Theor Appl Genet (1996) 93:22–29

Y. Tsumura · T. Kawahara · R. Wickneswari
K. Yoshimura

Molecular phylogeny of Dipterocarpaceae in Southeast Asia using RFLP of PCR-amplified chloroplast genes*

Received: 5 October 1995 / Accepted: 17 November 1995

Abstract Dipterocarpaceae is the dominant family of Southeast Asia's climax tropical rain forest region, and it contains the region's most important commercial timber species. A molecular phylogeny of the Dipterocarpaceae subfamily Dipterocapoideae was constructed using restriction fragment length polymorphisms of polymerase chain reaction-amplified specific genes in chloroplast DNA. A total of 141 site changes were detected among ten genera and 30 species in 11 different genes: *rbc*L, *psb*A, *psb*D, *rpo*B, *rpo*C, *pet*B, *atp*H, 16S, *psa*A, *pet*A and *trn*K. Phylogenetic trees constructed by Wanger parsimony and neighbor-joining methods, using *Upuna* as the outgroup, displayed five monophytelic groups that included *Upuna*: *Hopea-Shorea-Parashorea-Neobalanocarpus*; *Dryobalanops*; *Dipterocarpus*; *Anisoptera-Vatica-Cotylelobium*; and *Upuna*. The phylogenetic trees clearly separate species with two different base chromosome numbers: the first group is $x = 7$, and the other is $x = 11$. The $x = 7$ group is thought to be in a synapomorphic character state. *Parashorea lucida* is a sister to most *Shorea* species. *Neobalanocarpus heimii* and *Hopea* from a clade of a sister to two *Shorea* species, and *Cotylelobium* and *Vatica* are closely related species. Our conclusions agree with a phylogeny derived from wood anatomy data analysis, and with Symington's and Ashton's taxonomic classifications.

Key words Dipterocarpaceae · PCR-RFLP
Chloroplast gene · Phylogeny · Southeast Asia

Communicated by P. M. A. Tigerstedt

Y. Tsumura (✉) · T. Kawahara · K. Yoshimura
Genetics Section, Bio-resources Technology Division, Forestry and Forest Product Research Institute, Kukizaki, Ibaraki, 305 Japan
E-mail: ytsumu@ffpri.affrc.go.jp

R. Wickneswari
Forest Research Institute of Malaysia, Kepong, 52109 Kuala Lumpur, Malaysia

* The raw data of the PCR-RFLP analysis can be obtained from the authors

Introduction

Dipterocarpaceae is distributed mainly in Southeast Asia and its species represent a major canopy component in tropical forests. Dipterocarpaceae species are dominant in Southeast Asia's tropical rain forests, and family is one of the region's most important tree families, being represented by an extreme richness of species in many tropical rain forests (Federov 1966; Ashton 1969; Whitmore 1984). The family Dipterocarpaceae has been divided into three subfamilies: Dipterocarpoideae, Monotoideae and Pakaraimoideae. Monotoideae is represented in Africa and Madagascar by two genera. The monotypic genus Pakaraimaea is placed in the Pakaraimoideae and is distributed throughout Guyana, South America. According to recent classifications, Dipterocarpoideae includes 13 genera and 470 species (Symington 1943; Ashton 1982).

Phylogenetic relationships among dipterocarps have been addressed using only morphology data, fossil records and pollen analysis (see review of Ashton 1982). In particular, wood anatomy data has provided valuable taxonomic information at the subfamily, tribe, genetic, infrageneric and species levels (Desch 1941; Gottwald and Parameswaran 1966; Ashton 1982). Cytological data for the family are limited but provide important information relevant to cytotaxonomy. It has been established that Indo-Malesian species in the 9 genera studied have, for the most part, 7 or 11 chromosomes (Jong and Lethbridge 1967; Jong 1976; Somego 1978). Although genetic studies have been conducted in these species (Gan et al. 1977; Gan and Robertson 1981; Ashton et al. 1984; Ihara et al. 1986; Harada et al. 1994; Kitamura et al. 1994; Murawski and Bawa 1994), molecular phylogenetic relationships among genera and species have not been investigated. In the study described here, we developed hypotheses for the phylogeny of Dipterocarpaceae using restriction fragment length polymorphism (RFLP) of specific genes of chloroplast DNA (cpDNA) amplified by the polymerase chain reaction (PCR).

Material and methods

Leaf samples

Leaves from 30 dipterocarp tree species from ten genera were collected at the Forest Research Institute of Malaysia (FRIM) arboretum and nursery (Table 1).

Manipulation of DNA

Total DNA was extracted from leaves of each species by a slightly modified version of the CTAB method (Murray and Thompson 1980). Approximately 5 g of fresh leaf tissue was ground to a fine powder with liquid nitrogen in a Millser IFM-150 homogenizer (Iwatani Co.). Twenty milliliters of 2 × CTAB buffer (2% CTAB, 1.4 *M* NaCl, 0.2% β-mercaptoethanol, 20 m*M* EDTA, 100 m*M* Tris-HCl, pH 8.0, 1% PVP-40) was added to the ground powder and homogenized with the homogenizer again. The solution was incubated at 60 °C for 30 min, 20 ml of chloroform solution (chloroform:isomyl alcohol = 24:1) was then added and the solution was mixed gently for 15 min and then centrifuged at 4,500 rpm for 15 min. The aqueous phase was subsequently transferred to a clean tube to which a two-third volume of cold isopropanol was added and mixed gently to precipitate the nucleic acids. The DNAs were purified by RNase digestion and extracted two to four times with TE-saturated phenol.

Eleven genes in the cpDNA were amplified by PCR: *rbc*L, *rpo*B, *pet*B, *psb*A, *psb*D, *atp*H, 16S, *rpo*C, *psa*A, *pet*A and *trn*K (Table 2). Primers were designed using the DNASIS Program (Takara, Kyoto) and were based on nucleotide sequence data from the EMBL database. The primers were synthesized (Nippon Gene). PCR amplification of these cpDNA regions and PCR-RFLP analysis were performed by the methods of Tsumura et al. (1995).

Scoring of site changes and phylogenetic tree construction

Five microliters of the PCR product was used for estimating the sizes of fragment-amplified genes by gel electrophoresis on 2% agarose gels in TAE buffer before digestion with restriction endonucleases. The digested DNAs were also subjected to electrophoresis on 2% agarose gels, and the molecular size of each fragment was estimated (± 5% error) by comparing the size marker, *Hinc*II-digested ϕ 174, using the DNASIS Program (Takara, Kyoto). Site changes between specific genes in the various species was determined from these data.

