HYPER-SEQ SNP GENOTYPING REVEALS CRITICALLY LOW GENETIC DIVERSITY IN THE ENDANGERED TREE *MADHUCA PASQUIERI* (SAPOTACEAE)

JIANG, H. L.[#] – ZHONG, W. T.[#] – JIANG, Y. S. – CHEN, Z. Y. – YANG, Y. S. – GAO, L. M.[*] – ZOU, R.[*]

Guangxi Key Laboratory of Plant Functional Phytochemicals and Sustainable Utilization, Guangxi Institute of Botany, Guangxi Zhuang Autonomous Region and Chinese Academy of Sciences, Guilin, Guangxi 541006, China

[#]*These authors contributed equally.*

[*]*Corresponding authors*
e-mail: zr@gxib.cn, glm@gxib.cn; phone: +86-135-9732-0111

(Received 30th Jul 2025; accepted 14th Nov 2025)

Abstract. Wild populations of *Madhuca pasquieri*, a nationally protected tree, have become highly fragmented in Guangxi, China. To establish a genomic baseline for conservation, we genotyped 35 individuals from four populations using Hyper-seq genotyping. Hyper-seq genotyping is a low-cost, high-throughput approach that constructs barcoded, reduced-representation libraries in a single PCR, enabling flexible marker density without restriction digestion. Across populations, mean observed heterozygosity (H_O = 0.4318), expected heterozygosity (H_E = 0.2548) and nucleotide diversity (π = 0.4423) fall within the range documented for fragmented subtropical tree species. Tajima's D was positive in three populations, consistent with recent contraction and small effective size. Population-pair F_{ST} averaged 0.062, pointing to moderate genetic differentiation that was unrelated to geographic distance. ADMIXTURE and PCA resolved two genetic clusters; TXLJZ and JXCDX showed comparatively higher within-population diversity. These genomic baselines reveal a species already experiencing heterozygote surplus and signals of demographic contraction. These results provide a genomic baseline and actionable guidance for monitoring and conservation planning in *Madhuca pasquieri*.
Keywords: *karst landscape, RAD-type SNPs, population structure, conservation units, isolation-by-resistance, genetic monitoring*

Introduction

Madhuca pasquieri (Sapotaceae) is a perennial evergreen tree—also known as "Dian zijing wood," "milkwood," "Tiese," and "peanut wood." Its high-quality timber is used for machinery, boats, vehicles, bridges, and sporting goods. Seeds are valued for high oil content, underscoring considerable economic and scientific value. The species occurs mainly in southwestern Guangdong, southern Guangxi, and southeastern Yunnan, China, and in Vietnam (Kan et al., 2021). Owing to its narrow range, specialized habitat, and weak natural regeneration, *Madhuca pasquieri (M. pasquieri)* is endangered and listed as a National Class II protected wild plant in China.

Genetic differences between individuals of the same species and between species are the genetic diversity of the species (Hewitt, 2000). However, research on *M. pasquieri* has emphasized morphology (Cai et al., 2018), seed propagation (Li et al., 2014), and trunk resolution (Lin et al., 2018), with no SNP-based population study quantifying diversity or evaluating how karst landforms shape gene flow. Filling this gap is essential to delineate conservation units and set management priorities.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2401_449460

We therefore adopted Hyper-seq, a low-cost, flexible, and high-throughput library preparation and genotyping approach that generates abundant, genome-wide SNPs with robust performance in non-model plants (Zou and Xia, 2022). Hyper-seq has been successfully applied to population structure inference and DNA fingerprinting in wild and cultivated species, and its predictive accuracy can approach that of whole-genome resequencing while enabling highly multiplexed, single-PCR library construction—features advantageous for endangered trees in fragmented karst habitats.

Our objectives were to (i) quantify within- and among-population genetic diversity, (ii) resolve population structure and test whether karst barriers more strongly constrain gene flow than geographic distance, and (iii) identify conservation units and management priorities for *Madhuca pasquieri*.

Materials and methods

Plant materials and sampling

In 2023, we sampled 35 mature trees from four wild populations of *M. pasquieri* in Guangxi, China: CXSBZ (n = 7), JXCDX (n = 11), RXSSZ (n = 2), and TXLJZ (n = 15). Fresh leaves were silica-dried; locality data are listed in *Table 1*.

Table 1. *Detailed information of 4 Communities of M. pasquieri*

Name	Origin	Population	Sample size	Longitude (E)	Latitude (N)	Altitude (m)
M. pasquieri	Sanbao Town, Cenxi City	CXSBZ	7	110°57'28.4"	23°1'28.9"	103
	Changdong Township, Jinxiu County	JXCDX	11	110°6'50"	23°59'11.76"	889
	Songshan Town, Rong County	RXSSZ	2	110°30'20.8"	24°59'7.07"	127
	Lingjing Town, Teng County	TXLJZ	15	110°37'22.4"	23°8'15.36"	124

DNA extraction and HYPER-SEQ library preparation

Genomic DNA (>30 ng μL^{-1}; $A_{260/280}$ = 1.8-2.0) was isolated via modified CTAB. Hyper-seq libraries (150–200 ng μL^{-1} DNA) were barcoded, size-selected and PCR-amplified, then paired-end sequenced (2 × 150 bp) on Illumina NovaSeq 6000.

