PLOS ONE

RESEARCH ARTICLE

Genetic diversity and population structure in the endangered tree *Hopea hainanensis* (Dipterocarpaceae) on Hainan Island, China

Chen Wang[1], **Xiang Ma**[1], **Mingxun Ren**[2], **Liang Tang**[2]*

1 Key Laboratory of Tropical Biological Resources of Ministry of Education, School of Life and Pharmaceutical Sciences, Hainan University, Haikou, China, **2** College of Ecology and Environment, Hainan University, Haikou, China

* tangliang@hainanu.edu.cn

Abstract

Hopea hainanensis Merrill & Chun (Dipterocarpaceae) is an endangered tree species restricted to Hainan Island, China and a small part of Northern Vietnam. On Hainan Island, it is an important indicator species for tropical forests. However, because of its highly valued timber, *H. hainanensis* has suffered from overexploitation, leading to a sharp population decline. To facilitate the conservation of this species, genetic diversity and population structure were assessed using 12 SSR markers for 10 populations sampled across Hainan Island. Compared to non-threatened *Hopea* species, *H. hainanensis* exhibited reduced overall genetic diversity and increased population differentiation (AMOVA: $F_{ST} = 0.23$). Bayesian model-based clustering and principal coordinate analysis consistently assigned *H. hainanensis* individuals into three genetic groups, which were found to be widespread and overlapping geographically. A Mantel test found no correlation between genetic and geographical distances ($r = 0.040$, $p = 0.418$). The observed genetic structure suggests that long-distance gene flow occurred among *H. hainanensis* populations prior to habitat fragmentation. A recent population bottleneck was revealed, which may cause rapid loss of genetic diversity and increased differentiation across populations. Based on these findings, appropriate strategies for the long-term conservation of the endangered species *H. hainanensis* are proposed.

Introduction

Earth's biodiversity is rapidly declining as a consequence of agricultural expansion, overexploitation, deforestation, pollution and climate change [1–3]. Approximately 40% of plant species are threatened with extinction [4]. Conservation genetics, a new discipline that applies the concepts and tools of population genetics to biological conservation, is aimed at preserving endangered species from extinction [1]. Endangered species are commonly characterized by small, fragmented populations and restricted gene flow among populations [3]. In small, isolated populations, mating occurs more frequently among relatives, and a shift to selfing may be observed in hermaphroditic plants. Inbreeding leads to homozygosity in detrimental

recessive alleles and the consequent production of inferior offspring, a phenomenon known as inbreeding depression [5]. In addition, genetic drift is stronger in small populations, further contributing to the fixation of deleterious mutations and loss of genetic variation, which compromise the adaptive potential of a population and thereby increase its extinction risk [3, 6]. With the goal of studying the genetic diversity, population differentiation, mating system and historical demography of endangered species, the field of conservation genetics provides remarkable insights into the preservation of biodiversity in the real world [2].

Tree species from a single family, the Dipterocarpaceae, are key community members in Asian tropical forests, accounting for 20–50% of forest basal area and more than 50% of canopy trees [7, 8]. Many Dipterocarpaceae species represent important timber resources, and thus have been heavily exploited in tropical Asian countries. Due to massive logging for timber, as well as deforestation for agriculture, many dipterocarps are now classified as endangered or critically endangered [4, 8]. However, dipterocarp forests are far more than a mere resource for timber production. They are key components of Asian tropical rainforest ecosystems, serving as a foundation on which these highly diverse ecosystems are assembled. Indeed among 25 worldwide 'biodiversity hotspots', four are located in Southeast Asia [9]. Furthermore, dipterocarp forests provide a variety of ecosystem services and play a crucial role in maintaining the equilibrium of ecological processes at both regional and global scales [8, 10].

Dipterocarp forests flourish in the Malay Peninsula, Sumatra, Borneo, Java and wetter parts of the Philippines, and stretch to the northern limits of tropical Asia [8]. The species diversity of the Dipterocarpaceae is dramatically reduced at the range limits in contrast to the aseasonal equatorial zones of Malaysia and Indonesia [7, 11]. Hainan Island in China is located near the edge of the Asian tropics, and only two *Hopea* species, *H. hainanensis* Merrill & Chun and *H. reticulata* Tardieu, are found there. *H. hainanensis* grows in the tropical lowland forests of Hainan Island and Northern Vietnam. It is a large evergreen tree with a height up to 20 meters, and it blooms and fruits almost every year. This tree species serves as an indicator for tropical forests on Hainan Island [11, 12] and is known for its highly valued timber, which is extremely durable and suitable for building boats, bridges and houses. Polyploidy is infrequent in the family Dipterocarpaceae, but triploid and tetraploid species were recorded in genus *Hopea* [13, 14]. Although many dipterocarps have a mixed mating system, most mature seeds are outcrossed, indicating selective abortion of selfed fruits [8]. Neither ploidy nor mating system has been reported for *H. hainanensis*. The size of *H. hainanensis* populations on Hainan Island has been greatly reduced due to the overexploitation and habitat loss for rubber tree plantations and agriculture [15, 16]. The remaining populations are now severely fragmented, preserved only in a few natural reserves [11]. *H. hainanensis* was assessed as endangered in the IUCN Red List of Threatened Species [16] and classified as the first-class protective plant in the Information System of Chinese Rare and Endangered Plants (ISCREP) [17]. This species is quite scarce even in protected areas and the mature trees were estimated to be less than 250 [17]. However, the lack of information on the distribution of genetic variation in *H. hainanensis* has hampered the effective conservation and management of this endangered species.

In view of the ecological and silvicultural importance of the Dipterocarpaceae, there have been many studies of dipterocarp trees that have investigated patterns of genetic diversity, fine-scale spatial genetic structure, mating system and gene flow among populations [18–22] The majority of studied species have moderate to high levels of genetic variation, and low differentiation among populations, suggesting a historical pattern of outcrossing combined with large, stable populations [3, 8]. In addition, the genetic diversity and population structure of endangered species in the Dipterocarpaceae have been assessed for the purposes of conservation and management of genetic resources [23–26].

In this study, population genetic analyses of *H. hainanensis* were performed using 12 microsatellite markers newly developed for this species. The goal of the study was to quantify the amount of genetic variation in *H. hainanensis* populations on Hainan Island, and compare this to the genetic diversity of non-threatened *Hopea* species. Then, the geographic pattern of genetic variation was revealed and discussed. The effect of habitat fragmentation on genetic differentiation across populations was also evaluated. Our studies on genetic diversity and population structure should facilitate the long-term conservation of this endangered species.

Materials and methods

Sample collection

Totally 76 individuals were collected from 10 natural reserves for population genetic analyses, as *H. hainanensis* is very rare in tropical forests of Hainan Island [17] (Table 1, Fig 1). The field investigation was approved by the Administration Office of Nature Reserves, Forestry Department of Hainan Province. The habitat of *H. hainanensis* is primary or secondary tropical lowland forest, at altitudes ranging from 200 to 1000 meters [16]. Leaf samples were collected from mature trees with a diameter at breast height larger than 0.2 meter; all sampled trees were separated by a distance of at least ten meters. Young leaves free of disease and damage were collected and then immediately dried with silica gel. Voucher specimens for *H. hainanensis* were deposited in the Herbarium of Hainan University, Haikou, China.

