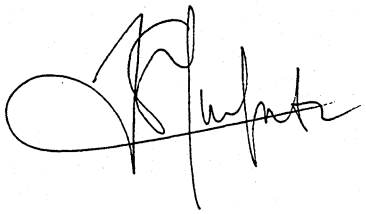
Generic Delimitation of *Hopea* Roxb. and *Shorea* Roxb. ex. C.F.Gaertn. (Dipterocarpaceae): Molecular and Morphological Evidence

Kusumadewi Sri Yulita

January 2001

A thesis submitted for the degree of Doctor of Philosophy of The Australian National University, Canberra, Australia

I, Kusumadewi Sri Yulita, certify that this thesis is my own original work and has not been submitted in a previous application for a higher degree.



Kusumadewi Sri Yulita

29 January 2001

ACKNOWLEDGEMENTS

I would like to express my appreciation to the Global Environment Facility-Indonesia Biodiversity Collection Projects (GEF-IBCP) for its financial support for my project, which was coordinated by the Research and Development Centre for Biology of the Indonesian Institute of Science. I would also like to convey my gratitude to my panel supervisors: Dr Judy G. West and Dr Randall J. Bayer of CSIRO Plant Industry, and Prof. Peter J. Kanowski and Dr Ann Gibson of the Department of Forestry A.N.U., for the supervision and advice that they rendered during my academic endeavours.

The project has been assisted by the collaboration of several institutions. These include the Centre for Plant Biodiversity Research, CSIRO Plant Industry, the Australian National Herbarium, the Herbarium Bogoriense, Bogor Botanic Garden, the Forestry Research Institute of Malaysia, the Harvard University Herbaria, Tropenbos-Wanariset Project, and Forestry Department of Universitas Mulawarman. I thank all their staff who have provided some direct assistance for my project.

I also thank my colleagues, Joe Miller and Bernard Pfeil for their valuable criticism and comments for improvement of my thesis and I thank Sucipto for his generous assistance in caring for my field collections from the Bogor Botanic Garden.

My deepest gratitude goes to some wonderful people who were part of my life during my agonising years of study, supporting me with their love and encouragement: my family; Nia Karina and Ari.

My last and greatest gratitude is conveyed to Judy West. This thesis could not have been completed without her slavery work during the past few months.

Canberra, January 2001.

Generic delimitation of *Hopea* Roxb. and *Shorea* Roxb. ex C.F.Gaertn. (Dipterocarpaceae): Molecular and Morphological Evidence

ABSTRACT

The two largest genera in the Dipterocarpaceae, *Hopea* and *Shorea*, have many similarities and exhibit continuous morphological variation at both the generic and specific levels and they are regarded as closely related genera (Ashton, 1982). The many similarities between the two genera leave very few discrete characters to separate them. The single and most conspicuous morphological character distinguishing the two genera is the comparative development of the fruit calyx. *Hopea* is characterised by two long and three short fruit calyx wings, while *Shorea* has three long and two short wings on the fruit (Ashton, 1982).

This study investigated the phylogenetic relationship of *Hopea* and *Shorea* to address the issue of their generic delimitation. Observations and measurements were made of morphological characters, and DNA sequences were obtained for the *trn*L-F region of the chloroplast and the ITS region of the nuclear genome.

Cladistic analyses were performed on a dataset of 40 selected morphological characters, categorised as either quantitative or qualitative. These analyses enabled the construction of a putative phylogeny of *Hopea* and *Shorea*, and the characters that define each genus were identified and examined. A detailed study of the inflorescence structure of some selected Dipterocarpaceae species was also carried out. The inflorescence was parsed into hierarchical nested units and the characters obtained were incorporated into the cladistic analyses. Several analyses were performed to test the effect of different parts of the data set on the robustness of the resultant topologies. Results from the morphological study showed that neither *Hopea* nor *Shorea* are monophyletic genera.

Analyses of the molecular data sets were performed to infer phylogenetic relationships using independent sources of evidence, the chloroplast and nuclear genomes.

Analyses that examined the effect of insertion-deletion events and of different putative outgroups on the robustness of the resultant topologies were also performed. The results suggested that *Hopea* is probably a monophyletic genus (albeit with some minor recircumscription) while *Shorea* is clearly non-monophyletic.

Since the study used two independent data sets—morphological and molecular–a combined analysis using both was also performed. This combination of data provided a better insight into the relationships of *Hopea* and *Shorea*. Results from this analysis were largely similar to those obtained from analyses of molecular data.

Table of Contents

Page

List of Figures

Page

Page

List of Tables

Page

CHAPTER 1
GENERAL INTRODUCTION

Outline

1.1 Scope of Project

The major aim of this study is to reconstruct the phylogeny and assess the evolutionary relationships of *Hopea* and *Shorea*, the two largest genera in the flowering plant family Dipterocarpaceae. This was achieved by examining the plant taxa to collect as many morphological and molecular features as possible and incorporating these characters into a series of cladistic analyses. The results of this research enable a re-evaluation of the currently accepted generic and infra-generic divisions of *Hopea* and *Shorea*. However, a formal taxonomic treatment of *Hopea* and *Shorea* was not within the scope of this study.

Hopea and *Shorea* contain the majority of species within Dipterocarpaceae, making up 60% of the total number. The Dipterocarps are generally regarded as the most successful angiosperm family found in the tropical forests of South East Asia, with regard to both tree size and biomass and number of species (Meijer, 1974). Most species are commercially important for timber, particularly in Malesia[1]. Indonesia supplies more than 70% of the world's demand for plywood made from hardwood, principally from Dipterocarpaceae (Choong and Achmadi, 1996).

Dipterocarpaceae are considered to be the most important timber family in Asia for its abundance and because of the workability, high finishing qualities, inherent beauty and strength of its timber (Choong and Achmadi, 1996). Industrial uses range from spools for textile factories to beams for heavy construction in bridges. The attractive grain and colours mean that the timber is also commonly used for

[1] Malesia includes Indonesia, Malaysia, Singapore, Brunei, Peninsular Thailand and the Philippines.

decorative veneers. The most common use of Dipterocarp timber, however, is for plywood manufacturing, moulding, and as furniture components.

1.2 Distribution: geography and ecology

Dipterocarpaceae was initially considered to consist of three subfamilies that were grouped according to their geographic distributions. However, the discovery of *Pseudomonotes tropenbosii*, which occurs in Amazonian Colombia in South America (Londono *et al.*, 1995; Morton, 1995), and its placement in a formerly African subfamily has caused a reassessment of these divisions. The first subfamily, Monotoideae, is represented by two genera (*Monotes* and *Marquesia*) in Africa (De Candolle, 1868; Baker, 1877; Dalziel, 1937; Verdcourt, 1989; Friedmann, 1994), and one genus (*Pseudomonotes*) in South America (Londono *et al.*, 1995; Morton, 1995; Morton *et al.*, 1999). The second subfamily is the South American Pakaraimoideae, which is represented by a monotypic genus, *Pakaraimaea* (Maguire and Ashton 1977). The third and largest subfamily is the Asian Dipterocarpoideae which includes 13 genera and around 470 species (Verdcourt, 1989; Smitinand *et al.*, 1990). Thus, the present distribution pattern of the Dipterocarpaceae represents a pan-tropical range (Figure 1.1) and the disjunct distributions of the three subfamilies are considered to be the result of geographical changes in the Tertiary (Ashton, 1982).

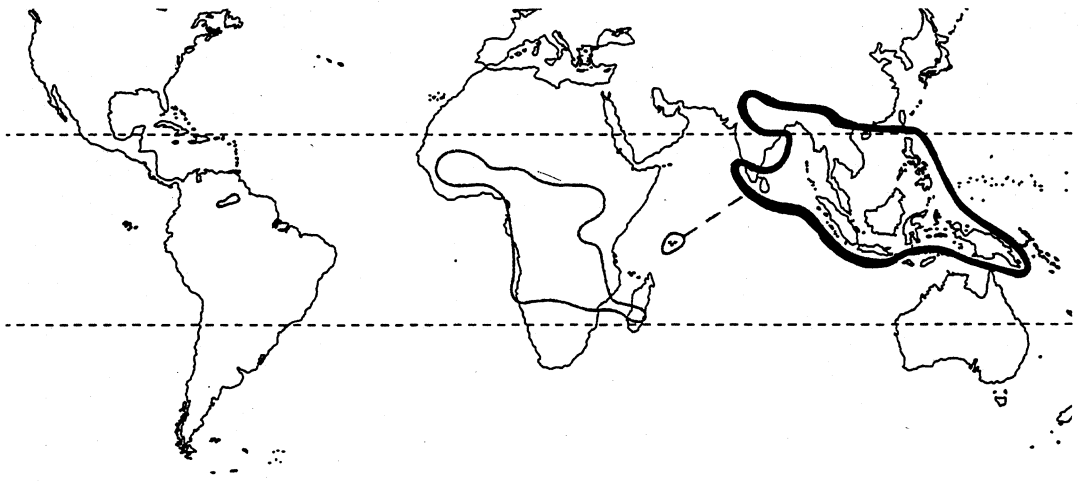


Figure 1.1 World distribution of the Dipterocarpaceae (after Ashton, 1982). The thickest outline represents the sub-family Dipterocarpoideae.

Species of Dipterocarpaceae occur in almost every edaphic and climatic zone including heath, swamp, riparian, lowland (granite, limestone and sandstone), and

sub-montane forests (Whitmore, 1975; Ashton, 1977; Ashton, 1980; Ashton, 1982; Curran and Leighton, 2000).

Most Dipterocarpaceae species have a tree form and most are valuable for timber production. The African species are represented by tall forest (*Marquesia*) and savannah (*Monotes*). This difference in habit may be related to ecological adaptation. The members of sub-family Monotoideae in tropical Africa may be prone to longer droughts than the allied sub-families which occur in tropical South America and Asia.

Hopea and *Shorea* species predominate in the sub-family Dipterocarpoideae. This sub-family occurs mainly in the tropical evergreen forest of the Far East where the climate is usually wet throughout the year. Such forest is often referred to as "rainforest". This forest formation is luxuriant and usually characterised by three distinct canopy layers. In the top layer, emergent Dipterocarp dominants form large timber trees to almost 40 m tall. The geographic distribution of this forest formation covers South West Sri Lanka, the West Ghats of Peninsula India, the southern wall of the Himalaya, tropical Indochina including Cambodia, Laos, Myanmar, South China, Hainan, Thailand, and almost the whole of Malesia including the whole of the island of New Guinea (Tardieu-Blot, 1950; Martin, 1971; Whitmore, 1975; Aubreville, 1976; Clunie, 1978; Ashton, 1980; Ashton, 1982; Saldanha and Rao, 1985; Smitinand *et al.*, 1990).

1.3 Reproductive biology and seed dispersal

1.3.1 Flowering

Like any other angiosperm family, the Dipterocarpaceae flower annually or in certain years but never throughout the year. An interesting phenomenon of Dipterocarp phenology occurs in the Malesian rainforest where most Dipterocarpaceae are found. Ashton (1982) recognised two seasonal zones within Malesia, seasonal and aseasonal forest zones. The seasonal forests are characterised by a short but regular dry season with the mean annual rainfall exceeding 2000 mm. This region includes NW Malaya, SE Peninsular Thailand (Pattani), NW Sumatra (Aceh), S Sumatra (Lampung), NE and SE Borneo (Kudat district) (Ashton, 1982). The aseasonal forest zone includes all the other regions in Malesia. It is characterised by a similar annual

rainfall as in the seasonal forest, but with no regular dry season (Ashton, 1982).

Within the seasonal zones, Dipterocarpaceae species usually flower annually (Non General Flowering Periods, NGFP, Sakai *et al.*, 1997) but in the aseasonal region they flower at irregular times, every 2–10 years (Ashton *et al.*, 1988; Appanah and Chan, 1981). This is termed mass flowering (General Flowering Periods, GFP, Sakai *et al.*, 1997). In mass flowering, a large number of species flower within a short period spanning three to four months. It has been suggested that 80% of forest canopy families take part in this event (Ashton, 1982; Ashton *et al.*, 1988; Appanah, 1993). Mass flowering may be triggered by a long drought season in which the assimilate accumulates instead of being used for growth. This accumulated assimilate can then be used to produce a mass of flowers when the "end of drought" trigger is recognised (Ashton *et al.*, 1988; Appanah, 1990).

During the flowering periods, the most important phenomenon to be noted is the pollination syndrome[2]. Since Dipterocarp species are characterised by cup-shaped rotate flowers with the petals enclosing the fertile parts, the pollinators are usually small insects whose morphology is co-adapted to such floral morphology. During the non-GFP, the common pollinators are small social bees: *Trigona* (stingless bees), *Braunsapis* (primitive eusocial bees) or diverse insects of the Orders Coleoptera, Diptera and Hymenoptera. Pollinator behaviour in responding to the GFP is generally related to the mass production of flowers. During this short flowering period, pollinators such as beetles (Momose *et al.*, 1997) and thrips (Chan, 1981) respond quickly, increasing their population using the massive floral resources (Ashton, 1982; Ashton *et al.*, 1988). This strategy is also thought to avoid competition among pollinator species (Janzen, 1974) or contamination of pollen among the species sharing the same pollinator (Appanah and Chan, 1981). Therefore, this flowering strategy acts to maintain the low level of inbreeding depression in Dipterocarpaceae in their natural populations.

---

[2] Relationship among pollination systems and multiple floral characters (Sakai *et al.*, 1997)

1.3.2 Seed dispersal

Dipterocarpaceae are short-distance dispersed trees, with dispersal restricted mainly by their fruit morphology and intolerance to salinity (Dayanandan *et al.*, 1999). Suzuki and Ashton (1996) suggested three dispersal systems in Dipterocarpaceae: wind-gyration, non-wind gyration and water dispersal. Most Dipterocarpaceae have prolonged and twisted fruit sepal lobes, or wings. These function as propeller blades, causing the seed to gyrate in the wind during dispersal (Ashton, 1982; Suzuki and Ashton, 1996). Such wind-gyrating species usually occur as emergents in the forest, while the species in the understorey usually have short fruit sepals and larger seed size and thus do not utilise wind gyration (Suzuki and Ashton, 1996). The dispersal of both wind gyrated and non-wind gyrated species is no more than 100 m from the parent trees (Ridley, 1930 in Suzuki and Ashton, 1996). The water-dispersed species are usually emergent species with non-gyrating seeds that fall into the current. The seed dispersal distances of these riparian species are much greater than those of wind gyrating and non-wind gyrating species.

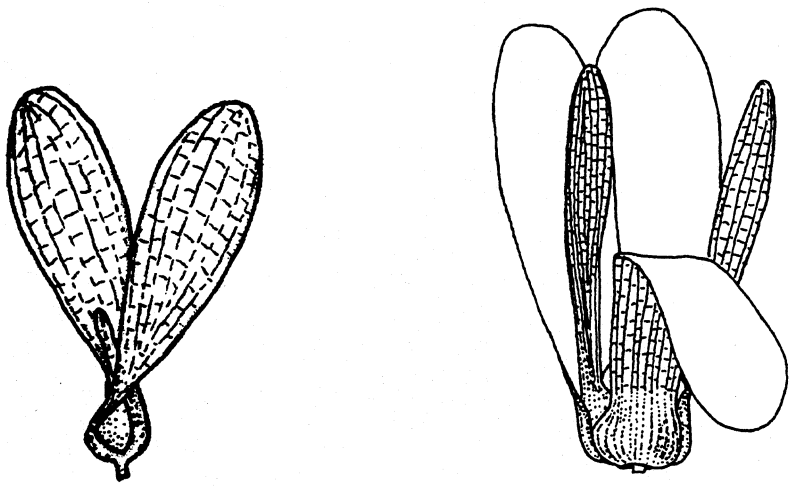
Dipterocarp seeds exhibit no dormancy period, germinating within a week after release (Curran and Leighton, 2000). Certain animals such as the bearded pig, parrakeet, and spiny rat are the main predators of Dipterocarp seeds, but it is unlikely they have any role in seed dispersal either while foraging or through digestion (Curran and Leighton, 2000; Curran and Webb, 2000).

The dispersal of waterborne seed is probably confined to relatively short distances due to their rapid germination but also the effects of immersion, for the family as a whole consists of salt intolerant species (Dayanandan *et al.*, 1999).

1.4 The genera *Hopea* and *Shorea*

The complexity of the Dipterocarps, evident in the distribution, ecology and reproductive behaviour may have resulted from adaptive radiation within each particular geographic region. Speciation may have occurred, resulting in a large number of diverse species centred in the island of Borneo. Of the 550 species of Dipterocarpaceae, ~500 are from sub-family Dipterocarpoideae. Within this sub-family, *Hopea* consists of 102 species and *Shorea* of 194 species.

Hopea and *Shorea* have numerous morphological characters in common. These include being large trees with buttresses, scalariform leaf nervation, paniculate inflorescences and 5-merous stamens (Ashton, 1982). The most conspicuous morphological character that separates the two genera is the comparative development of the fruit calyx lobes, which is also referred to as the "number of aliform fruit wings". *Hopea* is characterised by two long and three short wings (Figure 1.2a), and *Shorea* by three long and two short wings (Figure 1.2b). This difference was used by Ashton (1982) to delineate the two genera.



a **b**

Figure 1.2. Fruit wings distinguishing (a) *Hopea* and (b) *Shorea*

Nevertheless, as is often the case with large groups of taxa that exhibit continuous variation, several classification systems may be used to differentiate the groups. For instance, a classification of *Hopea* and *Shorea* has been developed using field characters. Foresters often find field characters (usually wood anatomy) useful to differentiate the groups. For example, the "Keruing" group refers to the genus *Dipterocarpus*, the "Kapur" group to *Dryobalanops* and the "Meranti" group includes both *Shorea* (divided into several sub-groupings) and some *Hopea* species. For example, the "Balau" subgroup within *Shorea* refers mostly to *Shorea* section *Anthoshorea* but often includes *Hopea dryobalanoides* and *H. mengarawan*. Conversely, *Hopea* is given the name "Bengkirai" although sometimes *Shorea laevis* is included. Since species of *Hopea* and *Shorea* are often placed in the same group under this field classification system, it is not congruent with the existing taxonomic

classification. However, both systems serve to illustrate the complexity of the family and each has its own merits.

As might be expected, when taxonomists attempted to define distinguishing characters for *Hopea* and *Shorea* many characters were found to occur in both genera, including subequal/equal development of fruit calyces, scalariform leaf nervation and several floral characters. Unfortunately, until now no evidence other than morphology has been available to clarify the limits of each genus or the relationship between them. Several more recent studies have investigated Dipterocarp phylogeny based upon molecular markers of the chloroplast genome, including the use of RFLPs (Tsumura *et al.*, 1996) and DNA sequences (Kajita *et al.*, 1998; Dayanandan *et al.*, 1999). Most of the results of these studies are in concordance with the classification of Dipterocarpaceae by Ashton (1982) and indicate a close relationship between *Hopea* and *Shorea*. These studies have also suggested that *Hopea* is monophyletic and *Shorea* is paraphyletic (Tsumura *et al.*, 1996; Kajita *et al.*, 1998; Dayanandan *et al.*, 1999). However, none of these studies contain samples from the whole geographic distribution of *Hopea* and *Shorea*, and do not provide detailed evidence on the relationship between the two.

The complexity and continuous variation of characters occurring within *Hopea* and *Shorea* raise a question as to whether the genera form two "natural" groups". More specifically, how do the evolutionary processes explain their close relationship? In addition, can evolutionary inferences help to clarify their taxonomic status? These questions prompted further investigation and reconstruction of a putative phylogeny, using both molecular markers (the *trnL-F* region of chloroplast DNA and ITS regions of nuclear DNA) and morphological data. The results of these studies are presented in the following chapters.

CHAPTER 2

TAXONOMIC HISTORY OF *HOPEA* AND *SHOREA*

Outline

2.1 Introduction

Taxonomic work on Dipterocarpaceae at the family level started in 1825 when Blume described the group as a family for the first time. Prior to this, other taxonomists had placed genera or species now considered to be part of Dipterocarpaceae into various other families. Descriptive work has continued since then with new species being discovered across the extensive distribution of the family.

Several classifications have been published in previous taxonomic works on Dipterocarpaceae. The differences in these classifications arise largely from the varying taxonomic interpretations of their authors. Defining the limits of *Hopea* and *Shorea* has been a long-standing problem in Dipterocarpaceae systematics and they are still in need of careful examination, since "taxonomic disagreement" regarding their circumscription remains. Nonetheless, several classifications have placed *Hopea* and *Shorea* as two closely-related genera (Miquel, 1820; Heim, 1891; Burck, 1887; Pierre, 1892; Symington, 1943; Hasskarl, 1858; Bentham and Hooker, 1865; de Candolle, 1868; Ashton, 1964a, b, 1968, 1982; Wight and Arnott, 1834; Endlicher, 1840, 1841; Lindley, 1853; Brandis, 1895; Whitmore, 1962; Meijer and Wood, 1964, 1976; Maury, 1978 and Maury-Lechon, 1978 in Maury-Lechon and Curtet, 1998). This chapter aims to review the classification systems previously used for these genera.

2.2 History of the classification of the Dipterocarpaceae

The family Dipterocarpaceae was erected by Blume in 1825 as "Dipterocarpeae" based mostly upon reproductive characters, including the campanulate calyx,

contorted corolla, numerous stamens, erect, elongate, bilocular anthers, 6-locular ovary, single and simple style, and single-seeded drupaceous fruit (Blume, 1825). Vegetative characters such as plant habit and leaf petiole were also described (Blume, 1825). In this work, Blume placed Dipterocarpeae in Guttiferae and recognised only the genus *Dipterocarpus*, containing four species: *D. trinervis, D. retusus, D. littoralis* and *D. gracilis*. He did not acknowledge genera previously described, such as *Vateria* (de Jussieu,1789), *Vatica* (de Candolle, 1868), or *Shorea* and *Hopea* (Sprengel, 1825).

Before Blume (1825) circumscribed taxonomic rank for Dipterocarpaceae, some Dipterocarp taxa had been identified and placed in other families (Linnaeus, 1771; de Jussieu, 1879; de Candolle, 1868; Sprengel, 1825). Linnaeus (1771) and Sprengel (1825) included *Vatica* (*V. chinensis*) in Dodecandria Monogynia based upon leaf, inflorescence and fruit characters and *Vateria* (*V. indica*) into Polyandria Monogynia based on leaf and inflorescence characters. In addition, genera *Shorea* Roxb. and *Dipterocarpus* Gärtn. were included within Polyandria Monogynia by Sprengel (1825) on the basis of similarities of staminal characters. The same author (Sprengel, 1825) included *Hopea* Roxb. (*H. odorata*) in Dodecandria Monogynia based on leaf and inflorescence characters. Moreover, de Jussieu (1789) placed *Vateria* and *Vatica* within *Aurantiis affinia* and *Hopea* Gard. L. into Guaiacanae.

After the work by Blume (1825) had been published, other taxonomists attempted to provide an ordinal rank for Dipterocarpaceae (Lindley, 1853 in Ashton, 1982; Meissner, 1837; Endlicher, 1840, 1841; de Candolle, 1868). Through his work, Blume (1825) noticed a similarity between Dipterocarpaceae and Tiliaceae in the contorted corolla, while affinities to Guttiferae were noted in the presence of resin canals, superior ovary, many stamens and single exalbuminous seed. Lindley (1853) considered "Dipteraceae", Ternströmiaceae, Rhizobolaceae, Clusiaceae, Marcgraviceae, Hypericaceae, Reaumuriaceae to have similarities due to

> "Hypogynous Exogens, with monodichlamydeous flowers, axilae placentae, an imbricate calyx, an imbricate corolla, ∞ stamens, and an embryo with little or no albumin"

Hence he classified these families in the order Guttiferales. In addition, Endlicher (1840) placed Dipterocarpeae in class Guttiferales (Guttiferae *sensu* Endlicher, 1841) and recognised three genera: *Dipterocarpus* Gärtn., *Dryobalanops* Gärtn., and *Vateria* L.

Different classifications of Dipterocarpaceae not only complicate assignment of the ordinal status of Dipterocarpaceae, as shown above, but also give rise to some differences in designation of lower taxonomic ranks for Dipterocarp taxa. At this level, taxonomists have different viewpoints when assigning taxonomic status to certain taxa and often place distantly related species or genera in the same taxonomic rank. This can be seen in earlier works such as those by Endlicher (1840) and Lindley (1853), who united *Shorea* Roxb. and *Vatica* L. Sprengel (1825) included *Shorea* Roxb. and *Dipterocarpus* Gärtn. within Polyandria Monogynia, while he included *Vatica* and *Hopea* Roxb. within Dodecandria Monogynia following Linneaus' work of 1771. Burck (1887), who studied the Indian Dipterocarps, included *Pentacme* DC. and *Monoporandra* Thw. in the genus *Vateria*, and included genera *Pachynocarpus* and *Sunaptera* in genus *Vatica*, and transferred *Petalandra micrantha* Hassk. to *Doona*.

That such different classifications arise may be due not only to different interpretations, but also to the use of a limited range of available herbarium specimens as well as to the great differences in vegetative characters between immature and mature stages of Dipterocarpaceae species. These two factors may have resulted in the use of only a few characters when describing Dipterocarp taxa. Before the 20th century, this was often the case in the taxonomic works on Dipterocarpaceae. Heim (1891) published the most detailed study on the Dipterocarps (Ashton, 1982), but made a mistake when describing *Cotylelobiopsis* Heim from a single sterile herbarium specimen that appears to represent fallen leaflets of *Pseudosindora palustris* Sym. of Leguminosae (Ashton, 1982).

Later taxonomic works on Dipterocarpaceae include comprehensive *floras* for South East Asia and Malesia (Burck, 1887; Symington, 1943; Ashton, 1982), for Indian species (Dyer, 1894), and for Indochinese species (Pierre, 1892). In addition, Burck (1887), Heim (1891, 1892), and Gilg (1960) produced excellent monographs. Most

of these taxonomic works employ macro-morphological characters (Gilg, 1960; Pierre, 1892; Dyer, 1894 ; Brandis, 1895; Symington, 1943; Ashton, 1982) and a few use anatomical features (Burck, 1887; Heim, 1891).

Among the taxonomic works described above, that of Ashton in *Flora Malesiana* (1982) and earlier his treatment of Sri Lankan species (Ashton, 1980) give the most comprehensive account of Dipterocarpaceae to date. Thus, this study follows the Dipterocarp classification of Ashton (1980, 1982). Ashton (1982) recognised diagnostic characters in some Dipterocarp taxa which enabled him to assign new taxonomic status for certain taxa. This is obvious when he delimits the terminology of infra-specific level such as "sub-species" and "variety", though there is no common agreement among taxonomists about the use of these categories. Since each angiosperm family may possess unique forms or variations it is difficult to generalise. This "character delimitation" is useful for developing a determination / identification key since some Dipterocarp genera or species are known to have highly continuous characters that may present difficulties when one attempts to delimit them. With regard to the higher taxonomic level, Ashton (1982) recognised the infra-generic variation in the complex genus *Shorea* based on floral characters. For instance, he reduced the genus *Pentacme* described earlier by de Candolle (1868) into genus *Shorea* section *Pentacme* and included genus *Doona* Thw. within genus *Shorea* as section *Doona*.

Furthermore, Ashton (1982) attempted to create a less artificial classification by employing characters other than morphological ones. He employs chromosome numbers (Jong, 1969 in Ashton, 1982; Somego, 1978) in his tribal system for the sub-family Dipterocarpoideae as well as considering previous wood anatomy classifications (Symington, 1943, Whitmore, 1962). Ashton's work on the Asian sub family Dipterocarpoideae recognised two tribes, *Dipterocarpeae* and *Shoreae*, following karyological evidence. He suggests that the tribe *Dipterocarpeae* has basic chromosome numbers of x =11, and the tribe *Shoreae* x = 7. At the lower taxonomic level, he seems to recognise natural groups such as *Shorea* section *Pachycarpae* and Section *Doona*. The former section is endemic to the island of Borneo and the latter is endemic in Sri Lanka.

2.3 The classification of *Hopea* and *Shorea*

2.3.1 The classification of *Hopea*

Hopea was first described by Roxburgh (1811) as a *nomen nudum*, with a formal description not published until 1819. The generic name was later conserved by Ashton (1982). Roxburgh recognised one species, *Hopea odorata*, which he described as a tall tree with dark brown flaky bark, a paniculate inflorescence, ovoid leaf buds, and a densely puberulent calyx and corolla.

Hopea Roxb., *Pl. Corom.* 3: 7 (1811)

Generic account (after Ashton, 1980, 1982)

Trees small to medium with tapering boles, frequently branching low with thin or sometimes thick buttresses, with flying buttresses and stilt roots sometimes present. ***Canopy*** lanceolate and monopodial; with more or less horizontal pendent branching in small trees, becoming hemispherical with straight branchlets in large trees. ***Bark*** variable for most species and dependent mostly on growth stage; smooth, chocolate and grey mottled, hoop-marked in early growth, remaining or becoming cracked and flaked or fissured. ***Twigs*** slender. ***Tertiary nervation*** scalariform, reticulate, reticulate with intramarginal vein, dryobalanoid, or sub-dryobalanoid. ***Domatia*** present or absent. ***Flower buds*** usually small, ovoid and rarely globose. ***Calyx lobes*** imbricate; 2 outer lobes ovate, obtuse or suborbicular; 3 inner lobes mucronate. ***Petals*** connate at base, falling in a rosette. ***Stamens*** 10, 15 or up to 20. ***Filaments*** variable, broad and compressed at base, tapering medially, filiform below anthers. ***Connective appendage*** slender, twice as long as anthers, glabrous or minutely glandular-tuberculate. ***Ovary*** glabrous or tomentose, ovoid, with or without stylopodium. ***Stigma*** usually minute. ***Fruit*** with 2 outer calyx lobes longer than 3 inner ones, spatulate, thin, or with all lobes equal/sub-equal, thickened, saccate at base. ***Nut*** ovoid, glabrous.

2.3.2 The classification of *Shorea*

The name *Shorea* was attributed to Roxburgh by Gaertner (1805), based on an Indian specimen that Gaertner described as *Shorea robusta*. Gaertner recognised only this one species, described as a medium-sized tree with large buttresses and without stilt roots. *S. robusta* was also described as having alternate leaves with scalariform venation, paniculate inflorescences, and three long and two short wings on the fruit.

Shorea Roxb. ex Gaertn.f., *Fruct. Suppl. Corp.* 3: 47 (1805)

Generic account (after Ashton, 1980, 1982)

Trees medium to large with large straight bole and buttresses, stilt roots absent, frequently branching low with thin or sometimes thick buttresses. ***Canopy*** large, hemispherical, dome shaped and sympodial; more or less horizontal pendent branches in small trees, and becoming hemispherical in large trees, in addition to having straight branchlets. ***Bark*** variable for most species and dependent mostly on growth stage; smooth, dimpled, fissured, scaly and laminated bark. ***Twigs*** slender. ***Tertiary nervation*** scalariform or reticulate. ***Domatia*** present or absent. ***Flower buds*** usually small, ovoid or rarely globose. ***Calyx lobes*** free to receptacle; with 3 thick outer and 2 narrow inner lobes. ***Petals*** connate at base on falling, sometimes free. ***Stamens*** 10 to ∞. ***Filaments*** variable, lorate to filiform. ***Connective appendage*** vestigial or prominent. ***Ovary*** tomentose or rarely glabrous, ovoid, with or without stylopodium. ***Stigma*** minute. ***Fruit*** with 3 outer calyx lobes longer than the two inner ones, thin, spatulate, or with all lobes sub-equal, base of lobes more or less thickened, expanded, saccate. ***Nut*** ovoid, free from calyx.

2.4 Classification systems of *Hopea* and *Shorea*

There are six major classification systems that have been followed by taxonomists working on the Dipterocarpaceae (Table 2.1). These are the systems established by Ashton (1982), Meijer and Wood (1964, 1976), Maury (1978 in Maury-Lechon and Curtet, 1998) and Maury-Lechon (1979 in Maury-Lechon and Curtet, 1998), Kostermans (1978, 1983, 1984), Heim (1892) and Symington (1943). The differences between these classifications are largely due to varying interpretations among the taxonomists. Four of the classifications—those by Ashton (1982), Maury (1978 in Maury-Lechon and Curtet, 1998), Meijer and Wood (1964) and Kostermans (1983, 1984)—largely followed the previous classifications of Symington (1943) and Heim (1892). Meijer and Wood (1964) based their work on Dipterocarpaceae of Sabah, while Kostermans (1984) concentrated on the Sri Lankan taxa. Maury (1978 in Maury-Lechon and Curtet, 1998) and Ashton (1982) worked on the entire Dipterocarpaceae using different approaches. Ashton (1982) used a taxonomic approach while Maury-Lechon (1978, 1979 in Maury-Lechon and Curtet, 1998)

Table 2.1 Comparison of major classification systems of *Hopea* and *Shorea*, and their putative members (after Maury-Lechon & Curtet, 1998)

Maury (1978)	Kostermans (1981-85)	Heim (1892)	Meijer and Wood (1964 and 1976)	Ashton (1982)	Symington (1943)
Hopea	**Hopea**	**Hopea**	**Hopea**	**Hopea**	**Hopea**
sect. Hopea		sect. Euhopea		sect. Hopea	
subsect. Hopea		sect. Petalandra		subsect. Hopea	sect. Euhopea
subsect. Pierrea		sect. Hancea		subsect. Pierrea	sect. Pierrea
sect. Dryobalanoides		sect. Dryobalanoides		sect. Dryobalanoides	sect. Dryobalanoides
subsect. Dryobalanoides		**Pierrea**		subsect. Dryobalanoides	sect. Bracteata
subsect. Sphaerocarpae		**Duvaliella**		subsect. Sphaerocarpae	
		Parahopea			
Balanocarpus	**Neobalanocarpus**	**Balanocarpus**		**Neobalanocarpus**	**Balanocarpus**
	Balanocarpus				
Pentacme		**Pentacme**	**Pentacme**	**Shorea**	**Pentacme**
Doona	**Doona**	**Doona**	**Doona**	sect. Pentacme	**Shorea**
Anthoshorea				sect. Doona	
sect. Bracteolatae				sect. Anthoshorea	timbgrp. Meranti Pa'ang
sect. Anthoshorea					
		Richetia			
Shorea	**Shorea**	**Shorea**	**Shorea**		
	incl. Pentacme				
sect. Shoreae		sect. Eushorea	subgen. Shorea	sect. Shorea	timbgrp. Balau
sect. Barbatae		sect. Anthoshorea	subgen. Anthoshorea	subsect. Shoreae	
		sect. Hopeioides		subsect. Barbatae	
				sect. Neohopea	

Table 2.1 (continued)

Maury (1978)	Kostermans (1981-85)	Heim (1892)	Meijer and Wood (1964 and 1976)	Ashton (1982)	Symington (1943)
Richetia		sect. Richetioides	subgen. Richetia	sect. Richetioides	timbgrp. Meranti damar hitam
sect. Richetioides				subsect. Richetioides	
sect. Maximae				subsect. Polyandrae	
Rubroshorea		sect. Rugosae	subgen. Rubroshorea		timbgrp. Meranti merah
sect. Muticae			subgrp. Parvifolia	sect. Mutica	subgen. Parvifolia
subsect. Muticae				subsect. Muticae	
subsect. Auriculatae				subsect. Auriculatae	
sect. Ovalis			subgrp. Ovalis	sect. Ovalis	subgen. Ovalis
sect. Rubellae				sect. Rubella	
sect. Brachypterae		sect. Brachypterae		sect. Brachypterae	subgen. Pauciflora
subsect. Brachypterae		subsect. Brachypterae	subgrp. Pauciflora	subsect. Brachypterae	
subsect. Smithianae		subsect. Smithianae	subgrp. Smithiana	subsect. Smithianae	
sect. Pachycarpae		sect. Pachycarpae	subgrp. Pinanga	sect. Pachycarpae	
Parashorea	**Parashorea**	**Parashorea**		**Parashorea**	**Parashorea**

examined the natural groups and provided some descriptive phylogenetic inferences.

As shown by the comparison of classification systems outlined in Table 2.1, only Ashton (1982) and Meijer and Wood (1964) recognised *Hopea* and *Shorea* as two broadly-defined genera. The four other treatments recognised more than two genera within *Hopea* and *Shorea sens. lat.* All of the authors excluding Kostermans also recognised infra-generic groupings within the genera. Maury and Heim recognised seven and eight genera respectively. In addition to *Hopea* and *Shorea*, both recognised *Pentacme* and *Doona* as separate genera. Maury recognised *Anthoshorea, Richetia* and *Rubroshorea* as additional genera related to *Shorea*, while Heim included these genera within *Shorea* and gave them sectional rank. Heim also recognised *Pierrea, Duvallelia* and *Parahopea* as additional genera related to *Hopea*. Both Kostermans and Symington recognised three genera respectively. The additional genus recognised by both authors was *Doona* by Kostermans and *Pentacme* by Symington.

The differences of the classification systems of *Hopea* and *Shorea* are mainly due to the use of different key characters to distinguish the genera or infra-generic groupings. This is a result of the remarkable similarities and highly continuous morphological variation at generic, infra-generic and specific levels. Some characters which overlap the generic boundaries of *Hopea* and *Shorea sens. strict.* have led to recognition of intermediate "forms" or taxa. In turn, the question arises as to whether the group should be regarded as a single genus or separated into two or more genera.

Parahopea appears to be an intermediate form between *Shorea* and *Hopea*. Heim (1891) compared *Parahopea* to *Shorea* and *Hopea* and found that sepals, receptacle, stamens and petiole anatomy are similar to *Shorea,* while the petals resemble *Hopea*. He considered that *Parahopea* was closer to *Hopea* than *Shorea*. However, Ashton (1982) suggested *Parahopea* shows similarities to *Shorea* on the basis of their fruit features. Brandis (1895) had already recognised these similarities and placed *Parahopea* into *Shorea* section *Anthoshorea* Heim on the basis of the same characters. However, Maury (1978) considered *Anthoshorea* as a genus separate from *Shorea* and comprising two sections—*Anthoshoreae* and *Bracteolatae*.

Heim (1891) showed that his newly described genus *Richetia*, has some similarities in anatomical characters to *Shorea* section *Richetioides*. After examining embryo characters, petiole anatomy and its sub-equal fruit sepals, he considered *Richetia* merited generic status. Meijer and Wood (1964) and Maury (1978 in Maury-Lechon and Curtet, 1998) followed this classification system. Maury (1978 in Maury-Lechon and Curtet, 1998) further divided *Richetia* into two sections, *Richetioides* and *Maximae*. On the other hand, Symington (1943) did not agree with Heim's segregation of *Richetia*, and therefore included *Richetia* in *Shorea*. This was followed by Ashton (1982) who placed *Richetia* as a section within *Shorea* (section *Richetioides*).

Isoptera was described by Burck (1887). It differed from *Shorea* in having a canal medularis in the petiole. Heim also recognised *Isoptera* as a separate genus, but Ashton (1982) placed *Isoptera* as section *Neohopea* within *Shorea*.

Doona, a genus endemic to Sri Lanka, was described by Thwaites (1835). He acknowledged that *Doona* shared similarities with *Shorea* in having 15 stamens, a contorted corolla, and in wood anatomy features. However, *Doona* has a distinct club-shaped connective staminal appendage that distinguishes it from *Shorea*. Heim (1892) and Burck (1887) combined certain species of *Hopea* and *Doona* since they lacked this distinct feature (Ashton, 1982). However, Ashton (1982) considered that the difference in connective appendage was not sufficient to place *Doona* as a separate genus. Thus, he placed *Doona* into *Shorea* at sectional rank. By contrast, Maury (1978), Meijer and Wood (1964) and Kostermans (1984) still recognised it as a genus separate from *Shorea*.

Beddome (1874) described a new genus *Balanocarpus* and its infra-generic divisions were provided by Heim (1891). Some taxonomists suggested that some sections of *Balanocarpus* resembled *Shorea*, and others included some of the sections in *Hopea*. Symington (1943) and Maury (1978 in Maury-Lechon and Curtet, 1998) maintained *Balanocarpus heimii* as a separate genus but indicated that it was closely related to *Hopea*. Subsequently, Ashton (1982) stated that *Balanocarpus* could not be placed in either *Hopea* or *Shorea*, as it has short equal fruit sepals and a unique androecium structure. Thus, he reinstated *Balanocarpus* and provided a new genus name,

Neobalanocarpus, containing only one species, *N. heimii*.

Genus *Parashorea* was considered by Maury (1978 in Maury-Lechon and Curtet, 1998) to belong within tribe *Shoreae* sub-tribe *Parashorinae*. She thus indicated a close relationship between *Parashorea* and *Shorea*. Nonetheless, other authors mentioned in Table 2.1 considered *Parashorea* to be a separate genus.

2.5 Problems in generic circumscription of *Hopea* and *Shorea*: a biosystematic review

Hopea and *Shorea* show remarkable similarities and highly continuous morphological variation at both generic and specific level and have long been regarded as closely related genera (Ashton, 1982). Most species of *Hopea* and *Shorea* are large timber trees but *Hopea* rarely becomes as large as *Shorea*. Stilt roots are present in *Hopea*, but not in *Shorea*. The presence of dryobalanoid and sub-dryobalanoid leaf nervation in *Hopea* alone is one of the few discrete characters that can separate the genera but the most distinctive character is the comparative development of fruit calyx.

Recent studies of phylogenetic relationships in Asian Dipterocarpaceae have been carried out by employing some molecular markers such as RFLP[1] (Tsumura *et al.*, 1996), RAPDs[2] (Rath, *et al.*, 1998) and DNA sequences of three regions in the chloroplast genome (Kajita *et al.*, 1998). Their results are quite consistent with the classical taxonomic work of Ashton (1982). However, they show that *Hopea* is a monophyletic group nesting within *Shorea* (Rath *et al.*, 1998) and *Shorea* thus forms a paraphyletic group (Tsumura *et al.*, 1996; Kajita *et al.*, 1998).

The independent approaches described above show that *Hopea* and *Shorea* cannot be readily separated. This may be due to some overlap of characters between the genera. The most obvious examples probably come from morphological studies. Even though the differentiation of the fruiting calyx is considered to be the only distinguishing character, this feature is not entirely reliable since many species of both genera have five sub-equal (vestigial) calyx lobes in fruit. Moreover, karyological studies (Jong, 1969 in Ashton, 1982 and Somego, 1978) and molecular research (Tsumura *et al.*,

[1] Restriction Fragment Length Polymorphism

1996; Kajita *et al.*, 1998; Rath *et al.*, 1998) also show some overlap of characters among the two.

The occurrence of overlapping characters between *Hopea* and *Shorea* has led to recognition of intermediate "forms" and to uncertainty about how many taxa should be recognised and at what rank. This uncertainty played an important role in the classification of *Hopea* and *Shorea sens. lat.* before Ashton's treatment (1982). The differing taxonomic concepts used by dipterocarp systematists has resulted in two divergent approaches: (1) a narrow generic concept, with assignment of generic status to many intermediate taxa such as *Doona, Parahopea, Richetia* and *Balanocarpus*, or (2) a broad generic concept, with few genera being recognised and the intermediate taxa being given only infra-generic rank, for example by Ashton (1982).

In conclusion, there are still areas requiring resolution in the classification of *Hopea* and *Shorea*, with issues of their infra-generic status and even their generic status. The assessment of the generic and infra-generic boundaries is not only important in determining the circumscription of the genera but also essential in drawing any evolutionary inferences.

In order to address these issues, this study will construct a morphological and molecular based phylogeny of *Hopea* and *Shorea* with the aim of determining the evolutionary relationships within these complex genera. The morphological study is the subject of the next chapter, which discusses the importance of using morphological characters in phylogenetic reconstructions and examines the characters that are useful in distinguishing between the two genera. Chapter 4 discusses the molecular analysis, while Chapter 5 will present the combined data sets from both the molecular and morphological studies. The thesis then concludes with a general discussion in Chapter 6.

[2] Random Amplified Polymorphic DNA

CHAPTER 3
MORPHOLOGICAL STUDY

Outline

3.1 Introduction

Hopea and *Shorea* are very similar in morphology, anatomy, and reproductive biology, as well as geographic distribution (as discussed in chapter 1). This high degree of similarity raises problems in identifying suitable characters to define the two genera. The previous chapter presented the difficulties associated with previous generic and infra-generic classifications of *Hopea* and *Shorea*. This chapter will describe and discuss the morphological work undertaken, which provides some of the evidence needed to reconstruct the phylogenetic relationships between *Hopea* and *Shorea*.

Classifications of Dipterocarpaceae place *Hopea* and *Shorea* as closely-related genera on the basis of only a few characters (Gaertner, 1805; Roxburgh, 1811 in Ashton, 1982; Ashton, 1982). The recent classification by Ashton (1982) suggested that the comparative development of the fruit calyx is the only morphological character that reliably distinguishes the two genera. In his classification system Ashton (1982) united some species that had previously been recognised as separate genera (Miquel, 1820; Burck, 1887; Heim, 1891, 1892; Brandis, 1895) and gave them infra-generic rank on the basis of a variety of often simple characters. The different taxonomic arrangements and ranks that have been proposed between and within *Hopea* and *Shorea* raise questions about their generic circumscription, and whether they form a natural group. This study will therefore test the monophyly of the genera and examine the existing infra-generic classifications.

The development of a cladistic approach to infer the evolutionary relationships of taxa reflects a shift towards a more explicit methodology in systematics (Freudenstein and Rasmussen, 1999). Several studies on Dipterocarpaceae have been carried out using a cladistic approach and molecular data (Tsumura *et al*, 1996; Kajita *et al.*, 1998; Dayanandan *et al.*, 1999). Since all these studies examined the relationships within the entire family, the relationship between *Hopea* and *Shorea* was not their main focus, though the two genera were shown to have a close evolutionary relationship.

This present morphological study aims to:

1. Reconstruct the phylogenetic relationships between the two genera using a cladistic analysis of morphological characters.
2. Examine the nature of the characters used in these cladistic analyses.
3. Determine shared and distinct characters for the two genera.
4. Investigate the taxonomic relationships between *Hopea* and *Shorea*.

Data were gathered to provide a more comprehensive assessment of the generic and infra-generic circumscription of *Hopea* and *Shorea* than hitherto. This chapter is divided into ten sections, with the first five sections (3.1–3.5) describing the theoretical background of the morphological study, the characters used to delimit the genera and a discussion of their selection. The next two sections (3.6–3.7) describe

the selection of taxa and the methods of data analysis. A further two sections (3.8–3.9) present the results and a discussion of these, with a final section (3.10) providing some conclusions about the morphological study.

3.2 Theoretical background of phylogenetic analysis using morphological data

Cladistic analysis is the most common method currently used to reconstruct phylogenetic trees. It involves two basic phases—exploration of characters (including selection and examination), followed by analysis of the data to generate a set of trees (Thiele, 1993). The type of data may affect the choice of analysis method. Conversely, the choice of analysis method may set bounds for the type of data that are able to be analysed, and for the format in which they are recorded. Therefore, the selection and resolution of characters and taxa is critical for the interaction between these two phases.

However, much of cladistic methodology (Wiley, 1981) is thought by Thiele (1993) to put too much emphasis on the analysis phase. Thiele's concern is understandable, since evolutionary analysis of morphological data almost always involves continuous characters. Certain authors have considered these to be "non-cladistic" characters (Crisp and Weston, 1987) because they are not discrete. This problem has caused Thiele and others (e.g. Archie, 1985) to pay more attention to means of analysing quantitative data. However, the inclusion of quantitative characters in cladistic studies is still debated, as shown below.

There are two types of characters used in morphological systematic studies: qualitative and quantitative. Qualitative characters are mostly obtained by observation without the need for measurement. An example of a qualitative character is the presence or absence of stipule scars (character 1, Table 3.1). By contrast, quantitative characters are obtained by measurement. These can be further divided into continuous and discrete (meristic) quantitative characters. When assessing continuous quantitative characters, each individual measurement is not necessarily an integer and potentially forms a "continuum", as exemplified by leaf length (character 2). For discrete quantitative characters, an individual measurement is an integer, with an example being stamen number. A qualitative character is unambiguous and is

therefore considered to be a "cladistic" character. It can be binary (such as the presence or absence of stipule scars, character 1) or multistate discrete (such as the comparative development of the fruit calyx, character 35).
Cranston and Humphries (1988) and Chappill (1989) showed that morphometric (continuous) data added "noise" rather than a phylogenetic signal, and had lower consistency indices on the trees obtained than did the qualitative data. Crisp and Weston (1987) had previously suggested that quantitative data should be excluded from the data set since they are less useful for cladistic analysis.

However, some authors have provided arguments (Thiele and Ladiges, 1988; Thiele, 1993) or even "theoretical" justifications (Rae, 1998) against the exclusion of continuous characters. Thiele and Ladiges (1988), in their study on *Angophora* (Myrtaceae), produced a single tree with a high consistency index (0.63) even though most of their data set consisted of quantitative characters. They showed that qualitative data alone resulted in a tree with little resolution. Another argument for the inclusion of quantitative characters is that the qualitative characters may be a collection or transformation of quantitative characters. For instance, leaf shape (a qualitative character) can be defined by the ratio of leaf length to leaf width, which are quantitative characters (Thiele and Ladiges, 1988; Thiele, 1993). Furthermore, some authors have shown the utility of continuous characters by proposing mathematical algorithms to code them (Archie, 1985; Thiele, 1993).

3.2.1 Method for coding quantitative characters

There are several methods to code quantitative characters, all of which basically rank the taxa along the scaled attribute axis and then divide the attribute axis into states (Mikevich and Johnson, 1976; Colless, 1980; Thorpe, 1984; Archie, 1985; Baum, 1988; Chappill, 1989; Thiele, 1993). The difference among the methods lies in the degree to which they divide the attribute axis. Simple gap coding (Mikevich and Johnson, 1976) divides the attribute axis at points where no values occur (gaps) or where the gaps between the means exceed a predefined value. Generalised gap coding (Archie, 1985) and segment gap coding (Colless, 1980; Thorpe, 1984; Chappill, 1989) divide the axis evenly so that such gaps do not occur. Weighted gap coding (Thiele, 1993) uses differential weighting of gaps between coded states within one character. This method allows some elements of raw quantitative data to be used

in a cladistic analysis. This present study follows the method proposed by Thiele (1993) as outlined below:

1. The raw data are arranged in order according to their mean values.
2. These means are tested for their normal distribution. If the variances are not equal, the data are standardised using log (x+1) transformation.
3. The data are then subjected to range-standardisation using this formula:

$$X_s = \frac{x - \min}{\max - \min} \times 10$$

where X_s is the standardised value

4 The resulting values are coded as the nearest integer to their standardised value.

5 The coded values are treated as ordered multistate characters in a cladistic analysis.

3.3 Character examination of *Hopea* and *Shorea*

In order for this analysis to be based on a broad set of characters, the morphology of *Hopea* and *Shorea* was carefully studied. In an initial pilot study, all characters used by Ashton (1982) were assessed. Then, morphological features that were not described by Ashton (mostly floral parts) were described. Thirdly, all the terminology used by Ashton (1982) was verified and modified for use in the analysis. This system was used to confirm the identity of herbarium specimens available from BO, CANB, FRIM and HUH and found to be workable and repeatable.

Following this verification, six genera of subfamily Dipterocarpoideae were examined in order to recognise useful diagnostic characters. These genera are *Dipterocarpus*, *Dryobalanops*, *Parashorea*, *Neobalanocarpus*, *Hopea* and *Shorea*. Previous taxonomists working on these genera had identified the comparative development of the calyx lobes in the fruit as the single morphological character that distinguished *Hopea* and *Shorea*. Determination of any other useful characters may result in the separation of some infra-generic taxa into their own genera, as has been the case of previous classifications (Miquel, 1820; Burck, 1887; Heim, 1891; Brandis, 1895; Symington, 1943). This study does not intend to suggest or provide taxonomic ranks for well-established sections, but rather to investigate the phylogeny of these groups. This will be achieved by examining as many characters used by previous

authors as possible and including those considered to be of evolutionary significance in a cladistic analysis.

The preliminary investigation of characters indicated that floral and inflorescence morphology provided diagnostically useful variation across the species examined. A decision was thus made to investigate inflorescence structure in greater detail, as discussed in the following section.

3.4 An investigation of the inflorescence structure of *Hopea* and *Shorea*

An initial morphological study of *Hopea* and *Shorea* revealed that there were some potentially informative characters associated with inflorescence structure and growth form. In order to understand the complex inflorescence structure in these genera, some features of inflorescence architecture are introduced below. These may provide characters capable of distinguishing *Hopea* from *Shorea*. The following is an analysis of inflorescence characters primarily as they relate to *Hopea* and *Shorea*.

3.4.1 Inflorescence structure

The structure of the inflorescence is an important and complex feature in plant systematic studies. The inflorescence cannot be described just from observation of its overall form—an insight into its development is also needed. This knowledge is important in understanding the evolution of the inflorescence, which in turn can provide the basis to explain the evolution of a particular taxon. The way in which inflorescence architecture can be utilised to explore the evolution of Dipterocarpaceae is examined below.

Linneaus was the first to introduce the term "inflorescence" and he described several types of inflorescences, including cyme, raceme, spike and umbel (Tucker, 1999). According to Weberling (1989) these are the four basic forms of simple inflorescences. Bentham (1892) defined an inflorescence as "the arrangement of flowering branches and the flowers upon them", but he also applied the term to the flowering branch itself. Weberling (1989) and Troll (1964) defined the term as "the shoot system which serves for the formation of flowers and which is modified accordingly".

The various types of inflorescence have been described by many authors (Eichler, 1878; Rickett, 1944; Troll, 1964; Briggs and Johnson, 1979; Barlow, 1989; Weberling, 1989). These authors introduced a range of terminology to account for differences due to specialisation of the inflorescence in particular taxa. Several authors have found these terminologies utilise well-known published terms that are unsuitable for their taxon of interest. For instance, Briggs and Johnson (1979) could not apply some of the terminology proposed by Troll (1964) in their own studies of Myrtaceae. They therefore proposed new terms and introduced a new concept of inflorescence architecture to aid in their understanding of the evolution of this family. However, Bradford (1998) did not find Briggs and Johnson's terminology could be satisfactorily used for his research on *Weinmannia* (Cunoniaceae). Moreover, Grimes (1999) in his study on Mimosoideae (Fabaceae) could not use the term "seasonal growth unit" (SGU) introduced by Briggs and Johnson (1979). Grimes argued that the term would cause confusion if applied to a tropical species that exhibits no seasonally-related growth and he thus introduced a new term, Repeating Growth Unit (RGU).

Notwithstanding these complications and inconsistencies, there are important aspects of inflorescence structure that can contribute to an understanding of the evolution of a group of taxa. The basic foundation for examining evolutionary relationships among taxa is homology. Homology itself can only be deduced by understanding both the inflorescence architecture (structural form *sensu* Tucker and Grimes, 1999) and the developmental pathways which led to the final structure. Grimes (1999) showed that the developmental pathways of the primordial leaves and inflorescence give rise to the inflorescence architecture of the Mimosoid Tribes Ingeae and Acacieae.

From a developmental point of view, the inflorescence can be seen as a shift from the vegetative to the reproductive phase (Frijters, 1978; Briggs and Johnson, 1979; Weberling, 1989; Bradley *et al.,* 1997; Diggle, 1999; Singer *et al.*, 1999; Tucker and Grimes, 1999). When a plant makes a developmental shift between vegetative and reproductive phases, the vegetative apical meristem may change in a variety of ways that reflect physiological change. An inflorescence apical meristem may arise either from conversion of a vegetative apical meristem or *de novo* from a bud in the axil of a leaf or bract, depending on the taxon in question.

Exploring the evolutionary relationships among inflorescence types can be confusing because most plant families show a shift from one particular type into another. For example, in Fabaceae, the basic inflorescence type is a raceme, but evolutionary shifts have led to a pseudoraceme in five tribes of Papilionidae (Tucker, 1987; Tucker and Grimes, 1999). Gesneriaceae, which is characterised by a terminal flower in each cyme (Weber, 1973 in Tucker and Grimes, 1999), also contains taxa that show shifts to "normal" cymes (Weber, 1978 in Tucker and Grimes, 1999). The parallelism of inflorescence structure led Stebbins (1974) to argue that such shifts are a result of adaptive selection, for instance to avoid competition for pollinators.

These shifts can be examined from an ontological perspective. Evans and Dickinson (1999) produced a comprehensive account of the ontogeny of four genera of the subfamily Spiraeoideae (Rosaceae), which included a description of the developmental stages in each. The four genera investigated have three different types of inflorescence: racemose, paniculate and corymbose. The morphology of the mature inflorescence can be better understood after examination of its development from floral primordia. For instance, in the genus *Spiraea*, the mature inflorescence is classified as indeterminate (or polythelic *sensu* Weberling, 1989) since there is no terminal flower. However, a study of ontological development in *S. sorbifolia* showed that the inflorescence meristem ceases to initiate bracts. Thus, it does not form a terminal flower, but instead leaves an apical residuum on the apex of the inflorescence. Ontological studies coupled with an understanding of the mature inflorescence architecture can thus ensure correct homology assessments are made when comparing inflorescences between taxa.

A need to understand inflorescence development leads to the question of whether specific genes control the form and expression of the inflorescence. The most recent studies that have been conducted to answer this question are those by Bradley *et al.* (1996; 1997) and Singer *et al.* (1999). Bradley *et al.* (1996, 1997) carried out a study on *Arabidopsis* (Brassicaceae) and *Antirrhinum* (Scrophulariaceae) and identified genes controlling the development of indeterminate inflorescences. Singer *et al.* (1999) conducted an experiment on mutated species of *Pisum sativ*um (Fabaceae) and showed how the mutant genes affected the developmental growth and overall

structure of the pea inflorescence. They compared two mutated species of *Pisum sativum* to two species of *Arabidopsis* in order to understand how conserved genes create diverse inflorescence architectural patterns. One of the reasons for choosing peas as the study species was that their inflorescences are one step more complex than the simple raceme of *Arabidopsis*. Comparisons between the two could thus be used to provide insight into the underlying genetic mechanisms that distinguish simple from compound racemes. Such genetic studies are useful in taxa where the patterns of inflorescence development have been comprehensively studied.

3.4.2 A hierarchical system for paniculate inflorescence of *Hopea* and *Shorea*

The classification of inflorescence structures provided by Weberling (1989), Weberling and Troll (1989) and Briggs and Johnson (1979) entails examination of the overall inflorescence display (Total Inflorescence Display, TID). Briggs and Johnson (1979) defined the TID through the presence of a subtending prophyll. Under this system, the total inflorescence display of both *Hopea* and *Shorea* is categorised as paniculate. This study included *Dipterocarpus confertus* and *D. retusus* as the outgroup, having a different total inflorescence display type, known as a botryoid.

Observation of the TID in this group of Dipterocarpaceae is limited when using herbarium materials to define the characters. Dipterocarpaceae inflorescences are generally large and often only parts of them are presented on a herbarium sheet. In view of this size problem, the description of inflorescences in this study is restricted to the smallest units of the total inflorescence that can be used to make consistent inferences about inflorescence structure in Dipterocarpaceae.

The decision to parse the TID into smaller units is particularly important when identifying cladistic characters. Many cladists have avoided including inflorescences in their analyses because it can be difficult to assess homologies. Grimes (1999) in his work on the *Pithecillobium* complex (Fabaceae) and Bradford (1998) in his work on *Weinmannia* (Cunoniaceae) have used inflorescence structure in cladistic analyses by breaking the total structure into nested repeated units. This study follows that concept.

The inflorescence structure of the Dipterocarps in this study can be classified into four nested units (Figure 3.1):

1. The organisation of individual flowers into a botryoid/racemose form, termed the "Basic Unit" (BU).
2. The development of the botryoid (BU) in conjunction with a supporting stem, termed the "Inflorescence Unit" (IU)
3. The position these botryoid-stem units (IU) occupy in relation to the secondary axis, termed the Total Unit (TU).
4. The position of the Total Unit (TU) in relation to the primary axis is termed the Total Inflorescence (TI) and is usually defined by the presence of a leaf-like structure (prophyll).

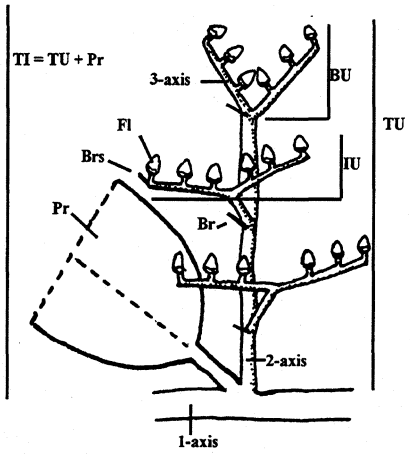


Figure 3.1 A diagram showing the inflorescence structure in Hopea and Shorea. Fl: individual flower, Br: bract, Brs: bracteole, BU: Basic Unit, IU: Inflorescence Unit, TU: Total Unit, Pr: prophyll, 1-axis: primary axis, 2-axis: secondary axis, 3-axis: tertiary axis.

Each of the hierarchical units of the inflorescence is defined by the presence of a "leaf-like" structure (frondose *sensu* Weberling, 1989 or prophyll *sensu* Briggs and Johnson, 1979) on each TI, a bract on each IU (in complex inflorescences) or on each BU (in simple inflorescences, see detailed discussion below), and two opposite bracteoles in each individual flower.

Weberling (1989) described this leaf-like structure as "frondose", while Briggs and Johnson (1979) suggested the use of the term "prophylls" used earlier by Troll (1964) for various bract-like structures that are not always equivalent to bracteoles. Weberling (1989) used the term "prophyll" in reference to paniculate inflorescences. The presence of a prophyll of this structure therefore defines the TI as shown in

Figure 3.1. In the taxa examined for this study, the TI is always subtended by a "leaf-like" rather than a "bracteate" organ, but in some cases this structure is smaller than an actual leaf. Thus, the present study uses the term prophyll in the sense of Briggs and Johnson (1979) as a structure that defines the TI, whereas a bract defines the BU for simple inflorescences or the IU for complex inflorescences. The bracteole then defines an individual flower. The term bract is used for two different units, the BU or the IU, because some forms of paniculate inflorescence do not possess any IUs (Figure 3.3). The BU is the smallest form in the hierarchical system, and in paniculate inflorescences the BU is basically a botryoid. This BU is then extended into IUs, and in essence this is consistent across the genera examined in this study. The BU and the IU to some degree can be used to deduce whether the inflorescence type is botryoid (Figure 3.2) or paniculate (Figure 3.3).

3.4.2.1 Botryoid

The botryoid is essentially an impoverished raceme (Weberling, 1989), meaning a "determinate raceme". Each individual flower is attached directly to the secondary axis. Hence, a series of individual flowers arranged along the secondary axis forms the TU and is terminated by a prophyll attached to the primary axis. The total display of the botryoid therefore does not contain BUs and IUs, and an individual flower is analogous to both the BU and the IU (Figure 3.2).

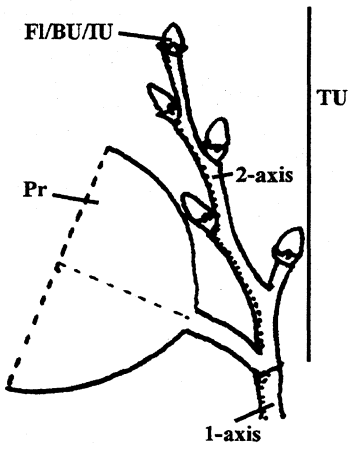


Figure 3.2 A diagram of the Total Inflorescence of botryoid-type nested units in *Hopea* and *Shorea*. Fl: individual flower, BU: Basic Unit, IU: Inflorescence Unit, TU: Total Unit, Pr: prophyll, 1-axis: primary axis, 2-axis: secondary axis.

3.4.2.2 Paniculate

In a panicle, individual flowers are arranged along the BU. The BU can be directly attached to the secondary axis, and the total inflorescence is then termed "simple paniculate" (Figure 3.3). Alternatively, two BUs may be joined by a supporting stem to form an IU, which is in turn attached to the secondary axis. In these inflorescences, the IU seems to be branched into two BUs (tertiary axis) to form a more complex structure. Thus, the BU/IU of the panicle is homologous to the individual flower of the botryoid type. The total inflorescence display for this form is called "complex paniculate" (Figure 3.4).

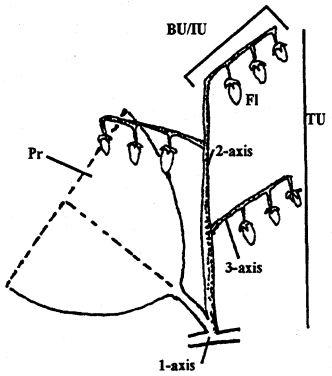


Figure 3.3 A diagram of a Total Inflorescence of simple paniculate-type nested units, as found in *Hopea* and *Shorea*. Fl: individual flower, BU: Basic Unit, IU: Inflorescence Unit, TU: Total Unit, Pr: prophyll, 1-axis: primary axis, 2-axis: secondary axis, 3-axis: tertiary axis.

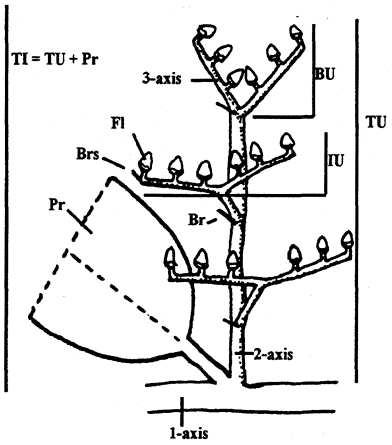


Figure 3.4 A diagram of a Total Inflorescence (TI) of complex paniculate-type nested units, as found in *Hopea* and *Shorea*. Fl: individual flower, Brs: bracteole, BU: Basic Unit, IU: Inflorescence Unit, Br: bract, TU: Total Unit, Pr: prophyll, 1-axis: primary axis, 2-axis: secondary axis, 3-axis: tertiary axis.

This complex paniculate form represents the most complex inflorescence structure recorded in this study. The various complex inflorescence forms found in *Hopea* and *Shorea* may have resulted from development of the branching system and differing orientations of the IU as well as varying numbers of TUs that constitute the TI.

