Research Article

Current genetic structure of teak (*Tectona grandis*) in Myanmar based on newly developed chloroplast single nucleotide polymorphism and nuclear simple sequence repeat markers

Thwe-Thwe-Win[1], Tomonori Hirao[2], Atsushi Watanabe[3] and Susumu Goto[1]

[1] Education and Research Centre, The University Forests, Graduate School of Agriculture and Life Science, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

[2] Forest Tree Breeding Center, Forestry and Forest Products Research Institute, Independent Administrative Institution, Ishi, Juo, Ibaraki 319-1301, Japan

[3] Forest Sciences, Agro-environmental Science, Faculty of Agriculture, Kyushu University, 610-1, Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

Corresponding author: Thwe-Thwe-Win; e-mail: ttw74kk@gmail.com

Abstract

Teak (*Tectona grandis*), which is among the most valuable tropical timber species, is under pressure from rapid deforestation and habitat fragmentation. Limited genetic information is available for Myanmar teak, which comes from the largest natural teak-bearing forest area in the world. To determine the phylogeographic patterns of Myanmar teak, we evaluated three newly developed chloroplast single nucleotide polymorphism (cpSNP) and 10 nuclear simple sequence repeat (nrSSR) markers in 480 individuals representing 20 natural populations. The cpSNP markers detected four haplotypes, each differing by a single mutation. The G_{ST} value was lower than the N_{ST} value and did not reveal a phylogeographic structure of Myanmar teak. Nuclear microsatellite analysis revealed high genetic diversity with the mean expected heterozygosity (0.652). The same level of genetic differentiation (4%) was observed for both cpSNP and nrSSR markers in different groups. Conservation of the HMB, TDG, KTA, and POL populations should receive highest priority because these contribute most to the total genetic diversity. The genetic boundaries of teak observed from combining the results of cpSNP and nrSSR marker barriers suggested four potential zones in the teak seed transfer guidelines of Myanmar. In light of our findings, we discussed appropriate gene conservation strategies and proposed seed zones to safeguard the current genetic resources of Myanmar teak.

Key words: gene conservation, genetic clustering, haplotypes, Myanmar, teak

Received: 8 June 2014; Accepted 22 January 2015; Published: 23 March 2015

Cite this paper as: Thwe-Thwe-Win, Hirao, T., Watanabe, A. and Goto, S. 2015 Current genetic structure of teak (*Tectona grandis*) in Myanmar based on newly developed chloroplast single nucleotide polymorphism and nuclear simple sequence repeat markers. *Tropical Conservation Science* Vol.8 (1): 235-256. Available online: www.tropicalconservationscience.org

Introduction

Many tropical forests are declining more drastically than ever before, given the increased use of forest resources; diminishing forestland threatens the sustainability of forest genetic resources and highlights the importance of conservation and sustainable management of valuable natural resources [1]. Teak (*Tectona grandis* L., Lamiaceae), which produces high-value wood, is among the most economically important tropical tree species. This species naturally occurs in the Indo-Pacific region that includes India, Myanmar, Thailand, and Laos [2]. Myanmar has the largest natural teak forest area, which represents approximately 60% of the total worldwide natural teak forest area [3]. Except for Myanmar, countries to which teak is native have banned teak harvesting from natural forests since the 1980s. By the 1850s, teak forests in the southern region of Myanmar had been depleted by overharvesting by the timber and shipbuilding industries [4]. During the last two decades, natural forests in Myanmar have dwindled because of overexploitation and are threatened by selective logging, a primary cause of both forest degradation and deforestation.

To restore natural teak forests, a teak plantation program was first implemented in Myanmar in 1700; this small-scale plantation program included enrichment planting in overexploited areas of natural forests [5]. Extensive Myanmar teak plantation programs were first implemented by the Forestry Department in 1997 and a private company in 2007. A widely distributed pattern of teak geographic variation would be expected in Myanmar, as teak grows in diverse climatic and edaphic conditions. In this case, restoration with non-local genotypes might result in a failure of forestation due to maladaptation and could harm the remnant natural populations through gene pollution [6]. Large-scale planting without considering the genetic composition of seed poses a risk of genetic disturbance if seeds from various sources and with different genetic compositions are mixed [7]. To develop scientifically sound conservation strategies and formulate the seed transfer guidelines, we must understand these geographic variations in the genetic composition of Myanmar teak. Large-scale planting will require populations with high genetic diversity or genetic divergence and minimal genetic disturbances.

Knowledge of the genetic variations of extant populations over the entire distribution range is essential for the conservation of genetic resources [8]. Intensive native teak genetic analyses have been conducted previously using materials from India, Thai, and Laos and genetic markers, including an isozyme for natural populations from India and Thai [9] and amplified fragment length polymorphisms (AFLP) to investigate genetic variations in teak from India, Thailand, and Indonesia [10]. Compared to isozymes and AFLP, highly polymorphic nuclear simple sequence repeat markers (nrSSRs) are useful for elucidating the spatial structure of genetic diversity and demographic patterns of variation that have resulted from migration [11] and drift, as well as evolutionary history. NrSSR markers for teak have been developed [12] and successfully used to analyze native teak from India, Thai, and Laos derived from a provenance trial [9, 13, 14]. These studies revealed that South Indian teak exhibited the highest genetic diversity, followed by North India, Thailand, and Laos teak populations [9]. Genetic information from Myanmar teak has been lacking, although a recent publication discussed geographic variations in Myanmar teak [15]. Minn et al. [15] applied nrSSR markers to eight populations from the upper and lower regions of Myanmar and found significant genetic differentiation between the two regions. However, the study did not assess geographic patterns in the Myanmar teak population with which to designate seed transfer guidelines.

Chloroplast marker-based genetic information is useful for phylogeographic studies and genetic resource conservation, because the haploid chloroplast genomes are maternally inherited in angiosperm cells and hence transmitted by seeds [16]. Chloroplast DNA (cpDNA) variation-based haplotype distribution reflects the phylogeographic structure [17]. Among these markers, chloroplast

single nucleotide polymorphisms (cpSNP) are very useful for clarifying phylogeographic patterns because of their slow rate of sequence evolution, small organelle DNA genome size, and absence of recombination [18], and cpSNP variations are known to be structured geographically in plant species [19]. Nevertheless, no attempts have been made to evaluate the genetic diversity and geographic patterns of teak using cpDNA markers. A cpSNP and nrSSR marker analysis might provide sound genetic information relevant to conservation, seed migration, and the phylogeographic variation patterns of teak in Myanmar.

This study of natural populations from various regions in Myanmar was conducted with two main objectives: first, to provide sound information about the genetic structure of Myanmar teak for the conservation of genetic resources, and second, to propose potential seed zones based on the genetic variation revealed by cpSNP and nrSSR markers. Finally, we will discuss appropriate conservation strategies and the seed transfer guidelines most likely to minimize genetic disturbances in Myanmar teak.

Methods

Study site and sampling design

Teak is an insect-pollinated, major outcrossing species [20]. The fruits are naturally dispersed by wind [21] or gravity, but dispersal is very limited due to the heavy weight and lack of wings on the seeds. This species is widely distributed throughout Myanmar, which lies between the latitudes of 9° 33'N and 28° 31'N and longitudes of 92° 10'E and 101° 10'E, although the northern and southern boundaries of the teak-growing area lie at the respective latitudes of 28° 30'N and 10°N approximately. In Myanmar, teak normally grows at altitudes ranging from sea level to 1,000 m and thrives best in warm, moist, tropical climates with an annual rainfall level between 1,300 and 3,800 mm. However, teak can grow at precipitation extremes of <800 mm and >5,000 mm, corresponding to the outer fringes of the central dry zone and the southern region of Myanmar, respectively.

Myanmar can be divided into five physiographic regions based on geology, topology, and terrain: the northern mountains, comprising the upper and northern sub-regions; western ranges; eastern plateau; central basin; and lowland Myanmar, which includes the lower region and southern sub-region [22]. These geographic regions comprise seven states and seven divisions. For example, the northern mountain region includes Kachin state; the western ranges include Yakhaing and Chin states; the central region comprises Mandalay, Magway, and a small part of the Sagaing division; the lower region includes the Yangon, Ayeyarwaddy, and Bago divisions; the southern sub-region includes Mon state, the southern part of Kayin state, and Tanintharyi division; and the eastern plateau includes Shan state, Kayar state, and the northern part of Kayin state.

Fresh leaves were collected from 480 individuals representing 20 natural populations that covered nearly the entire teak distribution area throughout Myanmar. Some samples were collected from fragmented and remnant natural forests (e.g., TZ, MLM). The sample size of each population ranged from 16 to 32 (Table 1), and sampled trees were spatially separated by at least 50 m to avoid sampling closely related individuals. The collected samples were dried with silica gel in plastic bags and kept at room temperature prior to DNA extraction.