DNA isolation, SSR amplification and genotyping

Total genomic DNA was extracted from the silica gel-dried leaves using a DNeasy Plant Mini Kit (QIAGEN, Shanghai, China) following the manufacturer's instructions, with the addition of a step to remove high levels of polysaccharides in the initial extracts. The concentration and quality of the extracted DNA were measured with a NanoDrop 2000 spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE, United States). Using second generation sequencing and computational screening, twelve polymorphic SSR primer pairs were developed to genotype *H. hainanensis*. PCR amplification was carried out using an Eppendorf Mastercycler ep Gradient S thermocycler (Eppendorf, Hamburg, Germany); the 20 μL final reaction volume contained 2 μL gDNA (at least 50 μg/mL), 0.2 μL of each primer (50 μM) 10 μL 2× Taq PCR MasterMix (TIANGEN Biotech, Beijing, China), and ddH$_2$O. The following cycling program was used: an initial five minute denaturation at 94˚C; followed by 32 cycles of denaturing at 94˚C for 20 s, annealing at 59–63˚C for 20 s, and extension at 72˚C for 60 s; with a final extension of seven minutes at 72˚C. PCR products were mixed with Hi-Di Formamide

Table 1. Sampling information for *Hopea hainanensis* populations on Hainan Island, China.

| Population code	Sample size	Collection locality	Geographic coordinates	Voucher collection no.
BL	8	Baolong Forestry Station	18.4855˚N, 109.4385˚E	HBL10
BW	8	Bawang Mountain	19.0982˚N, 109.1313˚E	HBW09
DL	7	Diaoluo Mountain	18.6961˚N, 109.8839˚E	HDL06
FJ	9	Fanjia Country	19.2722˚N, 109.6150˚E	HFJ05
JF	7	Jianfeng Mountain	18.7422˚N, 108.9902˚E	HJF04
JX	9	Jiaxi Country	18.8429˚N, 109.1662˚E	HJX02
KF	7	Kafa Mountain	18.6988˚N, 109.3303˚E	HKF03
LM	8	Limu Mountain	19.1909˚N, 109.7417˚E	HLM01
MR	6	Maorui Forestry Station	18.6724˚N, 109.4116˚E	HMR08
QW	7	Qinwang Mountain	18.9388˚N, 109.4468˚E	HQW07
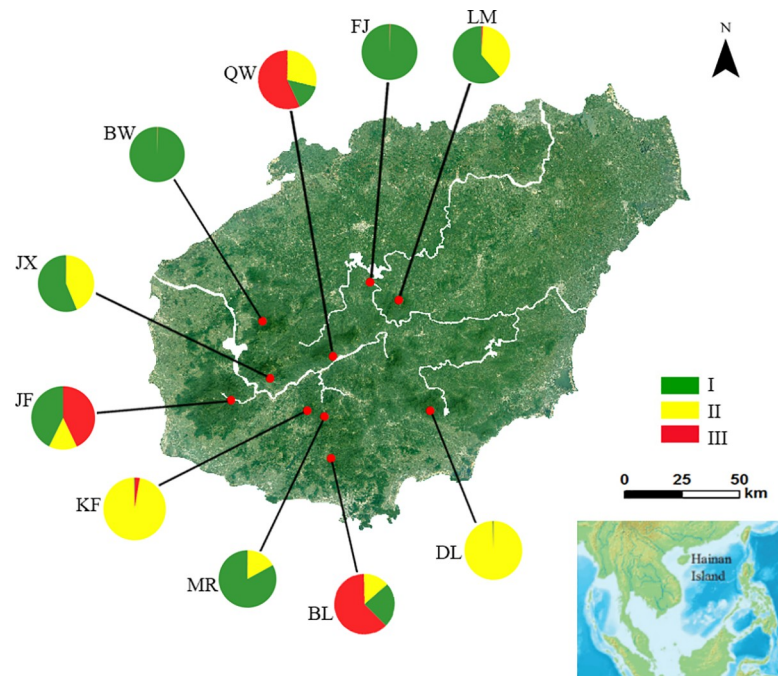
Fig 1. Geographic location of *Hopea hainanensis* populations used in this study (red dots). Pie charts illustrate the proportion of each of the three genetic clusters identified by STRUCTURE analyses for each population. Green, yellow and red represent genetic clusters I, II and III, respectively.

loading buffer (Applied Biosystems, USA) and GeneScan 500 LIZ Size Standard (Life Technologies, Carlsbad, California, USA), and then genotyped on an ABI 3730XL DNA Analyzer (Applied Biosystems) owned by the TIANYI Biotechnology Company (Beijing, China). Alleles were scored using the GeneMarker software (SoftGenetics, State College, Pennsylvania, USA).

Data analysis

In view of the autopolyploid nature of *H. hainanensis* (personal communication with Rong Wang of East China Normal University, who initiated the whole genome sequencing of *H. hainanensis*), allelic dosage was determined based on the ratios between peak intensities following the MAC-PR method [27] (S1 File). However, in the analysis of this autopolyploid, genotypic ambiguities caused by unknown allelic dosage could not be fully resolved with the MAC-PR method. In order to account for missing dosage information for partial heterozygotes, GenoDive version 3.04 [28] was used. Standard metrics of genetic diversity, such as the number of alleles (N_a), effective number of alleles (N_e), and observed (H_o) and expected (H_e) heterozygosity were then calculated. The heterozygosity-based G_{is} statistic was used to test for deviations from Hardy-Weinberg equilibrium (HWE). Another challenge posed by autopolyploidy is polysomic inheritance, under which double-reduction may occur, biasing the results of standard population genetic analyses [29]. A new software package, Polygene version 1.2b, was developed to analyze polyploid genetic data while taking into account both genotypic ambiguities and double-reduction [30]. Four polysomic inheritance models are implemented in the software, and the optimal model is selected based on the Bayesian information criterion (BIC). For each SSR locus, the possible genotypes and their posterior probabilities are inferred based on the allelic phenotype and inheritance model. The effects of null alleles, negative amplification and self-fertilization are also modeled by Polygene. To take advantage of these benefits,

Polygene version 1.2b was implemented in this study to conduct further population genetic analyses.

The observed (H_o) and expected (H_e) heterozygosity, polymorphic information content (PIC) and inbreeding coefficient (G_{is}) were estimated for each population and locus. Differentiation between populations was assessed using G_{ST} [31]. An analysis of molecular variance (AMOVA, [32]) was implemented in Polygene with modification for polysomic inheritance; AMOVA hierarchically partitions genetic variation among populations. Isolation by distance was assessed by regressing pairwise genetic distance against the natural logarithm of geographical distance (km) using the Mantel test [33] with 9,999 permutations. Slatkin's linearized F_{ST} was adopted as the measure of genetic distance [34]. Principal coordinate analysis (PCoA) was carried out based on Cavalli-Sforza's (1967) chord distances [35], which have been shown to be the least biased distance measure when lacking dosage information [36].

A model-based clustering method, STRUCTURE v. 2.3.4 [37], was implemented to infer population structure, using an admixture model with correlated allele frequencies. Individuals were assigned to groups without prior knowledge of their population affinities [38]. The number of populations (K) varied from one to ten, and for each value of K, ten independent replicates were run with 100,000 burn-in iterations followed by 1,000,000 MCMC (Markov chain Monte Carlo) iterations. The best K was inferred using STRUCTURE Harvester [39] following recommended procedures [40]. The Greedy algorithm of Clumpp version 1.1.2 [41] was used to permute the independent repetitions for the selected value of K, and a graphical representation of the population structure was generated with Distruct version 1.1 [42]. A neighbor-joining tree based on Cavalli-Sforza's (1967) genetic distance was created for the *H. hainanensis* individuals using the software package MEGA 5.0 [43].