3.4.2.2.1 Branching system of the IUs within the complex paniculate inflorescence

The complex paniculate inflorescence can be either of two types distinguished on the basis of the branching system of the IU. The IU can be branched consistently along the secondary axis (Figure 3.5a) or occur in combination with BUs (Figure 3.5b).

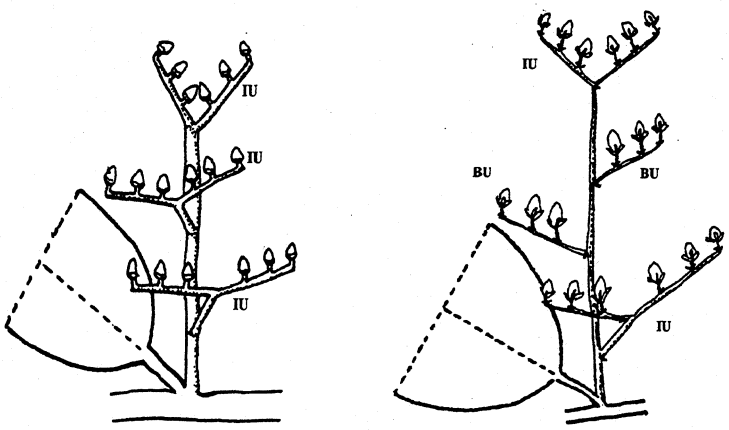
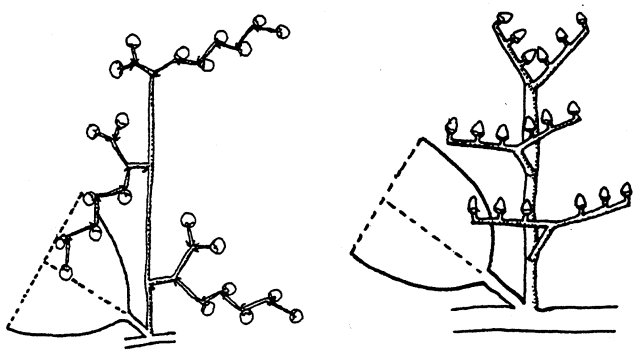


Figure 3.5 Type of branching system of IUs within a complex paniculate inflorescence. Along the secondary axis IUs may be (a) regularly arranged, or (b) occur in combination with BUs. BU: Basic Unit, IU: Inflorescence Unit.

3.4.2.2.2 Orientation of the IU within the complex paniculate inflorescence

The orientation of IUs along the secondary axis within the complex paniculate inflorescence is of two types, 2-dimensional (Figure 3.6a) or 3-dimensional (3.6b). This feature may be only an autapomorphy for certain species, since close observation revealed that one of the forms represents only a modification of the other. Complex paniculate inflorescences with 3-dimensional IU orientation were the most common type found in the species examined.

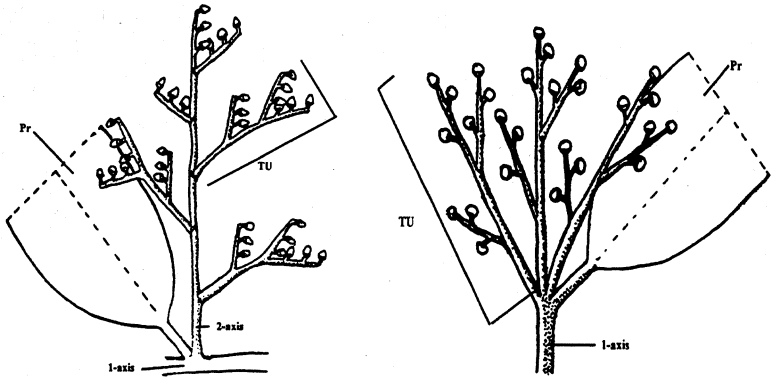


a b

Figure 3.6 Orientation of IUs within a complex paniculate inflorescence, (a) 2-dimensional, and (b) 3-dimensional.

3.4.2.2.3 Arrangement of TUs within the complex paniculate inflorescence

Another distinctive feature of the complex paniculate inflorescence is the arrangement of the TUs within the TI. They may be spiral or alternate (Figure 3.7a) or arranged in a "rosette" (Figure 3.7b). Alternately arranged inflorescences can consist of several TUs with the same or different organisations. This can be analogous to a synflorescence (*sensu* Weberling, 1989), where the TU consists of the main and lateral inflorescence, or to "co-florescences" with one or more laterally borne inflorescences which arise from a bi-axial primary axis (*sensu* Weberling, 1989) or an "enrichment zone" (*sensu* Troll, 1964). When spirally or alternately arranged TUs are condensed, several TUs appear to be borne from the same point and thus form a "rosette" inflorescence. The number of TUs within the TI varies among species.



a b

Figure 3.7 Arrangement of TUs within a complex paniculate, (a) alternate or spiral, and (b) "rosette". Pr: Prophyll, TU: Total Unit, 1-axis: primary axis, 2-axis: secondary axis

Parsing the inflorescence structures described above has shown consistency in arrangement and the smaller units can be used for cladistic analysis to some degree, though it is not yet clear whether such units can be used as diagnostic characters for taxonomic recognition.

3.5 Selection of morphological characters for *Hopea* and *Shorea*

Following consideration of those features used in previous studies on Dipterocarpaceae and examination of herbarium materials, 40 characters were selected for scoring. The resulting data matrix is contained in Appendix 3B.

1 Stipule scars

Scars on the stem left by stipules (Figure 3.8) have often been used as a diagnostic character at the specific and infra-generic levels for Dipterocarpaceae. This character consists of two discrete states:

0 = absent 1 = present

a b

Figure 3.8 Stipule scars, (a) absent and (b) present

Leaves

Leaf shape can be a diagnostic character at the specific level for Dipterocarpaceae (Ashton, 1982) and is therefore included in this analysis. The terminology most commonly used to describe leaf shape can be subjective. However, the leaf can provide discrete (qualitative) characters. For this study, leaf shape has been defined by two of its parameters, leaf length and leaf width.

2 Length of leaf lamina

Hopea and *Shorea* show considerable variation in leaf length (Figure 3.9). However, the size range in *Shorea* tends to be larger than *Hopea*, and this difference may provide a phylogenetic signal in the later analysis. Eleven states are recorded here, via the method of coding continuous characters described earlier in this chapter.

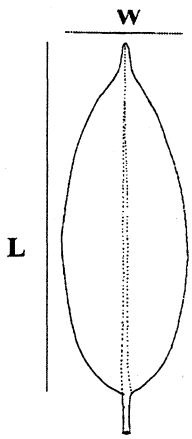


Figure 3.9 Leaf length (L) and width (W)

3 Width of leaf lamina

As for leaf length, this character (Figure 3.9) also comprises 11 multistate characters.

4 Tertiary leaf venation

Tertiary leaf venation is one of the most useful characters for delimiting genera of Dipterocarpaceae—for example, parallel venation is usual in *Dryobalanops* and dryobalanoid venation in *Hopea*. Four qualitative character states (Figure 3.10) are therefore recognised:

0 = scalariform
1 = dryobalanoid
2 = reticulate
3 = parallel

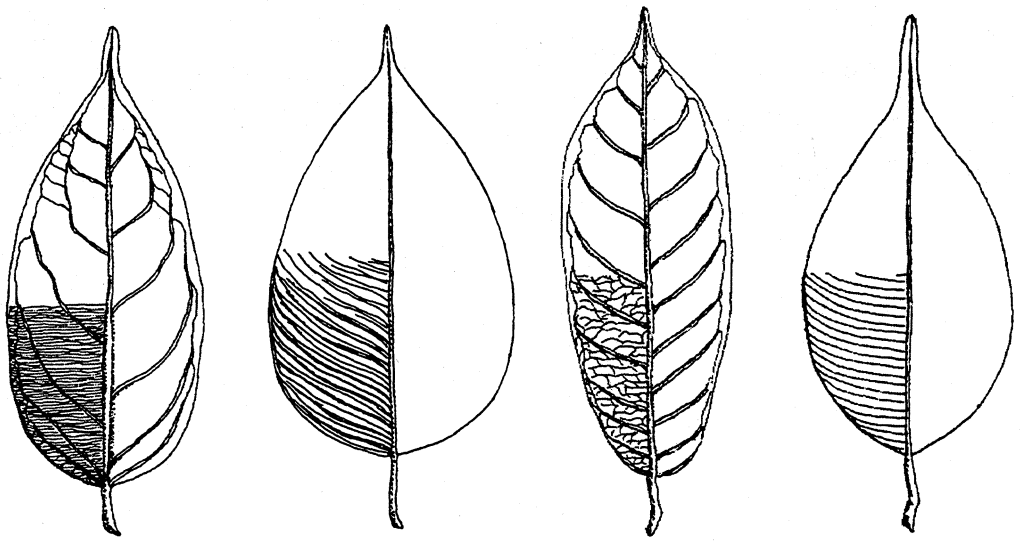


Figure 3.10 Type of leaf venation: (a) scalariform, (b) dryobalanoid , (c) reticulate and (d) parallel

5 Domatia

The presence of gland-like structures on the underside of the leaf surface (Figure 3.11) is common in *Hopea* and saplings of *Shorea* (Ashton, 1982). This character can be diagnostic for species but its presence depends on the stage of developmental growth. Homology may thus be difficult to deduce. Nevertheless, the presence of prominent domatia on certain species has proved useful in previous classifications and therefore this character is included in this study. Two states are recorded:

0 = absent

1 = present

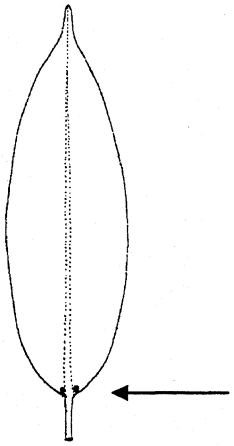


Figure 3.11 Leaf domatia, indicated by arrow

Inflorescence

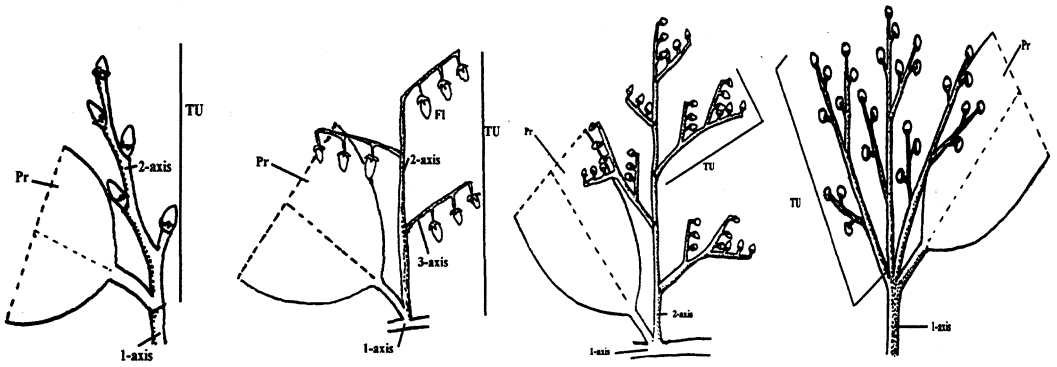
Some hierarchical units of inflorescence structure described earlier may produce informative characters for cladistic analysis.

6 Number of TUs within TI

This character describes the arrangement of the TUs within the TI (Figure 3.12). Two states are recognised:

0 = single

1 = multiple



a b c d

Figure 3.12 Number of TUs within TI, (a, b) single and (c, d) multiple. TU: Total Unit, Pr: prophyll, 1-axis: primary axis, 2-axis: secondary axis, 3-axis: tertriary axis.

7 Position of the inflorescence

This character describes the position of the inflorescence observed, and the states are:

0 = axillary

1 = terminal

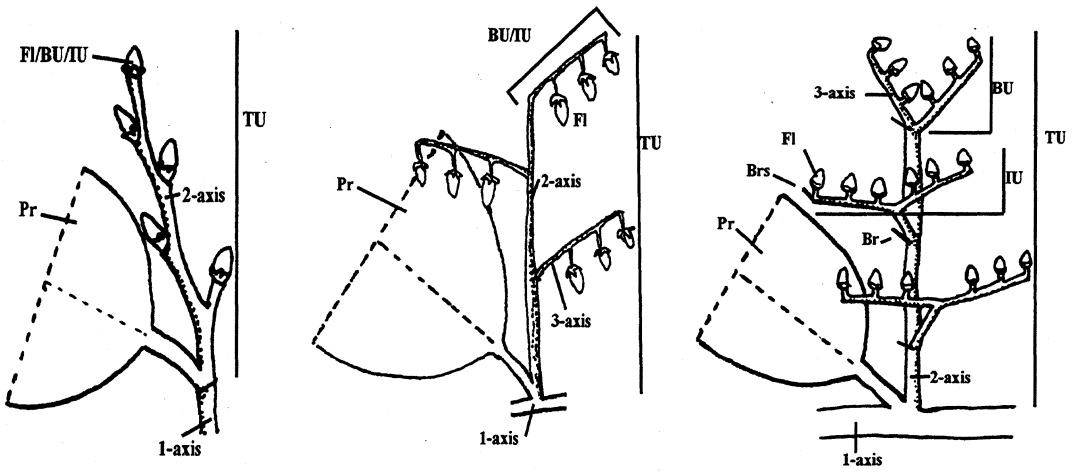
8 Type of TU

This character describes an aggregation of the total floral display and the description provided here is for each type of inflorescence (botryoid and panicle). This character features a simplified interpretation of the TU described in sections 3.4.2.1 and 3.4.2.2, and three states (Figure 3.13) are recognised:

0 = botryoid, where there are no actual BUs and IUs, and hence an individual flower is analogous to the BU of a simple paniculate inflorescence and the IU of a complex inflorescence.

1 = simple paniculate form, where there are no IUs and hence a BU is analogous to an IU of a complex inflorescence.

2 = complex paniculate form, where all hierarchical units exist.



a b c

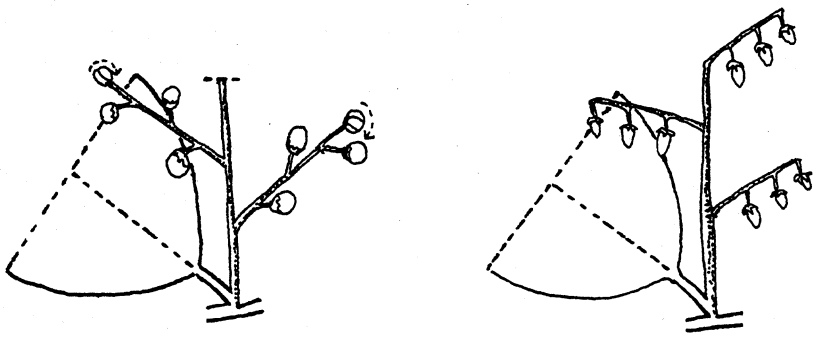
Figure 3.13 Type of the Total Unit (TU): (a) botryoid, (b) simple paniculate, and (c) complex paniculate. Fl: individual flower, BU: Basic Unit, IU: Inflorescence Unit, TU: Total Unit, Pr: prophyll, 1-axis: primary axis, 2-axis: secondary axis, 3-axis: tertiary axis

9 Termination of BU

This character refers to the presence of a terminal flower, and therefore there are two discrete characters: determinate and indeterminate (Figure 3.14). Determinate describes a situation where the BU ceases its growth with the formation of a terminal flower, and therefore there are a limited number of flowers in each unit. Indeterminate describes a situation where the BU meristem grows indefinitely, producing an indefinite number of flowers. Two states are recorded:

0 = determinate

1 = indeterminate



a b

Figure 3.14 Type of termination of BU, (a) indeterminate and (b) determinate

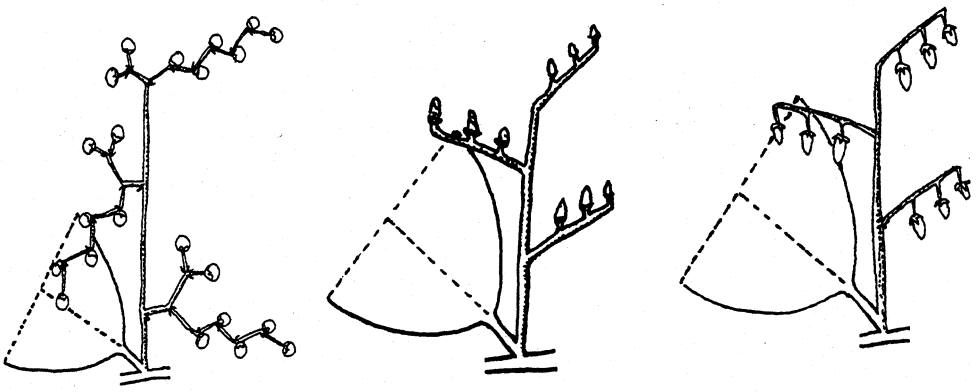
10 Orientation and directional growth of individual flowers on axis

The flowers can be differentiated on the basis of their position on the axis and their orientation (Figure 3,15). When a series of flowers has the same orientation, either pointing up (Figure 3.15b) or pointing down (Figure 3.15c), the term "secund" is applied. Non-secund flowers are arranged alternately on their axis (Figure 3.15a). The possible states for this character are therefore:

0 = non-secund

1 = secund, pointing up

2 = secund, pointing down



a b c

Figure 3.15 Orientation of flower growth: (a) non-secund, (b) secund, pointing up and (c) secund, pointing down

Flowers

The shape and size of the flower bud may be diagnostic for certain sections of *Shorea* (Ashton, 1982) and it is therefore included as a character in this analysis.

11 Length of individual flower bud

Hopea and *Shorea* show considerable variation in length of individual floral buds, and this variation may be informative for cladistic analysis. The measurements are taken on a minimum of five of the oldest flower buds (Figure 3.16) and 11 states recognised.

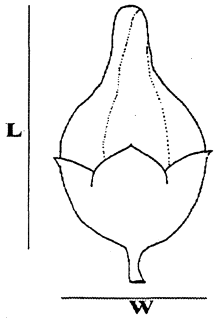


Figure 3.16 Length (L) and width (W) of flower bud

12 Width of individual flower bud

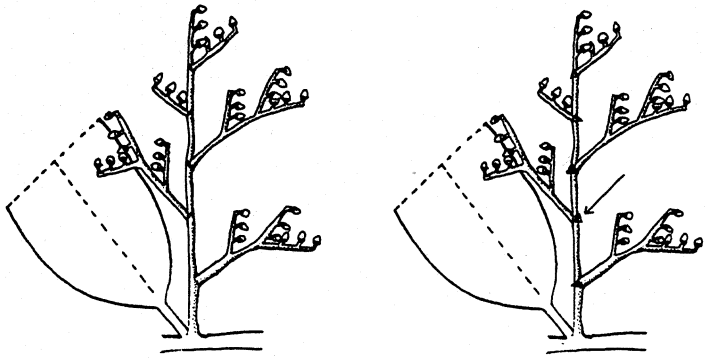
As with floral bud length (Figure 3.16), this character may contain phylogenetic information. Ten states are recorded.

13 Persistence of bracts

Several species possess a prominent bract-like organ (Figure 3.17b) on the inflorescence, but in some taxa it falls early (caducous, Figure 3.17a). This feature defines the BU-IU and consists of two discrete states:

0 = caducous

1 = persistent



a b

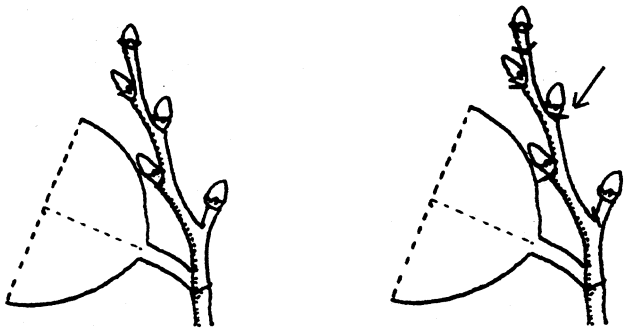
Figure 3.17 Persistence of bracts: (a) caducous and (b) persistent (indicated by arrow)

14 Persistence of bracteoles

As with bracts, in some taxa the bracteoles are persistent (Figure 3.18b), while in others the bracteoles fall early (caducous, Figure 3.18a). The bracteoles define the individual flower and two discrete states were scored:

0 = caducous

1 = persistent



a b

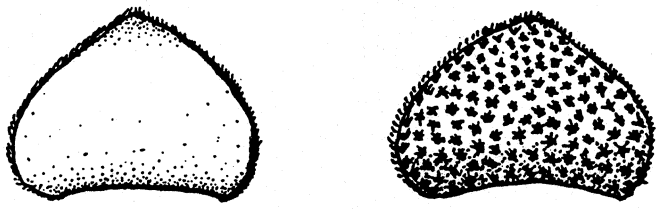
Figure 3.18 Persistence of bracteoles: (a) caducous and (b) persistent (indicated by arrow)

15 Presence of hairs on sepal

Preliminary analysis showed that the presence of hairs on the sepal (Figure 3.19) may be diagnostic for certain taxa. Two states are recorded:

0 = absent

1 = present



a b

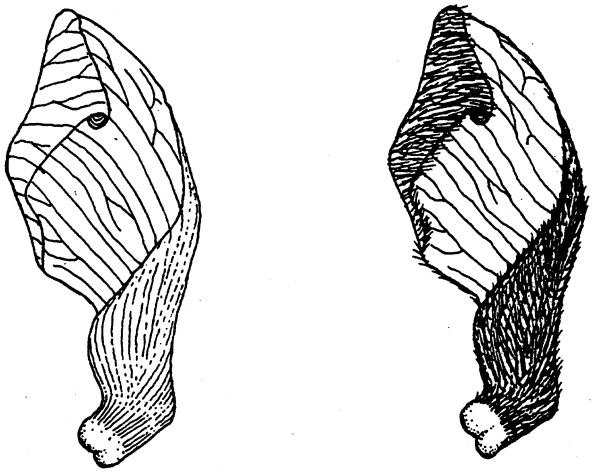
Figure 3.19 Presence of hairs on sepals: (a) absent and (b) present

16 Presence of hairs on petal

As with character 15, presence of hairs on the petals (Figure 3.20) may also be diagnostic for certain taxa. Two states are also recorded for this character:

0 = absent

1 = present



a b

Figure 3.20 Presence of hairs on petals, (a) absent and (b) present

17 Length of petal

Ashton (1982) considered that the shape of the perianth is diagnostic for some members of section *Shorea*, and therefore it was important to include it in the analysis. In order to avoid any subjectivity in deducing the shape, only quantitative measurements of petal length post-anthesis (Figure 3.21) were used and 11 states were recorded.

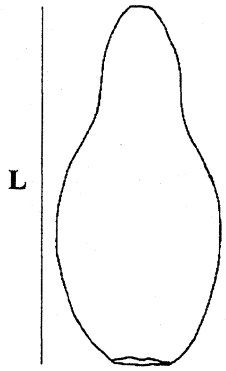


Figure 3.21 Length (L) of petal

Stamens

Different parts of the stamens may provide diagnostic characters and therefore it is important to parse the staminal parts. Certain taxa may possess different shapes of the appendage, anther and filament. These parts of the stamen appear to be independent characters. In addition, parsing the staminal parts is important to describe the hairiness on these parts, which also appears to be independent.

18 Number of stamens

Ashton (1982) suggested that the number of stamens, being commonly less than 20 in a "normal" flower, may be a diagnostic character at the species level. This character was treated as binary instead of qualitative, since some members of the outgroup have numerous stamens which were difficult to count. The character is divided into two states, up to 15 stamens and more than 15 stamens, since a statistical analysis (not shown) indicated that this character's distribution was polarised between these two states:

0 = stamens more than 15

1 = stamens up to 15

19 Number of rows of stamens

This character may be correlated with the number of stamens, since with more stamens present, the number of rows arranged around the ovary (Figure 3.22) also increases. This character comprises two states:

0 = more than two rows of stamens

1 = up to two rows of stamens

a b

Figure 3.22 Arrangement of stamens, (a) in more than two rows and (b) in up to two rows.

20 Type of anther connective appendage

Dipterocarpaceae is characterised by an additional protruding organ extending beyond the top of the anther. There are several types (shapes) of appendage recognised within the Sub-family Dipterocarpoideae and these shapes may be diagnostic, particularly in *Shorea*. However, the homology of the recorded types across the groups studied could only be divided into three states (Figure 3.23):

0 = broad appendage widest in the upper part

1 = semi broad, which is an intermediate form between states 0 and 2

2 = filiform

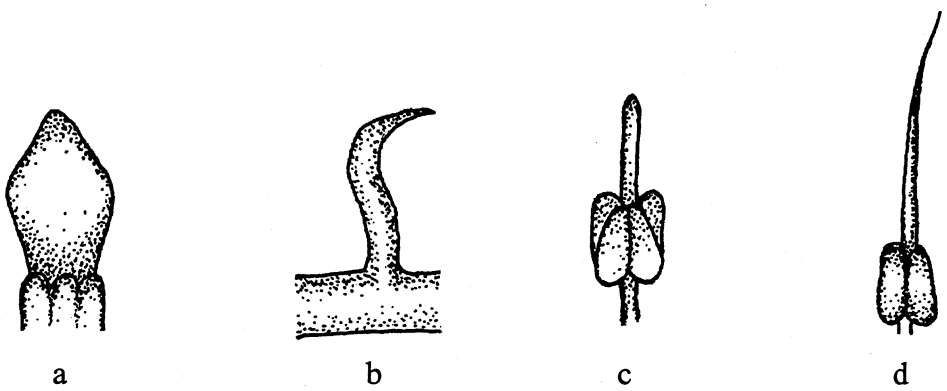


Figure 3.23 Type of anther connective appendage: (a) broad, (b, c) semi-broad, and (d) filiform.

21 Length of anther connective appendage

This measurement is determined as the average of the five longest appendages in each flower (Figure 3.24). Eleven states are recorded for this character.

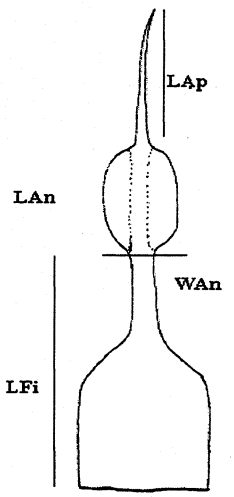


Figure 3.24 Anther measurements: length of anther connective appendage (LAp, character 21), length of anther (LAn, character 23), width of anther (WAn, character 24) and length of filament (LFi, character 28)

22 Presence of hairs on anther connective appendage

The presence of stellate hairs can be diagnostic for certain sections of *Hopea* and *Shorea*. It is represented by two states (Figure 3.25):

0 = absent

1 = present

a b

Figure 3.25 Presence of hairs on connective appendage, (a) absent and (b) present

23 Length of anther

Measurements of anther length were taken as illustrated in Figure 3.24. Eleven states are recorded for this character.

24 Width of anther

Measurements of anther width were taken as illustrated in Figure 3.24. Ten states are recorded for this character.

25 Presence of hairs on anthers

The presence of stellate hairs on the anthers (Figure 3.26) may provide a phylogenetic signal in the analysis. Two states are recorded:

0 = absent

1 = present

a b

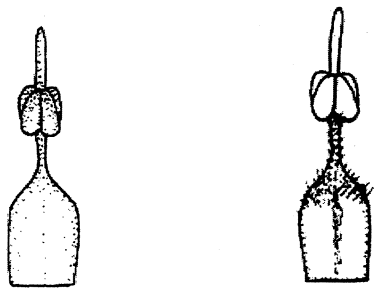
Figure 3.26 Presence of hairs on anther: (a) absent and (b)present.

26 Presence of hairs on filament

The presence of stellate hairs on the filament (Figure 3.27) may provide phylogenetic information for the analysis. Two states are recorded:

0 = absent

1 = present



a b

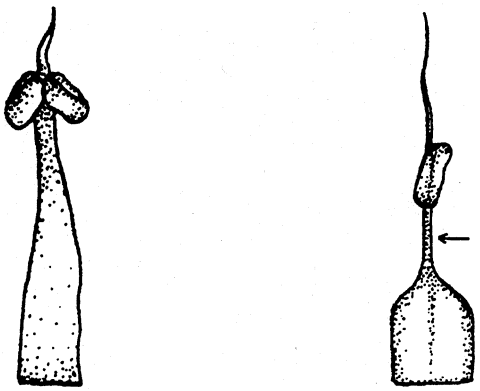
Figure 3.27 Presence of hairs on filament: (a) absent and (b) present.

27 Presence of neck in filament

Ashton (1982) considered the shape of the filament and the presence of a "neck" (Figure 3.28) to be a particularly important diagnostic character for some sections of *Shorea*. Binary states are recognised:

0 = absent

1 = present



a b

Figure 3.28 Presence of neck in the filament: (a) absent and (b) present, with arrow indicating the neck.

28 Length of filament

This continuous character is included because *Hopea* and *Shorea* show a wide range of filament size. The measurement was taken on a minimum of five of the longest filaments in each flower (Figure 3.24). Nine states are recorded.

Pistil

The ovary and style seem to provide independent characters that can be incorporated into a cladistic analysis. A discussion of exclusion of the shape of pistil is given in Section 3.6. Parsing the pistil into its various components is also important to

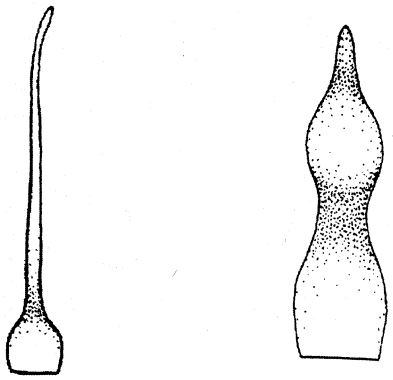
describe their hairiness, since the presence of hairs on different parts of pistil seems to be independent.

29 Presence of stylopodium

A stylopodium is an extended part of the style that is located abaxial to the ovary (Figure 3.29). The presence or absence of a stylopodium appears to be important in determining the shape of the pistil. Various forms of pistil are recognised among the sub-family Dipterocarpoideae (Ashton, 1982). However, the homology of these features cannot easily be inferred, since the pistil types are variable and continuous. A simple binary scoring was thus used for the stylopodium, with two states recognised:

0 = absent

1 = present



a b

Figure 3.29 Presence of stylopodium: (a) absent and (b) present.

30 Length of pistil

This continuous character is used since the two study genera exhibit a wide range of sizes. The measurement includes the ovary and style (Figure 3.30). Ten states are recognised.

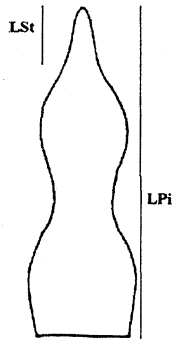


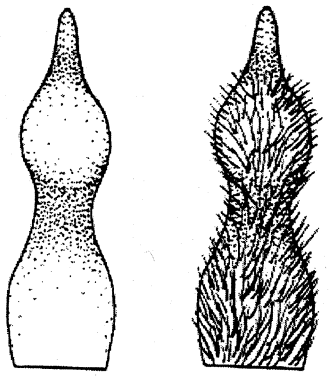
Figure 3.30 Length of pistil (LPi, character 30) and of style (LSt, character 32).

31 Presence of hairs on ovary

Since the ovary and stylopodium are not distinguished, the presence of hairs on the stylopodium is referred to as "hairiness of the ovary" (Figure 3.31). Two states are recognised:

0 = absent

1 = present



a b

Figure 3.31 Presence of hairs on ovary: (a) absent and (b) present.

32 Length of style

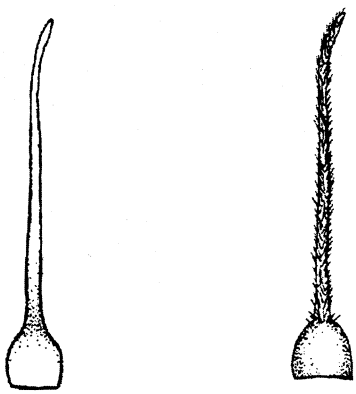
This continuous character (Figure 3.30) is used because *Hopea* and *Shorea* show a wide range of sizes. Ten states are recorded for this character.

33 Presence of hairs on style

The presence of hairs on the style (Figure 3.32) is recorded as two states:

0 = absent

1 = present



a b

Figure 3.32 Presence of hairs on the style: (a) absent and (b) present

Fruits

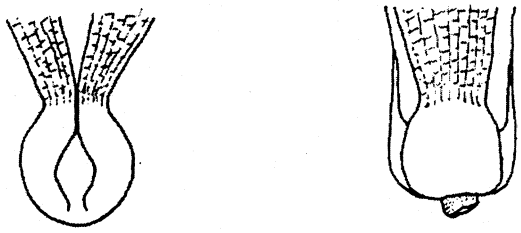
The morphology of the fruits provides particularly important characters for generic diagnosis of Dipterocarpaceae, including the fusion of fruit calyx lobes and the comparative development of the fruit calyx. In existing classifications, several genera are defined solely by using characters derived from the fruit.

34 Presence of fruit pedicel

Some taxa show a distinct pedicel, while others have sessile fruit (Figure 3.33). Two states are recorded for this character:

0 = absent

1 = present



a b

Figure 3.33 Presence of fruit pedicel: (a) absent or sessile and (b) present or prominent.

35 Comparative development of fruit calyx

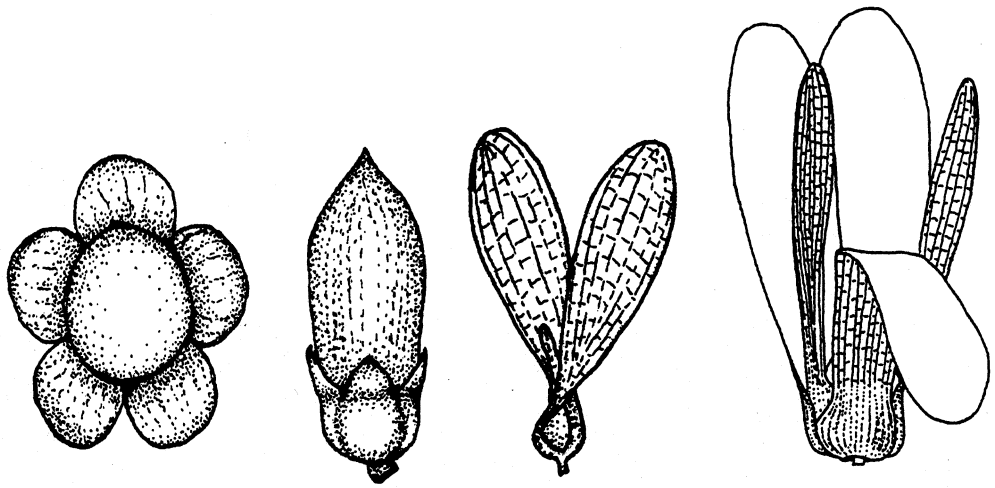
Dipterocarpaceae species have five sepals and their development is different among the genera (Figure 3.34). This comparative development of the sepals in the fruit is used as a diagnostic character at tribal rank (Ashton, 1982). This development includes the fusion of the base of the sepals into a calyx tube and the elongation of the sepals into wing-like structures. The fruit calyx of Tribe Dipterocarpeae is valvate at the base, while in Tribe Shoreae (which includes *Hopea* and *Shorea*) the fruit sepals expand and are imbricate at the incrassate cupped base (Ashton, 1982). The sepals of some genera develop into five equal "fruit wings" (e.g. *Dryobalanops* and *Vatica* section *Vatica*), while in other taxa the five sepals do not develop into significant wings (e.g. *Shorea multiflora*, *Hopea brevipetiolaris* and *Neobalanocarpus heimii*—see Figure 3.34a). However, most species of Dipterocarpaceae have unequal development of the sepals in the fruit, and two or three sepals thus develop to be larger than the others. This comparative development is often referred to as "unequal

fruit wings" (Figure 3.34 c, d). Three character states are scored for the comparative development of the sepals in fruit:

0 = subequal/equal

1 = unequal sepals: 2 long and 3 short

2 = unequal sepals: 3 long and 2 short



a b c d

Figure 3.34 Type of fruit wings: (a, b) equal/subequal and (c,d) unequal.

36 Shape of fruit wings

It was considered necessary to make the shape of the fruit wing a separate character, because it seems to be independent from character 35. For instance, taxa that possess equal development of the sepals often have a different sepal shape (compare Figure 3.34a and b) and the combination of the number of wings and wing shape is often used as a generic diagnostic feature. Hence, three states are recorded for this character:

0 = spatulate, linear (Figure 3.34c and d)

1 = triangular (Figure 3.34b)

2 = rounded (Figure 3.34a)

37 Length of nut

This continuous character is used since the taxa of interest show a wide range of sizes. Measurements are taken at the longest part of the nut, including the style remnant if present (Figure 3.35). Eleven states are recorded for this character.

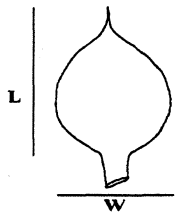


Figure 3.35 Nut measurements: length (L, character 37) and width (W, character 38).

38 Width of nut

Measurements are taken at the widest part of the nut (Figure 3.35) and 11 states are recorded for this character.

39 Presence of hairs on nut

This character seems to be consistent across species or genera and two states are recognised (Figure 3.36):

0 = absent

1 = present

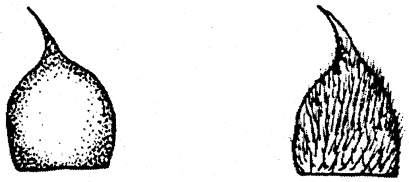


Figure 3.36 Presence of hairs on nut: (a) absent and (b) present

40 Presence of style remnant

The presence of a style remnant and its indumentum (Figure 3.37) may provide a diagnostic character for certain taxa. Hence, three states are recognised:

0 = absent

1 = present, glabrous

2 = present, not glabrous

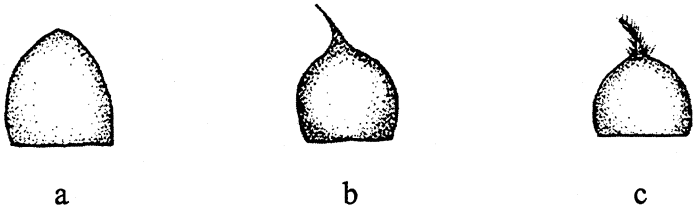


Figure 3.37 Presence of style remnant and indumentum of remnant: (a) absent, (b) present, glabrous and (c) present, hairy.

Excluded characters

Three characters, or groups of characters, have been excluded from this analysis, even though some authors have suggested these are important for diagnosing species or genera.

1 Indumentum of leaf surface (leaf texture)

The indumentum of the upper leaf surface may provide a useful character to distinguish *Hopea* and *Shorea*, though Ashton (1982) argued hairiness is likely to be affected by environment. The type of indumentum varies between taxa and leaves may be tomentose, scabrous or glabrous. To identify the different types of indumentum, a scanning electron microscope (SEM) would be needed, with only stellate hairs being visible under a light microscope. This study did not involve the use of SEM, and therefore scoring leaf indumentum may have provided incorrect homology determinations, since taxa that possess non-stellate hairs would be assigned a score of 0.

2 Wood and bark anatomy

Wood and bark anatomy have proven to be useful for generic, and to some extent infra-generic, differentiation among Dipterocarpaceae (Whitmore, 1962; Parameswaran and Gotwald, 1979 in Maury-Lechon & Curtet, 1998). A recent molecular phylogenetic study (Dayanandan *et al.*, 1999) used four wood anatomy characters to infer relationships among the genera of Dipterocarpaceae. They showed that these characters were useful for differentiation among certain genera of Dipterocarpaceae but there was no variation within *Hopea* and *Shorea*. Thus, these characters are not included in the present study.

3 Shape of the pistil

The pistil or gynoecium varies in shape among Dipterocarpaceae. Though several types of pistil are recognised among the sub-family Dipterocarpoideae, this feature is not generically informative. However, the shapes can be diagnostic at the infra-generic level for certain *Shorea* taxa. For example, section *Pachycarpae* has a spindle-like pistil, section *Neohopea* has a conical ovary and stylopodium, and section *Anthoshorea* does not have a distinct stylopodium. Based on observations of the pistil across the selected species and the compilation of these shapes among

Dipterocarpaceae, it is suggested that the homology of the shape of pistil can be inferred from the presence of stylopodium. For this reason, the shape of the pistil is excluded from the analysis.

Summary of characters selected

Forty characters were selected, with five vegetative and 35 reproductive (7 fruit and 28 floral) characters being used, as shown in Table 3.1. The 40 characters comprised 27 qualitative and 13 quantitative characters. All the 13 quantitative characters are continuous.

3.6 Selection of taxa for analysis

Since this study aimed to assess the evolutionary relationships between *Hopea* and *Shorea*, the species selected were those thought likely to have implications for their evolutionary history. The taxa selected for analysis were therefore chosen to represent the full range of morphological variation, to include previously proposed infra-generic divisions, and to fully cover the geographic distribution. The taxa selected for this study are shown in Table 3.2 and the herbarium voucher material studied is listed in Appendix 3A.

3.6.1 Ingroup

The most recent taxonomic work on Dipterocarpaceae was for the *Flora Malesiana* (Ashton, 1982) and it examined the sub-family Dipterocarpoideae in Malesia. Ashton's classification system was used as the basis for taxon selection in this study, for the reasons explained in Chapter 2.

All sections and sub-sections were represented in this study by including the type species of each (Table 3.2). Some of Ashton's (1982) infra-generic divisions are defined by a few distinctive morphological features—for example, *Shorea* section *Pachycarpae* is characterised by the largest seeds, and *Shorea* section *Doona* is characterised by a broad connective appendage. These two sections are narrowly endemic (section *Pachycarpae* to Borneo and section *Doona* to Sri Lanka), so in effect taxon selection in these groups was made based both on morphological features and ecological distribution. However, when a section is delimited by a number of well-defined morphological features, such as in *Shorea* section *Anthoshorea*, species

were selected according to their geographic distribution. For sections or sub-sections containing only one species, that species was selected.

Table 3.1 Characters used in the morphological study.

No	Name of character	Type of character	
		Qualitative	Quantitative
1	Stipule scars	•	
2	Length of leaf lamina		•
3	Width of leaf lamina		•
4	Tertiary leaf nervation	•	
5	Domatia	•	
6	Number of TUs within TI	•	
7	Position of the inflorescence	•	
8	Type of TU	•	
9	Termination of BU	•	
10	Orientation and directional growth of individual flowers on axis	•	
11	Length of individual flower bud		•
12	Width of individual flower bud		•
13	Persistence of bract	•	
14	Persistence of bracteoles	•	
15	Presence of hairs on sepal surface	•	
16	Presence of hairs on petal surface	•	
17	Length of petal		•
18	Number of stamens	•	
19	Number of rows of stamens	•	
20	Type of anther connective appendage	•	
21	Length of anther connective appendage		•
22	Presence of hairs on anther connective appendage	•	
23	Length of anthers		•
24	Width of anthers		•
25	Presence of hairs on anthers	•	
26	Presence of stellate hairs on filament	•	
27	Presence of neck in filament	•	
28	Length of filament		•
29	Presence of stylopodium	•	
30	Length of pistil		•
31	Presence of hairs on the ovary	•	
32	Length of style		•
33	Presence of hairs on style	•	
34	Presence of fruit pedicel	•	
35	Comparative development of fruit calyx	•	
36	Shape of fruit wings	•	
37	Length of nut		•
38	Width of nut		•
39	Presence of hairs on nut	•	
40	Presence of style remnant	•	
	TOTAL	27	13

Table 3.2 Taxa selected for morphological study

Putative outgroup			Abbreviation
Parashorea malaanonan			PMALA
Neobalanocarpus heimii			NHEMI
Dryobalanops aromatica			DAROM
Dryobalanops lanceolata			DLANC
Dipterocarpus confertus			DCONF
Dipterocarpus retusus			DRETU
Total number of the outgroup		6	

Genus *Hopea*

Section	Subsection	Species	Abbreviation
Dryobalanoides	*Dryobalanoides:*	*H. pubescens*	HPUBE
		H. dryobalanoides	HDRYO
		H. ferruginea	HFERR
		H. mengerawan	HMENG
		H. pierrei	HPIER
		H. dyeri	HDYER
	Sphaerocarpa	*H. bracteata*	HBRAC
		H. subalata	HSUBA
		H. sphaerocarpa	HSPHA
		H. nervosa	HNERV
		H. nigra	HNIGR
Hopea	*Hopea*	*H. celebica*	HCELE
		H. aptera	HAPTE
		H. celtidifolia	HCELT
		H. papuana	HPAPU
		H. odorata	HODOR
	Subsection Pierrea	*H. wyatt-smthii*	HWYAT
		H. apiculata	HAPIC
		H. philippinensis	HPHIL
		H. jucunda	HJUCU
		H. chinensis	HCHIN
		H. discolor	HDISC
		H. hainanensis	HHAIN
		H. brevipetiolaris	HBREV
		*H. wightiana***	HWIGH
Total number of *Hopea* species		25	

* monotypic section or subsection
** currently recognised as *Hopea ponga* (Dennst.) Mabb.
*** recognised as *Shorea worthingtoni* Ashton, Syn = *Doona venulosa*
Bold font shows type species of each genus, section or subsection.

Table 3.2 Taxa selected for morphological study (continued)

Genus *Shorea* Section	Subsection		Abbreviation
Shorea	*Shorea*	***S. robusta***	SROBU
		S. seminis	SSEMI
		S. guiso	SGUIS
		S. foxworthyii	SFOXW
		S. hypoleuca	SHYPL
	Barbata	*S. laevis*	SLAEV
		S. asahii	SASAH
		S. maxwelliana	SMAXW
*Pentacme**		***S. siamensis***	SSIAM
*Neohopea**		***S. isoptera***	SISOP
Richetioides	*Polyandrae**	***S. polyandra***	SPOLY
	Richetioides	*S. acuminatissima*	SACUM
		S. faguetiana	SFAGU
		S. longiflora	SLONF
		S. multiflora	SMULT
		S. richetia	SRICH
		S. hopeifolia.	SHOPE
		S. maxima	SMAXI
Anthoshorea		*S. javanica*	SJAVA
		S. virescens	SVIRE
		S. hypochra	SHYPC
		S. roxburghii	SROXB
		S. stipularis	SSTIP
*Rubella**		*S. rubella*	SRUBE
Brachypterae	*Smithiana*	***S. smithiana***	SSMIT
	Brachypterae	*S. balangeran*	SBALA
		S. johorensis	SJOHO
		S. parvistipulata	SPARV
		S. scaberrima	SSCAB
		S. kunstleri	SKUNS
		S. selanica	SSELA
		S. singkawang	SSING
		S. palembanica	SPALE
Pachycarpae		*S. amplexicaulis*	SAMPL
		S. macrophylla	SMACR
		S. mecisopteryx	SMECI
		S. pinanga	SPING
		S. splendida	SSPLE
		S. stenoptera	SSTEN
		S. pilosa	SPILO
		S. beccariana	SBECC
		S. rotundifolia	SROTU
Mutica	*Auriculatae**	*S. macroptera*	SMACT
	Mutica	***S. parvifolia***	SFOLI
		S. macrantha	SMACN
		S. leprosula	SLEPR
*Ovalis**		***S. ovalis***	SOVAL
Doona		*S. thorelii*	STHOR
		S. congestiflora	SCONG
		S. cordifolia	SCORD
		*S. venulosa****	SVENU
		S. trapezifolia	STRAP
		S. gardneri	SGARD
Total number of *Shorea* species		53	
Total number of selected species		84	

Twenty-five species of *Hopea* and 53 species of *Shorea* were selected using these methods. Of the 25 *Hopea* species, two are distributed over New Guinea, four are Indochinese, two are Indian and four are from Sri Lanka. The remainder of the species exhibit Malesian distributions. Of the 53 species of *Shorea* selected, six are Sri Lankan, two Indian, seven Indochinese and the rest have Malesian distributions. The selection of the ingroup represented all the infra-generic divisions of both *Hopea* and *Shorea* and covered most of their geographic distribution.

3.6.2 Outgroup

The outgroup should consist of species that are considered the closest relatives of the ingroup and some that are more distantly related. The more distantly related species selected in this study are *Dipterocarpus retusus* and *D. confertus* of Tribe Dipterocarpeae. These species show some characters not possessed by the ingroup, such as large flowers and a calyx tube that surrounds the fruit but is not fused to it. Such outgroup characters may therefore be used to polarise the ingroup characters for analysis. These species are clearly separated from Tribe Shoreae based on a molecular phylogeny of Dipterocarpaceae using the *rbc*L gene (Dayanandan *et al.*, 1999).

The genus *Dryobalanops* was selected to provide a putative sister taxon to a grouping of both *Hopea* and *Shorea*, since it was placed in a separate clade from *Hopea* and *Shorea* in the analysis by Dayanandan *et al.* (1999), albeit with an unresolved phylogenetic position. The putative sister genus to *Hopea* (*Neobalanocarpus*) and to *Shorea* (*Parashorea*) are also included, even though Trueman (1997) suggested that putative sister groups should be excluded from the analysis since their proximity to the ingroup may cause biases towards unpolarised characters. They are included in this analysis in order to identify the most likely sister groups to *Hopea* and *Shorea*, to test the nature of the putative sister taxa and to examine the discreteness of the characters that are used to separate them. These putative sister groups have been used in molecular phylogenies of Dipterocarpaceae, all of which have confirmed their close relationships to *Hopea* and *Shorea* (Tsumura *et al.*, 1996; Kajita *et al.* 1998; Dayanandan *et al.*, 1999).

3.7 Data analysis

Cladistic analyses were performed using PAUP* 4.0b4a (Swofford, 1998), with different versions of the data set in order to test the effect of the continuous characters, the inflorescence characters and missing data. The first analysis included the complete data set of 84 taxa and 40 characters. Using the same taxa included in the first analysis, the second analysis excluded the thirteen continuous characters and the third eliminated five inflorescence characters. The fourth analysis excluded five taxa (*Hopea discolor*, *H. jucunda*, *Shorea congestiflora*, *S. gardneri* and *S. javanica*) that had more than 50% missing data. All of these analyses were performed to examine the effect of different character sets on the robustness of the resultant phylogenies. The impact of continuous characters, inflorescence characters and of taxa with a significant amount of missing data were of particular interest.

The optimal tree—evaluated using the maximum parsimony criterion—was estimated using an heuristic search strategy. For all final tree searches, one thousand random addition sequence replicates were conducted in order to search for multiple "islands" of most parsimonious trees (Maddison, 1991). A maximum of 500 trees was saved. The TBR (Tree Biconnection Reconnection) search strategy was used, with steepest descent off, ACCTRAN (Accelerated Transformation), MULPARS (Multiple Parsimonious Trees), and with branches of length zero being collapsed. Ten equally parsimonious trees were saved from each replicate (Swofford, 1998).

The character states were treated as ordered (Wagner) only (Swofford, 1998). Multistate characters were coded as polymorphism (Kornet and Turner, 1999), since polymorphic characters can contain significant phylogenetic information (Weins, 1995). Statistical measures of the Consistency Index (CI), Homoplasy Index (HI) (Kluge and Farris, 1994), Rescaled Consistency Index (RC) and Retention Index (RI) (Farris, 1989) were also calculated.

Clade support was estimated by performing 100 bootstrap replicates (Felsenstein, 1985) with MULPARS off and then calculating the 50% majority-rule bootstrap tree. Trees were rooted using the outgroup taxa previously selected.

3.8 Results

3.8.1 The complete data set

This analysis includes all the 84 taxa and 40 morphological characters. Using the search strategy outlined above, two optimal trees were found. These have a length of 751 steps, CI of 0.23, HI of 0.77, RI of 0.46 and RC of 0.11.

One of the two cladograms obtained in the first analysis is shown in Figure 3.38, with the outgroup of *Dipterocarpus retusus* and *D. confertus* used to root the trees. Several of the ingroup taxa then form a grade outside the "core" ingroup clade. These lineages mainly consist of taxa from *Shorea* section *Shorea* and a few other species, including *Shorea thorelii* (section *Doona*), *S. foxworthyi* and *S. robusta* and *S. maxwelliana* (all from section *Shorea*), *S. polyandra* (from section *Richetioides*), and *S. parvifolia* and *S. singkawang* (both from section *Mutica*). A clade consisting of *S. foxworthyi* and *S. thorelii* forms the sister lineage to the remainder of the ingroup, suggesting that these two species are morphologically quite distinct from the others.

The ingroup (labelled A, synapomorphies 6, bootstrap 82%) includes a clade considered to be the "core" ingroup (labelled J), which consists of *Hopea*, *Shorea* and their putative sister taxa. The six synapomorphic changes defining the total ingroup clade include the complex paniculate type of Total Unit in the inflorescences, size of the flower bud, presence of the fruit pedicel and comparative development of the fruit calyx.

All the putative sister taxa included in the analysis are nested within the *Shorea* clade. This suggests a closer relationship between *Shorea* and these taxa than any of them share with *Hopea*. *Dryobalanops aromatica*, *D. lanceolata* and *Neobalanocarpus heimii* group together (clade labelled U) on the basis of eight synapomorphic changes, including type of leaf nervation and comparative development of the fruit calyx. This group falls within one of the *Shorea* subclades (T). *Shorea*'s putative sister taxon, *Parashorea malaanonan*, is grouped within another *Shorea* subclade (Q).

The core ingroup (J, synapomorphies 4, bootstrap <50%) consists of two further subgroupings, which each correspond to the generic divisions into *Hopea* (C) and *Shorea* (K). Species in this group share synapomorphic changes relating to flower

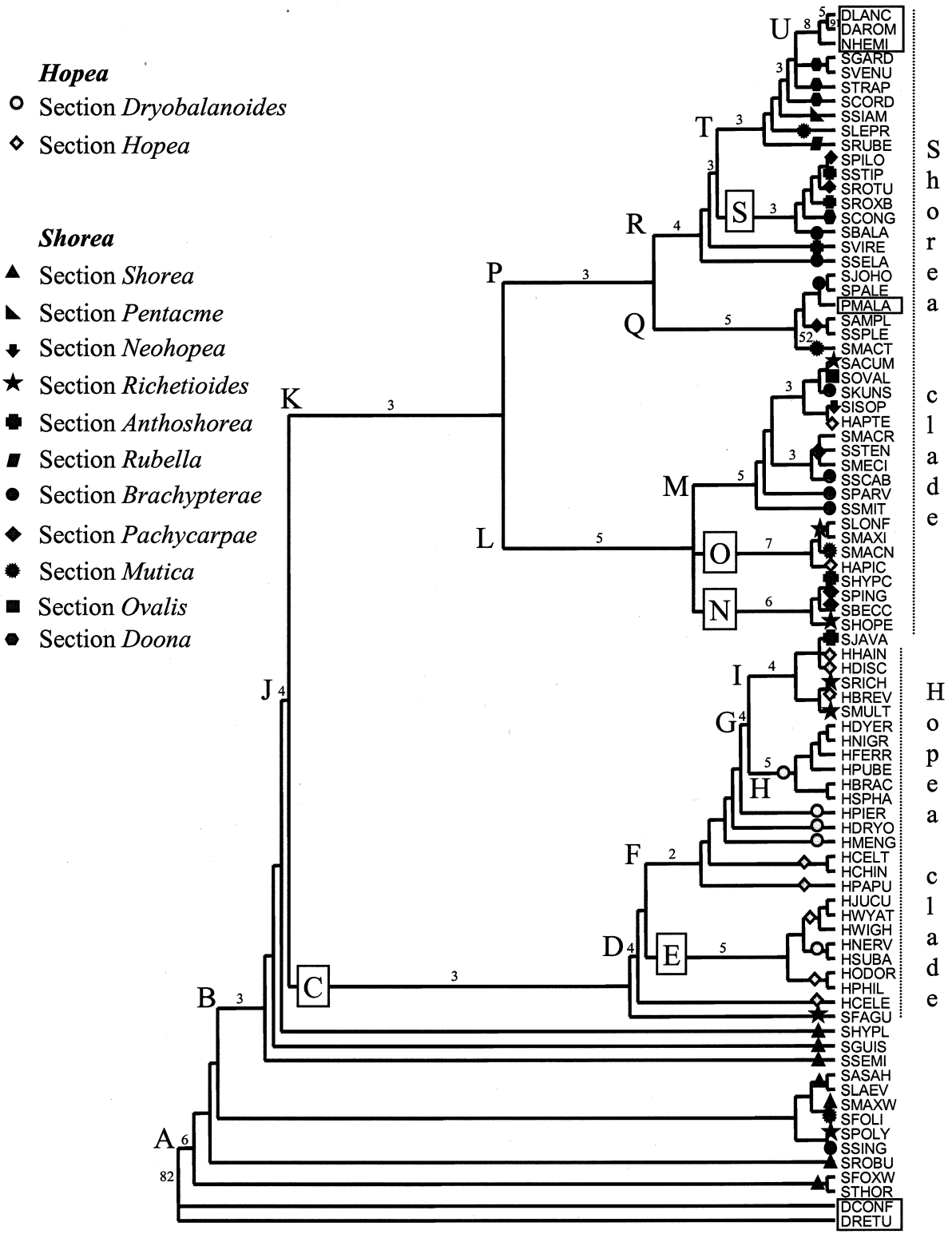


Figure 3.38 One of the most parsimonious trees obtained from cladistic analysis of the complete morphological data set, using 40 characters and 84 taxa. Numbers above the branches are branch lengths, and bootstrap values of 50% and greater are shown below. Taxon names in boxes are the putative outgroups.

bud size, petal size, stamen number and the absence of hairs on the anther connective appendage.

The *Shorea* clade (K, synapomorphies 3, bootstrap <50%) is defined by the size of the leaf lamina and the pistil length. Within this clade, all sections of *Shorea* are paraphyletic, other than the monotypic sections *Neohopea*, *Pentacme*, *Rubella* and *Ovalis*. The exclusion of section *Shorea* from this clade may suggest that this section has undergone separate diversification and thus that their morphology has diverged from that of the other taxa in *Shorea*. The endemic Sri Lankan *Shorea* section *Doona,* represented by six species, is paraphyletic with most of its species placed in group T (synapomorphies 3, bootstrap <50%). The Bornean endemic section *Pachycarpae*, represented by nine (of 10 total) species, also appears to be paraphyletic (groups M, N and Q) as is *Shorea* section *Brachypterae*, with all its taxa that were included in the analysis grouping with taxa from other sections. Most members of section *Richetioides* are excluded from the main *Shorea* clade, with species either grouping with *Hopea* or forming part of the unresolved grade of ingroup taxa. In addition, the phylogenetic position of two species from section *Mutica* is also unresolved.

The main *Hopea* clade (labelled C, synapomorphies 3, bootstrap <50%) is non-monophyletic, as it excludes *H. aptera* (which falls within the main *Shorea* clade (M)). It also includes four *Shorea* species– *S. richetia, S. multiflora* and *S. faguetiana* from section *Richetioides* and *S. javanica* from section *Anthoshorea*. The *Hopea* clade is defined by decreasing size of flower bud and nut, as well as increasing filament length. Most of the species from section *Dryobalanoides* are located in a *Hopea* subclade (labelled H, synapomorphies 5, bootstrap <50%), sister to a grouping containing both *Hopea* and *Shorea* species (labelled I, synapomorphies 4, bootstrap <50%). This subset of section *Dryobalanoides* (H) may be potentially monophyletic. However section *Dryobalanoides* as a whole is non-monophyletic, as two other species from this section (*H. nervosa* and *H. subalata*) form a group within clade E (synapomorphies 5, bootstrap <50%), while three more (*H. pierrei*, *H. dryobalanoides* and *H. mengerawan*) form a grade outside clade G. Most species of section *Hopea* do not group together either but show strong paraphyly, with its members distributed as single species or small groups throughout the *Hopea* clade.

The results of the first morphological analysis thus indicate that all the sections in *Hopea* and *Shorea*, apart from those that are monotypic, appear to be paraphyletic.

3.8.2 Exclusion of continuous characters

The second analysis excluded the 13 continuous characters but retained the original 84 taxa. The heuristic search found one optimal tree with a length of 317 steps, CI of 0.16, HI of 0.84, RI of 0.47 and RC of 0.08. This tree is presented in Figure 3.39.

The ingroup (A) is well defined (synapomorphies 6, bootstrap 78%), consisting of a "core" ingroup clade (labelled B) and a grade of 17 species of *Shorea* plus one species of *Hopea* falling outside this clade. The species in the grade possess the simple paniculate form of Total Inflorescence Unit, secund flowers pointing upwards, hairs on the anther connective appendage, a stylopodium and fruit pedicel, as well as a "*Shorea*-type" fruit calyx. These species falling outside the "core" ingroup clade include all the members of *Shorea* section *Shorea* included in the analysis, except for *S. guiso*. The grade also includes two species from *Shorea* section *Mutica*, two species of *S.* section *Doona*, three species of *S.* section *Richetioides*, one species of *S.* section *Anthoshorea* and *Hopea aptera*. Five of these species form a subclade (labelled C, synapomorphies 2, bootstrap <50%).

The "core" ingroup clade (labelled B, synapomorphies 3, bootstrap <50%) consists of two further groupings (labelled E and F) that does not correspond to the traditional generic divisions nor to the timber groupings. The paraphyly of *Shorea* is clearly shown by the placement of *Shorea* in both clades (E and F). All of the putative sister taxa included in the analysis are grouped within this core clade and are included within *Shorea*, suggesting their closer relationship to *Shorea* than to *Hopea*.

The first clade within the "core" ingroup (E, synapomorphies 0, bootstrap <50%) has further subgroupings (labelled G and H), with *Shorea selanica* being the sister to these two groups. Clade G (synapomorphies 0, bootstrap <50%) includes some species of *Shorea* section *Richetioides* and all the *Hopea* species used in the analysis,

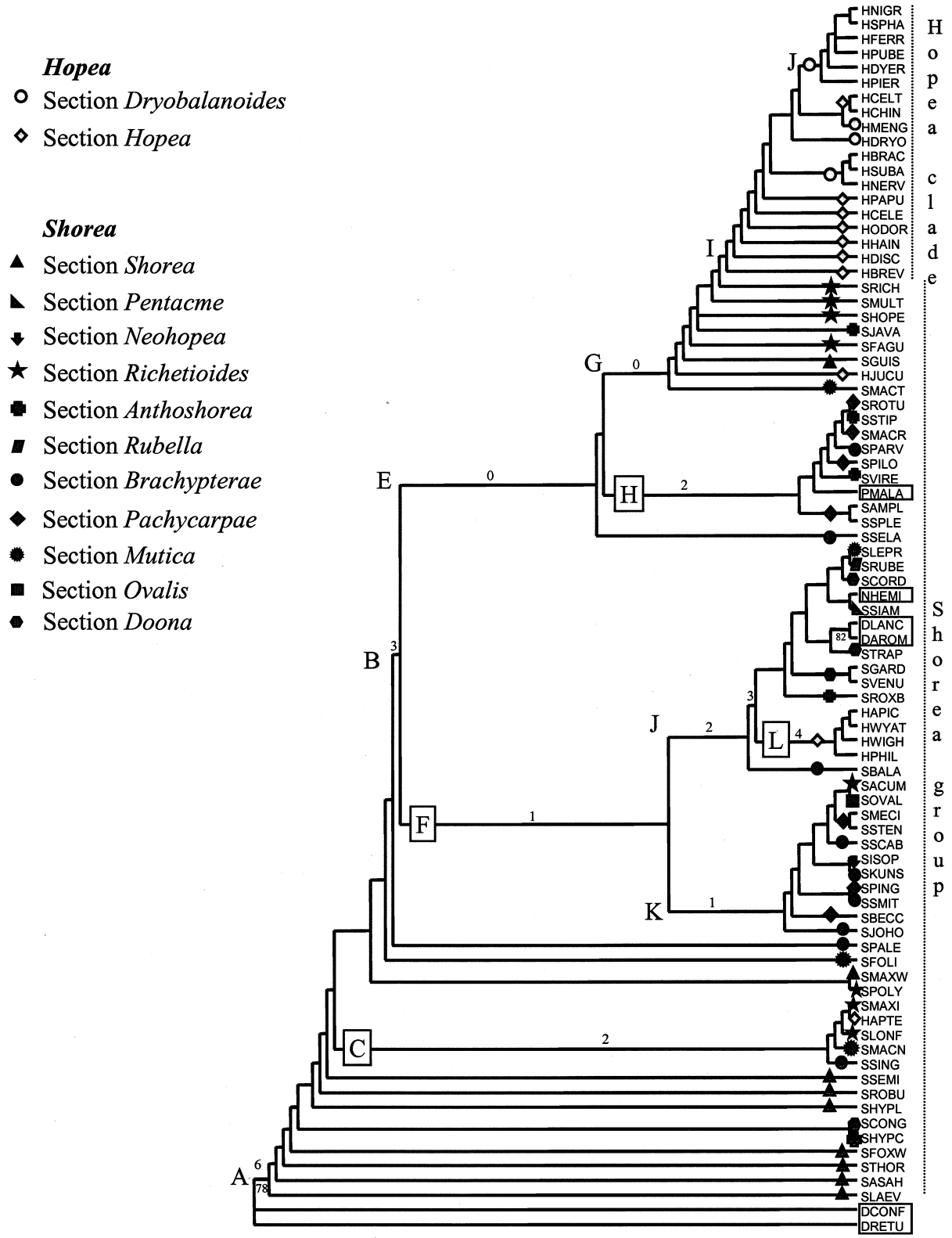


Figure 3.39 Single most parsimonious tree obtained from cladistic analysis of the morphological data set after excluding continuous characters, using 27 characters and 84 taxa. Numbers above the branches are branch lengths, and bootstrap values of 50% and greater are shown below. Taxon names in boxes are the putative outgroups.

except for a small group from section *Hopea* (clade L) and H. *aptera*. The final clade (I, synapomorphies 3, bootstrap <50%) includes only species of *Hopea*. However, neither of the two sections of *Hopea* are present in this clade as resolved monophyletic groupings. Clade I is a collection of small groups and single species from both sections. The majority of species from section *Hopea* are placed in a grade outside a clade of species primarily from section *Dryobalanoides*. Section *Hopea* thus appears to be strongly paraphyletic, given the placement of some species within clades J, C and G. The remaining species in clade G reflect single species lineages or form small groups, including *Hopea jucunda* and species from *Shorea* sections *Richetioides*, *Anthoshorea*, *Shorea* and *Mutica*.

