biologia plantarum
an international journal for experimental botany

BIOLOGIA PLANTARUM (2024) 68: 87-96 DOI: 10.32615/bp.2024.010

Microsatellite markers reveal genetic diversity and population genetic structure of the threatened Martaban camphor [*Cinnamomum parthenoxylon* (Jack) Meisn]

M.P. PHAM[1,2,+], T.T.X. BUI[3], D.G. VU[4], V.S. NGUYEN[3], M.D. NGUYEN[5], and D.D. VU[2,+,*]

[1] *Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Cau Giay, Hanoi 100000, Vietnam*
[2] *Join Vietnam-Russia Tropical Science and Technology Research Center, Cau Giay, Hanoi 100000, Vietnam*
[3] *Institute of Ecology and Biological Resource, Vietnam Academy of Science and Technology, Hanoi 100000, Vietnam*
[4] *Institute of Technology, Hanoi University of Industry, Hanoi 100000, Vietnam*
[5] *National Research Center for Medicinal Plant Germplasm and Breeding, National Institute of Medicinal Materials, Thanh Tri, Hanoi 100000, Vietnam*

[*]Corresponding author: E-mail: duydinhvu87@gmail.com

Abstract

Martaban camphor [*Cinnamomum parthenoxylon* (Jack) Meisn] is a woody tree in India, China, Indonesia, Thailand, and Vietnam and has been widely utilized for commercial purposes. It is threatened due to fragmented habitats, over-deforestation, and oil extraction. To conserve this species, the investigation of genetic diversity and population structure of this species is essential. Herein, we analyzed 192 adult trees from eight populations covering its natural distribution range in Vietnam using ten polymorphic EST-SSR markers. Medium levels of genetic diversity (R = 2.7, H_o = 0.399, H_e = 0.426) and genetic differences between populations (F_{st} = 0.223) were determined. Two populations, Cuc Phuong and Xuan Nha have undergone recent bottlenecks. These results indicated that anthropogenic activities may be the major factor for the low heterozygosity and influenced the number of alleles in all *C. parthenoxylon* populations. Clustering analyses revealed three genetic clusters that related to gene flow between different areas. We proposed *in situ* conservation for some populations with high levels of allelic richness, genetic diversity, or private alleles. The collecting of the seeds of the remaining populations for *ex-situ* conservation could be performed.

Keywords: admixture, bottleneck, genetic variability, human activities, species conservation.

Received 23 November 2023, *last revision* 6 May 2024, *accepted* 31 May 2024.

Abbreviations: A - number of alleles; *AMOVA* - analysis of molecular variance; CR - critically endangered; DAPC - discriminant analysis of principal components; E - effective alleles; EST - expressed sequence tag; F_{is} - coefficient of inbreeding; F_{it} - coefficient of total inbreeding; F_{is}IIM - corrected inbreeding coefficient for null alleles; F_{st} - genetic differentiation; H_e - expected heterozygosity; H_o - observed heterozygosity; H_t - total expected heterozygosity; HWE - Hardy-Weinberg equilibrium; ISSRs - inter simple sequence repeat markers; IUCN - International Union for Conservation of Nature; N - sample size; N_A - alleles per locus; NJ - neighbor-joining; N_P - private alleles; NS - not significant; PCoA - a principal coordinate analysis; PCR - polymerase chain reaction; PF - forward primer; PR - reverse primer; R - allelic richness; RAPD - random amplified polymorphic DNA; SMM - the stepwise mutation model; SSRs - simple-sequence repeats (microsatellite); T_m - PCR annealing temperature; TPM - the two-phase model.

Acknowledgements: This research is funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 106.06-2021.02. The first author was funded by the Ph.D. Scholarship Programme of Vingroup Innovation (VINIF), code VINIF.2023.TS.088.

[+]These authors contributed equally to this work and should be regarded as co-first authors.

Conflict of interest: The authors declare that they have no conflict of interest.

Introduction

Martaban camphor [*Cinnamomum parthenoxylon* (Jack) Meisn] is a large forest tree from the family *Lauraceae* with high economic value for its wood and essential oil. The species is naturally distributed in India, China, Indonesia, Thailand, and Vietnam (MOST and VAST 2007). According to JICA (1996) and Sein and Mitlöhner (2011), *C. parthenoxylon* is distributed in tropical evergreen forests and grows among other tree species such as *Michelia* spp., *Phoebe* spp., *Gironniera subaequalis*, *Vatica tonkinensis*, *etc*. Sometimes it can be found growing in groups of 5 - 7 trees in secondary forests. The species is found in the north and center of Vietnam at an elevation of less than 700 m on ancient alluvial rocks, granite, and basalt rocks with sheltered slopes and well-drained, fertile soils. Because of its high commercial value, Martaban camphor was cultivated in the Phu Tho province (Kha 2004). The wood is used for construction purposes such as household utensils, flooring strips, wood carvings, implements, and boats. The bark and leaves are used to treat liver aches, dyspepsia, backache, impotence, and amenorrhea and to improve blood circulation (Sein and Mitlöhner 2011). Martaban camphor populations are fragmented and overexploited due to human disturbance. Moreover, its habitats are also degraded. This species is listed as least concern under the IUCN Red List of Threatened Species (De Kok 2020) and as endangered (CR A1a,c,d) in the Vietnam Red Data Book (MOST and VAST 2007).

The adaptive and evolutionary potential of a species in nature is affected by its genetic diversity and genetic differentiation. Small populations consistently face an increased risk of inbreeding depression and decreased genetic variability, which can lead to increased vulnerability to environmental stochasticity (Barrett and Kohn 1991). Martaban camphor populations often suffer from fragmented habitats and overexploitation. Such populations become inbred and easy to genetically drift, which then enhances the extinction risk (Pauls *et al.* 2013). Therefore, it is important to acknowledge that understanding the genetic variability of Martaban camphor is essential to determining evolutionary characteristics, which can lead to establishing a species conservation program. In this respect, Sandigawad and Patil (2011) used the 11 RAPDs (random amplified polymorphic DNA) to analyze genetic relationships between 15 accessions of *C. zeylanicum* in the Western Ghats of South India. Gwari *et al.* (2016) used RAPDs and ISSRs (inter simple sequence repeats) to investigate the genetic diversity of *C. tamala* in Uttarakhand Himalaya and showed two main groups. Li *et al.* (2018) developed 21 genic SSRs from *C. camphora* and detected medium genetic diversity in this species. Zhong *et al.* (2019) used 22 EST-SSRs (expressed sequence tag-simple sequence repeats) to investigate *C. camphora* in China and also showed medium genetic diversity. Zhang *et al.* (2021) showed medium genetic diversity with an expected heterozygosity of 0.34 in the threatened *C. chago* in China using 4 of DNA markers (psbA-trnH, RPB2, CHS, and LFY) and 11 EST-SSRs. Cui *et al.* (2022) developed 15 EST-SSRs from *C. balansae* in Vietnam, and detected low genetic diversity (He = 0.262) in this species. To study genetic diversity, microsatellites have been widely chosen because of their co-dominance and polymorphism. Up to now, there has been little information on the genetic diversity of *C. parthenoxylon*. In the present study, we developed a set of EST-SSRs from this species to investigate the genetic diversity and population structure of *C. parthenoxylon* in Vietnam and determine the factors contributing to the reduction in genetic diversity in the species. The results will provide information on genetic diversity for its conservation and breeding programs.

