Nuclear microsatellites and chloroplast genes revealed strong genetic differentiation between Indian (*Azadirachta indica* A. Juss.) and Thai neem (*A. indica* A. Juss. var. *siamensis*) varieties

Suchitra Changtragoon[1*], Madhav Pandey[1,2], Chanpen Boontong[1,3]

[1] *Forest and Plant Conservation Research Office, Department of National Parks, Wildlife and Plant Conservation, 61 Phaholyothin, Chatuchak, Bangkok 10900, Thailand*
[2]*Molecular Research LP, 503 Clovis Rd, Shallowater, TX 79363, USA*
[3]*The Agricultural Research Development Agency, 2003/61 Phaholyothin, Ladyao, Chatuchak, Bangkok 10900, Thailand*

**Correspondence:* Email: suchitra.changtragoon@gmail.com

ABSTRACT

Knowledge about, within, and among population genetic diversity is important for efficient management of genetic resources of a target species. Genetic diversity and differentiation of 17 accessions of Indian neem and seven accessions of Thai neem varieties were investigated covering South Asia, Southeast Asia, and Africa by employing eight nuclear microsatellites and three chloroplast genes. Genetic differentiation among seventeen Indian and seven Thai neem accessions measured by AMOVA was 27 % and 25 %, respectively. No significant correlation between pairwise genetic and geographic distances was observed for both varieties. Several unique alleles for Indian and Thai neem varieties were detected. Bayesian cluster analyses grouped the accessions from Thai and Indian neem varieties in separated clusters. To elucidate the phylogenetic relationship between the two neem varieties, three genes from chloroplast DNA were also investigated. Overall genetic diversity was higher in Indian neem than in Thai neem variety. The sequence divergence and the phylogenetic analysis based on three cpDNA (matK, atpB-rbcl, and ycfl-b) genes revealed strong differences between the two varieties suggesting that the varieties possibly belong to two different neem species. Implications of the data presented in this study for the conservation and management of genetic resources are discussed.

Keywords: *Azadirachta indica*; conservation; cpDNA genes; genetic diversity; genetic differentiation; microsatellites

INTRODUCTION

Neem (*Azadirachta indica* A. Juss) is one of the most important tropical trees because of its multiple use characteristics such as firewood, timber, traditional medicines, and insecticides (Koul et al. 1990; Ketkar and Ketkar 1995). Additionally, because of its ability to grow in poor soils in arid and semi-arid areas, it is also used for rehabilitation of poor agricultural land. It is native to South and Southeast Asia and has been introduced to other tropical parts of the world (National Research Council 1992; Chamberlain 1999). Neem is mainly pollinated by insects and has been characterized as predominantly outcrossing species with the multilocus outcrossing rate of 90-92 % (Kundu 1999a; Singh et al. 1999).

There are two varieties of neem, i.e. Indian neem (*A. indica*) and Thai neem (*A. indica* var. *siamensis*) (Lauridsen et al. 1991). Indian neem occurs in most of the South Asia and in some parts of Africa, while Thai neem is mainly found in Thailand and Laos. The two varieties can be distinguished by various characteristics such as size, bitterness, color of leaves, shape of leaflets, inflorescences and flowers, size and color of seeds, and straightness of stem (Read 1993; Sombatsiri et al. 1995). Studies based on pollen morphology and allozyme patterns were also able to distinguish the two varieties (Lauridsen et al. 1991; Changtragoon et al. 1996). Based on these observations, earlier studies suggested to separate these two varieties into two separate species. However, so far, no data from the conserved region of the DNA such as chloroplast DNA are available. Genes within chloroplast DNA are widely used to study the systematics of angiosperm because of their high nucleotide substitution, non-synonymous mutations, and indels events (Johnson and Soltis 1994; Hilu and Liang 1997; Soltis and Soltis 1998; Cuenoud et al. 2002; Barthet and Hilu 2007).

Earlier studies based on isozyme or dominant DNA markers such as amplified fragment length

Received 2 February 2021, Revised 25 March 2021, Accepted 17 March 2022

polymorphism (AFLP) and random amplified DNA polymorphisms (RAPD) reported medium to high genetic diversity within and between populations of neem (Changtragoon et al. 1996; Farooqui et al. 1998; Kundu 1999b; Singh et al. 1999; Singh et al. 2002; Deshwall et al. 2005; Dhillon et al. 2007; Da Silva et al. 2013). So far, no genetic study in neem has been reported using species specific highly polymorphic codominant molecular genetic markers such as microsatellites. Additionally, earlier studies covered either only accessions from a single variety or very few international accessions. Therefore, a population genetic study of neem covering accessions from a wider range and the employment of species specific and highly variable codominant genetic markers, such as microsatellites is needed for a better understanding of the state of genetic diversity in the species.

The aims of the present study were; (1) to characterize the genetic diversity and population differentiation of neem accessions within and between Indian and Thai neem varieties, and (2) to resolve the taxonomic issue of the two neem varieties using sequence information from the conserved region such as cpDNA.

MATERIALS AND METHODS

Plant material and DNA extraction

Sampling was carried out in 24 neem accessions from FAO International Neem Provenance Trials established in Kanchanaburi, Thailand in August 1997. The trials were established from the seeds collected from 24 locations across nine countries from Asia and Africa (Figure 1; Table 1). Detailed information about sampling locations and seed source populations can be found in the International Neem Network booklet (FAO, 2000). Out of the 24 accessions, seven are from Thai neem and seventeen are from Indian neem varieties. Leaf samples were collected randomly from 12 to 24 individuals (in total 435 individuals) from each accession, and the leaf samples of each individual were packed in separate plastic bags and stored at -80 °C until DNA was extracted.

DNA extraction, genotyping and sequencing

Approximately 500 mg of leaf tissue from each of the individual plants was used to extract the DNA by employing the CTAB method (Doyle and Doyle 1990). Genotyping of individual trees was performed at eight microsatellite markers using the method described in Boontong *et al.* (2009). Aliquots of the amplification products were separated on 30 % polyacrylamide gel and were visualized by using Gel Scan 3000 (Corbett Robotics, Brisbane, Australia). Alleles were scored manually against 50 bp commercial DNA ladder (Invitrogen, Carlsbad, CA).

One to five samples from fourteen Indian and six Thai neem accessions were used to amplify *matK* (Cuenoud *et al.* 2002), *atpB-rbcL* (Small *et al.* 1998), and *ycf1-b* (Dong *et al.* 2012) genes from chloroplast DNA using the methods described in the respective articles to investigate the phylogenetic relationship between the two neem varieties. We selected these three genes based on their successful application in barcoding and species identification in plants including tropical trees (Dong *et al.* 2015; Kang *et al.* 2017). Sequencing of the amplified DNA fragments was performed using Big Dye Terminator v3.1 Cycle Sequencing Kit in ABI Prism 3100-Avant Genetic Analyzer (Applied Biosystem, Foster City, CA). Each sample was sequenced in both directions (forward and reverse) to create a consensus sequence.

Data analysis

Population genetic diversity parameters: mean number of alleles (N_A), the effective number of alleles (A_E), observed (H_O) and expected heterozygosity (H_E), inbreeding coefficient (*F*) were estimated using the program GENALEX (Peakall and Smouse 2006). Standardized allelic richness (A_R) was estimated based on the rarefaction method using the program HP-RARE (Kalinowski 2005). The significance of differences in mean genetic diversity parameters between accessions of Indian and Thai neem varieties was tested using analysis of variance (ANOVA) and Duncan's multiple range test (DMRT) based on linear model employing SAS 9.1 software package (SAS Inc., Carry, North Carolina). Overall genetic diversity measures for Indian and Thai neem were estimated considering all accessions of the same variety as one population. MICROCHEKER 2.2.3 program (Van Oosterhout *et al.* 2004) was used to check the presence of null alleles in the microsatellites used in this study.

To assess the genetic diversity among the accessions of the two neem varieties, genetic differentiation among populations was examined by Wright's (1965) F-statistics, hierarchical analysis of molecular variance (AMOVA; Excoffier *et al.* 1992) was performed using the GENALEX 6.5 program (Peakall and Smouse 2006). The Mantel test was used to perform an analysis of isolation by distance (IBD) by regressing pairwise geographical distance with pairwise $F_{ST}/1- F_{ST}$ between accessions of neem within each variety. The test was performed for Indian and Thai neem varieties separately. A significance test of

Figure 1 A Google map showing approximate locations of 24 neem accessions of Indian and Thai neem varieties (see also Table 1).