The sister group to clade G (labelled H, synapomorphies 2, bootstrap <50%) comprises a variety of species including *Parashorea malaanonan* and taxa from *Shorea* sections *Pachycarpae*, *Anthoshorea* and *Brachypterae*.

The second major clade within the core ingroup (labelled F, synapomorphies 1, bootstrap <50%) consists of two subgroups, J and K. Clade J is a variable group containing three of the putative sister taxa to *Hopea* and *Shorea* (*Neobalanocarpus heimii* and two *Dryobalanops* species. It also incorporates four species from *Hopea* section *Hopea* in addition to species from *Shorea* sections *Mutica*, *Rubella*, *Pentacme*, *Doona* and *Anthoshorea*. Two of the species from Section *Doona, S. gardneri* and *S. venulosa*, form an apparently monophyletic group. Another group of interest within clade J is a clade of four species from *Hopea* section *Hopea* (labelled L, synapomorphies 4, bootstrap <50%). The second subgroup (labelled K, synapomorphies 1, bootstrap <50%) is composed of four members of *Shorea* section *Brachypterae*, four species from *S.* section *Pachycarpae*, one species of S. section *Richetioides* and the monotypic sections *Ovalis* and *Neohopea*.

Overall, the topology obtained from the second analysis of morphological data (after excluding the continuous characters) suggests that *Hopea* and *Shorea* are non-monophyletic.

3.8.3 Exclusion of inflorescence characters

A third analysis was performed after excluding the five characters scored from the inflorescence (characters 6–10) from the data set. The heuristic search found 2 optimal trees with the length of 667, CI of 0.24, HI of 0.76, RI of 0.47 and RC of 0.11. One of these trees (#1, figure 3.40) was selected for discussion.

A branch defined by six synapomorphic changes (labelled A, bootstrap 75%) separates *Dipterocarpus confertus* and *D. retusus* from the ingroup. The overall topology is to some extent similar to that of the cladogram obtained after excluding continuous characters (Fig. 3.39).

The ingroup (A) consists of 12 divergent *Shorea* and *Hopea* species, and the "core" clade (B) containing a number of subgroupings. The species in the ingroup (A) share apomorphic changes relating to the length of leaf lamina and flower bud, the presence of a stylopodium and a "*Hopea*-type" fruit calyx. The species that are not included in the "core" ingroup are *Hopea aptera* and *Shorea* species mainly from sections *Shorea*, *Richetioides*, *Mutica*, and *Doona*. Eight of these species, including *H. aptera*, form a group (labelled D) on the basis of one synapomorphic change, which is the length of anther connective appendage.

The "core" ingroup clade (labelled B, synapomorphies 6, bootstrap <50%) consists of a number of subgroupings. It contains species of both *Shorea* and *Hopea*, and all of the putative sister groups included in the analysis. The clade is defined by synapomorphic changes relating to lengths of the flower buds, petals, filaments and pistils and by a broad connective appendage.

One of the larger subclades within the "core" ingroup (labelled H, synapomorphies 2, bootstrap <50%) consists of most of the species of *Hopea* included in this analysis, in addition to seven species of *Shorea* from sections *Richetioides*, *Mutica*, *Shorea* and *Pachycarpae*. The species in this main *Hopea* clade (H) possess synapomorphies relating to the width of the leaf and size of the flower bud. Within clade H, several species form an evolutionary grade sister to a clade containing most of the species of *Hopea* included in the analysis (labelled I, synapomorphies 1, bootstrap <50%). The

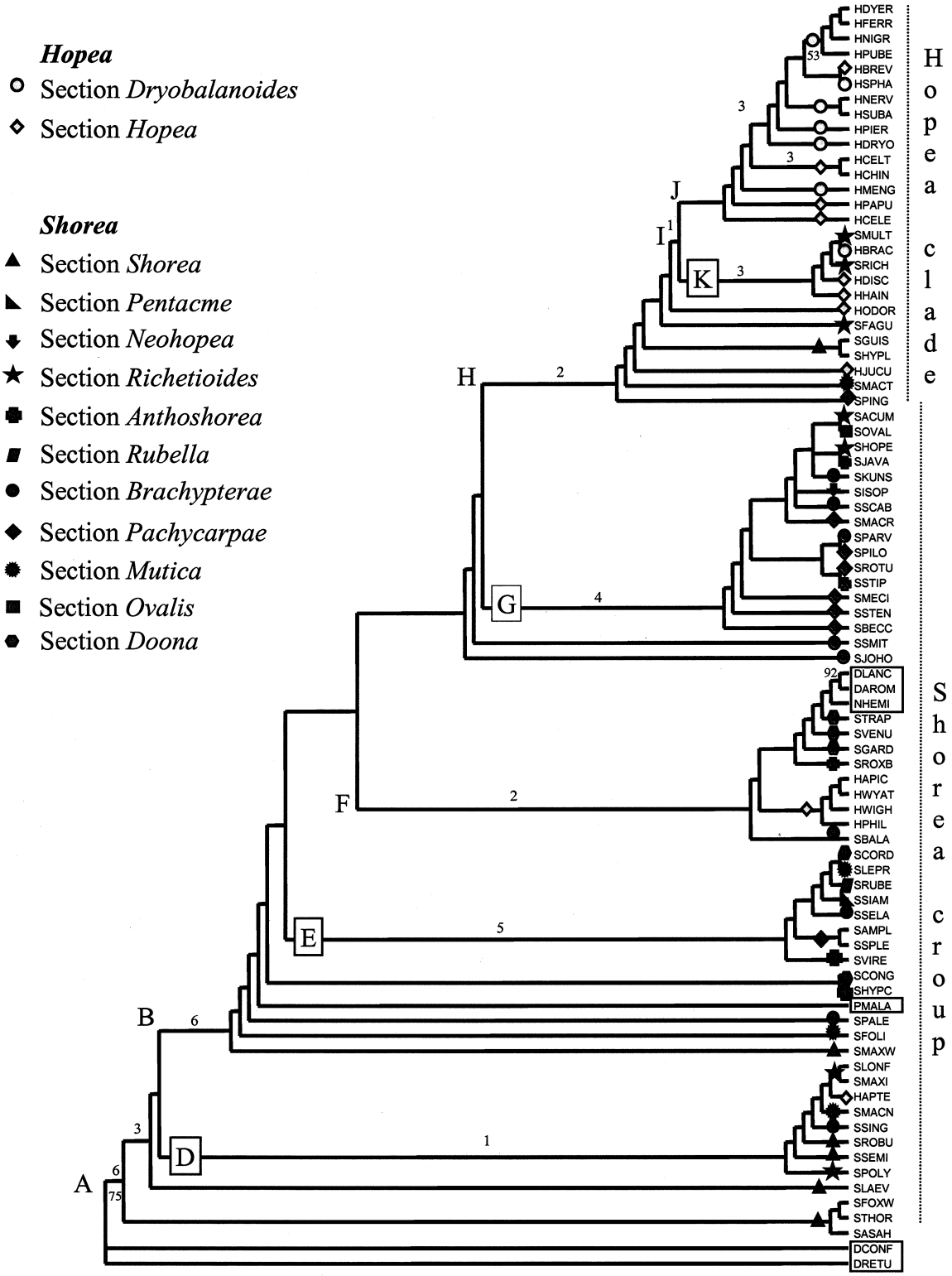


Figure 3.40 One of the most parsimonious trees obtained from cladistic analysis of the morphological data set after excluding inflorescence characters, using 35 characters and 84 taxa. Numbers above the branches are branch lengths, and bootstrap values of 50% and greater are shown below. Taxon names in boxes are the putative outgroups.

clade excludes *H. jucunda*, *H. apiculata*, *H. wyatt-smithii*, *H. wightiana*, *H. philippinensis*, *H. odorata* and *H. aptera* (all from *H.* section *Hopea*), but includes three *Shorea* species from section *Richetioides*. Within clade I, two further groupings (J and K) appear. All the species of *H.* section *Dryobalanoides* form a monophyletic group (synapomorphies 3, bootstrap <50%) within clade J, with the exception of *H. mengerawan*. The second grouping (K) contains the remaining members of section *Hopea*, as well as *Shorea multiflora* and *S. richetia* from *S.* section *Richetioides*).

The sister group of clade H (labelled G, synapomorphies 4, bootstrap <50%) contains the majority of taxa from *Shorea* section *Pachycarpae* included in this analysis. The clade also includes three species from *S.* section *Brachypterae,* two species from *S.* section *Richetioides* with one species from *S.* section *Anthoshorea*, as well as the monotypic *Shorea* sections *Ovalis* and *Neohopea.* Although seven species from section *Pachycarpae* fall within this group, they do not form a clade and instead are scattered throughout group G. This group is defined by prominent stipule scars, increasing length of the anthers, loss of hairs on the ovary, and length of the nut.

Clade F (synapomorphies 2, bootstrap <50%) is a variable group consisting of three of the putative sister taxa in addition to several species from *Shorea* sections *Doona*, *Anthoshorea* and *Brachypterae* and four from *Hopea* section *Hopea*. Within this grouping, species of *Hopea* section *Hopea* are united in a clade. The putative sister species within clade F (*Dryobalanops lanceolata*, *D. aromatica* and *Neobalanocarpus heimii*) form a monophyletic group in this analysis (synapomorphies 3, bootstrap <50%), and their position suggests that they may be part of the ingroup rather than being sister taxa.

The last clade of interest within the ingroup (labelled E, synapomorphies 5, bootstrap <50%), is also a variable group consisting of species from *Shorea* sections *Doona*, *Mutica*, *Rubella*, *Pentacme*, *Brachypterae* and *Anthoshorea*. Two species of section *Pachycarpae* form a pairing but none of the six sections of *Shorea* represented in this clade appear to be monophyletic, with the exception of the monotypic sections *Ovalis* and *Rubella*. Clade E is defined by synapomorphic changes relating to the size of the flower bud and the nut.

In summary, this analysis of morphological characters (which excluded those taken from the inflorescence) also indicates that *Hopea* and *Shorea* are not monophyletic, and that the traditional infra-generic groupings are also broadly non-monophyletic.

3.8.4 The "core" data set

Five taxa that could not be scored for a large number of the characters, and which thus may have presented a problem of missing data, were excluded from the dataset. After the analysis described above, the heuristic search found one optimal tree (shown in Figure 3.41). This tree has a length of 739 steps. The consistency index (CI) was fairly low at 0.24, but the retention index (RI) was relatively high at 0.46. A homoplasy index of 0.76 and rescaled CI of 0.11 suggests that, although there is a large amount of homoplasy in the data set, most of the characters used are synapomorphic features.

This analysis separated *Dipterocarpus retusus* and *D. confertus* from the rest of the taxa, and the ingroup (labelled A, synapomorphies 6, bootstrap 78%) consists of *Hopea*, *Shorea* and all putative sister taxa. This group is defined by possessing the simple paniculate form of Total Unit in the inflorescence, secund flowers pointing-upwards, the presence of hairs on the anther connective appendage and fruit pedicel, the size of the flower bud as well as the "*Shorea*-type" of fruit calyx. The ingroup consists of the "core" ingroup (B) and 11 species of *Shorea* and *Hopea aptera* that are excluded from the "core" clade. The *Shorea* species which fall outside the "core" ingroup include all the members of section *Shorea* included in the analysis, *S. longiflora* (from section *Richetioides*), *S. macrantha* (from section *Mutica*) and *S. thorelii* (from section *Doona*) and *Hopea aptera*.

The core ingroup clade (labelled B, synapomorphies 3, bootstrap <50%) consists of two further subgroupings, C and D. Both of these subgroups contain species of *Shorea*, *Hopea* and at least one of the putative sister taxa, thus suggesting that the two genera are not monophyletic. The "core" ingroup clade is defined by synapomorphic changes in the size of the leaf, flower bud, connective appendage and nut, as well as a differing number of stamens, shape of the anther connective appendage and the presence of a glabrous style remnant.

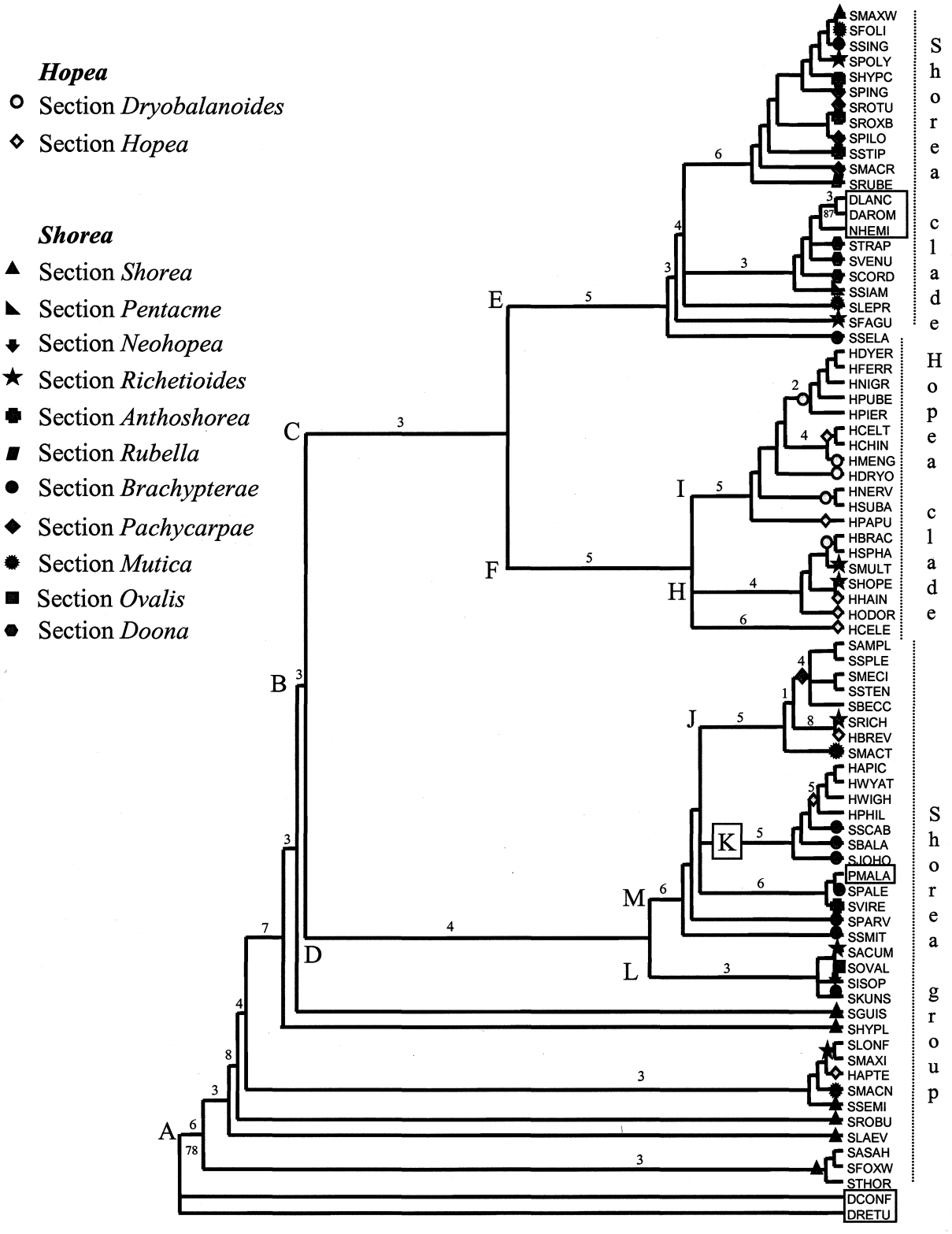


Figure 3.41 Single most parsimonious tree obtained from cladistic analysis of the morphological data set after excluding five taxa with a large amount of missing data, using 40 characters and 79 taxa. Numbers above the branches are branch lengths, and bootstrap values of 50% and greater are shown below. Taxon names in boxes are the putative outgroups.

The first major clade in the "core" ingroup (C) consists of two large sister groupings, and each of these groupings consists primarily of species from only one of the two genera. The first group (labelled E, synapomorphies 5, bootstrap <50%) consists mainly of *Shorea* taxa and includes species from sections *Pachycarpae*, *Mutica*, *Brachypterae*, *Doona*, *Anthoshorea*, *Rubella* and *Pentacme*. Three of the putative sister taxa also form a clade (synapomorphies 3, bootstrap 58%) within this group. The "*Shorea* clade" (E) is defined by synapomorphic changes in increasing size of the anthers, pistil and style and by the absence of a stylopodium. The inclusion of *Dryobalanops lanceolata*, *D. aromatica* and *Neobalanocarpus heimii* within the ingroup again suggests that these putative sister taxa might be considered to be part of *Shorea*.

The second major group within clade C (labelled F, synapomorphies 5, bootstrap <50%) is a polytomy containing two groups of *Hopea* species and *H. celebica*. This *Hopea* clade is defined by the decreasing size of the petals and nut, increasing length of the connective appendage and the "*Hopea*-type" of fruit calyx. One of the two groups in the polytomy (labelled I, synapomorphies 5, bootstrap <50%) consists largely of species from *H.* section *Dryobalanoides*, but the section is non-monophyletic due to the exclusion of two of its species and the inclusion of three species from section *Hopea*. The second group of *Hopea* (labelled H) that is involved in the polytomy is a variable clade comprised of two species from section *Hopea*, two from section *Dryobalanoides* and two from *Shorea* section *Richetioides*.

The second clade of the "core" ingroup is labelled D (synapomorphies 4, bootstrap <50%). This clade includes the majority of species from *Shorea* sections *Pachycarpae* and *Brachypterae* that were included in the analysis. The remaining species of *Shorea* sections *Richetioides*, *Mutica*, *Anthoshorea* and *Neohopea* are also included, in addition to the monotypic *S.* section *Ovalis* and four species from *Hopea* section *Hopea*. Clade D is defined by shortening of the flower bud and anther connective appendage and by increasing filament length. Within clade D, five species from *S.* section *Pachycarpae* form a polytomy (synapomorphies 4, bootstrap <50%). The four *Hopea* species also form a subclade (synapomorphies 5, bootstrap <50%), although the sister taxon to this group is a member of *S.* section

Brachypterae. *Shorea*'s putative sister taxon, *Parashorea malaanonan*, is also included within clade D, suggesting that it may be part of *Shorea*. This clade again suggests that the infra-generic divisions previously used for *Shorea* and *Hopea* are not monophyletic groups.

In summary, the cladogram obtained from analysis of morphological characters after excluding five taxa to minimise the amount of missing data suggests that both *Shorea* and *Hopea* are non-monophyletic groups.

3.9 Discussion

3.9.1 The information content of the morphological data set

Several classification systems for Dipterocarpaceae which considered the taxonomic placement of *Hopea* and *Shorea* have been put forward in the last 200 years. Most of the earlier classifications were based on only a few characters (Gaertner, 1805; Roxburgh, 1811 in Ashton, 1982). Subsequent classifications have been more elaborate and have included more characters (Heim, 1891; Pierre, 1892; Symington, 1943). However, most of these systems did not include explicit character analysis and it is therefore often unclear which characters supported the recognition of particular groups, such as infra-generic divisions. The main problem is that the traditional classifications usually focussed on different "key" characters in different parts of the genus, making it difficult to assess how characters interact in the genera as a whole. Ashton (1982) recognised only a single character that distinguished *Hopea* and *Shorea*, which was the number and development of fruit wings. The present study included detailed investigations of characters defining the genera and their infra-generic groupings.

As with many cladistic analyses of taxa that contain many groups of species, most of the groups are not supported by very many discrete characters, simply because the ratio of characters to taxa is small (Freudeinstein and Rasmussen, 1999). This was exemplified here, as the four analyses performed resulted in less than robust topologies. Even though all the analyses resulted in only one or two most parsimonious trees (MPTs), the skewness test (results not shown) indicated that eliminating some characters weakened the phylogenetic signal. Another indication of

this problem is that most of the branches in the trees obtained are not supported by many synapomorphic characters.

The resolution of the topologies obtained from analysis of equally-weighted characters has a distinct pattern: the genera may largely form monophyletic groups, but the infra-generic divisions are often poorly or not resolved. The groups that are largely resolved (e.g. *Hopea* section *Dryobalanoides*—see Figure 3.38, clade F and Figure 3.41, clade I) reflect patterns of previous classifications, regardless of the taxonomic rank assigned to them by the authors. This suggests that the previous classifications often placed weight on evolutionarily informative characters at this lower taxonomic level. However, the results of this study do not always agree with the previous classifications—for example, comparative development of fruit calyx did not separate *Hopea* and *Shorea* in the cladistic analyses. This character thus does not appear to be diagnostic at the generic level, nor is it useful as a cladistic character, mainly because it has a high level of homoplasy.

3.9.2 Effect of the continuous characters on the robustness of the resultant phylogeny

In the analyses presented above, the quantitative characters that have previously been asserted by many taxonomists to be "non-cladistic" provided greater phylogenetic resolution than that obtained when they were excluded. This was evident from the results of the three analyses that included continuous characters. These show that some of the quantitative characters define major lineages—in particular, the "core" ingroup and the *Hopea* and *Shorea* clades (Figures 3.38, 3.40 and 3.41).

The tree obtained from the dataset which excluded all continuous characters (Figure 3.39) had the lowest CI and highest HI of all four analyses. The low CI may be related to the number of changes within the continuous characters. The number of state changes is positively correlated with the number of character states. A greater number of possible states naturally provides more possibilities for character state change. Examination of the quantitative characters in the complete data set (characters 2, 3, 11, 12, 17, 21, 23, 24, 28, 30, 32, 37 and 38) shows that these characters change frequently throughout the tree and that these changes are often autapomorphic.

Reducing several multistate characters into binary characters will decrease the number of states but increase the number of characters overall. This is likely to decrease the number of state changes and increase the CI. Thus, it is difficult to ascertain whether quantitative characters will provide better resolution or result in a lower CI for the data set. Instead, including the quantitative characters merely added more characters into the data set.

3.9.3 Effect of the inflorescence characters on the robustness of the resultant phylogeny

The cladistic analysis of the dataset excluding inflorescence characters yielded a different topology (Figure 3.40) from that obtained in the first and fourth analyses (Figures 3.38 and 3.41), even though the heuristic search in these latter analyses yielded only two and one most parsimonious trees respectively. Excluding the inflorescence characters resulted in a higher CI than those obtained from analyses that included these characters (Figure 3.38 and 3.39). The inflorescence features incorporated here (characters 6–10 in Table 3.1) also have a high level of homoplasy. Among these five characters, only the type of TU and the orientation and directional growth of individual flowers on the inflorescence axis appear to change along branches defining the larger lineages rather than within the terminal taxa.

3.9.4 Effect of the missing data on the robustness of the resultant phylogeny

The exclusion of five taxa with a large amount of missing data did, to some degree, change the robustness of the resultant trees. This can be deduced from the lower CI of the topologies obtained when the missing data were included (Figure 3.38 and 3.39). Exclusion of the missing data resulted in a higher CI and the heuristic search only found one optimal tree. Hence, exclusion of the taxa with missing data produced a more robust topology than that obtained in analyses where these taxa are included. A large amount of missing data seems to have increased the level of homoplasy in the dataset.

The topology resulting from the analysis with troublesome taxa excluded (Figure 3.41) is different from those seen in the first three analyses (Figure 3.38, 3.39 and 3.40). The ingroup consists of a greater number of hierarchical groupings of *Hopea* and *Shorea* species when the taxa with missing data are excluded.

3.9.5 Evolution of inflorescences of the taxa used

Parsing the inflorescence into hierarchical units may provide better insight into the positional homology of the inflorescence structure and the evolutionary processes implied by the changes in its smaller units. In fact, the use of inflorescence information in systematics is often unclear because the level of organisation is not specified, as exemplified by the general terminology of "inflorescence terminal or not" in contrast to the description used here of "termination of Inflorescence Unit". This coding therefore specifies topographic information in a defined hierarchical context.

Only one character, the type of TU (#8) changing to the synapomorphic botryoid, separates the ingroup and the outgroup. The state changes of the inflorescence characters in terminal lineages tend to be autapomorphic for those taxa. Even though it is clear from the examination of these data that the botryoid TU type seems to have evolved early in the history of the ingroup, it is still difficult to infer the most primitive type of Dipterocarpaceae inflorescence because the sub-families Monotoideae and Pakaramoideae are not included. Two basic types of inflorescence are found in Dipterocarpaceae: racemose and paniculate. Weberling (1989) considered a raceme as a type of simple inflorescence, while a panicle is a compound inflorescence. A panicle is basically the "extended" form of a raceme or a "branching raceme". The *Dipterocarpus* species included within these analyses are the only taxa with raceme-based inflorescences and the results obtained here tend to suggest *Dipterocarpus* has a primitive type of inflorescence.

Admittedly, it is common to describe the basic Dipterocarpaceae inflorescence as a "raceme" (Ashton, 1982), since the inflorescence consists of a series of individual pedicellate flowers. However, the most distinct feature of a botryoid is the determinate apex, whereas a raceme is always marked by the "indeterminate" state of its inflorescence. Hence the term given to *Dipterocarpus* inflorescences is botryoid, which is basically a "determinate raceme" or "impoverished raceme" *sensu* Weberling (1989). The best example of this form is found in *Dipterocarpus retusus* (e.g. CANB specimen 109473).

Within the ingroup, "panicle" seems to be the most common inflorescence type found in Dipterocarpaceae and in fact this form is found throughout the genera examined, except for *Dipterocarpus*. The statement from Weberling (1980) that a panicle is a "branching raceme" is obvious when examining the typological change from the "botryoid" of the *Dipterocarpus* spp. to the "paniculate" type of the other genera. It is also obvious that several types of this form can be recognised, with their differences based only on the branching system. An individual flower of *Dipterocarpus* is therefore homologous to the basic unit (BU) in the other genera.

It can be concluded that the IU of basic and derived raceme type is the only definite inflorescence character from which the evolution of the inflorescence can be inferred. Other characters that seem to have evolved recently are unique and define only the terminal taxa.

3.9.6 Evolution of comparative development of fruit calyx

A closer examination of the synapomorphic state changes within the fruit calyx development character (#35) may provide some insights into its evolutionary history (Figure 3.42). *Dipterocarpus* spp. were used as the outgroup and to root the trees, and these taxa exhibit state 1 of character 35 (two long and three short fruit wings). This state was thus assumed to be plesiomorphic in the analyses. However, inclusion of other *Dipterocarpus* spp. that have subequal fruit wings (state 0) is necessary to clarify the plesiomorphic condition.

Over the course of evolutionary history, the fruits may have evolved into different forms with changes in the fusion of the calyx base and the relative development of the wings. This includes the change in fusion of the calyx tube, which differentiates Tribe Dipterocarpeae (with a valvate base) from Tribe *Shoreae* (imbricate base). The presence of an imbricate calyx tube within the ingroup may therefore be a unique derived apomorphic state for the taxa included in this study. Another change is the comparative development of the fruit calyx, which has resulted in the elongation of another sepal. Hence, taxa which are the sister groups to the remainder of the ingroup (*S. trapezifolia, S. venulosa, S. polyandra* and *S. ovalis*) possess three long fruit calyx wings and two short ones (state 2). This analysis suggests this state is the most primitive within the ingroup.

An unrooted cladogram topology (not shown) indicated that single species lineages of *Dipterocarpus* spp. were collapsed with the branch defining the ingroup, which suggests that no clear ancestral states have been defined. Hence, there is still a possibility that the ancestor of the ingroup may have possessed state 2, which was then lost in *Dipterocarpus* spp. To test this hypothesis, other members of Tribe Dipterocarpeae or members of other sub families would need to be incorporated into analyses to identify the plesiomorphic condition of every fruit character that is included within this study.

The cladogram in Figure 3.42 indicates that there are two major divergent lineages. One group (the *Shorea* clade) maintains state 2, with three long and two short wings on the fruit calyx. The second lineage (the *Hopea* clade) may have undergone a reversal, with the regaining of the plesiomorphic condition (state 1, with two long and three short fruit wings).

The analysis suggests that the third state of this character, equal/subequal fruit wings, is the derived condition relative to state 1 and 2 in the Tribe Shoreae. The tracing of the fruit calyx development character on the cladogram (Figure 3.42) indicates this relatively derived state may have arisen within both the *Hopea* and *Shorea* clades. Also, the placement of the putative sister taxa within relatively terminal lineages suggesting their recent origin may also have been the result of the parallel evolution. The parallelism of the third state within the *Hopea-Shorea* clades suggests that there may have been a common ancestor that possessed this state during the early divergence. This can be inferred from the position of *S. asahii*, which possesses equal/subequal fruit calyx wings, as the sister taxon to the majority of the ingroup (Figure 3.42).

3.9.7 Phylogenetic inferences of the putative outgroups, sister taxa and the ingroup

Several species from various genera of Dipterocarpaceae were selected as the sister groups in this morphological study. These are species of *Dipterocarpus*, *Dryobalanops*, *Parashorea* and *Neobalanocarpus heimii*. *Dipterocarpus* is considered to be relatively distantly related to the ingroup, and is therefore likely to have greater divergence from the ingroup compared to the other taxa.

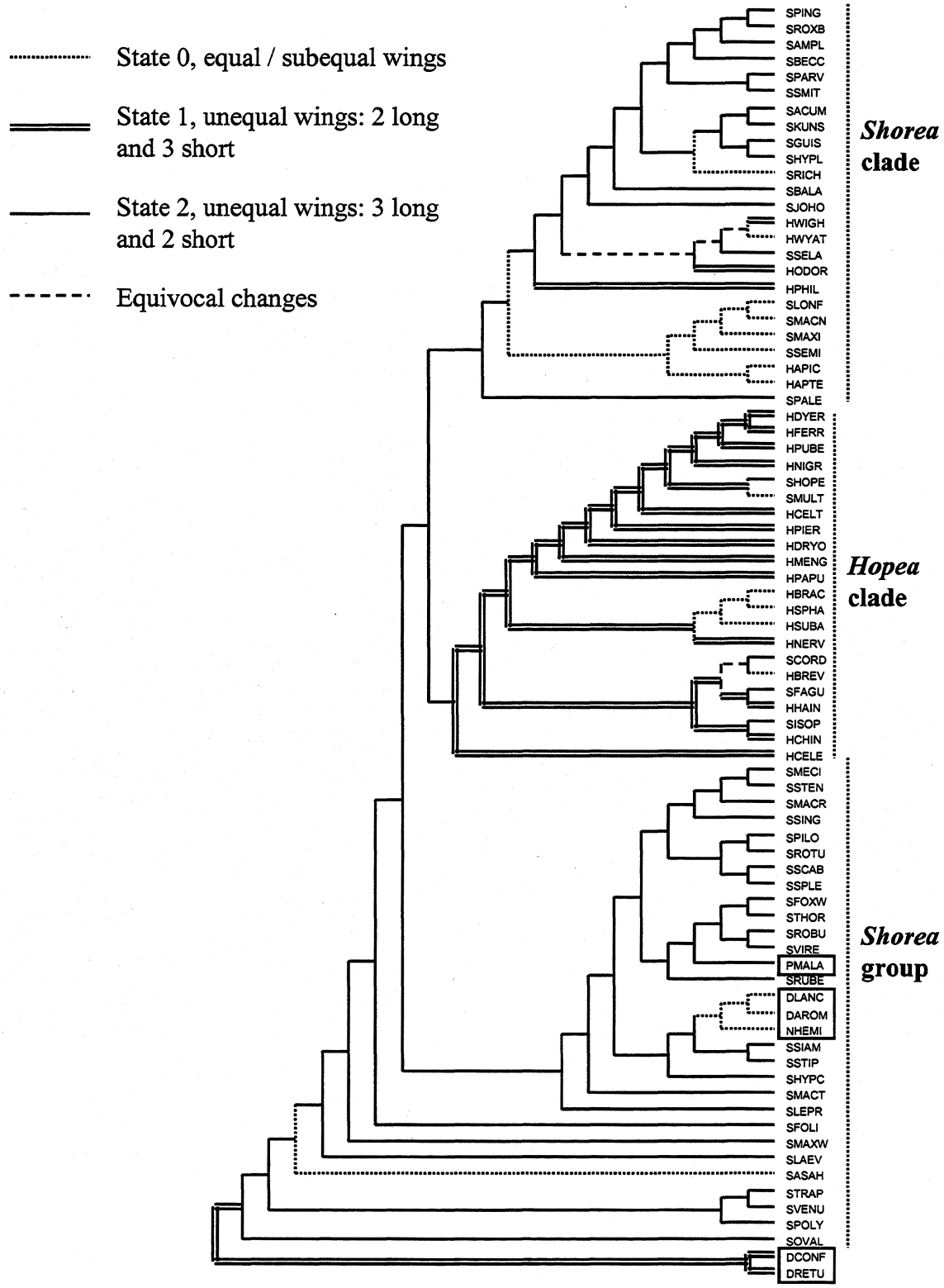


Figure 3.42 Cladogram showing changes in character #35 (comparative development of fruit calyx) obtained from cladistic analysis of the "core" morphological data of 40 characters and 79 taxa. Taxon names in boxes are the putative outgroups.

The four analyses of morphological data undertaken consistently placed the putative sister taxa (*Dryobalanops*, *Parashorea* and *Neobalanocarpus*) within clades of *Shorea* species, suggesting that they are actually part of the ingroup. This result is not in agreement with the results of previous molecular phylogenetic studies of Dipterocarpaceae using chloroplast DNA (Tsumura *et al.,* 1996; Kajita *et al.,* 1998; Dayanandan *et al,* 1999), which placed *Parashorea* as the sister to *Shorea* and *Neobalanocarpus* as the sister to *Hopea*. These putative sister taxa in the past have been placed in the same tribe (Shoreae) as *Hopea* and *Shorea*, and also share the same base chromosome numbers. However, they can be distinguished from *Hopea* and *Shorea* on the basis of comparative development of the fruit calyx. *Dryobalanops* has five equal (or subequal) spatulate wings and *Neobalanocarpus heimii* possesses five short triangular fruit wings. Nonetheless, results from this study show that character 35 (comparative development of fruit calyx) only undergoes a synapomorphic change within the lineage uniting the two *Dryobalanops* spp. with *N. heimii*, but not within the overall *Shorea* clade. The clade uniting these three putative sister taxa is defined by synapomorphic changes in characters 11 (length of flower bud), 17 (length of petal), 23 (length of anther), 30 (length of pistil), 35 (comparative development of fruit calyx) and 38 (width of nut).

The putative sister genus to *Shorea*, *Parashorea* (represented in this analysis only by *P. malaanonan*), falls within the *Shorea* clade D in the analysis of the "core" data set (Fig. 3.41). This genus has many similarities with *Shorea*, including the possession of two different types of comparative development of the fruit wings, and it is not yet clear which characters distinguish this genus from *Shorea*. Maury (1978 in Ashton, 1982) suggested that *Parashorea* has embryo and seedling characters which render it distinct from *Shorea*. However, results from the present analyses indicate that *Parashorea* is united with the other members of clade D on the basis of synapomorphic changes in the width of the leaf lamina (character 3), length of the flower bud (11), width of the flower bud (12), persistence of bracteoles (17), length of the pistil (30), and the length of the style (32).

3.9.8 Phylogenetic relationships of *Hopea* and *Shorea*

Neither *Hopea* nor *Shorea* appears to be monophyletic, with *Hopea* largely nested within *Shorea*. In two of the analyses (Figs 3.39 and 3.40) *Hopea* is nested within

Shorea, with some of the *Shorea* species forming a grade sister to *Hopea*. This may suggest that *Hopea* is relatively of more recent origin than that of *Shorea*. In the strict sense, some of the morphological characters that define *Hopea* as a genus may have been derived synapomorphic states from *Shorea*, even though this may include homoplasious changes in the form of parallel and reverse evolution of certain character states.

All analyses consistently place *Hopea* section *Dryobalanoides* nested within the Hopea clade, usually with species from section *Hopea* forming a grade outside this grouping. A number of the species from section *Dryobalanoides* also tend to form a monophyletic group in the cladistic analyses. The accepted classification defines this section on the basis of dryobalanoid leaf venation. This study confirms that this character may reflect a unique synapomorphy for this section but not for section *Hopea*. Its position in the cladograms suggests that this section may be a more recently divergent lineage than section *Hopea* or that it may even have originated from within section *Hopea*.

By contrast, *Hopea* section *Hopea* is a variable section with an equivocal phylogenetic position, as most of its members form single species lineages throughout the trees and four of its species tend to be included within *Shorea* (Fig. 3.39, clade L; Fig. 3.40, clade F; Fig. 3.41, clade K).

Shorea is clearly a variable genus. Results of the cladistic analyses derived from different sets of morphological characters do not reflect the groupings proposed in the existing infra-generic classifications, and no monophyletic sections are suggested by these analyses. Species from all the existing sections of *Shorea* were placed almost throughout the trees, with the exception of the monotypic sections *Pentacme*, *Rubella*, *Ovalis* and *Neohopea*. However, species of *S.* sections *Pachycarpae* and *Shorea* tend not to form larger groups, instead tending to occur as single species lineages. Other characters such as bark and wood anatomy may need to be incorporated into the analyses to obtain more resolution of the relationships of these sections.

However, there are some consistent phylogenetic positions reflected from all the analyses. The first is the phylogenetic position of *Shorea* section *Shorea* (Balau

group *sensu* Symington, 1943) as a grade of species which forms the sister group to the main part of the ingroup (Figs 3.38, 3.39 and 3.40). This may suggest the possession of plesiomorphic states in this section. This supports Maury-Lechon's argument (Maury-Lechon and Curtet, 1998) that *S.* section *Shorea* was one of the main taxa in sub-family Dipterocarpoideae from which new forms arise through diversification.

The phylogenetic position of *Shorea* section *Richetioides* is also of interest. This section is known as "yellow meranti" or "Damar Hitam" (Symington, 1943) and contains species with sub-equal calyx lobes, which has caused some debate as to whether it should be accorded generic status (Heim, 1891; Meijer and Wood, 1964; Maury-Lechon and Curtet, 1998). This section is clearly polyphyletic, with some of its members (*S. richetia*, *S. multiflora* and *S. hopeifolia*) forming a group with *Hopea* spp. and other species being scattered throughout the trees. Thus, the phylogenetic position of this section remains equivocal.

Shorea section *Anthoshorea* also has equivocal phylogenetic position and may be polyphyletic. The majority of the species in this section are included in the Meranti Pa'ang group (Symington, 1943). Results from the present analyses, however, do not confirm the integrity of this timber grouping for this taxon. Of all the sections/groups within *Shorea*, *Anthoshorea* has the most complex taxonomic history and this is reflected in the results of the present study. Maury-Lechon and Curtet (1998) considered that section *Anthoshorea* resembled other sections of *Shorea* such as *Doona* and *Pentacme*, as well as *Neobalanocarpus*, *Dryobalanops* and *Cotylelobium* on the basis of characters from the embryo, seedling and pollen surface. It is perhaps therefore unsurprising that these analyses indicate a close relationship of the members of this section to certain other groups/taxa of the ingroup clade.

The remaining sections within *Shorea* (i.e. sections *Mutica*, *Rubella*, *Brachypterae*, *Pachycarpae*, *Doona* and *Ovalis*) are placed within the Red Meranti group (Symington, 1943). Although none of these sections is strictly monophyletic, their members tend to be placed together in some larger groupings (Fig. 3.38, clade K; Fig. 3.39, clades F and H; Fig. 3.40, clades E, F and G; Fig. 3.41, clades E and D). Section *Pachycarpae*, represented in these analyses by nine species from a total of

ten, does not appear as a monophyletic group. This section is mainly characterised by its relatively large seed. However, reversal of the fruit size in this section may be a possible explanation, since this character is shown to be homoplasious. Section *Brachypterae* is also a non-monophyletic group. Most of its species included in the analyses form single species lineages within the *Shorea* clade. The remaining *Shorea* sections (with the exception of the monotypic sections *Ovalis* and *Neohopea*) represent non-monophyletic taxa with equivocal positions within the *Shorea* clade.

3.9.9 Taxonomic implications

A continuing debate among systematists is whether the results of cladistic analyses can be translated into taxonomy and thus used as natural classifications. A problem becomes evident when the results of cladistic analyses are incongruent with the classical taxonomy, even though the analysis was carried out using morphological characters. Morphological characters were utilised by classical taxonomists in order to classify and distinguish the taxa and such classifications do not necessarily reflect their evolutionary relationships.

Hopea and *Shorea* have been defined as separate genera using a single morphological character (based on character 35 in these analyses) in more traditional classifications. The cladistic analysis has shown that this character may have followed a different evolutionary course within each genus. Hence, this character may still be used to distinguish the two genera for taxonomic purposes.

The overall results from the cladistic analyses do reveal a close relationship between the two genera. However, a clear distinction between the two is difficult to identify. These difficulties arise mainly because all the characters used are not free from homoplasy, and the homoplasious changes in character states make it difficult to find diagnostic characters for certain groups. For example, the *Hopea* clade (Figure 3.38, clade H) is defined by synapomorphic state changes in the width of individual flower buds, the length of the nut, the absence of hairs on the ovary and in comparative development of the fruit calyx. Two of these are categorised as continuous characters. These are rather difficult features to use as diagnostic characters to define *Hopea* for the purpose of taxonomic identification. These characters can only be seen as reflecting evolutionary changes of their particular states (for instance, reduction of

the filament length from 0.7 to 0.5 mm) rather than discrete characters for defining a genus.

The only potential natural grouping within *Hopea* is section *Dryobalanoides*, so this section can be maintained. However, the taxonomic status of section *Hopea* cannot be deduced since it is obviously paraphyletic. It may require further detailed study by incorporating more characters to clarify and assign its taxonomic status.

With regard to the groupings within *Shorea,* the results of this study do not accord with the previously-suggested timber or infra-generic classifications (Meijer and Wood, 1964; Symington, 1943; Ashton, 1982). Further analyses incorporating more characters such as wood anatomy may be required to be clarify the relationships within this genus.

3.10 Conclusions

The results of the cladistic analyses of morphological data described above are at odds with the relationships proposed by previous taxonomists. In this study, *Hopea* and *Shorea* are not separated into two discrete genera as they are in traditional classifications. The present study shows that *Hopea* may have originated from within *Shorea*, with subsequent parallel evolution within the lineages.

The putative evolutionary relationships between *Hopea* and *Shorea* inferred from analyses of morphological data indicate that both genera are non-monophyletic. However, the morphological data exhibit high levels of homoplasy, which may bias any evolutionary inferences about the relationships of the genera. It is anticipated that greater resolution of the phylogenetic relationships of the two genera may be obtained by inclusion of data from an independent source. Hence, analyses of molecular data sets are presented in the next chapter and analyses of combined data follow in Chapter Five.

Appendix 3A Taxa selected for the morphological study

Species	Abbreviation	Herbarium number	Collection number	Location
Outgroup				
Dipterocarpus confertus Blume	DCONF	CANB 95949	Kost. 6379	E. Borneo
Dipterocarpu retusus Slooten	DRETU	CANB 132216		
		CANB 109473	Kost. 11197	Bogor B.G
		HUH	FRI 1353	FRIM
		HUH	Kost. 11197	Bogor B.G
		HUH	Dipter 88	W. Java
Dryobalanops aromatica C.F.Gaertn.	DAROM	CANB 00493972		New Guinea
		CANB 36017	A 3298	N. Borneo
Dryobalabops lanceolata Burck	DLANC	CANB 3217071		Sabah
Parashorea malaanonan Merrill	PMALA	CANB 105161	Sandakan 15508	Silam
		CANB 378021	A. 2832	Sandakan
		CANB 37860	A 2266	Sandakan
		CANB 37875	A. 896	Sandakan
		CANB 35949	A 1887	Sandakan
Neobalanocarpus heimii (King) Ashton	NHEMI	KEP 4454	FRI 28645	FRIM
		CANB 133924	Kep 79143	FRIM
		HUH	FRI 27613	FRIM
		HUH	FRI 13027	
		KEP 4433	FRI 22193	FRIM
Ingroup				
Genus *Hopea*				
Section *Dryobalanoides* sub-section *Dryobalanoides*				
H. pubescens Ridley	HPUBE	KEP 4238	10478	Pahang
		KEP 4215	5740	Pahang
***H. dryobalanoides* Miq.**	HDRYO	KEP 18428	FMS 49850	Pahang
		KEP 21000	FRI 27773	Pasoh
		CANB 37790		Lahad Datu
		BO-0114884		Sumatra
H. ferruginea Parijs.	HFERR	CANB 38964	4624	Sandakan
		CANB 37876	A 126	Lahad Datu
		HUH	San 83408	Sandakan
H. mengerawan Miq	HMENG	HUH	Sijo 36353	Singapore
		HUH	Sing 36187	Singapore
		HUH	Kost. 6752	Samarinda
		HUH	5986	Wanariset
		HUH	Singapore 36353	Singapore

Appendix 3A Taxa selected for the morphological study (continued)

Species	Abbreviation	Herbarium number	Collection number	Location
Genus *Hopea*				
Section *Dryobalanoides* sub-section *Dryobalanoides*				
H. pierrei Hance	HPIER	CANB 00491455	VII. B 50	
		BO-0114342	569 A	
		BO-0115595	17645	Philippines
		BO 0114755		Cambodia
H. dyeri Heim	HDYER	BO-0114718	9360	Kuching, Sarawak
		BO-0114720		
Section *Dryobalanoides* sub-section *Sphaerocarpae*				
H. bracteata Symington	HBRAC	HUH	SAN 83408	Sandakan
		HUH	13347	
		HUH	2225	Kuching
		HUH	13378	Sarawak
H. subalata Symington	HSUBA	KEP 4273	29770	Kuching
		KEP 4273	29770	
		KEP 4287	FRI 27606	FRIM
H. sphaerocarpa (Heim) Ashton	HSPHA	HUH	553	Sarawak
		HUH	15728	Sarawak
		HUH	15728	Sarawak
		HUH	554	Sarawak
H. nervosa King	HNERV	KEP 22402	SAN 1121800	Tengkulap, Tehupiu district
		KEP 22401	57304	FRIM
H. nigra Burck	HNIGR	CANB 103264		Bogor B. G
Section *Hopea* subsection *Hopea*				
H. celebica Burck	HCELE	BO-1257806		Sulawesi
		BO-1257805		Sulawesi
		BO-111261	bb 25.535	Celebes
H. aptera Ashton	HAPTE	HUH	BW 7414	West. N.G.
H. celtidifolia Kosterm.	HCELT	HUH	BW-6273	Irian Jaya
		CANB 95666		New Guinea
H. papuana Diels	HPAPU	CANB 180633		New Guinea
		CANB 51770	BW 2705	New Guinea
H. odorata Roxb.	HODOR	HUH	JIF mabel 94-14	Kanchapuri
		HUH		
		HUH		Annam
		HUH	Iwaling No 52	Peradeniya
		HUH	Iwaling no. 58	Peradeniya
		HUH	21657	Kedah
H. hainanensis Merrill & Chun	HHAIN	HUH	Lau. SK 3121	Hainan Is. Cha'ng King district
		HUH	65329	Hainan Is.
		HUH	27446	Hainan
H. discolor Thw.	HDISC	HUH	PSMh 2119	Permaoulk, Sri Lanka

Appendix 3A Taxa selected for the morphological study (continued)

Species	Abbreviation	Herbarium number	Collection number	Location
Genus *Hopea*				
Section *Hopea* subsection *Pierrea*				
H. wyatt-smithii Wood *ex* Ashton	HWYAT	HUH	BRUN 885	Brunei
		HUH	S-32275	Sarawak
		HUH	S-32288	Sarawak
		HUH	318	Sarawak
H. apiculata Symington	HAPIC	KEP 20998	77614	Bruas Forest Resort
		KEP 20999	KEP 80225	FRI, Kepong
		BO-0114737	KEP 77614	Perak, Malaya
		BO-0114734	Mohan 80225	FRI
H. philippinensis Dyer	HPHIL	HUH	9559	Philippines
		HUH	2099-2	Philippines
		HUH	21410	Luzon
		HUH	1706	Philippines
H. brevipetiolaris (Thw.)Ashton	HBREV	HUH	2053	Ceylon
		HUH	PS ashtor 2105	Peradeniya B.G.
H. jucunda Thw.	HJUCU	BO-1256727	Kost. 28351	Kelariya Rv. Sri Lanka
		BO-0032725	Kost. 28351	Kelariya Rv. Sri Lanka
H. chinensis (Merr.) Hand.-Mazz.	HCHIN	HUH	69802	Kwangtung Shruby
		HUH	22653	Kwangsi
		HUH	24288	Shangste, Kwangtung Border
		HUH	22223	Kwangtung
H. wightiana Miq. *ex* Dyer**	HWIGH	KEP 20997	76643	FRIM
		KEP 20996	KEP 80542	FRIM
		HUH	587	India
		HUH	Kepong Field no. 80225	FRI Kepong
		HUH	Kepong 76643	FRI Kepong
Genus *Shorea*				
Section *Shorea* subsection *Shorea*				
S. thorelii Pierre *ex* Laness	STHOR	HUH	2551	Burma
		HUH	36771	Indochina
		HUH	30059	Kamboja
		HUH	36772	Indochina
S. robusta A. DC.	SROBU	HUH	Sn	East Nepal
		HUH	179	East India
		HUH	Tonyo 8147	East India
S. seminis V. Slooten	SSEMI	CANB 37856	A 578	Sepilok, Sandakan
		CANB 38941	A 367	North Borneo
		HUH	7934	Sampit
		HUH	Uthman Ismani S.37049	Kuching
		HUH	S. 15576	Serawak
		HUH	SAN 21211	Sandakan

Appendix 3A Taxa selected for the morphological study (continued)

Species	Abbreviation	Herbarium number	Collection number	Location
S. guiso Blume	SGUIS	CANB 132320		Bogor B. G.
		CANB 111108	Kost. 6015	Kutai, Borneo
		CANB 132321	202	Sambas, Borneo
		BO-0107876	bb. 10.460	
		CANB 35966	A 3586	Sepilok
Genus *Shorea*				
Section *Shorea* subsection *Shorea*				
S. foxworthyi Symington	SFOXW	KEP 20995	76678	Jerangau
		KEP 20994	22180	FRIM
		BO-0114910	29041	FRI
S. hypoleuca Meijer	SHYPL	HUH S-63826		Sarawak
		HUH	Sarawak 15360	Sarawak
		HUH	SAN 36654	Sandakan
		HUH	SAN 16989	
Section *Shorea* subsection *Barbata*				
S. laevis Ridley	SLAEV	CANB 138840	Sand 26086	Sandakan
S. asahii Ashton	SASAH	HUH	S-57238	Sarawak
		HUH	S-32345	Sarawak
		HUH	S-41412	Kapit
		HUH	S41410	Kapit
***S. maxwelliana* King**	SMAXW	KEP 20992	Singapore 37680	Penang Hill
		BO-111622	bb. 22.338	Sumatra
		HUH	FRI 25032	Trengganu
		HUH	SAN 20948	Sandakan
Section *Pentacme**				
S. siamensis Miq.	SSIAM	HUH	J.F. Maxwell 94.279	Thailand
		HUH	2597	Thailand
		HUH	29102	Langkawi Is.
		HUH		Singapore B6
		HUH		Thailand
Section *Neohopea**				
***S. isoptera* Ashton**	SISOP	HUH	S 24690	Sarawak
		HUH	S 32659	Kuching
		HUH	41471	Sabah
		HUH	4394	Sandakan
		HUH	3018	Brunei
Section *Richetioides* Subsection *Polyandrae*				
S. polyandra Ashton	SPOLY	HUH	bb. 30201	Melawi
		HUH	Sandakan 15266	North Borneu
		HUH	S.24256	Marudi
		HUH	S.25298	Bukit Snibong

Appendix 3A Taxa selected for the morphological study (continued)

Species	Abbreviation	Herbarium number	Collection number	Location
Section *Richetioides* subsection *Richetioides*				
S. acuminatissima Symington	SACUM	CANB 327048	HP Nootebom 1260	Malaman, Sabah
		CANB 35879		Sandakan
		CANB 35908		Sandakan
***S. faguetiana* Heim**	SFAGU	CANB 118595	?	Sandakan
		HUH	Suih & Murry 24855	Sarawak
		HUH	FRI 28623	Pahang
		HUH	Ambri & Artin W 387	E. Borneo
		HUH	K. Sidaya 414	E. Borneo
S. longiflora (Brandis) Symington	SLONF	HUH	2441	Mt. Balawayah
		HUH	30455	Brunei
		HUH	2441	Mt. Balapau
		HUH	19425	Sarawak
S. multiflora (Burck) Symington	SMULT	CANB 105160	SAN 21306	Sandakan
		CANB 106867	Rastini 130	Bogor B. G.
		HUH	Jarvie 5101	Kalteng
		HUH	TL 1343	W. Borneo
		HUH	Jarvie 5320	Kalteng
S. richetia Symington	SRICH	KEP 20981	68699	Kuching
		KEP 20980	S.49990	Lundu
S. hopeifolia (Heim) Symington	SHOPE	CANB 00493368		Kalteng
		CANB 00493373		Kalteng
		CANB 105166	SAN 21497	Sandakan
S. maxima (King) Symington	SMAXI	KEP 5424	SA. 431	Pahang
		KEP 5402	19346	Pahang
Section *Anthoshorea*				
S. javanica Koord. & Valet	SJAVA	BO-1265083	bb. 20197	
		BO-1266302	bb. 8963	Teluk Betung
		HUH	bb. 8963	Sumatra
S. virescens Parijs	SVIRE	HUH	Jamie 5167	C. Borneo
		HUH	TL 1319	W. Borneo
		HUH	FRI 27639	FRI Kepong
		HUH	CLP no. 1553	Phillippines
S. hypochra Hance	SHYPC	HUH	Bejaud 300	
		HUH	94968	21.7.62
		HUH	15400	Silanyar
		HUH	Put-580	Thailand
		HUH	67812	MalayA

Appendix 3A Taxa selected for the morphological study (continued)

Species	Abbreviation	Herbarium number	Collection number	Location
***S. roxburghii* G. Don**	SROXB	KEP 20990	AZ. 4758	Perlis, Arau
		HUH	H12-108	FRIM
		HUH	LB&ECabbe 9665	Thailand
		HUH	KEP 99486	Kepong Arboretum
S. stipularis Thw.	SSTIP	HUH	PS Ashton 2330	Peradeniya
		HUH	PS Ashton2330	Peradeniya
Genus *Shorea*				
Section *Rubella*				
S. rubella Ashton	SRUBE	HUH	Sandakan A 1750	Sandakan
		HUH	F.21	
		HUH	S 15128	Sarawak
Section *Brachypterae* subsection *Smithiana*				
***S. smithiana* Symington**	SSMIT	CANB 119153	SAN 21483	Sandakan
		BO-0117217	bb.19.773	Nunukan
		BO-0117218	bb.19.985	Bokoi
		HUH	AA.1164A	East Borneo
		HUH	SAN-134957	Sandakan
		BO-0114328	bb.32.556	Kutai
		HUH	K. Sidayasa 435	Wanariset
Section *Brachypterae* Subsection *Brachypterae*				
S. balangeran Burck	SBALA	CANB 106869	Rastini 136	Bogor Botanic Garden
S. johorensis Foxw.	SJOHO	BO-0115593	196E 3P 950	Palembang
		BO-0115592	bb. 14596	Kutai
		BO-0115182	Endert 5254	Kutai
		CANB 00508572	Th 1221	West Borneo
S. parvistipulata Heim	SPARV	HUH	TL 1143	West Borneo
		HUH	TL 1415	Borneo
S. scaberrima Burck	SSCAB	CANB 004993961		Sarawak
S. kunstleri King	SKUNS	KEP 20989	FRI 32671	FRIM
		HUH	S. 38734	Kuching
		HUH	Sand 16822	Sandakan
		HUH	S. 32451	Kuching
S. selanica Blume	SSELA	BO-0114806	Sutrisno 87	Bogor B. G.
		BO-0114805	Sutrisno 87	
		HUH	3502	Maluku
		HUH	115	Bogor Botanic Garden
S. palembanica Miq.	SPALE	CANB 37733	A.51B	Sandakan
		CANB 37872	A 1459	Sandakan
		B0-73442	TH Erdert 742	Sumatra

Appendix 3A Taxa selected for the morphological study (continued)

Species	Abbreviation	Herbarium number	Collection number	Location
Section *Pachycarpae*				
S. amplexicaulis Ashton	SAMPL	HUH	Jarvie 5756	Kalteng
		HUH	Jarvie 5331	Kalteng
		HUH	S 29174	Sarawak
		HUH	S 46406	Sarawak
		HUH	S 30001	N. Borneo
***S. macrophylla* (De Vriese) Ashton**	SMACR	HUH	TL 1073	W. Borneo
		HUH	TL 1125	GP
		BO-71908		W. Borneo
		BO-117008		FRI, Kepong
S. mecistopteryx Ridley	SMECI	HUH	TL 1215	W. Borneo
		BO-0115594	Kepong 85298	FRIM
S. pinanga Scheff.	SPING	HUH	TL 1194	West Borneo
		BO-0111791	bb. 30.190	Melawi
		BO-111790	bb. 29. 636	Melawi
		BO-74699		W. Borneo
		BO-74830		Borneo
		BO-74751	bb. 20.021	Borneo
S. splendida (De Vriese) Ashton	SSPLE	HUH	Jurie 5270	Kalteng
		BO-72189	bb.29.674	W. Borneo
		BO-72180	bb. 31.436	W. Borneo
		HUH	Kepong 98877	Arbentu, FRI
S. stenoptera Burck	SSTEN	BO-0114891		Bogor B. G.
		BO-0114889		Bogor B. G.
		BO-71894	bb. 29.664	W. Borneo
		B0-71905	bb. 29.289	W. Borneo
		HUH	Peters.1036	W. Borneo
		HUH	L8-533	FRIM
S. pilosa Ashton	SPILO	KEP 20984	22380	Sarawak
		KEP 20985	23970	Kapit
		BO-0117204	bb. 30.208	Melawi
		HUH	SAN 97229	Sabah
		HUH	S.22380	Sabah
S. beccariana Burck	SBECC	KEP 20987	S 46468	Lambir N.P.
		KEP 20986	SAR 29174	Kapit
		HUH	S. 46468	Sarawak
		HUH	Ambra & Arifin, Berau 1088	East Borneo
		HUH	S. 29243	Sarawak
S. rotundifolia Ashton	SROTU	HUH	29207	
		HUH	S-29539	Sarawak
		HUH	S-29207	Kapit
		HUH	S-29226	Kapit
***S. macroptera* Dyer**	SMACT	KEP 20982	FRI 36629	Perak
		KEP 20983	23487	Pasoh
		CANB 133921		Kepong
		CANB 38934	4384	Sandakan
		CANB 38959	4389	Sandakan
		HUH		Singapore

Appendix 3A Taxa selected for the morphological study (continued)

Species	Abbreviation	Herbarium number	Collection number	Location
Section *Mutica* subsection *Mutica*				
***S. parvifolia* Dyer**	SFOLI	CANB 143332	21405	Sandakan,
		HUH	Jarvie 6017	C. Borneo
		HUH	TL 1128	W. Borneo
S. macrantha Brandis	SMACN	HUH	C.F. 3553	Kepong
		HUH		Sarawak
		HUH		Mersing
		HUH	S-33550	N. Borneo
S. leprosula Miq.	SLEPR	CANB 251360		Sandakan
		CANB 72061		Sumatra
		CANB 72060		
		CANB 334372		
***S. singkawang* Burck**	SSING	KEP 20979	FRI 21684	FRIM
		KEP 17360	FRI 25407	FRIM
Section *Ovalis**				
***S. ovalis* Blume**	SOVAL	BO-0117479	bb. 32. 562	E. Borneo
		HUH	SAN 18757	Sandakan
		HUH	Kessher 626	E. Borneo
		HUH	SAN 13717	N. Borneo
		BO-0115588	8949	Kuching
Section *Doona*				
S. trapezifolia(Thw.) Ashton	STRAP	CANB 0049 3383		Sri Lanka
		HUH	TB Worthy 4860 Ho.	Sri Lanka
		HUH	AJ Kost 24465	Sri Lanka
		HUH	Waas 1755	Sri Lanka
S. gardreri		CANB 327907	S. Waas 1788	Ratnapura district
		CANB 00493374		Sri Lanka
		HUH	PS Ashton 2130	Sri Lanka
		HUH	2957	Sri Lanka
		HUH	Kost 24921	Sri Lanka
S. venulosa Wood ex Meijer	SVENU	HUH	Swaas 1611	Sri Lanka
		HUH	Kost. 24975	Sri Lanka
S. congestiflora (Thw.) Ashton	SCONG	HUH	2026	Ratnapura District
S. cordifolia (Thw.) Ashton	SCORD	HUH	Kost. 24588	Peradeniya
		HUH	PS Ashton 2160	Peradeniya
		HUH	Waas 2087	Sri Lanka
		CANB 327906		Sri Lanka
		CANB 00493381		Peradeniya

* monotypic section or subsection
** currently recognised as *Hopea ponga* (Densst.) Mabb.
Bold font shows type species of each genus, section or subsection

Appendix 3B Morphological data matrix used for the analyses

Taxa	Characters																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
DCONF	0	9	a	0	0	0	?	?	1	0	5	1	0	?	0	0	?	0	0
DRETU	1	8	9	0	0	0	0&1	0	1	0	3	1	0	0	1	1	?	0	0
DLANC	0	3	2	3	0	0	?	?	1	0	8	6	0	0	0	1	4	0	0
DAROM	0	3	3	3	0	?	?	?	?	?	8	6	?	?	0	0	?	0	0
PMALA	1	6	6	0	0	0	1	2	0	1	8	4	0	0	1	1	7	1	1
NHEMI	0	6	4	2	0	0	?	?	1	0	5	4	0	0	1	1	5	1	1
SACUM	0	6	4	0	0	0	0&1	1	1	1	5	2	1	1	1	1	6	1	1
SAMPL	1	7	6	0	0	?	?	1	0	0	7	5	0	1	1	1	8	1	1
SBECC	1	7	6	0	0	0&1	0&1	1&2	1	1	6	4	0	1	1	1	9	1	1
SASAH	0	4	3	0	0	?	?	1	1	0	3	3	0	0	1	1	4	0	0
SBALA	0	6	5	0	0	?	?	1	?	0	7	4	0	0	1	1	?	1	1
SCONG	1	6	5	0&2	0	?	?	?	?	?	6	7	?	?	?	?	?	?	?
SCORD	0	5	4	0	0	0	0&1	1	?	2	?	?	0	0	0	1	0	1	1
SFAGU	0	5	4	2	0	0	0&1	1	1	1	3	4	0	0	1	1	0	1	1
SFOXW	0&1	5	4	0	0	?	?	?	1	1	6	3	0	0	1	1	?	0	0
SGARD	0	4	3	0&2	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?
SGUIS	0	5	4	0	0	0	0&1	1	1	1	5	2	0	0	1	1	8	0	0
SHOPE	0	3	3	2	0	0&1	0&1	1&2	0	1	1	1	0	1	1	1	?	1	1
SHYPC	1	5	5	0	0	0&1	0&1	1&2	1	0	6	5	0	0	1	1	5	0	0
SHYPL	0	5	4	0	0	0	?	1	1	1	6	3	0	0	1	1	a	0	1
SISOP	0	6	6	0	0	0	0	1	0	1	0	0	0	0	1	1	1	1	1
SJAVA	0	4	4	0	0	?	?	?	0	?	?	?	?	?	?	?	?	1	?
SJOHO	0	6	6	0	0	1	0&1	2	0	1	5	3	0	0	1	1	5	1	1
SKUNS	0	5	4	0&1	0	0	1	1	0	1	5	2	0	1	1	1	4	1	1
SLAEV	0	4	2	0	0	1	0	1	0	1	2	3	0	0	1	1	4	0	0
SLEPR	0	5	4	0	0	0	?	1	0	0&1	4	4	0	0&1	1	1	6	1	1
SLONF	1	a	6	2	0	0&1	0	2	?	?	7	4	0	0	0	1	a	1	1
SMACN	0	6	5	0	0	0	0&1	1	1	1	7	3	1	1	1	1	a	1	0
SMACR	1	9	a	0	0	0	?	1	?	0	7	7	0	1	1	1	?	1	?
SMACT	0	7	5	0	0	0	0&1	1	0	0	5	6	0	0	1	1	7	1	1
SMAXI	0	7	5	2	0	0&1	0	1	0	1	8	4	0	0	1	1	9	1	1
SMAXW	0	4	3	0	0	1	?	1	0	0&1	4	1	0	0	1	1	7	0	0
SMECI	1	a	7	0	0	?	?	?	?	0	9	6	?	?	1	1	?	1	?
SMULT	0	1	2	2	0&1	0&1	0&1	2	1	1	2	0	0	0	1	1	6	1	1
SOVAL	0	6	5	0	0	0	0&1	1	0	0	5	7	0	1	1	1	4	0	0
SPALE	1	6	3	0	0&1	0	?	?	0&1	1	4	3	0	0	1	1	9	1	1

Appendix 3B Morphological data matrix used for the analyses (continued)

Taxa	Characters																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
SFOLI	0	4	3	0	0&1	1	0	0	0	0	4	3	0	0	1	1	7	1	1
SPARV	0	6	6	0	0	1	?	?	0	1	7	5	0	1	1	1	3	1	1
SPILO	1	6	5	0	0	0	1	1	1	0	7	5	0	1	1	1	a	1	1
SPING	0&1	7	5	0	0	0	?	?	1	0	8	5	0	1	1	1	8	0	1
SPOLY	0	4	3	0	0	0	0&1	2	1	0	2	3	0	0	1	1	1	0	0
SRICH	0	4	4	2	0	1	1	1	1	0	7	3	0	0	1	1	7	1	1
SROTU	1	6	6	0	0	0	0&1	1	1	0	7	5	0	1	0	1	8	1	1
SROBU	0&1	5	6	0	1	0	0&1	2	1	0	7	5	0	0	1	1	a	0	0
SROXB	0&1	6	5	2	0	0	0	0	1	0	9	7	0	0	0	1	7	1	1
SRUBE	0	5	6	0	0	0	0&1	2	1	?	7	5	?	?	1	1	?	1	1
SSCAB	0	6	5	0	1	1	?	?	1	1	7	7	?	?	1	1	?	1	1
SSELA	0	7	6	0	0	1	?	?	1	1	3	2	0	0	1	1	7	1	1
SSEMI	0	7	6	0	0	0	0&1	1	1	1	7	3	0	0	1	1	a	0	0
SSIAM	0&1	7	8	0	0	0	0&1	1	1	0	A	8	0	0	1	1	0	1	1
SSING	0	9	6	0	0	?	0	1	1	0	8	6	?	?	1	1	8	0	0
SSMIT	0	6	6	0	0	0	0&1	2	0	1	8	4	0	1	1	1	5	1	1
SSPLE	1	7	6	0	0	0	?	?	0	1	5	4	0	1	1	1	7	1	1
SSTEN	1	a	9	0	0	0	1	1	0	0	8	7	0	1	1	1	?	1	1
SSTIP	1	5	6	0	0	0	0&1	1	1	0	0	5	1	1	0	1	0	1	1
STHOR	0	4	4	0	0	0	0&1	2	1	1	6	4	0	0	1	1	9	0	1
STRAP	0	3	2	0	0	0	0&1	1	1	0	3	4	0	0	0	1	2	1	1
SVENU	0&1	4	3	0	0	0	0&1	1	1	0	3	5	0	0	0	1	2	1	1
SVIRE	1	6	6	0	0	?	?	?	1	1	7	4	0	0	1	1	a	1	1
HAPIC	0	7	5	0	0	0	?	?	1	1	4	4	0	1	0	1	5	1	1
HAPTE	0	7	6	0	0	0	0	1	1	1	5	5	0	0	1	1	3	1	1
HBRAC	0	1	0	1	0	0	0	0	1	0	2	3	0	1	1	1	?	1	1
HBREV	0	3	4	2	0	1	0&1	1	0	0	3	2	0	0	0	0	1	1	1
HCELE	0	5	4	0	0	0	0&1	1	1	1	3	2	0	0	1	1	6	1	1
HCELT	0	3	3	0	0	0	0&1	1	1	2	1	2	0	0	0	1	?	1	1
HCHIN	0	5	3	0	0	0	?	?	0	0	3	0	0	0	0	1	1	1	1
HDISC	0	4	4	0	0	?	?	?	?	?	?	?	0	?	?	?	?	1	1
HDRYO	0	3	3	1	0	0	?	?	1	2	3	2	0	0	1	1	4	1	1
HDYER	0	2	1	1	1	0	?	?	1	2	1	1	0	1	1	1	?	1	1
HFERR	0	2	1	1	0	0	0	2	1	2	1	1	0	1	1	1	2	1	1
HHAIN	?	5	5	0	0	0	0	2	0	0	2	3	0	0	1	1	1	1	1

Appendix 3B Morphological data matrix used for the analyses (continued)

Taxa	Characters																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
HJUCU	0	5	5	0	0	0	?	?	1	?	5	5	0	?	?	?	?	1	1	
HMENG	0	5	4	1	0	0	0	1	1	2	3	2	0	0	0&1	1	?	1	1	
HNERV	0	6	5	1	0	0	?	?	1	1	3	4	0	1	1	1	?	1	1	
HNIGR	0	3	2	1	1	1	?	?	1	1	1	1	0	1	1	1	?	1	1	
HODOR	0	5	4	0	0&1	1	0&1	2	1	1	3	3	0	0	1	1	6	1	1	
HPAPU	0	6	4	0	0	0	?	?	1	1	3	2	0	0	1	1	?	1	1	
HPHIL	0	7	5	0	0	1	0	2	1	1	4	3	0	0	0	1	7	1	1	
HPIER	0	3	1	1	0	0	?	?	1	2	3	2	0	0	1	1	3	1	1	
HPUBE	0	0	0	1	0&1	0	?	?	1	0	1	0	0	1	1	1	?	1	1	
HSPHA	0	2	2	0	0	0	0	0	1	1	4	5	0	1	1	1	?	1	1	
HSUBA	0	4	3	1	0	0	?	?	1	1	3	2	0	1	1	1	3	1	1	
HWIGH	0	8	5	2	0	0	0&1	2	1	1	5	4	0	1	0	1	7	1	1	
HWYAT	0	5	5	0	0	0	0&1	2	1	2	2	2	0	1	0	1	5	1	1	

Taxa	Characters																				
	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
DCONF	0	2	0	9	9	1	1	0	3	0	?	1	1	1	0	1	0	?	?	?	?
DRETU	0	3	0	a	8	1	1	1	1	0	3	?	1	1	0	1	0	7	9	?	?
DLANC	0	?	?	?	?	?	?	0	?	0	?	0	?	0	?	0	0	6	8	0	?
DAROM	0	?	0	3	?	0	0	0	?	0	?	0	?	0	1	0	0	7	6	0	1
PMALA	0	0	0	0	3	0	0	0	?	0	6	1	3	0	1	2	0	5	5	1	2
NHEMI	0	1	0	3	0	0	0	0	3	0	6	0	6	0	1	0	1	6	6	0	1
SACUM	1	?	0	4	3	?	?	0	?	1	2	0	?	0	0	2	0	6	3	1	?
SAMPL	2	8	0	6	4	0	0	1	0	0	5	0	4	0	1	2	0	8	9	1	0
SBECC	2	7	0	4	3	0	0	1	0	0	4	0	2	0	0&1	2	0	6	7	1	1
SASAH	0	5	1	7	5	1	1	1	?	1	5	1	1	0	1	0	2	3	2	1	?
SBALA	2	?	0	?	?	?	?	1	?	0	?	1	?	0	1	2	0	2	2	1	1
SCONG	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	2	0	3	4	?	?
SCORD	0	0	0	?	?	0	0	1	?	0	?	0	?	1	?	?	?	?	?	?	?
SFAGU	2	?	0	?	?	0	0	1	?	0&1	?	1	?	0	1	2	0	5	2	1	1
SFOXW	0	?	1	?	?	1	1	1	?	0	?	1	?	0	1	2	0	3	5	?	?
SGARD	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	2	0	5	4	0	2
SGUIS	2	6	1	2	2	0	0	1	0	0	1	1	0	0	1	2	0	5	5	1	2
SHOPE	?	?	0	?	?	?	?	?	?	0	?	1	?	0	0	2	0	4	3	1	?