Materials and methods

Plant material and DNA extraction: We carried out a sample collection of eight populations of *Cinnamomum parthenoxylon* (Jack) Meisn from different areas in Vietnam (Table 1 Suppl., Fig. 1). Three populations in the provinces of Quang Ninh, Vinh Phuc, and Bac Giang are situated in the Northeast region; three populations in the provinces of Son La, Hoa Binh, and Ninh Binh in the Northwest region; and two populations in the provinces of Thanh Hoa and Nghe An are situated in the Central region. The populations are found in secondary forests. Seedlings prefer slight shading, and adults are light-demanding (JICA 1996). The genomic DNA was extracted from young leaves using a Plant Genomic DNA kit (*Tiangen Biotech*, Beijing, China) according to the manufacturer's procedures. The DNA quality was checked by 1% agarose gel electrophoresis and *NanoDrop 2000C* (*Thermo Scientific*, Wilmington, DE, USA) and then diluted to a concentration of 30 ng μL^{-1} for PCR amplification.

Development of microsatellite markers: To develop a set of microsatellites for *C. parthenoxylon*, we collected three plant tissues (leaves, flowers, and fruits) from one *C. parthenoxylon* individual in Tam Dao. Total RNA from each sample was isolated using the *OmniPlant* RNA kit (DNase I) following the manufacturer's instructions. Both quantity and quality of total RNA were checked through a *Nanodrop ND-2000* spectrophotometer (*Thermo Electron Corporation*, Waltham, Massachusetts, USA). Equal amount of total RNA from each sample was pooled together and sent to *Bioeditas Technology Corporation* (Shaanxi, China) for cDNA library construction and transcriptome sequencing using *Illumina HiSeq™4000*. The program *MIcroSAtellite* (*MISA*) was used to detect and locate SSRs in nucleotide sequences (Beier *et al.* 2017). Only unigenes that were longer than 1 kb were included in the EST-SSR detection. The parameters were a set for the detection of perfect mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide motifs with a minimum of 12, 6, 5, 5, 5, and 5 repeats, respectively. SSR primers were designed using *Primer v7.0* software (Clarke and Gorley 2015). The major parameters for primer design were as follows: primer length (18 - 24 bp), with an optimum of 20 bp, PCR product sizes of 100 - 300 bp, annealing temperature between 55

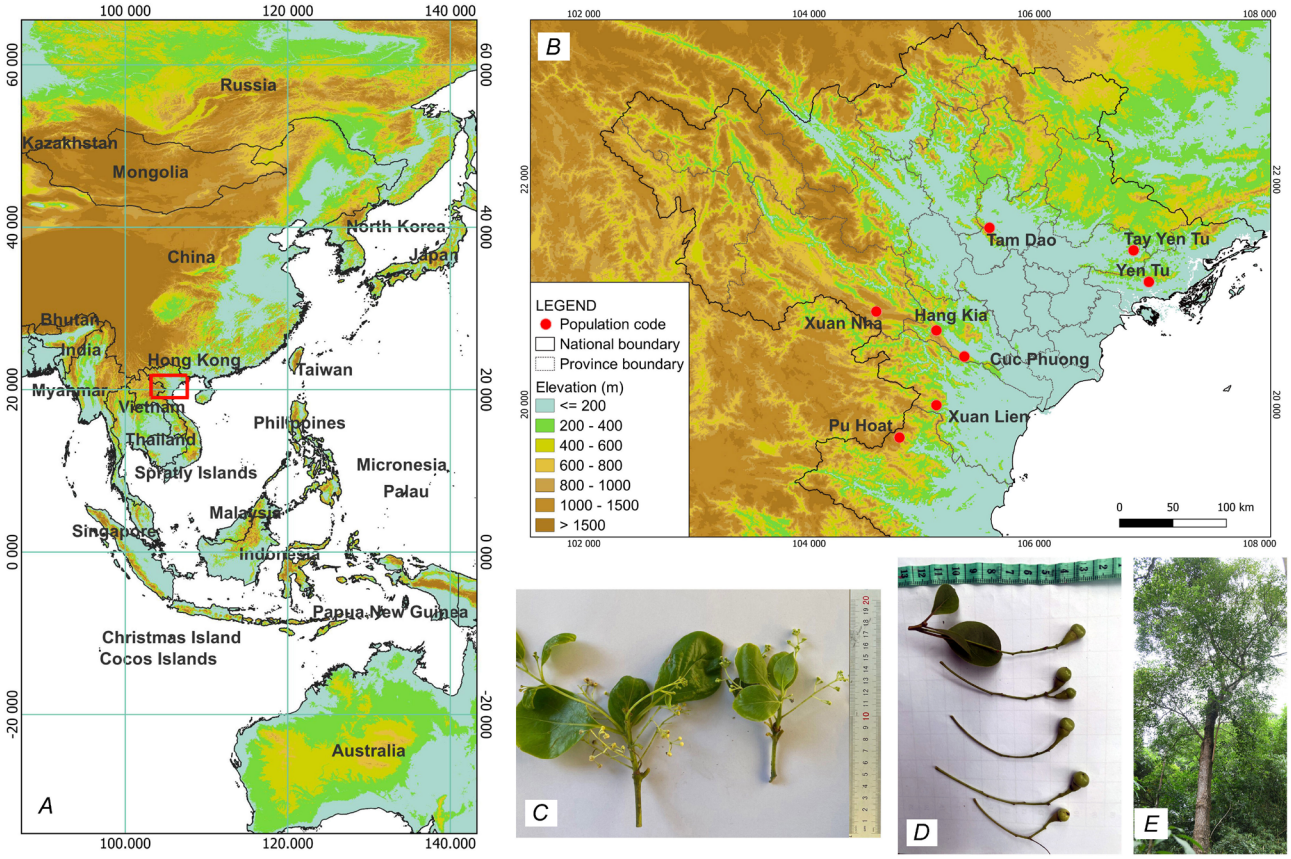

Fig. 1. Map of field survey locations and geographic distributions of *C. parthenoxylon* in Vietnam. Map showing the collection sites (*A*,*B*), leaves, flowers, and fruits (*C*,*D*), and adult plant (*E*).

and 65°C with an optimum of 60°C, and a GC content of 40 - 65%, with 50% being optimal. The SSR primer pairs were synthesized by *Breeding Biotechnologies Co.* We selected the set of ten polymorphic markers to analyze the genetic diversity and population structure of *C. parthenoxylon*. The nucleotide sequences of ten microsatellite markers were deposited in *GenBank* (Table 2 Suppl.).

