

RESEARCH

Open Access



Environment influences the genetic structure and genetic differentiation of *Sassafras tzumu* (Lauraceae)

Bicai Guan¹, Qian Liu¹, Xiang Liu¹ and Xi Gong^{1*}

Abstract

Background *Sassafras tzumu*, an elegant deciduous arboreal species, belongs to the esteemed genus *Sassafras* within the distinguished family Lauraceae. With its immense commercial value, escalating market demands and unforeseen human activities within its natural habitat have emerged as new threats to *S. tzumu* in recent decades, so it is necessary to study its genetic diversity and influencing factors, to propose correlative conservation strategies.

Results By utilizing genotyping-by-sequence (GBS) technology, we acquired a comprehensive database of single nucleotide polymorphisms (SNPs) from a cohort of 106 individuals sourced from 13 diverse *Sassafras tzumu* natural populations, scattered across various Chinese mountainous regions. Through our meticulous analysis, we aimed to unravel the intricate genetic diversity and structure within these *S. tzumu* populations, while simultaneously investigating the various factors that potentially shape genetic distance. Our preliminary findings unveiled a moderate level of genetic differentiation ($F_{ST}=0.103$, $p < 0.01$), accompanied by a reasonably high genetic diversity among the *S. tzumu* populations. Encouragingly, our principal component analysis painted a vivid picture of two distinct genetic and geographical regions across China, where gene flow appeared to be somewhat restricted. Furthermore, employing the sophisticated multiple matrix regression with randomization (MMRR) analysis method, we successfully ascertained that environmental distance exerted a more pronounced impact on genetic distance when compared to geographical distance ($\beta_E=0.46$, $p < 0.01$; $\beta_D=0.16$, $p < 0.01$). This intriguing discovery underscores the potential significance of environmental factors in shaping the genetic landscape of *S. tzumu* populations.

Conclusions The genetic variance among populations of *S. tzumu* in our investigation exhibited a moderate degree of differentiation, alongside a heightened level of genetic diversity. The environmental distance of *S. tzumu* had a greater impact on its genetic diversity than geographical distance. It is of utmost significance to formulate and implement meticulous management and conservation strategies to safeguard the invaluable genetic resources of *S. tzumu*.

Keywords *Sassafras tzumu*, Landscape genetics, Environmental distance, Genetic diversity, Genetic differentiation

Introduction

The genetic architecture and genetic variation serve as the underpinnings for plants to acclimate to their surroundings and ensure their survival and reproductive success [1, 2]. The greater the wealth of genetic diversity, the more adaptable a species becomes in navigating its environment, thus bestowing it with a heightened advantage in terms of survival and evolutionary potential

*Correspondence:

Xi Gong
gongxi@ncu.edu.cn

¹ School of Life Sciences, Nanchang University, Nanchang 330031, China



[3]. Factors such as the geographical distribution range, growth patterns, and reproductive strategies exert influence on the genetic structure and diversity of plant species [4]. Species with a broader range of distribution may exhibit a greater tapestry of genetic variability [5]. Self-pollinating species, due to reproductive assurance, possess a more extensive geographical reach compared to their hybrid counterparts [6–9]. The duration of a species' generational growth profoundly impacts its genetic diversity. Species with shorter generations are anticipated to possess more restricted gene pools, hence fostering population isolation. Conversely, species with lengthier generation times are expected to experience a more gradual decline in genetic diversity [10], albeit this trend could also lead to genetic homogeneity. Furthermore, species with limited dispersal capabilities tend to exhibit greater differentiation, while those with long-distance dispersal mechanisms are more prone to population homogenization [10]. Thus, the genetic structure and diversity of species can, to a certain extent, serve as a reflection of their genetic evolutionary potential and adaptive capacity to their environment [11].

