Journal of Tropical Forest Science 37(4): 484–501 (2025)
https://doi.org/10.26525/jtfs2025.37.4.484
ISSN: 0128-1283, eISSN: 2521-9847

INTEGRATING SSRS AND GENOME-WIDE SNPS TO ASSESS GENETIC DIVERSITY AND POPULATION STRUCTURE OF *RUBROSHOREA LEPROSULA* FOR CONSERVATION

Nur-Nabilah A[1, 2], Ng KKS[1, *], Lee SL[1], Ng CH[1], Lee CT[1], Tnah LH[1] & Zeti-Azura Mohamed-Hussein[2]

[1]*Forest Research Institute Malaysia (FRIM), 52109 Kepong, Selangor, Malaysia*
[2]*Faculty of Science and Technology (FST), Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia*

**kevin@frim.gov.my*

Submitted May 2025; accepted August 2025

Assessing genetic diversity and population structure with multiple marker systems provides critical insights for the conservation and sustainable management of tropical tree species. We applied simple sequence repeats (SSRs) and genome-wide single nucleotide polymorphisms (SNPs) to characterise genetic variation in *Rubroshorea leprosula*, a widely distributed dipterocarp in Southeast Asia. SSRs revealed higher allelic richness and heterozygosity, consistent with their multi-allelic nature and high mutation rates, whereas SNPs produced more conservative but genome-wide estimates of diversity. SNP analyses indicated consistently negative inbreeding coefficients and relatively high nucleotide diversity (π = 0.008–0.010), suggesting substantial standing variation and strong adaptive potential. Population differentiation indicates a higher value with SSRs (F_{ST} = 0.061) than with SNPs (F_{ST} = 0.027), reflecting the ability of SSRs to detect finer-scale versus the genome-wide connectivity captured by SNPs. Both marker systems consistently identified two major genetic clusters aligned with a north–south division across Peninsular Malaysia, with admixture indicating historical or ongoing gene flow. A localised inbreeding signal was detected in the Belum–Temenggor complex, highlighting populations at elevated risk under fragmentation. Together, these results demonstrate the complementary strengths of SSRs for detecting rare alleles and local differentiation, and SNPs for capturing broad genomic patterns. Integrating both marker systems strengthens conservation planning by informing strategies for maintaining connectivity, and guiding conservation effort to safeguard the adaptive capacity and genetic potential of *R. leprosula*.

Keywords: Conservation, Dipterocarpaceae, habitat fragmentation, simple sequence repeats, single nucleotide polymorphisms

INTRODUCTION

Tropical lowland forests of Southeast Asia are among the most biological diverse ecosystems globally, yet they are also experiencing some of the highest rates of deforestation and degradation (Sodhi et al. 2004). Land-use change and agricultural expansion have driven extensive habitat loss and fragmentation, threatening the persistence of many forest tree species (Boonman et al. 2024). The dipterocarps (family Dipterocarpaceae) dominate these forests in terms of the ecological importance, and significant economic value as a source of high-quality tropical timber. Consequently, the decline of dipterocarp populations poses serious ecological and socio-economic challenges, highlighting the urgency of developing effective strategies for their conservation and sustainable management.

Rubroshorea leprosula (formerly *Shorea leprosula*) is a widely distributed dipterocarp species that inhabits lowland tropical forests across Peninsular Malaysia, Sumatra and Borneo (Symington et al 2004). It is ecologically significant as a canopy-forming species and economically valuable due to its high-quality timber. However, ongoing logging and land conversion activities are expected to substantially reduce the effective population size of *R. leprosula* and increase population fragmentation over time. Hence, maintaining

the genetic diversity of this species is therefore critical safeguarding its adaptive potential, long-term survival and sustainable utilisation (Ng et al. 2004, Ng et al. 2006, Ng et al. 2009, Lee et al. 2016). Understanding population genetic structure therefore provides a scientific foundation for conservation planning, particularly in the context of ongoing habitat loss and climate change.

Molecular markers are powerful tools for evaluating genetic variation and inferring population structure. Traditionally, simple sequence repeat (SSR) markers have been widely applied in tropical tree studies due to their high polymorphism, codominant inheritance and effectiveness in capturing fine-scale diversity (Tautz & Renz 1984, Blankenship et al. 2002, Abdelkrim et al. 2009). More recently, advances in sequencing technologies have enabled genome-wide single nucleotide polymorphism (SNP) discovery, providing high-resolution insights into the genetic structure, demographic history and adaptive variation (Davey et al. 2011). However, SSRs and SNPs differ in their allelic nature, multiallelic versus biallelic, respectively, which in turn influences their capacity to detect genetic diversity and differentiation. While SSRs often report higher levels of heterozygosity, SNPs provide broader genomic coverage (Filippi et al. 2015, Van Inghelandt et al. 2010, Zavinon et al. 2020). Comparative studies that employ both marker types can therefore offer complementary perspectives, yet such studies remain limited for dipterocarp species.

In this study, we investigated the genetic diversity and population structure of *R. leprosula* across 22 natural populations in Peninsular Malaysia using both SSRs and SNPs. Specifically, we asked: (a) how do estimates of genetic diversity differ between multiallelic SSR markers and biallelic SNP markers? (b) to what extent do population differentiation (F_{ST}) estimates derived from allele frequency and genetic distance approaches converge or diverge between marker systems? and (c) do both marker types reveal consistent patterns of population genetic structure? By integrating SSR and SNP data, we aim to demonstrate the value of a multi-marker approach to support the conservation and long-term sustainable management of *R. leprosula*.

MATERIALS AND METHODS

Sample collection and DNA extraction

We collected fresh leaf samples from 714 *R. leprosula* individuals across 22 natural populations throughout its range in Peninsular Malaysia, ensuring broad representation of the species' geographical areas (Figure 1, Table 1). Total genomic DNA was extracted using modified CTAB method (Murray & Thompson 1980) and further purified using High Pure PCR Template Preparation Kit (Roche Diagnostics, Germany). The integrity and quality of the DNA were evaluated on 1% agarose gel electrophoresis and the concentration of the DNA samples was quantified using NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA). All the 714 samples were genotyped using SSR markers. A subset of 111 samples (three to six individuals per population) was selected for whole genome resequencing.

SSRs analyses

Ten SSR markers, as described in Ng et al. (2022), were used to genotype all the 714 individuals of *R. leprosula*. Multiplex PCR amplification and genotyping for each SSR marker were conducted following established protocols, with an annealing temperature at 55 °C (Ng et al. 2022). Fragment analysis was conducted on ABI 3130xl capillary sequencer (Applied Biosystems, USA), with GeneScan 400HD ROX as the size standard (Applied Biosystems, USA). Subsequently, GeneMarker v2.6.4 software was used for alleles scoring.

SNPs analysis

A total of 111 high quality DNA samples from 22 populations were selected and outsourced for whole genome resequencing. The DNA libraries were sequenced on Illumina Hiseq and NovaSeq platform with 8 Gb paired-end short reads and target depth (~16×) per sample. The raw reads were evaluated with FastQC version 0.11.8 (Andrews 2010) and filtered using Trimmomatic v0.40 (Bolger et al. 2014) to remove low-quality reads and adapter sequences with default parameters. All clean reads were mapped to the *R. leprosula* chromosome-level genome assembly v2

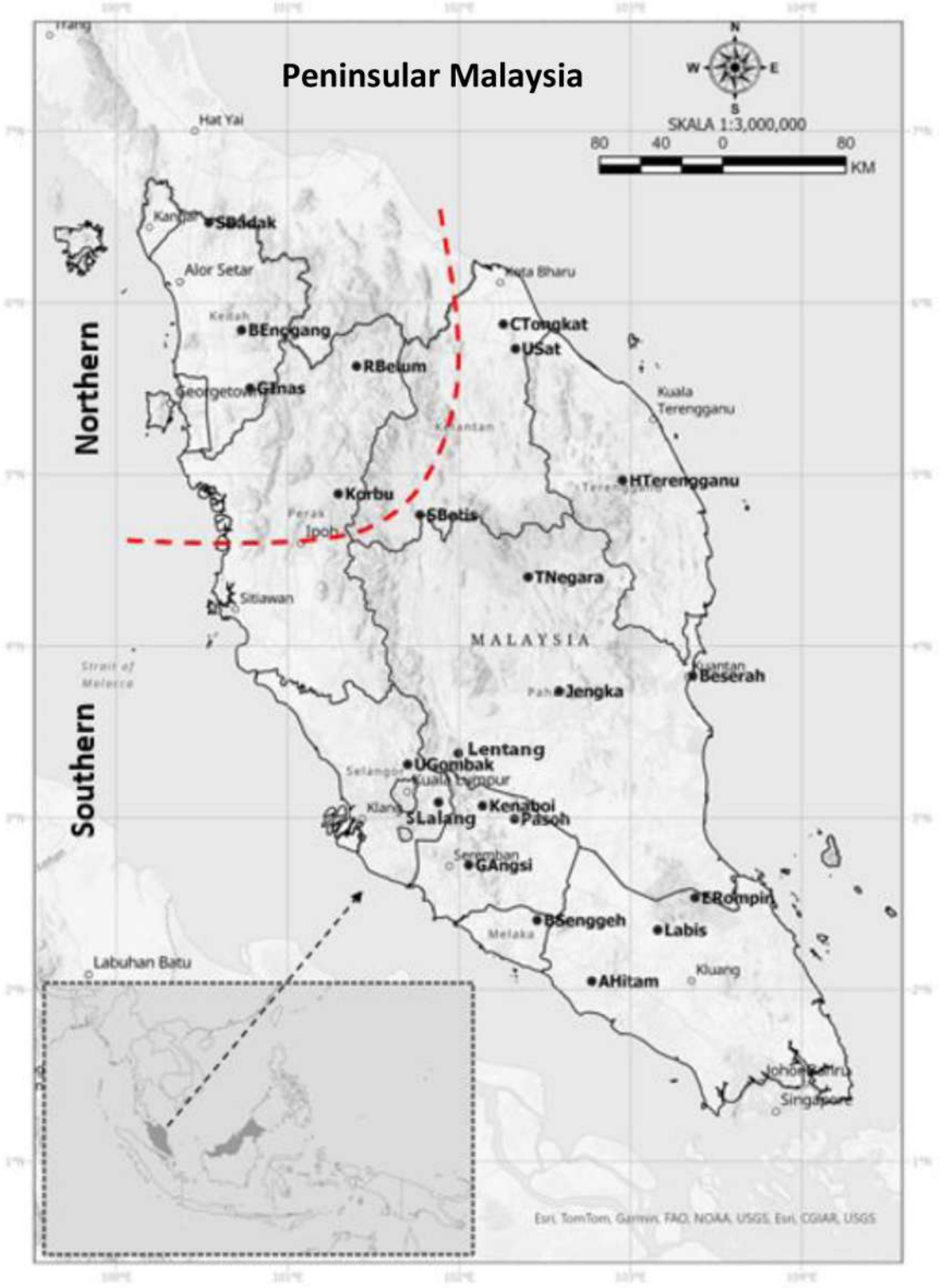

Figure 1 Geographical overview of 22 *Rubroshorea leprosula* populations sampled across Peninsular Malaysia. A red line indicating the division between the Northern and Southern regions