Microsatellite amplification: The PCR reaction was performed in a 25 µl volume comprising 2.5 µl of template DNA (20 ng total), 12.5 µl of 2× PCR Master Mix containing Taq DNA polymerase (0.05 U μL^{-1}), reaction buffer, 4 mM $MgCl_2$, and 0.4 mM of each dNTP (*Thermo Fisher Scientific*, Waltham, Massachusetts, USA), 1 µl of each primer (10 pmol μl^{-1}) (*Macrogen*, Seoul, Korea), and 8 µl of deionized water. PCR amplification was performed in the *GeneAmp PCR System 9700* (*Applied Biosystems*, Foster City, CA, USA) as follows: initial denaturation at 94°C for 3 min; 40 cycles of 1 min at 94°C, 30 s at a 55 - 56°C annealing temperature for primer pairs, 1 min of extension at 72°C and 10 min at 72°C for the final cycle; and holding at 4°C. The PCR products were sized and subjected to relative quantification between samples on a *5300 Fragment Analyzer* (*Agilent*, Santa Clara, California, USA) with an *Agilent DNF-905 dsDNA* kit (1 - 500 bp).

Data analysis: The genetic diversities, including the number of alleles (A), effective alleles (E), observed (H_o), expected (H_e) heterozygosity, and coefficient of inbreeding (F_{is}) were determined at locus and population levels using *GeneAlEx v.6.5* (Peakall and Smouse 2012). Gene flow (N_m), genetic differentiation among populations [the *F*-statistics F_{st} index of Weir and Cockerham (1984) and G_{st} index (Hedrick 2005)], and total expected heterozygosity (H_t) were also determined in *GenAlEx v.6.5*. *Fstat v.2.9.3* (Goudet 2001) was used to calculate allelic richness (R). Null alleles were checked in *Micro-Checker v.2.0* (van Oosterhout *et al.* 2004) using 1 000 bootstrap iterations. The F_{is} values for null allele frequency (F_{is}IIM) were calculated using *INEst* (Chybicki and Burczyk 2009), based on the individual inbreeding model. Deviation from Hardy-Weinberg equilibrium (HWE) for loci within populations was tested in *Cervus v.3.0* (Kalinowski *et al.* 2007) based on 1 000 permutations of alleles among individuals. The bottleneck event was detected using *Bottleneck v.1.2* (Piry *et al.* 1999), based on the evaluation *via* the one-tailed Wilcoxon signed-rank test. Analysis of molecular variance (*AMOVA*) was calculated in *Arlequin v.3.5* (Excoffier and Lischer 2010). A neighbor-joining (NJ) tree was implemented based on the F_{st} values using *Poptree2* (Takezaki *et al.* 2010). Analysis of population structure was performed using *Structure v.2.3.4*, and the Bayesian

clustering method was used (Pritchard *et al.* 2000). Establishing the admixture model with correlated allele frequencies required ten distinct runs for each number of groups in the dataset (K). These runs were carried out with 500 000 Markov chain Monte Carlo (MCMC) iterations and a burn-in time of 100 000 iterations. K ranged from 1 to 10. *Structure Harvester* (Earl and vonHoldt 2012) was used to detect the number of groups that best fit the dataset based on the K that was determined by Evanno *et al*. (2005) to determine the ideal value of K. This was done so that the optimal value of K could be determined. When the optimal K value had been determined, the duplicated findings were aligned with *Clumpp v.1.1.2* (Jakobsson and Rosenberg 2007), and the allocated cluster membership bar plots were created with *Distruct v.1.1.* (Rosenberg 2004). Discriminant analysis of principal components (DAPC) analysis (Jombart *et al.* 2010) was done using the *adegenet* package (Jombart 2008) for the software *R* (R Development Core Team).

Results

Genetic diversity: We determined 40 different alleles from the ten loci for 192 adult trees in the eight *Cinnamomum parthenoxylon* populations. The ten studied loci were polymorphic. Null alleles were detected in all populations (Table 3 Suppl.). Several loci in some populations show evidence for a null allele, as *xaxi2* and *xaxi8* in Xuan Lien, *xaxi1* and *2* in Pu Hoat, *xaxi2* and *4* in Hang Kia, *xaxi3* and *5* in Xuan Nha, and *xaxi3* in Cuc Phuong (Table 3 Suppl.). Linkage genotypic disequilibrium was examined for *C. parthenoxylon* and showed that 67 out of 387 tests were significant at the 5% level. The highest number of allelic richness (R = 4.1) per locus was observed at locus *xaxi2*, and the lowest (R = 2.0) was observed at loci *xaxi9* and *xaxi7*, with an average of 3.1 (Table 1). However, the number of effective alleles was the lowest (E = 1.4) at *xaxi7* and the highest (E = 2.5) at *xaxi6*, for an average of 1.9 effective alleles. The private alleles (N_p) were also located at the five loci. Three private alleles (one at *xaxi8* and two at *xaxi6*) were found in Hang Kia. Two private alleles were found in Tay Yen Tu (each one at *xaxi1* and *5*) and Pu Hoat (each one at *xaxi 1* and *6*). One private allele, one at *xaxi1*, and one at *xaxi4* was observed in Tam Dao and Xuan Lien, respectively. Similarly, the lowest observed and expected heterozygosity were detected at *xaxi1* and *xaxi7*, while the highest observed and expected heterozygosity was found at *xaxi6*. The total expected heterozygosity (H_t) averaged 0.554. The index of inbreeding coefficient (F_{is}) for each locus averaged 0.060. Positive F_{is} values were found at five loci and showed homozygosity excess at these loci under the Hardy-Weinberg equilibrium. The coefficient of total inbreeding (F_{it}) averaged 0.265 for each locus and indicated homozygosity excess in *C. parthenoxylon.* The tests of Hardy-Weinberg equilibrium showed no significant deviation at two loci (xaxi6 and xaxi9) and a significant deviation at the remaining eight loci, with Bonferroni correction.