Phylogenetic analysis was conducted using the heuristic search condition of the PAUP 3.0s Program (Phylogenetic Analysis Using Parsimony; Swofford 1991) with the Wagner parsimony option and also by the neighbor-joining (NJ) method (Saitou and Nei 1987). If phylogenetic trees constructed by several different methods agree, then confidence in the trees and the data used for their construction will increase. When the NJ method was used, a matrix of pairwise distances between species was calculated using the PAUP program. Because *Upana* displays primitive morphological characteristics (Ashton 1982), this taxon was treated as an outgroup for the other taxa. To estimate the nucleotide substitutions, we used the formula of Nei and Tajima (1983).

Results

Variation among species

We amplified the 11 genes of interest, including introns and/or spacers, of cpDNA. The predominant fragment sizes of the different genes showed no variation among species when examined on 2% agarose gel. The length of the amplified fragment of *rpo*B from dipterocarps was

Table 1 Species subjected to an analysis of cpDNA-based phylogeny, their abbreviations, voucher number and sources

Number	Species	Abbreviation	Voucher no.[a]	Source[b]
1	*Shorea macroptera* Dyar	SMA	LGF 59/95	FRIM, Kepong, Malaysia
2	*Shorea atrinervosa* Sym.	SAT	LGF 35/95	FRIM, Kepong, Malaysia
3	*Shorea bracteolata* Dyer	SBR	LGF 77/95	FRIM, Kepong, Malaysia
4	*Shorea singkawang* (Miq.) Miq.	SSI	LGF 64/95	FRIM, Kepong, Malaysia
5	*Shorea ovalis* (Korth.) Bl.	SOV	Kawahara-95025	FRIM, Kepong, Malaysia
6	*Shorea multiflora* (Burck) Sym.	SMU	LGF 79/95	FRIM, Kepong, Malaysia
7	*Shorea scaberrima* Burck	SSC	LGF 3/95	FRIM, Kepong, Malaysia
8	*Shorea kunstleri* King	SKU	LGF 66/95	FRIM, Kepong, Malaysia
9	*Shorea parvifolia* Dyer	SPA	LGF 70/95	FRIM, Kepong, Malaysia
10	*Shorea macrophylla* (De Vriese) Ashton	SMP	LGF 73/95	FRIM, Kepong, Malaysia
11	*Shorea lepidota* (Korth.) Bl.	SLE	LGF 55/95	FRIM, Kepong, Malaysia
12	*Hopea helferi* (Dyer) Brandis	HHE	LGF 68/95	FRIM, Kepong, Malaysia
13	*Hopea sangal* Korth	HSA	LGF 78/95	FRIM, Kepong, Malaysia
14	*Hopea latifolia* Sym.	HLA	LGF 63/95	FRIM, Kepong, Malaysia
15	*Hopea subalata* Sym.	HSU	LGF 58/95	FRIM, Kepong, Malaysia
16	*Hopea odorata* Roxb.	HOD	LGF 61/95	FRIM, Kepong, Malaysia
17	*Hopea wightiana*	HWI	LGF 59/95	FRIM, Kepong, Malaysia
18	*Hopea dyeri* Heim	HDY	LGF 80/95	FRIM, Kepong, Malaysia
19	*Hopea apiculata* Sym	HAP	LGF 65/95	FRIM, Kepong, Malaysia
20	*Hopea nervosa* King	HNE	LGF 74/95	FRIM, Kepong, Malaysia
21	*Dipterocarpus kerrii* King	DIK	LGF 71/95	FRIM, Kepong, Malaysia
22	*Dipterocarpus baudii* Korth	DIB	LGF 53/95	FRIM, Kepong, Malaysia
23	*Dipterocarpus oblongifolius* Bl.	DIO	LGF 72/95	FRIM, Kepong, Malaysia
24	*Neobalanocarpus heimii* (King) Ashton	NEO	LGF 57/95	FRIM, Kepong, Malaysia
25	*Dryobalanops aromatica* Gaertn.	DRY	LGF 51/95	FRIM, Kepong, Malaysia
26	*Parashorea lucida* (Miq.) Kurz	PAR	LGF 67/95	FRIM, Kepong, Malaysia
27	*Vatica odorata* (Griff.) Sym.	VAT	LGF 75/95	FRIM, Kepong, Malaysia
28	*Cotylelobium malayanum* V. Si.	COT	LGF 52/95	FRIM, Kepong, Malaysia
29	*Anisoptera oblonga* Dyer	ANI	Kawahara-95067	FRIM, Kepong, Malaysia
30	*Upuna bornensis* Sym.	UPU	LGF 76/95	FRIM, Kepong, Malaysia

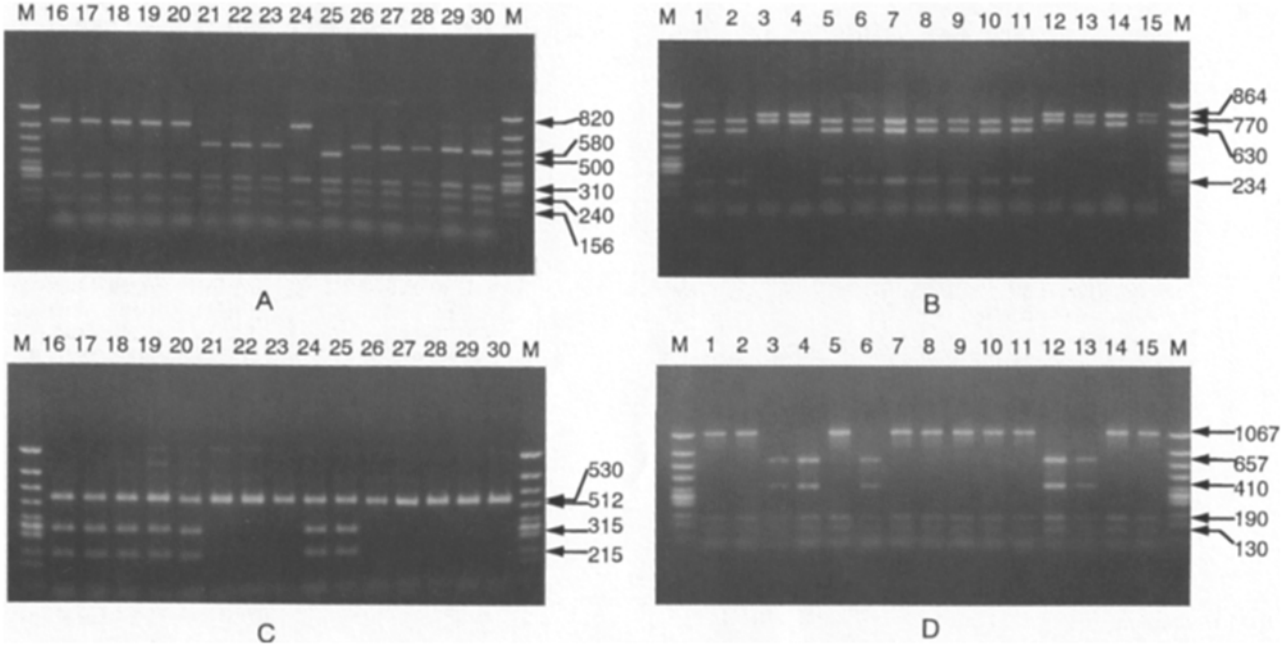
[a] FRIM Herbarium
[b] FRIM, Forest Research Institute of Malaysia dipterocarps arboretum

