TROPICS Vol. 14 (1) Issued August 31, 2004

Comparative study of AFLP and microsatellite variation in four Dipterocarp species from natural and artificial populations in Sarawak, Malaysia

Tomotaka KONISHI[1], Ko HARADA[1], Lucy CHONG[2], Ernest CHAI[2], Joseph Jawa KENDAWANG[3], Hua Seng LEE[3], Takuo YAMAKURA[4], Akira ITOH[4], Katsutoshi SAKURAI[5], and Kazuhiko OGINO[6]

[1] Faculty of Agriculture, Ehime University, Matsuyama 790-8566, Japan
[2] Forest Research Centre, Sarawak Foresty Corporation Kuching, Malaysia
[3] Forest Department Sarawak, Kuching, Malaysia
[4] Faculty of Science, Osaka City University, Osaka 558-8585, Japan
[5] Faculty of Agriculture, Kochi University, Nangoku 783-8502, Japan
[6] School of Environmental Science, Shiga Prefecture University, Hikone 522-8533, Japan
Corresponding author: Ko Harada, Faculty of Agriculture, Ehime University, 3-5-7 Tarumi, Matsuyama, Ehime 790-8566, Japan.
E-mail: kharada@agr.ehime-u.ac.jp

ABSTRACT Degrees of genetic variation accumulated in four dipterocarp species in Lambir Hills National Park, Sarawak, Malaysia were estimated and compared with those in a man-made forest in Bakam Experimental Reserve using microsatellite and AFLP markers.

By using 2 to 3 microsatellite loci, 9 to 14 alleles were detected. The average observed heterozygosities for *Dryobalanops lanceolata*, *D. aromatica*, *Shorea beccariana* and *S. macrophylla* in the natural population were 0.400, 0.400, 0.580 and 0.516, respectively, and expected heterozygosities were 0.603, 0.461, 0.685 and 0.745, respectively. In all of the four species, excess homozygotes were observed. Significant deviation from the Hardy-Weinberg equilibrium was detected in *D. lanceolata* and *S. beccariana*. AFLP analysis in the natural population revealed that 94.6 to 99.7% of the fragments were polymorphic. The average heterozygosities for these species were 0.220, 0.232, 0.221 and 0.211, respectively. These values were much larger than those obtained in tree populations such as Japanese beech and Japanese oak populations in temperate deciduous broad leaf forests.

Genetic variation in the artificial population was compared with that in the natural population. Both the average expected heterozygosities and the number of alleles in the microsatellite loci were reduced in the artificial population. Some distortions such as linkage disequilibrium and the deviation from the Hardy-Weinberg equilibrium appeared in the artificial population of *S. beccariana*. The use of seeds collected from a limited number of mother trees to make the experimental site is thought to be the cause of the loss of genetic variation.

Key words: genetic variation, Dipterocarps, microsatellite, AFLP

INTRODUCTION

The tropical rain forest in Borneo is known as one of the richest places of biodiversity in the world. Extensive speciation of dipterocarps, which includes major canopy and emergent trees, has occurred in this region. Two hundred and sixty seven dipterocarp species have been identified in Borneo, 155 of them being endemic, suggesting that Borneo could be the center of speciation of these species (Ashton, 1982). To provide a conventional comparative view of species composition and to examine demographic dynamics such as regeneration and growth in a low land dipterocarp forest, a joint program of long-term ecosystem research (LTER) with Malaysia, United States and Japan was initiated at Lambir Hills National Park, Sarawak since 1991 (Lee *et al.,* 2002b). Censuses have been carried out at five-year intervals in a plot of 52 hectares in the national park area (Lee *et al.,* 2002b).

Our primary research purpose was to determine how much genetic variation is maintained in tree populations in the LTER plot as a typical case of tropical rain forests. Genetic variation is a fundamental element of biodiversity and accumulates in a population during the evolutionary process of the species. However, studies on genetic variation have started only recently (Yamazaki, 2003), and there have been few studies on genetic variation in tropical rain forest trees. A high level of polymorphism in enzyme loci in a tropical rain forest was first reported by Gan *et al.* (1977). Lee *et al.* (2000) showed, by using enzyme loci, that heterozygosity in *Shorea leprosula* in the Malay Peninsula varied among populations, ranging from 0.360 to 0.400, and was much higher than the average for angiosperms (0.183) and the average for tropical plants (0.191) (Harmrick *et al.,* 1992). Whereas, Kanzaki *et al.* (1996) showed a moderate level of genetic diversity (0.209) in *Shorea macrophylla* in Sarawak. Using a RAPD marker, Harada *et al.* (1994) showed that the average values of nucleotide

diversity (π) of *Shorea curtisii*, *S. leprosula* and *S. acuminata* in the Malay Peninsula were 0.0207, 0.0183 and 0.0177, respectively. These values are much higher than that reported for specific nuclear genes in a temperate conifer species, *Criptmeria japonica* (Kado *et al.,* 2003). Although the RAPD method does not require previous knowledge of genomic sequences and is easy to use, its reproducibility is not always satisfactory (Jones, 1997). The AFLP (amplified fragment length polymorphism) method is a recently developed fingerprinting method (Vos *et al.,* 1995). The advantages of using the AFLP method are that data reproducibility can be achieved to a satisfactory level and that it produces a much greater number of amplified fragments, which cover the whole genome in a random fashion. This allows more detailed analysis of evolutionary dynamics in a population at the genomic level (Miyashita *et al.,* 1999). On the other hand, microsatellites comprise a class of a variable number of tandem repeats with a repeat unit of a few nucleotides and have been used as genetic markers (Jeffrays *et al.,* 1985). Their codominant character is suitable for studying reproductive systems. Because of their high mutation rate, microsatellite loci are usually highly polymorphic (Tautz & Renz,1984, Baird *et al.,* 1986, Chakraborty *et al.,* 1991 and others). This provides a means for studying the genetic constitution of a population such as rare allele composition and linkage disequilibrium. Comprehensive genetic analysis of populations is possible by combining these two methods.

