in successful placement and revealing of the relationship between Dipterocarpoideae's members. None of the four chloroplast regions showed a single DNA region as suitable to delineate the evolutionary relationships of dipterocarps, with every chloroplast region having its own specifities.

The *trn*L intron region was easy to amplify; it is the most-used region to infer the evolutionary relationship among plant species (Taberlet et al., 2007b; Zhou et al., 2008). However, this region was only able to resolve the taxa up to the generic level, separating tribes Dipterocarpeae and Shoreae most effectively using the maximum likelihood and maximum parsimony methods, although maximum parsimony could only clearly distinguish the *Dipterocarpus* genus into one monophyletic group.

The suitability of the *psb*C-*trn*S IGS region for tracing the evolutionary relationships between plants is controversial because of the limited research on this region. When combined with other chloroplast regions, however, this region succeeded in resolving the *Shorea* genus in agreement with phylogenies based on the wood color (Tsumura et al., 2011). Our results showed that this region works well in distinguishing species based on wood color and separated them into several monophyletic groups.

The *mat*K gene region is recommended in many phylogenetic studies because of its ability to resolve phylogenetic relationships at the intra and interspecific level. In our study, the *mat*K

gene was the best at revealing the evolutionary relationship of the members of Dipterocarpeae and could distinguish section *Doona* in tribe Shoreae, placing the members in a monophyletic group. However, this region could not reveal a clear distinction of Shoreae, because this tribe formed a paraphyletic group in which the members of *Shorea*, *Hopea*, *Parashorea* and *Neobalanocarpus* allied together in the clades.

The *rbc*L gene region was similar to the *trn*L intron in its ability to amplify easily and provide a satisfactory sequencing product. However, this gene region did not show an ability to infer the evolutionary relationships within tribe Shoreae.

The placement of the members of Shoreae was generally unclear in this study. Genus *Shorea* was paraphyletic because three other genera, *Hopea*, *Neobalanocarpus* and *Parashorea* were nested with it. However, the classification within *Shorea* could be revealed using the *psb*C-*trn*S IGS region and *mat*K gene region, since some of the subclades formed a monophyletic group based on the section, which corresponded to wood color. In this study, the *psb*C-*trn*S IGS placed *Neobalanocarpus* as a sister branch with *Hopea*, while the *mat*K gene tree showed this genus' affinity with *Shorea* (sections *Anthoshorea*, *Richetioides* and *Balau*) and *Hopea*. Our study showed that using the *mat*K gene resulted in genus *Dryobalanops* showing an affinity with genus *Dipterocarpus*, while the *trn*L intron tree showed that this genus is close to section *Balau* of the Shorea group

As the four chloroplast regions used in this study could not reveal unambiguous evolutionary relationships, particularly in tribe Shoreae, it is recommended that nuclear genes should be analyzed in a future study. In addition, it is also recommended that the status of the genera *Shorea*, *Hopea* and *Parashorea* should be revised because of their strong affinity in each investigated chloroplast region.

The *mat*K and *rbc*L regions were tested for their suitability as barcode DNA. Our study showed that the *mat*K gene region was difficult to amplify and showed a lesser discriminatory power at the species level particularly for tribe Shoreae. The *rbc*L gene was easy to amplify, while failing to provide enough information to discriminate until the species level. Both of these regions were only partially suitable to clarify the phylogeny of dipterocarps and to reliably identify species, possibly because closely-related species have many constraints that prevent them from being easily distinguished. These gene regions might be of use as barcode DNA for distant relatives if the *mat*K gene can be amplified.

It is suggested that another chloroplast region, *trn*H-*psb*A, should also be analyzed, as recommended by CBOL. Moreover, the nuclear gene ITS2 should also be tested as a barcode region, even though until now this region is recommended only for the fungi group. Since it is difficult to find a single universal barcode region for all land plants, I suggest that taxon-specific barcode regions are used instead.

This study could not provide a universal molecular taxonomic identification key for dipterocarps. Several haplotype sequences could not be unambiguously assigned to a single species.

Because this key aims to complement DNA barcoding analyses, besides applying nBLAST and using the results of phylogenetic tree analyses, I suggest that future studies develop a key from the phylogenetic tree of a barcode gene region. This phylogenetic tree will be a standard tree for each family, comprising as many members of the family as possible. I also recommend that a digital key is developed instead of a paper-based one to facilitate an easy way of species discrimination and identification.

7 Summary

Dipterocarpaceae is the main timber family of tropical forest trees in the Malesian region with a geographical distribution that extends to South America and Africa. The family comprises approximately 500 species in 17 genera and is subdivided into three subfamilies: Dipterocarpoideae, Monotoideae and Pakaraimoideae (Ashton, 1982). Dipterocarpoideae is the richest in species with a total of 470 species in 13 genera (Ashton, 1982). Dipterocarpoideae is divided into two tribes: Dipterocarpeae and Shoreae. The genera of Dipterocarpeae are *Anisoptera*, *Cotylelobium*, *Dipterocarpus*, *Stemonoporus*, *Upuna*, *Vateria* and *Vateriopsis*, while those of Shoreae are *Dryobalanops*, *Hopea*, *Neobalanocarpus*, *Parashorea* and *Shorea*. *Shorea* and *Hopea* contain most species; 169 in the former and 100 in the latter.

Molecular phylogenies of the subfamily Dipterocarpoideae have been studied since 1998, especially the genus *Shorea* and its sister genera in tribe Shoreae, because this genus has the highest number of species and the most valuable timber of the Dipterocarpaceae. Many of these species are endangered. The purpose of molecular phylogenies is to complement phylogenies based on morphology as there is still a debate on the placement of some genera in the tribe Dipterocarpoideae. The classification of *Shorea* in this research refers to Ashton (1982) and Symington (1943). Symington (1943) has divided *Shorea* based on wood color (White Meranti, Yellow Meranti, Balau and Red Meranti). Asthon (1982) has generally retained the classification by Symington (1943), but some of the groups were reclassified into lower taxonomical ranks.

The need for identification tools for Dipterocarpacae in order to avoid fraud in certifying the family's timber has led to an improvement in modern identification systems that use molecular data. Traditionally, Dipterocarpaceae are recognized based on their morphological characters, but sometimes these characters have constraints, particularly in the absence of a flower, the most useful taxonomic identification character for the dipterocarps.

The abundance of molecular data as well as advanced technologies in DNA sequencing have made DNA barcoding a widely-used practice in many different fields of taxonomic studies, not as a replacement but as a complement to traditional taxonomy and to accelerate the identification process. Another advantage of the large number of sequences available in public data bases as the NCBI database is that it can lead to a new concept of species

identification through the development of a molecular taxonomic key. However, since DNA barcoding methods are still in their infancy, the database for DNA barcoding is still being established.

This study aims to infer the phylogenetic relationships of the members of the subfamily Dipterocarpoideae and to study the placement of the genera based on four chloroplast regions (*trn*L intron, *psb*C-*trn*S IGS, *mat*K and *rbc*L). Furthermore, the suitability of the two barcoding regions (*mat*K and *rbc*L) will be evaluated, which were proposed by the Consortium for the Barcode of Life (CBOL) in 2009. This study also aims to develop a taxonomic identification key based on the phylogenetic analysis for species identification purposes.

Dipterocarpacae sequences that were deposited in the NCBI database were retrieved for four chloroplast regions (*trn*L intron, *psb*C-*trn*S IGS, *mat*K and *rbc*L). In addition to the analysis of the sequences from the NCBI database, we also sequenced samples of dipterocarps available at the section Forest Genetics and Forest Tree Breeding, Georg-August-University Göttingen, at the four chloroplast regions in order to analyze the highest possible number of species.

The phylogenetic analysis was done using MEGA 5 software and the statistical methods of maximum parsimony (MP), maximum likelihood (ML) and neighbor joining (NJ). For the DNA-based identification analyses, we evaluated the suitability of the two barcode regions using nBLAST, and performed the phylogenetic analysis using the neighbor joining method.

Our results succeeded in obtaining sequences for various numbers of species for each studied chloroplast region, namely 145 species for the *trn*L intron, 117 species for the *psb*C-*trn*S IGS, 116 species for the *mat*K region and 67 species for the *rbc*L region. The final length of the sequences varied for each region, 537 bp, 1136 bp, 653 bp and 647 bp for the *trn*L intron, *psb*C-*trn*S IGS, *mat*K and *rbc*L, respectively.

For the phylogenetic analyses, MP, ML and NJ analyses of cpDNA sequences produced similar tree topologies. As a result, our discussion is mostly based on the results of the MP analysis. Generally, the evolutionary relationships within the subfamily Dipterocarpoideae could not be clearly revealed by the four chloroplast regions. The regions were able to resolve the tribes Dipterocarpeae and Shoreae, but were less successful within the tribes, particularly Shoreae. For the genus *Dipterocarpus*, recent studies only provide sequence data for the regions *trn*L intron and *mat*K. We observed two distinct groups comprising species of this

genus for both gene regions. There is an assumption that *Dipterocarpus* may represent the basal clade of Dipterocarpoideae (Meijer, 1979). The name of this family was taken based on this genus, possibly because it is regarded as a primitive group among Dipterocarpaceae's members (Maury - Lechon, 1979). This genus is well defined in the Dipterocarpaceae family based on morphological characters and molecular analyses.

The *psb*C-*trn*S IGS region in this study agreed with previous research by Symington (1943) in its ability to form a monophyletic group based on wood color in the genus *Shorea*. The *mat*K region showed the best ability to delineate the relationships of the tribe Dipterocarpeae and succeeded in distinguishing section *Doona* of *Shorea* as a monophyletic group. However, this region failed to work as well in classifying other members of Shoreae. Despite the *rbc*L region's status as the first DNA region to be sequenced from a chloroplast region, there are few *rbcL* sequences available for dipterocarps in the NCBI database. The results based on the data from the laboratory showed that this region was unable to trace the evolutionary relationship of Dipterocarpoideae below the generic level. The *mat*K region in this study showed that the genus *Dryobalanops* has an affinity with genus *Dipterocarpus*, while the *trn*L intron tree showed that *Dryobalanops* is close to section *Balau* of the *Shorea* group. These contradictory results support the assumption of Indrioko (2005) that this genus is a basal clade of tribe Shoreae.

The DNA-based identification was studied using two approaches, namely DNA barcoding and a molecular taxonomic identification key. The two DNA barcode regions, *mat*K and *rbc*L, adopted from the Consortium for the Barcode of Life for land plants (Hollingsworth et al., 2009), were applied to assess the feasibility of these regions as barcodes to discriminate the Dipterocarpaceae. Most information for the *mat*K region was available in the NCBI database, but additional samples were also included in this study. In total, 119 and 67 samples were studied using the *mat*K and *rbc*L regions, respectively. The effectiveness of the barcode analysis in this study was assessed by the formation of monophyletic groups of the query sequences and the reference sequences which are deposited in NCBI using neighbor joining trees and then searching for the similarity of the query sequences from the laboratory against the available data in the NCBI database using nBLAST. Although the neighbor joining tree placed some of the sequences in the correct genus, this region could not clearly separate the genera *Shorea*, *Hopea* and *Parashorea* into one distinct group for each of them. The nBLAST analysis resulted in most of the query sequences leading to misidentification at the species level. Because of the low ability of the *mat*K region for species discrimination, as

indicated by nBLAST and phylogenetic analysis, along with the difficulty in amplifying it, makes this region unsuitable as a barcode region for Dipterocarpaceae. Regarding the *rbc*L region, we could not observe any affiliation of the query sequences from the laboratory since only several sequences of this region are available in the databases. However, based on our neighbor joining analysis, we observed that this region is able to discriminate above the generic level but not the specific level.

DNA-based identification using a taxonomic identification key indicated that the approach is not yet a suitable tool to discriminate species. Many species belonging to the same haplotype were detected when constructing the key. A possible reason for this is the use of the *trn*L intron region to construct the key. Taberlet et al., (2007) has reported that this region is not effective in distinguishing closely-related species.

8 Zusammenfassung

Die Arten der Familie der Dipterocarpaceaen (Flügelfruchtgewächse) sind in der Region Malesien die Hauptbaumarten in Bezug auf Holzgewinnung. Die geografische Verbreitung der Pflanzenfamilie erstreckt sich bis Südamerika und Afrika. Die Familie umfasst etwa 500 Arten in 17 verschiedenen Gattungen und ist unterteilt in drei Unterfamilien: Dipterocarpoideae, Monotoideae und Pakaraimoideae (Ashton, 1982). Dipterocarpoideae ist mit 470 Arten in 13 Gattungen die artenreichste Unterfamilie (Ashton, 1982). Sie ist noch einmal unterteilt in zwei Triben: Dipterocarpeae und Shoreae. Dipterocarpeae umfasst die Gattungen *Anisoptera*, *Cotylelobium*, *Dipterocarpus*, *Stemonoporus*, *Upuna*, *Vateria* und *Vateriopsis*, Shoreae die Gattungen *Dryobalanops*, *Hopea*, *Neobalanocarpus*, *Parashorea* und *Shorea*. *Shorea* und *Hopea* sind mit 169, bzw. 100 Arten die artenreichsten Gattungen.

Studien zur molekularen Phylogenie der Unterfamilie Dipterocarpoideae werden bereits seit 1998 durchgeführt, besonders an der Gattung *Shorea* und ihren Schwestergattungen im Tribus Shoreae, da diese Gattung die höchste Artenzahl aufweist und von allen Dipterocarpaceaen das wertvollste Holz liefert. Viele dieser Arten sind vom Aussterben bedroht. Ziel von Untersuchungen zur molekularen Phylogenie ist die Vervollständigung von Phylogenien, die auf morphologischen Merkmalen beruhen, da die Einordnung von einigen Gattungen im Tribus Dipterocarpoideae noch immer zur Diskussion steht. Die Klassifizierung von *Shorea* in dieser Untersuchung bezieht sich auf Ashton (1982) und Symington (1943). Symington (1943) unterteilt *Shorea* basierend auf der Farbe des Holzes (White Meranti, Yellow Meranti, Balau und Red Meranti). Ashton (1982) hat die Klassifizierung von Symington (1943) grundsätzlich beibehalten, aber einige der Gruppen wurden in niedrigere taxonomische Ränge neu klassifiziert.

Die Nachfrage nach Identifikationsmöglichkeiten für Dipterocapaceaen zur Vermeidung von Betrug bei der Zertifizierung von Holz hat zu einer Verbesserung moderner Identifikationssysteme geführt, die auch molekulare Daten nutzen. Traditionell werden Dipterocarpaceaen anhand von morphologischen Merkmalen identifiziert. Allerdings ist diese Art der Bestimmung ist oft nur eingeschränkt nutzbar, vor allem wenn keine Blüte vorhanden ist, da dies das eindeutigste taxonomische Bestimmungsmerkmal bei Dipterocarpaceaen ist.

Die große Menge molekularer Daten und die fortschrittlichen Technologien im Bereich der DNA-Sequenzierung ermöglichten es dem DNA-Barcoding zu einer weitverbreiteten

Technik für verschiedene taxonomische Studien zu werden. Dabei will es die traditionelle Taxonomie nicht ersetzen, sondern ergänzen und den Identifikationsvorgang beschleunigen. Zusätzlich ermöglicht die große Anzahl an verfügbaren Sequenzen in öffentlichen Datenbanken, wie z.B. die NCBI-Datenbank, die Entwicklung eines molekularen taxonomischen Schlüssels, einem neuen Konzept der Artidentifikation. Allerdings sind die Methoden des DNA-Barcoding noch immer in ihren Anfängen, so wird z.B. die Datenbank für das Projekt DNA Barcoding zurzeit noch eingerichtet.

Diese Studie hat zum Ziel, mithilfe von vier Chloroplastenregionen (*trn*L intron, *psb*C-*trn*S IGS, *mat*K und *rbc*L) die phylogenetischen Beziehungen in der Unterfamilie Dipterocarpoideae zu erschließen, sowie die Einordnung der verschiedenen Gattungen. Zusätzlich prüft diese Untersuchung auch die Eignung der beiden Barcoding-Regionen *mat*K und *rbc*L, die vom Konsortium Barcode of Life (CBOL) im Jahr 2009 vorgeschlagen wurden. Ein weiteres Ziel ist die Entwicklung eines taxonomischen Identifizierungsschlüssels für die Identifizierung von Arten basierend auf der phylogenetischen Analyse.

Alle Sequenzen von Dipterocarpaceaen, die in der NCBI-Datenbank hinterlegt sind, wurden für vier Chloroplastenregionen (*trn*L intron, *psb*C-*trn*S IGS, *mat*K und *rbc*L) abgerufen. Zusätzlich zu den Sequenzen aus der NCBI-Datenbank wurden für die Untersuchung auch Proben sequenziert, die in der Abteilung Forstgenetik und Forstpflanzenzüchtung der Universität Göttingen zur Verfügung standen, um eine höchstmögliche Zahl von unterschiedlichen Arten untersuchen zu können.

Für die phylogenetischen Analysen wurde die Software MEGA 5 verwendet und die statistischen Methoden maximum parsimony (MP), maximum likelihood (ML) und neighbor joining (NJ). Für die DNA-basierte Identifizierung wurde die Eignung von zwei Barcoding-Regionen mithilfe von nBLAST getestet. Die phylogenetische Analyse wurde unter Verwendung der neighbor joining-Methode durchgeführt.

Es war für eine große Anzahl von Arten möglich, Sequenzen von den oben genannten Chloroplastenregionen zu erhalten: 145 Arten für *trn*L intron, 117 Arten für *psb*C-*trn*S IGS, 116 Arten für *mat*K und 69 Arten für *rbc*L. Die Länge der Sequenzen für die verschiedenen Regionen variierte, 537 bp, 1136 bp, 653 bp und 647 bp für die Regionen *trn*L intron, *psb*C-*trn*S IGS, *mat*K bzw. *rbc*L.

Die verschiedenen Methoden MP, ML und NJ für die phylogenetischen Analysen erzeugten sehr ähnliche Baumtopologien. Daher basiert die Diskussion vor allem auf den Ergebnisse

der MP-Methode. Grundsätzlich war es nicht möglich, die evolutionären Beziehungen der Unterfamilie der Dipterocarpoideae anhand der vier Chloroplastenregionen eindeutig zu entschlüsseln. Die Regionen ermöglichten nur eine Aufklärung der Triben Dipterocarpeae und Shoreae, waren aber innerhalb der Triben deutlich weniger erfolgreich, vor allem in Bezug auf Shoreae. Für die Gattung *Dipterocarpus* stehen bisher nur Sequenzdaten der Regionen *trn*L intron und *mat*K zur Verfügung. In dieser Studie wurden für beide Regionen eindeutig abgrenzbare Gruppen von Arten entdeckt. Es wird vermutet, dass *Dipterocarpus* die basale Gruppe der Dipterocarpoideae repräsentiert (Meijer, 1979). Diese Gattung hat der Familie auch ihren Namen gegeben, möglicherweise weil sie als eine sehr ursprüngliche Gruppe innerhalb der Dipterocarpaceaen gilt (Maury – Lechon, 1979). Auch ist diese Gattung innerhalb der Familie der Dipterocarpaceaen eindeutig definiert, basierend auf morphologischen Merkmalen und molekularen Analysen.

