— **Proceeding** —
Nucleotide variation in the *GapC* region of four species of *Shorea* and their putative hybrids

Hiroko ISHIYAMA¹⁾, Tomoyuki KADO¹⁾, Mayumi IWASAKI¹⁾, Mieko MATSUOKA¹⁾, N. Ab SHUKOR²⁾, Alfred E. SZMIDT¹⁾
and Tsuneyuki YAMAZAKI^{1, 2, 3)}

¹⁾Department of Biology, Faculty of Sciences, Kyushu University, 10-1, Hakozaki 6, Higashi-ku, Fukuoka 812-8581, Japan.

²⁾Faculty of Forestry, University, Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia.

³⁾To whom reprint request should be addressed: Department of Biology, Faculty of Sciences, Kyushu University, Fukuoka, 10-1 Hakozaki 6, Higashi-ku, Fukuoka 812-8581, Japan TEL: +81-92-642-2614 E-mail: aszmiscb@mbox.nc.kyushu-u.ac.jp

⁴⁾Corresponding author

⁵⁾Present address: Research Institute of Evolutionary Biology, 4-28, Kamiyoga 2, Setagaya, Tokyo 158-0098, Japan.
E-mail: tyamascb@mbox.nc.kyushu-u.ac.jp

ABSTRACT Very little is known about the levels and patterns of DNA polymorphism in populations of tropical trees. In this study, we investigated population genetic structure of the four following *Shorea* species: *Shorea acuminata*, *S. curtisii*, *S. leprosula* and *S. parvifolia* from peninsular Malaysia using DNA sequences of Glyceraldehyde-3-phosphate dehydrogenase nuclear gene (*GapC*). The average nucleotide diversities of the investigated species for silent sites ($\pi_{\text{sil}} = 0.0027$) were lower than those reported for herbaceous plants, and similar or somewhat lower than those of other woody species. These low levels of nucleotide diversities may be caused by small effective population size of *Shorea* species. We found that some individuals harbored atypical haplotypes consisting of sequences from two different species. The segmented architecture of these haplotypes suggests past inter-specific gene exchange and fertility of the hybrids. As demonstrated in this study analysis of DNA sequences provides genetic evidence for the occurrence of inter-specific hybridization and gives important insights into its evolutionary effects.

Key words: *Shorea*, *Mutica*, Dipterocarpaceae, *GapC*, DNA polymorphism, hybrids

INTRODUCTION

Shorea acuminata, *S. curtisii*, *S. leprosula* and *S. parvifolia* are members of the genus *Shorea* section *Mutica* in the family Dipterocarpaceae. The family is dominant in lowland rain forest and forms mixed dipterocarp stands. The four *Shorea* species investigated in our study are distributed in peninsular Malaysia and Indonesia. Except for *S. acuminata*, they also occur on Borneo Island. They are predominantly outcrossing and their pollinators are mainly thrips and beetles, which can migrate only short distances (Appanah & Chan, 1981; Sakai *et al.*, 1999). They are major sources of timber for the hardwood market. For sustainable forestry, it is necessary to conserve genetic polymorphism of the species and to know genetic makeup of the natural forest. Genetic variation of some *Shorea* species was investigated using molecular markers. Lee *et al.* (2000) showed high heterozygosity in *Shorea leprosula* using allozyme markers. *Shorea curtisii* and *S. leprosula* were studied using microsatellites showing high (>90%) outcrossing rate (Lee *et al.*, 2000; Nagamitsu *et al.*, 2001; Obayashi *et al.*, 2002). Harada *et al.* (1994) investigated *S. acuminata*, *S. curtisii* and *S. leprosula* using RAPD markers and found low inter-specific nucleotide divergence and high within species diversity, suggesting that the diversification of the genus *Shorea* occurred only recently and/or that there is considerable gene exchange among species.

DNA sequence data provide information about past population structure of the species as well as evolutionary forces shaping DNA polymorphism. So far, most research on DNA sequence polymorphism in plants has been conducted on herbaceous species with very short generation times. Little is known about the DNA polymorphism and divergence in long-lived woody perennials (*Pinus sylvestris* (Dvornyk *et al.*, 2002; García-Gil *et al.*, 2003), *Betula pendula* (Järvinen *et al.*, 2003), *Cryptomeria japonica* (Kado *et al.*, 2003)). To our knowledge, there have been no such studies on tropical trees.