Appendix 3B Morphological data matrix used for the analyses (continued)

Taxa	Characters																				
	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
SHYPC	2	0	0	1	0	0	0	0	?	0	5	1	1	0	1	2	0	8	7	0	1
SHYPL	2	?	1	8	3	0	0	1	1	1	2	1	?	0	0	2	0	5	5	1	1
SLONF	1	?	0	6	6	0	1	0	?	1	7	0	4	0	0	0	1	0	7	0	0
SMACN	1	2	0	5	5	0	0	0	?	1	7	0	7	0	0	0	1	1	9	1	0
SMACR	1	?	0	?	?	?	?	1	?	0	?	1	?	0	1	2	0	7	1	1	1
SMACT	2	1	0	2	3	0	0	0	0	0	5	1	2	0	0	2	0	5	5	1	2
SMAXI	1	4	0	5	5	0	0	1	1	1	5	1	2	0	0	0	1	7	5	1	0
SMAXW	2	3	1	3	3	0	0	1	2	1	4	1	1	0	1	2	0	6	6	1	?
SISOP	0	?	0	4	4	0	0	?	?	1	1	1	0	0	0	2	0	3	6	0	1
SJAVA	?	?	?	?	?	?	?	?	?	1	?	1	?	0	1	2	0	4	3	1	?
SJOHO	2	6	0	5	3	0	0	0&1	0	0	4	1	1	0	1	2	0	2	4	1	1
SKUNS	2	6	0	4	4	0	0	1	0	1	3	1	1	0	1	2	0	6	6	1	1
SLAEV	1	2	1	0	5	1	1	1	?	1	3	1	1	1	1	2	0	5	4	1	?
SLEPR	1	1	1	3	5	0	0	1	1	0	4	0	4	0	0	2	0	4	4	1	1
SMECI	?	?	0	?	?	?	?	1	?	?	?	?	0	?	1	2	0	9	a	1	1
SMULT	2	1	0	2	2	0	0	1	2	0	1	1	1	1	1	0	1	4	2	1	1
SOVAL	1	6	0	6	6	0	0	0	1	1	3	0	0	0	1	2	0	6	6	0&1	1
SPALE	2	0	0	5	3	0	0	0	0	1	6	1	1	1	1	2	0	3	4	1	2
SFOLI	1	2	0	3	3	0	0	0	0	1	4	1	2	0	1	2	0	3	4	1	0
SPARV	2	5	0	4	2	0	0	1	2	1	5	1	?	1	1	2	0	7	8	1	2
SPILO	2	4	0	8	5	0	0	1	9	0	8	1	8	1	1	2	0	5	5	1	2
SPING	2	a	0	3	3	0	0	1	0	0	7	1	6	0	1	2	0	7	7	1	1
SPOLY	1	1	0	6	1	0	0	0	?	0	?	1	?	0	1	2	0	7	6	1	?
SRICH	2	7	0	4	2	0	0	1	0	0	2	0	2	0	1	0	1	7	4	1	?
SROTU	2	1	0	7	5	0	0	1	?	0	8	0	8	0	1	2	0	6	8	0	1
SROBU	1	4	1	0	7	0	0	1	8	1	7	0&1	1	1	1	2	0	3	4	?	?
SROXB	2	a	0	1	4	0	0	1	4	0	8	0	9	0	1	2	0	5	4	0	1
SRUBE	1	?	0	?	?	0	0	1	?	0	?	0	?	0	1	2	0	6	4	1	1
SSCAB	1	?	?	?	?	?	?	1	?	1	?	1	?	0	1	2	0	a	9	1	?
SSELA	2	9	0	7	5	0	0	1	5	0	4	1	4	0	1	2	0	3	5	1	?
SSEMI	1	1	1	2	2	0	0	1	4	1	3	1	0	0	1	0	2	3	5	1	0
SSIAM	1	?	0	7	1	0	0	0	?	0	9	0	4	0	1	2	0	6	6	0	1
SSING	1	1	0	3	5	0	0	0	6	1	7	1	2	0	0	2	0	6	7	1	0
SSMIT	2	6	0	3	2	0	0	1	3	1	3	1	1	0	1	2	0	7	7	1	1
SSPLE	2	1	0	5	4	0	0	1	5	0	5	0	4	0	1	2	0	a	9	1	2
SSTEN	?	?	0	?	?	0	0	?	?	1	6	0	2	0	1	2	0	8	8	1	1

Appendix 3B Morphological data matrix used for the analyses (continued)

Taxa	Characters																				
	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
SSTIP	2	1	0	0	0	0	0	1	3	0	8	1	6	0	?	?	0	?	?	?	?
STHOR	0	2	1	0	0	1	1	1	?	0	5	1	2	1	1	2	0	?	?	?	?
STRAP	0	6	0	1	0	0	0	0	?	1	4	0	1	0	1	2	0	6	4	0	1
SVENU	0	4	0	7	4	0	0	0	?	0	4	1	1	0	1	2	0	6	4	0	2
SVIRE	2	6	0	0	6	0	0	1	5	0	6	0	5	1	1	2	0	3	5	1	1
HAPIC	2	?	0	5	3	0	0	1	?	1	4	1	0	0	0	0	1	4	3	1	0
HAPTE	1	?	0	?	?	0	0	1	?	1	?	?	?	?	0	0	1	3	2	?	?
HBRAC	1	6	0	0	0	0	0	1	?	1	3	1	0	1	1	0	1	3	3	0	1
HBREV	1	1	0	1	3	0	0	1	?	0	?	0	0	?	1	0	1	4	5	?	?
HCELE	2	2	0	2	3	0	0	1	3	1	3	0&1	0	1	1	1	0	8	7	1	1
HCELT	2	3	0	?	?	0	0	1	?	1	2	0	0	0	1	1	0	5	2	?	?
HCHIN	2	4	?	3	2	0	0	1	?	1	2	0	0	0	1	1	0	5	4	0	1
HDISC	?	?	?	?	?	?	?	?	?	0	?	0	0	?	1	1	0	3	3	?	?
HDRYO	2	3	0	3	3	0	0	0&1	2	0	3	0	0	1	1	1	0	2	3	0	1
HDYER	2	?	0	?	?	0	0	0	?	0	?	0	0	?	1	1	0	0	0	0	1
HFERR	1	3	0	1	2	0	0	1	2	0	2	0	0	0	1	1	0	1	1	0	?
HHAIN	2	1	0	1	2	0	0	0	?	0	0	0	0	0	1	1	0	4	3	1	1
HJUCU	2	?	?	?	?	?	?	?	?	0	?	?	?	?	1	1	0	5	5	?	?
HMENG	1	2	0	3	2	0	0	1	2	1	2	0	0	0	1	1	0	2	2	0	1
HNERV	2	?	0	?	?	0	0	0	?	1	3	0	0	0	1	1	0	3	4	0	1
HNIGR	1	?	0	?	?	0	0	1	?	?	?	?	?	?	0&1	1	0	1	1	0	1
HODOR	2	5	0	8	4	0	0	1	?	0	4	1	0	2	0	1	0	2	1	1	1
HPAPU	2	?	0	2	2	0	0	1	3	1	3	0	0	0	0	1	0	2	2	0	1
HPHIL	2	5	0	4	3	0	0	1	?	1	4	0	1	0	0	1	0	2	4	0	1
HPIER	2	?	0	1	2	0	0	1	1	1	2	0	0	0	1	1	0	3	3	1	1
HPUBE	1	?	0	?	?	0	0	1	?	0	?	0	?	?	1	1	0	2	1	0	?
HSPHA	1	?	0	?	?	0	0	1	?	0	?	0	0	?	1	0	1	2	4	0	1
HSUBA	2	5	0	2	3	0	0	1	3	1	3	0	0	0	1	0	1	4	4	0	?
HWIGH	2	6	0	5	5	0	0	1	3	1	4	0&1	6	0	1	1	0	4	4	0&1	2
HWYAT	2	6	0	3	2	0	0	1	?	1	4	0	6	0	1	0	1	4	5	?	?

CHAPTER 4
MOLECULAR STUDY

Outline

4.1 Introduction

Molecular data are often considered to be more reliable when inferring phylogenetic relationships at a lower taxonomic level than are morphological data. This is mainly because molecular characters provide complementary information at the genetic level. Hence, for recently diverged taxa where morphological characters are prone to phenotypic plasticity, molecular characters such as those from DNA sequences can provide more resolution to infer phylogenetic relationships.

The large amount of variation found among the species of *Hopea* and *Shorea* may be why there appear to be limited homologous characters that can be used for cladistic analysis. Cladistic analyses performed on 40 morphological characters (Chapter 3) have shown that none of these characters are free from homoplasy, including the

comparative development of the fruit calyx—the single character used in most taxonomic treatments to distinguish the two genera.

Cladistic analysis using DNA sequences has provided useful insights into the systematics and evolution of plants (Baldwin *et al.*, 1995). This technique has been widely used for phylogenetic studies, and is expected to provide more resolution to determine the relative phylogenetic position of the two genera within Dipterocarpaceae as well as examining the infra-generic relationships within *Hopea* and *Shorea*. Hence, DNA sequences from the *trn*L-F region of cpDNA and the ITS region of nrDNA will be employed to reconstruct the phylogeny of *Hopea* and *Shorea*.

4.2 Homology and evolution of molecular features

The study of phylogenetic reconstruction requires data that are homologous[1] throughout the different species used. Each nucleotide position in the sequence can usually be assumed to evolve independently and therefore an inference regarding homology can be made at the basic level. This level is that of the nucleotide position, which consists only of four possible character states in deoxyribonucleic acids—Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) (Lewin, 1990).

Phylogenetic reconstructions using cladistic analysis requires correctly inferred homologies. This is particularly important when using molecular data because all possible evolutionary events may have occurred in the molecular (genetic) level. Sequences can be related in several ways within a lineage. However, only sequences that have been duplicated through speciation events (orthologous sequences, where the sequence remains present as a single copy in both daughter lineages) are useful in inferring the phylogeny of their host taxa. However, some sequences may result from a gene duplication event (paralogy) and lateral gene transfer (xenology), and cause incorrect phylogenetic inferences to be made (Quicke, 1993; Hillis *et al.*, 1996).

[1] Homology means derived from a common ancestry (Hillis *et al.*, 1996), therefore inferences of phylogeny from homologous characters should reflect evolutionary relationships. The term homology is commonly misused to mean similarity (Fitch, 1966; Reeck *et al.*, 1987 in Hillis *et al.*, 1996), which is an empirical observation and can be quantified, whereas homology must be inferred and is not usually a quantifiable relationship (Hillis *et al.*, 1996).

Paralogous sequences are problematic in molecular phylogenies, since the sequences within a lineage do not always evolve independently (Nei and Koehn, 1983; Quicke, 1993; Hillis *et al.*, 1996). Such phenomena occur in multi-gene families (such as nuclear ribosomal RNA genes), which mostly occur in tandemly repeated sequences that are widely known to undergo concerted evolution.

Another issue regarding homologous sequences that needs to be addressed is positional homology. This positional homology is important to deduce any mutation events that have occurred within target sequences. Mutations which occur in nucleotide sequences can be present in the form of transitions (from purine to pyrimidine), transversions (within purine or pyrimidine) and insertion/deletion events (indels). These events can provide useful information for phylogenetic inference and assist in examining the course of evolution within a lineage.

4.3 Target Sequences

In molecular-based phylogenetic reconstruction, selection of the target sequence for analysis is important. The chloroplast and nuclear genomes selected for this study have different mutation rates and modes of inheritance, and thus may provide differing evolutionary signals for the phylogenetic analysis. Independent sources of DNA sequences have been used to increase the confidence placed in an estimate of phylogeny (Donoghue and Ackerly, 1996; de Queiroz *et al.*, 1995; Hillis, 1995; Miyamoto & Fitch, 1995). In addition, the use of sequences from a single genome reflects gene trees and not species trees (Doyle, 1992), whereas chloroplast and nuclear DNA can provide complementary and independent sources of sequence data.

4.3.1 Chloroplast genome: *trn*L-F

The chloroplast is an intracellular organelle present in plants, which contains the entire enzymatic machinery for the process of photosynthesis (Shizonaki *et al.*, 1993). In higher plants, the chloroplast genome is a circular molecule with a size of 120–160 kb[2]. It contains two single copy regions: the Large Single-Copy region (LSC) and the Small Single-Copy region (SSC), which are separated by two Inverted Repeats (IRs).

[2] kilobase pairs

The chloroplast genome consists of all four rRNA genes and around 35 tRNA genes, which are more conserved among higher plant species than the ITS (Internal Transcribed Spacer) regions (Baldwin *et al*, 1995). However, the conservative nature of chloroplast evolution has limited the use of the coding regions for phylogenetic studies at higher taxonomic levels. Some studies with a focus on the non-coding chloroplast regions have shown that they are more variable, as they are more often the subject of mutation (Baldwin *et al*, 1995) in the form of transversions, transitions, insertions and deletions.

Some regions within the chloroplast (cp) genome have been widely used for phylogenetic purposes, since the genome is highly conserved across taxonomic levels (Baldwin *et al*, 1995). Such regions are overwhelmingly similar in size, conformation, repeat structure and gene content. Phylogenetic studies using cpDNA markers in the Dipterocarpaceae have concentrated on the coding regions such as *rbc*L (Dayanandan *et al.*, 1999), *atp*B (Kajita *et al.*, 1998) and non-coding regions such as *trn*L-F (Kajita *et al.*, 1998). Since non-coding regions are usually more variable than coding regions, they are more useful when inferring the evolutionary relationships at lower taxonomic levels (Bayer *et al.*, 2000).

The present study employed the *trn*L intron and *trn*L/*trn*F intergenic spacer (Figure 4.1). Both are non-coding regions of the cp genome and have been widely used to infer evolutionary relationships at generic and infra-generic levels (Bayer and Starr, 1998; Bayer *et al.*, 2000), as well as at the specific level. These regions were chosen for the present analysis because of their utility at low levels in other groups, where a paucity of morphological variation has limited the resolution obtainable with that data.

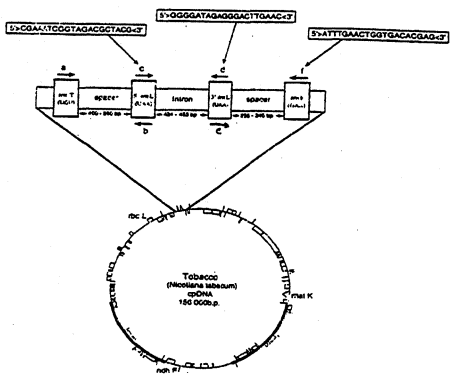


Figure 4.1 Diagram of chloroplast *trn*L-F region (after Bayer and Starr, 1998)

4.3.2 Nuclear genome: Internal Transcribed Spacers

The nucleus is an organelle with linear DNA. The eukaryotic nuclear ribosomal DNA gene unit is transcribed as a large precursor containing 17/18S[3], 5.8S and the internal and external transcribed spacers (ITS and ETS). It consists of small and large subunits of rRNA[4] forming a multigene family (Torres *et al.*, 1990; Liston *et al.*, 1996). Tandemly repeated nrDNA encodes for three ribosomal RNA (rRNA) genes (18S, 5.8S and 25S) and each copy contains a transcribed region that is separated by the long non-transcribed intergenic spacer (IGS). The transcribed region contains three rRNA coding genes along with two internal transcribed spacers (ITS regions), which occur in the following order: 5′–18S–ITS-1–5.8S–ITS2–25S (or 26S)–3′ (Kim and Jansen, 1994; Herskovitz and Lewis, 1996). Figure 4.2 shows the structure of ITS in nrDNA. These regions are transcribed as a single precursor rRNA. The ITS-1 and ITS-2 regions are subsequently spliced out and not incorporated into mature ribosomes, whereas the three coding regions eventually mature into rRNA. However, ITS-1 and ITS-2 appear to function, at least in part, in the maturation of nrDNAs (Baldwin *et al.*, 1995).

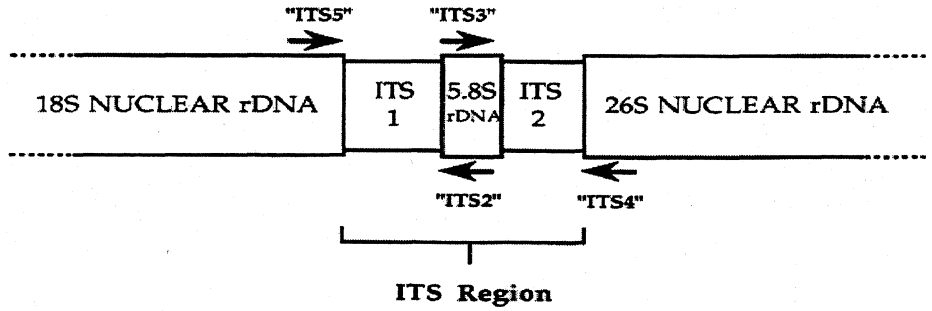


Figure 4.2 A diagram of the ITS regions in nrDNA, with arrows showing direction and approximate positions of some of the primers used to amplify these regions (after Baldwin *et al.*, 1995).

As with self-splicing group I introns, ITS is apparently under some evolutionary constraints in structure and sequence, as suggested by size and G+C content comparisons among angiosperm families (Baldwin *et al.*, 1995). G+C content may be a useful feature in assessing positional homology and in understanding the molecular evolution of such sequences, since the richness of G+C may be important in the formation of secondary structures as suggested by Baldwin *et al.* (1995) and Nickrent *et al.* (1993).

[3] subunits of RNA

The utility of the ITS sequence to infer evolutionary relationships at intrafamilial, interspecific and intergeneric levels has been widely acknowledged (Nickrent *et al.*, 1993; Baldwin *et al.*, 1995; Hershkovitz and Lewis, 1996; Liston *et al.*, 1996). It is not only useful because it is more variable than the three rRNA coding regions (Hershkovitz and Lewis, 1996), but also because of its other molecular properties. As part of nrDNA, whose sequence is biparentally inherited, ITS is useful in detecting hybridisation that may have occurred in related taxa as well as to infer the lineage of such taxa (Baldwin *et al.*, 1995; Campbell *et al.*, 1997). Secondly, its relatively small size[5] makes it relatively easy to amplify, even from old herbarium material (Baldwin *et al.*, 1995; Hillis *et al.*, 1996; Kim and Jansen, 1994). Lastly, as members of a gene family, the ITS regions are assumed to undergo rapid concerted evolution through unequal crossing-over and gene conversion (Baldwin *et al.*, 1995; Hillis *et al.*, 1996; Kim and Jansen, 1996). This particular property has some further implications. It promotes intragenomic uniformity between repeated units, even between nrDNA on non-homologous chromosomes (Baldwin *et al.*, 1995; Campbell *et al.*, 1997). Thus it reduces the presence of confounding paralogous copies of this region. Concerted evolution also promotes uniformity of nrDNA within inbreeding populations, thus minimising the need for intrapopulation sampling for phylogenetic studies.

However, some concerns arise when concerted evolution is incomplete, which will result in incomplete homogenisation. Incomplete homogenisation can arise when a hybridisation event was recent and nrDNA repeats are at different loci in the parental taxa (e.g. in different chromosomes), interlocus gene conversion is inoperative in the hybrid, or the hybrid is asexual (Baldwin *et al.*, 1995; Dubouzet and Shinoda, 1999; Steane *et al*, 1999; Kornkven *et al.*, 1998). Thus, recombination events in ITS sequences may occur and phylogenetic inferences from these sequences require examination of any evidence of recombination. When there is incomplete homogenisation, such sequences may be assumed to be paralogous instead of orthologous, which is common in multi-gene families (Zimmer *et al.*, 1980; Hillis *et al.*, 1996; Baldwin *et al.*, 1995).

Incomplete homogenisation can affect phylogenetic inferences, i.e. it may cause incorrect placement of taxa in topologies derived from the ITS data. Therefore the

---

[4] ribosomal RNA

[5] In angiosperms its size ranges from 565–700 bp (Baldwin *et al.*, 1995).

placement of taxa needs to be compared against topologies derived from other data sources. Incongruent placements may indicate that the ITS copies for these taxa are not fully homogenised. However, different modes of inheritance coupled with past hybridisation may also cause incongruent patterns. Therefore, some care needs to be taken when interpreting the phylogenetic history of taxa whose positions are incongruent among datasets.

4.4 Selection of taxa for analysis

The ingroup and outgroup taxa were selected to sample as widely as possible across the existing infrageneric divisions, geographic distribution and morphological variation. Selection of the taxa was made using the same criteria as those used for the morphological study. Table 4.1 gives the abbreviations used for taxon names and the region or regions for which each taxon was sequenced. Appendix 4A shows the herbarium voucher details and Genbank accession numbers for all taxa used in the molecular analyses. However, the number of species used in these molecular analyses is not as great as that in the morphological analyses due to difficulty in obtaining DNA samples, both from field and herbarium specimens. Some taxa were only represented by herbarium specimens and it proved difficult to isolate DNA particularly from older collections. The species present in the morphological analyses that are not included in the molecular data set are *Shorea rubella* from section *Rubella*, *S. hypochra* and *S. virescens* from section *Anthoshorea*, *S. thorelii* from section Shorea, and *S. congestiflora*, *S. venulosa*, *S. trapezifolia* and *S. gardneri* from section *Doona*.

Samples for DNA sequencing were obtained from two sources, fresh field samples and herbarium specimens. Field samples were either preserved on ice or preserved in a CTAB/ NaCl saturated solution (Rogstad, 1992). Where samples from fresh sources were unavailable, CANB and HUH herbarium materials were used. Vouchers for all field specimens are held at BO, CANB, FRIM, HUH and WAN (Appendix 4A).

Sixty species were used for the *trn*L-F analysis, including 53 species of the ingroup with 15 species of *Hopea* and 38 species of *Shorea*. The outgroup consisted of

Table 4.1 Taxa selected for the molecular study

Outgroup

Species	Abbreviation	*trn*L-F	ITS
Neobalanocarpus heimii	NHEMI	√	√
*Parashorea lucida**	PLUCI	√	
P. globosa	PGLOB		√
Dryobalanops aromatica	DAROM	√	
D. lanceolata	DLANC	√	√
Dipterocarpus retusus	DRETU	√	
D. confertus	DCONF	√	
*D. kerrii**	DKERI	√	
Anisoptera marginata	AMARG		√
Cotylelobium lanceolatum	CLANC		√
Total number of the outgroup species		7	5

Genus *Hopea*

Section	Subsection	Species	Abbreviation	*trn*L-F	ITS
Dryobalanoides	*Dryobalanoides*	*H. pubescens*	HPUBE	√	√
		H. mengerawan	HMENG	√	√
		H. cernua	HCERN	√	√
		H. dryobalanoides	HDRYO	√	√
		H. ferruginea	HFERR	√	√
		H. pierrei	HPIER	√	√
	Sphaerocarpa	*H. nervosa*	HNERV	√	
		H. nigra	HNIGR	√	√
		H. subalata	HSUBA		√
Hopea	*Hopea*	*H. celtidifolia*	HCELT	√	√
		H. celebica	HCELE	√	√
	Pierrea	*H. apiculata*	HAPIC	√	√
		H. wightiana	HWIGH	√	√
		H. brevipetiolaris	HBREV	√	√
		H. jucunda	HJUCU	√	√
		H. cordifolia	HCORD	√	√
Total number of *Hopea* species				15	15

Genus *Shorea*

Section	Subsection	Species	Abbreviation	*trn*L-F	ITS
Shorea	*Shorea*	*S. guiso*	SGUIS	√	√
		S. foxworthyi	SFOXW	√	√
		S. exelliptica	SEXEL	√	√
		S. seminis	SSEMI	√	√
		S. materialis	SMATE	√	√
	Barbata	*S. laevis*	SLAEV	√	√
		S. maxwelliana	SMAXW	√	√
Neohopea		*S. isoptera*	SISOP	√	√
Richetioides	*Richetioides*	*S. richetia*	SRICH	√	√
		S. multiflora	SMULT	√	√
		S. longisperma	SLONG	√	√
		S. hopeifolia	SHOPE	√	√
		S. maxima	SMAXI	√	√
		S. faguetiana	SFAGU	√	√
Anthoshorea		*S. roxburghii*	SROXB	√	√
		S. javanica	SJAVA	√	√
		*S. bracteolata**	SBRAC	√	

Table 4.1 Taxa selected for the molecular study (continued)

Section	Subsection	Species	Abbreviation	*trn*L-F	ITS
Brachypterae	*Smithiana*	*S. smithiana*	SSMIT	√	√
	Brachypterae	*S. selanica*	SSELA	√	√
		S. parvistipulata	SPARV	√	√
		S. johorensis	SJOHO	√	
		S. scaberrima	SSCAB	√	√
		S. balangeran	SBALA	√	
		S. palembanica	SPALE	√	
		S. kunstleri	SKUNS	√	√
Pachycarpae		*S. pilosa*	SPILO	√	√
		S. splendida	SSPLE	√	√
		S. stenoptera	SSTEN	√	
		S. macrophylla	SMACR	√	√
		S. amplexicaulis	SAMPL		√
		S. beccariana	SBECC	√	√
		S. pinanga	SPING	√	
Mutica	*Auriculatae*	*S. macroptera*	SMACT	√	√
	Mutica	*S. leprosula*	SLEPR	√	
		S. singkawang	SSING	√	√
		S. parvifolia	SFOLI	√	
Ovalis		*S. ovalis*	SOVAL	√	√
Doona		*S. cordifolia*	SCORD	√	
Total number of *Shorea* species				38	29
Number of taxa within each target region				60	49
Total number of taxa for both regions					44

* taken from the Genbank database

√species successfully sequenced for a given region

Neobalanocarpus heimii (a monotypic genus and putative sister to *Hopea*), *Parashorea lucida* (from the putative sister genus to *Shorea*), *Dryobalanops lanceolata* and *D. aromatica* (the putative sister taxa to the ingroup), and *Dipterocarpus confertus*, *D. kerrii* and *D. retusus* (the putatively most distantly related taxa in the outgroup).

Forty-nine species were used for the ITS analysis, including 44 ingroup species with 15 species of *Hopea* and 29 species of *Shorea*. The outgroup consisted of *Anisoptera marginata*, *Cotylelobium lanceolatum*, *Dryobalanops lanceolata*, *Parashorea globosa* and *Neobalanocarpus heimii*.

Different outgroup species were chosen for these analyses, partially due to the limited number of sequences obtained but also in order to examine the placements of the putative outgroup and sister taxa. It was thus also possible to investigate the effect of the outgroup used on the topologies resulting from the analyses.

4.5 Methods

4.5.1 DNA isolation

Total genomic DNA was isolated from specimens from fresh field collections following a modified DNA extraction procedure for small quantities (0.01–0.02 g) of tissue (Doyle and Doyle, 1987). Leaf tissue was ground in liquid nitrogen with the addition of preheated (65°C) CTAB[6] grinding buffer with 1% ß-mercaptoethanol, and then incubated in Eppendorf tubes at 55°C for 30 minutes. Then, 250 µl SEVAG[7] was added to the tubes, which were inverted for 10 minutes and then centrifuged at top speed for 10 minutes. The resulting supernatant was removed to another tube and 350 µL of 95% EtOH added. The DNA was precipitated by incubating the mixture at 4°C for 4 hours followed by 10 minutes centrifugation. The resulting pellet was washed with 350 µL of 70% EtOH, dried in a vacuum centrifuge for 15 minutes, and then resuspended in 40 µL TE buffer.

[6] CTAB: n-hexadecyl trimethylammonium bromide. CTAB extraction buffer: 100 ml 1 M Tris pH 8.0, 250 ml 5 M NaCl, 80 ml 0.25 M EDTA, 20 g CTAB, 0.5 ml β-mercaptoethanol. All these are diluted to 1 litre with distilled water.

[7] 20 chloroform : 1 isoamyl alcohol

Total genomic DNA from herbarium material was isolated following a modified DNA extraction procedure for large quantities (0.2–1 g) of tissue (Soltis lab procedure, unpubl.). This procedure is modified from Doyle and Doyle (1987) with an additional 4% PVP[8] mixed with the CTAB. The inorganic component of the cell tissue was removed via precipitation by adding 2/3 volume of 24:1 chloroform/isoamyl alcohol and inverting the tubes several times. The tubes were then centrifuged at 4000 rpm for 10 minutes and the supernatant removed. The total DNA was precipitated by adding 2/3 volume of cold[9] isopropanol. RNA was removed by digesting in 1/5 volume of RNase in a 55°C water bath for 30 minutes[10]. The DNA was then pelleted by pouring off the content of the tubes into microtubes followed by one minute of centrifugation. This process was repeated until all the solution in a 25 mL tube had been pelleted. Each pellet was washed in 1.5 mL cold DNA wash solution[11] and dried for 30 minutes in a vacuum centrifuge. This process was followed by resuspension in 200 µL of DNA storage buffer[12] and 55°C incubation. The suspended DNA was kept at -4°C for temporary storage or -80°C for longer-term storage.

The total genomic DNA extracted from the herbarium material was purified using the diatomite method of Gilmore *et al.* (1993) and Qiagen Quick PCR Preps DNA Purification System® (Qiagen Inc., Chatsworth, U.S.A.). This is a necessary step since dried herbarium materials contain inhibitors, which may interfere with the amplification process. The genomic DNA isolated by the above procedure, except the CTAB grinding buffer, was used with the addition of 1% ß-mercaptoethanol and 2% 0.5 M EDTA.

DNA was collected by centrifugation for two minutes at soft spin followed by phenol or chloroform extraction (1 volume) to remove any residual inorganic components. Following five minutes of centrifugation, DNA was precipitated by keeping the tubes at -20° C for two hours (or for 30 minutes at ultralow[13] temperature or 24 hours at

---

[8] Percent PVP = number of grams/100 ml CTAB extraction buffer.

[9] -4° C

[10] This process is termed RNasing.

[11] 76% EtOH/10mM NH_4Oac: 760 mL 100% EtOH, 1 mL 10 M Ammonium Acetate and 239 mL distilled water.

[12] 10 mM NH_4Oac 0.25 mM EDTA : 0.773 g CH_3OONH_4 and 0.095 g EDTA in 1 litre of distilled water.

[13] -80° C

room[14] temperature) and then adding 1/3 volume of 1 M NaCl and 2/3 volume of absolute EtOH. The DNA pellet was collected by centrifugation at top speed for 10 minutes. The pellet was then vacuum dried and resuspended in 50 μL TE. The working template for PCR amplification was 150 ng/μL. Quantification of DNA was performed by measuring absorbency at 260 nm with a spectrophotometer.

4.5.2 DNA amplification

The *trn*L intron region was amplified by PCR using primer pair "c" and "d" and the intergenic spacer of *trn*L-F was amplified with primer pair "e" and "f" (Taberlet *et al.*, 1991). The ITS-1 region was amplified using the primer pair ABI 101F and ITS2, while the ITS-2 region was amplified using primers ITS3 and ABI 102R (White *et al.*, 1990). Table 4.2 gives sequences for the primers used. Both the fresh and herbarium materials were amplified using the same procedure.

Standard PCR amplification as described by Bayer *et al.* (1996) was employed. The PCR program used to amplify the *trn*L-F regions included a three minute "hot start" period before the addition of Taq DNA polymerase, followed by 30 cycles of denaturation, annealing and extension. The temperature for denaturation of the double stranded DNA was set at 94°C for 1 minute. Annealing temperature varied according to a "touch down" procedure, in which the first five cycles were run at 55° C and then dropped one degree per cycle for 7 cycles to reach a minimum annealing of 48°C repeated for 18 cycles. Extension temperature was set at 72°C for two minutes. After 30 cycles were completed, the amplification was terminated by a final 7 minutes extension phase at 72°C. The amplification of the ITS regions was the same as that described for *trn*L-F, but the "touch down" technique was not used. Instead, annealing temperature was set at 60°C. PCR products were visualised by UV fluorescence after staining with ethidium bromide on a 1% agarose gel in 1x TE buffer. The target bands were purified using the Wizard Kit PCR Preps DNA Purification System® (Promega Inc.) and the Qiagen Quick PCR Preps DNA Purification System® (Qiagen Inc.).

[14] 20° C

Table 4.2 Primers used to amplify and sequence *trn*L-F and ITS regions in this study.

Target region and primer direction	Primer name	5′–3′ primer sequence
*trn*L intron		
forward	c	CGAAATCGGTAGACGCTAGG
reverse	d	GGGGATAGAGGGACTTGAAC
intergenic spacer of *trn*L-F		
forward	e	GTTCAAGTCCCTCTATCCCC
reverse	f	ATTTGAACTGGTGACACGAG
ITS-1		
forward	ABI101F	ACGAATTCATGGTCCGGTGAATTCG
reverse	ITS2	GCTGCGTTCTTCATCGATGC
ITS-2		
forward	ITS3	GCATCGATGAAGAACGCAGC
reverse	ABI102R	TAGAATTCCCCGGTTCGCTCGCCGTTAC

4.5.3 DNA sequencing

The sequence data were obtained through direct sequencing of double stranded DNA derived from the PCR procedure. The double stranded PCR products were purified in order to remove excess dye terminator by using the ABI PRISM dRhodamine Terminator Cycle Sequencing Ready Reaction Kit®[15]. Cycle sequencing reactions were carried out in a Perkin-Elmer thermocycler, using the purified PCR product and following the protocol of the DeDeoxy Terminator Cycle Sequencing Kit (Applied Biosystems Inc., Foster City, California). Sequencing primers "c", "d", "e" and "f" were used in the sequencing the trnL-F region and ABI101F, ITS2, ABI102R, and ITS3 for the ITS region. The samples were then run on an acrylamide gel using an ABI Automated Sequencer (Applied Biosystems Inc.) at the CSIRO Division of Plant Industry in Canberra.

[15] This kit contains AmpliTaq DNA Polymerase and dRhodamine dye terminators which are premixed into a single tube of Ready Reaction Mix.

4.6 Data analysis

4.6.1 Alignment of homologous sequences

Ingroup and outgroup sequences were aligned and edited using Sequencher version 3.0. All substitutions were double-checked after alignment to verify their accuracy. Ambiguous sequence alignments were aligned using Clustal-W (Thompson *et al.*, 1994) and refined by eye. Gaps were inferred in the alignment and unambiguous indels[16] were coded as binary characters. Data matrices were prepared in MacClade (Maddison and Maddison, 1992) and these were analysed using PAUP* 4.0b4a (Swofford, 1998). The alignments used for both ITS and trnL-F regions are given in Appendix 4B. The data sets were analysed with and without indels.

4.6.2 Cladistic analyses

Two analyses were performed on each region by including and excluding the indels to estimate their effect on the robustness of the topologies obtained. Another analysis was performed using only taxa that had been sequenced for both the *trn*L-F and ITS regions, in order to assess any incongruence between the data sets before combination into a single matrix.

The optimal tree—evaluated using maximum parsimony—was estimated using a heuristic search strategy. A thousand replicate search was conducted using random addition to search across multiple islands of trees (Maddison, 1991), and this strategy was used for all final tree searches. MAXTREES was set to 500 and not increased. Tree Bisection Reconnection (TBR) branch-swapping was used, with the steepest descent option off and using ACCTRAN (Accelerated Transformation) optimisation. The MULPARS (multiple parsimonious trees) option was on and minimum branches of zero were collapsed. Ten equally parsimonious trees were held following each replicate (Swofford, 1998).

The character states were treated as unordered (Fitch, 1971) only. Statistical measures of the Consistency Index (CI), Homoplasy Index (HI) (Kluge and Farris, 1994), Rescaled Consistency Index (RC) and Retention Index (RI) (Farris, 1989) were also calculated.

[16] insertion-deletion

Clade support was estimated by performing 100 bootstrap replicates (Felsenstein, 1985) by using 50% majority-rule of MPT input as trees but with MULPARS off. Trees were rooted using the selected outgroup taxa, as defined in each analysis.

4.7 Results

An outline of the characters used and of the topological features of the putative phylogenies obtained is given for each region.

4.7.1 *trn*L-F

4.7.1.1 Outline of the characters used

The entire aligned length of the *trn*L intron and *trn*L-F spacer is 910 bp, consisting of 514 bp of *trn*L intron and 396 bp of *trn*L-F spacer. Indels, coded to infer positional homology, range in size between 1 and 14 bp, excluding the long deletions that have occurred in *Hopea pubescens*, *H. mengerawan*, *H. pierrei* and *H. ferruginea* (Table 4.3). However, only indels whose size is more than 2 bp are included in the analyses, in order to avoid the high level of homoplasy that often characterises small (1 or 2 bp) indels.

Table 4.3 Position and size of the *trn*L-F indels

Indel no.	Size (bp)	Location	Type of event	Species
1	14	277–290	insertion	DKERI
2	14	291–304	insertion	DCONF, DKERI, DRETU
3	3	607–609	insertion	DCONF, DKERI, DRETU
4	3	615–617	deletion	DCONF, DKERI, DRETU
5	10	655–664	insertion	DLANC, DAROM
6	6	665–673	deletion	HFERR, HPIER, HPUBE, HMENG
7	4	674–677	insertion	DLANC, DAROM
8	10	684–747	deletion	HFERR, HPUBE, HPIER, HMENG

Eight indels were recorded, two from the *trn*L intron and six from the *trn*L-F spacer. However, there are only two long insertions that give phylogenetic information at the

generic level. These are the two insertions in the *trn*L intron recorded for *Dipterocarpus* (indels 1 and 2 in Table 4.3) and two insertions noted for *Dryobalanops* (indels 5 and 7 of Table 4.3).

In addition to these insertion events, there are three distinct deletions (indels 4, 6, and 8) that have occurred in four species of *Hopea* section *Dryobalanoides*, and these also provide a strong phylogenetic signal on the branch supporting these taxa (Figure 4.3, clade H).

To confirm the identity of the *trn*L-F sequences, they were subjected to a BLAST search in Genbank. The sequence of *Dipterocarpus kerrii* used was obtained from Genbank (accession number AB006409). The sequences of *D. confertus* and *D. retusus* obtained are similar to those of *D. kerrii* and *D. baudii* (AB006410). Sequences obtained from other species are also similar to angiosperm sequences in the Genbank database.

The number of parsimony informative characters (92) from the *trn*L-F regions used including the indels is comparatively low, given the total length of sequence obtained (910 bp). The parsimony informative characters occur in the form of base substitutions, with transitions occurring more frequently than transversions. The G+C content over 827.55 sites (average unaligned sequence length) is 32.46%, suggesting that the *trn*L-F region is A+T rich.

4.7.1.2 Topological features

The discussion of this topology is based on the results of analysis of a data set where indels were scored as binary characters. In this analysis, 110 most parsimonious trees with a length of 337 steps were obtained. These trees have a CI of 0.62, RC of 0.66, HI of 0.38 and RI of 0.82. These statistics suggest that even though homoplasy occurred in 38% of the characters, the changes are mostly apomorphic.

The topology of the cladogram (Figure 4.3) resulting from this analysis of *trn*L-F sequences from 60 species suggests that the *Dryobalanops* clade (labelled C, synapomorphies 8, bootstrap 100%) is the sister taxon to the ingroup (clade B). Also, the inclusion of the putative sister taxon to *Hopea* (*Neobalanocarpus heimii*) and the putative sister to *Shorea* (*Parashorea lucida*) within clade B suggests that these species may form part of the ingroup.

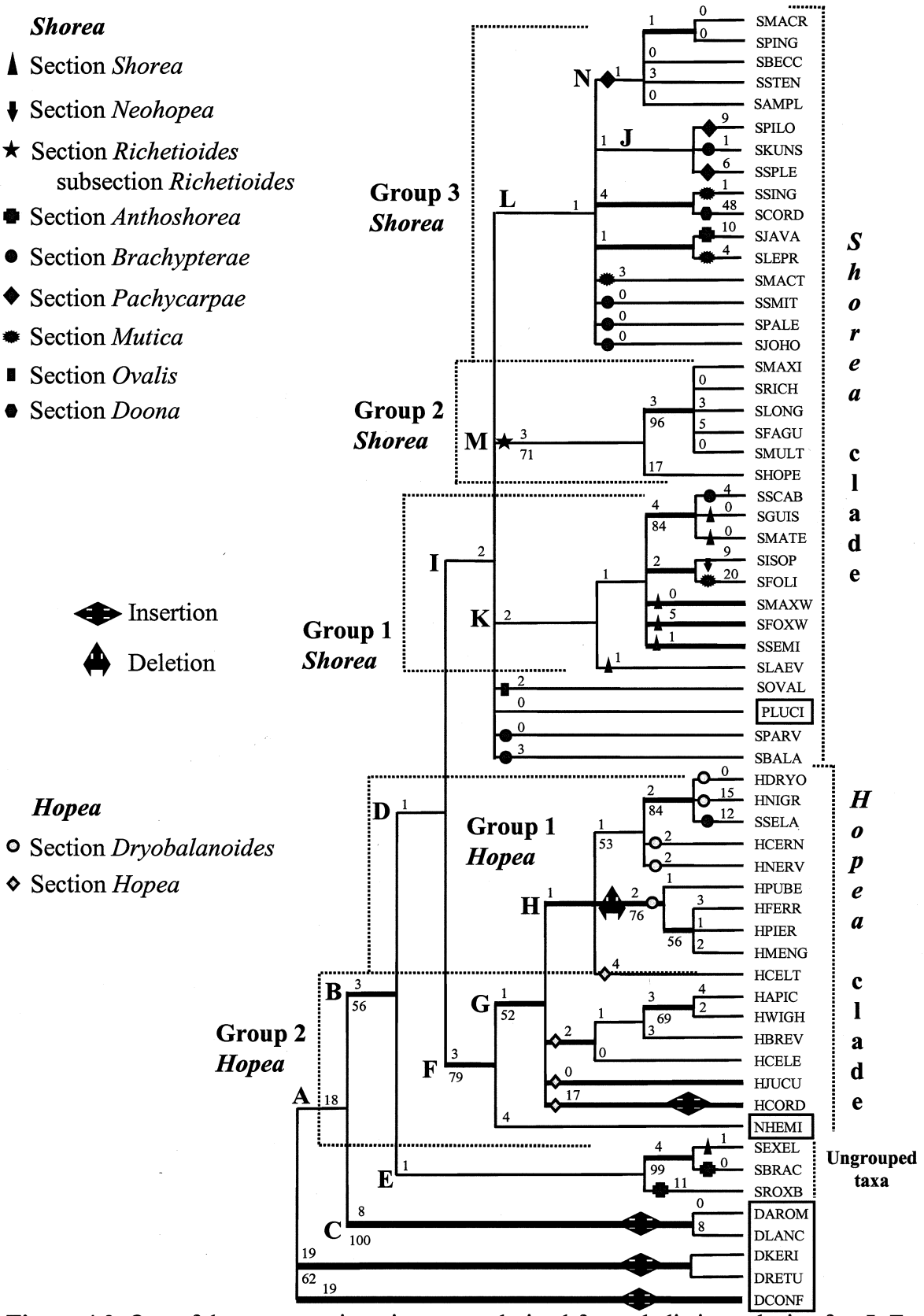


Figure 4.3 One of the most parsimonious trees derived from cladistic analysis of *trn*L-F sequences of selected Dipterocarpaceae taxa, excluding those characters and indels which were uninformative. Numbers above the branches are branch lengths, while bootstrap values of 50% and greater are shown below. Thicker branches are those that appear in the strict consensus of all trees. Taxon names in boxes are the putative outgroups.

The ingroup clade (B) was moderately supported by the bootstrap (56%) and defined by three synapomorphic changes. This group contains two paraphyletic clades, with a group (E) consisting of three species of *Shorea* sections *Anthoshorea* and *Shorea* being the sister taxon to the remaining ingroup taxa (D). However, the clade with *Shorea* species (E) only received weak bootstrap support (<50%), with only one synapomorphic base change. The phylogenetic position of these three species is thus equivocal.

The main ingroup clade (D) has two major subclades, with a clade containing all the *Hopea* species included in the analysis (F) appearing to be the sister group to the majority of the *Shorea* species (I).

The main *Shorea* group (I, synapomorphies 2, bootstrap <50%) forms a large polytomy consisting of three major groups (M, K and L) and four single species—*S. parvistipulata*, *S. balangeran*, *S. ovalis* and *Parashorea lucida*. Members of clade I share two synapomorphic base changes with little bootstrap support.

The first group within *Shorea* (K, synapomorphies 2, bootstrap <50%) is referred to as Group 1, and consists of species from sections *Shorea*, *Mutica*, *Neohopea* and *Brachypterae*. All the taxa of section *Shorea* included in the analysis fall within this group except for *S. exelliptica* (which is in clade E). The only clade within Group 1 with strong bootstrap support (84%, synapomorphies 4) is a polytomy containing *S. scaberrima* (section *Brachypterae*), *S. guiso* and *S. materialis* (both from section *Shorea*). *Shorea isoptera* (section *Neohopea*) and *S. parvifolia* (section *Mutica*) are also paired within Group 1 (synapomorphies 2, bootstrap <50%). It is apparent from these results that all the *Shorea* sections within Group 1 are non-monophyletic, with the exception of the monotypic section *Neohopea*.

The second group in the main *Shorea* clade (M, synapomorphies 3, bootstrap 71%) is referred to as Group 2 and consists entirely of species from Section *Richetioides* subsection *Richetioides*. This group is further divided with *S. hopeifolia* as the sister taxon to the remaining species. These remaining species form a polytomy that is strongly supported by the bootstrap (96%) and defined by three synapomorphic changes.

The third group in the main *Shorea* clade (L, synapomorphies 1, bootstrap <50%) is referred to as Group 3 and consists of a mixture of species from sections *Brachypterae*, *Pachycarpae*, *Anthoshorea*, *Mutica* and *Doona*. This group is a large polytomy, with four single species from sections *Brachypterae* and *Mutica* and four larger sub-groupings. All the seven species from section *Pachycarpae* included in the analysis are placed in two (N and J) of these four sub-groupings. The first *Pachycarpae* sub-group contains five species (N, synapomorphies 1, bootstrap <50%), and the second sub-group (J, synapomorphies 1, bootstrap <50%) consists of two species that are grouped with *S. kunstleri* (from section *Brachypterae*).

Another grouping within the large Group 3 polytomy consists of *Shorea cordifolia*, a member of the Sri Lankan endemic section *Doona*, paired with *S. singkawang* from section *Mutica* (synapomorphies 4, bootstrap <50%). The long branch (48 autapomorphic changes) leading to *S. cordifolia* suggests that this species has a *trn*L-F sequence that is very distinct from *S. singkawang*. In addition, this long branch may suggest that more species from section *Doona* should have been included in this analysis. Such species may "break up" the long branch, but with only one species included the large number of sequence changes in a section become autapomorphic for that species.

Also within *Shorea* Group 3 one member of section *Anthoshorea*, *S. javanica*, groups with *S. leprosula* from section *Mutica* based on a single base substitution in the *trn*L-F spacer. Moreover, with no or few unequivocal changes present in *S. smithiana*, *S. palembanica* and *S. johorensis* (all from section *Brachypterae*) or in *S. macroptera* (Section *Mutica*) these species all collapse into the main Group 3 polytomy. Except for section *Doona*, from which only one species was included in this analysis, all the *Shorea* sections within Group 3 appear to be non-monophyletic.

The cladogram based on *trn*L-F data suggests the potential monophyly of *Hopea* albeit with the inclusion of *Shorea selanica*. The *Hopea* clade (F, synapomorphies 3, bootstrap 79%) consists of a sub-clade (G, synapomorphies 1, bootstrap 52%) which combines *S. selanica* with all the species of *Hopea* included in the analysis. *Neobalanocarpus heimii* is then placed as the sister group to this entire sub-clade, which seems to confirm the suggestion that *N. heimii* is the closest relative to *Hopea*.

The first group within the *Hopea* clade is termed Group 1 (H, synapomorphies 1, bootstrap <50%) and consists of three subgroups forming a polytomy. The first of these subgroups (synapomorphies 1, bootstrap 53%) consists of four members of section *Dryobalanoides* (*H. dryobalanoides*, *H. nigra*, *H. cernua* and *H. nervosa*) grouped with *Shorea selanica*. A polytomy within this subgroup consisting of *H. dryobalanoides*, *H. nigra* and *S. selanica* is strongly supported by the bootstrap analysis (84%) although only two synapomorphic changes occur along the branch. The second of the *Hopea* subgroups (synapomorphies 2, bootstrap 76%) contains four other species from section *Dryobalanoides—H. ferruginea*, *H. pubescens*, *H. pierrei* and *H. mengerawan*. A single species from section *Hopea*, *H. celtidifolia*, is the final subgroup of the larger Group 1 *Hopea* polytomy.

The remaining taxa within the main *Hopea* clade (G) are all members of Section *Hopea* and are termed Group 2. Two Sri Lankan taxa, *H. jucunda* and *H. cordifolia*, form single species lineages. Another Sri Lankan species, *H. brevipetiolaris*, forms a larger grouping (synapomorphies 2, bootstrap <50%) with the remaining taxa, *H. celebica*, *H. apiculata*, and *H. wightiana*.

The overall topology obtained by analysis of the *trn*L-F sequences with indels included strongly suggests that *Shorea* is monophyletic only with the inclusion of *Parashorea lucida* and the exclusion of four other species (*S. selanica*, *S. exelliptica*, *S. bracteolata* and *S. roxburghii*). *Hopea* is only monophyletic with the inclusion of *S. selanica*.

Excluding the indels from the cladistic analysis of *trn*L-F sequences produces some changes in the topology obtained (Figure 4.4). The changes include the arrangement of the internal nodes in the *Shorea* clade and the arrangement of species within the *Hopea* clade. However, no change occurs in the sister clade to the ingroup (E, Figure 4.3).

Within the *Hopea* clade, *H. celebica* is now paired with *H. brevipetiolaris* but both species are still within the same general group as in the previous analysis including indels.

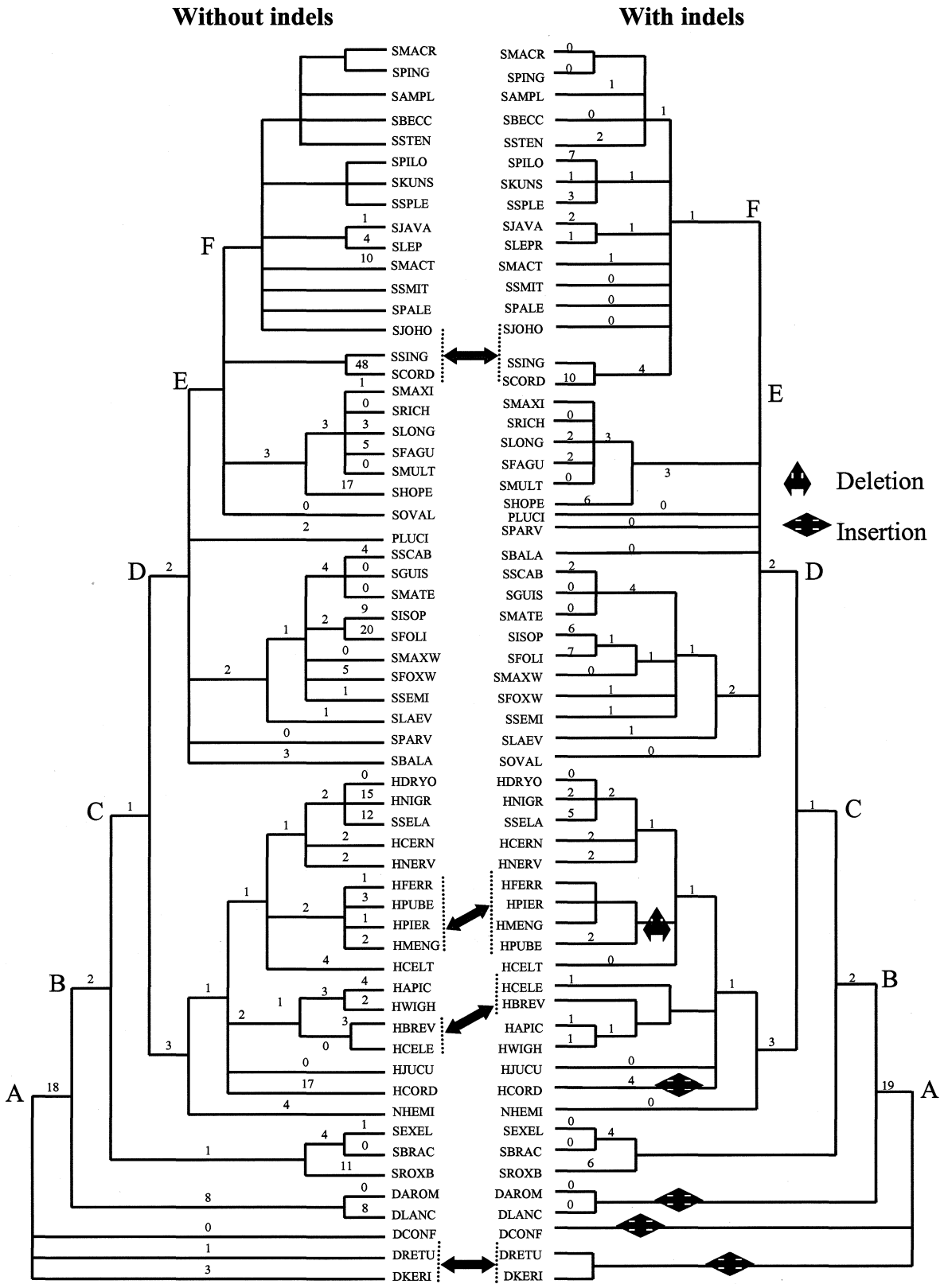


Figure 4.4 Comparison between topologies obtained from cladistic analysis of Dipterocarpaceae *trn*L-F sequences with indels included and excluded. Numbers above the branches are branch lengths and blue arrows indicate the major differences between the topologies.

The changes within the *Shorea* clade in the topology obtained from analysis of the *trn*L/F data excluding indels mostly occur at the internal nodes. For example, *Shorea ovalis* is now part of a polytomy with the Group 2 and Group 3 clades, rather than part of the larger *Shorea* polytomy. Nevertheless, there are few changes in the species arrangement within the groupings obtained in the first analysis.

4.7.2 ITS

4.7.2.1 Outline of the characters used

The alignment for the ITS region is 934 bp long. Mutation in the ITS regions is marked mostly by point substitutions rather than insertion or deletion events, and only 8 unambiguous indels were introduced to maintain the positional homology (Table 4.4). As with the *trn*L-F region, only indels longer than 2 bp are included in the analysis to avoid unnecessary homoplasy. Insertion events within ITS provided phylogenetic information in this analysis, but these mostly resolved species relationships rather than those at a higher level.

Table 4.4 Position and size of the ITS indels

Indel no.	Size of indel (bp)	Location in alignment	Type of event	Species
1	3	1162–1164	insertion	SSING, SMACT, SPILO, SKUNS, SMACR, SSPLE, SAMPL, SBECC, SROXB, SJAVA, SSMIT, SPING, SBALA, SACUM, SJOHO, SLEPR
2	4	1273–1276	deletion	SPILO, SSELA
3	5	1441–1445	insertion	SSELA, SMAXI
4	6	1456–1461	deletion	SSING, SMACT
5	4	1464–1467	insertion	HCELE, SSING, SMACT, SSELA, SMAXI, SRICH, NHEMI, AMARG, SLONG, SHOPE, SFAGU, SMULT
6	3	1475–1477	insertion	HCELE, SSELA, SMAXI, SRICH, SLONG, SHOPE, SFAGU, SMULT, SROXB
7	4	1493–1495	deletion	SMAXW, SISOP, SGUIS, SFOXW, SMATE, SJAVA, PGLOB
8	3	1505–1507	insertion	HCELE, SSELA, SMAXI, SRICH, SLONG, SHOPE, SFAGU, SMULT

The ITS sequence generally is more variable than the *trn*L-F, making it difficult to align. The alignment was checked by comparing the sequences of several representative species used in this study with the ITS sequences available on Genbank.

Of the total ITS nucleotides, 39.47% are parsimony informative (358 bp). This number is considerably higher than that for the *trn*L-F sequences (< 10%). The ITS regions are G+C rich, with these two bases accounting for 65% of the approximately 731.39 average unaligned sequence length.

4.7.2.2 Topological features

Eight most parsimonious trees were obtained from the analysis. These had a length of 1821, CI[17] of 0.40, HI of 0.60, RI of 0.51 and RC of 0.24. As with the earlier analyses of the *trn*L-F sequences, this discussion of the putative phylogeny is based on the topology obtained using the data set with indels included. The shortest trees obtained have a high level of homoplasy (60%), due to many changes occurring at each nucleotide position.

Figure 4.5 shows the cladogram obtained from sequence and indel data for the ITS regions of 49 species. The topology suggests that the ingroup is very well defined (A, synapomorphies 35, bootstrap 100%). Three putative outgroup taxa (*Dryobalanops lanceolata*, *Parashorea globosa* and *Neobalanocarpus heimii*) are nested within this ingroup clade. *Dryobalanops lanceolata* is grouped within Group 3 of *Shorea* (M), *P. globosa* is placed in Group 1 of *Shore*a (L) and *N. heimii* is the sister taxon to a clade containing manly of the species of *Hopea* included in the analysis (D).

The ingroup clade (A) consists of two distinct lineages, the "core" ingroup clade (C) and a clade (B) consisting of three species of *Hopea* section *Hopea*—*H. apiculata*, *H. wightiana* and *H. brevipetiolaris*. However, the monophyly of this group (B) was only weakly supported by the bootstrap analysis (<50%). The main ingroup (C) is further divided into two groups, the *Hopea* and *Shorea* clades, which thus accords with the generic divisions.

[17] after excluding uninformative characters

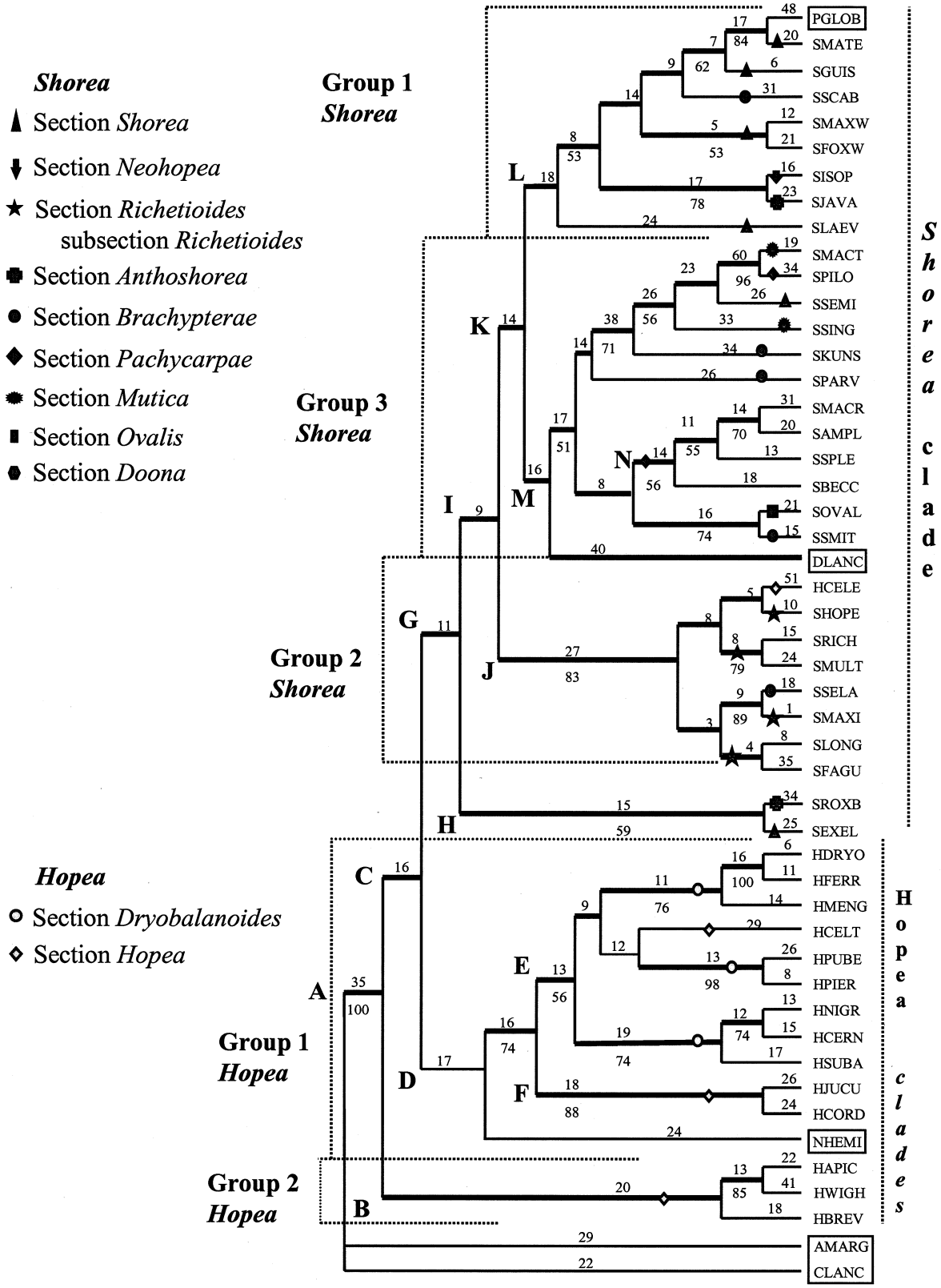


Figure 4.5 One of the most parsimonious trees derived from cladistic analysis of ITS sequences of selected Dipterocarpaceae taxa, excluding those characters and indels which were uninformative. Numbers above the branches are branch lengths, while bootstrap values of 50% and greater are shown below. Thicker branches are those that appear in the strict consensus of all trees. Taxon names in boxes are the putative outgroups.