Die Analysen der Region *psb*C-*trn*S IGS bestätigten die Ergebnisse von Symington (1943) basierend auf der Farbe des Holzes dahingehend, dass die Gattung *Shorea* eine monophyletische Gruppe bildet. Durch die Analyse der Region *mat*K war es am ehesten möglich, die Beziehungen innerhalb des Tribus Dipterocarpeae zu beschreiben und die Sektion *Doona* innerhalb der Gattung *Shorea* als monophyletische Gruppe abzugrenzen. Allerdings war diese Region nicht geeignet für die weitere Klassifizierung innerhalb des Tribus Shoreae. Obwohl die Region *rbc*L die erste Chloroplastenregion ist, die sequenziert wurde, sind in der NCBI-Datenbank nur wenige Sequenzen verfügbar. Die Ergebnisse basierend auf den eigenen Labordaten führten zu dem Schluss, dass diese Region nicht geeignet ist, um die evolutionären Beziehungen der Dipterocarpoideae unterhalb der Gattungsebene aufzuzeigen. Die Region *mat*K zeigte in dieser Untersuchung eine nahe Verwandtschaft zwischen den Gattungen *Dryobalanops* und *Dipterocarpus*, während die Region *trn*L intron eher darauf hindeutete, dass *Dryobalanops* eine Verwandtschaft zur Sektion *Balau* aus der *Shorea*-Gruppe aufweist. Diese gegensätzlichen Ergebnisse unterstützen die Annahme von Indrioko (2005), dass diese Gattung eine basale Gruppe des Tribus Shoreae ist.

Die Artidentifizierung basierend auf DNA-Daten wurde anhand von zwei Vorgehensweisen untersucht, DNA-Barcoding und ein molekularer taxonomischer Identifizierungsschlüssel. Die zwei Barcode-Regionen *mat*K und *rbc*L, übernommen vom Consortium for the Barcode of Life für Landpflanzen (Hollingsworth et al., 2009), wurden auf ihre Eignung als Barcoding-Regionen für die Unterscheidung der Dipterocarpaceae geprüft. Die meisten

benötigten Informationen für die Region *mat*K waren in der NCBI-Datenbank vorhanden, aber es wurden auch einige zusätzliche Proben in dieser Studie verwendet. Insgesamt wurden 119 bzw. 67 Proben für die Untersuchung der Region *mat*K bzw. *rbc*L, verwendet. Für die Beurteilung der Effektivität der Barcoding-Analyse in dieser Untersuchung wurden zunächst mithilfe von neighbor joining-Bäumen monophyletische Gruppen einmal für die Eingabesequenzen und einmal für die Referenzsequenzen, die in der NCBI-Datenbank hinterlegt sind, identifiziert. Unter Verwendung von nBLAST wurde dann nach Ähnlichkeiten zwischen den Eingabesequenzen aus dem Labor und den Sequenzen aus der NCBI-Datenbank gesucht. Obwohl der neighbor joining-Baum einige der Sequenzen in die korrekte Gattung eingeordnet hat, konnte diese Region keine drei klar abgetrennten Gruppen für die Gattungen *Shorea*, *Hopea* und *Parashorea* erstellen. Die nBLAST-Analyse ergab für die meisten Eingabesequenzen auf der Artebene eine falsche Identifizierung. Aufgrund der fehlenden Unterscheidung zwischen Arten durch die Region *mat*K, was nicht nur durch die Ergebnisse des nBLAST, sondern auch durch die phylogenetische Analyse deutlich wurde, und der Probleme bei der Amplifizierung ist diese Region ungeeignet als Barcoding-Region für die Familie der Dipterocarpaceaen. Über die Region *rbc*L kann keine weitere Aussage gemacht werden, da nur wenige Sequenzen für diese Region in der Datenbank verfügbar waren. Allerdings konnte die neighbor joining-Analyse zeigen, dass diese Region erfolgreich auf der Gattungsebene unterscheidet, aber nicht auf der Artebene.

Das DNA-basierte Identifizierungsverfahren unter der Verwendung eines taxonomischen Identifizierungsschlüssels kann noch nicht ausreichend zwischen Arten unterscheiden. Viele verschiedene Arten mit dem gleichen Haplotypen wurden bei der Erstellung des Schlüssels gefunden. Ein möglicher Grund ist die Verwendung der Region *trn*L intron für die Erstellung des Schlüssels. Taberlet (2007) berichtet, dass diese Region nicht effektiv ist bei der Unterscheidung zwischen nah verwandten Arten.

References

Apanah S. (1993). Mass flowering of dipterocarp forest in the aseasonal tropics. Journal of Bioscience 18 (4): 457 – 474

Ashton P.S. (1979) Phylogenetic speculation on dipterocarpaceae. In: Maury-Lechon G (ed) Dipterocarpacees, Taxonomie-Phylogenie-Ecologie Memoires du Museum National d'Histoire Naturelle. Paris. Series B, Botanique 26: 145-149

Ashton P.S. (1982) Dipterocarpaceae. In: Van Steenis CGGJ (ed) Flora malesiana. Series I. Spermatophyta. 9: 237 - 552

Ashton P.S. (1988) Dipterocarp Biology as a Window to the Understanding of Tropical Forest Structure. Annual Review of Ecology and Systematics 19:347-370

Ashton P.S. (2004) Dipterocarpaceae. In E. Soepadmo, L.G. Saw and R.C.K. Chung (ed) Tree Flora of Sabah and Sarawak, 5: 63-338. Forest Research Institute Malaysia, Kuala Lumpur

Ashton P.S., Gan Y.Y., Robertson F.W. (1984) Electrophoretic and Morphological Comparisons in 10 Rain-Forest Species of Shorea (Dipterocarpaceae), Botanical Journal of the Linnean Society. pp. 293-304.

Bell D., Long D.G., Forrest A.D., Hollingsworth M.L., Blom H.H., Hollingsworth P.M. (2012) DNA barcoding of European Herbertus (Marchantiopsida, Herbertaceae) and the discovery and description of a new species. Mol Ecol Resources 12:36-47

Birky C.W. (1995) Uniparental Inheritance of Mitochondrial and Chloroplast Genes - Mechanisms and Evolution. Proceedings of the National Academy of Sciences of the United States of America 92:11331-11338.

Brinegar C. (2009) Assessing Evolution and Biodiversity in Plants at The Molecular Level. Kathmandu University Journal of Science, Engineering and Technology. 5 (II) pp 149 - 159

Cao C.P., Gailing O., Siregar I., Indrioko S., Finkeldey R. (2006) Genetic variation at AFLPs for the Dipterocarpaceae and its relation to molecular phylogenies and taxonomic subdivisions. Journal of Plant Research 119:553-558

Cawthorn D.-M., Steinman H.A., Corli Witthuhn R. (2011) Establishment of a mitochondrial DNA sequence database for the identification of fish species commercially available in South Africa. Molecular Ecology Resources 11:979-991

Chao K., M. , Zhang L. (2009) Sequence comparison. Theory and Methods. Spinger. 210 pp

Chase M.W., Salamin N., Wilkinson M., Dunwell J.M., Kesanakurthi R.P., Haidar N., Savolainen V. (2005) Land plants and DNA barcodes: short-term and long-term goals. Philosophical Transactions of the Royal Society B-Biological Sciences 360:1889-1895

Cowan R.S., Fay M.F. (2012) Challenges in the DNA Barcoding of Plant Material Plant DNA Fingerprinting and Barcoding. In Methods Mol Biol 862 : 23-33

Cuenoud P., Savolainen V., Chatrou L.W., Powell M., Grayer R.J., Chase M.W. (2002) Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences. American Journal of Botany 89: 132-144

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., Rich T.C.G., Ford C.R., Trinder S.A., Long C., Moore C.W., Satterthwaite D., Davies H., Allainguillaume J., Ronca S., Tatarinova T., Garbett H., Walker K., Wilkinson M.J. (2012) DNA Barcoding the Native Flowering Plants and Conifers of Wales. PLoS ONE. 7 (6) .e37945

Dick C.W., Webb C.O. (2012) Plant DNA Barcodes, Taxonomic Management, and Species Discovery. In *Methods Mol Biol* 858: 379-393

Fazekas AJ, Burgess KS, Kesanakurti PR et al (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One 3:e2802

Finkeldey R., Leinemann L., Gailing O. (2010) Molecular genetic tools to infer the origin of forest plants and wood. Applied Microbiology and Biotechnology 85: 1251-1258

Gamage D.T., de Silva M.P., Inomata N., Yamazaki T., Szmidt A.E. (2006) Comprehensive molecular phylogeny of the sub-family Dipterocarpoideae (Dipterocarpaceae) based on chloroplast DNA sequences. Genes & Genetic Systems 81: 1-12

Gemeinholzer B, Christoph Oberprieler, Bachmann K. (2006) Using GenBank Data for Plant Identification: Possibilities and Limitations Using the ITS 1 of Asteraceae Species Belonging to the Tribes Lactuceae and Anthemideae. JSTOR Taxon, Vol. 55. (1): pp. 173-187

Gielly L., Taberlet P. (1994) The Use of Chloroplast DNA to Resolve Plant Phylogenies - Noncoding Versus Rbcl Sequences. Molecular Biology and Evolution 11: 769-777

Graham S.W., Olmstead R.G. (2000) Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. American Journal of Botany 87: 1712-30

Haider N. (2011) Chloroplast-specific universal primers and their uses in plant studies. Biologia Plantarum 55: 225-236

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

Hall B.G. (2011) Phylogenetic Trees Made Easy. A How-to Manual. Fourth edition. Sinauer Associates Inc. Sunderland. 282 pp

Hall T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95-98

Hawksworth D.L. (1995) Biodiversity: Measurement and Estimation. Chapman and Hall. London

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

Hilu K.W., Borsch T., Muller K., Soltis D.E., Soltis P.S., Savolainen V., Chase M.W., Powell M.P., Alice L.A., Evans R., Sauquet H., Neinhuis C., Slotta T.A.B., Rohwer J.G., Campbell C.S., Chatrou L.W. (2003) Angiosperm phylogeny based on matK sequence information. American Journal of Botany 90: 1758-1776

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

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and Using a Plant DNA Barcode. PLoS ONE 6 (5): e19254

Hollingsworth P.M., Forrest L.L., Spouge J.L., Hajibabaei M., Ratnasingham S., van der Bank M., Chase M.W., Cowan R.S., Erickson D.L., Fazekas A.J., Graham S.W., James K.E., Kim K.J., Kress W.J., Schneider H., van AlphenStahl J., Barrett S.C.H., van den Berg C., Bogarin D., Burgess K.S., Cameron K.M., Carine M., Chacon J., Clark A., Clarkson J.J., Conrad F., Devey D.S., Ford C.S., Hedderson T.A.J., Hollingsworth M.L., Husband B.C., Kelly L.J., Kesanakurti P.R., Kim J.S., Kim Y.D., Lahaye R., Lee H.L., Long D.G., Madrinan S., Maurin O., Meusnier I., Newmaster S.G., Park C.W., Percy D.M., Petersen G., Richardson J.E., Salazar G.A., Savolainen V., Seberg O., Wilkinson M.J., Yi D.K., Little D.P., Grp C.P.W. (2009) A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America 106: 12794-12797

Humphries E.M., Winker K. (2010) Working through polytomies: auklets revisited. Molecular Phylogenetics and Evolution 54: 88-96

Indrioko S. (2005). Chloroplast DNA variation in Indonesia Dipterocarpacae-phylogenetic, taxonomic and population genetic aspects. PhD thesis. University of Goettingen. Goettingen

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

Ishiyama H. T. Kado M. Iwasaki N. A B Shukor, A. E. Szmidt, and T.Yamazaki. (2003). Nucleotide variation in the GapC region of four species of Shorea and its putative hybrids. Tropics 13: 89–99

IUCN. (2011) IUCN Red List of Threatened Species. Version 2011.1. www.iucnredlist.org.

Johnson L.A., Soltis.D.E. (1994.) MatK DNA sequences and phylogenetic reconstruction in Saxifragaceae sensu stricto. Systematic Botany 19: 143–156

Kajita T., Kamiya K., Nakamura K., Tachida H., Wickneswari R., Tsumura Y., Yoshimaru H., Yamazaki T. (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

Kamiya K., Harada K., Tachida H., Ashton P.S. (2005) Phylogeny of PgiC gene in Shorea and its closely related genera (Dipterocarpaceae), the dominant trees in southeast Asian tropical rain forests. American Journal of Botany 92: 775-788

Kelchner S.A. (2000) The evolution of non-coding chloroplast DNA and its application in plant systematics. Annals of the *Missouri Botanical Garden*. 87: 482–498

Kerfeld C. S, Scott K. M. (2011) Using BLAST to teach "E-value-tionary" concepts. PLoS Biol 9 (2): e1001014

Kimura M. (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution. 16: 111-120

Kool A., de Boer H.J., Kruger A., Rydberg A., Abbad A., Bjork L., Martin G. (2012) Molecular identification of commercialized medicinal plants in southern Morocco. PLoS One 7:e39459.

Kostermans A.J.G. (1981) Stemonoporus Thw. (Dipterocarpaceae) - a Monograph 1. Bulletin Du Museum National D Histoire Naturelle-Section B Adansonia 3: 321-358

Kreft H., Jetz W. (2010) A framework for delineating biogeographical regions based on species distributions. Journal of Biogeography 37:2029-2053.

Kress W.J. (2005) Use of DNA barcodes to identify flowering plants. Proceedings of the National Academy of Sciences 102: 8369-8374

Kress WJ, Erickson DL (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:e508

Kress W.J., Erickson D.L., Jones F.A., Swenson N.G., Perez R., Sanjur O., Bermingham E. (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proceedings of the National Academy of Sciences 106:18621-18626

Lawton J.H., and , May R.M. (1995) Extinction rates. Oxford University Press, London

Lemey P., Salemi M., Vandamme A.M. (2009) The Phylogenetic Handbook. A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press. 723 pp

Li Q.-M., He T.-H., Xu Z.-F. (2004) Generic Relationships of Parashorea chinensis Wang Hsie (Dipterocarpaceae) Based on cpDNA Sequences. Taxon 53: 461-466

Little D.P., Stevenson D.W. (2007) A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. Cladistics 23: 1-21

Little D.P. (2011) DNA Barcode Sequence Identification Incorporating Taxonomic Hierarchy and within Taxon Variability. PLoS ONE 6: e20552

McLennan D.A. (2010) How to Read a Phylogenetic Tree. Evo Edu Outreach 3:506–519.

Madden T. (2002) The BLAST Sequence Analysis Tool. In: McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US)

Maury - Lechon G. (1979) Systematic and Phylogenetic Interest of Juvenile Characters (Germination, Mature Embryo, Seedlings) in Dipterocarpaceae. Bulletin De La Societe Botanique De France-Actualites Botaniques 126: 13-21

Maury G., Muller J., Lugardon B. (1975) Notes on Morphology and Fine-Structure of Exine of Some Pollen Types in Dipterocarpaceae. Review of Palaeobotany and Palynology 19: 241-289

Maury L.,G., Curtet L. (1998) Biogeography and evolutionay systematics of family Dipterocarpaceae. In Areview of Dipterocarps, taxonomy, ecology and silviculture (eds. S. Appanah and J,M. Turnbull). Center for forest research Institute, Malaysia. pp 5 - 44

Mayo S.J., Allkin,R., Baker,W, Blagoderov, V., Brake, I., Clark,B., Govaerts,R., Godfray,C, Haigh A., Hand,R., Harman,K., Jackson,M., Kilian, N., Kirkup,D.W., Kitching,I., Knapp,I, Lewis G.P., Malcolm,P., Raab-Straube, E.von., Roberts, D.M., Scoble,M., Simpson,D.A., Smith,C., Smith V., Villalba,S., Walley,L., & Wilkin,P. (2008) Alpha e-taxonomy responses from the systematics community. Kew Bulletin 63: 1–16

Meijer W. and Wood, G.H.S. (1964). Dipterocarps of Sabah. Pp. 110-112. Sabah Forest Record No. 5. Forest Departement Sabah. Sandakan. 344.pp

Meijer W. & Wood G. H. S. (1976) Keys to Dipterocarps of Sabah. BIOTROP, Bogor

Meijer W. (1979) Taxonomic studies in the genus Dipterocarpus. In G. Maury - Lechon (ed). Dipterocarpacees:Taxonomie Phylogenie-Ecologie, Memoires du Museum National d'Histoire Naturelle. Botanique 26: 50 - 56

Meyer C.P., Paulay G. (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biol 3: e422

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

Morton C. M, Dayanandan S, Dissanayake D. (1999). Phylogeny and biosystematics of Pseudomonotes (Dipterocarpaceae) based on molecular and morphological data. Plant Systematics and Evolution 212: 197–205

Mount D.W. (2007a) Steps Used by the BLAST Algorithm. Cold Spring Harbor Protocols

Mount D.W. (2007b) Using the Basic Local Alignment Search Tool (BLAST). Cold Spring Harbor Protocols

Muller K., Borsch a.T. (2005) Phylogenetics of Utricularia (Lentibulariaceae) and molecular evolution of the trnK intron in a lineage with high substitutional rates. Plant Syst. Evol 250:39–67

Nguyen P.N. (2009). Molecular Phylogeny of Southeast-Asian Dipterocarps belonging to tribe Dipterocarpaceae (Family Dipterocarpaceae) based on Non-Coding Sequence Data of Chloroplast and Nuclear DNA. PhD thesis. University of Goettingen. Goettingen

Nuroniah H.S. (2009). Diagnostic markers for the identification of the tree species Shorea leprosula Miq. and S. parvifolia Dyer and the geographic origin of S. leprosula Miq. PhD thesis. University of Goettingen. Goettingen

Nuroniah H.S., Gailing O., Finkeldey R. (2010) Development of SCAR Markers for Species Identification in the Genus Shorea (Dipterocarpaceae). Silvae Genetica 59:249-257.

Oginuma K., Kono Y., and Tobe H. (1999) Embryology of Vateriopsis (Dipterocarpaceae) and relationship of the family. In: Abstract for the XVI International Botanical Congress,http://www.biologie.unihamburg.de/bonline/ibc99/ibc/abstarct/listen/abstracts /3183.html.in Gamage, D. T., de Silva, M. P., Inomata, N., Yamazaki, T., & Szmidt, A. E. (2006). Comprehensive molecular phylogeny of the sub-family Dipterocarpoideae (Dipterocarpaceae) based on chloroplast DNA sequences. Genes & Genetic Systems, 81(1): 1-12

Olmstead, R. G., and Palmer, J. D. (1994). Chloroplast DNA systematics: a review of methods and data analysis. American Journal of Botany 81:1205-1224.

