Phylogenetic and molecular dating analyses of the tropical tree family Dipterocarpaceae based on chloroplast *matK* nucleotide sequence data

by

Nirosha Gunasekara

A Thesis

in

the Department

of Biology

Presented in partial fulfillment of the requirements

for the Degree of Master of Science at

Concordia University

Montreal, Quebec, Canada

June 2004

Your file Votre référence
ISBN: 0-612-94659-2
Our file Notre référence
ISBN: 0-612-94659-2

ABSTRACT

Phylogenetic and molecular dating analyses of the tropical tree family Dipterocarpaceae based on chloroplast *matK* nucleotide sequence data

Nirosha Gunasekara

The theories that explain the origin and evolution of the pantropical tree family Dipterocarpaceae have long been controversial, with two alternative hypotheses of Gondwanic origin and vicariance or Eurasian origin and dispersal. To test these two hypotheses, I have inferred the divergence dates of major clades through molecular phylogenetic and dating analyses using the nucleotide sequences of the chloroplast *matK* gene and fossil data of 49 dipterocarp species. Phylogenetic analyses produced highly consistent, well-resolved tree topologies with molecular clocklike evolution up to the level of the genera and sections. Independent calibration points based on dipterocarp fossil data were used to calibrate the molecular clock. The results revealed the early Cretaceous Gondwanic origin of the family and the radiation of the Pakaraimoideae (121 – 117 Mya), mid to late Cretaceous radiation of the Monotoideae (117 – 88 Mya) and the late Cretaceous to early Eocene origin and diversification of the Dipterocarpoideae (88 – 54 Mya). The rate of nucleotide substitution of the *matK* gene in dipterocarps is 2.5×10^{-10} per synonymous site per year. The absolute age estimates showed a gradual pattern of speciation in agreement with the "museum hypothesis" up to the level of genera and sections. The results were highly congruent with the paleogeographical and paleoclimatic events, morphological and ecological evidence. This study supports the view that ancestral stocks of dipterocarpoids dispersed from the African to Deccan plate and subsequently dispersed to the Eurasian plate and diversified in Southeast Asia giving rise to the present day species richness in the far eastern rain forests.

ACKNOWLEDGEMENTS

I would like to acknowledge my supervisor, Dr. S. Dayanandan for collecting and maintaining plant material as a part of a long-term research program in Dipterocarp evolution, supporting and giving me this opportunity to perform research work independently. I would also like to express my gratitude to the members of the thesis committee, Dr. Paul Widden and Dr. William Zerges for their support. This research was supported by grants from NSERC, FQRNT, Concordia University and the Canada foundation for innovation to Dr. Dayanandan. My special thanks go to Dr. Paul J. Albert, graduate program director at the department of Biology, Concordia University, for helping me with the installation of computer software, r8s. I am also grateful to Dr. Robert J. Morley at school of Geography and Environmental science, University of Monash, Australia, for his help in verifying some information about dipterocarp fossil data.

Many thanks go to my friends and colleagues, Catherine Calogeropoulos, Margo Marszalek, Kevin Lam, Mona Hamzeh, Jacquelyn Boys, Sandra Chiovitti, Nathalie Brodeur, Dominique Anzellotti and others in the biology department for making research at Concordia an intellectually stimulating and fun experience.

Finally and most importantly, I gratefully acknowledge my family members for their invaluable support and encouragement. I am especially indebted to my beloved parents and soul mate, Haris. The mere word 'thanks' will never be enough to show my deepest appreciation for you, mom, dad and Haris, and I never would accomplish this without your unwavering support.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0. INTRODUCTION

Despite their scattered distribution, far Eastern tropical rain forests are ranked highest among the most unique biodiversity hotspots in the world (Whitmore, 1998). The floristic wealth of these ecosystems has mostly been attributed to the tremendous species diversity of the dominant tree family, the Dipterocarpaceae (Ashton, 1982), which are commonly known as dipterocarps. Amongst lowland evergreen taxa, the ecological and biogeographical uniqueness of dipterocarps remain outstanding since they show a disjunct distribution pattern, exceptionally skewed to the far East (Fig. 1). The three sub families, Pakaraimoideae in South America, Monotoideae in Africa, Madagascar and South America, Dipterocarpoideae in South East Asia, represent 0.5%, 7.0% and 92.0% of the estimated total number of 500 species respectively (Ashton, 1982; Maury and Curtet, 1998). Apart from their ecological value, dipterocarps serve as a main source of timber and non-timber forest products extracted in the forms of resins, medicinal and edible plant parts (Appanah, 1998; FRA 2000).

1.1. THE FAMILY DIPTEROCARPACEAE

1.1.1. Classification

The existence of strikingly distinct groups of dipterocarps in South East Asian lowland evergreen forests, also known as dipterocarp forests, has captivated researchers for many decades (Ashton, 1982, 1988; Whitmore, 1989, 1998; Maury and Curtet, 1998). The original circumscription of the family Dipterocarpaceae was based on the type genus *Dipterocarpus* (Blume, 1825). The phylogenetic placement of the Dipterocarpaceae

either in the order Malvales or in the order Theales within the eudicot angiosperms had long been controversial. The members of the family Dipterocarpaceae show affinities to other families (Table 1) such as the Clusiaceae, Sarcolaenaceae, Theaceae and Tiliaceae (Ashton, 1982; Maury and Curtet, 1998). Cronquist (1968) placed it in the order Theales, based on morphological characters such as the persistent imbricate sepals, numerous hypogynous stamens and frequently contorted corolla. Conversely, the floral vascular supply and the presence of mucilage cells, glandular hairs and connectival appendages assign it to the order Malvales (Dalgren, 1975). Molecular phylogenetic analyses based on *rbcL* DNA sequence data (Chase *et al.* 1993; Dayanandan, 1996; Alverson, *et. al,* 1998) however, have confirmed the Malvalean affinity of Dipterocarpaceae.

The placement of the sub families, Monotoideae and Pakaraimoideae in the family Dipterocarpaceae, has been controversial. Blume (1825) and Heim (1892) mentioned close affinities of the genus *Monotes* to the Tiliaceae whereas Gilg (1925) assigned it to the sub family Monotoideae. The family Sarcolaenaceae, which is rare and endemic to Madagascar, is also considered to be a close ally of the sub family Monotoideae based on anatomical features of the wood (Ashton, 1982). After the circumscription of the monotypic sub family Pakaraimoideae, based on *Pakaraimaea dipterocarpacea* (Maguire *et al.,* 1977), the close affinities between the Dipterocarpaceae and Sarcolaenaceae were defensible with reference to the homologous characters of the fruit calyx (Maury and Curtet, 1998). Maury (1979) suggested close affinities between the Monotoideae and Dipterocarpoideae, primarily based on the fruit, embryo and seedling characters whereas Kostermans (1985) argued for the similar staminal characters in both *Monotes* and *Pakaraimaea* that are shared with the members of the Tiliaceae.

Kostermans (1985) therefore placed both the Monotoideae and the Pakaraimoideae in a new family called the Monotaceae, which was closer to the Tiliaceae than to the Dipterocarpaceae, based on the tilioid pollen exine surface. The recently identified South American Dipterocarp species, *Pseudomonotes tropenbosii* has been attributed to the sub family Monotoideae, considering its close affinities to the Monotoid genus, *Marquesia* (Londono *et al.,* 1995; Morton, 1995; Morton *et al.,* 1999).

Based on distinct floral morphology, reproductive biology and anatomical characters of the wood (Table 1), however, the Monotoideae, Pakaraimoideae and Dipterocarpoideae have been raised into three separate subfamilies within the family Dipterocarpaceae (Maguire and Ashton, 1977).

Classification of taxa within the species rich sub family Dipterocarpoideae has been based on fruit, embryo and seedling characters, chromosome number and anatomy of the wood (Ashton, 1982; Maury and Curtet, 1998). The Dipterocarpoideae comprises two tribes, 13 genera 17 sections and 12 sub sections (Table 2). The tribe Dipterocarpeae consists of more than 150 species in 8 genera (*Dipterocarpus, Upuna, Cotylelobium, Stemonoporus, Anisoptera, Vatica, Vateria* and *Vateriopsis*) and 4 sections characterized by a valvate arrangement of the fruit sepals and the chromosome number of 11. The species rich tribe Shoreae comprises over 300 species in 6 genera *(Hopea, Shorea, Balanocarpus, Neobalanocarpus, Parashorea* and *Dryobalanops*), 13 sections and 12 sub sections (about 200 species of *Shorea* and over 100 species of *Hopea*) classified based on the imbricate arrangement of fruit sepals and base chromosome number of 7. The genus *Dryobalanops* has been considered intermediate to the two groups and the type genus

Dipterocarpus is recognized as a basal lineage of the sub family Dipterocarpoideae (Meijer, 1979; Maury and Curtet, 1998).

Based on embryo and seedling characters Ashton (1982) identified two prominent groups within the large genus *Shorea*. *Shorea* sections *Doona*, *Anthoshorea* and *Pentacme* form one subgroup whereas sections *Rubella*, *Ovales*, *Mutica*, *Brachypterae* and *Pachycrpae* form the most heterogeneous other subgroup, *Rubroshorea* (Meijer, 1979). Subgroup *Rubroshorea*, genus *Parashorea* and the other two sections of *Shorea* (*Richetioides* and *Shorea*) are closely allied (Ashton, 1982).

Among the numerous morphological characters of dipterocarps, a very few characters are equally shared by all species (Maury and Curtet, 1998). Morphological characters are too ambiguous to define the cladogenic events as reversals and parallelisms obscure the relative evolutionary significance. Molecular data better resolves the phylogenetic relationships of taxa rather than morphological characters, as four nucleotide character states are discreet and less homoplastic compared to intricate morphological character states (Gielly and Taberlet, 1996; Page and Holmes, 1998).

As a first attempt at using molecular data, Chase *et al.* (1993) and Dayanandan (1996) have used nucleotide sequences of the chloroplast *rbcL* gene in dipterocarp phylogenetic studies. Tsumura *et al.* (1996) compared 30 species of the Asian sub family using a PCR-RFLP technique based on Chloroplast DNA. However, the taxonomic position of the monotypic *Upuna borneensis* was not evaluated as it was used to root the phylogenetic tree. Based on the nucleotide sequences of *matK*, *trnL* intron and *trnL-F* intergenic spacer regions of the chloroplast genome, Kajita *et al.* (1998) reconstructed the Dipterocarp phylogeny including 17 species of the sub family

Dipterocarpoideae, using *Tilia kiusiana* (Tiliaceae) as the outgroup. Nonetheless, the taxonomic positions of St*emonoporus*, *Vateria* and *Vateriopsis* remained unresolved. Dayanandan *et al,* (1999) reconstructed the phylogeny based on chloroplast *rbcL* sequences including more representative taxa from the three sub families (13 genera, 28 species) and confirmed the paraphyletic basal positions of the Pakaraimoideae and Monotoideae.

1.1.2. Evolutionary trends

The equal flower sepals that are longer than the petals, generalised tricolporate pollen and 4-5 celled ovary with 4 ovules that are typical of the members of the order Malvales suggest the Pakaraimoideae as the most primitive sub family of the family Dipterocarpaceae (Ashton, 1982). The subfamilies, Pakaraimoideae and Monotoideae generally share close affinities with each other in the anatomical characters of the wood, stamens and pollen as well as in fruit and seedling development traits such as continued growth of the cotyledons following germination, albumen in the ripe embryo and independent dehiscence of the pericarp that are distinctive to the Pakaraimoideae and Monotoideae compared to the Dipterocarpoideae (Ashton, 1982). The Monotoideae clearly differ from the Dipterocarpoideae in the anatomical characters of the wood (Table 1) where the former holds primitive characters, such as uniseriate rays and elongated medullary mucilage cells, instead of the multiseriate rays and resin canals of the latter.

The departure from the primitive actinomorphy in the Dipterocarp flowers indicates an advanced condition (Ashton, 1982). Generally the valvate flower and fruit sepal arrangement is considered to be more primitive than the imbricate arrangement

(Stebbins, 1974). Therefore, it is generally considered that some of the recently emerged species of the tribe Shoreae have evolved from the ancestors of the tribe Dipterocarpeae (Ashton, 1982). Most of the morphological characters however, may have re-evolved many times, depending on the selective pressures during the course of evolution (Stebbins, 1974).

Within the sub family Dipterocarpoideae, the majority of large flowered taxa, (*Dipterocarpus*, *Vateria*, *Vateriopsis, Anisoptera*, *Parashorea*, *Shorea* sections *Doona*, *Shorea* and *Anthoshorea*) are confined to the canopy (except for large flowered *Stemonoporus* and *Vatica*) whereas small flowered taxa (*Hopea* and some sections of *Shorea*) inhabit the sub canopy and understory (Maury and Curtet, 1998). The number of stamens follows the same trend (Ashton, 1982) indicated by the reduction from the generalised number of 15 (*Vateriopsis, Vateria*, *Stemonoporus, Anisoptera*) to 10 (*Hopea*, *Shorea* section *Richetioides*) and 5 (*Hopea* and some sections of *Shorea*). Flower characteristics illustrate lines of adaptive radiation for pollen vectors. Large flowers produce large yellow pollen grains and are pollinated by bees (e.g. *Dipterocarpus*) whereas small flowers with white anthers (some *Shorea* and *Hopea*) are often pollinated by thrips or other insects (Ashton, 1989; Maury and Curtet, 1998).

Winged fruits characterize the family and the wings are developed from persistent sepals (Ashton, 1982). The long twisted sepals of most of the members of Dipterocarpoideae generally facilitate dispersal by gyration and are characteristic of canopy dominants (basal lineages such as *Dipterocarpus* and *Shorea* section *Doona*) whereas short sepals are common in relatively short understorey trees (*Hopea*) (Ashton, 1982). However, wings as a character had become redundant with the reduction of the

propelling function in larger fruits of some emergent trees since they can easily reach the ground without gyration (Suzuki and Ashton, 1996). Therefore, dipterocarp fruit sepals are considered to have re-evolved with both the reduction of tree height and the enlargement of nut size (Suzuki and Ashton, 1996). Nevertheless, the long sepals of the Dipterocarpaceae in general, are considered to have evolved from ancestors without long sepals because of the fact that no family related to the Dipterocarpaceae has such long and persistent sepals (Suzuki and Ashton, 1996). The heavy seeds with poor dispersal ability, of the genera *Stemonoporus*, *Vateriopsis* and *Vateria* therefore suggest their archaic status within the sub family Dipterocarpoideae (Ashton and Gunatilleke, 1987a).

1.1.3. Ecology and phytogeography

The monotypic genus *Pakaraimaea*, the type genus of the sub family Pakaraimoideae, includes small trees or shrubs confined to the dry seasonal evergreen forests in the Guyana highlands of South America (Maury and Curtet, 1998; Ashton, 1982). Most of the species of Monotoideae (*Monotes*: 26 - 30 species; *Marquesia*: 3 or 4 species) are scattered in dry or deciduous forest formations of Africa and Madagascar except for some species of *Marquesia* such as *Marquesia excelsa*, an evergreen residual species found in the rain forests in Gabon (Maury and Curtet, 1998). Most of the species of *Marquesia* include evergreen, tall to medium sized, buttressed trees whereas *Monotes* are medium sized trees without buttresses (Maury and Curtet, 1998). The recently identified monospecific Monotoid genus *Pseudomonotes* inhabits the rain forests of the Araracuara (Colombia) region of South America (Londono *et al.*, 1995).

The members of the sub family Dipterocarpoideae are highly specialized to their biotic and physical environment and dominate the lowland evergreen rain forests of South East Asia (Ashton, 1969). Dipterocarpoids are resinous, often buttressed, usually evergreen trees, most of which are found in the emergent and canopy layers of the mature phase of primary forests (Ashton, 1982; Maury and Curtet, 1998). Dipterocarp seeds are generally not dispersed by wind or water (Ashton, 1982). Seeds lack dormancy and hardly colonize in the pioneer phase of the secondary vegetation but invade mid to late successional phases depending on seed availabilities (Ashton, 1988). Seeds remain viable only for a short period and thus the seed mortality rates are very high. They can only germinate in shade (except for some drought tolerant species) on the forest floor and grow slowly. Seedlings of the shade tolerant species depend on canopy gaps to initiate the growing phase (Ashton, 1998).

Extant Dipterocarpoid species show high levels of endemicity (Table 3) in the aseasonal tropical zone of Asia where mean annual rainfall exceed 1000 mm with less than 6 months of dry season (Ashton, 1982). Although the most common habitat of Dipterocarpoids is the lowland evergreen forests of the sunda shelf (Ashton, 1988), some species are found on ridges, slopes or valleys of dry to deciduous forests, coastal, riverine or swampy areas (Ashton, 1982).

The contemporary phytogeography of dipterocarps has long been attributed to historical biogeography, merely based on the pantropical disjunction of the three sub families (Ashton, 1988; Maury and Curtet, 1998; Whitmore, 1998). Dipterocarpoid diversity tends to decline eastwards (Ashton, 1982) with clear disjunctions of genera (Fig. 1; Table 2). Notably, the isolation of genera, *Vateriopsis* in the Seychelles, *Upuna* in

Borneo, *Stemonoporus* in Sri Lanka and monotypic *Vateria* in Southern India and Sri Lanka suggest historical extensions of their geographical ranges (Ashton and Gunatilleke, 1987a). The species rich genera, *Dipterocarpus, Shorea* and *Hopea* that rarely show disjunctions, demonstrate a widespread occurrence in the Sunda shelf (Java, Sumatra, Borneo and Malesia), west to the Wallace's line and vicarious with other sister taxa in South East Asia (Table 2). Only 26 species belonging to 4 genera (*Shorea*, *Hopea*, *Vatica* and *Anisoptera*) are found to the East of the Wallace's line in the Celebes, Moluccas and New Guinea (Ashton, 1982, 1988). At the same equatorial latitude there are 287 extant dipterocarp species in Borneo compared to only 7 species in Sulawesi, 80 km away but separated by Wallace's line (Ashton, 1988).

1.1.4. Fossil data

No fossils or living Monotoids or Pakaraimoids have been reported in Asia (Maury and Curtet, 1998). Fossil flowers in the upper Miocene sediments of Budapest and fruit fragments from Oligocene and Miocene sediments of Zevengebergte and Oeningen (the Netherlands) have uncertainly been referred to Monotes, although the evidence has not adequately supported their inclusion in the sub family Monotoideae (Ashton 1982).

The earliest distinct macrofossils of Dipterocarpoids have been recorded in the early Eocene (Ypresian; 52 Mya) London clay flora (Poole, 1993). Those fossil twigs have unambiguously been attributed to the genus *Anisoptera* (*A. ramunculiformis*) although their occurrence in that particular locality has been postulated to be fortuitous, due to the lack of other macrofossils (Poole, 1993). The convincing macrofossils of East

Asia (Lakhanpal, 1970; Awasthi, 1971) and Africa (Bancroft, 1935b) have also exclusively been attributed to the sub family Dipterocarpoideae (Table 4). Although there has been some evidence for fossilized winged fruits in the upper and lower Cretaceous strata of the Eastern United States and England (Bedfordshire) respectively, those fossils were too incomplete to include in the family Dipterocarpaceae (Ashton, 1982).

Tertiary fossils of East Africa suggest a historical extension of the geographical range of Dipterocarpoids. The African fossils, attributed to the genus *Dipterocarpus*, include Tertiary leaf impressions of *Dipterocarpophylum humei* and *D. zeraibense* in Nubian sandstone (Seward, 1935) and Miocene (25 Mya) fossil wood of *Dipterocarpoxylon africanum* from mount Elgon, Kenya (Bancroft, 1935b). The earliest South East Asian Dipterocarpoid fossils have emerged in the form of bicadinane, a geochemical biomarker that has been extracted from the middle Eocene (44 Mya) sediments of Myanmar (Aarssen *et al.*, 1990). The earliest fossil pollen of Oligocene (36 Mya) and Miocene (25 Mya) sediments of Brunei (Borneo) has been attributed to *Dipterocarpus* and *Dryobalanops* respectively (Muller, 1970). Distinct Dipterocarpoid macrofossils however, have not emerged from the fossil records of either Africa or Asia until the early Miocene (25 Mya).

1.2. MOLECULAR APPROACHES FOR INFERRING EVOLUTIONARY HISTORY

1.2.1. Rationale

Although the taxonomic positions of extant genera within the family Dipterocarpaceae have been established based on molecular and morphological data, imperative evolutionary issues that link their prehistoric origin, evolution and contemporary disjunction, remain to be elucidated by the knowledge of their absolute divergence times.

Disjunctions of extant taxa can be explained by either of two models, dispersal or vicariance, in which the former explains dispersals via existing barriers whereas the latter postulates isolation by the emergence of barriers fragmenting the ancestral lineages (Croizat *et al.* 1974; Humphries, 1999).

The origin of the family Dipterocarpaceae has long been debated based on two hypotheses both of which are founded on substantial evidence. The abundance of distinct macrofossils and the higher contemporary species diversity (Fig. 1) provoked many authors to suggest a Eurasian origin of dipterocarps and dispersal into Africa through West Asia and eastwards to the Malayan archipelago (Lakhanpal, 1970; Prakash, 1972; Meher-Homji, 1979). Based on the African fossils and the significant decline of species to the East of the Wallace's line (Fig. 1), the other hypothesis suggests a Gondwanic origin of dipterocarps and vicariance with the continental drift (Croizat, 1952; Ashton, 1982, 1988). Some have suggested a biphyletic origin, the Monotoideae in Gondwanaland and the Dipterocarpoideae in Laurasia (Aubreville, 1976), referring to both African and Asian fossils.

Dipterocarp seeds are incapable of crossing oceanic barriers (Ashton, 1982). Therefore, in order to prove either of the two hypotheses based on the available fossil record, the ages of fossils should match with the paleogeographical reconstructions that indicate the times of existence of historical land bridges linking South American, African and Asian continents.

However, no fossils have been recorded from South America to prove the existence of their ancestors (Maury and Curtet, 1998). The African fossils are too ambivalent and stratigraphically incomplete to support the dispersal of dipterocarps from Africa to the Deccan plate, since fossils have not been recorded from Africa until the younger Tertiary, by which time the Deccan plate already had collided with Eurasia (Ashton, 1982; Morley, 1999). Although distinct fossils of South East Asia suggest an Asian origin and westward dispersal (Lakhanpal, 1970), there was no evidence for direct land bridges that linked India and South America since the late Cretaceous (74 Mya) according to the paleogeographical reconstructions (Morley, 1999). Even though the earliest Eocene fossils of *Anisoptera* suggest the presence of Dipterocarpoids in the lower Tertiary, those twigs do not indicate their proper locality. Moreover the coexistence of extremely diverse groups of species within a given geographical location and the disjunct distribution of other closely related species are too ambiguous to merely interconnect them based on morphological characters (Ashton, 1982).

Despite the paucity of the fossil record and the ambiguities of morphological characters, molecular phylogenies could be used to infer evolutionary patterns of the derived taxonomic assemblages, when combined with available fossil data, historical

biogeography, morphological and ecological evidence (Avis, 1998; Moritz, 1998; Birmingham and Moritz, 1998; Humphries, 1999; Ford, 2002).

The molecular approach to estimating the absolute divergence time of speciation events within a taxonomic group relies on the molecular clock hypothesis (Zukerkandl and Pauling, 1965), which is pragmatically based on the neutral theory of molecular evolution. Neutral theory assumes constant rates of nucleotide substitutions across evolutionary lineages in a temporal scale (Kimura, 1968; Kreitman and Akashi, 1995; Doyle and Gaut, 2000). Although most of the genes of an organism deviate from constancy in evolutionary rates (Arbogast *et al.*, 2002), the selection of genes that evolve in a clocklike fashion greatly increases the accuracy of the divergence time estimates (Kumar and Hedges, 1998; Bromham and Penny, 2003). Molecular evolutionary analyses have revealed some important events in evolutionary history (Arbogast *et al.*, 2002), including the origin of Angiosperms (Wikstrom *et al.*, 2001; Sanderson and Doyle, 2001; Soltis *et al.*, 2002), vertebrate evolution (Kumar and Hedges, 1998) and human evolution (Cann *et al.*, 1987).

1.2.2. Testing for evolutionary neutrality

Of the coding and regulatory DNA sequence variation of an organism, only a fraction is involved in the species' biological differences (Fay and Wu, 2003), as degeneracy of the triplet code allows for synonymous (silent) substitutions, which do not change the function of proteins (Page and Holmes, 1998; Kreitman and Comeron, 1999; Nei and Kumar, 2000). The majority of DNA polymorphisms within a species therefore, are known to be neutral according to the neural theory of molecular evolution (Kimura,

1968). Genes or genomic regions that evolve in a neutral fashion follow a poisson pattern in accumulating mutations, with the variance of nucleotide substitutions equal to the mean with an approximately constant substitution rate per site per year, following a clocklike mode of evolution (Kimura, 1968; Hughes, 1999; Doyle, 2000).

Synonymous substitutions are considered to be free from natural selection whereas nonsynonymous substitutions that change the amino acids are apparently caused by positive selection (Nei and Kumar, 2000). In most nucleotide sequences however, there are more sites that potentially produce nonsynonymous mutations than sites that potentially produce synonymous mutations and the numbers of these sites vary from gene to gene (Nei and Kumar, 2000). Thus, the rates of synonymous and nonsynonymous substitution are defined as the number of synonymous substitutions per synonymous site (K_s) and the number of nonsynonymous substitutions per nonsynonymous site (K_n) (Nei and Kumar, 2000).

Within the framework of the neutral theory, the most robust and straightforward estimation among the quantitative approaches used to evaluate the neutrality of a gene is the ratio of K_n to K_s. Accordingly, it assumes neutral evolution when $K_n/K_s=1$, positive selection for selectively advantageous mutations when K_n/K_s ratio > 1.0 and purifying selection to prevent the spread of detrimental mutations when K_n/K_s <1.0 (Page and Holmes, 1998; Nei and Kumar 2000). Highly expressed and thus functionally constrained genes demonstrate very low values of K_n/K_s ratios and higher codon usage bias due to an excessive number of synonymous substitutions (Page and Holmes, 1998) and vice versa. Therefore, the K_n/K_s ratio could be used as a tool to evaluate the degree of selective neutrality for genes.

1.2.3. Chloroplast DNA as a tool in evolutionary studies of plants

In the past two decades, chloroplast DNA (cpDNA) has become a popular tool in phylogenetic analyses (Curtis and Clegg, 1984; Soltis and Soltis, 1998). The advantages of using cpDNA in molecular evolutionary studies has been emphasized not only as it facilitates straightforward PCR amplifications, due to the high copy number, but also due to their uniparental inheritance that produces unambiguous ancestor-descendant relationships where the confounding effect of recombination is alleviated (Clegg *et al.*, 1994; Birky, 1995; Bakker *et al.*, 1999).

The chloroplast genome is generally considered to be conservative in its evolution and most of its genes demonstrate slow nucleotide substitution rates (Soltis and Soltis, 1998). Highly constrained slow evolving chloroplast genes such as *rbcL, PsaA* and *PsaB* generally demonstrate elevated synonymous substitution rates that are biased towards the third codon position (Soltis and Soltis, 1998; Magallon and Sanderson, 2002). Evaluating the tempo and mode of accumulation of nucleotide substitutions in highly conserved genes over longer evolutionary spans becomes problematic, since the superimposed substitutions at the third codon position reduce the statistical reliability (Ayala, 1997). Although the *rbcL* gene has extensively been used in family level phylogenetic studies (Gielly and Taberlet, 1994, 1996; Judd *et al.* 1999), it has often produced inconsistent results in attempts at molecular dating analyses, primarily as a consequence of a highly constrained mode of evolution that conflicts with the null model of clocklike evolution (Bousquet *et al.*, 1992; Gaut *et al.* 1996; Sanderson and Doyle, 2001; Conti *et al.*, 2002). Nevertheless, different parts of the chloroplast genome evolve at different rates (Soltis and Soltis, 1998). The *matK* gene is known to have evolved

faster than many other chloroplast genes (Hilu and Liang, 1997) suggesting a low functional constraint.

1.2.4. The *matK* gene

Among the commonly used protein coding chloroplast genes in phylogenetic studies, *matK* stands out in resolving intrafamilial relationships (Hilu and Liang, 1997; Soltis and Soltis, 1998; Hilu *et al.,* 2003). Well-resolved intrafamilial phylogenies have been reconstructed using *matK* sequences in many plant families including the Polemoniaceae (Steel and Vilgalys, 1994), Saxifragaceae (Johnson and Soltis, 1994; Soltis *et al.*, 1996), Apiaceae (Plunkett *et al.*, 1996), Ericaceae (Kron, 1997), Juglandaceae (Stanford *et al.*, 2000), Crassulaceae (Mort *et al.,* 2001), Cactaceae (Nyffeler, 2002) and Characeae (Sanders *et al.*, 2003). Furthermore, a recently reconstructed angiosperm phylogeny based on the *matK* gene was highly congruent with many multigene angiosperm phylogenies (Hilu *et al.,* 2003).

The coding gene of *maturase* (*matK*), is involved in splicing introns from RNA transcripts (Kelchner, 2002) and is located in the large single copy region between two exons of the *trnK* intron (Fig. 2a), which is categorized under the family of group II (G2) introns of the chloroplast genome (Hilu and Liang, 1997). Unlike some prokaryotic and lower eukaryotic G2 introns, those of higher plants are known to have evolved by vertical inheritance as exemplified by the *matK* gene (Zimmerly *et al.,* 2001; Kelchner, 2002). Vertical inheritance, which signifies a higher likelihood of orthology, enhances the phylogenetic fidelity of a gene in inferring cladogenic events (Mindell, 2001; Kelchner, 2002). The gene phylogeny derived from such an orthologous gene follows the species

phylogeny of the specified group of taxa (Page and Holmes, 1998; Felsenstein, 2004). Phylogenies based on the *matK* gene are also considered to be less homoplastic, due to a reduced transition bias, as transitions only slightly outnumber transversions (Hilu and Liang, 1997). Relatively low values (closer to 1.0) of transition/transversion (ti/tv) ratios have been recorded for many angiosperm and gymnosperm taxa (Xiang *et al.,* 1998; Hilu *et al.,* 2003).

Protein coding genes are generally considered to be functionally constrained and subject to purifying selection (Hilu and Liang, 1997; Young, 2000) where a higher number of synonymous (neutral) rather than nonsynonymous (selective) mutations tend to accumulate during the course of evolution. However, *matK* is under a relatively lower evolutionary constraint (Hilu and Alice, 1999) as a fast evolving gene (Olmstead and Palmer, 1994). Thus, it demonstrates a relaxed and unbiased mode in accumulating synonymous and nonsynonymous mutations and purifying selection has become less significant (Hilu *et al.,* 2003). The nucleotide substitutions in the first, second and third codon positions are therefore approximately equal when compared to highly constrained genes (Johnson and Soltis, 1994; Hilu *et al.,* 1999).

Most importantly, the less constrained mode of evolution indicates increased evolutionary neutrality (Page and Holmes; 1998), which is central to the molecular clock hypothesis (Kimura, 1983; Kimura and Ohta, 1971; Ohta, 1992; Kreitman and Akashi, 1995; Nei and Kumar, 2000). A derived phylogenetic tree based on *matK* sequences therefore, could be used in estimating divergence times, assuming the neutral mode of evolution of a specified group of taxa. Constant substitution rates have been observed in the *matK* gene in the basal families of both gymnosperms and angiosperms, such as the

Pinaceae (Wang *et al.,* 2000) and Magnoliaceae (Azuma *et. al,* 2001), that agree with the molecular clock hypothesis.

