New Forests
DOI 10.1007/s11056-011-9288-2

Genetic diversity in natural populations of *Gmelina arborea*: implications for breeding and conservation

Alison K. S. Wee · ChunHong Li · William S. Dvorak · Yan Hong

Received: 10 March 2011 / Accepted: 29 August 2011

Abstract *Gmelina arborea* is a valuable plantation tree species that is native to South and Southeast Asia. In this study, 534 samples representing 19 natural populations in India, China, Thailand and Myanmar were analyzed with 10 polymorphic microsatellite markers. The genetic diversity analysis revealed highly polymorphic loci ($N_a = 16.4$), a good level of genetic diversity ($H_o = 0.56$; $H_e = 0.83$) and the deficiency of heterozygotes in *G. arborea* populations evidenced by positive fixation index and deviation from Hardy–Weinberg Equilibrium in all loci and most populations. The Analysis of Molecular Variance attributed 21, 10 and 69% of total genetic diversity to among-region, among-population (within region) and within-population variation. Unweighted Pair Group Method with Arithmetic Mean dendrogram and Principle Coordinate Analysis revealed three separate clusters composed of China, India and Thailand/Myanmar that were consistent with geographical distance and the presence of natural barriers to gene flow. Populations from within India grouped together genetically consistent with geographical locations, with the exception of the Nowgong population (eastern India), that might have originated from the Kasa area (western India) with which it has high genetic similarity. Understanding genetic diversity and structure of *G. arborea* populations serve as an important reference for tree breeding programs and conservation strategy. Breeding populations of *G. arborea* should include selections from each of the major geographic regions to maximize genetic diversity and heterosis. Vegetative propagated clones of elite trees can be used for plantation to address the issue of high level of segregation among seed derived plants.

A. K. S. Wee
Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore

C. Li · Y. Hong (✉)
Temasek Life Sciences Laboratory, National University of Singapore, 1 Research Link, Singapore 117604, Singapore
e-mail: hongy@tll.org.sg

W. S. Dvorak
Camcore, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC 27695, USA

Published online: 06 September 2011

Keywords SSR · *Gmelina arborea* · Genetic diversity · Genetic structure · Plantation · Tree improvement

Introduction

Gmelina arborea Roxb. (family: Verbenaceae) is a valuable timber tree species native to India and large parts of Southeast Asia and is an important plantation species in many tropical areas around the world. Natural stands of Gmelina are found from 5° to 30°N at 50–1,300 m elevation in semi-deciduous forests in tropical/subtropical regions of Bangladesh, Cambodia, China (Yunnan and Guangxi Provinces), India, Laos, Myanmar, Nepal, (west) Pakistan, Sri Lanka, Thailand and Vietnam (Fig. 1). Known locally in Asia as Gamar or gumhar, *G. arborea* is a medium to large tree that grows up to 35 m in height and 3 m in diameter, often with large multiple stems and a wide spreading crown under natural conditions (Dvorak 2004). During its juvenile stage, *G. arborea* has smooth, whitish-gray to yellow-gray bark, which turns dark with white mottling in adults.

Gmelina arborea has developed into an important commercial plantation species because it is easy to establish from either seeds or cuttings, grows quickly, has favorable wood characteristics and therefore offers growers rapid returns on investment. The wood of *G. arborea* is highly durable and yields reasonable quantity and quality of pulp. Although it has a moderately low wood density (380–430 kg/m^3), the wood is relatively uniform, stable and light in colour (Dvorak 2004). In some regions it is commonly named "White Teak" because of the whiteness of the wood and the fact that its grain pattern resembles

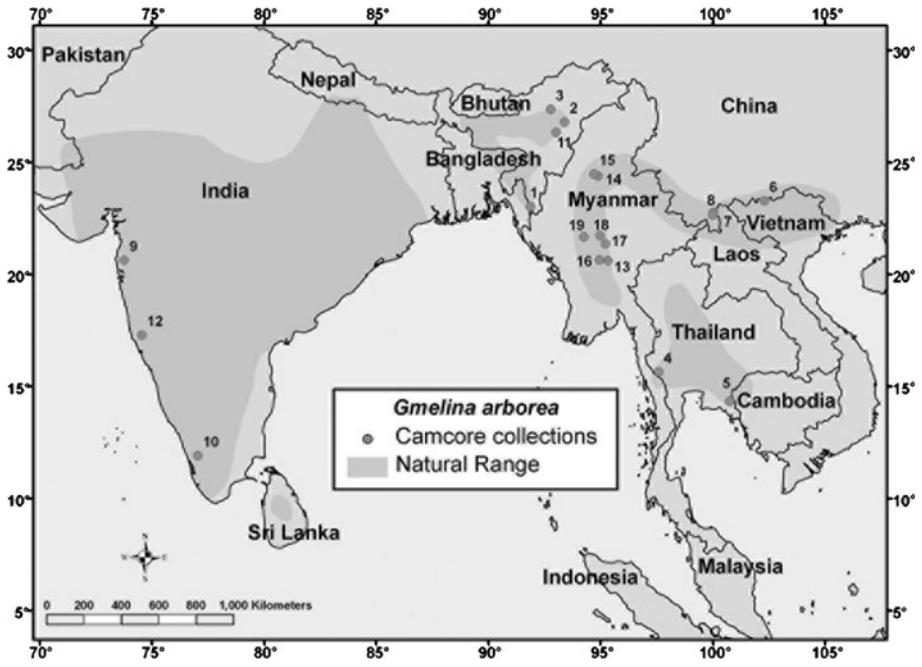

Fig. 1 Map of Asia showing the geographic range *of G. arborea* distribution as the grey area (adapted from Dvorak 2004) and locations (represented by *circles*) of natural populations under this study

that of natural Teak. These properties make *G. arborea* a multipurpose tree for a variety of uses like furniture, plywood, particle board and paper pulp. With the many positive attributes and a short rotation of 8–15 years (Hodge and Dvorak 2004), *G. arborea* was promoted by the Food and Agriculture Organization (FAO) Panel of Experts on Forest Genetic Resources (Lauridsen and Kjaer 2002) since 1969 to be an important tree species with high potential and utility. It was estimated that by 2020, 800,000 ha of *G. arborea* could be established in plantations in the tropical and subtropical region (Dvorak 2004). Despite the optimistic outlook on the potential commercial gain, the amount of improved seeds to develop *G. arborea* plantations is very limited. In order to make *G. arborea* plantation more commercially viable, improvements have to be made in growth, wood quality, wood productivity, pest resistance and uniformity.