Hopea as a whole is not monophyletic, given the previously mentioned placement of clade B and the presence of *H. celebica* within *Shorea* Group 2. *Neobalanocarpus heimii* is the sister taxon to the Group 1 *Hopea* clade, which supports the previous suggestion that *Neobalanocarpus* is the closest relative to *Hopea*. The *Hopea* Group 1 clade is supported by a 74% bootstrap value. Two groups are apparent within this clade. The first contains all the species from section *Dryobalanoides* included in the analysis (E) and is reasonably well defined (synapomorphies 13, bootstrap 56%), but is only monophyletic with the inclusion of *H. celtidifolia* (section *Hopea*). The second group (F) is a well supported pairing (synapomorphies 18, bootstrap 88%) of two Sri Lankan members of section *Hopea*, *H. jucunda* and *H. cordifolia*. Thus, the current sectional division of *Hopea* is not supported by the results of this analysis, as neither of the sections can be considered monophyletic.

The *Shorea* clade (G, synapomorphies 11, bootstrap <50%) contains two major groupings (H and I). Clade H (synapomorphies 15, bootstrap 59%) consists of two anomalous species, *S. roxburghii* (section *Anthoshorea*) and *S. exelliptica* (section *Shorea*). Clade I (synapomorphies 9, bootstrap <50%) contains the remainder of the *Shorea* species included in the analysis in two further subgroupings (K and J).

Clade K consists of two further groupings, L and M, that are only weakly supported by the bootstrap analyses (<50%). Group 1 *Shorea* (L, synapomorphies 18, bootstrap <50%) contains *Parashorea globosa* and *Shorea* species from sections *Shorea*, *Anthoshorea*, *Brachypterae* and *Neohopea*. *Parashorea globosa*, the putative sister to *Shorea*, is paired with *S. materialis* from section *Shorea* (synapomorphies 17, bootstrap 84%). *Parashorea* is nested well within the Group 1 clade and this placement contradicts the hypothesis that this taxon is the sister group to *Shorea*. With the exception of *P. globosa* and the monotypic section *Neohopea* (*S. isoptera*), all the *Shorea* sections included within Group 1 appear to be non-monophyletic.

Group 3 *Shorea* (M, synapomorphies 16, bootstrap <50%) consists of species from sections *Mutica*, *Shorea*, *Pachycarpae*, *Brachypterae* and *Ovalis*. It also includes *Dryobalanops lanceolata*. The placement of this taxon as the sister group to Group 3 suggests that it is more closely related to *Shorea* than *Hopea*, rather than *Dryobalanops* being the sister taxon to the ingroup as a whole. *Shorea* section

Pachycarpae (N, synapomorphies 14, bootstrap 56%) can only be considered monophyletic with the exclusion of *S. pilosa*.

The sister taxon to the clade containing Group 1 and 3 *Shorea* is Group 2 (J, synapomorphies 27, bootstrap 83%). It consists of *Shorea* species from section *Richetioides* subsection *Richetioides*, *S. selanica* (section *Brachypterae*) and *Hopea celebica* (Section *Hopea*). Two clades are apparent within Group 3. *Hopea celebica* falls within the first clade and *Shorea selanica* in the second clade. Even though *H. celebica* is nested within *Shorea* section *Richetioides* subsection *Richetioides*, it is defined by large number of autapomorphic changes (51) and this suggests it is very distinct from the *Shorea* species. Nonetheless, the section *Richetioides* subsection *Richetioides* is non-monophyletic in this analysis due to the inclusion of *S. selanica* and *H. celebica*.

Results from this analysis of ITS sequences show that *Shorea* is monophyletic only with the inclusion of *Parashorea globosa*, *Dryobalanops lanceolata* and *Hopea celebica*. *Shorea* contains four potential sub- groupings, but these do not accord with the existing infra-generic classification. *Hopea* is monophyletic in this analysis and has a sister group relationship with *Neobalanocarpus heimii*. However, the previously proposed infra-generic groupings in *Hopea* are also non-monophyletic.

A comparison of the topologies obtained from the two analyses with the inclusion and exclusion of indels (Figure 4.6) shows considerable differences in the arrangement of some internal clades. *Shorea exelliptica* now forms the sister lineage to the remainder of *Shorea*. Another significant difference between the two topologies is that when indels were included Group 2 *Shorea* (J, Figure 4.5) was the sister taxon to the Red Meranti group (L+M or Groups 1+3, Figure 4.5). In the analysis excluding indels, Group 2 instead appears as the sister to Group 3. Consequently, Group 1 *Shorea* moves to become the sister taxon to the pairing of Groups 2 and 3. However, the species arrangements within these groups does not differ between the two analyses, except that *Hopea nigra* no longer groups with *H. cernua* and *H. subalata,* but moves to a clade containing *H. pubescens* and *H. pierrei*.*

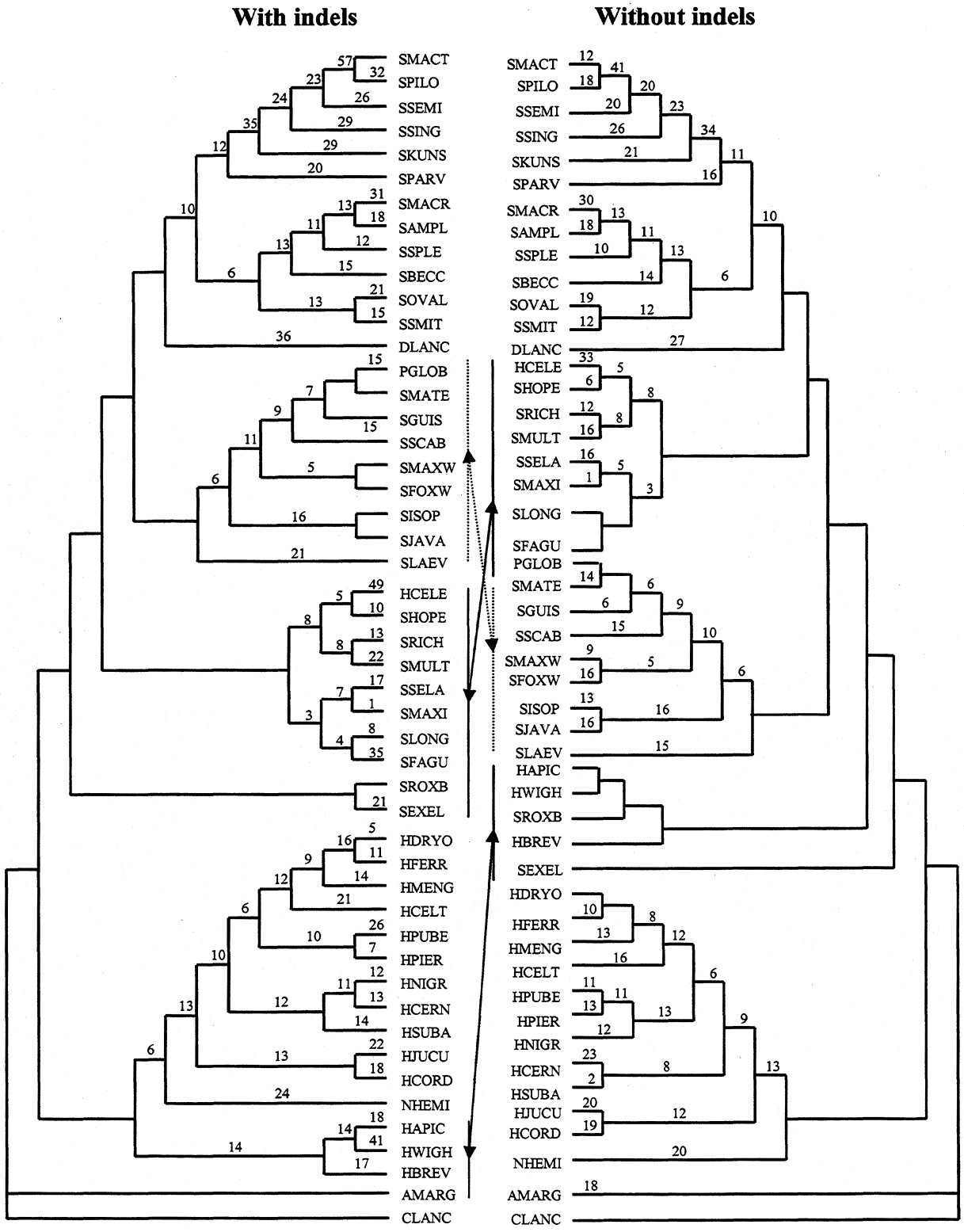


Figure 4.6 Comparison between topologies obtained from cladistic analysis of Dipterocarpaceae ITS sequences with indels included and excluded. Numbers above the branches are branch lengths and arrows indicate the major differences between the topologies.

4.7.3 Combined regions

4.7.3.1 Outline of the characters used

The combined alignment of both *trn*L-F and ITS consists of a total of 1844 characters, of which 388 are parsimony informative.

4.7.3.2 Topological features

A comparison was made between the topologies obtained with indels included in the analysis and those obtained after exclusion of the indels for this combined data set, since both the separate ITS and *trn*L-F analyses showed that inclusion of the indels resulted in significantly different topologies.

A total of 42 species were included in this combined analysis and a heuristic search as described above resulted in 53 most parsimonious trees (MPTs) with a length of 1885 steps. One of these shortest trees is presented in Figure 4.7. It has a CI of 0.42, HI of 0.58, RI of 0.53 and RC of 0.28 and many of the branches shown also appear in the strict consensus of all trees retained (indicated by heavy lines).

The two major ingroup clades (A and B) form a polytomy with the putative outgroup, *Dryobalanops lanceolata* (Figure 4.7). The first clade (A) consists of Group 3 *Shorea*. The second clade (B) includes Group 2 *Shorea*, the *Hopea* clade and Group 1 *Shorea* (C). The content of the groupings within *Shorea* is similar to those in the previous separate analyses of *trn*L-F and ITS.

Group 3 *Shorea* (A, synapomorphies 13, bootstrap 61%) consists of some species from sections *Mutica* and *Brachypterae,* all the species from section *Pachycarpae* included in the analysis and the monotypic section *Ovalis*. All the sections included in this clade appear to be non-monophyletic.

The second clade (B, synapomorphies 18, bootstrap <50%) consists of two further subgroupings—Group 1 *Shorea* (C) forms the sister taxon to a pairing of Group 2 *Shorea* (E) with the *Hopea* clade (F). Group 1 *Shorea* (synapomorphies 12, bootstrap 64%) has a similar arrangement to that shown in the ITS analysis with *S. laevis* placed as the sister taxon to the remainder of the species. Group 1 consists of some of

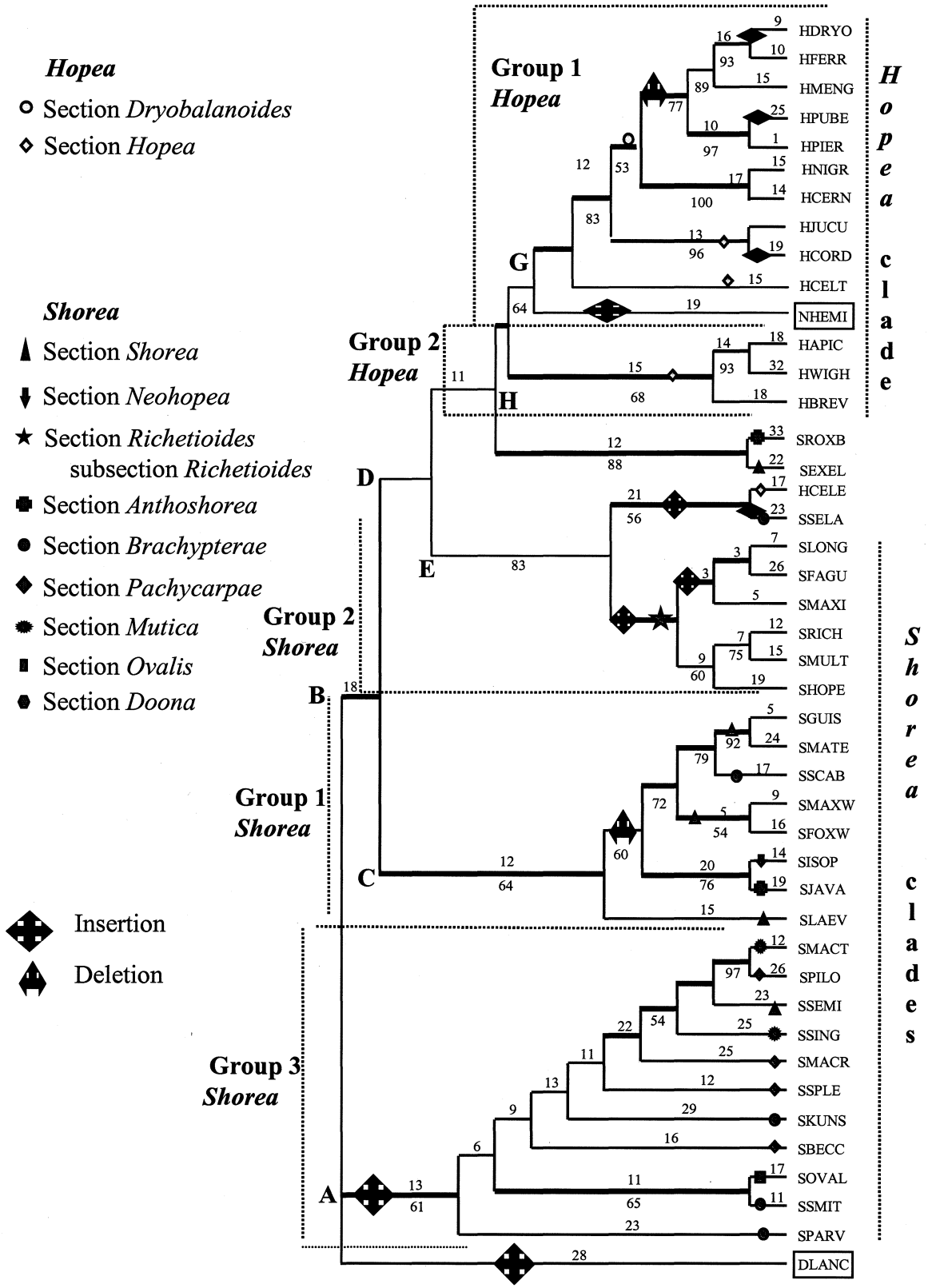


Figure 4.7 One of the most parsimonious trees derived from cladistic analysis of combined *trn*L-F and ITS sequences of selected Dipterocarpaceae taxa. Numbers above the branches are branch lengths, while bootstrap values of 50% and greater are shown below. Thicker branches are those that appear in the strict consensus of all trees. Taxon names in boxes are the putative outgroups.

the species from sections *Brachypterae* and *Anthoshorea,* the monotypic section *Neohopea* and most of section *Shorea*.

Group 2 *Shorea* (E, synapomorphies 21, bootstrap 83%) is composed of *Shorea selanica* (*S.* section *Brachypterae*), *Hopea celebica* (*H.* section *Hopea*) and the potentially monophyletic *Shorea* section *Richetioides* subsection *Richetioides*. The placements of *S. selanica* and *H. celebica* are problematic. *Shorea selanica* was included within Group 1 *Hopea* (containing all the species of section *Dryobalanoides*) in the *trn*L-F topology (Figure 4.3), but was nested within Group 2 *Shorea* in the ITS topology (Figure 4.5). *Hopea celebica* was part of section *Hopea* within the *trn*L-F topology and was grouped with Group 2 *Shorea* in the ITS topology. The results of this combined analysis are similar to those from the ITS topology, as both species are placed within Group 2 *Shorea.*

The next clade (F, synapomorphies 11, bootstrap <50%) consists of a clade which unites *Shorea exelliptica* and *S. roxburghii* with the *Hopea* clade. A pairing of two anomalous taxa, *Shorea exelliptica* and *S. roxburghii*, forms the sister taxon to the *Hopea* clade. The two *Shorea* taxa are problematic. They were excluded from the core ingroup clade in the *trn*L-F topology (Figure 4.3) and they were formed the sister group to Group 1 *Shorea* in the ITS topology (Figure 4.5). All analyses thus suggest that these two *Shorea* species are more closely related that previously thought, although this may simply be the result of long branch attraction.

Within the *Hopea* clade, there are two further subgroupings (G and H). The first clade (G, synapomorphies 0, bootstrap 64%) is referred to as Group 1 *Hopea* and contains all the members of section *Dryobalanoides* included in the analysis in addition to three species from section *Hopea* (*H. celtidifolia*, *H. jucunda* and *H. cordifolia*). *Neobalanocarpus heimii* is then placed as the sister taxon to Group 1. The second group of *Hopea* species (H, synapomorphies 15, bootstrap 68%) consists of the remaining members of section Hopea (H. *apiculata*, *H. wightiana* and *H. brevipetiolaris*).

The cladogram derived from an analysis of the combined molecular data thus suggests that both *Shorea* and *Hopea* are non-monophyletic. *Hopea* includes its

putative sister taxon, *Neobalanocarpus*, while *Shorea* is split into several groups and includes *H. celebica.*

4.8 Discussion

4.8.1 Sequence variability

The nature of the sequences obtained from the *trn*L-F and ITS regions differed. The *trn*L-F region is A+T rich, while ITS is a G+C rich region. The richness of A+T or G+C was mainly a result of the high number of repetitive regions. These repetitive motifs occurred in various sizes, ranging between one and 14 nucleotides occurring both in pure tandem or interspersed.. The short motifs are usually conserved among the taxa and even though there is mutation in the form of length polymorphism, it only accounted for less than 2 bp as exemplified by the poly-A and poly-T motifs. The conserved nature of the short motifs and very few mutations occurring within them may be due to slipped strand mispairing, which will result in homoplasious inferences of parallelism in the phylogenetic context (Mummenhoff *et al.*, 2001) and secondary structure formation (Nickrent *et al.*, 1993; Baldwin *et al.*, 1995; Muir *et al.*, 2001; Platas *et al.*, 2002; Costa *et al.*, 2002). Such short indels were thus not incorporated into the data scoring.

However, longer motifs (those of more than 2 bp) are usually phylogenetically informative. This was exemplified by indels 1, 2, 5 and 6 of *trn*L-F (Table 4.3) and indel 5 of ITS (Table 4.4). These indels were diagnostic for certain groups within genera (sections) or even at the generic level itself.

The variability of the sequences caused by repetitive nucleotides was therefore useful in inferring positional homology by introducing indels. However, most of the indels that were inferred within *trn*L-F and ITS sequences were compensated by non-repetitive sequences that were unique to certain groups. Some examples are indels 3, 4 and 7 of *trn*L-F (Table 4.3) and indels 1, 5, 6 and 7 of ITS (Table 4.4). Most of these non-repetitive indels of *trn*L-F are diagnostic at the generic level, while the ITS indels were diagnostic for infra-generic groupings.

Inclusion of the indels therefore contributed to considerable changes in clade arrangement, which to some extent provided better phylogenetic resolution in the analyses. It is therefore suggested that the indels in this analysis may have useful

additional phylogenetic signal and that they are a useful source of information for phylogenetic analysis at lower taxonomic levels. This has also been suggested by Ham *et al.* (1994), Mes and Hart (1994), and Sang *et al.* (1997).

Another type of mutational change that contributes to the sequence variability of *trn*L-F and ITS is nucleotide substitution. A comparison of the average pairwise divergence between species for the *trn*L-F and the ITS regions indicates that ITS has much higher rates of nucleotide substitution than *trn*L-F. Hence, distinguishing autapomorphic, synapomorphic and homoplasious substitutions should enable a comparison of the phylogenetic information yielded by these regions. The percentage of phylogenetically informative sites[18] is also much higher in the ITS region (31%) than in the *trn*L-F region (< 6%). This implies that the ITS has evolved more rapidly than *trn*L-F and should thus be a useful region for phylogenetic studies at lower taxonomic levels.

4.8.2 Incongruence between *trn*L-F and ITS topologies

Topological incongruence is commonly found in studies that include data from independent sources, such as nuclear and chloroplast genomes (Mason-Gamer and Kellogg, 1996). In order to be able to examine the incongruence between the topologies obtained from each region, two separate analyses were performed on the same taxa included in the combined data matrix and the level of bootstrap support assessed. Mason-Gamer and Kellogg (1996) suggested that comparisons between independent sources of data should not be made for nodes with bootstrap values of less than 70%, since weakly supported nodes only ambiguously represent patterns within each individual data set. The comparative topology is presented in Figure 4.8.

This topology is similar to those obtained by including all the species in separate analyses (Figures 4.3 and 4.5). There are three distinct observations that can be made when comparing the *trn*L-F and ITS topologies. The first is that the positions of the major clades differ between the two topologies. Group 3 *Shorea* is the sister taxon to the remainder of the ingroup in the ITS topology, whereas in the *trn*L-F topology the sister group to the main ingroup is a clade consisting of two anomalous *Shorea* species (*S. exeliptica* and *S. roxburghii*). Disregarding the position of these two taxa,

[18] The sites where substitutions are shared by two or more taxa.

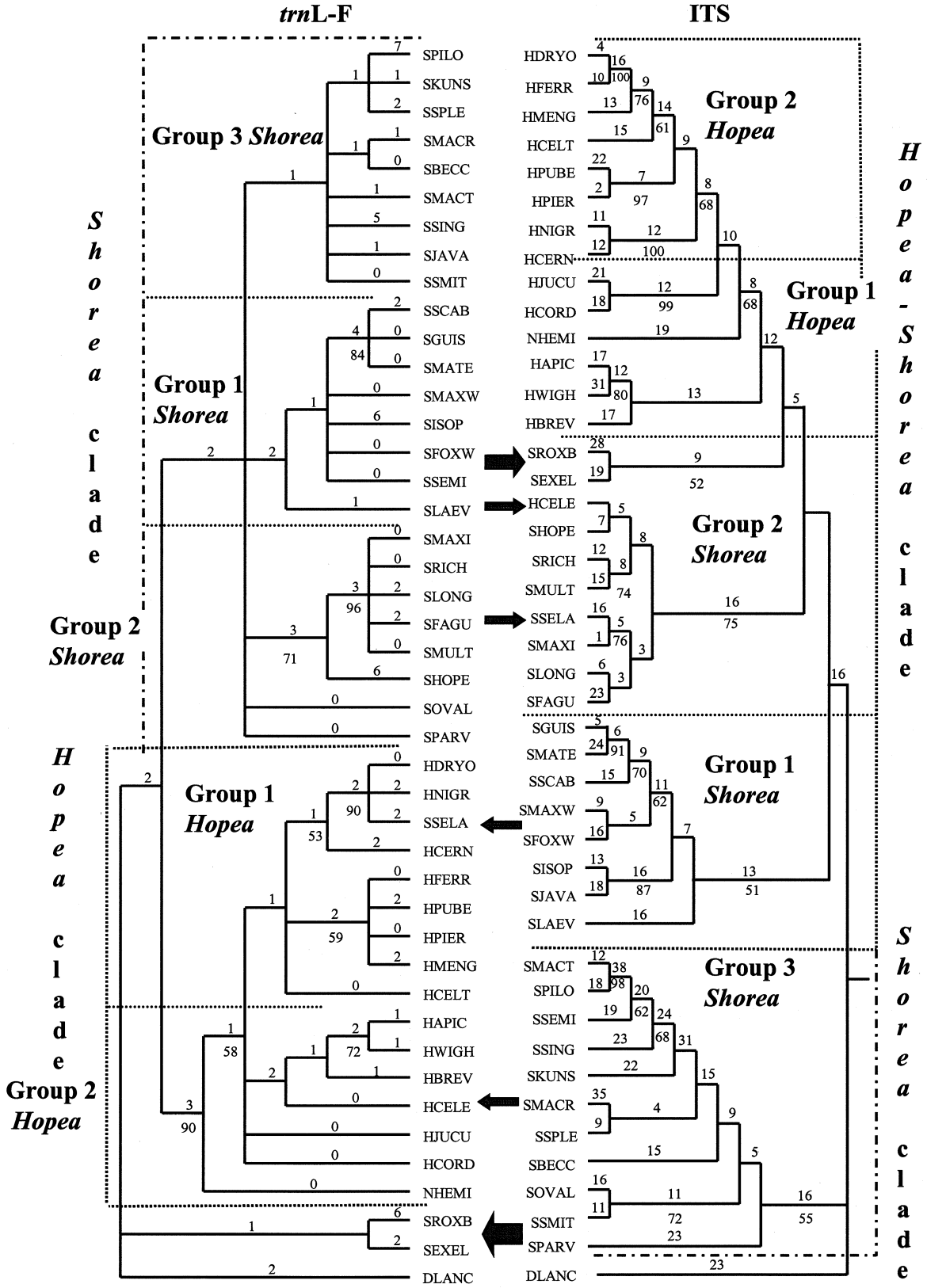


Figure 4.8 Comparison between topologies obtained from cladistic analysis of Dipterocarpaceae trnL-F (left-hand cladogram) and ITS (right-hand cladogram), excluding those characters and indels which were uninformative. Numbers above the branches are branch lengths and bootstrap values of 50% and greater are shown below. Arrows indicate the major differences between the topologies.

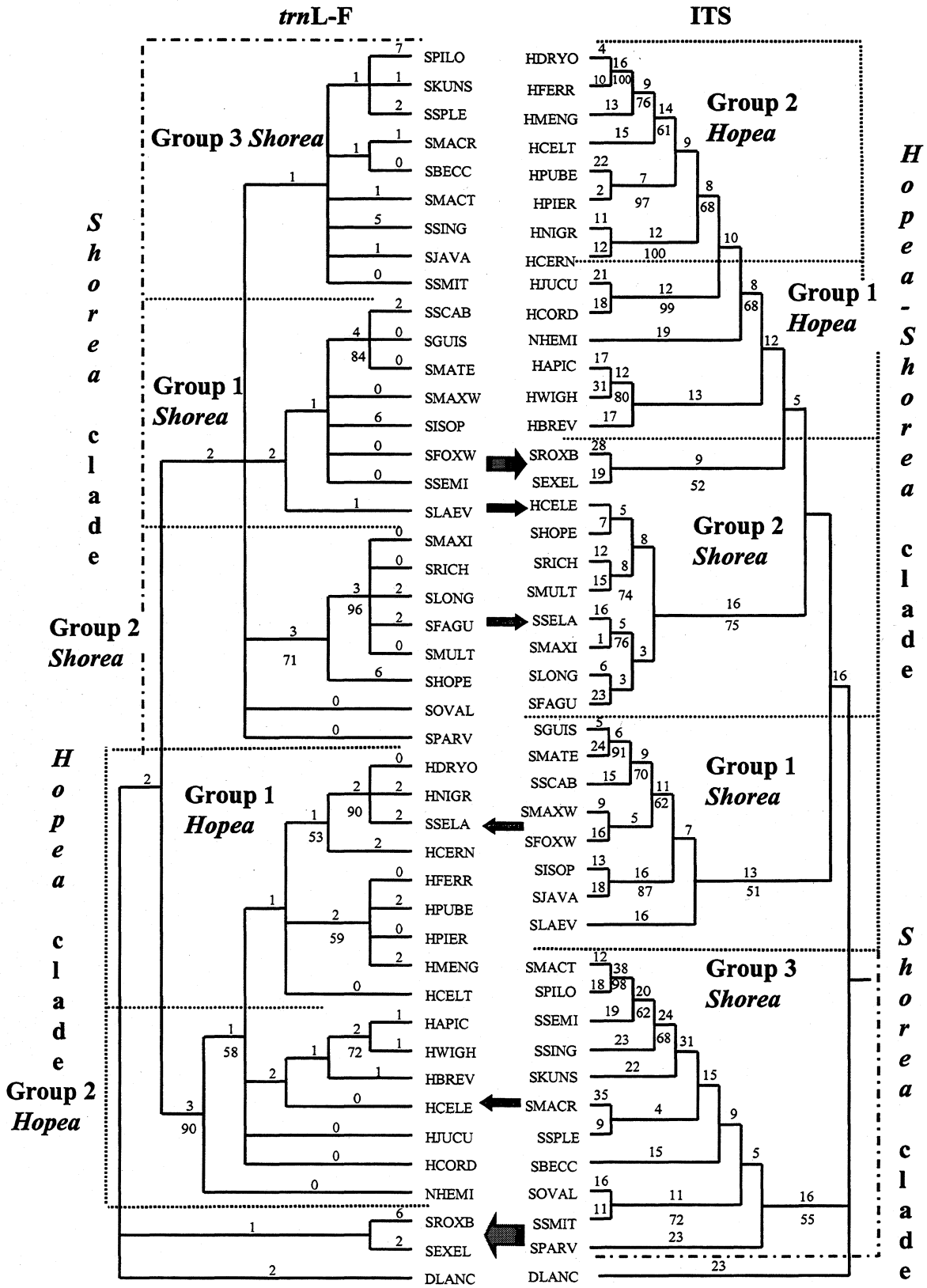


Figure 4.8 Comparison between topologies obtained from cladistic analysis of Dipterocarpaceae trnL-F (left-hand cladogram) and ITS (right-hand cladogram), excluding those characters and indels which were uninformative. Numbers above the branches are branch lengths and bootstrap values of 50% and greater are shown below. Arrows indicate the major differences between the topologies.

the major ingroup clade in the *trn*L-F topology contains *Hopea* and *Shorea* as two separate sister clades. In this analysis, *Hopea* is monophyletic with the inclusion of *Shorea selanica*. *Neobalanocarpus heimii* appears to to the closest relative to this *Hopea* clade, supporting its prior identification as the sister group to *Hopea*. The overall *Shorea* clade, however, is formed from a polytomy of 3 subgroups and two single species, *S. ovalis* and *S. parvistipulata*. The groupings are largely in accordance with the results from the other analyses (ITS and combined regions). The differences in the arrangement of the major clades may reflect the nature of the different markers used. Since the ITS is part of a multiple gene family, it is possible that the sequences used represent multiple copies. For species that contain multiple copies, there is likely to be more sequence variation and comparisons among the taxa may actually be being made from different sets of copies. By contrast, *trn*L-F is a single copy so that comparisons between sequences can only be made based on the same copy.

The second observation to be made when comparing the topologies from the two DNA regions is that the topology based on the ITS data set is more completely resolved than that from *trn*L-F data set, particularly at the internal nodes. For example, *Hopea cordifolia* and *H. jucunda* form a species pair within the main *Hopea* clade rather than forming part of a polytomy. Similarly, the positions of *Shorea parvistipulata* and *S. ovalis* are resolved in the ITS topology, rather than forming part of a larger *Shorea* polytomy. In addition, the positions of *Shorea maxwelliana, S. isoptera, S. foxworthyi* and *S. seminis* are also resolved in the ITS topology.

The third observation is the different phylogenetic positions of some taxa that have equivocal placements in both analyses. There are four taxa whose positions are equivocal in both topologies and in almost all the analyses, i.e. *Shorea roxburghii, S. exelliptica, S. selanica* and *Hopea celebica*. In the *trn*L-F cladogram shown in Figure 4.8, *S. roxburghii*, and *S. exelliptica* are excluded from the ingroup clade and form a polytomy with *Dryobalanops lanceolata* and the clade containing the majority of the ingroup taxa. The phylogenetic position of these two species is thus unresolved in the *trn*L-F topology. However, *S. roxburghii* and *S. exelliptica* form the sister taxon to the *Hopea* clade in the ITS cladogram.

Hopea celebica (*H.* section *Hopea*) and *Shorea selanica* (*S.* section *Brachypterae*) form part of *S.* section *Richetioides* in the ITS analysis, and thus invalidate the monophyly of this section. In the trnL-F topology, by contrast, *H. celebica* and *S. selanica* are nested within the *Hopea* clade in the *trn*L-F topology. *H. celebica* is grouped with other members of section *Hopea* and *S. selanica* is grouped with members of *H.* section *Dryobalanoides*. Thus, these two species cannot confidently be placed within either genus. Some possible explanations for these results include a taxonomic error or different phylogenetic histories. Misidentification of the taxa is one possibility, mainly because the sequences were obtained from vegetative material collected in the Bogor Botanic Garden. However, the identity of these two collections has been checked against several herbarium sheets obtained from other herbaria (Appendix 3A). In addition, if a misidentification had occurred it is more likely that these two species would be consistently "misplaced" in either the *Hopea* or *Shorea* clade. However, from the cladograms obtained it seems more likely that these species have had a complex evolutionary history (Table 4.5).

Table 4.5 Comparison of the phylogenetic position of *Hopea celebica* and *Shorea selanica* within the morphological and molecular topologies.

Taxon	Morphology	*trn*L-F	ITS
Hopea celebica	*Hopea*	*Hopea*	*Shorea*
Shorea selanica	*Shorea*	*Hopea*	*Shorea*

4.8.3. Phylogenetic inferences of the putative outgroup, sister taxa and the ingroup

Several species from various genera of Dipterocarpaceae were selected as putative outgroups for this molecular study. This includes species of *Dipterocarpus*, *Anisoptera*, *Cotylelobium*, *Dryobalanops*, *Neobalanocarpus* and *Parashorea*. *Dipterocarpus*, *Anisoptera* and *Cotylelobium* are relatively distantly related to the ingroup, and therefore have a greater divergence from the ingroup than the other putative sister taxa used (*Dryobalanops*, *Parashorea* and *Neobalanocarpus heimii*).

A pairing of *Dryobalanops aromatica* and *D. lanceolata* is the sister taxon to the remainder of the ingroup according to the *trn*L-F phylogeny. By contrast, a close relationship between *Dryobalanops* and the variable genus *Shorea*, particularly Group

3 *Shorea*, was indicated in the morphological study (Chapter 3). Currently, the morphological characters that define *Dryobalanops* are features of the wood anatomy, parallel leaf nervation and five equal spatulate fruit wings. These characters have been shown to be homoplasious and thus do not provide a clear phylogenetic signal. In addition, *Dryobalanops* is considered to be an intermediate between Tribe Shoreae and Dipterocarpeae (Maury-Lechon, 1979 in Maury-Lechon & Curtet, 1998), which may explain its unresolved phylogenetic position in previous analyses (Dayanandan *et al.*, 1999). The evidence from the analysis of *trn*L/F sequences supports a sister group relationship between *Neobalanocarpus* and a clade containing both *Hopea* and *Shorea*, indicating that the morphological similarities may be the result of homoplasy.

Neobalanocarpus heimii has been previously suggested to be the sister taxon of *Hopea* and these two taxa are always placed in a close relationship in these analyses. However, the nature of the relationship varies between analyses. The trnL-F results indicate a sister relationship between *N. heimii* and the *Hopea* clade, as do the results from ITS with indels included. However, the results from ITS without indels and from the combined analysis show a sister relationship between *N. heimii* and a subset of the *Hopea* species included. A close relationship of *Neobalanocarpus heimii* to *Hopea* section *Hopea* was suggested previously by Ashton (1982) on the basis of inflorescence features, fruit embryo and germination mode. A previous molecular phylogeny of Dipterocarpaceae suggested a close relationship between *Neobalanocarpus* and *Hopea* section *Dryobalanoides* subsection *Dryobalanoides* (Tsumura *et al.*, 1996). *Hopea* and *N. heimii* share similarities of wood anatomy in the possession of medium-sized vessels and storied rays (Parameswaran and Gotwald, 1979 in Maury-Lechon & Curtet, 1998), anthocyanin development (Bate-Smith and Whitmore, 1959) and bark morphology (Whitmore, 1962). However, *Neobalanocarpus* is endemic to Peninsular Thailand and has a distinctive "semi-broad" anther appendage that is not possessed by any *Hopea* species. Clearly, it is not possible to separate the two genera only on the basis of morphological characters since they have been shown to be homoplasious, and synapomorphic changes throughout evolutionary history may have biased the phylogenetic inferences. However, the inclusion of *Neobalanocarpus heimii* in *Hopea* section *Hopea* confirms its close relationship to the section. This monotypic genus may have the same ancestor as all of the *Hopea* species and its restriction to Peninsular Thailand may

suggest that it did not continue to disperse to other parts of the geographic region, such as Malesia.

Parashorea lucida forms part of the *Shorea* clade in the *trn*L-F topology and *P. globosa* is nested in the *Shorea* clade in the ITS topology. These results are not in agreement with those from previous studies of the molecular phylogeny of Dipterocarpaceae using cpDNA, which placed *Parashorea* as the sister to *Shorea* (Tsumura *et al.*, 1996; Kajita *et al.*, 1998).

4.8.4 Phylogenetic relationships of *Hopea* and *Shorea*

Neither *Hopea* and *Shorea* appear to be monophyletic based on the results of the molecular analyses. In the combined analysis, the *Hopea* clade is nested within *Shorea.* In the combined analysis, *Hopea* section *Dryobalanoides* forms a monophyletic group (Figure 4.7), although with only moderate support from the bootstrap (53%). By contrast the members of section *Hopea* form four distinct lineages in the cladogram.

Neobalanocarpus heimii is placed as the sister group to only a subset of the *Hopea* species (Group 1) in the combined analysis. A clade of species from section *Hopea* (Group 2) then forms the sister group to this combined clade of *Hopea* Group 1 and *Neobalanocarpus*. *Hopea* s.l. could be made monophyletic (fig. 4.7) with the inclusion of *Neobalanocarpus* within it, however, none of the six major classifications of Dipterocarpaceae (Table 2.1, Chapter 2) has ever suggested the inclusion of *Neobalanocarpus* in *Hopea*.

Shorea consists of 10 sections (Ashton, 1982) and eight of these sections are included in the combined *trn*L-F and ITS analysis. The taxa excluded are section *Doona* and the monotypic section *Pentacme*, due to difficulty in obtaining useful samples for DNA extraction. Of the eight sections included, the results indicate that only some are monophyletic. The overall arrangement within *Shorea* is to some extent in accordance with the timber groupings *sensu* Symington (1943) and Meijer and Wood (1964); and correlated with the infra-generic divisions proposed by Ashton (1982). Symington (1943) and Heim (1892) have previously suggested that floral characters and field characters of bark and wood anatomy were correlated. The placement of a

few of the species/sections in unexpected positions in the phylogeny may reflect past hybridization events.

Group 1 *Shorea* is dominated by members of section *Shorea* (Balau group) and this corresponds to the results of the morphological study. This results supports Maury-Lechon's argument that section *Shorea* is one of the main taxa of sub-family Dipterocarpoideae in which new forms arise though diversification. The monophyly of section *Shorea* is called into question by the inclusion of the monotypic section *Neohopea* (i.e. *Shorea isoptera*) and members of sections *Anthoshorea* and *Brachypterae*. The phylogenetic placement of section *Neohopea* remains unclear. Burck (1887) regarded this species as a separate genus, *Isoptera*, but Symington (1943) and Ashton (1982) transferred it into *Shorea* with sectional status. However, Meijer and Wood (1964) included *Isoptera* within *Eushorea* (which consists of taxa from both sections *Shorea* and *Anthoshorea*), and this placement is to a certain extent confirmed by this molecular phylogenetic analysis.

Group 2 *Shorea* is made up of section *Richetioides* subsection *Richetioides*, the monophyly of which is suggested by all the analyses. This may indicate that the evolution of differentiated and undifferentiated floral features postulated by Ashton (1982) is tracked by the molecular data. This group is also known as "Yellow Meranti" or "Damar Hitam" on the basis of wood and bark anatomy (Symington, 1943) and it exhibits certain morphological characters not shared with other sections of *Shorea*. It contains species with sub-equal calyx lobes and this feature has been debated among taxonomists, as some have considered the taxa to be distinct enough to merit generic status (Heim, 1891; Meijer and Wood, 1964; Maury, 1978 in Maury-Lechon and Curtet, 1998). The monophyly of subsection Richetioides suggests that a unique evolutionary pathway was followed by the group and this may support Heim (1892), who proposed recognition of this group at the generic level as *Richetia*.

Group 3 *Shorea* consists of taxa from sections *Mutica*, *Rubellae*, *Brachypterae*, *Pachycarpae* and *Ovalis*. Group 3 is nearly identical to *Rubroshorea* (Meijer and Wood, 1964) and Red Meranti (Symington, 1943) with a few exceptions—the inclusion of *Shorea seminis* (section *Shorea*) and the exclusion of *Shorea scaberrima* (section *Brachypterae*). Hence, section *Brachypterae* is a non-monophyletic group.

The non-monophyly of section Brachypterae is also indicated by the sister relationship between *Shorea smithiana* and the monotypic section *Ovalis* in the combined analysis (fig. 4.7).. Section *Pachycarpae*, represented in the combined molecular data set by only four of its 10 species. This section is assumed to have undergone rapid radiation on the island of Borneo. This section does not appear to be monophyletic, although an analysis including all 10 species may be required to test this.

The phylogenetic position of four anomalous taxa—*Shorea roxburghii* (*S.* section *Anthoshorea*), *S. exelliptica* (*S.* section *Mutica*), *Shorea selanica* (*S.* section *Shorea*) and *Hopea celebica* (*H.* section *Hopea*)—is interesting. The phylogenetic positions *of H. celebica* and *S. selanica* has already been examined in a previous section of this chapter (4.82). *Shorea* section *Anthoshorea* has been questioned by Brandis (1895) and Ashton (1982) for being intermediate between *Hopea* and *Shorea* and was recognised as the genus *Parahopea* by Heim (1892).

Other species of section *Anthoshorea* included within the molecular analyses are scattered throughout the cladograms obtained, which makes this section subject to debate. Section *Anthoshorea* appears to be polyphyletic, since species from this section are placed in various different lineages within *Shorea* and *Hopea*. The majority of the species in this section are included in Meranti Pa'ang (White Meranti) according to the timber groupings (Symington, 1943), so the results from the present analyses do not confirm the integrity of the timber grouping for this section. *Shorea* section *Anthoshorea* resembles sections *Doona* and *Pentacme*, *Dryobalanops*, *Neobalanocarpus heimii* and *Cotylelobium* based on similarities of the embryo, seedling and pollen surface (Maury-Lechon and Curtet, 1998). It is perhaps not surprising therefore that the molecular analyses found that the closest relatives of each member of section *Anthoshorea* are not other members of that section. Nevertheless, section *Anthoshorea* requires further examination in order to clarify both the phylogenetic relationships within the section and between it and other groups.

4.8.5 Taxonomic implications

It is often a cause for debate as to whether the results of phylogenetic studies can be applied to taxonomic rankings. It may not be a problem for a monophyletic group to be acknowledged in a taxonomic concept, but the taxonomic status of paraphyletic groups still causes debate. With the exception of four ambiguous taxa identified in all of the molecular analyses, the results of this study suggest that *Hopea* and *Shorea* as currently defined cannot be separated into two distinct taxa.

The only potential natural grouping within *Hopea* is section *Dryobalanoides*, thus this section can be maintained. The placement of *Neobalanocarpus heimii* into section *Hopea* is also proposed based on the results of this study. However, the taxonomic status of Section *Hopea* can not be ascertained, since the section is obviously non-monophyletic. A further detailed study incorporating more species may be required to clarify the taxonomic status of this group.

The groupings within *Shorea* are more complex. A comparison of the results from the phylogenetic analyses of molecular data to the existing taxonomic groupings shows that there may be a need to re-establish the Balau group or subgenus *Eushorea sensu* Meijer and Wood (1964). This group consists of section *Shorea* and some species from section *Anthoshorea sensu* Ashton (1982). It is also proposed to recognise Meranti Damar Hitam *sensu* Symington (1943) or subgenus *Richetia sensu* Meijer and Wood (1964) or to maintain Section *Richetioides sensu* Ashton (1982). Finally, it is proposed that the Red Meranti group *sensu* Symington (1943) with the possible inclusion of *Parashorea* or subgenus *Rubroshorea sensu* Meijer and Wood (1964) be recognised. The above three groups of *Shorea* taxa can be either assigned infrageneric or generic rank.

4.9 Conclusions

Using a broad generic delimitation, data from the ITS and *trn*L-F regions yielded similar trees with a few exceptions in the placement of two anomalous taxa, *Hopea celebica* and *Shorea selanica*. Despite these exceptions, this molecular study has provided strong evidence of the broad non-monophyly of *Shorea* and the potential monophyly of *Hopea* with some recircumscription. The only natural grouping to be recognised within the *Hopea* clade is section *Dryobalanoides sensu* Ashton (1982). It

is also possible that *Neobalanocarpus heimii* should be subsumed into *Hopea*. However, the phylogenetic position of Section *Hopea* is still unresolved.

With regard to the groupings within *Shorea*, the results of this molecular study support the classifications by Meijer and Wood (1964) and Symington (1943). Both systems used mainly timber characters and thus provided the practical advantage of an identification system for taxa which are important timber species. The putative sister taxon to *Shorea*, *Parashorea,* may in reality be part of *Shorea*. However, further analyses that incorporate taxa from *Shorea* sections *Doona* and *Pentacme* and more species from section *Anthoshorea* are required in order to make confident inferences about the relationships among infrageneric groupings as well as those at the generic level.

Appendix 4A Taxa selected for the molecular study

Outgroup

Species	Abbreviation	Voucher; Herbarium; Locality	trnL intron; trnL/F spacer	ITS-1; ITS-2
Neobalanocarpus heimii (King) Ashton	NHEMI	28645; FRIM,; Tree no.36, Dipterocarp arboretum FRIM	AB006400*; AB006417*	AY026657; AY026713
*Parashorea lucida**Kurz.	PLUCI	—	AB006399;AB006416	—; —
Parashorea globosa Symington	PGLOB	KY 867; BO, CANB, WAN; Bogor Botanic Garden, W. Java	—; —	AY026658; AY026714
Dryobalanops aromatica Gaertn. f.	DAROM	KY 805; BO, CANB, WAN; Lempake. E. Borneo	AY026530; AY026585	—; —
Dryobalanops lanceolata Burck	DLANC	KY 806; BO, CANB, WAN; Lempake. E. Borneo	AY026531; AY026586	AY026640; AY026698
Dipterocarpus retusus Blume	DRETU	CANB 109473; CANB;-	AY026529; AY 026584	—; —
Dipterocarpus confertus Slooten	DCONF	KY 854; BO, CANB, WAN; Bukit Bengkirai, E. Borneo	AY 026528; AY 026583	—; —
Dipterocarpus kerrii *King	DKERI	—	AB006392; AB006409	—; —
Anisoptera marginata Korth.	AMARG	KY 851, BO, CANB, WAN; PT ITCI Arboretum, E. Borneo	—; —	AY 026638; AY 026695
Cotylelobium lanceolatum Craib	CLANC	KY 871, BO; CANB, Bogor Botanic Garden, W. Java	—; —	AY 026639; AY 026696
Total number of the outgroup species			7	5

Appendix 4A (continued)

Hopea

Section	Subsection	Species	Abbreviation	Voucher; Herbarium; Locality	trnL intron; trnL/F spacer	ITS-1; ITS-2
Dryobalanoides	*Dryobalanoides*	*H. pubescens* Ridley	HPUBE	10478; FRIM; Tree no. 45 Dipterocarp Arboretum FRIM	AY 026544; AY 026599	AY 026654;-
		H. mengerawan Miq.	HMENG	KY 817; BO, CANB, WAN; Wanariset, E. Borneo	AY 026541; AY 026596	AY 026650; AY 026708
		H. cernua Teijsm. & Binn.	HCERN	KY 802; BO, CANB, WAN; Lempake, E. Borneo	AY 026536; AY 026591	AY 026645; AY 026703
		***H. dryobalanoides* Miq.**	HDRYO	FRI 27773; FRIM; Tree no. 784, Dipterocarp Arboretum FRIM	AY 026538; AY 026593	AY 026647; AY 026705
		H. ferruginea Parijs.	HFERR	KY 826; BO, CANB, WAN; Meratus Mt, E. Borneo	-; AY 026594	AY 026648; AY 026706
		H. pierrei Hance	HPIER	KY 858; BO; CANB, Bogor Botanic Garden, W. Java	AY 026543; AY 026598	AY 026653; AY 026710
	Sphaerocarpa	*H. nervosa* King	HNERV	SAN 121800; FRIM; Tree no. 21, Dipterocarp Arboretum FRIM	AB006401*; AB006418*	—; —
		H. nigra Burck	HNIGR	SC 04; BO; Bogor Botanic Garden, W. Java	AY 026542; AY 026597	AY 026652; AY 026709
		H. subalata Symington	HSUBA	29770; FRIM; Tree no. 149 Dipterocarp Arboretum FRIM -	—; —	AY 026655; -

Appendix 4A (continued)

Hopea

Section	Subsection	Species	Abbreviation	Voucher; Herbarium; Locality	trnL intron; trnL/F spacer	ITS-1; ITS-2
Hopea	*Hopea*	*H. celtidifolia* Kosterm.	HCELT	CANB 95666; CANB; -	AY 026535; AY 026590	AY 026644; AY 026702
		H. brevipetiolaris (Thw.) Ashton Ashton	HBREV	KY 860; BO; CANB, Bogor Botanic Garden, W. Java	AY 026533; -	AY 026642; AY 026700
		H. jucunda Thw.	HJUCU	KY 870; BO, CANB; Bogor Botanic Garden, W. Java	AY 026540; AY 026595	AY 026649; AY 026707
		H. cordifolia Trim	HCORD	SC 01, BO; Bogor Botanic Garden, W. Java	AY 026537; AY 026592	AY 026646; AY 026704
		H. celebica Burck	HCELE	KY 869; BO; CANB, Bogor Botanic Garden, W. Java	AY 026534; AY 026589	AY 026643; AY 026701
	Pierrea	*H. apiculata* Symington	HAPIC	80225; FRIM; Tree no 118, Dipterocarp Arboretum FRIM	AY 026532; AY 026587	AY 026641; AY 026699
		H. wightiana Miq. ex Dyer	HWIGH	KEP 76643; FRIM; Tree no. 141, Dipterocarp Arboretum FRIM	AY 026545; AY 026600	AY 026656; AY 026712
Total number of *Hopea* species					15	15

Appendix 4A (continued)

Shorea

Section	Subsection	Species	Abbreviation		trnL intron; trnL/F spacer	ITS-1; ITS-2
Shorea	*Shorea*	*S. guiso* Blume	SGUIS	SC 03; BO; Bogor Botanic Garden, W. Java	AY 026551; AY 026609	AY 026667; AY 026720
		S. foxworthyi Symington	SFOXW	FRI 22180; FRIM; Tree no. 81, Dipterocarp Arboretum FRIM	AY 026550; AY 026608	AY 026666; AY 026719
		S. exelliptica Meijer	SEXEL	KY 837; BO, CANB, WAN; Meratus Mt., E. Borneo	AY 026548; AY 026606	AY 026664; AY 026717
		S. seminis V. Slooten	SSEMI	KY 857; BO; Bogor Botanic Garden, W. Java	AY 026576; AY 026633	AY 026690; AY 026742
		S. materialis Ridley	SMATE	KY 862; BO; Bogor Botanic Garden, W. Java	AY 026561; AY 026619	AY 026678; AY 026729
	Barbata	*S. laevis* Ridley	SLAEV	KY 810; CANB, BO, WAN; Bukit Suharto, E. Borneo	AY 026557; AY 026615	AY 026673; AY 026725
		***S. maxwelliana* King**	SMAXW	5428; FRIM; Tree no. 75, Dipterocarp Arboretum FRIM	AY 026563; AY 026621	AY 026680; AY 026731
*Neohopea***		*S. isoptera* Ashton	SISOP	-; -FRIM; Tree no. 540, Dipterocarp Arboretum FRIM	AY 026553; AY 026611	AY 026669; AY 026722

Appendix 4A (continued)

Shorea

Section	Subsection	Species	Abbreviation		trnL intron; trnL/F spacer	ITS-1; ITS-2
Richetioides	*Richetioides*	*S. richetia* Symington	SRICH	-; FRIM; Tree no. 551, Dipterocarp Arboretum FRIM	AY 026572; AY 026629	AY 026686; AY 026738
		S. multiflora (Burck) Symington	SMULT	KY 859, BO; Bogor Botanic Garden, W. Java	AY 026565; AY 026622	AY 026681; AY 026732
		S. longisperma Roxb.	SLONG	KY 840; BO, CANB, WAN; Meratus Mt., E. Borneo	AY 026559; AY 026617	AY 026675; AY 026726
		S. hopeifolia (Heim) Symington	SHOPE	KY 846; BO, CANB, WAN; Meratus Mt., E. Borneo	AY 026552; AY 026610	AY 026668; AY 026721
		S. maxima (King) Symington	SMAXI	FRI 19346; FRIM; Tree no. 148, Dipterocarp Arboretum FRIM	AY 026562; AY 026620	AY 026679; AY 026730
		***S. faguetiana* Heim**	SFAGU	KY 819; BO, CANB, WAN; Wanariset, E. Borneo	AY 026549; AY 026607	AY 026665; AY 026718
Anthoshorea		***S. roxburghii* G. Don**	SROXB	Lei 2397; FRIM; Tree no. 424, Dipterocarp Arboretum FRIM	AY 026573; AY 026630	AY 026687; AY 026739
		S. javanica Koord. & Valet	SJAVA	KY 864, BO; CANB; Bogor Botanic Garden, W. Java	AY 026554; AY 026612	AY 026670; AY 026723
		*S. bracteolata**Dyer	SBRAC	—	AB006398; AB006415	—; —

Appendix 4A (continued)

Shorea

Section	Subsection	Species	Abbreviation		trnL intron; trnL/F spacer	ITS-1; ITS-2
Brachypterae	*Smithiana***	***S. smithiana* Symington**	SSMIT	KY 832; CANB, BO, WAN; Meratus Mt., E. Borneo	AY 026578; AY 026635	AY 026692; AY 026744
	Brachypterae	*S. selanica* Blume	SSELA	KY 832; BO; Haurbentes, W. Java	AY026575; AY026632	AY026689; AY026741
		S. parvistipulata Heim	SPARV	KY 825; BO, CANB, WAN; Meratus Mt., E. Borneo	AY 026569; AY 026626	AY 026683; AY 026736
		S. johorensis Foxw.	SJOHO	KY 811; BO, CANB, WAN; Wanariset, E. Borneo	AY 026555; AY 026613	—; —
		S. scaberrima Burck	SSCAB	KY 863; BO; Bogor Botanic Garden, W. Java	AY 026574; AY 026631	-; AY 026740
		S. balangeran Burck	SBALA	KY 873; BO; CANB; Bogor Botanic Garden, W. Java	AY 026546; AY 026604	—; —
		S. palembanica Miq.	SPALE	-; -; Bogor Botanic Garden W. Java	AY 026567; AY 026624	—; —
		S. kunstleri King	SKUNS	FRI 32671; FRIM; Tree no. 61, Dipterocarp Arboretum FRIM	AY 026556; AY 026614	AY 026672; AY 026724

Appendix 4A (continued)

Shorea

Section	Subsection	Species	Abbreviation		trnL intron; trnL/F spacer	ITS-1; ITS-2
Pachycarpae		*S. pilosa* Ashton	SPILO	S 22380; FRIM; Dipterocarp Arboretum FRIM	AY 026570; AY 026627	AY 026684; AY 026737
		S. splendida (De Vriese) Ashton	SSPLE	KEP 98877; FRIM; Tree no. 592, Dipterocarp Arboretum FRIM	AY 026579; AY 026636	AY 026693; AY 026745
		S. stenoptera Burck	SSTEN	KY 861; BO; CANB; Bogor Botanic Garden, W. Java	AY 026580; AY 026637	—; —
		***S. macrophylla* (De Vriese) Ashton**	SMACR	TL 1073, HUH; Gunung Palung, W. Borneo	AY 026560; AY 026618	AY 026676; -
		S. amplexicaulis Ashton	SAMPL	5331; HUH; Central Kalimantan	—; —	AY 026660; AY 026715
		S. beccariana Burck	SBECC	S 29174; FRIM; Dipterocarp Arboretum FRIM	AY 026547; AY 026605	AY 026662; AY 026716
		S. pinanga Scheff.	SPING	KY 866, BO, CANB; Bogor Botanic Garden, W. Java	AY 026571; AY 026628	—; —

Appendix 4A (continued)

Shorea

Section	Subsection	Species	Abbreviation		trnL intron; trnL/F spacer	ITS-1; ITS-2
Mutica	*Auriculatae*	*S. macroptera* Dyer	SMACT	FRI 39929; FRIM; Tree no. 44, Dipterocarp Arboretum FRIM	AB006396*; AB006413* .	AY 026677; AY 026728
	Mutica	*S. leprosula* Miq.	SLEPR	KY 815; BO, CANB, WAN; Wanariset, E. Borneo	AY 026558; AY 026616	—; —
		S. singkawang Burck	SSING	FRI 25407; FRIM; Dipterocarp Arboretum FRIM	AY 026577; AY 026634	AY 026691; AY 026743
		***S. parvifolia* Dyer**	SFOLI	KY 849; BO, CANB, WAN; PT ITCI Arboretum E. Borneo	AY 026568; AY 026625	—; —
*Ovalis***		***S. ovalis* Blume**	SOVAL	KY 814; CANB, BO, WAN; Wanariset, E. Borneo	AY 026566; AY 026623	AY 026682; AY 026733
Doona		*S. cordifolia* (Thw.) Ashton	SCORD	SC 08; BO; Bogor Botanic Garden, W. Java	**	—; —
Total number of *Shorea* species					38	29
Number of taxa within each target region					60	49
Total number of taxa for both regions					44	

*: taken from Genbank database.
**: sequences are not submitted to the Genbank but available upon request.