Sassafras tzumu, an arboreal species, pertains to the botanical family Lauraceae and falls under the genus *Sassafras*. It thrives abundantly across the southern region of the Yangtze River in China, flourishing at altitudes ranging from 100 to 1900 m amidst both sparse and dense woodland habitats [12]. Although the International Union for Conservation of Nature (IUCN) classifies *S. tzumu* as least concern (LC) and this species is not listed in the China Species Red List, in 2017, *S. tzumu* was included in the Reference List of Major Cultivated Precious Tree Species in China by the State Forestry Administration [13]. It usually begins to bloom from an early spring. With its vibrant crimson foliage in the autumn, ever-changing leaf patterns, and exceptional quality, *S. tzumu* not only served as a splendid ornamental tree but also boasted remarkable timber properties [14]. Moreover, the roots and bark possessed medicinal properties, known for their ability to enhance blood circulation, disperse blood stasis, alleviate wind and dampness, and treat contusions and lumbar muscle strains [13]. Regrettably, due to its immense commercial value, certain regions had suffered from rampant deforestation (http://qjwb.thehour.cn/html/2017-07/18/content_3547987.htm?div=-1) and theft of *S. tzumu* specimens (http://www.jiande.gov.cn/art/2018/6/27/art_1468690_18903167.html). In recent decades, escalating market demands and unforeseen human activities within its natural habitat had emerged as new threats to *S. tzumu* [15].

Presently, investigations pertaining to the genetic diversity and genetic structure of *S. tzumu* populations predominantly center around the analysis of genetic

diversity utilizing isoenzymes and the development of microsatellite loci [14, 16–18]. Wang et al. [17] used 27 polymorphic nSSRs of *S. tzumu* to conduct genetic diversity, genetic structure analysis and mantel tests, thus providing corresponding theoretical basis for the protection and utilization of *S. tzumu*. A study conducted an analysis of the genetic structure of three diminutive natural populations of *S. tzumu* in Hubei Province, wherein the findings explicated that the proportion of polymorphic loci in the three populations stood at 49.5%. These populations were characterized by an inadequacy of heterozygotes and a surplus of homozygotes, thus signifying a state of disequilibrium [18]. Jiang et al. [14] found that five wild populations of *S. tzumu* at different altitudes had high genetic diversity ($H_E=0.84$) and moderate genetic differentiation with 13 polymorphic SSR markers from *S. randaiense* and *Cinnamomum camphora*. Employing ArcGIS 10.2 and MaxEnt 3.3.2, the distribution pattern of *S. tzumu* in different timeframes was simulated. The outcomes demonstrated that precipitation of driest month, precipitation of wettest month, temperature seasonality, and mean temperature of wettest quarter chiefly influenced the distribution. Furthermore, a comparison between the simulated outcomes of the distribution range during distinct periods revealed that the suitable habitat for *S. tzumu* contracted and shifted from the southern to the northern regions [19].

Genotyping-by-sequence (GBS) technology represented one of the streamlined genome sequencing methodologies, aiming to reduce genome complexity through the application of restriction enzymes and the incorporation of single nucleotide polymorphisms (SNPs) [20]. This technique yielded a substantial number of SNPs, which were harnessed for exploring interspecies diversity, constructing haplotype maps, conducting genome-wide association studies, and facilitating genome selection [21]. Notably, GBS offered the advantage of requiring fewer steps for database construction and enabled the establishment of databases for a large number of samples [22]. In certain genomic studies, the utilization of genotyping-by-sequencing (GBS) technology allowed for the elucidation of species' genetic diversity and hybridization characteristics [23].

The fragmentation of habitats caused by anthropogenic activities and other contributing factors, coupled with the escalating effects of global climate change, has resulted in a significant decline in the genetic diversity of numerous indigenous plant species. Such diminishment in genetic diversity has profound implications for the long-term viability and survival of these plants [24, 25]. In order to establish a robust theoretical framework for future judicious development and utilization of germplasm resource collections, this study employed cutting-edge

genotyping-by-sequencing technology to delve into the intricacies of genetic diversity and genetic structure across thirteen distinct populations. The primary objectives of this research endeavor encompass the following: (1) elucidating the genetic diversity and genetic structure within the species *S. tzumu*; (2) discerning the interplay between genetic distance, geographical distance, and environmental factors; (3) formulating pertinent conservation policies. The findings of this study hold the potential to enhance our understanding of the genetic diversity and structural dynamics within the thirteen populations of *S. tzumu*, thereby furnishing the necessary scientific foundation and conservation strategies.