Read processing, SNP calling and filtering

Raw paired-end reads (Illumina NovaSeq 6000, 2×150 bp) were demultiplexed to single-sample files and quality-filtered with fastp v0.23.4, including adapter removal, poly-G/ poly-X tail trimming, exclusion of reads with excessive ambiguous bases, and base-quality filtering (Chen et al., 2018). Variant discovery used Stacks v2.6.4 following the standard RAD/RRL workflow (ustacks → cstacks → sstacks → tsv2bam → gstacks) to produce a multi-sample VCF (Rochette et al., 2019).

For population-genetic analyses, variants were filtered to: limit within-locus linkage by retaining a single SNP per RAD locus; exclude rare variants below a minor-allele frequency of 0.02 (Linck and Battey, 2019); remove sites with insufficient mean sequencing depth across individuals; and require a minimum site call rate across samples. These criteria balance marker number against genotyping reliability and reduce inflation of diversity and differentiation estimates due to linkage or spurious rare alleles. For

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2401_449460

model-based clustering, we further prepared an LD-reduced subset from the filtered panel to ensure approximately independent markers for ancestry inference. Sequencing quality metrics were evaluated for each individual; detailed per-sample values are provided in *Table S1*.

Data analysis

Genetic diversity metrics, including nucleotide diversity (π), number of private alleles (N_p), expected heterozygosity (H_E), observed heterozygosity (H_O), and inbreeding coefficient (F_{IS}), were calculated for each population using the population module of Stacks. Tajima's D values were calculated using Vcftools v0.1.13 (Danecek et al., 2011), and a neutrality test was performed on the SNPs obtained with 95% confidence intervals, with the sliding window size of the operation set to 3000 bp.

Pairwise genetic differentiation and absolute sequence divergence were calculated from the filtered SNPs. Where appropriate, unbiased estimators robust to missing data were used to minimize downward bias in diversity and differentiation. Pairwise genetic differentiation and absolute sequence divergence among populations were estimated from the filtered SNPs using unbiased estimators robust to missing data, as implemented in pixy v1.2.7 (Korunes and Samuk, 2021).

Ancestry clustering was inferred with ADMIXTURE v1.3.0 (Alexander et al., 2009) using an LD-reduced SNP subset derived from the filtered panel. The parameters were set to K values = 2-20, and the optimal number of clusters was determined using the minimum CV error.

Combine the genetic structure and geographic distribution (latitude and longitude) information of populations to create an interactive map. Use the geosphere analysis package to perform Mantel tests, and conduct distance isolation analysis (IBD) tests for Euclidean geographic distances and Edwards' genetic distances, with the relevant calculations and plotting completed using R statistical software.

By transforming a set of variables with correlation into linearly uncorrelated variables, the transformed linearly uncorrelated variables are the principal components. In this study, VCF2PCACluster software (Patterson et al., 2006) was used to obtain the clustering of the samples based on the SNP data by performing principal component analysis. TASSEL v5.0 (Bradbury et al., 2007) was used which can be used to calculate the kinship between two individuals of a natural population.

Results

Within-population genetic diversity

Genetic-diversity parameters for the four *M. pasquieri* populations are summarized in *Table 2*. Using the polymorphic SNP panel (one SNP per locus), populations harbored on average 15,105 private alleles (N_p range 7,971–21,172). Mean diversity metrics were H_O = 0.4138 (range 0.3854–0.4518), H_E = 0.2548 (0.2040–0.2985), and π = 0.4423 (0.3947–0.4877). In all populations H_O exceeded H_E, and Weir–Cockerham F_{IS} values were small and positive. Because RXSSZ had n = 2, Tajima's D was not estimated for this population; the remaining three populations showed positive Tajima's D (TXLJZ = 1.7718, JXCDX = 1.7150, CXSBZ = 1.6103), indicating an excess of intermediate-frequency variants.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2401_449460

Table 2. *Indicators of genetic diversity of populations*

Population	N_p	H_O	H_E	π	F_{IS}	Tajima's D
CXSBZ	13,584	0.4026	0.2463	0.4306	0.0494	1.6103
JXCDX	17,693	0.4156	0.2702	0.4565	0.0764	1.7150
RXSSZ	7,971	0.3854	0.2040	0.3947	0.0140	-
TXLJZ	21,172	0.4518	0.2985	0.4877	0.0776	1.7718
Mean	15,105	0.4138	0.2548	0.4423	0.0543	1.6990

Note: Calculation of Tajima's D values requires a minimum of 4 samples from the population

Between-population differentiation and isolation by distance

Estimation of inter-population F_{ST} can reveal the extent of population genetic variation and distinguish the magnitude of relative genetic variation among populations. In this study, PIXY v1.2.7 (Korunes and Samuk, 2021) was used to calculate the coefficient of genetic differentiation (F_{ST}) and the degree of nucleotide disambiguation (D_{XY}) between the two populations (*Fig. 1*).