Appendix 4B List of aligned sequence of *trn*L-F and ITS

```
          [1]
NHEMI     CCTGCTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
PLUCI     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
DAROM     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAA---GGGCAATCCTGAGCCAAATCCTGTTTAA
DLANC     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTCA
DRETU     CCTACTGAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
DKERI     CCTACTGAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
DCONF     CCTACTGAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
HPUBE     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
HMENG     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAAA-GGGCAATCCTGAGCCAAATCCTGTTTAA
HCERN     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
HDRYO     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
HFERR     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
HPIER     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAAA-GGGCAATCCTGAGCCAAATCCTGTTTAA
HNERV     CCT-CTAAGTGATA-CTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
HNIGR     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAAAAGGGCATTCTTGAGCCAAATCCTGTTTAA
HCELT     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
HCELE     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAAA-GGGCAATCCTGAGCCAAATCCTGTTTAA
HAPIC     CCTACTGAGTGATTACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAAA-GGGCAATCCTGAGCCAAATCCTGTTTAA
HWIGH     CCTACTGAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
HBREV     CCTACTGAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
HJUCU     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
HCORD     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAA---GGGCAATCCTGAGCCGAATCCTGTTTMA
SMACT     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAAA-GGGCAATCCTGAGCCAAATCCTGTTTAA
SPILO     CCTACTAAGTGATAACTTTCCA-TTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SKUNS     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAAA-GGGCATTCCTGAGCCAAATCCTGTTTAA
SSELA     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAAA-GGGCAATCCTGAGCCAAATCCTGTTTAA
SMAXI     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAA---GGGCAATCCTGAGCCAAATCCTGTTTAA
SSCAB     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SSING     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAAA-GGGCAATCCTGAGCCAAATCCTGTTTAA
SRICH     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SMAXW     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAAA-GGGCAATCCTGAGCCAAATCCTGTTTAA
SISOP     CCT-CTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SGUIS     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SMACR     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SSPLE     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SLONG     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SFOXW     CCTACTAAGCGATC-CTTTCAA-TTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATTCTGTTCCAAATCCTGTTTAA
SHOPE     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SFAGU     CCTACTTAGTGATGACTCTCCAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SBECC     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SMULT     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
```

Appendix 4B (continued)

```
SOVAL     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGGAAAAAAAAA-GGGCAATCCTGAGCCAAATCCTGTTTAA
SROXB     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SMATE     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SEXEL     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SJAVA     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAAA-GGGCAATCCTGAGCCAAATCCTGTTTAA
SSMIT     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SSEMI     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SPARV     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA-GGGCAATCCTGAGCCAAATCCTGTTTAA
SPALE     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SFOLI     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAAA-GGGCAATCCTGAGCCAAATCCTGTTTAA
SSTEN     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SLAEV     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCA-T-CTGAGCCAAATCCTGTTTAA
SPING     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SCORD     CCTACTAAG-GATAACTTTCAAATTCAGAGAAACCCTGG-AAAAAAAA--GGGCGACTCTGAGCCAAATCCTGTTGAA
SBALA     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SJOHO     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAAAAAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SLEPR     CCTACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAAAA-AAAA--GGGCAATCCTGAGCCAAATCCTGTTTAA
SBARC     CCTACTAAGTGATAACTT-CAA-TTCAGAGAAACCCTGGAAAAA------GGGCAATCCTGAGCCAAATC-TGTTTAA
AMARG     ----------------------------------------------------------------------------
PGLOB     ----------------------------------------------------------------------------
CLANC     ----------------------------------------------------------------------------
HSUBA     ----------------------------------------------------------------------------
SAMPL     ----------------------------------------------------------------------------

                                [100]
NHEMI     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCATTGGTAAAAGAATACTT
PLUCI     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
DAROM     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACCGCTTTTGCGTTGGTAAAAGAATACTT
DLANC     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACCGCTTTTGCGTTGGTAAAAGAATACTT
DRETU     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGACCAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
DKERI     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGACCAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
DCONF     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGACCAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
HPUBE     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAA-AATACTT
HMENG     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
HCERN     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
HDRYO     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
HFERR     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
HPIER     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
HNERV     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
HNIGR     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCCTTGGCAAAAGAATACTT
HCELT     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
HCELE     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
HAPIC     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAAAATACTT
```

Appendix 4B (continued)

```
HWIGH     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAAAACTT
HBREV     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
HJUCU     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
HCORD     AAAAGCGTTCAAAAAGAAATGGGATAGGTGCAAAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCRTTGGTAAAAGAATTCTT
SMACT     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGGTTTTGCGTTGGTAAAAGAATACTT
SPILO     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGGTTTTGCGTTGGTAAAAGAATACTT
SKUNS     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGGTTTTGCGTTGGTAAAAGAATACTT
SSELA     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCGGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
SMAXI     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATATTT
SSCAB     AAAAGCATTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTTACTGCTTTTGCGTTGGTAAAAGAATACTT
SSING     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGGTTTTGCGTTGGTAAAAAAATACTT
SRICH     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATATTT
SMAXW     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
SISOP     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
SGUIS     AAAAGCATTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTTACTGCTTTTGCGTTGGTAAAAGAATACTT
SMACR     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGGTTTTGCGTTGGTAAAAGAATACTT
SSPLE     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTAGACTGGTTTTGCGTGGGTAAAGGAAAACTC
SLONG     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATATTT
SFOXW     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
SHOPE     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATATTT
SFAGU     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATATTT
SBECC     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGGTTTTGCGTTGGTAAAAGAATACTT
SMULT     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATATTT
SOVAL     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
SROXB     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
SMATE     AAAAGCATTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTTACTGCTTTTGCGTTGGTAAAAGAATACTT
SEXEL     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
SJAVA     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGGTTTTGCGTTGGTAAAAGAATACTT
SSMIT     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGGTTTTGCGTTGGTAAAAGAATACTT
SSEMI     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
SPARV     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
SPALE     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGGTTTTGCGTTGGTAAAAGAATACTT
SFOLI     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGGTTTTGCGTTGGTAAAAGAATCCTT
SSTEN     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGGTTTTGCGTTGGTAAAAGAATACTT
SLAEV     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
SPING     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGGTTTTGCGTTGGTAAAAGAATACTT
SCORD     AAAAGCGTGCAAAAGGAAATGGGATTGGTGCAGAGACTCAATGGAAGTTGGACGAACAAATGGGGCTGACTGCTTTGGCGTTGGTAAAGAAAT-CTT
SBALA     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
SJOHO     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGGTTTTGCGTTGGTAAAAGAATACTT
SLEPR     AAAAGCGTTCAAAAAGAAATAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGGTTTTGCGTTGGTAAAAGAATACTT
SBARC     AAAAGCGTTCAAAA-GAA-TAGGATAGGTGCAGAGACTCAATGGAAGTTGTTCGAACAAATGGGGTTGACTGCTTTTGCGTTGGTAAAAGAATACTT
AMARG     ----------------------------------------------------------------------------------------------------
PGLOB     ----------------------------------------------------------------------------------------------------
```

Appendix 4B (continued)

```
CLANC     ----------------------------------------------------------------------------------------------------
HSUBA     ----------------------------------------------------------------------------------------------------
SAMPL     ----------------------------------------------------------------------------------------------------

                                  [200]
NHEMI     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCCTTTGATTTATA
PLUCI     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
DAROM     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAATTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTCTA
DLANC     CTATCGAAACTTCAGCAATGATAAACATATAGAATACGTAATGAAAAATTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTCTA
DRETU     CTATCGAAATTTCAGCAAGGAGAAACATATAGAATACGTAATGAAAAACGATCTCAAAAAAGA-CGACTGGGACACCTATTCTTTTGATTTCTA
DKERI     CTATCGAAATTTCAGCAAGGAGAAACATATAGAATACGTAATGAAAAACGATCTCAAAAAAGA-CGACTGGGACACCTATTCTTTTGATTTCTA
DCONF     CTATCGAAATTTCAGCAAGGAGAAACATATAGAATACGTAATGAAAAACGATCTCAAAAAAGA-CGACTGGGACACCTATTCTTTTGATTTCTA
HPUBE     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTGATA
HMENG     CTATCGAAACTTCGGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTCTA
HCERN     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTCTA
HDRYO     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTATA
HFERR     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTATA
HPIER     TTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTATA
HNERV     CTATCGAAACTTCAGCAAGGATAAACATATGGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTATA
HNIGR     CTATCCAAACTTCAGCAAGGGTAACCCTATCCATTACGTAATGAAAACCTATCTTAAAAAAAA-CGACTGGGACACCTATTCTTTGGATTCATA
HCELT     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTATA
HCELE     CTATCGAAACTTCGGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTCTA
HAPIC     CTATCGAAACTTCGGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTATA
HWIGH     CTATCGAAACTTCGGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTATA
HBREV     CTATCGAAACTTCGGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTCTA
HJUCU     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTATA
HCORD     CTATCGAAACTTGAGCAAGGATTAACATATGGAATGCGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTATA
SMACT     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SPILO     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SKUNS     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SSELA     CTATCGAAACTTTAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTACA
SMAXI     CTATCGAAACTTCAGCAAGGAGAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SSCAB     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SSING     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTCCA
SRICH     CTATCGAAACTTCAGCAAGGAGAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SMAXW     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SISOP     CTAGCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTAT-TCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SGUIS     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SMACR     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SSPLE     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTGGATTTACA
SLONG     CTATCGAAACTTCAGCAAGGAGAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SFOXW     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SHOPE     CTATCGAAACTTCAGCAAGGAGAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
```

Appendix 4B (continued)

```
SFAGU     CTATCGAAACTTCAGCAAGGAGAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SBECC     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SMULT     CTATCGAAACTTCAGCAAGGAGAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SOVAL     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SROXB     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTCTA
SMATE     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SEXEL     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CAACTGGGACACCTATTCTTTTGATTTCTA
SJAVA     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SSMIT     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SSEMI     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SPARV     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SPALE     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SFOLI     TTATCGAAACTTCAGCAAGGATAAACATATAAAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SSTEN     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SLAEV     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SPING     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SCORD     CTCTCTAACCTTCAGCAAGGTTTAACATATAGAATACGGAATGAAAAACTATCTCAAAAAAAAGCGCCCGGGATACCTATTCTTATGA------
SBALA     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SJOHO     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAA--CGACTGGGACACCTATTCTTTTGATTTACA
SLEPR     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAA-CTATCTCAAAAAAA-CGACTGGGACACCTATTCTTTTGATTTACA
SBARC     CTATCGAAACTTCAGCAAGGATAAACATATAGAATACGTAATGAAAAACTATCTCAAAAAAAA-CAACTGGGACACCTATTCTTTTGATTTCTA
AMARG     ----------------------------------------------------------------------------------------------
PGLOB     ----------------------------------------------------------------------------------------------
CLANC     ----------------------------------------------------------------------------------------------
HSUBA     ----------------------------------------------------------------------------------------------
SAMPL     ----------------------------------------------------------------------------------------------

                                      [300]
NHEMI     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGGTGTGAATCGATT
PLUCI     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
DAROM     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
DLANC     TTTTTTT----------------------------ATAGGTTATTGCAAAAAGAATTATATGAAAAATTCAAAGAATCGTTGTGAATCGGTT
DRETU     TTTTTTT--------------ATAGGTTATAGCAAATAGGTTATAGCAAAAAGAATTATAGGAAAAATCAAAAGAATCGTTGTGAATCGATT
DKERI     TTTTTTTTATAGGTTATAGCAAATAGGTTATAGCAAATAGGTTATAGCAAAAAGAATTATAGGAAAAATCAAAAGAATCGTTGTGAATCGATT
DCONF     TTTTTTT--------------ATAGGTTATAGCAAATAGGTTATAGCAAAAAGAATTATAGGAAAAATCAAAAGAATCGTTGTGAATCGATT
HPUBE     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGGGGTGAATCGATT
HMENG     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGGTGTGAATCGATT
HCERN     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGGTGTGAATCGATT
HDRYO     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGGTGTGAATCGATT
HFERR     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGGTGTGAATCGATT
HPIER     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGGTGTGAATCGATT
HNERV     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGGTGTGAATCGATT
HNIGR     TTTTTTC----------------------------ATAGGTTATAGCAAAAAGAATTATTTGAAAAATCAAAAGAATCGGTGTGAATCGATT
```

Appendix 4B (continued)

```
HCELT     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGGTGTGAATCGATT
HCELE     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGGTGTGAATCGATT
HAPIC     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGGGGGGAATCGATT
HWIGH     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGGGG-GAATCGATT
HBREV     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGGTGTGAATCGATT
HJUCU     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGGTGTGAATCGATT
HCORD     TTTTGTT----------------------------ATAGGTTATAGCATAATGAATTATATGAAAAATCTAAAGAATCGGTGTGAATCGATT
SMACT     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SPILO     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SKUNS     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SSELA     TTTTTT-----------------------------ATAGGTTATAGCGAAAAGAATTATATGAAAAATTAAAA-AATCGGTGTGAATCGATT
SMAXI     TTTTTTT----------------------------AGAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SSCAB     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SSING     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SRICH     TTTTTTT----------------------------AGAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SMAXW     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SISOP     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SGUIS     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SMACR     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SSPLE     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SLONG     TTTTTTT----------------------------AGAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGGTGTGAATCGATT
SFOXW     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SHOPE     TTTTTTT----------------------------AGAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SFAGU     TTTTTTT----------------------------AGAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SBECC     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SMULT     TTTTTTT----------------------------AGAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SOVAL     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SROXB     TTTTTTT----------------------------ATAGGTTATATCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SMATE     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SEXEL     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SJAVA     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SSMIT     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SSEMI     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SPARV     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SPALE     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SFOLI     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SSTEN     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SLAEV     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAACCAAAAGAATCGTTGTGAATCGATT
SPING     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SCORD     -----------GTTATAAC---TTGCT--------ATAGGTTATAGCAAAAAGAATTATATGAAACATCAAAAGAATCGTTGTGAATCGATT
SBALA     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCG-TGTGAAACGATT
SJOHO     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
SLEPR     TTTTTTT----------------------------ATAGGTTATAGCAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
```

Appendix 4B (continued)

```
SBARC     TTTTTTT----------------------------ATAGGTTATA-CAAAAAGAATTATATGAAAAATCAAAAGAATCGTTGTGAATCGATT
AMARG     ----------------------------------------------------------------------------------------------
PGLOB     ----------------------------------------------------------------------------------------------
CLANC     ----------------------------------------------------------------------------------------------
HSUBA     ----------------------------------------------------------------------------------------------
SAMPL     ----------------------------------------------------------------------------------------------

                                                  [400]
NHEMI     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
PLUCI     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
DAROM     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATAAAACCATTCACTCCATAGTCTGATAAATCAATG-ATTAATTGGAC-GAGAATA
DLANC     TCAAATTGAAGAAAGAATCCAATAGAATATTCATTAATTAAACCATTCACTTCATAGTCTGATAAATCCATGGATTAATTGGAC-GAGAATA
DRETU     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACCCCATAGTCTGATAAATCAATGGATTAATTGGAC-AAGAATA
DKERI     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACCCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
DCONF     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACCCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
HPUBE     TCAAATTGACGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
HMENG     TCAAATTGACGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
HCERN     TCAAATTGACGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
HDRYO     TCAAATTGACGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
HFERR     TCAAATTGACGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
HPIER     TCAAATTGACGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
HNERV     TCAAATTGACGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
HNIGR     TCAAATTGACGAAAGAATCGAATCGAATATTCATTAATCAAACCATTCCCTCCATAGTCTGATAAATCAATGGATTAATTGG-C-GAGAATA
HCELT     TCAAATTGACGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
HCELE     TCAAATTGACGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
HAPIC     TCAAATTGACGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
HWIGH     TCAAATTGACGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
HBREV     TCAAATTGACGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
HJUCU     TCAAATTGACGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
HCORD     TCAAATTGACGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGGTTCATTGGAC-GAGAATA
SMACT     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SPILO     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SKUNS     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SSELA     TCAAATTGACGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SMAXI     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SSCAB     TCAAATAGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SSING     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SRICH     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SMAXW     TCAAATAGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SISOP     TCAAATAGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SGUIS     TCAAATAGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SMACR     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SSPLE     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
```

Appendix 4B (continued)

```
SLONG     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SFOXW     TCAAATAGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGACCGAGAATA
SHOPE     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SFAGU     TCACATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SBECC     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SMULT     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SOVAL     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SROXB     TCAAATTTAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SMATE     TCAAATAGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SEXEL     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SJAVA     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SSMIT     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SSEMI     TCAAATAGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SPARV     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SPALE     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SFOLI     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SSTEN     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SLAEV     TCAAATAGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SPING     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SCORD     TCAAATTGAAGCAAGAAGCGAATGCACTATCCTATAATCAAACCATACACTCCATAAAA-------------------------SAGAATA
SBALA     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SJOHO     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SLEPR     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
SBARC     TCAAATTGAAGAAAGAATCGAATAGAATATTCATTAATCAAACCATTCACTCCATAGTCTGATAAATCAATGGATTAATTGGAC-GAGAATA
AMARG     ----------------------------------------------------------------------------------------
PGLOB     ----------------------------------------------------------------------------------------
CLANC     ----------------------------------------------------------------------------------------
HSUBA     ----------------------------------------------------------------------------------------
SAMPL     ----------------------------------------------------------------------------------------

                                                                                          [500]
NHEMI     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTAATCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
PLUCI     AAGATAGAGTCCCATTCTACATGTCAATATMAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
DAROM     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTTC-TTAAGA-GGTCAAAATTTCT
DLANC     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTTC-TTAAGA-GGTCAAAATTTCT
DRETU     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCT-----------GTCAAAATTTCT
DKERI     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCT-----------GTCAAAATTTCT
DCONF     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCT-----------GTCAAAATTTCT
HPUBE     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCTTTTCTCT-GG-TAAGA-GGTCAAAATTTCT
HMENG     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCTTTTCTCTT-GTTAAGA-GGTCAAAATTTCT
HCERN     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCTCTT-GTTAAGA-GGTCAAAATTTCT
HDRYO     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCTCTT-GTTCAGA-GGTCAAAATTTCT
HFERR     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCTTTTCTCTT-GTTAAGA-GGTCAAAATTTCT
```

Appendix 4B (continued)

```
HPIER     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCTTTTCTCTT-GTTAAGA-GGTCAAAATTTCT
HNERV     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTAATCCTACCCTTTCTCTT-GTTAAGA-GGTCAAAATTTCT
HNIGR     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTAATCCTACCCTTTCTCTT-GTTCAGA-GGTCAAAATTTCT
HCELT     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTGCTATTGG-TACGA-GGTCAAAATTTCT
HCELE     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
HAPIC     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTAATCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
HWIGH     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTAATCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
HBREV     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
HJUCU     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
HCORD     AAGATCGAGTCCCATTCTACATGTCAATCTCACTACC-GACAACAATGAAATTTATAGTAATCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SMACT     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTAATCCTACCCGTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SPILO     AAGATAGAGGCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTAATCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SKUNS     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTAATCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SSELA     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTGTCTCTT-GTTCAGA-GGTCAAAATTCTT
SMAXI     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTAATCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SSCAB     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTATCCTTTC-CTT-GTTAAGA-GGTCAAAATTTCT
SSING     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-G-CAACAATGAAATTTATAGTAATCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SRICH     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTAATCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SMAXW     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTAATCCTATCCTTTC-CTT-GTTAAGA-GG-CAAAATTTCT
SISOP     AAGATAGAGTCCCATTCTACATGTCAATATCAATACCCGACAACAATGAAATTTATAGTA-TCCTATCCTTTCC-TTGGT-AAGAAGG-CAAAATTTCT
SGUIS     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTAATCCTATCCTTTC-CTT-GTTAAGA-GGTCAAAATTTCT
SMACR     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SSPLE     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTAATCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SLONG     AAGATAGAGTCCCCTTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SFOXW     AAGATAGAGTCCCATTCTACATGTCAATATCAATACCCGACAACAATGAAATTTATAGTAATCCTATCCTTTC-CTT-GTTAAGA-GGTCAAAATTTCT
SHOPE     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACACCAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SFAGU     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SBECC     AAGATAGAGTCCCATTCTACATGTCAATATCAATACCCGACAACAATGAAATTTATAGTAATCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SMULT     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACA-CAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SOVAL     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SROXB     AAGATAGAGTCCCATTCTACATGTCACTATCAATACC-GACAACAATGAAATTTATAGTAATCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SMATE     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTATCCTTTC-CTT-GTTAAGA-GGTCAAAATTTCT
SEXEL     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGG-GGTCAAAATTTCT
SJAVA     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTTGG-TAAGA-GGTCAAAATTTCT
SSMIT     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SSEMI     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTATCCTTTC-CTT-GTTAAGA-GGTCAAAATTTCT
SPARV     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SPALE     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTAATCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SFOLI     AAGATAGAGTCCCATTCTACATGTCAATTTCAATACC-GACAACAATGAAATTTATAGTA-TCCTATCCTTTMACTTGGT-AACA-GGTCAARAAACTC
SSTEN     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SLAEV     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTATCCTTTC-CT-CGTTAAGA-GGTCAAAATTTCT
SPING     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SCORD     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCT--CCCTT-CCCTTGGG—AGA-GGTCAAAATT-GT
```

Appendix 4B (continued)

```
SBALA     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTTCCTT-GTTAAGA-GGTCAAAATTTCT
SJOHO     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
SLEPR     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATCTCT
SBARC     AAGATAGAGTCCCATTCTACATGTCAATATCAATACC-GACAACAATGAAATTTATAGTA-TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
AMARG     ------------------------------------------------------------TCCTACCCTTTTCCTT-GTTAAGA-GGTCAAAATTTCT
PGLOB     ------------------------------------------------------------TCCTATCCTTTC-CTT-GTTAAGA-GGTCAAAATTTCT
CLANC     ------------------------------------------------------------TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT
HSUBA     ------------------------------------------------------------------------------------------------
SAMPL     ------------------------------------------------------------TCCTACCCTTTCCCTT-GTTAAGA-GGTCAAAATTTCT

                                                                  [600]
NHEMI     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
PLUCI     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
DAROM     TATGTTCATCCTATTCTTTTACATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTATCTTAT
DLANC     TATGTTCATCCTATTCTTTTACATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTATCTTAT
DRETU     TATGTTCATCCTATTCTTTTACATTTTATACATTTTACAAACGTATCC-GAGTAACAAAA---TTTCTCTCTTTTAT
DKERI     TATGTTCATCCTATTCTTTTACATTTTATACATTTTACAAACGTATCC-GAGTAACAAAA---CTTCTTTCTTTTAT
DCONF     TATGTTCATCCTATTCTTTTACATTTTATACATTTTACAAACGTATCC-GAGTAACAAAA---TTTCTCTCTTTTAT
HPUBE     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
HMENG     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
HCERN     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-CAG---CAAAAAATTTTCTCTCTCTTAT
HDRYO     TATGTTCATCCTATTCTTTTAGATTTTATCCATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
HFERR     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
HPIER     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
HNERV     TATGTTCATCCTATTCTTTTCGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
HNIGR     TATGTTCATCCTATTCTTTTAGATTTTATCCATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
HCELT     TATGTTCATCCTATTCTTTTAGATTTTATATATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
HCELE     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
HAPIC     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
HWIGH     TATGTTCATCCAATTCTTTTAGATTTTATACATTTTACAAACGGATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
HBREV     TATGTTCATCCTATTCTTTTCGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
HJUCU     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
HCORD     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SMACT     TATGTTCAT-CTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SPILO     TATGTTCATTCTATTCTTTTAGATTTTATACATTTTACA--GGTATCC-GAG---CCAAAAATTTTTTCTCCCTTAT
SKUNS     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CCAAAAATTTTCTCTCTCTTAT
SSELA     AATGTTATTTCTAT-CTTT-AGATTTTATCCATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SMAXI     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SSCAB     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTGTCTCTCTCTTAT
SSING     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SRICH     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SMAXW     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SISOP     TATGGTCATCCTATTCTTTTAAAATTTATACATTTTACAAACGTATCC-CA----C-AAAAATTTT-T-T-T-TTAT
```

Appendix 4B (continued)

```
SGUIS     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTGTCTCTCTCTTAT
SMACR     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SSPLE     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CCAAAAATTTTCTCTCTCTTAT
SLONG     TATGTTCATCCTATTCTTTTAGATTTTATTCATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SFOXW     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SHOPE     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCT-TTAT
SFAGU     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SBECC     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SMULT     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SOVAL     TATGTTCATCCTATTCTTTTAGATTTGATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SROXB     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAATGTATCC-GAG---CAAAAAATTTTTTCTCCCTTAT
SMATE     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTGTCTCTCTCTTAT
SEXEL     TATGTTCATCCTATTCTTTTCGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SJAVA     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SSMIT     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SSEMI     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SPARV     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SPALE     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SFOLI     TAGGGTCATCCGATCATTTTAGATTATATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SSTEN     TATGTTTTTTCTATTCTTTTAGATTTTATACATTTTACAAACGTATCCCGAG---CAAAAAATTTTCTCTCTCTTAT
SLAEV     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SPING     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SCORD     TATGTTCATCCTATTCTTTTCGATTCTATACATTTTACAAACGTATCC-GAG---CAAAAAATGTTCTCTCTCTTAT
SBALA     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SJOHO     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
SLEPR     TATGTTCATTCTATTCTTTTAGATTT-ATACATTTTTCAAACGTATTC-GAG---CAAAAAATTTTCTCTCTCTTAT
SBARC     TATGTTCATCCTATTCTTTTCGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT
AMARG     TATGTTCATCCTATTCTTTTCCATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTTTCTTAT
PGLOB     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTGTCTCTCTCTTAT
CLANC     TATGTTCATCCTATTCTTTTACATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTTTCTTAT
HSUBA     ----------------------------------------------------------------------------
SAMPL     TATGTTCATCCTATTCTTTTAGATTTTATACATTTTACAAACGTATCC-GAG---CAAAAAATTTTCTCTCTCTTAT

                                                                                  [700]
NHEMI     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
PLUCI     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
DAROM     CAGATATCATACACAAGTCTTGTGATATATTATGATATATATGTATGATATACGTACAAATGAGCATCGGAATATATACCCCT
DLANC     CAGATATCATACACAAGTCTTGTGATATATTATGATATATATGTATGATATACGTACAAATGAGCATCGGAATATATACCCCT
DRETU     CAGATATCATACGCAAGCCCTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
DKERI     CAGATATCATACGCAAGCCCTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
DCONF     CAGATATCATACGCAAGCCCTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
HPUBE     CAGATATCATACACAAGTCTTGT----------GATATA----------------------------------------
HMENG     CAGATATCATACACAAGTCTTGT----------------------GATATA-------------------------------
```

Appendix 4B (continued)

```
HCERN     CCGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
HDRYO     CCGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
HFERR     CAGATATCATACACAAGTCTTGT----------------------GATATA------------------------------
HPIER     CAGATATCATACACAAGTCTTGT----------------------GATATA------------------------------
HNERV     CCGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
HNIGR     CCGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
HCELT     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
HCELE     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
HAPIC     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATCCCCCT
HWIGH     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
HBREV     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
HJUCU     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
HCORD     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SMACT     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SPILO     CAGATATCATACATAAGTCTTGT----------GATATATAT----GATATACGTACAAATAAGCATCGGAATACATACCCCT
SKUNS     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SSELA     CCGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SMAXI     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATATCCCT
SSCAB     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SSING     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SRICH     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATATCCCT
SMAXW     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SISOP     -A-A-ATCATAC-C-AG-CTTGG----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SGUIS     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SMACR     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SSPLE     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SLONG     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATATCCCT
SFOXW     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SHOPE     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGCACAAATGAGCATCGGAATACATACCCCT
SFAGU     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATATCCCT
SBECC     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SMULT     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATATCCCT
SOVAL     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SROXB     CAGATATCATACATAAGTCTTGT----------GATATATAT----GATATACGTACAAATAAGCATCGGAATACATACCCCT
SMATE     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SEXEL     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATAAGCATCGGAATGCATACCCCT
SJAVA     CAGATTCATACCCCAAGTCTTGG----------GATATATAT----GATATCCGTCCAAATGAGCATCGGAATCCAT-CCCC-
SSMIT     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SSEMI     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SPARV     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SPALE     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SFOLI     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SSTEN     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
```

Appendix 4B (continued)

```
SLAEV     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SPING     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SCORD     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SBALA     TAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SJOHO     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
SLEPR     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATCCATACCCCT
SBARC     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATAAGCATCGGAATGCATACCCCT
AMARG     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATCCATATCCCT
PGLOB     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT
CLANC     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATATCCCT
HSUBA     --------------------------------------------------------------------------------
SAMPL     CAGATATCATACACAAGTCTTGT----------GATATATAT----GATATACGTACAAATGAGCATCGGAATACATACCCCT

                                [900]
NHEMI     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTCTTTTCTTTGAAGATCCAAGAAAATCCCCTGC
PLUCI     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
DAROM     T-TGAA-GATTCACAATCCATATCATTACTCATACTGAAACTAACAACGTATTCTCTTTGAAGATCCAAGAAAATCCCCTGC
DLANC     T-TGAA-GATTCACAATCCATATCATTACTCATACTGAAACTAACAACGTATTCTCTTTGAAGATCCAAGAAAATCCCCTGC
DRETU     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATCCCCTAC
DKERI     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATCCCCTAC
DCONF     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATCCCCTAC
HPUBE     ----------------------------------ACTGAAACTGACAAAGTATTTTCTTTGAAGATCCAAGAAAATCCCCTGC
HMENG     ----------------------------------ACTGAAACTGACAAAGTCTTTTCTTTGAAGATCCAAGAAAATCCCCTGC
HCERN     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTCTTTTCTTTGAAGATCCAAGAAAATCCCCTGC
HDRYO     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTCTTTTCTTTGAAGATCCAAGAAAATCCCCTGC
HFERR     ----------------------------------ACTGAAACTGACAAAGGCTTTTCTTTGAAGATCCAAGAAAATCCCCTGC
HPIER     ----------------------------------ACTGAAACTGACAAAGTCTTTTCTTTGAAGATCCAAGAAAATCCCCTGC
HNERV     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTCTTTTCTTTGAAGATCCAAGAAAATCCCCTGC
HNIGR     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTCTTTTCTTTGAAGATCCAAGAAAATCCCCTGC
HCELT     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTCTTTTCTTTGAAGATCCAAGAAAATCCC----
HCELE     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTCTTTTCTTTGAAGATCCAAGAAAATCCCCTGC
HAPIC     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTCTTTTCTTTGAAGATCCAAGAAAATCCCCTGC
HWIGH     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTCTTTTCTTTGAAGATCCAAGAAAATCCCCTGC
HBREV     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTCTTTTCTTTGAAGATCCAAGAAAATCCCCTGC
HJUCU     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTCTTTTCTTTGAAGATCCAAGAAAATCCCCTGC
HCORD     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTCTTTTCTTTGAAGATCCAAGAAAATCCCCTGC
SMACT     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SPILO     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SKUNS     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SSELA     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTCTTTTCTTTGAAGATCCAAGAAAATCCCCTGC
SMAXI     T-TGAATGATTCACAATCCAGATCATTACTCATACTGAAGCTGAGAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SSCAB     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATC-AAGAAAATTCCCTGC
SSING     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATT-TCTTTGAAGATCCAAGAAAATTCCCTGC
```

Appendix 4B (continued)

```
SRICH     T-TGAATGATTCACAATCCAGATCATTACTCATACTGAAACTGAGAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SMAXW     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SISOP     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SGUIS     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SMACR     T-TGAATTATTCACAATCCATATCATTACTCATACTGAAACTGACAACGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SSPLE     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SLONG     T-TGAATGATTCACAATCCAGATCATTACTCATACTGAAACTGAGAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SFOXW     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SHOPE     T-TGCATTATTCACAATCCATATTATTACTCATACTGAAACTGACAACGAATTGTCTTTGAAGAACCACGAAAATTCCCTGC
SFAGU     T-TGAATGATTCACAATCCAGATCATTACTCATACTGAAACTGAGAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SBECC     T-TGAATTATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SMULT     T-TGAATGATTCACAATCCAGATCATTACTCATACTGAAACTGAGAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SOVAL     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SROXB     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATCCCCTGC
SMATE     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SEXEL     TATGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTATCTTTGAAGATCCAAGAAAATCCCCTGC
SJAVA     ----------------------------------------------------------------------------------
SSMIT     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SSEMI     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SPARV     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SPALE     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SFOLI     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SSTEN     T-TGAATTATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SLAEV     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SPING     T-TGAATTATTCACAATCCATATCATTACTCATACTGAAACTGACAACGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SCORD     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATCCCCTGC
SBALA     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SJOHO     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SLEPR     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
SBARC     TATGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTATCTTTGAAGATCCAAGAAAATCCCCTGC
AMARG     T-TGAATGATTCATAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATCCCCTGC
PGLOB     T-TGAATGATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC
CLANC     T-TGAATGATTCATAATCCATATCATTACTCATACTGAAACTGACAA-----TCTCTTTGAAGATCCAAGAAAATCCCCTGC
HSUBA     ----------------------------------------------------------------------------------
SAMPL     T-TGAATTATTCACAATCCATATCATTACTCATACTGAAACTGACAAAGTATTCTCTTTGAAGATCCAAGAAAATTCCCTGC

               [800]
NHEMI     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTTTAATTG-ACATAGACCCAA-TC
PLUCI     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAGACCCAA-TC
DAROM     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAAACCCAA-TC
DLANC     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAAACCCAA-TC
DRETU     GAGACTCTTCTTT-TAAGACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAGACCAAA-TC
DKERI     GAGACTCTTCTTT-TAAGACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCAAA-TC
```

Appendix 4B (continued)

```
DCONF     GAGACTCTTCTTT-TAAGACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAGACCAAA-TC
HPUBE     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAGACCCA-GTC
HMENG     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAGACCCA-GTC
HCERN     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCA-GTC
HDRYO     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCA-GTC
HFERR     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAGACCCA-GTC
HPIER     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAGACCCA-GTC
HNERV     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTTTAATTG-ACATA-ACCCA-GTC
HNIGR     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATA-ACCC--GTC
HCELT     -----TCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCA-GTC
HCELE     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
HAPIC     GAGACTCTTCTTT-TAATACCTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
HWIGH     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
HBREV     GAGACTCGTCGTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
HJUCU     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
HCORD     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SMACT     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SPILO     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SKUNS     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SSELA     GAGACTCTTCTTT-TAATACTTTTTT-G-CTCTTTTTTTTTT-AATTG-ACATAGACCCA-GTC
SMAXI     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAGACCCAA-TC
SSCAB     GAGACTCTTCTTT-TAATACTCTTTTGG-CTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SSING     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAGACCCAA-TC
SRICH     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAGACCCAA-TC
SMAXW     GAGACTCTTCTTT-TAATACTCTTTT-GTCTCTTTTTTTTTTTAATTG-ACATAGACCCAA-TC
SISOP     GAGACTCTTCTTT-TAATACTCTTTTGG-CTCTTTTTTTTTT-AATTG-ACATAAACCCAA-TC
SGUIS     GAGACTCTTCTTT-TAATACTCTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SMACR     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SSPLE     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SLONG     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAGACCCAA-TC
SFOXW     GAGACTCTTCTTT-TAATACTCTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SHOPE     GAGACTCTTCTTT-TAATACTTYWGTG-CAGCTTTTTTTTTTAAA--GGACAGAAACCCAA-TC
SFAGU     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAGACCCAA-TC
SBECC     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SMULT     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAGACCCAA-TC
SOVAL     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SROXB     GAGACTCTTCTTT-TAATATTTTTTT-GC-TCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SMATE     GAGACTCTTCTTT-TAATACTCTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SEXEL     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAGACCCAA-TC
SJAVA     ------------------------------------------------------------
SSMIT     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SSEMI     GAGACTCTTCTTT-TAATACTCTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGGCCCAA-TC
```

Appendix 4B (continued)

```
SPARV     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SPALE     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SFOLI     GAGACTCTTCTTT-TAATACTCTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGGCCCAA-TC
SSTEN     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SLAEV     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SPING     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SCORD     GAGACTCTTTTTTATAA--CTTTTTG-GCCT-TTTTTTTTT-AAAT-GGACATAA-CCCA--TC
SBALA     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTT--AATTG-ACATAGACCCAA-TC
SJOHO     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SLEPR     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
SBARC     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
AMARG     GAGACTCTTCTTT-TAATACTTTTTTTGTATCTTTTTTTTTTTTAATTG-ACATAGACCAAA-TC
PGLOB     GAGACTCTTCTTT-TAATACTCTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC
CLANC     GAGACTCTTCTTT-TAATACTTTTTT-GTATCTTTTTTTTT--AATTG-ACATAGACCAAA-TC
HSUBA     ------------------------------------------------------------
SAMPL     GAGACTCTTCTTT-TAATACTTTTTT-GTCTCTTTTTTTTTT-AATTG-ACATAGACCCAA-TC

                                  <------trnL/F region    |  ITS--->
NHEMI     CTCTAGTAAAATGAGGATGATGGA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
PLUCI     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGX----------------------------------------------
DAROM     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGX----------------------------------------------
DLANC     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
DRETU     CTCTAGTAAAATGAGGATGATACA-TC-GTTAATGGTCAGGATAGCTCAGX----------------------------------------------
DKERI     CTCTAGTAAAATGAGGATGATACA-TC-GTTAATGGTCAGGATAGCTCAGX----------------------------------------------
DCONF     CTCTAGTAAAATGAGGATGATACA-TC-GTTAATGGTCAGGATAGCTCAGX----------------------------------------------
HPUBE     CTCTAGTAAAATCAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
HMENG     CTCTAGTAAAATCAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
HCERN     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
HDRYO     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
HFERR     CTCTAGTAAAATCAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
HPIER     CTCTAGTAAAATCAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCG-CGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
HNERV     CT-TAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCTGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
HNIGR     CT-TAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
HCELT     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
HCELE     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
HAPIC     CT-TAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
HWIGH     CT-TAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CG-CGCGAGAAGTCCACTGAACCTTATCAT
HBREV     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
HJUCU     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
HCORD     CT-TAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SMACT     CTCTAGTAAAATGAGGATGATC-AATAAGTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SPILO     CTCTAGTAAAATGAGGATGATGCA-TC-CTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SKUNS     CTCTAGTAAAATGAGGATGATGC--TC-GTTAATGGTCAGGATAGCTCAGX----CGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
```

Appendix 4B (continued)

```
SSELA     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SMAXI     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SSCAB     CTTTAGTAAAATGGGGAGGATCCA-TC-GTTAATGGGCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCACGAGAAGTCCACTGAACCTTATCAT
SSING     CTTT-GTAAAA-G-GG-TGATGC--T--G-TAATGGGCAGGATAGCT-AAXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SRICH     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SMAXW     CT-TAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGX-CGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SISOP     CTTTAGTAAAATGAGGATGATGCA-TC-GTTAATGGGCAGGATAGCTCAAXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SGUIS     CTCTAGTAAAATGAGGATGATCCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGTCCGCGA-CGTCACGAGAAGTCCACTGAACCTTATCAT
SMACR     CT-TAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SSPLE     CT-TAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCACGAGAAGTCCACTGAACCTTATCAT
SLONG     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SFOXW     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCC-CTGAACCTTATCAT
SHOPE     CTTTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCC-AGAACGTCACGA----------GACGCTTATCAT
SFAGU     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTTGACGTCCGCGA-CGTCGCGAGAAGTCCCCTGAACCTTATCAT
SBECC     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SMULT     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SOVAL     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SROXB     CTCTAGTAAAATGAGGATGATGCA-TC-CTTAATGGTCAGGATAGCTCAAXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SMATE     CTCTAGTAAAATGAGGATGATCCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCACGAGAAGTCCACTGAACCTTATCAT
SEXEL     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCCCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SJAVA     ----------------------------------------------XTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SSMIT     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGG-CAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SSEMI     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGGGCGGCCCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SPARV     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SPALE     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGX------------------------------------------
SFOLI     CTTTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTAAGX------------------------------------------
SSTEN     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGX------------------------------------------
SLAEV     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCACCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SPING     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SCORD     -TTTA-TAAAA-G-GGA-GA-GC--TC-GTTAA-GGTCAGGATAG-T-ACXTCGAC-TCCGC-A-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SBALA     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTTCGCCCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SJOHO     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SLEPR     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SBARC     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGX------------------------------------------
AMARG     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
PGLOB     CTCTAGTAAAATGAGGATGATCCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCACGAGAAGTCCACTGAACCTTATCAT
CLANC     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGX----------GA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
HSUBA     ----------------------------------------------XTCGCCGCCTGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT
SAMPL     CTCTAGTAAAATGAGGATGATGCA-TC-GTTAATGGTCAGGATAGCTCAGXTCGCCGCCCGCGA-CGTCGCGAGAAGTCCACTGAACCTTATCAT

                                                        [1000]
NHEMI     TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC---AAAGCAGAACG-ACCCGCGA
PLUCI     ----------------------------------------------------------------------------------------------------
```

Appendix 4B (continued)

```
DAROM  ----------------------------------------------------------------------------------------------------
DLANC  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
DRETU  ----------------------------------------------------------------------------------------------------
DKERI  ----------------------------------------------------------------------------------------------------
DCONF  ----------------------------------------------------------------------------------------------------
HPUBE  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC---AAAGCAGAACG-ACCCGCGA
HMENG  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC---AAAGCAGAACG-ACCCGCGA
HCERN  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC---AAAGCGGAAGG-ACCCGCGA
HDRYO  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC---AAAGCAGAACG-ACCCGCGA
HFERR  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC---AAAGCAGAACG-ACCCGCGA
HPIER  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC---AAAGCAGAACG-ACCCGCGA
HNERV  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC---AAAGCAGAACG-ACCCGTGA
HNIGR  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC---AGAGCAGAATG-ACCCGCGA
HCELT  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC---AAAGCGGAACG-ACCTGCGA
HCELE  TTA-GAGGAAGGAGAAGTCGTAWCACGGTTTCCGTAGGTGAACCTACGGAAGGATCACATTGTGCATGCCTGCCC---AAGGCAGA-CGCACCCGCGA
HAPIC  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCTAAT---GCAGAAC-CACCCGCGA
HWIGH  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCTACTAAAGCAGAAC-CACCCGCGA
HBREV  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCTA-T--AGCAGAAC-CACCCGCGA
HJUCU  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC---AAAGCAGAACG-ACCCGCGA
HCORD  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC---AAAGCAGAACG-ACCCGCGA
SMACT  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SPILO  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SKUNS  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SSELA  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC---AGAGCAGAATG-ACCCGCGA
SMAXI  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SSCAB  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCT-----AGCAGAACG-ACCCGCGA
SSING  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SRICH  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SMAXW  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SISOP  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCT-----AGCAGAACG-ACCCGCGA
SGUIS  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCT-----AGCAGAACG-ACCCGCGA
SMACR  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SSPLE  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SLONG  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SFOXW  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SHOPE  TTACGAGGAAGGAGAAGT-GTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SFAGU  TTA-GAAGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGAAAGGATCA--TTGTCGATGCCTGCCC-----AGTAGAATG-ACCCGCGA
SBECC  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SMULT  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SOVAL  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SROXB  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCT---AAAGCAGAACG-ACCCGCGA
SMATE  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCT-----AGCAGAACG-ACCCGCGA
SEXEL  TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCT-----AGCAGAATG-ACCCGCGA
```

Appendix 4B (continued)

```
SJAVA     TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SSMIT     TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SSEMI     TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SPARV     TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SPALE     ----------------------------------------------------------------------------------------------------
SFOLI     ----------------------------------------------------------------------------------------------------
SSTEN     ----------------------------------------------------------------------------------------------------
SLAEV     TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SPING     TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SCORD     TTA-GAGGAAGGTTTTGTCGTTACAAGGTTTCCGTAGGTGAACC-GCGGAAGGATCA--TTGTCGATGCCTGCCCCG---AGCAGAACG-ACCCGCGA
SBALA     TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AACAGAACG-ACCCGCGA
SJOHO     TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SLEPR     TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
SBARC     ----------------------------------------------------------------------------------------------------
AMARG     TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
PGLOB     TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCT-----AGCAGAACG-ACCCGCGA
CLANC     TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA
HSUBA     TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC---AAAGCAGAACG-ACCCGTGA
SAMPL     TTA-GAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA--TTGTCGATGCCTGCCC-----AGCAGAACG-ACCCGCGA

                                                                [1100]
NHEMI     ACGCGTTTTCAAA--AA----C-GGGGC------GGGAGGGGG-AGGG---CGCCGGGCTCGCGAGA-------GCAC-GTCGTCCCAT---CCCCTGCC
PLUCI     ----------------------------------------------------------------------------------------------------
DAROM     ----------------------------------------------------------------------------------------------------
DLANC     ACCCGTTTTAAA---AA--GGC-GGGGCGCGGGGGGGTGGGGGAAA-----C-TC---C-----A-A--------CCC-GGGGTCCCTTCCATCCCC-TC
DRETU     ----------------------------------------------------------------------------------------------------
DKERI     ----------------------------------------------------------------------------------------------------
DCONF     ----------------------------------------------------------------------------------------------------
HPUBE     ATGTGTTTTCA---GAAA---C-GAAGC------GGGAGGGG-AAGGG---CGTCGGGC----GAGA-------GCCC-GGGGTCCCCC--TCCCCTGCC
HMENG     ACGTGTTTTCCAAAA------C-GGAGC-----GGGAGGGGG-AAGGG---CGTCGGGC----GAGA-------GCCC-GGGGTCCC---GACCCTTGCC
HCERN     ACGTGTTTTCA---GAAA----TGGAAC------GGGAGAGG-AAGGG---CAGCGGGC----GAGA-------GCCC-GGGGTCCCC---TCCCCTGCC
HDRYO     ACGCGTTTTCAAAAA------C-GGAGC-----GGGAGGGGG-AAGGG---CGTCGGGC----GAGA-------GCCT-GGGGTCCC---GTCCCTTGCC
HFERR     ACGCGTTTTCAAAAA------C-GGAGC-----GGGAGGGGG-AAGGG---CGTCGGGC----GAGA-------GCCT-GGGGTCCC---GTCCCTTGCC
HPIER     ATGTGTTTTCA---GAAA---C-GAAGC------GGGAGGGG-AAGGG---CGTCGGGC----GAGA-------GCCC-GGGGTCCCC---TCCCCTGCC
HNERV     ACGTGTTTTCAAAAA-------TGGAGC------GGGAGAGG-AAGGG---CGTCGGGC----GAGA-------GCCC-AGGGTCCCCT---CCCCTGCC
HNIGR     ACGTGTTTTCA---GAAA---C-GGAAC------GGGAGAGG-AAGGG---CATCGGGC----GAGA-------GCCC-GGGGTCCCC---TCCCCTGCC
HCELT     ACGCGTTTTTA---GAAA---C-GGAGC------GGGAGGGG-AAGGG---CGTCGGGC----GAGA-------GCTC-GGGGTCCCGT---CCCCTGCC
HCELE     ACACCTCAGTA---GAAA---C-AGAG-----TGGGGGGGGG-ATGGG---TGCTGGG-T---GAAA-------GCTC-GGGGTCC--T-GTCCCTTGCC
HAPIC     ACACGTTT-CA----CAAA--CCGGAGCGC------GAGGGG-AAGGG---CGCCGGGC----GAGA-------GCCC-GGGGTCCC---GTCCCCTGCC
HWIGH     ACGCGTTTACACA--CAAA--C-AGAGC------GGGAGGGG-AAGGG---TGCCGGG-T---GACA-------GCCC-GGGGTCCC---GTCCCCTGCC
HBREV     ACGCGTTT-CA----TAAA--C-AGAGC------GGGAGGGGAAAGGG---TGCCGGGCA----ACAAA-----GCCC-GGGGTCCC---GTCCCCTGCC
HJUCU     ACGCGTTTTCA---GAAA---C-AGAGC------GGGAGGGG-AAGGG---CGCCGGG-T---GAGA-------GCCT-GGGGTCC--TG-CCCCCTGCC
HCORD     ACGCGTTTTCA----TAAA--C-AGAGC-----GGGAGGGG--AAGGG---CGCCGGG-T---GAGA-------GCCT-GGGGTCC--T-GCCCCCTGCC
```

Appendix 4B (continued)

```
SMACT     ACCCGTTTTAAA--GAA--GGC-GGGGCGC---GGGGGGGGG-AAC-------TCC-CA----------------CCC-CCCGTCC-CTCC-TTCC-T-T
SPILO     ACCCGTTTCAAA--GAA--GGC-GGGGCGC----GGGGGGGG-AAGGG---CAT-TGC--------AA-------CCC-GG-GTCCCCTCCCTTCCCT-A
SKUNS     ACCCGTTT-CGAA-GAAA--GC-GGGGCGCCGACGAGGGGGG-AAGGG---CATCGGGC----GAGA-------GCCC-GGGGTCCCCT---CCC-TCCC
SSELA     ACGTGTTTTCA---GAAA---C-GGAAC------GGGAGAGG-AAGGG---CATCGGGC----GAGA-------GCCC-GGGGTCCCC----TCCCC-TG
SMAXI     ACTCGTTTTTAAA-GAAAA--CTGGGGCGC----GGGTGGG--AACGGTGTCGTCGGGC----GAGAAA-----GCCC-GGGGTCCCC---TCCCCTCCT
SSCAB     ACTCGTTTAAAAAAAA-----CTGGGG-TGTGGGGGGGGGG---AGG----C-CC---C----GAGG-------GCCC-GGGGTCCCC---TCCCTTCCC
SSING     ACCCGTTT-CAAA-GAA--GGC-AGGGCGC---GGGGGGGGGGGGGGGG----GTC---C-----AGA-------GCCC-AGGGT-CCCT--TCC-TTCCC
SRICH     ACTCGTTTCTTAA-GAAAA--CTGGG-C-----GGGGGTGGG-AAGGG---CGTCGGGC----CAGAAA-----GCCC-GGGGTCCCC---TCCCCTCCC
SMAXW     ACTCGTTTTAAA--GAAA---CTGGGG-----TGGGGGGGG--AAGGG---CGTCGGGC----GAGG-------GCCC-GGGGTCCCC---TCCCTTCCT
SISOP     ACTCGTTTTTAAA-GAAA---TTGGGG-----TGGGGGGGG-AAAGGG---CGTCGGGC----GAAG-------GCCC-GGGGTCCAT---TCCCCTCCC
SGUIS     ACTCGTTTAAAAA--AAA---CTGGGG--TGTGGGGGGGGGG--AGGG---CCCCGGGC----GAGG-------GCCC-GGGGTCCCC---TCCCTTCCC
SMACR     ACCCGTTT-CAAA-GAA--GGC-GGGGCGC---GGGGGGGGG-AAGGG---CGTCCCTC----GAAAAA------CCC--G--T-CCCT--TCC-TTCCC
SSPLE     ACCCGTTT-CAAA-GAA--GGC---GGCGC----GGGGGGGG-AAGGG---CGTCCGGC----GAGA-------GCCC-GGGGTCCCCT----YCYTYCC
SLONG     ACTCGTTTTTTAA-GAAAA-CC-GGGGC-----GGGGGTGGG-AACGGCGTCGTCGGGC----GAGAAATTCTCGCCC-GGGGTCCCC---TCCCCTCCT
SFOXW     ACTCGTTTTAAA--GAAA---CTGGGG-----TGGG-GGGGG-AAGGG---CGCCCG-C----CAGG-------GCCC-GGGGTCCCC---TCCCTT-CC
SHOPE     ACTCGTTTCTTAA-GCAAA--CTGGG-C-----AGGGGTGGG-AAGGG---CGTCGGGC---CGGAAA------GCCC-GGGGTCCCC---TCCCCTCCC
SFAGU     ACTCGTTTCTTAA-GAAAA--CTGGG-C-----GGGGGTGGG-AAGGG---CGTCGAGC----CAAAAA-----GCCT-AGGGTCCCC---TCCCCTCCT
SBECC     ACCCGTT-CCAAA--CAA-GGC-GGGGCGCGGGGGGGGGGGG--AGG----CC-CC--------AGA-------GCCCC-GGGTCCCCT----CCCTCCC
SMULT     ACTCGTTTCCTAA-GAAAA---TGGG-C-----GGGGGTGGG-AAGGG---CGTCGGGC----GAGAAA-----GCCC-GGGGTCCCC---TCCCCTCCT
SOVAL     ACCCGTTTT-AAA-GAA--GGC-GGGGCGC--GGGGGGGGGG-AAGGG----T-----C-----A-AAA------CCCC-GGGTCCCCT----CCCTCCC
SROXB     ACGCGTTTTCA----CAAA--C-GGAGC-----GGGAGGGGG-AAGGGG--CGCCCGGC----GAGA-------GCCC-GGCGTCC--TGGGCCCC-CCC
SMATE     ACTCGTTT-AA--AAAAAA--CTGGGG---TGTGGGGGGGGGG-AGGG---CTCC--GC----GAGG-------GCCCC-GGGTCCCCT---CCCTTTCC
SEXEL     ACGCGTTCTCA---GAAA---C-GGGGC-----GGGAGGGG--AAGGG---CGCCGGGT----GAGAA-------CCC-GGGGTCC---GGGCCCCTCCC
SJAVA     ACCCGTTT-CGAA-GAAA--GC-GGGGCGCC-GAGGGGGGG--AAGGG---CGTCGGGC----GAGA-------GCCC-GGGGTCCCCT----CCCTCCC
SSMIT     ACCCGTTTAAA---GAA--GGC-GGGGCGC-GGGGGGGGGGG-AAGGG---CCT-C------CCA-AAA-----GCCC-GGGGTCCCCT---CCATTCCC
SSEMI     ACCCGTTTT-AAA-GAAA---CTGGGG------TGGGAGGGG-AAGGG---CGCCGGGC----GAG-------CGCCCGA-GGTCC--TC-TCCCTTCCC
SPARV     ACCCGTTT-CAAA-GAA--GGC-GGGGCGC----GGGGGGGG-AAGGG---CGTCGGGC----GAGA-------GCCC-GGGGTCCCCT--T--CATCCC
SPALE     ----------------------------------------------------------------------------------------------------
SFOLI     ----------------------------------------------------------------------------------------------------
SSTEN     ----------------------------------------------------------------------------------------------------
SLAEV     ACTCGTTTTAAA--GAAA--TC-------GGGGTGGGGGGGG-AA---CC---CT---CCCC-A-A--------CCT-TCC-T--CCC---C-CCT-CT
SPING     ACCCGTTT-CAAA-GAA--GGC-GGGGCGC----GGGGGGGG-AAGGGCGTCTCCC-------CAGA-------GCCCC-GGGTTCCCT----CCCTCTC
SCORD     ACGCGT---CAAAA-CA----C-GGAGC-----GGGAGGGGG--AGGGG--CGCCGGGC----GAGA-------GCCC-GGGGTCCAA---CCCCGTCCC
SBALA     ACCCGTTTTAATTAAAAAAGGC-GGAGCCC----GGGGGGGG-AAGGG---CGTCCGGC----GAAA-------GCCC-GGGGTTCCCT--TCC-TTCCC
SJOHO     ACCCGTTT-CGAA-GAAA--GC-GGGGCGCC-GAGG-GGGGG-AAGGG---CGTCGGGC----GAGA-------GCCC-GGGGTCCCCT---CCC-TCCC
SLEPR     ACCCGTTTTTAAAAGAA--GGC-GGGGCGC--GGGGGTGGGG-AAGGACTT---C--------TATA-------GCCCC-GGGTCCC-TG---CCTTCCC
SBARC     ----------------------------------------------------------------------------------------------------
AMARG     ACCCGT---GAA---CAA---C--GA-CGC-GGGGGGGCGGAGGAGGG---CGCCGGGC----GAGA-------GCCC-GGGG-CCCC--GTCCCCTCCC
PGLOB     ACTCGTTT-AAAAAAAA----CTGGGG---TGTGGGGGGGGG--AGGG---CACCC-GC----GAGG-------GCCCC-GGGTCCCCT---CCCTT-CC
CLANC     ACCCGTTAACAAA-GA-----------CGC-GGGGGGGCGGG--ACGG---CGTCGGGC----GAGA-------GCCC-GGGGTC---GCGCCCCCTCCC
HSUBA     ACGTGTTTTCAAAAA-------TGGAGC------GGGAGAGG-AAGGG---CGTCGGGC----GAGA-------GCCC-AGGGTCCCC---TCCCCTGCC
SAMPL     ACCCGTTT-CAAA--CAA-GGC-GGGGCGC---GGGGGGGGG-AAGGG---CATCCGCC----GAGA-------GCCC-GGGGTCCCCT----CCCTCCC
```

Appendix 4B (continued)

```
                                                  [1200]
NHEMI     GTGCCAC-------------GGGGGCGTGCATGAGCGCCC---------------TGCC-----GCCCGCGCA-CGCCCCCCAGCT-AACTAACGAACC-
PLUCI     ----------------------------------------------------------------------------------------------------
DAROM     ----------------------------------------------------------------------------------------------------
DLANC     GCGCACCC-----GCACCCCGGGGGTGCCCGCGCGC--GGGAATTCAAACCGCCCGCCTGC--------------GCCCCCG-------CTAAC-AACC-
DRETU     ----------------------------------------------------------------------------------------------------
DKERI     ----------------------------------------------------------------------------------------------------
DCONF     ----------------------------------------------------------------------------------------------------
HPUBE     GCGCCCC-------------GGGGGCACGTGCGGG--------------------TGCCCTGTTGCCCGCCC-GTGCCC--TTGCT------ACGAACC-
HMENG     GTGCCCT-------------GGGGGCATGTGCGGG--------------------TGCCTTGCTGCCCGCCC-GTGCCCCC--G-----CTAACGAACC-
HCERN     GTGCCTC-------------GGGGGCACGTGCGGG--------------------CGCCTCGTCGCTCGCCCCG-GCCCCC--G-----CTAACAAACC-
HDRYO     GTGCCCC-------------GGGGGCATGTGCGGG--------------------TGCCCTGCT-CCCGCC-TGTGCCCCC-A------CTAACGAACC-
HFERR     GTGCCCC-------------GGGGGCATGTGCGGG--------------------TGCCCTGCTGCCCGCC-TGTGCCCCC--G-----CTAACGAACC-
HPIER     GTGCCCC-------------GGGGGCACGTGCGGG--------------------TGCCCTGTCGCCCGCTC-GTGCCCTC--G-----CTAACGAACC-
HNERV     GTGCCCT-------------GGGGGCACGTGCGGGCGCCC---------------------TGCCGCTCGCCTGTGCCCCC-A------CTAACGAACC-
HNIGR     GTGCCTC-------------GGGGGCACGTGCGGGCGCCC----------------TGCC-----GCTCGCCCGTGCCCCC-A------CTAACGAACC-
HCELT     GTGCCAT-------------GGGGGCATGTGCGGGCGC-AC-------------------TGCTGCC-GCCC-GTGCCCC--TG-----CTAACGAACC-
HCELE     GTGTCTC-------------GGCAGTGCGTGT----GAGCGA--------------CGCCCTGCCCCTA-CTTGCGCCCCC--G-----CTAA-GAACC-
HAPIC     GTGCCCCC------------GGGGGCGCGCG----CGAGCGAG-------CGCCCTGCCGCCC-GCCCGCC-TGTGCCCCC--G-----CTAACGAACC-
HWIGH     CCGCCCT-------------GGGGGCGTGAGCAAGCGACC-----------GCCCTGCCACCC--GCCTG-CTGCGCCCCC--G-----CTAACGAACC-
HBREV     GTGCCCC-------------GGGGGYGCGCG----CGAGCGAG-------CCCCCTGCCCC--CSCCCS-S-TGTGSCCCCS------CTAACGAACC-
HJUCU     GTGCCTC-------------GGGGGCGTGCG--------CGA-------------TGCCCTGTCGCCCGCCC-GCGCCCCC-A--TTAAC----GAACC-
HCORD     GTGCCTC-------------GGGGGTGTGCG--------CGAG-------------TGCCCTGTCGCTCGCCCGCGCCCCC-A--TTAAC----GAACC-
SMACT     CGT--CCCCCCCC--CCCCCGGGGG---CC-CCCC----GCA-T-----AA--------------CCCCCCCCCCCCCCCC--------TAA-TAA-CCC
SPILO     CGC-CCCGAACACCGGGGG-GGGGCC--TTTTGGGCCAA-AAAT-T-TTT--TTC-C-GGTTTAT-A--CCCCCTC--TCC--------TAA--GGGTT-
SKUNS     CCGAGGCACCCG--CACACCGGGGGCGCGCGCGCGCGA--GAGTTTA--------------CCGCGCCCGCGCGCGCCC---A-CT--A-TAACGAACCT
SSELA     CCGTGCCTC-----------GGGGGG--------------------------------------------------------------------------
SMAXI     GCGCCCC-------------GGGGGCACGCGG-CGGGCGACAG-----------------------CCCGCCCGCGCCCCC--G-----CTAACGAACC-
SSCAB     GGGCCTC-------------GGGGGCGTGC-TGCGCGTGCGA-TA----------------CGTCGTCTGCCCACGCCCCC-T------CTAACGAACC-
SSING     TTGCCCTGCCGGCCCAC-GCGGGGGCGTGCGCGCTG----GAATTTAAACCCAC--------------TGCCCACGCCCCCC-G-TTAAC-----AACC-
SRICH     GCGCCCC-------------GGGGGCGCGCGCGGTCG-GCGAC---------------------GGCCCGCCCACGCCCCC--G-----CTAACGAACC-
SMAXW     GCGCCTC-------------GGGGGCGTGCCGCGGG----GA-TT----------------CGTCGTCCGCCCGCGCCCCC-T------CTAACGAACC-
SISOP     GCGCCTC-------------GGGGGCGTGCGCCTGC----GAAT--ATT----------------GTGCGCCCGCGCCCCC--GCTTA-----CGAACC-
SGUIS     GGGCCTC-------------GGGGGCGTGCCGGCCCGTGCGA-TA----------------CGTCGTCTGCCCACGCCCCC-T------CTAACGAACC-
SMACR     -TTC--TCCCGGCCCAC-CCGGGGGGCGC---CGCCG---GAATTTAAACCCACT--------------GCCCACGCCCCCC--CTTAACTAA----CC-
SSPLE     CTYACGCACCCG--CACACCGGGGGCGCGCGCGCACGG--GAATTTAAACCGCCC--------------GCCCACGCCCCC--G-TTAAC----GAACC-
SLONG     GCGCCTC-------------GGGGGCACGCG-GTGGGCGACA-----------------------GCCCGCCTGCGCGCCC--G-----CTAACGAACC-
SFOXW     GCGCCTT-------------GGGG-CGTGCCGCGTG---CGA-TT----------------CGC-GTC-GCCCACGCCCCC-T------CTAACGAACC-
SHOPE     GCGCCCC-------------GGGGGCGCGCGG-TCGGCGACAG-----------------------CCAGCCCGCGCCCCC--G-----CTAACGAACC-
SFAGU     GCGCCCC-------------GAGGGCACGCAGTCGG---CGA----------------------CGCCCACCCACGCCCC--AG-----CTAATGAACC-
SBECC     GTTACGCACCCG--CACCCCGGGGGTGCGCGCGCCCCC--GAGTTTAAATCGCCCGCCCGC--------GCCCGCGCCACC--G-TTAAC----GAACC-
SMULT     GCGCCCC-------------GGGGGCGCGCG-GTCGGCGACAG-----------------------CCCGCCCACGCCCCC--G-----CTAACGAACC-
SOVAL     CTTGGGCACCCG--CACACCGGGGGTGCGCGCGCGCGG--GAATTTAAACCGCCC--------------GCCCG-SCCCCC--GG-----TAACGAACC-
```

Appendix 4B (continued)

```
SROXB     GTGCC--A-------A----GGGGGCGCGTGCGCGC-------------TCGCGCGCCATGTGCCGCCCGCCTGCGCCCCC--G-----CTAACGAACC-
SMATE     GGGCCTCA------------GGGG-CGTGCTTGCCCGTTC-----TATTCCGCC-GGT----------TGCCACCGCCCCC----TTT--TA-CGA-CC-
SEXEL     GTGCC--AC-----------GGGGGCACGCGCGGACGC------------GCCCTGY----GCCGCCTGCCCGCGCCCCC--G-----CTAACGAACC-
SJAVA     CCGACGCACCCG--CACACCGGGGGTGCGCGCGCGCGCGCGAGTTT--------------CGACCGCCCGCCCGTGCCCCC--G-----CTAACGAACC-
SSMIT     TTGCCCCACCCG--CACCCCGGGGGTGCCCGC-----------------------------------------GCCCCC--G-T----TAACGAACC-
SSEMI     G-------------CACCTCGGGGGCGTGC-CGCGCG--CGA-TTMA---------------TCGTCCGCCCTCCCCCCC-T-CTTA-CT---GAACC-
SPARV     CTCGCGCACCCG--CACCCCGGGGGTGCCCGC-----------------------------------------GCCCCC-A------CTAACGAACC-
SPALE     ----------------------------------------------------------------------------------------------------
SFOLI     ----------------------------------------------------------------------------------------------------
SSTEN     ----------------------------------------------------------------------------------------------------
SLAEV     G--CCCCCC-----------GGGC-C-----------------------------------------------------------------------
SPING     CCCACGCACCCG--CACACCGGGGGGGCGCGCGCCCGG--GAATTTAAACCGCCC-------------GCCCCCGCCCCCCC-------TAACGAACC-
SCORD     GTGCCCC-------------GGGGGCGCGCGCGCGCCGC-------------------------------GT-CCCCTG-----CTAAC-AACC-
SBALA     TT-ACGCACCCG--GACACCGGGGGTGCGCGCGCGCGG--GAATTTAAACCCCCCGGCC----------GYCTG-GCCCCC-A------CTAACGAACC-
SJOHO     CCGACGCACCCGCGCACACCGGGGGTGCGCGCGCGCGCGCGAGTTT--------------CGACCGCCCGCCCGCGCCCCC--G-----CTAACGAACC-
SLEPR     TTG-CGCACCCG--CACCCCGGGGGGGCCG-C-----------------------------------------GCCCCC--GCTTA-----CGAACC-
SBARC     ----------------------------------------------------------------------------------------------------
AMARG     C-GCCGC-----CCGAC-CCGGG-------------------------------------------------CGCGC------------TAACGAACC-
PGLOB     GGGCCTCA------------GGGG-CGTGCTTGCCCGTCC-------AATCCGCCGGT---------TGCCCC-CGCCCCC----TT---TAACAAACC-
CLANC     C-GCCGC-------------------------------------GCCC---------CGCTCGGGC-GCGC-------------TAACGAACC-
HSUBA     GTGCCCT-------------GGGGGCACGTGCGGG---------------CGCCCTGCC-----GCTCGCC-TGTGCCCCC-A------CTAACGAACC-
SAMPL     CTTATGCCCCCG--GACACCGGGGGGCCCCCCCCCCGG--GAATTTAAACC-CCC--------------CCTGCGCCCCC--GGTTA-CTAAC---CC-

                                                                [1300]
NHEMI     CCGGCGTGGA--CCACGCCAAGGAAAACCAACCGGG-AGGGTCGCGCCCGCGCACGC--GGGC--CGCGCCTT-CCGC-AA-C----TA-C----AAAAA
PLUCI     ----------------------------------------------------------------------------------------------------
DAROM     ----------------------------------------------------------------------------------------------------
DLANC     CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCCGGC-CCCGCGCACGC--GGGCCAC-C-C-T--CCTCTA--C-C-TTA-C----AAAAA
DRETU     ----------------------------------------------------------------------------------------------------
DKERI     ----------------------------------------------------------------------------------------------------
DCONF     ----------------------------------------------------------------------------------------------------
HPUBE     CCGGCGTGGA--CCACGCCAAGGAAAACCAAT-GGGGAGGCTCGTGCCTGCCCACG--TGGGC--CGCGCCT-CCCGC-AA-C----TA-C----AAAAA
HMENG     CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCTCGCGCCTGCGCATG--TGGGCCG--CGCCT-CCCGC-AA-TGC-TTA-C----AAAAA
HCERN     CCGGCGTGGA--CCACGCCAAGGAAAACCAACAGGG-AGGCTTG----------------------TGCCT-CC-GC-AA-CAC-TTAAC---G-AAAA
HDRYO     CCGGCGTGGA--CCACGCCAAGGAAAACCAA-GGGGGAGGCTCGTGCCTGCGCATG--TGGGCCG--CGCCT-CCCGC-AA-CGC-TTA-C----AAAAA
HFERR     CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCTCGTGCCTGCGCATGY--GGGCCG--TGCCT-CCCGC-AA-CGC-TTA-C---G-AAAA
HPIER     CCGGCGTGGA--TCACGCCAAGGAAAACCAAT-GGGGAGGCTCGTGCCTGCGCACG--TGGGCCG--CGCCT-CCCGC-AA-CGC-TTA-C----AAAAA
HNERV     CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAAGGTTGGGCCCGCACACC--TGGGCCGGG--CCTT-CCGC-AA-CGC-TTA-C----AAAAA
HNIGR     CCGGCGTGGA--CCACGCCAAGGAATACCAAC-GGGGAGGCTCGTGCCTG--CACGCGTGGGC--CGCGCCT-CCCGC-AA-CGC-TTA-C---G-AAAA
HCELT     CCGGCGTGGA--CAACGCCAAGGAAAACCAAC-GGGGAGGCTCGCGCCCGCGCACG--TGGGC--TGCGCCT-CCCGC-AA-CGC--------------
HCELE     CCAGCGTGG---CCACGCCAAGGAAAACCAAC-TAGGAGGCTCGTGCCCC-GCCC-C--AAACCG--TGCCT--CCGC-AA-CGC-TTA-C----CAAAA
HAPIC     CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGATCGCGCCCGCGCCCGC--GGGCCGC--GCCT-CCCGC-AA-C----TA-C----AAAAA
HWIGH     CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCTCGTGCCCGCGCACGC--GGGCCG--TGCCT-CCTAC-AA-C----TA-C----AAACT
```