Although the diversity of tropical rain forests in Sarawak is high, these forest areas have been reduced in size and fragmented rapidly by human activities such as nontraditional slash and burn and deforestation for timber production (Lee, 1997). In order to restore the forest resources a reforestation and rehabilitation program began in Sarawak under the Fourth Malaysia Plan (1981 to 1985) which aimed at shifting cultivation areas (Lee, 1997). A joint study by the Sarawak Forestry Department and a team of Japanese scientists to examine the applicability of various ecology-based silvicultural techniques for the rehabilitation of the degraded ecosystem was carried out from 1995 to 2000 at the Bakam Experimental Reserve, approximately 18 km south of Miri, Sarawak (Sakurai *et al.,* 2000). An important issue for the construction of artificial forests aimed at the restoration of biological diversity of original forests is how much of the genetic variation will be retained in the reforestation. Most of the seeds used for planting in Bakam were collected from Lambir Area including the national park (Sakurai *et al.,* 2000). In order to evaluate the method for restoration and rehabilitation of the tropical rain forest, we examined the genetic variation of tree populations in this area. We examined four dipterocarp species that appeared in both the 52-ha plot and the Bakam Experimental Reserve: *Dryobalanops lanceolata*, *Dryobalanops aromatica*, *Shorea beccariana* and *Shorea macrophylla*. These species were planted in the Bakam Experimental Reserve because they are the major canopy trees in the 52-ha LTER plot and are commonly used for reforestation programs in the Sarawak area.

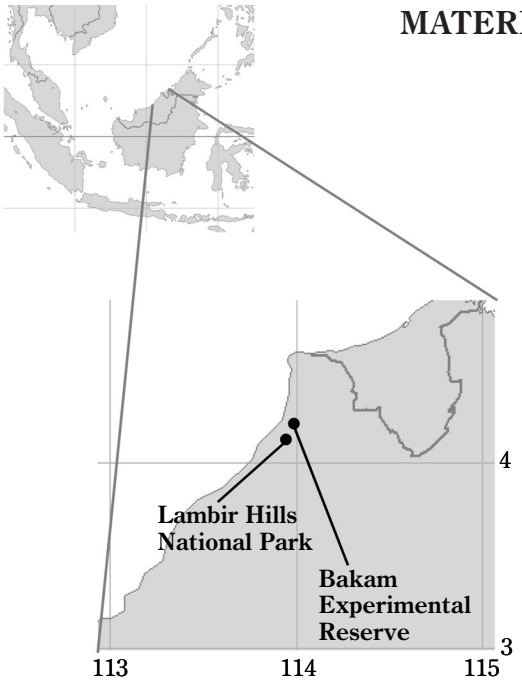

Fig. 1. The study sites in Sarawak, Malaysia. The X axis is longitude and the Y axis is latitude.

MATERIALS AND METHODS

Sampling methods

Natural forest

We collected samples from Lambir Hills National Park (4° 20'N, 113° 50' E) in the suburb of Miri (Fig. 1). This site is 6,800 hectares in area. Most of the area comprises lowland mixed dipterocarp forest with an extensive area of disturbed forest on the margins of the park and patches of heath (kerangas) forest on the highest ridges as well as small patches of swampy forest associated with streams (Watson, 1985). The sampling site was set in the 52-ha LTER plot in Lambir Hills National Park. Sampling was done from June 1999 to August 2002. Leaves were collected from 13 to 32 randomly chosen trees for *D. lanceolata*, *D. aromatica*, *S. beccariana* and *S. macrophylla* trees with diameters 1 to 10 cm. Two to five leaves were collected from each tree and kept in a cooler with dry ice at the sampling site. They were transported to a laboratory at the Forest Research Center (FRC) in Kuching and stored in a freezer at −20℃ until used for further analysis.

Artificial forest

The following nine tree species have been planted in the Bakam

Experimental Reserve: *Callophyllum* sp., *Koompassia malaccensis*, *Dryobalanops lanceolata*, *Dryobalanops aromatica*, *Shorea argentifolia*, *Shorea beccariana*, *Parashorea smythiesii*, *Dipterocarpus tempehes* and *Shorea macrophylla* (Sakurai *et al.,* 2000). Seeds were collected from the whole area of Lambir Hills National Park (Fig. 1) at mass flowering time in late 1996 and early 1997 (Sakurai *et al.,* 2000) and raised for planting in subsequent years. Sampling was done in 2001 in a manner similar to that for the natural population.

DNA isolation

DNA was extracted using a modified CTAB method (Murry & Thompson, 1980). Leaf tissues (0.5 gm) were ground with liquid nitrogen. Cold isolation buffer I (50 mM Tris-HCl, pH 8.0, 5 mM EDTA, 350 mM sorbitol, 0.5 % 2-mercaptoethanol, 0.1 % bovine serum albumin, 10 % polyethylene glycol) and cold isolation buffer II (50 mM Tris-HCl, pH 8.0, 5 mM EDTA, 350 mM sorbitol, 0.5 % 2-mercaptoethanol, 1 % sodium sarkosyl) were used for the washing process, and then samples were treated with 2X CTAB solution (2 % CTAB, 0.1 M Tris-HCl; pH 8.0, 20 mM EDTA, 1.4 M NaCl, 2 % 2-mercaptoethanol). Extracted DNA was further purified by Binding Matrix (Bio 101). Finally it was dissolved in 100 µl of TE. One microliter of the DNA solution was electrophoresed in a mini-gel apparatus (Mupid-2, ADV). The presence of high-molecular-weight DNA was checked by agarose gel with the marker λ/HindIII under UV illumination (Mini Transilluminator, Funakoshi).

Microsatellite analysis

We used microsatellite primers developed for *Shorea curtisii* by Ujino *et al.* (1998) and for *Dryobalanops lanceolata* by Terauchi (1994). After an initial screening of a total of 10 primer sets, 2 - 3 primer combinations that yielded a moderate number of alleles were chosen for the genetic analysis (Table 1). The reaction mixture for PCR contained 10 x reaction buffer (100 mM Tris-HCl, 500 mM KCl, pH 8.3, 15 mM $MgCl_2$, 0.01 % gelatin), 2 mM deoxynucleotide mix, 0.2 µM forward primers and 0.5 µM reverse primers, 0.1 units *Taq* DNA polymerase (Promega) and 10 ng genomic DNA in a total volume of 10 µl. The forward primers were FITC-labeled (GENSET OLIGOS). Microsatellite amplification was carried out using the following cycling parameters: preheating for 2 min at 94 ºC followed by 40 cycles of denaturing at 94 ºC for 45 s, annealing at primer-specific temperatures of 52 ºC to 56 ºC for 30 s and extension for 45 s at 72 ºC. Reactions were completed by incubation at 72 ºC for 5 min and holding at 4 ºC. Twenty microliters of deionised formamide was added to the PCR products and denatured for 3 min at 94 ºC, and then they were analyzed using a DSQ- 1000L auto-sequencer (Shimadzu).

Table 1. Microsatellite loci and conditions studied for dipterocarp species.