In this study, the partial region of the *GapC* gene was sequenced to survey the levels and patterns of DNA polymorphism and divergence in four *Shorea* species. This is the first report on genetic polymorphism using nuclear gene sequence in *Shorea* species.

MATERIALS AND METHODS

Plant materials

Leaf samples were collected from three populations of *Shorea acuminata*, three populations of *S. curtisii*, one population of *S. leprosula*, and one population of *S. parvifolia* in different parts of peninsular Malaysia (Table 1). Species identification in the field was done by a professional taxonomist. Herbarium specimens were examined again by an independent taxonomist. The sequences of the *GapC* gene were determined for 12 to 16 individuals from each population.

Table 1. List of populations analyzed in this study.

Population (Individual number)	Species	Sample size	Locality
Sa-1 (80-93)	<i>S. acuminata</i>	14	Mersing, Johor
Sa-2 (394, 398, 401-404)	<i>S. acuminata</i>	6	44 km from Mersing to Keluang, Johor
Sa-4 (341-344, 346-350, 352)	<i>S. acuminata</i>	10	Seremban, Negeri Sembilan
Sc-1 (1-12)	<i>S. curtisii</i>	12	Mersing, Johor
Sc-2 (100, 102, 104, 106-111)	<i>S. curtisii</i>	9	57 km from Mersing to Keluang, Johor
Sc-3 (260-273, 275)	<i>S. curtisii</i>	15	Port Dickson, Negeri Sembilan
Sl-4 (312-323, 325-326)	<i>S. leprosula</i>	14	Seremban, Negeri Sembilan
Sp-4 (281-293, 296)	<i>S. parvifolia</i>	14	Seremban, Negeri Sembilan

Molecular methods

Genomic DNA was isolated from ~300mg of leaves with a modified CTAB method (Murray & Thompson, 1980). A partial region of the Glyceraldehyde-3-phosphate dehydrogenase nuclear gene (*GapC*) was amplified by PCR. The forward and reverse primers for PCR were designed using the *GapC* region from *Arabidopsis thaliana* sequenced by Shin *et al.* (1991). In addition, a set of primers for PCR (5' ATTATTAGTTTTTGCTTTACAGG 3' and 5' CAATCTTTACCTGCTATCAC 3') was designed from the sequences of the *GapC* gene of *S. acuminata*. Subsequently, PCR products were excised from agarose gel and purified by GENECLAN III kit (BIO101).

The PCR products were sequenced using ABI Prism 377 automatic sequencer (Applied Biosystems). For sequencing three additional internal primers (5' CAGGCATCAGAGCTTCATCCAAAAC 3', 5' CATCATGGTGAAATTAACCTCCTA 3', 5' GGAACGGACCCTGCTTACTACAA 3') were designed based on the *GapC* sequence from *S. acuminata*. The PCR products from individuals sa85, sa86, sa347, sa350, sc275, sl326, and sp296, which had many heterozygous sites and sp282, sp286, sp289, sp292, which had heterozygous insertion/deletions (indels), were cloned into the pGEM T-easy vector (Promega) and then sequenced. The PCR products from other individuals were directly sequenced and analyzed without haplotype phase information.

Data analysis

The sequences were aligned using the ClustalX program (Thomson *et al.*, 1997), and corrected manually. The gene tree was constructed using maximum parsimony method with the PAUP version 4.0b10 program (Swofford, 1991). To assess genetic variation of populations the number of pairwise differences between sequences (per base estimates of nucleotide diversity, π ; Nei (1987), equations 10.5 or 10.6) and the nucleotide polymorphism (θ w; Watterson (1975)) were estimated. The extent of DNA divergence between species was estimated as the average number of nucleotide substitutions per site between pairs of populations (D_{xy}), and the number of net nucleotide substitutions per site between pairs of populations (D_a). Data analyses were performed using DnaSP version 3.51 program (Rozas & Rozas, 1999).

RESULTS

DNA sequences in the *GapC* region of four *Shorea* species

The partial *GapC* region sequenced for 94 individuals was 756-814bp in length. It contained four exons and three introns (Fig 1). The sequences of each species were distinguishable, because they had specific indels or fixed mutations (nucleotide polymorphisms) (Fig 1). *Shorea acuminata* had fixed mutations at positions 172, 261, 339 and 416. *Shorea curtisii* had fixed mutations at positions 218, 603 and 782. *Shorea leprosula* had fixed mutations at positions 43, 103, 160, 265, 405, 582, 797, fixed large deletion at positions 276-346 and fixed insertion at positions 347-359. *Shorea parvifolia* had fixed mutations at positions 5, 198, 294, 309, 333, 388, 432, 803 and fixed single-base deletion at position 285.