Parameswaran and Gottwald H. (1979) Problematic taxa in the Dipterocarpaceae. Their anatomy and taxonomy. In: Dipterocarpacees, Taxonomie-Phylogenie-Ecologie, 14 -17 Juin 1977. Paris First International Round Table on Dipterocarpaceae. Maury - Lechon G. (ed). Memoires du Museum National d'Histoire Naturelle, Serie B. Botanique 26: 69 - 75

Pons J. (2006) DNA-based identification of preys from non-destructive, total DNA extractions of predators using arthropod universal primers. Molecular Ecology Notes 6:623-626

Rachmayanti Y, Leinemann L, Gailing O, Finkeldey R. (2009) Extraction, amplification and characterization of wood DNA from Dipterocarpaceae. Plant Molecular and Biology Report. 24: 45–55

Rath P., Rajaseger G., Goh C.J., Kumar P.P. (1998) Phylogenetic Analysis of Dipterocarps Using random Amplified Polymorphic DNA Markers. Annals of Botany 82: 61-65

Ratnasingham S., Hebert P.D. (2007) Bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7: 355-364

Raubeson, L. A., and Jansen R. K. (2005).Chloroplast genomes of plants. pp. 45–68. In R.Henry,ed. Diversity and evolution of plants-genotypic and phenotypic variation in higher plants. CABI Publishing. Oxfordshire. UnitedKingdom

Ridgway K . P., Duck J. M and Young J. P. W. (2003). Identification of roots from grass swards using PCR-RFLP and FFLP of the plastid *trn*L (UAA) intron. BMC Ecology 3:8

Roy S., Tyagi A., Shukla V., Kumar A., Singh U.M., Chaudhary L.B., Datt B., Bag S.K., Singh P.K., Nair N.K., Husain T., Tuli R. (2010) Universal Plant DNA Barcode Loci May Not Work in Complex Groups: A Case Study with Indian Berberis Species. PLoS ONE 5: e13674

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. 2nd ed. Cold Spring Harbor N.Y., Cold Spring Harbor Laboratory

Sanger F., Nicklen S., Coulson A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74: 5463-7

Schuh R. T. 2000. Biological Systematics: Principles and Applications. Cornell University Press, Ithaca. 236 pp

Shaw J. (2007) A molecular phylogenetic approach to the evolution of bryophytes. Bryologist 110: 178-178

Shaw J., Lickey E.B., Beck J.T., Farmer S.B., Liu W., Miller J., Siripun K.C., Winder C.T., Schilling E.E., Small R.L. (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. American Journal Botany 92: 142-66

Simpson, M.G. (2006). Plant Systematics. 1st editon. Elsevier-Academic Press. 590 pp

Singh G. (2004) Plant systematics: an integrated approach. Plants Systematics: An Integrated Approach. Enfield, N.H: Science Publishers

Soltis D.E., Soltis P.S. (2000) Contributions of plant molecular systematics to studies of molecular evolution. Plant Molecular Biology 42: 45–75

Soltis D.E., Soltis P.S. (2003) The role of phylogenetics in comparative genetics. Plant Physiol 132: 1790-800

Soltis P.S., Soltis D.E., Doyle J.J. (1998) Choosing an approach and an appropriate gene for phylogenetic analysis. D. E. Soltis, P. S. Soltis, and J. J. Doyle, eds. Molecular Systematics of Plants II. DNA Sequencing. pp 1-42

Stuessy T.F. (2008) Plant Taxonomy. The systematic Evaluation of Comparative data. Columbia Univeristy Press. 539 pp

Sun X.-Q., Zhu Y.-J., Guo J.-L., Peng B., Bai M.-M., Hang Y.-Y. (2012) DNA Barcoding the *Dioscorea* in China, a Vital Group in the Evolution of Monocotyledon: Use of *mat*K Gene for Species Discrimination. PLoS ONE 7: e32057

Suzuki E., and Ashton P.S. (1996). Sepal and nut size ratio of fruits of Asian Dipterocarpaceae and its implications for dispersal. Journal of Tropical Ecology, 12, pp 853-870

Symington C.F. (1943) Forester's Manual of Dipterocarps. Malayan forester records no 16. Penerbit Universiti Malaysia. Kuala Lumpur

Taberlet P., Gielly L., Pautou G., Bouvet J. (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17: 1105-9

Taberlet P., Coissac E., Pompanon F., Gielly L., Miquel C., Valentini A., Vermat T., Corthier G., Brochmann C., Willerslev E. (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Research 35: e14-e14

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28:2731-2739

Tautz D., Arctander P., Minelli A., Thomas R.H., Vogler A.P. (2003) A plea for DNA taxonomy. Trends in Ecology & Evolution 18: 70-74

Thompson D.J., Desmond G.H.,Toby J.G. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22: 4673-4680

Tnah L.H., Lee S.L., Ng K.K.S., Faridah Q.Z., Faridah-Hanum I. (2010) Highly Variable Str Markers of Neobalanocarpus Heimii (Dipterocarpaceae) for Forensic DNA Profiling. Journal of Tropical Forest Science 22: 214-226

Tsumura Y., Kawahara T., Wickneswari R., Yoshimura K. (1996) Molecular phylogeny of Dipterocarpaceae in Southeast Asia using RFLP of PCR-amplified chloroplast genes. Theoretical and Applied Genetics 93: 22-29

Tsumura Y., Kado T., Yoshida K., Abe H., Ohtani M., Taguchi Y., Fukue Y., Tani N., Ueno S., Yoshimura K., Kamiya K., Harada K., Takeuchi Y., Diway B., Finkeldey R., Na'iem M., Indrioko S., Ng K.K.S., Muhammad N., Lee S.L. (2011) Molecular database for classifying Shorea species (Dipterocarpaceae) and techniques for checking the legitimacy of timber and wood products. Journal of Plant Research 124: 35-48

Vijayan K., Tsou C.H. (2010) DNA barcoding in plants: taxonomy in a new perspective. Current Science 99: 1530-1541

Wesselink M., Kuiper I. (2008) Species identification of botanical trace evidence using molecular markers. Forensic Science International: Genetics Supplement Series 1:630-632

Wiley E.O., Lieberman B.S. (2011) Phylogenetics.Theory and Practice of Phylogenetic Systematics (2nd ed). 424 pp. John Wiley and Sons. New York etc.

Whitmore T. C. (1962). Studies in systematic bark morphology. III. Bark taxonomy in Dipterocarpaceae. Gardens Bulletin, Singapore 19: 321–371

Yu J., Xue J.H., Zhou S.L. (2011) New universal matK primers for DNA barcoding angiosperms. Journal of Systematics and Evolution 49:176-181

Yulita K.S., Bayer R.J., West J.G. (2005) Molecular phylogenetic study of Hopea and Shorea (Dipterocarpaceae): Evidence from the trnL-trnF and internal transcribed spacer regions. Plant Species Biology 20:167-182

Zhang C.Y., Wang F.Y., Yan H.F., Hao G., Hu C.M., Ge X.J. (2012) Testing DNA barcoding in closely related groups of Lysimachia L. (Myrsinaceae). Molecular Ecology Resources 12: 98-108

Zhou Y., Lu C., Wu Q. J., Wang Y., Sun Z.T., Deng J.C., Zhang Y. (2008) GISSD: Group I Intron Sequence and Structure Database. Nucleic Acids Resources 36: D31-D37

Zhu W., Liao B., Li R. F. (2010) A Method for Constructing Phylogenetic Tree Based on a Dissimilarity Matrix. Match-Communications in Mathematical and in Computer Chemistry 63: 483-492

Ziegenhagen B., Fladung M. (1997) Variation in the psbC gene region of gymnosperms and angiosperms as detected ky a single restriction site polymorphism. Theoretical and Applied Genetics 94: 1065-1071

Zulkifli J., F.,, Ariffin N., A., Sherif S., MD. (2012) Molecular Characteristic of Partial Cytochrome Oxidase Subunit 1 (CO1) on Tenualosa toli from Sarawak.In UMT 11th International Annual Symposium on Sustainability Science and Management 09th – 11th July 2012, Terengganu, Malaysia

Zurawski G., Perrot B, Bottomley W, Whitfeld W. (1981) The structure of the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase from spinach chloroplast. NucleicAcids Res 9:3252-3270

Appendices

Appendix 1. List of plant species and corresponding GenBank accession numbers retrieved from the database for *trn*L intron

No	GeneBank identifier no	Accession number	Species
1	gi 226236582	AB451982	*Shorea acuminata*
2	gi 226236585	AB451979	*Shorea acuminata*
3	gi 226236578	AB451986	*Shorea acuminatissima*
4	gi 226236579	AB451985	*Shorea acuminatissima*
5	gi 226236580	AB451984	*Shorea acuminatissima*
6	gi 226236572	AB451988	*Shorea acuta*
7	gi 226236573	AB451987	*Shorea acuta*
8	gi 226236577	AB451990	*Shorea acuta*
9	gi 226236575	AB451992	*Shorea agami*
10	gi 226236576	AB451991	*Shorea agami*
11	gi 208609662	AB458531	*Shorea albida*
12	gi 208609666	AB458535	*Shorea albida*
13	gi 226236542	AB451994	*Shorea almon*
14	gi 226236543	AB451995	*Shorea almon*
15	gi 226236545	AB451997	*Shorea amplexicaulis*
16	gi 226236546	AB451998	*Shorea amplexicaulis*
17	gi 226236547	AB451999	*Shorea andulensis*
18	gi 226236548	AB452000	*Shorea argentifolia*
19	gi 226236549	AB452001	*Shorea assamica*
20	gi 226236550	AB452002	*Shorea assamica*
21	gi 226236551	AB452003	*Shorea assamica*

No	GeneBank identifier no	Accession number	Species
33	gi 226236563	AB452015	*Shorea confusa*
34	gi 226236564	AB452016	*Shorea confusa*
35	gi 226236566	AB452018	*Shorea crassa*
36	gi 226236567	AB452019	*Shorea curtisii*
37	gi 226236568	AB452020	*Shorea curtisii*
38	gi 226236569	AB452021	*Shorea dasyphylla*
39	gi 226236570	AB452022	*Shorea dasyphylla*
40	gi 226236612	AB452023	*Shorea domatiosa*
41	gi 22034068	AY026548	*Shorea exelliptica*
42	gi 22034069	AY026549	*Shorea faguetiana*
43	gi 226236613	AB452024	*Shorea faguetiana*
44	gi 226236614	AB452025	*Shorea faguetiana*
45	gi 226236616	AB452027	*Shorea faguetioides*
46	gi 226236617	AB452028	*Shorea falcifera*
47	gi 226236618	AB452029	*Shorea falciferoides*
48	gi 226236619	AB452030	*Shorea fallax*
49	gi 226236620	AB452031	*Shorea fallax*
50	gi 226236622	AB452033	*Shorea fallax*
51	gi 226236625	AB452036	*Shorea ferruginea*
52	gi 226236626	AB452037	*Shorea ferruginea*
53	gi 226236627	AB452038	*Shorea flaviflora*

22	gi 226236552	AB452004	*Shorea atrinervosa*
23	gi 226236553	AB452005	*Shorea atrinervosa*
24	gi 22034066	AY026546	*Shorea balangeran*
25	gi 22034067	AY026547	*Shorea beccariana*
26	gi 226236554	AB452006	*Shorea beccariana*
27	gi 226236555	AB452007	*Shorea biawak*
28	gi 226236556	AB452008	*Shorea biawak*
29	gi 226236558	AB452010	*Shorea bracteolata*
30	gi 226236559	AB452011	*Shorea bracteolata*
31	gi 226236560	AB452012	*Shorea bullata*
32	gi 226236562	AB452014	*Shorea collina*
65	gi 226236640	AB452051	*Shorea isoptera*
66	gi 22034074	AY026554	*Shorea javanica*
67	gi 226236641	AB452052	*Shorea javanica*
68	gi 226236642	AB452053	*Shorea johorensis*
69	gi 226236644	AB452055	*Shorea johorensis*
70	gi 226236645	AB452056	*Shorea johorensis*
71	gi 22034075	AY026555	*Shorea johorensis*
72	gi 22034076	AY026556	*Shorea kunstleri*
73	gi 226236649	AB452060	*Shorea kunstleri*
74	gi 226236651	AB452062	*Shorea kunstleri*
75	gi 22034077	AY026557	*Shorea laevis*
76	gi 226236652	AB452063	*Shorea laevis*
77	gi 226236653	AB452064	*Shorea laevis*
78	gi 226236655	AB452066	*Shorea lepidota*
79	gi 22034078	AY026558	*Shorea leprosula*
80	gi 226236656	AB452067	*Shorea leprosula*
81	gi 226236657	AB452068	*Shorea leprosula*
82	gi 226236660	AB452071	*Shorea leprosula*

54	gi 22034070	AY026550	*Shorea foxworthyi*
55	gi 226236628	AB452039	*Shorea foxworthyi*
56	gi 226236629	AB452040	*Shorea gibbosa*
57	gi 22034071	AY026551	*Shorea guiso*
58	gi 226236630	AB452041	*Shorea guiso*
59	gi 226236631	AB452042	*Shorea havilandii*
60	gi 226236632	AB452043	*Shorea havilandii*
61	gi 226236635	AB452046	*Shorea henryana*
62	gi 22034072	AY026552	*Shorea hopeifolia*
63	gi 226236639	AB452050	*Shorea inappendiculata*
64	gi 22034073	AY026553	*Shorea isoptera*
98	gi 226236678	AB452089	*Shorea macroptera subsp. sandakanensis*
99	gi 22034081	AY026561	*Shorea materialis*
100	gi 226236679	AB452090	*Shorea materialis*
101	gi 226236680	AB452091	*Shorea materialis*
102	gi 22034082	AY026562	*Shorea maxima*
103	gi 226236682	AB452093	*Shorea maxima*
104	gi 226236683	AB452094	*Shorea maxima*
105	gi 22034083	AY026563	*Shorea maxwelliana*
106	gi 226236684	AB452095	*Shorea maxwelliana*
107	gi 22034084	AY026564	*Shorea mecistopteryx*
108	gi 226236685	AB452096	*Shorea mecistopteryx*
109	gi 226236686	AB452097	*Shorea mecistopteryx*
110	gi 226236687	AB452098	*Shorea mujongensis*
111	gi 22034085	AY026565	*Shorea multiflora*
112	gi 226236689	AB452100	*Shorea obscura*
113	gi 226236690	AB452101	*Shorea ochracea*
114	gi 226236691	AB452102	*Shorea ochracea*
115	gi 226236692	AB452103	*Shorea ochracea*

83	gi 226236661	AB452072	*Shorea longiflora*
84	gi 226236662	AB452073	*Shorea longiflora*
85	gi 22034079	AY026559	*Shorea longisperma*
86	gi 226236663	AB452074	*Shorea longisperma*
87	gi 22034080	AY026560	*Shorea macrophylla*
88	gi 226236668	AB452079	*Shorea macrophylla*
89	gi 226236670	AB452081	*Shorea macrophylla*
90	gi 226236673	AB452084	*Shorea macroptera*
91	gi 226236676	AB452087	*Shorea macroptera*
92	gi 4210582	AB006396	*Shorea macroptera*
93	gi 226236664	AB452075	*Shorea macroptera subsp. baillonii*
94	gi 226236665	AB452076	*Shorea macroptera subsp. baillonii*
95	gi 226236666	AB452077	*Shorea macroptera subsp. macropterifolia*
96	gi 226236667	AB452078	*Shorea macroptera subsp. macropterifolia*
97	gi 226236677	AB452088	*Shorea macroptera subsp. sandakanensis*
131	gi 226236707	AB452118	*Shorea parvistipulata*
132	gi 226236708	AB452119	*Shorea patoiensis*
133	gi 226236709	AB452120	*Shorea patoiensis*
134	gi 226236712	AB452123	*Shorea pauciflora*
135	gi 226236713	AB452124	*Shorea pauciflora*
136	gi 226236714	AB452125	*Shorea pauciflora*
137	gi 226236716	AB452127	*Shorea peltata*
138	gi 22034090	AY026570	*Shorea pilosa*
139	gi 226236717	AB452128	*Shorea pilosa*
140	gi 226236718	AB452129	*Shorea pilosa*
141	gi 22034091	AY026571	*Shorea pinanga*
142	gi 226236720	AB452131	*Shorea pinanga*
143	gi 226236721	AB452132	*Shorea platycarpa*
144	gi 226236722	AB452133	*Shorea platyclados*

116	gi 226236693	AB452104	*Shorea ochrophloia*
117	gi 22034086	AY02656	*Shorea ovalis*
118	gi 226236694	AB452105	*Shorea ovalis*
119	gi 226236695	AB452106	*Shorea ovalis*
120	gi 4210583	AB006397	*Shorea ovalis*
121	gi 226236697	AB452108	*Shorea ovata*
122	gi 22034087	AY026567	*Shorea palembanica*
123	gi 226236698	AB452109	*Shorea palembanica*
124	gi 226236699	AB452110	*Shorea palosapis*
125	gi 22034088	AY026568	*Shorea parvifolia*
126	gi 226236700	AB452111	*Shorea parvifolia*
127	gi 226236702	AB452113	*Shorea parvifolia*
128	gi 226236703	AB452114	*Shorea parvifolia*
129	gi 22034089	AY026569	*Shorea parvistipulata*
130	gi 226236706	AB452117	*Shorea parvistipulata*
162	gi 22034095	AY026575	*Shorea selanica*
163	gi 22034096	AY026576	*Shorea seminis*
164	gi 226236739	AB452150	*Shorea seminis*
165	gi 22034097	AY026577	*Shorea singkawang*
166	gi 226236740	AB452151	*Shorea singkawang*
167	gi 226236742	AB452153	*Shorea singkawang*
168	gi 226236743	AB452154	*Shorea slootenii*
169	gi 226236744	AB452155	*Shorea slootenii*
170	gi 226236745	AB452156	*Shorea slootenii*
171	gi 22034098	AY026578	*Shorea smithiana*
172	gi 226236746	AB452157	*Shorea smithiana*
173	gi 226236749	AB452160	*Shorea smithiana*
174	gi 22034099	AY026579	*Shorea splendida*
175	gi 226236755	AB452166	*Shorea splendida*

145	gi 226236723	AB452134	*Shorea platyclados*
146	gi 226236724	AB452135	*Shorea platyclados*
147	gi 226236725	AB452136	*Shorea pubistyla*
148	gi 226236726	AB452137	*Shorea quadrinervis*
149	gi 226236727	AB452138	*Shorea quadrinervis*
150	gi 226236728	AB452139	*Shorea quadrinervis*
151	gi 226236729	AB452140	*Shorea resinosa*
152	gi 22034092	AY026572	*Shorea richetia*
153	gi 22034093	AY026573	*Shorea roxburghii*
154	gi 226236730	AB452141	*Shorea roxburghii*
155	gi 226236733	AB452144	*Shorea roxburghii*
156	gi 226236734	AB452145	*Shorea rubra*
157	gi 226236735	AB452146	*Shorea rubra*
158	gi 226236736	AB452147	*Shorea rubra*
159	gi 226236737	AB452148	*Shorea rugosa*
160	gi 22034094	AY026574	*Shorea scaberrima*
161	gi 226236738	AB452149	*Shorea scaberrima*
176	gi 226236756	AB452167	*Shorea splendida*
177	gi 22034100	AY026580	*Shorea stenoptera*
178	gi 226236757	AB452168	*Shorea stenoptera*
179	gi 226236758	AB452169	*Shorea stenoptera*
180	gi 226236764	AB452175	*Shorea sumatrana*
181	gi 226236765	AB452176	*Shorea sumatrana*
182	gi 226236763	AB452174	*Shorea sumatrana*
183	gi 226236766	AB452177	*Shorea superba*
184	gi 226236767	AB452178	*Shorea superba*
185	gi 226236768	AB452179	*Shorea superba*
186	gi 226236769	AB452180	*Shorea symingtonii*
187	gi 226236770	AB452181	*Shorea teysmanniana*
188	gi 22034101	AY026581	*Shorea virescens*
189	gi 226236772	AB452183	*Shorea virescens*
190	gi 226236773	AB452184	*Shorea virescens*
191	gi 226236775	AB452186	*Shorea xanthophylla*