Table 1 Names and geographical coordinates of seed source populations of 24 accessions of Indian and Thai neem varieties from the international provenance trial of neem established at Kanchanaburi and Kamphaengphet provinces in 1997.

Variety	Country	Accession	Accession ID	Accession code	No. of Samples	Geographical position	
						Latitude	Longitude
Thai neem	Thailand	Ban Bo, Kanlasin	TB3	A	15	16º17'N	103º35'E
		Ban Nong Hoi, Kanchanaburi	TN3	B	16	14º09'N	99º19'E
		Sai Ngram*		B1	1	16°45′N	99°52′E
		Doi Tao, Chiang Mai	TD3	C	18	17º57'N	98º41'E
		Khao Laung, Nakhon Sawan	TK2	D	23	15º32'N	99º57'E
		Tung Luang, Suratthani	TT3	E	12	09º09'N	99º07'E
		Ban Nong Rong, Kanchanaburi	TNN2	F	20	14º05'N	99º40'E
	Laos P.D.R	Vientiane	LV1	G	23	18º00'N	102º45'E
Indian neem	Myanmar	Yezin	MY2	H	19	19º51'N	96º16'E
		Myene	MM2	I	24	22º03'N	95º13'E
	India	Ramannaguda, Orissa	IR3	J	18	19º05'N	83º49'E
		Sagar, Madhya Pradesh	IS3	K	20	21º51'N	78º45'E
		Balharshah, Maharashtra	IB2	L	20	19º51'N	79º25'E
		Ghaati Subramanya, Karnataka	IG2	M	21	13º22'N	77º34'E
		Chitradurga, Karnataka	IC2	N	15	14º02'N	76º04'E
		Mandore, Jodhpur	IM2	O	12	26º18'N	73º01'E
		Annur, Tamil Nadu	IA2	P	19	11º17'N	77º07'E
		Allahabad Town, Uttar Pradesh	IAI1	Q	13	25º28'N	81º54'E
	Nepal	Lamahi	NL2	R	22	27º52'N	82º31'E
		Geta	NG3	S	19	28º46'N	80º34'E
	Pakistan	Tibbi Laran, Rahimyar Khan	PT2	T	22	28º24'N	70º18'E
		Multan, Cantonment Area	PM2	U	12	30º11'N	71º29'E
	Sri Lanka	Kuliyapitiya	SK1	V	13	07º08'N	80º00'E
	Ghana	Sunyani	GS2	W	17	07º21'N	02º21'W
	Senegal	Bandia	SB2	X	22	14º30'N	17º02'W

*Only one sample was collected from this accession and used only for cpDNA analysis.

the analysis was performed with 9999 bootstraps. As there is a large geographical gap between the African and Asian accessions, the African accessions were excluded from the Mantel test analysis. The GENALEX 6.5 program (Peakall and Smouse 2006) was used to perform afore-mentioned analyses.

To resolve inter population genetic relationships, Bayesian approach-based cluster analysis was performed using STRUCTURE program (Pritchard *et al.* 2000). For this analysis, the number of subpopulations (K) was set 1 to 24 and analysis for each K was performed 20 times and burn-ins were set to 50000 followed by 100000 iterations. The optimum number of clusters was estimated based on ΔK parameter described by Evanno *et al.* (2005) using the program STRUCTURE HARVESTER (Earl 2009).

To construct phylogenetic trees, sequences of each cpDNA marker were aligned separately using Clustal W (Thompson *et al.* 1994) within MEGA X (Kumar *et al.* 2018). The aligned sequences were used to construct a phylogenetic tree using the same program. Before constructing tree, a test was performed to select the best model using find best model option in MEGA X. The test revealed that Jukes-Cantor model (Jukes and Cantor 1969) is the best model for all cpDNA sequence data and therefore the model was used to construct the trees with 1000 bootstrap replicates. To construct the trees the uniform rates among sites was applied. Additionally, mean pairwise sequence distance (*p*; Kimura 1980) within and between the two neem varieties was estimated for the three chloroplast genes based on a combined transition and transversion substitution model using the program MEGA X (Kumar *et al.* 2018). The distance was estimated using only the variable sequence sites (Figure S1, Supplementary Material).

RESULTS

A total of 57 alleles at eight microsatellite loci were detected in 453 individuals of 24 accessions of Indian and Thai neem varieties. The number of alleles per locus ranged from three to ten across 24 neem accessions (Table 2). The average number of alleles per locus in Indian and Thai neem varieties were 4.04 and 3.52, respectively. Observed heterozygosity (H_O) for Indian and Thai neem varieties ranged from 0.25 to 0.62 and from 0.44 to 0.61, respectively. Similarly, the expected heterozygosity (H_O) for Indian and Thai neem varieties ranged from 0.44 to 0.69 and from 0.45 to 0.57, respectively. The mean expected heterozygosity (H_E) was higher in Indian (average H_E = 0.58) than in Thai neem (average H_E = 0.53). Although DMRT in ANOVA did not show any significant differences in the mean observed (H_O) and expected heterozygosities (H_E) between the two varieties, the mean values of allelic diversity measures (N_A, A_E and A_R) were significantly higher in Indian than in Thai neem varieties, respectively (Table 2). The average fixation index was relatively higher in accessions of both Indian (F = 0.17) and Thai (F = 0.12) neem varieties. Eight and five unique alleles were found in Indian and Thai neem varieties, respectively (Table 3). The frequency of unique alleles in the Indian neem variety ranged from 0.03 to 0.75 with six out of the eight alleles showing a frequency ≥0.15, while in the Thai neem variety, it ranged from 0.02 to 0.75 with only one out of the five alleles showing a frequency below 0.07. Additionally, several of the unique alleles found between varieties were also found within the two varieties (Table S1; Supplementary Material). Within the Indian neem variety, five (accessions J, L, M, R and T) of the 17 accessions possessed at least one unique allele. In contrast three (accessions A, F and G) of the seven accessions of Thai neem variety have at least one unique allele. None of the eight microsatellites used in this study showed null alleles in all accessions. Seven microsatellites showed the presence of null alleles in three to nine accessions while one microsatellite (Npct_5) showed null alleles in 14 accessions.

Genetic differentiation (F_{ST}) between the two varieties was about 20 %, while it was 15 % and 12 % among accessions within Indian and Thai neem varieties, respectively. When the Fst was estimated using the null alleles corrected genotype data, it was 19% between the two varieties, and 13 % and 12 % among accessions within Indian and Thai neem varieties, respectively. Hierarchical AMOVA showed that 31 % of the genetic variation of Indian and Thai neem was distributed among populations (Φ_{PT} = 0.312, P = 0.010) and 69 % among individuals within populations (Table 4). In Indian neem, 27 % genetic variation was among populations (Φ_{PT} = 0.271, P = 0.010) and 73 % of the genetic diversity was within populations. Similarly, in Thai neem 25 % genetic variation was among populations (Φ_{PT} = 0.253, P = 0.010) and 75 % of the genetic variation was within populations. AMOVA estimated based on null allele corrected data showed that 38 % of the genetic variation of Indian and Thai neem was distributed among populations (Φ_{PT} = 0.375, P = 0.000) and 62 % among individuals within populations. In Indian neem, it was 28 % variation among populations (Φ_{PT} = 0.282, P = 0.000) and 72 % of the genetic diversity was within populations. Similarly, 26 % genetic variation was among populations (Φ_{PT} = 0.260, P = 0.000) and 75 %

Table 2 Estimates of population genetic diversity measures in 24 accessions of Indian and Thai neem varieties.