Table 1 Geographical areas and sample size for 22 natural populations of *Rubroshorea leprosula* in Peninsular Malaysia used for SSR and SNP analyses

Population	Abbreviation	State	Coordinates (Latitude, Longitude)	No. of samples (*N*)	
				SSR	SNP
Sungai Badak	SBadak	Kedah	6.46619, 100.53942	34	5
Bukit Enggang	BEnggang	Kedah	5.84055, 100.73063	35	6
Gunung Inas	GInas	Kedah	5.50258, 100.78000	27	5
Royal Belum	RBelum	Perak	5.63040, 101.40138	39	4
Korbu	Korbu	Perak	4.88693, 101.29481	35	5
Ulu Gombak	UGombak	Selangor	3.31140, 101.70010	30	5
Sungai Lalang	SLalang	Selangor	3.09083, 101.87953	23	6
Gunung Angsi	GAngsi	Negeri Sembilan	2.72516, 102.05612	32	6
Kenaboi	Kenaboi	Negeri Sembilan	3.07017, 102.13643	28	5
Pasoh	Pasoh	Negeri Sembilan	2.99364, 102.32272	39	5
Bukit Senggeh	BSenggeh	Melaka	2.40398, 102.45575	30	5
Sungai Betis	SBetis	Kelantan	4.76461, 101.77117	34	5
Ulu Sat	USat	Kelantan	5.73222, 102.32856	26	6
Chabang Tongkat	CTongkat	Kelantan	5.87543, 102.25806	32	5
Hulu Terengganu	HTerengganu	Terengganu	4.96598, 102.95417	32	6
Taman Negara	TNegara	Pahang	4.40228, 102.40273	39	3
Beserah	Beserah	Pahang	3.82756, 103.36250	30	5
Jengka	Jengka	Pahang	3.73798, 102.58297	34	5
Lentang	Lentang	Pahang	3.37562, 101.99458	33	6
Endau Rompin	ERompin	Johor - Pahang	2.53479, 103.37729	40	5
Labis	Labis	Johor	2.34681, 103.15914	27	3
Air Hitam	AHitam	Johor	2.04838, 102.77435	35	5

(Ng et al. unpublished data) using BWA-MEME with default parameters (Jung & Han 2022). Then, SAMtools v1.19 (Danecek et al. 2021) was used to convert the mapping results to BAM and sorted BAM format. Following this, Picard v3.1.1 (https://broadinstitute.github.io/picard/) was used to remove PCR duplicates. SNP calling was performed using 'bcftools mpileup' (Danecek et al. 2021). To minimize false positives in variant detection and retain only high-quality variants for downstream analysis, a hard filtration was applied to remove reads with the following criteria: (i) genotype quality < 20, (ii) coverage depth < 5, (iii) minor allele frequency < 0.01 and missing rates > 20%. Besides, multiallelic and monomorphic SNPs were also removed, retaining only biallelic SNPs for the subsequent analyses. The filtered SNPs were further phased and imputed using Beagle v5.4 (Browning et al. 2021). To minimize the effect of linkage disequilibrium (LD) on population structure inference, SNPs were subsequently LD-pruned using PLINK v1.90 (Purcell et al. 2007) with the parameter "– indep pairwise 50 10 0.2". The

resulting LD-pruned SNP dataset was specifically used for population structure analyses, whereas the unpruned dataset was retained for genetic diversity analyses to maximize genome-wide information (Schlḍtterer et al. 2014, Malomane et al. 2018).

Genetic diversity

To assess and compare genetic diversity by SSR and SNP markers, standard diversity indices were calculated for each of the 22 natural populations. These indices included the number of alleles (A), observed heterozygosity (H_O), expected heterozygosity (H_E), inbreeding coefficient (F_{IS}), and an additional nucleotide diversity (π) for genome-wide SNP marker.

For the SSR dataset, genetic diversity parameters and their corresponding standard deviations were calculated using Microsatellite Toolkit (Park 2008) and R packages such as *adegenet* (Jombart 2008) and *hierfstat* (Goudet 2005). The genepop-formatted dataset was imported using read.genepop() function in *adegenet* and summary statistics were obtained via basic.stats() function in *hierfstat*. Hardy-Weinberg Equilibrium (HWE) tests were conducted separately using the hw.test() function in *pegas* (Paradis 2010). F_{IS} values with $p < 0.05$ were considered significant deviations from HWE.

For the SNP dataset, genetic diversity was assessed using the unpruned dataset, in line with the dataset processing strategy described above. H_O, H_E as well as HWE statistics were computed using the – hardy function in PLINK v1.90, which is optimized for biallelic markers (Purcell et al. 2007). In addition, genome-wide nucleotide diversity (π) was calculated using Pixy (Korunes & Samuk 2021), which provides unbiased estimates of genetic diversity by accounting for missing data. Both variant and invariant sites were included in the analysis, with π calculated in non-overlapping 10 kb windows across the genome. The genome-wide mean π values were then obtained for each population. To measure genetic variability, standard deviations (SDs) for each parameter (H_O, H_E, F_{IS} and π) were calculated in R using the sd() function applied to per-locus values within each population. Statistical significance was determined using False Discovery Rate (FDR)-adjusted p-values based on the Benjamini-Hochberg method (Benjamini & Hochberg 1995), with a significance threshold of 0.05 to account for multiple testing and reduce the likelihood of false positives across the large number of loci. Pearson's correlation tests were conducted using the cor.test() function in R to assess the relationship between genetic diversity estimates derived from SSR and SNP markers.

Population genetic structure and differentiation

Population structure was inferred using both model-based clustering and multivariate approaches tailored to each marker type. For the SSR dataset, Bayesian clustering was performed using STRUCTURE v2.3.4 (Pritchard et al. 2000) under the admixture ancestry model with the LOCPRIOR option. Ten independent runs were conducted for each K value (1 to 6), with a burn-in of 100,000 iterations followed by 200,000 Markov Chain Monte Carlo (MCMC) steps. For the large SNP dataset, we applied ADMIXTURE v.1.3.0 (Alexander et al. 2009), a program that employs a maximum-likelihood approach similar to STRUCTURE but enables faster inference. The optimal number of genetic clusters (K) was determined using the Delta K method (Evanno et al. 2005) in STRUCTURE SELECTOR (Li & Liu 2018) for the SSR dataset and the cross-validation (CV) method (Alexander & Lange 2011) for the SNP dataset. Data from ten independent runs of STRUCTURE and ADMIXTURE analyses were graphically presented by CLUMPAK (Kopelman et al. 2015). In addition, to complement the Bayesian analysis, Unweighted Pair Group Method with Arithmetic Mean (UPGMA) trees were constructed to visualize genetic relationships among inferred clusters. For the SSR dataset, Nei's D_A genetic distance was calculated and used to construct a tree with POPTREE2 (Takezaki et al. 2014) with 1,000 bootstrap replicates. For the SNP dataset, a UPGMA tree was constructed using the *phangorn* package in R (Schliep 2011) and visualised in the Interactive Tree of Life (iTOL) online tool (Letunic & Bork 2021).

Overall and pairwise genetic differentiation was assessed among all populations and between the major genetic clusters identified by STRUCTURE and ADMIXTURE using two

approaches which are fixation index (F_{ST}) (Weir & Cockerham 1984) and Jost's *D* (Jost 2008). F_{ST} quantifies differentiation relative to total genetic variation but can be biased by high within-population heterozygosity, whereas Jost's *D* measures allelic differentiation independently of within-population diversity. Using both approaches therefore provides a more robust assessment and understanding of population genetic structure.

For SSR data, overall and pairwise estimates of F_{ST} were calculated in Genepop 4.7 (Raymond & Rousset 1995), while Jost's *D* was computed in R with the *mmod* package (Winter 2012). While for SNP data, pairwise F_{ST} was obtained using the –weir-fst-pop option in VCFtools and Jost's *D* was estimated with *mmod*. Statistical confidence for both approaches was evaluated by 1000 bootstrap replicates in R with *hierfstat* (Goudet 2005), from which standard deviations, 95% confidence intervals and *p*-values were derived. Results with $p < 0.001$ were considered statistically significant. The pairwise F_{ST} results were visualized as bar plots with error bars and histograms. To assess concordance between marker types, correlation between SSR- and SNP-derived F_{ST} matrices was evaluated using both Pearson's correlation and a Mantel test (10,000 permutations) implemented in the *vegan* package (Oksanen et al. 2025).