Table 2 shows information on genetic diversity at the population level. The lowest number of different alleles (A) were found in the Tam Dao, while the highest was in the Hang Kia population. The average number was 2.9 alleles per population. The mean allelic richness (R) was 2.7, whereas the mean of effective alleles (E) was 1.9. The highest number of private alleles ($N_p = 3$) was observed in the Hang Kia populations. The two populations of Tay Yen Tu and Pu Hoat had two private alleles, and the two populations of Xuan Lien and Tam Dao had one private allele. The lowest levels of observed heterozygosity (H_o = 0.341) and expected heterozygosity (H_e = 0.338) were observed in Tam Dao. The observed heterozygosity and the expected heterozygosity averaged 0.399 and 0.426, respectively. The high inbreeding coefficient index was detected in Hang Kia (F_{is} = 0.203), which points to a high inbreeding in this population. Based on the individual inbreeding model, the inbreeding coefficient corrected for null alleles (F_{is}IIM) averaged 0.021, ranging from 0.007 in Hang Kia to 0.09 in Xuan Lien. The heterozygosity excess was significant ($P < 0.05$) under the two-phase mutation in these populations under Wilcoxon sign-rank tests.

Genetic structure: Genetic differentiation at the locus level overall *C. parthenoxylon* populations was medium (F_{st} = 0.223) (Table 1). Genetic differentiation for population pairs was presented in Table 3. Significant genetic differentiation (F_{st}) was detected between each of the populations except Hang Kia/Cuc Phuong and Xuan Nha/Cuc Phuong pairs ($P > 0.05$). The *AMOVA* analysis showed that most of the genetic variation was detected within individuals (68.89%) (Table 4). Based on the F_{st} values matrix, the NJ analysis showed the three main groups (Fig. 2). The first group included the three populations of Yen Tu, Tay Yen Tu, and Tam Dao in the Northeast area, with a bootstrap of 99%. The second group included the three populations of Hang Kia, Cuc Phuong, and Xuan Nha in the Northwest area with a bootstrap of 78%, and the third group included two populations of Xuan Lien and Pu Hoat in the Central area with a bootstrap of 98%.

The Structure analysis determined two peaks of deltaK at K = 2 and 3 (Fig. 3*A*) and indicated that the optimal number of genetic groups was two, with the highest deltaK (739.23). The percentage of ancestry of individuals and populations in each group was presented in each color (Fig. 3*B*). At K = 2, all eight populations were divided into two main groups. The first group included most individuals from the three Northeast populations of Tam Dao, Yen Tu, and Tay Yen Tu, with strong ancestry values of more than 90% (Table 4 Suppl.). The second group included the remaining five populations with strong ancestry values (>90%). At K = 3, two Central populations of Pu Hoat and Xuan Lien were separated from the second group at K = 2 to form a distinct group. Therefore, the first group included most individuals from the three Northeast populations of Yen Tu, Tam Dao, and Tay Yen Tu with strong ancestry (91 - 95%). The second group included most individuals from the three Northwest populations of Hang Kia, Xuan Nha, and Cuc Phuong, with an ancestry range of 88.1 - 90.3%, and the third group included two Central

Table 1. Genetic parameters of the ten SSR loci for *Cinnamomum parthenoxylon*. A - number of alleles, R - allelic richness, E - effective alleles, N_P - private alleles, H_o and H_e - observed and expected heterozygosity, H_t - total expected heterozygosity, F_{is} - coefficient of inbreeding, F_{it} - coefficient of total inbreeding, F_{st} - genetic differentiation index of Weir and Cockerham (1984), SE - standard error, ns - no significance with Bonferroni correction, * - $P < 0.05$, *** - $P < 0.001$.

Loci	A	R	E	Null allele	N_P	H_o	H_e	H_t	F_{is} (SE)	F_{it} (SE)	F_{st} (SE)	P_{HWE}
xaxi1	5	3.6	1.7	0.109	1	0.371	0.364	0.561	–0.017	0.339	0.350	***
xaxi2	6	4.1	2.3	no	2	0.454	0.539	0.664	0.156	0.316	0.189	***
xaxi3	4	3.9	1.9	0.105	0	0.214	0.446	0.618	0.520	0.653	0.277	***
xaxi4	4	3.1	2.0	0.071	1	0.436	0.482	0.657	0.095	0.336	0.267	***
xaxi5	4	3.0	1.5	0.119	1	0.229	0.287	0.306	0.203	0.253	0.062	***
xaxi6	6	3.4	2.5	no	3	0.602	0.596	0.664	–0.010	0.094	0.103	ns
xaxi7	2	2.0	1.4	no	0	0.285	0.259	0.493	–0.099	0.421	0.474	***
xaxi8	4	3.1	1.7	0.127	1	0.319	0.399	0.616	0.200	0.482	0.353	***
xaxi9	2	2.0	1.9	no	0	0.544	0.474	0.486	–0.147	–0.119	0.024	ns
xaxi10	3	3.0	1.7	no	0	0.530	0.408	0.469	–0.300	–0.131	0.130	*
Mean	4	3.1	1.9			0.399 (0.023)	0.426 (0.018)	0.554 (0.037)	0.060 (0.072)	0.265 (0.080)	0.223 (0.046)	

Table 2. Genetic diversity values and results of bottleneck tests for eight *C. parthenoxylon* populations. N - sample size, A - alleles per locus, E - effective alleles, R - allelic richness, N_P - number of private alleles, H_o and H_e - observed and expected heterozygosity, F_{is} - coefficient of inbreeding, F_{is}IIM - corrected inbreeding coefficient for null alleles, SMM - stepwise mutation model, TPM - two-phase model, SE - standard error; * - $P < 0.05$, ** - $P < 0.01$, *** - $P < 0.001$.

Populations	N	A	E	R	N_p	H_o (SE)	H_e (SE)	F_{is} (SE)	F_{is}IIM	*P* value of bottleneck	
										SMM	TPM
Hang Kia	35	3.4	2.1	3.2	3	0.406	0.500	0.203***	0.007	0.539	0.065
Yen Tu	21	2.6	1.7	2.5	0	0.371	0.344	–0.057	0.013	0.715	0.500
Xuan Nha	22	2.9	2.1	2.8	0	0.423	0.508	0.190**	0.013	0.065	0.005
Cuc Phuong	19	2.8	2.1	2.8	0	0.432	0.482	0.131*	0.013	0.216	0.016
Xuan Lien	24	2.7	1.7	2.7	1	0.358	0.385	0.091	0.090	0.615	0.278
Tam Dao	27	2.4	1.6	2.3	1	0.341	0.338	0.010	0.009	0.500	0.213
Tay Yen Tu	20	2.8	1.8	2.8	2	0.445	0.408	–0.065	0.014	0.539	0.385
Pu Hoat	24	2.7	1.9	2.7	2	0.413	0.439	0.082	0.010	0.216	0.053
Mean		2.9	1.9	2.7		0.399 (0.023)	0.426 (0.018)	0.073	0.021		

Table 3. Genetic differentiation between the eight populations of *C. parthenoxylon*. ns - not significant, * - $P < 0.05$, ** - $P < 0.01$, *** - $P < 0.001$ based on 999 permutations.