DNA polymorphism within species

The DNA polymorphism of the investigated species is summarized in Table 2. The average nucleotide diversities (π) and the nucleotide polymorphism (θw) of the four investigated *Shorea* species for silent sites were 0.0027 and 0.0034 respectively.

The nucleotide diversities and polymorphisms within species for the whole sequenced region were highest in *S. curtisii* ($\pi = 0.0029$, $\theta w = 0.0039$), and lowest in *S. leprosula* ($\pi = 0.0013$, $\theta w = 0.0017$). Replacement polymorphisms were rarely found. Only individual sa82 had replacement of threonine with proline and another individual sl316 had replacement of serine with threonine.

Variation between species

Estimates of divergence between species are presented in Table 3. The average number of nucleotide substitutions per site between pairs of populations (D_{xy}) and the number of net nucleotide substitutions per site between pairs of populations (D_a ; which does not include genetic diversity within species) were calculated. These estimators show the distance between two species. The largest divergence was found between *S. leprosula* and *S. parvifolia* ($D_a = 0.0172$), while the smallest divergence was found between *S. acuminata* and *S. curtisii* ($D_a = 0.0093$). Nucleotide diversity among species was clearly higher than that within species. The average number of nucleotide substitutions per site between each pair of species (D_{xy}) was four to 12 times higher than the average number of pairwise differences per site within species (π) (Table 3). On the gene tree, haplotypes of the four *Shorea* species formed clearly separated clades (Fig 2).

Individuals with unusual haplotype composition

In most individuals the number of heterozygous sites in the *GapC* region ranged from zero to three. However, seven individuals had many (seven to 17) heterozygous sites (Fig 1). Haplotypes of these individuals were determined by subcloning. We found that the haplotype compositions of these individuals did not conform to their taxonomic identification based on morphology. One of their haplotypes differed from typical haplotypes of the corresponding species. Here, typical haplotypes means the sequences placed in the same clade on the gene tree, which differed by zero to three base pairs (Fig 2). The atypical haplotypes were placed outside the corresponding clade and differed by more than seven base pairs from the typical haplotypes. The sequences of these atypical haplotypes had a segmented structure where individual segments (more than 100 bp long) resembled typical sequences of different species. The positions of these atypical haplotypes in the gene tree are marked in Fig 2. Note that six of the seven individuals had one atypical haplotype and one typical haplotype (Fig 2). Only in the individual sl326, both haplotypes did not conform to morphological identification.

An individual sa85 had one atypical haplotype containing two segments, the positions 25-200 resembled *S. acuminata* sequences while the positions 225-800 resembled *S. leprosula* sequences. Moreover, this haplotype harbored *S. leprosula* specific large indel (Fig 2, Fig 3a). Three *S. acuminata* individuals (sa86, sa347, sa350) had one atypical haplotype, which differed by seven bp from the most common haplotype observed in *S. acuminata* (Fig 2). The positions 275-500 of this haplotype resembled *S. curtisii* sequences, positions 450-575 resembled *S. leprosula* sequences and the remaining positions did not resemble any of the investigated species (Fig 3b). One *S. curtisii* individual (sc275) had one atypical haplotype, whose positions 400-700 resembled *S. parvifolia* sequences and the remaining positions did not resemble any of the investigated species, (Fig 3c). One *S. parvifolia* individual (sp296) and one *S. leprosula* individual (sl326) had the same haplotype composition, where one haplotype was typical sequence of *S. parvifolia* and the other haplotype was atypical (Fig 2). The positions 25-175 and 550-700 of this atypical haplotype resembled *S. acuminata* and the other positions did not resemble any