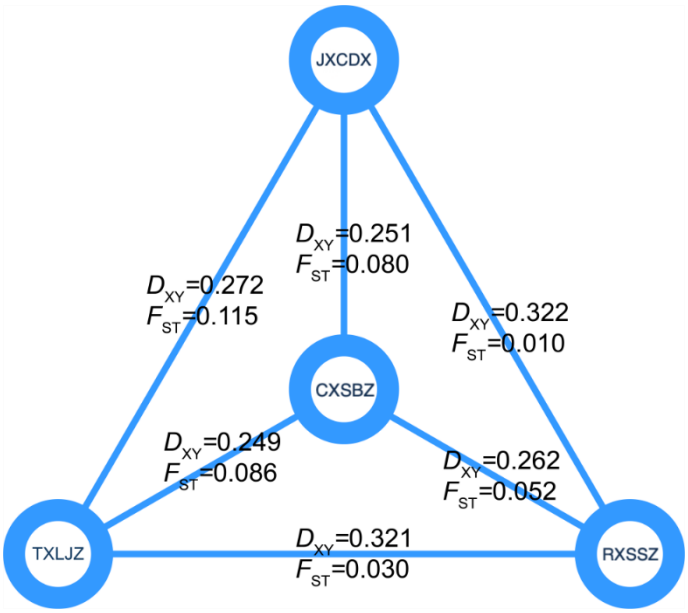

Figure 1. *Pairwise genetic differentiation (F_{ST}) and absolute divergence (D_{XY}) among four populations of M. pasquieri, estimated from the 56,847-SNP dataset. TXLJZ is the most differentiated from other populations, whereas RXSSZ is the least; rankings are concordant across F_{ST} and D_{XY}*

The results showed that the genetic differentiation coefficients among the four *M. pasquieri* populations were in the range of 0.010 (JXCDX-RXSSZ)-0.115 (JXCDX-TXLJZ), with a grand mean of 0.062. The TXLJZ population showed the most differentiated from the other populations (F_{ST} = 0.030-0.115), whereas RXSSZ showed the smallest (F_{ST} = 0.010- 0.052); estimates involving RXSSZ are uncertain due to n=2.

Isolation-by-distance (IBD) was not supported: The Mantel correlation between Edwards' genetic distance and geographic distance was r = 0.0286, P = 0.583(*Fig. 2*).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2401_449460

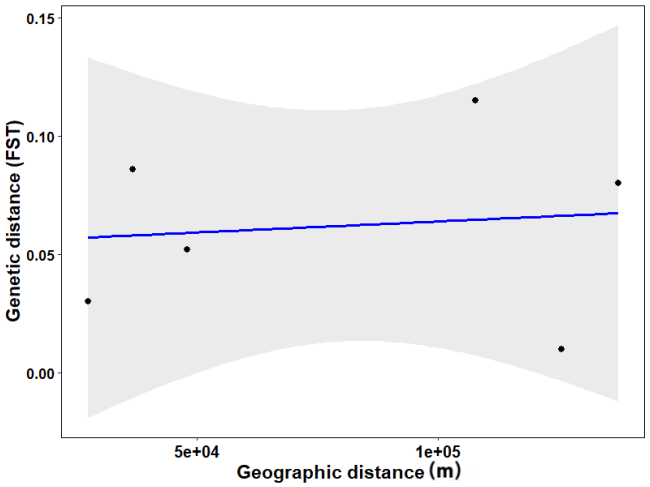

Figure 2. *Isolation-by-distance test: linearized F_{ST} vs geographic distance among population pairs (m). Mantel r = 0.0286, P = 0.583*

Population structure

ADMIXTURE analyses on the LD-reduced panel (730 SNPs) indicated a clear primary split at K = 2, where individuals partitioned into two ancestry components (*Figs. 3,4*). Additional, finer substructure appeared at higher K; although the minimum cross-validation error occurred at K = 19, we interpret K = 2 as the major axis of structure and use higher-K results descriptively (*Fig. 3*).

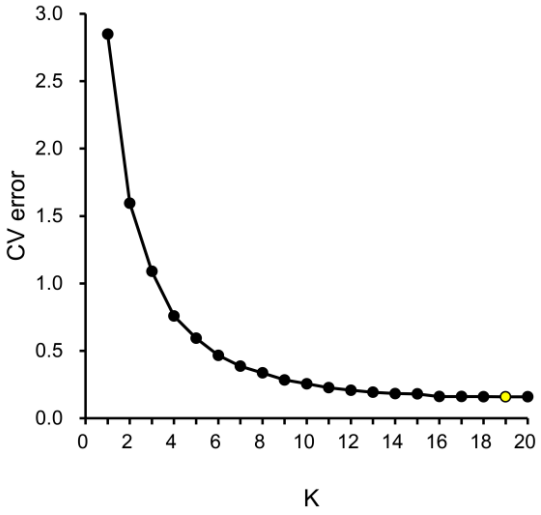

Figure 3. *ADMIXTURE cross-validation error across K for the 730-SNP LD-reduced panel. The primary split occurs at K = 2; the CV minimum appears at high K=19*

Principal component analysis (PCA)

According to the principal component analysis (*Fig. 5*), it can be seen that the contribution rates of PCA1 and PCA2 are 47.2% and 32.7%, respectively, with a cumulative contribution rate of 79.9%, which indicates that these two principal components reflect most of the information of the original data better.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2401_449460