A qualitative graphical method was performed to identify potential population bottleneck in *H. hainanensis* [44]. Allele frequencies were estimated using Polygene under the best polysomic inheritance model. SSR alleles were grouped into each of 10 allele frequency classes and then a frequency histogram was plotted. The 10 allele frequency classes are 0.001–0.100, 0.101–0.200, 0.201–0.300, etc. The rationale of the graphical method is that compared to common alleles, rare alleles are expected to be lost rapidly during a bottleneck. As a result, alleles at low frequency (<0.1) would be less abundant than alleles at intermediate frequency (e.g. 0.101–0.200) after population bottleneck, regardless of the mutation rate and model [44]. For comparison, SSR data of three dipterocarp species, a critically endangered tree of the Seychelles, *Vateriopsis seychellarum* [23], and two non-threatened species in the rainforests of Southeast Asia, *Shorea leprosula* [45] and *S. macrophylla* [21], were further analyzed with the qualitative graphical method.

Results

Based on the BIC, the optimal polysomic inheritance model selected for Polygene analyses is RCS (S1 Table). For the 12 SSR markers genotyped across the ten *Hopea hainanensis* populations on Hainan Island, a total of 45 alleles were detected, for an average of 3.75 alleles per locus; the minimum number of alleles detected per locus was two (at loci Hha1, Hha9 and Hha10) and the maximum was six (Hha7 and Hha8). The polymorphism information content (PIC) varied from 0.205 (Hha10) to 0.707 (Hha7), with an average of 0.477. A large number of loci were found to deviate from HWE ($P < 0.05$) across the study populations (Table 2). At the population-level, the effective number of alleles (N_e) ranged from 1.550 in DL to 2.351 in JF, averaging 1.964 alleles per population. The expected heterozygosity (H_e) as estimated by GenoDive ranged from 0.265 (DL) to 0.525 (JF), while the H_e estimated by Polygene ranged from

Table 2. Summary statistics for the 12 SSR markers across all *Hopea hainanensis* populations with metrics of genetic diversity estimated by GenoDive 3.04 and Polygene version 1.2b.

Locus	Sequences (5'–3')	Repeat motif	Allele size	GENODIVE					POLYGENE			
				Na	*Ne*	*Ho*	*He*	dHWE	*Ho*	*He*	PIC	*G*$_{is}$
Hha1	F: AGTTGGAGATTAAAGAAAGTGGCT	(TTTTA)$_6$	103–108	2	1.476	0.423	0.335	7	0.428	0.336	0.280	-0.272
	R: TTCAATTTAGACCCGTGGACCTC											
Hha2	F: ACATGGTCTTTGTTATCTGCTTA	(TTCT)$_7$	155–163	3	1.899	0.552	0.579	5	0.566	0.573	0.507	0.012
	R: CCATGGTGCTACAACCTTTCTTG											
Hha3	F: TTCATGGTCATTGAGTCATAGGT	(AT)$_{10}$	124–134	4	2.010	0.392	0.629	6	0.393	0.579	0.532	0.321
	R: GCCTCTACCTAGTGTATGAAGGC											
Hha4	F: ACCTGGTAAGCCATAACACTGAA	(TTC)$_6$	144–150	3	2.712	0.755	0.661	9	0.756	0.655	0.581	-0.155
	R: TGATGCAAGCTCCAGAAACAAAG											
Hha5	F: AGTCAATGAGAAGGAGACATGTT	(TA)$_8$	116–132	3	1.318	0.000	0.422	5	0.000	0.381	0.339	1.000
	R: AAGTCATTTGGTAAAAGGTGCCC											
Hha6	F: GCTTTCTGCATTTCCTTGAGAGA	(AT)$_9$	141–153	4	1.200	0.060	0.504	4	0.064	0.455	0.416	0.858
	R: TGATTAGCTGCTGAATTTGGCTG											
Hha7	F: ACGAATGGAGGTTTGTAATTGGA	(AT)$_{10}$	127–137	6	2.144	0.405	0.772	8	0.411	0.747	0.707	0.449
	R: AGAGTACAATCGGGATCAATGGA											
Hha8	F: TCAAACGCAACATGGAATAAGGA	(AT)$_9$	222–230	6	2.555	0.503	0.751	8	0.505	0.735	0.691	0.313
	R: AGCCATTAACTCAGAACACGAGA											
Hha9	F: GATGAGGGATAATGGTGCGTTTG	(AAG)$_5$	126–129	2	1.383	0.111	0.468	5	0.104	0.448	0.348	0.767
	R: CAACTCACGCCTCTGTGTTATTG											
Hha10	F: TCAATCGTTTTGAACCACAGGTG	(AT)$_8$	157–159	2	1.244	0.211	0.237	1	0.202	0.232	0.205	0.129
	R: AGCTATTGCCTAGAAGATTTCACAC											
Hha11	F: GGCATCGTAATACCGCATAGAGA	(AT)$_{10}$	157–165	5	1.965	0.455	0.680	4	0.458	0.657	0.607	0.303
	R: CTACCAACAACACTAGGCGCTGT											
Hha12	F: ATTACTAACCTTTGCCCACTCCT	(GT)$_{10}$	78–94	5	1.865	0.530	0.565	4	0.530	0.568	0.510	0.066
	R: ACCAGCTTTAGCCAATTCAAACC											

N_a: number of alleles for each marker across all populations, N_e: effective number of alleles for each marker across all populations, H_o: observed heterozygosity, H_e: expected heterozygosity, dHWE: number of populations deviating from Hardy–Weinberg equilibrium ($P < 0.05$), PIC: polymorphic information content, G_{is}: inbreeding coefficient.

0.233 (DL) to 0.495 (JF). The inbreeding coefficients (G_{is}) were all greater than zero, ranging from 0.138 (FJ) to 0.408 (JF), and averaging 0.243 (Table 3).

The overall G_{ST} across all populations and loci was 0.229. Pairwise comparisons of genetic differentiation between populations showed that G_{ST} ranged from 0.001, between populations BW and FJ, up to 0.271, between populations BW and DL. Some populations, for example QW, KF and DL, were highly divergent from the other populations (S2 Table). The population genetic structure of *H. hainanensis* on Hainan Island did not conform to the isolation by distance model. No correlation was detected (Mantel test: $r = 0.040$, $p = 0.418$) between the genetic distance (as measured by Slatkin's linearized F_{ST}) and the natural logarithm-transformed geographical distance among populations (S1 Fig). An analysis of molecular variance (AMOVA) revealed that 77.43% of total genetic variation was found within populations, while a small portion of variation (22.57%) was attributed to among population differentiation (Table 4).

In model-based STRUCTURE analysis, LnP(K) was found to increase from $K = 1$ to 10 and ΔK was maximized at $K = 3$, suggesting that there are three distinct genetic clusters in *H. hainanensis*. The results of the individual assignment indicated that only a few individuals were

Table 3. Genetic diversity of *Hopea hainanensis* populations based on 12 SSR markers.