Locus	Core sequence	Primer sequence (5'-3')	Species	Annealing temperature
DL(GT)202	$(GT)_6GC(GT)_5(GA)_5$	CAGCACTTTTCTTTGACACA AACTTTCTATTACACCATTG	*D. lanceolata, D. aromatica*	50
Shc03	$(CT)_8$	TTGAAGGGAAGGCTATG CTTCTCAACTACCTTACC	*D. lanceolata, D. aromatica, S. beccariana*	54
Shc04	$(CT)_{16}$	ATGAGTAACAAGTGATGAG TATTGACGTGGAATCTG	*S. beccariana, S. macrophylla*	52
Shc11	$(CT)_4(A/T)T(CT)_5$	ATCTGTTCTTCTACAAGCC TTAGAACTTGAGTCAGATAC	*D. lanceolata, S. macrophylla*	54

Parameters for genetic variation, such as average number of alleles, unbiased expected heterozygosity (H_E) and observed proportion of heterozygotes (H_O), were calculated using POPGENE Version 1.31 (Yeh *et al.,* 1999). Deviation from the Hardy-Weinberg equilibrium (HWE) was examined using the Markov chain method with GENPOP version 3.1b (Raymond and Roussett, 1995) for overall loci and using the chi-square test for each locus with POPGENE Version 1.31 (Yeh *et al.,* 1999). The default parameters in GENEPOP with 100 batches for 1000 iterations were used for the Markov chain test. The inbreeding coefficient (F_{IS}) was calculated using GDA version 1.0 (Lewis & Zaykin, 1999). The unbiased estimates of H_E and F_{IS} were estimated for each locus. The 95 % confidence interval (CI) of F_{IS} was determined by bootstrap analysis using

1000 replicates. Linkage disequilibrium between microsatellite loci was examined, and its significance level was determined using Fisher's method with GENPOP version 3.1b (Raymond & Roussett, 1995).

AFLP analysis

AFLP analysis was performed by the method of Vos *et al.* (1995). Total DNA was digested with two restriction enzymes, *Eco*RI and *Mse*I. Site-specific adapters for *Eco*RI and *Mse*I restriction sites were then ligated. Primary and secondary selective amplifications were done using PCR primers complementary to the adapter sequences. Primers with additional nucleotide sequences were used for the secondary amplification. A total of 36 different selective amplification primer sets (*Mse*I + 3 selective nucleotides for 24 sets, *Mse*I + 4 selective nucleotides for 12 sets) were screened and 2 to 4 primer combinations that yielded a moderate number of bands were chosen for the genetic analysis. The following primer pairs were used for both the natural and artificial populations: *Eco*RI-ACA/*Mse*I-CTT for *D. lanceolata* and *D. aromatica*, *Eco*RI-ACA/*Mse*I-CAT for *D. lanceolata* and *S. macrophylla*, *Eco*RI-AAG/*Mse*I-CTT for *D. lanceolata* and *D. aromatica*, and *Eco*RI-AAG/*Mse*I-CAT for *D. lanceolata* and *S. macrophylla*. For *S. beccariana* *Eco*RI-ACA/*Mse*I-CAT and *Eco*RI-AAC/*Mse*I-CTG were used for the natural population and *Eco*RI-ACA/*Mse*I-CTT and *Eco*RI-AAC/*Mse*I-CTG were used for the artificial population. *Eco*RI primers were labeled with FAM for *Eco*RI-ACA, with JOE for *Eco*RI-AAG and with NED for *Eco*RI-AAC (Applied Biosystems). AFLP Core Reagent Kits (Life Technology, Inc.) were used according to the instructions of the manufacturer with minor modification. Fifty ng of genomic DNA was digested with a pair of restriction enzymes (*Mse*I/*Eco*RI) and then ligated to double-stranded *Mse*I and *Eco*RI adapters. The ligation mixture was diluted to 1/10, and 1 μl of the solution was used for the template for pre-amplification. PCR pre-amplification was performed with 30 cycles at 94 ℃ for 30 s, 56 ℃ for 1 min and 72 ℃ for 1 min. The pre-amplification products were diluted to 1/20 and were used as templates for the secondary amplification. The following PCR conditions were used: preheating at 94 ℃ for 2 min, followed by 10 cycles of 94 ℃ for 1 min, 65 ℃ reducing by 1 ℃ for 30 s and 70 ℃ for 2 min, and 20 cycles of 94 ℃ for 1 min, 56 ℃ for 30 s, and 72 ℃ for 2 min. The amplification products were analysed in an ABI 310 Genetic Analyser with GeneScan 500 ROX (Applied Biosystems) as a size standard.

Amplification products were scored as discrete character states (present/absent), and non-discernible fragments were excluded from the analysis using GENESCAN ver. 3.1 (Applied Biosystems) and GENOTYPER ver. 2.5 (Applied Biosystems). Gene diversity was estimated as expected heterozygosity (Nei, 1978) for each population and overall populations. Heterozygosity was calculated after correcting for the inbreeding coefficient obtained from microsatellite data using POPGENE Version 1.31 (Yeh *et al.,* 1999). Pairwise nucleotide diversity (π) within a population was estimated by the method of Innan *et al.* (1999).

RESULTS

Microsatellite analysis

Two to three microsatellite primers enabled detection of 6 - 14 alleles for each type of population. Locus Shc04 showed the largest number of alleles (8) of *S. macrophylla* in the artificial population, while DL(GT)202 of *D. aromatica* in the artificial population and Shc03 in *D. lanceolata* and in *D. aromatica* in the artificial population and in *S. beccariana* in both populations showed the smallest number of alleles (3) corresponding to the number of di-nucleotide repeats (Table 1). The observed heterozygosity (H_O) ranged from 0.214 in locus DL(GT)202 to 0.719 in locus Shc11. The average observed heterozygosities (H_O) for *D. lanceolata*, *D. aromatica*, *S. beccariana* and *S. macrophylla* in the natural population were 0.400, 0.400, 0.580 and 0.516, respectively, while those of the expected heterozygosities (H_E) were 0.603, 0.461, 0.685 and 0.745, respectively. Expected heterozygosities were larger than the observed heterozygosities in all the species in the natural population. The average observed heterozygosities for these species in the artificial population were 0.451, 0.462, 0.415 and 0.641, respectively, while those of the expected heterozygosities were 0.459, 0.394, 0.607 and 0.681, respectively (Table 2). All the average expected heterozygosities in the natural population were larger than those of the artificial population (Table 2).