Breeding programs will help achieve these improvements. Over the last 15 years, Camcore (International Tree Breeding & Conservation Program), North Carolina State University has made a number of mother-tree seed collections in China, India, Myanmar, and Thailand with local collaborators and its industrial members have now established 47 genetic tests in six countries. The most technically advanced Camcore members are now in their 2nd and 3rd generation of breeding (Camcore 2007). To complement and direct these breeding programs, there is the need to understand genetic diversity present in natural populations for this species. It is also crucial to examine the ancestries of various provenances and their relationship to each other across countries and geographic regions. The knowledge of genetic diversity will enable inclusion of a wide gene pool of desirable traits and a broad genetic base in a breeding program. It will also facilitate efficient seed exchange to maximize the diversity of genetic bases of *G. arborea* plantation programs.

One way to assess genetic diversity in forest trees is using genetic markers like microsatellites. Microsatellites are short tandemly repeated motifs found throughout the eukaryotic genomes (Zane et al. 2002). Known also as simple sequence repeats (SSRs) or simple tandem repeats (STRs), microsatellites vary among each other in terms of repeat sequence and unit length (mono- to hexa-nucleotide), thus they are characterized by a high degree of length polymorphism (Byrne et al. 1996; Lowe et al. 2000) when amplified by PCR primers located in flanking regions. Microsatellites are gaining popularity in genetic diversity analysis and marker-assisted breeding because of their high level of polymorphism and being co-dominant and neutral Mendelian markers. Microsatellite markers were used (Moriguchi et al. 2010) to study mating pattern of *Cyptomeria japonica*. Butcher et al. (2005) employed microsatellite markers to examine the genetic diversity among populations of *Eucalyptus benthamii*, which provided a foundation for further investigation into the genetic divergence and outcrossing rates of fragmented populations. In the context of forest fragmentation, White et al. (2002) evaluated pollen-mediated gene flow through microsatellite genotyping of *Swietenia humilis* from a highly fragmented forest mosaic. Another recent study looked into the genetic structure and mating system of *Manilkara huberi*, an endangered Amazonian tree species, by using several microsatellite loci (Azevedo et al. 2007).

In this study, we utilized microsatellite markers previously developed by us (Liao et al. 2010) to study genetic diversity in 534 individuals from 19 *G. arborea* populations in India, China, Thailand and Myanmar that were originally sampled by Camcore. The goals were to determine trends in genetic diversity in the species across its geographic range and identify populations of high and low genetic diversity to facilitate developing genetic base populations, breeding strategies and ex situ gene conservation activities.

Materials and methods

Plant materials

Leaf samples were collected from 534 individuals of *G. arborea* in ex situ progeny tests or nursery plantings that originated from 19 locations; 15 from natural stands and four from plantations in four countries—India, China, Thailand and Myanmar (Table 1). Plantation refers to trees planted by villagers who collected seeds from distant natural stands and planted seedlings around the village. The leaf samples taken from progeny of the original mother trees in 15 natural stands represent 1st generation breeding material. The genetic history of selections in the small plantings around villages is unknown. The 19 populations were sampled by Camcore between 1994 and 2003. Camcore grew seedlings in its greenhouse at North Carolina State University and sent leaf samples from three populations in India (Chanmari, Dimapur and Chessa) to Temasek Life Sciences Laboratory (TLL; Singapore) as dried leaves which were stored at 4°C until DNA extraction. Camcore member PT Sumalindo Lestari Jaya (Indonesia) also sent leaf samples to TLL that were collected in progeny trials of China (YiWu, ManNa and NanXi) and Thailand sources (Kanchanaburi and Chantaburi). Camcore member Pizano S. A. (Colombia) sent DNA extracts of the Myanmar and the remaining Indian populations.

Genomic DNA was extracted from approximately 80 mg fresh leaf tissue or 30–40 mg dried leaf tissue. The tissue was ground in liquid nitrogen by using mortar and pestle to

Table 1 Details of the 19 *G. arborea* populations used in this study

Country of origin	Population of origin	Type	Code on map	Latitude	Longitude	No. of samples
India	Kasa	NS	9	20°17′N	73°15′E	63
	Sholayar	NS	10	10°22′N	76°37′E	27
	Nowgong	NS	3	27°00′N	93°50′E	33
	Amboli	NS	12	15°54′N	73°46′E	33
	Chanmari	NS	1	22°24′N	92°24′E	30
	Dimapur	NS	11	26°29′N	94°18′E	30
	Chessa	NS	2	27°05′N	93°38′E	30
Myanmar	Sin Thaut	NS	13	19°47′N	96°23′E	15
	Kyun Taw	P	14	23°52′N	95°55′E	20
	Waibon	P	15	23°58′N	95°43′E	20
	Ladagyi	NS	16	19°49′N	95°58′E	19
	Kin Tha	P	18	20°36′N	96°16′E	20
	Pyin Oo Lwin	NS	17	21°00′N	96°00′E	30
	Poppa	NS	19	20°56′N	95°14′E	30
China	NanXi	P	7	22°40′N	103°56′E	27
	ManNa	NS	8	22°07′N	101°28′E	26
	YiWu	NS	6	21°59′N	101°27′E	26
Thailand	Kanchanaburi	NS	4	14°25′N	98°50′E	28
	Chantaburi	NS	5	13°00′N	102°15′E	27

Fifteen of the collections came from natural stands (*NS*) and four from plantations (*P*) around the villages that originated from natural stands

fine powder. gDNA was isolated with DNeasy® Plant Mini Kit (QIAGEN GmbH, Hilden, Germany), following the protocol provided by QIAGEN. The quality and yield of isolated gDNA was verified with 0.8% agarose gel electrophoresis visualized by ethidium bromide staining. DNA was run alongside with λ DNA-*Hin*dIII Digest (New England BioLabs Inc., Ipswich, MA, USA). Concentration of the isolated gDNA was obtained using ND-1000 spectrophotometer (Nanodrop Technologies, Rockland, Delaware, USA).

Microsatellite analysis

Ten microsatellite loci previously developed for *G. arborea* (Liao et al. 2010) were used to genotype all individuals. Forward primers were labeled with 6-Carboxyfluorescein (6-FAM). PCR amplifications were performed in 96-well microtitre plates using either PTC-100® or PTC-200 DNAEngine® Thermal Cycler machines (MJ Research, Massachusetts, USA). PCR reaction (20 µl) contained 1× PCR buffer [10 mM Tris–HCl (pH 8.4), 1.5 mM $MgCl_2$] (Qiagen GmbH, Strasse, Germany), 0.1 mM of each dNTP and 1U Taq polymerase. For individual microsatellite marker analysis, 0.1 µM of each forward and reverse primer and approximately 40 ng of DNA template was used. PCR products were separated by an ABI 3730XL DNA sequencer (Applied Biosystems, California, USA) and automatically sized using GeneScan 500 ROX size standard (Applied Biosystems, Foster City, USA) and analyzed by GeneMapper 3.7 (Applied Biosystems, Foster City, USA).