Appendix 1. List of plant species and corresponding GenBank accession numbers retrieved from the database for *psb*C-*trn*S IGS

No	GeneBank Identifier no	Accession number	Species
1	gi 226237206	AB452617	*Anisoptera laevis*
2	gi 226237226	AB452637	*Cotylelobium lanceolatum*
3	gi 226237258	AB452669	*Hopea dryobalanoides*
4	gi 226237260	AB452671	*Hopea mengarawan*
5	gi 226237261	AB452672	*Hopea mengarawan*
6	gi 226237262	AB452673	*Hopea mengarawan*
7	gi 226237312	AB452723	*Neobalanocarpus heimii*
8	gi 226237192	AB452603	*Shorea acuminata*
9	gi 226237193	AB452604	*Shorea acuminata*
10	gi 226237194	AB452605	*Shorea acuminata*
11	gi 226237195	AB452606	*Shorea acuminata*
12	gi 226237196	AB452607	*Shorea acuminatissima*
13	gi 226237198	AB452609	*Shorea acuminatissima*
14	gi 226237199	AB452610	*Shorea acuminatissima*
15	gi 226237197	AB452608	*Shorea acuminatissima*
16	gi 226237200	AB452611	*Shorea acuta*
17	gi 226237201	AB452612	*Shorea acuta*
18	gi 226237202	AB452613	*Shorea acuta*
19	gi 226237203	AB452614	*Shorea acuta*
20	gi 226237204	AB452615	*Shorea agami*
21	gi 226237205	AB452616	*Shorea agami*
22	gi 208609665	AB458534	*Shorea albida*
23	gi 208609669	AB458538	*Shorea albida*
24	gi 226237207	AB452618	*Shorea almon*
25	gi 226237208	AB452619	*Shorea almon*

No	GeneBank Identifier no	Accession number	Species
37	gi 226237220	AB452631	*Shorea biawak*
38	gi 226237221	AB452632	*Shorea biawak*
39	gi 226237222	AB452633	*Shorea biawak*
40	gi 226237223	AB452634	*Shorea bracteolata*
41	gi 226237224	AB452635	*Shorea bracteolata*
42	gi 226237225	AB452636	*Shorea bullata*
43	gi 226237227	AB452638	*Shorea collina*
44	gi 226237228	AB452639	*Shorea confusa*
45	gi 226237229	AB452640	*Shorea confusa*
46	gi 226237230	AB452641	*Shorea confusa*
47	gi 226237231	AB452642	*Shorea crassa*
48	gi 226237232	AB452643	*Shorea curtisii*
49	gi 226237233	AB452644	*Shorea curtisii*
50	gi 226237234	AB452645	*Shorea dasyphylla*
51	gi 226237235	AB452646	*Shorea dasyphylla*
52	gi 226237236	AB452647	*Shorea domatiosa*
53	gi 226237237	AB452648	*Shorea faguetiana*
54	gi 226237238	AB452649	*Shorea faguetiana*
55	gi 226237239	AB452650	*Shorea faguetiana*
56	gi 226237240	AB452651	*Shorea faguetioides*
57	gi 226237241	AB452652	*Shorea falcifera*
58	gi 226237242	AB452653	*Shorea falciferoides*
59	gi 226237243	AB452654	*Shorea fallax*
60	gi 226237244	AB452655	*Shorea fallax*
61	gi 226237245	AB452656	*Shorea fallax*

No	GeneBank Identifier no	Accession number	Species
26	gi 226237209	AB452620	*Shorea almon*
27	gi 226237210	AB452621	*Shorea amplexicaulis*
28	gi 226237211	AB452622	*Shorea amplexicaulis*
29	gi 226237212	AB452623	*Shorea andulensis*
30	gi 226237213	AB452624	*Shorea argentifolia*
31	gi 226237214	AB452625	*Shorea assamica*
32	gi 226237215	AB452626	*Shorea assamica*
33	gi 226237216	AB452627	*Shorea assamica*
34	gi 226237217	AB452628	*Shorea atrinervosa*
35	gi 226237218	AB452629	*Shorea atrinervosa*
36	gi 226237219	AB452630	*Shorea beccariana*
73	gi 226237257	AB452668	*Shorea havilandii*
74	gi 226237259	AB452670	*Shorea henryana*
75	gi 226237263	AB452674	*Shorea inappendiculata*
76	gi 226237264	AB452675	*Shorea isoptera*
77	gi 226237265	AB452676	*Shorea javanica*
78	gi 226237266	AB452677	*Shorea johorensis*
79	gi 226237267	AB452678	*Shorea johorensis*
80	gi 226237268	AB452679	*Shorea johorensis*
81	gi 226237269	AB452680	*Shorea johorensis*
82	gi 226237270	AB452681	*Shorea johorensis*
83	gi 226237271	AB452682	*Shorea johorensis*
84	gi 226237272	AB452683	*Shorea johorensis*
85	gi 226237273	AB452684	*Shorea kunstleri*
86	gi 226237274	AB452685	*Shorea kunstleri*
87	gi 226237275	AB452686	*Shorea kunstleri*
88	gi 226237276	AB452687	*Shorea laevis*
89	gi 226237277	AB452688	*Shorea laevis*
90	gi 226237278	AB452689	*Shorea laevis*

No	GeneBank Identifier no	Accession number	Species
62	gi 226237246	AB452657	*Shorea fallax*
63	gi 226237247	AB452658	*Shorea fallax*
64	gi 226237248	AB452659	*Shorea fallax*
65	gi 226237249	AB452660	*Shorea ferruginea*
66	gi 226237250	AB452661	*Shorea ferruginea*
67	gi 226237251	AB452662	*Shorea flaviflora*
68	gi 226237252	AB452663	*Shorea foxworthyi*
69	gi 226237253	AB452664	*Shorea gibbosa*
70	gi 226237254	AB452665	*Shorea guiso*
71	gi 226237255	AB452666	*Shorea havilandii*
72	gi 226237256	AB452667	*Shorea havilandii*
109	gi 226237288	AB452699	*Shorea macroptera subsp. baillonii*
110	gi 226237289	AB452700	*Shorea macroptera subsp. baillonii*
111	gi 226237290	AB452701	*Shorea macroptera subsp. macropterifolia*
112	gi 226237291	AB452702	*Shorea macroptera subsp. macropterifolia*
113	gi 226237301	AB452712	*Shorea macroptera subsp. sandakanensis*
114	gi 226237302	AB452713	*Shorea macroptera subsp. sandakanensis*
115	gi 226237304	AB452715	*Shorea materialis*
116	gi 226237305	AB452716	*Shorea materialis*
117	gi 226237303	AB452714	*Shorea materialis*
118	gi 226237306	AB452717	*Shorea maxima*
119	gi 226237307	AB452718	*Shorea maxima*
120	gi 226237308	AB452719	*Shorea maxwelliana*
121	gi 226237309	AB452720	*Shorea mecistopteryx*
122	gi 226237310	AB452721	*Shorea mecistopteryx*
123	gi 226237311	AB452722	*Shorea mujongensis*
124	gi 226237313	AB452724	*Shorea obscura*
125	gi 226237314	AB452725	*Shorea ochracea*
126	gi 226237315	AB452726	*Shorea ochracea*

No	GeneBank Identifier no	Accession number	Species
91	gi 226237279	AB452690	*Shorea lepidota*
92	gi 226237280	AB452691	*Shorea leprosula*
93	gi 226237281	AB452692	*Shorea leprosula*
94	gi 226237282	AB452693	*Shorea leprosula*
95	gi 226237283	AB452694	*Shorea leprosula*
96	gi 226237284	AB452695	*Shorea leprosula*
97	gi 226237285	AB452696	*Shorea longiflora*
98	gi 226237286	AB452697	*Shorea longiflora*
99	gi 226237287	AB452698	*Shorea longisperma*
100	gi 226237292	AB452703	*Shorea macrophylla*
101	gi 226237293	AB452704	*Shorea macrophylla*
102	gi 226237294	AB452705	*Shorea macrophylla*
103	gi 226237295	AB452706	*Shorea macrophylla*
104	gi 226237296	AB452707	*Shorea macrophylla*
105	gi 226237297	AB452708	*Shorea macroptera*
106	gi 226237298	AB452709	*Shorea macroptera*
107	gi 226237299	AB452710	*Shorea macroptera*
108	gi 226237300	AB452711	*Shorea macroptera*
145	gi 226237334	AB452745	*Shorea patoiensis*
146	gi 226237335	AB452746	*Shorea patoiensis*
147	gi 226237336	AB452747	*Shorea pauciflora*
148	gi 226237337	AB452748	*Shorea pauciflora*
149	gi 226237338	AB452749	*Shorea pauciflora*
150	gi 226237339	AB452750	*Shorea pauciflora*
151	gi 226237340	AB452751	*Shorea peltata*
152	gi 226237341	AB452752	*Shorea pilosa*
153	gi 226237342	AB452753	*Shorea pilosa*
154	gi 226237343	AB452754	*Shorea pilosa*
155	gi 226237344	AB452755	*Shorea pinanga*

No	GeneBank Identifier no	Accession number	Species
127	gi 226237316	AB452727	*Shorea ochracea*
128	gi 226237317	AB452728	*Shorea ochrophloia*
129	gi 226237318	AB452729	*Shorea ovalis*
130	gi 226237319	AB452730	*Shorea ovalis*
131	gi 226237320	AB452731	*Shorea ovalis*
132	gi 226237321	AB452732	*Shorea ovata*
133	gi 226237322	AB452733	*Shorea palembanica*
134	gi 226237323	AB452734	*Shorea palosapis*
135	gi 226237324	AB452735	*Shorea parvifolia*
136	gi 226237325	AB452736	*Shorea parvifolia*
137	gi 226237326	AB452737	*Shorea parvifolia*
138	gi 226237327	AB452738	*Shorea parvifolia*
139	gi 226237328	AB452739	*Shorea parvifolia*
140	gi 226237329	AB452740	*Shorea parvifolia*
141	gi 226237330	AB452741	*Shorea parvistipulata*
142	gi 226237331	AB452742	*Shorea parvistipulata*
143	gi 226237332	AB452743	*Shorea patoiensis*
144	gi 226237333	AB452744	*Shorea patoiensis*
178	gi 226237367	AB452778	*Shorea slootenii*
179	gi 226237368	AB452779	*Shorea slootenii*
180	gi 226237369	AB452780	*Shorea slootenii*
181	gi 226237376	AB452787	*Shorea smithiana*
182	gi 226237378	AB452789	*Shorea smithiana*
183	gi 226237370	AB452781	*Shorea smithiana*
184	gi 226237371	AB452782	*Shorea smithiana*
185	gi 226237372	AB452783	*Shorea smithiana*
186	gi 226237373	AB452784	*Shorea smithiana*
187	gi 226237374	AB452785	*Shorea smithiana*
188	gi 226237375	AB452786	*Shorea smithiana*

No	GeneBank Identifier no	Accession number	Species
156	gi 226237345	AB452756	*Shorea platycarpa*
157	gi 226237346	AB452757	*Shorea platyclados*
158	gi 226237347	AB452758	*Shorea platyclados*
159	gi 226237348	AB452759	*Shorea platyclados*
160	gi 226237349	AB452760	*Shorea pubistyla*
161	gi 226237350	AB452761	*Shorea quadrinervis*
162	gi 226237352	AB452763	*Shorea quadrinervis*
163	gi 226237351	AB452762	*Shorea quadrinervis*
164	gi 226237353	AB452764	*Shorea resinosa*
165	gi 226237354	AB452765	*Shorea roxburghii*
166	gi 226237355	AB452766	*Shorea roxburghii*
167	gi 226237356	AB452767	*Shorea roxburghii*
168	gi 226237357	AB452768	*Shorea roxburghii*
169	gi 226237358	AB452769	*Shorea rubra*
170	gi 226237359	AB452770	*Shorea rubra*
171	gi 226237360	AB452771	*Shorea rubra*
172	gi 226237361	AB452772	*Shorea rugosa*
173	gi 226237362	AB452773	*Shorea scaberrima*
174	gi 226237363	AB452774	*Shorea seminis*
175	gi 226237364	AB452775	*Shorea singkawang*
176	gi 226237365	AB452776	*Shorea singkawang*
177	gi 226237366	AB452777	*Shorea singkawang*

No	GeneBank Identifier no	Accession number	Species
189	gi 226237377	AB452788	*Shorea smithiana*
190	gi 226237379	AB452790	*Shorea splendida*
191	gi 226237380	AB452791	*Shorea splendida*
192	gi 226237381	AB452792	*Shorea stenoptera*
193	gi 226237382	AB452793	*Shorea stenoptera*
194	gi 226237383	AB452794	*Shorea stenoptera*
195	gi 226237384	AB452795	*Shorea stenoptera*
196	gi 226237385	AB452796	*Shorea stenoptera*
197	gi 226237386	AB452797	*Shorea stenoptera*
198	gi 226237387	AB452798	*Shorea sumatrana*
199	gi 226237388	AB452799	*Shorea sumatrana*
200	gi 226237389	AB452800	*Shorea sumatrana*
201	gi 226237390	AB452801	*Shorea superba*
202	gi 226237391	AB452802	*Shorea superba*
203	gi 226237392	AB452803	*Shorea superba*
204	gi 226237393	AB452804	*Shorea symingtonii*
205	gi 226237394	AB452805	*Shorea teysmanniana*
206	gi 226237396	AB452807	*Shorea virescens*
207	gi 226237397	AB452808	*Shorea virescens*
208	gi 226237399	AB452810	*Shorea xanthophylla*
209	gi 226237395	AB452806	*Vatica bella*
210	gi 226237398	AB452809	*Vatica oblongifolia*

Appendix 1. List of plant species and corresponding GenBank accession numbers retrieved from the database for *rbc*L

No	GeneBank Identifier no	Accession number	Species
1	gi 2897113	AF030238	*Pseudomonotes tropenbosii*
2	gi 14595085	AJ247623	*Shorea talura*
3	gi 37790902	AY328198	*Hopea hainanensis*
4	gi 2654338	Y15144	*Anisoptera marginata*
5	gi 37790904	AY328199	*Vatica mangachapoi*

Appendix 1. List of plant species and corresponding GenBank accession numbers retrieved from the database for *mat*K

No	GeneBank Identifier no	Accession no	Species
1	gi 292679842	AB295878	*Anisoptera laevis*
2	gi 292679844	AB295879	*Anisoptera marginata*
3	gi 292679846	AB295880	*Anisoptera oblonga*
4	gi 292679848	AB295881	*Cotylelobium malayanum*
5	gi 292679850	AB295882	*Cotylelobium malayanum*
6	gi 292679852	AB295883	*Cotylelobium scabriusculum*
7	gi 292679854	AB295884	*Dipterocarpus alatus*
8	gi 292679856	AB295885	*Dipterocarpus baudii*
9	gi 292679858	AB295886	*Dipterocarpus cornutus*
10	gi 292679860	AB295887	*Dipterocarpus glandulosus*
11	gi 292679862	AB295888	*Dipterocarpus hispidus*
12	gi 292679864	AB295889	*Dipterocarpus insignis*
13	gi 292679866	AB295890	*Dipterocarpus kerrii*
14	gi 292679868	AB295891	*Dipterocarpus palembanicus*
15	gi 292679870	AB295892	*Dipterocarpus zeylanicus*
16	gi 292679872	AB295893	*Dryobalanops aromatica*
17	gi 292679874	AB295894	*Dryobalanops oblongifolia*
18	gi 292679876	AB295895	*Hopea discolor*
19	gi 292679878	AB295896	*Hopea helferi*
20	gi 292679880	AB295897	*Hopea jucunda*
21	gi 292679882	AB295898	*Hopea jucunda subsp. modesta*
22	gi 292679884	AB295899	*Hopea latifolia*
23	gi 292679886	AB295900	*Hopea nervosa*
24	gi 292679888	AB295901	*Hopea odorata*

No	GeneBank Identifier no	Accession no	Species
36	gi 4210561	AB006376	*Shorea bullata*
37	gi 4210562	AB006377	*Shorea bullata*
38	gi 4210563	AB006378	*Shorea bullata*
39	gi 4210564	AB006379	*Shorea bullata*
40	gi 4210565	AB006380	*Shorea bullata*
41	gi 4210566	AB006381	*Shorea bullata*
42	gi 4210567	AB006382	*Shorea bullata*
43	gi 4210568	AB006383	*Shorea bullata*
44	gi 4210570	AB006384	*Shorea congestiflora*
45	gi 4210571	AB006385	*Shorea cordifolia*
46	gi 71891362	AJ581409	*Shorea curtisii*
47	gi 94966499	AB246414	*Shorea disticha*
48	gi 94966501	AB246415	*Shorea dyeri*
49	gi 94966503	AB246416	*Shorea elliptica*
50	gi 94966505	AB246417	*Shorea fallax*
51	gi 94966507	AB246418	*Shorea fallax*
52	gi 94966509	AB246419	*Shorea fallax*
53	gi 94966511	AB246420	*Shorea fallax*
54	gi 94966513	AB246421	*Shorea fallax*
55	gi 94966515	AB246422	*Shorea fallax*
56	gi 94966517	AB246423	*Shorea fallax*
57	gi 94966519	AB246424	*Shorea fallax*
58	gi 94966523	AB246426	*Shorea gardneri*
59	gi 94966525	AB246427	*Shorea kunstleri*

