

Genome-wide assessment of genetic diversity and population structure in *Magnolia odoratissima* based on SLAF-Seq

Tao Zhang (✉ 2469928100@qq.com)

Yunnan Agricultural University

Xue Li

Yunnan Agricultural University

Shuilian He

Yunnan Agricultural University

Research Article

Keywords: *Magnolia odoratissima*, SNP, SLAF-seq, Population structure, Genetic diversity

Posted Date: August 4th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-747821/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Genome-wide assessment of genetic diversity and population structure in *Magnolia odoratissima* based on SLAF-Seq

Tao Zhang¹, Xue Li¹, Shuillian He^{1,2*}

1 College of Horticulture and Landscape, Yunnan Agricultural University, Kunming, 650201, China

2 Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Populations, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, 650201, China

Author to whom correspondence should be addressed; e-mail:

heshuilian2006@163.com; Tel.: +86-871-65220399

Abstract

Magnolia odoratissima is a highly threatened species with small populations and scattered distribution due to habitat fragmentation and human activity. The species is recognized as a Plant Species with Extremely Small Populations (PSESP) and is endemic to China. In the current study, the population structure and levels of genetic diversity of *M. odoratissima* in the five remaining natural populations and three cultivated populations were evaluated using single nucleotide polymorphisms (SNPs) derived from Specific-Locus Amplified Fragment Sequencing (SLAF-seq). A total of 180,650 SNP loci were found in seventy *M. odoratissima* individuals. The genome-wide Nei's and Shannon's nucleotide diversity indexes of the total *M. odoratissima* population were 0.3035 and 0.4695, respectively. The observed heterozygosity (H_o) and expected heterozygosity (H_e) were 0.1122 and 0.3011. Our results suggest that *M. odoratissima* has relatively high genetic diversity at the genomic level. F_{ST} and AMOVA indicated that high genetic differentiation existed among populations. A phylogenetic neighbor-joining tree, Bayesian model-based clustering and principal components analysis (PCA) all divided the studied *M. odoratissima* individuals into three distinct clusters. The Treemix analysis showed

that there was low gene flow among the natural populations and a certain gene flow from the wild populations to the cultivated population (LS to KIB, and GN to JD). In addition, a total of 36 unique SNPs were detected as being significantly associated with environmental parameters (altitude, temperature and precipitation). These candidate SNPs were found to be involved in multiple pathways including several molecular functions and biological process, suggesting they may play key roles in environmental adaptation. Our results suggested that three distinct evolutionary significant units (ESUs) should be set up to conserve this critically endangered species.

Keywords: *Magnolia odoratissima*; SNP; SLAF-seq; Population structure; Genetic diversity

Introduction

Global biodiversity has drastically decreased over the last several decades due to anthropogenic activities such as deforestation, habitat fragmentation and climate change (Attard et al. 2016). Thousands of species have declined in numbers or gone extinct (Khan et al. 2016). Plant Species with Extremely Small Populations (PSESP), is a concept referring to species with low number populations and a narrow geographical distribution that has been disturbed and stressed by external factors over a long time. The number of individuals in a PSESP are smaller than the minimum required to prevent extinction (Ren et al. 2012). Small remaining populations, restricted habitat, serious human disturbance and extremely high risk of extinction are the key characteristics of PSESP, and nearly half of all PSESP face a high risk of extinction (John 2013). Precise protection strategies for the rescuing of PSESP are considered essential for ecological, economic, and human health, and this topic has been a major focus in the field of international conservation biology and restoration ecology. A conservation action concept focusing on PSESP was first proposed in 2012.

Magnolia odoratissima Law et R. Z. Zhou is a perennial, tree in the family Magnoliaceae, and is endemic to southwest China. The flower and leaves of *M. odoratissima* have different fragrances, and the species is used in landscaping, and planted in urban and rural areas, scenic spots, high-grade residential areas and private villas. It is considered to be an excellent landscaping tree species (Jiang et al. 2005) and its flower extract (pure oil) is a natural fragrance worthy of development (Li et al. 1996); *M. odoratissima* has also strong wind and pollution resistance, and can be used in environmental purification.

M. odoratissima is an extremely endangered plant, and is a national second-class key protected wild plant in China. The species has a narrow distribution and is restricted to a few localities in Yunnan Province, China. An exhaustive field survey across the whole of the natural *M. odoratissima* distribution range of was conducted between 2017 and 2019. Six remnant populations were found and all were severely fragmented. Population sizes in these populations currently range from 1 to approximately 200 individuals (He et al. 2018). The species is experiencing a rapid demographic decline, mainly due to increasing anthropogenic pressures (e.g. deforestation and urbanisation) in its natural habitat. *M. odoratissima* is currently ranked as having a high conservation priority in China and considered to be a PSESP (Li et al. 2003) and it is therefore vitally important to develop conservation strategies for this species.

The wild population of *M. odoratissima* are relatively isolated and far from nature reserves, and the protection of these areas is not strong. Protection of this species has not therefore been studied. To date, the wild growth, seed dormancy and germination of *M. odoratissima* resources have been studied by He et al (2020), with the results suggesting that there are substances in the seeds that inhibit germination, and that the inhibition effect is significant (He et al. 2020). Chai (2010, 2013) investigated the soil structure, population ecological characteristics and population pattern distribution characteristics of wild *M. odoratissima* populations, and analyzed

the threats facing *M. odoratissima* from the perspective of ecology (Chai et al. 2013; Dong et al. 2010; Qi et al. 2010). Jin et al (2014) analyzed the genetic diversity of four populations of *M. odoratissima* using the random amplified polymorphic DNA (RAPD) technique, with the results suggesting that *M. odoratissima* has higher genetic diversity than other endangered and endemic plants in the Magnoliaceae (Jin et al. 2014). Xu (2016) analyzed the height structure, age structure and spatial distribution pattern of *M. odoratissima*, compiled a specific time life table and drew a survival curve (Xu et al. 2016). Yang (2008, 2009) preliminarily discussed seedling and cutting propagation techniques in *M. odoratissima* (Yang 2009; Yang 2008). However, although the research to date has made progress in the field of *M. odoratissima* conservation, there is still a lack of comprehensive research into genetic diversity and population genetics in this species.

Genetic diversity is a major focus of conservation genetics. The population structure, which refers to patterns of distribution of genetic diversity in plant populations, and the levels of genetic diversity can be influenced by a variety of biotic and abiotic factors. These factors include plant breeding systems, genetic drift, pressure from natural selection, gene flow, genetic mutations and anthropogenic destruction (Chang et al. 2020), and are closely related to the adaptability and evolutionary potential of the species. Therefore, assessment of genetic variation and the spatial structure of a species is an important premise for the exploration of various evolutionary factors, especially for rare and endangered species and species impacted by direct human interference. It is difficult to develop reasonable conservation and rescue measures in the absence of basic genetic information.