Data analysis

Microsatellite loci were characterized mainly by MS Excel add-in, "Genetic Analysis in Excel 6" GeneAIEx 6 (Peakall and Smouse 2006). Allele information for ten loci from 534 individuals was compiled as co-dominant data. Individuals were classified according to their population and country of origin. Countries were referred to as "regions" in the analysis. Intrapopulation genetic diversity was characterized by the average number of alleles per locus (N_a), the number of effective alleles per locus (N_e), observed heterozygosity (H_o), expected heterozygosity (H_e) under the Hardy–Weinberg expectation, fixation index (F) and Wright's F-statistics. To test for the departure from Hardy–Weinberg equilibrium, a score test was conducted by considering both the hypotheses of heterozygote deficiency and excess (Raymond and Rousset 1995). The p-value was calculated by a Markov chain method, with 20 batches and a demmorization number of 10,000. Calculation of the Analysis of Molecular Variance (AMOVA) test and Wright's fixation index, F_{st}, were performed with Arlequin 3.11 to examine the partitioning of genetic variation. The number of migrants per generation, N_m, was calculated as $(1 - F_{st})/4\ F_{st}$ (Wright 1951). F_{st} describes the reduction of heterozygosity within populations relative to the total population (Lowe et al. 2004). A Principal Coordinate Analysis (PCA) was performed by GenAlEx to visualize the genetic distance between populations in two hierarchies. First, a primary PCA was conducted at the regional level (based on country of origin). A secondary PCA was then conducted for each cluster obtained from the primary PCA. In addition, Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering were performed with the program NTSYS-pc version 2.0 (Rohlf 1998) based on pairwise Nei's unbiased genetic identity calculated between populations with GenAlEx.

Results

Genetic diversity and Hardy–Weinberg equilibrium

The analysis of 534 *G. arborea* individuals using ten microsatellite loci revealed a total of 164 alleles (Table 2), with the mean number of alleles detected per locus as 16.4 (range 13–23). All loci were polymorphic for the species. The frequencies of alleles for all loci were 0.5 or below (range 0.001–0.500). At the species level, the mean expected heterozygosity (H_e) estimated for each locus was higher than the mean observed heterozygosity (H_o) for all loci. The overall average of H_e for each locus at the species level was 0.83, ranging from 0.65 (2E7) up to 0.91 (2F4 and 2F10) among loci; while overall average of H_o for each locus was 0.56, ranging from 0.21 (2E7) up to 0.75 (2G12) among loci. As a result, all loci were positive for fixation index, which was calculated as $F = (H_e - H_o)/H_e$. For all loci, effective alleles (N_e) were all lower than number of alleles (N_a). This discrepancy suggests uneven frequencies among the alleles.

When genetic diversity parameters were compared among populations (summarized in Table 3), the mean observed heterozygosity (H_o) for all populations was 0.60, with the highest values from Amboli (India, 0.67), Sin Thaut (Myanmar, 0.65), Sholayar (India, 0.64) and Kasa (India, 0.64) populations. The lowest H_o values were from ManNa (China, 0.42), Nanxi (China, 0.45) and Poppa (Myanmar, 0.47) populations. The mean number of alleles per locus for all 19 populations studied was 5.96, ranging from 3.70 (Chessa) to 9.60 (Kassa) (Table 3). The mean effective number of alleles (N_e) was lower, averaging 3.19 for all populations. N_e estimates the number of equally frequent alleles in an ideal population under random mating (Kimura and Crow 1964), therefore, the lower value that we obtained here indicates the presence of rare alleles within each population. Five out of nineteen populations—Nowgong, Kyun Taw, Kanchanaburi, NanXi and ManNa—displayed no private alleles, while Pyin Oo Lwin (from Myanmar) had the most number of private alleles, with six in total.

Test of conformity to Hardy–Weinberg proportions conducted for each polymorphic locus in each population found that 83 out of 187 tests (44%) showed a significant

Table 2 Summary of the ten microsatellite loci in all *G. arborea* samples

	N	Na	Ne	Ho	He	F
1H4	534	15	8.08	0.70	0.88	0.20
1F4	533	15	3.12	0.39	0.68	0.43
1H10	534	19	7.62	0.59	0.87	0.32
1H12	534	13	6.82	0.51	0.85	0.41
2E7	512	14	2.87	0.21	0.65	0.67
1G8	534	16	5.84	0.57	0.83	0.32
2F4	534	23	10.66	0.63	0.91	0.30
2G12	534	16	6.87	0.75	0.85	0.13
2F10	533	19	11.09	0.70	0.91	0.23
1C12	534	14	6.60	0.52	0.85	0.39
Mean (±SE)	531.60 (2.18)	16.40 (0.97)	6.96 (0.85)	0.56 (0.05)	0.83 (0.03)	0.34 (0.05)

n total number of samples analyzed for each locus, N_a number of alleles per locus, N_e number of effective allele per locus, H_o observed heterozygosity, H_e expected heterozygosity, *F* fixation index, *Mean (±SE)* averaged across all loci with standard error

Table 3 Estimates of genetic parameters in 19 populations of *G. arborea* averaged over all loci