1.2.5. Modes of speciation in tropical trees

Derived molecular clocklike phylogenetic trees could be used to distinguish between the two models that have been used to describe the speciation in the tropical flora, the gradual accumulation of species versus rapid and recent diversification. Stebbins (1974) pointed out that the majority of species in the tropical rain forests belong to families that are considered to be intermediate in their phylogenetic position between the most primitive and the most advanced families. Thus, active speciation is considered to be rare in stable tropical ecosystems as compared to marginal ecosystems. Therefore, he suggested that the rain forests are not the centers of origin of angiosperms, but to greater extent museums, in which the species are preserved under stable equatorial climates and new species are accumulated gradually over long geological periods. This theory is commonly known as the 'museum hypothesis'.

Recently Haffer (1987) and Richardson *et al.* (2001) suggested that the intermittent cool and warm periods of the Pleistocene (< 2 Mya) fragmented the rain forest species into small isolated pockets. This model therefore states that the allopatric divergence during the Pleistocene period acted as a species pump, triggering recent speciation in separate refugia. Although the evidence for rapid speciation has been documented in the neotropics, its role in the species richness in South East Asia has not been widely discussed (Morley, 1999).

The clustering pattern of species with reference to their most recent common ancestors in the derived molecular clocklike phylogenetic trees could be used to assess their mode of speciation (Richardson *et al.,* 2001). If a gradual speciation is in effect, the phylogenetic tree should be well resolved, with a substantial number of substitutions along the branches as shown in Fig. 3a. In contrast, a rapid recent speciation would be indicated by a phylogenetic tree with unresolved short branches emerging from the most recent common ancestor (Fig. 3b). Using a chronogram, which is a phylogenetic tree with branches proportional to time, these effects could be identified with reference to the estimated divergence times. Thus, the mode of speciation could apparently be distinguished.

1.2.6. Objectives

The objective of the present study is to reconstruct the phylogeny of the family Dipterocarpaceae using *matK* nucleotide sequence data and use the resulting phylogenetic tree(s) and available Dipterocarp fossil data in molecular dating analyses to; 1) elucidate the evolutionary history of the family, integrating morphological, biogeographical and ecological data, 2) infer the absolute divergence times for the origin and radiation of dipterocarps in order to test the two alternative hypotheses, Gondwanic origin and vicariance or Eurasian origin and dispersal, and 3) evaluate the relative importance of the museum hypothesis (gradual accumulation of species through time) versus recent rapid speciation hypothesis in the speciation of dipterocarps.

2.0. MATERIALS AND METHODS

2.1. TAXON SAMPLING

Taxa representing the three sub families and 16 out of the 18 genera of the family Dipterocarpaceae were selected for the present study; *viz.* three species (2 genera) from the Monotoideae; 2 individuals of monotypic *Pakaraimaea* from the Pakaraimoideae; 33 species from the Dipterocarpoideae (11 genera belonging to two tribes, Shoreae and Dipterocarpeae). From the species rich clades of the tribe Shoreae, 11 species of *Shorea* belonging to 7 sections (*Rubella*, *Ovales*, *Mutica*, *Brachypterae*, *Neohopea*, *Doona* and *Anthoshorea*) and 7 species of *Hopea* were chosen.

Except for the monospecific genera, *Upuna, Vateriopsis*, *Neobalanocarpus* and the monotypic genus *Balanocarpus*, at least two species were selected from each of the remaining genera. Twelve other sequences including, *Anisoptera* and *Dryobalanops* were obtained from GenBank (Table 5). Altogether, 49 ingroup species were included in the molecular phylogenetic and dating analyses.

In addition to the outgroups, *Durio zibathinus* and *Pseudobombax sp.* (Bombacaceae), selected for the present study, GenBank sequences of *Tilia kiusiana* (AB006386), *Acer campestre* (AJ438793) and *Acer opalus* (AJ438787) were used to determine the rooting position of basal dipterocarps (section 2.7.2.). A list of species with the accession or voucher numbers is given in the Table 5.

2.2. DNA EXTRACTION, PCR AMPLIFICATION AND SEQUENCING

Total genomic DNA was extracted from the silica gel dried samples according to the methods of Doyle and Doyle (1987), Dayanandan *et al.*, (1997) and standard protocols of the QIAGEN Dneasy plant tissue extraction kit (QIAGEN). Several oligonucleotide primers (Table 6) were used for PCR amplification of an approximately 1200 bp fragment from the 5' end of the chloroplast *matK* gene (Fig. 2b), which is homologous to the 2129 - 3329 bp region of the *Nicotina tabacum* (GenBank accession number: NC_001879) complete chloroplast genome sequence (Shinozaki, *et al.*, 1986). Out of the four internal primers, three (392R, 602F, 892R) were selected based on Kajita, *et al.*, (1998), while the primer, 265F was designed from a conserved region after sequencing. Polymerase chain reactions were performed using a Mastercycler gradient thermal cycler applying a touchdown method (Appendix 1).

Amplified DNA was purified using a QIAQUICK PCR purification kit (QIAGEN), electrophoresed on a 1% agarose gel stained with ethidium bromide (0.4 µg/ml) at 3.5 v/cm for 90 minutes, visualized under UV and quantified (ng/µl) using a Mass Ruler DNA ladder. The Gene Tools software as implemented in the Syngene Gene Genius gel imaging system (a division of Synoptics Ltd.) was used to quantify DNA. Depending on the quantity of DNA (ng) and the specific molecular weight, the volume per sample was adjusted to a standard concentration of 10ng/100bp per each sequencing reaction. Cycle sequencing reactions were performed in an ABI thermal cycler; 25 cycles, each at 96^0C for 10 seconds, 50^0 C for 5 seconds and 60^0 C for 4 minutes. The total volume of each sequencing reaction was 20 µl, which contained, 3.0 µl of ABI sequencing reaction buffer, 1.0 µl of (3.2 pmol) primer, 1.0 to 14.0 µl DNA sample, 0.0

to 13.0 µl of distilled water and 2.0 µl of ABI BigDye™ Terminator Cycle Sequencing Ready Reaction mix. Amplified fragments were purified and directly sequenced using an ABI 310 automated genetic analyser (Applied Biosystems, USA) following the standard protocols.

2.3. SEQUENCE ALIGNMENT

Sequenced DNA chromatograms were edited and assembled using the STADEN software package (Staden group, MRC Laboratory of molecular biology, Cambridge, UK) to obtain approximately 1200 bp homologous contigs of *matK* for each taxon, including the outgroups. The assembled consensus sequences were aligned using ClustalW global alignment software (Thompson, *et al.,* 1994) with default parameters. The aligned DNA sequence data file was converted to Phylip format and imported to MacClade version 4.02 (Maddison and Maddison, 2001) for manual editing. Nucleotide sequences were converted to amino acids using MacClade after detecting the relative codon positions using the DNA to protein translation tool of the expert protein analysis system, *ExPASy* (www.expasy.org).

2.4. PRELIMINARY SEQUENCE ANALYSES

Initially, the preliminary tests were carried out to infer the tempo and mode of evolution of the *matK* nucleotide sequence data of dipterocarps. Unless otherwise noted, Phylogenetic analysis using Parsimony* (* and other methods), PAUP version 4β10 (Swofford, 1999) software was used for DNA sequence data analyses. Based on the nucleotide data matrix, a chi square (X^2) test was performed to rule out the null

hypothesis of homogeneous base distribution across sequences. The phylogenetic signal was evaluated by calculating skewness (g_1) and kurtosis (g_2) based on the distribution of lengths of 10^6 random trees (Hillis and Huelsenbeck, 1992). Numbers of constant, variable and parsimony informative sites were also calculated based on amino acid and nucleotide data matrices by using the computer software, Molecular Evolutionary Genetics Analysis (MEGA; Kumar *et al.,* 2001). Transition/transversion ratios were calculated based on all positions and separately on the first, second and third codon positions using MEGA.

To evaluate the level of neutrality with respect to the amino acid content of *matK* sequences, numbers of synonymous and nonsynonymous sites and substitutions, codon usage and neutrality indices were estimated using the computer software MEGA and DNAsp (DNA polymorphism analysis; Rozas and Rozas, 1995).

Numbers of synonymous and nonsynonymous sites, differences between synonymous and nonsynonymous substitutions and the values of K_s and K_n were calculated using Nei and Gojobori's method (1986), based on all ingroup taxa, as implemented in MEGA. Differences between nonsynonymous to synonymous substitution ratios (K_n/K_s) among the three subfamilies were compared by the Fisher's exact test and Neutrality index (Appendix 2) as implemented in the MacDonald Kreitman neutrality test (MacDonald and Kreitman, 1991) using DNAsp. To evaluate random or non random use of synonymous codons in the *matK* sequences of ingroup taxa, the values of effective codon number (ECN; Wright, 1990), codon bias index (CBI; Morton, 1993) and G+C content were calculated and compared using DNAsp.

In order to select the best fitting nucleotide substitution model for *matK* sequence data, the goodness of fit of nested nucleotide substitution models was tested using a hierarchical likelihood ratio test (LRT), calculating the chi square distributed test statistic, δ (Appendix 3) using PAUP and MODELTEST version 2.0 (Huelsenbeck and Crandall, 1997; Posada and Crandall, 1998). The parameters selected by Akaike information criterion (AIC) as implemented in the MODELTEST, was also used where necessary (Appendix 3). The likelihood ratio test was repeatedly applied in each maximum likelihood analysis performed later during molecular dating analyses.

2.5. INFERRING THE PHYLOGENY

Phylogenetic trees of dipterocarps based upon nucleotide sequences of the *matK* gene of 49 ingroup taxa and 5 outgroups (Table 5) were reconstructed using four different phylogeny inferring methods; parsimony, maximum likelihood, distance and Bayesian inference using PAUP* and MrBayes (version. 2.01) software packages. These methods differ from each other in their assumptions and algorithms for character state optimization (Felsenstein 1981, 1988; Swofford *et al.*, 1996; Huelsenbeck *et al.,* 2001). A detailed description of these methods is given in appendix 4.

Maximum parsimony (MP) analysis was carried out using a random heuristic search with 100 replicates and a TBR (tree bisection reconnection) branch swapping algorithm. All characters were equally weighted since there was no significant transition bias in the *matK* dataset. Since there were no indels in the ingroups except for two (6 bp) deletions, all gaps were treated as missing data. The relative support for the majority rule consensus tree was evaluated using consistency, retention and rescaled consistency

indices (CI, RI and RCI). Maximum likelihood (ML) analyses (Felsenstein, 1981; Huelsenbeck and Rannala, 1997) were performed based on the selected substitution model with estimated base frequencies, rate matrices, invariable sites and the gamma distributed site to site rate variation (Nei and Kumar, 2001). Maximum likelihood analyses were carried out to obtain all combinations of topologies used in bootstrap (section 2.7.3.) and sensitivity (section 2.7.4.) analyses. The neighbour joining (NJ) method (Saitou and Nei, 1987) was applied with the minimum evolution algorithm for corrected distances based on the selected substitution model. The robustness of MP and NJ topologies was evaluated using 2000 bootstrap replicates (Hedges, 1992).

Bayesian analysis was carried out using MrBayes software (Huelsenbeck and Ronquist, 2001; Huelsenbeck *et al.,* 2001) applying the estimated values of substitution rates, base frequencies, invariable sites and gamma shape parameter with 8 rate categories. Four simultaneous Metropolis Coupled Monte Carlo Markov Chains (MC^3) were run for 1,000,000 generations, sampling a tree once every 100 generations, to obtain a 50% majority rule consensus tree for 4200 trees by omitting the first 1000 trees of each chain.

The Shimodaira Hasegawa test (Shimodaira and Hasegawa, 1999) based on the resample estimated log likelihood (RELL) method with 1000 bootstrap replicates of sites (Appendix 5) was used to statistically evaluate the congruence among the resulting MP ML, NJ and Bayesian topologies.

2.6 TESTING THE MOLECULAR CLOCK HYPOTHESIS

Based on the results of phylogenetic analyses, a subset of DNA sequence data was selected to obtain the maximum likelihood topology to use as the dipterocarp backbone phylogenetic tree, which included 1-2 species per genus and the major sections of *Shorea*, *Hopea, Vatica* and one outgroup taxon (29-taxon dataset). The Dipterocarp backbone tree topology was used as the basis for estimating the divergence times of genera and major sections using alternative outgroups (section 2.7.3.).

First, the variance to mean ratio (*R(t)* = index of dispersion) across the branches of the 29-taxon tree was calculated to test the overall rate constancy assuming the poisson distribution of nucleotide substitution rates. The topology independent Tajima's relative rate test (as implemented in MEGA) was used to rule out the null hypothesis of rate constancy between bifurcating taxon pairs (Tajima, 1993) with respect to their immediate basal taxon (Appendix 6).

To evaluate whether the dipterocarp backbone evolved in a clocklike fashion, a likelihood ratio test was used (Appendices 5 and 8). Log likelihood values were calculated with and without enforcing the molecular clock (clocklike substitution rates) and compared using a chi square (X^2) test with *n-2* degrees of freedom (n = number of taxa). While keeping the dipterocarp backbone constant, subsequent likelihood ratio tests were performed with and without enforcing the molecular clock in the sequential and random additions of ingroup taxa (section 2.7.4.).

2.7. MOLECULAR DATING ANALYSES

Molecular dating analyses were performed using the Langley Fitch (Langley and Fitch, 1974) and penalized likelihood (Sanderson, 2002) methods as implemented in the r8s software (version 1.50; Sanderson, 2002). A detailed description of these methods is given in the appendix 7. Based on the results of molecular clock tests (section 2.6.), the Langley Fitch (LF) method was used to estimate the divergence times when the null hypothesis of rate constancy was accepted at the alpha level of 0.05, whereas the penalized likelihood (PL) method was used when the null hypothesis was rejected.

Although the molecular clock hypothesis is accepted at the alpha level of 0.05, molecular clock tests sometimes fail to detect the rate variation due to superimposed nucleotide substitutions within and between lineages (Bromham, *et al.*, 2000). Thus, there is a chance of generating a Type II error, which is accepting the null hypothesis of clocklike evolution when the rates are actually heterogeneous. Therefore, an alternative analysis was carried out using the PL method instead of the LF method based on the clocklike 29-taxon maximum likelihood tree. To calculate the absolute substitution rates across the tree, the Truncated Newton (TN) algorithm (Sanderson, 2002) was used as the optimization criterion. In order to prevent the weak optimizations that produce inaccurate age estimates, multiple starting points were used to obtain 5 random sets of initial age estimates by perturbing the solutions and checking whether each returns to the same point.

2.7.1. Calibration of the molecular clock

In a phylogenetic tree two ages of an extant lineage (Soltis, 2002) have to be taken into account for the calibration of the molecular clock (Sanderson and Doyle, 2001); a) the age of the extant crown group; b) the age of the stem lineage where it connects to the extant sister group (most recent common ancestor of the crown group). For instance, in Figure 9, nodes B and C denote the stem and crown lineages of the sub family Dipterocarpoideae respectively.

As fossils were unavailable to calibrate the root node, the age of the root node was first set to 1.0 and the relative branch lengths for each chronogram were obtained in r8s analyses. In order to distinguish the calibration points that produce less biased age estimates, dipterocarp fossils were first mapped on the molecular clocklike 29-taxon maximum likelihood tree, using distinct fossils as illustrated in the appendix 8.

Consequently, the earliest fossils of *Anisoptera, Hopea* and *Vatica* were independently used to calibrate the molecular clock based on the maximum likelihood trees of 29 and 49 taxa. The node K was calibrated based on the early Eocene (54 – 49 Mya; mid point = 51.5 Mya) fossils of *Anisopteroxylon ramunculiformis* (Poole, 1993). The other two nodes, Q and O were calibrated based on mid Miocene (16 – 10 Mya; mid point =13 Mya) and mid Pliocene (5 – 1.6 Mya; mid point = 3.3 Mya) fossils of *Hopenium* (North India) and *Vaticoxylon* (Java) respectively (Awasthi, 1971). Node K (51.5 Mya) was calibrated in the other ML trees generated in bootstrap (section 2.7.3.) and sensitivity analyses (section 2.7.4.) in order to estimate the ages.

2.7.2. Establishment of the rooting position

Elevated nucleotide substitution rate variation provoked by distantly related outgroups compared to ingroups substantially alters the polarity and branch lengths of ingroup lineages (Graham *et al.*, 2002), which consequently alter the absolute age estimates (Sanderson and Doyle, 2001). To establish the rooting position among the basal branches of the 29-taxon ML tree and to compare the effect of an outgroup on the divergence time estimates, three alternative outgroups from the closely related Bombacaceae (*Durio zibathinus*) and Tiliaceae (*Tilia kiusiana*; AB006386) and the distantly related Sapindaceae (*Acer campestre*; AJ438793) were used. Once the rooting position was established, the outgroup was excluded from molecular dating analysis in order to set the root node to 1.0. Ages were estimated separately to evaluate the effect of the outgroup on the polarity of ingroup lineages.

2.7.3. Calculation of confidence limits for the estimated ages

In order to calculate the confidence limits for the estimated ages, each maximum likelihood tree was fixed and the characters were bootstrapped repeatedly 100 times. The ages were estimated for each bootstrap replicate (Baldwin and Sanderson, 1998; Sanderson and Doyle, 2001). Bootstrap replicates were generated using the SEQBOOT program in PHYLIP (Felsenstein, 1993), translated into NEXUS format using the r8s bootstrap kit (Eriksson, 2002) and a PAUP commands block was inserted for maximum likelihood analyses. The resulting output with 100 ML tree topologies was used as the input file in r8s. Ages were estimated for each replicate and the mean and variance of the

100 replicates were obtained by profiling across the 100 age estimates using the 'profile taxon' command in the r8s program (Baldwin and Sanderson, 1998; Sanderson, 2002).

2.7.4. Sensitivity analyses

In molecular dating analyses, the phylogenetic topology can lead to wrong interpretations due to autocorrelated rates of putative sister taxa commonly known as lineage effects (Donoghue, *et al.*, 1989; Sanderson and Doyle, 2001). Therefore, to test whether the estimated ages were sensitive to lineage effects of different tree topologies, a sensitivity analysis was performed. Likelihood ratio tests with and without enforcing the molecular clock followed by maximum likelihood analyses and molecular dating procedures (LF for clocklike; PL for non clocklike) were applied for each dataset starting with a molecular clocklike backbone topology of 23 taxa and then; a) sequentially adding one taxon at a time to *Stemonoporus* and the species rich *Shorea* and *Hopea* clades b) randomly adding a few taxa at a time to all clades until all (49) taxa were added.

3.0. RESULTS

3.1. SEQUENCE ALIGNMENT

The length of *matK* sequences after alignment was 1219 bp, homologous to the base positions 2129 – 3348 of the *Nicotina tabacum* reference sequence. This includes two insertions (7 and 9 bp in *Acer* spp.) and one deletion (9 bp) in all outgroup species. No indels were found in the ingroup alignments except for two (6 bp) deletions near the 5' end in the *Monotes* and *Pakaraimaea* sequences. The average amino acid and nucleotide compositions at conserved, variable and parsimony informative sites for the 29 and 49 taxon datasets are given in Table 7. The amino acid and nucleotide compositions at all sites and the conserved and variable sites of the first, second and third codon positions in both datasets were approximately equal. The number of parsimony informative sites at the third codon position was slightly higher than those of the first and second codon positions (Table 7).

3.2. PRELIMINARY SEQUENCE ANALYSES

The negatively skewed frequency distribution of the length of 10^6 random trees indicated a strong phylogenetic signal ($g_1 = -1.92$; $g_2 = 6.36$). Base distribution across the sequences was homogeneous and the null hypothesis was accepted at the alpha level of 0.05 ($X^2 = 22.76$; $P = 1.00$). Transitions slightly outnumbered transversions at all positions (Table 8).

Based on the likelihood ratio test (LRT) and akaike information criterion (AIC), the transversion model (TVM) was selected for the 49-taxon dataset with one transition rate, four transversion rates, unequal base frequencies and gamma distributed site-to-site

rate variation (Table 8). For the 29-taxon dataset, LRT selected the transversion model with zero invariable sites and a gamma shape parameter of 0.878 (TVM + G), whereas AIC selected the transversion model with 0.414 invariable sits (TVM + I).

The synonymous and nonsynonymous sites based on all ingroup taxa were 235.7 and 813.7 and the synonymous and nonsynonymous differences were 14.6 and 19.5 respectively. Thus, the values of K_n and K_s were 0.239 and 0.619 respectively. The overall nonsynonymous to synonymous substitution ratio (K_n/K_s) was 0.387. There were no significant differences between K_n/K_s ratios among sub families [South American (SA), African (AF) and South East Asian (SEA)] according to the Fisher's exact test ($P_{SA \text{ and } AF}$=1.000; $P_{AF \text{ and } SEA}$=0.517; $P_{SA \text{ and } SEA}$=0.516; α=0.05). The results also indicated the neutrality of dipterocarp *matK* sequences with a slight trend for purifying selection among sub families, as the resulting values of the neutrality index were only slightly greater than 1.0 for all comparisons ($P_{SA \text{ and } AF}$=1.09; $P_{AF \text{ and } SEA}$=1.64; $P_{SA \text{ and } SEA}$=1.4).

Codon usage indices revealed a random utilization of codons in *matK* sequences, as the overall values of ENC and CBI for 49 ingroup taxa were 51.122 and 0.393 (section 4.1). The average G + C contents of second (0.283) and third codon positions (0.276) were approximately equal.

3.3. PHYLOGENETIC ANALYSES

The maximum parsimony analysis based on 49-taxon dataset resulted in 45680 equally parsimonious trees. The maximum parsimony tree (Fig. 4) was highly supported by CI, RI and RC values (length = 626; CI = 0.87; RI = 0.95; RC = 0.83). The majority rule (50%) and the strict consensus trees were identical. Maximum likelihood analyses

retained a single tree with a log likelihood (*-lnL*) value of 5268.69 (Fig. 5) which was identical to the parsimony (Fig. 4) and Bayesian majority rule consensus trees (Fig. 6), in which all the clades were supported by > 60% bootstrap values. With the partial exception of some clades in the neighbour joining tree (Fig. 7), phylogenetic trees obtained from all analyses were similar to each other and the Shimodaira Hasegawa test accepted the null hypothesis for each pair of topologies at the alpha level of 0.05 (Table 9). Therefore, the phylogenetic relationships are discussed based on the maximum likelihood tree and majority rule consensus trees of parsimony and Bayesian analyses, which showed the same clustering pattern of taxa.

Monophyly of the Dipterocarpaceae in relation to the selected outgroups was strongly supported by all analyses (Figs. 4, 5, 6 and 7). The South American sub family Pakaraimoideae occupied the basal position (100% bootstrap support) and was paraphyletic to the African sub family Monotoideae (99% bootstrap support). The Monotoideae was paraphyletic to Dipterocarpoideae. Within the monophyletic Asian sub family, relative taxonomic positions of genera (except *Dipterocarpus* and *Dryobalanops*) and species within the two tribes, Shoreae and Dipterocarpeae were well resolved, each with >70% bootstrap support. The tribe Shoreae included *Hopea, Shorea, Balanocarpus, Neobalanocarpus,* and *Parashorea*, while tribe Dipterocarpeae included *Anisoptera, Cotylelobium, Stemonoporus*, *Upuna, Vatica, Vateria* and *Vateriopsis*.

Within the tribe Shoreae, the *Shorea* species in general showed a polyphyletic origin in two distinct sub clades (Figs. 4, 5 and 6). The clustering pattern (with > 90% bootstrap support) within each sub clade suggested paraphyletic origin of genera and some sections of *Shorea* such as *Doona*, *Anthoshorea* and *Neohopea*. At the specific

level, within each sub clade the clustering pattern of *Shorea* and *Hopea* species reflected shared derived character states (synapomorphies). The phylogenetic relationships were highly congruent with the Ashton's (1982) classification based on embryo and seedling characters; *viz.* sub clade 1: - *Shorea* section *Doona* occupied the basal position paraphyletic to section *Anthoshorea*; section *Anthoshorea* was basal and paraphyletic to the genus *Neobalanocarpus,* which occupied an intermediate position between *Shorea* and *Hopea*; sub clade 2: - the genus *Parashorea* followed by *Shorea* section *Neohopea*, were clustered basal to the subgroup *Rubroshorea* that included *Shorea* section(s), *Rubella*, *Ovales*, *Mutica* and *Brachypterae*.

Parsimony, likelihood and Bayesian analyses (Figs. 4, 5 and 6) produced polytomies of the genera, *Vateriopsis*, *Stemonoporus* and *Anisoptera* of the tribe Dipterocarpeae each with 100% bootstrap support. The monotypic genera *Upuna* and *Vateria* were paraphyletic to *Stemonoporus*, *Anisoptera* and *Vateriopsis* although the branches were collapsed into polytomies in parsimony and neighbour joining bootstrap (>80%) analyses (Figs. 4 and 7). *Cotylelobium* and *Vatica* together formed a monophyletic group, which was paraphyletic to *Vateria* and *Upuna*.

3.4. MOLECULAR CLOCK HYPOTHESIS

The likelihood ratio test performed with (Fig. 8a) and without (Fig. 8b) enforcing the molecular clock, based on the 29-taxon dataset, accepted the molecular clocklike evolution of the Dipterocarp backbone at the alpha level of 0.05 ($-lnL = 3939.31$; $-lnL_{cl} = 3953.44$; $2[lnL_{cl} - lnL] = 28.26$; P = 0.397; DF = 27). The results of the likelihood ratio

tests performed with and without enforcing the molecular clock, based on the other datasets used in sensitivity analysis (section 2.7.4.) are included in the section 3.5.

The variance to mean ratio [*R (t)* = index of dispersion] of the sum of nucleotide substitutions based on the 29-taxon ML tree (Fig. 8b) was 0.95 indicating an underdispersed molecular clock with neutrality. Tajima's relative rate test (RRT) also accepted the null hypothesis of a clocklike evolution of the Dipterocarp backbone (Fig. 8b) with constant substitution rates between each pair of taxa with respect to their immediate outgroup (Table 10).

3.5. MOLECULAR DATING ANALYSES

Independent fossil calibration points based on the 29-taxon (Fig. 9) clocklike topology (Dipterocarp backbone) produced approximately equal age estimates with narrow confidence limits of less than 2.0 Mya in both Langley Fitch (Table 11) and penalized likelihood (PL) analyses (Table 12). Estimated ages were approximately equal when *Durio zibathinus* and *Tilia kiusiana* were used as alternative outgroups (Table 11). However, the age of the root node was relatively underestimated and the ages of the internal nodes were slightly overestimated when a distantly related outgroup, *Acer campestre* was used (Fig. 10).

The average ages of the origin of the family estimated from the independent calibration points (nodes K, O and Q) based on 29-taxon ML trees (Table 11) with outgroups, *D. zibathinus* and *T. kiusiana* were 121.2 Mya and 123.4 Mya, respectively. Based on those average values of divergence time (T), the estimated nucleotide substitution rates at synonymous sites (Rs) were 2.47×10^{-10} and 2.43×10^{-10} per site

per year respectively [Rs = Ks/2T; synonymous differences = 14.61; number of synonymous sites = 235.7; Ks = 0.06; T = 121.2 and 123.4 Mya]. The values of local substitution rates were approximately equal among lineages in the Langley Fitch analysis based on molecular clocklike 29-taxon ML trees and the values ranged between 3.1 x 10^{-10} to 2.4 x 10^{-10} per site per year and the mean value was 2.7 x 10^{-10} per site per year. In the PL analysis, which is based on a nonclocklike 49-taxon ML tree (Fig.12), the values of local rates ranged between 1.3 x 10^{-10} to 7.4 x 10^{-10} per site per year (Table 13) and the mean value was 4.3 x 10^{-10} per site per year.

Deviation from the molecular clocklike evolution was higher with the random addition of putative sister taxa (Table 14) when compared to the sequential addition of species into the *Shorea*, *Hopea* and *Stemonoporus* sub clades (Table 15). The 49-taxon tree showed a non-clocklike evolution (Fig. 11; Table 14). Deviations of the estimated ages based on non-clocklike trees when compared to those of molecular clocklike trees were proportional to the degree of rate heterogeneity of the topology under consideration. A calibration point close to the root node (node K) of 49-taxon penalized likelihood tree resulted in divergence times that are approximately equal to those of the 29-taxon Langley Fitch tree (Table 11 vs. Table 13). The other calibration points (nodes Q and O) overestimated the ages at the root part of the 49-taxon tree. The degree of overestimation of a given clade increased with its distance from the root node (Fig. 12; Table 13).

4.0. DISCUSSION

4.1. MODE OF EVOLUTION OF THE *matK* GENE IN DIPTEROCARPS

The results based on the preliminary sequence analyses showed that the *matK* gene is less functionally constrained in dipterocarps. Average nucleotide compositions of the first, second and third codon positions were approximately similar and were not biased towards the third codon position (Table 7). Similar patterns of nucleotide substitution have been observed in other studies employing the *matK* gene (Azuma *et al.*, 2001; Hilu *et al.*, 2003). The synonymous mutation rates were slightly higher than the nonsynonymous mutation rates (section 3.2) indicating a neutral mode of evolution or relaxed purifying selection with the neutrality index closer to 1.0.

Less functionally constrained genes use more synonymous codons in a randomized fashion (Sharp *et al.*, 1993). The values for the effective codon number, ENC (51.12) and codon bias index, CBI (0.393) of *matK* sequences demonstrated a less constrained mode of evolution. The value of 51.12 of ENC denotes increased randomization in overall codon bias in the scale of 18 to 61, where 18 indicates the use of only one synonymous codon and 61 indicates completely random usage (Wright, 1990). The codon bias index measures the extent to which a gene uses a subset of optimal codons (Morton, 1993). Therefore, the resulting value of 0.393 indicates a fairly random codon use in the scale of 0.0 to 1.0, where the value of 0.0 for random codon usage and 1.0 for extreme bias. Randomized codon usage in Dipterocarp *matK* sequences was also reflected by the unbiased nucleotide composition (P=1.00) across the sequences. More functionally constrained genes tend to have higher G+C contents in the third codon positions (Comeron and Aguade, 1998). Approximately equal G+C contents of the

second and the third codon positions of the Dipterocarp *matK* sequences, therefore, indicated a relatively low functional constraint.

4.2. PHYLOGENETIC RELATIONSHIPS OF DIPTEROCARPS

The assumptions of phylogenetic methodology are sometimes violated due to strong rate variation across taxa (Felsenstein, 1988; 2004), base composition and transition bias (Lockhart *et al.*, 1994; Wakeley, 1996), long branch attraction (Felsenstein, 1981; Stewart, 1993; Steel, 2002), wrong model selections (Yang *et al.*, 1994; Yang, 1997; Bruno, 1999) or a combination of these factors (Appendix 4). Nonetheless, higher congruency among the topologies inferred by different methods indicated that the *matK* sequences of dipterocarps were less likely to violate the underlying assumptions, which was largely due to the unbiased nucleotide substitution rates, as revealed by the preliminary sequence analyses. Particularly when the coding gene used in phylogenetic analyses shows a neutral mode of evolution, phylogeny reconstruction methods tend to produce more accurate results (Page and Holmes, 1998), as demonstrated by the present study.