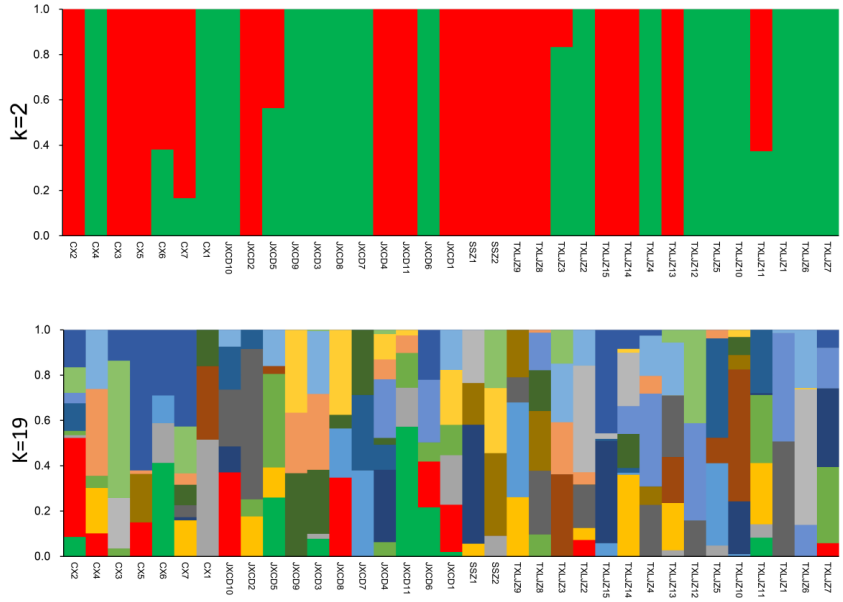

Figure 4. *ADMIXTURE ancestries inferred from the 730-SNP LD-reduced panel. At K = 2, individuals split into two main components; TXLJZ is comparatively cohesive, whereas other populations show admixture. Samples are ordered by population*

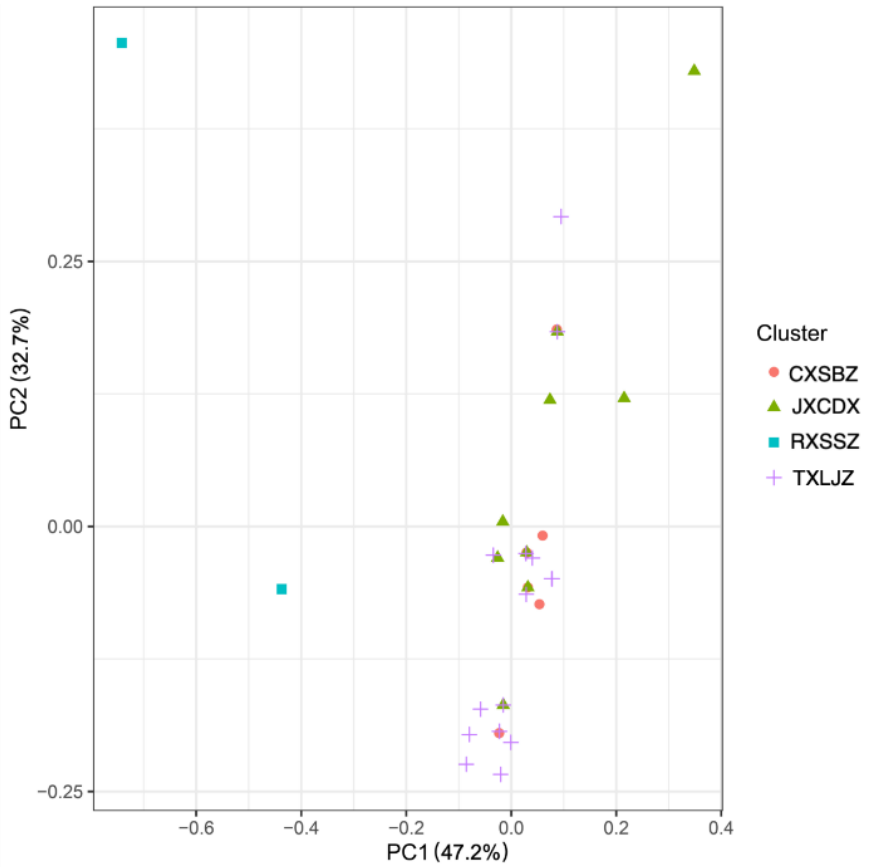

Figure 5. *PCA based on the 730-SNP LD-reduced panel. PC1 = 47.2% and PC2 = 32.7% (cumulative 79.9%). The ordination recapitulates the K=2 split; TXLJZ clusters tightly, whereas CXSBZ and JXCDX partially overlap*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2401_449460

It can be seen that except for the TXLJZ population, which was mostly clustered in the fourth quadrant, the individuals of the other populations were loosely distributed, indicating that except for TXLJZ individuals, which tend to cluster toward one quadrant, samples from the other populations show substantial overlap in the PCA space. This pattern is consistent with the primary split recovered by ADMIXTURE at K = 2, indicating discernible structure but overall moderate among-population separation; the scatter also likely reflects individual-level admixture and the small sample size for RXSSZ.

Kinship

The kinship relationships among the 35 samples were inferred from the SNP dataset and are illustrated in *Fig. 6*, where darker shading denotes a closer relationship between two individuals.