Gene structure		Exon										Exon										No. of heterozygous sites	
Position	Category	Consensus																				sites	
Consensus		CCCGTCAGGACATCTGAAGGCGCTTGAAGGGCCGATATCGG-TCTCAGCGAGTGCTGTTGTTATTTCAAA																					
sa80																						2	
sa81																						1	
sa82																						1	
sa83																						0	
sa84																						0	
sa85-hap1																						11	
sa85-hap2																						7	
sa86-hap1																						0	
sa86-hap2																						2	
sa87																						2	
sa88																						1	
sa89																						0	
sa90																						2	
sa91																						1	
sa92																						0	
sa93																						2	
sa394																						2	
sa398																						2	
sa401																						0	
sa402																						0	
sa403																						0	
sa404																						2	
sa341																						0	
sa342																						3	
sa343																						0	
sa344																						3	
sa346																						3	
sa347-hap1																						8	
sa347-hap2																						3	
sa348																						2	
sa349																						7	
sa350-hap1																						1	
sa350-hap2																						1	
sa352																						0	
sc01																						3	
sc02																						0	
sc03																						3	
sc04																						0	
sc05																						3	
sc06																						0	
sc07																						3	
sc08																						0	
sc09																						1	
sc10																						0	
sc11																						3	
sc12																						2	
sc100																						1	
sc102																						1	
sc104																						1	
sc106																						3	
sc107																						2	
sc108																						0	
sc109																						2	
sc111																						0	
sc260																						4	
sc261																						2	
sc262																						5	
sc263																						3	
sc264																						2	
sc265																						3	
sc266																						3	
sc267																						4	
sc268																						0	
sc269																						2	
sc270																						2	
sc271																						0	
sc272																						3	
sc273																						17	
sc275-hap1																						0	
sc275-hap2																						0	
si312																						0	
si313																						1	
si314																						0	
si315																						1	
si316																						1	
si317																						1	
si318																						1	
si319																						1	
si320																						0	
si321																						1	
si322																						0	
si323																						2	
si323																						1	
si326-hap1																						13	
si326-hap2																						2	
sp281																						3	
sp282-hap1																						1	
sp282-hap2																						1	
sp283																						1	
sp284																						1	
sp285																						1	
sp286-hap1																						3	
sp286-hap2																						0	
sp287																						1	
sp288																						5	
sp288-hap1																						0	
sp288-hap2																						0	
sp289																						3	
sp291																						1	
sp292-hap1																						0	
sp292-hap2																						3	
sp293																						1	
sp296-hap1																						13	
sp296-hap2																						0	
sp296-hap1																						13	

Fig. 1. Polymorphic sites in the partial *GapC* region of the four *Shorea* species.

Each site is categorized in the upper row as; F: fixed mutation or indel in one species, P: polymorphic site in one species, P*: singleton, S: shared polymorphic site, X: mutation specific to atypical sequences. Both haplotypes are shown when they were determined by subcloning. The atypical sequences are shaded. Dots indicate sites identical with consensus sequence and Y, M, K, R, S, W, -, + indicate C or T, A or C, T or G, A or G, C or G, A or T, deletion and insertion, respectively. The sequence of insertion from position 347 is CTGCTCATTTATG.

Fig. 2. Maximum parsimony tree based on the partial *GapC* sequences of four *Shorea* species. The atypical haplotypes are shaded.

(a) sa85-hap2

(b) sa86-hap2
(or sa347-hap2, or sa350-hap2)

(c) sc275-hap2

(d) sp296-hap2
(or sl326-hap2) **Fig. 3.** *Da* between atypical and typical haplotypes of each species. Window size is 100 bp, and step size is 25 bp

Table 2. DNA polymorphisms within species

Species	No. of sequences	S^*	Total		Intron		Silent (Intron & synonymous)		Synonymous		Non-synonymous	
			814bp		544bp		611.83bp		67.83bp		199.17bp	
			π	θ_w	π_{int}	θ_{w-int}	π_{sil}	θ_{w-sil}	π_{syn}	θ_{w-syn}	π_a	θ_{w-a}
<i>S. acuminata</i>	52	7	0.0016	0.0019	0.0018	0.0012	0.0020	0.0022	0.0046	0.0098	0.0002	0.0011
<i>S. curtisii</i>	68	15	0.0029	0.0039	0.0030	0.0042	0.0039	0.0051	0.0108	0.0123	0.0000	0.0000
<i>S. leprosula</i>	26	5	0.0013	0.0017	0.0012	0.0016	0.0016	0.0019	0.0040	0.0039	0.0004	0.0013
<i>S. parvifolia</i>	24	10	0.0026	0.0033	0.0028	0.0039	0.0034	0.0043	0.0076	0.0077	0.0000	0.0000

* S is the number of segregating sites**Table 3. The average number of nucleotide substitutions per site between species (D_{xy}), the number of net nucleotide substitution per site between pairs of populations (D_a) and the number of pairwise differences between sequences within species (π).**