Variety	Accession	N_A	N_E	A_R	H_O	H_E	*F*
Thai neem	A	3.50	2.32	3.13	0.46	0.53	0.19*
	B	3.13	2.14	2.87	0.44	0.49	0.12
	C	3.63	2.45	2.89	0.45	0.55	0.18*
	D	4.00	2.66	3.24	0.50	0.56	0.14
	E	3.13	2.45	3.04	0.47	0.54	0.08*
	F	3.63	2.55	3.04	0.61	0.57	-0.06*
	G	3.63	2.27	4.23	0.34	0.45	0.16*
Indian neem	H	4.00	3.00	3.64	0.43	0.59	0.30*
	I	4.00	2.76	3.31	0.45	0.56	0.15
	J	4.75	3.34	3.56	0.50	0.66	0.23*
	K	4.25	3.02	3.84	0.49	0.65	0.28*
	L	4.38	2.90	3.64	0.52	0.58	0.09*
	M	5.00	3.68	4.33	0.54	0.69	0.16*
	N	4.13	2.95	3.73	0.54	0.62	0.11*
	O	3.63	2.14	3.33	0.36	0.51	0.29*
	P	4.88	3.34	4.21	0.62	0.64	0.01
	Q	3.63	2.51	3.45	0.57	0.57	-0.01*
	R	4.25	2.59	3.50	0.54	0.54	0.08*
	S	3.25	1.93	3.59	0.34	0.44	0.23*
	T	4.00	2.74	2.78	0.58	0.59	0.01*
	U	3.00	2.32	3.52	0.36	0.48	0.35*
	V	4.25	3.04	2.89	0.57	0.64	0.09*
	W	4.25	3.08	4.01	0.61	0.65	0.05*
	X	3.00	2.40	3.77	0.25	0.50	0.45*
Mean Thai neem		3.52^A	2.40^A	3.21^A	0.47^A	0.53^A	0.12^A
Mean Indian neem		4.04^B	2.81^B	3.59^B	0.49^A	0.58^A	0.17^A
Mean total		3.89	2.69	3.48	0.48	0.57	0.16

N_A, mean number of alleles per locus; N_E, effective number of alleles A_R, rarefied allelic richness; H_O, observed heterozygosity; H_E, expected heterozygosity; *F*, fixation index. * indicates significant deviation ($P \leq 0.05$) from Hardy-Weinberg expectation. Means followed by different letters are statistically different at $P \leq 0.05$ according to Duncan's multiple range tests.

Table 3 Unique alleles and their frequency found only in Indian or Thai neem varieties.

Variety	Locus	Unique allele (bp)	Highest frequency	Accession/s
Thai neem	Ai_5	170	0.020	G
	Ai_6	134	0.700	A, B, C, D, E, F
		138	0.750	A, B, C, D, E, F
	Ai_11	259	0.067	A, B, D
	Ai_48	101	0.306	A, B, C, D, E, F, G
Indian neem	Ai_5	130	0.147	M, N, O, V, X
	Ai_13	162	0.211	K, M, P, Q, R, X
		194	0.294	I, J, M, N, V, W, X
		232	0.029	J
	Ai_14	236	0.025	M
	Ai_48	119	0.750	H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X
		121	0.520	H, I, J, K, L, M, N, P, Q, R, S, T, U, V, W, X
		125	0.370	H, I, J, L, M, N, P, Q, R, S, T, U, V, X

of the genetic variation was within populations. Although the Mantel test performed according to the varieties showed positive correlation between the pairwise $F_{ST}/1\text{-}F_{ST}$ and geographic distances (Indian neem; $R^2 = 0.006$, $P = 0.287$, Thai neem; $R^2 = 0.159$, $P = 0.164$), they were not statistically significant (Figure S2, Supplementary Material). The cluster analysis performed based on Bayesian approach using STRUCTURE grouped the accessions according to variety (Figure 2). The UPGMA dendrogram constructed based on Nei's (1978) genetic distance showed a strong grouping of accessions according to variety (Figure S3, Supplementary Material).

All maximum likelihood phylogenetic trees estimated based on cpDNA markers clearly separated the 24 accessions into two major clades according to the two varieties (Figure 3-5). The trees from *matK* and *ycf1-b* genes grouped all accessions from Indian neem into a single clade while that of Thai neem were grouped into three and four sub-clades, respectively. Although the tree from *atpB-rbcL* gene grouped the two accessions into two major separate clades, four sub-clades were found in both accessions. The maximum likelihood tree constructed based on the concatenated sequences of all three chloroplast genes also grouped all accessions according to the varieties (Figure 6).

The mean pairwise distance among sequences of all three chloroplast genes were higher among the individuals of Thai neem than in Indian neem (Table S2, Supplementary Material). When the distance was estimated between the varieties, a strong differentiation between the two varieties was observed for all three chloroplast genes (Table S2, Supplementary Material).

DISCUSSION

A major finding of this study is that the genetic differentiation between Indian and Thai neem varieties is strong at both nuclear and chloroplast DNA levels. This is probably due to independent evolution of these two varieties as a result of restricted gene flow. To explain the variation in nuclear DNA, the geographic isolation due to the distance barrier could be one of the main reasons for the limited gene flow between these varieties. Moreover, the two varieties have non-overlapping flowering times. Indian neem produces flowers in November/December, while Thai neem flowers mainly in March (Sombatsiri *et al.* 1995).

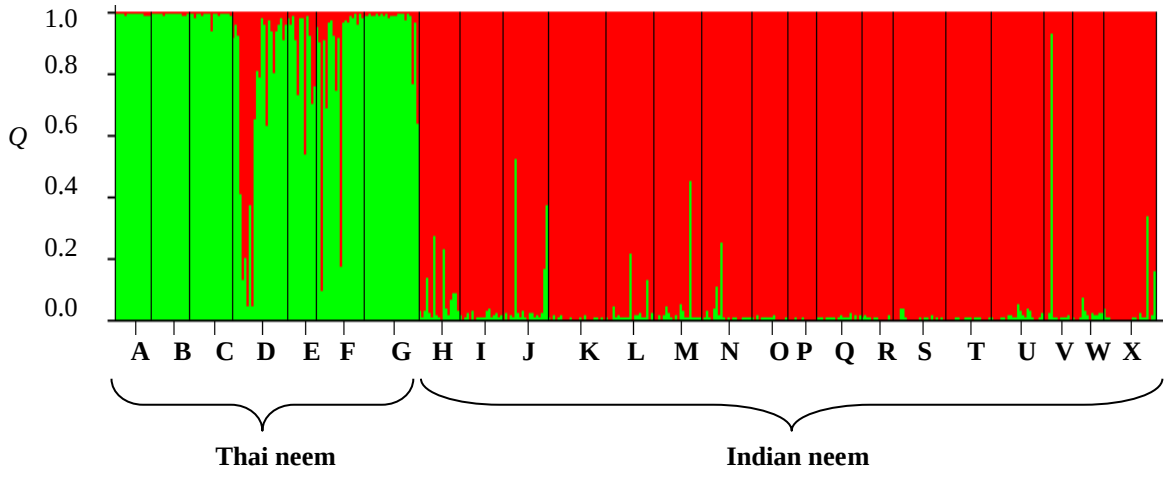


Figure 2 Summary bar plot of estimation of the membership coefficient (Q) of 24 neem accessions of Indian and Thai neem varieties. Each individual is represented by a single vertical line of two-colored segments (K = 2). The Figure shows two clusters of accessions: cluster one, Thai neem (A, B, C, D, E, F and G), cluster two, Indian neem (H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W and X).