In addition, an analysis of molecular variance (AMOVA) test was conducted to quantify the distribution of genetic variation within and among populations of *R. leprosula*. Analysis was done according to Excoffier et al. (1992) with the function poppr.amova from *poppr* package (Kamvar et al. 2014) and the significance of the analysis was tested with 999 permutations using the randtest() function in the *ade4* package (Dray & Dufour 2007). Results with $p < 0.001$ were considered significant for both SSR and SNP datasets.

RESULTS

Genetic diversity

A total of 714 *R. leprosula* individuals from 22 populations were genotyped using 10 SSR markers, with sample sizes per population ranging from 23 (SLalang) to 40 (ERompin) (Table 1). All 10 SSR loci were polymorphic in the studied *R. leprosula* populations and harboured a mean number of alleles (*A*) ranging from 5.60 (CTongkat) to 8.90 (Pasoh) (Table 2). The mean H_E and H_O varied among populations, H_E ranged from 0.557 ± 0.063 (CTongkat) to 0.717 ± 0.045 (GAngsi), and H_O ranged from 0.578 ± 0.028 (CTongkat) to 0.689 ± 0.024 (Pasoh and TNegara) (Table 2). Negative F_{IS} values were observed in six populations (SBadak, CTongkat, TNegara, SLalang, ERompin and AHitam), but none deviated significantly from HWE, suggesting no strong evidence of inbreeding within these populations. In contrast, the RBelum population exhibited a statistically significant positive mean F_{IS} value ($F_{IS} = 0.144 \pm 0.094$, Table 2), indicating a heterozygote deficit consistent with recent or ongoing inbreeding.

In the SNP dataset, a total of 642,335 high-quality genome-wide SNPs (unpruned) were identified from a subset of 111 of the 714 *R. leprosula* individuals genotyped with SSRs. Sample size per population ranged from 3 (TNegara and Labis) to 6 individuals (BEnggang, SLalang, GAngsi, USat, HTerengganu and Lentang) (Table 1). SNP-derived diversity estimated for H_E values ranged from 0.224 ± 0.186 (GInas) to 0.260 ± 0.185 (SLalang; Table 2). However, the H_O was consistently higher across all populations (ranging from 0.328 ± 0.322 to 0.397 ± 0.342), resulting in uniformly negative F_{IS} values ranging from -0.371 ± 0.300 to -0.509 ± 0.340 (Table 2). Nevertheless, these negative F_{IS} values were not statistically significant after applying the Benjamini-Hochberg false discovery rate (FDR) correction at 0.05, indicating no strong deviation from HWE in SNP data. In addition, the high and consistent genome-wide average π values across *R. leprosula* populations ranging from 0.008 ± 0.007 to 0.010 ± 0.010 (Table 2) are considered evidence of substantial genetic variation and potentially higher heterozygosity.

Pearson correlation analysis was conducted to evaluate the concordance between genetic diversity estimates (H_E, H_O and F_{IS}) derived from SSR and SNP markers across populations of *R. leprosula* (Figure 2). The correlation between SSR-based H_E and SNP-based H_E was very low and not statistically significant ($r = 0.082$, $p = 0.716$, 95% CI = [-0.352, 0.487]) (Figure 2A). Similarly, H_O values yielded a weak, non-significant positive

Table 2 Summary of genetic diversity estimates and Hardy-Weinberg Equilibrium (HWE) tests of *Rubroshorea leprosula* based on SSR and SNP analyses

Population	SSR				SNP			
	A (±) SD	H_E (±) SD	H_O (±) SD	F_{IS} (±) SD	π (±) SD	H_E (±) SD	H_O (±) SD	F_{IS} (±) SD
SBadak	6.20 ± 2.82	0.626 ± 0.050	0.635 ± 0.026	-0.003 ± 0.078	0.009 ± 0.009	0.237 ± 0.196	0.364 ± 0.352	-0.437 ± 0.330
BEnggang	7.00 ± 2.87	0.644 ± 0.053	0.634 ± 0.026	0.006 ± 0.090	0.009 ± 0.008	0.259 ± 0.185	0.395 ± 0.341	-0.412 ± 0.324
GInas	7.20 ± 2.82	0.683 ± 0.037	0.659 ± 0.029	0.035 ± 0.140	0.009 ± 0.009	0.224 ± 0.186	0.328 ± 0.322	-0.371 ± 0.300
RBelum	7.40 ± 2.91	0.701 ± 0.043	0.609 ± 0.025	0.144 ± 0.094*	0.009 ± 0.008	0.241 ± 0.196	0.373 ± 0.354	-0.452 ± 0.331
Korbu	7.60 ± 3.41	0.646 ± 0.057	0.609 ± 0.026	0.039 ± 0.139	0.009 ± 0.009	0.253 ± 0.188	0.384 ± 0.342	-0.419 ± 0.323
SBetis	7.20 ± 2.90	0.669 ± 0.058	0.668 ± 0.026	0.016 ± 0.103	0.010 ± 0.009	0.246 ± 0.192	0.378 ± 0.349	-0.428 ± 0.330
USat	7.20 ± 2.70	0.641 ± 0.058	0.596 ± 0.030	0.078 ± 0.116	0.010 ± 0.008	0.237 ± 0.184	0.351 ± 0.326	-0.374 ± 0.307
CTongkat	5.60 ± 2.46	0.557 ± 0.063	0.578 ± 0.028	-0.030 ± 0.133	0.009 ± 0.009	0.238 ± 0.196	0.368 ± 0.356	-0.443 ± 0.334
HTerengganu	7.00 ± 2.40	0.658 ± 0.061	0.647 ± 0.027	0.003 ± 0.085	0.008 ± 0.007	0.227 ± 0.191	0.341 ± 0.337	-0.392 ± 0.318
TNegara	7.20 ± 2.70	0.678 ± 0.049	0.689 ± 0.024	-0.009 ± 0.119	0.009 ± 0.008	0.229 ± 0.204	0.361 ± 0.367	-0.499 ± 0.337
Beserah	6.30 ± 2.75	0.648 ± 0.074	0.627 ± 0.028	0.025 ± 0.069	0.010 ± 0.009	0.241 ± 0.193	0.369 ± 0.349	-0.429 ± 0.330
Jengka	8.00 ± 2.71	0.674 ± 0.063	0.665 ± 0.026	0.008 ± 0.088	0.010 ± 0.009	0.249 ± 0.187	0.377 ± 0.340	-0.411 ± 0.324
Lentang	6.50 ± 2.37	0.649 ± 0.068	0.636 ± 0.027	0.034 ± 0.090	0.009 ± 0.008	0.253 ± 0.187	0.386 ± 0.342	-0.413 ± 0.325
UGombak	6.80 ± 3.39	0.656 ± 0.053	0.653 ± 0.028	0.008 ± 0.138	0.009 ± 0.009	0.248 ± 0.190	0.380 ± 0.347	-0.427 ± 0.328
SLalang	7.20 ± 3.36	0.683 ± 0.061	0.670 ± 0.031	-0.003 ± 0.181	0.010 ± 0.009	0.260 ± 0.185	0.397 ± 0.342	-0.413 ± 0.326
GAngsi	7.90 ± 3.11	0.717 ± 0.045	0.662 ± 0.027	0.079 ± 0.137	0.008 ± 0.007	0.245 ± 0.196	0.381 ± 0.357	-0.445 ± 0.334
Kenaboi	7.80 ± 3.26	0.682 ± 0.052	0.618 ± 0.029	0.071 ± 0.146	0.010 ± 0.009	0.248 ± 0.190	0.378 ± 0.345	-0.424 ± 0.326
Pasoh	8.90 ± 2.88	0.696 ± 0.051	0.689 ± 0.024	0.010 ± 0.053	0.010 ± 0.010	0.252 ± 0.187	0.382 ± 0.341	-0.413 ± 0.326
BSenggeh	7.20 ± 2.97	0.681 ± 0.060	0.667 ± 0.027	0.017 ± 0.117	0.009 ± 0.009	0.244 ± 0.194	0.377 ± 0.354	-0.440 ± 0.333
ERompin	5.90 ± 2.38	0.610 ± 0.084	0.642 ± 0.024	-0.046 ± 0.079	0.010 ± 0.009	0.244 ± 0.190	0.373 ± 0.345	-0.421 ± 0.327
Labis	7.40 ± 2.76	0.677 ± 0.053	0.667 ± 0.029	0.038 ± 0.159	0.009 ± 0.009	0.240 ± 0.204	0.380 ± 0.371	-0.509 ± 0.340
AHitam	7.50 ± 3.03	0.659 ± 0.068	0.686 ± 0.025	-0.013 ± 0.167	0.010 ± 0.009	0.244 ± 0.192	0.373 ± 0.347	-0.426 ± 0.328

A = mean of number of alleles; H_E = mean of expected heterozygosity; H_O = mean of observed heterozygosity; F_{IS} = mean of inbreeding coefficient; π = nucleotide diversity; SD = standard deviation; * = populations with a significant proportion of SSRs deviating from HWE ($p < 0.05$)