	Hang Kia	Yen Tu	Xuan Nha	Cuc Phuong	Xuan Lien	Tam Dao	Tay Yen Tu	Pu Hoat
Hang Kia		***	*	ns	***	***	***	***
Yen Tu	0.211		***	***	***	***	**	***
Xuan Nha	0.022	0.202		ns	***	***	***	***
Cuc Phuong	0.017	0.202	0.019		***	***	***	***
Xuan Lien	0.150	0.235	0.134	0.133		***	***	***
Tam Dao	0.202	0.085	0.177	0.193	0.269		***	***
Tay Yen Tu	0.169	0.027	0.154	0.160	0.190	0.039		***
Pu Hoat	0.112	0.230	0.101	0.105	0.036	0.257	0.191	

populations of Pu Hoat and Xuan Lien, with an ancestry value of 96.3 and 94.1%, respectively.

The DAPC analysis with prior information indicated relationships among populations (Fig. 4*A*). A high degree of overlap among populations was found in some geographical distribution areas in relation to low genetic differentiation among these populations. The highest overlap was found for the Central population pair of Xuan Lien and Pu Hoat (F_{st} = 0.036) or for the Northeast population pair of Tay Yen Tu and Yen Tu (F_{st} = 0.027).

Table 4. Analysis of molecular variance for eight *C. parthenoxylon* populations. df - degree of freedom; F_{is} - inbreeding coefficient; F_{st} - genetic differentiation; F_{it} - total expected heterozygosity, *** - $P < 0.001$, significance tests with 1023 permutations.

	df	Sum of squares	Variance components	Total variation [%]	Fixation indices
Among populations	7	246.709	0.688	23.92	$F_{is} = 0.094$*** $F_{st} = 0.239$*** $F_{it} = 0.311$***
Among individuals within populations	184	440.747	0.207	7.19	
Within individuals	192	380.500	1.082	68.89	
Total	383	1 067.956	2.877		

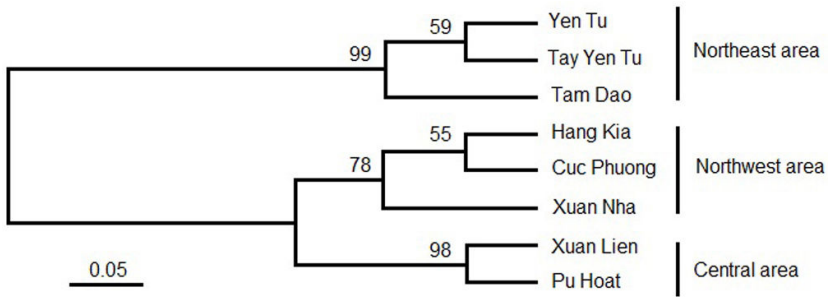

Fig. 2. The Neighbor-joining (NJ) tree of eight populations of *C. parthenoxylon* based on the F_{st} values.

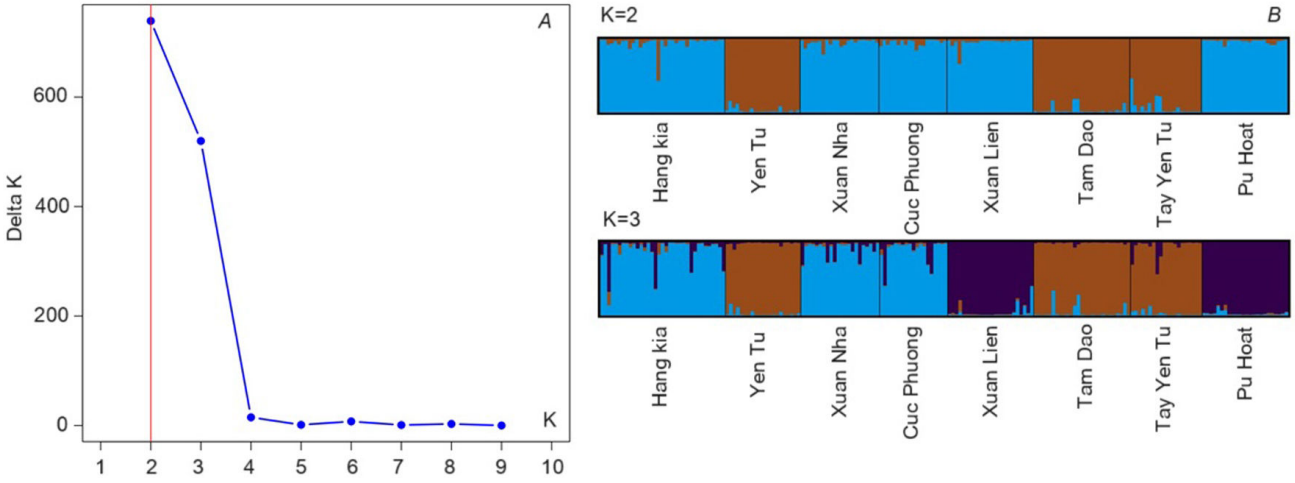

Fig. 3. Distribution of DeltaK over K = 1 - 10 (*A*) and barplot of admixture assignment for 192 trees of eight *C. parthenoxylon* populations with K = 2 and 3 (*B*).

No overlap was observed between different areas, showing high genetic differentiation in these population pairs. The DAPC analysis, without prior information on population origin, also presented the three genetic groups (Fig. 4*B*). All individuals from the three Northeast populations of Yen Tu, Tam Dao, and Tay Yen Tu were assigned to cluster 1 (Table 5 Suppl., Fig. 1 Suppl.). Cluster 2 included the most individuals from the Northwest populations of Hang Kia, Xuan Nha, and Cuc Phuong. One individual from Xuan Lien was also assigned to this cluster. All individuals from Pu Hoat and most individuals from Xuan Lien, together with some individuals from Hang Kia, Xuan Nha, and Cuc Phuong, were included in cluster 3. No individuals from the three Northeast populations were included in cluster 3.