Table 4 Estimated genetic variation among populations based on hierarchical AMOVA (Φ_{PT})

Variety/Combined	Φ_{PT} before null allele correction	P-value	Φ_{PT} after null allele correction	P-value
Thai neem	0.253	0.010	0.260	0.000
Indian neem	0.271	0.010	0.282	0.000
Two varieties combined	0.310	0.010	0.375	0.000

atpB-rbcL

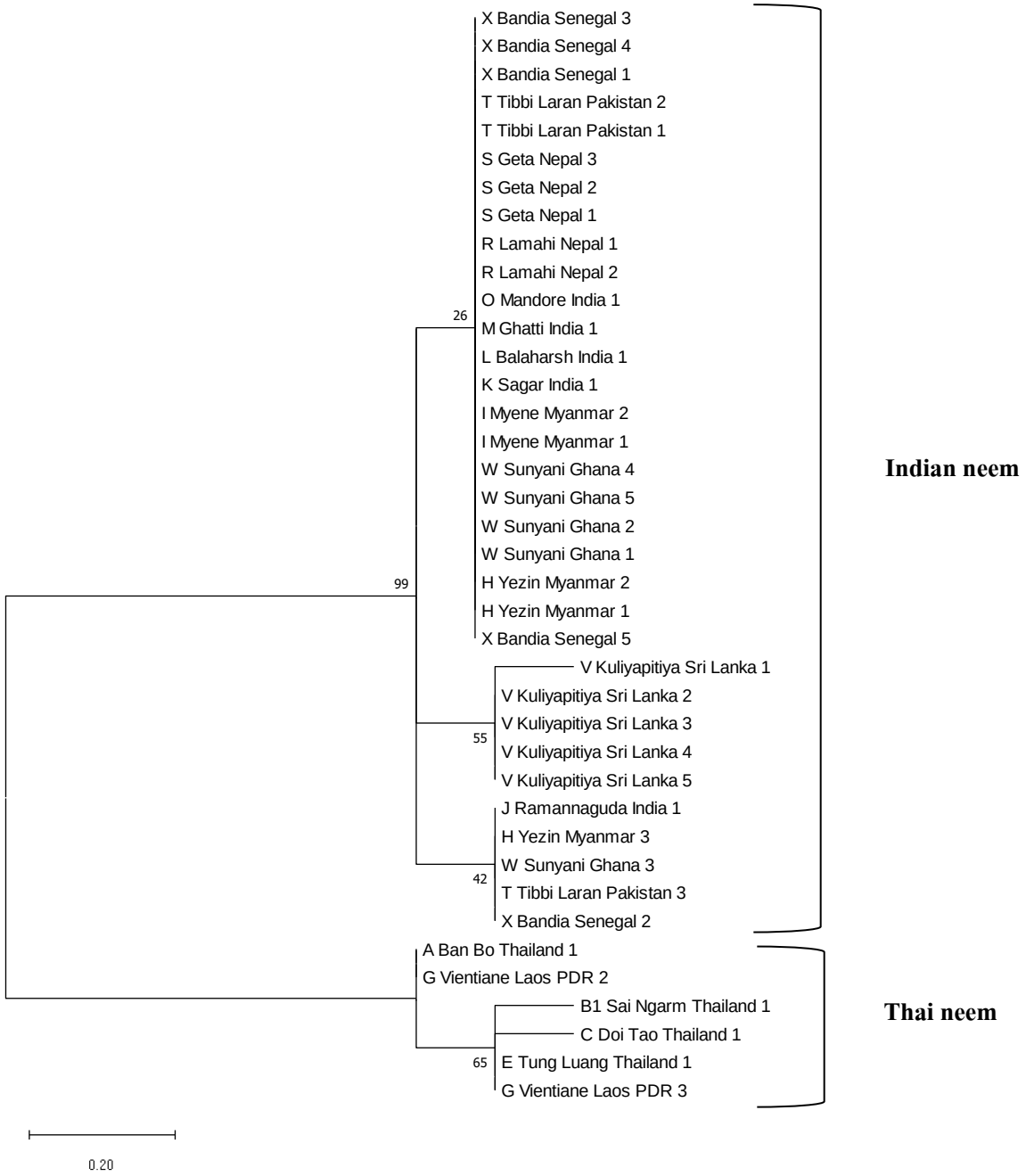


Figure 3 Maximum likelihood trees constructed based on atpB-rbcL (chloroplast DNA sequences of Indian and Thai neem accessions**.** The trees were constructed based on Jukes-Cantor model (Jukes and Cantor 1969). Values on the sides of the branch nodes of the trees are the percentage of 1000 bootstrap.

matK

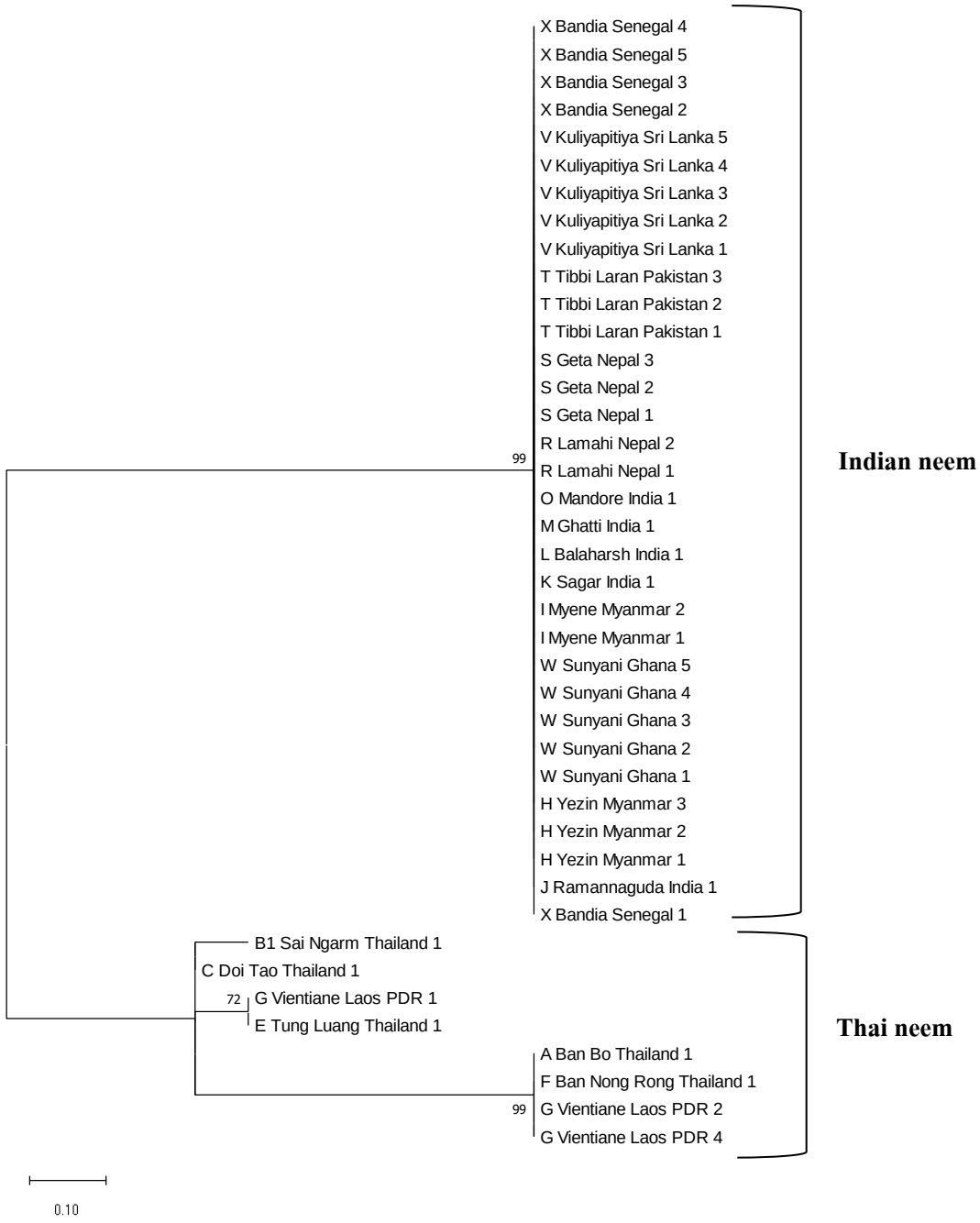


Figure 4 Maximum likelihood trees constructed based on matK chloroplast DNA sequences of Indian and Thai neem accessions**.** The trees were constructed bassed on Jukes-Cantor model (Jukes and Cantor 1969). Values on the sides of the branch nodes of the trees are the percentage of 1000 bootstrap.