Results

Genotyping and sequencing analysis results

In this study, GBS technology was used to genotype 106 *S. tzumu* samples collected from 13 different distributions in China, and a total of 134,935 SNPs were obtained. A total of 11,862 SNPs were obtained after Vcftools filtering. Meanwhile, clean reads ranged from 2,946,429 (TMS10) to 7,999,731 (GZS03), and the mean value of clean reads of 106 samples was 4,693,475. HighQ reads ranged from 2,436,391 (TMS10) to 6,541,952 (JHS06), and the mean value was 3,815,536. HighQ reads rates ranged from 77% (SS01) to 88% (LS05), and the mean value of the HighQ reads rates was 83.22%. HighQ bases ranged from 345,967,522 (TMS10) to 928,957,184 (JHS06), and its mean value was 541,806,161. HighQ bases rates were from 73% (SS01) to 84% (LS05), and its average value was 78.77%. HighQ Q20 bases rates were from 98.934% (SS01) to 99.4015% (GZS08), HighQ Q30

bases rates were from 95.742% (GZS17) to 97.654% (GZS08), and their average values were 99.19% and 96.83%, respectively. HighQ N content ranged from 13 (FD08)—1738 (LS05), HighQ GC content ranged from 43.554 (WYS01) to 49.0725 (JGS02), and their mean values were 630.55 and 45.0424 (Supplementary file 1), respectively.

Genetic diversity and genetic differentiation

The populations exhibited a number range of 2.020 (JGS) to 2.571 (SS) alleles (N_A), with an average of 2.428. The effective number of alleles (N_E) ranged from 1.912 (JGS) to 2.129 (SS), with an average of 2.057. The Shannon's information index (I) varied from 0.628 (JGS) to 0.779 (SS), with a mean of 0.734. Expected heterozygosity (H_E) ranged from 0.426 (JGS) to 0.490 (SS), while observed heterozygosity (H_O) ranged from 0.492 (JHS) to 0.595 (JGS). The average values of H_E and H_O for *S. tzumu* were 0.469 and 0.515, respectively. The polymorphism information content (PIC) ranged from 0.8396 (JGS) to 0.9676 (SS), with a mean of 0.9459 (Table 1). The results of AMOVA revealed that the genetic variation of different populations of *S. tzumu* was 10% ($p < 0.01$) among population, 19% ($p < 0.01$) among individual, and the primary variation in *S. tzumu* was attributed to differences within populations (71%, $p < 0.01$). Therefore, there was greater genetic differentiation within populations (Table 2). Moderate genetic differentiation was observed among the 13 populations of *S. tzumu* ($F_{ST} = 0.103$, $p < 0.01$). The inbreeding coefficient (F_{IS}) was calculated to be 0.215 ($p < 0.01$). The number of migrants (N_m) between the 13 populations exceeded 1 ($N_m = 2.172$; Table 3). This

Table 1 Genetic diversity index of the 13 populations of *Sassafras tzumu*

Population	N_A	N_E	I	H_E	H_O	PIC
FD	2.458	2.054	0.735	0.469	0.495	0.9459
GZS	2.486	2.069	0.745	0.473	0.502	0.9555
HS	2.425	2.071	0.742	0.474	0.522	0.9439
JGS	2.020	1.912	0.628	0.426	0.595	0.8396
JHS	2.447	2.043	0.729	0.466	0.492	0.9448
LCS	2.450	2.056	0.735	0.469	0.498	0.9446
LS	2.391	2.073	0.739	0.474	0.534	0.9363
ML	2.511	2.083	0.752	0.476	0.503	0.9559
MS	2.524	2.076	0.750	0.474	0.501	0.9610
SS	2.571	2.129	0.779	0.490	0.519	0.9676
TMS	2.506	2.081	0.751	0.476	0.505	0.9546
TTS	2.504	2.078	0.750	0.476	0.503	0.9566
WYS	2.274	2.019	0.703	0.458	0.532	0.9063
Mean	2.428	2.057	0.734	0.469	0.515	0.9459

Abbreviations: N_A Number of alleles, N_E Effective number of alleles, I Shannon's information index, H_O Observed heterozygosity, H_E Expected heterozygosity, PIC Polymorphism information content

Table 2 The results of analyses of molecular variance (AMOVA) for *Sassafras tzumu*