Research into genetic diversity has traditionally involved DNA-based molecular markers, such as simple sequence repeats (SSR), inter-simple sequence repeats (ISSR), random amplified polymorphic DNA (RAPD), and amplified fragment length polymorphisms (AFLP) (He et al. 2016; Vangestel et al. 2016; Yang et al. 2018). However, genome-wide single nucleotide polymorphisms (SNP) are more effective

than traditional molecular markers at revealing genetic diversity and distinguishing genetic relationships (Lu et al. 2020; Zhang et al. 2018). SNP markers have a wide variety of applications in different plant species, including identification of plant varieties and cultivars, QTL analysis, construction of high-density genetic maps, and genome-wide association analysis (Delourme et al. 2008). At present, next-generation sequencing (NGS) technology can generate huge amounts of sequence data, which make the identification of high-throughput SNP markers at the species level. In order to reduce the cost of library construction when dealing with large numbers of samples, several methods of reduced-representation genome sequencing have been developed, including genotype-by-genotype (GBS) (Elshire et al. 2011), restriction-site associated DNA (RAD) sequencing (Bus et al. 2012), reduced representation libraries (RRLs) (Hyten et al. 2010), and single locus amplified fragment sequencing (SLAF-seq). Specific-Locus Amplified Fragment sequencing (SLAF-seq) is a simplified genome deep sequencing technology based on high-throughput sequencing. A large number of SNPs are developed at a genome-wide scale by constructing an SLAF-seq library and screening for specific amplified fragments. SNP markers have been widely used in genetic relationship analysis, genetic map construction, genome-wide association analysis and gene mapping of plant germplasm resources, with the advantages of high accuracy, high flux and low cost (Liu et al. 2004; Liu et al. 2019; Singh et al. 2020).

In this study, seventy individuals from the five known remaining natural populations and three cultivated populations of *M. odoratissima* were collected and analyzed using SNP data derived from SLAF-seq. Key aims of this study included: (1) to evaluate population genetic diversity of *M. odoratissima* at the genomic level and to determine the genetic differentiation and population structure in and between these eight populations, in order to provide scientific theoretical guidance for the protection of *M. odoratissima*; (2) to reveal potential correlation between the genetic data and

climatic variables and then to identify potential genes responsible for local adaptation .

Materials and Methods

Study species and sampling procedure

M. odoratissima is an understory tree species with hermaphroditic flowers, of temperate, evergreen broadleaf forests, and usually occurring in limestone forests at altitudes of 1300 m or above. Mature trees of *M. odoratissima* may reach 8 m in height and 9 cm in diameter at breast height. The bisexual flowers of the population cultivated in Kunming bloom in May, but in their original native habitat perpetual flowering is usual, and the fruits mature in September to October. This species has serious reproductive disadvantages, including low seed productivity and low germination rates. Seeds of *M. odoratissima* lack long-distance dispersal structures are dispersed by gravity and hence the vast majority of the seeds remain within 3 m of the mother tree.

This study was conducted in Wenshan Prefecture, Yunnan Province, China, which has an annual rainfall of approximately 1294 mm and an annual mean temperature of 5.9 °C. The topography of the study site is planar. From 2017 to 2020, the study site was subject to intensive demographic monitoring. During this period, the status of resources, habitat condition, companion species, population survival and age grade were assessed in a field survey by sampling approach across the whole area of distribution of *M. odoratissima* in Wenshan.

Nearly 600 individuals occur in the six populations in the study location. Only a single individual was found in the population at Boli Village, Xichou county. This could not be used in our population genetic analyses, so that this population was excluded from our study. From the other five populations, we collected new leaves from individuals and dried them immediately in fine-grain silica gel. Because of its fragrant flowers, *M. odoratissima* is also planted in gardens, so we also investigated

three cultivated populations (Jindian Garden, Kunming Institute of Botany, and Yunnan Agricultural University). Approximately ten individuals were randomly selected from each of the study populations. To ensure the sample uniformity and to minimize the relationships between the sampled plants, an average distance between individuals of greater than 10 m was ensured. Plants with normal growth, no serious defects, no obvious diseases, no insect pests and fresh leaves were chosen for sampling. A total of 70 samples were collected. Summaries of the sampling and localities are given in Table 1.

Table 1 The located sites of samples of *M. odoratissima*

Locality	Number of samples	Longitude(E)	Latitude(N)	Altitude(m)
Duimen(DM)	10	104.51	23.21	1383
Longshu(LS)	10	104.52	23.21	1365
Chengzi(CZ)	10	104.59	23.11	1556
Xiaoqing(XQ)	10	104.60	23.11	1556
Jindian Garden(JD)	10	102.77	25.09	1975
Guangnan(GN)	10	105.11	23.62	1496
Kunming Institute of Botany (KIB)	7	102.74	25.15	2040
Yunnan Agricultural University (YNAU)	3	102.75	25.13	2040

High-Throughput Sequencing and SNP calling

Genomic DNA was isolated from young leaves of *M. odoratissima* using the cetyltrimethylammonium bromide (CTAB) method (Allen et al. 2006). DNA concentration and quality were assessed using a Nanodrop 1000 Spectrophotometer and visualized on 2% agarose gel electrophoresis. Quantified DNA samples were diluted to 100 ng/ μ L for the subsequent SLAF-seq analysis. According to the genomic size and GC content of *M. odoratissima*, the *Liriodendron chinense* (Magnoliaceae) genome was selected as the reference genome for enzyme digestion prediction. To get > 100,000 SLAF tags that were evenly distributed in the genome, two restriction

enzymes, RsaI and EcoRV-HF, were selected. According to the selected optimal digestion scheme, the genomic DNA from each qualifying sample was digested separately. The obtained SLAF tags (414-464bp) were treated with a single nucleotide (A) at the 3'end, connected to the Dual-Index sequencing joint (Kozich et al. 2013), amplified using PCR, purified, mixed and gelled to select the target fragment, and then, if they passed the library quality inspection, sequenced on an Illumina platform. In order to evaluate the accuracy and effectiveness of enzyme digestion efficiency, *Oryza sativum* ssp. *Japonica* was selected as the control for sequencing (Lu et al. 2020).

The raw reads generated from the sequencing were first processed by removal of the cohesion subsequence contained in the original reads, and by removal of low quality reads (quality scores < 20), and empty reads (containing only the adapter sequence). Based on sequence similarity, BWA software was used to cluster high quality paired end reads (Kent 2002). Sequences with more than 90% similarity between different individuals were identified as SLAF loci (Sun et al. 2013). The Genome Analysis Toolkit (GATK) (McKenna, et al. 2010) and SamTools (Li et al. 2009) were used for SNP calling, as their intersection is considered to represent reliable SNPs. For the phylogenetic analysis, SNPs with a minor allele frequency (MAF) ≥ 0.05 and integrity $\geq 50\%$ were retained.

Genetic Diversity Analysis

A total of 2,998,167 SNPs from 70 individuals were developed to investigate the genetic diversity and population structure. The commonly used indexes of genetic diversity, including the observed allele number (N_a), expected allele number (N_e), observed heterozygous number (H_o), expected heterozygous number (H_e), Nei's diversity index (H), Shannon's wiener index (I), and polymorphism information content (PIC), were calculated in POPGENE (Lu et al. 2020). These indexes were calculated to estimate the degree of allele distribution (N_a and N_e), genomic heterozygosity (H_o and H_e), gene diversity (H and I) and DNA polymorphism (PIC).

In order to assess the population differentiation, analysis of molecular variance (AMOVA) was performed to estimate the partitioning of genetic variance among populations. Meanwhile, pairwise fixation index (F_{ST}) among populations was also computed to detect how gene diversity was partitioned at each level. Inter-individual fixation index (F_{IS}) was analyzed to determine the deviation of genotype frequencies from Hardy–Weinberg proportions within each population. AMOVA, F_{ST} , and F_{IS} were estimated using Arlequin (Laurent et al. 2007).