Species	<i>S. acuminata</i>	<i>S. curtisii</i>	<i>S. leprosula</i>	<i>S. parvifolia</i>
<i>S. acuminata</i>	0.0016	0.0115	0.0161	0.0171
<i>S. curtisii</i>	0.0093	0.0029	0.0161	0.0162
<i>S. leprosula</i>	0.0149	0.0144	0.0013	0.0192
<i>S. parvifolia</i>	0.0150	0.0135	0.0172	0.0026

 D_{xy} : above the diagonal, D_a : below the diagonal, π : boxed**Table 4. Comparison of the levels of nucleotide polymorphism in various plants.**

Species		gene	Total π	Silent π_s	Replacement π_a	Source
woody plants	<i>S. acuminata</i> (Dipterocarpaceae)	<i>GapC</i>	0.0016	0.0020	0.0002	This study
	<i>S. curtisii</i> (Dipterocarpaceae)	<i>GapC</i>	0.0029	0.0039	0.0000	This study
	<i>S. leprosula</i> (Dipterocarpaceae)	<i>GapC</i>	0.0013	0.0016	0.0004	This study
	<i>S. parvifolia</i> (Dipterocarpaceae)	<i>GapC</i>	0.0026	0.0034	0.0000	This study
	Average for four <i>Shorea</i>	<i>GapC</i>	0.0021	0.0027	0.0002	This study
	<i>Pinus sylvestris</i> (Pinaceae)	<i>GapC</i>	0.0058	0.0078	0.0002	Szmidt, unpublished
	<i>Pinus sylvestris</i> (Pinaceae)	<i>Gln1a</i>	0.0018	0.0024	0.0000	Szmidt, unpublished
	<i>Pinus sylvestris</i> (Pinaceae)	<i>CHS</i>	0.0089	0.0252	0.0015	Szmidt, unpublished
	<i>Pinus sylvestris</i> (Pinaceae)	<i>pal1</i>	0.0014	0.0049*	0.0003	Dvornyk <i>et al.</i> , 2002
	<i>Pinus sylvestris</i> (Pinaceae)	<i>PHYP</i>	0.0002	0.0020*	0.0003	García-Gil <i>et al.</i> , 2003
	<i>Betula pendula</i> (Betulaceae)	<i>BpMADS2</i>	-	0.0043*	0.0001	Järvinen <i>et al.</i> , 2003
	<i>Cryptomeria japonica</i> (Taxodiaceae)	**	-	0.0038	0.0008	Kado <i>et al.</i> , 2003
	<i>Dipterocarpus alatus</i> (Dipterocarpaceae)	<i>PgiC</i>	0.0088	0.0108	0.0027	Tsuida, unpublished
herbaceous plants	<i>Arabidopsis thaliana</i> (Brassicaceae)	<i>PgiC</i>	0.0038	0.0051*	0.0016	Kawabe <i>et al.</i> , 2000
	<i>Arabidopsis thaliana</i> (Brassicaceae)	<i>Adh</i>	0.0080	0.0192*	0.0022	Innan <i>et al.</i> , 1996
	<i>Leavenworthia stylosa</i> (Brassicaceae)	<i>PgiC</i>	0.0440	0.0554	0.0062	Filatov & Charlesworth, 1999
	<i>Leavenworthia stylosa</i> (Brassicaceae)	<i>Adh1</i>	-	0.0307	-	Liu <i>et al.</i> 1998
	<i>Dioscorea tokoro</i> (Dioscoreaceae)	<i>PgiC</i>	-	0.0015*	0.0010	Terauchi <i>et al.</i> , 1997
	<i>Dioscorea tokoro</i> (Dioscoreaceae)	<i>Adh1</i>	0.0040	0.0059	-	Terauchi <i>et al.</i> , 1997
	<i>Zea mays ssp. parviglumis</i> (Poaceae)	<i>Adh1</i>	-	0.0270	0.0013	Zhang <i>et al.</i> , 2002
	<i>Zea mays ssp. parviglumis</i> (Poaceae)	<i>hm1</i>	-	0.0414	0.0204	Zhang <i>et al.</i> , 2002

* π for synonymous sites** average for seven loci: *Ad1*, *Chi1*, *Ferr*, *GapC*, *HemA*, *Lcyb*, *Pat*