DNA extraction

DNA was isolated from dried leaf tissues using the hexadecyltrimethylammonium bromide (CTAB) method [23]. Approximately 100 mg of dry leaves were pulverized and then mixed with 1 mL of CTAB buffer (100 mM Tris-HCl, pH 8.0, 20 mM EDTA, 2% CTAB) with 0.1% β-mercaptoethanol added immediately prior to use. The mixture was incubated at 65°C for 60 min and centrifuged for 2 min at

13,000 ×*g*, after which 600 µL of the supernatant were then transferred to a 1.5-mL microcentrifuge tube. The supernatant was mixed twice with phenol/chloroform/isoamyl alcohol (25:24:1) and shaken for 10 min, followed by centrifugation for 10 min at 12,000 ×*g*. DNA was precipitated from the aqueous phase by adding a 0.1 volume of 3 M sodium acetate and 2.5 volumes of ethanol. The DNA precipitate was washed twice with 70% ethanol and dissolved in water. The quality of randomly selected DNA samples was determined using a NanoVue Plus Spectrophotometer (GE Healthcare and Life Sciences, Little Chalfont, UK).

Table 1. Locations and sample size of 20 natural teak populations in Myanmar. Letters in parentheses indicate abbreviated population names. The small m in parentheses indicates the altitude in meters.

Region	Population Name	Sample size	Latitude	Longitude	Altitude (m)
(1) Lower	1. Hlawkha (HKA)	16	17.020	96.118	34
	2. Hmawbi (HMB)	20	17.352	96.057	99
	3. Yangon (YGN)	20	16.792	96.159	23
(2) Central	4. Kyaukpadaung (KPG)	20	20.806	95.574	364
	5. Popa (PPA)	20	20.914	95.228	889
	6. Seitphyu(SP)	24	21.063	94.477	286
	7. Taungdwingyi (TDG)	32	20.081	95.617	189
	8. Tharzi(TZ)	24	20.779	96.392	372
(3) Upper	9. Gangaw (GGW)	32	22.216	94.158	226
	10. Hteechaik (HTK)	24	23.925	96.184	306
	11. Indaw (IND)	24	23.617	95.719	334
	12. Katha (KTA)	24	24.299	96.000	306
	13. Pyinoolwin (POL)	32	21.817	96.300	520
(4) Eastern	14. Naungkhio (NKO)	32	22.352	96.831	572
	15. Moemeik (MMK)	24	23.073	96.643	267
	16. Mabein (MB)	26	23.466	96.601	124
(5) Southern	17. Hpa-an (HPN)	24	16.867	97.653	108
	18. Halinebwe (HB)	24	17.122	97.857	42
	19. Bilin (BLN)	25	17.225	97.228	134
	20. Mawlamein (MLM)	26	16.468	97.627	17

DNA sequencing for the development of chloroplast markers

The complete teak chloroplast genome nucleotide sequence was previously determined (GenBank Accession No. *NC020098*). To detect intra-specific genetic variation, partial genome walking was conducted using eight individuals and 58 PCR genome walking primers (Appendix 1). One individual was selected from each of eight populations**,** HMB, SP, TDG, NKO, GGW, KTA, HLB and MLM, which nearly cover the entire geographical and altitudinal range of the teak-growing area of Myanmar (Table 1)**.** DNA amplification was performed in a GeneAmp® PCR System 9700 (Applied Biosystems, Carlsbad, CA, USA) using the following program: 95°C for 1 min; 30 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 60 s; and a final extension at 72°C for 10 min. PCR was performed in a volume of 10 µL, which contained 10 mM Tris-HCl, pH 9.0, 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.25 µM of each primer, 20 ng of genomic DNA, and 0.1 U of *Taq* DNA polymerase (Life Technologies, Carlsbad, CA, USA). To remove excess dNTPs and primers, aliquots of each PCR product (5 µL) were purified with 2 µL of ExoSAP-IT reagent (Affymetrix, San Diego, CA, USA) during a 1 h incubation at 37°C, followed by exposure at 80°C for 15 min. Cycle sequencing was performed using the BigDye® 3.1 Terminator Cycle Sequencing kit (Applied Biosystems) as follows: 96°C for 1 min; 25 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min; and a final hold at 4°C according to the manufacturer's instructions. The sequencing primer (3.3 pmol**;** same sequence as the PCR primer), 1.0 µL of ABI Dye Terminator Ready-Reaction sequencing premix, and 1.5 µL of 5× sequence buffer were added to the template. Excess dye terminators were removed via ethanol precipitation. The extension products were evaporated to dryness under vacuum, resuspended in Hi-Di™ formamide (Applied Biosystems), incubated for 2 min at 95°C, and loaded onto an ABI PRISM® model 3130xl DNA sequencer (Applied Biosystems) according to the manufacturer's instructions. For sequence analysis and alignment, we used the software package Sequencher® 4.10.1 (Gene Codes Corporation, Ann Arbor, MI, USA).

Chloroplast SNP genotyping

We successfully designed three new teak chloroplast SNP markers based on polymorphisms identified in our analysis of intraspecific variation in the teak chloroplast genome. SNP genotyping was conducted via a primer extension-based method with the ABI PRISM SNaPshot Multiplex Kit (Applied Biosystems). Based on the flanking sequences of the target SNPs, a locus-specific primer pair (LSP) and extension primer (EP) were designed for each SNP (Appendix 2). The LSPs were designed to anneal at 58–60°C. Multiplex PCR was performed in a 10 µL total volume with the QIAGEN multiplex PCR Kit (QIAGEN, USA); each reaction contained 2x QIAGEN Multiplex PCR Master Mix, 0.25 µM of each primer, and 20 ng of template DNA. PCR amplification and purification were performed using the same protocols as described above. A single-base extension reaction was conducted using the ABI PRISM SNaPshot Multiplex Kit (Applied Biosystems). The reaction was conducted in a total volume of 10 µL, which contained 2 µL of ExoSAP-treated PCR product, 0.2 µM of each EPs, 1× SNP reaction buffer, and 0.50 µL of ABI PRISM SNaPshot Multiplex Kit. The thermal cycler program was run according to the manufacturer's recommendations. To dissolve the unincorporated fluorescent ddNTP, the SNaPshot product cleanup was performed with shrimp alkaline phosphatase (SAP; USB Corp., USA): 1 µL (1U) was added to 5 µL of SNaPshot product; the mixture was briefly centrifuged and incubated at 37°C for 1 h for digestion, followed by 80°C for 15 min to terminate the reaction. For analysis, 1 µL of SAP-treated product was mixed with 9.8 µL of HiDi™ formamide and 0.2 µL of GeneScan 120 LIZ size standard (Applied Biosystems). Data were generated after performing capillary electrophoresis on an automated sequencer (ABI 3130 Genetic Analyzer; Applied Biosystems) and analyzed using GeneMapper analysis software, version 4.0 (Applied Biosystems).

Nuclear microsatellite genotyping

For nrSSR analysis, eleven nuclear microsatellite primer pairs [14] that yielded high levels of polymorphism and clear peaks (Teak-A06, Teak-B07, Teak-B03, Teak-H10, Teak-Da12, Teak-F01, Teak-F02, Teak-A11, Teak-F05, Teak-B02, and Teak-C03) were used in this study. Multiplex PCR was performed according to the manufacturer's protocol (QIAGEN Multiplex PCR kit; QIAGEN, Venlo, The Netherlands) in a final volume of 10 µL. Four fluorescently labeled (labels: FAM, NED, PET, and VIC) primers were multiplexed. Amplifications were performed in a thermal cycler (Bio-Rad T100^{TM}; Bio-Rad, Tokyo, Japan) with an initial denaturation step at 95°C for 15 min; 30 cycles of denaturation at 94°C for 30 s, annealing at 57°C for 90 s, and extension at 72°C for 1 min; and a final extension at 60°C for 30 min for all primers. Multiplexed PCR products were diluted 10-fold with sterilized water, and 1 µL of the diluted product was mixed with Hi-Di^{TM} formamide and GeneScan 600 LIZ size standard (Applied Biosystems). PCR product electrophoresis was performed on an ABI 3130xl sequencer (Applied Biosystems), and alleles were scored using GeneMapper software. If many artifacts and peaks with abnormally high fluorescence intensities were detected, independent PCR reactions were conducted to confirm the amplification.