Discussion

Genetic variation: For a threatened species, genetic diversity is essential to maintaining its evolutionary potential and adapting to its environment (Dick *et al.* 2008) and it is related to geographic distribution range, longevity, population size, and reproductive system (Nybom 2004, White *et al.* 2007). In this study, the genetic diversity and genetic structure of the threatened *C. parthenoxylon*

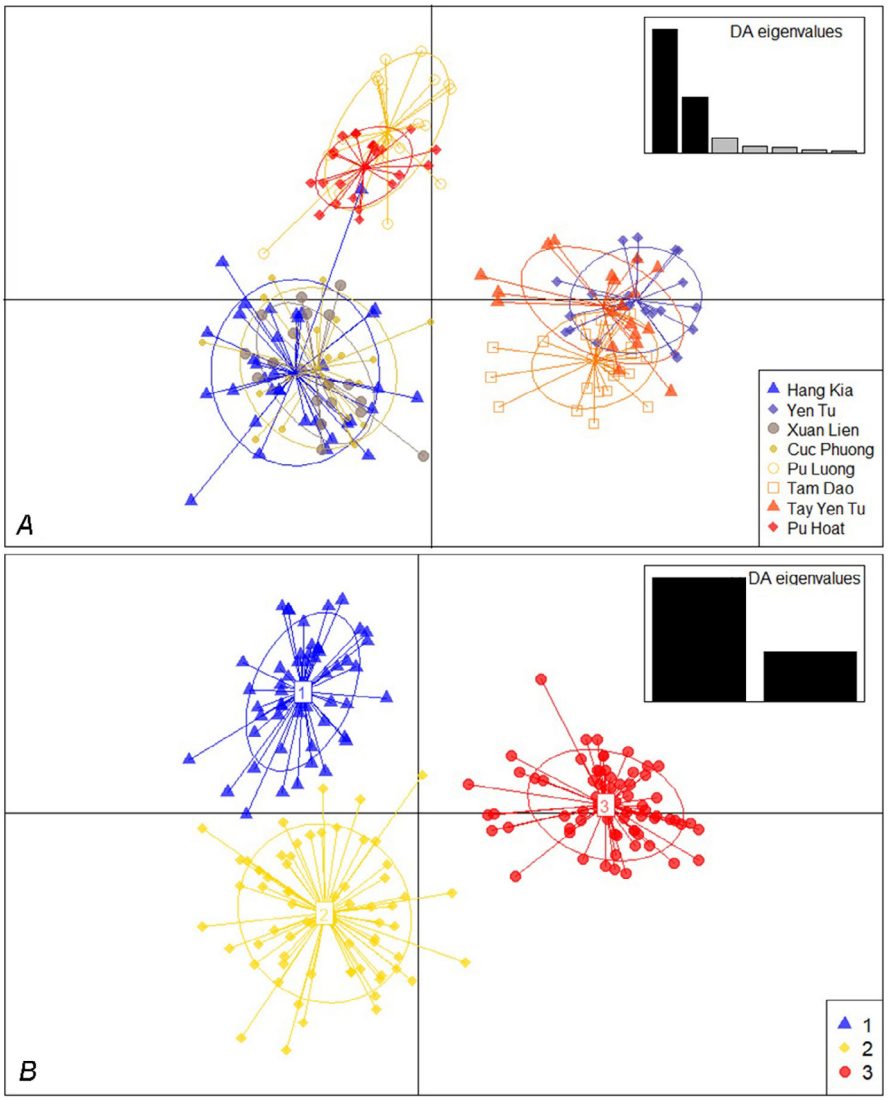

Fig. 4. Scatter plot of DAPC with prior information (*A*) and without prior information (*B*).

species were investigated to provide genetic information for conservation plans for this species. The species maintained a moderate genetic diversity (R = 2.7, H_o = 0.399, H_e = 0.426). Our findings were low compared to some *Lauraceae* species with widespread distribution. In previous studies, Martins *et al*. (2015) reported high genetic diversity of three species in the Brazil rainforests using nuclear microsatellites, such as *Ocotea catharinensis* (H_e = 0.73), *O. odorifera* (H_e = 0.78), and *O. porosa* (H_e = 0.64). High genetic diversity was also observed for *Lindera benzoin* (Mooney *et al*. 2010), *C. camphora* (Kameyama *et al*. 2017), and *Machilus thunbergia* in Japan (Watanabe *et al*. 2017) using microsatellites. However, the genetic diversity results of our study were similar to those reported for some threatened species, such as *C. camphora* in China based on EST-SSRs (H_e = 0.41; Li *et al*. 2018), *C. chago* in China using 11 EST-SSRs (H_e = 0.34; Zhang *et al*. 2021), and *C. chago* in China based on ISSRs (H_e = 0.46; Dong *et al*. 2016). Genetic diversity can be reduced through genetic drift and increased homozygotes for common alleles due to the loss of rare alleles (Honnay and Jacquemyn 2007, Gijbels *et al*. 2015). The low number of alleles is often related to a population size reduction and then decreased heterozygosity within populations (Young *et al*. 1996, Lowe *et al*. 2005, Leimu *et al*. 2006). Among the eight populations, three populations Yen Tu, Tam Dao, and Xuan Lien had lower genetic diversity than others. The low genetic diversity in these populations may be related to anthropogenic disturbance. The destroyed forests and overexploitation in the 1980s and 1990s (MOST and VAST 2007) can be the main factors decreasing genetic diversity within all populations. To prove the direct connection between these negative anthropogenic interventions and decreased genetic diversity would require a different set-up study. Anthropogenic activities can cause a decrease in effective population size and an increase in the geographic distance among populations. The decrease in genetic diversity

may be a consequence of genetic drift, inbreeding, and bottlenecks. The bottleneck events were found in two populations, Xuan Nha and Cuc Phuong. Additionally, the limited gene exchange among different areas (N_m = 0.871) also contributes to a reduction in the genetic diversity of C. *parthenoxylon.*