The phylogenetic relationships among the family Dipterocarpaceae based on *matK* sequences (Figs. 4 to 6) were congruent with the current classification based on morphological characters (Maguire and Ashton, 1977; Ashton, 1982) and the previous studies based on molecular data (Tsumura *et al.*, 1996; Kajita *et al.*, 1998; Dayanandan *et al.*, 1999). In the derived phylogenetic trees (Figs. 4 to 7), the Pakaraimoideae and Monotoideae formed two distinct clades paraphyletic to the Dipterocarpoideae. The relative positions of the Pakaraimoideae and the Monotoideae demonstrated their close

affinities to the Dipterocarpoideae than to the Tiliaceae. This is in agreement with the recent classification of Maguire *et al.* (1977) that suggest close affinities of the Pakaraimoideae and Monotoideae to Asian dipterocarps. The similarities in the anatomical characteristics of the wood of Monotoids and Dipterocarpoids (Bancroft, 1935d) are also in agreement with the results. Therefore, Kosterman's (1985) suggestion for the new family Monotaceae (section 1.1.1.) was not supported by the present study.

Ancestral characters (plesiomorphic) constrain the ecological ranges of species whereas derived (apomorphic) characters are more adaptive to the specific microclimatic conditions (Maury and Curtet, 1998). The clustering patterns of most of the genera and sections of *Shorea* in parsimony, likelihood and bayesian trees (Figs. 4 to 6) suggested derived and ancestral character states for the species within the tribes Shoreae and Dipterocarpeae, respectively. The polytomies of *Vateriopsis*, *Stemonoporus*, *Anisoptera*, *Upuna* and *Vateria* in parsimony, likelihood and Bayesian trees (Figs. 4 to 6) reflected plesiomorphic character states and were indicative of parallel evolution. This is congruent with Ashton's classification (Ashton, 1982; Ashton and Gunatilleke, 1987a) that suggests the archaic positions of *Stemonoporus*, *Vateriopsis* and *Vateria*, based on primitive characters such as large, heavy, wingless fruits. The genera, *Vatica* and *Cotylelobium* formed a monophyletic group demonstrating a relatively recent origin when compared to other genera in the tribe Dipterocarpeae, which is also in agreement with Ashton's classification (1982).

The derived phylogenetic tree topologies (Figs. 4 to 6) illustrate the major evolutionary trends within the family (section 1.1.2.) particularly within the tribe Shoreae; from large to small flowers, canopy to understorey dominance and occurrence

in seasonal to aseasonal climates agreed with Ashton's classification (1982, 1989). The large flowered canopy dominant genus *Dipterocarpus* occupied the basal position in the sub family Dipterocarpoideae relative to the terminal lineages, *Shorea* and *Hopea*. Although smaller than those of *Dipterocarpus*, the flowers of *Shorea* with many stamens are comparatively larger relative to the fewer stamened smaller flowers of *Hopea* (Ashton, 1989). The clustering of taxa having longer (e.g. *Dipterocarpus*) and shorter (e.g. *Hopea*) fruit sepals followed the same trend in the *matK* phylogeny supporting the re-evolution of shorter sepals from longer sepals (Suzuki and Ashton, 1996).

In the *matK* topologies, *Shorea* section *Doona* occupied the basal position of one of the two monophyletic clades of the tribe Shoreae, which is in agreement with Ashton's (1982) classification. *Shorea* section *Doona* has been identified as the most primitive section of *Shorea* (Ashton, 1982). Section *Anthoshorea*, followed by *Neobalanocarpus heimii* followed the same trend, being paraphyletic to each other. Ashton (1982) stated that *Neobalanocarpus* has close affinities with both *Shorea* and *Hopea* occupying an intermediary position between them. Floral biology assigns *Neobalanocarpus* to *Shorea* whereas bark morphology and anthocynin development place it in *Hopea*. The floral morphologies of *Shorea* sections, *Anthoshorea*, *Hopea* and *Neobalanocarpus* also show similarities. Jong and Kaur (1979) speculated that *Neobalanocarpus* might be a result of hybridization between *Shorea* and *Hopea* species. Supporting these views, the phylogenetic trees based on *matK* sequences (Figs. 4 to 7) showed a position of *Neobalanocarpus* intermediary between *Shorea* section *Anthoshorea* and the *Hopea* crown lineage.

My results do not resolve the position of *Dipterocarpus* in the sub family

Dipterocarpoideae. *Dipterocarpus* appears closer to the tribe Shoreae and *Dryobalanops* occupies a position between the two tribes (Figs. 4 to 7). The unresolved positions of *Dipterocarpus* and *Dryobalanops* were also detected in previous studies based on molecular data (Kajita *et al.,* 1998; Dayanandan *et al.,* 1999). Morphological characteristics indicate the intermediate position of the genus *Dryobalanops* (Ashton, 1982; Maury and Curtet, 1998). Although it has been classified in the tribe Shoreae, based on chromosome number, thickened fruit sepal base, connate petals and the grouped resin canals (Ashton, 1979), the presence of solitary vessels and valvate fruit sepals assign it to the tribe Dipterocarpeae (Maury, 1978). Similarly, the presence of solitary resin canals and the base chromosome number of 11, support the placement of *Dipterocarpus* in the tribe Dipterocarpeae. Ashton (1982) stated that the similarity in chromosome number, however, is not necessarily an unequivocal indication of taxonomic affinity. Besides, morphological evidence supports considering *Dipterocarpus* as a basal lineage of the sub family Dipterocarpoideae (Meijer, 1979). Large flowers, winged free calyx tube and pollination syndromes characterize *Dipterocarpus* as a unique lineage retaining advanced characters, as a canopy dominant (Whitmore, 1998; Ashton, 1982).

Overall, the phylogenetic tree topologies based on *matK* nucleotide sequences were concordant with Ashton's (1982) speculations based on morphological characters and provided a solid skeleton for the estimation of species divergence times.

4.3. CONFIDENCE LEVELS FOR THE ABSOLUTE DIVERGENCE TIME ESTIMATES

The limitations of molecular dating methods are reviewed by Sanderson and Doyle (2001). Major sources of error in molecular dating analyses are deviations from

the molecular clock (Kumar and Hedges, 1998) and inaccurate fossil calibrations (Donoghue, *et al.*, 1989). The results of the tests that I used to evaluate molecular clocklike evolution strongly suggested the clocklike evolution of the *matK* gene in dipterocarps up to the levels of genera and sections of *Shorea* (Fig. 9; Table 11). The local substitution rates of different lineages in the 29-taxon dipterocarp backbone topology ranged from 2.4×10^{-10} to 3.1×10^{-10} per site per year and were relatively constant among lineages compared to the rate variation among the lineages of the 49-taxon tree, which ranged from 1.3×10^{-10} to 7.4×10^{-10} per site per year. Thus, the ages of dipterocarps estimated using molecular clocklike *matK* topologies based on independent calibration points were less prone to errors associated with rate heterogeneity (Table 11 vs. Table 13).

Despite the high heterogeneity of substitution rates observed in the 49-taxon PL tree, the earliest fossils of *Anisoptera* used to calibrate the node K notably minimised calibration errors and thus resulted in equal ages with narrowed confidence limits (<2 Mya) in both LF (29 taxa; Table 11) and PL (49 taxa; Table 13) analyses. Although the other two calibration points, based on *Hopenium* (node Q) and *Vaticoxylon* (node O), generated highly consistent age estimates based on the 29-taxon clocklike ML trees (Table 11), ages estimated based on the 49-taxon penalized likelihood tree were inconsistent at the root part, relative to the ages estimated from the node K (Table 13). Similar results obtained in angiosperm molecular phylogenetic and dating studies (Wikstrom *et al.*, 2001) pinpoint the inaccuracy of calibrating within the parts of the tree that are far away from the root node (Bremer, 2002). Sanderson and Doyle (2001) suggested that the rate heterogeneity of some basal branches in the angiosperm

phylogeny based on the *rbcL* gene, led to an overestimation of the ages of the angiosperm ancestors. As demonstrated in the present study, the limitation of using terminal calibration points in dating distant nodes of a phylogenetic tree may be the lineage specific heterogeneity of substitution rates or deviation from molecular clocklike evolution.

Results of the sensitivity analyses showed that the deviation from 'clockness' was greater with random (Table 14) rather than sequential (Table 15) addition of terminal taxa into the species rich clades. In sequential addition, higher deviations from the molecular clock were apparent upon the addition of sister taxa to the *Shorea* and *Hopea* clades than to *Stemonoporus*. Upon addition of sister species to the tribe Shoreae, the species generally clustered with them into sub clades of *Shorea* and *Hopea* species. In contrast, the tree topology did not change upon addition of sister species into the tribe Dipterocarpeae. Thus, it was evident that the species in the tribe Dipterocarpeae were less affected by substitutional noise (lineage effects) compared to the species of the tribe Shoreae. This may be due to higher variation of nucleotide substitution rates among lineages with derived (apomorphic) characters (*Shorea, Hopea*) compared to those with primitive (plesiomorphic) characters (*Vateriopsis, Stemonoporus*). Some characters such as the valvate fruit sepals of the members of tribe Dipterocarpeae are considered to be primitive when compared to those of the members of the tribe Shoreae (section 1.1.2.).

However, the sequential addition of *Shorea* and *Hopea* species resulted in molecular clocklike topologies up to the total number of 34 taxa (Table 15), suggesting that the lineage specific substitution rates of most of the *Shorea* and *Hopea* species alone, do not contribute to the rejection of the molecular clock in the 49-taxon tree. Thus, it is

evident that the random sampling of lineages adds more rate heterogeneity to the topology, despite the clocklike evolution of some taxa clustered among them. However, the internal nodes were not essentially sensitive to the rate heterogeneities triggered by either sequential or random additions of taxa. Therefore, the ages estimated from node K based on the 49-taxon tree were similar to those of the 29-taxon tree and those of the other topologies generated in the sensitivity analyses (Table 14; Table 15).

4.4. EVOLUTIONARY BIOGEOGRAPHY OF THE FAMILY DIPTEROCARPACEAE

The contemporary disjunct distributions of many taxonomic units reflect the historical influence of geologic and climatic events. The historical range expansions or reductions are explicitly correlated with a sequence of favourable or detrimental climatic events and the emergence or elimination of geographical barriers (Croizat, 1974; Humphries, 1999). In order to understand the biogeographic history, therefore, the ecological and biological traits of a taxonomic unit should be assessed together with paleontological and geological evidence in the context of evolution. Phylogeneticists often hypothesize that the historical patterns could repeatedly be tested with other data and alternative analyses (Skelton, 1993). Morley and Dick (2003) suggest that the dispersal routes and the paucity of Gondwanic plant fossils should be key considerations in reconstructing the biogeographic history of pantropical taxa. One of the major implications of phylogeny and molecular dating analyses in historical biogeography is the testing of congruency between inferred cladogenic events and the sequence of geological and ecological events, employing paleontological information.

In order to corroborate the evolutionary biogeography of the family Dipterocarpaceae, therefore, the results of phylogenetic and molecular dating analyses were compared with morphology and ecology (Ashton, 1982, 1988; Maury and Curtet, 1998), historical floristic interplate dispersal routes based on paleontological evidence (Morley, 1998, 1999, 2003) and the results of similar studies based on the angiosperm origin and evolution (Sanderson and Doyle, 2001; Wikstrom *et al.,* 2001; Soltis *et al.,* 2002).

4.4.1. The origin and interplate dispersal of family Dipterocarpaceae

Molecular dating analyses along with present biogeography and fossil data revealed the early Cretaceous origin of the Dipterocarpaceae in the Western part of Gondwanaland, supporting the Gondwanan vicariance hypothesis for the evolution of the Asian rain forest flora. The warm and humid climatic conditions that are considered to have prevailed during the Cretaceous to early Eocene periods of western Gondwanaland may have been highly favourable for the radiation of tropical and sub tropical flora (Furley and Newey, 1983; Morley, 1999).

The origin of ancestral dipterocarps and divergence of the three sub families coincide with the historical interplate dispersal events, paleoclimatic events and with the fossil records. According to the palynological records (Doyle, 1978; Morley, 1998, 2003; Soltis, 2001), angiosperm pollen was not common until the Hauterivian, 135 - 132 millions of years ago (Mya) or Barremian (132 - 124 Mya). Although the Cretaceous fossil pollen of ancestral dipterocarps has not been reported, it is classified in the triaperturate class (Ashton, 1982). The emergence of eudicots has been demarcated as

the second wave of angiosperm radiation, based on the appearance of triaperturate-derived (e.g. tricolporate, tricolpate) pollen that has commonly been reported in the Aptian (113 – 108 Mya; Fig. 12a). The age of the first Dipterocarp ancestors (node A: Pakaraimoideae crown lineage), 121 – 117 Mya, nearly coincides with the appearance of triaperturate pollen, which was highly congruent with the time of emergence of the eudicot angiosperms. The generalised tricolporate pollen of *Pakaraimaea* proves its archaic status within the family compared to the specialized walls and exine structures of Monotoid and Diptrocarpoid pollen (Ashton, 1982).

The adaptation of extant Pakaraimoids to thrive in drier conditions may be related to their time of origin under drier paleoclimatic conditions (Scotese, 2000), which prevailed in the early to mid Cretaceous. The estimated ages on the Dipterocarp origin were also congruent with the estimated ages on the origin of angiosperms based on *rbcL* sequences (190 – 140 Mya; Sanderson and Doyle, 2001) and the origin of tricolpates based on *18S* rDNA sequences (147 – 131 Mya; Wikstrom *et al.*, 2001).

The age of 119 Mya for the Pakaraimoideae – Monotoideae split (node A) and the emergence of the Monotoideae crown lineage (node B) supported the existence of direct land connections between the South American and African plates, which facilitated direct dispersals until the end of Albian (96 Mya) and stepping stone dispersals up until late Maastrichian (70 Mya). Isolation of the Monotoid genus *Pseudomonotes* in South America was also in agreement with the estimated ages that support the origin of 'dipterocarp ancestral stock' (Ashton, 1982) before the complete separation of South American and African plates (Fig. 12).

The appearance of the pollen of the members of plant families such as the Euphorbiaceae, Bombacaceae and Leguminosae in the Paleocene and Eocene sediments of Africa (Morley, 1999) suggests the existence of luxurious tropical rain forests throughout the Cretaceous and early Tertiary periods in the African continent (Raven and Axelrod, 1974; Davis *et al.*, 2002). Most of these forests have been replaced by Savannas since the middle Miocene (Fig 13d), as a consequence of accelerated aridification (Davis *et al.*, 2002). The fossils of *Dipterocarpoxylon africanum* from East Africa (Bancroft, 1935b) suggest the presence of Dipterocarpoids in African forests until the middle Miocene. Rain forest species that dominated aseasonal forests in Africa and Madagascar, therefore, might have become extinct in the upper Tertiary while other taxa, such as *Monotes*, may have adapted to thrive in dry deciduous forests.

At the time of the initial radiation of angiosperms (>132 Mya) the Indian plate was located too far south (Fig. 12a) for interplate dispersals (Morley, 2003). Once separated from Southern Gondwanaland in the Aptian (113 –108 Mya), the Indian plate had drifted rapidly northward and by the Turonian to early Maastrichian (92 – 74 Mya) it was positioned in a close proximity to Madagascar (Fig. 12b) facilitating many plant dispersals from Africa via Madagascar (Morley, 2003). The approximate age of the origin of Dipterocarpoids (stem lineage, node B) of 88.0 Mya therefore, undoubtedly is in agreement with the existence of Dipterocarpoid ancestors in Africa before the Maastrichian (74 Mya), that have dispersed from Africa to the Deccan plate.

4.4.2. Diversification of the sub family Dipterocarpoideae

The extensive volcanism at the Cretaceous-Tertiary (KT) boundary resulted in mass extinctions of many fauna and flora leading to population bottlenecks, evidenced by indiscernible fossil data or so-called 'missing links' (Bossuyt, 2001). The unresolved ancestral status of the extant basal lineages, *Dipterocarpus* and *Dryobalanops* suggest plausible missing links from the Turonian (late Cretaceous) to the earliest Eocene (from node B to C). The approximate ages of 88.0 – 54.0 Mya (node D to C) of the Dipterocarpoideae stem and crown lineages (Fig. 15) are highly congruent with the warmest periods of the late Cretaceous and early Tertiary (Morley, 2003). Despite the paucity of dipterocarp fossils in the Cretaceous strata, the divergence times of major clades therefore, correspond to the late Cretaceous (100 – 70 Mya; Fig. 12b) to early Eocene (60 – 49 Mya; Fig. 13b) thermal maximum (warmest period) that facilitated extensive dispersal of tropical and sub tropical flora as high as 40^0 N polewards (Morley, 2003).

The emergence of the tribe Dipterocarpeae (stem lineage: node K, 57.1 – 52.7 Mya) also correspond with the Eocene thermal maximum, which agrees with the age of the most recent common ancestor of *Vateriopsis*, *Stemonoporus* and *Anisoptera*. As discussed in the previous sections, the polytomies detected through parsimony, likelihood and Bayesian analyses suggest parallel evolution of those taxa from a single common ancestor. Dating estimates indicate the presence of their common ancestor (Node K) in the India-Seychelles landmass in the Early Eocene before the collision. The divergence times and the isolation of *Vateriopsis* ancestors in the Seychelles corroborate the theory that a common ancestor, present in the Eocene era, has given rise to other plesiomorphic

ancestors throughout the India's northward movement (Fig. 13). Moreover the earliest fossils of *Vateria* (*Vaterioxylon*) found in early Miocene sediments of North India together with the estimated age of 21.5 Mya suggest parallel evolution of *Vateria* and *Upuna*, which most probably had shared ancestors with *Stemonoporus* and *Vateriopsis*. The independent evolution of shared characters in these lineages may be correlated with their microhabitats and microclimatic conditions, to which they may have been exposed independently during the historical range reduction.

The ancestral Gondwanan elements of the Deccan plate (later the Indian plate) had extremely affected by early Tertiary volcanism, latitudinal and climatic changes and extensive late Tertiary aridification following the uplift of Himalayan chain and, later, the cycles of aridity during the Quaternary glaciations (Raven and Axelrod, 1974; Ashton and Gunatilleke, 1987a; Morley, 2003). Therefore, its vegetation has changed erratically from the early Tertiary onwards (Morley, 1999). Consequently, dipterocarp ancestors adapted intermittently to seasonal conditions rather than aseasonal tropical forests. This is in agreement with the hypothesized evolutionary trend of dipterocarps from seasonal to aseasonal climates as evidenced by the synchronous mass flowering and mast fruiting of many extant dipterocarps and their correlation with El-Nino southern oscillation events, which is superfluous in the aseasonal tropics where they currently flourish (Ashton, 1988).

The estimated divergence times indicate the existence of a common ancestor of the *Shorea* section *Doona* in the Deccan plate before its collision with Eurasia. This suggests that the other *Shorea* species should also have dispersed to the Eurasian plate through the Deccan plate (Fig. 13, *c* and *d*). The age of the *Shorea* stem lineage that

consists of two monophyletic groups, in one of which *Doona* was basal, extended back to 54.0 Mya (node D) and the minimum age of the crown lineage was not less than 34.4 Mya (node E). This agrees with Ashton (1982, 1988) who stated that section *Doona,* endemic to Sri Lanka, is the most primitive lineage of *Shorea.*

The Deccan plate provides a classic example of a geographically isolated Noahs' ark' since it has rafted the late Cretaceous dipterocarp elements (Figs. 12 and 13) from Africa and Madagascar to South East Asia (Morley, 1999). The same pattern of dispersal from Africa or Madagascar via the Deccan plate to South East Asia has been observed in the dispersal history of other tropical families such as the Dilleniaceae, Clusiaceae, Monimiaceae, Myrtaceae and Crypteroniaceae (Ashton and Gunatilleke, 1987a; Morley, 1999; Conti *et al.,* 2003).

4.4.3. Dispersal of dipterocarps from the Deccan plate to South East Asia

Dipterocarp fossils have commonly been reported from the Neogene sediments of the Deccan intertrappean beds of India (Lakhanpal, 1970). The widespread occurrence of *Dipterocarpus* and *Dryobalanops* pollen in the late Oligocene and the earliest Miocene sediments (30 – 25 Mya) of Borneo (Muller, 1970) depict the dispersal of dipterocarp elements from the Deccan plate (India) to South East Asia via the dispersal corridors of East Asia (Figs. 13c and 13d). Geochemical fossils that have become common in the middle Eocene (44 Mya) sediments of Myanmar (Aarssen *et al.,* 1990; Morley, 1999) demarcate a possible path for those historical dispersal events.

Geological records indicate that the lowering of sea levels linked the islands of the Sunda shelf until the mid Holocene (0.01 Mya/10,000 years ago) (Morley, 1999;

Holloway and Hall, 1998). It has been postulated that, in the absence of physical barriers and under favourable climatic conditions, the tropical species including dipterocarps dispersed throughout the Sunda shelf from the Oligocene onwards. During the Pleistocene glaciations, many flora and fauna dispersed or migrated from higher latitudes to refugial areas in South East Asia (Morley, 1998). According to the refuge concept (Haffer, 1987), the tropical forest species have evolved by isolation in areas that remained stable. This may also be attributed to the fact that most tropical biodiversity hotspots are situated in climatically stable areas that have been buffered during the Pleistocene climatic fluctuations (Holloway and Hall, 1998). The melting of polar ice caps at the dawn of the present interglacial period however, led to the fragmentation of Sundaland into a series of islands (e.g. Borneo, Java, Sumatra, Celebes, Moluccas) fragmenting dipterocarps into several land areas, as they appear today.

My study has demonstrated that the family Dipterocarpaceae had already diverged into its genera and major sections by the end of the Miocene. The approximate ages of 32.4 Mya (node G) and 28.5 Mya (node I) of the ancestors of species rich *Rubroshorea* stem lineage and the *Anthoshorea* crown lineage respectively, can be considered as initial demarcations of the Oligocene radiation and subsequent dispersal of dipterocarps from the Deccan plate to South East Asia (Fig. 13c). Interestingly the Sri Lankan representative of *Shorea* section *Anthoshorea* (*S. stipularis*) occupied the basal position (node AN1, 10.4 Mya), showing an early divergence (Fig. 11; Table 13), whereas the Malaysian species, *S. assamica* and *S. bracteolata* showed a relatively recent radiation (node AN2, 6.1 Mya). *Rubroshorea* stem and crown lineages showed a similar trend.

The monospecific *Shorea* section *Neohopea* (*S. isoptera*), restricted to Northern Borneo, showed an early divergence (28 Mya) compared to the other sections (*Rubella, Ovales, Mutica, Brachypterae*). The dispersal of a few Dipterocarp elements to the East of Wallace's line is explained by the formation of the Makasar straits between Borneo and Sulawesi in the middle Miocene, followed by the collision of the Sahul shelf with the Sunda shelf and the formation of New Guinea (Morley, 1999), by which time the major speciation events had already taken place. This is in agreement with the view that the dipterocarp flora to the west of the Wallace's line may include recent formations (Ashton, 1969, 1982).

4.5. DIPTEROCARP ENDEMISM IN THE TEMPORAL SCALE

The origin and maintenance of a high degree of endemism in the tropical flora have been controversial (Whitmore, 1998). The isolation of taxa in a spatial scale alone does not provide causal evidence for endemism without the relative times of divergence. The present study provides a framework to assess the distribution of endemics in the temporal scale in relation to their spatial patterns, which is highly valuable for formulating conservation strategies. Endemic species are categorized as recently derived endemic species (neoendemics) or survivors of ancient lineages (paleoendemics) that have formerly had a wider range and geographically limited due to vicariance events (Stace, 1989). The members of the family Dipterocarpaceae are classic examples that exhibit both paleoendemism (e.g. *Stemonoporus, Vateriopsis, Upuna*) and neoendemism (e.g. *Hopea, Shorea*).

The estimated ages were highly congruent with the spatial distribution of endemic species (Fig.11; Table 13). Some species of *Hopea* (*Hopea* crown lineage, HO4 in Sri Lanka) and *Shorea* (*Rubroshorea* crown lineage in East Asia) are considered to be neoendemics that demonstrated shorter evolutionary spans (less than 5.0 Mya) relative to the longer evolutionary spans of paleoendemics such as *Upuna borneensis* (node M, 20.5 Mya) and *Vateriopsis seychellarum* (node L, 25.0 Mya). The isolation of *Upuna* in Borneo can thus be attributed to its ancient origin and subsequent range reduction. The study shows that the genus *Stemonoporus* and *Shorea* section *Doona* are unequivocal examples of paleoendemism. The monospecific *Shorea* section *Neohopea* that shows an early divergence (node NH; 26.6 Mya), which is isolated in Northern Borneo could also be considered as a paleoendemic species.

4.6. THE PATTERN OF SPECIATION IN THE FAMILY DIPTEROCARPACEAE

Species richness in the tropics has mostly been ascribed to the gradual accumulation of species over a long geological period in stable equatorial climates (museum model; Stebbins, 1974; Barmley *et al.*, 2004). On the other hand, some studies revealed a rapid and recent speciation in response to late tertiary geological events and unstable Pleistocene climates (Richardson *et al.*, 2001). Estimated divergence times provide a solid basis for testing between these two hypotheses (section 1.2.5.).

The molecular dating analyses revealed low rates of mutation in the *matK* gene and proved that the family Dipterocarpaceae had already evolved into genera and sections by the Miocene/early Pliocene (Fig. 14). This is congruent with Ashton's (1969) speculations on the speciation of dipterocarps. He mentioned that dipterocarps are long-

lived species with a more than 60-year life cycle that maintain a high numerical consistency in their chromosomes. In the species with longer generation times and lower rates of mutation, the probability of fixation of deleterious mutations is relatively low (Ohta, 1992). Therefore, the molecular clocklike evolutionary pattern closely matches with the speciation of taxa with longer generation times (Page and Holmes, 1998), such as the dipterocarps.

Different lineages within the Dipterocarpaceae, however, have substantially different levels of species diversity, which show distinct phytogeographical distribution patterns with several series of endemic groups, and the size of these series often correlates with the total species richness in a given habitat (Ashton, 1988). The present day tree species diversity in the South East Asian rainforests could be attributable to species richness in the genera *Shorea* and *Hopea*, which may have diversified recently. The expansion of dipterocarp ancestors into vacant niches in the South East Asia upon collision of the Deccan plate, as well as climatic fluctuations during the Pleistocene era, may have played a significant role in contemporary species richness.

Thus it can be concluded that the museum model is applicable to speciation in the family Dipterocarpaceae up to the level of sections of *Shorea*. Rapid and recent speciation may be prevalent in the terminal lineages, particularly within sections of *Shorea* and *Hopea*.

5.0. CONCLUSIONS

The evolutionary patterns and the divergence times for dipterocarps estimated using molecular data are highly congruent with ecological and morphological data (Ashton, 1969, 1982) as well as with the Cretaceous and Tertiary paleogeographical and paleoclimatic events that facilitated historical interplate dispersal of angiosperms (Morley, 1999; 2003). The origin of dipterocarp ancestors in seasonal environments in the early to mid Cretaceous and diversification in the late Cretaceous to early Eocene was parallel with the second wave of angiosperm radiation. Late Cretaceous and early Tertiary warm climatic conditions, tectonic settings and interplate dispersal routes influenced the speciation events within the species rich subfamily Dipterocarpoideae. It can be concluded that the current diversification of some of the dipterocarps is a result of gradual evolution from widespread historical populations into stable but narrow ecological niches. The integration of fossil evidence comparative morphology and biogeography with molecular phylogenetic and dating analyses of dipterocarps indicated a Cretaceous Gondwanic origin of the Asian rain forest flora.

Even though the present distribution pattern is different from that of the Cretaceous period, some dipterocarps in the aseasonal tropics of South East Asia represent their historical counterparts, being the nearest living relatives or so called 'living fossils', as evidenced by their gradual pattern of speciation. Thus, the ecological, biological and biogeographical uniqueness of dipterocarps is not paralleled by any other taxonomic group of comparable size.

Dipterocarp forests however, have rapidly been declining, despite the increasing awareness of conservation issues, since forestry related activities play a major economic

role in the far Eastern tropics (Appanah, 1998; FRA 2000). Although these are exploited extensively for non-timber forest products (Shiva and Jantan, 1998), large-scale timber extraction has become the most destructive process in the South East Asian Dipterocarp forests (Appanah, 1998).

Conservation of rain forests not only involves the assessment of existing floristic diversity but also, most importantly, it entails an understanding of the origin and evolution of the rain forest flora and fauna, which has shaped the contemporary biodiversity. Molecular level phylogenetic structuring of species and populations attentively addresses several questions involved in conserving groups of taxa as evolutionary significant units or ESUs (Moritz *et al.*, 1994; Ortega *et al.*, 1996; Avis, 1998; Birmingham and Moritz, 1998). Phytogeographical information of ESUs in the spatial scale alone is of limited value in evaluating their relative importance. The present study demonstrated the significance of assessing the evolutionary history of an organism in a temporal scale, which opens up multidisciplinary research areas that would help to explain ESUs in the phylogenetic context.