Data analysis

For chloroplast data analysis, the gene diversity of each population was calculated using CONTRIB ver. 1.02 [24]. Haplotypes occurring at a frequency of <5% of the total samples were defined as rare haplotypes [25]. Two measures of population differentiation, G_{ST} and N_{ST}, were compared using the programs PERMUTE and CPSSR version 2.0 [26]. G_{ST} incorporates haplotype frequencies, whereas N_{ST} accounts for differences between haplotypes. To infer the genealogical relationships among haplotypes, a network analysis based on parsimony methods was conducted using TCS [27]. The proportion of each haplotype per population was computed and mapped to visualize the haplotype distribution. The CONTRIB version 1.02 software package [24] was used to assess the conservation value of a particular population; this estimates the contribution of each population to the total diversity in terms of genetic diversity and differentiation.

For nrSSR analysis, FSTAT ver. 2.9.3 [28] was used for the calculation of genetic diversity parameters such as the expected heterozygosity (H_E), allelic richness (R) and fixation index (F_{IS}) of each population. Moreover, genetic diversity parameters (H_E and R), fixation indices (F_{IS}), genetic differentiation among populations (F_{ST}) of each group, and the significance of differences in each parameter among regions were measured using the same software. The Lositan software package [25] was used to detect outlier loci (non-neutral loci) in order to evaluate the neutral genetic diversity of teak. Genetic clusters of 20 natural populations and F_{ST} estimates between the clusters were generated using the program STRUCTURE ver. 2.3 [29], which was run for 15 iterations using the admixture model and a number of clusters (K) ranging from 1 to 10 with a burn-in length of 100,000 and a run length of 100,000 iterations. The most suitable number of clusters was inferred from the delta (K) statistic [30]. The average proportion of ancestry among individuals in each population was calibrated. Bottleneck version 1.2.02 [31] was used to examine the bottleneck effect on studied populations.

Putative genetic boundaries within the sampling regions were identified using Barrier software version 2.2 [32] according to the Euclidean distance in the cpSNP dataset and Nei's genetic distance in the nrSSR dataset. For an analysis of molecular variance (AMOVA), populations were grouped based on the combined cpSNP and nrSSR barrier results, and hierarchical AMOVAs were performed to determine the different genetic hierarchies among groups, among populations within a group, and among individuals within populations for both markers using Arlequin ver. 2.0 [33]. Significance testing was conducted with 1,000 permutations at each level. Spatial genetic structure was investigated using the logarithm of geographic distance in the range of 32–889 km and the pair-wise genetic distance

$F_{ST}/(1 - F_{ST})$, which was based on Mantel's test as performed with 1,000 permutations using GeneAlex 6.5 [34].

Results

Chloroplast polymorphism

A total of 43,734 base pairs, excluding missing data and gaps, were sequenced in eight individuals using 58 walking primers (Appendix 1). Three variable sites, all SNPs, were detected; two were in intergenic spaces (IGS), a G/T mutation in the *psbK-psbI* IGS and AAA/TTT in the *trnK-rps*16 IGS and one, C/A, occurred in the *rpl*16 intron. Fragments of these three variable sites were used to develop chloroplast markers with which to investigate geographic variations in Myanmar teak (Appendix 2).

Geographic variation revealed by cpSNP markers

A cpSNP variation analysis identified four haplotypes in 20 natural populations (Table 2). The frequency of H1 was highest (84%), and this haplotype was the most common in all populations except TDG and HMB, whereas H2 was the rarest haplotype (3%). Although the frequencies of H3 (6%) and H4 (7%) were relatively low, these haplotypes were detected in many populations. No private haplotype was found in any population. The gene diversity ranged from 0.000 to 0.565, with a mean value of 0.173 (Table 3, Fig. 1). The highest level of gene diversity was observed in the KTA population from the upper region, followed by the HMB and TDG populations from the lower and central regions, respectively. At the regional level, the lower region and eastern regions harbored the highest (0.249) and lowest gene diversity (0.075), respectively. Half of the 20 populations, mostly those from the central region, were monotypic and fixed for the common haplotype (H1), whereas the remaining populations were polytypic and contained two to four haplotypes. The overall F_{ST} value using cpDNA markers was 0.293 (Table 3). The measure of population differentiation, G_{ST} = 0.346, was higher than the N_{ST} value of 0.302. The geographic distribution of the chloroplast haplotype frequency in each population and the star-shaped parsimony haplotype network, in which each haplotype is distinguished by a single mutation, are shown in Fig. 1. The rare haplotype H2 occurred mainly in the southern region; it was also present in HMB, TDG, and POL from the lower, central, and upper regions, respectively, but not in the eastern region. The H3 haplotype was absent from the central populations, and H4 was not found in the southern and eastern regions. The highest frequency of H4 was observed in TDG, the only polytypic population in the central region. The assessment of populations to prioritize for conservation depicted stronger contributions of the HMB, TDG, KTA, and POL populations to the total diversity in terms of their high levels of genetic diversity (Fig. 2).

Table 2. Polymorphic sites and chloroplast DNA haplotypes according to a SNaPshot analysis.

Polymorphic sites	*psbK-psbI*	*rpl* 16	*trnK-rps* 16
Haplotye			
H1	G	C	T
H2	T	C	T
H3	G	A	T
H4	G	C	A

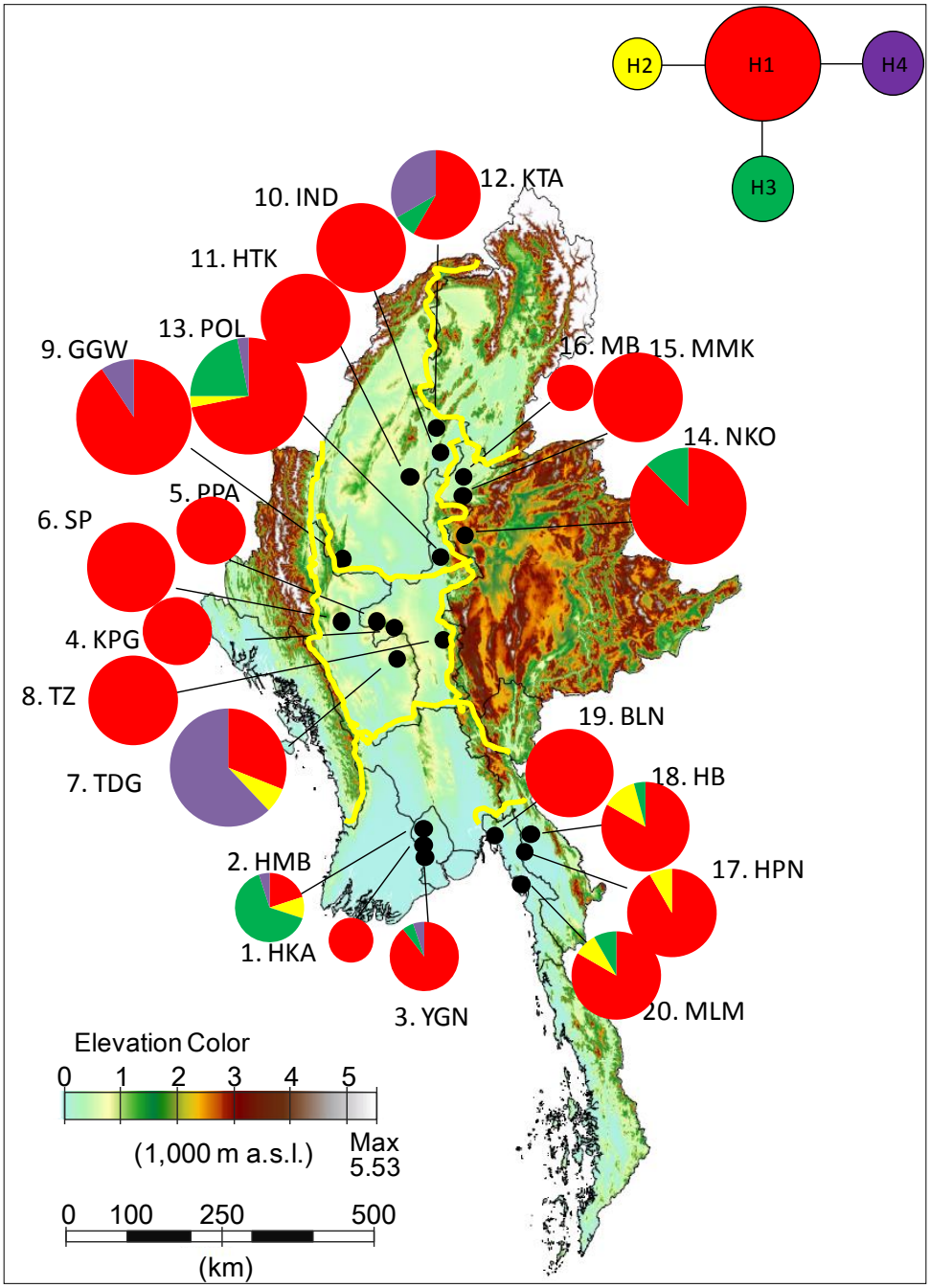


Fig. 1. Myanmar teak haplotype network and geographic variation patterns as revealed by chloroplast single nucleotide polymorphism markers. The population numbers and abbreviations correspond to those used in Table 1. Black lines show the administrative boundaries of states and divisions. Yellow lines show the regional boundaries and overlap some administrative boundary lines. The circle sizes are proportional to the sample size of each population.