Population	N	GenoDive					Polygene			
		Na	*Ne*	*Ho*	*He*	dHWE	*Ho*	*He*	PIC	G_{is}
BL	8	2.917	2.045	0.357	0.495	7	0.356	0.477	0.414	0.258
BW	8	1.833	1.780	0.444	0.355	6	0.444	0.345	0.278	0.139
DL	7	1.583	1.550	0.294	0.265	6	0.292	0.233	0.188	0.367
FJ	9	1.833	1.792	0.441	0.357	7	0.440	0.341	0.275	0.138
JF	7	2.917	2.351	0.295	0.525	8	0.295	0.495	0.433	0.408
JX	9	2.750	2.109	0.373	0.459	8	0.372	0.445	0.387	0.205
KF	7	2.667	1.900	0.348	0.419	5	0.347	0.402	0.345	0.197
LM	8	2.667	2.100	0.425	0.487	6	0.423	0.487	0.418	0.153
MR	6	2.500	1.970	0.385	0.419	7	0.384	0.399	0.347	0.223
QW	7	2.917	2.041	0.304	0.491	6	0.302	0.469	0.405	0.340
Average (SD)	7.2	2.458 (0.511)	1.964 (0.221)	0.367 (0.058)	0.427 (0.081)		0.366 (0.058)	0.409 (0.084)	0.349 (0.080)	0.243 (0.098)

N_a: number of alleles in a population, N_e: effective number of alleles in a population, H_o: observed heterozygosity, H_e: expected heterozygosity, dHWE: number of populations deviating from Hardy–Weinberg equilibrium ($P < 0.05$), PIC: polymorphic information content, G_{is}: inbreeding coefficient.

https://doi.org/10.1371/journal.pone.0241452.t003

found to have mixed ancestry of more than one genetic cluster (Fig 2). Five populations, FJ, BW, DL, KF and MR, were dominated by one genetic subpopulation. In LM and JX, two genetic subpopulations were represented in roughly equal proportions. Lastly, in JF, QW and BL, all three genetic subpopulations were represented, with one cluster (red, III) dominated (Fig 1). Geographically, individuals assigned to the same genetic cluster were widespread, and the distributions of the three genetic clusters were overlapping (Fig 1).

The genetic relationships among *H. hainanensis* individuals were further explored using PCoA, and the results were consistent with the STRUCTURE analysis (Fig 3). The first two principal coordinates explained 43.89% of the total genetic variance (27.69% and 16.21%, respectively). Three genetic groups with clear boundaries were identified, corresponding to the three genetic components determined in the STRUCTURE analysis. In addition, using Cavalli-Sforza's (1967) chord distances, a neighbor-joining tree was constructed. In the tree, individuals were divided into three groups, again in agreement with the three genetic groups identified by both PCoA and STRUCTURE (S2 Fig).

The distribution of allele frequencies in *H. hainanensis* is obvious distorted and different from the other three dipterocarp species (Fig 4). The number of alleles at low frequency (<0.1) was equal to the number of alleles at intermediate frequency (0.101–0.200) in *H. hainanensis*. However, the alleles at low frequency (<0.1) are much more abundant than those at intermediate frequency (0.101–0.200) in the other two non-threatened *Shorea* species. Such a mode-shift distortion of allele frequency distribution is a characteristic of bottlenecked populations, revealing that *H. hainanensis* on Hainan Island may have undergone a recent population bottleneck.

Table 4. Analysis of molecular variance (AMOVA) for *Hopea hainanensis* populations.

Source	df	Sum of squares	Variance	Percentage of variation	*F*-statistic
Among populations	108	7980.81	2.06	22.57	$F_{ST} = 0.23$
Among individuals within population	764	10111.39	2.05	22.47	$F_{IS} = 0.29$
Within individuals	2652	13319.49	5.02	54.96	$F_{IT} = 0.45$

All variance components were statistically significant ($P < 0.005$); df: degrees of freedom.

https://doi.org/10.1371/journal.pone.0241452.t004

Delta K = mean ($|\mathrm{L}''(K)|$)/sd(L(K))

(a)

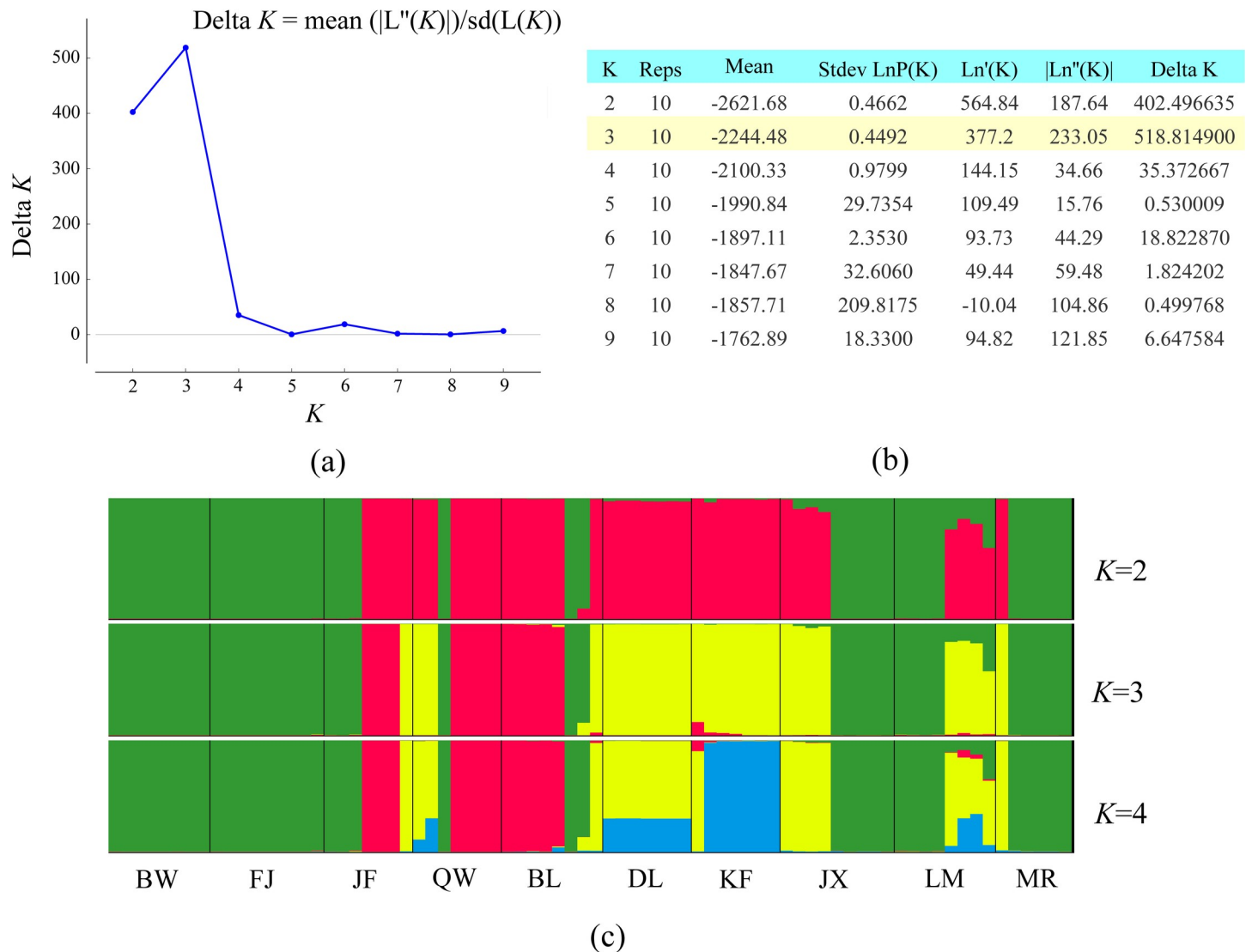| K	Reps	Mean	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
2	10	-2621.68	0.4662	564.84	187.64	402.496635
3	10	-2244.48	0.4492	377.2	233.05	518.814900
4	10	-2100.33	0.9799	144.15	34.66	35.372667
5	10	-1990.84	29.7354	109.49	15.76	0.530009
6	10	-1897.11	2.3530	93.73	44.29	18.822870
7	10	-1847.67	32.6060	49.44	59.48	1.824202
8	10	-1857.71	209.8175	-10.04	104.86	0.499768
9	10	-1762.89	18.3300	94.82	121.85	6.647584

(b)

(c)

Fig 2. Results of the STRUCTURE analysis. (a) Using the ΔK method, which is based on the rate of change in the log probability of the data between successive K values, the most likely number of clusters was $K = 3$. (b) Log probabilities and ΔK values for K from two to nine. (c) The results of individual assignment at $K = 2$, 3 and 4. Each vertical bar represents an individual, and the proportion of the colors corresponds to the posterior probability of assignment to one of the three genetic clusters.