Appendix 4B (continued)

```
HBREV   CCGGCGTGGA--CCACCCCAAGGAAAACCA-C-GGGGAGGCTTGCGCCCGCGCACGC--GGGCCG--TGCCTTT-CGC-AA-CGC-TTA-C-----AAAA
HJUCU   CCGGCGTGGA--CCACGCCAAGGAAAACCAAT-GGGGAGGCTCGTGCCCGCGCACGC--GGGC--TGCGACT-CCCGC-AA-CGC-TTA-C---G-AAAA
HCORD   CCGGCGTGGA--CCACGCCAAGGAAAACCAAT-GGGGAGGCTCGTGCCCGCGCATG--TGGGC--TGCGACT-CCCGC-A-GTGC-TTA-C---G-AAAA
SMACT   CCCCGGGG---CCACCCCACGGAA-AC-A-CCGGGGGGG--GGCGTCCTCCTC-CCCC--GCCCACGC-CCT--CCTTTTT-CTTTTT---TA----AA-
SPILO   A---G-GGGG--TTTGT-AC-------------------------------------------GA-----------------------------------
SKUNS   CCGG--TGGACCCCACGCCAAGGAAAACCAAC-GGAGAGGCCC--GCCCGCGCACGCGCGGGCCGCGCGCCT-CCCGAGA--CGC-TTA-C----AAACA
SSELA   ------------------------------------------------------------------------------------------------
SMAXI   CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCTCGCGCCCGCGCACGC--GGGTCGTGCGCCT-CCCGCGA--CCC-TTA-C---G-AAAA
SSCAB   CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCTCGTGCCCGCGCACGC--GGGCCGCGTGCCTCTT-GC-AA-TGC-TTA-C----AAAAA
SSING   CCGGGGTGGA--CCACCCCCGGGAAAACCAACCGGG-AGGGCGGGCCCCGCCCATC-GCGGGCGCCGCGCCT-CCC--TACACGC-TTA-CT----AAAA
SRICH   CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCTCGCGCCCGCGCCCGC--GGGCCGTGCGCCT-CCCGCGA--C---TT--CT----AAAA
SMAXW   CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCTCGTGCCCGCGCACGC--GGGCTGCGTGCCTCT-CGC-AA-C---TT-------CAAAA
SISOP   CCGGGGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCTTGTGCTCGCGCACCC--GGGCTGCATGCCTTT-CG-GAA-CGC-TTA-----AAAAAA
SGUIS   CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCTCGTGCCCGCGCACGC--GGGCCGCGTGCCTCTT-GC-AA----TTT--C----AAAAA
SMACR   CCGG-GTGGGCCCCACCCC-GGGAAAACCAACCGGG-AGGGTCG-GCCCGCGCACGCC-GGGC-GGGCGCCTTCCCG-GAA--GC-TTACCT-----AAC
SSPLE   CCGGCGTGGA--CCACCCCAAGGAAAACCAAC-GGGGAGGCCGGCGCCCGCGCACGC--GGGCAGCGCACCT-CCCGCGA--CGC-TTA-C---G-AAAA
SLONG   CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCTCGCGCCCGCGCACGC--GGGCCGTGCGCCT-CCCGCGA--CCC-TTA-C---G-AAAA
SFOXW   CCGGCGTGGA--CCACGCCAAGGAAAACTAAC-GGGGAGGCTTGTGCCCCCCCACGC--GGGCCGCGTGCCTCTT-GC-AA-CGC-TTA-C----AAAAA
SHOPE   CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCTCGCGCCCGCGCCCGC--GGGCCGTGCGCCT-CCCGCGA--CTC-TTA-C---G-AAAA
SFAGU   CCGGCGTGGA--CCACTCCAAGGAAAACCAAC-GGGGAGGCTCGTGCCTGC--------GGGCCGTGCGCCT-CCCGCGA--C-TTTTA-C---G-AAAA
SBECC   CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGTCGGTGCCCGCGCACGC--GGGCGGCTCGCCT-CCCTCGA--CGC-TTA-C---G-AAAA
SMULT   CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCTCGCGCCCGCACCCGC--GGGCCGTGCGCCT-CCCGCGA--CTC-TTA-C---G-AAAA
SOVAL   CCGGGG-GGGCCCC-CCCCAGGGAAAAC-AAC-GGGGGGGGCGG-GCCC-CGCACCG--GGGCCG--GCCCTTTCTT-GAA--GG-TTTA----GAAAAA
SROXB   CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GAGGAGGCTCGTGCCCGCGCACGC--GGGCTTCGTGCCT-CCCGC-AA-CGC-TTA-CTA-G---AA
SMATE   CCG-CCTGGG--CCACCCCCAGGAAAACCCACCGGGGAGGCTCGTGCCCCCGCCCCC--GGGCCGTGCGCCTGT-CGC-A--TGG-TT--CC---AAAAC
SEXEL   CCGGCGTGGA--CCACGCCAAGGAAAACCGAC-GAGGAGGCTCGCGCCCGCGTGCG--TGGGCGTCGTGCCT-CCCGC-AA-CGC-TTA-CTA-G---AA
SJAVA   CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGACGGCGCCCGCGCACGC--GGGCCGCTGGCCT-CCCGCGA--CGC-TTA-CT----AAAA
SSMIT   CGGGGG-GGG-CCCACGCCAAGGAAAACCAAC-GGGGAGGCCGGC-CCCGGCAACGC--GGGCCACCC-CCT-CC-GC-AAACGC-TTA-C-----AAAA
SSEMI   CCGCG-TGGA--CCACCCCAAGGAAAACAAAC-GGGGAGGCTGGTGCCCCCCCACCC--CGCCCGCGTGCCT-CCCGC-AA-CGC-TTA-CT----AAAA
SPARV   CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCCGGCGCCCGCGCACGC--GGGCCGCACGCCTT-CCTTTA--CGC-TTA-C---G-AAAA
SPALE   ------------------------------------------------------------------------------------------------
SFOLI   ------------------------------------------------------------------------------------------------
SSTEN   ------------------------------------------------------------------------------------------------
SLAEV   ------------------------------------------------------------------------------------------------
SPING   CCCGCGTGGG--CCACCCCCAGGAAAACCAACAGGGG-GGGCGGGGCCCCCCCCACGC--GGGCAGAGCGCCT-CCCTCGA--CGC-T---C---G-AAAA
SCORD   CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCTCCTGCCTGCGCACG--TGGGCCGC--TCCT-CCCGC-AA-CGC-TTA-C----AAAAA
SBALA   CCGGGGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGACGGGGCCCGCGCACGC--GGGCCGGGCGCCTT-CCGTGA--CGC-TT--C---G-AAAA
SJOHO   CCGGCGTGGA--CCACGCCAAGGAAAACGAAC-GGGGAGGACGGCGCCCGCGCACGC--GGGCCGCTGGCCT-CCCGCGA--CGC-TTA-CT----AAAA
SLEPR   CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGGCGGAGCCCCCGCACCC--GGGCCGCGTGCCT--CCG-TAT-CCC-TTA-C---G-AAAA
SBARC   ------------------------------------------------------------------------------------------------
AMARG   CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGAAGGCACGCCACCCGCGACGGC-GGGCGGCGCGCATCCT-GCGAA-CGCATTA-CT--G--AAA
PGLOB   CCGGCGTGGA--CCACCCCAAGGAAAACCAAC-GGGGAGGTTTGGGCCCCCCCACCCC--GGCCCCTTTCCTTT-CGC-AA----TTTA-C----AAAAA
CLANC   CCGGCGTGGA--CCACGCCAAGGAAAACGAAA-GGGAAGGCTCGCCCCCGCGCAGGC--GGGCGGCGCGCCT-CCCGC-A--CTTTTTA-CTAAGCAAAA
```

Appendix 4B (continued)

```
HSUBA     CCGGCGTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCTTGGGCCCGCACACG--TGGGCCGTGC--CTT-CCGC-AA-C----TA-C----AAAAA
SAMPL     CGGG-GTGGA--CCACGCCAAGGAAAACCAAC-GGGGAGGCCCGTGCCCGCCCACGC-TGG-CAGGGCGCCTT-CCGCTAC-----TT--C---G-AAAA

                                                                [1400]
NHEMI     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAGTCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
PLUCI     ----------------------------------------------------------------------------------------------------
DAROM     ----------------------------------------------------------------------------------------------------
DLANC     C-ACTCTC-GCAAC-GATATCTC-GCTCTCT-CATCGATGAAGAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
DRETU     ----------------------------------------------------------------------------------------------------
DKERI     ----------------------------------------------------------------------------------------------------
DCONF     ------------------------------------------TTAAACTCAGCGGGTGGTCCCGCCTGACCTGGGATCGCGTTGC-----GAGGCCG
HPUBE     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAAGATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTGGC-----GAGGCCG
HMENG     TGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGA-----GAGGCCA
HCERN     CGACTCTCG-CAACGGATATCTCGG-TCTTG-CATCG-TGAAAAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
HDRYO     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATAAATGAAGAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GGGGCCA
HFERR     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GGGGCCA
HPIER     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
HNERV     TGACTCTCGGCAACGGATATCTCGGGTCTCG-CATTGATGAAAAA----------------------------------------------------
HNIGR     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCA
HCELT     ------------------------------------------TTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGT-----GAGGCCA
HCELE     TGACTCTTA-CAATGGATATCT-GGCTCTCG-CATCGATGAAAAATTAAACTCAGCGGGTAGTCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
HAPIC     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAGTCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
HWIGH     CGACTCTCGGCATCGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAAACCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
HBREV     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAG----
HJUCU     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
HCORD     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAAGATTAAACTCAGCGGGTATTCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
SMACT     TTACT-TT--C-A---T-TTTTTCGGTTTTTT-TTCT-TTAAAAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGCG--
SPILO     ------------------------------------------TTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
SKUNS     CCACTCTCGGCAACGGATATCTCGTCTCTCG-TATCGATGAAGAATTAAACTCAGCGGGTAATCCCGCCTGAT-TGGGGTCGCGTTGC-----GAGGCCG
SSELA     ------------------------------------------TTAAACTCAGCGGGTAGTCCCGCCTGACCTGGGGTCGCGTTGCGATGCGAGGCCG
SMAXI     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAGTCCCGCCTGACCTGGGGTCGCGTTGCGATGCGAGGCCG
SSCAB     CAACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
SSING     CGACTCTCGTCTACGGATATCTCGGCTCTCG-CATCGATGAAAA-TTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGCG--
SRICH     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAGTCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
SMAXW     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGTCG
SISOP     CCACTCTCGGCAACGGATATCTCGGCTCTTG-CATCGATGAAAAATTAAACTCAGCGGGTAATCCCGCCTGACCCGGGGTCGCGTTGC-----GAGGTCG
SGUIS     CAACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGTCG
SMACR     CGACTCTCTGCTACGGATATGTCGGGTCTCG-CATTCATGAAAAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
SSPLE     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
SLONG     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAGTCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
SFOXW     CAACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGTCG
SHOPE     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAGTCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
SFAGU     TGACTCTCAGGAACGGATATCTCGACTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAGTCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
```

Appendix 4B (continued)

```
SBECC     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
SMULT     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAGT-CCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCA
SOVAL     ACACTTTGGCGAAGGGATATGTGGGCTCTC--C-T---TGAAGAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAG----
SROXB     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAGTCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
SMATE     CAACTCTCGGAACCGGAT-TCT-GGCTCTGGCCATGAATGAAAAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGTCG
SEXEL     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAGTCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCA
SJAVA     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAAAATTAAACTCAGCGGGTAATCCCGCCTGACCCGGGGTCGCG--GC-----GAGGTCG
SSMIT     AAACTCTCGGCAACGGATATGTCGGCTCTCG-CATCAA-AAAAA-TTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAG----
SSEMI     CAACTCTTGTCAACGGATATCTCGGCTCTTG-TATCGATGAAGAA---AACTCAGCGGGTAGTCCCGCCTGACCTGGGGTCGCGTTGA-----GAGGCCG
SPARV     CGACTCTTGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAA----------------------------------------------------CG
SPALE     ----------------------------------------------TTAAACTCAGCGGGTAGTCCCGCTTGACCTGGGGTCGCGTTGC-----GAGGCCA
SFOLI     ----------------------------------------------TTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAG----
SSTEN     ----------------------------------------------TTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG
SLAEV     ----------------------------------------------TTAAACTCAGCGGGTAGTCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGTCG
SPING     CGACTCTCGGTAACGGATATCTCGGCTCTCG-CATCGATGAAAAA----------------------------------------------------
SCORD     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAA----------------------------------------------------
SBALA     CGACTCTCTGGAACGGATATCTCGGGTCTCT-CATTGATGAAGAA----------------------------------------------------
SJOHO     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAA----------------------------------------------------
SLEPR     CCACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAA----------------------------------------------------
SBARC     ----------------------------------------------------------------------------------------------------
AMARG     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CAT---TGAAAAATTAAACTCAGCGGGTAGTCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGTCG
PGLOB     CCACTTTCGGCAACGGATATCTGGGCTCTC--CAT---TGAAAAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGT--CGTTGC-----GAGGTCG
CLANC     CGACTCTCGGCAACGGATATCTCGGCTCTCG-CATCGATGAAGAATTAAACTCAGCGGGTAGTCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGTCG
HSUBA     TGACTCTTGGCAACGGATATCTCGGCTCTCG-CATTGATGAAAAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCA
SAMPL     CGACTCTCGGCAACGGATATCTCGGGTCTCG-CATCGATGAAAAATTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGTTGC-----GAGGCCG

                                                                                [1500]
NHEMI     TCCTGAGAGCAGGACGGCG---CGTTAGGG-TCCGGACGA---GCTCCCCAC---GGGACGCCGGGGGTGCGCAGGGG-ACA---CATGGACACG---CGA
PLUCI     ----------------------------------------------------------------------------------------------------
DAROM     ----------------------------------------------------------------------------------------------------
DLANC     -CCGGGCGGAC-----GCGC-T----G-G-G-CCA-ACTA----CTCCCCAC----G-ACTCCGGA-T-CTCAAGGGCACA-GACATGGACACAA-----
DRETU     ----------------------------------------------------------------------------------------------------
DKERI     ----------------------------------------------------------------------------------------------------
DCONF     TCA-TGCAGAC----GGCCGCT----GGGGGTCCA-ACAA---GGTCCCCGCG---TGGCGACGGG-CGCGC-CGGGCACGGGACGTGGAC------C-A
HPUBE     TCCTCACGGAC----GGCGCAT---CAGGG-TCCGAACGA---GCTCCCCAC---GAGACGCCGGG-AGCGCAGGGGCACA---TATGGACA---CACGA
HMENG     TCCGCACGGAC----GGCGCAT---CAGGGG-CCGGACGA---GCTCCCCAC---GAGACGCC-AGGAGCACAGGGG-ACA---TATGGACAC---ACGA
HCERN     TCCGCACGGAC----GGCACCT---CAGGG-TCCGGACGA---GCTCCCCAC---GGGACGCCGAG-TGCGCAGGCG-ACA---CATGGACACG---CGA
HDRYO     TCCGCATGGAC----GGCGCAT---CAGGG-TGCAAACGA---GCTCCCCAC---GAGACGCCGAG-AGCACAGGGGCACA---TATGGACACG---CGA
HFERR     TCCGCATGGAC----GGCGCAT---CAGGG-TGCAAACGA---GCTCCCCAC---GAGACGCCGAG-AGCACAGGGGCACA---TATGGACACG---CGA
HPIER     TCCGCACGAAC----GGCGCAT---CAGGG-TCCGAACGA---GCTCCCCAC---GAGACGCCGAG-AGCGCAGGGGCACA---TATGGACA---CACGA
HNERV     ----------------------------------------------------------------------------------------------------
HNIGR     TCCGCACGGAC----GGCACCT---CAGGG-TCCAGACGA---GCTCCCCAC---GGGACACCGAG-TGCACAGGCG-ACA---TATGGAC---ACACGA
HCELT     TCCGCACGGAC----AGCGCAT---CAGGA-TCCAGACAAGAAGCTCCCCAC---GAGACGCCG-GGAGCGCAGGGGCACA---TATGGACA---CACGA
```

Appendix 4B (continued)

```
HCELE     TCCGCGCGGACGGACGGCGCATCGTCGGGG-TCCGGACGA---GCTCCCGCAAATGGGACGCCGGG-TGCGCAGGGGCACAGGACGCGGACACGACACGA
HAPIC     TCCGCGCGGAC----GGCGCAT---CAGGG-TCCGGACGA---GTTCCCCAC---GGGACGCCGGG-TGCGCAGGGGCACAGGACATGGACA-----CGA
HWIGH     TCCCCACGGAC----GGGGCAAA--CAGGG-TCCGGACGA---GCTCCCCAC---GGGACGCCGGG-TGCGCAAGGGCACAAGACAWGGACA-----CGA
HBREV     ----CGCGGAC----GGCGCTTCG--GGGG-TCCTGACGA---GCTCCCCAC---GGGACGCCGGG-TGCGCAGGGGCACGGGACACGGAC--G---CGA
HJUCU     TCCGCACAGAC----GGCACATT----GGGGTCCTAACGA---GCTCCCCACA---GGACGCC-AGGAGCGCAGGGG-ACA---TATGGACACG---CGA
HCORD     TCCGCACGGAC----GGCACAT---C--GGGTCCGAACGA---GCTCCCCACA---GGACGCC-AGGAGCGCAGGGG-ACA---TATGGACACG---CGA
SMACT     ----------CGGACGGCGCTT---CGGGGGTCCTGACGA---GCTCCCCAC---GGGACGCCGGG-TGCGCAGGGGCACGGGACACGGAC-----GCGA
SPILO     TCCACGCGGAC----GGCGCTT---CGGGGGTCCTGACGA---GCTCCCCAC---GGGACGCCGGG-TGCGCAGGGGCACGGGACACGGAC-----GCGA
SKUNS     TCCGCGCGGAC----GGCGCTT---CGGGGGTCCTGACGA---GCTCCCCAC---GGGACGCCGGG-TGCGCAGGGGCACGGGACACGGAC-----GCGA
SSELA     TCCGCGCGGACGGACGGCGCATCGTCGGGG-TCCGGACGA---GCTCCCGAAAATGGGACGCCGGG-TGCGCAGGGGCACAGGACGCGGACACGACACGA
SMAXI     TCCGCGCGGACGGACGGCGCATCGTCGGGG-TCCGGACGA---GCTCCCGAAAATGGGACGCCGGG-TGCGCAGGGGCACAGGACGCGGACACGACACGA
SSCAB     TCCGCGCGGAC----GGCGCAT---CAGGG-TCCGGACGA---GCTCCCCA---TGGGACGCCGGG-TGCGCAGGGGCACAGGACATGGACA-----CGA
SSING     ----------CGGACGGCGCTT---CGGGGGTCCTGACGA---GCTCCCCAC---GGGACGCCGGG-TGCGCAGGGGCACGGGACACGGAC-----GCGA
SRICH     TCCGCGCGGACGGACGGCGCATCGTCGGGG-TCCGGACGA---GCTCCCGGAAATGGGACGCCGGG-TGCGCAGGGGCACAGGACGCGGACACGACACGA
SMAXW     TCCGCGCGGAC----GACGCTT----AAGGGTCAAGAC----------CCAC---GGGACGCCGAG-TGCGCAGGGGCACAGGACACGGAC--G---CGA
SISOP     GCCGCGGGGAC----GACGCTT---AAGGGG-CAAAA----------CCCAC---GGGACGCCGAGG-GCGCAGGGGCACAAGACACGGAC--S---CGA
SGUIS     TCCGCGTGGAC----GACGCTT---AAGGG-TCAAAA----------CCCAC---GGGACGCCGAG-TGCGCAGGGGCACAAGACACGGAC--G---CGA
SMACR     TCCGCGCGGAC----GGCGCTT---CGGGGGTCCTGACGA---GCTCCCCAC---GGGACGCCGGG-TGCGCAGGGGCACGGGACACGGAC--G---CGA
SSPLE     TCCGCGCGGAC----GGCGCTT---CGGGG-TCCTGACGA---GCTCCCCAC---GGGACGCCGGG-TGCGCAGGGGCGCGGGACACGGAC--G---CGA
SLONG     TCCGCGCGGACGGACGGCGCATCGTCGGGG-TCCGGACGA---GCTCCCGCGAATGGGACGCCGGG-TGCGCAGGGGCACAGGACGCGGACACGACACGA
SFOXW     TCCCCGCGGAC----GACGCTT---CAGGG-TGCAAAA----------CCAC---GGGACGCCGGG-TGCGCAGGGGCACAGGACACGGAC--G---CGA
SHOPE     TCCGCGCGGACGGACGGCGCATCGTCGGGG-TCCGGACGA---GCTCCCGCAAATGGGACGCCGGG-TGCGCAGGGGCACAGGACGCGGACACGACACGA
SFAGU     TCCGCACGGACGGACGGCGCATCGTCGGGG-TCCGGACGA---GCTCCCGCGAATGGGACGCCGGG-TGCGCAGGGGAACAGGACGCGGACACG------
SBECC     TCCGCGCGGAC----GGCGCTT---CGGGGGTCCTGACGA---GCTCCCCAC---GGGACGCCGGG-TGCGCAGGGGCACGGGACACGGAC--G---CGA
SMULT     TCCGCGTGGACGGACGGCGCATCGTCGGGG-TCCGGACGA---GCTCCCGAAAATGGGACGCCGGG-TGCGCAGGGGCACAGGACGCGGACACGACACGA
SOVAL     ----CGCGGAC----GGCGCTT---CGGGGGTCWTGACGA---GCTCCCCAC---GGGACGCCG-GGTGCGCAGGGGCACGGGACACGGAC--G---CGA
SROXB     TCCGCGCGGAC----GGCG---CGTCAGGG-TCCGGACGA---GCTCCCCG---TGGGACGCCG-GGTGCGCAGGGGCACAGGA----------CACGA
SMATE     TCCGCGTGGAC----GACGCTT---AAGGG-TCAAAAC----------CCAC---GGGACGCCGAG-TGCGCAGGGGCACAAGACACGGAC--G---CGA
SEXEL     TCCGCGAGGAC----GGCGCAT---CAGGG-TCCGAACGA---GCTCCCCG---TGGGACGCC-AGGTGCGCAGGGGCACAAGACACGGA-----CACGA
SJAVA     GCCGCGGGGAC----GACGCTT---AAGGGG-CAAAA----------CCCAC---GGGACGCCGAGG-GCGCAGGGGCACAAGACACGGAC--S---CGA
SSMIT     ----CGCGGAC----GGCGCTT---CGGGGGTCATGACGA---GCTCCCCAC---GGGACGCCGGG-TGCGCAGGGGCACGGGACACGGAC--G---CGA
SSEMI     -CCGCGCGGAC----GGCGCTT---CGGGGGGCCTGACGA---GCTCCCCAC---GGGAC-CCGGGG-GCGCAGGGGCACGGGACACGGAC--G---CGA
SPARV     GACGGCGGGAC----GGCGCT----CCGGCGTCCTGACGAA--GCTCCCCAC---GGGACGCCGGGG-GCGCAGGGGCACGGGACACGGAC--G---CGA
SPALE     TCCGCGCGGAC----GGCACTT---CGGGGGTCCTGACGA---GCTCCCCAC---GGGACGCCG-GGTGCGCAGGGGCACGGGACACGGAC--G---CGA
SFOLI     ----CGCGGAT----GGCGCTT---CGGGGGTCCTGACGA---GCTCCCCAC---GGGACGCCG-GGTGCGCAGGGGCACGGGACACGGAC--G---CGA
SSTEN     TCCGCGCGGAC----GGCGCTT---CGGGGGTCCTGACGA---GCTCCCCAC---GGGACGCCG-GGTGCGCAGGGGCACGGGACACGGAC--G---CGA
SLAEV     TCCGCGCGGAC----GACGCTT---CAGGG-TCCGAAGGA---GCTCCCCAC---GGGACGC--GGGTGCGCAGGGGCACAAGACACGGAC-----GCGA
SPING     ----------
SCORD     ----------
SBALA     ----------
SJOHO     ----------
SLEPR     ----------
SBARC     ----------
```

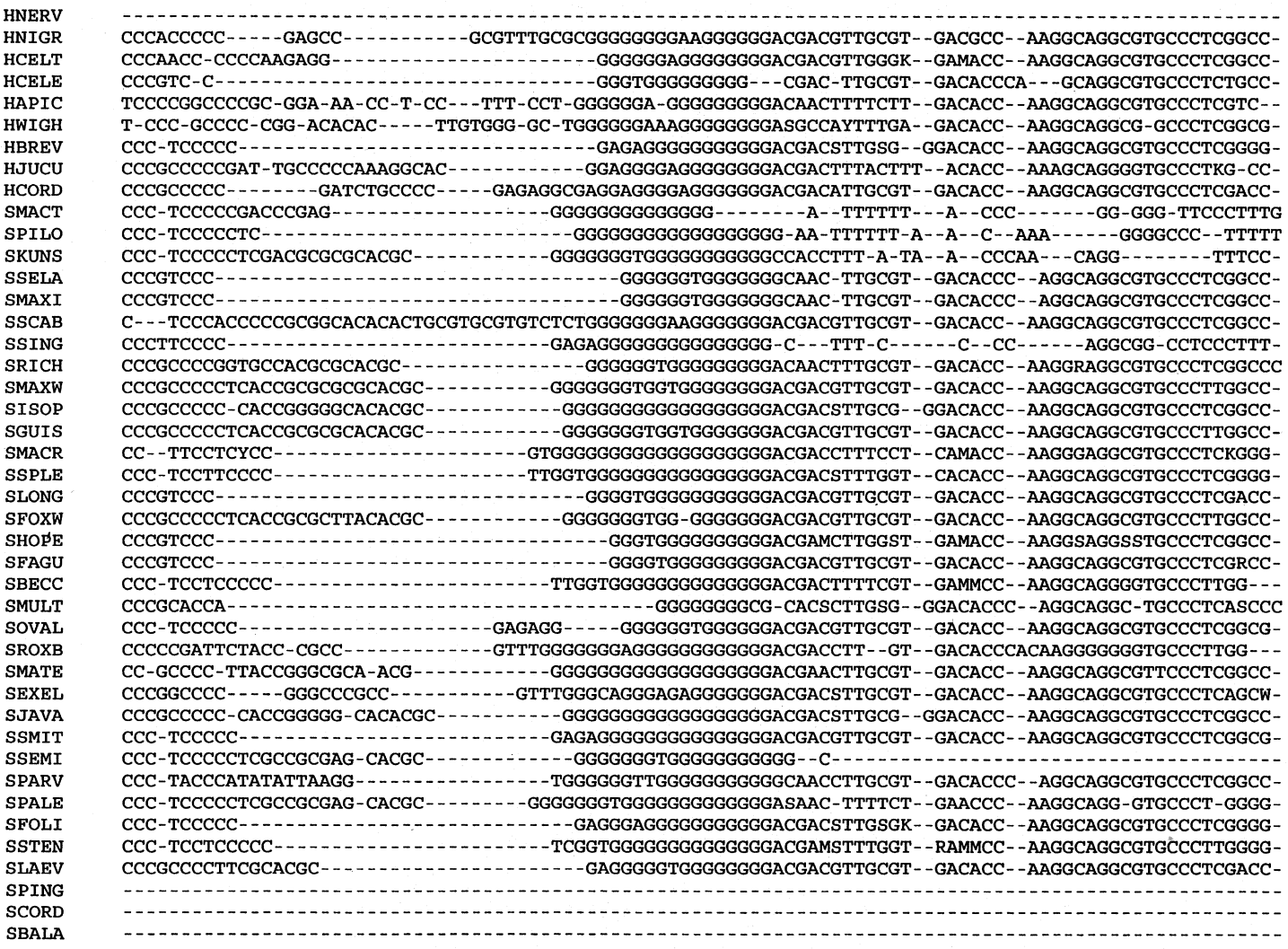
Appendix 4B (continued)

```
AMARG     TCCGCCGTG---CGCGGCGC---GACGGGG-TCCGAACGA---GTTCCCTGCC---SGACGCCGAGG-GCGCAGGGGCACTGGACATGGAC--G---TGA
PGLOB     TCCGCGTGGAC----GACGCTT---AAGGG-TCAAAAC----------CCAC---GGGACGCCGAG-TGCGCAGGGGCACCAGGCAACGACCCCAAGCAA
CLANC     -CGGCGC-GAC---------------GGGGTCCGAACGA---GTT----------GACGCCGAGG-GCGCAGGGGCACTTGACATGGAC--G---TGG
HSUBA     TCCGCACGGAC----GGCGCCT---CAGGG-TCCAAACGA---GCTCCCCAC---GGGACACCGAG-TGCACAGGGG-ACA---TATGGACAC---ACGA
SAMPL     TCCGCGCGGAC----GGCGCTT---CGGGGGTCCTGACGA---GCTCCCCAC---GGGACGCCG-GGTGCGCAGGGGCACGGGACACGGAC--G---CGA

                                                                                [1600]
NHEMI     G--CGGCCACCGCT-TGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGAAG--TGCGCACGGGAGGCCAGCATCCGCC
PLUCI     ----------------------------------------------------------------------------------------------------
DAROM     ----------------------------------------------------------------------------------------------------
DLANC     ---CGGCCACCG-T-TGCCCTTGCGCTCGGCGCTT-ATGACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCGTCCGCC
DRETU     ----------------------------------------------------------------------------------------------------
DKERI     ----------------------------------------------------------------------------------------------------
DCONF     G--CGGCCACCACT-TGCCCGTGCGCCCGGCGCCGG-GGACTC--TTTT-GGGCCAACCGCGAGC--GGAGCGAG-GCACGCGGGAGGCCA-CATCCGCC
HPUBE     G--CGGCCACCGCT-TGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCACACGGGAGGCCAGCATCCGCC
HMENG     G--CGGCCACCGCT-TGCCCTTGCGCCTGGCGCCGAGG-ACTCAATTTT-GGGCCAACCGCGAGC--GGGTAG--CGCACACGGGAGGCCAGCATCCGCC
HCERN     G--CGGCCACCGCT-TGCCCTTGCACCCGGCACCGGGG-ACTCGCTTTT-GGGCCAACCGCGAGC--GAGGAG--CGCACACGGGAGGCCAGCATCCGCC
HDRYO     G--CGACCACCGCT-TGCCCTTGCGCCTGGCGCCGGGG-ACTCAATTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGCC
HFERR     G--CGACCACCGCT-TGCCCTTGCGCCTGGCGCCGGGG-ACTCAATTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGCC
HPIER     G--TGGCCACCGCT-TGCCCTTGCGCCCGGTGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCACACGGGAGGCCAGCATCCGCC
HNERV     ----------------------------------------------------------------------------------------------------
HNIGR     G--CGGCCACCGCT-TGCCCTTGCACCCGGCACCGGGG-ACTCGCTTTT-GGGCCAACCGCGAGC--GACGAG--CGCACACGGGAGGCCAGCATCCGCC
HCELT     G--CGACCACYGGT-TGYCCTTGCACCCGGCGCCGGGG-AGTYACTTTT-GGGCCAACCGCGAGC--GGGGAG--CCCCCACGGGAGGGCATCATCCGCC
HCELE     G--CGGCCACCGCT-TGCCCTTGCGCCCGGCGCCGGGGGGCTCATTTTT-CGGCCAACCGCGAGC--GGAGAG--CGCGCACGGGAGGCCAGC-TCCGCC
HAPIC     G--CGGCCACCGCT-TGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGCC
HWIGH     A--CGGCCACCGCT-TGCCCTTGCGCCCGGCGCCAGGG-ACTCACTTTT-GGGCCAACCGCAAGC--GGGGAG--CGCACACGGGAGGSCAACATCCGCC
HBREV     G--CGACCACAGCTCTGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGAC
HJUCU     G--CGGCCACCGCT-TGCCCTTGCACCCGGCGCCAGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GCAGAG--CGCACACGGGAGGCCAGCATCCGCC
HCORD     G--CGGCCACCGCT-TGCCCTTGCACCCGGCGCCAGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GCAGAG--CGCACACGGGAGGCCAGCATCCGCC
SMACT     G--CGACCACAG-TCTGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGAC
SPILO     G--CGACCACAGCTCTGCCCTTGCGCCCGGCGCCGGGG-ACT-ACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCATGGGAGGCCAGCATCCGAC
SKUNS     G--CGGCCACAGCTCTGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAAGCCAGCATCCGAC
SSELA     G--CGGCCACCGCT-TGCCCTTGCGCCCGGCGCCGGGGGGCTCA-TTTT-GGGCCAACCGCGAGC--GGAAAG--CGCGCACGGGAGGCCAGC-TCCGCC
SMAXI     G--CGGCCACCGCT-TGCCCTTGCGCCCGGCGCCGGGGGGCTCA-TTTT-GGGCCAACCGCGAGC--GGAAAG--CGCGCACGGGAGGCCAGC-TCCGCC
SSCAB     G--CGGCCACCGCT-TGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGCC
SSING     G--CGACCACAGCTTT-CCCTT--GCCCGG-GCCGGGG-ACTTACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGA-G-CAACATTC-AC
SRICH     G--CGGCCACCGCT-TGCCCTTGCGCCCGGCGCCGGGGGGCTCACTTTT-CGGCCAACCGCGAGC----GGAGAGCGCGCACGGGAGGCCAGCATCCGCC
SMAXW     G--CGACCACCGCT-TGCCCTTGCGCCCGGCGCCGAGG-ACTCT-TTTT-GGGCCAACCGCGAGCACGGGGAGAGCGCGCACGGGAGGCCAGCATCCGCC
SISOP     G--CGACCACCGCT-TGCCCYTGCGCCCGGCGGCGGGG-ACCCCCYTTT-GGGCCAACCGCGAGC--GGGGAGAGCGCSCACGGGAGGSCAGCATCCGCC
SGUIS     G--CGACCACCGCT-TGCCCTTGCGCCCGGCGTCGGGG-ACTCTCTTTT-GGGCCAACCGCGAGC--GGGGAGAGCGCGCACGGGAGGCCAGCATCCGCC
SMACR     G--CGACCACAGCTCTGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGAC
SSPLE     G--CGACCACAGCTCTGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGAC
SLONG     G--CGGCCACCGCT-TGCCCTTGCGCCCGGCGCCGGGGGGCTCACTTTT-GGGCCAACCGCGAGC--GCAGAG--CGCGCACGGGAGGCCAGGATCCGCC
```

Appendix 4B (continued)

```
SFOXW     G--CGACCACCGCT-TGCCCTTGCGCCCGGCGCCGGGG-ACTCTCTTTT-GGGCCAACCGCGAGC--GGGGAGAGCGCGCACGGGAGGCCAGCATCCGCC
SHOPE     G--CGGCCACCGCT-TGCCCTTGCGCCCGGCGCCGGGGGGCTCATTTTT-CGGCCAACCGCGAGC--GGAGAG--CGCGCACGGGAGGCCAGCATCCGCC
SFAGU     -TCGTGCCACCGCT-TGCCCTTGCGCCCGGCGCCGGGGGGCTCACTTTT-GGGCCAACCGCGAGC--GGAGAG--CGCGCACGGGAGGCCAGAATCCGCC
SBECC     G--CGACCACAGCTCTGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGAC
SMULT     G--CGGCCACCGCT-TGCCCTTGCGCCCGG-GCCGGGGGGCT-ACTTTT--GGCCAACCGCGAGC--GGAGAG--CGCGCACGGGAGGCCAGCATCCGCC
SOVAL     G--CGACCACAGCTCTGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGAC
SROXB     G--CGGCCACCGCT-TGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-CGGCCAACCGCGAGC--GGGGAGAG--CGCACGGGAGACCAGCATCCGCC
SMATE     G--CGACCACCGCT-TGCCCTTGCGCCCGGCGTCGGGG-ACTCTCTTTT-GGGCCAACCGCGAGC--GGGGAGAGCGCGCACGGGAGGGCAGCATYCGCC
SEXEL     G--CGGCCACCGYT-TGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTTTGGGCCAACCGCGAGC--GAGGAGAG--CGCACGGGAGGCCAGCATCCGCC
SJAVA     G--CGACCACCGCT-TGCCCYTGCGCCCGGCGGCGGGG-ACCCCCYTTT-GGGCCAACCGCGAGC--GGGGAGAGCGCSCACGGGAGGSCAGCATCCGCC
SSMIT     G--CGACCACAGCTCTGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGAC
SSEMI     G--CGGCCACAGCTCTGCCCTTGCGACCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGAC
SPARV     G--CGTCCACAGTTCCGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGCTTGGAGAG----CGCACGGGAGGCCAGCATTCGAC
SPALE     G--CGGCCACAGCTCTGCCCTTGCGACCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGAC
SFOLI     G--CGACCACAGCTCTGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGAC
SSTEN     G--CGACCACAGCTCTGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGAC
SLAEV     G--CGACCACCGGT-TGCCCTTGCGCCCAGCGCCGGGG-ACTCGCTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGCC
SPING     ----------------------------------------------------------------------------------------------------
SCORD     ----------------------------------------------------------------------------------------------------
SBALA     ----------------------------------------------------------------------------------------------------
SJOHO     ----------------------------------------------------------------------------------------------------
SLEPR     ----------------------------------------------------------------------------------------------------
SBARC     ----------------------------------------------------------------------------------------------------
AMARG     G--ACGCCACCGCT-TGCCCGTGCGCTCGGCGCCGAGG-ACTCACTTTT-GGGCCAACCGCGGGC--GGGGAA--CGCACGCGGGAGGCCAGCTTCCGCC
PGLOB     CAACGGTTGGCCTT-TGCCC------CCGGGGT-GGGGAACTTTTTTTT-GGGCCAACCGCAACC--GGGGAAAGCGCCCACGGAAGGCCAGATTCCGCC
CLANC     G--CGACCACCGCT-TGCCCGTGCGCTCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGGGC--GGGGAA--CGCGCGCGGGAGGCCAGCATCCGCC
HSUBA     G--CGGCCACCGCT-TGCCCTTGCACCCGGCACCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GAGGAG--CGCACACGGGAGGGCAACATYCGCC
SAMPL     G--CGACCACAGCTCTGCCCTTGCGCCCGGCGCCGGGG-ACTCACTTTT-GGGCCAACCGCGAGC--GGGGAG--CGCGCACGGGAGGCCAGCATCCGAC

                                                                                                    [1700]
NHEMI     CTCGTCCC-----TGAGCCACGCCTCCA---------GGCGCGGAGGGAGGGGGGGACGACGTTGCGT--GACACC--AAGGCAGGCGTGCCCTCGGCC-
PLUCI     ----------------------------------------------------------------------------------------------------
DAROM     ----------------------------------------------------------------------------------------------------
DLANC     CCTCCGCCGATCCCCGGCGCGCACC-GTGTGCGTGCGCGGGGGGGAAGGGGGGG---CGACGTTGCGT--GACGCCCA--GGCAGGCGTGCCCTCGGCC-
DRETU     ----------------------------------------------------------------------------------------------------
DKERI     ----------------------------------------------------------------------------------------------------
DCONF     CCCTCCCCGCGCCAC----------G-GTTG-TGGGGGGGAGGT-TCGGGGGG---CGACGTTGCGT--GACACCCA--GGCAGGCGTGCCCTTGGCC-
HPUBE     CCCGTCCCC--------GAG-C-CCCC-------GAGA-GGAGGGGGAA-GGGGGGAYGACATT--GAGGGACACC--AACSRTGGGGTTGCCTGATTST
HMENG     CCCGCCCCCAAGCCACCCCGAGAGGC--------------GGGGAGAGGGGGGTGACGACGTTGTGT--GACACC--AAGGCAGGCGTGCCCTCGGCC-
HCERN     CCCACCCCCGGGCCGCG-----------------TTTGCGGGGGGGGAAGG-GGGACGACGTTGCGT--GACCCC--AAGGCAGGCGTGCCCTCGGCC-
HDRYO     CCCACCCCCAAGCAGCCCCGAGAGG---------------GGGGAGAGGGGGGGGACGACCTTTCTT--GACACC--AACGCAGGCGTGCCCTCCGCC-
HFERR     CCCACCCCCAAGCAACCCCGAGAGG---------------GGGGAGAGGGGGGGGACGAMCTTTTTT---ACACC--AMAGCMGGGGTGCCCTTCGGC-
HPIER     CCCATCCCC--------GAGCCA----------GAGAGGTAGGGGGAAGGGAGGGACGACGTTGCGT--GACACC--AAGGCAGGCGTGCCCTCGGCC-
```

Appendix 4B (continued)



Appendix 4B (continued)

```
SJOHO     ----------------------------------------------------------------------------------------------------
SLEPR     ----------------------------------------------------------------------------------------------------
SBARC     ----------------------------------------------------------------------------------------------------
AMARG     CCC-TCCCC-GCTCCGCCCCCCGTGGAACGGCGGATC--GAGAGGGGGGTTGGGGGACGACGTTGCGT--GACACC--AAGGCAGGCGTGCCCTCGGCC-
PGLOB     CCCGCCCCTTAACCGGGCGCACACCC-----------GGGGGGGGGGGGGGGGGCCAACTTTGCTT--GACACCC--AGGCAGGCGTGCCCTCGGCC-
CLANC     CCC-TCCCCGC-------------------------T-GGGGAGGTCGTGGGGGACGACGTTGCGT--GA-ACC--AAAGCAGGCGTGCCCTTG-CC-
HSUBA     CCCACCCCCGAACCG-GTTTG-G-----------------GGGGG-AAGGGGGGACAACTTTGCGT--GACACC--AAGGCAGGCGTGCCCTCGGCC-
SAMPL     CCC-TCCTCCCCC----------------------TTGGGGGGGGGGGGGGGGGGACGAMCTTTGGK--RAMACC--AAGGAAGGCGTGCCCTCGGGG-

                                                                [1800]                                [1842]
NHEMI     ---TGATGGC----TTGGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
PLUCI     ----------------------------------------------------------------------------------------------
DAROM     ----------------------------------------------------------------------------------------------
DLANC     ---TGACGGC----TTGGGGCGCA-ACTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
DRETU     ----------------------------------------------------------------------------------------------
DKERI     ----------------------------------------------------------------------------------------------
DCONF     ---TGACGGCC---TTGGG-CGCA-ACTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TT-TGCAATTCACACCAA-GTATCGCATT
HPUBE     GGTGGA--------TTGGGGCGCAAA--TGGCTTTAAAAGACTCGA-TGGGTT--ACGGGA--TTCTGCAATTCACACCTA-G-ATTGCATT
HMENG     ---TGATGGC----TTGGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
HCERN     ---TGGTGGC----TTGGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CGCGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
HDRYO     ---TGGTGGG----TTGGGGCGCAA-CTTGCG-TTCAAAGACTTGA-TGGG-T-CACGGGA--TTGTTCAATTCACACCAA-GTATTGCATT
HFERR     ---TGGTGGG----TTGGGGGCCCA-CTTGGGTTTCAA-GACTCGA-TGGGTT-G-CGGGAA-TTTT-CAATTTACACCAA-GTATTGSATT
HPIER     ---TGATGGC----TTGGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
HNERV     ----------------------------------------------------------------------------------------------
HNIGR     ---TGATGGC----TTGGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
HCELT     ---TGGTGG-----TTGGGGCGCAA-CTTGCG-TTCAAAGACTCGG-TGG-TT-CACGGGA--TTGTGCAATTCACACCAA-GTGTCGCATT
HCELE     ---TGACGG-----TTGGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TT-TGCAATTCACACCAA-GTATCGCATT
HAPIC     ---TTAAGGG----TTGGGGCGCAA-CTTGGG-TTCAAAGACTCTA--GGGGT-CACGGGA--TTTTGCAATTCACACCCA-ATTTCGCATT
HWIGH     -TTTGACGGC---TTTGGGGCACAAA-GTGTGATTAAAACACTCGA--GGGG-ACACGG--AATACTCCAAAACACACCCAAAATTC-CA--
HBREV     -TTTGACGGC----TTGGGGCCCAA-CTTGCG-TTCAAAGACTCGA--GGGTT-CACGGGAA--TCTGCAAATCACACCAA-GTATCGCATT
HJUCU     --TTGACGGG----TTGGGGC-CAA-CTCGCC-TTAAAAGACTAGA-TGG-TT-CAGGGGA--TTTGGAAATTCACACCAA-G-CTCGCATT
HCORD     ---TGATGGC----TTCGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
SMACT     G-TTTT---GG---TT-GGG-CC---TTT---TTTAAAAAA-TT----GGG-T----GG--ATTTTTC-ATTT---CC-AAATTT---TTTT
SPILO     TTTTA--G------TT-GGG-CC---TTTTTTTT-AAAAAA-TTT------TT----GG---TTTTTTTTTT------AAA-TT--TTTTT
SKUNS     TTTCCCAGG-----TTTGGG-CCAA-TTT--GTTTAAAAAA-T--A--GGGGT----GG-A--TTTTTCA-TTC---CCAA--TTTT---TT
SSELA     ---TGACGG----TTTGGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
SMAXI     ---TGACGG----TTTGGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
SSCAB     ---TGACGGC----TTGGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
SSING     -G-TTG---GG----TGGGG-CC-A--TTTTCA-AAAAAAATT------GGT-GGTG-A-G--TTTTTT-AAT------CAA-G-TTTT-TT
SRICH     GCTTGACAACGG-TTTCGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTTTGCAATTAACACCAA-GTTT-G--TT
SMAXW     ----GA-GGC----TTAGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GT-TCGCATT
SISOP     ----GA-GGC----TTAGGGCGCAA-CTTGCG-TTCAAAGACTCGAATGG--T-CACGGGAA--TCCGCAAATCACACCAA-GGATCGCAT-
SGUIS     ----GA-GGC----TTAGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
```

Appendix 4B (continued)

```
SMACR     -TTTGACGGC----TTGGGGCCCAA-CTTGG-TTTCAAACACGCGG-TGGGT-ACAGGGGA--TTCTTAAATTCACACCA-------CCTTT
SSPLE     -TTTGACGGC----TTGGGGCCCAA-CTTGGG-TTCAAAGACTCGA--GGGTT-GSCGGGA--TTTTTTTTTTCA-ACCCA------CCTTT
SLONG     ---TGACGGC---TTTGGG-CGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
SFOXW     ----GA-GGG----TTAAGGCGCAA-CCTGCG-TTCAAAGACTCGA-TGGG-T-CACGGGAA--TCTGCAATTCACACCAA-GATT-GCATT
SHOPE     ---TGACGGC----TTKGGGGGCAA-CTTGCG-TTCAAAGACTCGAATGG-TT-CACGG--AATTTT-CAATTCACACCAAAGTATGG-ATT
SFAGU     ---TGACGGG----TTCGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
SBECC     -TTTGACGG----TTTGGGGCGCAC-CTTGGG-TTCAAAGACTCAA-TGGGT--CACGGGA--TTTTTCAATTAACACCAA-GTATTGCTTT
SMULT     GACTGACGAGCYGTTTACCGCC-TT-CTTGCACTTTAA-GACTCAA-TGG-TTT-ACAAAAA--TTTGCAATTTACACCAA-GTATCGCATT
SOVAL     -TTTGACGGC----TTGGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
SROXB     TTTTGGGG-----TTTGGGGC-CAA-CTTTGG-TTCAAAGACTC--TTGGGGTTCACGGGAA-TTCTGCA-TTCACCCCCAA-TTTT-CTTT
SMATE     ----GA-GGC----TTARGGCGCAA-CTTGCG-TTCAAAGACTCGAATGG-TT--ACGGGAA--TCTTCAATTTAAACCAA-GTATCGCATT
SEXEL     -TTTGACGGC----TTGGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
SJAVA     ----GA-GGC----TTAGGGCGCAA-CTTGCG-TTCAAAGACTCGAATGG--T-CACGGGAA--TCCGCAAATCACACCAA-GGATCGCAT-
SSMIT     -TTTGACGGC----TTGGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
SSEMI     ----------------------------------------------------------------------------------------
SPARV     ---TGACGG----TTTGGGGCGAAA-CTTGCG-TTAAAAAACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTTTCGCATT
SPALE     --TTGGGGGK----TTTGGGCCCAA-CTTGGG-TTCAAAGACTC-TTTGGGTT-GACGGGA--TTTTGCAATTCACACCAA-GTTTTGGTTT
SFOLI     -TTTGACGGC----TTGGGGCCCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTTTGCAATTCACACCAA-GTATCGCATT
SSTEN     -TTTGACGGC----TTGGGGGCCAA-CTTGGG-TTCAAAGACTCGA--GGGTTTCACGGGA--TTTTGAAATTCAACCCAA-TTTTTT-ATT
SLAEV     ---TGACGGC----TTGGGGCGCAA-CTTGCG--TCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
SPING     ----------------------------------------------------------------------------------------
SCORD     ----------------------------------------------------------------------------------------
SBALA     ----------------------------------------------------------------------------------------
SJOHO     ----------------------------------------------------------------------------------------
SLEPR     ----------------------------------------------------------------------------------------
SBARC     ----------------------------------------------------------------------------------------
AMARG     ---TGACGGC----TTGGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATTCACACCAA-GTATCGCATT
PGLOB     ----GA-GG----TTTAGGGCGCAC-CTTGCG-TTAAAAAACTCGA-TGG-TT-CACGG--AATTCTGCAATTCACACCAA-GTATCCGATT
CLANC     ---TGACGGC----TTCGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGA--TTCTGCAATT-ACACCCAAG-ATCGAATT
HSUBA     ---TGAKGG----TTTGGGGCGCAA-CTTGCG-TTCAAAGACTCGA-TGG-TT-CACGGGAA-TTCTGCAATT-ACACCAA-GTATCG-ATT
SAMPL     -TTTGACGGC---TTTGGGGC-CAA-CTTGGG-TTCAAAGACTCCAA-GGGG-CCACGGGAAATTTTACAAA-CA-ACCAA-GTCTTCCATT
```

CHAPTER 5
COMBINED ANALYSIS

Outline

5.1 Introduction

A question that troubles evolutionary biologists is whether the data sets available to them are sufficiently comprehensive to enable them to make inferences concerning evolutionary processes. The extent to which the topology can be seen as representative of the true phylogeny is of major concern. There are always restrictions imposed on phylogenetic analyses by the type of data sets used to reconstruct the phylogeny. Morphological data are often criticised as being "plastic", which thus make it difficult to deduce the homologies between characters. The main limitation of molecular data is that they reflect the gene tree rather than the taxon tree, even though advances in molecular studies have undoubtedly provided new insights into evolutionary processes. The different natures of morphological and molecular data sets therefore often result in differing phylogenetic inferences.

It is thus not unexpected that the results of the cladistic analyses of molecular and morphological data presented here show rather different topologies. The morphological analysis suggests the non-monophyly of both *Hopea* and *Shorea* (Chapter 3), while the molecular data show in at least some of the analyses that *Hopea* is potentially monophyletic while *Shorea* is broadly non-monophyletic (Chapter 4). However, this incongruence between the topologies should not be interpreted as an error, but rather explored to search for biologically meaningful

information. This is exemplified by the incongruence between the analyses of the *trn*L-F and ITS molecular data (refer to Chapter 4, section 4.8.2).

Since the nature of morphological and molecular data is different, combining such data sets is often the subject of debate among systematists (Bull *et al.*, 1993; de Queiroz *et al*, 1993; Weiblen, 2000). This is particularly the case when there is conflict between data sets. This conflict can occur as a result of systematic error, rate heterogeneity, or because data sets do not share the same phylogenetic history (Bull *et al.*, 1993; Weiblen, 2000). However, when it is likely that datasets share the same phylogenetic history, phylogenetic inferences made from all the available data are considered more likely to be accurate than those made using only a subset of the data (Kluge, 1989; Barret *et al.*, 1991). Another advantage of combining data is that it provides the best estimate of phylogeny when incongruence resulting from conflict between subsets of data is due to random error (de Queiroz *et al.*, 1995).

5.2 Selection of characters for analysis

All of the characters used in the morphological and molecular analyses in this study were combined and a cladistic analysis performed on the combined matrix. All 40 morphological characters and the 1844 molecular characters were included in the analysis, providing a total of 1884 characters.

5.3 Selection of taxa for analysis

Only those taxa which had characters scored for all three data sets—morphology, *trn*L-F and *ITS*—were included in the combined analysis. A total of 42 species was included, consisting of *Dryobalanops lanceolata*, *Neobalanocarpus heimii*, 13 species of *Hopea* and 27 species of *Shorea* (Table 5.1).

Table 5.1 Taxa included in the combined analysis

Outgroup

Species	Abbreviation
Neobalanocarpus heimii (King) Ashton	NHEMI
Dryobalanops. lanceolata Burck	DLANC
Total number of outgroup species	2

Hopea

Section	Subsection	Species	Abbreviation
Dryobalanoides	*Dryobalanoides*	*H. pubescens* Ridley	HPUBE
		H. mengerawan Miq.	HMENG
		***H. dryobalanoides* Miq.**	HDRYO
		H. ferruginea Parijs.	HFERR
		H. pierrei Hance	HPIER
	Sphaerocarpa	*H. nigra* Burck	HNIGR
		H. subalata Symington	HSUBA
Hopea	*Hopea*	*H. celtidifolia* Kosterm.	HCELT
		H. celebica Burck	HCELE
	Pierrea	*H. apiculata* Symington	HAPIC
		H. wightiana Miq. ex Dyer	HWIGH
		H. brevipetiolaris (Thw.) Ashton	HBREV
		H. jucunda Thw.	HJUCU
Total number of *Hopea* species			13

Shorea

Section	Subsection	Species	Abbreviation
Shorea	*Shorea*	*S. guiso* Blume	SGUIS
		S. foxworthyi Symington	SFOXW
		S. seminis V.Slooten	SSEMI
	Barbata	*S. laevis* Ridley	SLAEV
		***S. maxwelliana* King**	SMAXW
*Neohopea**		***S. isoptera* Ashton**	SISOP
Richetioides	*Richetioides*	*S. richetia* Symington	SRICH
		S. multiflora (Burck) Symington	SMULT
		S. hopeifolia (Heim) Symington	SHOPE
		S. maxima (King) Symington	SMAXI
		***S. faguetiana* Heim**	SFAGU
Anthoshorea		***S. roxburghii* G.Don**	SROXB
		S. javanica Koord. & Valeton	SJAVA
Brachypterae	*Smithiana**	***S. smithiana* Symington**	SSMIT
	Brachypterae	*S. selanica* Blume	SSELA
		S. parvistipulata Heim	SPARV
		S. johorensis Foxw.	SJOHO
		S. scaberrima Burck	SSCAB
		S. kunstleri King	SKUNS
Pachycarpae		*S. pilosa* Ashton	SPILO
		S. splendida (De Vriese) Ashton	SSPLE
		S. amplexicaulis Ashton	SAMPL
		***S. macrophylla* (De Vriese) Ashton**	SMACR
		S. beccariana Burck	SBECC
Mutica	*Auriculatae*	*S. macroptera* Dyer	SMACT
	Mutica	***S. singkawang* Burck**	SSING
*Ovalis**		***S. ovalis* Blume**	SOVAL
Total number of *Shorea* species			27
Total number of included taxa			42

* monotypic section or subsection.

Names in bold font are the type species of each section or subsection.

5.4 Data analysis

Cladistic analyses were performed under maximum parsimony criteria with all characters treated as unordered (Fitch, 1971) using PAUP* 4.0b4a (Swofford, 1998). The most parsimonious trees (MPTs) were found by using a heuristic search, with 1000 random addition sequence replicates to search across multiple islands of trees (Maddison, 1991). These settings were used for all final tree searches. MAXTREES was set to 500 and not increased. The TBR (Tree Bisection Reconnection) branch-swapping algorithm was employed with the "steepest descent" option off. ACCTRAN (Accelerated Transformation) character optimisation was used with the MULPARS (Multiple Parsimonious Trees) option on, and branches of zero length were collapsed. Ten equally parsimonious trees were held following each replicate (Swofford, 1998).

Statistical measures of the Consistency Index (CI) and Homoplasy Index (HI, Kluge and Farris, 1994), and the Retention Index and Rescaled Consistency Index (RC, Farris, 1989) were also calculated. Clade support was estimated by performing 100 bootstrap replicates (Felsenstein, 1985) and using the 50% majority-rule MPTs as input trees but with the MULPARS option off. Trees were rooted using the pre-defined outgroup.

5.5 Results

5.5.1 Outline of the characters used

The total number of characters in the combined data matrix is 1884, but only 407 of these are parsimony informative. Of these, 40 are from the morphological dataset and 367 from the molecular data.

Close examination of the morphological data matrices shows that the homoplasy of the morphological characters is lessened when used in combination with the molecular data than when the morphological data are used alone. The large number of molecular characters makes it difficult to examine the homoplasy of each one.

5.5.2 Topological features

A heuristic unconstrained search using equally weighted characters yielded 3 MPTs with a length of 2214 steps. These trees have a CI of 0.49, HI of 0.51, RI of 0.48 and RC of 0.24. Tree 1 was selected for discussion (Figure 5.1).

The topology obtained contains two ingroup clades labelled A and B, although the relationship of these groups to each other is unresolved. The first clade (A or Group 3 *Shorea*, synapomorphies 33, bootstrap <50%) primarily contains species from the Red Meranti timber grouping . This clade contains taxa from *Shorea* sections *Mutica*, *Pachycarpae*, *Brachypterae* and *Ovalis*. Section *Pachycarpae* appears to be monophyletic (synapomorphies 40, bootstrap 73%), albeit with the exception of *S. pilosa*. The sister group to clade J is a pairing of *S. ovalis* (section *Ovalis*) with *S. smithiana* (section *Brachypterae*). *Shorea parvistipulata* (section *Brachypterae*) then forms the sister taxon to the remainder of species in Group 3.

The second major ingroup clade, which is unresolved with respect to Group 3 *Shorea*, is clade B (synapomorphies 69, bootstrap <50%). This clade consists of two further subgroups, Group 1 *Shorea* (C) and a clade uniting *Shorea* section *Richetioides* (F) with *Hopea* (D).

Clade C (synapomorphies 17, bootstrap <50%) contains taxa primarily from *Shorea* section *Shorea*, and is referred to as Group 1 *Shorea*. Section *Shorea* could be regarded as monophyletic in this analysis, with the inclusion of *S. scaberrima*, *S. javanica*, *S. johorensis* and *S. isoptera*. This clade, to some extent, represents the Balau timber group proposed by Symington (1943).

Clade D (synapomorphies 24, bootstrap <50%) consists of species from both *Hopea* and *Shorea*, with *S.* section *Richetioides* subsection *Richetioides* (Group 2 *Shorea*) being the sister group of the *Hopea* clade. The combined data set thus supports the monophyly of *S.* section *Richetioides* subsection *Richetioides* (F, synapomorphies 25, bootstrap 95%).

Two taxa that fell within Group 2 *Shorea* in the ITS topology and formed a pair in the combined *trn*L-F/ITS analysis, *Hopea celebica* and *Shorea selanica*, are now placed

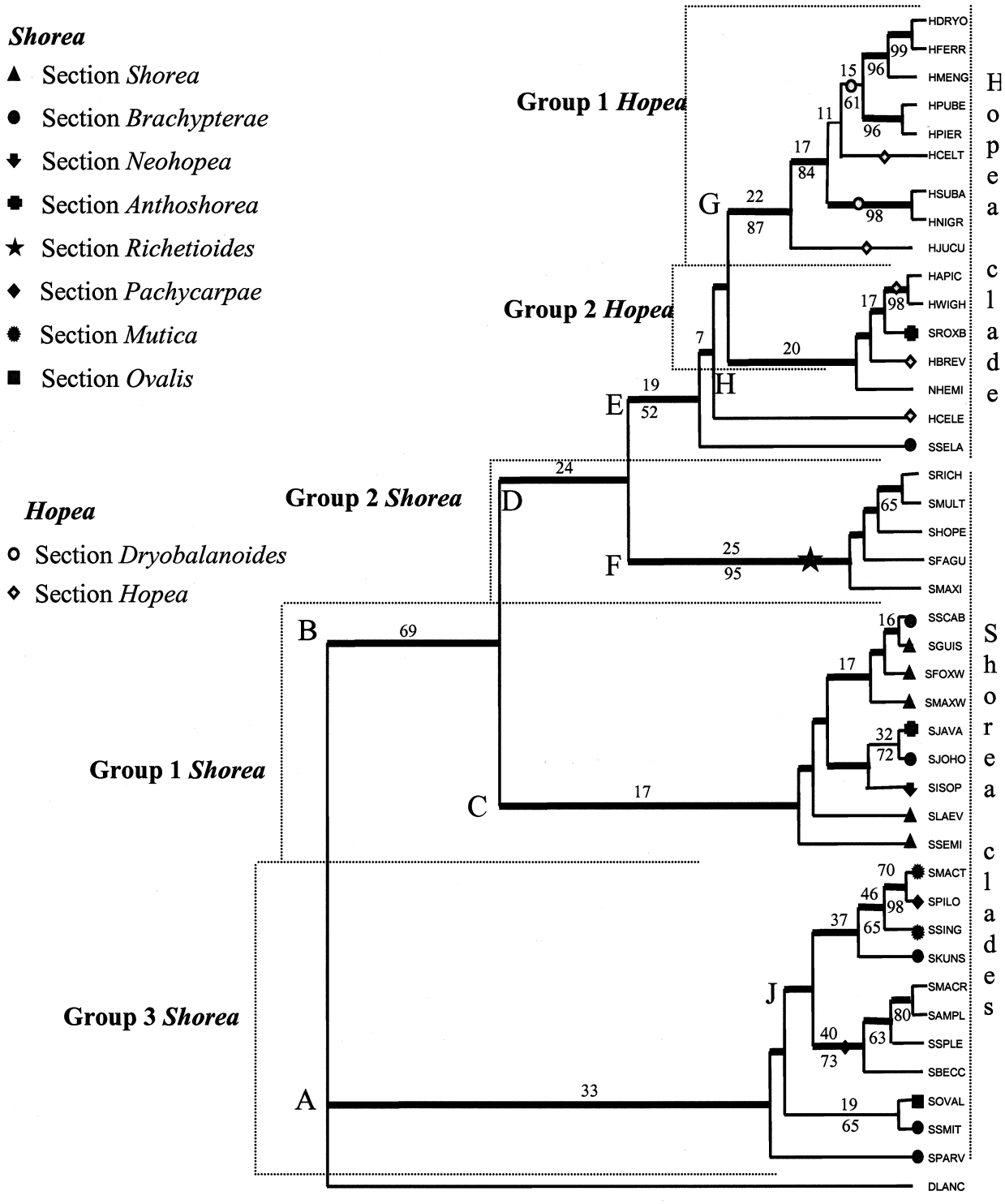


Figure 5.1 One of the most parsimonious trees obtained from cladistic analysis of the combined morphological and molecular data sets for 42 Dipterocarpaceae taxa. Numbers above the branches are branch lengths, and bootstrap values of 50% and greater are shown below. Thicker branches are those that appear in the strict consensus of all shortest trees.

within the *Hopea* clade (E). *Hopea celebica* forms the sister taxon to a clade consisting of *Neobalanocarpus heimii*, *Shorea roxburghii* and all the other *Hopea* species included in the analysis. *Shorea selanica* is then the sister taxon to all the other species within the *Hopea* clade.

As with the *Shorea* clade, the arrangement of the *Hopea* clade (E, synapomorphies 19, bootstrap 52%) remains very similar to that shown in the separate analyses of morphological and molecular data. Two major groups (G and H) form part of this clade, but the content of these is not in accordance with the accepted classification. Both *Hopea* section *Hopea* and section *Dryobalanoides* appear to be non-monophyletic.

The type of leaf nervation, used by Ashton (1982) to separate the two *Hopea* sections *Dryobalanoides* and *Hopea*, does not provide an informative signal in the topology obtained from the combined dataset since this character does not unambiguously support any of the branches within the *Hopea* clade. Examining the apomorphic changes in the cladogram suggests that the type of leaf nervation is an autapomorphic character that defines *Hopea celtidifolia*. Even though this character may be of use to diagnose the sections as currently defined, results from this study show that it is confounded with considerable homoplasy and may provide limited information when inferring phylogeny.

The first group within the *Hopea* clade (G, synapomorphies 22, bootstrap 87%) consists of all the members of section *Dryobalanoides* included in the analysis in combination with two species from section *Hopea* (*H. celtidifolia and H. jucunda*). *Hopea jucunda* forms the sister group to a clade containing *H. celtidifolia* and the members of section *Dryobalanoides*.

The second group in the *Hopea* clade (H, synapomorphies 20, bootstrap <50%) contains most of the remaining members of section *Hopea* with *Shorea roxburghii* and *Neobalanocarpus heimii*. The placement of *S. roxburghii* (*Shorea* section *Anthoshorea*) within this clade is interesting result, since this species and *S. exelliptica* formed a species pair that was the sister taxon to the *Hopea* clade in the analysis of the combined *trn*L-F/ITS regions. Homoplasious changes in the form of

reversals may have occurred in *S. roxburghii*. The placement of *Neobalanocarpus heimii*, the putative sister taxon to *Hopea*, within Clade H with members of section *Hopea* may suggest that *N. heimii* is actually part of *Hopea* with a closer relationship to section *Hopea* than to section *Dryobalanoides*.

In summary, the phylogenetic topology obtained from the analysis of combined morphological and molecular data indicates that neither *Shorea* nor *Hopea* are monophyletic as currently circumscribed. *Shorea* is split into three major groups which are not sister taxa and *Hopea* is monophyletic only with the inclusion of *S. roxburghii* and *Neobalanocarpus heimii*. With regard to the infra-generic divisions of *Hopea* and *Shorea*, only *Shorea* section *Richetioides* subsection *Richetioides* appears to be monophyletic.

5.6 Discussion

5.6.1 The information content of the combined data sets

The topology resulting from this combined analysis of molecular and morphological data is to some extent similar to those from the molecular analyses. This may be due to the molecular characters simply outweighing those from the morphology, as the number of nucleotide characters is so much greater.

By combining the data sets, we can examine which characters define a grouping in the topologies obtained. It is important in systematics studies to determine which characters are definitive for particular groups, so that a clear pattern is shown by the character changes from the base towards the terminal nodes. Examination of the character changes over the topologies obtained in this combined analysis suggests that both morphological and molecular characters provide useful phylogenetic signal at various taxonomic levels. The comparative development of fruit wings, thought to be the single morphological character which distinguishes *Hopea* and *Shorea*, does not appear to be a consistent diagnostic character for the two genera. An examination of the evolutionary changes within this character was made earlier in the morphological chapter (Chapter 3). Briefly, this character does not appear to provide a clear phylogenetic signal to differentiate *Hopea* and *Shorea*. Other morphological characters that seemed to be capable of distinguishing between the two genera are the presence of an indumentum on the nut, presence of a fruit pedicel, length of the ovary,

shape of the anther appendages, flower size and the type of Inflorescence Unit. The resolution provided by the combined data was also somewhat better than that shown by the separate data sets, recognising that inclusion of fewer taxa may have influenced the result. The combined analysis resolved the placement of some potentially monophyletic groups and the placement of taxa that may have undergone recombination. Some examples of improved resolution are the clades containing *Shorea* section *Richetioides* (with 95% bootstrap) and *Hopea* section *Dryobalanoides* including *H. celtidifolia* (with 87% bootstrap). The second example of improved resolution is the position of two species previously showing uncertain placement, *Hopea celebica* and *Shorea selanica*. Before combining the two data sets these taxa were placed in different lineages, perhaps due to possession of recombinant sequences. Merging the data sets provides more characters to resolve their "actual" positions within the topology (Figure 5.1).

On the other hand, the placement of *Shorea roxburghii* (Section *Anthoshorea*) within the *Hopea* clade is interesting, since the earlier analyses (discussed in Chapter 3 and 4) did not indicate the inclusion of this species in the *Hopea* clade. However, the monophyly of *Shorea* section *Anthoshorea* and *Hopea* was supported by an analysis of *rbc*L data (Dayanandan *et al.*, 1999). *Hopea* and *Shorea* section *Anthoshorea* have similarities in floral morphology, with both having an urceolate corolla and an acicular anther connective appendage.

5.6.2 Incongruence between molecular and morphological analyses

Amongst several different methods available to combine two data sets (Chippindale and Weins, 1994; Farris *et al.*, 1994; Mason-Gamer and Kellogg, 1996), the tree-based comparison outlined by Mason-Gamer and Kellogg (1996) suggests that trees be examined for conflict involving nodes with bootstrap values of over 70%.

> "Weakly supported nodes only ambiguously represent patterns within individual data sets, and therefore conflict among data sets cannot be inferred from comparisons involving weak nodes" (Mason-Gamer and Kellogg, 1996).

In the present study, independent analyses of the morphological and molecular data sets produced different topological arrangements of *Hopea* and *Shorea*, and hence

suggested different phylogenetic conclusions. In order to be able to examine the incongruence between the topologies, two separate analyses were performed. The first estimated the incongruence between data sets using the Partition Homogeneity test (Farris *et al.*, 1994) and the second examined any conflict occurring among the clades from the morphological and molecular topologies.

The partition homogeneity test indicates that the morphological and molecular data sets are not congruent (P=0.01). According to this result, the data sets should not be combined. Bull *et al.* (1993) and Weiblen (2000) suggested that conflict between data sets can occur from systematic error, rate heterogeneity, or because the data sets do not share the same phylogenetic history (Weiblen, 2000). Systematic or taxonomic error is unlikely to explain the incongruence among the lineages recovered from the two data sets, mainly because the taxa used appear to have been correctly identified. Instead, the considerable number of incongruent clades suggests that the two data sets may have different phylogenetic histories. In order to identify these conflicts, close examinations are made of the clade support measures in each topology following Mason-Gamer and Kellogg (1996).