Table 2 Primers, fragments sizes and sources of information for amplification by PCR of 11 specific genes found in cpDNA

Gene	Primers (forward and reverse)	Size (bp)	Source
*rbc*L	5′-TGTCACCAAAAACAGAGACT-3′ 5′-TTCCATACTTCACAAGCAGC-3′	1,387	*Pseudotsuga menziesii* (Hipkins et al. 1990)
*rpo*B	5′-CTAAGGGGTTGTTGTGTAAC-3′ 5′-AATATGCAACGTCAAGCAGT-3′	1,286	*Nicotiana tabacum* (Shinozaki et al. 1986)
*pet*B	5′-TGGGGAACTACTCCTTTGAT-3′ 5′-CCCGAAATACCTTGCTTACG-3′	1,634	*Nicotiana tabacum* and *Oryza sativa* (Shinozaki et al. 1986; Hiratsuka et al. 1989)
*psb*A	5′-TACGTTCGTGCATAACTTCC-3′ 5′-CTAGCACTGAAAACCGTCTT-3′	939	*Pinus contorta* (Lidholm and Gustafsson 1991)
*psb*D	5′-TATGACTATAGCCCTTGGTA-3′ 5′-TAGAACCTCCTCAGGGAATA-3′	1,042	*Nicotiana tabacum* (Shinozaki et al. 1986)
*atp*H (*atp*H-*atp*I)	5′-TTGACCAACTCCAGGTCCAA-3′ 5′-CCGCAGCTTATATAGGCGAA-3′	1,283	*Spinacia oleracea* (Hennig and Herrmann 1986)
16S	5′-ACGGGTGAGTAACGCGTAAG-3′ 5′-CTTCCAGTACGGCTACCTTG-3′	1,375	*Nicotiana tabacum* (Shinozaki et al. 1986)
*rpo*C (*rpo*C1-*rpo*C2)	5′-GCAGTTTCTTGAAAACTCGC-3′ 5′-TGTACACGCGGTAGAAAAAT-3′	3,603	*Nicotiana tabacum* (Shinozaki et al. 1986)
*psa*A	5′-AAGAATGCCCATGTTGTGGC-3′ 5′-TTCGTTCGCCGGAACCAGAA-3′	2,218	*Nicotiana tabacum* (Shinozaki et al. 1986)
*pet*A	5′-TATGAAAATCCACGAGAAGC-3′ 5′-TATCAGCAATGCAGTTCATC-3′	2,314	*Nicotiana tabacum* (Shinozaki et al. 1986)
*trn*K	5′-AACCCGGAACTAGTCGGATG-3′ 5′-TCAATGGTAGAGTACTCGGC-3′	2,569	*Oryza sativa* (Hiratsuka et al. 1989)

1,286 bp (± 5%) and was the same as that of tobacco *rpo*B. When the fragments were digested by *Msp*I, three patterns were identified, two of which we denoted pattern A (820 bp, 310 bp and 156 bp; ± 5%) and pattern B (580 bp, 310 bp, 240 bp and 156 bp; ± 5%). If the 820-bp (± 5%) fragment of pattern A were to gain a new site for *Msp*I, pattern A would be transformed into pattern B (Fig. 1 A). The length of the amplified dipterocarp fragment of *psb*D was 1,042 bp (± 5%) and was the same as that of tobacco *psb*D. When this fragment was digested by *Mae*I, we detected two patterns (Fig. 1C). In pattern A, there were two fragments, 530 bp and 512 bp (± 5%). If the 530-bp fragment were to gain a new site for *Mae*I, pattern A would change to yield two fragments of 315 bp and 215 bp (± 5%), which are the size of the fragments of pattern B. The RFLP patterns of the *Msp*I-digested

Fig. 1A–D RFLP patterns of specific PCR-amplified genes of cpDNA. **A** *Msp*I-digested *rpo*B region, **B** *Hae*III-digested *pet*A region, **C** *Mae*I-digested *psb*D region, **D** *Taq*I-digested *rbc*L region. *Numbers above* each photograph refer to the numbering in Table 1 and *M* indicates the molecular size marker, namely, a *Hin*cII digest of ϕ 174



*rpo*B region and the *Taq*I-digested *rbc*L region can also easily be explained by site changes (Fig. 1B, D). Except for *atp*H, the other amplified genes from other species were shown to be similar in size to tobacco cpDNA genes. Since the amplified *atp*H fragment was 385 bp (± 5%) long, dipterocarp *atp*H may not have a spacer region between *atp*H and *atp*I.