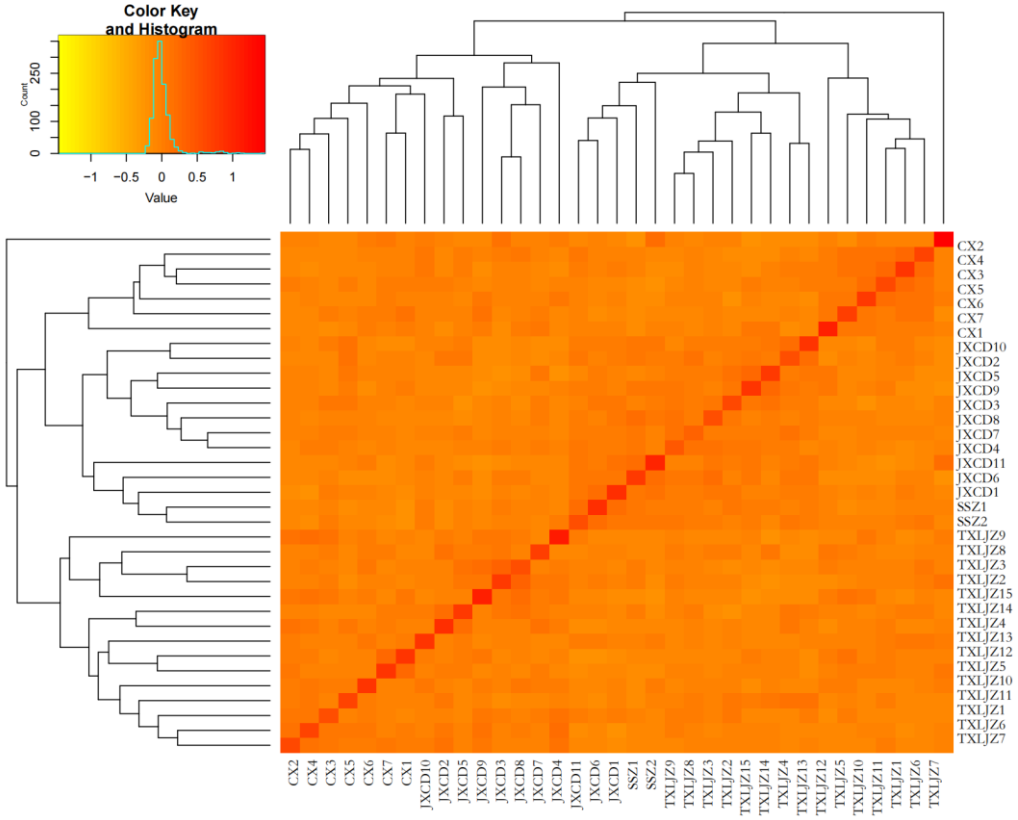

Figure 6. *Pairwise relatedness (IBS) among 35 individuals computed from the 730-SNP LD-reduced panel. Warmer colors indicate higher relatedness. Clustering is stronger within populations and uniformly low across populations; the closest pair is JXCDX11–JXCDX7*

Overall, the four populations—CXSBZ, JXCDX, RXSSZ and TXLJZ—form distinct clusters, highlighting their genetic separation. The closest pair is JXCDX11 and JXCDX7, with a kinship coefficient of 0.303. Apart from such within-population pairs, the heat-map shows uniformly low kinship values across populations, confirming that the four populations are only distantly related.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2401_449460

Discussion

Genetic diversity of M. pasquieri

Across populations, our polymorphic SNP panel indicates H_O = 0.3854–0.4518, H_E = 0.2040–0.2985, π = 0.3947–0.4877, and small positive Weir–Cockerham F_{IS} (0.014–0.078). For instance, the island shrub–tree *Viburnum japonicum* shows H_O = 0.181–0.248, H_E = 0.213–0.290, π = 0.229–0.304 across four populations. the critically endangered conifer *Glyptostrobus pensilis* has H_O = 0.086, H_E = 0.035, π = 0.072. and the long-lived Cryptomeria *japonica var. sinensis* exhibits H_O = 0.143 (Cai et al., 2020). Compared with these, *M. pasquieri* retains comparable expected heterozygosity but shows a higher Ho (heterozygote excess), consistent with our positive Tajima's D and expectations for recently bottlenecked or small populations where allelic richness erodes faster than heterozygosity. For context among subtropical trees sampled in fragmented forests, *Camellia tachangensis* reports Ho = 0.08–0.09, π = 0.22–0.23 with positive Tajima's D, again indicating modest diversity under human disturbance and habitat edges (Huang et al., 2024).

Ecologically, these cross-system contrasts align with range size, dispersal, and habitat connectivity. Compared with other endangered trees, *M. pasquieri* shows expected heterozygosity within the typical range for such systems, while its higher H_O than H_E, together with small positive Weir–Cockerham F_{IS} and positive Tajima's D, points to retention of heterozygosity alongside erosion of rare alleles—this pattern is consistent with retention of heterozygosity alongside erosion of rare alleles, as expected under recent contraction or population subdivision; nonetheless, balancing selection at a subset of loci cannot be excluded.