ycf1-b

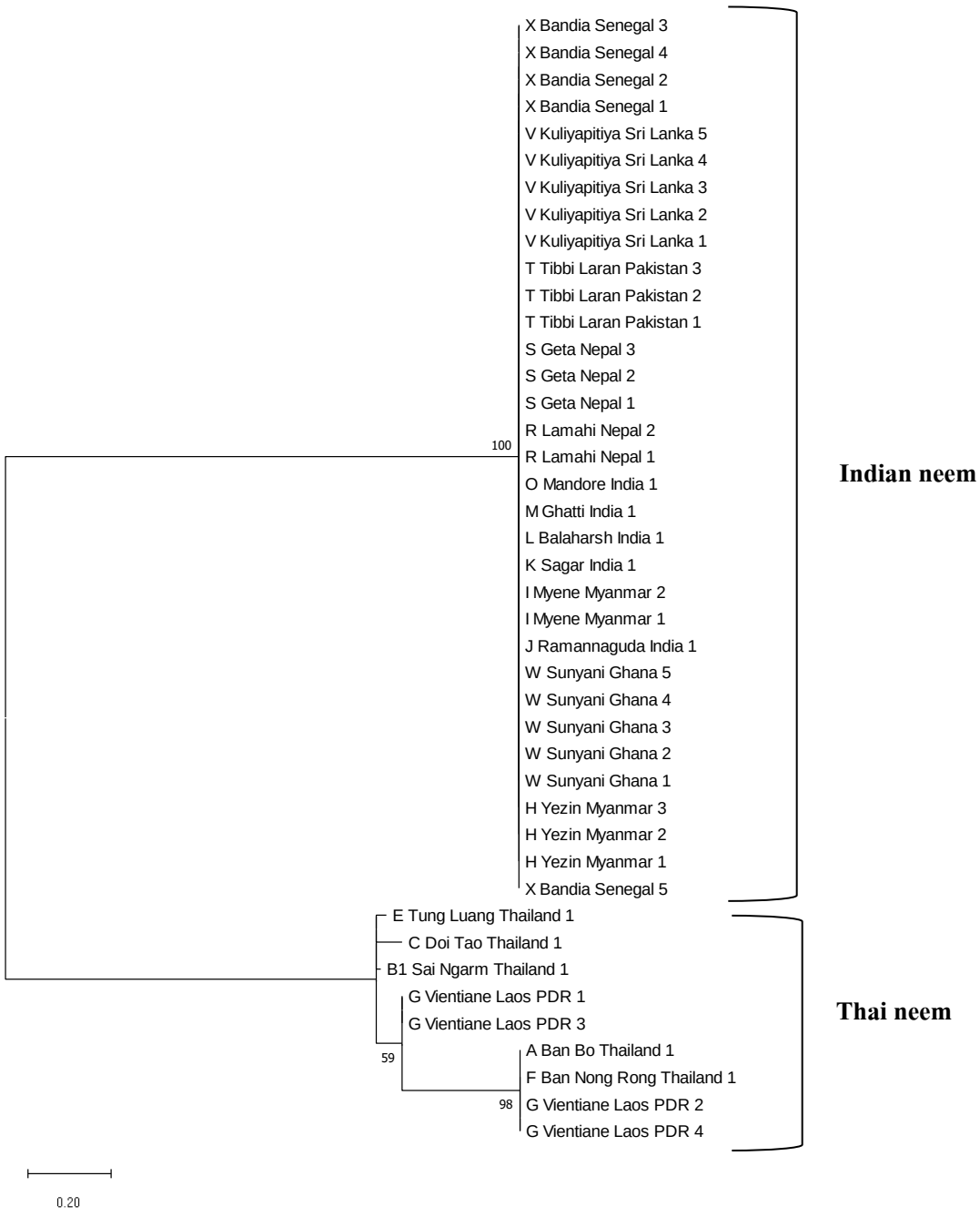


Figure 5 Maximum likelihood trees constructed based on ycf1-b chloroplast DNA sequences of Indian and Thai neem accessions**.** The trees were constructed bassed on Jukes-Cantor model (Jukes and Cantor 1969). Values on the sides of the branch nodes of the trees are the percentage of 1000 bootstrap.

concatenated

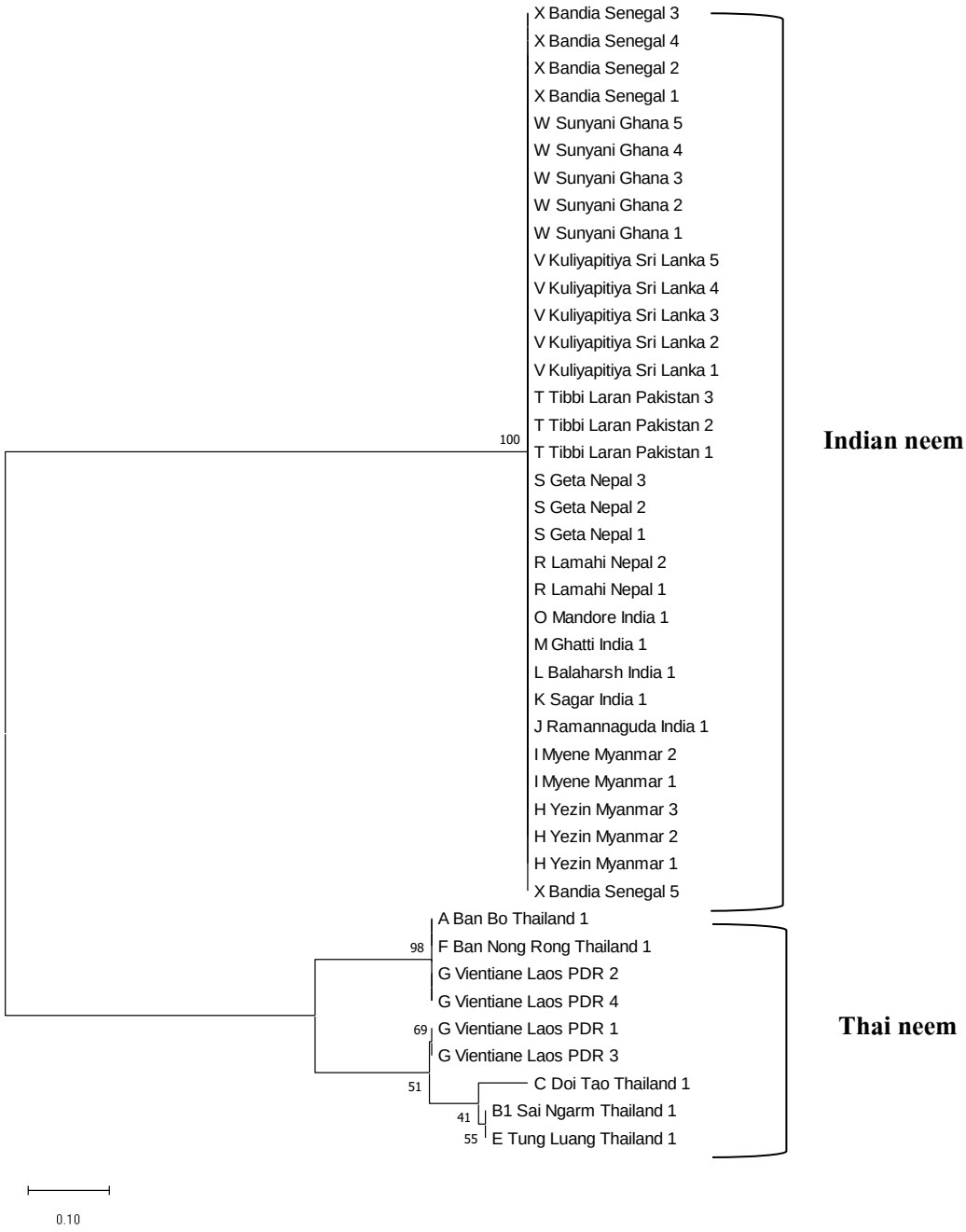


Figure 6 Maximum likelihood trees constructed based on concatenated chloroplast DNA sequences of Indian and Thai neem accessions**.** The trees were constructed bassed on Jukes-Cantor model (Jukes and Cantor 1969). Values on the sides of the branch nodes of the trees are the percentage of 1000 bootstrap.

Together, the distance barrier as well as the difference in flowering time may have prevented gene flow, thus increasing the genetic differentiation between the varieties. This is also evident from the results of cluster analyses based on STRUCTURE, that showed distinct groups for Indian and Thai neem accessions. Similar results were also reported by Changtragoon *et al.* (1996), Singh *et al.* (1999, 2002) using isozyme, and AFLP and SAMPL markers, respectively.