Population structure and analyses

A neighbor-joining phylogenetic tree was constructed using MEGA X software (Sudhir et al. 2018) with the following parameters, Kimura 2-parameter model, and 1000 bootstrap replicates. The population structure was analyzed in Admixture (Alexander et al. 2009), on the basis of the maximum-likelihood method. A number of populations (K) between 1 to 10 was tested, and each individual tree was assigned to its respective population according to the maximum membership probability. Genetic relationships among the studied individuals were also assessed using principal coordinates analysis (PCoA) in EIGENSOFT (Price et al. 2006) based on the Euclidian distances between individual genotypes.

Gene Flow among populations

Gene flow (Nm) at the species level was estimated using the software Genepop v4.0, and the pairwise Nm values at the population level were measured using the formula $Nm = (1 - F_{ST}) / 4 F_{ST}$ (Wright 1950) based on the F_{ST} values. TreeMix v1.13 (Pickrell and Pritchard 2012) was used to construct maximum-likelihood trees that best described the historical relationships between these populations and to infer migration events (mixtures) between them. We set the SNP block size parameter to 10. TreeMix was run iteratively for values of the migration parameter (-5).

Climatic Association Analysis

First, the chosen 19 climatic variables from the studied sites were extracted from the WorldClim data set (<http://www.worldclim.org/>) interpolated to 30-arcsec (ca. 1

km) resolution using ArcGIS. Then, Two programs, BayeScan (Matthieu and Oscar 2008) and LFMM (Frichot et al. 2013), were used to detect outlier loci that were possibly associated with climatic variables. Candidate loci under natural selection can be identified from genetic data with BayeScan, using inter-population allele frequency differences.

The other software used, LFMM, can also be used for gene-climate association analysis. LFMM estimates the hidden impact of population structure, and permits the presence of background levels of population structure (latent factors). The detected SNPs that exhibit an association with the climatic variables were determined according to the z-score. Bonferroni adjustment was used on the z-score values for multiple tests. Markers with calibration of p-values ≤ 0.01 based on chi-square tests were considered to be significant. Putative functions for the identified outlier loci were annotated using the NCBI and UniProt databases.

Results

SNP Detection

High-throughput sequencing based on SLAF generated a total of 229.54 M paired-end reads, with a mean GC content of 43.01%, and an average Q30 of 86.96%. The clean data derived from each sample ranged from 476,062 to 1,392,468, with most being approximately 700,569. An average of 3,279,207 Mb reads were assigned to each individual with a minimum and maximum read number of 1,499,485 Mb and 7,274,846 Mb, respectively (Table S1). We obtained a total of 843,082 high-quality SLAF tags for the 70 samples, with an average coverage of 11.78x for each SLAF (Table 2). Of the SLAF tags, 287,843 were polymorphic. These polymorphic SLAFs contained 2,998,167 SNPs in total, and following application of the filtering criteria, 180,650 of these were utilized in the further analyses.

Table 2 Summary of specific locus amplified fragment sequencing (SLAF-seq)

	No. of Reads	GC Content (%)	Q30(%)	No. of SLAF	Coverage
Total	229,544,493			843,082	
Avg	3,279,207.04	43.01	86.96	12,044.03	11.78

Genetic Diversity and Genetic Differentiation

The observed allele number (N_a) had a value of 2 across all the studied populations, and the value of the expected allele number (N_e) ranged from 1.4899 (JD) to 1.6553 (YNAU), with a mean value of 1.5759. The observed heterozygosity (H_o) between 0.1363 (JD) and 0.4075 (YNAU), with an average of 0.2869. The expected heterozygosity (H_e) across the eight populations was highest in YNAU (0.3818) lowest at GN (0.2961), with an average value of 0.3422. Nei's diversity index (H) was within the range from 0.3126 (GN) to 0.4581 (YNAU), and had a mean value of 0.3739. Shannon's wiener index (I) varied from 0.4553 to 0.5656 for the GN and YNAU populations. The PIC values of the eight populations ranged between 0.2398 and 0.3045, with an average of 0.2745. The maximum value of PIC was found in the YNAU population, while the minimum value was found in the GN population. The F_{IS} (a measure of intragametophytic selfing) of this species is low (0.1610), which means that *M. odoratissima* is a predominantly outcrossing species. The F_{ST} was found to be 0.4620 and the Nm was 0.2910, which means the degree of genetic differentiation is high and that gene flow among the populations is low (Table 3).

Table 3 Genetic diversity indexes of *M.odoratissima* populations

Population	N_a	N_e	H_o	H_e	H	I	PIC	F_{IS}	F_{ST}	Nm
DM	2	1.5726	0.3107	0.3378	0.3566	0.5092	0.2715	-	-	-
LS	2	1.5925	0.3217	0.3465	0.3659	0.5196	0.2774	-	-	-
CZ	2	1.6015	0.3522	0.3500	0.3691	0.5234	0.2796	-	-	-
XQ	2	1.6004	0.3551	0.3497	0.3688	0.5232	0.2795	-	-	-
JD	2	1.4899	0.1363	0.3107	0.3280	0.4814	0.2562	-	-	-
GN	2	1.4966	0.2711	0.2961	0.3126	0.4553	0.2398	-	-	-
KIB	2	1.5985	0.1406	0.3574	0.3863	0.5359	0.2875	-	-	-
YNAU	2	1.6553	0.4075	0.3818	0.4581	0.5656	0.3045	-	-	-
Overall	2	1.4714	0.1122	0.3011	0.3035	0.4695	0.2491	0.1610	0.4620	0.2910

N_a , observed allele number; N_e , expected allele number; H_o , observed heterozygosity; H_e , expected heterozygosity; H , Nei's diversity index; I , Shannon's wiener index; PIC, polymorphism information content; F_{IS} , inter-individual fixation index. F_{ST} , genetic distance. Nm , gene flow.

The pairwise fixation index (F_{ST}) is a measure of genetic differentiation among

populations. The pairwise Wright's F_{ST} values between populations in our study was a little high, with the largest value being 0.781 between the XQ and YNAU populations and the smallest being -0.022 between the JD and KIB populations. We also see that GN and YNAU are genetically similar (-0.020), and that these two populations are genetically quite different from the other populations (F_{ST} value > 0.380) (Table 4). AMOVA revealed that the greatest diversity was found within individuals is 45.12%, while the diversity found among individuals within populations was 8.68%, and a total of 46.20% of the genetic variation occurred among populations (Table 5).

Table 4 Pairwise fixation index (F_{ST}) values among eight populations of *M. odoratissima*

	DM	LS	CZ	XQ	JD	GN	KIB	YNAU
DM	0.000							
LS	0.089	0.000						
CZ	0.176	0.214	0.000					
XQ	0.205	0.234	0.142	0.000				
JD	0.088	0.125	0.157	0.172	0.000			
GN	0.750	0.757	0.756	0.760	0.525	0.000		
KIB	0.169	0.192	0.218	0.231	-0.022	0.458	0.000	
YNAU	0.765	0.776	0.775	0.781	0.475	-0.020	0.389	0.000

Table 5 Analysis of molecular variance (AMOVA) of genetic diversity in *M. odoratissima*

populations				
Source of Variation	df	Sum of Squares	Variance Components	Percentage of Variation (%)
Among populations	7	310,248.98	2,374.33	46.20
Among individuals within populations	62	199,107.33	446.20	8.68
Within individuals	70	162,330.50	2,319.01	45.12

Population Structure and Phylogenetic Analysis

Population structure also known as population stratification, refers to the existence of subgroups with different gene frequencies in the studied populations, where the individuals within the same subgroup are closely related but the subgroups are more distantly related. Population genetic structure analysis can give an estimate

of the number of ancestors of the studied population and infer the origin of each sample. The population structure of the 70 samples in our study was analyzed based on SNPs using the ADMIXTURE software (Alexander et al. 2009). The number of ancestors was determined using numbers of subgroups (K) between 1 and 10 (Figure 1A), with the value of K with the lowest predicted number of ancestors being chosen as the most likely. To generate an alternative view of population stratification, we used the population clustering program ADMIXTURE. The cross-validation error (CV) showed the lowest peak at $K=3$, which indicated that the optimal number of genetic clusters comprising the 70 *M. odoratissima* accessions was 3 (Figure 1B). Thus, the 70 accessions were classified into three major subgroups (Figure 1C): cluster I was formed from individuals from the two populations CZ and XQ; cluster II comprised all individuals from the GN and YNAU populations, as well as two individuals from JD and two individuals from KIB; cluster III comprised individuals from the DM and LS populations and the remaining individuals from JD and KIB. Natural populations all formed independent lineage.