We detected a total of 141 site changes by RFLP analysis of PCR-amplified cpDNA genes, but a clear length mutation was not detected in this study. The total sequence surveyed was 1,126 bp and was composed of 6% of all the amplified sequences of the genes (Table 3). The length of the amplified genes ranged from 385 bp for *atp*H to 3,603 bp (both ± 5%) for *rpo*C1-C2. Amplified fragments were coding regions except for *rpo*C1-C2 and *atp*H, each of which has an intergenic spacer region. Therefore, we excepted to detect more site changes in these cases. However, the number of site changes in *rpo*C1-C2 was smaller despite the amplified fragment's considerable length. Fewer changes were detected in the other genes, *psb*A, *psb*D, *atp*H and 16S. The results, however, depended on the restriction enzymes used. As we could not easily compare differences between genes, we calculated the nucleotide subtitutions between species (Table 4) for each gene. The sequence divergence values ranged from 0.00% between *Shorea macrophylla* and *S. scaberrima* to 7.67% between *S. singkawang* and *Anisoptera oblonga*. The mean value was 3.83% (SD = 0.0627). All amplified DNA fragments were digested by 29 restriction endonucleases, averaging 10.64 restriction endonucleases per gene. As the fragment from the *atp*H gene was smaller (385 bp; ± 5%) than those from the other genes (939–3,603 bp), we detected only 10 sites using 10 restriction endonucleases. Ultimately, we obtained 485 fragments from digesting amplified fragments of the 11 genes: 62 fragments from *rbc*L, from *rpo*B, 64 from *pet*B, 29 from *psb*A, 39 from *psb*D, 38 from 16S, 93 from *rpo*C, 36 from *psa*A, 21 from *pet*A and 47 from *trn*K. The amplified region of *pet*B, *psb*A, *atp*H, *pet*A and *trn*K were relatively diverse among species, and the mean rate of nucleotide substitutions exceeded 6.00% (Table 3), with lower SD values than in other cases. The rates for *rbc*L and *psa*A were 4.05% (SD = 0.0229) and 4.05% (SD = 0.0237), respectively.

Phylogenetic trees

Twelve maximally parsimonious trees were obtained when the Wagner parsimony method was used, and a strict consensus tree was obtained from them (Fig. 2). For each tree, the consistency index was 0.675, the retention index 0.835 and the tree length was 209 steps. The NJ tree is shown in Fig. 3. Data set analysis produced similar basic patterns when both the Wagner parsimony and the NJ methods were applied to separate major taxomic groups (Fig. 2, 3). As the monotypic genus *Upuna* is primitive in several aspects (Ashton 1982), we used this genus as an outgroup when applying the Wagner parsimony method. Our results indicated that *Cotylelobium* and *Vatica* are closely related genera, and that *Shorea*, *Hopea*, *Parashorea* and *Neobalanocarpus* are also closely related genera forming a sister grouping to *Dryobalanops* and *Dipterocarpus*, with *Dryobalanops* lying between the first two (Fig. 2). In this phylogenetic tree, *Dipterocarpus*, *Dryobalanops* and *Anisoptera* are clearly separated from the other genera. *Parashorea* is the sister to a large group of *Shorea* species and to *Neobalanocarpus heimii*, *Hopea latifolia* and *H. dyeri*, which are closely related species. Most *Shorea* and *Hopea* species are clearly separated, but *Shorea* consists of two lineages: the *S. bracteolata* and *S. singkawang* group and all other *Shorea* species. *Dryobalanops* has the same basal lineage as *Shorea*, *Parashorea*, *Hopea* and *Neobalanocarpus*.

The NJ tree is similar to the strict consensus of the 12 most parsimonious Wagner trees except for the position of *S. multifolia* (Fig. 3). In the NJ tree, *Cotylelobium* forms a group with *Anisoptera* and *Vatica*, and the

Table 3 The mean estimates of nucleotide substitutions, the corresponding standard deviations, site change between species, the surveyed sequence and the sizes of the amplified fragments for each gene

Gene	Fragment size (bp)	Site change	Surveyed sequence[a] (bp)	Nucleotide substitution (%)	SD
*rbc*L	1,387	15	143 (10.3%)	4.05	0.0229
*rpo*B	1,286	17	124 (9.6%)	3.09	0.0197
*pet*B	1,634	23	133 (8.1%)	6.22	0.0476
*psb*A	939	3	79 (8.4%)	6.48	0.0096
*psb*D	1,042	8	85 (8.2%)	3.11	0.0254
*atp*H	385	3	24 (6.2%)	6.21	0.0682
16S	1,375	6	78 (5.7%)	1.87	0.0176
*rpo*C	3,603	33	220 (6.1%)	4.21	0.0321
*psa*A	2,218	8	86 (3.9%)	4.05	0.0237
*pet*A	2,314	8	46 (2.0%)	8.57	0.0790
*trn*K	2,569	17	108 (4.2%)	6.83	0.0390
Total	18,752	141	1,126 (6.0%)	3.83	0.0627

[a] Surveyed sequence means the surveyed number of nucleotide sequence. The value in parentheses is the percentage of the amplified gene sizes as inferred from restriction endonuclease analysis

Table 4 Estimated percentage of nucleotide substitutions (above diagonal) and the number of changes in restriction sites (below diagonol) for species of Dipterocarpaceae (as identified in Table 1)