Population	N	N_a (±SE)	N_e (±SE)	N_p	H_o (±SE)	H_e (±SE)	F (±SE)	Departure from HWE (score test)
Kasa	63	9.60 (0.58)	4.03 (0.47)	2	0.64 (0.06)	0.71 (0.04)	0.13 (0.05)	**
Sholayar	27	7.10 (0.60)	3.64 (0.41)	3	0.64 (0.06)	0.68 (0.05)	0.09 (0.07)	**
Nowgong	33	7.10 (0.60)	3.76 (0.43)	0	0.62 (0.07)	0.70 (0.04)	0.14 (0.08)	**
Amboli	33	9.20 (0.61)	4.72 (0.45	1	0.67 (0.06)	0.76 (0.03)	0.14 (0.06)	**
Chanmari	30	3.80 (0.59)	2.54 (0.39)	3	0.49 (0.10)	0.49 (0.09)	−0.01 (0.04)	ns
Dimapur	30	3.90 (0.43)	2.44 (0.27)	1	0.51 (0.09)	0.53 (0.06)	0.06 (0.12)	ns
Chessa	30	3.70 (0.45)	2.34 (0.29)	3	0.46 (0.07)	0.52 (0.05)	0.08 (0.13)	*
Sin Thaut	15	6.30 (0.78)	3.97 (0.61)	2	0.65 (0.08)	0.66 (0.07)	0.02 (0.06)	*
Kyun Taw	20	6.00 (0.80)	3.04 (0.46)	0	0.57 (0.09)	0.58 (0.08)	0.03 (0.05)	ns
Waibon	20	6.00 (0.78)	3.35 (0.50)	1	0.62 (0.07)	0.61 (0.07)	−0.02 (0.04)	ns
Ladagyi	19	5.10 (0.81)	3.31 (0.58)	1	0.58 (0.09)	0.59 (0.08)	0.02 (0.04)	ns
Kin Tha	20	6.70 (0.76)	3.68 (0.60)	1	0.58 (0.07)	0.64 (0.07)	0.11 (0.03)	**
Pyin Oo Lwin	30	7.50 (0.93)	3.31 (0.51)	6	0.60 (0.07)	0.62 (0.06)	0.05 (0.05)	*
Poppa	30	6.50 (0.31)	2.70 (0.41)	3	0.47 (0.06)	0.56 (0.06)	0.18 (0.04)	**
Kanchanaburi	28	4.80 (0.59)	2.96 (0.43)	0	0.55 (0.10)	0.56 (0.08)	0.04 (0.09)	ns
Chantaburi	27	6.60 (0.72)	3.08 (0.47)	2	0.51 (0.09)	0.60 (0.07)	0.19 (0.08)	**
NanXi	27	3.80 (0.33)	2.57 (0.32)	0	0.45 (0.07)	0.55 (0.06)	0.15 (0.10)	ns
ManNa	26	4.60 (0.50)	2.23 (0.32)	0	0.42 (0.08)	0.46 (0.08)	0.10 (0.08)	*
YiWu	26	5.00 (0.67)	2.87 (0.46)	2	0.51 (0.09)	0.55 (0.08)	0.07 (0.07)	*
Overall populations	534	5.96 (0.19)	3.19 (0.11)	1.63	0.55 (0.02)	0.60 (0.02)	0.08 (0.02)	**

n number of samples from each population, N_a number of alleles per locus, N_e number of effective allele per locus, N_p number of private alleles, H_o observed heterozygosity, H_e expected heterozygosity, *F* fixation index

Significant departures from Hardy–Weinberg expectations at * $P < 0.05$; ** $P < 0.01$ and *ns* not significant

departure from the expected proportions at the 5% level. Global Hardy–Weinberg tests (Score test) revealed that deficit in heterozygotes was significant in 12 out of 19 populations. On the other hand, 17 out of 19 populations had positive fixation index ranging from 0.02 to 0.19. Since only two populations had negative fixation index values, the average fixation index across populations was positive. Together with the much lower values of N_e (3.19) as compared to N_a (5.96), these results suggest heterozygote deficit in *G. arborea*.

Genetic structure

The AMOVA procedure was carried out to partition the genetic variation among populations and regions. Results from the AMOVA analysis showed that the majority of genetic variation resided within populations (69%), whereas the variation among region and among populations within the same region was only 21 and 10% respectively (Table 4). All variance components were found to be highly significant ($P < 0.001$). This implies that although each population still retained a considerable amount of genetic variation, they had significantly differentiated from each other at a population and region level. When populations were analyzed regionally (based on country of origin), the within population component increased to above 80% (Table 4). This further reiterated the fact that populations from the same region share more genetic similarities than those in different regions. The F_{st} value among all Gmelina populations is 0.31, with the overall level of inferred gene flow (N_m) estimated as 0.56 individual per generation among populations. Among the four regions, N_m varied from 4.06, 4.56 and 5.19 in China, Thailand and Myanmar respectively, to 1.01 in India. This indicates that India has a lower level of gene flow among its populations than the other regions.

Table 4 Population structure of *G. arborea* as shown by AMOVA with three levels of hierarchy—among regions, among populations and within populations

Source of variation	F_{ST}	N_m	*df*	SSD	Variance component	Percentage of variation (%)	*P* value
Overall	0.31	0.56					
Among regions			3	744.11	0.92	20.62	<0.001
Among pops			15	429.81	0.46	10.35	<0.001
Within pops			1,049	3,214.36	3.06	69.03	<0.001
India	0.20	1.01					
Among pops			6	327.53	0.75	19.86	<0.001
Within pops			485	1,462.12	3.02	80.14	
Myanmar	0.05	5.19					
Among pops			6	57.69	0.15	4.63	<0.001
Within pops			301	928.72	3.09	95.37	
Thailand	0.05	4.56					
Among pops			1	11.94	0.16	5.25	<0.01
Within pops			108	318.86	2.95	94.75	
China	0.06	4.06					
Among pops			2	22.29	0.16	5.77	<0.001
Within pops			155	408.80	2.64	94.23	

F_{st} Wright's fixation index, N_m number of migrants per generation, *df* degree of freedom, *SSD* sum of squared deviation

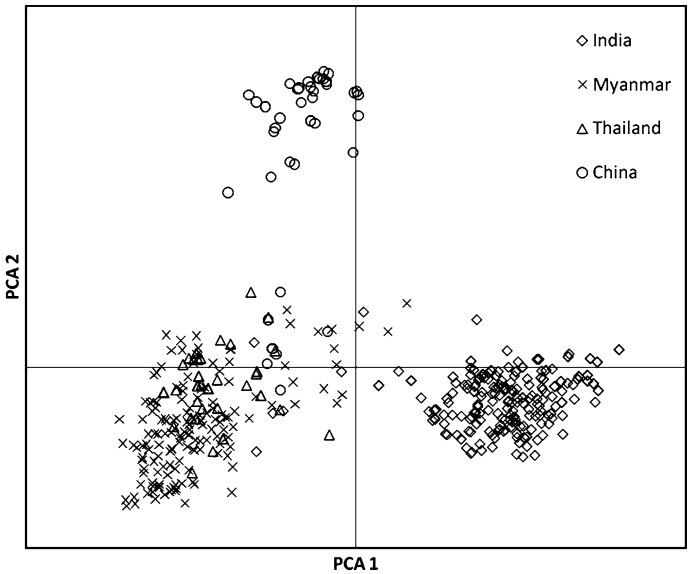

Fig. 2 Primary Principle Coordinate Analysis (PCA) of genetic distances among *G. arborea* individuals from four regions

The PCA analysis further illustrates the regional differences. Populations were clustered into three distinct groups—(1) China, (2) Thailand-Myanmar and (3) India (Fig. 2). Apart from the overlap of a few individuals from China and India with the Thailand-Myanmar cluster, the groups are clearly segregated. A secondary-level PCA was done for each of these groups to disect the within-group genetic similarity. No distinct clustering was observed within the China group. Within the Thailand-Myanmar group, there is an indication of aggregation among individuals from the Thai populations Kanchanaburi and Chantaburi. Clustering was more apparent in India, where three subgroups were observed. Individuals from Chanmari formed a subgroup (east coast population), individuals from Dimapur and Chessa formed another subgroup (highland populations), and the remaining individuals from other populations formed the third subgroup (all western coast populations, except for Nowgong) (Fig. 3).