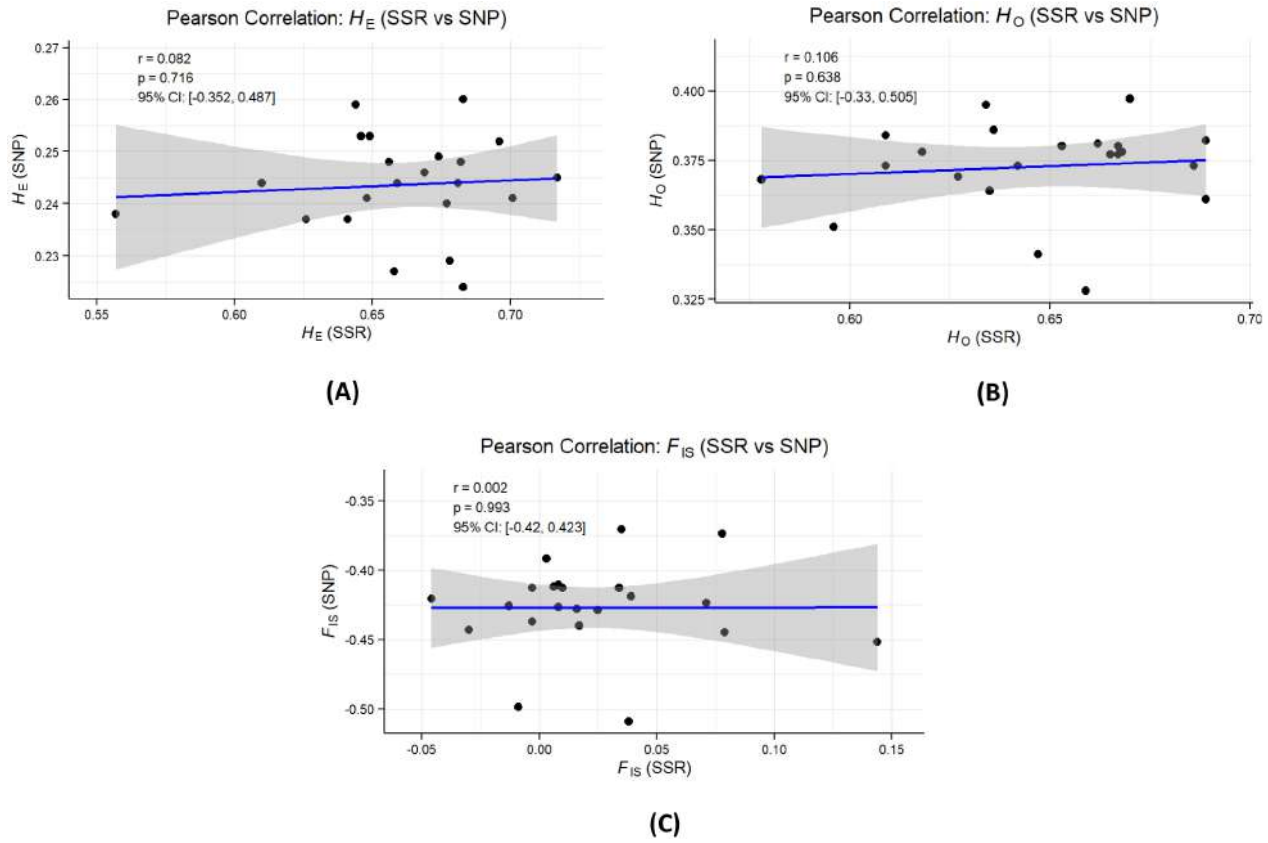

Figure 2 Pearson correlation analyses between genetic parameters estimated using SSR and SNP markers across 22 populations of *Rubroshorea leprosula*. (A) Expected heterozygosity (H_E), (B) Observed heterozygosity (H_O) and (C) Inbreeding coefficient (F_{IS}). Each black point represents a population. The blue line shows the fitted linear trend with the grey shading showing the 95% confidence interval

correlation between SSR and SNP datasets (r = 0.106, p = 0.638, 95% CI = [-0.330, 0.505]) (Figure 2B). The correlation for F_{IS} was also weak and not significant (r = 0.002, p = 0.993, 95% CI = [-0.420, 0.423]) (Figure 2C). Collectively, these findings indicate poor concordance between SSR- and SNP-derived estimates of heterozygosity and inbreeding (H_E, H_O and F_{IS}), suggesting that the two marker systems may capture different aspects of the underlying genetic structure in *R. leprosula* populations.

Population genetic structure and differentiation

Both clustering analysis (STRUCTURE and ADMIXTURE) produced the highest likelihood scores for both markers when the number of populations was set at K = 2, dividing 22 populations in Peninsular Malaysia into two main genetic clusters; Cluster A and Cluster B. Cluster A consists of five northern populations, namely SBadak, BEnggang, GInas, RBelum and Korbu, while Cluster B consists of the remaining 17 populations, namely UGombak, SLalang, GAngsi, Kenaboi, Pasoh, BSenggeh, TNegara, SBetis, USat, CTongkat, HTerengganu, Beserah, Jengka, Lentang, ERompin, Labis and AHitam (Figure 3A and Figure 3B). Notably, several populations showed clear signs of admixture in both datasets, with more pronounced patterns observed in the SSR analysis.

Further sub-clustering within Cluster B was detected in both SSR and SNP datasets, resulting in two sub-clusters; Sub-cluster B1 and Sub-cluster B2 (Figure 3), with K = 2 being the optimal value for both Delta K (SSR) and minimum CV error value (SNP). Using 266,270 SNPs (LD-pruned), the SNP dataset provided higher resolution than SSRs, grouping individuals from USat, CTongkat and HTerengganu into Sub-cluster B1 (Figure 3A),

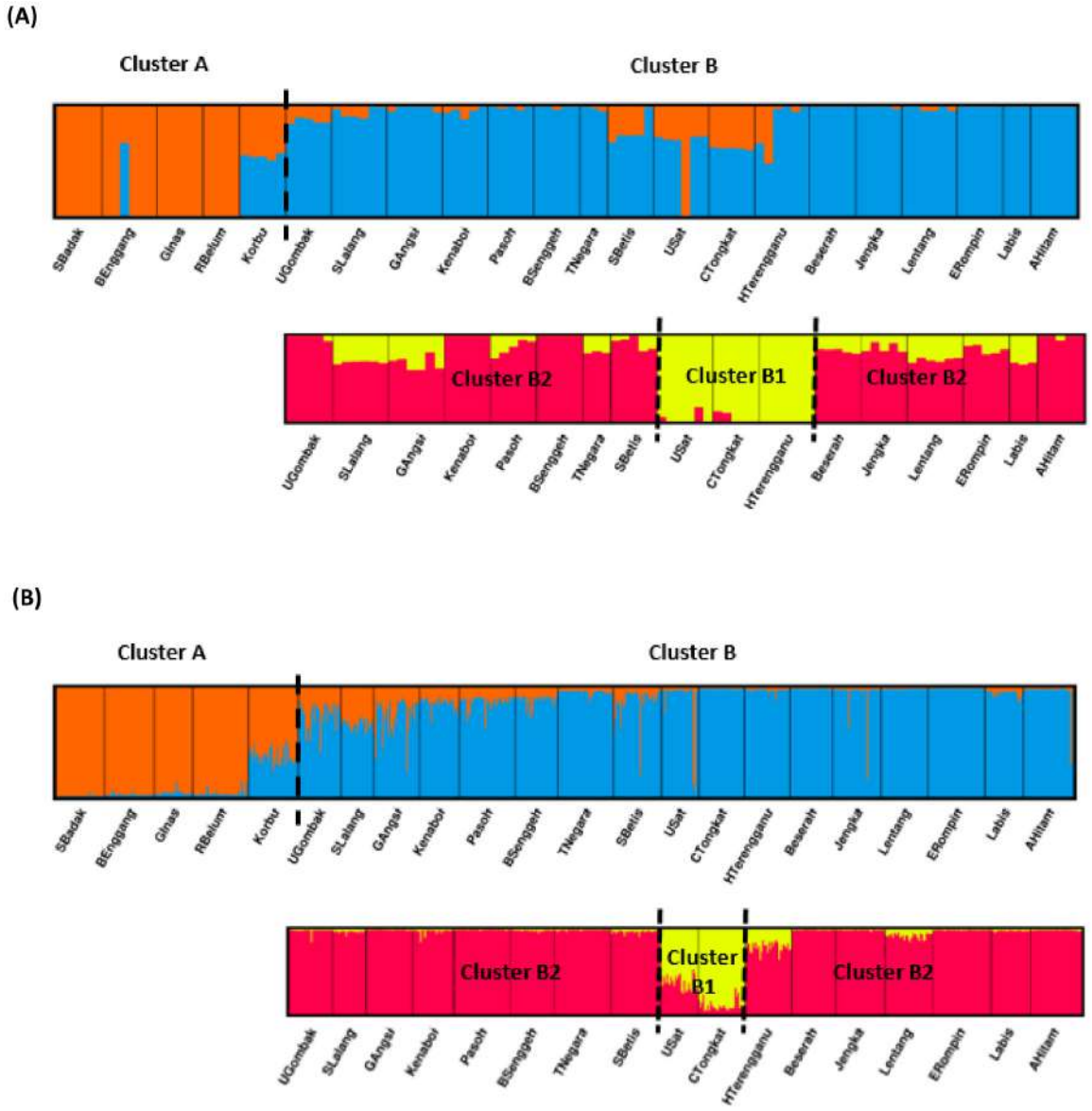

Figure 3 Population structure analysis of *Rubroshorea leprosula* inferred using different marker systems. (A) ADMIXTURE analysis of 111 individuals based on 266,270 LD-pruned SNPs and (B) STRUCTURE analysis of 714 individuals genotyped with 10 SSR markers. Each column represents a population. Dashed lines indicate main cluster and sub-cluster boundaries identified by each marker system

whereas SSRs grouped only USat and CTongkat into this sub-cluster (Figure 3B).

In addition, clustering analysis by UPGMA was performed to ratify the groupings. The result showed that the phylogenetic trees derived from both marker systems were in agreement with the findings of the population structure analysis (K = 2) (Figure 4A and Figure 4B). This clustering indicated that the populations can be divided into two major distinct clusters (Cluster A and Cluster B), suggesting substantial genetic differentiation between them. Despite a clear separation between Cluster A and Cluster B across both marker types, some individuals and populations were assigned differently, particularly in the SNP dataset. However, the consistency between the UPGMA trees and the population structure analysis supports the validity of these results.