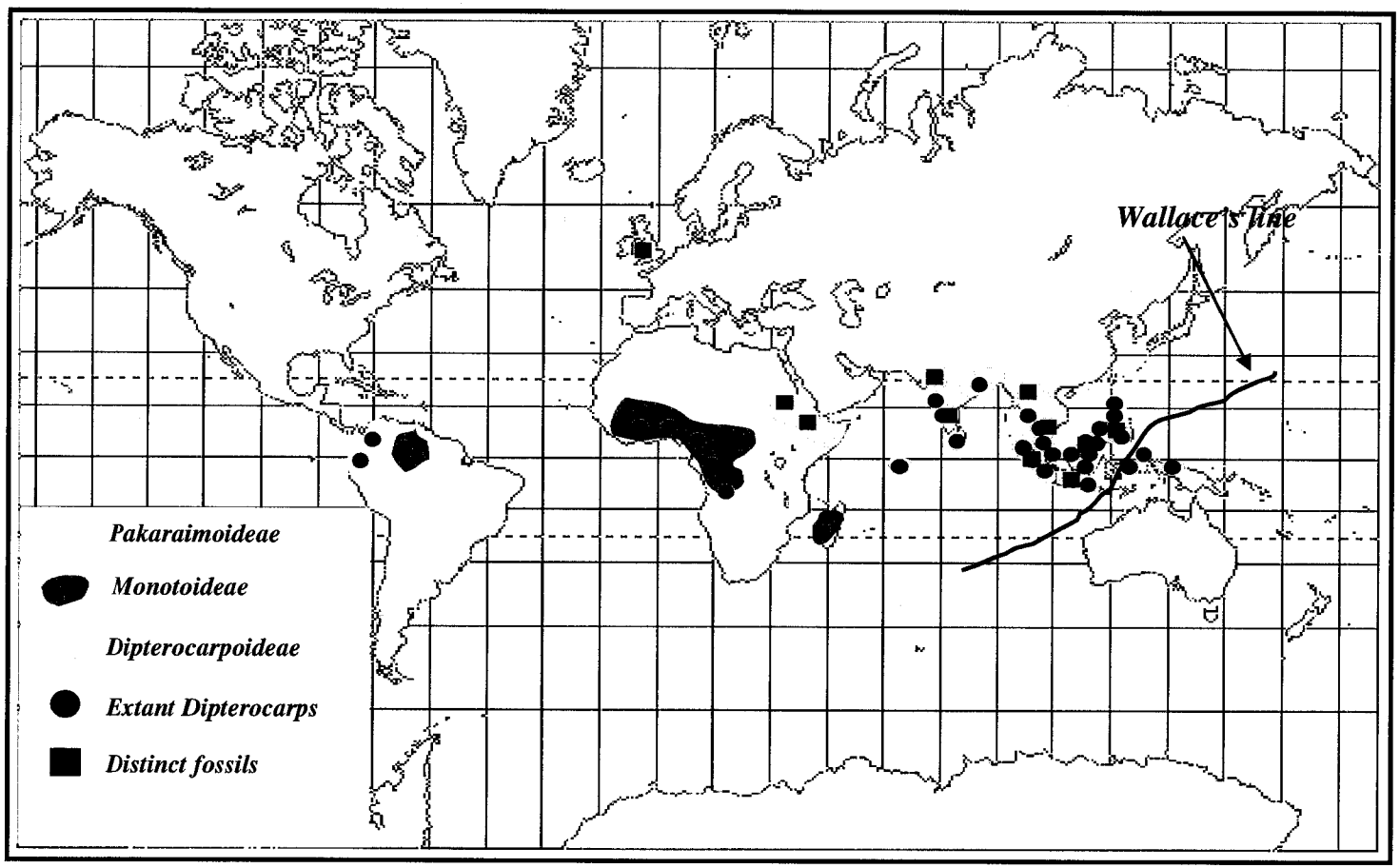
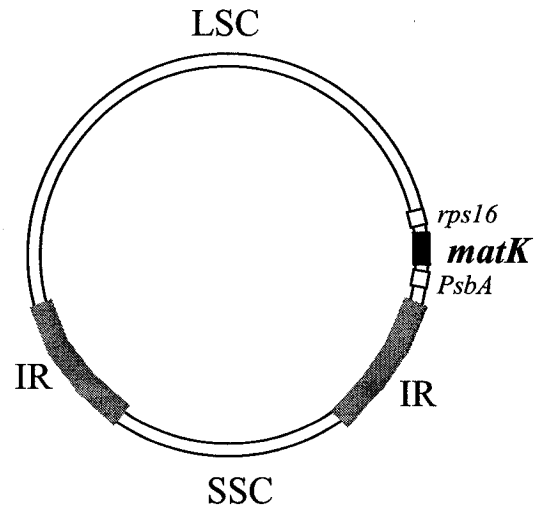


Fig.1 Contemporary distribution of extant members and fossils of the family Dipterocarpaceae.
Extant Dipterocarps:- Pakaraimoideae in South America. Monotoideae in South America, Africa and Madagascar. Dipterocarpoideae in Seychelles, India, Sri Lanka, Chittagong, Burma, Indochina, Malesia, Borneo, Thailand, Philippines, Indonesian islands, Sulawesi and New Guinea. *Dipterocarpoid fossils:-* East Africa, London, North and South India, Burma, Thailand, Vietnam, Java, Sumatra and Borneo

a)



b)

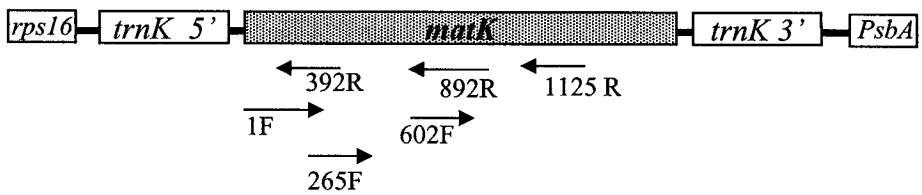


Fig. 2 *a)* Diagram of the chloroplast genome typical to most land plants, showing the relative location of the *matK* gene within *trnK* intron; neighboring genes= *rps16* and *PsbA*; LSC = Large single copy region; SSC = Small single copy region; IR = Inverted repeat.

b) Relative positions and directions of primers used to amplify and sequence 1200 bp region from the 5' end of the *matK* gene of dipterocarp species.

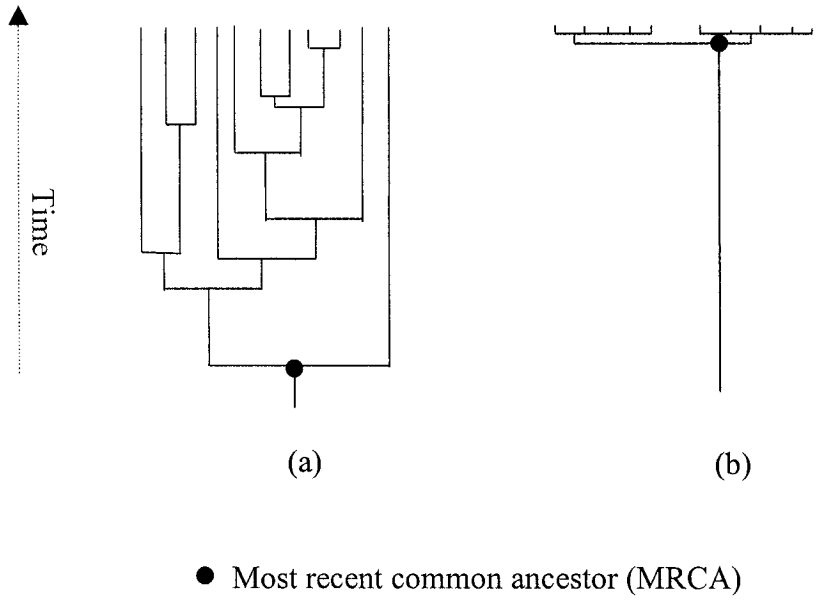


Fig. 3 Phylogenetic tree shapes and structures showing, *a)* gradual accumulation of species through time (museum model) from the most recent common ancestor of the extant species that results a well resolved phylogeny and *b)* rapid and recent diversification from the MRCA resulting a poorly resolved phylogeny (Richardson, *et. al.*, 2001).

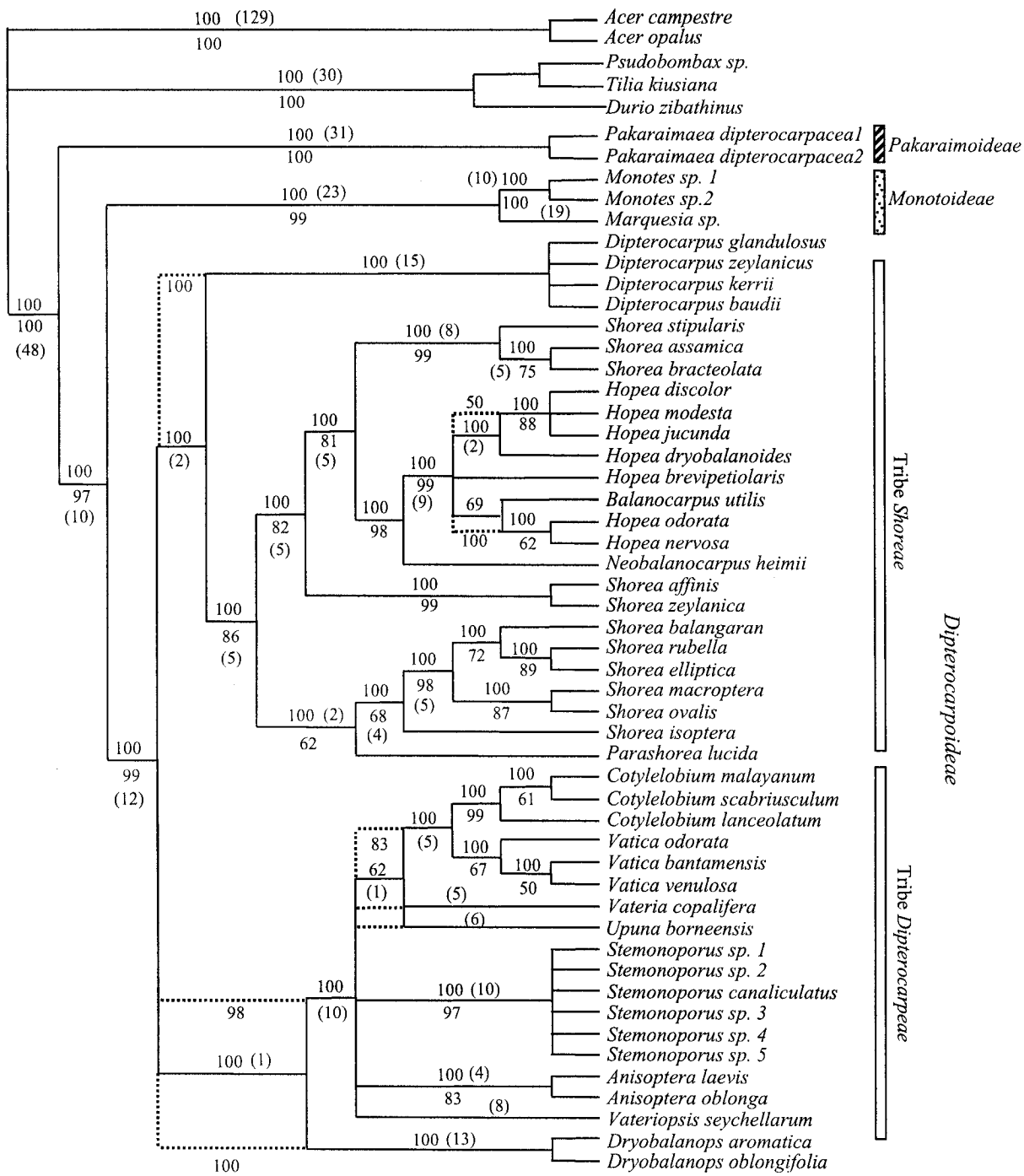


Fig. 4 The maximum parsimony tree of 45680 equally parsimonious trees based on *matK* nucleotide sequences of 49 dipterocarp species and five outgroups. Numbers above the branches show the clade credibility values of 50 % majority rule consensus and numbers below the branches indicate bootstrap values (%). Numbers within brackets are branch lengths of major clades. Dotted lines indicate the branches collapsed in the bootstrap analysis. See Table 5 for a detailed classification of species.

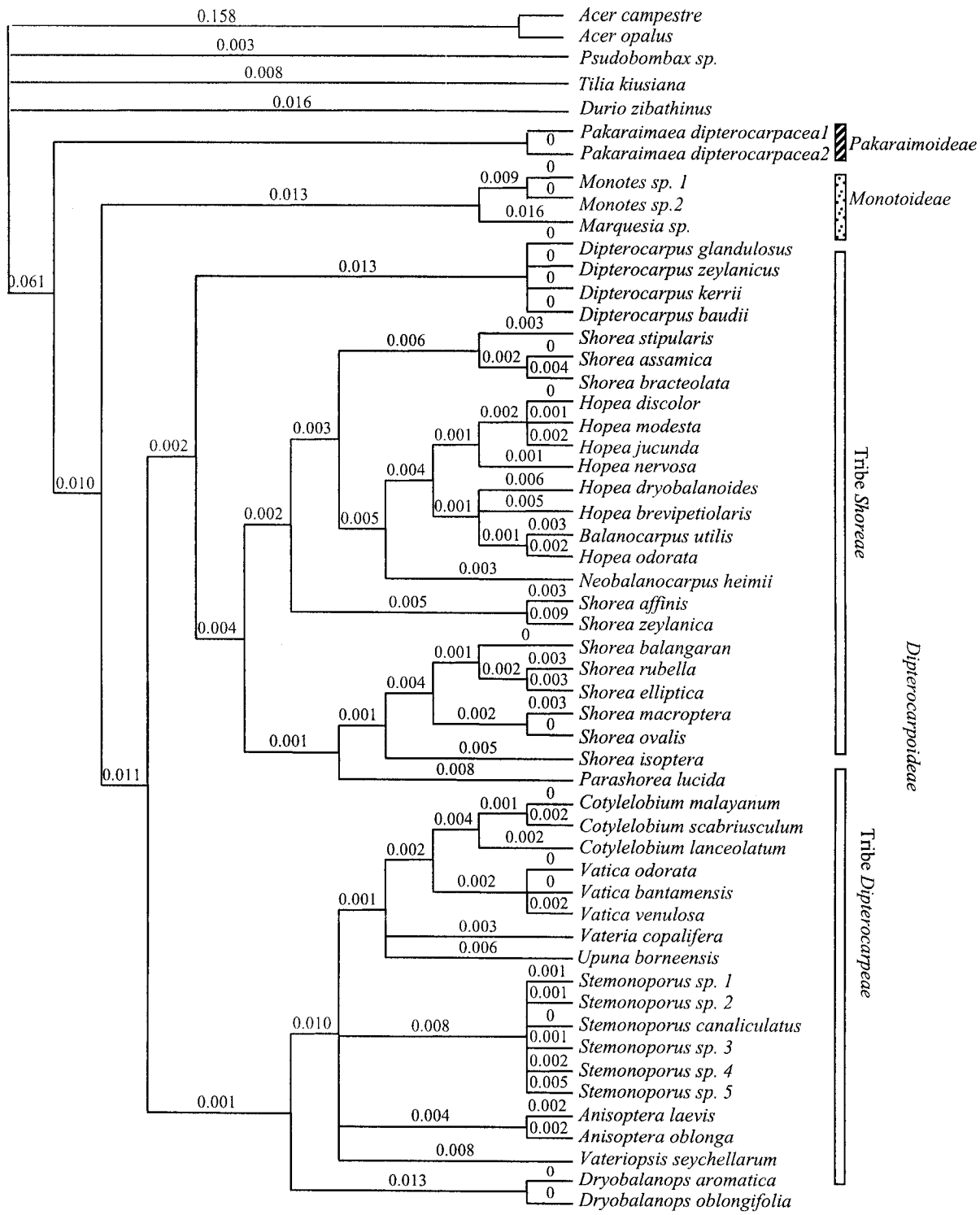


Fig. 5 The maximum likelihood tree based on *matK* nucleotide sequences of 49 dipterocarp species and five outgroups. Numbers on the branches are nucleotide substitutions. See Table 5 for a detailed classification of species.

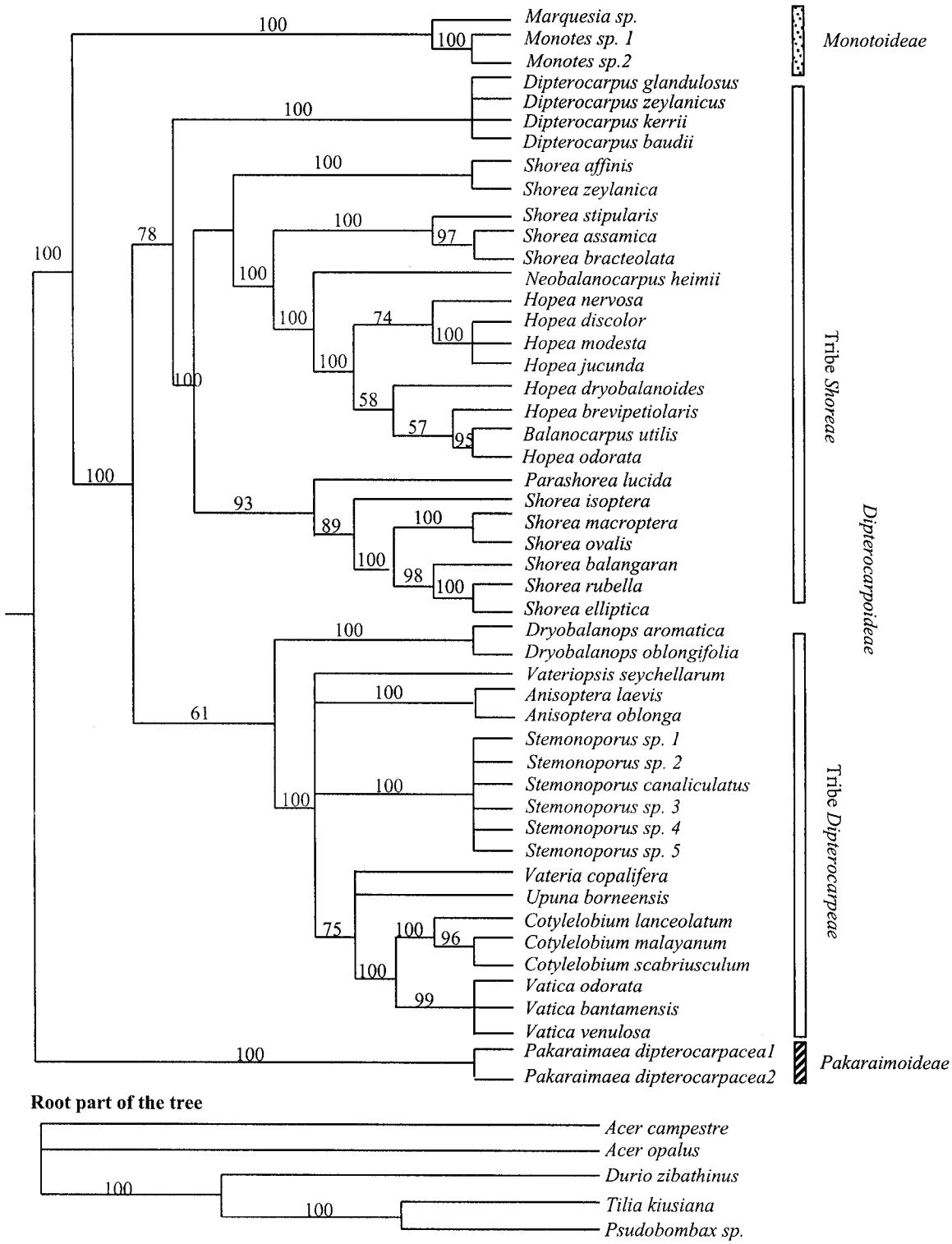


Fig. 6 The Bayesian 50% majority rule consensus tree based on the *matK* nucleotide sequences of 49 dipterocarp species and five outgroups. Numbers on the branches show the percentage clade credibility values (frequency of occurrence).

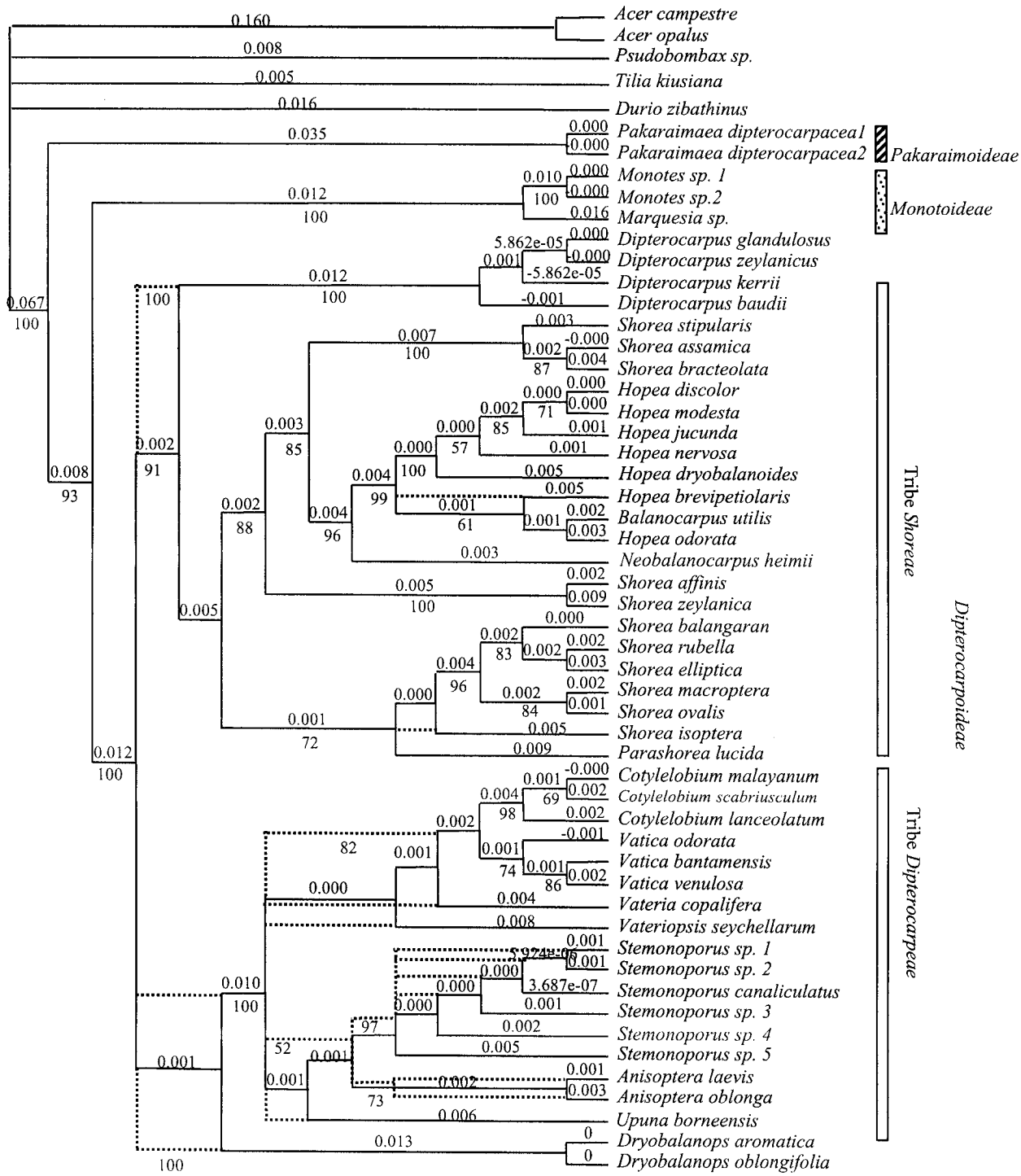


Fig. 7 The neighbour joining tree based on *matK* nucleotide sequences of 49 dipterocarp species and five outgroups. Numbers above the branches show the branch length and numbers below the branches indicate bootstrap values (%). Dotted lines indicate the branches collapsed in the bootstrap analysis. See Table 5 for a detailed classification of species.

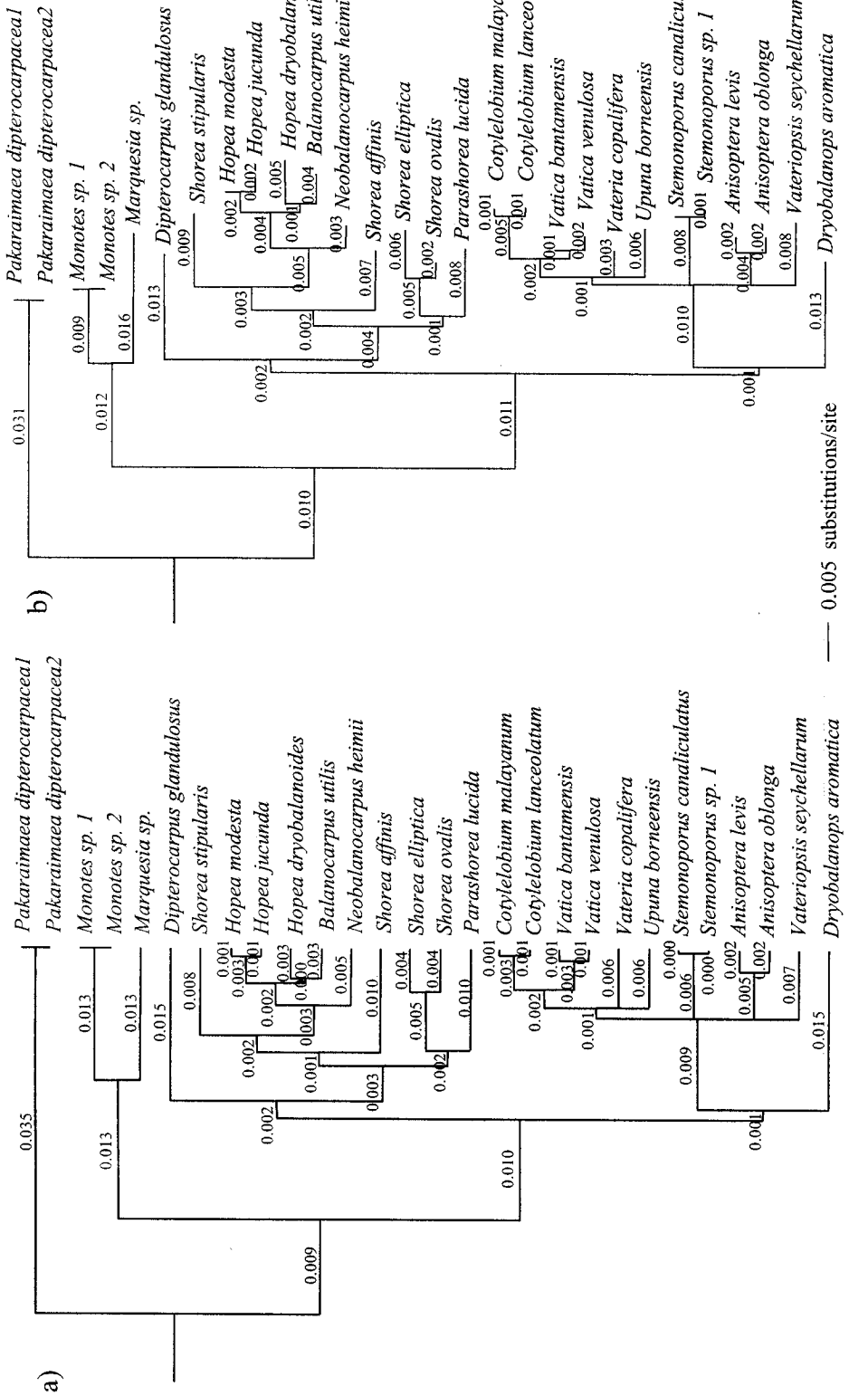


Fig. 8 Maximum likelihood trees based on *matK* nucleotide sequences of 29 taxa obtained with (a) and without (b) enforcing the molecular clock. Numbers on the branches indicate nucleotide substitutions.

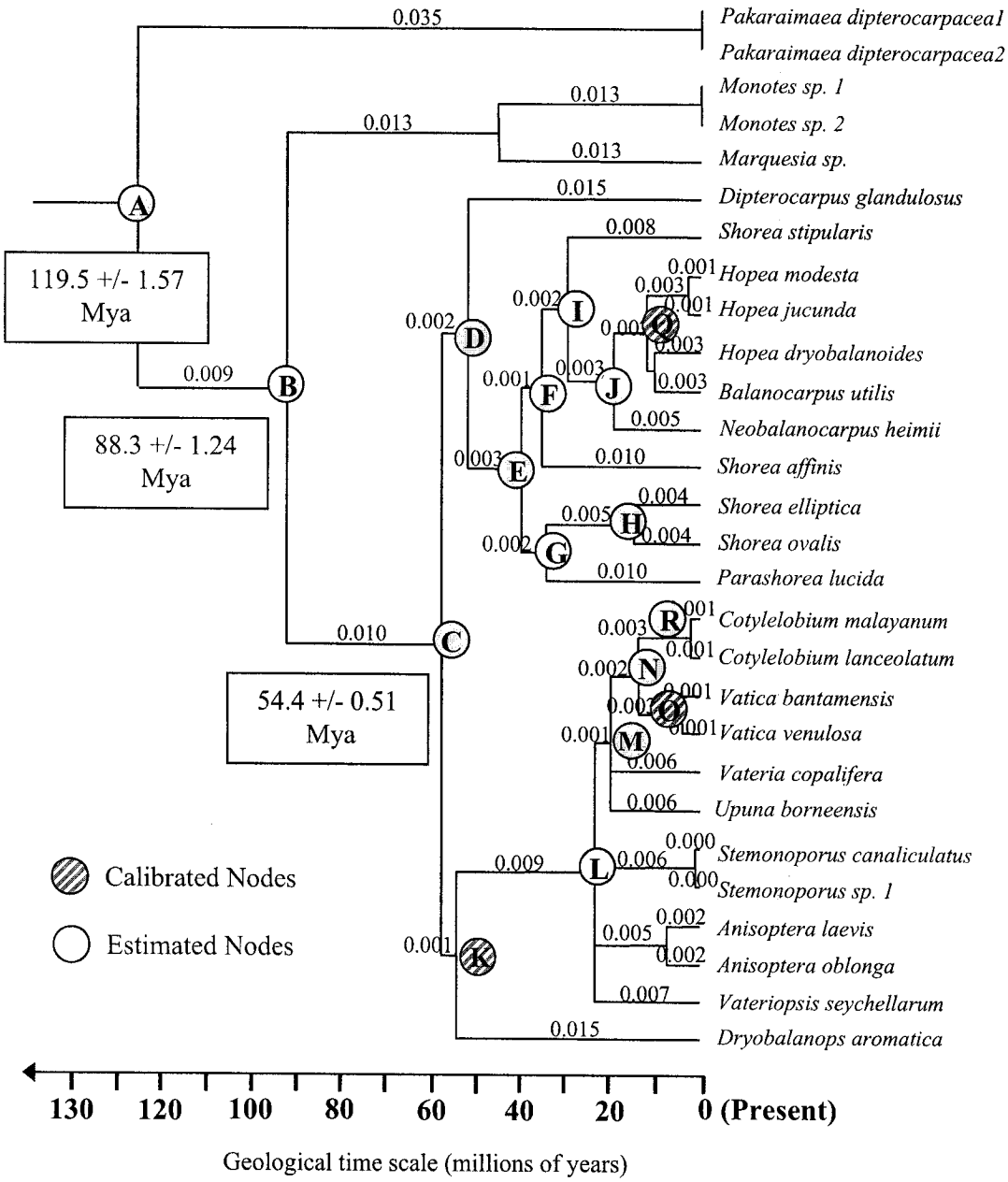


Fig. 9 Molecular clocklike maximum likelihood tree based on *matK* nucleotide sequences of 29 taxa with branches proportional to time. Numbers on the branches are nucleotide substitutions. Numbers in the boxes show the estimated ages (millions of years ago) with confidence limits of the crown lineages of, Pakaraimoideae (A), Monotoideae (B) and Dipterocarpoideae (C). See Table 11 for the ages of the other nodes.