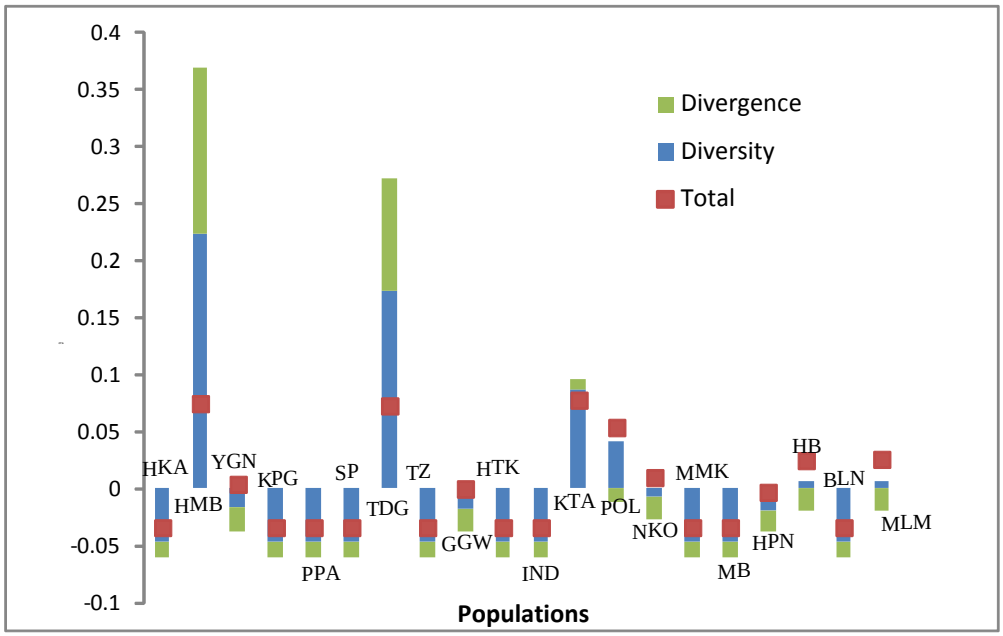


Fig. 2. Contribution (CT%) of each population to total diversity in terms of genetic diversity and differentiation from other populations. The population abbreviations correspond to those used in Table 1.

Genetic structure revealed by nrSSR

Of the 11 nuclear microsatellite loci, one outlier (Teak-B03) was excluded from the nuclear microsatellite dataset analysis. Regarding the average genetic diversity statistics, the expected heterozygosity (H_E) and allelic richness (R) were 0.652 and 5.76, respectively (Table 3). The range of genetic diversity indices in these populations among 10 loci varied with respect to both the expected heterozygosity (H_E), which ranged from 0.530 to 0.763, and the allelic richness (R), which ranged from 3.09 to 8.29. The minimum and maximum fixation indices (F_{IS}) were −0.148 for the HMB population and 0.076 for the MMK population. The genetic differentiation among 20 populations was 0.094. At the regional level, the genetic diversity estimates and F-statistics of each group are shown in Table 3. The lower region possessed the highest genetic diversity in terms of H_E (0.712) and R (7.28), whereas the southern region exhibited the lowest for both values (Table 3). The highest F_{ST} value (0.097) and lowest F_{IS} (−0.091) were found in the eastern populations, whereas the lowest F_{ST} value (0.017) and highest F_{IS} (0.040) were found among the lowland populations. Significant differences among regional groups were detected in terms of allelic richness (p = 0.012), expected heterozygosity (p = 0.011), and F_{IS} (p = 0.02).

When a Bayesian clustering approach was implemented via STRUCTURE analysis, a maximum delta K value of 4 was detected. Using this value, we grouped the 20 populations of Myanmar teak into four genetic clusters (Fig. 3). The greatest genetic differentiation between clusters (F_{ST}) was observed in cluster III (0.310), followed by clusters IV (0.264), II (0.128), and I (0.037). Cluster I predominated in the lower and central regions, cluster II was dominant among populations from the upper and eastern regions apart from the KTA and NKO populations, and cluster IV was present mainly in the southern region and TDG population. Cluster III was predominant in two populations (POL and NKO) that were from different regions but were geographically close. Of these 20 populations, 16 exhibited a significant bottleneck effect under the stepwise mutation model, but only two populations (HMB and YGN) showed significant bottleneck effect under infinite assumption model.

Table 3. Statistical summary of the diversity revealed by the chloroplast single nucleotide polymorphism and nuclear simple sequence repeat markers in 20 teak populations. *A,* number of alleles*; R*, allelic richness; H_O, observed heterozygosity; H_E, expected heterozygosity; F_{IS}, fixation index; F_{ST}, genetic differentiation among populations. Populations with asterisks indicate those with bottleneck effects at different significance levels. * $p < 0.05$, ** $p < 0.001$, *** $p < 0.001$.

No.	Region	Pop.	N	cpSNP			nrSSR					
				No. of haplotype	*h*	F_{ST}	*A*	*R*	H_O	H_E	F_{IS}	F_{ST}
	Lower											
1		HKA**	16	1	0		5.90	5.90	0.631	0.660	0.044	
2		HMB**	20	4	0.55		8.20	7.64	0.650	0.703	0.076	
3		YGN**	20	3	0.2		8.90	8.29	0.760	0.763	0.004	
		Mean	19	2.67	0.25			7.28	0.684	0.712	0.04	0.017
	Central											
4		KPG*	20	1	0		6.70	6.22	0.640	0.677	0.055	
5		PPA	20	1	0		5.70	5.48	0.685	0.661	-0.036	
6		SP	24	1	0		6.00	5.51	0.633	0.648	0.022	
7		TDG*	32	4	0.54		8.20	6.61	0.675	0.683	0.011	
8		TZ*	24	1	0		8.10	7.03	0.675	0.718	0.060	
		Mean	24	1.6	0.108			6.17	0.663	0.678	0.023	0.067
	Upper											
9		GGW	32	2	0.18		5.40	4.77	0.588	0.618	0.049	
10		HTK*	24	1	0		6.20	5.60	0.629	0.633	0.007	
11		IND*	24	1	0		6.30	5.73	0.609	0.640	0.048	
12		KTA*	24	3	0.57		7.50	6.67	0.750	0.719	-0.043	
13		POL	32	4	0.45		9.40	7.50	0.693	0.730	0.050	
		Mean	27.2	2.2	0.24			6.05	0.652	0.669	0.025	0.078
	Eastern											
14		NKO	32	2	0.23		4.60	4.24	0.672	0.634	-0.059	
15		MMK	24	1	0		5.20	4.77	0.717	0.624	-0.148	
16		MB*	26	1	0		5.00	5.00	0.613	0.574	-0.068	
		Mean	27.3	1.3	0.08			4.67	0.674	0.618	-0.091	0.097
	Southern											
17		HPN	24	2	0.16		4.60	4.27	0.638	0.576	-0.107	
18		HB*	24	3	0.3		7.20	6.28	0.629	0.671	0.062	
19		BLN	25	1	0		3.20	3.09	0.529	0.530	0.002	
20		MLM*	26	3	0.3		5.00	4.54	0.500	0.577	0.134**	
		Mean	24.8	2.25	0.19			4.54	0.574	0.588	0.025	0.080
	Overall (mean)		25	2	0.173	0.293	6.37	5.76	0.646	0.652	0.002	0.094