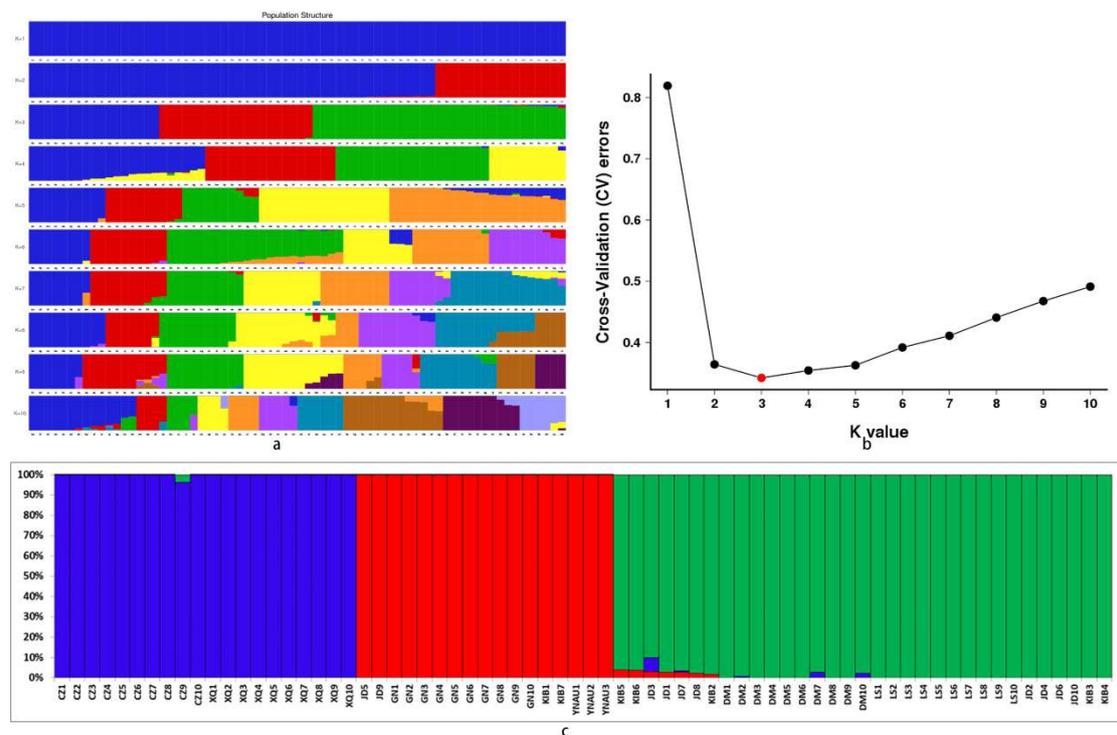


Fig. 1 Analysis of the population structure of the 70 studied *M. odoratissima* accessions using ADMIXTURE. a: Admixture graph showing the individual cluster values corresponding to each K

value. The different samples are shown on the x-axis. The y-axis quantifies the membership probability of samples belonging to different groups. Colors in each row represent structural components. b: Admixture estimation of the number of groups for K values ranging from 1 to 10. c: population structure based on $K = 3$

Our results suggested that the first and second principal components divided our 70 *M. odoratissima* samples into three groups (Figure 2A), which was highly consistent with the results of the ADMIXTURE analysis. As shown in Figure 2A, individuals from the CZ and XQ population formed a cluster; individuals of GN and YUAN formed a cluster together with some of the individuals from JD and KIB; and individuals from the populations DM and LS individuals as well as the remaining JD and KIB individuals formed a cluster. There was good distinction between the three clusters. The same conclusion was also suggested by the PCA analysis in a two dimensional map of the second principal component and the third principal component (Figure 2B). The first and third principal components were also able to distinguish between CZ, XQ and the other six regions to a great extent (Figure 2C). The first, second, and third principal components explained 56.29%, 3.52%, and 1.82% of the genetic diversity respectively (Figure 2D).

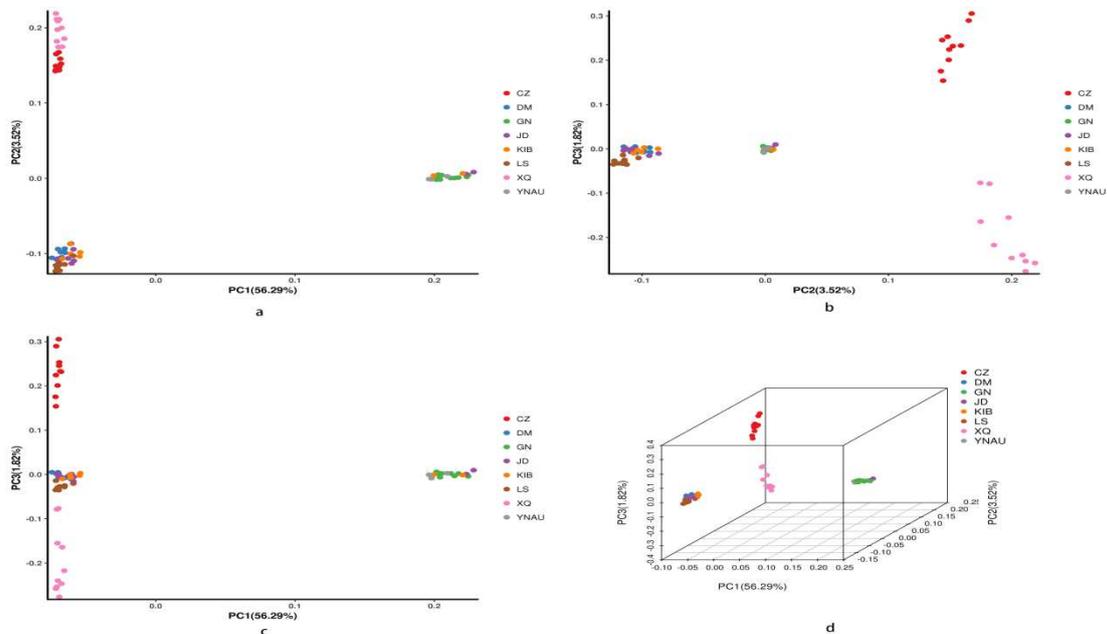


Fig. 2 PCA plots of 70 accessions based on the analysis of 180,650 SNPs

To further understand the evolutionary history of *M. odoratissima*, we used a neighbor-joining clustering algorithm to estimate the phylogenetic relationships between our study populations. Based on the analysis of the 180,650 SNP loci developed in this study, a *M. odoratissima* phylogenetic tree was constructed using the MEGA X (Sudhir et al. 2018) software, based on neighbor-joining method, and using a Kimura 2-parameter model with 1,000 bootstrap repeats. Neighbor-joining cluster analysis clearly divided the 70 samples into three clusters (Figure 3). This result was consistent with the results obtained from the ADMIXTURE software. From the results of the clustering experiment, it can be seen that there are some differences between the 70 samples of *M. odoratissima*. The phylogenetic tree divided the 70 study individuals into 3 lineages. The first group consisted of all individuals from the GN and YUAN population together with and 4 individuals from JD and KIB. The second group comprised all individuals from CZ and XQ, and was well separated from the other groups. The third group consisted of all individuals from the DM and LS population, together with 13 individuals from JD and KIB. The relationships between most of the individuals was inconsistent with their geographical location.