	SMA	SAT	SBR	SSI	SOV	SMU	SSC	SKU	SPA	SMP	SLE	HHE	HSA	HLA	HSU	HOD	HWI	HDY	HAP	HNE	DIK	DIB	DIO	NEO	DRY	PAR	VAT	COT	ANI	UPU
SMA	–	1.23	2.39	3.39	0.56	2.25	0.30	0.43	0.82	0.30	0.56	2.93	2.81	3.32	2.23	2.65	2.67	3.12	2.94	2.36	5.62	5.15	5.29	2.94	3.88	2.52	6.01	5.13	6.20	5.78
SAT	11	–	2.31	3.61	1.54	2.55	1.54	1.68	2.08	1.54	1.81	3.05	2.92	3.44	2.34	2.77	2.78	2.94	3.06	2.47	5.35	4.88	5.02	3.06	3.91	2.73	6.38	5.28	6.35	5.93
SBR	17	18	–	1.38	2.42	3.29	2.42	2.47	2.69	2.42	2.70	2.81	2.40	2.82	2.11	2.25	2.26	2.32	2.54	2.24	6.38	6.20	6.24	2.82	4.37	3.06	6.45	5.76	6.53	6.20
SSI	25	28	10	–	3.71	3.92	3.42	3.47	3.68	3.42	3.70	3.53	3.12	3.83	3.10	2.96	2.98	3.43	3.25	3.23	7.18	7.00	7.04	3.74	5.12	4.07	7.36	6.65	7.67	7.11
SOV	3	13	19	29	–	2.57	0.26	0.38	0.77	0.26	0.51	2.96	2.84	3.35	2.26	2.69	2.70	3.15	2.97	2.39	5.85	5.37	5.52	2.97	3.91	2.37	6.35	5.46	6.53	6.11
SMU	16	20	24	30	19	–	2.29	2.43	2.83	2.29	2.56	2.58	2.17	3.26	2.44	2.31	2.04	3.05	2.59	2.57	5.07	4.90	5.04	3.26	3.93	3.40	6.96	6.14	7.15	6.38
SSC	1	13	19	27	2	17	–	0.13	0.51	0.00	0.25	2.96	2.84	3.35	2.26	2.69	2.70	3.15	2.97	2.39	5.96	5.48	5.62	2.97	3.91	2.46	6.35	5.46	6.53	6.11
SKU	1	14	20	28	3	18	1	–	0.64	0.13	0.38	3.11	2.99	3.50	2.41	2.83	2.85	3.30	3.12	2.54	6.13	5.64	5.79	3.12	4.07	2.60	6.52	5.63	6.70	6.28
SPA	5	17	21	29	6	21	4	5	–	0.51	0.25	2.66	3.10	3.61	2.52	2.95	2.96	3.41	3.23	2.65	5.91	5.74	5.89	3.52	3.88	2.72	6.30	5.73	6.48	6.27
SMP	1	13	19	27	2	17	0	1	4	–	0.25	2.96	2.84	3.35	2.26	2.69	2.70	3.15	2.97	2.39	5.96	5.48	5.62	2.97	3.91	2.46	6.35	5.46	6.53	6.11
SLE	3	15	21	29	4	19	2	3	2	2	–	2.67	3.11	3.62	2.53	2.96	2.98	3.43	3.25	2.66	6.25	5.77	5.91	3.25	4.19	2.73	6.64	5.75	6.83	6.40
HHE	21	23	21	27	22	19	22	23	20	22	20	–	1.19	2.23	1.18	0.52	0.79	2.03	0.79	1.05	5.69	5.83	5.87	1.96	4.23	3.41	7.26	6.77	7.23	6.79
HSA	20	22	18	24	21	16	21	22	23	21	23	9	–	2.10	1.05	0.66	0.39	1.90	0.92	1.18	5.25	5.39	5.44	2.11	3.81	3.29	6.71	6.33	6.79	6.35
HLA	24	26	22	30	25	24	25	26	27	25	27	17	16	–	1.27	1.69	1.69	0.75	1.96	1.40	5.48	5.62	5.66	1.87	4.63	3.70	6.82	6.02	6.80	6.36
HSU	16	18	16	24	17	18	17	18	19	17	19	9	8	10	–	0.65	0.66	1.06	0.92	0.13	4.92	4.76	4.80	1.28	3.50	2.71	6.45	5.66	6.32	6.00
HOD	19	21	17	23	20	17	20	21	22	20	22	4	5	13	5	–	0.26	1.48	0.26	0.52	5.39	5.53	5.58	1.42	3.65	3.14	6.63	6.15	6.61	6.17
HWI	19	21	17	23	20	15	20	21	22	20	22	6	3	13	5	2	–	1.48	0.52	0.79	5.11	5.25	5.29	1.70	3.67	3.15	6.66	6.18	6.64	6.21
HDY	22	22	18	26	23	22	23	24	25	23	25	15	14	6	8	11	11	–	1.75	1.19	5.50	5.65	5.69	1.66	4.44	3.51	6.86	6.05	6.84	6.40
HAP	21	23	19	25	22	19	22	23	24	22	24	6	7	15	7	2	4	13	–	0.78	5.71	5.85	5.89	1.69	3.95	3.42	6.96	6.47	6.93	6.50
HNE	17	19	17	26	18	20	18	19	20	18	20	9	10	12	2	5	7	10	7	–	5.06	4.89	4.93	1.14	3.63	2.84	6.59	5.80	6.46	6.14
DIK	42	41	47	54	44	40	44	45	44	44	46	43	40	44	38	41	39	42	43	38	–	0.39	0.52	5.91	4.84	5.81	5.66	5.30	5.43	5.95
DIB	39	38	46	52	41	38	41	42	43	41	43	43	40	44	36	41	39	42	43	38	4	–	0.13	5.74	4.98	5.64	5.81	4.83	5.57	6.09
DIO	41	40	48	54	43	40	43	44	45	43	45	45	42	46	38	43	41	44	45	38	4	2	–	5.79	5.12	5.68	5.95	4.97	5.72	6.23
NEO	22	24	22	30	23	24	23	24	27	23	25	15	16	16	10	11	13	14	13	10	46	44	46	–	4.24	3.61	7.17	6.04	7.14	6.71
DRY	31	33	35	43	32	33	32	33	32	32	34	34	31	37	29	30	30	35	32	29	37	39	39	35	–	4.37	5.53	5.07	5.30	5.30
PAR	17	19	23	31	18	23	18	19	20	18	20	24	23	27	19	22	22	25	24	21	45	43	45	27	36	–	5.47	4.72	5.55	5.34
VAT	45	49	49	57	48	51	48	49	48	48	50	54	51	53	49	50	50	51	52	49	43	45	45	55	44	46	–	1.53	2.20	2.02
COT	39	41	43	51	42	45	42	43	44	42	44	50	47	47	43	46	46	45	48	43	39	37	37	47	40	40	12	–	1.70	1.61
ANI	47	49	49	59	50	53	50	51	50	50	52	54	51	53	49	50	50	51	52	48	40	43	43	55	42	46	18	14	–	2.10
UPU	44	46	46	54	47	48	47	48	49	47	49	51	48	50	46	47	47	48	49	46	44	46	46	52	43	45	17	13	17	–