DISCUSSION

Low levels of DNA polymorphism within species

The average nucleotide diversity of the four investigated *Shorea* species for silent sites ($\pi_{sil} = 0.0027$) was lower than that reported for both herbaceous species with a very short generation time and for other woody species (Table 4). The nucleotide diversity (π) is an estimator for $4N\mu$ where N is the effective population size and μ is mutation rate. Thus, the small effective population size and/or low mutation rate per generation may be responsible for the observed low levels of nucleotide diversity. In the absence of information about the mutation rates in *Shorea* it is difficult to decide how much low mutation rate contributed to the low levels of polymorphism revealed in this study. On the other hand, some insights about the effective population size could be gleaned from the ecological characteristics of *Shorea* species. First, *Shorea* species are predominantly outcrossing and they are partially self-incompatible (Chan, 1981; Nagamitsu *et al.*, 2001; Obayashi *et al.*, 2002; Sakai *et al.*, 1999). However, their pollinators are mainly thrips and beetles, which can migrate only limited distance (Appanah & Chan, 1981; Sakai *et al.*, 1999). Second, seed dispersal in *Shorea* species is limited to app. 50m (Chan, 1980). Third, *Shorea* species do not form dense stands, instead they are usually scattered among other species and individual trees are often widely separated. All these factors suggest that the effective population size of *Shorea* species may be quite small.

Variation in the levels of polymorphism among different nuclear genes even at silent sites, which are considered selectively neutral was reported in *P. sylvestris* (Szmidszt unpublished), *Cryptomeria japonica* (Kado *et al.*, 2003) and *Arabidopsis thaliana* (Innan *et al.*, 1996; Kamiya *et al.*, 2002; Kawabe & Miyashita, 1999; Kawabe *et al.*, 1997; Kawabe *et al.*, 2000; Purugganan & Suddith, 1998; Purugganan & Suddith, 1999; Savolainen *et al.*, 2000). Therefore, it is also possible that the low levels of polymorphism found in our study were caused by the nature of the *GapC* gene. Evidently, additional sequence data are necessary to clarify this issue.

Diversification: recent or ancient?

The largest divergence was found between *S. leprosula* and *S. parvifolia* ($D_a = 0.0172$), and the smallest divergence was found between *S. acuminata* and *S. curtisii* ($D_a = 0.0093$). This suggests that *S. leprosula* and *S. parvifolia* are relatively diverged, whereas *S. acuminata* and *S. curtisii* appear to be more closely related. In fact, these two species had shared polymorphic site at position 11 (Fig 1). *Shorea curtisii* and *S. leprosula* also had shared polymorphic site at position 404 (Fig 1). However, this polymorphism was singleton in *S. leprosula*, therefore it may represent a recent mutation.

Previous analysis of the RAPD variation in *S. acuminata*, *S. curtisii*, and *S. leprosula* revealed that the average number of nucleotide substitutions among species was only 1.5 times higher than that within species (Harada *et al.*, 1994). Therefore, the authors suggested recent speciation of the genus and/or considerable gene exchange among species. In the present study, the average number of nucleotide substitutions per site between species (D_{xy}) was on the average 8.5 times higher than the average number of pairwise differences per site within species (π) (Table 3). Furthermore, on the constructed gene tree, haplotypes belonging to individual species were grouped in separate distinct clades (Fig 2). These results do not support recent divergence hypothesis proposed by Harada *et al.* (1994). If hybrid individuals harboring sequences from different species similar to those found in this study were also present in populations analyzed by Harada *et al.* (1994), they could account for the increased estimates of within species diversity obtained in their study.

Possibility of hybridization

Seven individuals harboring atypical sequences were found in this study. Several explanations can be considered for the presence of these sequences. First, individuals harboring such sequences were misidentified in the field and belong to a different species. Second, the atypical sequences can represent ancestral polymorphism. Third, they are due to inter-specific hybridization.

The four *Shorea* species investigated in this study are distinct morphologically and relatively easy to identify in the field. As mentioned before, identification of the individuals harboring atypical haplotypes was carried independently by two experts and six of the seven individuals with atypical haplotype also harbored one typical haplotype that conformed to morphological identification. Moreover, haplotype composition of all remaining individuals was concordant with their identification based on morphology. We therefore believe that incorrect field identification is not a likely explanation for the presence of atypical haplotypes detected in this study.