Discussion

Genetic diversity in *Hopea hainanensis*

Species in the family Dipterocarpaceae are generally characterized by moderate to high levels of genetic diversity and low differentiation among populations [8]. This pattern of genetic variation is expected for long-lived, perennial woody plants with predominantly outcrossing mating systems [46]. For example, using SSR markers, most *Shorea* species have been found to have high levels of genetic variation (as measured by expected heterozygosity [H_e] and the number of alleles per locus) [19, 47–49]. *Hopea dryobalanoides* is a non-threatened species widely distributed in Malaysia, Sumatra and Borneo [50]. A single population of *H. dryobalanoides* was collected from the Pasoh Forest Reserve in Peninsular Malaysia, and five SSR markers were used to evaluate its genetic diversity. Expected heterozygosity (H_e) was estimated at
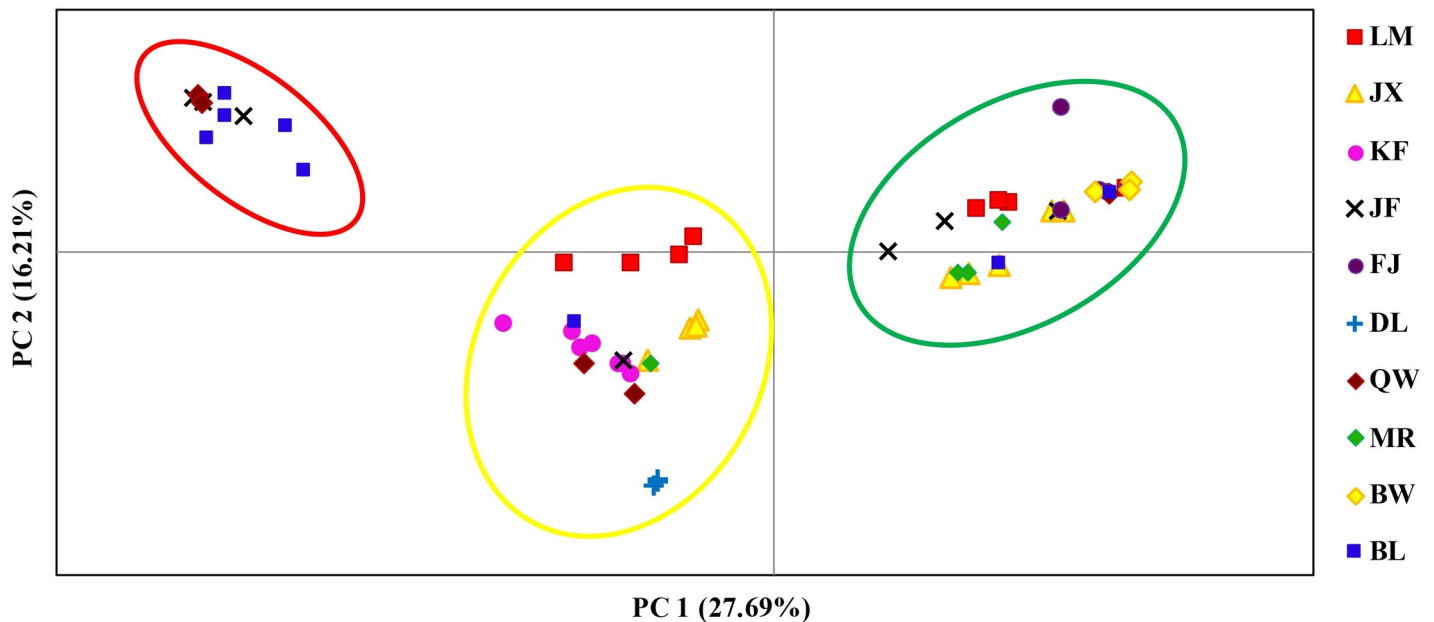
Fig 3. Principal coordinate analysis (PCoA) based on Cavalli-Sforza's (1967) chord distance among individual samples of *Hopea hainanensis*. The three genetic groups identified by PCoA (highlighted by green, yellow and red ellipses) were consistent with the STRUCTURE results when *K* = 3.

0.678 in *H. dryobalanoides*, which is comparable with that found for a number of *Shorea* species [21, 22, 49]. In contrast, in this study, the average H_e for *H. hainanensis* populations from Hainan Island was 0.409 as estimated by Polygene and 0.427 as estimated by GenoDive, which is markedly lower than the H_e measured for *H. dryobalanoides* [50]. Furthermore, fewer alleles were discovered in *H. hainanensis* (N_a = 2.46) versus *H. dryobalanoides* (N_a = 5.60). In most cases, reductions in the effective population size and shifts of mating system from outcrossing to selfing cause the loss of genetic diversity within species [51, 52]. The low genetic diversity within populations of *H. hainanensis* is most likely due to a severe demographic bottleneck, as this species has undergone approximately a 70% population reduction over the last three hundred years [16]. On Hainan Island, deforestation is greatly accelerated in the 20th century. About 80% to 95% of the primary forests were destroyed due to massive logging for timber, transitions to rubber tree and *Eucalyptus* plantations, as well as urban expansion [15, 53]. The *H. hainanensis* populations would shrink proportionally, perhaps even more, due to the high quality of its timber.