The UPGMA dendrogram revealed clusters (Fig. 4) which were consistent with the groupings by the first and secondary-level PCA analysis. Pairwise population matrix of Nei's Unbiased Genetic Identity showed that at the population level, genetic identity between the populations ranged from 0.07 (see Table 5 Chessa-Ladagyi) to 1.00 (see Table 5 Amboli-Kasa, Amboli-Nowgong and Kyun Taw-Waibon).

Discussion

Genetic diversity

The genetic diversity of *G. arborea* was evaluated using ten microsatellite markers. The number of loci utilized in this study is considered comparable to numbers used in other studies on tropical timber species—five loci in *Dinizia excelsa* (Dick et al. 2003), seven

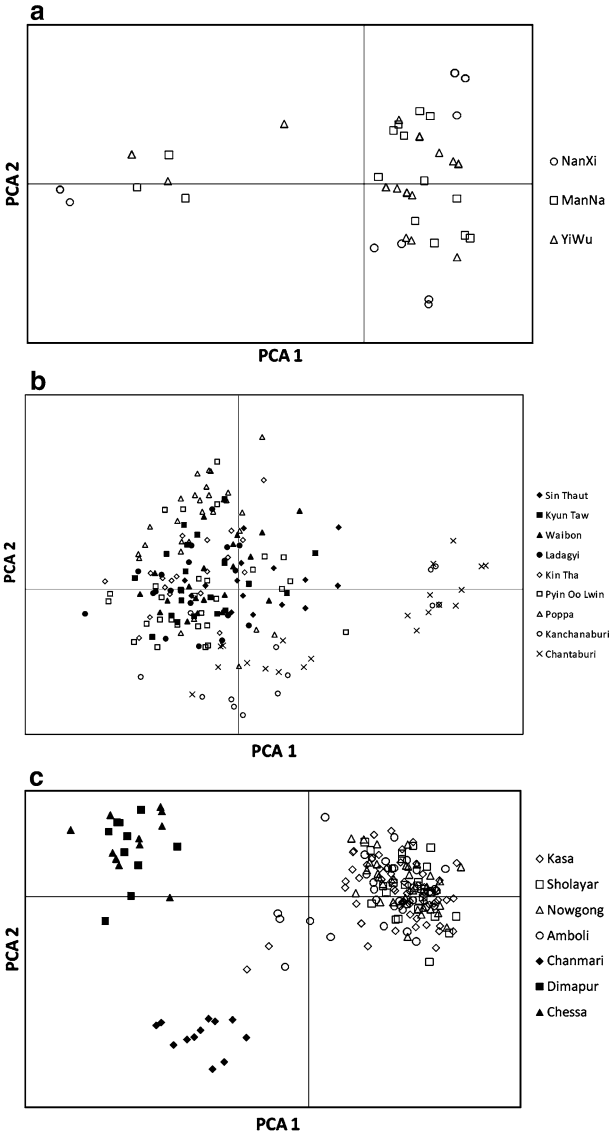

Fig. 3 Secondary Principle Coordinate Analysis (PCA) of genetic distances among *G. arborea* individuals based on the three clusters obtained from the primary PCA. **a** Populations from China; **b** populations from Thailand and Myanmar; **c** populations from India

loci in *M. huberi* (Azevedo et al. 2007), eight loci in *S. macrophylla* (Lemes et al. 2003) and 15 loci in *Tectona grandis* (Fofana et al. 2008).

Our finding of genetic diversity parameters as average values of all loci for *G. arborea* ($N_a = 16.4$; $H_o = 0.56$; $H_e = 0.83$) were comparable to those for Teak (*T. grandis*), its geographically closest relative that had been examined with microsatellite markers ($N_a = 12.6$; $H_o = 0.52$; $H_e = 0.49$) (Fofana et al. 2008). The heterozygosity found in our study was much higher than the heterozygosity detected for *G. arborea* using inter-simple

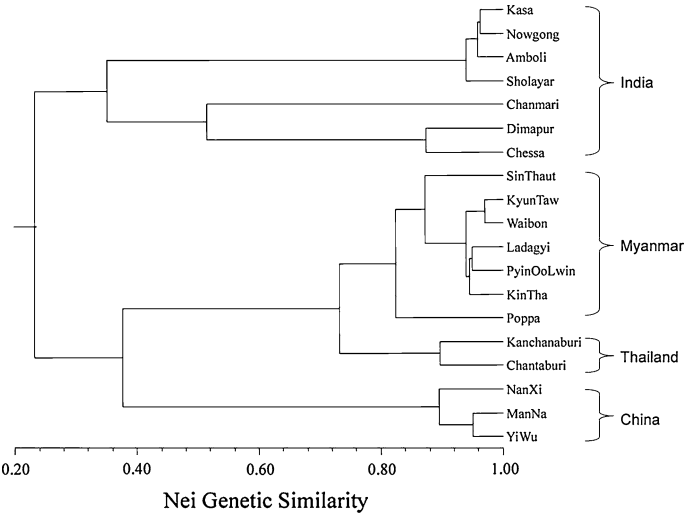

Fig. 4 Dendrogram from UPGMA cluster analysis based on Nei's genetic identity among all *G. arborea* populations

sequence repeat (ISSR) markers (Naik et al. 2009), most probably due to the higher number of individuals used in our study or higher level of polymorphism revealed by microsatellite markers, although strict comparison is difficult. Also, Naik et al. (2009) included samples from plantations and fragmented forests which might have experienced genetic drift or bottlenecks, thus the genetic diversity obtained was lower. Based on our study, *G. arborea* also exhibited an overall higher level of allelic richness but slightly lower H_o and higher H_e values as compared to the results from 106 intraspecific genetic diversity studies in wild plants based on microsatellite markers ($N_a = 9.9$; $H_o = 0.58$; $H_e = 0.61$) (Nybom 2004). Similarly, *G. arborea* showed a higher allelic richness, lower H_o and higher H_e when compared to other tropical tree species, such as *Carapa guianensis* ($N_a = 7.6$; $H_o = 0.68$; $H_e = 0.67$; Dayanandan et al. 1999), *Dicorynia guianensis* ($N_a = 7.3$; $H_o = 0.62$; $H_e = 0.63$; Latouche-Halle et al. 2003), *D. excelsa* ($N_a = 18.2$; $H_o = 0.65$; $H_e = 0.61$; Dick et al. 2003) and *M. huberi* ($N_a = 6.4$; $H_o = 0.69$; $H_e = 0.81$; Azevedo et al. 2007). The high level of allelic richness in *G. arborea* confirmed that the ten microsatellite markers in this study have identified sufficient genetic variation.