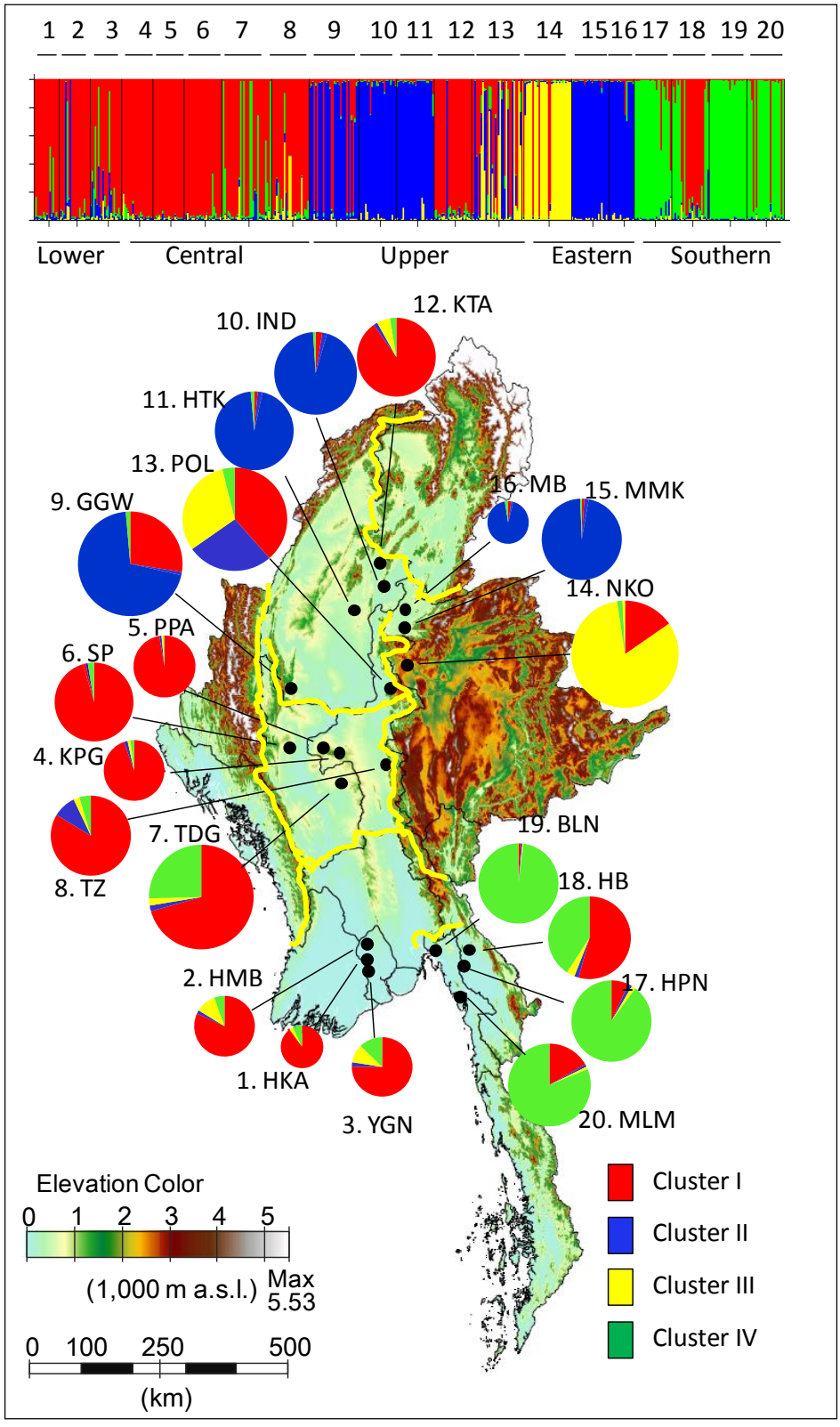


Fig. 3. Myanmar teak genetic structure and geographic variation as detected using nuclear simple sequence repeat markers. The population numbers and abbreviations correspond to those used in Table 1. Yellow lines show the regional boundaries and overlap some black lines that indicate the administrative boundaries of states and divisions. The sizes of the circles are proportional to the sample size of each population.

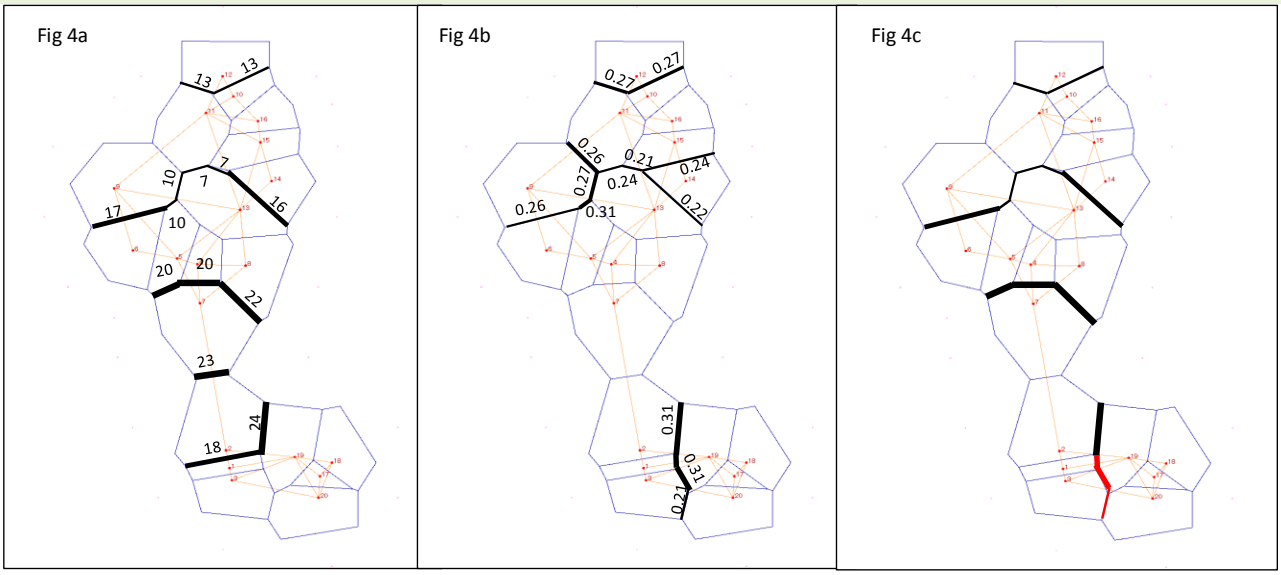
Geographic variation and genetic differentiation

The KTA, TDG, and HMB populations were not included in any cpSNP marker groups (Fig. 4a), and the KTA, GGW, and NKO populations also were separated from the nrSSR marker groups (Fig. 4b) according to the barrier results. The boundaries between the lower and southern regions as determined with nrSSR markers were used for grouping because of the high cluster IV F_{ST} value, although the genetic boundaries between those regions were incongruent. Therefore, an identification of genetic boundaries based on cpSNP markers and nrSSR generally yielded four groups (Fig. 4c) that were used for AMOVA. Those groups likely corresponded to geographic regions (e.g., upper and eastern regions group, central populations group, lower region group, and southern region group). The AMOVA results revealed that the same amount of genetic variation accounted for the different groups generated using both types of markers. Significant genetic variations were observed at all hierarchical levels (Table 4). No significant isolation by distance (IBD) pattern was detected using cpSNP markers (r = −0.049, p = 0.503), whereas the nrSSR markers revealed a significant pattern (r = 0.333, p = 0.002).

Table 4. Results of an analysis of molecular variance (AMOVA) involving the chloroplast single nucleotide polymorphism (cpSNP) and nuclear simple sequence repeat (nrSSR) markers. Groups are identified based on the genetic boundaries from the cpSNP and nrSSR marker barrier results.

Source of variation	d.f	Sum of squares	Variance components	Percentage of variation	Fixation Indices
cpSNP					
Among groups	3	4.93	0.006	4	F_{CT} = 0.044, p < 0.001
Among populations within group	15	14.10	0.035	26	F_{ST} = 0.303, p < 0.001
Within populations	440	41.70	0.095	70	F_{SC} = 0.271, p < 0.001
Total	458	60.72	0.136	100	
nrSSR					
Among groups	3	135.12	0.131	4	F_{CT} = 0.037, p < 0.001
Among populations within group	15	227.01	0.252	7	F_{ST} = 0.108, p < 0.001
Within populations	893	2824.58	3.163	89	F_{SC} = 0.074, p < 0.001
Total	911	3186.71	3.546	100	

Fig. 4. a: Identification of genetic boundaries based on chloroplast single nucleotide polymorphism (cpSNP) markers. The segment values indicate the Euclidean distances between the separated populations. The segment thicknesses correspond to the Euclidean distance values. b: Identification of genetic boundaries based on nuclear simple sequence repeat (nrSSR) markers. The segment values represent Nei's genetic distances between the separated populations. c: Identification of genetic boundaries based on the combined cpSNP and nrSSR barrier results. Black lines indicate the congruent boundary segments for both the cpSNP and nrSSR barrier results. Red lines indicate the boundary segments demonstrated by the nrSSR barrier results.