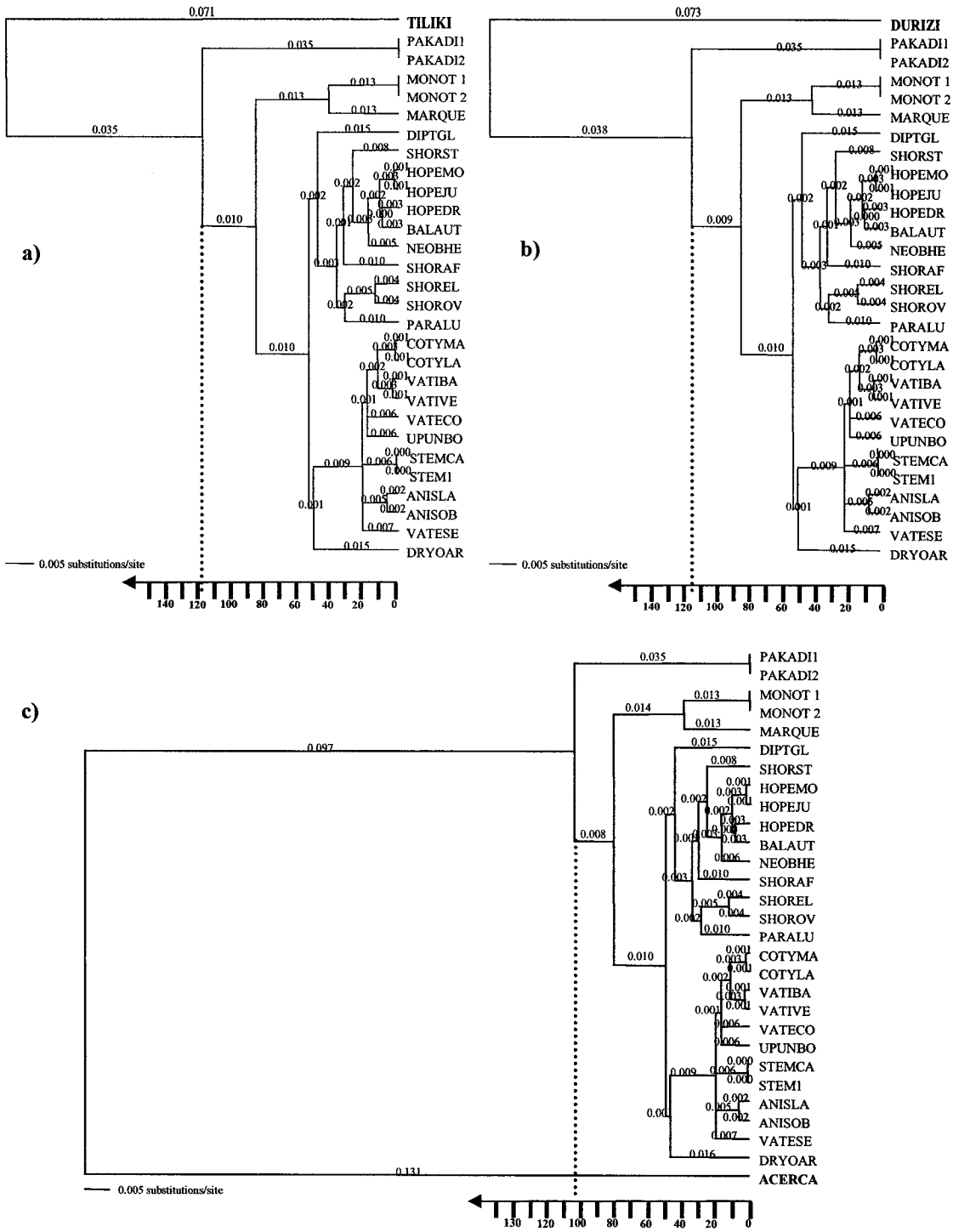


Fig. 10 Molecular clocklike 29-taxon ML tress with 3 different outgroups, *a*) *Tilia kiusiana* (family Tiliaceae; order Malvales), *b*) *Durio zibathinus* (family Bombacaceae; order Malvales), *c*) *Acer campestre* (family Sapindaceae; order Sapindales) that show the effect of outgroup in molecular dating analysis; note the compressed branch lengths at the root part of the tree in ***c,*** compared to ***a*** and ***b;*** geological time scale (millions of years) is included below each tree. See Table 11 for estimated ages based on *a*, *b* and *c*; names denoted by species codes are listed in Table 5.

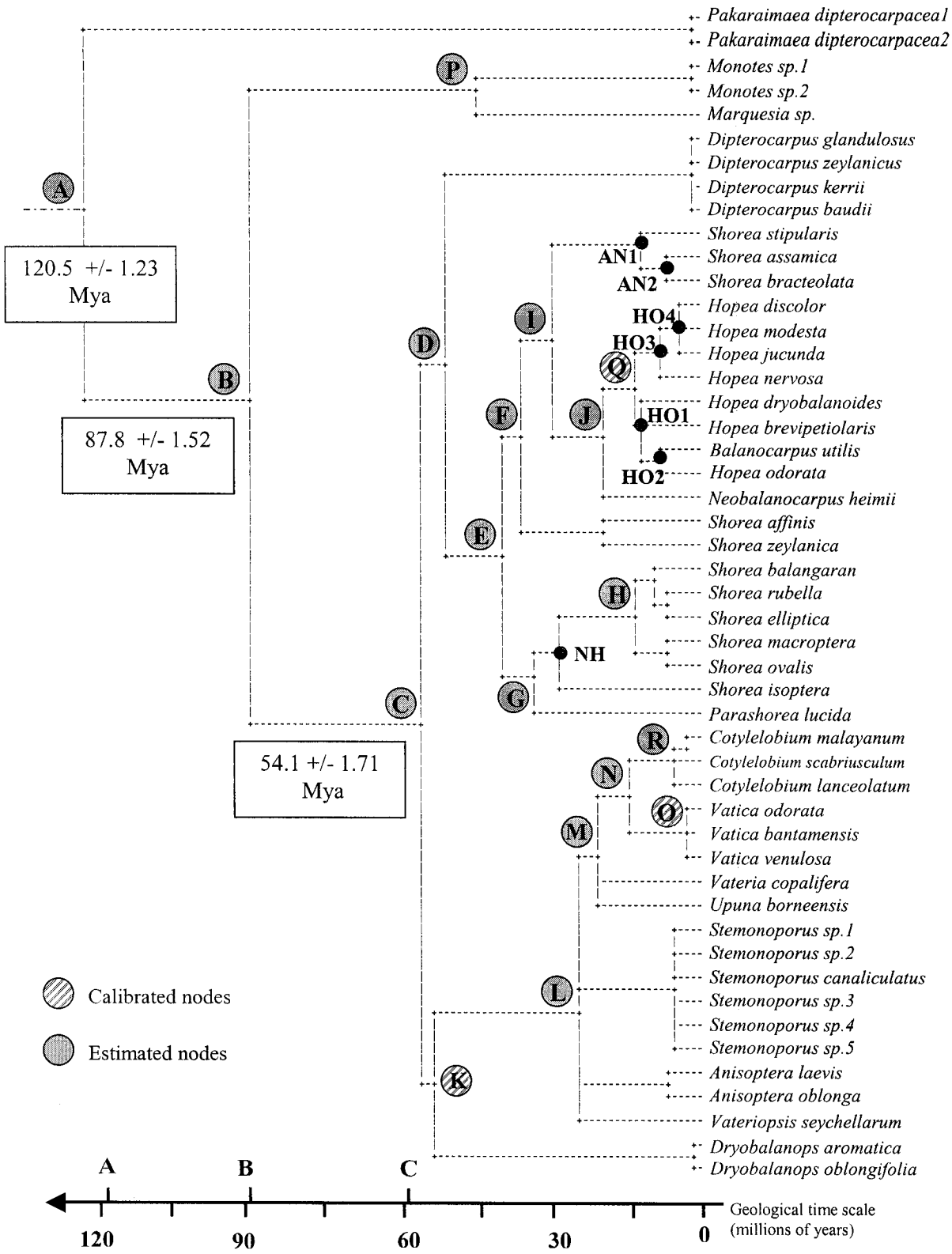
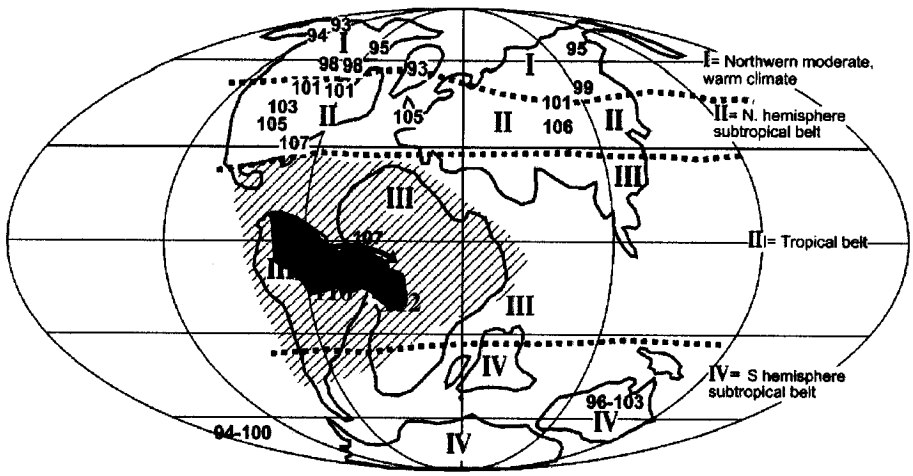
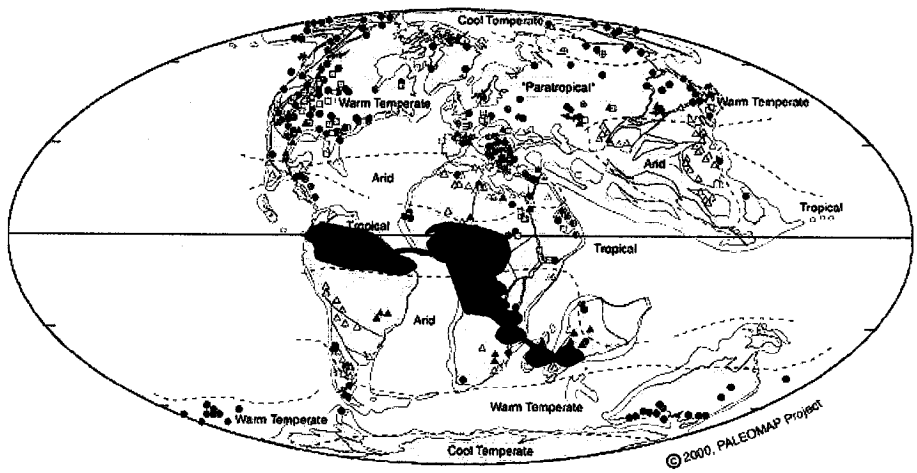


Fig. 11 The chronogram obtained by penalized likelihood analysis of the ML tree based on *matK* nucleotide sequences of 49 dipterocarp species. Numbers in the boxes are the estimated ages (millions of years ago) of the crown lineages of Pakaraimoideae (A), Monotoideae (B) and Dipterocarpoideae (C). See Table 13 for the estimated ages of the other nodes.

a)



b)



(Scotese, 2000)

Fig. 12 Paleogeographical reconstructions of, *a)* early Cretaceous (144 - 110 Mya) showing approximate locations and ages of triapertuarte pollen; and *b)* paleoclimatic reconstructions of late Cretaceous (110 – 65 Mya); black and green striped areas indicate probable historical distribution of dipterocarp forests; red arrows indicate interplate dispersal events of ancestral dipterocarps; climate sensitive rock types are indicative of **dry** [*Calcrete* (△), *Evaporite* (▲)], **warm** [*Bauxite* (●), *Laterite* (●) and *Coal* (●)] and **cool** [*Glendonite* (●), *Dropstone* (⊕)] paleoclimatic conditions.

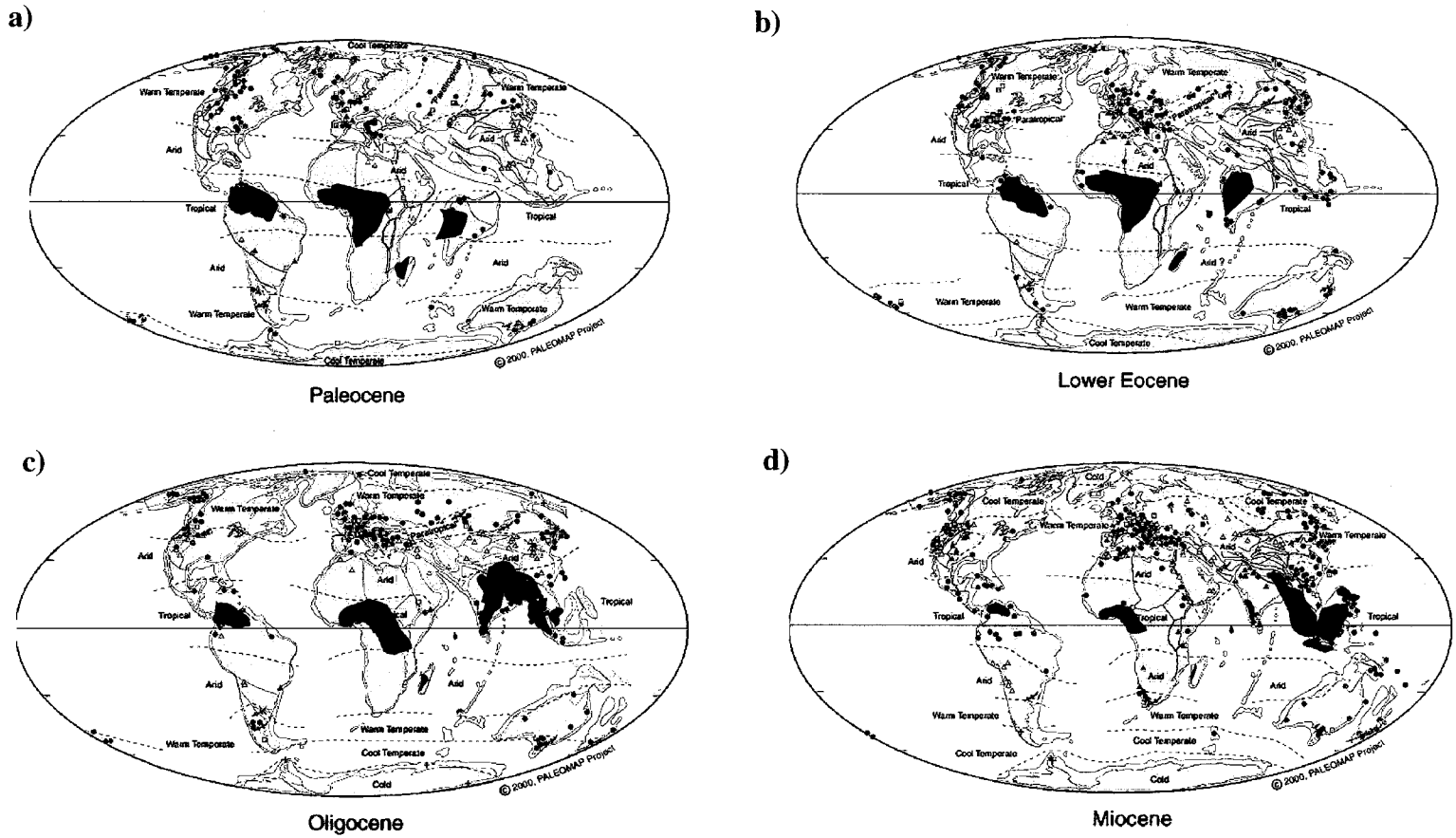


Fig.13 Paleogeographical and paleoclimatic reconstructions of a) Paleocene (65 – 54 Mya), b) early Eocene (54 – 49 Mya), c) Oligocene (36 – 25 Mya) and d) Miocene (25 – 5 Mya); red arrows indicate dispersal of dipterocarps from Deccan plate to SE Asia; black and green striped areas indicate probable historical distribution of dipterocarp forests; climate sensitive rock types are indicative of **dry** [*Calcrete* (△), *Evaporite* (▲)], **warm** [*Bauxite* (●), *Laterite* (●) and *Coal* (●)] and **cool** [*Glendonite* (●), *Dropstone* (⊕)] paleoclimatic conditions.

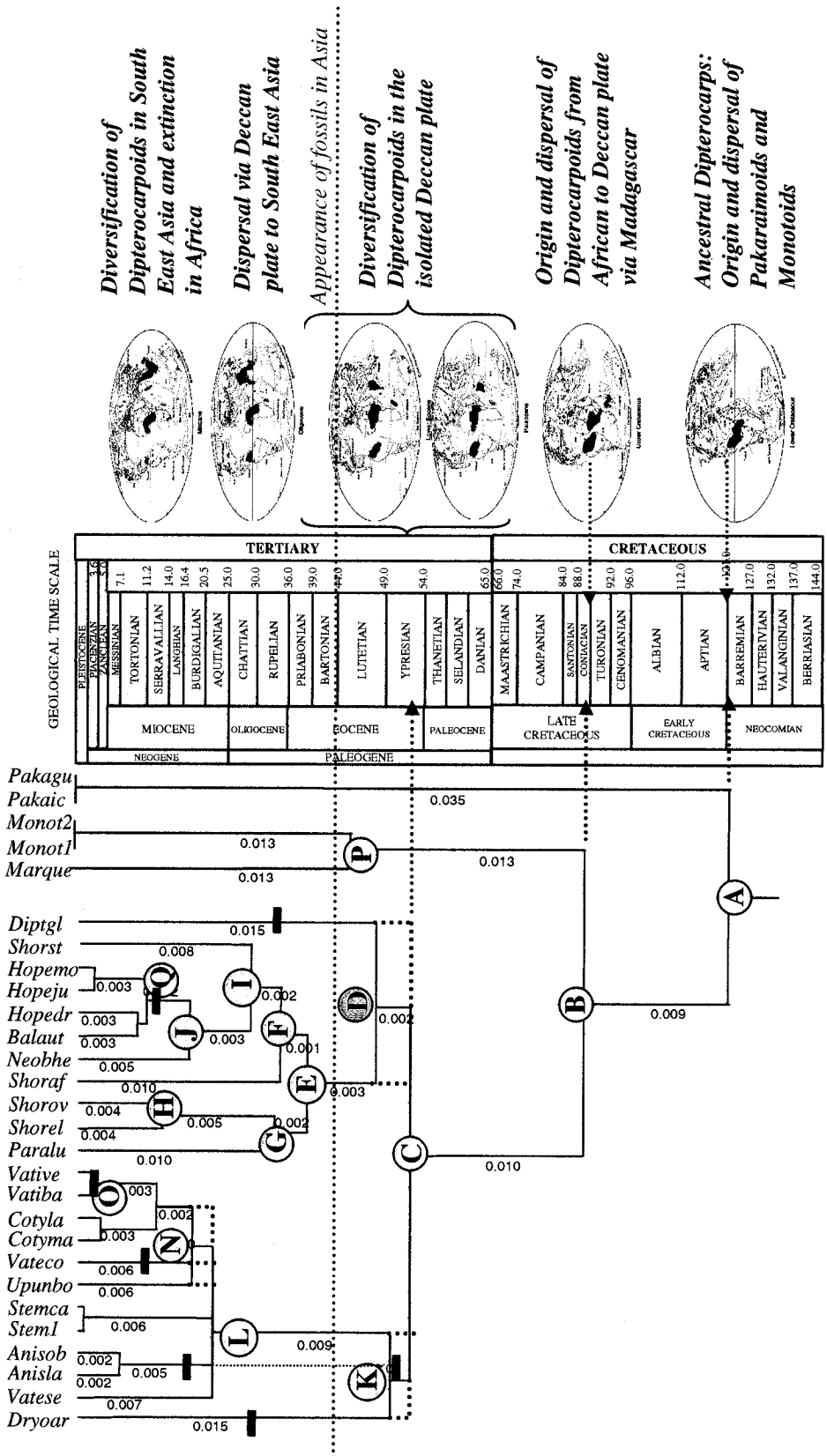


Fig.14 The results of phylogenetic and molecular dating analyses together with fossil data, paleogeographical and paleoclimatic reconstructions that illustrate the gradual pattern of speciation from the most recent common ancestors of extant dipterocarps. Black areas with green dots on the maps indicate probable historical distribution of dipterocarp forests; letters within gray circles indicate calibrated (blue) and estimated (black) nodes (MRCA) of the molecular clocklike 29-taxon ML tree; dotted blue lines indicates the branches collapsed in the bootstrap analysis; the earliest distinct macro/micro fossils are mapped on the branches (▬). See Fig. 18 for fossil descriptions and Table 5 for species code descriptions

Table 1. Affinities between family Dipterocarpaceae and closely related angiosperm families.

Botanical Characters		Dipterocarpaceae			Clusiaceae	Theaceae	Tiliaceae	Sarcolaenaceae
		Diperocarpoideae	Monotoideae	Pakaraimoideae				
Inflorescence	compound racemi	+	+		+/--			
	paniculate racemi	(+)	+	+				+
	cyme	(+)						
Bisexual flowers		+	+	+	+	+	+	+
5-merous perianth		+	+	+		+/--	+/--	+/--
Bud-flower sepals	imbricate	+	--	+		+	--	+
	valvate	+	+	--		--	+	--
Open-flower sepals	imbricate	+	--	+				
	valvate	+	+	--				
Contorted corolla		+/--			+/--	+	+/--	+
Persistent sepals and calyx								
Fruit sepals	imbricate	+	+	+/--	+/--	+		+
	valvate	+	+	+		+	+	
Centrifugal stamens		+	+	+	+	+	+	
Hypogynous stamens	numerous	+	+	+	+	+		
	many	+	+	+	+			
2-celled anthers dehising longitudinally		+/--	+	+		+		
Subversatile anthers		+	+	+		+		
Connectival appendages		+/--	+/--	+/--	--	--	+	
Pollen	tricolporate	--	+	+			+	+
	tricolpate	+						
Pollen exine	2-3 layers	+						
	4 layers	--	+	+			+	
Ovary	(2) - 3 locular	+				+		
	(2) - 3 - (5) locular		+					
	4 - 5 locular			+				
	superior	+	+	+				
	semi inferior	(+)	(+)					
	generally 2 ovules/cell	+	(+)					+

Axile placentation		+	+	+		+		
Aliform fruit sepals		+	+	+				
Short-sepaled fruit calyx		+				--		
Leaves	opposite	--	--	--	+			
	alternate	+	+	+		+	+	+
Leaf venation	prominently pinnate	+	+	+	+			
	vertically transcurrent	+	+	+	+		+	
	entirely transcurrent							
	indistinct leaf venation	+		+		+/--		
Stipule		+/--	+/--	+/--	--	--	+	+
Hair	stellate	+/--			--	--		
	tufted	+/--				--		+
	glandules	+/--			--	--		
Complex indumentum		+	+	+			+	
Indumentum + Malvales type petiole anatomy		+	+	+				
Wood rays	uniseriate	--	+	--				
	biseriate		--	+				
	multiseriate	+	--	--			--	
	mixed uni/multiseriate	+					+	
Resinous wood		+	+	+			--	
Intercellular resin canals		+	--	--	+		--	
Mucilage canals in cortex and epidermis		+		+			+	+
Elongated medullary mucilage cells		--	+	+			+	+
Pith and primary cortex with indumentum		+	+	+			+	
Anomocytic stomata		+	+	+			+	
Complex petiolar vascular supply		+/--	--	--			--	
Chromosome number (n)	n = 7	+	+		(+)	(+)	(+)	
	n = 11	+						

Table abbreviations:- ' + ' = present; ' -- ' = absent; ' +/-- ' = present and other possibilities; ' (+) ' = present but exceptions

Adapted from Maury and Curtet, 1998:

Table 2. Genera, sections, subsections and total number of species per genus in the family Dipterocarpaceae and their geographical distribution.

Sub Family/Tribe	Genus	Section	Sub section	Total # of species	Geographical area
Sub Family: Pakaraimoideae	*Pakaraimaea*			1	Guyana Highland, South America
Sub Family: Monotoideae	*Marquesia*			3	Africa
	Monotes			30	Africa and Madagascar
	Pseudomonotes			1	South America (Colombia)
Sub Family: Dipterocarpoideae					
Tribe Shoreae	*Balanocarpus*			?	Malesia, Peninsular Thailand
	Dryobalanops			7	Malesia, Borneo, Central Sumatra and intervening islands
	Hopea	*Dryobalanoides*	*Dryobalanoides*	102	Sri Lanka, Andamans, South and East India, Burma, Thailand, Indochina, continental south China, Hainan; 84 spp. throughout Malesia except lesser Sunda islands
			Sphaerocarpae		
		Hopea	*Hopea*		
			Pierrea		
	Neoblanocarpus			1	Malesia
	Parashorea			14	Burma, Thailand, Indochina, South China; 10 spp. in Malesia, Sumatra, Borneo, Philippines and intervening islands
	Shorea	*Anthoshorea*		194	Sri Lanka, India, Burma, Thailand, Indochina; 163 spp. in Malesia, Sumatra, Borneo and intervening islands, Java, Philippines and Moluccas
		Doona			
		Neohopea			
		Ovales			
		Pachycarpae			
		Pentacme			
		Rubellae			
		Brachypterae	*Brachypterae,*		
			Smithiana		
		Mutica	*Mutica*		
			Auriculatae		
		Shorea	*Shorea*		
			Barbata		
		Richetioides	*Richetioides*		
			Polvandrae		

Tribe Dipterocarpeae				
	Anisoptera	*Anisoptera*	11	Chittagong, Burma, Indochina; 10 spp. in Malesia, Peninsular Thailand, Sumatra, Borneo, Philippines; Celebes and Moluccas (1), Sulawesi (1), New Guinea (1)
		Glabrae		
	Cotylelobium		6	Sri Lanka, Peninsular Thailand, Malesia, Borneo, Sumatra
	Dipterocarpus		69	Sri Lanka, India, Burma, Peninsular Thailand, Indochina; 54 spp. in Malesia, Borneo, Philippines, Bali, Sumatra, Java
	Stemonoporus		15	Sri Lanka
	Upuna		1	Malesia and Borneo
	Vateria		2	Sri Lanka and India
	Vateriopsis		1	Seychelles
	Vatica	*Sunaptea*	65	10 spp. throughout Sri Lanka, India, Burma, Peninsular Thailand, South Chaina and Indochina; 55 spp. throughout Malesia except lesser Sunda Islands, Celebes (1), New Guinea (1)
		Vatica		

Adapted from Ashton (1982), Maury and Curtet (1998) and Dayanandan *et al.*, (1999); *Table abbreviations* :- ' ? ' = uncertain number

Table 3. Distribution of endemic species in the family Dipterocarpaceae.

Geographical areas	Total number of species	Number of Endemics	% of endemicity
Seychelles	1	1	100
Sri Lanka	44	43	98
South India	13	11	85
North India	13	4	40
Andamans	8	1	12
Burma	32	0	0
China	5	3	60
Hainan	1	0	0
Vietnam	35	3	9
Laos	19	1	5
Cambodia	27	0	0
Thailand	63	0	0
Peninsular Malesia	156	26	17
North Peninsular Malesia	47	23	49
Sumatra	95	10	10
Java	10	2	20
Lombok	3	0	0
Borneo	267	158	55
Philippines	45	21	47
Celebes	7	2	29
Moluccas	6	1	16
New Guinea	15	11	73

Adapted from Maury and Curtet (1998).

Table 4. Distribution of dipterocarp fossils in spatial and temporal scales.

Fossil Genera	Country	Number of fossils	TERTIARY											QUATERNARY
			Epoch and age (millions of years)											
			Paleocene		Eocene			Oligocene			Miocene		Pliocene	Pleistocene (< 1.6)
			66 - 60	60 - 54	54 - 49	49 - 39	39 - 36	36 - 30	30 - 25	25 - 16	16 - 10	10 - 5	5 -1.6	
Dipterocarpophyllum	Egypt	1									M			
Dipterocarpoxylon africanum	Kenya	1										U		
Geochemical fossils	Myanmar					M								
Dipterocarpophyllum	Nepal, N. India	1							M		M			
Dipterocarpus type pollen	Nepal, N. India	1									M			
Dipterocarpoxylon ??	Ethiopia	1								E ??				
Dipterocarpoxylon ??	Somalia	2											U ??	E ??
Dipterocarpus type pollen	Vietnam	1							M					
Dipterocarpoxylon	North India	11										U	U	
"	Burma	2							M					
"	Vietnam	2							M					
"	Sumatra	3							M					M
"	Java	7							M					
Anisopteroxylon	London Clay				E									
Anisopteroxylon	N.W. India										MU		E	
Vaterioxylon	North India	2									MU		E	
Vaticoxylon	Sumatra	1												M
"	Java	1											M	
Shoreoxylon	India-Assam	13									MU		E	
"	Northwest India	1									MU		E	
"	North India	1							M					
"	South India	3									MU		E	
"	Cambodia	1												M
"	Thailand	1											E	E
"	Sumatra	7									M		M	M
Hopenium	North India	2									M			
Dryobalanoxylon	Cambodia	1												M
"	South Vietnam	1								E			E	
"	Borneo	1									M			
"	Borneo	1							M					
"	Sumatra	6											M	M
"	Java	5											MU	E

Table abbreviations : E = Early Period; M = Mid period; MU = Mid to Upper period; U = Upper period; ?? = Uncertain fossils

Adapted from Bancroft (1935), Aarssen *et. al.* (1990, 1992), Maury and Curtet (1998) and Poole (1993).

Table 5. List of species selected for the present study with voucher and GenBank accession numbers

Family/ sub family/tribe	Genus	Section	Species	Species Code	GenBank accession number * or Voucher number
Bomabacaceae	*Durio*		*Durio zibathinus*	DURIZI	B3
	Pseudobombax		*Pseudobombax sp.*	PSEUDO	B4
Tiliaceae	*Tilia*		*Tilia kiusiana*	TILIKI	AB006386*
Sapindaceae	*Acer*		*Acer campestre*	ACERCA	AJ438793*
			Acer opalus	ACEROP	AJ438787*
Dipterocarpaceae					
Sub Family:					
Pakaraimoideae	*Pakarimaea*		*Pakarimaea dipterocarpacea 1*	PAKADI1	D55
			Pakarimaea dipterocarpacea 2	PAKADI2	D56
Sub Family:					
Monotoideae	*Marquesia*		*Marquesia sp.*	MARQUE	D54
	Monotes		*Monotes sp1*	MONOT1	D51
			Monotes sp2	MONOT2	D52
Sub Family:					
Dipterocarpoideae					
Tribe Shoreae	*Balanocarpus*		*Balanocarpus utilis*	BALAUT	D31
	Hopea	*Dryobalanoides*	*Hopea dryobalanoides*	HOPEDR	D41
		Dryobalanoides	*Hopea nervosa*	HOPENE	AB006384*
		Hopea	*Hopea brevipetiolaris*	HOPEBR	D05
		Hopea	*Hopea discolor*	HOPEDI	D29

		Hopea	*Hopea jucunda*	HOPEJU	D30
		Hopea	*Hopea modesta*	HOPEMO	D35
		Hopea	*Hopea odorata*	HOPEOD	AB006385*
	Neoblanocarpus		*Neoblanocarpus heimii*	NEOBHE	D22
	Parashorea		*Parashorea lucida*	PARALU	AB006382*
	Shorea	*Anthoshorea*	*Shorea bracteolata*	SHORBR	AB006381*
		Anthoshorea	*Shorea stipularis*	SHORST	D06
		Anthoshorea	*Shorea assamica*	SHORAS	D42
		Brachypterae	*Shorea balangaran*	SHORBA	D16
		Doona	*Shorea zeylanica*	SHORZE	L5
		Doona	*Shorea affinis*	SHORAF	L4
		Mutica	*Shorea macroptera*	SHORMA	AB006379*
		Neohopea	*Shorea isoptera*	SHORIS	D57
		Ovales	*Shorea ovalis*	SHOROV	AB006380*
		Rubella	*Shorea elliptica*	SHOREL	D58
		Rubella	*Shorea rubella*	SHORRU	D25
	Dryobalanops		*Dryobalanops aromatica*	DRYOAR	AB006377*
			Dryobalanops oblongifolia	DRYOOB	AB006378*
Tribe : Dipterocarpeae	*Anisoptera*	*Anisoptera*	*Anisoptera oblonga*	ANISOB	AB006372*
		Glabrae	*Anisoptera laevis*	ANISLA	AB006370*
	Cotylelobium		*Cotylelobium lanceolatum*	COTYLA	D20
			Cotylelobium scbriusculum	COTYSC	D34
			Cotylelobium malayanum	COTYMA	AB006371*
	Dipterocarpus		*Dipterocarpus glandulosus*	DIPTGL	D32

		Dipterocarpus zeylanicus	DIPTZE	D02
		Dipterocarpus kerrii	DIPTKE	AB006375*
		Dipterocarpus baudii	DIPTBA	AB006376*
Stemonoporus		*Stemonoporus canaliculatus*	STEMCA	D60
		Stemonoporus sp1	STEM1	D67
		Stemonoporus sp2	STEM2	D68
		Stemonoporus sp3	STEM3	D69
		Stemonoporus sp4	STEM4	D63
		Stemonoporus sp5	STEM5	D64
Upuna		*Upuna borneensis*	UPUNBO	D18
Vateria		*Vateria copalifera*	VATECO	D36
Vateriopsis		*Vateriopsis seychellarum*	VATESE	D66
Vatica	*Sunaptea*	*Vatica bantamensis*	VATIBA	D12
	Vatica	*Vatica venulosa*	VATIVE	D11
	Vatica	*Vatica odorata*	VATIOD	AB006373*

Table 6. Nucleotide sequences of primers used to amplify 1200 bp length region of the chloroplast *matK* gene

Primer's name	DNA Sequence
matK 1F	ACT GTA TCG CAC TAT GTA TCA
matK 265F	TGT ATT CTC AAA TAA TAT CGG C
matK 602F	CCA TTT TCC TTT TCC GT
matK 392R	GAT GGA TGG GAT GAG GTA TTA GT
matK 892R	ATC CTT CCT TGA TTG AGA CCA
matK 1125R	TCC AGA TCG GCT TAC TAA TG

See Fig. 3 for the locations of primers

Table 7. The average amino acid composition and average nucleotide compositions (in first second and third codon positions) at all sites, conserved, variable and parsimony informative sites of the *matK* gene in 29 and 49 taxon datasets.