Genetic differentiation: Genetic variance and genetic structure of species are affected by the range of natural distribution, gene flow, population size, evolution history, and reproductive systems (Hamrick and Godt 1990). Species with widespread distribution, large longevity, and outcrossing maintain low genetic differences and high gene flow between populations (Finkeldey and Hattemer 2007, Kettle *et al*. 2011, Zhao *et al*. 2012). Genetic difference is strongly influenced by gene flow and genetic drift (Slatkin and Barton 1989, Momose *et al*. 1994, de Morais *et al*. 2015). In the present study, the genetic differentiation among populations was 0.223 and is consistent with the result of the *AMOVA* analysis (23.92% of genetic variation among populations). This finding was consistent with previous studies, such as *C. camphora* in South China using EST-SSRs (F_{st} = 0.22; Zhong *et al*. 2019), *C. chago* in Yunnan, China using ISSRs (G_{st} = 0.217; Dong *et al*. 2016), and EST-SSRs (F_{st} = 0.2198; Zhang *et al*. 2021). Gene flow contributes to genetic differentiation and population structure and is determined by the dispersal of pollen grains and seeds (Finkeldey and Hattemer 2007). As a woody species, *C. parthenoxylon* has large longevity and is predominately outcrossing. Its pollen grains can be dispersed by the wind and insects (bees). Due to the fact that dispersal depends on insects, an increase in geographic distance will decrease the transmission of pollen grains and seeds between populations. Therefore, habitat fragmentation can lead to increasingly isolated populations and then decrease gene exchange among the populations. This can affect the genetic structure of *C. parthenoxylon.* Our study showed low genetic difference in the same areas, such as Cuc Phuong and Hang Kia (F_{st} = 0.017), Xuan Nha and Hang Kia (F_{st} = 0.022) in the Northwest area, Yen Tu and Tay Yen Tu (F_{st} = 0.027), and Tam Dao and Tay Yen Tu (F_{st} = 0.039) in the Northeast area. High genetic differences were recorded between populations located in different areas, such as F_{st} = 0.257 between Tam Dao and Pu Hoat, F_{st} = 0.269 between Tam Dao and Xuan Lien, all of which are separated by large geographic distances. Moreover, the dispersal of seeds through the wind may be restricted due to their heavy mass. In our study, the genetic structure of *C. parthenoxylon* was detected by the NJ tree, Structure analysis, and DAPC in which the 192 studied individuals of the eight populations were clustered based on their geographic origin. The genetic structure could be a consequence of the gene exchange and lead to the formation of different groups. Large geographic distances may increase population isolation and then restrict gene migration between populations.

Conclusions and conservation implications: Assessing genetic diversity and genetic structure is important to constructing a suitable plan for the preservation of threatened species (Forest *et al*. 2007). In this study, three different gene pools with a medium level of genetic diversity were detected for *C. parhenoxylon*, which could contribute to the conservation of this threatened species. Lower genetic diversity was found in the two populations, Yen Tu and Tam Dao, than in the remaining six populations, suggesting a consequence of anthropogenic activities in these two populations. Low genetic differentiation among populations in the same areas was recorded for population pairs located in the same area compared with the genetic differentiation found between the populations located in different areas. A higher allelic richness was found in the six populations, Hang Kia, Xuan Nha, Cuc Phuong, Tam Dao, Tay Yen Tu, and Pu Hoat than in the remaining two populations. Private alleles were observed in five populations, Hang Kia, Tay Yen Tu, Pu Hoat, Xuan Lien, and Tam Dao. Thus, some populations such as Hang Kia, Xuan Nha, Cuc Phuong, Tay Yen Tu, and Pu Hoat have high genetic diversity or private alleles and might be prioritized for *in situ* conservation. Simultaneously, seeds from remaining populations should be collected for *ex-situ* conservation activities. An increase in population sizes can prevent a reduction in genetic diversity *via* genetic drift, homozygosity for common alleles, and limited gene migration across populations.

References

Barrett S.C.H., Kohn J.R.: Genetic and evolutionary consequences of small population size in plants: Implications for conservation. - In: Falk D.A., Holsinger K.E. (ed.): Genetics and Conservation of Rare Plants. Pp. 3-30. Oxford Academic, New York 1991.

Beier S., Thiel T., Münch T. *et al.*: MISA-web: a web server for microsatellite prediction. Bioinformatics **33**: 2583-2585, 2017.

Chybicki I.J., Burczyk J.: Simultaneous estimation of null alleles and inbreeding coefficient. - J. Hered. **100**: 106-113, 2009.

Clarke K.R., Gorley R.N.: PRIMER v7: User Manual/Tutorial. Pp. 296. PRIMER-E, Plymouth 2015. *Available at*: https://learninghub.primer-e.com/books/primer-v7-user-manual-tutorial/page/download.

Cui B., Vu D.D., Vu D.G. *et al.*: Genetic diversity and population structure of *Cinnamomum balansae* Lecomte inferred by microsatellites. - Open Life Sci. **17**: 323-332, 2022.

De Kok R.: *Cinnamomum parthenoxylon.* The IUCN Red List of Threatened Species 2020, e.T33198A2834736, 2020.

de Morais C.T., Ghazoul J., Maycock C.R. *et al.*: Understanding local patterns of genetic diversity in dipterocarps using a multi-site, multi-species approach: implications for forest management and restoration. - Forest Ecol. Manag. **356**: 153-165, 2015.

Dick C.W., Hardy O.J., Jones F.A., Petit R.J.: Spatial scales of pollen and seed-mediated gene flow in tropical rain forest trees. - Trop. Plant Biol. **1**: 20-33, 2008.

Dong W., Zhang X., Guansong Y. *et al.*: Biological characteristics and conservation genetics of the narrowly distributed rare plant *Cinnamomum chago* (Lauraceae). - Plant Divers. **38**: 247-252, 2016.

Earl D.A., vonHoldt B.M.: Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. - Conserv. Genet. Resour.