All loci under this study had lower H_o values than those for H_e, which translate into positive fixation index (Table 2). For the nineteen populations we studied, 44% of population/locus tests were found significantly deviated from Hardy–Weinberg equilibrium. Seventeen populations had positive fixation index values ranging from 0.02 to 0.19. There were also the much lower values of N_e (3.19) as compared to N_a (5.96). All these results together strongly suggest heterozygote deficit in *G. arborea*, in other words, nonrandom mating favoring inbreeding. This is inconsistent with the finding of self incompatibility in *G. arborea* (Bolstad and Bawa 1982) and the suggestion of a mixed mating system which favors out-crossing (Raju and Rao 2006). We won't be able to reach a strong conclusion since several other possibilities can not be excluded. Pollen limitation may encourage selfing. Biparental inbreeding could also have occurred when two genetically related individuals were crossed. Type and availability of pollinators might also change pollination

Table 5 Nei's genetic similarity between the 19 *G. arborea* populations

	Kasa	Sholayar	Nowgong	Amboli	Chanmari	Dimapur	Chessa	Sin Thaut	Kyun Taw
Sholayar	0.99								
Nowgong	0.99	0.97							
Amboli	1.00	0.97	1.00						
Chanmari	0.41	0.37	0.37	0.43					
Dimapur	0.35	0.30	0.32	0.38	0.53				
Chessa	0.35	0.30	0.35	0.40	0.52	0.89			
Sin Thaut	0.37	0.32	0.31	0.39	0.43	0.14	0.12		
Kyun Taw	0.33	0.31	0.30	0.38	0.32	0.11	0.09	0.94	
Waibon	0.32	0.29	0.30	0.37	0.33	0.11	0.11	0.96	1.00
Ladagyi	0.32	0.29	0.29	0.37	0.32	0.10	0.07	0.90	0.98
Kin Tha	0.32	0.29	0.29	0.38	0.31	0.11	0.09	0.91	0.98
Pyin Oo Lwin	0.33	0.29	0.29	0.39	0.32	0.13	0.11	0.90	0.97
Poppa	0.29	0.23	0.23	0.33	0.26	0.13	0.10	0.78	0.84
Kanchanaburi	0.28	0.25	0.26	0.34	0.30	0.17	0.12	0.78	0.81
Chantaburi	0.32	0.28	0.27	0.35	0.34	0.18	0.12	0.76	0.72
NanXi	0.23	0.20	0.20	0.25	0.27	0.17	0.16	0.41	0.37
ManNa	0.19	0.15	0.18	0.20	0.17	0.09	0.11	0.38	0.31
YiWu	0.23	0.19	0.21	0.25	0.20	0.11	0.11	0.42	0.33

	Waibon	Ladagyi	Kin Tha	Pyin Oo Lwin	Poppa	Kanchanaburi	Chantaburi	NanXi	ManNa
Sholayar									
Nowgong									
Amboli									
Chanmari									
Dimapur									

Table 5 continued

	Waibon	Ladagyi	Kin Tha	Pyin Oo Lwin	Poppa	Kanchanaburi	Chantaburi	NanXi	ManNa
Chessa									
Sin Thaut									
Kyun Taw									
Waibon									
Ladagyi	0.99								
Kin Tha	0.98	0.99							
Pyin Oo Lwin	0.96	0.98	0.98						
Poppa	0.84	0.87	0.89	0.89					
Kanchanaburi	0.81	0.82	0.78	0.83	0.70				
Chantaburi	0.72	0.73	0.71	0.75	0.68	0.92			
NanXi	0.40	0.40	0.45	0.41	0.38	0.49	0.53		
ManNa	0.34	0.33	0.38	0.35	0.30	0.41	0.45	0.93	
YiWu	0.35	0.35	0.39	0.38	0.29	0.44	0.44	0.90	0.97

pattern. More studies are necessary to have a better understanding on reasons for reduced heterozygosity in *G. arborea*.

Genetic structure

Our study showed a significant genetic differentiation among *G. arborea* populations, similar to other tropical trees which occur in low density forests (Collevatti et al. 2001; Latouche-Halle et al. 2003; Lemes et al. 2003). A considerable amount of variation was observed among regions, which accounted for 21% of the total variation. In natural populations, *G. arborea* is distributed sparsely, with one to several mature trees found per km^2 in a mixed broadleaf forests (Dvorak 2004). Low abundance, coupled with isolation of populations by limited gene flow may be the driving force behind population differentiation (Alves et al. 2007). Also, as mentioned previously, microsatellites have a high mutation rate. This marker could be more sensitive than other molecular markers in capturing differentiation among recently isolated populations. Despite the significant level of among-population and among-region variation, the amount of variation within population in this study was higher than Naik et al. (2009) observed in another study of *G. arborea*. Again, this reflected that the fragmented forests and plantation that had been sampled might have undergone genetic drift or bottlenecks.

The high level of regional diversity is also evident in both the PCA and the UPGMA dendrogram produced by cluster analysis, where populations are clustered mostly according to their geographic regions. Based on the UPGMA dendrogram and the secondary-level PCA analysis, Indian populations could be subgrouped according to their geographical proximity into three clusters: highland, west coast and east coast subgroups. Groups of populations are distinct possibly due to restricted gene flow over large distances. Isolation by distance has been demonstrated for several tropical trees (Azevedo et al. 2007; Lemes et al. 2003), though it is yet to be confirmed in *G. arborea*. The only exception to the consensus between geographical and genetic clustering is the population from Nowgong—located in the Assam state, south of the eastern Himalayas—which did not cluster together with Chessa and Dimapur despite its geographical proximity. The reason for clustering of the Nowgong population among the west coast populations remains unclear. The only other example where populations from the highland areas in India shared high degree of genetic similarity with coastal populations was in *Phyllanthus amarus* from the family Euphorbiaceae (Jain et al. 2003). It could be possible that similar to *P. amarus*, the *G. arborea* populations from Nowgong originates from the west coast, most likely from Kasa since these two populations shared the most genetic similarity. However, the mechanism in which individuals from Kasa area were introduced to Nowgong is still speculative at this stage. Further studies would be needed to provide more concrete explanation behind this exceptional case.