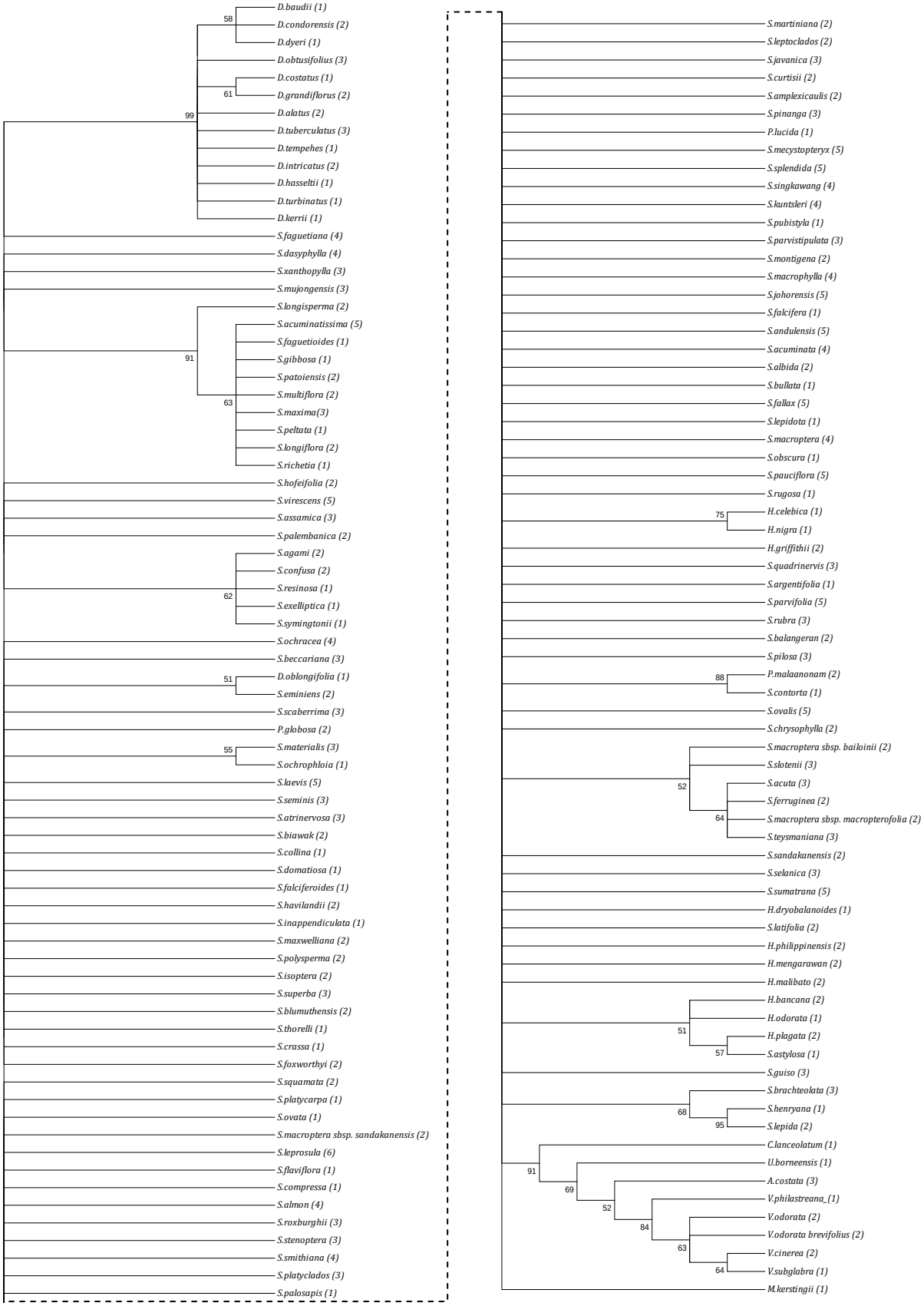
No	GeneBank Identifier no	Accession no	Species
25	gi 292679890	AB295902	*Hopea subalata*
26	gi 292679892	AB295903	*Hopea wightiana*
27	gi 34597658	AY305717	*Monotes madagascariensis*
28	gi 34597658	AY305717	*Neobalanocarpus heimii*
29	gi 34597660	AY305718	*Parashorea chinensis*
30	gi 4210551	AB006370	*Parashorea chinensis var. kwangsiensis*
31	gi 4210555	AB006371	*Parashorea lucida*
32	gi 4210556	AB006372	*Shorea acuminata*
33	gi 4210558	AB006373	*Shorea affinis*
34	gi 4210559	AB006374	*Shorea assamica*
35	gi 4210560	AB006375	*Shorea bracteolata*
71	gi 94966549	AB246439	*Shorea multiflora*
72	gi 94966551	AB246440	*Shorea ovalifolia*
73	gi 94966553	AB246441	*Shorea ovalis*
74	gi 94966555	AB246442	*Shorea pallescens*
75	gi 94966557	AB246443	*Shorea parvifolia*
76	gi 94966559	AB246444	*Shorea pinanga*
77	gi 94966561	AB246445	*Shorea quadrinervis*
78	gi 94966563	AB246446	*Shorea richetia*
79	gi 94966565	AB246447	*Shorea seminis*
80	gi 94966569	AB246449	*Shorea smithiana*
81	gi 94966571	AB246450	*Shorea smithiana*
82	gi 94966573	AB246451	*Shorea smithiana*
83	gi 94966575	AB246452	*Shorea smithiana*
84	gi 94966577	AB246453	*Shorea smithiana*
85	gi 94966579	AB246454	*Shorea smithiana*
86	gi 94966581	AB246455	*Shorea splendens*

No	GeneBank Identifier no	Accession no	Species
60	gi 94966527	AB246428	*Shorea kunstleri*
61	gi 94966529	AB246429	*Shorea kunstleri*
62	gi 94966531	AB246430	*Shorea kunstleri*
63	gi 94966533	AB246431	*Shorea kunstleri*
64	gi 94966535	AB246432	*Shorea laevis*
65	gi 94966537	AB246433	*Shorea leprosula*
66	gi 94966539	AB246434	*Shorea lissophylla*
67	gi 94966541	AB246435	*Shorea macrophylla*
68	gi 94966543	AB246436	*Shorea macroptera*
69	gi 94966545	AB246437	*Shorea macroptera*
70	gi 94966547	AB246438	*Shorea megistophylla*
91	gi 94966591	AB246460	*Shorea zeylanica*
92	gi 94966593	AB246461	*Stemonoporus acuminatus*
93	gi 94966595	AB246462	*Stemonoporus bullatus*
94	gi 94966597	AB246463	*Stemonoporus gilimalensis*
95	gi 94966599	AB246464	*Stemonoporus kanneliyensis*
96	gi 94966601	AB246465	*Stemonoporus lancifolius*
97	gi 94966603	AB246466	*Stemonoporus reticulatus*
98	gi 94966605	AB246467	*Stemonoporus scalarinervis*
99	gi 94966607	AB246468	*Stemonoporus wightii*
100	gi 94966609	AB246469	*Upuna borneensis*
101	gi 94966611	AB246470	*Vateria copallifera*
102	gi 94966613	AB246471	*Vateriopsis seychellarum*
103	gi 94966615	AB246472	*Vatica affinis*
104	gi 94966617	AB246473	*Vatica bella*
105	gi 94966619	AB246474	*Vatica chinensis*
106	gi 94966621	AB246475	*Vatica coriacea*

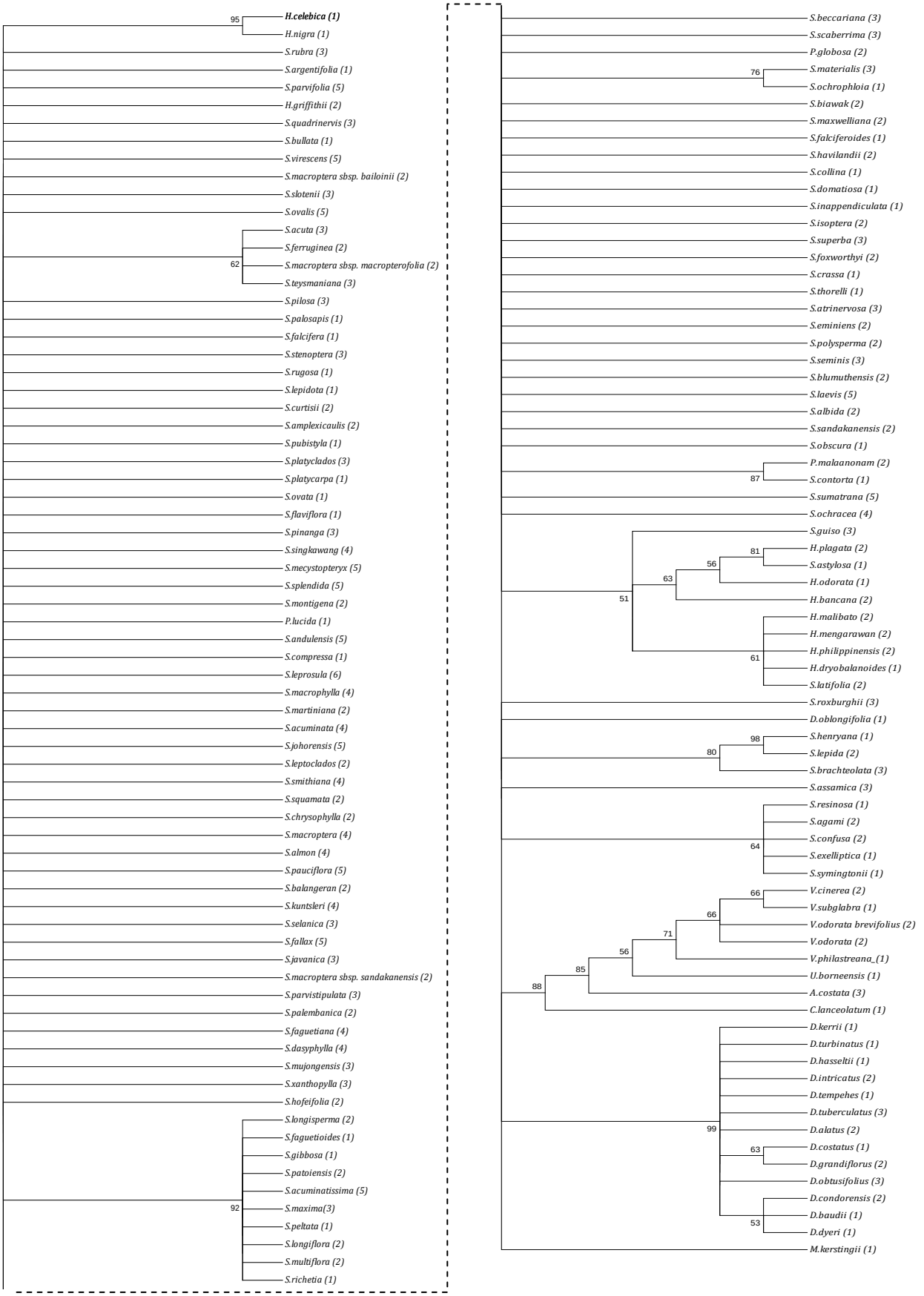
No	GeneBank Identifier no	Accession no	Species
87	gi 94966583	AB246456	*Shorea stipularis*
88	gi 94966585	AB246457	*Shorea trapezifolia*
89	gi 94966587	AB246458	*Shorea worthingtonii*
90	gi 94966589	AB246459	*Shorea xanthophylla*

No	GeneBank Identifier no	Accession no	Species
107	gi 94966623	AB246476	*Vatica micrantha*
108	gi 94966625	AB246477	*Vatica odorata*
109	gi 94966627	AB246478	*Vatica pauciflora*

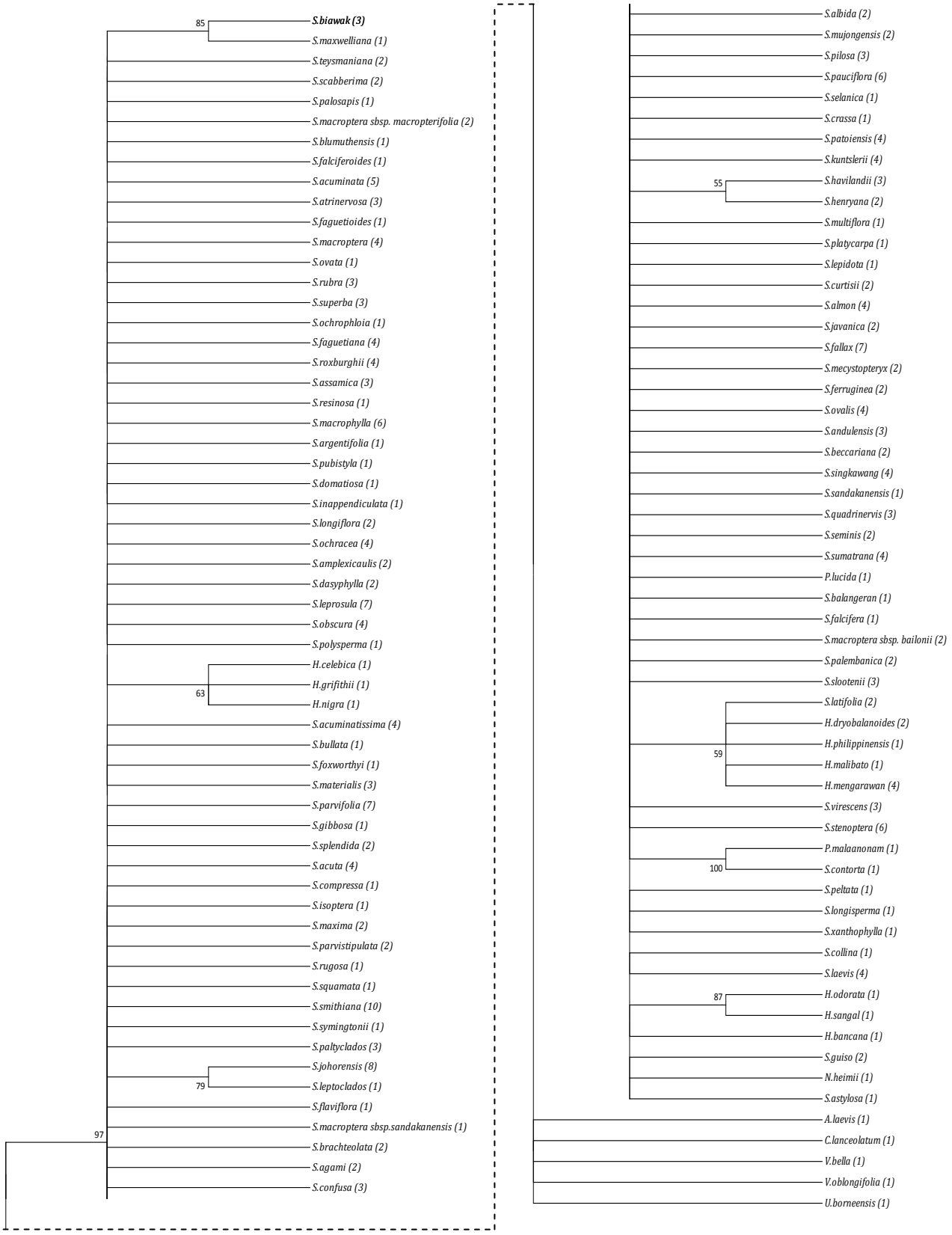
Appendix 2. The tree of *trn*L intron using Maximum Likelihood method based on the Kimura 2-parameter model. The percentage of bootstrap value is shown next to the branches. The analysis involved 145 nucleotide sequences. The number in bracket means number of species tested.



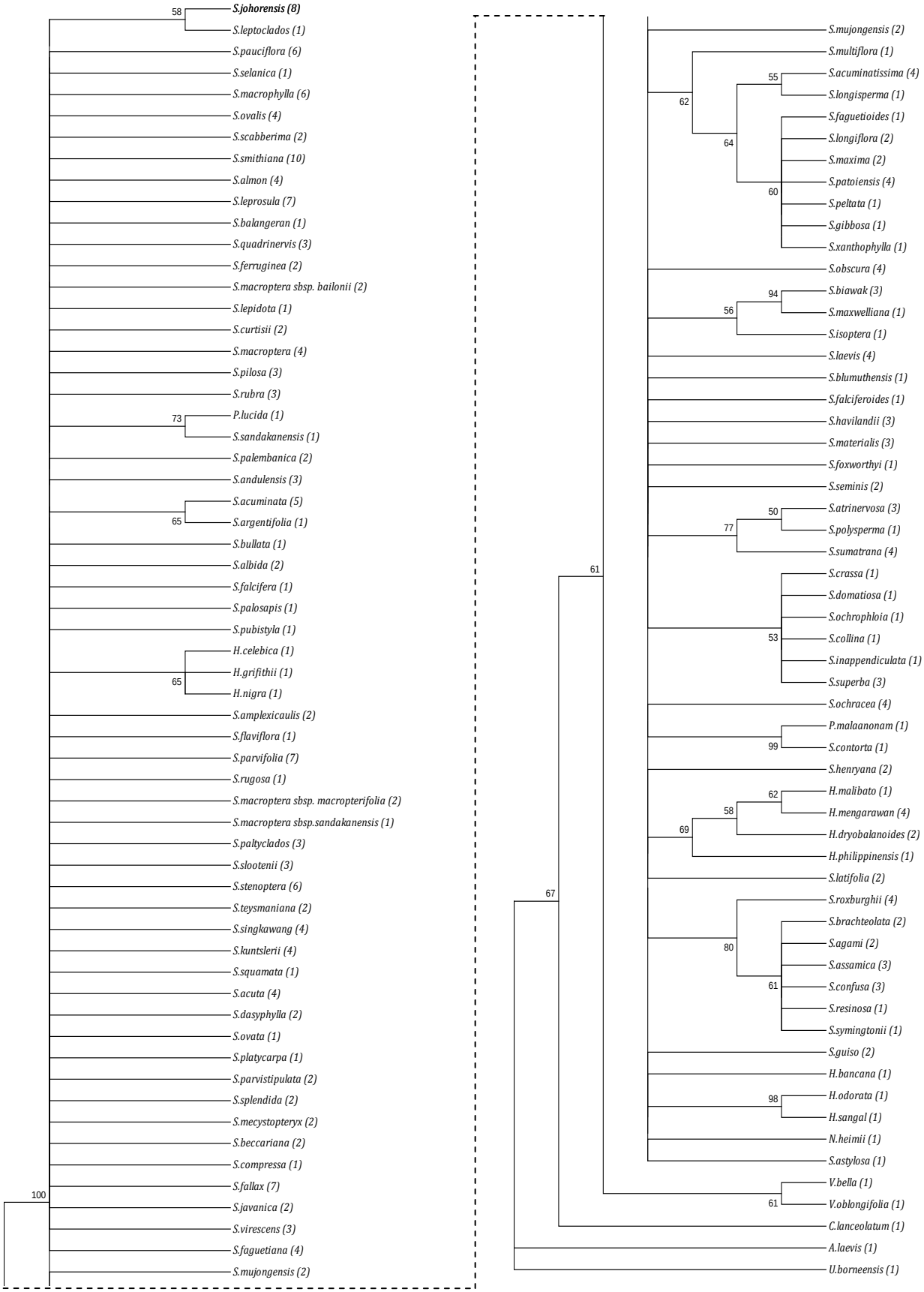
Appendix 3. The tree of *trn*L intron using the neighbor joining method and using the Kimura 2-parameter for genetic distance. The percentages of the bootstrap test (1000 replicates) are shown next to the branches. The analysis involved 145 nucleotide sequences. The number in bracket means number of species tested.