Source	df	SS	MS	Est. Var	Percentage of variation (%)	P-Value
Among Pops	12	46677.17	3889.764	145.62	10	<0.01
Among Indiv	93	142978.7	1537.405	272.438	19	<0.01
Within Pops	106	105208	992.528	992.528	71	<0.01
Total	211	294863.8		1410.587	100	-

Abbreviations: df degrees of freedom, SS sums of squares, MS mean squares, Est. Var. estimated variance

Table 3 The change in genetic differentiation (F_{ST}) (above diagonal) where $p < 0.01$, inbreeding coefficient (F_{IS}), and number of migrants (N_m) among 13 populations of *Sassafras tzumu*

F-Statistics	Value	P-Value
F_{ST}	0.103	<0.01
F_{IS}	0.215	<0.01
N_m	2.172	

indicated that the 13 *S. tzumu* populations had more gene exchange overall at the species level.

Genetic structure analysis was conducted on the population data of 11,862 SNPs from the 13 populations. The results obtained from structure harvester indicated that the most suitable clustering was achieved when $K=4$ (Fig. 1 a, b). When $K=2$, 13 populations of *S. tzumu* could be divided into two genetic clusters, specifically, Hengshan Mountains (HS), Shaoshan Mountains (SS), Meiling Mountains (ML) and Lushan Mountains (LS) were divided into one cluster, and other populations were clustered into the other one. When $K=3$, 13 populations of *S. tzumu* could be divided into three genetic clusters, in which HS, SS, ML and LS were still one cluster, Jiuhua Mountains (JHS), Longchi Mountains (LCS), Maoshan Mountains (MS) and Tianmu Mountains (TMS) were clustered into one cluster, Fuding Mountains (FD), Guanzhai Mountains (GZS), Jigong Mountains (JGS), Tiantai Mountains (TTS) and Wuyi Mountains (WYS) were clustered into one cluster. The findings of the genetic structure analyses revealed that the populations were organized into distinct clusters when $K=4$. Cluster 1 consisted of samples from HS and SS, while cluster 2 comprised samples from ML and LS. Cluster 3 encompassed a significant number of samples from FD, GZS, JGS, MS, TMS, TTS and WYS. Additionally, a few specimens sourced from the JHS and LCS were consolidated within cluster 4 (Fig. 1c).

The first principal component (PC_1) and the second principal component (PC_2) of the PCA captured 7.68% and 11.94% of the total variance, respectively. The proportions of inertia associated with PC_1 and PC_2 were

statistically significant ($p < 0.01$). The results showed that HS, SS, ML and LS clustered, and others segregated from them (Figs. 2 and 3).

The multiple matrix regression with randomization (MMRR) analysis

The utilization of MMRR analysis had unveiled a noteworthy positive correlation between genetic and geographical distances amongst the 13 populations ($r=0.58$, $p < 0.01$; Fig. 4a). Within the 13 *S. tzumu* populations, the regression coefficient of environmental distance ($\beta_E=0.46$, $p < 0.01$) exceeded that of geographical distance ($\beta_D=0.16$, $p < 0.01$) by approximately threefold, implying that IBE predominantly accounted for the genetic distance. However, IBD also made a considerable contribution (Fig. 4b). Additionally, a significant association between genetic and environmental distances was observed ($r=0.56$, $p < 0.01$; Fig. 4c). Furthermore, the outcomes of MMRR analyses demonstrated a significant association between environmental distances and geographical distances ($r=0.97$, $p < 0.01$; Fig. 4d). In conclusion, the genetic distance among the 13 *S. tzumu* populations was jointly influenced by environmental and geographical distances (Table 4). Notably, environmental distance exhibited a more pronounced impact on genetic distance.