Deviation from the Hardy-Weinberg equilibrium (HWE) was examined and its significant level was determined overall and for each locus. Significant deviation from the HWE was found in 4 of the 18 cases: Shc03 and Shc11 for *D. lanceolata* and Shc04 and Shc11 for *S. macrophylla* in the natural population. Significant deviation from the HWE at overall loci was found in 3 cases: *D. lanceolata* and *S. macrophylla* in the natural population and *S. beccariana* in the artificial population (Table 2).

Since a linkage disequilibrium (LD) was expected in the artificial population, LD tests were conducted. A significant LD

was found only in *S. beccariana* in the artificial population (Table 3).

Table 2. Genetic variation and inbreeding coefficients by microsatellite analysis.

Species	Population	N^a	Locus	A^b	$H_O{}^c$	$H_E{}^d$	P^e	$F_{IS}{}^f$
D.lanceolata	Natural	20	DL(GT)202	4	0.55	0.56	0.225	
		20	Shc03	4	0.25	0.604	0.003**	
		20	Shc11	3	0.4	0.645	0.004**	
	Overall			11	0.4	0.603	0.020*	0.342
		14	DL(GT)202	4	0.214	0.206	0.999	
	Artificial	14	Shc03	3	0.5	0.611	0.118	
		14	Shc11	3	0.64	0.558	0.562	
	Overall			10	0.451	0.459	0.43	0.014
D.aromatica	Natural	20	DL(GT)202	4	0.45	0.389	0.961	
		20	Shc03	5	0.35	0.534	0.455	
	Overall			9	0.4	0.461	0.205	0.136
	Artificial	13	DL(GT)202	3	0.462	0.394	0.814	
		13	Shc03	3	0.462	0.394	0.814	
	Overall			6	0.462	0.394	0.223	−0.18
S.beccariana	Natural	13	Shc03	3	0.62	0.6	0.717	
		13	Shc04	6	0.54	0.769	0.28	
	Overall			9	0.58	0.685	0.303	0.086
	Artificial	18	Shc03	3	0.44	0.527	0.219	
		18	Shc04	5	0.39	0.687	0.144	
	Overall			8	0.415	0.607	0.021*	0.286
S.macrophylla	Natural	32	Shc04	7	0.406	0.75	0.001**	
		32	Shc11	7	0.625	0.739	0.005**	
	Overall			14	0.516	0.745	0.015*	0.311
	Artificial	32	Shc04	8	0.563	0.732	0.875	
		32	Shc11	6	0.719	0.629	0.355	
	Overall			14	0.641	0.681	0.218	0.063

**P < 0.01, *P < 0.05
[a]*N*, number of individuals
[b]*A*, number of alleles
[c]H_O, observed proportion of heterozygotes
[d]H_E, expected proportion of heterozygotes
[e]*P*, fitness to the Hardy-Weinberg equilibrium
[f]F_{IS}, inbreeding coefficient

Table 3. Probability of linkage equilibrium between microsatellite loci.

Species	Population	DL(GT)202-Shc03	DL(GT)202-Shc11	Shc03-Shc04	Shc03-Shc11	Shc04-Shc11
D. lanceolata	Natural	0.4344	0.2096	—	0.4864	—
	Artificial	0.7511	0.591	—	0.0575	—
D. aromatica	Natural	0.79	—	—	—	—
	Artificial	0.9024	—	—	—	—
S. beccariana	Natural	—	—	0.6858	—	—
	Artificial	—	—	0.0018**	—	—
S. macrophylla	Natural	—	—	—	—	0.3879
	Artificial	—	—	—	—	0.034

**P < 0.01
"—" Sign shows that the test was not applicable

AFLP analysis

Four primer combinations were used for *D. lanceolata*, and two primer combinations were used for the other species. From the combined data of natural and artificial populations, 483, 171, 206 and 284 fragments for *D. lanceolata*, *D. aromatica*, *S. beccariana* and *S. macrophylla*, respectively were scored in the natural population and 436, 286, 260 and 324 fragments for these species, respectively were scored in the artificial population (Table 4). We considered the bands that showed variation at least in one individual to be polymorphic. Both the natural and artificial populations are highly polymorphic (Table 4).

AFLP bands were scored as present or absent. The genotype of the null bands, which are absent on the gel, is considered to be recessive homozygote, and that of the bands present on the gel is either homo or heterozygote for the wild type. Here, wild type is defined as a band that possesses intact nucleotide sequences for the restriction sites and three additional bases. The average heterozygosities (gene diversity by Nei (1977)) for each species were estimated after correction by the inbreeding coefficient obtained by microsatellite analysis and are listed in Table 5. The value in the natural population was significantly higher than that in the artificial population in *S. macrophylla* ($t = 3.55$, $df = 26$, $P < 0.005$), but was not significantly different in the other species. The frequency distribution of heterozygosities is shown in Fig. 2. The distribution showed a U shape and a large fraction located near 0.

Table 4. Number of polymorphic bands in AFLP analysis.

	Natural			Artificial		
Species	N^a	M^b	$K(\%)^c$	N^a	M^b	$K(\%)^c$
D.lanceolata	27	483	469 (97.1)	27	436	434 (99.5)
D.aromatica	8	171	167 (97.7)	16	286	284 (99.3)
S.beccariana	13	206	197 (95.6)	18	260	246 (94.6)
S.macrophylla	12	284	282 (99.3)	16	324	323 (99.7)

[a] *N*, number of individuals
[b] *M*, total number of bands
[c] *K*, number of polymorphic bands