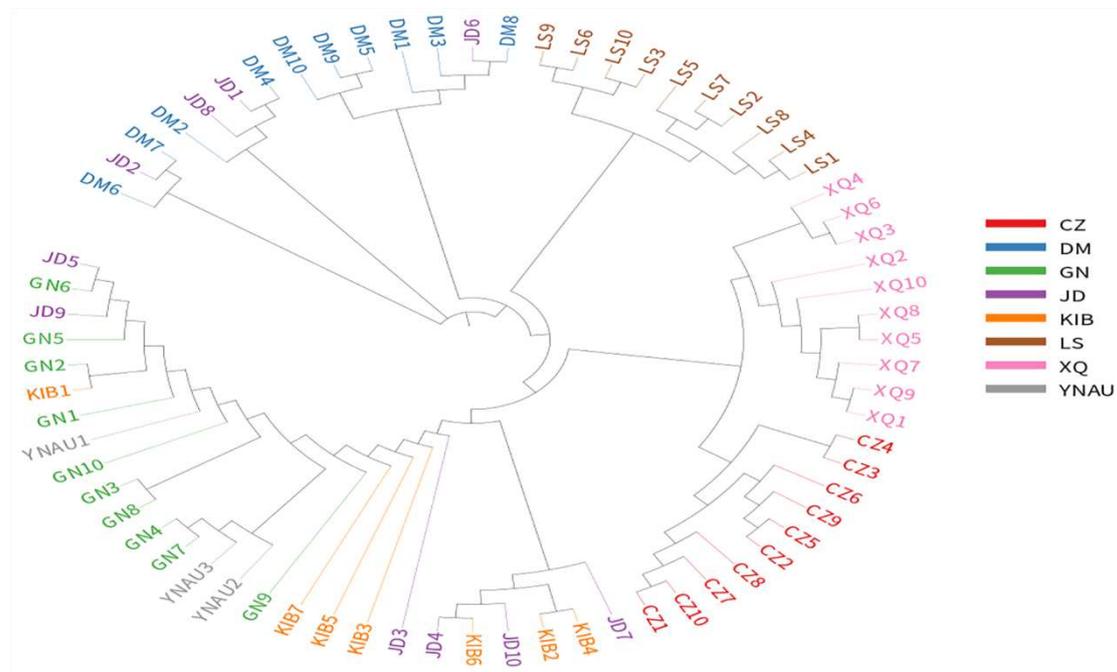


Fig. 3 Phylogenetic tree generated from 70 *M. odoratissima* samples and based on the analysis of 180,650 SNPs

Gene Flow among Populations

The pairwise population Nm values calculated from Wright's analysis indicated that the levels of gene flow between populations was quite low, with the lowest value being -12.883 between the GN and YNAU populations, and the highest being 2.597 between the JD and DM populations (Table 6).

To further describe the historical relationships between these populations and to infer migration events (mixtures) between them, we applied TreeMix to the 8 populations. TreeMix uses genome wide allele frequency data and a Gaussian approximation to genetic drift to analyze whether or not the migration events have occurred. The results obtained show population splits and gene flow between populations. From the TreeMix output (Figure 4), it can be clearly seen that the YUAN, GN and KIB populations clustered together as one group, and that there was strong gene flow to the JD population. However, individuals from the XQ and CZ populations clustered together in one group. Considering that JD, YNAU and KIB are cultivated populations, we think that the JD and KIB populations are probably therefore not single origin populations with some individuals in the JD population having been introduced from the natural population DM, and other individuals from GN. The individuals comprising the KIB population may have come from the LS and GN populations.

Table 6 Pairwise gene flow (Nm) values between the eight populations of *M. odoratissima*

	DM	LS	CZ	XQ	JD	GN	KIB	YNAU
DM	0.000							
LS	2.552	0.000						
CZ	1.167	0.919	0.000					
XQ	0.969	0.821	1.511	0.000				
JD	2.597	1.748	1.341	1.203	0.000			
GN	0.083	0.080	0.081	0.079	0.226	0.000		
KIB	1.233	1.052	0.896	0.833	-11.671	0.296	0.000	
YNAU	0.077	0.072	0.073	0.070	0.276	-12.883	0.393	0.000

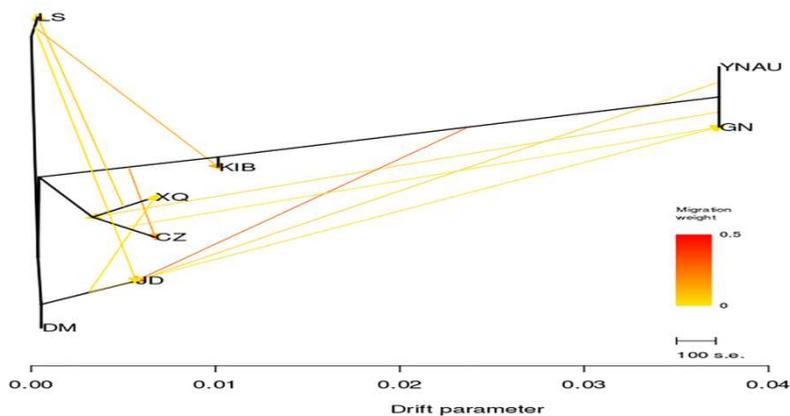


Fig. 4 TreeMix analysis of 70 *M. odoratissima* samples divided into 8 populations. Arrows correspond to migration events, and the darkness of the arrow color indicates the weight of the migration edge

Associations between SNP Markers and Environmental Variables

Analysis of the associations between SNPs and environmental variables was conducted in the Bayenv2 and LFMM programs. A total of 387 and 14,369 SNP markers were identified as being significantly correlated with climate variables by Bayenv2 and LFMM, respectively. 36 SNP markers showed a significant correlation with climate variables in both of the analyses (Table 7). The highest degree of correlation was reached with “isothermality”, which could be applied to 18 markers in our study. The other climate variables, altitude and latitude, correlated with between 1 and 9 markers..

A Blast search revealed that 25 related SNP markers were annotated, with molecular functions listed as enabling DNA binding and RNA binding, and enhancing the activity of RNA-directed, DNA polymerase, and aspartic acid type endopeptidases. These markers also seem to be involved in biological processes that involve RNA dependent, DNA biosynthesis, DNA recombination and proteolysis (Table S2).