Discussion

Geographic variation revealed by chloroplast SNP markers

Due to teak's limited seed dispersal ability, we expected clear geographic patterns in Myanmar teak would be detected using cpSNP markers. However, the significantly higher N_{ST} relative to G_{ST} indicated the lack of a phylogeographic structure and randomized teak haplotype distribution at cpSNP markers. Using maternally inherited markers, tree species are typically shown to exhibit strong genetic differentiation among populations, with F_{ST} values exceeding 0.6 (average: 0.764) [18]. Therefore, the overall cpSNP-based F_{ST} of Myanmar teak populations (0.293) was quite low in term of tree species. The lack of phylogeographic structure, star-shaped haplotype network pattern, and low F_{ST} values indicated recent colonization from a small number of Myanmar teak founders and therefore insufficient time for the occurrence of mutation in long-lived species. The predominant distribution of a single haplotype (H1) throughout much of the range suggested a bottleneck effect at the time of postglacial recolonization. In fact, we detected significant bottleneck effects in 16 of the 20 populations under stepwise mutation model in this study. In contrast to the distribution of H1, the rare haplotype H2 occurred mainly in the southern region, but not the eastern region. The haplotype H3 was absent from the central region, whereas the haplotype H4 was not found in southern and eastern regions. Therefore, rare haplotypes exhibit geographic patterns, despite the wide distribution of the major haplotype throughout Myanmar.

Genetic diversity and genetic structure revealed by nrSSR markers

In the nrSSR analysis, the populations in the lower region exhibited the greatest genetic diversity and allelic richness (Table 3). The lower populations are located near the Bago Mountains, an area that is considered the home of Myanmar teak [35]. In temperate or boreal forests, populations that

possessed high levels of genetic diversity, allelic richness, and geographic clustering of rare alleles acted as refuges during the LGM [36] and might have derived from an ancestral population or an admixture of different refuges [37]. In contrast, the southern populations exhibited the lowest diversity found in Myanmar (Table 3). A small population size and isolation generally result in low genetic diversity [37]. Forest fragmentation has occurred in the southern region and resulted in decreased population sizes and increased isolation, which might explain the low genetic diversity in southern populations.

Four genetic clusters that basically corresponded to geographic regions were detected in our Myanmar teak sample. The genetic structures of tree species are profoundly influenced by geographical isolation and environmental factors such as topography (including the effects of tectonic shifts), river course dynamics, and sea level fluctuations [38]. For instance, the southern populations were separated from the other groups by rivers and high mountain ranges. Those geographic features might have acted as genetic barriers between the southern and other populations.

Comparison of findings obtained using cpSNP and nrSSR markers

In this study, we observed four main genetic groups of Myanmar teak using both types of markers. Genetic variation, which accounted for the different groups (Table 4), was relatively low in terms of cpSNP markers when compared to the variation observed in other tree species such as *Neobalanocarpus heimii* [39] and *Caryocar brasilense* [40]. A genetic bottleneck in Myanmar teak might account for the low level of cytoplasmic variation. In the chloroplast dataset, the rare haplotype H2 was distributed throughout all southern populations; these populations possessed the lowest genetic diversity and represented the major component of cluster IV, with a high F_{ST} value according to nrSSR data. The haplotype H2 might have derived from the common ancestor H1, and cluster IV might similarly have derived from cluster I.

Implication for conservation

The results revealed by combining both cpSNP and nrSSR markers were useful in terms of conserving the genetic resources of Myanmar teak. Our results indicate that the HMB, TDG, KTA, and POL populations should be given the highest priority for *in situ* conservation because these provided strong contributions to total diversity, given their high levels of genetic diversity and divergence from other populations (Fig. 2). Total genetic diversity is a key factor of conservation measures designed to ensure the long-term survival of a species in a changing environment [41]. Furthermore, most of the southern populations, which featured a rare haplotype and predominance of cluster IV, should also be considered not only for *in situ* but also *ex situ* conservation. To retain cluster III and haplotype H3, the POL and NKO populations, which exhibited genetic divergence, should also be included in the conservation program.

Our findings provide essential information to support conservation efforts and establish potential seed and planting zones appropriate for Myanmar teak. We observed at least four genetic Myanmar teak groups that basically corresponded to geographic regions (Fig. 4). Future studies should conduct provenance trials to clarify the potential zones identified from the cpSNP and nrSSR data.

Acknowledgements

We thank C. Lian for sharing facilities to perform the experiments. We extend our thanks to H. Iwata and M. Takahashi for their helpful comments regarding this study. We convey special thanks to W. Ishizuka, S. Sakaguchi, D. Kusumoto, H. Kurokochi, K. Uchiyama, and four anonymous reviewers for their productive comments on an earlier version of this manuscript. We thank the staff of the Forestry Department, MOECAF, Myanmar, and FTBC, Japan for their contributions to this research study.

References

[1] Rajora, O. P. and Mosseler A. 2001. Challenges and opportunities for conservation of forest genetic resources. *Euphytica* 118:197-212.

[2] Pandy, D. and Brown, C. 2000. Teak: a global overview. Unasylva 201:1-14.

[3] Gyi, K. K. and Tint, K. 1998. Management status of natural teak forests in Teak for future. In: Masakazu, K. and Kevin, W. (eds) *Teak for the future*, Dharmasarn, Bangkok, pp. 27-48.

[4] Bryant R. L. 1997. The political ecology of forestry in Burma, 1824-1994. *Honolulu*, Hawaii: University of Hawaii Press.

[5] Saw, E. D. (2003) Sustainable management of teak forests in Myanmar. *Proceeding of the International Conference on Quality Timber Products of Teak from Sustainable Forest Management*, Peechi, India, 2-5 December 2003, pp. 135-142.

[6] Saltonstall, K. 2002. Cryptic invasion by non-native genotype of common reed *Phragmites australis*, into North America. *Proceedings of the National Academy of Sciences*, USA 99:2445-2449.

[7] Hufford, K. M. and Mazer, S. J. 2003. Plant ecotypes: genetic differentiation in the age of ecological restoration. *Trends in Ecology and Evolution* 18:147-155.

[8] Neale, D. B. and Kremer, A. 2011. Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* 12:111-122.

[9] Kertadikara, A. W. S. and Prat, D. 1995. Isozyme variation among teak (*Tectona grandis* L.f) provenances. *Theoretical Applied and Genetics* 90:803-810.

[10] Shrestha, M. K., Volkaert, H. and Straeten, D. V. D. 2005. Assessment of genetic diversity in *Tectona grandis* using amplified fragment length polymorphism markers. *Canadian Journal of Forest Research* 4:1017-1022.

[11] Neale, D. B. and Ingvarsson, P. K. 2008. Population, quantitative and comparative genomics of adaptation in forest trees. *Current Opinion in Plant Biology* 11:149-155.

[12] Verhaegen, D., Ofori, D., Fofana, I., Poitel, M. and Vaillant, A. 2005. Development and characterization of microsatellite markers in *Tectona grandis* (Linn.f). *Molecular Ecology Notes* 5:945-947.

[13] Alcântara, B. K. and Veasey, E. A. 2013. Genetic diversity of teak (*Tectona grandis* LF) from different provenances using microsatellite markers. *Revista Árvore* 37:747-758.

[14] Fofana, I. J., Ofori, D., Poitel, M. and Verhaegen, D. 2009. Diversity and genetic structure of teak (*Tectona grandis* L.f) in its natural range using DNA microsatellite markers. *New Forests* 37:175-195.

[15] Minn Y., Prinz, K. and Finkeldey, R. 2014. Genetic variation of teak (*Tectona grandis* Linn.) in Myanmar revealed by microsatellite markers. *Tree Genetics and Genomes*. DOI 10.1007/s11295-014-0772-7.

[16] Duran, C., Appleby, N., Vardy, M., Imelfort, M., Edwards, D. and Batpey, J. 2009. Single nucleotide polymorphism discovery in barley using autoSNPdb. *Plant Biotechnology Journal* 7:326-333.

[17] Ohatani, M., Kondo, T., Tani, N., Ueno, S., Lee, L. S., Ng, K. S., Muhammad, N., Finkeldey, R., Natem, M., Indrioko, S., Kamiya, K., Harada, K., Diway, B., Khoo, E., Kawamura, K. and Tsumura, Y. 2013. Nuclear and chloroplast DNA phylogeography reveals Pleistocene divergence and subsequent secondary contact of two genetic lineages of the tropical rainforest tree species *Shorea leprosula* (Dipterocarpaceae) in South-East Asia. *Molecular Ecology* 22:2264-2279.

[18] Petit, R. J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D. and Vendramin, G.G. 2005. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology* 14:689-701.

[19] King, R. A and Ferris, C. 1998. Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular Ecology* 7:1151-1161.

[20] Tangmitcharoen, S. and Owens, J. N. 1997. Floral biology, pollination, pistil receptivity, and pollen tube growth of teak (*Tectona grandis* Linn.f). *Annals of Botany* 79:227-241.

[21] Koonkhunthod, N., Sakurai, K. and Tanaka, S. 2007. Composition and diversity of woody regeneration in a 37-year-old teak (*Tectona grandis* L.) plantation in Northern Thailand. *Forest Ecology and Management* 247:246-254.