Overall population genetic differentiation for F_{ST} was 0.061 for SSRs and 0.027 for SNPs (Table 3, Figure 5A). Pairwise F_{ST} comparisons were subsequently estimated between major population clusters using both marker types

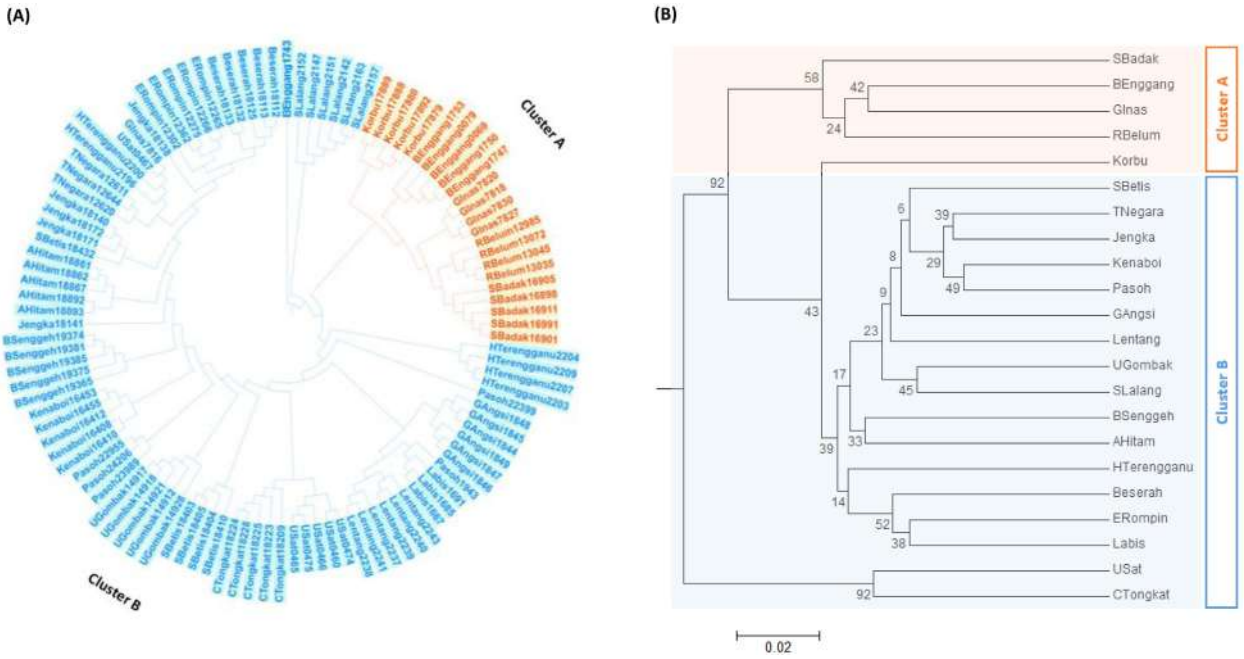

Figure 4 UPGMA clustering of *Rubroshorea leprosula*. (A) 111 accessions based on 266,270 LD-pruned SNPs and (B) 714 accessions based on 10 SSRs. Colours correspond to ADMIXTURE and STRUCTURE assignments assuming $K = 2$

Table 3 Estimates of genetic differentiation (F_{ST} and Jost's D) using SSR and SNP markers in *Rubroshorea leprosula*

Comparison	Statistic	SSR		SNP	
		Mean ± SD	95% CI	Mean ± SD	95% CI
Cluster A vs B	F_{ST}	0.031 ± 0.008	0.018 - 0.049	0.011 ± 7.06e-05	0.011 - 0.011
	Jost's D	0.072 ± 0.048	0.016 - 0.204	0.001 ± 2.73e-05	0.003 - 0.003
Overall	F_{ST}	0.061 ± 0.006	0.051 - 0.073	0.027 ± 9.67e-05	0.027 - 0.028
	Jost's D	0.122 ± 0.036	0.087 - 0.229	0.025 ± 3.69e-05	0.025 - 0.025

Values are presented as mean ± SD with 95% confidence intervals (1000 bootstrap replicates).
All comparisons are significant ($p < 0.001$). SD = standard deviation; CI = confidence interval.

(Table 3, Figure 5A). Pairwise F_{ST} estimates between clusters (Cluster A vs. Cluster B) for SSR and SNP was 0.031 ± 0.008 and 0.011 ± 7.06e-05, respectively.

To complement the F_{ST} results, which may underestimate divergence in cases of high within-population diversity, we further examined allelic differentiation with Jost's D. For SSR markers, the overall Jost's D was 0.122 ± 0.036 (95% CI: 0.087–0.229; $p < 0.001$), indicating moderate allelic differentiation across populations. In contrast, the SNP dataset yielded a considerably lower overall Jost's D of 0.025 ± 3.69e-05 (95% CI: 0.025–0.025; $p < 0.001$) (Table 3, Figure 5B). The absolute allelic differentiation by SNP markers was consistently lower than SSRs, with Cluster A vs. Cluster B showing minimal divergence (Jost's $D = 0.001 \pm 2.73e\text{-}05$; $p < 0.001$) (Table 3, Figure 5B).

Population genetic differentiation analysis on variance component based AMOVA revealed consistent patterns across both marker systems

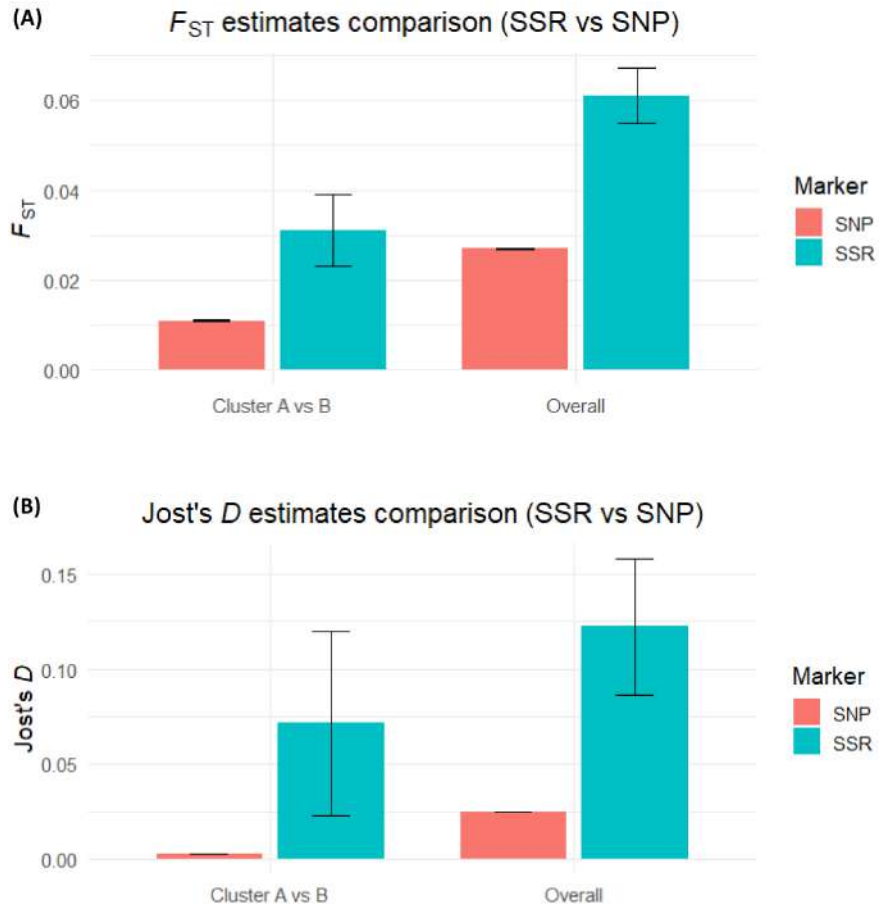

Figure 5 Comparison of genetic differentiation (F_{ST} and Jost's *D*) between SSR and SNP markers in *Rubroshorea leprosula*. (A) F_{ST} and (B) Jost's *D*. Bars represent mean values with error bars indicating ± SD. All comparisons are significant ($p < 0.001$) and SD = standard deviation

at the population level (Table 4). In the SSR dataset, the majority of genetic variance was found within populations (93.94%), and 6.07% among populations (Table 4). The overall genetic differentiation among populations has a ΦST value = 0.061. Similarly, the SNP dataset showed 93.47% of the variation occurred within populations and 6.53% among populations, with a relatively similar ΦST value of 0.065 (Table 4). These results indicate that most genetic variation in *R. leprosula* resides within populations, reflecting the species' outcrossing nature and high within-population diversity.

The histogram patterns of SNP-based pairwise F_{ST} and Jost's *D* values were unimodal, with most F_{ST} values concentrated below 0.02 (mean = 0.016; Figure 6A) and most Jost's *D* values below 0.030 (mean = 0.029; Figure 7A), reflecting a narrow range of low genetic differentiation among most population pairs. In comparison, the histogram of SSR-based pairwise F_{ST} and Jost's *D* values showed a higher mean value of 0.06 (Figure 6B) and 0.122 (Figure 7B) respectively. Majority of SSR-based pairwise F_{ST} comparisons fell below the 0.05 threshold, indicating that most population pairs exhibited low genetic differentiation. However, a considerable proportion of comparisons approached and exceeded 0.10, with some reaching the moderate differentiation threshold of 0.15. For Jost's *D*, most values ranged between 0.05 and 0.15, with the highest frequency around 0.10–0.12, suggesting overall low to moderate differentiation.