Table 7 A summary of putative adaptive markers displaying associations with different climate variables, identified by Bayenv2 and LFMM analysis

SNP ID	Pos	log10(PO)	qval	b1	b3	b5	b8	b10	b12	b14	b16	b17	b18	b19	Alt	Lat
Markr75684	182	3.2215	0.0004	*												
Markr93856	99	1.2750	0.0198	*												
Markr173639	217	0.8450	0.0446	*												
Marker113261	76	0.8603	0.0404	*					*							
Marker40767	158	1.5278	0.0084	*				*								
Marker79135	175	1.0187	0.0306	*	*		*	*								*
Marker38338	23	1.1220	0.0264		*											
Marker41163	237	2.3961	0.0009		*											
Marker62186	139	2.2191	0.0015		*											
Marker75096	35	1.6055	0.0096		*											
Marker80697	154	3.3977	0.0002		*											
Marker88635	70	1.0725	0.0268		*											
Marker94916	32	1.0154	0.0310		*											
Marker97928	258	1.5567	0.0110		*											
Marker101966	253	2.1909	0.0027		*											
Marker155158	47	1.6901	0.0069		*											
Marker125197	66	1.0653	0.0276		*										*	
Marker87087	54	0.8554	0.0418		*							*		*		*
Marker48307	94	1.0642	0.0293		*							*		*		
Marker71656	59	2.5215	0.0013		*							*		*	*	
Marker73151	48	1.8540	0.0044		*				*							
Marker80861	60	0.8587	0.0409		*				*		*		*			

SNP ID	Pos	log10(PO)	qval	b1	b3	b5	b8	b10	b12	b14	b16	b17	b18	b19	Alt	Lat
Marker20490	51	0.8982	0.0453		*										*	
Marker37773	17	2.5215	0.0005			*									*	
Marker71622	64	2.3961	0.0016			*			*						*	
Marker60162	24	2.4670	0.0007				*	*							*	
Marker17623	4	1.2074	0.0199				*	*				*		*		
Marker172800	191	2.1044	0.0032				*	*	*							
Marker202350	173	3.3977	0.0002				*	*								
Marker202965	5	1.5534	0.0112				*	*								
Marker26135	165	1.4353	0.0113						*							
Marker49381	237	1.6438	0.0060						*							
Marker29458	60	1.2768	0.0172							*		*		*		
Marker35825	220	1.8940	0.0041												*	
Marker60205	96	1.2572	0.0182												*	
Marker63297	239	0.9678	0.0399												*	

(1) b1, Annual Mean Temperature; b3, Isothermality; b5, Maximum Temperature of Warmest Month; b8, Mean Temperature of Wettest Quarter; b10, Mean Temperature of Warmest Quarter; b12, Annual Precipitation; b14, Precipitation in Driest Month; b16, Precipitation in Wettest Quarter; b17, Precipitation in Driest Quarter; b18, Precipitation in Warmest Quarter; b19, Precipitation in Coldest Quarter; Alt, Altitude; Lat, Latitude. (2)* suggests that the SNP showed an association with that specific climate variable

Discussion

Genetic diversity in *M. odoratissima* populations

Genetic diversity may occur as the result of long-term accumulation of neutral substitutions during the evolution of species, or may also be a result of diversifying selection from different environments. Genetic diversity is also linked to the long-term survival and development of populations and species. It is widely believed that endangered and endemic plants must have low genetic diversity (Soltis et al. 1992), however some studies have shown that, in contrast, certain rare and endangered plants may have high genetic diversity (Jin and Li 2007; Matthew and Pamela 2000). At the species level, *M. odoratissima* has higher genetic diversity ($H = 0.3035$, $I = 0.4695$) than other endangered species in the same family, including *Manglietia grandis* ($H = 0.2433$, $I = 0.3651$) (Chen et al. 2010), and *Tsoongiodendron odorum* ($H = 0.2597$, $I = 0.3565$) (Huang and Zhuang 2002), and levels similar to those in *Liriodendron chinensis* ($H = 0.3995$, $I = 0.5806$) (Luo et al. 2000). Our results indicate that *M. odoratissima* has relatively high genetic diversity.

Hamrick et al. (Hamrick and Loveless 1992) pointed out that the level of genetic diversity in a species cannot be separated from its life history characteristics and ecological characteristics. This has led to the belief that rare or endemic species and species with narrow distributions have low levels of genetic diversity, and those with wide distributions, outcrossing and animal seed dispersal would have high genetic diversity. However, although the distribution of *M. odoratissima* is narrow and the number of existing populations is small, the genetic diversity of *M. odoratissima* is high, which may be related to its biological characteristics, life history and habitat characteristics. First of all, as with many members of the Magnoliaceae, *M. odoratissima* is insect pollinated, and the flower division, color, sweet nectar and dichogamy of the flowers are all characteristic of insect pollination. Furthermore, *M. odoratissima* is a perennial tree, and the generations therefore overlap. Under the effects of long term natural selection, the genes that confer an advantage and therefore

higher rates of survival are continuously accumulated and retained, thus forming a broad genetic basis.

Genetic differentiation and gene flow

According to Wright (Wright 1950), if the Nm is less than 1, genetic drift and differentiation may occur. If the F_{ST} is 0.05 or less, differentiation is thought to be negligible; an F_{ST} range of between 0.15 and 0.25 indicates a moderate level of differentiation; while $F_{ST} = 0.25$ means subpopulations were highly differentiated. The results of our genetic analysis showed that there is high genetic differentiation ($F_{ST} = 0.462$) but low gene flow ($Nm = 0.291$) at genomic level between populations of *M. odoratissima*. At the same time, AMOVA analysis showed that 46.02% of the genetic variation mainly occurred among populations. Levels of genetic differentiation in *M. odoratissima* are therefore much higher than in other endangered species of Magnoliaceae, such as *Manglietia decidua* ($F_{ST} = 0.207$) (Liao et al. 2004), *Michelia chapensis* ($F_{ST} = 0.222$) (Jiang et al. 2005), and *Magnolia wufengensis* ($F_{ST} = 0.171$) (Chen et al. 2014). Thus, it is very likely that the *M. odoratissima* populations have undergone either genetic drift or any apparent degree of genetic differentiation to date. The natural populations of *M. odoratissima* showed serious fragmentation, and their disjunctive distribution and the lack of pollinators visiting more than one population maybe the main reason for the high genetic differentiation and possible genetic drift observed in the populations.

Inter-population differentiation is related to the existing barriers to gene flow, as well as selection from environmental factors. Some studies have shown that variations in ecological micro-environments can lead to significant differences in genetic structure among different populations (Douglas and Lonnie 1990; Turkington and Harper 1979). Therefore, the levels of genetic variation observed between different populations may indicate the extent to which an organism adapts to different environments. Analysis of the associations between SNPs and environmental variables conducted in the Bayenv2 and LFMM programs suggested that there was a

correlation between genetic markers and environmental variables, and therefore, the variation in ecological micro environments in the separate *M. odoratissima* populations may lead to significant differences in genetic structure in different populations.

Implications for conservation

Maintenance of genetic variation is one of the major objectives for conserving endangered and threatened species (Wallis 1996). Knowledge of genetic variation between and within populations provides essential information on the formulation of appropriate management strategies directed towards their conservation (Milligan et al. 2010). From the results obtained in this study it is possible to draw inferences on the conservation of *M. odoratissima*. Given that *M. odoratissima* has moderate to high genetic diversity, high population differentiation and low gene flow suggest that genetic drift is currently of great concern for this species. During the past few decades, this tree has been over-exploited for timber and ornamental use because of its fragrant flowers and beautiful shape. Many of the trees in the remaining wild populations are located near villages in limestone mountains and are easily accessible, which may heighten the possibility of genetic drift. Furthermore, population fragmentation is associated with decreases in population size and increasing geographic separation of the populations, which reduces rates of reproductive success (Suzuki et al. 2013). We found that gene flow between populations was low, and that the effects of the population structure and the size of current *M. odoratissima* populations were high, meaning that the species is likely to be susceptible to the loss of genetic polymorphisms in the future because of random genetic drift. The medium-term goal is therefore to establish a new generation of individuals to improve gene flow among populations, and thereby reduce genetic drift. Field surveys have shown that natural regeneration in *M. odoratissima* populations is slow, because of low fruit set and low seed germination rates. We therefore suggest that appropriate anthropogenic measures, including artificial pollination between populations, should be carried out to promote

the natural regeneration of the population, so that the populations of *M. odoratissima* are able to recover and gene flow between the populations can be rebuilt.