Accessions from the Indian neem variety possessed higher mean genetic diversity than those of the Thai neem. This is fairly obvious as the geographical distribution of the Indian neem variety is wider than that of the Thai neem variety. In general, a species that has a wide distribution range possesses higher genetic diversity as compared to a species with narrow distribution range to persist in a variety of environmental conditions (Hamrick and Godt 1996).

As expected, mean genetic diversity of neem observed in this study using eight microsatellite markers (N_A = 3.89, H_O = 0.48, H_E = 0.57) was considerably higher than the genetic diversity reported in earlier studies using isozyme markers (N_A = 1.82, Ho = 0.29, H_E = 0.23, Changtragoon *et al.* 1996; N_A = 2.08, H_O = 0.25, H_E = 0.34, Kundu 1999b). When the microsatellite genetic diversity of neem and other multipurpose tropical agroforestry trees was compared, it was higher in neem than in the shea tree (*Vitellaria paradoxa*) (N_A = 2.33, H_O = 0.31, H_E = 0.32, Sanou *et al.* 2005). However, the genetic diversity in neem was lower than that found in *Acacia mellifera* (N_A = 9.09, H_O = not reported, H_E = 0.67, Ruiz-Guajardo *et al.* 2007), *A. senegal* (N_A = 6.85, H_O = 0.71, H_E = 0.69, Omondi *et al.* 2010) and *Eucalyptus globulus* (N_A =17.08, H_O = 0.66, H_E = 0.85, Steane *et al.* 2001).

In addition to the nuclear microsatellite markers, the analysis of cpDNA genes clearly separated the accessions of the two varieties into two groups suggesting their long-term independent evolutionary history. The genetic distinctness of the two neem varieties found in this and previous studies (e.g., Changtragoon et al. 1996, Singh et al. 999; 2002) are concurrent with their phenological and morphological characteristics (Lauridson et al. 1991; Boontawee et al. 1993). Consistent with our findings, earlier studies using cpDNA markers had shown a strong genetic divergence between plant species within a family (e.g., Ge et al. 2002; Yang et al. 2004; Liu et al. 2013). These results indicate that these two neem varieties can be designated as two separate species i.e. *A. indica* and *A. siamensis* for Indian neem and Thai neem, respectively. However, to confirm the species status, further investigation about number of chromosomes, karyotype, and occurrence of hybrid are recommended.

Conservation and management implications

Unique or diagnostic alleles were observed in both Indian (8 unique alleles) and Thai (5 unique alleles) neem varieties at six of the eight microsatellite loci. These microsatellites can be used for seed or planting material identification of both varieties. The unique alleles found in both Indian and Thai neem varieties can be used to check the source of planting material such as seeds or clones. As the microsatellites used in this study showed medium to high levels of allelic diversity, they could also be used to decide to select certain seed collection populations by investigating the genetic variation of the populations under consideration. Out of the seven accessions of Thai neem variety, Ban Nong Hoi from Kanchanaburi (B) was the most representative accession because it showed no significant genetic differentiation with the other five of the six accessions. Vientiane (G) from Laos was the genetically most distinct accession as it shows significant differentiation with all other accessions of the Thai neem variety, and the same accession possesses a unique allele that was not found in any of the studied accessions (Table 3). Three accessions from Thailand (C, D and F) as well as seven (J, K, M, N, P, V and W) out of the 17 accessions of the Indian neem variety showed higher than mean genetic diversity in both the varieties. Therefore, the afore-mentioned accessions should get priority for conservation of genetic resources (Table 2). As accessions J, M and X possess unique alleles (Table 3), they can also be prioritized for conserving the resources. Further studies about fine-scale spatial genetic structure and paternity analysis in natural populations of neem would provide more insight about gene flow, conservation, and management of genetic resources of the species.

ACKNOWLEDGEMENT

This study was financially supported by the Department of National Parks, Wildlife and Plant Conservation, Bangkok, Thailand. The excellent technical assistance of Miss Kittiya Singthong is highly appreciated.

REFERENCES

Barthet MM, Hilu KW. Expression of *matK*: functional and evolutionary implications. Am J Bot. 2007;94:1402-1412.

Boontong C, Pandey M, Changtragoon S. Isolation and characterization of microsatellite markers in Indian neem (*Azadirachta indica* var. *indica* A.Juss) and cross-amplification in Thai neem (*A. indica* var. *siamensis* Valenton). Conserv Genet. 2009;10:669-671.

Boontawee B, Kanchanaburagura C, Boonsermsuk S. Studies on neem in Thailand. In: Read MD, French JH (eds) Genetic Improvement of Neem: Strategies for the Future. Proceedings of the International Consultation on Neem Improvement held at Kasetsart University, Bangkok, Thailand. 18-22 January 1993. Craftsman Press Ltd. 1993. pp 41-45.

Chamberlain JR. Improvement of neem (*Azadirachta indica*) and its potential benefits to poor farmers in developing countries. Report on a visit to India. CNRD, Green College, Oxford, UK. 1999.

Changtragoon S, Szmidt AE, Boontawee B. The study of genetic diversity of *Azadirachta spp*. by using isozyme analysis, In: The proceeding of the Third Asia-Pacific Conference on Agriculture Biotechnology, 10-15, November, 1996. Hau Hin, Prachuapkhirikhan, Thailand. 1996. pp. 353-360.

Cuenoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. Am J Bot. 2002;89:132-144.

Da Silva AVC, Rabbani ARC, Almeida CS, Clivati D. Genetic structure and diversity of the neem germplasm bank from Brazil Northeast. African J Biotech. 2013;12:2822-2829.

Deshwall RPS, Singh R, Malik K, Randhawa GJ. Assessment of genetic diversity and genetic relationships among 29 populations of *Azadirachta indica* A. Juss. using RAPD markers. Genet Resour Crop Evol. 2005;52:285-292.

Dhillon RS, Mohapatra T, Singh S, Boora KS, Singh K. Assessment of genetic diversity in *Azadirachta indica* based on DNA fingerprinting. Indian J Biotech. 2007;6:519-524.

Dong W, Liu J, Yu J, Wang L, Zhou S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS ONE. 2012;7:e35071.

Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J, Zhou S. ycf1, the most promising plastid DNA barcode of land plants. Sci Rep. 2015;5:8348.

Doyle JJ, Doyle JL. Isolation of plant DNA from fresh tissue. Focus. 1990;12:13-15.

Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol. 2005;14:2611-2620.

Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992;131:479-491.

Farooqui N, Ranade SA, Sane PV. RAPD profile variation amongst provenances of neem. Biochem Mol Biol Int. 1998;45:931-939.

Ge S, Li A, Lu BR, Zhang SZ, Hong DY. A phylogeny of the rice tribe Oryzeae (Poaceae) based on *matK* sequence data. Am J Bot. 2002;89:1967-1972.

Hamrick JL, Godt MJW. Effects of life history traits on genetic diversity in plant species. Phil Trans R Soc Lond B. 1996;351:1291-1298.

Hilu KW, Liang H. The *matK* gene: sequence variation and application in plant systematics. Am J Bot. 1997;84:830-839.

Johnson LA, Soltis DE. *matK* DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. Syst Bot. 1994;19:143-156.

Jukes TH, Cantor CR. Evolution of protein molecules. In Munro HN, editor, Mammalian Protein Metabolism, 1969; pp. 21-132, Academic Press, New York.

Kalinowski ST. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Mol Ecol Notes. 2005;5:187-189.

Kang Y, Deng Z, Zang R, Long W. DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. Sci Rep. 2017;7:12564.

Ketkar CM, Ketkar MS. Soap production from mixtures of neem oil with other non-edible or edible oils. In: Schmutterer H (ed) The Neem Tree: Source of Unique Natural products for Integrated Pest Management, Medicine, Industry and other Purposes. Weinheim, Germany. 1995. pp 549-554.

Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980;16:111-120.

Koul O, Isman M B, Ketkar CM. Properties and uses of neem, *Azadirachta indica*. Can J Bot. 1990;68:1-11.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol.2018;35:1547-1549.

Kundu SK. The mating system and genetic significance of polycarpy in the neem tree (*Azadirachta indica*). Theor Appl Genet. 1999a;99:1216-1220.

Kundu SK. Comparative analysis of seed morphometric and allozyme data among four populations of neem (*Azadirachta indica*). Genet Resource Crop Evol. 1999b;46:569-577.