Dataset	Average amino acid composition	Average nucleotide composition			
		Position 1	Position 2	Position 3	All sites
29-taxon					
All sites	377	381	380	379	1139
Conserved sites	255	316	321	288	925
Parsimony informative sites	69	36	28	58	121
Parsimony uninformative variable sites	53	28	30	34	92
49-taxon					
All sites	372	375	374	373	1122
Conserved sites	225	297	305	271	872
Parsimony informative sites	92	48	40	78	165
Parsimony uninformative variable sites	56	30	29	25	85

Table 8. The parameters of nucleotide substitution models used in the maximum likelihood, distance and bayesian analyses of 29 and 49 taxon datasets

Parameters		Dataset	
		29-taxon	**49-taxon**
Nucleotide substituion model	LRT	TVM + Γ	TVM + Γ
	AIC	TVM + I	TVM + Γ
Gamma shape parameter (Γ)		0.878	1.018
Proportion of invariable sites (I)		0.4142	0.00
Log likelihood value		-3884.90	-5268.69
Base frequencies	A	0.31	0.31
	C	0.17	0.18
	G	0.15	0.15
	T	0.36	0.36
Transition (ti) / transversion (tv) ratio	all sites	1.30	1.30
	position 1	1.20	1.10
	position 2	1.10	1.00
	position 3	1.60	1.70
Nucleotide substitution rate matrix	*R (a)* A-C	0.982	1.177
	R (b) A-G	1.788	1.768
	R (c) A-T	0.211	0.239
	R (d) C-G	0.827	0.810
	R (e) C-T	1.788	1.768
	R (f) G-T	1.000	1.000

Table abbreviations : LRT = Likelihood ratio test; AIC = Akaike information criterion; TVM = Transversion model; R = nucleotide substitution rate (see appendix 3 for rate classification); A = Adenine; C = Cytosine; G = Guanine; T = Thiamine;

Table 9. Results of the Shimodaira Hasegawa test showing the differences of log likelihood values of the pairwise comparisons of alternative topologies.

	MP	ML	NJ	BI
MP	-			
ML	1.333 (0.36)	-		
NJ	5.609 (0.22)	4.275 (0.24)	-	
BI	0.011 (0.48)	1.344 (0.29)	5.620 (0.15)	-

Probability values of one tailed test are within brackets (alpha level = 0.05).

Table abbreviations :- MP = Maximum Parsimony; ML = Maximum Likelihood; NJ = Naighbor Joining; BI = Baysian inference

Table 10. Results of Tajima's relative rate test based on 29-taxon maximum likelihood tree

Taxa compared		Outgroup	Chi-square (X^2) value	P
Ingroup 1	Ingroup 2			
(MONOT1,	PAKAIC)	DURIZI	0.02	0.90
(MONOT2,	PAKAIC)	DURIZI	0.44	0.51
(MARQUE,	PAKAIC)	DURIZI	0.69	0.41
(MONOT1,	PAKAGU)	DURIZI	0.02	0.90
(MONOT2,	PAKAGU)	DURIZI	0.44	0.51
(MARQUE,	PAKAGU)	DURIZI	0.69	0.41
(MONOT1,	MARQUE)	PAKAIC	2.29	0.13
(SHORST,	DIPTGL)	MARQUE	0.76	0.38
(SHORST,	SHORAF)	DIPTGL	1.32	0.25
(SHOROV,	PARALU)	DIPTGL	0.29	0.59
(SHOREL,	PARALU)	DIPTGL	0.22	0.64
(SHORST,	SHORAF)	PARALU	2.25	0.13
(NEOBHE,	SHORST)	SHORAF	0.00	1.00
(HOPEMO,	NEOBHE)	SHORST	3.60	0.06
(HOPEDR,	HOPEMO)	NEOBHE	2.00	0.16
(BALAUT,	HOPEMO)	NEOBHE	0.11	0.74
(DRYOAR,	DIPTGL)	MARQUE	0.04	0.84
(VATESE,	DRYOAR)	DIPTGL	0.31	0.58
(ANISLA,	VATESE)	DRYOAR	0.69	0.41
(STEMCA,	VATESE)	DRYOAR	0.33	0.56
(UPUNBO,	ANISLA)	VATESE	0.09	0.76
(UPUNBO,	STEMCA)	VATESE	0.33	0.56
(VATECO,	UPUNBO)	STEMCA	1.00	0.32
(VATECO,	UPUNBO)	ANISLA	0.11	0.74
(COTYLA,	VATECO)	UPUNBO	1.60	0.21
(VATIBA,	COTYLA)	UPUNBO	2.67	0.10
(VATIVE,	COTYLA)	VATECO	2.78	0.10

Table abbreviaions : P = Probability of Chi Square (X^2) statistic at alpha level of 0.05

Table 11. Ages of nodes with confidence limits in the molecular clocklike maximum likelihood trees of 29 taxa (with alternative outgroups) inferred from three independent fossil calibration points using Langley Fitch method.

Node	Ancestral lineages	Estimated ages (millions of years) with confidence limits								
		Calibrated node : K = 51.5 Mya			Calibrated node : Q = 13.0 Mya			Calibrated node : O = 3.3 Mya		
		D. zibathinus	*T. kiusiana*	*A. campastre*	*D. zibathinus*	*T. kiusiana*	*A. campastre*	*D. zibathinus*	*T. kiusiana*	*A. campastre*
A	Pakaraimoideae crown lineage	119.5 +/- 1.57	121.8 +/- 1.59	114.1 +/- 1.78	117.3 +/- 1.72	118.4 +/- 1.75	115.6 +/- 1.71	127.2 +/- 1.66	131.4 +/- 1.70	122.2 +/- 1.67
B	Monotoideae crown lineage	88.3 +/- 1.24	88.0 +/- 1.16	89.6 +/- 1.47	88.0 +/- 1.32	86.6 +/- 1.39	96.1 +/- 1.38	94.3 +/- 1.50	95.0 +/- 1.61	99.8 +/- 1.62
C	Diptrocarpoideae crown lineage	54.4 +/- 0.51	54.4 +/- 0.48	54.1 +/- 0.59	54.8 +/- 0.94	56.0 +/- 0.83	57.9 +/- 0.91	58.1 +/- 0.97	59.0 +/- 1.10	60.1 +/- 1.11
D	*Diptrocarpus* stem lineage	49.1 +/- 0.95	49.2 +/- 0.62	48.5 +/- 0.72	50.1 +/- 0.95	51.3 +/- 0.62	51.9 +/- 0.72	52.3 +/- 0.99	53.0 +/- 1.13	54.0 +/- 1.12
E	Tribe Shoreae (crown lineage)	37.3 +/- 0.63	37.7 +/- 0.60	37.3 +/- 0.66	39.3 +/- 0.80	39.3 +/- 0.79	39.9 +/- 0.70	40.0 +/- 0.79	40.4+/- 0.90	41.5 +/- 0.91
F	*Doona* stem lineage	34.5 +/- 0.61	34.6 +/- 0.59	34.3 +/- 0.72	35.3 +/- 0.77	36.2 +/- 0.75	36.8 +/- 0.74	36.9 +/- 0.77	37.2 +/- 0.87	38.2 +/- 0.87
G	*Rubroshorea* stem lineage	32.4 +/- 0.79	32.5 +/- 0.77	32.0 +/- 0.95	32.7 +/- 0.85	33.2 +/- 0.83	34.3 +/- 0.81	34.2 +/- 0.92	34.4 +/- 1.06	35.7 +/- 1.06
H	*Rubroshorea* crown lineage	13.7 +/- 0.74	13.7 +/- 0.73	13.4 +/- 0.70	14.7 +/- 0.72	14.9 +/- 0.72	14.4 +/- 0.55	15.2 +/- 0.76	15.5 +/- 0.79	14.9 +/- 0.78
I	*Anthoshorea* stem lineage	28.5 +/- 0.66	28.5 +/- 0.62	28.7 +/- 0.66	29.3 +/- 0.71	29.8 +/- 0.70	30.8 +/- 0.83	30.6 +/- 0.72	30.8 +/- 0.81	32.0 +/- 0.82
J	*Hopea* stem lineage	19.5 +/- 0.58	19.5 +/- 0.57	18.1 +/- 0.58	17.5 +/- 0.56	17.8 +/- 0.56	17.5 +/- 0.74	18.3 +/- 0.58	18.4 +/- 0.64	20.2 +/- 0.63
K	Tribe Dipterocarpeae (stem lineage)	*Calibrated node*			52.7 +/- 1.09	54.2 +/- 1.10	55.0 +/- 1.09	55.0 +/- 1.15	56.0 +/- 1.40	57.1 +/- 1.40
L	Tribe Dipterocarpeae (crown lineage)	22.5 +/- 0.51	22.4 +/- 0.26	22.0 +/- 0.26	22.9 +/- 0.37	22.0 +/- 0.37	23.8 +/- 0.35	24.0 +/- 0.26	23.8 +/- 0.60	23.8 +/- 0.61
M	*Vateria/Upuna* crown lineage	19.5 +/- 0.20	19.4 +/- 0.48	19.0 +/- 0.47	19.4 +/- 1.35	18.0 +/- 1.22	19.8 +/- 1.32	20.8 +/- 0.48	20.7 +/- 0.20	20.7 +/- 0.21
N	*Cotylelobium/Vatica* stem lineage	13.2 +/- 0.53	13.2 +/- 0.52	12.8 +/- 0.51	13.5 +/- 0.49	13.6 +/- 0.49	13.6 +/- 0.45	14.1 +/- 0.53	14.1 +/- 0.55	14.1 +/- 0.54
O	*Vatica* crown lineage	3.0 +/- 0.24	3.1 +/- 0.24	2.9 +/- 0.24	3.1 +/- 0.24	3.2 +/- 0.24	3.2 +/- 0.23	*Calibrated node*		
P	*Marquesia/Monotes* crown lineage	42.6 +/- 1.21	42.4 +/- 1.12	43.5 +/- 1.28	43.6 +/- 1.07	44.3 +/- 1.06	46.6 +/- 1.04	45.6 +/- 1.15	45.9 +/- 1.36	48.4 +/- 1.35
Q	*Hopea* crown lineage	13.3 +/- 0.41	13.3 +/- 0.40	12.6 +/- 0.46	*Calibrated node*			13.9 +/- 0.68	14.1 +/- 0.87	14.9 +/- 0.45

See Fig. 9 for the locations of nodes

Table 12. Ages of nodes with confidence limits in the molecular clocklike maximum likelihood trees of 29 taxa (with alternative outgroups) inferred from three independent fossil calibration points using penalized likelihood method *.

Node	Ancestral lineages	Estimated ages (millions of years) with confidence limits								
		Calibrated node : K = 51.5 Mya			Calibrated node : Q = 13.0 Mya			Calibrated node : O = 3.3 Mya		
		D. zibathinus	*T. kiusiana*	*A. campastre*	*D. zibathinus*	*T. kiusiana*	*A. campastre*	*D. zibathinus*	*T. kiusiana*	*A. campastre*
A	Pakaraimoideae crown lineage	119.1 +/- 1.24	121.8 +/- 1.42	114.1 +/- 1.65	117.5 +/- 1.58	118.3 +/- 1.42	115.5 +/- 1.50	127.2 +/- 1.52	130.2 +/- 1.15	126.1 +/- 1.33
B	Monotoideae crown lineage	88.0 +/- 1.14	88.0 +/- 1.19	89.6 +/- 1.40	85.5 +/- 1.22	85.6 +/- 1.39	92.3 +/- 1.38	94.1 +/- 1.34	94.1 +/- 1.01	99.1 +/- 1.03
C	Diptrocarpoideae crown lineage	54.4 +/- 0.53	54.4 +/- 0.17	54.1 +/- 0.46	52.8 +/- 0.51	52.8 +/- 0.83	55.7 +/- 0.91	58.1 +/- 0.57	58.1 +/- 1.38	59.8 +/- 1.15
D	*Diptrocarpus* stem lineage	49.1 +/- 0.89	49.0 +/- 0.82	48.5 +/- 0.69	47.6 +/- 0.45	47.6 +/- 0.62	50.0 +/- 0.72	52.3 +/- 0.79	52.3 +/- 1.93	53.7 +/- 1.52
E	Tribe Shoreae (Crown Lineage)	37.3 +/- 0.62	37.7 +/- 0.72	37.3 +/- 0.65	36.6 +/- 0.71	36.6 +/- 0.79	38.4 +/- 0.70	40.3 +/- 0.72	40.3+/- 0.83	41.2 +/- 0.82
F	*Doona* stem lineage	34.5 +/- 0.69	34.6 +/- 0.81	34.3 +/- 0.71	33.6 +/- 0.75	33.6 +/- 0.75	35.4 +/- 0.74	37.0 +/- 0.88	37.0 +/- 0.79	38.0 +/- 0.67
G	*Rubroshorea* stem lineage	32.4 +/- 0.74	32.5 +/- 0.66	32.0 +/- 0.90	31.6 +/- 0.89	31.6 +/- 0.83	33.6 +/- 0.81	34.8 +/- 0.82	34.8 +/- 1.45	35.4 +/- 1.05
H	*Rubroshorea* crown lineage	13.7 +/- 0.71	13.7 +/- 0.72	13.4 +/- 0.69	13.3 +/- 0.62	13.2 +/- 0.72	13.8 +/- 0.55	14.7 +/- 0.75	14.7 +/- 0.68	14.8 +/- 0.73
I	*Anthoshorea* stem lineage	28.5 +/- 0.60	28.5 +/- 0.57	28.7 +/- 0.64	27.7 +/- 0.51	27.7 +/- 0.70	29.6 +/- 0.83	30.6 +/- 0.47	30.5 +/- 0.80	32.0 +/- 0.32
J	*Hopea* stem lineage	19.5 +/- 0.54	19.5 +/- 0.44	18.1 +/- 0.53	18.9 +/- 0.59	18.8 +/- 0.56	19.5 +/- 0.74	18.3 +/- 0.52	20.8 +/- 0.78	20.1 +/- 0.43
K	Tribe Dipterocarpeae (stem lineage)	*Calibrated node*			50.0 +/- 1.06	50.0 +/- 1.10	53.0 +/- 1.09	55.0 +/- 1.12	55.0 +/- 1.10	56.9 +/- 1.26
L	Tribe Dipterocarpeae (crown lineage)	22.5 +/- 0.49	22.4 +/- 0.23	22.0 +/- 0.22	21.8 +/- 0.32	21.8 +/- 0.37	22.6 +/- 0.35	24.0 +/- 0.16	24.0 +/- 0.54	24.3 +/- 0.90
M	*Vateria/Upuna* crown lineage	19.5 +/- 0.19	19.4 +/- 0.42	19.0 +/- 0.46	18.9 +/- 1.25	18.9 +/- 1.22	19.5 +/- 1.32	20.9 +/- 0.44	20.8 +/- 0.56	21.0 +/- 0.81
N	*Cotylelobium/Vatica* stem lineage	13.2 +/- 0.42	13.2 +/- 0.50	12.8 +/- 0.50	12.8 +/- 0.24	12.8 +/- 0.49	13.2 +/- 0.45	14.2 +/- 0.51	14.1 +/- 0.51	14.2 +/- 0.84
O	*Vatica* crown lineage	3.0 +/- 0.16	3.1 +/- 0.29	2.9 +/- 0.21	3.0 +/- 0.21	3.0 +/- 0.24	3.1 +/- 0.23	*Calibrated node*		
P	*Marquesia/Monotes* crown lineage	42.6 +/- 1.81	42.4 +/- 1.23	43.5 +/- 1.15	41.3 +/- 1.03	41.3 +/- 1.06	44.8 +/- 1.04	45.5 +/- 1.12	45.5 +/- 1.74	48.4 +/- 1.25
Q	*Hopea* crown lineage	13.3 +/- 0.21	13.3 +/- 0.20	12.6 +/- 0.36	*Calibrated node*			13.6 +/- 0.63	14.3 +/- 0.82	13.9 +/- 0.35

* Smoothing value = 3.16; See Fig. 9 for the locations of nodes

Table 13. Ages of nodes with confidence limits in the maximum likelihood tree of 49 taxa inferred from three independent fossil calibration points using penalized likelihood method *.

Node	Ancestral Lineage	Estimated ages (Mya) with confidece limits			Local rates (SSMY)
		Node K = 51.5	Node Q = 13.0	Node O = 3.3	
A	Pakaraimoideae crown lineage	120.5 +/- 1.23	135.8 +/- 1.61	160.0 +/- 1.85	
B	Monotoideae crown lineage	87.8 +/- 1.52	99.3 +/- 1.34	116.8 +/- 1.46	0.00027
C	Diptrocarpoideae crown lineage	54.1 +/- 1.71	61.9 +/- 1.65	71.9 +/- 1.12	0.00028
D	*Diptrocarpus* stem lineage	49.0 +/- 1.30	56.0 +/- 0.89	65.1 +/- 0.43	0.00026
E	Tribe Shoreae (crown lineage)	37.7 +/- 0.38	43.2 +/- 0.71	50.0 +/- 0.23	0.00030
F	*Doona* stem lineage	34.5 +/- 0.78	39.3 +/- 0.85	45.8 +/- 0.56	0.00037
I	*Anthoshorea* stem lineage	28.1 +/- 0.76	32.4 +/- 0.93	37.2 +/- 0.84	0.00042
AN1	*Anthoshorea* crown lineage 1	10.4 +/- 0.81	11.0 +/- 0.76	13.1 +/- 0.85	0.00032
AN2	*Anthoshorea* crown lineage 2	6.1 +/- 1.38	6.5 +/- 0.91	8.0 +/- 0.91	0.00038
J	*Hopea* stem lineage	17.8 +/- 1.22	20.2 +/- 1.10	23.5 +/- 0.78	0.00046
Q	*Hopea* crown lineage	11.3 +/- 0.53	*Calibrated node*	14.9 +/- 0.67	0.00063
HO1	*Hopea* crown lineage 1	10.2 +/- 0.34	13.0 +/- 0.54	13.5 +/- 0.49	0.00074
HO2	*Hopea* crown lineage 2	7.2 +/- 0.21	8.5 +/- 0.38	9.5 +/- 0.56	0.00028
HO3	*Hopea* crown lineage 3	7.0 +/- 0.54	8.9 +/- 0.67	9.3 +/- 1.21	0.00019
HO4	*Hopea* crown lineage 4	2.5 +/- 1.84	2.0 +/- 1.61	3.4 +/- 1.35	0.00037
G	*Rubroshorea* stem lineage	31.5 +/- 1.71	36.3 +/- 1.63	41.9 +/- 1.45	0.00013
NH	*Neohopea* stem lineage	26.6 +/- 0.92	30.4 +/- 0.45	38.9 +/- 0.64	0.00016
H	*Rubroshorea* crown lineage	11.5 +/- 1.68	13.1 +/- 1.58	15.3 +/- 1.25	0.00021
K	Tribe Dipterocarpeae (stem lineage)	*Calibrated node*	58.8 +/- 0.78	68.4 +/- 0.98	0.00031
L	Tribe Dipterocarpeae (crown lineage)	22.9 +/- 1.34	26.2 +/- 1.23	30.4 +/- 0.87	0.00034
M	*Vateria/Upuna* crown lineage	18.9 +/- 0.88	21.6 +/- 1.24	25.1 +/- 0.56	0.00020
N	*Cotylelobium/Vatica* stem lineage	13.5 +/- 0.71	15.5 +/- 0.56	18.0 +/- 0.67	0.00031
O	*Vatica* crown lineage	2.3 +/- 0.73	2.5 +/- 0.71	*Calibrated node*	0.00015
P	*Marquesia/Monotes* crown lineage	43.3 +/- 0.65	49.6 +/- 0.63	57.7 +/- 0.78	0.00028
R	*Cotylelobium* crown lineage	4.5 +/- 0.57	5.2 +/- 0.56	6.0 +/- 0.43	0.00034

Table abbreviations:- SSMY = nucleotide substitutions per site per million years; MYA = Million years ago

* Smoothing value = 316.6: See Fig. 11 for the locations of nodes.

Table 14. Results of the likelihood ratio tests of the sensitivity analysis and the ages (millions of years) estimated by randomly adding taxa into species rich clades starting with a molecular clocklike tree of 23 taxa.

Phylogenetic Tree	OTU	DF	*lnL*	lnL_{CL}	lnL_{CL} - *lnL*	2 x (lnL_{CL} - *lnL*)	P	Molecular dating method	Estimated Ages (calibrated node, K = 51.5 Mya)		
									node A	node B	node C
23-taxon tree	23	21	3869.56	3881.28	11.72	23.44	0.32	LF	119.52	88.36	54.48
23 (+2)	25	23	3883.37	3895.20	11.83	23.65	0.42	LF	119.00	88.21	54.81
25 (+3)	28	26	3980.72	3999.17	18.45	36.90	0.08	LF	120.10	88.39	54.76
28 (+7)	35	33	4174.54	4198.25	23.71	47.41	0.05	LF	121.02	88.40	54.51
35 (+3)	38	36	4236.20	4262.93	26.73	53.46	0.03*	PL (10.0)	120.32	87.53	54.23
38 (+2)	40	38	3786.15	3814.91	28.77	57.53	0.02*	PL (316.6)	120.45	87.23	54.19
49-taxon tree	49	47	5263.28	5508.33	245.04	490.09	0.00*	PL (316.6)	120.50	87.80	54.10

Table abbreviations :- OTU= Operational taxonomic unit = number of taxa; DF =Degrees of freedom = [OTU - 2]; *lnL* = Log likelihood values obtained without enforcing the molecular clock; lnL_{CL} = log likelihood values obtained with enforcing the molecular clock; P = Probability of Chi Square (X^2) statistic; ' * ' = $P < 0.05$ = significant rate heterogeneity; Mya = Million years ago; LF = Langley Fitch; PL = Penalized Likelihood with smoothing value within brackets.

Table 15. Results of the likelihood ratio tests of the sensitivity analysis and the ages (millions of years) estimated by sequentially adding taxa to, a) *Hopea* and *Shorea* and b) *Stemonoporus* clades starting with a molecular clocklike tree of 23 taxa.

	Phylogenetic Tree	OTU	DF	*lnL*	lnL_{CL}	lnL_{CL} - *lnL*	2 x (lnL_{CL} - *lnL*)	P	Molecular dating method	Estimated Ages (calibrated node, K = 51.5 Mya)		
										node A	node B	node C
a)	23-taxon tree	23	21	3869.56	3881.28	11.72	23.44	0.32	LF	119.52	88.36	54.48
	23 (+ HOPEDI)	24	22	3876.88	3888.81	11.93	23.86	0.35	LF	119.78	88.51	54.52
	24 (+ HOPEJU)	25	23	3894.59	3908.62	14.03	28.06	0.21	LF	119.48	88.31	54.47
	25 (+ HOPENE)	26	24	3903.81	3919.29	15.48	30.96	0.15	LF	120.31	88.86	54.60
	26 (+ HOPEOD)	27	25	3922.60	3937.15	14.54	29.09	0.26	LF	118.84	87.90	54.37
	27 (+ HOPEBR)	28	26	3968.08	3983.53	15.45	30.89	0.23	LF	118.86	87.91	54.37
	28 (+ SHORAS)	29	27	3995.85	4011.42	15.57	31.14	0.27	LF	118.49	87.67	54.31
	29 (+ SHORBR)	30	28	4029.46	4047.43	17.97	35.94	0.14	LF	117.31	86.86	54.47
	30 (+ SHORZE)	31	29	4104.27	4123.79	19.52	39.04	0.10	LF	118.30	87.55	54.28
	31 (+ SHORIS)	32	30	4144.36	4165.28	20.92	41.84	0.07	LF	118.29	87.54	54.28
	32 (+ SHORRU)	33	31	4171.19	4192.11	20.92	41.83	0.09	LF	117.65	87.13	54.18
	33 (+ SHORMA)	34	32	4194.44	4216.76	22.32	44.63	0.07	LF	118.74	87.83	54.35
	34 (+ SHORBA)	35	33	4196.68	4222.48	25.80	51.60	0.02*	PL (10.0)	118.36	89.60	55.20
b)	23-taxon tree	23	21	3869.56	3881.28	11.72	23.44	0.32	LF	119.52	88.36	54.48
	23 (+ STEM1)	24	22	3939.34	3945.51	6.17	12.34	0.43	LF	119.40	88.21	54.42
	24 (+ STEM3)	25	23	3955.55	3962.23	6.68	13.36	0.38	LF	119.20	88.90	54.31
	25 (+ STEM6)	26	24	3964.15	3970.54	6.39	12.78	0.41	LF	119.45	88.21	54.34
	26 (+ STEM5)	27	25	3971.91	3978.71	6.80	13.60	0.39	LF	119.10	88.30	54.20
	27 (+ STEMCA)	28	26	3982.12	3991.35	9.23	18.46	0.36	LF	119.44	88.60	54.17
	28 (+ STEM4)	29	27	3988.13	3997.61	9.48	18.96	0.35	LF	119.51	88.00	54.23

Table abbreviations :- OTU= Operational taxonomic unit = number of taxa; DF =Degrees of freedom = [OTU - 2]; *lnL* = Log likelihood values obtained without enforcing the molecular clock; lnL_{CL} = log likelihood values obtained with enforcing the molecular clock; P = Probability of Chi Square (X^2) statistic; ' * ' = P < 0.05 = significant rate heterogeneity; Mya = Million years ago; LF = Langley Fitch; PL = Penalized Likelihood with smoothing value within brackets.

6.0. REFERENCES

Aarssen, B. G. K. V., Cox, H. C., Hoogendoorn, P. and Leeuw, J. W. D. (1990). A cadinane bipolymer present in fossil and extant dammar resins as a source for cadinanes and bicadinanes in crude oils from Southeast Asia. *Geochimica et Cosmochimica Acta.* 54: Pp. 3021 – 3031.

Alverson, W. S., Karol, K. G., Baum, D. A., Chase, M. W., Swensen, S. M., McCourt, R. and Sytsma, K. J. (1998). Circumscription of the Malvales and relationships to other Rosidae: Evidence from *rbcL* sequence data. *American Journal of Botany.* 85: Pp. 876 – 887.

Appanah, S. (1998). Management of natural forests. *In* A review of Dipterocarps, taxonomy, ecology and silvicultrue (*eds.* S. Appanah and J. M. Turnbull). Center for forest research institute, Malesia. Pp. 133 –149.

Arbogast, B. S., Edwards, S. V., Wakeley, J., Beerli, P. and Slowinski, J. B. (2002). Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics.* 33: Pp. 707 – 740.

Ashton, P. M. S. (1998). Seedling ecology of mixed Dipterocarp forest. *In* A review of Dipterocarps, taxonomy, ecology and silvicultrue (*eds.* S. Appanah and J. M. Turnbull). Center for forest research institute, Malesia. Pp. 87 – 98.

Ashton, P. S., Givnish, T. J. and Appanah, S. (1988). Staggered flowering in the Dipterocarpaceae. New sights into floral induction and the evolution of mast fruiting in the aseasonal tropics. *The American Naturalist.* 132 (1): Pp. 44 – 66.

Ashton, P. S. and Gunatilleke, C. V. S. (1987a). New light on the plant geography of Ceylon 1. Historical plant geography. *Journal of Biogeography.* 14: Pp. 249 – 285.

Ashton, P. S. (1969). Speciation among tropical forest trees: some deductions in the light of recent evidence. *Biological Journal of the Linnaean Society.* 1: Pp. 155 – 196.

Ashton, P. S. (1982). Dipterocarpaceae, in Flora Malesiana, Series 1, Spermatophyta (*ed.* C. G. G. J. Van Steenis), Martinus Nijhoff Publishers, The Hague. 1: Pp. 237 – 289.

Ashton, P. S. (1988). Dipterocarp biology as a window to the understanding of tropical forest structure. *Annual Review of Ecology and Systematics.* 19: Pp. 347 – 370.

Ashton, P. S. (1989). Dipterocarp reproductive biology. *In* Tropical rain forest Ecosystems. Biogeographical and ecological studies. (*eds.* H. Leith and M. J. A. Werger). Chapter 11: Pp. 219 – 240.

Aubreville, A. (1976). Essai d'interpretation nouvelle de la distribution des Dipterocarpacees. *Adansonia.* 12 (2): Pp. 205 – 210.

Avis, J. C. (1998). The history and preview of phylogeography: a personal reflection. *Molecular Ecology.* 7: Pp. 371 – 379.

Awasthi, N. (1971). Revision of some Dipterocarpaceous wood previously described from the Tertiary of South India. *Paleobotanist.* 18: Pp. 226 – 233.

Ayala, F. J. (1997). Vagaries of the molecular clock. *Proceedings of the National Academy of Science.* USA. 94: Pp. 7776 – 7783.

Azuma, H., Franco, J. G. G., Gray, V. R. and Thien, L. B. (2001). Molecular phylogeny of the Magnoliaceae: The biogeography of tropical and temperate disjunctions. *American Journal of Botany.* 88 (12): Pp. 2275 – 2285.