Genetic structure and differentiation

The impact of the genetic structure and genetic diversity of a population depends on its population size, range, growing environment, reproduction mode, and biological adaptability, which can have an impact on species formation, environmental resilience, and evolutionary potential (Stronen et al., 2022). Observed heterozygosity exceeded expected heterozygosity (Ho > H_E) in all *M. pasquieri* populations, whereas locus-wise, weighted Weir–Cockerham F_{IS} is small and positive (0.014–0.078). This indicates at most a subtle heterozygote deficit at informative loci (Cornuet and Luikart, 1996; Balloux, 2004). Allelic richness scaled positively with sample size, as expected from rarefaction theory (Kalinowski, 2004): N_p ranged from 7,971 alleles in the smallest sample (RXSSZ, n = 2) through 13,584 (CXSBZ, n = 7) and 17,693 (JXCDX, n = 11) to 21,172 alleles in the largest sample (TXLJZ, n = 15). This pattern confirms that larger samples capture more alleles and higher mean alleles-per-locus. The Tajima's D values of three populations, CXSBZ, JXCDX, and TXLJZ, were 1.6103, 1.7150, and 1.7718, respectively, a pattern consistent with an excess of intermediate-frequency variants; we hypothesize recent contraction and/or population subdivision as plausible drivers, while balancing selection at some loci cannot be excluded (Cornuet and Luikart, 1996). Nei et al. (1975) showed that the bottleneck effect causes a surplus of heterozygotes.

According to classic guidelines (Wright, 1978), F_{ST} values below 0.05 indicate little genetic differentiation, values between 0.05 and 0.15 reflect moderate differentiation, values from 0.15 to 0.25 denote great differentiation, and values above 0.25 suggest very great differentiation. In this study, the mean F_{ST} of *M. pasquieri* populations was 0.0622, indicating a moderate level of genetic differentiation. This interpretation is consistent

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2401_449460

with the PCA/ADMIXTURE results, which reveal a primary split (K = 2) but substantial overlap among populations, pointing to limited rather than strong subdivision. Isolation by distance was not detected, suggesting that landscape permeability and karst barriers may be more influential than straight-line distance (Van Strien et al., 2015). This hypothesis requires denser spatial sampling and resistance-surface analyses for testing.

ADMIXTURE based on the LD-reduced 730-SNP panel identifies a primary split at K = 2, broadly mirrored by PCA (*Figs. 3-5*). Assignments at higher K (e.g., K = 19) represent model-based granularity and should be treated as hypotheses pending replication with complementary approaches. Individuals from RXSSZ show near-complete membership in one cluster at K = 2, but inferences for this population remain uncertain due to its small sample size (n = 2).

The dispersed ancestry components at higher K are consistent with recent, multisource gene exchange under fragmented habitats, in line with the global synthesis that anthropogenic disturbance reduces allelic richness and can foster admixed individual patterns (González et al., 2020). Congeners in *M. pasquieri* involve mixed animal-mediated pollination and bat-assisted fruit set, indicating biological plausibility for occasional long-distance movement, although its frequency in *M. pasquieri* remains unknown (Nathan et al., 2009; Prasanna Kumar et al., 2013). Together with the absence of IBD, these results are compatible with an isolation-by-resistance scenario in karst landscapes, where landscape permeability and barriers outweigh straight-line distance (Van Strien et al., 2015).

Conservation strategies for M. pasquieri

The two diversity-rich populations, TXLJZ and JXCDX, should be treated as core genetic reservoirs: they harbor higher π and H_E and the greatest private-allele richness, and TXLJZ shows the strongest pairwise differentiation from other sites (F_{ST} up to 0.115), which—together with the primary split recovered at K = 2—argues for safeguarding these stands in situ as the backbone of the species' allelic diversity. Within these cores, we recommend establishing strictly protected plots, undertaking micro-habitat restoration, and implementing systematic seed collection to build ex-situ germplasm resources across climatic and elevational gradients following IUCN/SSC good practice and sampling guidance (IUCN/SSC, 2013).

At the same time, the signal of greater separation around TXLJZ suggests it should provisionally be managed as a distinct lineage (a candidate MU) until cross-lineage trials demonstrate that mixing does not impose genetic or fitness risks (Palsbøll et al., 2007; Coates et al., 2018). Any translocations or augmentation should therefore proceed cautiously and be preceded by small-scale compatibility assessments, rather than large-scale admixture by default. In contrast, RXSSZ currently provides insufficient information (n = 2) to anchor firm decisions; priority should be given to resampling ($\geq$ 15 trees) to stabilize diversity and structure estimates. If subsequent data confirm low diversity and elevated F_{IS}, a phased, closely monitored genetic-rescue program drawing on geographically proximate sources (e.g., TXLJZ/CXSBZ) could be considered to increase heterozygosity while tracking recruitment and inbreeding metrics through time.

Because isolation by distance was not detected, corridor design and restoration planting should be guided by resistance-based connectivity rather than straight-line geography, acknowledging that karst permeability and barriers likely govern movement pathways (McRae and Beier, 2007). Finally, to evaluate conservation effectiveness and enable adaptive management, we recommend setting explicit genetic monitoring

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2401_449460

targets—periodic DNA surveys (every 5–10 years) tracking H_E, π and F_{IS} and aiming to maintain effective population sizes above widely used benchmarks (Hoban et al., 2020). Together, these actions align management units with the observed patterns of diversity and differentiation, prioritize protection where genetic value is greatest, and create a feedback loop in which monitoring informs course corrections as new evidence accrues.