Lauridsen EB, Kanchanabufagura C, Boonsermusuk S. Neem (*Azadirachta indica* A. Juss.) in Thailand. For Genet Resources Info. 1991;19:25-33.

Liu Y, Wang K, Liu Z, Luo K, Chen S, Chen K. Identification of medical plants of 24 Ardisia species from China using the *matK* genetic marker. Pharmacogn Mag. 2013;9:331-337.

National Research Council. Neem: A tree for solving global problems. National Academy Press, Washington D.C. USA. 1992.

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

Omondi S, Kireger E, Dangasuk O, Chikamai B, Odee D, Cavers S, Khasa D. Genetic diversity and population structure of *Acacia senegal* (L) Willd. in Kenya. Trop Plant Biol. 2010;3:59-70.

Van Oosterhout C, Hutchinson B, Wills D, Shipley P. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes. 2004;4:535–538.

Peakall R, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes. 2006;6:288-295.

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945-959.

Read MD. Gaps in the knowledge: supplemental studies needed to support international provenance trials. In: Read MD, French JH (eds) Genetic improvement of neem: strategies for the future. Winrock International, Bangkok, Thailand. 1993. pp 179-186.

Ruiz-Guajardo JC, Otero-Arnaiz A, Taylor T, Stone G, Glenn CT, Otero-Arnaiz NA, Miller JT, Preuss S, Schnabel A. Isolation of polymorphic microsatellite markers in the sub-Saharan tree, *Acacia* (*senegalia*) *mellifera* (Fabaceae: Mimosoideae). Mol Ecol Notes. 2007;7:1138-1140.

Sanou H, Lovett PN, Bouvet J-M. Comparison of quantitative and molecular variation in agroforestry populations of the shea tree (*Vitellaria paradoxa* C.F. Gaertn) in Mali. Mol Ecol. 2005;14:2601-2610.

Singh A, Negi MS, Rajagopal J, Bhatia S, Tomar UK, Srivastava PS, Lakshmikumaran M. Assessment of genetic diversity in *Azadirachta indica* using AFLP markers. Theor Appl Genet. 1999;99:272-279.

Singh A, Chaudhury A, Srivastava PS, Lakshmikumaran M. Comparison of AFLP and SAMPL markers for assessment of intra-population genetic variation in *Azadirachta indica* A.Juss. Plant Sci. 2002;162:17-25.

Soltis DE, Soltis PS. Choosing an approach and an appropriate gene for phylogenetic analysis. In Soltis PS, Soltis DE, Doyle JJ [eds.], Molecular systematics of plants, vol. II, 1–42. Kluwer, Boston, Massachusetts, USA. 1998.

Sombatsiri K, Ermel K, Schmutterer H. Other meliaceous plants containing ingredients for integrated pest management and further purposes. Azadirachta spp. The Thai neem tree: *Azadirachta siamensis* (Val.). In: Schmutterer H (ed) The Neem Tree (*Azadirachta indica* A.Juss.) and Other Meliaceous Plants: Sources of Unique Natural Products for Integrated Pest Management, Medicine, Industry and Other Purposes. VCH Publishers Inc. Weinheim, Germany. 1995. pp 585-597.

Small RL, Ryburn JA, Cronn RC, Seelann T, Wendel JF. The tortoise and the hare: choosing between noncoding plastome and nuclear Adh sequences for phylogenetic reconstruction in a recently diverged plant group. Am J Bot. 1998;85:1301–1315.

Steane DA, Vaillancourt RE, Russell J, Powell W, Marshall D, Potts BM. Development and characterization microsatellite loci in *Eucalyptus globulus* (Myrtaceae). Silvae Genet. 2001;50:89-91.

Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993;10:512-526.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673-4680.

Wright S. The interpretation of population structure by F-Statistics with special regard to systems of mating. Evolution. 1965;19:395-420.

Yang DY, Fushimi H, Cai SQ, Komatsu K. Molecular analysis of *Rheum* Species used as Rhei Rhizoma based on the chloroplast *matK* gene sequence and its application for identification. Biol Pharm Bull. 2004;27:375-383.

Earl D. Structure Harvester v0.3. 2009. Available at: http://users.soe.ucsc.edu/"dearl/software/struct_harvest/

FAO. Description of international provenance trials of neem (Azadirachta indica A. Juss) established by collaborators of the International Neem Network. International Neem Network Booklet. 2000. (http://www.fao.org/docrep/006/AD233E/AD233E01.htm)

SUPPLEMENTARY MATERIAL

Table S1. Unique alleles and their frequency within accessions of Indian or Thai neem varieties. A, Ban Bo; F, Bang Nong Rong; G, Vientiane; J, Ramannaguda; L, Balharshah; M, Ghatti Subramanya; R, Lamahi; T, Tibbi Laran.

Unique alleles within Indian neem variety				Unique alleles within Thai neem variety			
Locus	**Allele (bp)**	**Highest frequency**	**Accession**	**Locus**	**Allele (bp)**	**Highest frequency**	**Accession**
Ai_11	247	0.024	R	Ai_4	210	0.132	F
Ai_13	158	0.095	T	Ai_5	170	0.022	G
	232	0.022	J	Ai_6	122	0.065	G
Ai_14	236	0.025	M		124	0.630	G
Ai_48	105	0.025	L	Ai_11	191	0.111	G
				Ai_13	158	0.107	A

Table S2. Mean pairwise sequence distances (*p*; Kimura 1980) of variable sites, estimated based on combined transitions and transversions substitution model, within Indian and Thai neem, and between the two varieties. Values in parentheses are standard errors (± SE) obtained after 1000 bootstraps.

Locus	**Number of variable sites**	**Locus**	**Sequence distance (*p*)**		
			Indian neem	**Thai neem**	**Between Indian and Thai neem**
atpB-rbcL	10	*atpB-rbcL*	0.064 (0.048)	0.138 (0.080)	1.050 (0.454)
matK	15	*matK*	0.000 (0.00)	0.332 (0.145)	1.480 (0.570)
ycf1-b	17	*ycf1-b*	0.000 (0.000)	0.218 (0.110)	2.120 (0.763)