Two separate cladistic analyses of the morphological and molecular data sets were performed, using a subset of the taxa included in the combined analysis. These analyses revealed that the majority of the clades from both topologies have different species arrangements (Figure 5.2). There are no clear groups in the morphological topology (Figure 5.2, left-hand cladogram) that correspond to the groupings from the combined analysis, but there are clades in common between the topology from combined data and the molecular topology (Figure 5.2, right-hand cladogram). *Hopea* is monophyletic in the molecular topology with the inclusion of *Neobalanocarpus heimii*, *Shorea selanica and S. roxburghii*. By contrast, *Hopea* is non-monophyletic in the morphological topology since four species from section *Hopea*—*H. apiculata*, *H. jucunda*, *H. brevipetiolaris* and *H. wightiana* (indicated by underlined taxon names in Figure 5.2, left-hand cladogram) are placed within a clade of *Shorea* species. The possibility of parallel evolution of morphological characters in these four *Hopea* species is therefore not supported by the molecular analysis, since the molecular topology included these taxa within *Hopea*.

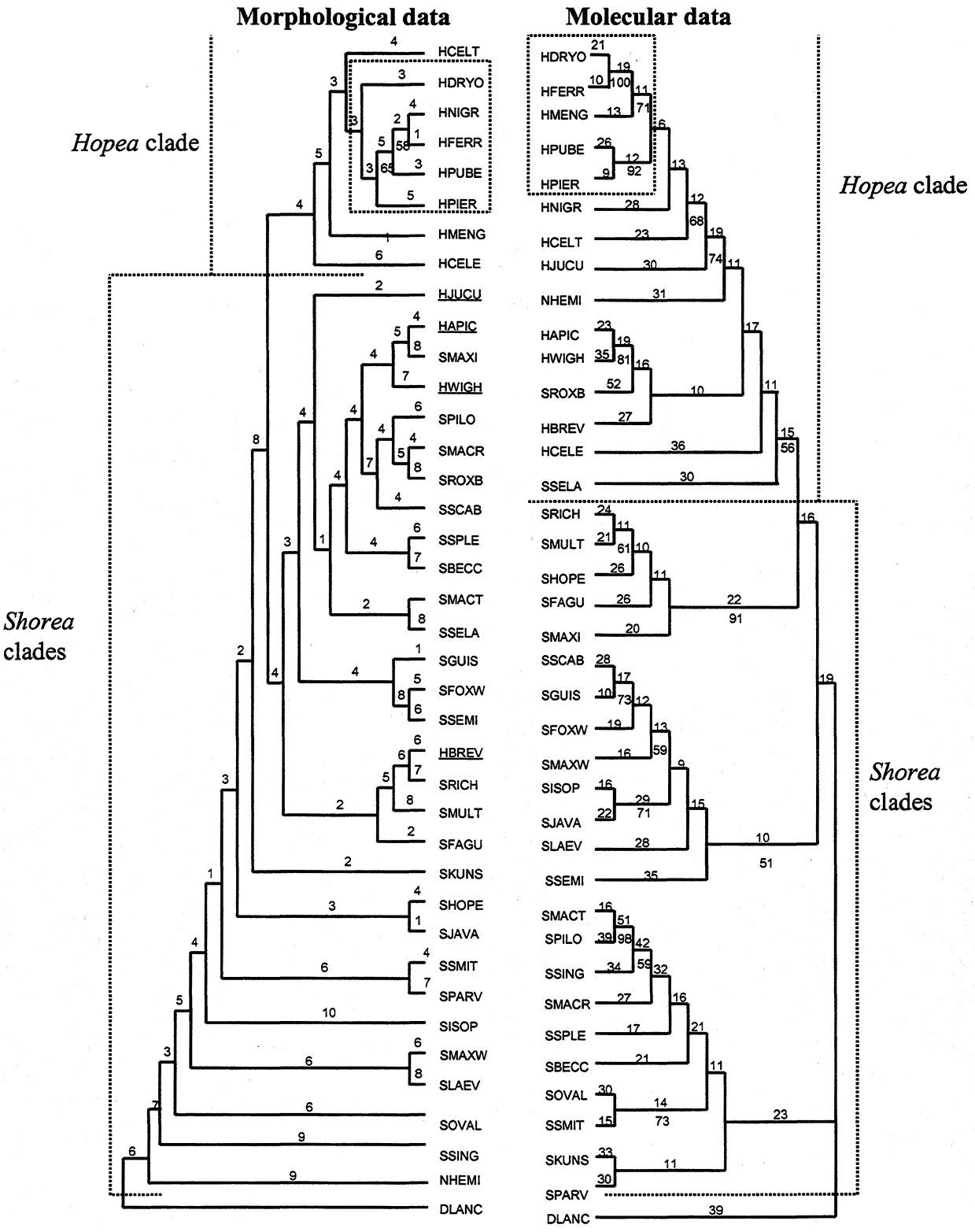


Figure 5.2 Comparison between topologies obtained from cladistic analysis of morphological data (left-hand cladogram) and molecular data (right-hand cladogram) for selected Dipterocarpaceae taxa. Numbers above the branches are branch lengths and bootstrap values of 50% and greater are given below. Congruent clades are indicated by boxes.

Almost all the clades in both topologies are weakly supported by the bootstrap (Figure 5.2), so clade conflicts cannot be inferred from the two topologies (Mason-Gamer and Kellogg, 1996). Even though examination of the conflict involving lineages that have similar clade arrangements (marked by dotted boxes), the bootstrap support between the two topologies are different strengths (B <50% in the morphological and B 92% in the molecular data set). Hence, no well-supported conflict occurs between the two data sets and, following the arguments of Mason-Gamer and Kellogg (1996), the two data sets were therefore justifiably combined.

5.6.3 Phylogenetic relationships of Hopea and Shorea

The close relationship between the variable genus *Shorea* and *Dryobalanops lanceolata*, which was clearly indicated in the separate morphological and molecular analyses, is confirmed by this combined analysis. Even though *Shorea* and *D. lanceolata* are very distinct morphologically, this analysis showed that the two morphological characters that distinguish *D. lanceolata* from the ingroup are homoplasious. The relationship of *Hopea*, *Shorea* and *Dryobalanops lanceolata* is readily discerned from the analysis.

Neither *Hopea* nor *Shorea* appear to be monophyletic, with *Hopea* mostly nested within *Shorea*. In the topology obtained from the combined analysis, *Hopea sens. lat.* is the sister taxon to Group 2 *Shorea*. The clade containing both *Hopea* and Group 2 *Shorea* is then most closely related to Group 1 *Shorea*. Thus, this may suggest that *Hopea* originated from *Shorea*, and the character states possessed by *Hopea* are of relatively more recent origin than those of *Shorea*.

The groupings within the *Hopea* clade do not accord with the currently accepted infra-generic divisions. Section *Dryobalanoides* (Figure 5.1, clade G) forms a clade with strong support from the bootstrap, but this group is monophyletic only with the inclusion of two species from section *Hopea*. Most members of section *Dryobalanoides* have a Malesian distribution. Character state changes shared between this section and two species from section *Hopea* suggest diversification events among these taxa. The sister group to clade G contains three other members of section *Hopea* grouped with *Shorea roxburghii*, although this receives only weak support from the bootstrap.

Hopea section *Hopea* is a variable group. Two of its members, *H. celtidifolia* and *H. jucunda*, are more closely related to section *Dryobalanoides* than they are to other members of their own section. Exclusion of these two species from Group 2 *Hopea*, which is made up largely of members of section *Hopea*, suggests that they may have undergone separate adaptive radiations within their confined distributions and hence had different phylogenetic histories from the rest of section *Hopea*. *Hopea celtidifolia* is endemic to New Guinea Island and *H. jucunda* is endemic to Sri Lanka.

The putative sister taxon to *Hopea*, *Neobalanocarpus heimii*, is nested within Group 2 *Hopea* (consisting mainly of taxa from section *Hopea*) in the topologies obtained from both the molecular and combined analyses. However, the results from a previous phylogenetic analysis of Dipterocarpaceae based on molecular data suggested that there was a close relationship between *N. heimii* and *Hopea* section *Dryobalanoides* subsection *Dryobalanoides* (Tsumura *et al.*, 1996). *Hopea* and *N. heimii* both possess medium-sized vessels and storied rays (Parameswaran and Gotwald, 1979 in Maury-Lechon and Curtet, 1998), and share similarities in their anthocyanin development (Bate-Smith and Whitmore, 1959) and bark morphology (Whitmore, 1962). However, *Neobalanocarpus* is endemic to Peninsular Thailand and has a distinct "semi-broad" anther appendage that is not present in any *Hopea* species. Clearly, it is not sufficient to separate *Neobalanocarpus* from *Hopea* solely on the basis of these two morphological characters, since they have been shown to be homoplasious. However, the inclusion of *Neobalanocarpus heimii* within section *Hopea* in at least some of the phylogenies indicates its close relationship to this section. *Hopea* may share a common ancestor with the monotypic *Neobalanocarpus*. The restriction of the latter genus to Peninsular Thailand suggests that it did not continue to diversify into other parts of the geographic region, such as Malesia. This may indicate that *Neobalanocarpus heimii* is a taxon with limited potential for diversification.

In the combined analysis, *Shorea* contains three groups that largely do not correspond to the current infra-generic classification. These groups mostly contain the same arrangement of species seen in the separate analyses. As currently circumscribed, *Shorea* consists of 10 sections and eight of these are included in this analysis. The

excluded sections are section *Doona* and the monotypic section *Pentacme*. Of the eight sections included, only some appear to be monophyletic groups.

Group 1 *Shorea* consists mainly of taxa from section *Shorea* (Balau group), which is in accordance with the results of the analyses of morphological data (Chapter 3) and the molecular analyses (Chapter 4). The placement of Group 1 *Shorea* (Figure 5.1, clade C) as the sister group to a clade that unites Group 2 *Shorea* with *Hopea* (Figure 5.1, clade D) may suggest that the members of section *Shorea* possess relatively plesiomorphic character states. Section *Shorea* (C) is monophyletic only with the inclusion of a species each from sections *Brachypterae*, *Anthoshorea* and *Neohopea*. The placement of the monotypic *Shorea* section *Neohopea* (i.e. *S. isoptera*) is therefore resolved. This supports the hypothesis that this taxon is part of *Shorea* (Symington, 1943; Meijer and Wood, 1964; Ashton, 1982), rather than Heim's (1891) suggestion that it be recognised as a separate genus *Isoptera*.

Group 2 *Shorea* consists of all the taxa from section *Richetioides* subsection *Richetioides* included in the analysis. This apparently monophyletic section is also known as "Yellow Meranti" or "Damar Hitam" on the basis of wood and bark anatomy (Symington, 1943, Whitmore, 1962). It contains species with sub-equal calyx lobes, a character that has prompted some taxonomists to propose the group be recognised as a separate genus (Heim, 1891; Meijer and Wood, 1964; Maury, 1978 in Maury-Lechon and Curtet, 1998). The placement of *Shorea* section *Richetioides* subsection *Richetioides* as the sister taxon to the *Hopea* clade suggests that this group is closely related to *Hopea*. The monophyly of section *Richetioides* subsection *Richetioides* also suggests that unique evolutionary events occurred in the group.

Group 3 *Shorea* is a variable group containing taxa from sections *Mutica*, *Pachycarpae*, *Brachypterae* and *Ovalis*. This group is nearly analogous to the previously recognised genus *Rubroshorea* (Meijer and Wood, 1964) and to the Red Meranti timber grouping (Symington, 1943). The phylogenetic position of this group in relation to the rest of the ingroup taxa is unresolved. All the sections included in Group 3 *Shorea* appear to be non-monophyletic, with the exception of the monotypic section *Ovalis*. Section *Pachycarpae*, which is endemic to the island of Borneo, is monophyletic only when *S. pilosa* is excluded. Section *Pachycarpae* is assumed to

have undergone rapid diversification. The reduction of fruit size which defines this section may be a subject for speculation, since analyses of morphological data have suggested that this character is homoplasious.

The phylogenetic positions of three anomalous taxa—*Shorea roxburghii* (*S.* section *Anthoshorea*), *Shorea selanica* (*S.* section *Shorea*) and *Hopea celebica* (*H.* section *Hopea*)—are of interest. An explanation of the phylogenetic positions of *H. celebica* and *S. selanica* has been incorporated into a previous chapter (Chapter 4, section 4.8.2). These two species are placed within the *Hopea* clade.

The results of this combined analysis suggest that there is a close relationship between *Shorea roxburghii* (section *Anthoshorea*) and *Hopea*. According to Maury-Lechon (1979; in Maury-Lechon and Curtet, 1998), *S. roxburghii* is a highly variable species that contains a whole suite of the variation occurring in other species (Maury-Lechon, 1979). To some degree, *Shorea* section *Anthoshorea* resembles *Cotylelobium*, *Neobalanocarpus heimii* and sections *Doona*, *Dryobalanops* and *Pentacme* based on similarities in characters from the embryo, seedling and pollen surface (Maury-Lechon and Curtet, 1998). Section *Anthoshorea* has also been suggested by Brandis (1895) and Ashton (1982) to be an intermediate taxon between *Hopea* and *Shorea,* and was recognised as the genus *Parahopea* by Heim (1892). This may explain why the members of section *Anthoshorea* included in this analysis are distributed in many disparate clades in the topologies obtained from analyses of the morphological and molecular datasets (Chapter 3 and 4).

Section *Anthoshorea* is recognised as Meranti Pa'ang (White Meranti) by Symington (1943). However results from this present study do not confirm the integrity of the timber grouping for this section since they may have been a polyphyletic section.

5.64 Taxonomic implications

Results from the combined data set of morphological and molecular characters suggest that *Hopea* and *Shorea* cannot be separated into two distinct taxa and that *Hopea* is nested within the variable *Shorea* group.

However, if the two genera are still to be recognised, only a few of the existing infra-generic groupings can be considered "natural". Within *Hopea*, only section *Dryobalanoides* can be considered monophyletic, albeit with the inclusion of *Shorea celtidifolia*. Thus, this section can be maintained. Reduction of *Neobalanocarpus heimii* into *Hopea* section *Hopea* is also proposed. However, the taxonomic status of section *Hopea* cannot be deduced since the section is obviously not monophyletic. Further detailed studies incorporating more species may be required to clarify the taxonomic status of this section. Re-establishment of the classification of Meijer and Wood (1964), which recognised *Hopea* as a single genus, can also be considered if the monophyly of section *Hopea* can be established.

The groupings within *Shorea* are more complex. When the phylogeny obtained from the analysis of the combined dataset is used to assess the previously suggested taxonomic groupings, it appears there is a need to recognise the Balau group s*ensu* Symington (1943) or subgenus *Eushorea sensu* Meijer and Wood (1964), both of which consist of section *Shorea* and some species of section *Anthoshorea sensu* Ashton (1982). Based on the present results, it would be valid to recognise the Meranti Damar Hitam timber grouping *sensu* Symington (1943), which is equivalent to subgenus *Richetia sensu* Meijer and Wood (1964) and section *Richetioides sensu* Ashton (1982). In addition, the Red Meranti group *sensu* Symington (1943) or subgenus *Rubroshorea sensu* Meijer and Wood (1964) could be re-established. The three groups of *Shorea* listed above can be assigned either infra-generic or generic rank.

5.7 Conclusions

Analysis of the combined morphological and molecular data yielded similar species trees to those obtained from analysis of the molecular data alone, with the exception of the placement of three anomalous taxa (*Hopea celebica*, *Shorea selanica* and *S. roxburghii*). Despite this exception, the combined data set provided strong evidence for the broad non-monophyly of *Shorea* and the potential monophyly of *Hopea* following some recircumscription. The only largely natural grouping within *Hopea* which accords with the existing classifications is section *Dryobalanoides*. The inclusion of *Neobalanocarpus heimii* within *Hopea* may also be required. However, the phylogenetic position of section *Hopea* remains complex.

With regard to the groupings within *Shorea*, the results of this analysis of the combined data sets suggest that the classifications by Meijer and Wood (1964) and Symington (1943) be recognised. Two exceptions to this are that Meranti Pa'ang *sensu* Symington and subgenus *Anthoshorea sensu* Meijer and Wood should not be recognised. Both classification systems used mainly timber characters, with important benefits for practical identification of the large timber species which make up the genus. However, further analyses incorporating *Shorea* sections *Doona* and *Pentacme* are required to provide better resolution of the relationships at the infra-generic level of *Shorea* as well as at the generic level.

CHAPTER 6
GENERAL DISCUSSION

Outline

This chapter discusses the overall results from the morphological, molecular and combined analyses in order to draw some general conclusions about phylogenetic patterns. It discusses the phylogenetic relationships among the taxa used and the implications for classification, with a particular focus on *Hopea* and *Shorea*.

6.1 Phylogenetic relationships among the outgroup and the ingroup taxa

Six genera of Dipterocarpaceae in addition to *Hopea* and *Shorea* were included in the analyses as putative outgroups. These were *Dipterocarpus*, *Anisoptera*, *Cotylelobium*, *Dryobalanops*, *Parashorea* and *Neobalanocarpus*. All the genera included belong to the Asian subfamily Dipterocarpoideae, which was further divided by Ashton (1982) into two tribes distinguished by their base chromosome numbers. The two tribes are Dipterocarpeae, which has a base number of x=11 (*Vateria*, *Vateriopsis*, *Stemonoporus*, *Vatica*, *Cotylelobium*, *Upuna*, *Anisoptera*, *Dipterocarpus*), and Shoreae, with a base number of x= 7 (*Dryobalanops*, *Parashorea*, *Hopea*, *Neobalanocarpus*, *Shorea*).

Several analyses were performed to identify the relationship of these hypothetical outgroups to the ingroup and to examine the effect of including these taxa on the resulting topology. The results from the analysis of morphological data (Figure 3.41) and from analysis of the *trn*L-F sequences (Figure 4.3) using *Dipterocarpus* as an outgroup are in accordance with Meijer (1974), Tsumura *et al.* (1996), Kajita *et al.* (1998) and Dayanandan *et al.* (1999). These authors considered *Dipterocarpus* to be the sister taxon to the remainder of sub-family Dipterocarpoideae. The results of the present study are consistent with the hypothesis that *Dipterocarpus* retains primitive states of particular characters or unspecialised morphological features, exemplified

particularly by its floral characters. The plesiomorphic nature of its sequences may suggest that *Dipterocarpus* diverged relatively early from the rest of sub-family Dipterocarpoideae.

Examination of the phylogenetic position of another putative outgroup, *Anisoptera* and *Cotylelobium*, showed that these two genera are relatively distantly related to *Hopea* and *Shorea*. Tsumura *et al.* (1996), Kajita *et al.* (1998) and Dayanandan *et al.* (1999) suggested that taxa in *Anisoptera* and *Cotylelobium* form a group within tribe Dipterocarpeae, and indicated they have a relatively distant relationship to *Hopea* and *Shorea*. This finding is supported in the present study by the results of the analysis of the ITS data set (Figure 4.5). Hence, the results of this analysis are congruent with the tribal division by Ashton (1982).

The present study has provided information that may help to clarify the phylogenetic placement of *Dryobalanops*, which to date has been difficult to resolve (Dayanandan *et al.*, 1999). Analyses performed in order to resolve the phylogenetic position of several putative outgroup taxa (Chapter 4) have confirmed the placement of *Dryobalanops* as the sister taxon of Group 3 *Shorea*, a group largely made up of members of the Red Meranti timer grouping. Three earlier studies (Tsumura *et al.*, 1996; Kajita *et al.*, 1998; Dayanandan *et al.*, 1999) had indicated a close relationship between *Dryobalanops*, as a member of Tribe Shoreae, with the ingroup (*Hopea* and *Shorea*) but had failed to clarify its phylogenetic placement. As highlighted by Ashton (1982), these earlier studies also suggest the possibility of intermediate morphological characters between *Dryobalanops* and Tribe Dipterocarpeae. Ashton (1982) and Tsumura *et al.* (1996) suggested that the wood anatomy exhibited some intermediate characters, while Gotwald and Parameswaran (1966 in Dayanandan *et al.*, 1999) suggested the presence of solitary vessels to be an intermediate character. However, the analysis of the combined data set including both morphological and molecular data did not indicate any apomorphic changes in morphological characters that defined *Dryobalanops* (Figure 5.1). The molecular analyses have consistently provided evidence that this genus is the sister taxon to the Group 3 *Shorea* species.

The putative sister to *Hopea*, *Neobalanocarpus heimii*, is nested within the *Hopea* clade in this present study. This placement is in agreement with the findings of

previous studies on the phylogeny of Dipterocarpaceae (Tsumura *et al.*, 1996; Kajita *et al.*,1998). Analyses performed in this present study have consistently placed *Neobalanocarpus* within *Hopea* section *Hopea*, indicating a closer relationship to section *Hopea* than to section *Dryobalanoides* as was also suggested previously by Tsumura *et al.* (1996). The position of *Neobalanocarpus* within section *Hopea* suggests that this genus may have arisen from a common ancestor with that section.. There are no synapomorphic changes in morphological characters on the branch that unites this genus with section *Hopea*. However, there is a substitutional change in one base pair from the *trn*L intron and in nine base pairs from the ITS regions. The close relationship of *Neobalanocarpus heimii* and *Hopea* was previously suggested by Parameswaran and Gotwald (1979 in Dayanandan *et al.*, 1999) on the basis of the presence of medium-sized vessels and storied rays and the absence of silica. Moreover, anthocyanin development and bark morphology have led Whitmore (1962) to group this genus with *Hopea*. It is therefore suggested that *Neobalanocarpus heimii* may be part of the genus *Hopea*.

The putative sister taxon to *Shorea*, *Parashorea*, appears to be part of *Shorea*. Results from separate molecular and morphological analyses have consistently shown that this genus is not the sister taxon to *Shorea,* which is not in accord with the results of previous studies (Tsumura *et al.*, 1996; Kajita *et al.*, 1998). Cladistic analyses of data derived from morphology and the ITS region show the genus is nested within a group of *Shorea* species (Figures 3.38, 3.39, 3.41 and 4.5. *Parashorea* is characterised by globose, verrucose fruit with subequal and unequal aliform calyx wings, and it has plicate venation that is similar to *Shorea*. Ashton (1982) and Maury (1978 in Ashton, 1982) considered that differences in embryo and seedling characteristics could be used to distinguish *Parashorea* from *Shorea*. The analysis of morphological data in this present study, however, did not include seedling and embryo characters and therefore could not examine the phylogenetic signal yielded by these features. Nevertheless, the six synapomorphic changes in morphological characters uniting *P. malaanonan*, *S. palembanica* and *S. virescens* (Figure 3.41) are homoplasious. Hence, parallel or reverse evolution of these characters is likely to have occurred early in the evolutionary history of these taxa.

6.2 Phylogenetic relationships of *Hopea* and *Shorea*

Taxonomically, *Hopea* and *Shorea* have long been acknowledged as being problematic taxa. A large number of similarities and the continuity of morphological variation in the genera have led to several different classifications being proposed, including detailed infra-generic divisions. The morphological characters that were used by earlier taxonomists to distinguish *Hopea* and *Shorea* have been shown to be homoplasious, and they thus cannot provide a clear signal to enable confident phylogenetic inference of the relationship between the two genera. Cladistic analyses of morphological characters yielded topologies with unclear and unresolved groupings within *Hopea* and *Shorea*. The comparative development of fruit sepals that was assumed to be an apomorphic character to distinguish the two genera was shown to be homoplasious due to reversals and parallelism. The morphological characters that were used by previous taxonomists are useful for identification purposes but not to deduce the phylogenetic relationship between *Hopea* and *Shorea*.

The phylogenetic position of the two genera was thus unclear until several analyses of molecular data were performed. Better resolution gained from these analyses resulted from there being fewer homoplasious changes in the molecular characters than shown by the morphological data. Other recent studies using molecular data have also produced results that generally agree with this study in regard to the phylogenetic relationship between *Hopea* and *Shorea*. *Hopea* forms a mostly monophyletic genus that is nested within the broadly non-monophyletic *Shorea* group.

An important result from this study is the identification of a problem with two anomalous taxa, *Hopea celebica* and *Shorea selanica*. These taxa are endemic to areas on the eastern side of the Wallace's line, with *Shorea selanica* being endemic to the Moluccas and *Hopea celebica* to Sulawesi. The effects of reproductive isolation may have been more extreme in these populations and interbreeding within their narrow distribution is thus likely to occur. The phylogenetic placement of these two taxa in the topology obtained from the analysis of combined data, with both forming separate lineages within the *Hopea* clade, suggests that each has some autapomorphic changes in the morphological and molecular characters. Reliance on morphological characters alone, however, may lead to a different phylogenetic inference, as the two may be undergoing parallel evolution.

Another finding relevant to any explanation of the biogeographic patterns observed in the dipterocarp taxa is the frequent monophyletic pairing of *Hopea jucunda* with *H. cordifolia*, both of which are endemic to Sri Lanka. Interestingly, *H. brevipetiolaris* is also endemic to Sri Lanka but does not group with *H. cordifolia* and *H. jucunda*. Instead it forms a clade with the South Indian *H. wightiana* and *Shorea roxburghii*. The phylogenetic patterns of the taxa that are distributed outside Malesia suggest that local adaptive radiation may have occurred. Hence, the DNA sequences and morphological characters of these taxa may be unique.

6.3 Classification of *Hopea* and *Shorea*

Phylogenetic studies are always based on the principle of descent and the concept of monophyly. Cladistic results are focused on the question of monophyly or non-monophyly of groups of taxa. The concept of monophyly has now become central to evolutionary biology (Gordon, 1999) and when referring to this concept, one must return to the theory of common descent. Thus, results gained from the cladistic approach are based on the idea of a common origin of related groups of organisms

Nevertheless, if *Hopea* and *Shorea* are considered to be separate genera, *Shorea* is clearly seen as a non-monophyletic group, since clades arising from the putative common ancestor of all *Shorea* species also contain species of *Hopea*. To what extent paraphyletic groups should be recognised as separate entities is a question that systematists debate passionately (Freudenstein, 1998; Gordon, 1999).

The Linnaean hierarchical system has served as an important taxonomic methodology in classifying and naming organisms for almost 250 years (De Queiroz, 1997). The system provides a series of ranked taxonomic categories, based on those adopted by Linnaeus (1771), to which taxa (named groups or organisms) are assigned. The classifications of *Hopea* and *Shorea* described in Chapter 2 are all based on the Linnaean system.

Recently, as new disciplines have developed, phylogeneticists have attempted to produce classifications that reflect phylogenetic relationships according to the hypothesis of common descent. It is common for results from a phylogenetic analysis to conflict with classifications based on the Linnaean system (de Queiroz, 1997 ;

Stuessy, 1997). Where a phylogenetic analysis results in two monophyletic groups, they may be recognised at a number of taxonomic levels that correspond with the Linnaean hierarchy. In this case, *Hopea* and *Shorea* could be recognised as two separate genera or united as one genus. A second option is that the two monophyletic groups within the expanded genus could be given a lower taxonomic rank, e.g. infra-generic rank or species rank. In contrast, under the ancestor-descent hypothesis, two monophyletic sister taxa are simply considered to form a clade containing all the extant descendants of their nearest common ancestor, rather than being given a taxonomic rank.

To clarify the position before re-examining the taxonomic status of *Hopea* and *Shorea*, these two taxa should be referred to as "groups". This terminology is important because there is no universal generic concept. Monophyly adopts the principle of descent from a common ancestor and while the monophyly of *Hopea* a less contentious issue, the problem of non-monophyly of *Shorea* is more acute. The concept of "pluralism" is thus adopted in reviewing the problem of the non-monophyly of *Shorea*.

The concept of pluralism was discussed by Horvath (1997), when he realised that the phylogenetic species concept was not concordant with the principle of descent or monophyly. Pluralism allows taxonomists to group organisms using various criteria, depending upon the biology of the organisms being classified or upon the goal of the researchers constructing the classification (Kitcher, 1984; Ereshefsky 1992; Standford, 1995). By contrast, monophyly obliges taxonomists to construct classifications solely based on biology. Paraphyletic lineages do not contain all the descendants of a given ancestor, and polyphyletic lineages contain groups that are not direct descendants of a common ancestor. Thus, both paraphyletic and polyphyletic lineages are the product of human classification, not of an evolutionary process (Horvath, 1997). Therefore, only monophyletic lineages qualify as natural groups under the phylogenetic species concept. It follows that *Hopea* may be a natural group with some minor recircumscription. However, *Shorea* is non-monophyletic and thus not a natural group.

Gordon's (1999) philosophies about monophyly are speculative. Reviewing macro scale evolutionary processes, he surmises that

> "the highest level of evolutionary categories might be interpreted to indicate the apparent monophyly of surviving descendants, but only if one ignores the horizontal genetic transfers, transposons, symbionts, hybrid etc.".

By this argument, he concedes that paraphyly and polyphyly can be a result of evolutionary processes. This reasoning was followed in this study, and thus *Shorea* is recognised as a separate taxonomic group resulting from evolutionary processes just as *Hopea* is.

The next stage is to examine the taxonomic ranks of *Hopea* and *Shorea*. Results of this present study are incongruent with the current infra-generic classification of *Hopea* and *Shorea* (Ashton, 1982), except for Group 2 *Shorea* which consists of *Shorea* section *Richetioides* subsection *Richetioides*. It has been argued by many cladists that recognition of a clade should be dependent on branch support. Only Group 2 *Shorea* is strongly supported as a monophyletic group by bootstrap analysis (95%, Figure 5.1). Hence, other apparently monophyletic taxa cannot be formally recognised. The monophyly of *Hopea* section *Dryobalanoides, including H. celtidifolia* from sect. Hopea, is well supported by the bootstrap (84%, Figure 5.1). However, this group is a subclade within Group 1 *Hopea* (Figure 5.1, clade G), which is also strongly supported by the bootstrap (87%) and contains members of both section *Dryobalanoides* and section *Hopea*. Thus, section *Dryobalanoides* cannot be recognised as a natural group without significant recircumscription of section Hopea.

Another author (Sosef, 1997) has suggested that the "branch support" indicated by a high RI and low HI is sufficient to define a genus. However, Freudenstein (1998) argues that RI and HI do not warrant recognising a particular group as a genus. In even the most stable cladogram, the CI and HI for at least some of the characters at a particular node could be low, depending upon how many times they appear in parallel or are reversed in other parts of the tree.

Following the arguments outlined above, there are three options to assign taxonomic ranks to *Hopea* and *Shorea*. Firstly, they could be combined into a single genus, *Shorea*. The second option is to maintain *Hopea* and *Shorea* as separate genera as in

the traditional classification. The third option is to divide *Hopea* and *Shorea* into five groups. Within *Shorea*, results from the analyses of the molecular and combined data sets have consistently suggested recognition of the timber groupings *sensu* Symington (1943) and Meijer and Wood (1964). Many of the clades obtained, other than Group 2 *Shorea* and *Hopea* section *Dryobalanoides*, are heterogeneous assemblages of taxa.

The first option would be the most conservative approach, as the clade containing both genera is a monophyletic group in almost all the analyses undertaken. Secondly, the results of the analysis of morphological characters were not congruent with the current infrageneric groupings within both genera. The decision to combine *Hopea* and *Shorea* will result in a rather cumbersome genus of more than 300 species. However, it is realised that this present study is limited by the relatively low number of species included. Inclusion of more taxa will provide further evidence for this provisional hypothesis on the phylogenetic relationship between *Hopea* and *Shorea*. Moreover, as more studies are carried out incorporating more traits, the taxonomic status of *Hopea* and *Shorea* will become clearer.

CONCLUDING REMARKS

The results from this study do not provide the definitive answer to the question of whether *Hopea* and *Shorea* are natural groups. The term "natural" could only be used if *Hopea* and *Shorea* had been shown to be monophyletic in the phylogenetic analyses. Although the recognition of non-monophyletic groups is debated, as discussed in Chapter 5, it is accepted that neither genus as currently circumscribed is monophyletic and that the current classification is thus at odds with the evolutionary history of the taxa.

The morphological and molecular analyses produced differing results. According to the morphological study, both genera are non-monophyletic. Results from the molecular study, however, indicated the potential monophyly of *Hopea* (with minor recircumscription) and the non-monophyly of *Shorea*. These results were confirmed by the combined analysis. The overall results indicate that *Hopea* is nested within *Shorea* and therefore neither can be considered monophyletic.

If the concept of monophyly is adopted in assessing the taxonomic status of *Hopea* and *Shorea*, then *Hopea* is actually part of *Shorea* and the two genera should perhaps be combined. However, if paraphyletic genera are recognised as acceptable, then *Hopea* could be maintained as a separate genus and *Shorea* divided into three genera.

Although there are definite indications from the results of the present study that both *Hopea* and *Shorea* are non-monophyletic, the formal taxonomic and nomenclatural changes that entails are not made here. The level of support for some of the findings is rather low and several important taxa could not be included in all analyses. For these reasons, the large amount of circumscription required will not be made until further evidence is available to support these preliminary results.

REFERENCES

APPANAH, S. 1990. Plant-pollinator interactions in Malaysian rain forests. *In* K. S. Bawa and M. Hadley [eds.], Reproductive ecology of tropical forest plants, Man in the Biosphere Series. Volume 7. UNESCO-MAB, Paris, France. Pp. 85-102.

_____. 1993. Mass flowering of dipterocarp forests in the aseasonal tropics. *Journal of Bioscience* 18: 457 – 474.

_____, AND H. T. CHAN. 1981. Thrips: the pollinators of some dipterocarps. *The Malaysian Forester* 44: 234 – 252.

APPLIED BIOSYSTEMS Inc., Foster City, California

ARCHIE, J. W. 1985. Methods for coding variable morphological features numerical taxonomic analysis. *Systematic Zoology*: 326-345.

ASHTON, P. S. 1964*a* *Ecological studies in the mixed Dipterocarp forests of Brunei State*. Oxford Clarendon, Oxford.

ASHTON, P. S. 1964b *Manual of the Dipterocarp trees of Brunei State*. Oxford University Press, London.

ASHTON, P. S. 1977. A contribution of rainforest research to evolutionary theory. *Annals of the Missoury Botanic Garden* 64: 694-705

_____ . 1980. Dipterocarpaceae. *In* M. D. Dassanayake and F. R. Fosberg [eds.]. A revised handbook to the Flora of Ceylon. Volume 1. Amerind Publishing Co. Pvt. Ltd., New Delhi, India. Pp. 364-423.

ASHTON P. S. 1982. Dipterocarpaceae. *Flora Malesiana Series I-Spermatophyta Flowering Plants* 9(2): 237 - 552.

_____, T. J. GIVNISH AND S. APPANAH. 1988. Staggered flowering in the Dipterocarpaceae: new insight into floral induction and the evolution of mast fruiting in the aseasonal tropics. *American Naturalist* 132: 44 – 66.

AUBRÉVILLE, A. 1976. Essai D'interprétation nouvelle de la Distribution des Diptérocarpacées. *Adansonia*, ser. 2, 16 (2): 205-210

BAKER, J. G. 1877. Flora of Mauritius and the Seychelles: A description of the flowering plants and ferns of those islands. L. Reeve & Co., London. U.K.

BAKKER, F. T., Hellbrugge D, Culham A & Gibby M. 1998. Phylogenetic relationships within *Pelargonium* sect. *Peristera* (Geraniaceae) inferred from nrDNA and cpDNA sequence comparisons. *Plant Systematic and Evolution* 211: 273 - 287.

BALDWIN, B. G., J. M. SANDERSON, J. M. PORTER, WOJCIECHOWSKI, C. S. CAMPBELL AND M. J. DONOGHUE. 1995. The ITS region of ribosomal DNA: a valuable source of evidence of Angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247-277.

BARLOW, P. W. 1989. Meristems, metamers and modules, and the development of shoot and root systems, *Journal of the Linnean Society, Botany* 100: 255-279.

BARRETT, M., M. J. DONOGHUE, AND E. SOBER. 1991. Against consensus. *Systematic Zoology* 40: 486-493.

BATE-SMITH, E. C. & T. C. WHITMORE. 1959. Chemistry and taxonomy in Dipterocarpaceae. Nature 184: 759-796.

BAUM, B. R. 1988. A simple procedure for establishing discrete characters from measurement data, applicable to cladisctics. *Taxon* 37: 63-70.

BAYER, R. J., E. SOLTIS D. E. AND P.S. SOLTIS. 1996. Phylogenetic inferences in *Antennaria* (Asteraceae, Gnaphalieae, Cassiniinae) based on sequences from nuclear ribosomal dna internal transcribed spacers (ITS). *American Journal of Botany* 83 (4): 516-527.

BAYER, R. J. AND J.R. STARR. 1998. Tribal phylogeny of the Asteraceae based on two non-coding chloroplastsequences, the trnl intron and *trn*L/*trn*F intergenic spacer. *Annals of the Missouri Botanical Garden* 85 (2): 242-256.

BAYER, R. J. B., C.F. PUTTOCK, AND S.A KELCHNER. 2000. Phylogeny of South African Gnaphalieae (Asteraceae) based on two-coding sequences. *American Journal of Botany* 87: 259-272.

BEDDOME, R. H. 1874. *The flora sylvatica for Southern Indian Delhi*. Volume 3. Periodical Expert Book Agency.

BENTHAM, G. 1892. Handbook of the British Flora. Sixth edition. J. D. Hooker [rev.], Lovell Reeve & Co., London. U.K.

BENTHAM, G. AND J.D. HOOKER. 1865. *Genera Plantarum ad exemplaria impiris in Herbariis Kewensibus servata definita*. Volume 1. Sistens Dicotyledonum Polypetalarum Ordines LXXXIII: Ranunculaceas - Cornaceas. Reeve & Co., London.

BLUME, C. L. 1825. Dipterocarpaceae. *Bijdragen tot de Flora von Nederlandische Indie* 4: 222-224.

BRADFORD, J. C. 1998. A cladistic analysis of species groups in *Weinmannia* (Cunoniaceae) based on morphology and inflorescence structure. *Annals Missouri Botanical Garden* 85: 565-593.

BRADLEY, D., R. CARPENTER, L. COPSEY, C. VINCENT, S. ROTHSTEIN, AND E. COEN. 1996. Control of inflorescence architecture in *Antirrhinum*. *Nature* 379 (29): 791-797.

BRADLEY, D., O. RATCLIFFE, C. VINCENT, R. CARPENTER, AND E. COEN. 1997. Inflorescence commitment and architecture in *Arabidopsis*. *Science* 273: 80-83.

BRANDIS, S. D. 1895. An enumeration of the Dipterocarpaceae. *Journal of the Linnean Society* 31: 1-148.

BRIGGS, B. G., AND L. A. S. JOHNSON. 1979. Evolution in the Myrtaceae - evidence from inflorescence structure. *Proceedings of the Linnean Society of the New South Wales* 102 (4): 157-256.

BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, AND P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. *Systematic Biology* 42: 348-397.

BURCK, W. 1887. Sur les Dipterocarpacees de Indes Nerrlandaises. *Annales du Jardin Botanique de Buitenzorg* 6 (20): 221.

CAMPBELL, C. S., M. F. WOJCIECHOWSKI, B. G. BALDWIN, L. A. ALICE, AND M. J. DONOGHUE. 1997. Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier* agamic complex (Rosaceae). *Molecular Biology and Evolution* 14 (1): 81-90.

CANDOLLE, A. P. de. 1868. Dipterocarpaceae. In: *Prodomus systematis naturalis regni vegetabilis* (Ed. A. P. de Candolle). XVI (2) Pariisis argentorati londini treuttel wurstz. pp. 604-637. Vitoris Masson et Filii. Pariis.

CHAN, H. T. 1981. Reproductive biology of some Malaysian dipterocarps. III. Breeding systems. *The Malaysian Forester* 38: 160 – 170.

CHAPPILL, J. A. 1989. Quantitative characters in phylogenetic analysis. *Cladistics* 5: 217-234.

CHIPPENDALE, P. T., AND J. J. WEINS. 1994. Weighting, partitioning, and combining characters in phylogenetic analysis. *Systematic Biology* 43 (2): 278-287.

CHOONG, E. T., AND S. S. ACHMADI. 1996. Utilisation potential of the dipterocarps resource in International trade. *In* A. Schulte and D. Schone [eds.], Dipterocarp Forest Ecosystems Towards Sustainable Management. World Scientific, Singapore. Pp. 481-525.

CLUNIE, N. M. U. 1978. The vegetation. *In* J. S. Womersley [ed.], Handbooks of the Flora of Papua New Guinea. Volume I. Melbourne University Press.

COLLES, D. H. 1980. On "character" and related terms. *Systematic Zoology* 29: 288-299

COSTA, J.L., P. PAULSRUD AND P. LINDBLAD. 2002. The cyanobacterial t*RNA*(Leu) (UAA) intron: Evolutionary patterns in a genetic marker. *Molecular Biology & Evolution*. 19(6):850-857

CRANSTON, P. S., AND C. J. HUMPHRIES. 1988. Cladistics and computers: a chironomoid conundrum. *Cladistics* 4: 72-92.

CRISP, M. D. AND WESTON, P. H. 1987. Cladistics and legume systematics, with an analysis of the Bossiaceae, Brongniaetieae and Mirbelieae. *In* C. H. Stirton [ed.]. Advances in Legumes Systematics, Part 3. Royal Botanic Gardens, Kew. Pp. 65-130.

CURRAN, L. M., AND M. LEIGHTON. 2000. Vertebrate responses to spatiotemporal variation in seed production of mast-fruiting Dipterocarpaceae. *Ecological Monographs* 70 (1): 101-128.

CURRAN, L. M., AND C. O. WEBB. 2000. Experimental tests of the spatiotemporal scale of seed predation in mast fruiting Dipterocarpaceae. *Ecological Monograph* 70 (1): 129-148

DALZIEL, J. M., 1937. The useful plants of west Tropical Africa. The Crown Agents for Colonies. London. U.K.

DAVIS ,C. C., C. D. BELL, P. W. FRITSCH AND& S. MATHEWS. 2002 Phylogeny of *Acridocarpus-brachylophon* (Malpighiaceae): Implications for tertiary tropical floras and Afroasian biogeography. *Evolution* 56 (12): 2395-2405.

DAYANANDAN S., P. S. ASHTON, S. M. WILLIAMS AND& R. B. PRIMACK. 1999. Phylogeny of the tropical tree family Dipterocarpaceae based on nucleotide sequences of the chloroplast *rbc*L gene. *American Journal of Botany* 86 (8): 1182-1190.

DIGGLE, P. K. 1999. Heteroplasty and the evolution of flowering phenologies. *International Journal of Plant Sciences*. 160 (6 Suppl.): S123-S134.

DONOGHUE M. J. AND D.A. ACKERLY. 1996 Phylogenetic uncertainties and sensitivity analyses in comparative biology. *Philosophical Transactions of The Royal Society of London* Series B 351: 1241-1249.

DOYLE, J. J. AND J.L. DOYLE. 1987. A rapid DNA isolation procedure for smaller quantitaties of fresh leaf tissue. *Phytochemical Bulletin* 19: 11-15.

DOYLE, J. J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany* 17: 144-163.

DUBOUZET, J. G., AND K. SHINODA. 1999. ITS DNA sequence relationship between *Lilium concolor* Salisb., *L. dauricum* Ker-Gawl. and their putative hybrid, *L. maculatum* Thunb. *Theoretical and Applied Genetics* 98: 213-218.

EICHLER, A. W. 1878. Bluthendiagramme. Pts. 1,2. W. Engelmann, Leipzig. Reprinted 1954. O. Koeltz, Eppenhain, Germany.

ENDLICHER, S. 1840. *Genera Plantarum Secundum Ordines Naturales Disposita.* Vol. II. Vindobonae. Apud Fr. Beck Universitatis Bibliopolam.

ENDLICHER, S. 1841. Enchiridion botanicum exhibens classes et ordines plantarum accedit nomenclator generum et officinalum vel usualium Lipsiae. Sumptibus Guil. Engelmann. Viennae, Austria.

ERESHEFSKY, M. 1992. Eliminative Pluralism. *Philosophy of Science* 59: 671-690.

EVANS, R. C., AND T. A. DICKINSON. 1999. Floral ontogeny and morphology in subfamily Spiraeoideae Endl. (Rosaceae). *International Journal of Plant Sciences* 160 (5): 981-1012.

FARRIS, J.S. 1989. The retention index and the rescaled consistency index. *Cladistics* 5: 417-419.

FARRIS, J. S., M. KALLERSJO, A. G. KLUGE, AND C. BULT. 1994. Testing significance of incongruence. *Cladistics* 10: 315-319.

FITCH, W. M. 1971. Toward defining the course of evolution: minimum change for specified tree topology. *Sytematics Zoology* 20: 404-416.

FELSENSTEIN, J. 1985. Confidence limits on phylogenetics: an approach using the bootstrap. *Evolution* 39: 783-791.

FREUDENSTEIN, J. V. 1998. Paraphyly, ancestors and classification - a response to Sosef and Brummit. *Taxon* 47: 95-104.

FREUDENSTEIN, J. V., AND F. N. RASMUSSEN. 1999. What does morphology tell us about Orchid relationships? - a cladistic analysis. *American Journal of Botany* 86 (2): 225 - 248.

FRIEDMANN, F. 1994. Flore des Seychelles. Dicotylédones. Edition de l'orstom. Institute Française de Recherche Scientifique pour le dévelopment en coopération. Collection actiques. Paris, France. Pp 148-150.

FRITJERS, D. 1978. Mechanism of developmental integration of *Aster novae-angliceae* L. and *Hieracium murorum* L. *Annals of Botany* 42: 561-575.

GÄERTNER, C. F. 1805. Fructibus et seminibus plantarum. Volume 3. Supplementum Carpologiae. Sumtibus Carol. Frid. Enoch Richter. Bibliopolae Lipsiensis. Pp. 47-49.

GILMORE, S., P. H. WETSON, AND J. A. THOMSON. 1993. A simple, rapid, inexpensive and widely applicable technique for purifying plant DNA. *Australian Systematic of Botany* 6: 139-148.

GOLENBERG, E. M., M. T. CLEGG, M. L. DURBIN, J. DOUBLEY AND D.P. MA. 1993. Evolution of non coding region of the chloroplast genome. *Molecular Phylogenetics and Evolution* 2: 52-64.

GONZALES, D. AND A. P. VOVIDES. 2002. Low intralineage divergence in *Ceratozamia* (Zamiaceae) detected with nuclear ribosomal DNA ITS and chloroplast DNA *trn*L-F non-coding region. *Systematic Botany* 27 (4): 654-661.

GORDON, M. S. 1999. The concept of monophyly: a speculative essay. *Biology and Philosophy* 14: 331-348.

GRIMES, J., 1999. Inflorescence morphology, heterochrony, and phylogeny in the Mimosoid Tribes Ingeae and Acacieae (Leguminosae: Mimosoideae). *The Botanical Review* 4: 317-340.

HAM R. C. H. J., H. T. HART, T. H. M. MES AND J. M. SANDBRINK. 1994. Molecular evolution of noncoding regions of the chloroplast genome in Crassulaceae and related species. *Current Genetics* 25: 558-566.

HASSKARL, J. K. 1858. *Hortus Bogoriensis descriptus sive Retziae*. Part. 1. F. Gunst, Amsterdam.

HEIM, F. 1891. Dipterocarpacees. *Bulletin mensuel de la societe linneene de Paris* 119: 994.

HEIM, F. 1892. *Recherches sur les Dipterocapacées*, 1-186. Ph.D. dissertation à la faculte des sciences de Paris. France.

HERSHKOVITZ, M. A. AND L.A. LEWIS. 1996. Deep-level diagnostic value of the rDNA-ITS region. *Molecular Biology and Evolution* 13(9): 1276-1295.

HILLIS, D. M., C. MORITZ AND B. K. MABLE. 1996. *Molecular systematics*. Second Edition. Sinauer Associates Inc., Sunderland.

HORVATH, C. D. 1997. Discussion: Phylogenetic species concept: Puralism, Monoism, and History. *Biology and Philosophy* 12: 225-232.

JANZEN, D. H. 1974. Tropical blackwater rivers, animals, and mast fruiting by the Dipterocarpaceae. *Biotropica* 6 (2): 69-103

KAJITA, T., K., KAMIYA, K. NAKAMURA *et al.* 1998 Molecular phylogeny of Dipterocarpaceae in South East Asia based on nucleotide sequences of *mat*K, *trn*L intron, *trn*L-*trn*F intergenic spacer region in chloroplast DNA. *Molecular Phylogenetic and Evolution* 10 (2): 202-209.

KIM, K. J., AND R. K. JANSEN, 1994. Comparisons of phylogenetic hypotheses among different data sets in dwarf dandelions (*Krigia*, Asteraceae): additional information from internal transcribed spacer sequences of nuclear ribosomal DNA. *Plant Systematics and Evolution* 190: 157-185.

KITCHER, P. 1984. Species. *Philosophy of Science* 51: 308-333.

KLUGE, A. J. 1989. A concern for evidence and a phylogenetic relationship among *Epicrates* (Boidae, Serpentes). *Systematic Zoology* 38: 7-25.

KLUGE, A. G. AND J. S. FARRIS. 1994. Quantitative phyletics and the evolution of Anurans. *Systematic Zoology* 18: 1-32.

KONG, H. Z., Z. D. CHEN AND A.M. LU. 2002. Phylogeny of *Chloranthus* (Chloranthaceae) based on nuclear ribosomal ITS and plastid *trn*L-F sequence data. *American Journal of Botany* 89 (6): 940-946.

KORNET, D. J. AND H. TURNER. 1999. Coding polymorphism reconstruction. *Systematic Biology* 48 (2): 365-379.

KORNKVEN, A. B., L. E. WATSON, AND J. R. ESTES. 1998. Phylogenetic analysis of *Artemesia* section *Tridentatae* (Asteraceae) from the Internal Transcribed Spacers (ITS) of nuclear ribosomal DNA. *American Journal of Botany*. 85 (12): 1787-1795.

KOSTERMANS, A. J. G. H. 1978. *Pakaraimaeae dipterocarpaceae* Maguire and Ashton belongs to Tiliaceae not Dipterocarpaceae. *Taxon* 27: 357-359.

KOSTERMANS, A. J. G. H. 1983. The genus *Shorea* (Dipterocarpaceae). *Botanische Jahrbuecher fur Systamatik Pflanzengeschichte und Pflanzengeographie* 104: 183-201.

KOSTERMANS, A. J. G. H. 1984. Monograph of the genus *Doona* (Dipterocarpaceae). *Botanische Jahrbuecher fur Systamatik Pflanzengeschichte und Pflanzengeographie* 104: 425-435.

LEWIN, B. 1990. Genes IV. Oxford University Press. Oxford, U.K.

LINDLEY ,J. 1853. *The Vegetable Kingdom; or the structure, classification, and uses of plants, Illustrated upon the natural system.* Third Edition. Bradbury & Evans, London.

LINNAEUS, C. 1961. Mantissa plantarum 1761 and 1771, with an introduction by William T. Stearn. Facsimile ed. Weinheim Cramer, New York, U.S.A. PAGE

LISTON, A., W. A. ROBINSON AND J. M. OLIPHANT. 1996. Length variation in the nuclear ribosomal DNA Internal Transcribed Spacer region of non-flowering seed plants. *Systematic Botany* 21 (2): 109-120.

LONDONO, A. C., E. ALVAREZ, E. FERERO, AND C. M. MORTON. 1995. A new genus and species of Dipterocarpaceae from the Neotropics. 1. Introduction, taxonomy, ecology and distribution. *Brittonia* 47 (3): 225 – 236.

MADDISON, D. R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. *Systematic Zoology* 40: 315-328.

MADDISON, W. P. AND D. R. MADDISON. 1992. *Mac Clade, analysis of phylogeny and character evolution.* Version 3.04. Sinauer, Sunderland.

MAGUIRE, B. P. C., AND P. S. ASHTON. 1977. Pakaramoideae, Dipterocarpaceae of the Western Hemisphere II. Systematic, geographic and phyletic considerations. *Taxon* 26: 359-368.

MASON-GAMER R. AND A. KELLOG. 1996. Testing for phylogenetic conflict among molecular data sets in Triticeae (Gramineae). *Systematic Biology* 45: 524-525.

MARTIN, M. A. 1971. Intoduction A l ' ethnobotanique du Cambodge. Edition du centre national de la Recherche Scientifique. Paris. France. Pp 69-71.

MAURY-LECHON, G. AND L. CURTET. 1998. Biogeography and evolutionary systematics of Dipterocarpaceae. In: *A Review of Dipterocarps taxonomy, ecology and silviculture* (Eds S. Appanah & J. M. Turnbull) pp. 5-44. CIFOR, Bogor.

MEIJER, W. 1974. Plant geographic studies on Dipterocarpaceae in Malesia. *Annals of the Missouri Botanic Garden* 61: 806-818.

MEIJER W. AND G. H. S. WOOD. 1964. Dipterocarps of Sabah (North Borneo). *Sabah Forest Record* 4: 1-344.

MEIJER W. AND G. H. S. WOOD. 1976. Keys to Dipterocarps of Sabah. BIOTROP, Bogor.

MES T. H. M. AND H. T. HART. 1994. *Sedum surculosum* and *S. jaccardianum* (Crassulaceae) share a unique 70 bp deletion in the chloroplast DNA *trn*L (UAA)-*trn*F (GAA) intergenic spacer. *Plant molecular and Evolution* 193: 213-221.

MICKEVICH, M. F., AND M. F. JOHNSON. 1976. Congruence between mophological and allozyme data in evolutionary inference and character evolution. *Systematic Zoology* 25: 260-270

MIQUEL, F. A. W. 1820. Dipterocarpaceae. In: *Annales musei Botanici Lugduno-Batavi*. Volume 2 (Ed F. A. Guil). Amsterdam.

MIYAMOTO M. M. AND W. M. FITCH. 1995. Testing species phylogenies and phylogenetic methods with congruence. *Systematic Biology* 44: 64-67.

MOMOSE, K., T. YUMOTO, T. NAGAMITSU, M. KATO, H. NAGAMASU, S. SAKAI, R. D. HARRISON, A. A. HAMID, AND T. INOUE. 1997. Pollination biology in a lowland dipterocarp forest in Sarawak, Malaysia I. Characteristics of the plant-pollinator community in a lowland dipterocarp forest. *In* T. Inoue and A. A. Hamid [eds.], General flowering of tropical rainforest in Sarawak, 142-162, Kyoto University, Kyoto, Japan.

MORTON B. R. AND M. T. CLEGG. 1993. A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbc*L in the grass family (Poaceae). *Current genetics* 24: 357-365.

MORTON, C. M. 1995. A new genus and species of Dipterocarpaceae from the Neotropics. 2. Stem anatomy. *Brittonia* 47 (3): 237 – 247.

MUIR, G., C.C. FKEMING AND C. SCHLOTTERER. 2001. Three divergent rDNA clusters predate the species divergence *in Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. *Molecular biology and evolution* 18 (2): 112-119.

MUMMENHOFF, K., H. BRUGGEMANN AND J.L. BOWMAN. 2001. Chloroplast DNA phylogeny and biogeography of Lepidium (Brassicaceae) *American Journal of Botany*. 88(11):2051-2063, 2001.

NEI, M., AND R. K. KOEHN. 1983. Evolution of genes and proteins. Sinauer, Sunderland, Massachusetts, MA, U.S.A.

Nickrent D., Schuette K. P. & Starr E. M. 1993. A molecular phylogeny of *Arceuthobium* (Viscaceae) based on nuclear ribosomal DNA Transcribed Spacer sequences. *American Journal of Botany* 81 (9): 1149-1160.

OKAURA, T. AND K. HARADA. 2002. Phylogeographical structure revealed by chloroplast DNA variation in Japanese Beech (*Fagus crenata* Blume). *Heredity* 88 (4): 322-329.

PADGETT D. J., D. H. LES AND G. E. CROW. 1999. Phylogenetic relationships in *Nuphar* (Nymphaceae): evidence from morphology, chloroplast DNA, and nuclear ribosomal DNA. *American Journal of Botany* 86 (9): 1316-1324.

PLATAS, G., J.D. POLISHOOK AND F. PELAEZ. 2002. VNTR polymorphism in the ITS1 region (rDNA) in isolates of a Nodulisporium species. *Mycological Research*. 106(11):1293-1298

PIERRE, J. B. L. (1892) *Flore Forestière de la Conchin-Chine*. 4 (83): 246-249.

PROMEGA INC.

QIAGEN INC., Chatsworth, U.S.A

QUICKE, D. L. J. 1993. Principles and techniques of contemporary taxonomy. Blackie Academic & Professional. Chapman & Hall. St. Edmundsbury Press Ltd., St. Edmunds, Suffolk, U.K. PAGE

QUEIROZ, K. DE, M. J. DONOGHUE AND J. KIM. 1995. Separate versus combined analysis of phylogenetic evidence. *Annals Review of Ecology and Systematics* 26: 657-681.

_____. 1997. The Linnaean hierarchy and the evolution of taxonomy, with emphasis on the problem of nomenclature. *Aliso* 15(2): 125-144.

QUICKE, D. L. J. 1993. Principles and techniques of contemporary taxonomy. Blackie Academic & Professional. Chapman & Hall. St. Edmundsbury Press Ltd., St. Edmunds, Suffolk, U.K. PAGE

RAE, T. C. 1998. The logical basis for the use of continuous characters in phylogenetic systematics. *Cladistics* 14: 221-228.

RATH, P., G. RAJASEGER, C. J. GOH AND P. KUMAR (1998) Phylogenetic analysis of Dipterocarps using Random Amplified Polymorphic DNA markers. *Annals of Botany* 82: 61-65.

RICKETT, H. W. 1944. The Classification of inflorescences. *Botanical Review* 10: 183-231.

ROGSTAD S. H. 1992. Saturated NaCl-CTAB solution as a means of field preservation of leaves for DNA analyses. *Taxon* 41: 701-708.

ROXBURGH, W. 1814. Hortus Bengalensis, or a catalogue of the plants growing in the Honourable East India Company's botanic garden at Calcutta. Seranpore: Mission Press, Calcutta, India. PAGE

SAKAI, S. K. MOMOSE, T. NAGAMITSU, R. D. HARRISON, T. YUMOTO, T. ITINO, M. KATO, H. NAGAMASU. A. A. HAMID, AND T. IMOUE.. 1997. An outline of plant reproductive phenology in one episode of general flowering cycle in 1992 - 1996 in Sarawak, Malaysia. *In* T. Inoue, and A. A. Hamid [eds.], General flowering of tropical rainforest in Sarawak. Kyoto University, Kyoto, Japan. Pp. 112-141.

SALDANHA, C. J., AND M. S. E. RAO. 1985. Dipterocarpaceae. *In* C. J. Saldanha [ed.], Flora of Karnataka, Volume 1. Magnoliaceae to Fabaceae. Indian edition series 9. A. A. Balkema, Rotterdam, The Netherlands. Pp. 191-196.

SANG, T., D. J. CRAWFORD AND T. F. STUESSY. 1997. Chloroplast DNA phylogeny, reticulate evolution and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84 (8): 1120-1136.

SHIZONAKI, K., M., OHME, M. TANAKA, T. WAKASUGI, N. HAYASHIDA, T. MATSUBAYASHI, N. ZAITA, J. CHUNWONGSE, J. OBOKATA, K. YAMAGUCHI-SHIZONAKI, C. OHTO, K. TORAZAWA, B. Y. MENG, M. SUGITA, H. DENO, T. KAMOGASHIRA, K. YAMADA, J. KUSUDA, F. TAKAIWA, A. KATO, N. TOHDOH, H. SHIMADA, AND M. SUGUIRA. 1986. The complete nucleotide sequence of tobacco chloroplast genome: Its gene organization and expression. *EMBO Journal* 5: 2043-2049.

SINGER, S., J. SOLLINGER, S. MAKI, J. FISHBACH, B. SHORT, C. REINKE, J. FICK, L. COX, A. MCCALL, AND H. MULLEN. 1999. Inflorescence architecture: a developmental genetic approach. *Botanical Review* 4: 385-409.

SMITINAND, T., J. E. VIDAL, AND P. H. HO. 1990. Dipterocarpacees. *In* P. Morat [ed.], Flore du Cambodge, du Laos et du Viêtnam. Muséum National D ' histoire Naturelle, Paris, France. Pp. 3-10.

SOMEGO, M. 1978. Cytogenetical study of Dipterocarpaceae. *The Malaysian Forester* 41 (4): 358-265.

SOSEF, M. S. M. 1997. Hierarchical models, reticulate evolution and the inevitability of paraphyletic supraspecific taxa. *Taxon* 46: 75-85.

STANDFORD, P. 1995. For Pluralism and against Realism about species. *Philosophy of Science* 62: 70-91.

STEANE, D. A., R. W. SCOTLAND, D. J. MABBERLEY, AND R. G. OLMSTEAD. 1999. Molecular systematics of *Clerodendrum* (Lamiaceae): ITS sequences and total evidence. *American Journal of Botany* 86 (1): 98-107.

STEBBINS, G. L. 1974. Flowering plants: evolution above the species level. Harvard University Press, Cambridge, Massachusetts, MA, U.S.A. PAGE

STUESSY, T. F. 1997. Classification: more than just branching patterns of evolution. *Aliso* 15 (2): 113-124.

SUZUKI, E., AND P. S. ASHTON. 1996. Sepal and nut size ratio of fruits of Asian Dipterocarpaceae and its implications for dispersal. *Journal of Tropical Ecology* 12: 853 – 870.

SWOFFORD, D. L. 1998. *Phylogenetic analysis using parsimony (*and other methods).* Sinauer, Sunderland.

SYMINGTON, C. F. 1943. Foresters' Manual of the Dipterocarps. Forest Department, Kuala Lumpur.

TABERLET, P., L. GIELLY, G. PATOU AND J. BOUVET. 1991. Universal primers for amplification three non coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105-1109.

TARDIEU-BLOT, M. 1950. Dipterocarpacees. In: Flore Générale de l'Indo-Chine. Supplement. Imprimerie Protat Frères, Macon. CONTRY. Pp. 334-361.

THIELE, K. 1993. The holy grail of the perfect character: the cladistic treatment of morphometric data. *Cladistics* 9: 275-304.

THIELE, K, AND P. Y. LADIGES. 1988. A cladistic analysis of *Angophora* Cav. (Myrtaceae). *Cladistics* 4: 23-42.

TORRES, R. A., M. GANAL, AND V. HEMLEBEN. 1990. GC balance in the Internal Transcribed Spacers ITS 1 and ITS 2 of nuclear ribosomal RNA genes. *Journal of Molecular Evolution* 30: 170 – 181.

THORPE, R. S. 1984. Coding morphometric characters for constructing distance Wagner networks. *Evolution* 38: 244-355.

TROLL, W. 1964. Die Infloreszenzen, Typologie um Stellung im Aufbau des Vegetationskorpers. Vol. 1. Gustav Fischer Verlag, Jena, Germany. PAGE

TUCKER, S. C. 1987. Pseudoraceme in Papilionoid legumes: Their nature, development and variation. *Journal of the Linnean Society of Botany* 95: 181-206.

TUCKER, S. C., AND J. GRIMES, 1999. The inflorescence: Introduction. *The Botanical Review*. 4: 303-316.

THOMPSON J. D., D. G. HIGGINS AND T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of pregressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acid Research* 22: 4673-4680.

THWAITES, G. H. K. 1835. Enumeratio plantarum Zeylanica: an enumeration of Ceylon plants, with descriptions of the new and little known genera and species, observations on their habitats, uses, native names, etc. London

TSUMURA, Y., T. KAWAHARA, R. WICKNESWARI AND Y. YOSHIMURA Y. 1996. Molecular Phylogeny of Dipterocarpaceae in South East Asia using RFLP PCR-amplified chloroplast genes. *Theoritical and Applied Genetics* 93 (1-2): 22-29.

VERDCOURT, B. 1989. Dipterocarpaceae. *In* R. M. Polphill [ed.], Flora of Tropical East Africa. A. A. Balkema, Rotterdam, Brookfield, The Netherlands. Pp. 1-11.

WEBERLING, F. 1989. Morphology of Flowers and Inflorescences. Cambridge University Press, Cambridge, London, U.K. PAGE

WEIBLEN, G. D. 2000. Phylogenetic relationships of functionally dioecious *Ficus* (Moraceae) based on ribosomal DNA sequences and morphology. *American Journal of Botany* 87 (9): 1342-1357.

WEINS, J. J. 1995. Polymorphism characters in phylogenetic systematics. *Systematic Biology* 44 (4): 482-500.

WHITE, T. J., T. BRUNS, S. LEE AND J. TAYLOR. 1990. Amplification and direct sequencing for fungal ribosomal rRNA genes for phylogenetics. In: *PCR Protocols* (Eds M. A. Innis, D. H. Gelfard, J. J. Sninsky & T. J. White) pp. 315-322. Academic Press San Diego.

WHITMORE, T. C. 1962. Studies in systematic bark morphology. III. Bark taxonomy in Dipterocarpaceae. *Gardens Bulletin Singapore* 19: 321-371.

_____. 1975. Tropical rainforest of the far east. Clarendon Press, Oxford, London, U.K.

WIGHT, R. AND G. A. ARNOTT. 1834. *Prodromus florae Peninsulae Indiae Orientalis, containing abridged descriptions of the plants found in the Peninsula of British India, arranged according to the Natural System.* Volume 1. Allen & Co., London.

WILEY, E. O. 1981. Phylogenetics: the theory and practice systematics. J. Wiley & Sons, New York.

ZIMMER, E. A., S. L. MARTIN, S. M. BEVERLEY, Y. W. KAN, AND A. C. WILSON. 1980. Rapid duplication and loss genes for the alpha chains of hemoglobin. *Proceeding of the National Academy of Sciences U. S. A.* 77: 2158-2162.

ZIMMER, E. A., E. H. ROALSON, L. E. SKOG, J. K. BOGGAN AND A. IDNURM. 2002. Phylogenetic relationships in the Gesnerioideae (Gesneriaceae) based on nrDNA ITS and cpDNA *trn*L-F and *trn*E-T spacer region sequences. *American Journal of Botany* 89 (2): 296-311.