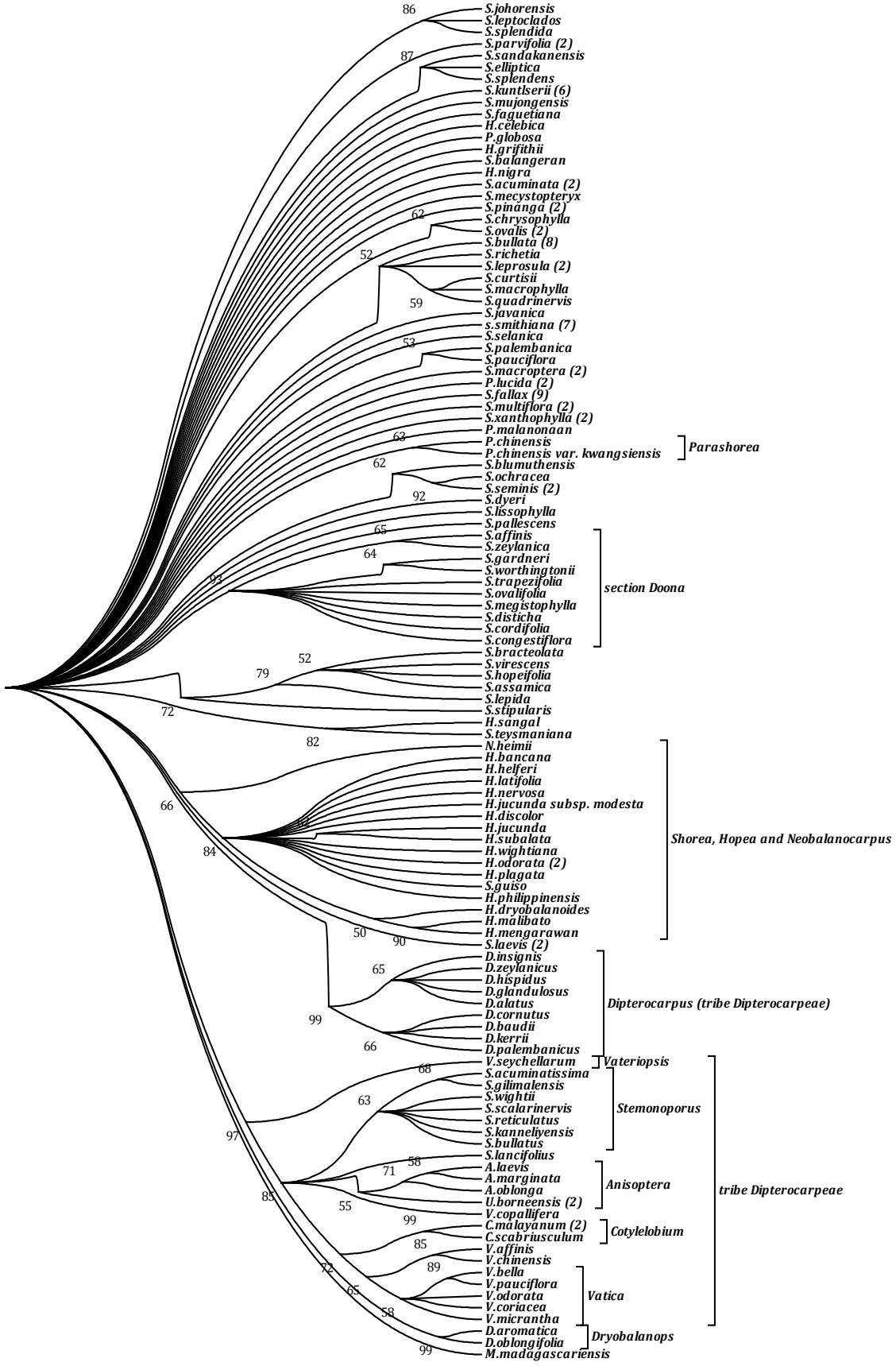
Appendix 4. The tree of *psb*C-*trn*S IGS using the maximum likelihood method based on the Kimura 2-parameter model. The percentage of the bootstrap value shown next to the branch. There were a total of 1137 positions in the final dataset. The number in parenthesis means number of species tested.



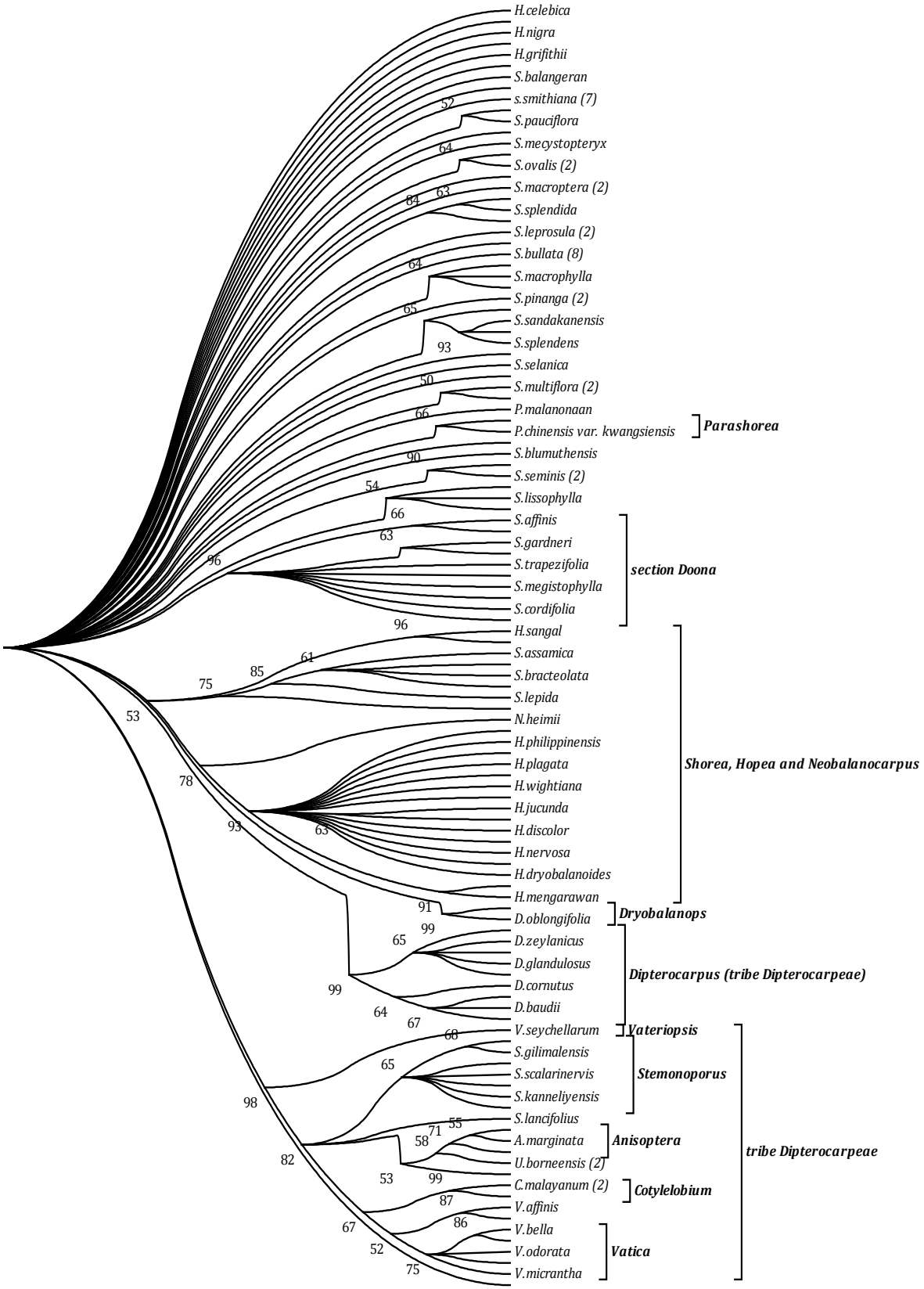
Appendix 5. The psbC-*trn*S IGS tree using neighbor joining method and Kimura 2-parameter. The percentages of bootstrap value are shown next to the branches. This analysis involved 117 nucleotide sequences and 1137 positions. The number in parentheses means number of species tested.



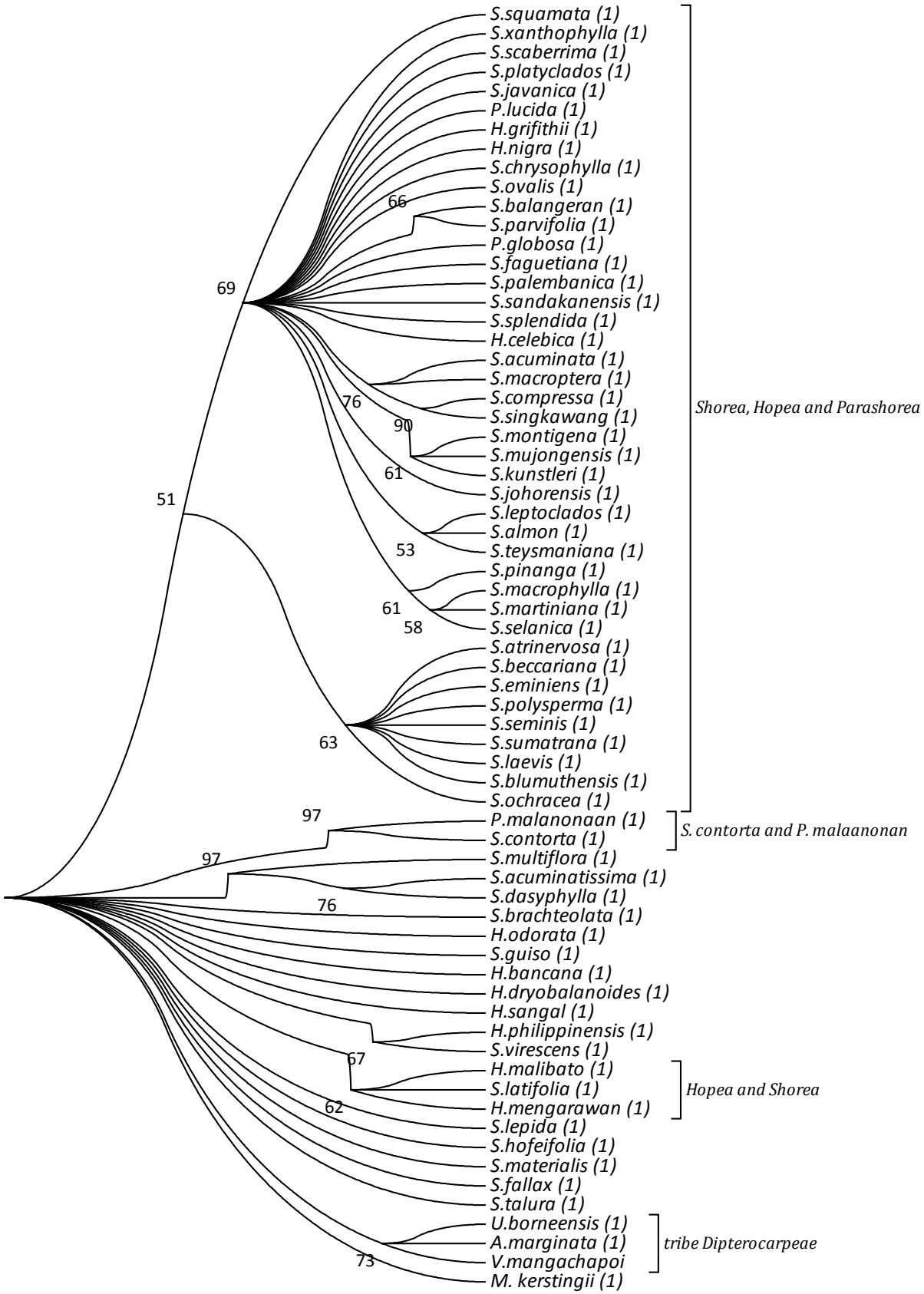
Appendix 6. The tree of *mat*K gene using Maximum Likelihood method based on the Kimura 2-parameter model. The percentage of bootstrap value is shown next to the branches. The analysis involved 116 nucleotide sequences of 635 positions in the final dataset. The number in parentheses means number of species tested.



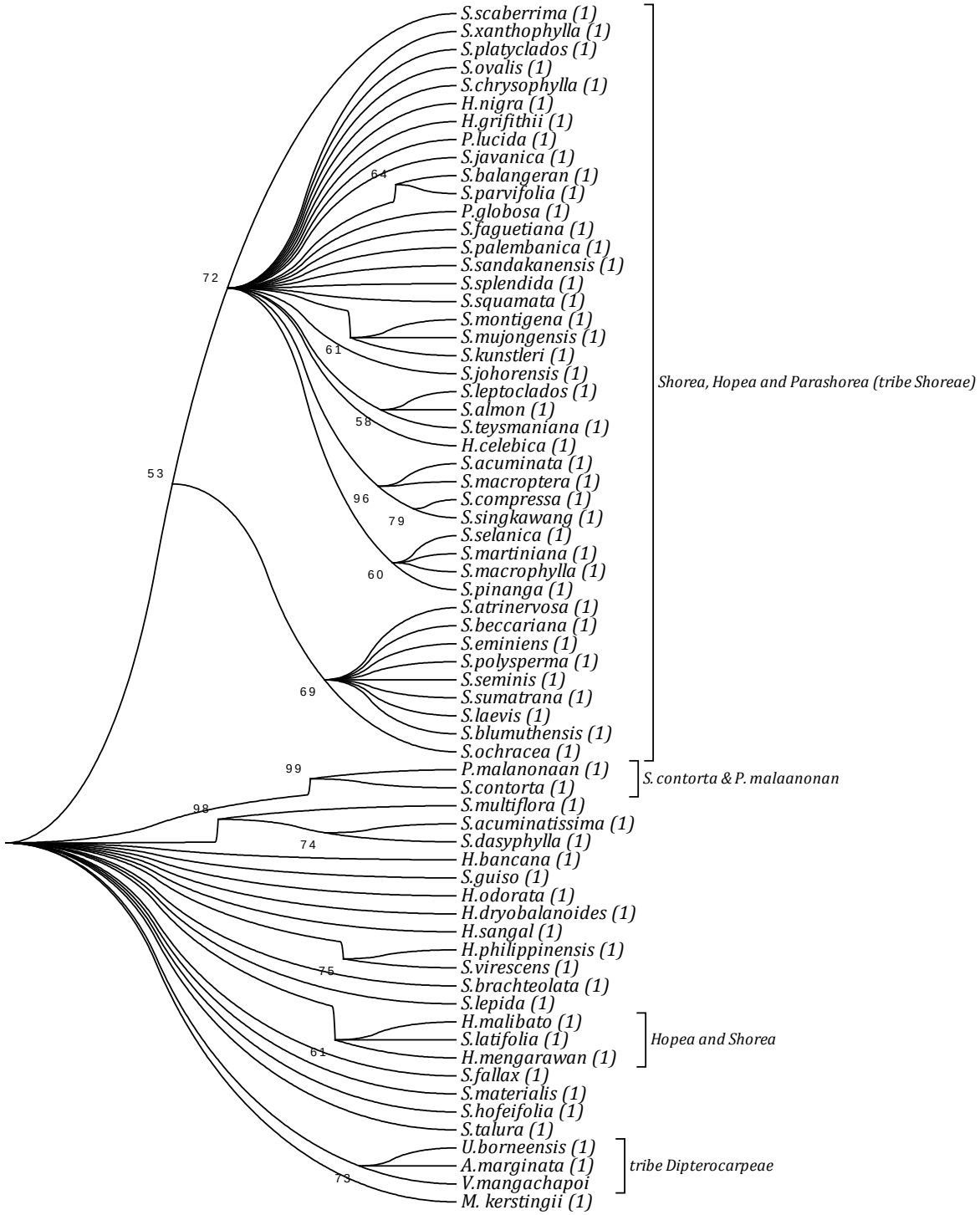
Appendix 7. The tree of *mat*K gene using the neighbor joining method and Kimura 2-parameter. The percentages of bootstrap value are shown next to the branches. The analysis involved 116 nucleotide sequences of 635 positions in the final dataset. The number in parentheses means number of species tested.



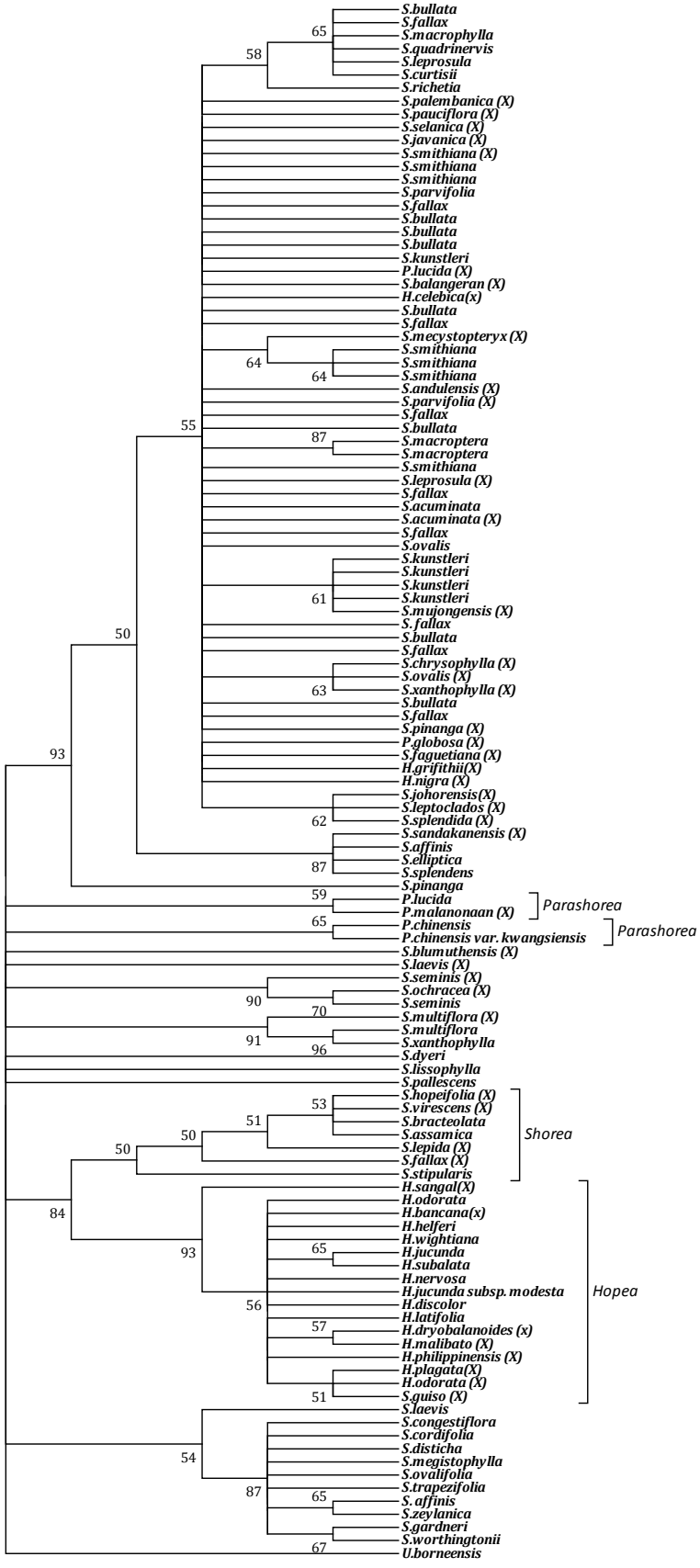
Appendix 8. The tree of *rbc*L gene using the maximum likelihood method based on the Kimura 2-parameter model. The percentage of bootstrap value is shown next to the branches. The analysis involved 67 nucleotide sequences. There were a total of 647 positions in the final dataset. The number in brackets is the number of species tested.