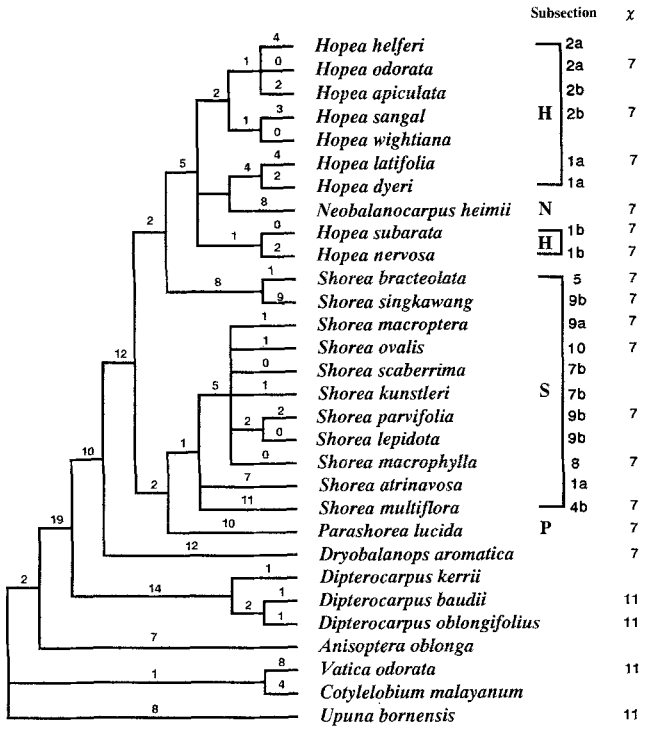


Fig. 2 Strict consensus tree of 12 maximally parsimonious Wagner trees from RFLP analysis of 11 PCR-amplified genes in cpDNA of Dipterocarpaceae. Subsection numbers in *Shorea* and *Hopea* are based on Ashton (1982). *S Shorea, P Parashorea, H Hopea* and *N Neobalanocarpus*. χ indicates basic chromosome numbers reported by Jong and Lethbridge (1967) and Jong (1976). The *numbers above* the branches indicate branch distances

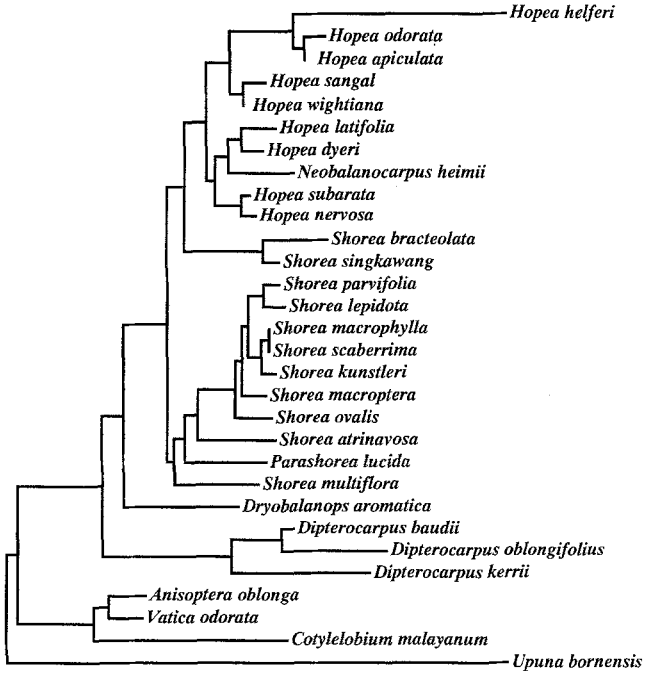


Fig. 3 Neighbor-joining tree constructed from RFLP analysis of 11 PCR-amplified genes in cpDNA of Dipterocarpaceae

position of *Parashorea* and some *Shorea* species are slightly different from a strict consensus of the 12 parsimonious Wagner trees.

Discussion

Evolution of Dipterocarpaceae in Southeast Asia

Ashton (1982) classified two tribes in Dipterocarpoideae. The first tribe, Dipterocarpeae, consist of *Vateria, Vateriopsis, Stemonoporus, Vatica, Cotylelobium, Upuna, Anithoptera* and *Dipterocarpus*. The first three of these genera occur in India and/or Sri Lanka. The other tribe, Shoreae, consist of *Dryobalanops, Parashorea, Hopea, Neobalanocarpus* and *Shorea*. These genera are distinguished by floral and wood anatomy characteristics (Desch 1941; Gottward and Parameswaran 1966) and a basic chromosome number of 7 (Shorea) or 11 (Dipterocarpeae) (Jong and Lethbridge 1967; Jong 1976; Somego 1978). Our results indicate the tribe Dipterocarpeae (*Vatica, Cotyleobium, Upuna, Anithoptera* and *Dipterocarpus* in our study) can be clearly differented from the group of the tribe Shorease including *Dryobalanops, Parashorea, Hopea, Neobalanocarpus* and *Shorea* (Figs. 2 and 3). Our data also identified Shoreae a monophyletic group and Dipterocarpeae as being paraphyletic. Therefore, Dipterocarpeae's characteristics can be considered as plesiomorphy. The data from our tree suggest that these are four natural taxa in the tribe Dipterocarpeae: *Dipterocarpus, Anithoptera, Vatica-Cotylelobium* and *Upuna*. Ashton (1982) suggested that there is a close affinity between most of the morphological traits of *Vatica* and *Cotylelobium* which is also consistent with our tree. The monotipic genus *Upuna*, which is primitive in several respects, is endemic to Borneo, especially with respect to its wood anatomy (Ashton 1982). Of the 267 most abundant species of Dipterocarpeae that have been identified within Southeast Asia, 155 (59%) of them are endemic to Borneo, with the Malaysian Peninsula and Sumatora following Borneo, with 127 (19% endemic) and 95 (11.5% endemic) species, respectively (Ashton 1982). This might suggest that Borneo and the surrounding regions are the center of diversity of this family after the ancestors reached Southeast Asia.

According to our cpDNA-based tree, the tribe Shoreae is a monophyletic taxa. Therefore, a series of characteristic changes, including a change in the basic chromosome number from $x = 11$ to $x = 7$, must have occured at the base of this taxon. *Dryobalanops* is situated outside the other genus, a position consistent with the wood anatomy data indicating the intermediacy of *Diterocarpus* between Shoreae and Dipterocarpeae as reviewed and classfied by Ashton (1982). *Dryobalanops* has five equally long calyx lobes, as does *Vatica* sect. *Vatica*, but other Dipterocarpaceae genera have two or three aliform fruits. Our results show that the evolution of the long calyx lobes occurred at the clade of *Dryobalanops* and independent of *Vatica* sect. *Vatica*.