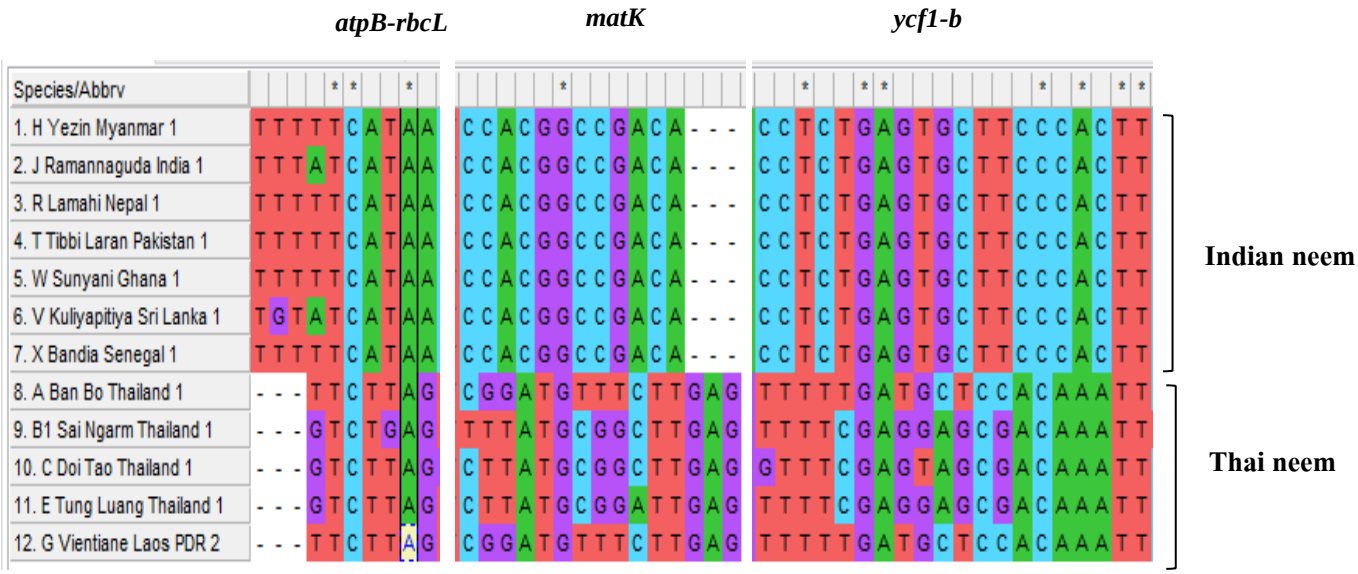
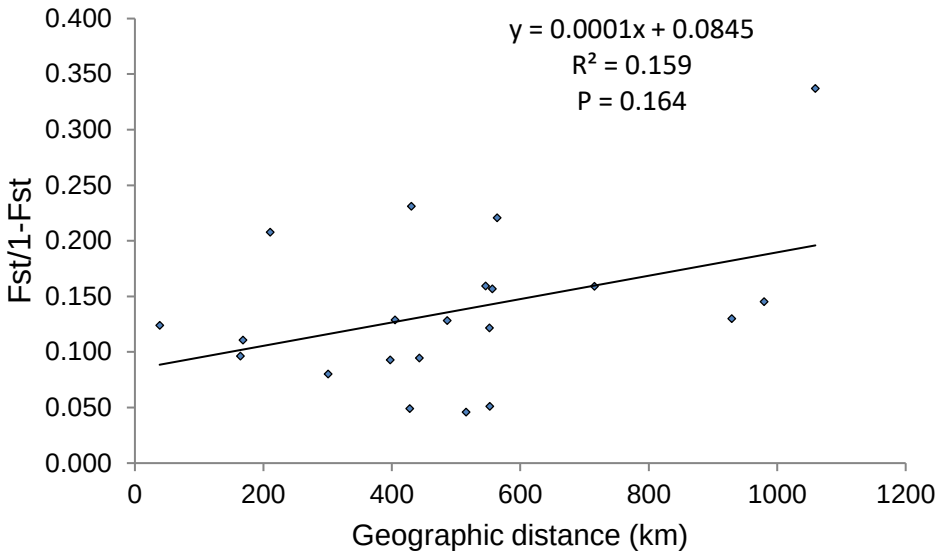


Figure S1 Variable sites of selected cpDNA seqeuences (*atpB-rbcL*, *matK, and ycf1-b*) of Indian and Thai neem varieties.

a.



b.

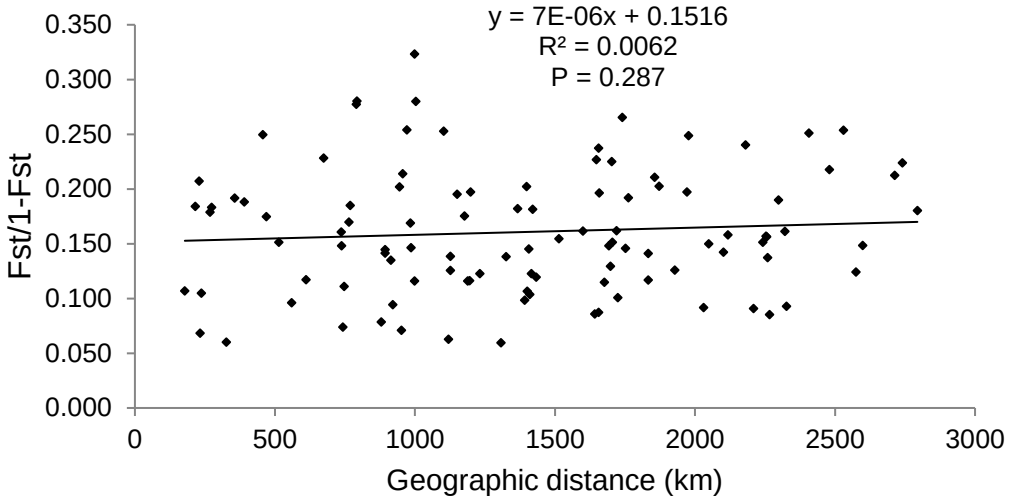


Figure S2 Pairwise F_{ST} /1- F_{ST} regressed on geographic distances of the accessions of Thai (a) and Indian neem (b) varieties.

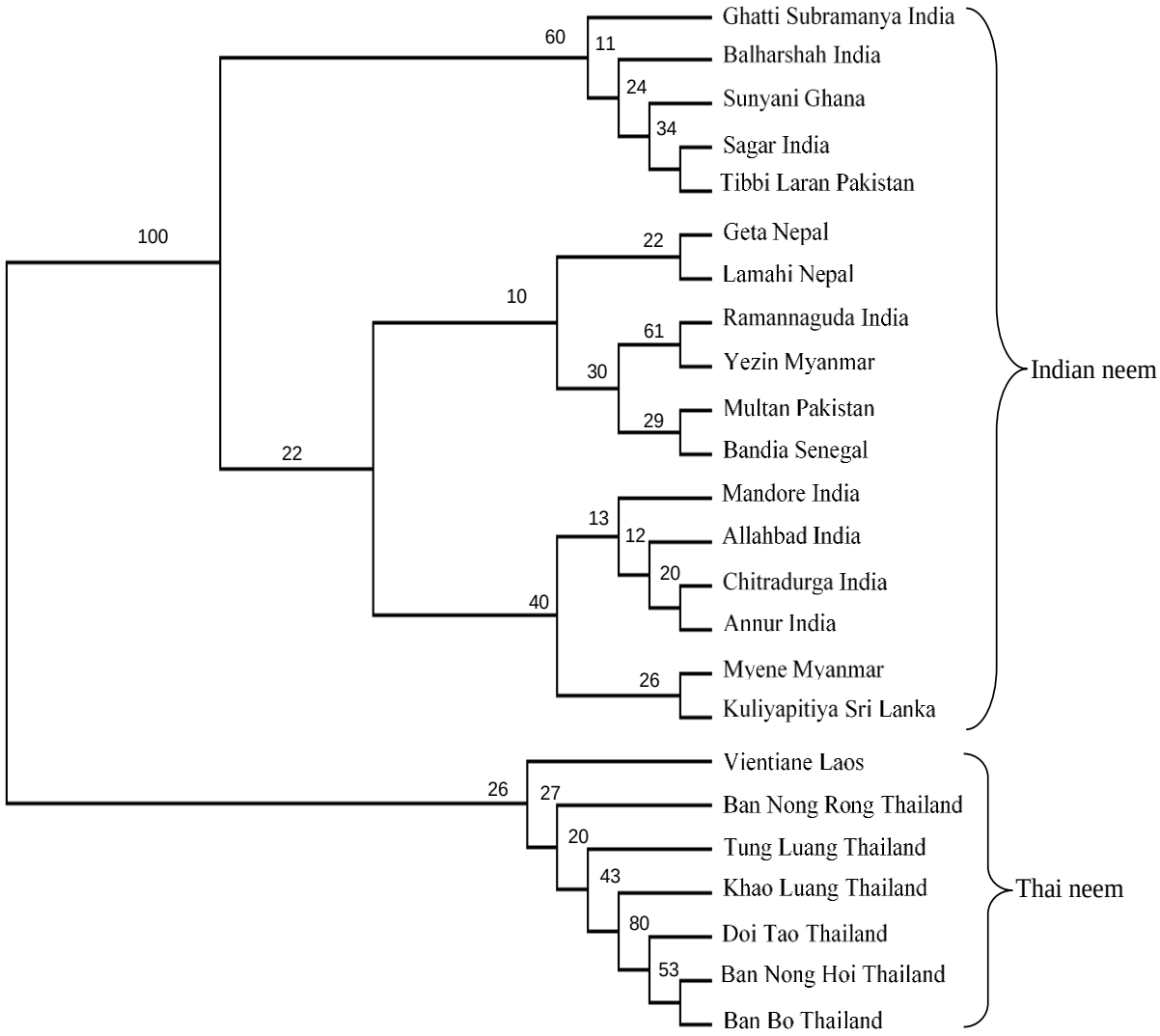


Figure S3. UPGMA tree constructed based on Nei's standard genetic distances of 24 accessions of Indian and Thai neem varieties using the eight nuclear microsatellites. Values at the left of the branch or brances nodes are the percentage of 1000 bootstraps support.