Considering that the main threats driving declines in the remaining six *M. odoratissima* populations are deforestation, over-exploitation of exotics, and habitat destruction from nearby residents, the most urgent measures required in the near future are in-situ and ex-situ conservation. There are six wild populations left, according to our field investigation, and we found that the extent of the *M. odoratissima* distribution is relatively concentrated, so in-situ protection is not difficult and natural reserves should be easy to established in these wild populations. Population structure analyses showed that the individuals of *M. odoratissima* clustered naturally into three groups, so we can look them as three distinct evolutionarily significant units (ESUs), and in this case, one representative population from each ESUs could be selected and set up as a key nature reserve. Using the information both from our field work and genetic analyses, we suggest that populations XQ, DM and GN are the best choices to set up nature reserve. At the same time, information and campaigns about the protection of endangered species should be strengthened, and protective support from the local government is necessary. The ex-situ protection of *M. odoratissima* is already underway, with conserved populations established in the endangered plant resource garden of the Kunming Institute of Botany (KIB), Yunnan Agricultural University (YNAU), and a famous park in Kunming (JD), and all three cultivated populations are well protected by the corresponding institutions. From the genetic structure analyses, we found that plants from the wild population GN clustered in a group containing some individuals from the cultivated populations at KIB and JD and all individuals from YNAU. Furthermore, individuals from the wild populations DM and LS cluster closely together with certain individuals from JD and KIB. This suggests that the wild populations DM, LS and GN are well represented in the ex situ conservation. However, wild populations CZ and XQ display a certain degree of genetic diversity

that is not represented in the cultivated populations, and ex-situ conservation of representative individuals from these populations should be implemented as soon as possible.

On top of the protection of existing living plants, another strategy for propagating future generations is to collect seeds in order to cultivate young seedlings and establish breeding in an off-site location. Germplasm resources can also be propagated by grafting, and genetically dissimilar individuals can be crossed by artificial pollination to produce highly heterozygous seedlings. Furthermore, in view of the difficulty of colonization and reproduction in this species, future research should encompass somatic embryogenesis for artificial reproduction.

Conclusions

Information on the genetic diversity, genetic structure, and demographic history of endemic species can aid the development of appropriate conservation and management strategies. In the present study, the genetic variation, genetic differentiation and population structures of five wild and three cultivated *M. odoratissima* populations were first investigated with Specific-Locus Amplified Fragment Sequencing (SLAF-seq). The plant showed high levels of genetic diversity and high levels of genetic differentiation among populations, three distinct evolutionary significant units (ESUs) were revealed and population XQ, DM and GN were suggested to set up nature reserve. The obtained results will certainly help us to promote the conservation of *M. odoratissima* in the near future.

Author Contributions: T. Z. directed most of the experimental and analytical work and wrote the manuscript. S. L. H. designed the analytical workflow and revised the manuscript. X. L. participated in the experimental and analytical work.

Financial support: The project was supported by the Fund of Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small

Populations (Project number: PSESP2021F02), and the National Natural Science Foundation of China (Project number: 32060083 & 31500459).

Conflicts of Interest: The authors declare no conflicts of interest.

References:

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Cold Spr Harb Lab Pr* 19

Allen GC, Flores VMA, Krasynanski S, Kumar S, Thompson WF (2006) A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat Protoc* 1: 2320-2325

Attard CRM, Möller LM, Sasaki M, Hammer MP, Bice CM, Brauer CJ, Carvalho DC, Harris JO, Beheregaray LB (2016) A novel holistic framework for genetic based captive breeding and reintroduction programs. *Conserv Biol* 30: 1060-1069

Bus A, Hecht J, Huettel B, Reinhardt R, Stich B (2012) High-throughput polymorphism detection and genotyping in *Brassica napus* using next-generation RAD sequencing. *BMC Genom* 13: 281

Chai Y, Qi RP, Meng GT, Shao JP, Li GX (2013) Fractal characteristics of spatial pattern of *Magnolia odoratissima* endemic to Yunnan Province (in Chinese). *J NE For Univ* 41: 25-28

Chang YL, Peng MW, Chu GM, Wang M (2020) Genome wide assessment of population structure and genetic diversity for *Anabasis aphylla* based on specific length amplification fragment sequencing. *J Plant Int* 15: 75-82

Chen LY, Chen FJ, He SC, Ma LY (2014) High genetic diversity and small genetic variation among populations of *Magnolia wufengensis* (Magnoliaceae), revealed by ISSR and SRAP markers. *Electr J Biotechnol* 17: 268-274

Chen SY, Han Y, Wu T, Fu YP, SiMa YK, Hao JB (2010) Genetic diversity of endangered species *Manglietia grandis* detected by inter-simple sequence repeat markers (in Chinese). *J For Envir* 30: 56-60

Delourme R, Piel N, Horvais R, Pouilly N, Domin C, Vallée P, Falentin C, Manzanares-Dauleux MJ, Renard M (2008) Molecular and phenotypic characterization of near isogenic lines at QTL for quantitative resistance to *Leptosphaeria maculans* in oilseed rape (*Brassica napus* L.). *Theor Appl Genet* 117: 1055-1067

Dong L, Liu XQ, Yu DH, Chai Y (2010) A preliminary study on community characteristics

of *Magnolia odoratissima*, a critically endangered species in Yunnan Province (in Chinese). *J Anhui Agr Sci* 38: 12068-12069+12297

Douglas RT, Lonnie WA (1990) Complex competitive relationships among genotypes of three perennial grasses: implications for species coexistence. *Am Nat* 136: 305-327

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *Plos One* 6

Frichot E, Schoville SD, Bouchard G, Francois O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol* 30: 1687-1699

Hamrick JL, Loveless MD (1992) Factors in influencing levels of genetic diversity in woody plant species. *New For* 6: 95-124

He SL, Wang YS, Li DZ, Yi TS (2016) Environmental and historical determinants of patterns of genetic differentiation in Wild Soybean (*Glycine soja* Sieb. et Zucc). *Sci Rep* 6: 131-140

He SL, Wang ZY, Wang W, Yang Y (2018) Research on population of *Magnolia odoratissima* Law et R. Z. Zhou, a Wild Plant Species with Extremely Small Population (in Chinese). *J W CN Fore Sci* 47: 64-68

He SL, Zhang T, Wang W, Yang Y (2020) Seed dormancy and germination of *Magnolia odoratissima*, a wild plant with extremely small population (in Chinese). *Agr Sci Jiangsu* 48: 133-136

Huang JX, Zhuang XY (2002) Genetic diversity of the populations of *Tsoongiodendron odorum* (in Chinese). *Chin J Plant Ecol* 4: 413-419

Hyten DL, Cannon SB, Song Q, Weeks N, Fickus EW, Shoemaker RC, Specht JE, Farmer AD, May GD, Cregan PB (2010) High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genom* 11: 38

Jiang JM, Teng HJ, Yuan JL, Luan QF, Tan ZF (2005) Genetic diversity of *Michelia chapensis* Dandy populations (in Chinese). *For Res* 2: 109-113

Jiang WB, Cao J, Li G, Wong ML (2005) Development of new ornamental tree species of

Magnolia family in China and Its application in landscaping (in Chinese). *Acta Agr Shanghai* 2: 68-73

Jin R, Shi YX, Xu T, Yang MZ, SiMa YK (2014) Study on genetic diversity of endangered and endemic plant of *Magnolia odoratissima* (Magnoliaceae) in Yunnan Province (in Chinese). *J W CN Fore Sci* 43: 80-84.