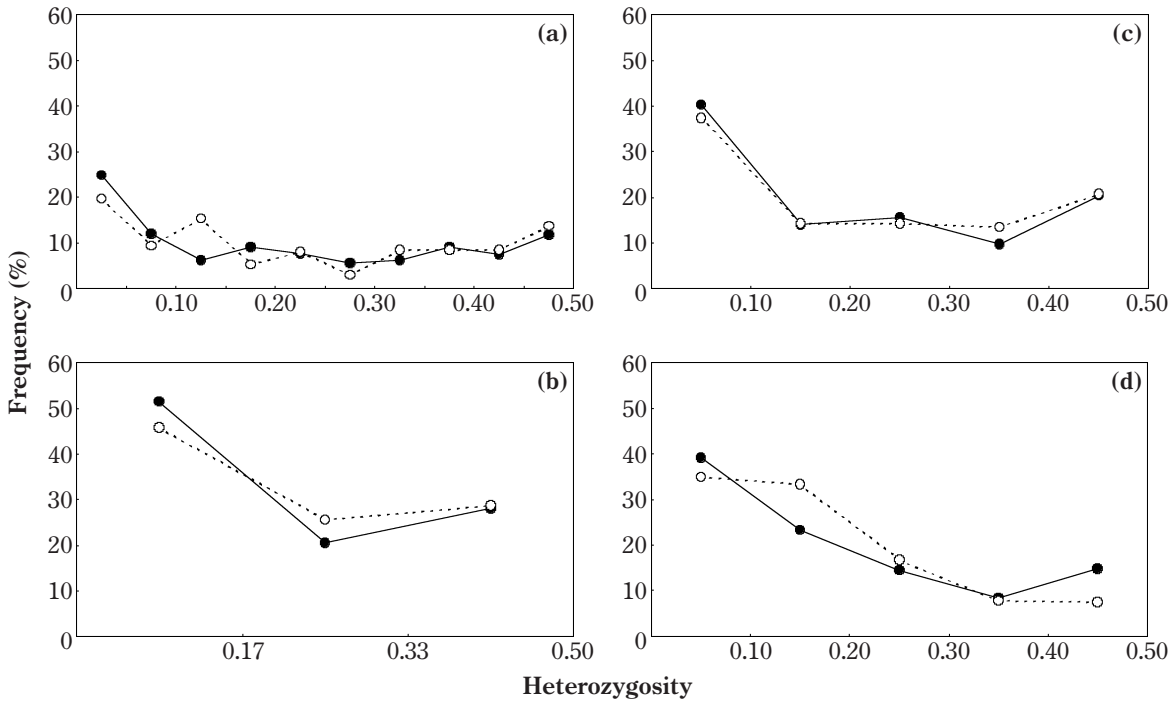

Fig. 2. Frequency distribution of heterozygosities in AFLP markers. Solid lines show the distribution in the natural population, and dotted lines show the distribution in the artificial population. a, *Dryobalanops lanceolata*; b, *Dryobalanops aromatica*; c, *Shorea beccariana*; d, *Shorea macrophylla*.

Table 5. Average heterozygosity revealed by AFLP analysis

Species	Natural population	Artificial population
D. lanceolata	0.220 ± 0.0073	0.231 ± 0.0078
D. aromatica	0.232 ± 0.0101	0.233 ± 0.0085
S. beccariana	0.221 ± 0.0111	0.222 ± 0.0102
S. macrophylla	0.211 ± 0.0072	0.174 ± 0.0068

Values following "±" sign denote standard error

Pairwise nucleotide diversity (π) is the proportion of nucleotide difference between two sequences. AFLP data were converted to pairwise nucleotide diversity (π) using the program developed by Innan *et al.* (1999). The average nucleotide diversities are shown in Table 6. The highest nucleotide diversity was obtained in *S. macrophylla* in both the natural and artificial populations. The smallest values were obtained in *D. lanceolata* in the natural population and in *S. beccariana* in the artificial population. The frequency distribution of pairwise nucleotide diversity is shown in Fig. 3. In *D. lanceolata* and *S.macrophylla*, both the natural and the artificial population have one peak with a similar distribution spectrum; however, several unusual values appeared in the artificial population. In *D. aromatica* and *S. beccariana*, the peaks for the artificial population shift to the left of the natural population.

Table 6. Nucleotide diversity (π) revealed by AFLP analysis

Species	Natural population	Artificial population
D. lanceolata	0.0310 ± 0.00041	0.0322 ± 0.00057
D. aromatica	0.0596 ± 0.00243	0.0556 ± 0.00105
S. beccariana	0.0320 ± 0.00102	0.0244 ± 0.00059
S. macrophylla	0.0711 ± 0.00190	0.0813 ± 0.00213

Values following "±" sign denote standard error

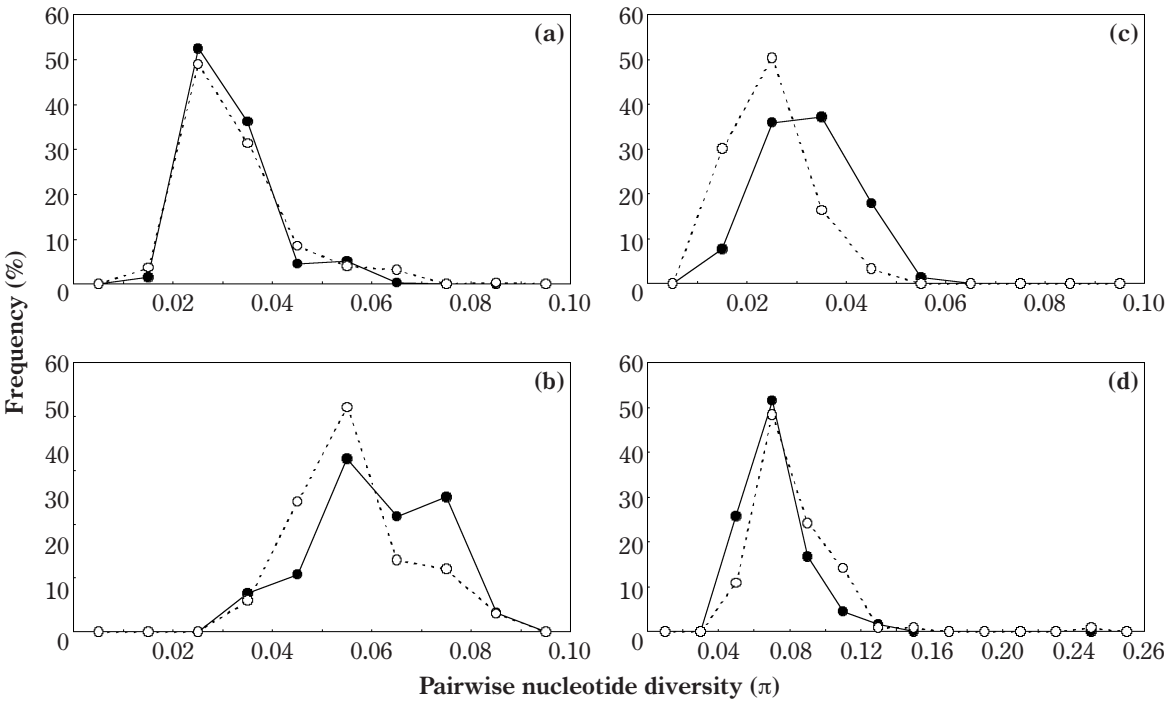

Fig. 3. Frequency distribution of nucleotide diversities (π) in AFLP markers. Solid lines show the distribution in the natural population, and dotted lines show the distribution in the artificial population. a, *Dryobalanops lanceolata*; b, *Dryobalanops aromatica*; c, *Shorea beccariana*; d, *Shorea macrophylla*.