For China group and Myanmar/Thailand group, many adjacent populations were clustered together (i.e. Waibon and Kyun Taw; ManNa and YiWu). However, geographical proximity did not explain the clustering of several other cases. For example, Pyin Oo Lwin (Myanmar) has a shorter distance to Man Na (China) than to Kanchanaburi (Thailand). Yet it is included in the Thailand-Myanmar subgroup but not in the China subgroup. Apart from geographic distance, restricted gene flow could also be attributed to isolation of populations caused by the presence of a barrier to gene flow (Butcher et al. 2005). Once the barrier is established between isolated populations, spatially variable selection or localized genetic drift could both result in genetic differentiation. The Gaoligong mountain range, which borders Yunnan Province in China and Myanmar is located between Pyin Oo Lwin

and Man Na. It might act as an effective barrier to gene flow since the mountain terrain is not a habitat for *G. arborea*. On the other hand, the terrain between Pyin Oo Lwin and Kanchanaburi is mostly lowland area which poses an ineffective barrier to gene flow. Therefore, geologic barriers to gene flow could be another important factor influencing the genetic differentiation in *G. arborea*.

Another population pair that was inconsistent with the isolation by distance model was populations from Kanchanaburi and Chantaburi in Thailand. Both populations were more than 400 km apart and yet shared the most genetic similarity with each other. One possible reason is the presence of similar selection pressures among locations. This applies when the environmental factors, such as topology or forest type, are similar in such a way that genotypes from one environment are more likely to survive in the same environment at different localities simply due to advantageous adaptation (McGowen et al. 2001; Ramakrishnan et al. 2006). Kanchanaburi and Chantaburi are located at approximately the same latitude, which presumably results in similar climatic conditions and forest types. Therefore, genetic differentiation between populations from Kanchanaburi and Chantaburi might be prevented through selection for corresponding adaptations. Another possibility is human intervention, such as carrying Gmelina seeds from one geographic location to another.

Impact on tree breeding

A practical utility of information on genetic diversity and structure is to facilitate planning seed collections for breeding and also the genetic composition of seed orchards. The higher the genetic diversity, the more variation there is for selection and improvement (Lee et al. 2000). Based on our study, populations in Thailand/Myanmar and China were genetically more similar to each other than with India populations. Therefore, a seed collection effort to maximize genetic diversity should include those populations with higher level of heterozygosity such as Amboli (India, $H_o = 0.67$), Sin Thaut (Myanmar, $H_o = 0.65$), Sholayar (India, $H_o = 0.64$) and Kasa (India, $H_o = 0.64$). For breeding through controlled pollination, parent plants should come from populations that are genetically distinct from each other (see Table 5 for a summary of Nei genetic identity between populations). We recommend a systematic sampling which should include populations from all three primary groups—(1) India, (2) China, and (3) Thailand/Myanmar.

For seed orchards, we suggest that a mixture of selections made in progeny trials from different geographic groups, placed in a clonal seed orchard together, will promote outcrossing vigor and improvement in productivity. One such example is the open-pollinated clonal seed orchard of *G. arborea* operated by Sumalindo (Indonesia) with selections from Thailand and Costa Rica (probably of Indian origin). The realized field gain from the orchard was 25% in wood productivity over local seed orchard of unknown diversity. Alternatively, seed collections of Gmelina from various countries (Thailand/Myanmar, China, India) can be placed in sublines and bred for several generations before being placed together in a breeding orchard to promote outcrossing vigor.

Impact on conservation

Unlike timber trees such as *Caesalpinia echinata* and *C. brasiliense*, *G. arborea* is not an endangered species (Cardoso et al. 1998; Collevatti et al. 2001; Dvorak 2004). At species level, *G. arborea* has a satisfactory diversity level with average observed heterozygosity ($H_o = 0.56$) comparable to other tropical tree species. Therefore, in terms of conservation,

the existence of this species is not a concern. Rather, the long-term evolutionary fitness and future commercial gain from the conserved diversity (Lee et al. 2000) will become the priority. By using results from international provenance trials, together with genetic diversity assessments, practical *ex situ* conservation programs can be developed. Considering significant genetic structuring detected in this study, the selection of candidates for ex situ conservation should take the genetic clusters into account, in order to ensure that all major clusters are represented. For example one ex situ conservation strategy might be to choose a few populations in each genetic cluster with maximum genetic diversity, fitness and good commercial potential to capture broad adaptability. In addition, one might heavily sample the Myanmar populations for their good growth (Hodge and Dvorak 2004) but also include the fringe and distinct populations in southwestern China because some are critically endangered and might include useful rare alleles. In this case, individuals from YiWu would be a good candidate for conservation since the population has moderately high level of genetic diversity and is the only one in southwestern China found to harbour private alleles. The approach for using field results coupled with molecular information was also practiced in the case of *E. urophylla* from Indonesia (Dvorak et al. 2008). The decision to include or exclude any populations for conservation should be based upon well-supported and comprehensive data on the productivity and genetic similarity among populations.

The understanding of genetic diversity of *G. arborea* forms a significant platform on which further studies can build upon. These highly polymorphic microsatellite markers could be further used to construct a genetic map, followed by marker assisted breeding for *G. arborea*. For plantations with vegetative clones, genetic profiling can help quality control of plantation. These microsatellite markers can be used to study mating system of *G. arborea* and to clarify if inbreeding indeed happens for *G. arborea*. In one recent study, paternity analysis of *M. huberi* populations showed that mating among relatives contributed considerably to the inbreeding that was observed (Azevedo et al. 2007). The study sampled and analyzed both adult trees and seedlings to compute several key parameters such as outcrossing rate, paternity correlation, self-pollination correlation and number of pollen donor per mother tree. A similar study could be conducted for *G. arborea.* Coupled with research on its reproductive ecology—such as pollination, pollen vectors, fertilization, seed dispersal and seed vectors—we can better understand the gene flow and migration pattern in this species. Studies can also compare continuous and fragmented forest to determine the effect of disturbance on genetic diversity. This is particularly important in strategizing conservation plans.

In conclusion, there is a moderately high level of genetic diversity in *G. arborea* populations that we studied. An excess of homozygotes was found within most *G. arborea* populations. Genetic diversity was observed to segregate geographically according to regions. Populations from China, Thailand/Myanmar and India form three genetically distinct groups. There was further geographical clustering among Indian populations, with only one exception. Findings from this study can be immediately applied to plant breeding programs and conservation planning. Future studies into the mating system, population fragmentation and evolution of this species can also build upon the results obtained.