Bakker, F. T., Culham, A. and Gibby, M. (1999). Phylogenies and diversification of Pelargonium. *In* Molecular systematics and plant evolution (*eds*; P.M. Hollingsworth, R. M. Bateman and R. J. Gornall). Taylor & Francis Publishers, London. Pp. 354 – 374.

Baldwin, B. G. and Sanderson, M. J. (1998). Age and rate diversification of the Hawaiian Silversword alliance (Compositae). *Proceedings of the National Academy of Science.* USA. 95: Pp. 9402 – 9406.

Bancroft, H. (1935b). Some fossil dicotyledonous woods from Mont Elgon, East Africa. *American Journal of Botany.* 22: Pp. 164 – 183.

Bancroft, H. (1935d). The taxonomic history and geographical distribution of the Monotoideae. *American Journal of Botany.* 22 (5): Pp. 505 – 519.

Barmley, G. L. C., Pennington, R. T., Zakaria, R., Tjitrosoedirdjo, S. S., and Cronk, Q. C. B. (2004). Assembly of tropical plant diversity on a local scale: *Cyrtandra* (Gesneriaceae) on Mount Kerinci, Sumatra. *Biological Journal of the Linnaean Society.* 81: Pp. 49 – 62.

Benton, M. J. and Ayala, F. J. (2003). Dating the tree of Life. *Science.* 300: Pp. 1698 – 1700.

Birky, C. W. (1995). Uniparental inheritance of mitochondral and chloroplast genes. Mechanisms and evolution. *Proceedings of the National Academy of Science.* USA. 92: Pp. 11331 – 11338.

Birmingham, E. and Moritz C., (1986). Comparative phylogeography: concepts and applications. *Molecular Ecology.* 7: Pp. 367 – 369.

Bittkau, C. and Mueller-Starck, G. (Unpublished). Chloroplast DNA variation of *Acer campestre*, *Acer monspessulanum* and related species in Europe. (Submitted to the Forest Genetics section, Technical University of Munich, Germany).

Blume, K. L. (1825). Dipterocarpaceae. *In* Bijdragen tot de Flora van Nederlandish Indie. *Batavia.* 1: Pp. 1 – 42.

Bossuyt, F. and Milinkovitch, M. C. (2001). Amphibians as indicators of early Tertiary out-of-India dispersal of Vertebrates. *Science.* 292: Pp. 93 – 95.

Bousquet, J., Strauss, S. H., Doerksen, A. H. and Price, R. A. (1992). Extensive variation in evolutionary rate of *rbcL* gene sequences among seed plants. *Proceedings of the National Academy of Science.* USA. 89: Pp. 7844 – 7848.

Bremer, K. (2002). Gondwanan evolution of the grass alliance of families (Poales). *Evolution.* 56 (7): Pp. 1374 – 1387.

Bromham, L. and Penny, D. (2003). The modern molecular clock. *Nature reviews.* 4: Pp. 216 – 224.

Bruno, W. J. and Halpern, A. L. (1999). Topological bias and inconsistency of maximum likelihood using wrong models. *Molecular Biology and Evolution.* 16 (4): Pp. 564 – 566.

Cann, R. L., Stoneking, M. and Wilson, A. C. (1987). Mitochondrial DNA and human evolution. *Nature.* 325: Pp. 31 – 36.

Chase, M. W. *et al.,* (1993). Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbcL. Annals of Missouri Botanical Garden.* 80: Pp. 528 – 580.

Clegg, M. T., Gaut, B. S., Learn, G. H. and Morton, B. (1994). Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Science.* USA. 91: Pp. 6795 – 6801.

Comeron, J. M. and Aguade, M. (1998). An evaluation of measures of synonymous codon usage bias. *Journal of Molecular Evolution.* 47: Pp. 268 – 274.

Conti, E., Eriksson, T., Schonenberger, J., Sytsma, K. J. and Baum, D. A. (2002). Early Tertiary out-of-India dispersal of Crypteroniaceae: evidence from phylogeny and molecular dating. *Evolution.* 56 (10): Pp. 1931 – 1942.

Croizat, L. (1952). Manual of phytogeography. The Hague; D. W. Junk BV publishers.

Croizat, L., Nelson, G. and Rosen, D. E. (1974). Centers of origin and related concepts. *Systematic Zoology.* 23: Pp. 265 – 287.

Cronquist, A. (1968). The evolution and classification of flowering plants. Nelson, London and Edinburgh.

Curtis, S. E. and Clegg, M. T. (1984). Molecular evolution of chloroplast DNA sequences. *Molecular Biology and Evolution.* 1(4): Pp. 291 – 301.

Cutler, D. J. (2000). Estimating divergence times on the presence of an overdispersed molecular clock. *Molecular Biology and Evolution.* 17(11): Pp. 1647 - 1660.

Dalgren, R. (1975). A system of classification of the angiosperms to be used to demonstrate the distribution of characters. Botaniska Notiser. 128: Pp. 119 – 146.

Dayanandan, S. (1996). Phylogeny of the tropical tree family Dipterocarpaceae based on nucleotide sequences of the chloroplast *rbcL* gene and morphology. Ph.D. dissertation, Boston University, Boston, MA.

Dayanandan, S., Bawa, K. S. and Kesseli, R. V. (1997). Conservation of microsatellites among tropical trees (Leguminosae). *American Journal of Botany.* 84: Pp. 1658 – 1663.

Dayanandan, S., Ashton, P. S., Williams, S. M. and 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 (8): Pp. 1182 – 1190.

Davis, C. C., Bell, C. D., Fritsch, P. W. and Mathews, S. (2002). Phylogeny of Acridocarpus-Brachylophon (Malphigiaceae): Implications for Tertiary tropical floras and Afroasian biogeography. *Evolution.* 56 (12): Pp. 2395 – 2405.

Donoghue, M. J., Doyle, J. A., Gauthier, J., Kluge, A. G. and Rowe, T. (1989). The importance of fossils in phylogeny reconstruction. *Annual Review of Ecology and Systematics*. 20: Pp. 431 – 460.

Doyle, J. A. (1978). Origin of angiosperms. *Annual Review of Ecology and Systematics*. 4: Pp. 365 – 393.

Doyle, J. A. and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*. 19: Pp. 11 – 15.

Doyle, J. J., and Gaut, B. S. (2000). Evolution of genes and taxa: a primer. *Plant Molecular Biology*. 42: Pp. 1 – 23.

Doyle, J. A., Sanderson, M. J. and Magallon, S. (2001). Integrating fossil and molecular data on the age of angiosperms: effects of fossil age constraints and rate smoothing methods. *American Journal of Botany*.

Eriksson, T. (2002). The r8s-bootstrap-kit. Bergianska University. Stockholm, Sweden.

Fay, J. C. and Wu, C. I. (2003). Sequence divergence, functional constraint and selection in protein evolution. Department of Genome Science. Lawrence Berkley national laboratory, Berkley, California.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*. 17: Pp. 368 – 376.

Felsenstein, J. (1983). Parsimony in systematics: biological and statistical issues. *Annual Review of Ecology and Systematics*. 14: Pp. 313 – 333.

Felsenstein, J. (1988). Phylogenies from molecular sequences: Inferring and reliability. *Annual Review of Genetics*. 22: Pp. 521 – 565.

Felsenstein, J. (1993). PHYLIP (phylogenetic inference package). Department of Genetics, University of Washington, Seattle, Washington.

Felsenstein, J. (2004). Inferring Phylogenies. Sinauer Associates, Sunderland. Massachusetts.

FRA 2000. Global forest resources assessment 2000. FAO Forestry paper 140 (Main Report). Food and Agriculture Organization (FAO) of the United Nations. Publishing and multimedia service, Information division, FAO, Rome.

Ford, M. J. (2002). Applications of selective neutrality tests to molecular ecology. *Molecular Ecology*. 11: Pp 1245 – 1262.

Furley, P. A. and Newey, W. W. (1983). Evolution of the plant system. Geography of the biosphere. An introduction to the nature, distribution and evolution of the World's life zones. Butterworth and Company publishers Ltd. Pp. 142 – 163.

Gaut, B. S., Morton, B. R., McCaig, B. C. and Clegg M. T. (1996). Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proceedings of the National Academy of Science*. USA. 93: Pp. 10274 – 10279.

Gilg, E. (1925). Dipterocarpaceae. *In* Engler. Naturlichen Pflanzenfamilien. 121: Pp. 237 – 269.

Gielly, L. and Taberlet, P. (1994). The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcL* sequences. *Molecular Biology and Evolution*. 11(5): Pp. 769 – 777.

Gielly, L. and Taberlet, P. (1996). Chloroplast DNA sequencing to resolve plant phylogenies between closely related taxa. *In* Molecular genetic approaches in conservation (*eds.* T. B. Smith and R. K. Wayne). Oxford university press. New York. Pp. 143 – 153.

Graham, S. W., Olmstead, R. G. and Barrett, C. H. (2002). Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. *Molecular Biology and Evolution.* 19(10): Pp. 1769 – 1781.

Haffer, J. (1987). Biogeography and Quaternary history in tropical America. *In* Biogeography and Quaternary history in tropical America (*eds.* T. C. Whitmore and G. T. Prance). Clarendon press. Oxford. Pp. 1 – 18.

Hasegawa, M., Kishino, H. and Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution.* 22: Pp.160 – 174.

Hedges, S. B. (1992). The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. *Molecular Biology and Evolution.* 9 (2): Pp. 366 – 369.

Heim, F. (1892). Sur un nouveau genre de Dipterocarpaceae: *Vateriopsis seychellarum* Heim, *Vateria seychellarum* Dyer in Baker. *Bulletine de la Societe Botanique de France* 39: Pp. 149 – 154.

Hillis, M. D. and Huelsenbeck, J. P. (1992). Signal, noise and reliability in molecular phylogenetic analyses. *Journal of Heredity.* 83: Pp. 189 – 195.

Hilu, K. W. and Liang H. (1996). The *MatK* gene: sequence variation and application in plant systematics. *American Journal of Botany.* 84(6): Pp. 830 – 839.

Hilu, K. W. and Alice, L. A., (1999). Systematic and evolutionary implications of *matK* indels in Poaceae. *American Journal of Botany.* 86: Pp. 1735 – 1741.

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. and Chatrou, L. W. (2003). Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany.* 90 (12): Pp. 1758 – 1776.

Holloway, J. D. and Hall, R. (1998). South East Asian geology and biogeography: an introduction. *In* Biogeography and geological evolution of SE Asia (*eds.* R. Hall and D. Holloway). Backbuys Publishers, Leiden, The Netherlands. Pp. 1 – 23.

Huelsenbeck, J. P. and Crandall, K. A. (1997). Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics.* 28: Pp 437 – 466.

Huelsenbeck, J. P., Larget, B., Miller, R. E. and Ronquist, F. (2002). Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology.* 51: Pp. 673 – 688.

Huelsenbeck, J. P. and Rannala, B. (1997). Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science.* 276: Pp. 227 – 232.

Huelsenbeck, J. P. and Ronquist, F. R. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics.* 17: Pp. 754 – 755.

Huelsenbeck, J. P., Ronquist, F. R., Nielsen, R. and Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science.* 294: Pp. 2310 – 2314.

Hughes, A. L. (1999). What is adaptive evolution? Adaptive evolution of genes and genomes. Oxford university press. Chapter 1.

Humphries, C. J. and Parenti, L. R. (1999). Interpreting patterns of plant and animal distributions. Cladistic Biogeography. Second edition. Oxford university press.

Johnson, L. A. and Soltis, D. E. (1994). *matK* DNA sequences and phylogenetic reconstruction in Saxifragaceae *s. str. Systematic Botany.* 19: Pp. 143 – 156.

Jong, K. and Kaur, A. (1979). A cytotaxonomic view of Dipterocarpaceae with some comments on polyploidy with apomixes. *In* Dipterocarpacees: Taxonomie-Phylogenie-Ecologie, Memoires du Museum National d'Histoire Naturelle, (*ed.* G. Maury-Lechon) serie B, Botanique. Editions du Museum, Paris. 26: Pp. 50 – 56.

Judd, W. S., Campbell, C. S., Kellogg, E. A. and Stevens, P. F. (1999). Plant systematics; a phylogenetic approach. Sinauer Associates, Sunderland. Massachusetts.

Jukes, T. H. and Cantor, C. (1969). Evolution of protein molecules. *In* Mammalian Protein Metabolism, (*ed.* M. N. Munro). Academic Press, New York. Pp. 21 – 132

Kajita, T., Kamiya, K., Nakamura, K., Tachida, H., Wickneswari, R., Tsumura, Y., Yoshimaru, H. and Yamazaki, T. (1998). Molecular phylogeny of Dipterocarpaceae 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 (2): Pp. 202 – 209.

Kelchner, S. A. (2002). Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. *American Journal of Botany.* 89(10): Pp. 1651 – 1669.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature.* 217: Pp. 624 – 626.

Kimura, M. (1980). A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: Pp.111 – 120.

Kimura, M. (1981). Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Science.* USA. 78: Pp. 454 – 458.

Kimura, M. and Ohta, T. (1971). Protein polymorphism as a phase of molecular evolution. *Nature.* 229: Pp. 467 – 469.

Kishino, H., Miyata, T. and Hasegawa, M. (1990). Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution.* 31: Pp. 151 – 160.

Kostermans, A. J. G. H. (1985). Family status of the Monotoideae, Gilg. and the Pakaraimoideae, Ashton, Maguire and de Zeeuw (Dipterocarpaceae). *Taxon.* 34: Pp. 426 – 435.

Kreitman, M. and Akashi, H. (1995). Molecular evidence for natural selection. *Annual Review of Ecology and Systematics.* 26. Pp. 403 – 422.

Kreitman, M. and Comeron, J. M. (1999). Coding sequence evolution. *Current Opinion in Genetics and Development.* 9: Pp. 637 – 641.

Kron, K. A. (1997). Phylogenetic relationships of Rhododendroideae (Ericaceae). *American Journal of Botany.* 84: Pp. 973 – 980.

Kumar, S. and Hedges, S. B. (1998). A molecular timescale for vertebrate evolution. *Nature.* 392: Pp. 917 – 920.

Kumar, S., Tamura, K., Jakobsen, B. and Nei, M. (2001). MEGA: Molecular Evolutionary Genetics Analysis software, Arizona state university, Tempe, Arizona, USA.

Langley, C. H. and Fitch, M. W. (1974). An estimation of the constancy of the rate of molecular evolution. *Journal of Molecular Evolution.* 3: Pp. 161 – 177.

Lakhanpal, R. N. (1970). Tertiary floras of India and their bearing on the historical geology of the region. *Taxon.* 19: Pp. 675 – 694.

Lockhart, P. J., Steel, M. A., Hendy, M. D. and Penny, D. (1994). Recovering Evolutionary trees under a more realistic model of sequence evolution. *Molecular Biology and Evolution.* 11: Pp. 605 – 612.

Londono, A. C. E., Alvarez, E., Forero, E. and Morton, C. M. (1995). A new genus and species of Dipterocarpaceae from the Neotropics. Introduction, taxonomy, ecology and distribution. *Brittonia.* 47: Pp. 225 – 236.

McDonald, J. H. and Kreitman, M. (1991). Adaptive protein evolution at the *Adh* locus in Drosophila. *Nature* 351: Pp. 652 – 654.

Maddison, W. P. and Maddison D. R. (2001). *MacClade 4.02.* Analysis of phylogeny and character evolution. Sinauer Associates, Sunderland.

Magallon, S. and Sanderson, M. J. (2002). Relationships among seed plants inferred from highly conserved genes: sorting conflicting phylogenetic signals among ancient lineages. *American Journal of Botany.* 89 (12): Pp. 1991 – 2006.

Maguire, B. P. C. and Ashton, P. S. (1977). Pakaraimoideae, Dipterocarpaceae of the western hemisphere II. Systematic, geographic and phyletic considerations. *Taxon.* 26: Pp. 341 – 385.

Maguire, B. P. C., Giannasi, D. E., Niklas, K. J. and Zeeuw, C. D. (1977). Pakaraimoideae, Dipterocarpaceae of the western hemisphere. *Taxon.* 26: Pp. 341 – 385.

Maury-Lechon, G. and Curtet, L. (1998). Biogeography and evolutionary systematics of family Dipterocarpaceae. *In* A review of Dipterocarps, taxonomy, ecology and silvicultrue (*eds.* S. Appanah and J. M. Turnbull). Center for forest research institute, Malesia. Pp. 5 – 44.

Maury-Lechon, G. (1979). Interpretation Phylogenique des caracters des pollens, fruits-germinations, Embryos et plantules des Dipterocarpacees. *In* Dipterocarpacees: Taxonomie-Phylogenie-Ecologie, Memoires du Museum National d'Histoire Naturelle, (*ed.* G. Maury-Lechon). Editions du Museum, Paris. serie B, Botanique. 26: Pp.139 – 144.

McPherson, M. J. and Moller, S. G. (2000). PCR. BIOS Scientific publishers ltd. USA. Chapter 4.

Meher-Homji, V. M. (1979). Distribution of the Dipterocarpaceae: some phytogeographic considerations on India. *Phytocoenologia.* 6: Pp. 85 – 93.

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

Mindell, D. P. and Meyer, A. (2001). Homology evolving. *Trends in ecology and evolution.* 16(8): Pp. 434 – 440.

Moreira, D. and Philippe, H. (2000). Molecular phylogeny: pitfalls and progress. *International Microbiology.* 3: Pp. 9 – 16.

Moritz. C., Wilmer, J. W., Pope, L., Sherwin, W. B., Taylor, A. C. and Limpus, C. J. (1996). Applications of genetics to the conservation and management of Australian fauna: four case studies from Queensland. *In* Molecular genetic approaches in conservation (*eds.* T. B. Smith and R. K. Wayne). Oxford university press. New York. Pp. 442 – 456.

Moritz, C. and Faith, D. P. (1998). Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology.* 7: Pp. 419 – 429.

Morley, R. J. (1998). Palynological evidence for Tertiary plant dispersals in the South East Asian region in relation to plate tectonics and climate. *In* Biogeography and geological evolution of SE Asia (*eds.* R. Hall and D. Holloway). Backbuys Publishers, Leiden, The Netherlands. Pp. 211 – 234.

Morley, R. J. (1999). Origin and evolution of tropical rain forests. John Wiley and Sons, Chichester, U. K.

Morley, R. J. (2003). Interplate dispersal paths for megathermal angiosperms. *Perspective in Plant Ecology, Evolution and systematics.* 6: Pp 5 –20.

Morley, R. J and Dick, C. W. (2003). Missing fossils, molecular clocks, and the origin of the Melastomataceae. *American Journal of Botany.* 90 (11): Pp. 1638 – 1644.

Mort, M. E., Soltis, D. E., Soltis, P. S., Ortega, J. F. and Guerra, A. S. (2001). Phylogenetic relationships and evolution of Crassulaceae inferred from *matK* sequence data. *American Journal of Botany.* 88 (1): Pp. 76 – 91.

Morton, B. R. (1993). Chloroplast DNA codon use: Evidence for selection at *psbA* locus based on *tRNA* availability. *Journal of Molecular Evolution.* 37: Pp. 273 – 280.

Morton, C. M. (1995). A new genus and species of Dipterocarpaceae from the neotropics. II. Stem anatomy. *Brittonia.* 47: Pp. 237 – 247.

Morton, C. M., Dayanandan, S. and Dissanayake, D. (1999). Phylogeny and biosystematics of Pseudomonotes (Dipterocarpaceae) based on molecular and morphological characters. *Plant Systematics and Evolution.* 216: Pp. 197 – 205

Muller, J. (1970). Palynological evidence on early differentiation of angiosperms. *Biological Reviews of the Cambridge Philosophical society.* 45: Pp. 417 – 450.

Muse, S. V. and Weir, B. S. (1992). Testing for equality of evolutionary rates. *Genetics*. 132: Pp. 269 – 276.

Nei, M. and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*. 3 (5): Pp. 418 – 426.

Nei, M. and Kumar, S. (2000). Molecular evolution and Phylogenetics. Oxford University Press. Oxford. U.K.

Nyffeler, R. (2002). Phylogenetic relationships in the Cactus family (Cactaceae) based on evidence from *trnK/matK* and *TrnL-trnF* sequences. *American Journal of Botany*. 89 (2): Pp. 312 – 326.

Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*. 23: Pp. 263 – 286.

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

Ortega, J. F., Jansen, R. K., Gamer, R. J. M. and Wallace, R. S. (1996). Application of chloroplast DNA restriction site studies for conservation genetics. *In* Molecular genetic approaches in conservation (*eds.* T. B. Smith and R. K. Wayne). Oxford University press. New York. Pp. 183 – 201.

Page, R. D. M. and Holmes, E. C. (1998). Molecular evolution: A phylogenetic approach. Blackwell publishers, Oxford. U. K.

Poole, I. (1993). A Dipterocarpaceous twig from the Eocene London clay formation of Southeast England. *Special Papers in Palaeontology*. 49: Pp. 155 – 163.

Plunkett, G. M., Soltis, D. E. and Soltis, P. E. (1996). Evolutionary patterns in Apiaceae: inference based on *matK* sequence data. *Systematic Botany*. 21: Pp. 477 – 495.

Posada, D. and Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics*. 14: Pp. 817 – 818.

Prakash, U. (1972). Paleoenvironmental analysis of Indian Tertiary floras. *Geophytology*. 2: Pp. 178 – 205.

Rand, D. M., Weinreich, D. M. and Cezairliyan, B. O. (2000). Neutrality tests of conservative radical amino acid changes in nuclear and mitochondrially encoded proteins. *Gene*. 291: Pp.115 –125.

Raven, P. H. and Axelrod, D. I. (1974). Angiosperm biogeography and past continental movements. *Annals of the Missouri Botanical Garden.* 61: Pp. 39 – 637.

Richardson, J. E., Pennington, T. R., Pennington, T. D. and Hollingsworth, P. M. (2001). Rapid diversification of a species rich genus of Neotropical rain forest trees. *Science.* 293: Pp. 2242 – 2245.

Rodriguez, F., Oliver, J. F., Marin, A. and Medina, J. R. (1990). The general stochastic model of nucleotide substitutions. *Journal of Theoretical Biology.* 142: Pp. 485 – 501.

Rozas, J. and Rozas, R. (1995). DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Computer Applications in Bioscience.* 11: Pp. 621 – 625.

Saitou, N. and Nei, M. (1997). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution.* 4: Pp. 406 – 425.

Sanders, E. R., Karol, K. G. and McCourt, R. M. (2003). Occurrence of *matK* in a *trnK* group II intron in Charophyte green algae and phylogeny of the Characeae. *American Journal of Botany.* 90 (4): Pp. 628 – 633.

Sanderson, M. J. and Doyle, J. A. (2001). Sources of error and confidence intervals in estimating the age of Angiosperms from *rbcL* and *18S rDNA* data. *American Journal of Botany.* 88 (8): Pp. 1499 – 1516.

Sanderson, M. J. (1997). A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution.* 14: Pp. 1218 – 1232.

Sanderson, M. J. (1998). Estimating rate and time in molecular phylogenies: beyond the molecular clock? *In* Molecular Systematics of Plants II. Second edition, (*eds.* P. S. Soltis, D. E. Soltis and J. Doyle). Chapman and Hall, London. Pp. 242 – 264.

Sanderson, M. J. (2002). Estimating Absolute Rates of Molecular Evolution and Divergence Times: A Penalized Likelihood Approach. *Molecular Biology and Evolution.* 19 (1): Pp.101 –109.

Sanderson, M. J. (2002). r8s user's manual version 1.50. Available via http://ginger.ucdavis.edu/r8s/r8s.manual.pdf

Sanderson, M. J. (2003). R8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics Application Note.* 19 (2): Pp. 301 – 302.

Scotese, C. R. (2001). Paleoclimatic maps. PALEOMAP project. http://www.scotese.com

Seward, A. C. (1935). Leaves of dicotyledons from the Nubian sandstones of Egypt. Geological Survey, Egypt.

Sharp, P., Stenico, M., Peden, J. and Lioyd, A. (1993). Codon usage: mutation bias, translational selection or both? *Biochemical Society of Transactions.* 21: Pp. 835 – 841.

Shimodaira, H. and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution.* 16: Pp. 1114 – 1116.

Shinozaki, K., Ohme, K., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N, Chunwongse, J., Obokata, J., Yamaguchi, S. K., Ohto, C., Torazawa, K., Meng, B. Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986). The complete nucleotide sequence of tobacco chloroplast genome. *The EMBO Journal.* 5: Pp. 2043 – 2049.

Shiva, M. P. and Jantan, I. (1998). Non-timber forest products from Dipterocarps. *In* A review of Dipterocarps; Taxonomy, ecology and silvicultrue (*eds.* S. Appanah and J. M. Turnbull). Center for forest research institute, Malesia. Pp. 188 – 197.

Skelton, P. W. (1993). Evolution. A biological and Paleontological approach. First edition. Addison-Wesley publishing company. Bath press, Open University. Great Britain.

Smith, A. G., Smith, D. G. and Funnel, B. M. (1994). Atlas of Mesozoic and Cenozoic coastlines. Cambridge University press, Cambridge.

Soltis, D. E., Kuzoff, R. K., Conti, E., Gornall, R., and Ferguson, K. (1996). *MatK* and *rbcL* gene sequence data indicate that *saxifraga* (Saxifragaceae) is polyphyletic. *American Journal of Botany.* 83: Pp. 371 – 382.

Soltis, D. E. and Soltis, P. S. (1998). Choosing an approach and an appropriate gene for phylogenetic analysis. *In* Molecular Systematics of Plants II. Second edition (*eds.* P. S. Soltis, D. E. Soltis and J. Doyle). Chapman and Hall, London. Pp. 1 – 42.

Soltis, P. S., Soltis, D. E., Savolainen, V., Crane, P.R. and Barraclough, T. (2002). Rate heterogeneity among lineages of land plants: integration of molecular and fossil data and evidence for molecular living fossils. *Proceedings of the National Academy of Science.* USA. 99: Pp. 4430 – 4435.

Stanford, M., Harden, R. and Parks, C. R. (2000). Phylogeny and biogeography of *Juglans* (Juglandaceae) based on *matK* and ITS sequence data. *American Journal of Botany.* 87(6): Pp. 872 – 882.

Stace, C. A. (1989). Plant taxonomy and biosystematics. Second edition. Edward Arnold educational, academic and medical publishing division. Hodder and Stoughton ltd. Bedford square, London. Pp. 164 – 165.

Stebbins, G. L. (1974). The Ecological basis of diversity. Flowering plants. Evolution above the species level. The Belknap press of Harvard university press. Cambridge, Massachusetts. Pp. 155 – 170.

Steel, K. P. and Vilgalys, R. (1994). Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *matK*. *Systematic Botany.* 19: Pp. 126 –142.

Steel, M. A. *et al.*, (1988). Loss of information in genetic distances. *Nature.* 336: Page 118.

Steel, M. A., (2002). Some statistical aspects of the maximum parsimony method. *In* Molecular systematics and evolution: Theory and Practice (*eds.* R. DeSalle, G. Giribet and W. Wheeler), Birkhauser publishers, Basel, Switzerland. Pp. 125 – 139.

Stewart, C. B. (1993). The powers and pitfalls of parsimony. *Nature* 361: Pp. 603 – 607.

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.

Swofford, D. L., Olsen, G. J., Waddell, P. J. and Hillis, D. M. (1996). Phylogenetic inference. *In* Molecular Systematics, second edition (*eds.* D. M. Hillis, C. Moritz and B. K. Mable), Sinauer Associates, Sunderland. Pp. 407 – 514.

Swofford, D. L. (1999). PAUP* phylogenetic analysis using parsimony (* and other methods). Version, 4β10. Sinauer Associates, Sunderland, Massachusetts.

Tamura, K. and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution.* 10: Pp. 512 – 526.

Tateno, Y., Takezaki, N. and Nei. M. (1994). Relative efficiencies of the maximum-likelihood, neighbour joining, and maximum parsimony methods when substitution rate varies with site. *Molecular Biology and Evolution.* 11 (2): Pp. 261 – 277.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). ClastalW, improving the sensitivity of progressive multiple sequence alignment through sequence weighing, position, specific gap penalties and matrix choice. *Nucleic Acid Research*. 22: Pp. 4673 – 4680.

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

Vakhrameev, V. A. (1991). Jurassic and Cretaceous floras and climates of the earth. Cambridge University press. Cambridge.

Wakeley, J. (1996). The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore it significance. *Trends in Ecology and Evolution*. 11(4): Pp. 158 – 163.

Wang, X. Q., Tank, D. C. and Sang T. (2000). Phylogeny and divergence times in Pinaceae: evidence from three genomes. *Molecular Biology and Evolution.* 17 (5): Pp. 773 – 781.

Wikstrom, N., Savolanien, V. and Chase, M. W. (2001). Evolution of angiosperms: calibrating the family tree. *Proceedings of the Royal Society of London.* B 268: Pp. 2211 – 2220

Whitmore, T. C. (1989). South East Asian tropical forests. *In* Tropical rain forest ecosystems: biogeographical and ecological studies (*eds.* H. Lieth and M. J. A. Werger). Elsevier Science Publishers, B. V. Amsterdam, The Netherlands.

Whitmore, T. C. (1998). An introduction to tropical rain forests. Oxford university press.

Wright, F. (1990). The effective number of codons used in a gene. *Gene.* 87: Pp. 23 – 29.

Xiang, Q. Y., Soltis, D. E. and Soltis, P. S. (1998). Phylogenetic relationships of Cornaceae and close relatives inferred from *matK* and *rbcL* sequences. *American Journal of Botany.* 85: Pp. 285 – 297.

Yang, Z., Goldman, N. and Friday, A. (1994). Comparison of models for nucleotide substitution used in maximum likelihood phylogenetic estimation. *Molecular Biology and Evolution.* 11(2): Pp. 316 – 324.

Young, N. D. and Pamphilis, C. W. D. (2000). Purifying selection detected in the plastid gene *matK* and flanking ribozyme regions within a Group II intron of Nonphotosynthetic plants. *Molecular Biology and Evolution.* 17 (12): Pp. 1933 – 1941.

Zharkikh, A. (1994). Estimation of evolutionary distances between nucleotide sequences. *Journal of Molecular Evolution.* 39: Pp. 315 – 329.

Zimmerly, S., Hausner, G. and Wu., X. C. (2001). Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Research.* 29 (5): Pp. 1238 – 1250.

Zuckerkandl, E. and Pauling, L. (1965). Evolutionary divergence and convergence in Proteins. *In* Evolving Genes and Proteins (*eds.* V. Bryson and H. J. Vogel) Academic Press, New York. Pp. 97 – 166.

APPENDIX 1

Polymerase Chain Reaction (PCR) Protocol

Touchdown PCR

Touchdown PCR starts with an annealing temperature higher than the melting temperature of primers and then the annealing temperature is gradually lowered by 1^0 C or 2^0 C at each cycle* during the first ten or five cycles. Therefore, annealing of primers to their target sequence is more specific (McPherson and Moller, 2000).