DISCUSSION

Mating system in the natural forest

In the natural population, observed heterozygosities for the microsatellite loci Shc03, Shc04 and Shc11 ranged from 0.250 to 0.620, 0.390 to 0.563 and 0.400 to 0.719, respectively (Table 2). The reported heterozygosities of these loci in *Shorea critesii* were 0.591, 0.844 and 0.640, respectively (Ujino *et al.,* 1998). Nagamatsu *et al.* (2001) reported values of 0.568, 0.795 and 0.614 in *Shorea leprosula* in Pasoh Forest Reserve for the loci Shc03, Shc07 and Shc09, respectively. The heterozygosities observed in our study are similar to the values in these studies. All of the four species showed more homozygotes than expected. Significant departure from HWE was observed in *D. lancelolata* and *S. macrophylla*. The mating system of some tropical tree species has been investigated using isozymes (Murawski & Hamrick, 1991; Bawa, 1992; O'Malley & Bawa, 1987; Kitamura *et al.,* 1994) and microsatellites (Nagamatsu *et al.,* 2001). Shiraishi *et al.* (1990) studied the mating system of *D. aromatica* using isozymes in the Andulau Forest Reserve in Brunei and found that surviving seedlings were the products of random mating. Nagamatsu *et al.* (2001) studied the mating system in *Shorea leprosula* in Pasoh Forest Reserve in Malaysia using microsatellites and found no significant deviations from HWE in any of the loci. These studies showed that these species are basically outbreeding species.

What is the mechanism by which inbreeding is promoted in *D. lancelolata* and *S. macrophylla* in Lambir Hills National Park? The current data set of tree censuses in the 52-ha plot includes 358,452 individual trees with diameters ≥1 cm dbh (mean = 6,898 trees/ha) (Lee *et al.,* 2002a). There are 956, 10,650, 3,785 and 109 trees with diameters ≥ 1 cm dbh for *D. lanceolata*, *D. aromatica*, *S. beccariana* and *S. macrophylla*, respectively (Lee *et al.,* 2002a), showing that tree density is much less in *D. lanceolata* and *S. macrophylla*. The census data also show that *D. lanceolata* and *S. macrophylla* have clumped distributions around limited numbers of mother trees and that *D. aromatica* and *S. beccariana* have fairly random distributions with high densities. Murawski and Hamrick (1991 and 1992) reported that outcrossing rates of tropical trees were positively correlated with flowering density. According to their report, trees with low density may lead to low flowering tree density and increase crosses among genetically related flowers. On the other hand, some of the Malaysian dipterocarps species, including *D. aromatica* and *S. macrophylla*, have been found to exhibit sporophytic agamospermy (Kaur *et al.,* 1978). Although the frequency of apomixis has not been determined for these and other species, this might also be the mechanism by which inbreeding is promoted.

Genetic variation in the natural forest

Genetic variation in tropical rain forest trees determined by isozymes has been reported by Hamrick & Loveless (1989) and Hamrick *et al.* (1992). According to Hamrick *et al.* (1992), the gene diversity of tropical tree species is not necessarily high compared with that of trees in a temperate forest. However, there have only been a few detailed investigations on dipterocarp species. Among them, an exceptionally high level of genetic variation has been found in *Shorea curtisii* using isozyme markers (Gan *et al.,* 1977; Lee *et al.,* 2000). Similarly, a high level of genetic variation was found in other *Shorea* species using RAPD markers (Harada *et al.,* 1994). In the present study, a high level of genetic variation was found by AFLP analysis (Table 5) compared with the level of genetic variation in other plant species, such as *Pinus pinaster* in France and Portugal (0.131 to 0.173: Ribeiro *et al.,* 2002), *Pterocarpus officinalis* in the Neotropics (0.15 to 0.27: Rivera-Ocasio *et al.,* 2002), *Trigonobalanus verticillata* in Malaysia (0.134 to 0.170: Kamiya *et al.,* 2002), *Avicennia marina* in Vietnam (0.037 to 0.083: Giang *et al.,* 2003), *Fagus crenata* in Japan (0.114 to 0.138: Okaura & Harada, 1998) and *Quercus mongolica var. crispula* in Japan (0.104 to 0.163: Okaura & Harada, 1998).

The frequency distribution of heterozygosities showed a U-shaped distribution (Fig. 2). This indicates that DNA variations at intermediate frequencies are fairly common. This pattern was not observed in temperate deciduous forest trees such as Japanese beech (*Fagus crenata*)(Harada *et al.,* 2000) and Japanese oak (*Quercus mongolica* var. *crispula*) (Harada *et al.,* unpublished data). In a study of *Drosophila*, the amount of additive genetic variance on viability in a northern population could be explained by the mutation-selection balance, whereas an excess amount of additive genetic variance was observed in southern populations (Kusakabe & Mukai, 1984). The excess amount of variance could be explained by the operation of some kind of balancing selection on non-coding regions (Kusakabe & Mukai, 1984). Most of the fragments detected by AFLP analysis are thought to be sampled from non-coding regions. Such regions are, however, not necessarily uninformative. Recent studies have demonstrated that there are many non-coding but transcribed regions. It has been revealed that those regions, so-called non-coding RNA (ncRNA) regions, play various roles in the regulation of metabolism through control of

transcription, replication and RNA processing (Storz, 2002 for review). Although it is not known how much of the DNA regions is occupied by these specific regions, these regions can also be the sources of natural selection.

Nucleotide diversity (π) is the average number of nucleotide differences per site between two sequences and one of the most commonly used parameters for evaluating genetic variation at the molecular level (Nei, 1987). The average values of nucleotide diversity (π) for *D. lanceolata*, *D. aromatica*, *S. beccariana* and *S. macrophylla* in the natural populations were 0.0310, 0.0596, 0.0320 and 0.0711, respectively. These values are about two to four-times higher than those of *Shorea curtisii*, *S. leprosula* and *S. acuminata* obtained from RAPD analysis (0.0207, 0.0183 and 0.0177, respectively: Harada *et al.,* 1994). The value obtained by AFLP analysis for ecotypes of *Arabidopsis thaliana* collected worldwide was 0.0106 (Miyashita *et al.,* 1999) and that for Japanese beech was 0.0130 (Harada *et al.,* 2000). A large amount of nucleotide diversity may indicate that the tree population in the tropical rain forest is evolutionary much older than those in forests in temperate and cold climates or that the population size of the former is much larger. In this study, nucleotide diversity (π) was estimated by the procedure of Innan *et al.* (1999). This method, however, can be used only under the assumption of HWE. In our study, deficiency of observed heterozygosity was detected by microsatellite analysis in all of the species in the natural population. This will cause an overestimation of heterozygosity. Thus, care must be taken, especially if the deficiency is statistically significant, such as in *D. lanceolata*, *S. beccariana* and *S. macrophylla*.