Conclusion

Using a Hyper-seq SNP panel, we established the first population-genomic baseline for the endangered karst tree Madhuca pasquieri in Guangxi. Populations show low standing diversity and discernible yet limited structure, with a primary split consistent across ADMIXTURE and PCA. Allelic richness is concentrated in TXLJZ and JXCDX, identifying these stands as core reservoirs of remaining variation. These results support immediate in-situ protection of TXLJZ/JXCDX, resistance-based planning to restore connectivity in karst landscapes, and cautious, closely monitored augmentation for data-poor or diversity-depleted sites (e.g., RXSSZ) once additional sampling confirms need. Key limitations include sparse sampling at RXSSZ and the reduced-representation nature of the marker set. Future work should expand spatial coverage across the species' range, combine landscape-resistance modeling with field ecology of pollination/seed dispersal, and implement periodic SNP-based monitoring to evaluate management outcomes over time.

Acknowledgements. This research was funded by the National Key Research and Development Program of China (No. 2022YFF1300703), the Guangxi Natural Science Foundation (No. 2024GXNSFAA010452), the Guangxi Forestry Science and Technology Extension Demonstration Projects (2024GLKX10, 2023LYKJ03, and [2022]GT23), the Fund of Guangxi Key Laboratory of Plant Functional Phytochemicals and Sustainable Utilization (ZRJJ2024-3, ZRJJ2024-4, and ZRJJ2024-11), the Scientific Research and Technological Development Program of Guilin (No. 20220134-3 and 20240102-3), and the Fundamental Research Program of Guangxi Institute of Botany (No. 25008 and 25009).

REFERENCES

[1] Alexander, D. H., Novembre, J., Lange, K. (2009): Fast model-based estimation of ancestry in unrelated individuals. – Genome Research 19(9): 1655-1664. DOI:10.1101/gr.094052.109.

[2] Balloux, F. (2004): Heterozygote excess in small populations and the heterozygote-excess effective population size. – Evolution 58(9): 1891-1900. DOI:10.1111/j.0014-3820.2004.tb00477.x.

[3] Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., Buckler, E. S. (2007): TASSEL: software for association mapping of complex traits in diverse samples. – Bioinformatics 23(19): 2633-2635. DOI:10.1093/bioinformatics/btm308.

[4] Cai, L.-Y., Zhang, X.-Y., Zhang, L., Ma, D. (2018): Research advances on ecological characteristics of the rare and endangered plant *Madhuca pasquierin* [in Chinese]. – Guihaia 38(7): 866-875.

[5] Cai, M., Wen, Y., Uchiyama, K., Onuma, Y., Tsumura, Y. (2020): Population Genetic Diversity and Structure of Ancient Tree Populations of *Cryptomeria japonica* var. *sinensis* Based on RAD-seq Data. – Forests 11(11): 1192. DOI:10.3390/f11111192.

[6] Chen, S., Zhou, Y., Chen, Y., Gu, J. (2018): fastp: an ultra-fast all-in-one FASTQ preprocessor. – Bioinformatics 34(17): i884-i890. DOI:10.1093/bioinformatics/bty560.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2401_449460

[7] Coates, D. J., Byrne, M., Moritz, C. (2018): Genetic diversity and conservation units: Dealing with the species-population continuum in the age of genomics. – Frontiers in Ecology and Evolution 6(OCT): 417538. https://doi.org/10.3389/fevo.2018.00165.
[8] Cornuet, J. M., Luikart, G. (1996): Description and Power Analysis of Two Tests for Detecting Recent Population Bottlenecks from Allele Frequency Data. – Genetics 144(4): 2001-2014. DOI:10.1093/genetics/144.4.2001.
[9] Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R. (2011): The variant call format and VCFtools. – Bioinformatics 27(15): 2156-2158.
[10] González, A. V., Gómez-Silva, V., Ramírez, M. J., Fontúrbel, F. E. (2020): Meta-analysis of the differential effects of habitat fragmentation and degradation on plant genetic diversity. – Conservation Biology 34(3): 711-720.
[11] Hewitt, G. M. (2000): The genetic legacy of the Quaternary ice ages. – Nature 405(6789): 907-913. https://doi.org/10.1038/35016000.
[12] Hoban, S., Bruford, M., D'Urban Jackson, J., et al. (2020): Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. – Biological Conservation 248: 108654. https://doi.org/10.1016/j.biocon.2020.108654.
[13] Huang, D., Niu, S., Bai, D., Zhao, Z., Li, C., Deng, X., Wang, Y. (2024): Analysis of population structure and genetic diversity of *Camellia tachangensis* in Guizhou based on SNP markers. – Molecular Biology Reports 51(1): 715.
[14] IUCN/SSC (2013): Guidelines for Reintroductions and Other Conservation Translocations. – Version 1.0. Gland, Switzerland: IUCN Species Survival Commission, viiii + 57 pp.
[15] Kalinowski, S. T. (2004): Counting Alleles with Rarefaction: Private Alleles and Hierarchical Sampling Designs. – Conservation Genetics 5(4): 539-543.
[16] Kan, L., Liao, Q., Chen, Z., Wang, S., Ma, Y., Su, Z., Zhang, L. (2021): Dynamic Transcriptomic and Metabolomic Analyses of *Madhuca pasquieri* (Dubard) H. J. Lam During the Post-germination Stages. – Frontiers in Plant Science 12(September): 1-22.
[17] Korunes, K. L., Samuk, K. (2021): pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. – Molecular Ecology Resources 21(4): 1359-1368. https://doi.org/10.1111/1755-0998.13326.
[18] Li, J.-F., Li, R.-F., Yan, R.-B., et al. (2014): Effect of NaCl stress on seed germination of *Madhuca pasquieri*. – Journal of Guangxi Agriculture 29(4): 27-29.
[19] Lin, L., Zeng, D.-Q., Chen, F.-F., Liu, M.-H. (2018): Research summary on current status and suggestions of *Madhuca hainanensis*. – Tropical Forestry 46(3): 26-29.
[20] Linck, E., Battey, C. J. (2019): Minor allele frequency thresholds strongly affect population structure inference with genomic datasets. – Molecular Ecology Resources 19: 639-647.
[21] McRae, B. H., Beier, P. (2007): Circuit theory predicts gene flow in plant and animal populations. – Proceedings of the National Academy of Sciences of the United States of America 104(50): 19885-19890. https://doi.org/10.1073/pnas.0706568104.
[22] Nathan, P. T., Karuppudurai, T., Raghuram, H., Marimuthu, G. (2009): Bat Foraging Strategies and Pollination of *Madhuca latifolia* (Sapotaceae) in Southern India. – Acta Chiropterologica 11(2): 435-441. DOI:10.3161/150811009X485657.
[23] Nei, M., Maruyama, T., Chakraborty, R. (1975): The bottleneck effect and genetic variability in populations. – Evolution 29(1): 1-10.
[24] Palsbøll, P. J., Bérubé, M., Allendorf, F. W. (2007): Identification of management units using population genetic data. – Trends in Ecology and Evolution 22(1): 11-16.
[25] Patterson, N., Price, A. L., Reich, D. (2006): Population structure and eigenanalysis. – PLoS Genetics 2(12): e190. https://doi.org/10.1371/journal.pgen.0020190.
[26] Prasanna Kumar, C. N., Shivaprasad, D., Somashekar, R. K., Nagaraja, B. C. (2013): Reproductive Phenology and Pollination Biology of *Madhuca neriifolia* in wet evergreen forest of Western Ghats, South India. – International Journal of Advanced Research 1: 296-306.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2401_449460