[22] Hadden, R. L. 2008. The Geology of Burma (Myanmar): *An Annotated Bibliography of Burma's Geology, Geography and Earth Science*. DTIC Document, pp. 312. Accession Number: ADA487552.

[23] Shiraishi, S. and Watanabe, A. 1995. Identification of chloroplast genome between *Pinus densiflora* Sieb. et Zucc. and P. thunbergii Parl. based on the polymorphism in rbcL gene. *Journal of Japanese Forest Society* 77:429-436 (in Japanese with English summary).

[24] Petit, R. J., Mousadik, E. A. and Pons, O. 1998. Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* 12:844-855.

[25] Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A. and Luikart, G. 2008. Lositan: A workbench to detect molecular adaptation based on Fst-outlier method. *BMC Bioinformatics* 9:323-328.

[26] Pons, O. and Petit R. J. 1996. Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* 144:1237-1245.

[27] Clement, M., Posada, D. and Crandall, K. A. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9:1657-1659.

[28] Goudet, J. 2001. FSTAT, version 2.9.3, A program to estimate and test gene diversities and fixation indices. *Journal of Heredity* 86:485-486.

[29] Pritchard, J. K., Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.

[30] Evanno, G., Regnaut, S. and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611-2620.

[31] Piry, S., Luikart, G. and Cornuet, J. M. 1999. Bottleneck: a computer program for detecting recent reduction in the effective population size using allele frequency data. *Journal of Heredity* 90:502-503.

[32] Manni, F., Guérard, E. and Heyer, E. 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by "Monmonier's algorithm." *Human Biology* 76:173-190.

[33] Schneider, S., Roessli, D. and Excoffier, L. 2000. Arlequin ver. 2.000. A software for population genetics data analysis. *Genetics and Biometry Laboratory*, University of Geneva, Switzerland, pp. 111. http://anthro.unige.ch/arlequin

[34] Peakall, R. and Smouse, P. E. 2012. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Bioinformatics Application Notes* 28:2537-2539.

[35] Maung, T. M. and Yamamoto, M. 2008. Exploring the socio-economic situation of plantation villagers: a case study in Myanmar Bago Yoma. *Small-scale Forestry* 7:29-48.

[36] Lowe A. J., Harris, D., Dormontt, E. and Dawson, I. K. 2010. Testing putative African tropical forest refugia using chloroplast and nuclear DNA phylogeography. *Tropical Plant Biology* 3:50-58.

[37] Petit, R. J., Aguinagalde, I., Beaulieu, J. L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Muller-Starck, G., Demesure-Musch, B., Palme, A., Martin, J. P., Rendell, S. and Vendramin, G. G. 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300:1563-1565.

[38] Hewitt, G. 2004. Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society of London. Series B: *The Royal Society* 359:183-195.

[39] Tanah, L. H., Lee, S. L., Ng, K. K., Lee, C. T., Bhassu, S. and Othman, R. Y. 2013. Phylogeographical pattern and evolutionary history of an important Peninsular Malaysian timber species, *Neobalanocarpus heimii* (Dipterocarpaceae). *Journal of Heredity* 104:115-126.

[40] Rosane, G. C., Grattapaglia, D. and John, D. H. 2003. Evidence for multiple maternal lineages of *Caryocar brasilense* populations in the Brazilian Cerrado based on the analysis of chloroplast DNA sequences and microsatellite haplotype variation. *Molecular Ecology* 12:105-115.

[41] Hebel, I., Haas, R. and Dounavi, A. 2006. Genetic variation of common ash (Fraxinus excelsior L.) populations from provenance regions in southern Germany by using nuclear and chloroplast microsatellites. *Silvae Genetica* 55:38-43.

Appendix 1. List of 58 walking primers used for cpSNP marker development.

Pimer Name	sequence (5'-3')	Primer Length	Start position	End position
TgCpGw_R5-7_001F	CAATAGCAGCTAGATCCAGAGG	22	467	488
TgCpGw_R5-7_001R	CCCTACCTTATTGACCGCAACTTC	24	1369	1392
TgCpGw_R5-7_002F	GATCCAGAAACAGGTTCACGAATACC	26	1297	1322
TgCpGw_R5-7_002R	GCGAGGAGTATATAGAAGTCGGG	23	2110	2132
TgCpGw_R5-7_003F	GATTTGTGATTCGCCAGATCATTG	24	2061	2084
TgCpGw_R5-7_003R	CATTTACGATCAACATCTTCTGGAG	25	2866	2890
TgCpGw_R5-7_004F	CCTTGACCAACCATAGGTTCGCC	23	2767	2789
TgCpGw_R5-7_004R	CGCACTATGTATCATTTGATAACCCCC	27	3631	3657
TgCpGw_R5-7_005F	GTGGATATAGGAAGTCGTGTTGTTG	25	3532	3556
TgCpGw_R5-7_005R	CGGGTTGCTAACTCAACGGTAG	22	4311	4332
TgCpGw_R5-7_006F	CTTTCAGGATCAGTCGCGGTC	21	4204	4224
TgCpGw_R5-7_006R	CCGTCGAGTCGGATGGTCTAG	21	5105	5125
TgCpGw_R5-7_007F	GTACATTCCTTGTTGGTTAGTCGAG	25	4977	5001
TgCpGw_R5-7_007R	GCACCGAAGTAATGTCTAAACCC	23	5880	5902
TgCpGw_R5-7_008F	GTCTGATCCAGTTATTGAGACTATCG	26	5796	5821
TgCpGw_R5-7_008R	CCTATTGAGTCACTTAAAGATGCGG	25	6628	6652
TgCpGw_R5-7_009F	GGTTCAGCTGTCCATAGACATCG	23	6427	6449
TgCpGw_R5-7_009R	CGACGTAACGATTGAGGCAGG	21	7290	7310
TgCpGw_R5-7_010F	ATTACTAGCCACCGGCCACTATG	23	7121	7143
TgCpGw_R5-7_010R	GTATGACTGGCATAACATCTACGATTGG	28	8037	8064
TgCpGw_R5-7_011F	CTATTCTCCTTTGAGAATTGGAGG	24	7786	7809
TgCpGw_R5-7_011R	CGAGTTATTCGTACCGAGGGTTC	23	8752	8774
TgCpGw_R5-7_012F	GGCGTAATCCTGGACGTGAAG	21	8591	8611
TgCpGw_R5-7_012R	GAGATCCAAATTTAATGAGATGCGG	25	9400	9424
TgCpGw_R5-7_013F	GAAGCCTCTTTTCCGCATCTC	21	9388	9408
TgCpGw_R5-7_013R	GGGAATCGAACCCGCATCG	19	10257	10275
TgCpGw_R5-7_014F	GAGACGTTCAAAGCACTGAATCG	23	10201	10233
TgCpGw_R5-7_014R	CCGCTATTAATGTGGGTATCTCCG	24	11092	11115
TgCpGw_R5-7_015F	GCTTCTAATTCTGCAAATTGCGC	23	10998	11020
TgCpGw_R5-7_015R	GAAGTAATGGCGGGTGAATTAGTCG	25	12019	12043
TgCpGw_R6_001F	GCAATCCTGAGCCAAATCCTG	21	48476	48496
TgCpGw_R6_001R	GGGGTATATCTCGCCAAGGTAG	22	49402	49423
TgCpGw_R6_002F	CGGGATAGCTCAGCTGGTAG	20	49234	49253
TgCpGw_R6_002R	CCGTCCGTTTTCTGGGTTTGG	21	50132	50152
TgCpGw_R6_003F	GATACGTTTCAAGCGCGGATG	21	50051	50071
TgCpGw_R6_003R	TGTTATTGCTATGGGAGCCTGTAC	24	50901	50924