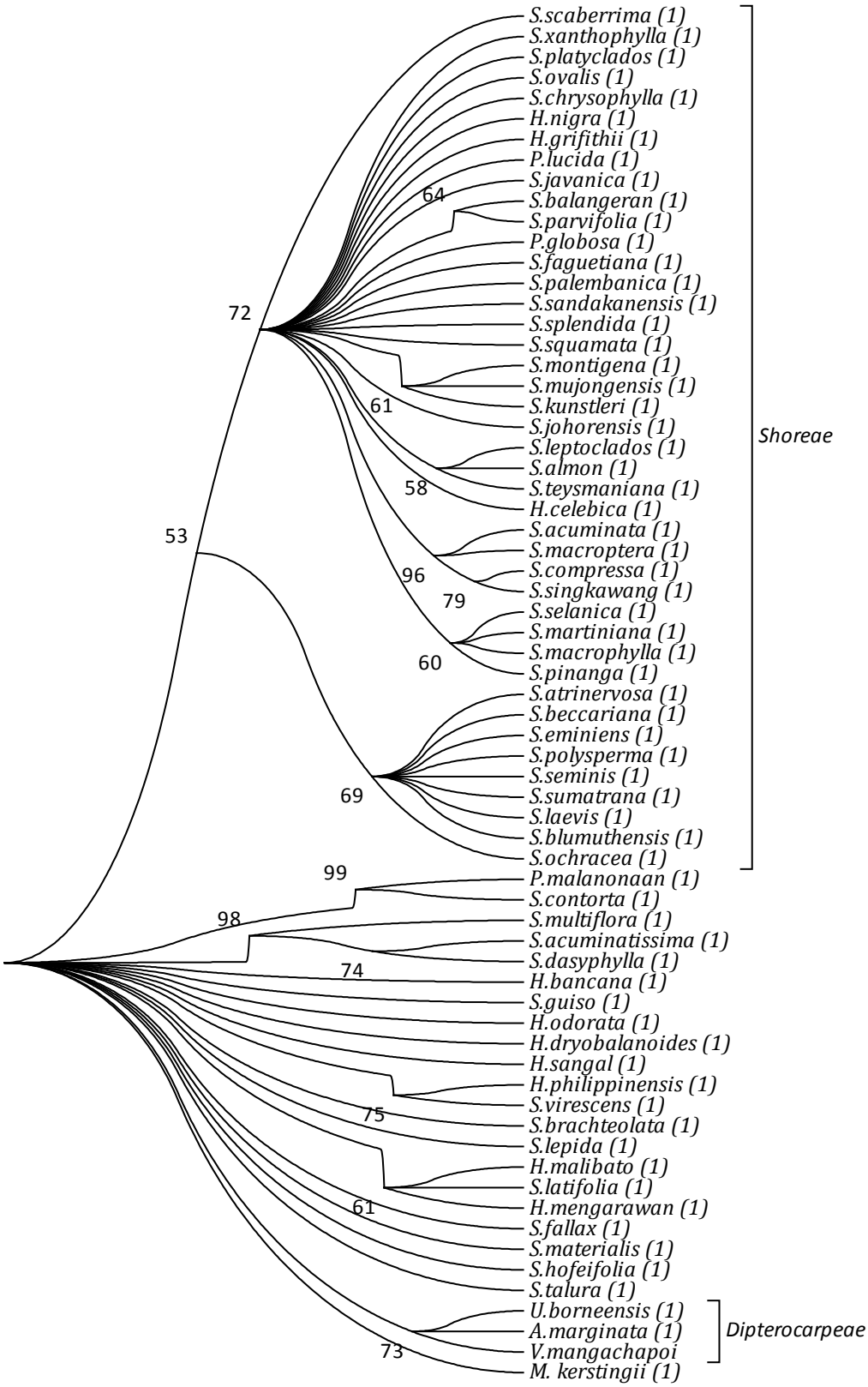
Appendix 9. The tree of *rbc*L gene using the neighbor joining method based on the Kimura 2-parameter. The percentages of bootstrap value are shown next to the branches. The analysis involved 67 nucleotide sequences. There were a total of 647 positions in the final dataset. The number in brackets is the number of species tested.



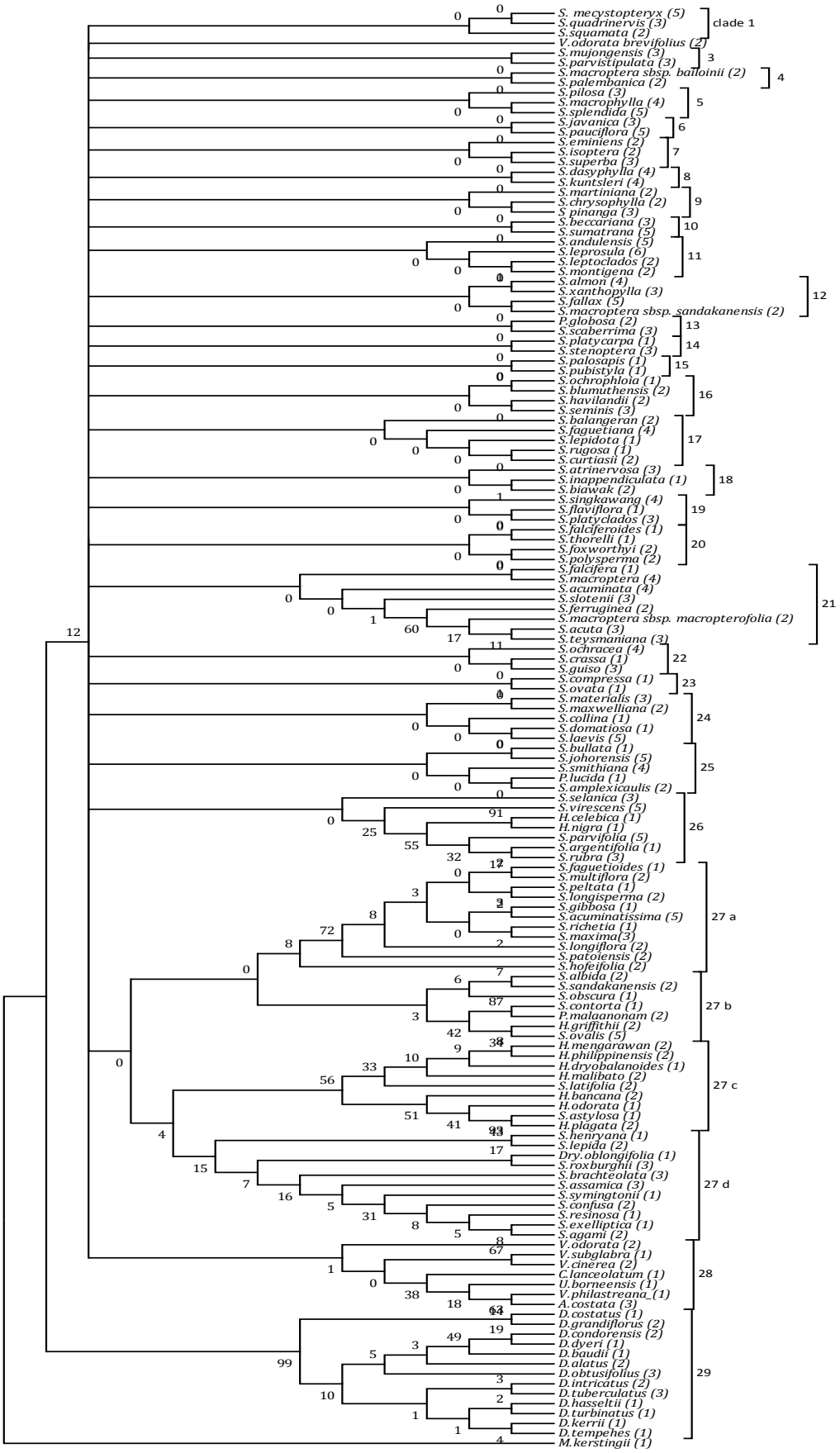
Appendix 10. The identification test of *mat*K (barcode region) using the neighbor joining analysis method with K2P formula as a parameter for genetic distance. The (X) label behind the species name indicating the sequences from laboratory examined in this study.



Appendix 11. The identification test of *rbc*L (barcode region) using the neighbor joining analysis method with K2P formula as a parameter for genetic distance. The number in the bracket indicated the species number tested.



Appendix 12. The cladogram of *trn*L intron tree using maximum parsimony analysis and the molecular taxonomic identification key based on the clades of the tree



The molecular taxonomic identification key based on clade of the *trn*L intron phylogenetic tree using maximum parsimony analysis.

Clade 1

Species	Position of polymorphic sites	
	52	212
S. squamata	-	G
S. mecystopteryx	A	G
S. quadrinervis	A	(R) A

1. a. Site 52 is (-) …… 2a
   b. If (A) …… 2bc
2. a. Site 212 is (G) …… *S. squamata*
   b. Site 212 is (G) …… *S. mecystopteryx*
   c. If (A) …… *S. quadrinervis*

Clade 2

V. odorata parvistipulata

Clade 3

There is no differentiation between *S. mujongensis* and *S. parvistipulata*

Clade 4

Species	Position of polymorphic sites			
	244	246	275	276
S. macroptera sbsp. bailonii	-	G	A	C
S. palembanica	A	R	M	Y

1. a. Site 244 is (-) …… 2a
   b. If (A) …… 2b

2. a. Site 246 is (G) …… 3a

   b. If (A) …… 3b

3. a. Site 275 is (A) …… 4a

   b. If (C) …… 4b

4. a. Site 276 is (C) …… *S. macroptera sbsp. Bailonii*

   b. If (Y) …… *S. palembanica*

Clade 5

Species	Position of polymorphic sites								
	22	51	150	165	172	176	180	266	474
S. macrophylla	A	-	T	T	A	T	T	T	T
S. pilosa	M	-	T	T	A	T	T	T	K
S. splendida	A	A	W	K	R	W	Y	K	T

1. a. Site 22 is (A) …… 2ab

   b. If (M) …… 2c

2. a. Site 51 is (-) …… 3a

   b. If (A) …… 3b

   c. If (-) …… 3c

3. a. Site 150 is (T) …… 4a

   b. If (W) …… 4b

   c. If (T) …… 4C

4. a. Site 165 is (T) …… 5a

   b. If (K) …… 5b

   c. If (T) …… 5c

5. a. Site 172 is (A) …… 6a

   b. If (R) …… 6b

   c. If (A) …… 6c

6. a. Site 176 is (T) …… 7a

	b.	If (W)	……	7b
	c.	If (T)	……	7c
7.	a.	Site 180 is (T)	……	8a
	b.	If (Y)	……	8b
	c.	If (T)	……	8c
8.	a.	Site 266 is (T)	……	9a
	b.	If (K)	……	9b
	c.	If (T)	……	9c
9.	a.	Site 474 is (T)	……	*S. macrophylla*
	b.	If (T)	……	*S. splendida*
	c.	If (K)	……	*S. pilosa*

Clade 6

There is no differentiation between *S. javanica* and *S. pauciflora*

Clade 7

Species	**Position of polymorphic sites**		
	184	404	502
S. eminiens	T	A	-
S. isoptera	G	C	C
S. superba	T	C	-

1.	a.	Site 184 is (T)	……	2ab
	b.	If (G)	……	2c
2.	a.	Site 404 is (A)	……	3a
	b.	If (C)	……	3b
	c.	If (C)	……	3c
3.	a.	Site 502 is (-)	……	*S. eminiens*
	b.	If (-)	……	*S. superba*
	c.	If (C)	……	*S. isoptera*

Clade 8

Species	Position of polymorphic sites		
	51	52	195
S. chrysophylla	A	A	C
S. pinanga	A	-	G
S. martiniana	-	-	G

1. a. Site 53 is (-) …… 2a
   b. If (A) …… 2b
2. a. Site 60 is (A) …… *S. dasyphylla*
   b. If (W) …… *S. Kuntslerii*

Clade 9

Species	Position of polymorphic sites	
	53	60
S. dasyphylla	-	A
S. kuntslerii	A	W

1. a. Site 51 is (A) …… 2ab
   b. If (-) …… 2c
2. a. Site 52 is (A) …… 3a
   b. If (-) …… 3b
   c. If (-) …… 3c
3. a. Site 195 is (C) …… *S. chrysophylla*
   b. If (G) …… *S. pinanga*
   c. If (G) …… *S. martiniana*

Clade 10

Species	Position of polymorphic sites						
	7	81	194	244	276	344	375
S. beccariana	A	A	A	-	C	T	A
S. sumatrana	R	M	R	A	Y	K	M

1. a. Site 7 is (A) …… 2a
   b. If (G) …… 2b
2. a. Site 81 is (M) …… 3a
   b. If (C) …… 3b
3. a. Site 194 is (R) …… 4ab
   b. If (G) …… 4b
4. a. Site 244 is (-) …… 5a
   b. If (A) …… 5b
5. a. Site 276 is (C) …… 6a
   b. If (Y) …… 6b
6. a. Site 344 is (T) …… 7a
   b. If (K) …… 7b
7. a. Site 375 is (A) …… *S. beccariana*
   b. If (M) …… *S. sumatrana*

Clade 11

Species	Position of polymorphic sites	
	51	258
S. andulensis	-	-
S. leprosula	-	-
S. leptoclados	A	-
S. montigena	-	A

1. a. Site 51 is (-) …… 2ab
   b. If (A) …… 2c
2. a. Site 258 is (-) …… *S. andulensis*, *S. leprosula*
   b. If (A) …… *S. montigena*
   c. If (-) …… *S. leptoclados*

Clade 12

Species	Position of polymorphic sites					
	52	178	202	244	276	287
S. almon	-	C	T	-	C	T
S. xanthophylla	-	Y/A	K/G	-	C	K/G
S. fallax	A	C	T	A	Y/T	T
S. macroptera	A	C	T	-	C	T

1. a. Site 52 is (-) …… 2ab
   b. If (A) …… 2c
2. a. Site 178 is (C) …… 3a
   b. If (T) …… 3b
   c. If (C) …… 3c
3. a. Site 202 is (K) …… 4a
   b. If (G) …… 4b
   c. If (T) …… 4cd

4. a. Site 244 is (-) …… 5a

   b. If (-) …… 5b

   c. If (A) …… 5c

   d. If (-) …… 5d

5. a. Site 276 is (C) …… 6ab

   b. If (C) …… 6b

   c. If (Y) …… 6c

   d. If (C) …… 6d

6. a. Site 287 is (T) …… *S. almon, S. xanthophylla*

   b. If (G) …… *S. xanthophylla*

   c. If (T) …… *S. fallax*

   d. If (T) …… *S. macroptera*

Clade 13

Species	**Position of polymorphic sites**			
	51	52	53	212
P. globosa	-	-	-	G
S. scaberrima	A	A	A	R

1. a. Site 51 is (-) …… 2a

   b. If (A) …… 2b

2. a. Site 52 is (-) …… 3a

   b. If (A) …… 3b

3. a. Site 53 is (-) …… 4a

   b. If (A) …… 4b

4. a. Site 212 is (G) …… *P. globosa*

   b. If (R) …… *S. scaberrima*

Clade 14

There is no differentiation between *S. platycarpa* and *S. stenoptera*

Clade 15

There is no differentiation between *S. palosapis* and *S. pubistyla*

Clade 16

Species	**Position of polymorphic sites**		
	90	151	283
S. blumuthensis	G	G	T
S. havilandii	G	G	T
S. ochroploia	A	T	T
S. seminis	G	G	K

1.	a.	Site 90 is (G)	……	2a
	b.	If (A)	……	2b
2.	a.	Site 151 is (G)	……	3ab
	b.	If (T)	……	3c
3.	a.	Site 283 is (T)	……	*S. blumuthensis*, *S. havilandii, S. seminis*
	b.	If (G)	……	*S. seminis*
	c.	If (T)	……	*S. ochroploia*

Clade 17

Species	Position of polymorphic sites								
	7	15	19	22	178	202	287	333	360
S. balangeran	A	A	T	A	C	T	T	C	A
S. faguetiana	W	R	Y	M	Y	K	K	M	M
S. lepidota	A	A	T	A	C	T	T	C	A
S. rugosa	A	A	T	A	C	T	T	C	A
S. curtisii	A	A	T	A	C	T	T	C	A

1. a. Site 7 is (A) …… 2a
   b. If (T) …… 2b
2. a. Site 15 is (A) …… 3a
   b. If (R) …… 3b
3. a. Site 19 is (T) …… 4a
   b. If (Y) …… 4b
4. a. Site 22 is (A) …… 5a
   b. If (M) …… 5b
5. a. Site 178 is (C) …… 6a
   b. If (Y) …… 6b
6. a. Site 202 is (T) …… 7a
   b. If (K) …… 7b
7. a. Site 287 is (T) …… 8a
   b. If (K) …… 8b
8. a. Site 333 is (C) …… 9a
   b. If (M) …… 9ab
9. a. Site 360 is (A) …… *S. balangeran*, *S. lepidota, S. rugosa, S. Curtisii, S. faguetiana*
   b. If (C) …… *S. faguetiana*

Clade 18

Species	Position of polymorphic sites
	51
S. atrinervosa	-
S. biawak	A
S. inapendiculata	-

1. a. Site 51 is (-) …… *S. atrinervosa, S. inapendiculata*

   b. If (A) …… *S. biawak*

Clade 19

Species	Position of polymorphic sites	
	173	275
S. flaviflora	G	A
S. platyclados	G	A
S. singkawang	R	M

1. a. Site 173 is (G) …… 2a

   b. If (A) …… 2ab

2. a. Site 275 is (A) …… *S. flaviflora, S. Platyclados, S. singkawang*

   b. If (C) …… *S. singkawang*

Clade 20

Species	Position of polymorphic sites								
	11	16	62	66	67	364-372	380	475	502
S. falciferoides	T	A	C	A	G	-		-	-
S. foxworthyii	Y	M	Y	W	K	-		C	C
S. polysperma	T	A	C	A	G	-		-	-
S. thorelii	T	A	C	A	G	TTTCAAATA	A	-	-

1. a. Site 11 is (T) …… 2a
   b. If (C) …… 2b
2. a. Site 16 is (A) …… 3a
   b. If (M) …… 3b
3. a. Site 62 is (C) …… 4a
   b. If (Y) …… 4b
4. a. Site 66 is (A) …… 5a
   b. If (W) …… 5b
5. a. Site 67 is (G) …… 6ab
   b. If (K) …… 6c
6. a. Site 364-372 is (-) …… 8a
   b. If (TTTCAAATA) …… 7
   c. If (-) …… 8b
7. Site 380 is (A) …… 8c
8. a. Site 475 is (-) …… 9a
   b. If (C) …… 9b
   c. If (-) …… 9c
9. a. Site 502 is (-) …… *S. falciferoides, S. polysperma*
   b. If (C) …… *S. foxworthyii*
   c. If (-) …… *S. thorelii*

Clade 21

Species	Position of polymorphic sites			
	52	275	294	500
S. falcifera	A	A	T	C
S. macroptera	A	A	T	C
S. acuminata	-	A	T	C
S. slotenii	-	M	K	C
S. ferruginea	-	A	G	C
S. macroptera sbsp. macropterifolia	A	A	G	Y
S. acuta	A	A	G	C
S. teysmaniana	-	A	G	C

1.	a.	Site 52 is (A)	……	2a
	b.	If (-)	……	2bc
2.	a.	Site 275 is (A)	……	3ab
	b.	If (A)	……	3cd
	c.	If (C)	……	3e
3.	a.	Site 294 is (T)	……	4a
	b.	If (G)	……	4bc
	c.	If (T)	……	4d
	d.	If (G)	……	4e
	e.	If (K)	……	4f
4.	a.	Site 500 is (C)	……	*S. falcifera, S. macroptera*
	b.	If (T)	……	*S. macroptera sbsp. macropterifolia*
	c.	If (C)	……	*S. acuta, S. macroptera sbsp. macropterifolia*
	d.	If (C)	……	*S. acuminate, S. slotenii*
	e.	If (C)	……	*S. teysmaniana, S. ferruginea, S. slotenii*
	f.	If (C)	……	*S. slotenii*