[27] Rochette, N. C., Rivera-Colón, A. G., Catchen, J. M. (2019): Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. – Molecular Ecology 28(21): 4737-4754. DOI:10.1111/mec.15253.
[28] Stronen, A. V., Norman, A. J., Vander Wal, E., Paquet, P. C. (2022): The relevance of genetic structure in ecotype designation and conservation management. – Evolutionary Applications 15(1). DOI:10.1111/eva.13339.
[29] Van Strien, M. J., Holderegger, R., Van Heck, H. J. (2015): Isolation-by-distance in landscapes: Considerations for landscape genetics. – Heredity 114(1): 27-37. DOI:10.1038/hdy.2014.62.
[30] Wright, S. (1978): Evolution and the Genetics of Populations, Vol. 4: Variability within and among Natural Populations. – University of Chicago Press, Chicago.
[31] Zou, M., Xia, Z. (2022): Hyper-seq: a novel, effective, and flexible marker-assisted selection and genotyping approach. – Innovation 3(4): 100254. DOI:10.1016/j.xinn.2022.100254.

APPENDIX

SampleID	CleanReads	Q20 content/%	Q30 content/%	GC content/%
CXSBZ1	6045274	97.18	92.41	41.17
CXSBZ2	12095268	97.83	94.19	39.81
CXSBZ3	8698760	97.41	93.13	39.16
CXSBZ4	9311166	97.53	93.38	40.45
CXSBZ5	8406420	97.53	93.43	39.67
CXSBZ6	7071988	97.26	92.66	38.42
CXSBZ7	6561790	97.21	92.53	39.27
JXCDX10	4257310	97.27	92.72	38.66
JXCDX11	8493582	97.54	93.44	38.19
JXCDX1	7085312	97.49	93.28	40.2
JXCDX2	5811536	97.8	94.11	40.21
JXCDX3	7522762	97.31	92.78	39.13
JXCDX4	5348754	97.08	92.24	38.6
JXCDX5	6084358	97.54	93.42	39.23
JXCDX6	6329814	97.31	92.21	39.25
JXCDX7	7278446	97.32	92.83	38.93
JXCDX8	11312212	97.55	92.47	39.96
JXCDX9	5620362	97.35	92.85	40.88
RXSSZ1	15351854	97.64	92.58	39.35
RXSSZ2	4705402	96.92	91.78	42.37
TXLJZ10	5946816	97.19	92.5	37.96
TXLJZ11	10010990	97.58	93.51	40.56
TXLJZ12	11196994	97.72	93.93	39.51
TXLJZ13	7051656	97.39	93.04	39.54
TXLJZ14	6141536	97.43	93.12	42.52
TXLJZ15	5841416	97.49	93.29	39.49
TXLJZ1	5659380	97.12	92.31	38.52
TXLJZ2	11804096	97.81	94.14	39.54
TXLJZ3	14051880	97.79	94.13	40.4
TXLJZ4	6099726	97.66	93.74	39.67
TXLJZ5	5986106	97.39	93.04	39
TXLJZ6	6767710	97.37	92.93	39.87
TXLJZ7	6822224	97.47	93.31	39.3
TXLJZ8	7974336	97.63	93.71	39.5
TXLJZ9	8714622	97.57	93.52	39.83

n_samples	CleanReads_mean	CleanReads_min	CleanReads_max	Q20_mean	Q30_mean	GC_mean
35	7813196	4257310	15351854	97.448	93.10457	39.66057

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2401_449460