All atypical haplotypes detected in this study showed characteristic segmented structure, where individual segments

resembled typical sequences of different species. This was particularly evident in the individual sa85, whose atypical haplotype contained two segments resembling *S. acuminata* and *S. leprosula* respectively. The structure of atypical haplotypes in individuals sc275, sp296 and sl326 was similar except that one of the two segments did not resemble any of the investigated species. The structure of the atypical haplotype found in individuals sa86, sa347 and sa350 was more complex because it consisted of three segments two of which resembled *S. curtisii* and *S. leprosula* respectively while the third segment did not resemble any of the species investigated in this study. Furthermore, individual sa86 was collected in the population separated by 200 km from the population where the other two individuals (sa347 and sa350) were sampled. In our opinion an independent origin of this atypical haplotype in each of these two populations is not likely. Thus its appearance must have preceded separation of the two populations.

Based on our data alone it is difficult to decide unequivocally whether the observed atypical haplotypes are simply ancestral polymorphisms or rather they originated through inter-specific hybridization. More sequence data from other related species are necessary to clarify this issue. Nevertheless, several independent lines of evidence suggest that at least some of these haplotypes might have arisen as a result of hybridization. First, mutations are expected to occur randomly along the sequence. In other words, polymorphic sites are expected to occur randomly along the sequence rather than in segments. In contrast, the atypical haplotypes found in this study appeared to consist of segments resembling different species. Such segmented structure is consistent with their origin through recombination between chromosomes of different species.

For hybridization to occur it is necessary that corresponding species are distributed sympatrically. Generally, the habitats of the investigated species are overlapping. In fact, in the area where the individuals harboring atypical haplotypes were collected parental species were distributed in close proximity. In addition, distributions of several other members of the *Mutica* section (*Shorea dasyphylla*, *S. lepidota*, *S. macroptera* ssp. *macroptera*, *S. ovata*, *S. singkawang* ssp. *singkawang*) are overlapping with species used in this study. Therefore, these species could also be donors of some of the atypical haplotypes. *Shorea* species flower at the same time and are pollinated by the same insect groups (Appanah & Chan, 1981). Thus, the occurrence of the inter-specific gene exchange is possible. In fact, Chan (1981) showed successful fruit formation resulting from a cross between *Shorea splendida* and *S. stenoptera* further indicating the potential for natural hybridization among closely related species.

If we assume that at least some of the atypical haplotypes detected in our study were produced by past recombination between sequences from different species, then we can also conclude that individuals harboring such haplotypes represent advanced generation hybrids, which in turn would indicate fertility of hybrids. The presence of such fertile hybrids may potentially enrich gene pool in the population, decrease divergence among species or evolve into new species. Nevertheless, additional sequences should be studied to confirm the origin of the atypical haplotypes found in this study. If some of the individuals which had atypical haplotype composition are F1 hybrids they are expected to have sequences from two different species also in the other regions of the genome. Moreover, additional evidence such as for example, morphological difference of hybrids, evidence for cross compatibility between species, fertility of hybrids etc. is necessary to support our present hybridization hypothesis.

ACKNOWLEDGMENTS We are grateful to the members of our laboratory for technical support and useful advice during our study. We also thank Tadashi Kajita for providing samples and information for this study, and Yoko Tsuida for supplying the sequence data. This study was supported by grants Nos 11490026 and 12575002 to T. Y. and 13490022 and 13575002 to AES from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

REFERENCES

- Appanah, S. & Chan, H.T. 1981. Thrips: The pollinators of some Dipterocarps. *The Malaysian Forester* **44**: 234-252.
- Chan, H.T. 1980. Reproductive biology of some Malaysian Dipterocarps 2. Fruiting biology and seedling studies. *The Malaysian Forester* **43**: 438-451.
- Chan, H.T. 1981. Reproductive biology of some Malaysian Dipterocarps 3. Breeding systems. *The Malaysian Forester* **44**: 28-36.
- Dvornyk, V., Sirvio, A., Mikkonen, M. & Savolainen, O. 2002. Low nucleotide diversity at the *pal1* locus in the widely distributed *Pinus sylvestris*. *Molecular Biology & Evolution* **19**: 179-188.