Genetic variation in the artificial forest

In the microsatellite analysis, expected heterozygosity was larger in the natural forest than in the artificial forest in all of the four species. Similarly, there were reduced numbers of alleles in three of the four species in the artificial forest (Table 2). The magnitude of reduction in the heterozygosity ranged from 8.6% (*S. macrophylla*) to 24% (*D. lanceolata*). In the AFLP analysis, slightly larger heterozygosity in the artificial population, though not statistically significant, was observed for all species except *S. macrophylla*. In *S. macrophylla*, average heterozygosity was significantly larger in the natural population than in the artificial population. Contradicting to these results, nucleotide diversity was larger in the artificial population in *D. lanceolata* and *S. macrophylla* (Table 6). This is thought to be caused by the unusually high values in some of the combinations in the artificial forest (Fig. 3).

Reduction of heterozygosity occurs due to a decrease in population size. A limited number of mother trees as a seed source may have caused the reduction of genetic variation in the artificial forest. Effective sizes of the mother trees are estimated from the reduction of heterozygosities in the microsatellite loci to be 2.0, 3.3, 4.5 and 5.8 for *D. lanceolata*, *D. aromatica*, *S. beccariana* and *S. macrophylla*, respectively. Sampling of seeds from a limited number of mother trees results in a biased sampling of genes and does not represent the genetic constitution of the original population. Loss of rare alleles and linkage disequilibrium among the loci may appear in the seed population. Only *S. beccariana* in the artificial population exhibited a significant level of linkage disequilibrium ($P = 0.002$). *Shorea beccariana* showed significant deviation from the HWE in the artificial population even though random mating is attained in the natural population. This may be caused by a biased sampling of genes in *S. beccariana* in the artificial forest.

Heterozygosity is a commonly used parameter to evaluate genetic variation; however, it is known to be rather robust against population size changes. Allele frequency is much more sensitive to population size change (Yamazaki, 2003). We examined rare allele frequencies in *D. lanceolata* using AFLP analysis, because sample sizes of the other species were small. Proportions of rare alleles with frequencies of less than 0.05 in *D. lanceolata* were 34.0% in the natural population and 28.7% in the artificial population. A fraction of rare alleles must have been lost in the artificial population. The results of our study suggest that genetic parameters, such as allele frequency and linkage disequilibrium, are useful for evaluating the degree to which the original genetic constitution of a natural forest has been retained in a man-made forest.

ACKNOWLEDGMENTS We thank the staff of the Forest Research Centre and Forest Department of Sarawak, especially Mr. Patsipun Rosen, who recently passed away, Mr. Tinjan Kuda, Ms. Latifah Teh, Mr. Augustine Jee, Mr. Awang Enjah, Ms. Bibian Michael Diway, Ms. Patriscia Lenger and Ms. Lisa Sualu for their technical support. We also thank Mr. Sylvester Tan for his helpful discussions. This research was supported by a grant from the Nissei Foundation to K. H. and by a grant-in-aid (No. 13575038) from the Japan Society for the Promotion of Science to K. S.

REFERENCES

Ashton, P. S. 1982. Dipterocarpaceae. *Flora Malesiana*, Ser. 1. **9:** 237-552.

Baird, M., Balazs, I., Giusti, A., Miyazaki, I., Wexler, K. *et al.,* 1986. Allele frequency distribution of two highly polymorphic DNA sequencies in three ethnic groups and its application to the determination of paternity. *American Journal of Human Genetics* **39:** 489-501.

Bawa, K. S. 1992. Mating system, genetic differentiation and speciation in tropical rain forest plants. *Biotropica* **24:** 250-255.

Chakraborty, R., Fornage, M., Gueguen, R. & Boerwinkle, E. 1991. Population genetcs of hypervariable loci: analysis of PCR based VNTR polymorphism within a population. In: *DNA Fingerprinting: Approaches and Applications*. (Burke, T., Dolf, G., Jeffreys, A. J. & Wolff, R. eds.), Pp. 127-143. Birkhauser Verlag, Basel.

Gan, Y. Y., Robertson, F. W., Ashton, P. S., Soepadmo, E. & Lee, D. W. 1977. Genetic variation in wild populations of rain forest trees. *Nature* **269:** 323-325.

Giang, L. H., Harada, K., Hong, P. N. & Tuan M. S. 2003. Genetic variation of *Avicennia marina* (Forsk.) Vierh. (Avicenniaceae) in Vietnam revealed by microsatellite and AFLP markers. *Genes and Genetic Systems* **78:** 399-407.

Hamrick, J. L., Godt, M. J. W. & Sherman-Broyles, S. L. 1992. Factors influencing levels of genetic diversity in woody plant species. *New Forests*. **6:** 95-124.

Hamrick, J. L. & Loveless, M. D. 1989. The genetic structure of tropical trees populations: associations with reproductive biology. In: *Evolutionary Ecology of Plants*. (Bockand, J. H. & Lindhart, Y. B. eds.), Pp. 129-146. Westview Press, Boulder, CO.

Harada, K., Kinoshita, A., Shukor, N. A. A., Tachida, H. & Yamazaki, T. 1994. Genetic variation estimated in three *Shorea* species by the RAPD analysis. *The Japanese Journal of Genetics* **69:** 713-718.

Harada, K., Okaura, T. & Fujiwake, H. 2000. AFLP analysis of the genetic variation in natural populations of Japanese beech (*Fagus crenata*, Blume). *Shimadzu Review* **57:** 115-119.

Innan, H., Herauchi, R., Kahl, G. & Tajima, F. 1999. A method for estimating nucleotide diversity from AFLP data. *Genetics* **151:** 1157-1164.

Jeffreys, A. J., Wilson, V. & Thein, S. L. 1985. Individual-specific "fingerprints" of human DNA. *Nature* **316:** 76-79.

Jones, C. J., Edwards, K. J., Castaglione, S., Winfield, M. O., Sala, F. *et al.* 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Mol. Breed*. **3:** 381-390.

Kado, T., Yoshimaru, H., Tsumura, Y. & Tachida, H. 2003. DNA variation in a conifer *Cryptomeria japonica* (Curpressaceae sensu lato) *Genetics* **164:** 1547-1559.