The Mantel test, which assesses the correlation between distance matrices, revealed a moderate and statistically significant positive correlation (F_{ST}: $r = 0.431$, $p = 0.008$), suggesting overall concordance in genetic distance patterns between SSR and SNP markers (Figure 6C). In

Table 4 AMOVA results for *Rubroshorea leprosula* based on SSRs and SNPs dataset. The AMOVA partitions genetic variation among and within populations

	SSR						SNP					
Source of Variation	Df	SS	Mean Square	Variance Component (σ^2)	Variation (%)	Phi Statistic	Df	SS	Mean Square	Variance Component (σ^2)	Variation (%)	Phi Statistic
Among populations	21	723.4052	34.4479	0.4269	6.07	ΦST = 0.061	21	274809.6	13086.171	676.3783	6.53	ΦST = 0.065
Within populations (individuals)	1406	9292.5123	6.6093	6.6116	93.94		89	861330.8	9677.874	9677.8743	93.47	
Total	1427	10,015.917	-	7.0385	100.00		110	1136140.4	-	10354.2525	100.00	

Df = degree of freedom; SS = sum of squares
All results are statistically significant ($p < 0.001$)

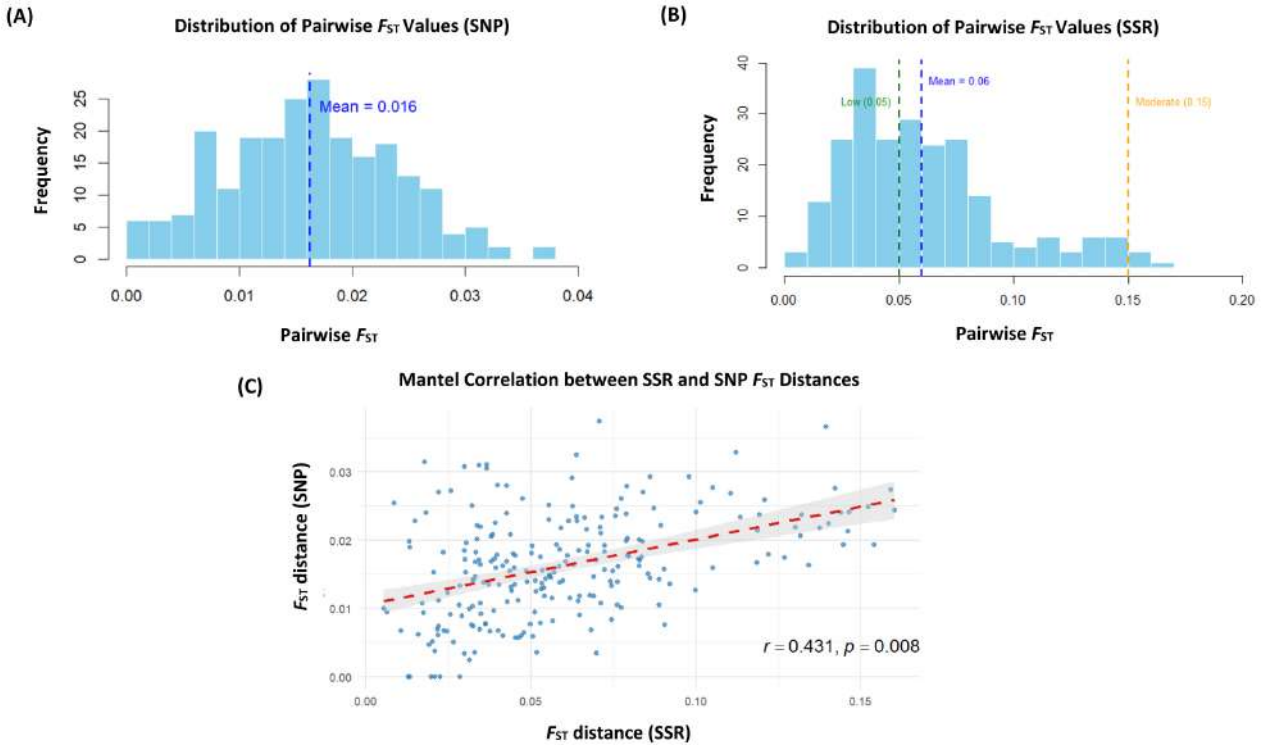

Figure 6 Distributions and correlations of pairwise F_{ST} values based on SNP and SSR markers. (A–B) Histograms show a unimodal distribution for SNPs (mean F_{ST} = 0.016) and a broader spread for SSRs (mean F_{ST} = 0.06), with blue dashed lines indicating mean values. (C) Correlation analyses between pairwise SNP- and SSR-based F_{ST} values, with a significant positive Mantel correlation (r = 0.431, p = 0.008). Regression lines indicate the direction and strength of associations.

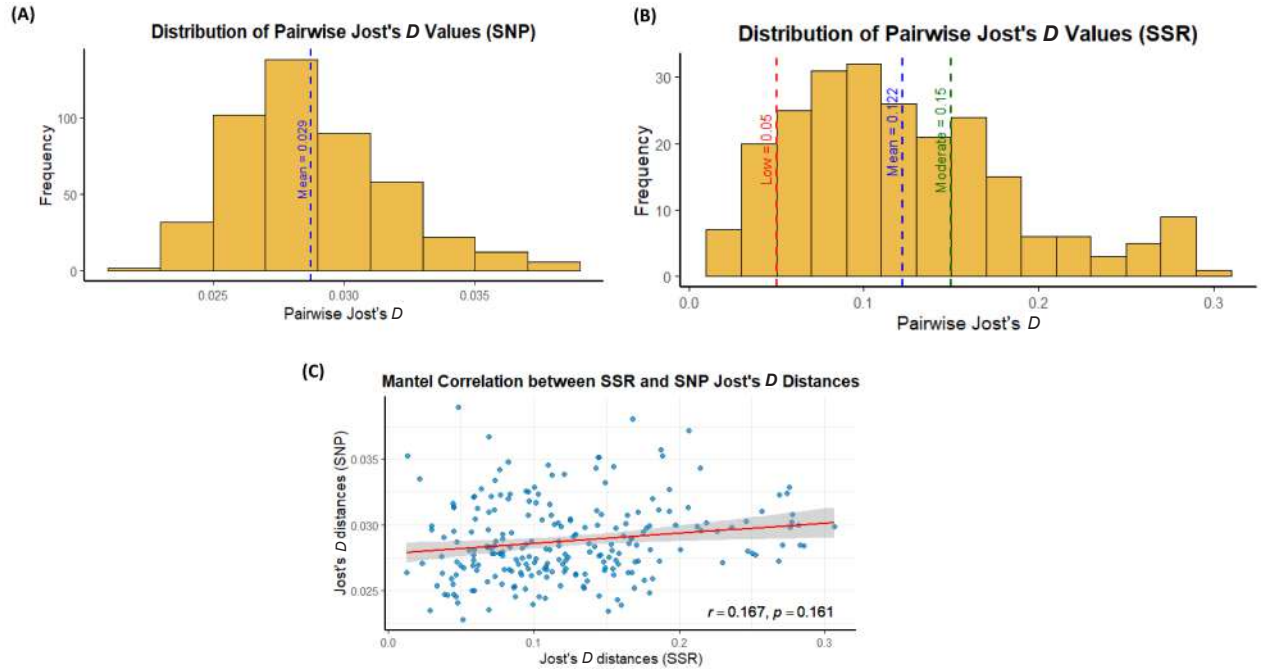

Figure 7 Distributions and correlations of pairwise Jost's *D* values based on SNP and SSR markers. (A–B) Histograms show a unimodal distribution for SNPs (mean Jost's *D* = 0.029) and a broader spread for SSRs (mean Jost's *D* = 0.122), with blue dashed lines indicating mean values. (C) Scatterplot showing the correlation between SNP- and SSR-based Jost's *D* values. A regression line indicates the direction and strength of the relationship, with a positive but weak Mantel correlation (r = 0.167, p = 0.161)

contrast, Jost's *D* showed a weaker and non-significant correlation (r = 0.167, p = 0.161), suggesting limited consistency between marker types in capturing differentiation (Figure 7C).

DISCUSSION

This study highlights the contrasting but complementary insights provided by SSRs and genome-wide SNPs in assessing genetic diversity of *R. leprosula*. SSRs revealed higher mean number of alleles and mean heterozygosity indices (A, H_O and H_E), these are consistent with reports of other dipterocarps (Ng et al. 2004, Ng et al. 2006, Ng et al. 2019, Ng et al. 2022, Ng et al. 2024, Lee et al. 2006), reflecting their high mutation rates and multi-allelic nature (Selkoe & Toonen 2006, Putman & Carbone 2014). In contrast, SNPs-based estimates provided a complementary but more conservative perspective on genetic diversity. Genome-wide SNP expected heterozygosity was relatively lower (H_E = 0.224 to 0.260) compared with SSR estimates, reflecting the biallelic nature and lower mutation rate of SNPs relative to SSRs. Observed heterozygosity (H_O = 0.328 to 0.397) was consistently higher than expected, leading to uniformly negative F_{IS} values. Although these deviations were not statistically significant after FDR correction, the consistent trend suggests that SNP datasets may capture genome-wide signals of outcrossing and balancing forces that help maintain heterozygosity. In addition, the estimates of nucleotide diversity (π = 0.008 to 0.010) for *R. leprosula* were relatively high, consistent with previous report (Ng et al. 2021). This reflects high standing genetic variation among *R. leprosula* populations, suggesting that the species retains considerable evolutionary potential. Such genetic variation provides a buffer against future environmental change and emphasizes the importance of preserving population connectivity to sustain adaptive capacity. The nucleotide diversity was found to be homogeneous across populations, with only marginal differences between sites, implying a relatively even distribution of standing genetic variation at the genomic scale. The narrow variance in nucleotide diversity further suggests that no single *R. leprosula* population holds a

disproportionately large share of the species' genomic diversity, which has implications for conservation strategies (Petit et al. 1998, Hoban & Schlarbaum 2014, Willoughby et al. 2015).