Jin ZX, Li JM (2007) ISSR analysis on genetic diversity of endangered relic shrub *sinocalycanthus chinensis* (in Chinese). *Chin J Appl Ecol* 2: 247-253

John FO, Primate (2013) conservation: unmet challenges and the role of the International Primatological Society. *Int J Primatol* 34: 234-245

Kent WJ (2002) Blat—The Blast-Like alignment tool. *Genome Res* 12: 656-664

Khan S, Nabi G, Ullah MW, Yousaf M, Manan S, Siddique R, Hou H, Cao HX (2016) Overview on the role of Advance Genomics in Conservation Biology of Endangered Species. *Int J Genom* 2016

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79: 5112-5120

Laurent E, Guillaume L, Stefan S (2007) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinformat Onl* 1

Li H, Bob H, Alec W, Tim F, Jue R, Nils H, Gabor M, Goncalo A, Richard D (2009) The sequence alignment/map format and samtools. *Bioinformatics* 25

Li YY, Li DX, Mao YL, Li J, Xu JH (1996) Study on extraction and chemical constituents of extract and essential oil from aromatic plants of Yunnan orchidaceae (in Chinese). *Flav Fragr Cosmet* 3: 2-8

Li YY, SiMa YK, Fang B, Guo LQ, Jiang H, Zhao WS (2003) Current situation and evaluation of natural resources of the priority protection wild plants in Yunnan Province of China (in Chinese). *Plant Div* 2: 181-191

Liao WF, Xia NH, Deng YF, Zheng QY (2004) Study on genetic diversity of *Manglietia decidua* (Magnoliaceae) (in Chinese). *Plant Div* 1: 58-64

Liu DY, Chu L, Yang YH (2004) Genetic diversity of rare and endangered plant *Magnolia amoena* (in Chinese). *Chin J Appl Ecol* 7: 1139-1142

Liu K, Li KX, Wei XJ, Liang WH, Wang K (2019) Development and genetic analysis on SNP sites from *Camellia nitidissima* based on SLAF-seq technology (in Chinese). *Non-wood For Res* 37: 78-83

Lu RS, Chen Y, Tamaki I, Sakaguchi S, Ding YQ, Takahashi D, Li P, Isaji Y, Chen J, Qiu YX (2020) Pre-quaternary diversification and glacial demographic expansions of *Cardiocrinum* (Liliaceae) in temperate forest biomes of Sino-Japanese Floristic Region. *Mol Phylogenet Evol* 143

Lu YZ, Dong XY, Huang LB, Zheng JW, He XD, Sun HN, Jiang ZP (2020) SLAF-seq uncovers the genetic diversity and adaptation of Chinese Elm (*Ulmus parvifolia*) in Eastern China. *Forests* 11

Luo GZ, Shi JS, Yin TM, Huang MR, Wang MX (2000) Comparison of genetic diversity between *Liriodendron tulipifera* Linn. and *Liriodendron chinense* (Hemsl.) Sarg. by means of RAPD markers (in Chinese). *J Plant Res Env* 2: 9-13

Matthew AG, Pamela SS (2000) Patterns of genetic variation in rare and widespread plant congeners. *Am J Bot* 87: 783-792

Matthieu F, Oscar G (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180: 977-993

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Cold Spr Harb Lab Pr* 20: 1297-1303

Milligan BG, Leebens MJ, Strand AE (2010) Conservation genetics: beyond the maintenance of marker diversity. *Mol Ecol Resour* 3: 423-435

Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome wide allele frequency data. *Plos Genet* 8: e1002967-e1002967

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:

904-909

Qi RP, Chai Y, Wu L, Li NY, Yang W, Fu, YB (2010) Study on soil physical and chemical properties in the habitat of the concentrated distribution of *Magnolia odoratissima* (in Chinese). *J W CN Fore Sci* 39: 95-98

Ren H, Zhang HX, Lu HF, Liu HX, Guo QF, Wang J, Jian SG, Bao HO (2012) Wild Plant Species with Extremely Small Populations require conservation and reintroduction in China. *Ambio* 41: 913-917

Singh RB, Mahenderakar MD, Jugran AK, Singh RK, Srivastava RK (2020) Assessing genetic diversity and population structure of sugarcane cultivars, progenitor species and genera using microsatellite (SSR) markers. *Gene* 753

Soltis PS, Soltis DE, Tucker TL (1992) A llozyme variability is absent in the narrow endemic *Bensoniella oregona* (saxifraga gaceae). *Conserv Biol* 6: 131-134

Sudhir K, Glen S, Michael L, Christina K, Koichiro T (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35: 1547-1549

Sun XW, Liu DY, Zhang XF, Li WB, Liu H, Hong WG, Jiang CB, Guan N, Ma CX, Zeng HP (2013) SLAF-seq: An Efficient Method of Large-Scale De Novo SNP Discovery and Genotyping Using High-Throughput Sequencing. *Plos One* 8

Suzuki S, Teruyoshi N, Nobuhiro T (2013) Pollen flow and effects of population structure on selfing rates and female and male reproductive success in fragmented *Magnolia stellata* populations. *BMC Ecol* 13: 10

Turkington R, Harper JL (1979) The growth, distribution and neighbor relationships of *Trifolium repens* in a permanent pasture. IV. Fine scale biotic differentiation. *J Ecol* 67: 201-218

Vangestel C, Vázquez-Lobo A, Martínez-García PJ, Calic I, Wegrzyn JL, Neale DB (2016) Patterns of neutral and adaptive genetic diversity across the natural range of sugar pine (*Pinus lambertiana* Dougl.). *Tree Genet Genomes* 12

Wallis GP (1996) Conservation genetics: Case histories from nature. *Tr Ecol Evol* 11: 308-308

Wright S (1950) Genetical structure of populations. *Nature* 166: 247-249

Xu JH, SiMa YK, Xu T, Mao YB, Yu H (2016) Population structure and dynamics of *Liriodendron odoratissimum* (Y. W. Law et R. Z. Zhou) N. H. Xia et C. Y. Wu, a Plant Species with Extremely Small Populations (in Chinese). *J W CN Fore Sci* 45: 35-40+51

Yang YH (2009) Propagation experiment on cuttings of *Magnolia odoratissima* (in Chinese). *J For Eng* 23: 103-105

Yang Z, Wang LJ, Zhao TT (2018) High genetic variability and complex population structure of the native Chinese hazelnut. *Br J Bot* 41: 687-697

Yang ZS (2008) Seedling technology of *Magnolia odoratissima* (in Chinese). *J For Eng* 4: 109-111

Zhang HX, Li HY, Li YX (2018) Identifying evolutionarily significant units for conservation of the endangered *Malus sieversii* using genome-wide RAD-seq data. *Nord J Bot* 36: n/a-n/a

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterials.pdf](#)