Kamiya, K., Harada, K., Clyde, M. M. & Mohamed, A. L. 2002. Genetic variation of *Trigonobalanus verticillata*, a primitive species of Fagaceae, in Malaysia revealed by chloroplast sequences and AFLP markers. *Genes and Genetic Systems*. **77:** 177-186.

Kanzaki, M., Watanabe, M., Kuwahara, J., Kendawang, J.J. Lee, H. S., Serizawa, S. & Yamakura, T. 1996. Genetic structure of *Shorea macrophylla* in Sarawak, Malaysia. *Tropics* **6:** 153-160.

Kaur, A., Ha, C. O., Jong, K., Sands, V. E., Chan, H. T., Soepadmo, E. & Ashton, P. S. 1978. Apomixis may be widespread among trees of the climax rain forest. *Nature* **271:** 440-442.

Kitamura, K., Rahman, M. Y. B. A., Ochiai, Y. & Yoshimaru, H. 1994. Estimation of the outcrossing rate on *Dryobalanops aromatica* Gaertn. F. in primary and secondary forests in Brunei, Borneo, Southeast Asia. *Plant Species Biol*. **9:** 37-41.

Kusakabe, S. & Mukai, T. 1984. The genetic structure of natural populations of *Drosophila melanogaster*. XVIII. Clinal and uniform genetic variation over populations. *Genetics* **108:** 617-632.

Lee, H. S. 1997. Restoration of Deforested and Depredated sites in Sarawak, Malaysia. Ph. D. Thesis. Ehime University.

Lee, H. S., Ashton, P. S., Yamakura, T., Tan, S., Davies, S. J., Itoh, A., Chai, E. O. K., Ohkubo, T. & Lafrankie, J.V. 2002a. *The 52-Hectare Forest Research Plot at Lambir Hills, Sarawak, Malaysia: Tree Distribution Maps, Diameter Tables and Species Documentation*. Forest Department Sarawak & The Arnold Arboretum-CTFS Asia Program.

Lee, H. S., Davies, S. J., Lafrankie, J. V., Tan, S., Yamakura, T., Itoh, A., Ohkubo, T. & Ashoton, P. S. 2002b. Floristic and structual diversity of mix dipterocarp forest in Lambir Hills National Park, Sarawak, Malaysia. *Journal of Tropical Forest Science* **14:** 379-400.

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.

Lewis, P. O. & Zaykin, D. 1999. Genetic Data Analysis: Computer program distributed over the internet from the GDA home page at http://alleyn.eeb.uconn.edu/gda/.

Miyashita, N. T., Kawabe, A. and Innan, H. 1999. DNA variation in the wild plant *Arabidopsis thaliana* revealed by amplified fragment length polymorphism analysis. *Genetics* **152:** 1723-1731.

Murawski, D. A. & Hamrick, J. L. 1991. The effect of the density of flowering individuals on the mating systems of nine tropical tree species. *Heredity* **67:** 167-174.

Murawski, D. A. & Hamrick, J. L. 1992. The mating system of *Cavanillesia platanifolia* under extremes of flowering tree density: A test of predictions. *Biotropica*. **24:** 99-101.

Murry , M. G. and Thompson, W. F. 1980. Rapid isolation of high molecular weight plant DNA. *Nuc. Acids Res.* **8:** 4321-4325.

Nagamatsu, 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 Science* **162:** 155-159.

Nei, M. 1977. F-statistics and analysis of gene diversity in subdivided populations. *Ann. Hum. Genet.* **41:** 225-232.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89:** 583-590.

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Okaura, T. & Harada, K 1998. Genetic structure of Japanese beech and a kind of oak in the populations in western part of Japan. *Genes and Genetic Systems* **73:** 451.

O'Malley, D. M. & Bawa, K. S. 1987. Mating system of a tropical rain forest tree species. *American Journal of Botany* **74:** 1143-1149.

Raymond, M. & Roussett, F. 1995. GENEPOP (version 3.1b): population genetics software for exact tests and ecumenicism. *Heredity* **86:** 248-249.

Ribeiro, M. M., Mariette, S., Vendramin, G. G., Szmidt, A. E., Plomion, C. & Kremer, A. 2002. Comparison of genetic diversity estimates within and among populations of maritime pine using cpSSR and AFLP data. *Molecular Ecology* **11:** 869-877.

Rivera-Ocasio, E., Aide, T. M. & Mcmillan, W. O. 2002. Patterns of genetic diversity and biogeographical history of the tropical wetland tree, *Pterocarpus officinalis* (Jacq.), in the Caribbean basin. *Molecular Ecology* **11:** 67: –683.

Sakurai, K., Ninomiya, I., Harada, K., Kendawang, J. J., Lee, H. S. and & Ogino, K. 2000. Tree planting as ecosystem initiation on land degraded by Shifting cultivation at the Bakam Experimental Reserve, Sarawak. Proceedings of workshop on forest ecosystem rehabilitation, Kuching, Sarawak, Malaysia. 5-17.

Shiraishi, S., Latiff, N. H., Miyatake, S., Ochiai, Y. & Furukoshi, T. 1990. Population-genetical studies on Kapur (Dryobalanops) species. *The Tropical Forestry* **19:** 10-17 (in Japanese).

Storz, G. 2002. An expanding universe of noncoding RNAs. *Science* **296:** 1260-1263.

Tautz, D. & Renz, M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research* **12:** 4127-4138.

Terauchi, R. 1994. A polymorphic microsatellite marker from the tropical tree *Dryobalanops lanceolata* (Dipterocarpaceae). *The Japanese Journal of Genetics* **69:** 567-576.

Ujino, T., Kawahara, T., Tsumura, Y., Nagamitsu, T., Wickneswari, R. & Yoshimaru, H. 1998. Development and polymorphism of simple sequence repeat DNA markers for *Shorea curtisii* and other Dipterocarpaceae species. *Heredity* **8:** 422-428.

Vos, R., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T. *et al.* 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23:** 4407-4414.

Watson, H. 1985. *Lambir Hills National Park: Resource inventory with management recommendations*. National Parks and Wildlife Office, Forest Department Sarawk, Kuching, Malaysia.

Yamazaki, T. 2003. Estimating the genetic variation in tropical tree populations: Heterozygosity or allele number? *Tropics* **13:** 67- 68.

Yeh, F., Yang, R. C. & Boyle T. 1999. POPGENE Version 1.31. Microsoft Window-based freeware for population genetic analysis. Molecular Biology and Biotechnology Center, University of Alberta, Edmonton.

Received 13th Jan. 2004
Accepted 14th May 2004