Correlation analyses revealed weak and non-significant associations between SSR- and SNP-derived estimates of H_E, H_O and F_{IS}, confirming that the two marker systems capture different dimensions of diversity. This lack of concordance is expected, given the high mutation rate, multi-allelic nature and locus-specific informativeness of SSRs, which make them particularly sensitive to detecting rare alleles and local demographic shifts (Hauser et al. 2011, Oliveira et al. 2006, Fischer et al. 2017). Conversely, SNPs provide broad genome-wide resolution but with reduced per-locus variability. Such marker-dependent differences have been documented in other long-lived tree taxa, like *Quercus* (Reutimann et al. 2020), suggesting the importance of integrating marker systems rather than relying on a single data type to infer population genetic processes.

The differences between F_{ST} and Jost's D estimates across markers further suggest the importance of methodological context in conservation genomics. Despite differences in absolute values of population differentiation (SSR F_{ST} = 0.061 vs SNP F_{ST} = 0.027), both datasets showed that the majority of variation resided within populations, consistent with the outcrossing reproductive system of *R. leprosula* (Lee et al. 2000, Ng et al. 2004, Ng et al. 2006, Dick et al. 2008, Crawford et al. 2012). The higher Jost's D for SSRs indicates greater allelic turnover per locus, whereas SNPs, by averaging across hundreds of thousands of loci, revealed lower absolute divergence but provided greater precision in delineating subtle structure not fully resolved by SSRs (Jost 2008, Meirmans & Hedrick 2011, Morin et al. 2004, Allendorf et al. 2010). These results suggest that while populations are not strongly differentiated, they retain measurable divergence that may represent local adaptation or historical isolation. The relatively low SNPs F_{ST} values imply substantial gene flow among populations, which is advantageous for maintaining genetic connectivity and reducing risks of inbreeding depression. However, the population differentiation detected by SSRs highlights that regional genetic variants could be at risk of excessive divergence if populations experience demographic decline, habitat fragmentation or overexploitation. Similar patterns have been reported in forest trees, where SNPs often capture genome-wide connectivity while SSRs detect finer-scale due to higher mutation rates and multi-allelic nature (Ellegren 2004, Ouborg et al. 2010, Wang et al. 2018, Rossetto et al. 2019). These findings underscore the complementary value of using multiple marker systems in assessing genetic structure for conservation management.

Both SSRs and SNPs consistently revealed a genetic partitioning of *R. leprosula* into two main clusters (K = 2) across Peninsular Malaysia, suggesting a history of population divergence likely influenced by geographic barriers such as the Titiwangsa mountain range, combined with limited dispersal and gene flow. Sub-structuring within the southern cluster (B1 and B2) further indicates finer-scale differentiation that may reflect localized demographic histories or restricted connectivity among populations. The presence of admixed individuals suggests that gene flow has occurred historically and may still be ongoing, with important implications for the distribution and maintenance of genetic diversity. Admixed populations can function as reservoirs of allelic variants and facilitate connectivity across the landscape, whereas non-admixed populations harbour unique allelic combinations that contribute to the overall genomic variation of the species. Notably, the low SNP-based F_{ST} and high levels of allele sharing imply that genome-wide connectivity remains substantial, underscoring that conservation measures should focus on preserving and enhancing gene flow, rather than enforcing strict separation of clusters (Allendorf et al. 2013).

The significantly positive F_{IS} detected in the RBelum population suggests localized inbreeding, potentially reflecting restricted gene flow and mating among related individuals. This pattern is consistent with the geographic isolation imposed by the construction of the Temenggor Dam in 1974, which fragmented the formerly continuous Belum-Temenggor rainforest into discrete forest "islands" on higher ground. Such anthropogenic fragmentation may reduce effective population size, increases genetic drift, and elevates the probability of biparental inbreeding. While the signal of inbreeding is

inferred from molecular data, direct validation would require controlled breeding experiments in which progeny of known relatedness are assessed for fitness effects across environmental conditions (Naito et al. 2005). If persistent, inbreeding may reduce heterozygosity, limit adaptive potential, and increase extinction risk through inbreeding depression. Consequently, the RBelum population should be prioritized for detailed genomic monitoring and considered for proactive interventions, including genetic rescue to counteract ongoing loss of genetic diversity (Frankham et al. 2017).

The complementary strengths of SSRs and SNPs support their combined use in conservation planning. SSRs remain highly effective for individual assignment, parentage, and long-term monitoring where historical datasets already exist (Guichoux et al. 2011, Lee et al. 2006, Selkoe & Toonen, 2006), while SNPs are better suited for genome-wide analyses of diversity, demographic inference, and detection of adaptive variation (Allendorf et al. 2010, Harrisson et al. 2014, Ng et al. 2021).

In practical terms, movement of reproductive material should generally remain within clusters to minimise maladaptation risk, though mixing across nearby sub-clusters may be justified where local populations are depleted (Lee et al. 2017, Weeks et al. 2011). Future studies may consider expanding SNP sampling across distribution range populations and integrating environmental data will further refine these recommendations, ensuring that conservation strategies for *R. leprosula* safeguard both its adaptive potential and genetic potential.

DATA ARCHIVING STATEMENT

Raw reads and genome assembly for version 1.0 have been deposited to DDBJ under BioProject accession numbers PRJDB8161, PRJDB8182 and for version 2.0 have been deposited to NCBI under BioSample accession number SAMN46626940 and BioProject accession number PRJNA1223480.

ACKNOWLEDGEMENTS

This research was funded by the Government of Malaysia through the 11th Malaysia Plan (2016–2020) and the 12th Malaysia Plan (2021–2025). We extend our gratitude to the State Forest Departments of Kedah, Perak, Kelantan, Terengganu, Pahang, Selangor, Negeri Sembilan, Melaka and Johor, the Department of Wildlife and National Parks, Royal Belum State Park and the Johor National Parks Corporation for their permission and logistical support during sample collection in the forests. Our heartfelt thanks go to Ghazali Jaafar, Yahya Marhani, Ramli Ponyoh, Yasri Baya, Ahmad Farhan Nuri Rasli, Mohamad Izham Mohd Ariffin and Sharifah Talib for their outstanding assistance in the field and laboratory. Special appreciation is given to the Malaysia Genome and Vaccine Institute (MGVI), particularly Mohd Faizal Abu Bakar for his dedicated technical support.

REFERENCES

Abdelkrim J, Robertson BC, Stanton JAL & Gemmel NJ. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques* 46: 185–192. https://doi.org/10.2144/000113084.

Alexander DH & Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* 12: 246. https://doi.org/10.1186/1471-2105-12-246.

Alexander DH, Novembre J & Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655–1664. 10.1101/gr.094052.109.

Allendorf FW, Hohenlohe PA & Luikart G. 2010. Genomics and the future of conservation genetics. *Nature Reviews Genetics* 11: 697–709.

Allendorf FW, Luikart G & Aitken SN. 2013. *Conservation and the Genetics of Populations*. 2nd ed. Wiley-Blackwell, Oxford.

Andrews S. 2010. FastQC: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Benjamini Y & Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57: 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Blankenship SM, May B & Hedgecock D. 2002. Evolution of a perfect simple sequence repeat locus in the context of its flanking sequence. *Molecular Biology and Evolution* 19: 1943–1951. https://doi.org/10.1093/oxfordjournals.molbev.a004018.

Bolger AM, Lohse M & Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Boonman CCF, Serra-Diaz JM, Hoeks S et al. 2024. More than 17,000 tree species are at risk from rapid global change. *Nature Communications* 15(166). https://doi.org/10.1038/s41467-023-44321-9.

Browning BL, Tian X, Zhou Y & Browning SR. 2021. Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics* 108: 1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005.

Crawford NG, Faircloth BC, McCormack, et al. 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biology Letters* 8: 783–786.

Danecek P, Bonfield JK, Liddle J et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10: giab008. https://doi.org/10.1093/gigascience/giab008.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM & Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12: 499–510. https://doi.org/10.1038/nrg3012.

Dick CW, Hardy OJ, Jones FA & Petit RJ. 2008. Spatial scales of pollen and seed-mediated gene flow in tropical rain forest trees. *Tropical Plant Biology* 1: 20–33.

Dray S & Dufour AB. 2007. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* 22: 1–20. https://doi.org/10.18637/jss.v022.i04.

Ellegren H. 2004. Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics* 5: 435–445. https://doi.org/10.1038/nrg1348.

Evanno G, Regnaut S & Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology* 14: 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

Excoffier L, Smouse PE & Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to genetic structure studies in plant and tree populations. *Genetics* 131: 479–491. https://doi.org/10.1093/genetics/131.2.479.

Filippi CV, Aguirre N, Rivas JG et al. 2015. Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. *BMC Plant Biology* 15: 52. https://doi.org/10.1186/s12870-014-0360-x.

Fischer MC, Rellstab C, Leuzinger M et al. 2017. Estimating genomic diversity and population differentiation - an empirical comparison of microsatellite and SNP variation of *Arabidopsis halleri*. *BMC Genomics* 18: 69. https://doi.org/10.1186/s12864-016-3459-7.

Frankham R, Ballou JD, Ralls K et al. 2017. *Genetic Management of Fragmented Animal and Plant Populations*. Oxford University Press, Oxford.

Goudet J. 2005. hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* 5: 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x.

Guichoux E, Lagache L, Wagner S et al. 2011. Current trends in microsatellite genotyping. *Molecular Ecology Resources* 11: 591–611.

Harrisson KA, Pavlova A, Telonis-Scott M & Sunnucks P. 2014. Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications* 7: 1008–1025.

Hauser L, Baird M, Hilborn R et al. 2011. An empirical comparison of SNP and microsatellite variation in two sockeye salmon populations. *Molecular Ecology Resources* 11: 150–168.