4: 359-361, 2012.
Evanno G., Regnaut S., Goudet J.: Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. - Mol. Ecol. **14**: 2611-2620, 2005.
Excoffier L., Lischer H.E.L.: Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. - Mol. Ecol. Resour. **10**: 564-567, 2010.
Finkeldey R., Hattemer H.H.: Tropical Forest Genetics. Pp. 316. Springer, Berlin-Heidelberg 2007.
Forest F., Grenyer R., Rouget M. *et al.*: Preserving the evolutionary potential of floras in biodiversity hotspots. - Nature **445**: 757-760, 2007.
Gijbels P., De Hert K., Jacquemyn H., Honnay O.: Reduced fecundity and genetic diversity in small populations of rewarding versus deceptive orchid species: a meta-analysis. - Plant Ecol. Evol. **148**: 153-159, 2015.
Goudet J.: FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3), 2001. *Available at*: https://www2.unil.ch/popgen/softwares/fstat.htm.
Gwari G., Bhandari U., Naik G. *et al.*: Genetic diversity in *Cinnamomum tamala* Nees. accessions through DNA fingerprinting using molecular markers. - Indian J. Agr. Res. **50**: 446-450, 2016.
Hamrick J.L., Godt M.J.W.: Allozymes diversity in plant species. - In: Brown A.H.D., Clegg M.T., Kahler A.L., Weir B.S. (ed.): Plant Population Genetics, Breeding and Genetic Resources. Pp. 43-63. Sinauer Associates, Sunderland 1990.
Hedrick P.W.: A standardized genetic differentiation measure. - Evolution **59**: 1633-1638, 2005.
Honnay O., Jacquemyn H.: Susceptibility of common and rare plant species to the genetic consequence of habitat fragmentation. - Conserv. Biol. **21**: 823-831, 2007.
Jakobsson M., Rosenberg N.A.: CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. - Bioinformatics **23**: 1801-1806, 2007.
JICA: [Vietnam Forest Trees.] Pp. 788. Agricultural Publishing House, Hanoi 1996. [In Vietnamese]
Jombart T.: *adegenet*: a R package for the multivariate analysis of genetic markers. - Bioinformatics **24**: 1403-1405, 2008.
Jombart T., Devillard S., Balloux F.: Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. - BMC Genet. **11**: 94, 2010.
Kalinowski S.T., Taper M.L., Marshall T.C.: Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. - Mol. Ecol. **16**: 1099-1106, 2007.
Kameyama Y., Furumichi J., Li J., Tseng Y.-H.: Natural genetic differentiation and human-mediated gene flow: the spatiotemporal tendency observed in a long-lived *Cinnamomum camphora* (Lauraceae) tree. - Tree Genet. Genom. **13**: 38, 2017.
Kettle C.J., Hollingsworth P.M., Burslem D.F.R.P. *et al.*: Determinants of fine-scale spatial genetic structure in three co-occurring rain forest canopy trees in Borneo. - Perspect. Plant Ecol. Evol. Syst. **13**: 47-56, 2011.
Kha L.D.: [Investigation of Indigenous Seedling Production Capacity in Some Nurseries in the North of Vietnam and Recommendations of Seedling Production Technical Measures.] Pp. 100. Ministry of Agriculture and Rural Development, Hanoi 2004. [In Vietnamese]
Leimu R., Mutikainen P., Koricheva J., Fischer M.: How general are positive relationships between plant population size, fitness and genetic variation? - J. Ecol. **94**: 942-952, 2006.
Li Y.-R., Fu C.-S., Yang W.-J. *et al.*: Investigation of constituents from *Cinnamomum camphora* (L.) J. Presl and evaluation of their anti-inflammatory properties in lipopolysaccharide stimulated RAW 264.7 macrophages. - J. Ethnopharmacol. **221**: 37-47, 2018.
Lowe A.J., Boshier D., Ward M. *et al.*: Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. - Heredity **95**: 255-273, 2005.
Martins E.M., Lamont R.W., Martinelli G. *et al.*: Genetic diversity and population genetic structure in three threatened *Ocotea* species (Lauraceae) from Brazil's Atlantic rainforest and implications for their conservation. - Conserv. Genet. **16**: 1-14, 2015.
Momose K., Nagamitsu T., Inoue T.: Reproductive ecology of an emergent tree, *Dryobalanops lanceolata*, Dipterocarpaceae, in a non-general flowering period in Sarawak. - In: Inoue T., Hamid A.A. (ed.): Plant Reproductive Systems and Animal Seasonal Dynamics: Long Term Study of Dipterocarp Forests in Sarawak. Canopy Biology Programme in Sarawak. Series I. Kyoto University, Kyoto 1994.
Mooney E., Edwards M., Niesenbaum R.: Genetic differentiation between sun and shade habitats in populations of *Lindera benzoin* L. - Popul. Ecol. **52**: 417-425, 2010.
MOST, VAST: [Vietnam Red Data Book. Part II. Plants.] Pp. 289-290. Natural Science and Technology Publishing, Hanoi 2007. [In Vietnamese]
Nybom H.: Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. - Mol. Ecol. **13**: 1143-1155, 2004.
Pauls S.U., Nowak C., Bálint M., Pfenninger M.: The impact of global climate change on genetic diversity within populations and species. - Mol. Ecol. **22**: 925-946, 2013.
Peakall R., Smouse P.E.: GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. - Bioinformatics **28**: 2537-2539, 2012.
Piry S., Luikart G., Cornuet J.-M.: Bottleneck: a computer program for detecting recent reductions in the effective size using allele frequency data. - J. Hered. **90**: 502-503, 1999.
Pritchard J.K., Stephens M., Donnelly P.: Inference of population structure using multilocus genotype data. - Genetics **155**: 945-959, 2000.
Rosenberg N.A.: DISTRUCT: a program for the graphical display of population structure. - Mol. Ecol. Notes **4**: 137-138, 2004.
Sandigawad A.M., Patil C.G.: Genetic diversity in *Cinnamomum zeylanicum* Blume. (Lauraceae) using random amplified polymorphic DNA (RAPD) markers. - Afr. J. Biotechnol. **10**: 3682-3688, 2011.
Sein C.C., Mitlöhner R.: *Cinnamomum parthenoxylon* (Jack) Meisn: Ecology and Silviculture in Vietnam. Pp. 18. CIFOR, Bogor 2011.
Slatkin M., Barton N.H.: A comparison of three indirect methods for estimating average levels of gene flow. - Evolution **43**: 1349-1368, 1989.
Takezaki N., Nei M., Tamura K.: POPTREE2: software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. - Mol. Biol. Evol. **27**: 747-752, 2010.
van Oosterhout C., Hutchinson W.F., Wills D.P.M., Shipley P.: MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. - Mol. Ecol. Notes **4**: 535-538, 2004.
Watanabe S., Kaneko Y., Maesako Y., Noma N.: Detecting the early genetic effects of habitat degradation in small size remnant populations of *Machilus thunbergii* Sieb. et Zucc. (Lauraceae). - Int. J. For. Res. **2017**: 9410626, 2017.
Weir B.S., Cockerham C.C.: Estimating *F*-statistics for the

analysis of population structure. - Evolution **38**: 1358-1370, 1984.
White T.L., Adams W.T., Neale D.B.: Forest Genetics. Pp. 682. CABI Publishing, Boston 2007.
Young A., Boyle T., Brown T.: The population genetic consequences of habitat fragmentation for plants. - Trends Ecol. Evol. **11**: 413-418, 1996.
Zhang X., Yang L., Liu Y.-H.: Genetic diversity, genetic structure, and demographic history of *Cinnamomum chago*, a plant species with extremely small populations in China. - Glob. Ecol. Conserv. **31**: e01808, 2021.
Zhao X.F., Ma Y.P., Sun W.B. *et al.*: High genetic diversity and low differentiation of *Michelia coriacea* (Magnoliaceae), a critically endangered endemic in southeast Yunnan, China. - Int. J. Mol. Sci. **13**: 4396-4411, 2012.
Zhong Y., Yang A., Li Z. *et al.*: Genetic diversity and population genetic structure of *Cinnamomum camphora* in South China revealedby EST-SSR markers. - Forests **10**: 1019, 2019.

© The authors. This is an open access article distributed under the terms of the Creative Commons BY-NC-ND Licence.