PCR Program used to amplify *matK* gene of dipterocarp species

1) *Initial denaturation*: - 94^0 C; 3 minutes
2) 5 cycles, each with
   - *Denaturation:* 94^0 C; 1 minute
   - *Initial annealing:* 60^0 C; 30 seconds
     * decrease annealing temperature by 2^0 C at each cycle
   - *Terminal annealing:* 50^0 C; 30 seconds
   - *Extension:* 72^0 C; two minutes
3) 34 cycles, each with
   - *Annealing:* 60^0 C; 30 seconds
   - *Denaturation:* 94^0 C; 1 minute
   - *Annealing:* 50^0 C; 30 seconds
   - *Extension:* 72^0 C; two minutes
4) *Terminal extension:* 72^0 C; 5 minutes
5) *Hold:* 4^0 C

Each PCR reaction contained 2.3 µl of dNTP, 2.5 µl of $MgCl_2$, 2.5 µl of PCR buffer, 2.0 µl of 5 pmol/µl primers (1.0 µl each, forward and reverse), 0.2 µl of Taq DNA Polymerase and 2.0 µl of the DNA sample in a total volume of 25.0 µl with distilled water.

APPENDIX 2

The McDonald Kreitman (MK) test

H_0 = K_n/K_s ratios of a given gene do not differ significantly among species

H_A = K_n/K_s ratios of a given gene are significantly different among species

Under the neutral theory, the ratio of replacement (nonsynonymous) to silent (synonymous) substitutions should remain approximately the same within and between species (Page and Holmes, 1998). The McDonald Kreitman test (McDonald and Kreitman, 1991) compares the relative amount of amino acid fixations and polymorphisms using the ratios of K_n (nonsynonymous substitutions for nonsynonymous site) and K_s (number of synonymous substitutions per synonymous site) between species using Fisher's exact test and the neutrality index (NI).

Fisher's exact test compares amino acid fixations and polymorphisms using a 2 x 2 contingency table and calculates the probability of accepting or rejecting the null hypothesis of neutral evolution (Nei and Kumar, 2001).

The Neutrality index (NI) indicates the direction and magnitude of a gene's departure from neutrality (Rand *et al.*, 2000) based on the relative amounts of amino acid fixations and polymorphisms.

NI < 1: - excess of amino acid fixations (positive selection)

NI > 1: - excess of amino acid polymorphisms (negative/purifying selection)

NI = 1: - amino acid polymorphisms equals to amino acid fixations (neutral evolution)

APPENDIX 3

The Likelihood Ratio test (LRT)

H_0 = Observed pattern of nucleotide substitutions fit with the expected nucleotide substitution model

H_A = Observed pattern does not fit with the expected nucleotide substitution model

The likelihood ratio test (LRT) is a user defined hierarchical criterion used to select the best nucleotide substitution model for a given dataset (Swofford *et al.*, 1996; Huelsenbeck and Crandall, 1997; Huelsenbeck and Rannala, 1997).

LRT test statistic (δ) = – 2 log Λ; (Λ = Likelihood ratio)

$$\Lambda = \frac{\textbf{Max } [L_0 (\textit{Null Model} \mid \textit{Data})]}{\textbf{Max } [L_1 (\textit{Alternative Model} \mid \textit{Data})]}$$

$$\delta = 2\,([-\ln L_{null}] - [-\ln L_{alternative}])$$

Test statistic, 'δ' is chi-square (X^2) distributed

DF (degrees of freedom) = difference in the number of free parameters of the two models compared

The LRT tests different combinations of nucleotide substitution rates. The nested hypotheses that can be tested using LRT include, *1)* equal versus unequal base frequencies, *2)* 1 to 6 types of nucleotide substitutions, *3)* absence versus more than one invariable sites, *4)* variable sits with or without gamma distributed rate variation, *5)* molecular clocklike or nonclocklike evolution of DNA sequences (Huelsenbeck and Crandall, 1997). LRT uses the log-likelihood score to evaluate the same tree under

progressively parameter rich models, from the simplest Jukes-Cantor model to the most complex general time reversible (GTR) model.

Nucleotide substitution models tested in hierarchical likelihood ratio tests as implemented in Modeltest version 2.0 (Posada and Crandall, 1998) include,

1. **Jukes-Cantor (JC; nst = 1): -** Equal base frequencies, all substitutions equally likely (Jukes and Cantor, 1969). PAUP rate matrix: - ***aaaaaa***

2. **Felsenstein 1981 (F81; nst = 1): -** Variable base frequencies, all substitutions equally likely (Felsenstein, 1981). PAUP rate matrix: - ***aaaaaa***

3. **Kimura 2 parameter (K80; nst = 2): -** Equal base frequencies, variable transition and transversion frequencies (Kimura, 1980).
PAUP rate matrix: - ***abaaba***

4. **Hasegawa Kishino Yano (HKY; nst = 2): -** Variable base frequencies, variable transition and transversion frequencies (Hasegawa *et al.,* 1985).
PAUP rate matrix: - ***abaaba***

5. **Tamura Nei (TrN; nst = 3): -** Variable base frequencies, equal transversion frequencies, variable transition frequencies (Tamura and Nei, 1993).
PAUP rate matrix: - ***abaaea***

6. **Kimura 3 parameter (K3P; nst = 3): -** Variable base frequencies, equal transition frequencies, variable transversion frequencies (Kimura, 1981).
PAUP rate matrix: - ***abccba***

7. **Transition Model (TIM; nst = 4): -** Special case of Tamura Nei model Variable base frequencies, variable transitions, transversions equal (Posada and Crandall, 1998). PAUP rate matrix: - ***abccea***

8. **Tranversion Model (TVM; nst = 5): -** Special case of Kimura 3 parameter model. Variable base frequencies, variable transversions, transitions equal (Posada and Crandall, 1998). PAUP rate matrix: - ***abcdbe***

9. **Symmetrical Model (SYM; nst = 6): -** Equal base frequencies, symmetrical substitution matrix (Zharkikh, 1994). PAUP rate matrix: - ***abcdef***

10. **General Time Reversible (GTR, nst = 6): -** Variable base frequencies, symmetrical substitution matrix (Rodriguez *et al.,* 1990).
PAUP rate matrix: - ***abcdef***

(nst = number of substitution types)

Nucleotide substitutions are grouped hierarchically in the models described above. Those groupings are symbolized as rate classifications according to the following rate matrix implemented in PAUP ver. 4β10

	A	C	G	T
A	--	***a***	***b***	***c***
C	--	--	***d***	***e***
G	--	--	--	***f***
T	--	--	--	--

Akaike information criterion (AIC)

The Akaike information criterion penalizes the increase in the number of parameters when the addition of a new parameter does not increase the log likelihood by at least one unit (Arbogast *et al.,* 2002).

$AIC = 2m - 2L$; m = number of parameters in the model; L = maximum log likelihood value (Posada and Crandall, 1998).

APPENDIX 4

Phylogeny inferring methods

Methods used in phylogenetic analyses are evaluated based on five criteria (Page and Holmes, 1998); *viz.* 1) *power* (amount of data needed to produce correct results), 2) *efficiency* (how fast), 3) *consistency* (converge towards accurate solutions), 4) *robustness* (minor violations do not affect the results) and 5) *falsifiability* (method indicates violations of its assumptions). Different methods are sensitive to different combinations of the above criteria.

Maximum Parsimony (MP) methods are efficient, powerful, robust and consistent when the site-to-site rate heterogeneity is moderate with enough parsimony informative sites. Nevertheless, it fails to produce a true phylogeny when there are numerous substitutions along bifurcating branches, which outnumber the substitutions that show the common ancestry (Felsenstein, 1988). This phenomenon is commonly known as 'long branch attraction' where parsimony methods treat true parallelisms as shared derived characters (Judd *et al.,* 1999; Moreira and Philippe, 2000). Although distance methods are highly efficient in some instances, the conversion of a DNA sequence matrix to a distance matrix has a tendency to loose some important evolutionary information that would eventually produce inconsistent phylogenies (Steel *et al.,* 1988).

The maximum likelihood (ML) method stands out among other phylogeny inferring methods (Felsenstein, 1981; Swofford *et al.,* 1996; Tateno *et al.,* 1994) mainly because of its statistical reliability. It is powerful, robust, consistent and also falsifiable when it is based on an accurate model of sequence evolution (Yang, *et. al.,* 1994; Bruno and Halpern, 1999) but computationally less efficient. Bayesian methods are highly

efficient and statistically reliable (Huelsenbeck and Ronquist, 2001). Although ML methods tend to take a longer processing time, Bayesian inference takes less time to search the entire tree space and is thus considered to be the most efficient phylogeny inferring method for larger datasets. However, the accuracy of Bayesian methods depends on the number of generations used to run MCMC analysis and the appropriate selection of prior parameters (Huelsenbeck *et al.,* 2002).

The relative performance of phylogeny inferring methods depends on the tempo and mode of evolution of a given set of DNA sequence data (Page and Holmes, 1998), which satisfy or violate the underlying assumptions of the method under consideration. Since each method is based on a particular set of assumptions that differs from one another (Page and Holmes, 1998), DNA sequence data must be analysed using different methods and the relative congruence of the resulting phylogenetic trees should be evaluated with respect to known classification systems or evolutionary events. The basic concepts of the phylogeny inferring methods used in the present study are given below.

Parsimony method

Parsimony is a method of inferring phylogenies by finding the phylogeny for which the observed characters could have evolved with the least evolutionary change (Felsenstein, 1983). The maximum parsimonious tree is the one with the minimum number of evolutionary changes and the total number of changes between sequences is attributed to parsimony branch length (Felsenstein, 1983, Page and Holmes, 1998). Different sites are identified as parsimony informative if they provide information that distinguishes between different topologies. However, some variable sites that do not

involve distinguishing between topologies are ignored by the method. There are several variants of the parsimony method such as Wagner, Fitch and Dollo parsimony that allow different possible ways to represent nucleotide substitutions among species (Felsenstein, 1983).

If the observed substitution rate heterogeneity or transition-transversion bias is low in the DNA sequence dataset, all characters are given equal costs and equally weighed parsimony method could be used.

In equally weighed parsimony, for k number of characters, each with l number of changes, the sum of the lengths L is simply given by,

$$* \mathbf{L} = \sum_{i=1}^{k} l_i$$

Neighbour-joining (NJ) method with minimum evolution (ME) algorithm

The neighbour joining method uses a converted distance matrix of DNA sequence data to build a phylogenetic tree. It clusters pairs of neighboring taxa in an additive starlike tree minimizing the total branch length at each stage of clustering (Saitou and Nei, 1987). The minimum evolution algorithm allows for the estimation of minimum possible number of changes based on a distance matrix when employed with a distance method such as neighbour joining (Page and Holmes, 1998; Felsenstein, 2004). For an unrooted tree with n number of terminal nodes and $(2n - 3)$ branches, each with length, e_i, the sum of branch lengths, **L,** is given by,

$$* \mathbf{L} = \sum_{i=1}^{2n-3} e_i$$

Maximum likelihood (ML) method

The maximum likelihood method evaluates a hypothesis about evolutionary history, which is based on a model (Appendix 5) of DNA or protein evolution (Felsenstein, 1988; Page and Holmes, 1998). Likelihood is the probability (P) that the DNA sequence data (D) will generate a specified phylogenetic tree, given the hypothesis, H (model). Based on the best fitting nucleotide substitution model (Appendix 3) therefore, likelihood methods describe the evolutionary pattern of DNA sequence data.

$$* \; L_H = P\,(D/H)$$

Each character in a dataset has an independent likelihood (L). For a large number of characters, therefore, the sum of the likelihood for all characters (1 to k) has to be considered. Since the actual value of the maximum (total) likelihood is very large, log values (*ln* L) are used to interpret the maximum likelihood.

$$* \; ln\, L = \sum_{i=1}^{k} ln\,(Li)$$

Bayesian inference

Bayesian methods provide the probability for a hypothesis (evolutionary pattern) given the sequence data, commonly known as posterior probability (Huelsenbeck and Ronquist, 2001) by using a Marcov Chain Monte Carlo method (MCMC), which searches the tree space in a stepwise fashion. Each step represents one generation and each new step is either accepted or rejected according to the Metropolis-Hastings-Green algorithm that calculates the probability for each step and counts it as a generation.

The posterior probability for the i^{th} phylogenetic tree, τ_i, based on the observed matrix of aligned DNA sequences (X) is obtained using Bayes formula,

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{B(s)} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

$f(\tau_i \mid X)$ => The posterior probability of the i^{th} phylogeny
(probability that τ_i is the correct tree given the DNA sequence data).

$f(X \mid \tau_i)$ => The likelihood for the i^{th} tree

$f(\tau_i)$ => The prior probability for the i^{th} tree

The summation in the denominator, ***B(s)*** is the total number of trees that are possible for ***(s)*** number of species (Huelsenbeck and Ronquist, 2001).

APPENDIX 5

Shimodaira Hasegawa test

H_0 = All trees are equally good explanations of the DNA sequence data. Therefore, the expected log likelihood values of sites do not significantly differ from the observed log likelihood values

H_A = Some or all trees are not equally good explanations of the data because the difference between the expected and observed log likelihood values of sites are significant.

The expected log likelihood for a given site is the average log likelihood per site. If evolution at different sites is independent and the two trees that are compared have equal likelihoods for having sites with the same pattern of evolution, the difference between expected and observed log likelihood values should be equal to zero (Felsenstein, 2004). The Shimodaira Hasegawa test based on the RELL (resample estimated log likelihood) method (Kishino *et al.,* 1990) generates 1000 (or more) nonparametric bootstrap replicates of sites and evaluates each tree based on the sum of differences for the log likelihood values for each replicate (Shimodaira and Hasegawa, 1999). The null hypothesis is rejected if the probability of having different likelihood values is significant at the alpha level of 0.05. The RELL test is considered to be more reliable because it responds to the differences for the sites at the tail part of the log likelihood distribution, which have a large effect on the differences in log likelihood values (Felsenstein, 2004). Therefore, it is sensitive enough to detect small differences in the alternative topologies that are compared.

APPENDIX 6

Testing the molecular clock hypothesis

H_0 = Substitution rates among lineages are relatively constant over time (clocklike)

H_A = Substitution rates among lineages are heterogeneous (non-clocklike)

Index of Dispersion [R (t)]

The index of dispersion is a simple and straightforward method to test for molecular clocklike evolution in a given group of taxa. It measures the substitution rate variation between lineages and tests whether there is a deviation from the expected value of the Poisson process (Sanderson, 1998). The Poisson distribution is the statistical distribution that describes the variation in accumulation of substitutions in the context of neutral theory and is commonly known as the Poisson clock (Page and Holmes, 1998). If the data fit the Poisson clock the variance in substitution rates should not be greater than the mean rate of substitution.

*** *R (t)* = Variance/Mean**

R(t) = 1.0 :- molecular clocklike evolution

R(t) < 1.0 :- clocklike evolution with underdispersed molecular clock

R(t) >>1.0 :- rate heterogeneity with overdispersed molecular clock

However, the utility of this test has been controversial (Sanderson, 1998; Cutler, 2000) because it assumes a starlike topology. The deviation of the true phylogeny from a starlike phylogeny, therefore, erroneously increases the value of *R (t)*. Moreover, the rate

variation introduced by independent sequences cannot be determined since the test is topology dependent.

Relative rate test

The relative rate test is a topology independent test that helps to identify the lineage specific rate heterogeneity (Sanderson, 1998). The rate variation due to the other lineages does not affect this test. It measures the rate variation between a given pair of taxa with respect to its most recent common ancestor (Sanderson, 1998). If the rates are constant, the accumulated changes from the common ancestor of two taxa are expected to be the same compared to their most recent common ancestor.

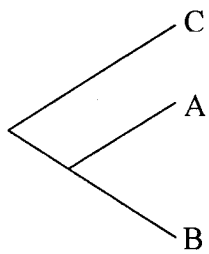


Fig. 15. Schematic phylogenetic tree used to illustrate Tajima's relative rate test

Tajima (1993) describes a version of the relative rate test that measures the difference in number of sites that have different nucleotide substitutions when two taxa (A and B) are compared with their most recent common ancestor, C (Fig. 15).

m_A = The observed number of sites in which nucleotides in sequence A are different from those in sequences B and C

m_B = The observed number of sites in which nucleotides in sequence B are different from those in sequences A and C

Therefore, $m_A - m_B$ is expected to be zero under rate constancy.

Since $(m_A - m_B)^2/(m_A + m_B)$ is chi-square (X^2) distributed with DF = 1, the statistical significance of rate variation between two taxa, A and B with respect to C can be evaluated.

Nevertheless, relative rate tests are considered to be liberal in accepting the null hypothesis of rate constancy and are not sensitive to the rate heterogeneity of the other parts of the tree (Sanderson, 1998). Therefore, it is problematic to combine the statistical conclusions from multiple relative rate tests that have nodes of a phylogenetic tree in common (Sanderson, 1998).

Likelihood ratio test

The likelihood ratio test is considered to be the most robust method for testing the molecular clock because it rejects the rate constancy even if a single branch deviates from the molecular clock (Sanderson, 1998).

In this method, log likelihood values with and without enforcing the molecular clock are calculated (Felsenstein, 1988) and the statistical significance of the difference in log likelihood values is compared using the likelihood ratio test statistic, δ

$$\delta = -2 (\log L_{\text{clock}} - \log L_{\text{no clock}})$$

See appendix 3 for a detailed description of the likelihood ratio test.

APPENDIX 7

Molecular dating methods used in the present study

Langley Fitch method

Langley and Fitch (1974) developed a straightforward global test that directly uses the branch lengths of a phylogenetic tree to estimate substitution rates. Based on the assumption of a globally constant molecular clock, it estimates one substitution rate across the entire tree (from root to tips) and a set of absolute divergence times for all unfixed nodes, using the calibration points based on fossil data (Sanderson, 2002). Then the expected branch lengths are calculated and the observed and expected branch lengths are statistically compared using a chi-square test.

Penalized Likelihood method

Upon the rejection of the molecular clock hypothesis, the computer software r8s integrates autocorrelated substitution rate variation into molecular dating algorithms through nonparametric rate smoothing (NPRS; Sanderson, 1997) and penalized likelihood (PL) methods. Both methods relax the assumption of rate constancy by smoothing the rapidity of rate variation across lineages. The smoothing criteria used in these methods are analogous to the smoothing techniques in regression analyses (Sanderson, 2003). Nonparametric rate smoothing is an entirely non-parametric method that estimates ages and times using a least square-smoothing criterion (Sanderson, 1997). Nevertheless, NPRS tends to overestimate ages (Doyle *et al.,* 2001; Sanderson, 2002, 2003) when the substitution rates substantially differ across taxa.

Penalized likelihood is a semiparametric method that is more reliable in handling the rate variation than NPRS, since the roughness penalty added to the smoothing criteria compromises the statistical power of the parametric branch length estimations of the maximum likelihood method and the strength of the nonparametric approach (Sanderson, 2002). A cross validation procedure (CV) is used to determine the optimal smoothing parameter that represents the substitution rate variation from root to tips (Sanderson, 2002). Cross validation prunes taxa from the tree based on a given smoothing value (λ) estimating parameters from the new matrix and then predicts the removed information by using the estimated parameters. The output of cross validation analysis lists the chi square (X^2) error associated with the observed and predicted data, where the optimal smoothing level corresponds to the lowest X^2 error estimate (Sanderson, 2002).

Semiparametric rate smoothing criteria

used in the penalized likelihood method (Sanderson, 2002)

Consider a rooted phylogenetic tree (e.g. Fig. 16), in which the node k has an age t_k, and its most recent common ancestor, *anc(k),* has an age of $t_{anc(k)}$.

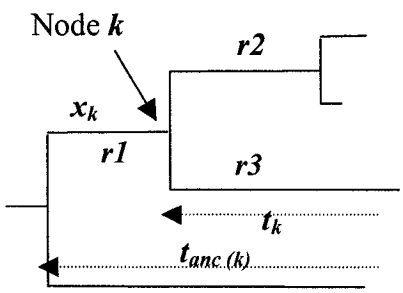


Fig. 16 Schematic phylogenetic tree that shows the local rates and duration of time between two nodes.

The duration of time for the branch defined by these two nodes => $t_{anc(k)} - t_k$; the number of substitutions along a branch is x_k.

Thus, for a given node such as k, the local rate, $r_k = x_k / t_{anc(k)} - t_k$

Assume that a phylogenetic tree has M taxa and $S + 1$ internal nodes. The root node is labeled by 0 and the remaining internal nodes are labeled by integers from {1,.., S}. The terminal nodes are labeled with integers {S + 1,..., S + M}. There are two different models at the opposite extremes that justify the rate variation among those lineages. One model explains molecular clocklike evolution (CL model), in which the rate parameters are the same for every branch ($r_k = r$). At the other extreme is the substitution rate saturated (SAT) model, in which each branch has a unique substitution rate (r_k).

Unknown parameters for CL model => $\theta_{CL} = \{t_0, \ldots\ldots, t_s; r\}$

Number of free parameters for CL model = S + 2.

Unknown parameters for SAT model => $\theta_{SAT} = \{t_0, \ldots\ldots, t_s; r_1, \ldots\ldots r_{s+m}\}$

Number of free parameters for saturated model = 2S + M + 1

According to the standard probability function used in the maximum likelihood analysis, the probability of an observation x (number of substitutions along a branch) taken from a poisson distribution with parameter Y is given by, $P(x|Y) = Y^x \exp(-Y)/x!$

Therefore, the log likelihood of 'θ' for the SAT model can be written as,

$$\log L(\theta_{SAT} \mid x_1, \ldots, x_{S+M}) = \sum_{k-1}^{S+M} \log P(x_k \mid r_k [t_{anc(k)} - t_k])$$

For the clocklike (CL) model, r is substituted for r_k in the above equation. In the CL model, there is one rate from root to tips with a known number of S. Therefore, θ can be estimated by numerical methods such as that of Langley Fitch (1974).

However, in the *SAT* model $t_0 - t_s$ indicates unknown times and $r_1 - r_{S+M}$ indicates unknown rates that are unique to the specified branch of the tree. In this model, therefore, unknown parameters are higher than the actual number of observations (S + M). Since several parameter values can produce the same likelihood, the model is uncertain. Thus it is not possible to estimate the unknown parameters of the saturated model (θ_{SAT}), without imposing some constraints on substitution rate variation.

In order to **penalize** the strict likelihood functions, therefore, a roughness penalty (Φ) is introduced to force rates to change smoothly from branch to branch. The roughness penalty should be designed to represent the changes in rate between neighboring branches of the tree.

Roughness penalty (Φ) = $\ldots+(r_1 - r_2)^2 + (r_1 - r_3)^2 + \ldots$

Thus, the objective function (Ψ) used in penalized likelihood could be explained as indicated in the following equation,

$$\Psi(\theta_{SAT} \mid x_1, \ldots, x_{S+M}) = \log L(\theta_{SAT} \mid x_1, \ldots, x_{S+M}) - \lambda \Phi(r_1, \ldots, r_{S+M})$$

Here 'λ' is the smoothing parameter that controls the tradeoff between smoothness and goodness-of-fit of the data to the SAT model. At one extreme, λ=0 (*SAT* model) and at the other extreme, as smoothing reaches infinity ($\lambda \rightarrow \infty$), the parameter estimates are expected to converge to those of the CL model because no variation in rate is tolerated.

APPENDIX 8

Calibration of the molecular clock using fossil data

Fossils generally constrain the nodes of a phylogenetic tree with the minimum ages of the most recent common ancestors of extant lineages and are often considered as extinct sister lineages of extant taxa rather than their ancestors (Donoghue *et al.*, 1989; Cutler, 2000; Sanderson and Doyle, 2001; Benton and Ayala, 2003). Therefore the earliest fossils unequivocally attributed to the specified extant lineage should be used to estimate divergence dates (Sanderson, 1998).

To determine appropriate nodes to which the fossils could unambiguously be assigned, the phylogenetic relationships of a given group of taxa in the derived phylogenetic tree should be well resolved (Sanderson, 1998). For instance, consider the phylogenetic tree in Fig. 17. Sanderson (1998) describes that a fossil, which is used to infer the minimum age of a stem lineage such as *X*, should possess synapomorphies (***b***) as well as apomorphies (***a***) that define both *X* and *Y* (e.g. fossil 1). In order to constrain the age of the crown group, *Y*, the fossil should have synapomorphies such as ***c*** of a sub clade of *Y*.

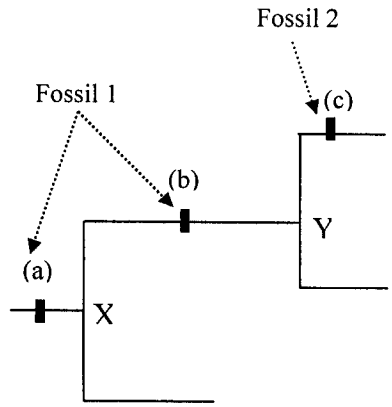


Fig.17 Assigning fossils to respective nodes of a phylogenetic tree (Sanderson, 1998)

The criteria illustrated above were used in the present study to assign distinct fossils of dipterocarps to the nodes of well-resolved clades in the phylogenetic trees based on *matK* nucleotide sequences. Paleobotanical characteristics (wood anatomy and palynological data) of *Anisopteroxylon ramunculiformis* (Poole, 1993), *Dipterocarpoxylon* (Bancroft, 1935), *Dryobalanoxylon* (Muller, 1970), *Hopenium* and *Vaticoxylon* (Maury and Curtet, 1998) suggest close affinities between these fossils and the extant genera *Anisoptera, Dipterocarpus*, *Dryobalanops, Hopea* and *Vatica* respectively.

The earliest fossils of *Vaticoxylon* and *Hopenium* could easily be assigned to their crown lineages (nodes O and Q; Figs. 9 and 11) because all species that clustered in the sub clades of O and Q belong to the genera, *Vatica* and *Hopea* respectively, suggesting shared derived character states among species within each sub clade. Therefore, *Vaticoxylon* and *Hopenium* were used to map other available fossils in the molecular clocklike 29-taxon ML tree calibrated by their relative ages (Fig. 18).

Although the distinct fossils of *Dipterocarpoxylon* and *Dryobalanoxylon* could be assigned to nodes D and K (Fig. 9) the branches extended from those nodes were collapsed in the bootstrap analyses (Figs. 4 and 7) suggesting uncertainties in the phylogenetic placements of *Dipterocarpus* and *Dryobalanops*. Therefore, those fossils were not considered as distinct calibration points.

Assigning *Anisopteroxylon ramunculiformis* to the crown lineage of the Dipterocarpeae (node L) was problematic because of the other clades emerge from node L that do not relate to *Anisopteroxylon*. However the polytomies of node L strongly suggested parallel evolution of *Stemonoporus*, *Vateriopsis, Anisoptera, Vateria* and

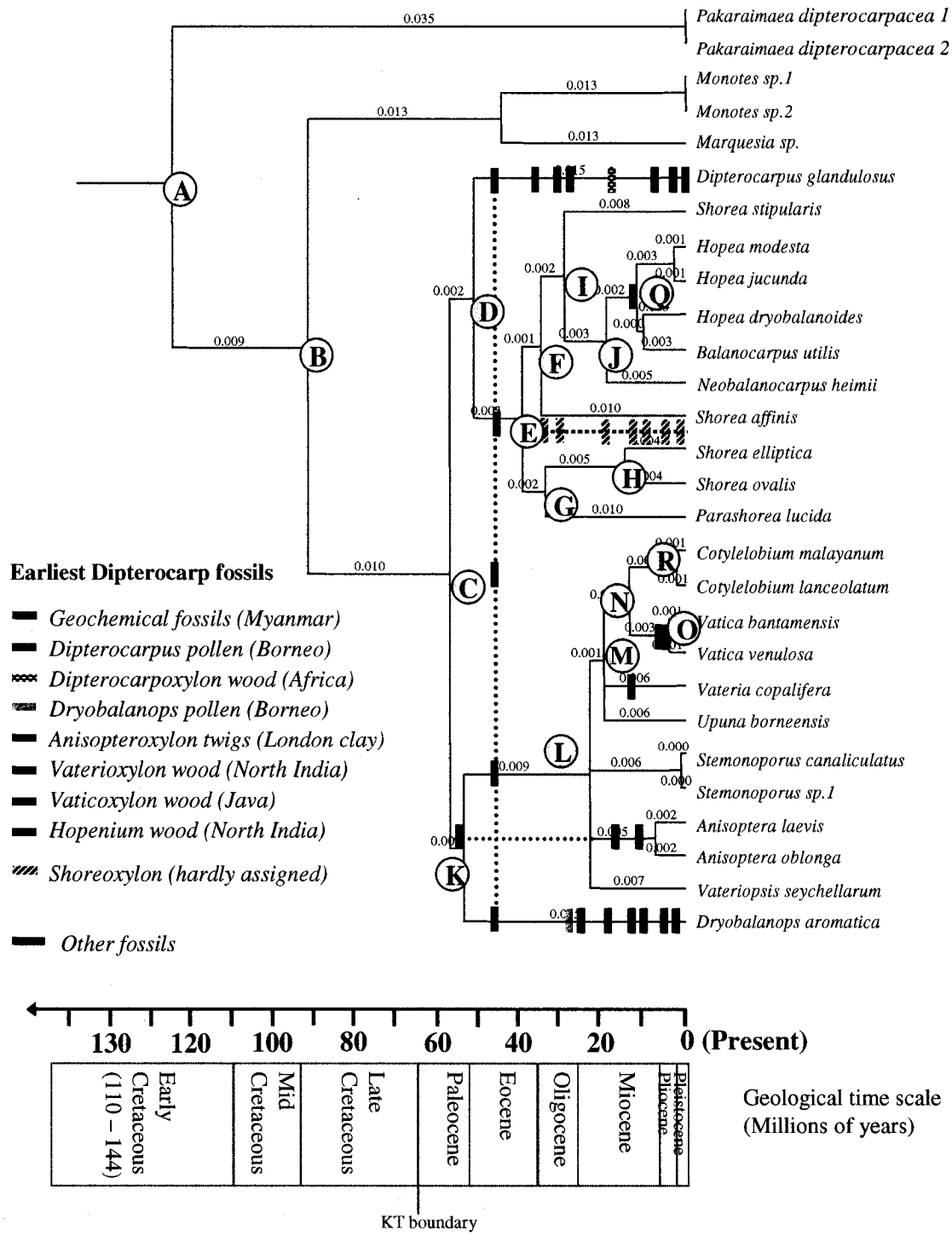


Fig. 18 Dipterocarp fossils mapped on the molecular clocklike maximum likelihood tree based on *matK* nucleotide sequences of 29 taxa

Upuna, with > 80% bootstrap support (Figs. 4 and 7). Based on the assumption of parallel evolution, therefore, *A. ramunculiformis* was alternatively assigned to the stem and crown lineages of the tribe Dipterocarpeae (nodes K and L) and the ages were estimated separately in order to distinguish between these two calibration points.

If the ages estimated by a particular fossil, assuming molecular clocklike evolution are less biased, the documented ages of the other fossils should be closer to the estimated ages. The results indicated an overestimation of >30 million years for the documented ages of the other fossils when *A. ramunculiformis* was used to calibrate the node L. These ages were inconsistent with the particular epochs in the geological time scale to which the other fossils have been attributed by authors, Bancroft (1935), Muller (1970) and Awasthi (1971). Nevertheless, based on the node K, *Anisoptera ramunculiformis* produced less biased age estimates, most of which were similar to the independent age estimates of *Vaticoxylon* and *Hopenium*. Therefore, *A. ramunculiformis*, *Vaticoxylon* and *Hopenium* were used as independent calibration points to estimate the minimum divergence dates of the unknown cladogenic events.