A population bottleneck was manifested by the qualitative graphical method based on the distorted distribution of allele frequencies in *H. hainanensis* populations (Fig 4). In non-bottle-necked populations near mutation-drift equilibrium, selectively neutral loci are expected to have a large proportion of alleles at low frequency. However, in recently bottlenecked populations, the expected distribution of allele frequencies would be distorted, with fewer alleles in the low frequency class (<0.1) than in one or more intermediate frequency classes [44, 54]. The graphical method makes no assumption about the ploidy of the studied species [44] and thus could be used in cases where common bottleneck testing tools, such as BOTTLENECK [55], could not be applied. We found a large number of low-frequency alleles (<0.1) are maintained in populations of the two non-threatened *Shorea* species. The ratio of the number of alleles at low frequency (<0.1) to the number of alleles at intermediate frequency (0.1–0.2) is 11.35 for *S. leprosula* and 7.10 for *S. macrophylla*. However, the ratios are 3.86 and 1.0 for *V. seychellarum* and *H. hainanensis*, respectively. Obviously, many low-frequency alleles have
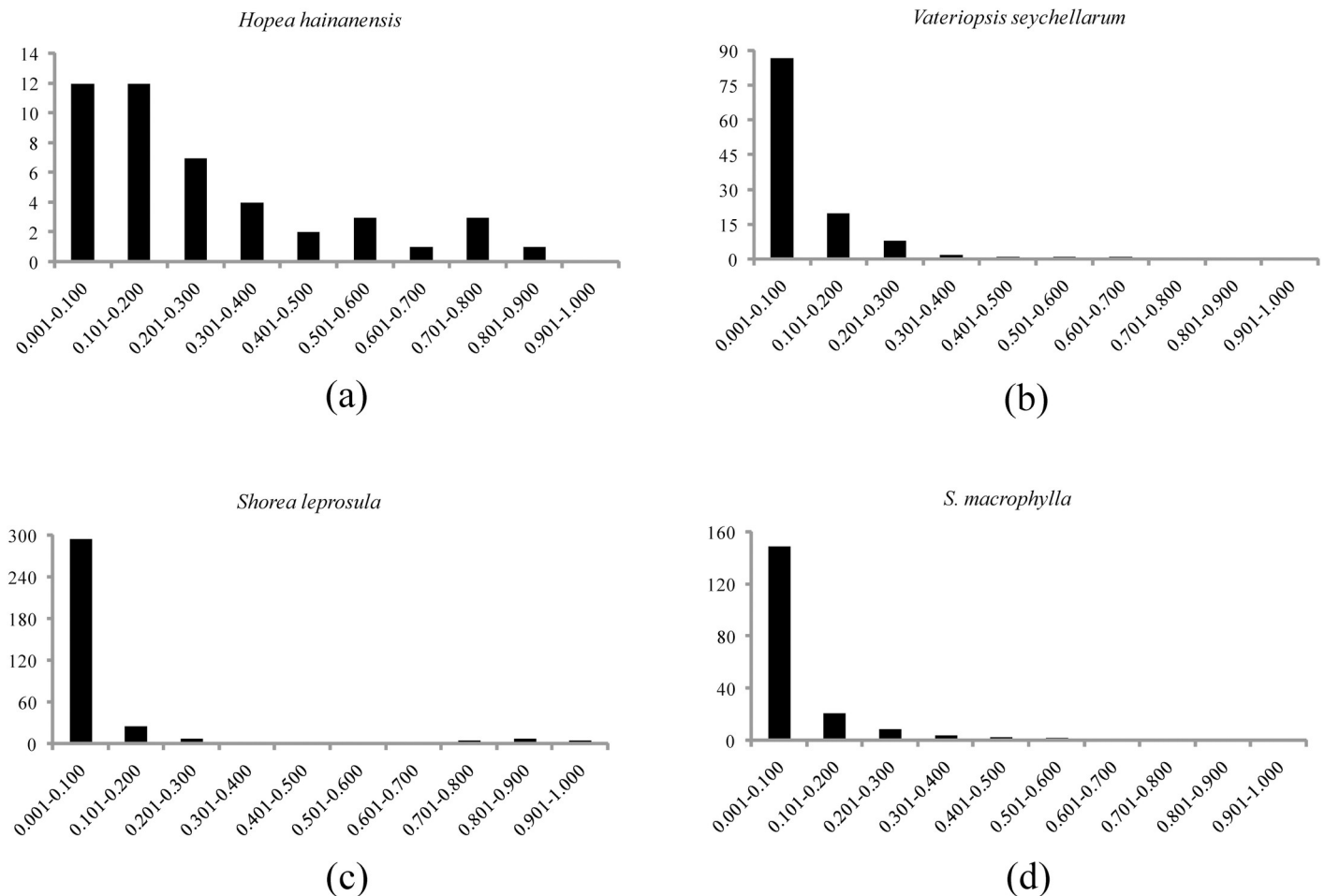
Fig 4. The allele frequency distributions based on SSR data from *Hopea hainanensis*, *Vateriopsis seychellarum*, *Shorea leprosula* and *S. macrophylla*.

been lost during recent bottlenecks in the populations of *V. seychellarum* and *H. hainanensis*. A higher proportion of low-frequency alleles are maintained in *V. seychellarum* than in *H. hainanensis*, which may arise from different sample sizes used for the two species. Since the graphical method requires only samples of 5 to 20 polymorphic loci and approximately 30 individuals, the SSR data used in this study should be adequate [44]. To conclude, the *H. hainanensis* populations on Hainan Island may have experienced a recent demographic bottleneck, resulting in relative fewer alleles at low frequency and lower genetic diversity detected in this species. The genetic diversity of *H. hainanensis* in Northeast Vietnam was recently evaluated using 10 SSR markers [56]. The populations collected in Northeast Vietnam (H_e = 0.382, allelic richness = 2.08) and Hainan Island have similar levels of genetic variation (Table 3). A population bottleneck was reported and suggested to be responsible for the low genetic diversity of these populations [56].

When the size of a population is drastically reduced, genetic drift may override natural selection and act as the main evolutionary force, leading to a loss of genetic variation [3]. Moreover, inbreeding is inevitable in small populations [3, 5]. The inbreeding level of a population can be quantified with the inbreeding coefficient (G_{is}). In this study, G_{is} ranged from 0.138 to 0.408, higher than the G_{is} estimates for many other dipterocarps [48–50]. The proposed bottleneck event in *H. hainanensis* and resulting population fragmentation likely

produced small and isolated populations, in which enhanced inbreeding would be expected [3]. Significant deviation from HWE was detected in a large portion of the loci from the 10 study populations (Table 3). Nonrandom mating may be the main cause of the departure from HWE in those populations. In conclusion, the low genetic diversity, increased inbreeding and deviations from HWE found in the *H. hainanensis* populations on Hainan Island are typical of endangered species, and are likely the consequence of intensified inbreeding and genetic drift acting in small *H. hainanensis* populations.

Population genetic structure and differentiation

Three genetic groups were identified in *H. hainanensis* using both Bayesian model-based clustering and principal coordinate analysis (PCoA) (Figs 2 and 3). The assignment of individuals to genetic groups was consistent between the two methods. Additionally, the topology of a neighbor-joining tree was concordant with the clustering results from STRUCTURE and the PCoA (S1 Fig). The three genetic clusters were widely distributed across the mountainous area of Hainan Island, without any clear geographical pattern (Fig 1). In addition, a Mantel test did not detect a correlation between genetic divergence (as Slatkin's linearized F_{ST}) and geographical distance (S2 Fig), suggesting that genetic differentiation in *H. hainanensis* does not follow a pattern of isolation by distance. Such population genetic structure implies the existence of long-distance gene flow among populations before the species-wide bottleneck. Long-distance gene flow has been reported in *Neobalanocarpus heimii*, *Dipterocarpus tempheses* and several *Shorea* species [56, 57]. *Neobalanocarpus heimii* is an emergent tree species endemic to the Malay Peninsula. Using paternity analysis, the average distance of pollen flow in *N. heimii* was estimated to be 191 m, while several pollination events were found to exceed 400 m [58]. In addition, seeds produced by *N. heimii* may be transported by squirrels. Interestingly, as with *H. hainanensis*, no positive correlation between genetic relatedness and spatial distance was detected in *N. heimii*. Long-distance gene flow via pollen and seed migration are likely responsible for the weak genetic structure in this species [58]. In summary, the widespread geographical distribution of genetic subpopulations and the failure to find isolation by distance in *H. hainanensis* are likely due to long-distance gene flow among populations before the species-wide bottleneck.

Endangered dipterocarp species are generally highly differentiated across populations based on the analysis of SSR markers. For example, the overall G_{ST} across adult populations of the endangered *V. seychellarum* is 0.20 [23]. Habitat fragmentation and restricted gene flow were suggested as possible reasons for the strong genetic divergence of this species [23]. Genetic differentiation was quantified for a few *Hopea* species. A moderate level of differentiation between two geographically adjacent populations was revealed in *H. bilitonensis* (mean value of G_{ST} is 0.116), which is an extremely rare and predominantly selfing dipterocarp in Peninsular Malaysia [24]. Two threatened *Hopea* species were also reported to have low to moderate levels of genetic differentiation (G_{ST} = 0.009 for *H. chinensis*; G_{ST} = 0.102 for *H. odorata*) [56]. Compared to the three above *Hopea* species, *H. hainanensis* showed higher overall genetic divergence (G_{ST} = 0.229). The AMOVA analysis also revealed high differentiation among populations (F_{ST} = 0.23). Population genetic theory predicts that under Island model of complete isolation, reduction in population size could promote divergence among populations [59], which is likely the reason of elevated differentiation observed in *H. hainanensis*. However, other factors, such as the number of generations after population isolation (*t*), geographic distance between sampled populations, mating system and evolutionary history also influence the level of population divergence [46]. In summary, further research is needed to explore the effect of habitat fragmentation on genetic differentiation across *H. hainanensis* populations.