Our results reveal a close affinity among *Hopea, Neobalanocarpus, Parashorea* and *Shorea* with no im-

portant morphological differences among them. Two large genera, *Hopea* and *Shorea*, can be distinguished by a single characteristic, number of long fruit calyxes: *Hopea* has two and *Shorea* three. In our data, *Hopea* and *Shorea* were separate, with only two exceptions, *S. bracteolata* and *S. singkawang*. These 2 species combined with *Hopea* not *Shorea*. *Neobalanocarpus*, a monotypic genus composed of *N. heimii*, was identified in 1978 by Ashton and is characterized by short, equal fruit sepals and linear anthers. Ashton (1982) studied its close affinity with *Hopea* sect. *Hopea* based on the appearance of the inflorescence, fruit embryo and germination mode. But our tree shows *Neobalanocarpus* with a closer affinity to *Hopea* sect. *Dryobalanoides* subsect. *Dryobalanoides* (subsect. 1a in Fig. 2) than to sect. *Hopea* (subsect. 2a,b in Fig. 2). Thus, it is at least certain that *Neobalanocarpus* nests in *Hopea* in our cpDNA-based analysis.

Dryobalanops and *Parashorea* (with about 10 species) are both characterized by an unequal aliform fruit that exhibits globose, verrucose and plicate vernation. However, other characteristics are very similar to *Shorea*. Ashton (1982) regarded the difference as that of section level in embryo and seeding characteristics. Data from our tree indicate that *Parashorea* is a very close relative of *Shorea*, but as a group, it is placed mostly outside of the *Shorea* species except for 2 species that are strangely combined with *Hopea*. Therefore, *Parashorea* is considered a sister genus of *Shorea*.

Hopea and *Shorea* contain 186 and 357 species, respectively (Ashton 1982). Our molecular phylogenetic tree showed that speciation between *Hopea* and *Shorea*, including *Neobalanocarpus* and *Parashorea*, occurred relatively recently because in our tree the branch distances of these genera are small. Although it is difficult to estimate relationships when small genetic divergence occurs, several clusters are recognizable. *Hopea* consists of two sections and four subsections, sect. *Dryobalanoides* subsect. *Dryobalanoides* (shown as subsect. 1a in Fig. 2) and subsect. *Sphaerocarpae* (1b) and sect. *Hopea* subsect. *Hopea* (2a) and subsect. *Pierra* (2b), as reviewed and classified by Ashton (1982). He seems to doubt that subsections are well-defined and might sometimes include heterogenous species. We recognized four natural clades (Fig. 2) corresponding to the classical subsections described by Ashton (1982). Although an exception, *Hopea apiculata*, combines with *H. helferi* and *H. odorata* to form a group with subsect. *Hopea* (2a) despite the existence of subsect. *Pierra* (2b). The other species in *Hopea* match Ashton's (1982) classification. The relationship among subsections is not clear from our data. However, there are ten sections and four subsections in the Malaysian *Shorea*. The infrageneric relationships of *Shorea* are unclear since there are few synapomorphic states. Classical assignments of subsections or sections correspond poorly to our cpDNA-based tree. *Shorea* bracteolata and *S. singkawang* are closer to *Hopea* than to other *Shorea*. Our data shows *Shorea* might be heterogenous and should be reexamined. Due to an insufficient number of site change we could not estimate infrageneric relationships in this study, but if more amplified genes and restriction enzymes were used in future studies, the analysis of infrageneric relationships would be possible.

Which cpDNA genes are useful for studying molecular phylogeny?

We used RFLP of 11 cpDNA genes to construct molecular phylogenetic trees. These genes are mostly conserved among plant species. The *rbc*L, *psb*A, *psb*D, 16S, *trn*K and *psa*A genes are, in general, highly conserved among plant species, and the homology of these genes between liverwort and rice exceeds 90% (Shimada and Sugiura 1991). Since the homology between *rpo*B from liverwort and rice is 64% (Shimada and Sugiura 1991), we would expect more site changes in *rpo*B than in the more highly conserved genes. In fact, we detected 17 site changes in the *rpo*B region. The *pet*B gene usually contains an intron, for example, similar to that found in the *pet*B genes of rice, tobacco and liverwort. Thus, this region was also expected to contain more site changes and, indeed, we found 23. The *rpo*C region contains a spacer region between the *rpo*C1 and *rpo*C2 genes, and the *rpo*C1 gene also contains an intron. The homology of *rpo*C1 and *rpo*C2 genes between rice and liverwort is not high, 62% and 47%, respectively (Shimada and Sugiura 1991). We detected 33 differences among our dipterocarps. The *atp*H region, containing the *atp*H and *atp*I genes, usually has a spacer region. For example, its length is 794 bp in rice, 1,158 bp in tobacco and 377 bp in liverwort (Shimada and Sugiura 1991). However, our amplified *atp*H fragment was 385 bp long, including the presence of only the *atp*H and *atp*I genes. In rice, tobacco and liverwort the *atp*H and *atp*I gene coding regions are 81 bp and 247–248 bp long, respectively. Therefore, the *atp*H region in dipterocarps lacks a large spacer region between genes. We estimated the spacer region in dipterocarps to be less than 100 bp long, and we detected only 3 site changes. In rice, tobacco and liverwort the *trn*K gene has a large intron that is 2,504 bp, 2,526 bp and 2,111 bp long, respectively. Because the dipterocarp *trn*K gene also has a large intron (about 2,500 bp), we therefore expected to find many variations between species – but detected only 17.

Sequence divergence was estimated for comparing the divergence of genes between species. The *rbc*L, *pet*B, *psb*A and *rpo*C genes were suitable for molecular phylogeny analysis, as were *psa*A and *trn*K, because the sequence divergence values were relatively high and the SD values were lower than those of the other amplified genes (Table 3). However, in *psb*A and *psa*A, there were only 3 and 8 site changes, respectively. Consequently, based on our results, the *rpo*C gene was the most informative with *rbc*L, *pet*B and *trn*K also being suitable for molecular phylogeny analysis. The region containing an intron and a spacer region, like the *rpo*C region,

was also informative. Liston (1992) used RFLP data from this *rpo*C region to construct a molecular phylogenetic tree within the genus *Astragalus* (Fabaceae). The PCR-RFLP method for studying molecular phylogeny allows us to survey many genes relatively easily. RFLP analysis of cpDNA using Southern hybridization is also useful, but when this protocol is used one sometimes encounters difficulty in assigning site changes. While sequencing is the most effective method for studying a molecular phylogeny, it currently is difficult to investigate multiple-gene sequences in many species because of high costs and labor intensity. Thus, most phylogenetic studies sequencing have involved only a very few genes.