- García-Gil, M.R., Mikkonen, M. & Savolainen, O. 2003. Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. *Molecular Ecology* **12**: 1195-1206.
- Harada, K., Kinoshita, A., Abshukor, N.A., Tachida, H. & Yamazaki, T. 1994. Genetic variation estimated in three *Shorea* species by the RAPD analysis. *Japanese Journal of Genetics* **69**: 713-718.
- Innan, H., Tajima, F., Terauchi, R. & Miyashita, N.T. 1996. Intragenic recombination in the *Adh* locus of the wild plant *Arabidopsis thaliana*. *Genetics* **143**: 1761-1770.
- Järvinen, P., Lemmetyinen, J., Savolainen, O. & Sopanen, T. 2003. DNA sequence variation in *BpMADS2* gene in two populations of *Betula pendula*. *Molecular Ecology* **12**: 369-384.
- Kado, T., Yoshimaru, H., Tsumura, Y. & Tachida, H. 2003. DNA variation in a conifer, *Cryptomeria japonica* (Cupressaceae sensu lato). *Genetics* **164**: 1547-1559.
- Kamiya, T., Kawabe, A. & Miyashita, N.T. 2002. Nucleotide polymorphism at the *Atmyb2* locus of the wild plant *Arabidopsis thaliana*. *Genetical Research* **80**: 89-98.
- Kawabe, A. & Miyashita, N.T. 1999. DNA variation in the basic chitinase locus (*ChiB*) region of the wild plant *Arabidopsis thaliana*. *Genetics* **153**: 1445-1453.
- Kawabe, A., Innan, H., Terauchi, R. & Miyashita, N.T. 1997. Nucleotide polymorphism in the acidic chitinase locus (*ChiA*) region of the wild plant *Arabidopsis thaliana*. *Molecular Biology & Evolution* **14**: 1303-1315.
- Kawabe A., Yamane, K. & Miyashita, N.T. 2000. DNA polymorphism at the cytosolic phosphoglucose isomerase (*PgiC*) locus of the wild plant *Arabidopsis thaliana*. *Genetics* **156**: 1339-1347.
- Lee, S.L., Wickneswari, R., Mahani, M.C. & Zakri, A.H. 2000. Genetic diversity of a tropical tree species, *Shorea leprosula* Miq. (Dipterocarpaceae), in Malaysia: Implications for conservation of genetic resources and tree improvement. *Biotropica* **32**: 213-224.
- Purugganan, M.D. & Suddith, J.I. 1998. Molecular population genetics of the *Arabidopsis CAULIFLOWER* regulatory gene: Nonneutral evolution and naturally occurring variation in floral homeotic function. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 8130-8134.
- Purugganan, M.D. & Suddith, J.I. 1999. Molecular population genetics of floral homeotic loci: Departures from the equilibrium-neutral model at the *APETALA3* and *PISTILLATA* genes of *Arabidopsis thaliana*. *Genetics* **151**: 839-848.
- Murray, M.G. & Thompson, W.F. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* **8**: 4321-4325.
- Nagamitsu, T., Ichikawa, S., Ozawa, M., Shimamura, R., Kachi, N., Tsumura, Y. & Muhammad, N. 2001. Microsatellite analysis of the breeding system and seed dispersal in *Shorea leprosula* (Dipterocarpaceae). *International Journal of Plant Sciences* **162**: 155-159.
- Nei, M. 1987. *In*: Molecular evolutionary genetics. Columbia University Press, New York, 512 pp.
- Obayashi, K., Tsumura, Y., Ihara-Ujino, T., Niiyama, K., Tanouchi, H., Suyama, Y., Washitani, I., Lee, C.T., Lee S.L. & Muhammad, N. 2002. Genetic diversity and outcrossing rate between undisturbed and selectively logged forests of *Shorea curtisii* (Dipterocarpaceae) using microsatellite DNA analysis. *International Journal of Plant Sciences* **163**: 151-158.
- Rozas, J. & Rozas, R. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174-175.
- Sakai, S., Momose, K., Yumoto, T., Kato, M. & Inoue, T. 1999. Beetle pollination of *Shorea parvifolia* (section *Mutica*, Dipterocarpaceae) in a general flowering period in Sarawak, Malaysia. *American Journal of Botany* **86**: 171-179.
- Savolainen, O., Langley, C.H., Lazzaro, B.P. & Freville, H. 2000. Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. *Molecular Biology and Evolution* **17**: 645-655.
- Shin, M.C., Heinrich, P. & Goodman, H.M. 1991. Cloning and chromosomal mapping of nuclear genes encoding chloroplast and cytosolic glyceraldehyde-3-phosphate-dehydrogenase from *Arabidopsis thaliana*. *Gene* **104**: 133-138.
- Swofford, D.L. 1991. PAUP: phylogenetic analysis using parsimony. *In*: ed. 3.0 ed. Champaign, Ill: Illinois Natural History Survey, ed. vol
- Thomson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequences alignment aided by quality analysis tools. *Nucleic Acids Research* **25**: 4876-4882.

Watterson, G. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**: 256-276.

Received 4th Oct. 2003

Accepted 1th Dec. 2003