Appendix 1. continued

TgCpGw_R6_004F	CTGGCAAATAGACATCCACAGGG	23	50815	50837
TgCpGw_R6_004R	CCACGTTCATAGGAGTGCGTC	21	51643	51663
TgCpGw_R6_005F	TCCGGCCCTTTCCTAATCGG	20	51526	51546
TgCpGw_R6_005R	CTTCGTCATTAATCTTCGGAGGG	23	52385	52407
TgCpGw_R6_006F	GCCTAACCCCTTGATTGTGGAC	22	52342	52363
TgCpGw_R6_006R	GCTCAGTTGGTAGAGCACCTC	21	53259	53279
TgCpGw_R6_007F	GGCACTCAAATCGGACCGAACC	22	53121	53142
TgCpGw_R6_007R	GGTTAACGATGGCTCTGATGGGTG	24	53987	54010
TgCpGw_R6_008F	GAGTTTGCTGAGCTTCTTGTGG	22	53894	53915
TgCpGw_R6_008R	CGGCCTTATTGGGTAGAATGCC	22	54795	54816
TgCpGw_R6_009F	GGTACTCAGGGTGGGTTGATAACCG	25	54761	54785
TgCpGw_R6_009R	ATCCTACTACTTCTGGTTCTGGGG	24	55639	55662
TgCpGw_R6_010F	AGGCTACATCTAGTACCGGACC	22	55570	55591
TgCpGw_R6_010R	CCCGCTTTGAATCCAACACTTG	22	56483	56504
TgCpGw_R6_011F	GCCTAATTCATTCGTGTCGAGTAG	24	56385	56408
TgCpGw_R6_011R	CCATTATCTCGGCAATAATGAGCC	24	57297	57320
TgCpGw_R6_012F	GAATTGGGAGTTCCTATCGTAATGC	25	57232	57256
TgCpGw_R6_012R	CTCTTTCCATACTTCACAAGCAGC	24	57826	57849
TgCpGw_R4_001F	CAGAAGCAATAACCGTAAATGGACCC	26	122976	123001
TgCpGw_R4_001R	ACCTAGTATCGTCATAATGTCAGCC	25	122088	122112
TgCpGw_R4_002F	GGAGATGGAAAATTCGACCGCC	22	122164	122185
TgCpGw_R4_002R	GTAAACAGTCTCGACTGCTTGTG	23	121268	121290
TgCpGw_R4_003F	CGGGTTGATGAATTAAACCAGATAG	25	121379	121403
TgCpGw_R4_003R	GCGCTTCAACTATATCAACTGTAC	24	120491	120514
TgCpGw_R4_004F	CGGTTCTGTAATAGCGATGGG	21	120557	120577
TgCpGw_R4_004R	CGGGTAGATCTATAGGACATACACG	25	119673	119697
TgCpGw_R4_005F	GTTTCCGAGGCCGAATCCAC	20	119737	119756
TgCpGw_R4_005R	GTAACTCCATCCCCAACGGTCC	22	118765	118786
TgCpGw_R4_006F	GCAGCTACTTATTTATGTAGGAGC	24	118858	118881
TgCpGw_R4_006R	CGATAAATTGATGAAACAATAGCCG	25	118040	118604
TgCpGw_R4_007F	GATAATCGCCAATTAAGAGGAGAC	24	118117	118140
TgCpGw_R4_007R	ATAGCAGAATGCATAAGTCGTTAG	24	117213	117236
TgCpGw_R4_008F	AGGGGAAATAGAGTAATCCGTTGG	24	117270	117293
TgCpGw_R4_008R	CACTCATACCCGGTAATGCAAGG	23	116195	116217
TgCpGw_R4_009F	CCTTGCATTACCGGGTATGAGTG	23	116195	116217
TgCpGw_R4_009R	CTCCAAGGAATTATTACAGGTCGC	24	115231	115254
TgCpGw_R4_010F	CCTAGACTAATAACCCGATAACTCC	25	115279	115303
TgCpGw_R4_010R	CGCTATGGTGAAATTGGTAGACAC	24	114446	114469
TgCpGw_R4_011F	ACTCGGACTCGAACCGAGATG	21	114496	114516
TgCpGw_R4_011R	GAACAATAGATGTCTTTCACATCCAGC	27	113475	113501
TgCpGw_R4_012F	GATTACTCATTGGTATAGCTGGATGTG	27	113492	113518
TgCpGw_R4_012R	CCCAACATGGAAGTACTGAAAAGGC	25	112588	112612

Appendix 1. continued

TgCpGw_R4_013F	ATGTATGGTCTTGGACCCTCG	21	112766	112786
TgCpGw_R4_013R	CCATAGCATCAGGTAACCAGAC	22	112235	112256
TgCpGw_R4_014F	GTGCTGCTCTATTATTTGCCG	21	112292	112312
TgCpGw_R4_014R	GAAATACGGCCACCCCCTAC	20	110836	110855
TgCpGw_R3_001F	ACTATCAAGATCAAGGAAGAGGTG	24	65213	65236
TgCpGw_R3_001R	CGATGTATTTGGAAGCCCTCG	21	66098	66118
TgCpGw_R3_002F	GTTCCAAAGGATCAAAACGGCC	22	66027	66048
TgCpGw_R3_002R	TTTCTATCAACCGAGTCGTTGAG	23	66919	66941
TgCpGw_R3_003F	CTCCCCTTTGATTTTAAGGAATTGG	25	66623	66647
TgCpGw_R3_003R	CTGATCACCACGCCTGTATTG	21	67521	67541
TgCpGw_R3_004F	GTCTAATTCCTATTACTTTGGCTGG	25	67474	67498
TgCpGw_R3_004R	GGGATTAGTCACAGATATTTGATTG	25	68451	68475
TgCpGw_R3_005F	CCCGGATGCGTTGACATTCCC	21	68377	68397
TgCpGw_R3_005R	GGCAAACGCCTACGAAAAGATC	22	69304	69325
TgCpGw_R3_006F	CCATGGATAAATCCAAGCGACC	22	69265	69286
TgCpGw_R3_006R	CAATGACCAGAATTAAACGCGG	22	70174	70195
TgCpGw_R3_007F	TGAGACCCTCGAAAACTTGATGC	23	70111	70133
TgCpGw_R3_007R	GAACATGTACTAGGGTGTATGTGCG	25	71025	71049
TgCpGw_R3_008F	GATCATTCATACTGGAAGCGCT	22	70969	70990
TgCpGw_R3_008R	GGAAATCCAATTCAGCGTCACAAAC	25	71948	71972
TgCpGw_R3_009F	TATTGTTAAGAGCCCTTCGGTGTG	24	71913	71936
TgCpGw_R3_009R	TTACCCGGAGTAGAGCATAAACC	23	72815	72837
TgCpGw_R3_010F	GAGATTCAATCCAAATCACGATGG	24	72752	72775
TgCpGw_R3_010R	CTATGGACTTCCCTTGTTCAGTG	23	73650	73672
TgCpGw_R3_011F	GTATCAATCCACCATTGCGTATTGG	25	73505	73529
TgCpGw_R3_011R	GTTGTGGTGGTCGAACGCTAAG	22	74398	74419
TgCpGw_R3_012F	CTCATCATATTGCAGCAGGCACG	23	74348	74370
TgCpGw_R3_012R	GGGGTCTTGTATCAGACTGCCTG	23	75265	75287
TgCpGw_R3_013F	CTCAAGTAGAGTTTGGAGCATTC	23	75212	75234
TgCpGw_R3_013R	GGCCAGAAATTCGAGTTCCAAC	22	76134	76155
TgCpGw_R3_014F	TCGCAATGGCTCTATTTGCGG	21	76008	76028
TgCpGw_R3_014R	CAATCGCTTGAATCTCGAGACG	22	76991	77012
TgCpGw_R3_015F	CGAGATGAAAGTCTCATATACGG	23	76913	76935
TgCpGw_R3_015R	GATTCTCTTCAAGCGAACCAG	21	77923	77943
TgCpGw_R3_016F	GTAGATATATGTCTGCCGCATTGG	24	77839	77862
TgCpGw_R3_016R	CGATCAGTTCATCGATGGTCG	21	77207	77227
TgCp_rpl16_For	GCTTAGTGTGTGACTCGTTGC	20	83746	83766
TgCp_rpl16_Rev	GGTTCAAGTGCCTGAAGAGC	20	82770	82789

Appendix 2. Locus-specific and extension primers used for SNaPshot genotyping. *In the multiplex primer extension, poly-t tails of different lengths were added to the 5' ends of each probe to separate the extension products by size.

Primer Name	Sequence (5'-3')*	Tm (°C)	Primer Length (bp)	Amplicon size (bp)	Mutation	Target SNP	Position (*NC020089*)	Region
TgCp_SNP_001F	GCATTACACAATCTCCAAGATG	62	22	332	G/T	G/T	8,316	*psbK* -IGS-*psbI*
TgCp_SNP_001R	GCCATAAATCAGGCTTACCC	62	20					
TgCp_SNP_001	ttCAGTGAAAGACCCTAACCCTATTAG	60	27					
TgCp_SNP_002F	GGTGTAGATAAATGGAAAGATGAGAG	60	26	392	C/A	C/A	83,525	*rpl16* intron
TgCp_SNP_002R	TTTATTGGCTCGAAGCTCTTG	60	21					
TgCp_SNP_002	CCTTACATATTGGAATTCTATATCATTAATATT	59	33					
TgCp_SNP_003F	CTCGACTAACCAACAAGGAATGTAC	60	25	610	AAA/TTT	A/T	4,797	*trnK* (UUU)-IGS-*rps16*
TgCp_SNP_003R	GACCCGGTTCACTTATCACG	60	20					
TgCp_SNP_003	ttttGAATATTATTGTATTGCATTGTTAAATTGAAATTC	60	39					