Acknowledgments Temasek Life Sciences Laboratory provided funding to this research. We thank Mr. Willi Woodbridge, Camcore, USA for his development of the Gmelina map. We thank Mr. Mike Tighe, (formerly Camcore, USA), Mr. Arif Purwanto (Sumalindo, Indonesia) and Mr. Miguel Rodriguez (Pizano, Colombia) for sending plant materials or DNAs for this study.

References

Alves RM, Sebbenn AM, Artero AS, Clement C, Figueira A (2007) High levels of genetic divergence and inbreeding in populations of cupuassu (*Theobroma grandiflorum*). Tree Genet Genomes 3(4):289–298

Azevedo NCR, Kanashiro M, Ciampi AY, Grattapaglia D (2007) Genetic structure and mating system of *Manilkara huberi* (Ducke) A. Chev., a heavily logged Amazonian timber species. J Hered 98(7):646–654

Bolstad PV, Bawa KS (1982) Self incompatibility in *Gmelina arborea* L. (Verbenaceae). Autoincompatibilidad en *Gmelina arborea* L. (Verbenaceae). Silvae Genetica 31(1):19–21

Butcher PA, Skinner AK, Gardiner CA (2005) Increased inbreeding and inter-species gene flow in remnant populations of the rare *Eucalyptus benthamii*. Conserv Genet 6(2):213–226

Byrne M, Marquezgarcia MI, Uren T, Smith DS, Moran GF (1996) Conservation and genetic diversity of microsatellite loci in the genus Eucalyptus. Aust J Bot 44(3):331–341. doi:10.1071/BT9960331

Camcore (2007) The Gmelina breeding program at PT Sumalindo Lestari Jaya. Camcore annual report. Department of Forestry & Environmental Science, NC State University, Raleigh, North Carolina, USA

Cardoso MA, Provan J, Powell W, Ferreira PCG, De Oliveira DE (1998) High genetic differentiation among remnant populations of the endangered *Caesalpinia echinata* Lam. (Leguminosae-Caesalpinioideae). Mol Ecol 7(5):601–608

Collevatti RG, Grattapaglia D, Hay JD (2001) Population genetic structure of the endangered tropical tree species *Caryocar brasiliense*, based on variability at microsatellite loci. Mol Ecol 10(2):349–356

Dick CW, Etchelecu G, Austerlitz F (2003) Pollen dispersal of tropical trees (*Dinizia excelsa*: Fabaceae) by native insects and African honeybees in pristine and fragmented Amazonian rainforest. Mol Ecol 12(3):753–764

Dvorak WS (2004) World view of *Gmelina arborea*: opportunities and challenges. New Forests 28(2):111–126

Dvorak WS, Hodge GR, Payn KG (2008) The conservation and breeding of *Eucalyptus urophylla:* A case study to better protect important populations and improved productivity. South Forests 70(2):77–85

Fofana IJ, Lidah YJ, Diarrassouba N, N'guetta SPA, Sangare A, Verhaegen D (2008) Genetic structure and conservation of Teak (*Tectona grandis*) plantations in Côte d'Ivoire, revealed by site specific recombinase (SSR). Trop Conserv Sci 1(3):279–292

Hodge GR, Dvorak WS (2004) The CAMCORE international provenance/progeny trials of *Gmelina arborea*: genetic parameters and potential gain. New Forests 28(2):147–166

Jain N, Shasany AK, Sundaresan V, Rajkumar S, Darokar MP, Bagchi GD, Gupta AK, Kumar S, Khanuja SPS (2003) Molecular diversity in *Phyllanthus amarus* assessed through RAPD analysis. Curr Sci 85(10):1454–1458

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population genetics 49(4):725–738

Latouche-Halle C, Ramboer A, Bandou E, Caron H, Kremer A (2003) Nuclear and chloroplast genetic structure indicate fine-scale spatial dynamics in a neotropical tree population. Heredity 91(2):181–190

Lauridsen EB, Kjaer ED (2002) Provenance research in *Gmelina arborea* Linn., Roxb. A summary of results from three decades of research and a discussion of how to use them. Int For Rev 4(1):20–29

Lee SL, Wickneswari R, Mahani MC, Zakri AH (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(2):213–224

Lemes MR, Gribel R, Proctor J, Grattapaglia D (2003) Population genetic structure of mahogany (*Swietenia macrophylla* King, Meliaceae) across the Brazilian Amazon, based on variation at microsatellite loci: implications for conservation. Mol Ecol 12(11):2875–2883

Liao MH, Li CH, Qi JF, Hong Y (2010) Isolation and characterization of 11 polymorphic microsatellite loci for *Gmelina arborea*. Mol Ecol Resour 10:1106–1108

Lowe AJ, Gillies ACM, Wilson J, Dawson IK (2000) Conservation genetics of bush mango from central/west Africa: implications from random amplified polymorphic DNA analysis. Mol Ecol 9(7):831–841

Lowe A, Harris S, Ashton P (2004) Ecological genetics: design, analysis and application. Blackwell, London

McGowen MH, Wiltshire RJE, Potts BM, Vaillancourt RE (2001) The origin of *Eucalyptus vernicosa*, a unique shrub eucalypt. Biol J Linn Soc 74(3):397–405

Moriguchi Y, Yamazaki Y, Taira H, Tsumura Y (2010) Mating patterns in an indoor miniature *Cryptomeria japonica* seed orchard as revealed by microsatellite markers. New Forests 39(3):261–273. doi:10.1007/s11056-009-9169-0

Naik D, Singh D, Vartak V, Paranjpe S, Bhargava S (2009) Assessment of morphological and genetic diversity in *Gmelina arborea* Roxb. New Forests 38(1):99–115. doi:10.1007/s11056-009-9134-y

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol 13(5):1143–1155

Peakall R, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6(1):288–295

Raju AJS, Rao SP (2006) Pollination by bees and passerine birds and seed dispersal by monkeys in the white teak *Gmelina arborea* Roxb., a commercially important timber tree species in the Eastern Ghats. Curr Sci 90(2):232–236

Ramakrishnan AP, Meyer SE, Fairbanks DJ, Coleman CE (2006) Ecological significance of microsatellite variation in western North American populations of *Bromus tectorum*. Plant Species Biol 21(2):61–73

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86(3):248–249

Rohlf FJ (1998) NTSYS-pc version 2.0. Numerical taxonomy and multivariate analysis system, version 2.0 edn. Exeter Software, Setauket, New York

White GM, Boshier DH, Powell W (2002) Increased pollen flow counteracts fragmentation in a tropical dry forest: an example from *Swietenia humilis* Zuccarini. Proc Natl Acad Sci USA 99(4):2038–2042

Wright S (1951) The genetical structure of populations. Ann Eugen 15:323–354

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11(1):1–16