Implications for conservation. Because the loss of genetic variation is a major threat to the survival of endangered species, an important conservation action is to preserve and restore genetic variation [3, 60]. Three *H. hainanensis* populations, BW, FJ and DL, were found to have relatively lower levels of genetic variation. These three populations may therefore be more vulnerable to abiotic and biotic stresses and should be the focus of special conservation attention. Meanwhile, the BL, JF, JX, LM and QW populations had relatively higher levels of genetic diversity and contained more than one genetic subpopulation. These five populations could therefore serve as seed sources for the propagation of seedlings and saplings, to be used in the restoration of previously logged lowland rainforests on Hainan Island. The natural regeneration of *H. hainanensis* populations is challenging: seedlings and saplings grow very slowly and often fail to establish in the highly shaded understory. Therefore, selecting populations with high genetic diversity (e.g. BL, JF, LM and QW) to produce seedlings may facilitate the restoration of endangered *H. hainanensis* populations on Hainan Island.

Supporting information

S1 File. SSR genotypes of 76 *H. hainanensis* individuals used in this study.
(XLSX)

S1 Table. The BIC scores of the four polysomic inheritance models implemented in Polygene version 1.2. Two options, "Consider negative PCR" and "Consider selfing", were combined along with optimal model selection.
(XLSX)

S2 Table. Pairwise divergence among *Hopea hainanensis* populations based on Nei's G_{ST}.
(XLSX)

S1 Fig. Mantel test of the correlation between genetic distance (as Slatkin's linearized F_{ST}) and the natural logarithm of geographical distance (km).
(TIF)

S2 Fig. Neighbor joining tree for individual samples based on Cavalli-Sforza's (1967) genetic distance. Colored bars represent individual assignments by STRUCTURE when $K = 3$.
(TIF)

Acknowledgments

We would like to thank Dr. Emily Drummond at the University of British Columbia for her editing of the manuscript.

Author Contributions

Conceptualization: Xiang Ma, Mingxun Ren, Liang Tang.

Data curation: Chen Wang.

Funding acquisition: Xiang Ma, Liang Tang.

Investigation: Chen Wang.

Resources: Mingxun Ren.

Supervision: Liang Tang.

Visualization: Chen Wang, Liang Tang.

Writing – original draft: Liang Tang.

Writing – review & editing: Xiang Ma, Liang Tang.

References

1. Frankham R. Conservation genetics. Annu Rev Genet. 1995; 29:305–27. https://doi.org/10.1146/annurev.ge.29.120195.001513 PMID: 8825477

2. DeSalle R, Amato G. The expansion of conservation genetics. Nat Rev Genet. 2004; 5(9):702–12. https://doi.org/10.1038/nrg1425 PMID: 15372093

3. Frankham R, Ballou JD, Briscoe DA. Introduction to conservation genetics. 2nd ed. Cambridge: Cambridge University Press; 2010.

4. IUCN. Red List of Threatened Species 2020. Available from: https://www.iucnredlist.org/

5. Hedrick PW, Kalinowski ST. Inbreeding depression in conservation biology. Annu Rev Ecol Syst. 2000; 31:139–62.

6. Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma R, Hedrick PW. Conservation genetics in transition to conservation genomics. Trends Genet. 2010; 26(4):177–87. https://doi.org/10.1016/j.tig.2010.01.001 PMID: 20227782

7. Ashton PS. Dipterocarp biology as a window to the understanding of tropical forest structure. Annu Rev Ecol Syst. 1988; 19(1):347–70.

8. Ghazoul J. Dipterocarp Biology, Ecology, and Conservation. Oxford: Oxford University Press; 2016. https://doi.org/10.1073/pnas.1517092112 PMID: 26621730

9. Sodhi NS, Koh LP, Brook BW, Ng PKL. Southeast Asian biodiversity: an impending disaster. Trends Ecol Evol. 2004; 19(12):654–60. https://doi.org/10.1016/j.tree.2004.09.006 PMID: 16701328

10. Edwards DP, Tobias JA, Sheil D, Meijaard E, Laurance WF. Maintaining ecosystem function and services in logged tropical forests. Trends Ecol Evol. 2014; 29(9):511–20. https://doi.org/10.1016/j.tree.2014.07.003 PMID: 25092495

11. Li XW, Li J, Ashton PS. Dipterocarpaceae. In Wu ZY, Raven PH, Hong DY, editors. Flora of China, Volume 13. Beijing: Science Press and St. Louis: Missouri Botanical Garden Press; 2007.

12. Zhu H. Tropical flora of southern China. Biodiversity Science. 2017; 25(2):204–17. https://doi.org/10.1002/ece3.3561 PMID: 29238563

13. Choong CY, Wickneswari R, Norwati M, Abbott RJ. Phylogeny of *Hopea* (Dipterocarpaceae) inferred from chloroplast DNA and nuclear PgiC sequences. Mol Phylogenet Evol. 2008; 48(3):1238–43. https://doi.org/10.1016/j.ympev.2008.01.004 PMID: 18280183

14. Ng CH, Lee SL, Tnah LH, Ng KKS, Lee CT, Madon M. Genome size variation and evolution in Dipterocarpaceae. Plant Ecol Divers. 2016; 9(5–6):437–46.

15. Lin SL, Jiang YZ, He JK, Ma GZ, Xu Y, Jiang HS. Changes in the spatial and temporal pattern of natural forest cover on Hainan Island from the 1950s to the 2010s: implications for natural forest conservation and management. PeerJ. 2017; 5: e3320. https://doi.org/10.7717/peerj.3320 PMID: 28533968

16. Ly V.; Nanthavong K.; Pooma R.; Hoang V.S.; Khou E.; Newman M.F. Hopea hainanensis. The IUCN Red List of Threatened Species 2018: e.T32357A2816074. Availabe online: https://dx.doi.org/10.2305/IUCN.UK.2018-1.RLTS.T32357A2816074.en.

17. Information System of Chinese Rare and Endangered Plants (ISCREP). Available from: http://www.iplant.cn/rep/.

18. Kettle CJ, Hollingsworth PM, Burslem D, Maycock CR, Khoo E, Ghazoul J. Determinants of fine-scale spatial genetic structure in three co-occurring rain forest canopy trees in Borneo. Perspect Plant Ecol Evol Syst. 2011; 13(1):45–54.

19. de Morais CT, Ghazoul J, Maycock CR, Bagchi R, Burslem D, Khoo E, et al. Understanding local patterns of genetic diversity in dipterocarps using a multi-site, multi-species approach: Implications for forest management and restoration. For Ecol Manage. 2015; 356:153–65.

20. Widiyatno, Indrioko S, Na'iem M, Purnomo S, Hosaka T, Uchiyama K, et al. Effects of logging rotation in a lowland dipterocarp forest on mating system and gene flow in *Shorea parvifolia*. Tree Genet Genomes. 2017; 13(4):85.

21. Utomo S, Uchiyama K, Ueno S, Matsumoto A, Widiyatno, Indrioko S, et al. Effects of Pleistocene climate change on genetic structure and diversity of *Shorea macrophylla* in Kalimantan rainforest. Tree Genet Genomes. 2018; 14(4):44.