Our molecular phylogenetic tree was based on cpDNA only. Therefore, we think that if other genome mitochondrial DNA and nuclear DNA were also investigated and morphological data included in the phylogenetic tree construction, then we might be able to describe the species phylogeny more precisely. A cpDNA-based phylogenetic tree, like that developed in this study, might exhibit some bias because cpDNA is only one of the organelle DNAs and is usually inherited uniparentally.

Acknowledgements The authors thank the staff of the Forest Research Institute of Malaysia (FRIM) for collecting materials from the FRIM dipterocarp arboretum and to Dr. S. Kobayashi for helpful comments about dipterocarp phylogeny. The authors also thank Drs. S. H. Strauss and A. Liston for helpful comments and criticism of the manuscript and Ms. M. Koshiba for technical assistance with experiments. The present study is part of a joint project between the Forest Research Institute of Malaysia, University Pertanian Malaysia and the National Institute for Environmental Studies of Japan. The study was supported by grant No. E-4 under the Global Environment Research Program of the Japan Environment Agency.

References

Ashton PS (1969) Speciation among tropical forest trees: some deductions in the light of recent evidence. Bot J Linn Soc 1:155–196

Ashton PS (1982) Flora Malesiana. Series I-Spermatophyta. Flowering plants, vol. 9, part 2, Dipterocarpaceae. Martinun Nijhoff Pub, The Netherlands

Ashton PS, Gan Y-Y, Robertson FW (1984) Electrophoretical and morphological comparisons in ten rain forest species of Shorea (Dipterocarpaceae). Bot J Linn Soc 89:293–304

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu Y-L, Kron KA, Retting JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clarl WD, Hedren M, Gaut BS, Jansen RK, Kim K-J, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang Q, Plunkett GM, Soltis PS, Swensen SM, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn GH, Graham SW, Barrett SCH, Dayanandan S, Albert VA (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Ann Mo Bot Gard 80:528–580

Desch HE (1941) Manual of Malayan timbers. Malayan Forest Records No 15, vol. I, vol. II

Fedorov AA (1966) The structure of the tropical rain forest and speciation in the humid tropics. J Ecol 54:1–11

Gan Y-Y, Robertson FW (1989) Isozyme variation in some rain forest trees. Biotropica 13:20–28

Gan Y-Y, Robertson FW, Ashton PS, Soepadmo E, Lee DW (1977) Genetic variation in wild populations of rain-forest trees. Nature 269:323–325

Gottwald H, Parameswaran N (1966) Das sekundare Xylem der Familie Dipterocarpaceae, antomische Untersuchungen zur Taxonomie und Phylogenie. Bot Jahrb 85:410–508

Harada K, Kinoshita A, Shukor NAA, Tachida H, Yamazaki T (1994) Genetic variation estimated in three *Shorea* species by the RAPD analysis. Jpn J Genet 69:713–718

Henning J, Herrmann RG (1986) Chloroplast ATP synthase of spinach contains nine nonidentical subunit species, six of which are encoded by plastid chromosomes in two operons in a phylogenetically conserved arrangement. Mol Gen Genet 203:117–128

Hipkins VD, Tsai CH, Strauss SH (1990) Sequence of the gene for large subunit of ribulose 1,5-bisphosphate carboxylase from a gymnosperm, Douglas fir. Plant Mol Biol 15:505–507

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Kondo C, Honji Y, Sun CR, Meng BY, Li YQ, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of cereals. Mol Gen Genet 217:185–194

Ihara M, Gadrinab LU, Siregar UJ, Iyama S (1986) Genetic control of alcohol dehydrogenese and estimation of some population parameters in *Hopea odorata* Roxb. (Dipterocarpaceae). Jpn J Genet 61:127–136

Jong K (1976) Cytology of the Dipterocarpaceae. In: Burly J, Styles BT (eds) Tropical tree: variation, breeding and conservation. Linnean Soc Symp Ser 2:79–84

Jong K, Lethbridge A (1967) Cytological studies in the Dipterocarpaceae. I: Chromosome number of certain Malaysian genera. Notes R Bot Gard Edinburgh 27:175–184

Kitamura K, Rahman MYBA, Ochai Y, Yoshimaru H (1994) Estimate of the outcrossing rate on *Dryobalanops aromatica* Gaertn. F. in primary and secondary forests in Burnei, Borneo, Southeast Asia. Plant Species Biol 9:37–41

Lidholm JA, Gustafsson P (1991) A three-step model for the rearrangement of the chloroplast trnK-*psb*A region of gymnosperm *Pinus contrta*. Nucleic Acids Res 19:2881–2887

Liston A (1992) Variation from restriction chloroplast genes *rpo* C1 and *rpo* C2 of the genus *Astragalus* (Fabaceae): Evidence from restriction site mapping of a PCR-amplified fragment. Am J Bot 79:953–961

Murawsaki PA, Banea KS (1994) Genetic structure and mating system of *Stemonoporous oblongifolius* (Dipterocarpaceae) in Sri Lanka. Am J Bot 81:155–160

Murray M, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

Nei M, Tajima F (1983) Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics 105:207–217

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Shimada H, Sugiura M (1991) Fine structural features of the chloroplast genome comparison of the sequenced chloroplast genomes. Nucleic Acids Res 19:983–995

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohta C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043–2049

Somega M (1978) Cytogenetical study of the Dipterocarpaceae. Malayan For 41:358–365

Symington CF (1943) Malayan Forest Records No. 16. Foresters' manual of Dipterocarps. (Reprinted with plates and historical introduction, University of Malaya Press, Kuala Lumpur, 1074)

Swofford DL (1991) PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0s. Computer program distributed by the Illinois Natural History Survey, Champaign, Ill.

Tsumura Y, Yoshimura K, Tomaru N, Ohba K (1995) Molecular phylogeny of conifers using RFLP analysis of PCR-amplified specific chloraplast genes. Theor Appl Genet 91:1222–1236

Whitmore TC (1984) Tropical rain forests of the Far East, 2nd edn. Oxford University Press, Oxford