Hoban S & Schlarbaum S. 2014. Optimal sampling of seeds from plant populations for ex-situ conservation of genetic biodiversity, considering realistic population structure. *Biological Conservation* 177: 90–99. https://doi.org/10.1016/j.biocon.2014.06.014

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405. https://doi.org/10.1093/bioinformatics/btn129.

Jost L. 2008. G_{ST} and its relatives do not measure differentiation. *Molecular Ecology* 17: 4015–26. https://doi.org/10.1111/j.1365-294X.2008.03887.x.

Jung Y & Han D. 2022. BWA-MEME: BWA-MEM emulated with a machine learning approach. *Bioinformatics* 38: 2404–2413. https://doi.org/10.1093/bioinformatics/btac137.

Kamvar ZN, Tabima JF & Grünwald NJ. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2: e281. https://doi.org/10.7717/peerj.281.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA & Mayrose I. 2015. Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* 15: 1179–1191. https://doi.org/10.1111/1755-0998.12387.

Korunes KL & Samuk K. 2021. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources* 22: 1228–1229. https://doi.org/10.1111/1755-0998.13326.

Lee CT, Tnah LH, Ng KKS et al. 2017. *Plant material transfer guidelines for tropical forest species*. Research Pamphlet No. 137, Forest Research Institute Malaysia, Kepong, Selangor.

Lee SL, Ng KKS, Ng CH, Tnah LH, Lee CT, Tani N & Tsumura Y. 2016. Spatial studies of *Shorea parvifolia* (Dipterpcarpaceae) in a lowland and hill dipterocarp forest. *Journal of Tropical Forest Science* 28(Special Issue): 309–317.

Lee SL, Ng KKS, Saw LG et al. 2006. Linking the gaps between conservation research and conservation management of rare dipterocarps: A case study of *Shorea lumutensis*. *Biological Conservation* 131: 72–92.

Lee SL, Wickneswari R, Mahani MC & Zakri AH. 2000. Mating system parameters in a tropical tree species, *Shorea leprosula* Miq. (Dipterocarpaceae), from Malaysian lowland dipterocarp forest. *Biotropica* 32: 693–702. https://doi.org/10.1111/j.1744-7429.2000.tb00517.x.

Letunic I & Bork P. 2021. Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* 49(W1): W293–W296. https://doi.org/10.1093/nar/gkab301.

Li YL & Liu JX. 2018. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular*

Ecology Resources 18: 176–177. https://doi.org/10.1111/1755-0998.12719.

Malomane DK, Reimer C, Weigend S, Weigend A, Sharifi AR & Simianer H. 2018. Efficiency of different strategies to mitigate ascertainment bias when using SNP panels in diversity studies. *BMC Genomics* 19: 22. https://doi.org/10.1186/s12864-017-4416-9.

Meirmans PG & Hedrick PW. 2011. Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources* 11: 5–18.

Morin PA, Luikart G & Wayne RK. 2004. SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution* 19: 208–216.

Murray M & Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* 8: 4321–4326. https://doi.org/10.1093/nar/8.19.4321.

Naito Y, Konuma A, Iwata H et al. 2005. Selfing and inbreeding depression in seeds and seedlings of *Neobalanocarpus heimii* (Dipterocarpaceae). *Journal of Plant Research* 118: 423–430. https://doi.org/10.1007/s10265-005-0245-z.

Ng CH, Lee SL, Tnah LH, Ng KK, Lee CT, Diway B & Khoo E. 2019. Genetic diversity and demographic history of an upper hill dipterocarp (*Shorea platyclados*): implications for conservation. *Journal of Heredity* 110: 844–856. https://doi.org/10.1093/jhered/esz052.

Ng CH, Ng KKS, Lee SL, et al. 2024. Genetic diversity and population structure of *Rubroshorea curtisii*, a hill dipterocarp species: implications for conservation. *Tree Genetics & Genomes* 20: 16. https://doi.org/10.1007/s11295-024-01649-7.

Ng CH, Ng KKS, Lee SL, Nurul-Farhanah Z, Lee CT & Tnah LH. 2022. DNA databases of an important tropical timber tree species *Shorea leprosula* (Dipterocarpaceae) for forensic timber identification. *Scientific Reports* 12: 9546. https://doi.org/10.1038/s41598-022-13697-x.

Ng KKS, Kobayashi MJ, Fawcett JA, et al. 2021. The genome of *Shorea leprosula* (Dipterocarpaceae) highlights the ecological relevance of drought in aseasonal tropical rainforests. *Communications Biology* 4: 1166. https://doi.org/10.1038/s42003-021-02682-1.

Ng KKS, Lee SL & Koh CL. 2004. Spatial structure and genetic diversity of two tropical tree species with contrasting breeding systems and different ploidy levels. *Molecular Ecology* 13: 657–669. https://doi.org/10.1046/j.1365-294X.2004.02094.x.

Ng KKS, Lee SL, Saw LG, Plotkin JB & Koh CL. 2006. Spatial structure and genetic diversity of three tropical tree species with different habitat preferences within a natural forest. *Tree Genetics & Genomes* 2: 121–131. https://doi.org/10.1007/s11295-006-0035-3.

Ng KKS, Lee SL & Ueno S. 2009. Impact of selective logging on genetic diversity of two tropical tree species with contrasting breeding systems using direct comparison and simulation methods. *Forest Ecology and Management* 257: 107–116. https://doi.org/10.1016/j.foreco.2008.08.035.

Oksanen J, Simpson G, Blanchet F et al. 2025. *vegan*: Community Ecology Package. R package version 2.8-0. https://vegandevs.github.io/vegan/.

Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R & Vieira MLC. 2006. Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology* 29: 294–307. https://doi.org/10.1590/S1415-47572006000200018.

Ouborg NJ, Pertoldi C, Loeschcke V et al. 2010. Conservation genetics in transition to conservation genomics. *Trends in Genetics* 26: 177–187. https://doi.org/10.1016/j.tig.2010.01.001.

Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26: 419–420. https://doi.org/10.1093/bioinformatics/btp696.

Park SDE. 2008. Excel Microsatellite Toolkit. http://www.animalgenomics.ucd.ie/sdepark/ms-toolkit/.

Petit RJ, El Mousadik A & Pons O. 1998. Identifying populations for conservation on the basis of genetic markers. *Genetics* 149: 967–981. https://doi.org/10.1093/genetics/149.2.967.

Pritchard JK, Stephens M & Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959. https://doi.org/10.1093/genetics/155.2.945.

Purcell S, Neale B, Todd-Brown K et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81: 559–575. 10.1086/519795.

Putman AI & Carbone I. 2014. Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution* 4: 4399–4428.

Raymond M & Rousset F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248–249. https://doi.org/10.1093/oxfordjournals.jhered.a111573.

Reutimann O, Gugerli F & Rellstab C. 2020. A species-discriminatory single-nucleotide polymorphism set reveals maintenance of species integrity in hybridizing European white oaks (*Quercus* spp.) despite high levels of admixture. *Annals of Botany* 125: 663–676, https://doi.org/10.1093/aob/mcaa001.

Rossetto M, Bragg J, McPherson H, & Wilson PD. 2019. From SSRs to SNPs: Trend towards multi-allelic molecular markers in conservation genetics. *Conservation Genetics Resources* 11: 243–255. https://doi.org/10.1007/s12686-018-1044-2.

Schliep KP. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* 27: 592–593. https://doi.org/10.1093/bioinformatics/btq706.

Schlötterer C, Tobler R, Kofler R & Nolte V. 2014. Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics* 15: 749–763. https://doi.org/10.1038/nrg3803.

Selkoe KA & Toonen RJ. 2006. Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. *Ecology Letters* 9: 615–629.

Sodhi NS, Koh LP, Brook BW & Ng PKL. 2004. Southeast Asian biodiversity: an impending disaster. *Trends in Ecology & Evolution* 19: 654–660. https://doi.org/10.1016/j.tree.2004.09.006.

Symington CF. 2004. *Foresters' manual of dipterocarps*: Forest Research Institute Malaysia & Malaysian Nature Society, Kepong.

Takezaki N, Nei M & Tamura K. 2014. POPTREEW: Web version of POPTREE for constructing population trees from allele frequency data and computing some other quantities. *Molecular Biology and Evolution* 31: 1622–1624. https://doi.org/10.1093/molbev/msu093.

Tautz D & Renz M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research* 12: 4127–4138. https://doi.org/10.1093/nar/12.10.4127.

Van Inghelandt D, Melchinger AE, Lebreton C & Stich B. 2010. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theoretical and Applied Genetics* 120: 1289–1299. https://doi.org/10.1007/s00122-009-1256-2.

Wang H, McArthur ED, Sanderson SC, et al. 2018. Narrow genetic base in forest tree populations despite high gene flow: Implications for conservation. *Heredity* 120: 70–80. https://doi.org/10.1038/s41437-017-0002-6.

Weeks AR, Sgro CM, Young AG et al. 2011. Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evolutionary Applications* 4: 709–725.

Weir BS & Cockerham CC. 1984. Estimating F-Statistics for the analysis of population structure. *Evolution* 38: 1358–1370. https://doi.org/10.2307/2408641.

Willoughby JR, Harder AM, Tennessen JA et al. 2015. Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. *Ecology Letters* 18: 772–779. https://doi.org/10.1111/ele.12461

Winter DJ. 2012. MMOD: an R library for the calculation of population differentiation statistics. *Molecular Ecology Resources* 12: 1158–1160. https://doi.org/10.1111/j.1755-0998.2012.03174.x.

Zavinon F, Adoukonou-Sagbadja H, Keilwagen J, Lehnert H, Ordon F & Perovic D. 2020. Genetic diversity and population structure in Beninese pigeon pea [*Cajanus cajan* (L.) Huth] landraces collection revealed by SSR and genome wide SNP markers. *Genetic Resources and Crop Evolution* 67: 191–208. https://doi.org/10.1007/s10722-019-00864-9.