22. Ng CH, Lee SL, Tnah LH, Ng KKS, Lee CT, Diway B, et al. Genetic diversity and demographic history of an upper hill dipterocarp (*Shorea platyclados*): implications for conservation. J Hered. 2019; 110 (7):844–56. https://doi.org/10.1093/jhered/esz052 PMID: 31554011

23. Finger A, Kettle CJ, Kaiser-Bunbury CN, Valentin T, Mougal J, Ghazoul J. Forest fragmentation genetics in a formerly widespread island endemic tree: *Vateriopsis seychellarum* (Dipterocarpaceae). Mol Ecol. 2012; 21(10):2369–82. https://doi.org/10.1111/j.1365-294X.2012.05543.x PMID: 22463385

24. Lee SL, Chua LSL, Ng KKS, Hamidah M, Lee CT, Ng CH, et al. Conservation management of rare and predominantly selfing tropical trees: an example using *Hopea bilitonensis* (Dipterocarpaceae). Biodivers Conserv. 2013; 22(13–14):2989–3006.

25. Ismail SA, Ghazoul J, Ravikanth G, Kushalappa CG, Shaanker RU, Kettle CJ. Fragmentation genetics of *Vateria indica*: implications for management of forest genetic resources of an endemic dipterocarp. Conserv Genet. 2014; 15(3):533–45.

26. Nguyen TM, Vu DD, Nguyen DM, Dang HP, Phan LK, Bui PX. Microsatellite analysis reveals genetic diversity of the endangered species *Dipterocarpus dyeri*. J. Forest Res. 2020; 25(3):198–201.

27. Esselink GD, Nybom H, Vosman B. Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting-peak ratios) method. Theor Appl Genet. 2004; 109(2):402–8. https://doi.org/10.1007/s00122-004-1645-5 PMID: 15085263

28. Meirmans PG. genodive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Mol Ecol Resour*. 2020; 20(4): 1126–31. https://doi.org/10.1111/1755-0998.13145 PMID: 32061017

29. Huang K, Wang TC, Dunn DW, Zhang P, Cao XX, Liu RC, et al. Genotypic frequencies at equilibrium for polysomic inheritance under double-reduction. G3-Genes Genomes Genet. 2019; 9(5):1693–706. https://doi.org/10.1534/g3.119.400132 PMID: 30910817

30. Huang K, Dunn DW, Ritland K, Li BG. polygene: Population genetics analyses for autopolyploids based on allelic phenotypes. Methods Ecol Evol. 2020; 11(3):448–56.

31. Nei M. Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci U S A. 1973; 70 (12):3321–3. https://doi.org/10.1073/pnas.70.12.3321 PMID: 4519626

32. 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 (2):479–91. PMID: 1644282

33. Mantel N. The detection of disease clustering and a generalized regression approach. Cancer Research. 1967; 27(2 Part 1):209–20. PMID: 6018555

34. Rousset F. Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. Genetics. 1997; 145(4):1219–28. PMID: 9093870

35. Cavalli-Sforza LL, Edwards AWF. Phylogenetic analysis: Models and estimation. Evolution. 1967; 21 (9):550–70.

36. Meirmans PG, Liu SL, van Tienderen PH. The analysis of polyploid genetic data. J Hered. 2018; 109 (3):283–96. https://doi.org/10.1093/jhered/esy006 PMID: 29385510

37. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155(2):945–59. PMID: 10835412

38. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics. 2003; 164(4):1567–87. PMID: 12930761

39. Earl DA, Vonholdt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012; 4(2):359–61.

40. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005; 14(8):2611–20. https://doi.org/10.1111/j.1365-294X.2005.02553.x PMID: 15969739

41. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007; 23 (14):1801–6. https://doi.org/10.1093/bioinformatics/btm233 PMID: 17485429

42. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes. 2004; 4(1):137–8.

43. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011; 28(10):2731–9. https://doi.org/10.1093/molbev/msr121 PMID: 21546353

44. Luikart G, Allendorf FW, Cornuet JM, Sherwin WB. Distortion of allele frequency distributions provides a test for recent population bottlenecks. J Hered. 1998; 89(3):238–47. https://doi.org/10.1093/jhered/89.3.238 PMID: 9656466

45. Ohtani M, Kondo T, Tani N, Ueno S, Lee LS, Ng KKS, et al. 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. Mol Ecol. 2013; 22(8):2264–79. https://doi.org/10.1111/mec.12243 PMID: 23432376

46. Hamrick JL, Godt MJW. Effects of life history traits on genetic diversity in plant species. Philos Trans R Soc Lond Ser B-Biol Sci. 1996; 351(1345):1291–8.

47. Lee SL, Tani N, Ng KKS, Tsumura Y. Isolation and characterization of 20 microsatellite loci for an important tropical tree *Shorea leprosula* (Dipterocarpaceae) and their applicability to *S. parvifolia*. Molecular Ecology Notes. 2004; 4(2):222–5.

48. Obayashi K, Tsumura Y, Ihara-Ujino T, Niiyama K, Tanouchi H, Suyama Y, et al. Genetic diversity and outcrossing rate between undisturbed and selectively logged forests of *Shorea curtisii* (Dipterocarpaceae) using microsatellite DNA analysis. Int J Plant Sci. 2002; 163(1):151–8.

49. Pandey M, Geburek T. Genetic differences between continuous and disjunct populations: some insights from sal (*Shorea robusta* Roxb.) in Nepal. Conserv Genet. 2010; 11(3):977–84.

50. Takeuchi Y, Ichikawa S, Konuma A, Tomaru N, Niiyama K, Lee SL, et al. Comparison of the fine-scale genetic structure of three dipterocarp species. Heredity. 2004; 92(4):323–8. https://doi.org/10.1038/sj.hdy.6800411 PMID: 14735142

51. Charlesworth B. Effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 2009; 10(3):195–205. https://doi.org/10.1038/nrg2526 PMID: 19204717

52. Ellegren H, Galtier N. Determinants of genetic diversity. Nat Rev Genet. 2016; 17(7):422–33. https://doi.org/10.1038/nrg.2016.58 PMID: 27265362

53. Zhou J, Wei F, Li M, Zhang JF, Wang D, Pan RL. Hainan black-crested gibbon is headed for extinction. Int J Primatol. 2005; 26(2):453–65.

54. Peery MZ, Kirby R, Reid BN, Stoelting R, Doucet-Beer E, Robinson S, et al. Reliability of genetic bottleneck tests for detecting recent population declines. Mol Ecol. 2012; 21(14):3403–18. https://doi.org/10.1111/j.1365-294X.2012.05635.x PMID: 22646281

55. Piry S, Luikart G, Cornuet JM. BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered. 1999; 90(4):502–3.

56. Nguyen PTT, Triest L. The genetics structure of three threatened *Hopea* species (Dipterocarpaceae) in the protected areas of Vietnam. Int J Appl Nat Sci. 2019; 8(3):191–204.

57. Kenta T, Isagi Y, Nakagawa M, Yamashita M, Nakashizuka T. Variation in pollen dispersal between years with different pollination conditions in a tropical emergent tree. Mol Ecol. 2004; 13(11):3575–84. https://doi.org/10.1111/j.1365-294X.2004.02345.x PMID: 15488013

58. Konuma A, Tsumura Y, Lee CT, Lee SL, Okuda T. Estimation of gene flow in the tropical-rainforest tree *Neobalanocarpus heimii* (Dipterocarpaceae), inferred from paternity analysis. Mol Ecol. 2000; 9 (11):1843–52. https://doi.org/10.1046/j.1365-294x.2000.01081.x PMID: 11091320

59. Hedrick PW. Genetics of populations. 4th ed. Sudbury: Jones & Bartlett Learning; 2010.

60. Keller LF, Waller DM. Inbreeding effects in wild populations. Trends Ecol Evol. 2002; 17(5):230–41.