Clade 22

Species	Position of polymorphic sites												
	7	90	151	194	244	246	275	276	336-341	344	363	375	378
S. crassa	A	G	G	A	-	G	A	C	-	T	A	A	A
S. guiso	R	R	K	R	A	G	M	Y	AAGAAT	K	W	M	R
S. ochracea	A	G	G	A	A	R	M	Y	AAGAAT	T	W	A	A

1. a. Site 7 is (A) …… 2a
   b. If (R) …… 2b
2. a. Site 90 is (G) …… 3a
   b. If (R) …… 3b
3. a. Site 151 is (G) …… 4a
   b. If (K) …… 4b
4. a. Site 194 is (A) …… 5ab
   b. If (R) …… 5c
5. a. Site 244 is (-) …… 6a
   b. If (A) …… 6b
   c. If (A) …… 6c
6. a. Site 246 is (G) …… 7a
   b. If (R) …… 7b
   c. If (G) …… 7c
7. a. Site 275 is (A) …… 8a
   b. If (M) …… 8b
   c. If (M) …… 8c
8. a. Site 276 is (C) …… 9a
   b. If (Y) …… 9b
   c. If (Y) …… 9c
9. a. Site 336-341 is (-) …… 10a

b. If (AAGAAT) …… 10b

c. If (AAGAAT) …… 10c

10. a. Site 344 is (T) …… 11a

b. If (T) …… 11b

c. If (K) …… 11c

11. a. Site 363 is (A) …… 12a

b. If (W) …… 12b

c. If (W) …… 12c

12. a. Site 375 is (A) …… 13a

b. If (A) …… 13b

c. If (M) …… 13c

13. a. Site 378 is (A) …… *S. crassa*

b. If (A) …… *S. ochracea*

c. If (R) …… *S. guiso*

Clade 23

Species	**Position of polymorphic sites**
	51
S. compressa	-
S. ovata	A

1. a. Site 51 is (-) …… *S. compressa*

b. If (A) …… *S. ovata*

Clade 24

Species	Position of polymorphic sites				
	51	90	151	265	332
S. colina	-	G	G	C	T
S. domatiosa	-	G	G	C	T
S. laevis	A	G	G	Y	Y
S. materialis	-	R	K	Y	T
S. maxwelliana	A	G	G	C	T

1. a. Site 51 is (-) …… 2ab
   b. If (A) …… 2c
2. a. Site 90 is (G) …… 3a
   b. If (R) …… 3b
   c. If (G) …… 3c
3. a. Site 151 is (G) …… 4a
   b. If (K) …… 4b
   c. If (G) …… 4cd
4. a. Site 265 is (C) …… 5a
   b. If (Y) …… 5b
   c. If (Y) …… 5c
   d. If (C) …… 5d
5. a. Site 332 is (T) …… *S. colina, S. domatiosa*
   b. If (T) …… *S. materialis*
   c. If (Y) …… *S. laevis*
   d. If (T) …… *S. maxwelliana*

Clade 25

Species	Position of polymorphic sites		
	51	52	53
P. lucida	-	-	-
S. amplexicaulis	-	-	-
S. bullata	A	A	A
S. johorensis	A	-	-
S. smithiana	A	-	-

1. a. Site 51 is (-) 2a
   b. If (A) 2bc
2. a. Site 52 is (-) 3a
   b. If (A) 3b
   c. If (-) 3c
3. a. Site 53 is (-) ..... *P. lucida, S. amplexicaulis*
   b. If (A) ..... *S. bullata*
   c. If (-) ..... *S. johorensis, S. smithiana*

Clade 26

Species	Position of polymorphic sites														
	5	10	17	21	22	58	115	156	177	181	193	212	231	244	246
H. celebica	A	A	C	T	C	C	A	G	A	C	C	A	T	-	G
H. nigra	C	G	A	T	C	C	A	G	A	C	C	A	T	-	G
S. argentifolia	C	G	C	C	A	C	A	G	A	C	C	A	T	-	G
S. parvifolia	C	G	C	C	A	Y	A	G	M	Y	C	A	T	-	G
S. rubra	C	G	C	C	A	C	A	G	A	C	C	A	T	-	G
S. selanica	C	G	C	C	A	C	R	S	A	C	Y	G	T	A	G
S. virescens	C	G	C	C	A	C	A	S	A	C	C	G	K	A	R

Species	Position of polymorphic sites									
	275	276	298	333	344	375	493	534	537	333
H. celebica	A	C	A	C	T	A	A	-	C	C
H. nigra	A	C	A	C	T	A	A	C	C	C
S. argentifolia	A	C	A	C	T	A	A	N	C	C
S. parvifolia	A	C	A	C	T	A	W	T	C	C
S. rubra	A	C	A	C	T	A	A	N	C	C
S. selanica	A	C	R	Y	K	M	A	T	G	Y
S. virescens	M	Y	A	C	T	A	A	T	C	C

1. a. Site 5 is (A) …… 2a
   b. If (C) …… 2b
2. a. Site 10 is (A) …… *H. celebica*
   b. If (G) …… 3
3. a. Site 17 is (A) …… 4a
   b. If (C) …… 4b
4. a. Site 21 is (T) …… 5a
   b. If (C) …… 5b
5. a. Site 22 is (C) …… *H. nigra*
   b. If (A) …… 6
6. a. Site 58 is (C) …… 7
   b. If (Y) …… *S. parvifolia*
7. a. Site 115 is (A) …… 8
   b. If (R) …… *S. selanica*
8. a. Site 156 is (G) …… *S. argenti folia*
   b. If (S) …… *S. virescens*

Clade 27a

Species	Position of polymorphic sites												
	51	178	202	244	246	275	276	282	287	326	344	487	507
S. acuminatissima	-	T	G	-	G	A	C	T	G	G	T	A	A
S. faguetioides	-	T	G	-	G	A	C	T	G	G	T	A	A
S. gibbosa	-	T	G	-	G	A	C	T	G	G	T	A	A
S. hopeifolia	A	Y	K	A	R	A	C	G	K	G	T	A	A
S. longisperma	-	T	G	-	G	M	Y	T	G	G	T	M	A
S. longiflora	-	T	G	-	G	A	C	T	G	G	T	A	A
S. maxima	-	T	G	-	G	A	C	T	G	G	K	A	A
S. multiflora	-	T	G	-	G	A	C	T	G	G	T	A	M
S. patoiensis	-	T	G	-	G	A	C	T	G	G	T	A	A
S. peltata	-	T	G	-	G	A	C	T	G	T	T	A	A
S. rihetia	-	T	G	-	G	A	C	T	G	G	T	A	A

1. a. Site 51 is (-) …… 2a

   b. If (A) …… 2b

2. a. Site 178 is (T) …… 3a

   b. If (Y) …… 3b

3. a. Site 202 is (G) …… 4a

   b. If (K) …… 4b

4. a. Site 214 is (-) …… 5

   b. If (A) …… *S. hopeifolia*

5. a. Site 275 is (A) …… 6a

   b. If (M) …… 6b

6. a. Site 276 is (C) …… 7a

8. a. Site 344 is (T) …… 9a

   b. If (K) …… *S. maxima*

9. a. Site 487 is (M) …… *S. longisperma*

   b. If (A) …… 10

9. a. Site 507 is (M) …… *S. multiflora*

   b. If (A) …… *S. patoiensis, S. richetia, S. longiflora, S. gibbosa*, *S. faguetioides*, *S*. *acuminatissima*,

Clade 27b

Species	Position of polymorphic sites																			
	5	22	38	42	53	195	217	218	257	294	318	322	358	522	523	525	529	534	535	537
H. grifithii	C	A	-	A	A	G	C	G	C	T	A	A	C	G	W	A	G	T	C	T
P. malaanonan	M	M	-	A	G	G	C	G	C	T	G	T	Y	G	W	A	G	T	Y	Y
S. albida	C	A	-	A	A	G	C	G	C	T	A	A	C	G	W	A	G	T	C	G
S. contorta	A	A	-	A	A	G	C	G	C	T	G	T	C	G	G	A	G	C	C	T
S. obscura	C	A	-	A	A	G	C	G	C	T	A	A	C	G	W	A	G	N	C	N
S. ovalis	C	A	C	R	A	S	Y	R	Y	K	A	A	C	R	K	R	R	C	C	T
S. sandakanensis	C	A	-	A	A	G	C	G	C	T	A	A	C	G	W	A	G	C	C	G

1. a. Site 5 is (C) …… 3b
   b. If (M) …… 2a
   c. If (A) ……
2. a. Site 5 is (A) …… 3a
   b. If (M) …… *P. malaanonan*
3. a. Site 22 is (A) …… 4a
   b. If (-) …… 4b
4. a. Site 214 is (-) …… 5a
   b. If (A) …… 5b
5. a. Site 246 is (G) …… 6a
   b. If (R) …… 6b
6. a. Site 275 is (A) …… 7a
   b. If (M) …… 7b
7. a. Site 276 is (A) …… 8ac
   b. If (M) …… 8b
8. a. Site 282 is (T) …… 9a
   b. If (T) …… 9a
   c. If (G) …… 9c
9. a. Site 287 is (G) …… 10a
   b. If (K) …… *S*

10. a. Site 326 is (G) …… 11a

    b. If (T) …… *S. multiflora*

11. a. Site 344 is (K) …… *S. patoiensis*

    b. If (T) …… *S. richetia, S. gibbosa*

12. a. Site 487 is (A) …… 13a

    b. If (M) …… *S. longisperma*

13. a. Site 507 is (A …… *S. acuminatissima*

Clade 27c

Species	Position of polymorphic sites									
	7	14	52	194	275	293	322	333	378	397
H. bancana	A	T	-	G	C	A	A	C	A	C
H.dryobalanoides	A	T	A	A	A	A	A	C	A	C
H. malibato	A	T	-	A	A	A	A	C	A	C
H. mengarawan	A	T	-	A	A	A	A	C	A	C
H. odorata	G	T	-	G	C	A	A	C	G	C
H. philippinensis	A	T	-	A	A	A	A	C	A	C
H. plagata	G	A	-	G	A	A	A	C	A	C
S. astylosa	G	T	-	G	A	C	C	A	A	A
S. latifolia	A	T	-	A	A	A	A	C	A	C

1. a. Site 7 is (A) …… 2a

   b. If (G) …… 2b

2. a. Site 14 is (T) …… 3a

   b. If (A) …… 8b

3. a. Site 52 is (-) …… 4a

   b. If (A) …… *H. dryobalanoides*

4. a. Site 194 is (A) …… 5a

   b. If (G) …… 5b

5. a. Site 275 is (A) …… 6ac

   b. If (C) …… 6b

6. a. Site 293 is (A) …… 7a

	b.	If (C)	……	*S. astylosa*
7.	a.	Site 378 is (A)	……	8a
	b.	If (G)	……	*H. odorata*
8.	a.	Site 379 is (C)	……	*S. latifolia*, *H. malibato*, *H. mengarawan*, *H. philippinensis*

Clade 27d

Species	Position of polymorphic sites														
	51	52	53	54	154	163	230	246	256	296	317	362	373	404	491
S.exelliptica	-	-	-	-	T	G	C	A	C	G	G	T	G	C	A
S.agami	-	-	-	-	T	G	-	A	C	G	G	T	G	C	A
S.resinosa	A	-	-	-	T	G	-	A	C	G	G	T	G	C	A
S.confusa	-	-	-	-	T	G	-	A	C	G	G	T	G	C	A
S.symingtonii	-	-	-	-	T	G	-	A	C	G	G	T	G	C	A
S.assamica	-	-	-	-	T	G	-	R	C	G	G	T	G	C	A
S.brachteolata	A	A	A	A	T	G	-	R	Y	G	K	Y	G	C	A
Dry.oblongifolia	-	-	-	-	C	G	T	G	C	G	G	T	G	A	A
S. roxburghii	A	-	-	-	T	G	-	G	C	T	G	T	T	C	C
S. henryana	-	-	-	-	T	A	-	G	T	T	K	C	C	C	A
S. lepida	-	-	-	-	T	A	-	G	T	T	K	C	C	C	A

1	a	Site 51 is (-)	……	2a
	b	If (A)	……	2b
2	a	Site 52 is (-)	……	3a
	b	If (A)	……	3b
	c	If (-)	……	3c
3	a	Site 53 is (-)	……	4a
	b	If (A)	……	4b
	c	If (-)	……	4c
4	a	Site 54 is (-)	……	5a
	b	If (A)	……	5b
	c	If (-)	……	5c
5	a	Site 154 is (T)	……	6a

	b	If (C)	……	*Dry. oblongifolia*
6	a	Site 163 is (A)	……	*S.henryana, S. lepida*
	b	If (G)	……	7a
	c	If (G)	……	7c
7	a	Site 230 is (-)	……	8a
	b	If (C)	……	*S. exelliptica*
	c	If (-)	……	8c
8	a	Site 246 (R)	……	*S. assamica*
	b	If (A)	……	*S. agami, S. confusa, S. symingthonii*
	c	If (A)	……	*S. resinosa*
	d	If (G)	……	*S. roxburghii*

Clade 28

Species	Position of polymorphic sites															
	51	140	197	241	248	272	273	274	275	279	280	283	286	289	288	289
A. costata	-	C	A	G	G	C	-	-	C	T	T	T	A	A	A	A
C.lanceolatum	A	C	C	A	A	C	T	T	A	A	A	C	T	T	T	T
V. cinerea	-	C	A	A	A	A	-	-	C	T	T	T	A	A	A	A
V. odorata	-	C	A	A	A	C	-	-	C	T	T	T	A	A	A	A
V.philastreana	-	C	A	A	A	C	-	-	C	T	T	T	A	A	A	A
V. subglabra	-	C	A	A	A	A	-	-	C	T	T	T	A	A	A	A
U. borneensis	-	T	A	A	A	C	-	-	C	T	T	T	A	A	A	A

291	296	298	299	315	316	317	320	321	330-331	325	327	342	343	344	351	353	357	380	381	382	384-388	391-392	481
T	G	A	A	A	A	G	T	T	AA	T	A	C	G	T	T	G	T	G	A	A	CAAAT	AA	G
C	A	T	T	T	A	A	A	A	--	-	A	A	G	G	G	A	C	-	-	C	0	0	C
T	G	A	A	A	A	G	T	T	AA	T	M	C	C	G	T	G	T	G	A	A	CAAAT	AA	C
T	G	A	A	A	A	G	T	T	AA	T	A	C	C	G	T	G	T	G	A	A	CAAAT	AA	C
T	G	A	A	A	A	G	T	T	AA	T	A	C	G	G	T	G	T	G	A	A	CAAAT	AA	C
T	G	A	A	A	A	G	T	T	AA	T	A	C	C	G	T	G	T	G	A	A	CAAAT	AA	C
T	G	A	A	A	G	G	T	T	AA	T	A	C	G	G	T	G	T	G	A	A	CAAAT	AA	C

1.	a.	Site 51 is (A)	……	2a
	b.	If (-)	……	2b
2.	a.	Site 84 – 89 is (AAAAGC)	……	3a
	b.	If (-)	……	3b
3.	a.	Site 140 is (T)	……	*U. borneensis*
	b.	If (C)	……	4
4.	a.	Site 197 is (A)	……	5
	b.	If (C)	……	*C. lanceolatum*
5.	a.	Site 241 is (G)	……	*A.costata*
	b.	If (A)	……	6
6.	a.	Site 272 is (A)	……	7a
	b.	If ((C)	……	7b
7.	a.	Site 327 is (M)	……	*V. cinerea*
	b.	If (A)	……	*V. suglabra*
	c.	If (A)	……	8
8.	a	Site 343 is (C)	……	*V. odorata*
	b	If (G)	……	*V. philastreana*

Clade 29

Species	Position of polymorphic sites							
	8	51	81	155	193	301-314	340	525
D. alatus	-	-	A	G	C	----------	A	A
D. baudii	-	A	A	G	C	TAGGTTATAGCAAA	A	A
D. condorensis	-	A	A	G	C	TAGGTTATAGCAAA	A	A
D. costatus	-	-	A	G	C	TAGGTTATAGCAAA	G	A
D. dyerii	-	-	C	G	C	TAGGTTATAGCAAA	A	A
D. grandiflorus	-	-	A	G	C	TAGGTTATAGCAAA	A	A
D. haseltii	-	-	A	G	C	TAGGTTATAGCAAA	G	A
D. intricatus	-	-	A	R	T	TAGGTTATAGCAAA	A	A
D. kerii	-	-	A	G	C	TAGGTTATAGCAAA	A	A
D. obtusifolius	G	-	A	G	C	TAGGTTATAGCAAA	A	A
D. tempehes	-	-	A	G	C	TAGGTTATAGCAAA	A	A
D. tuberculatus	-	-	A	G	C	TAGGTTATAGCAAA	A	R
D. turbinatus	-	-	A	G	C	TAGGTTATAGCAAA	A	A

1.	a.	Site 8 is (-)	……	2ab
	b.	If (G)	……	2c
2.	a.	Site 51 is (-)	……	3a
	b.	If (A)	……	*D. baudii*, *D. condorensis*
	c	If (-)	……	3c
3.	a.	Site 81 is (A)	……	4a
	b.	If (C)	……	*D. dyerii*
	c.	If (A)	……	*D. obtusifolius*
4.	a.	Site 155 is (G)	……	5a
	b.	If (R)	……	5b
5.	a.	Site 193 is (C)	……	6a
	b.	If (T)	……	*D. intricatus*
6.	a.	Site 301 - 314 is (----------)	……	*D. alatus*
	b.	If (TAGGTTATAGCAAA)	……	7
7.	a.	Site 340 is (A)	……	8a
	b.	If (G)	……	8b
8.	a	Site 525 is (A)	……	*D. grandiflorus, D. kerii, D. tempehes, D. turbinatus*
	b	If (G)	……	*D. costatus, D. haseltii*

CURRICULUM VITAE

PERSONAL

Name	:	Essy Harnelly
Place/date of birth	:	Banda Aceh, 9 January 1975
Sex	:	Female
Marital status	:	Married
Official Position	:	Lecturer at Biology Department in Mathematics and Natural Science Faculty, Syiah Kuala University- Banda Aceh-Indonesia
Home address	:	Jl. Mesjid Ashadaqah Lr. Assyura no 12 Lamlagang, Banda Aceh – Indonesia. 23239
E-mail	:	eharnelly@gmail.com or ehusin@gwdg.de

EDUCATIONAL BACKGROUND

1. 1993-1998	:	Bachelor Degree in Biology (Faculty of Mathematic and Natural Sciences, University of Syiah Kuala, Banda Aceh)
2. 1998-2003	:	Magister of Sains in Biology (Bogor Agricultural Institute)
3. 2009-2013	:	Doctoral study at the Institute of Forest Genetics and Forest Tree Breeding, Georg-August University of Goettingen, Germany

WORKING EXPERIENCES

2000 - Present	:	Lecturer staff in Biology Department, Syiah Kuala University, Banda Aceh Indonesia.

PUBLICATION

Harnelly, Essy., Prinz, Kathleen., and Finkeldey, Reiner., 2012. Suitability of the Two Barcoding Regions *mat*K and *rbc*L to dicriminate Dipeterocarpaceae. Poster Presented at Third European Conference For the Barcode of Life. Belgium 17 – 20 September 2012.