Discussion

Genetic diversity is the main factor leading to population survival and evolution. It determines the coping ability of natural populations under the stress of various biotic and abiotic factors, which is crucial for the long-term survival of populations [26]. Genetic diversity is directly affected by species' own factors, including mating system, meiotic behavior, gene flow, mutation and natural selection [27–29]. In addition, human activities and environmental factors can also affect the level of genetic diversity of species, such as temperature, water, light, wind, soil salinity and nutrients, and the surrounding biological communities may exert different selective pressures on plant populations, thereby determining the evolution process of the population and shaping the genetic diversity and structure of populations [30, 31]. The greater the genetic

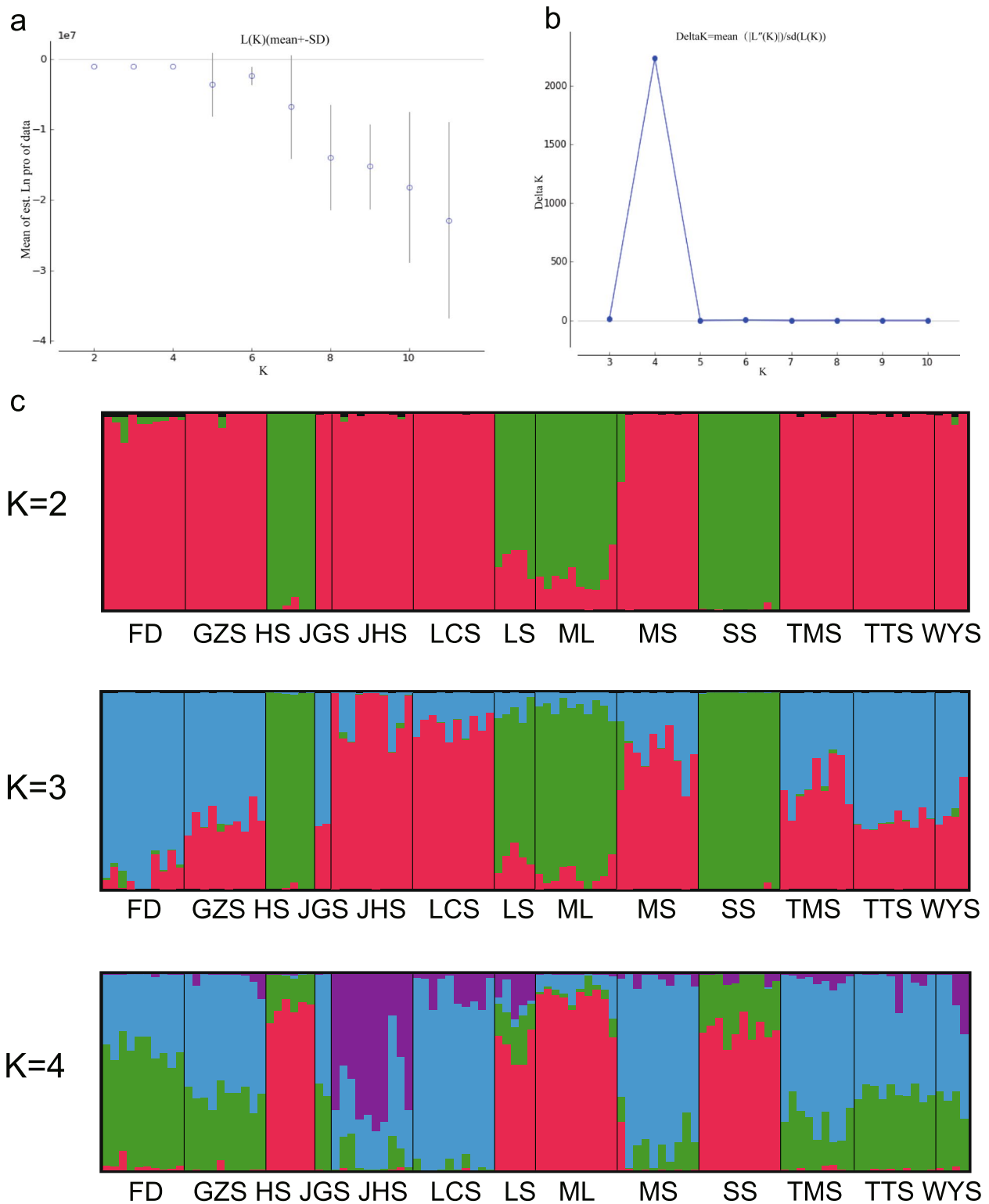


Fig. 1 Genetic structure analyses of 13 populations of *Sassafraz tzumu*. **a** Relationship between mean LnK and K value. **b** Relationship between Delta K and K value. **c** Genetic structure plots based on SNPs data (K=2, 3 and 4)

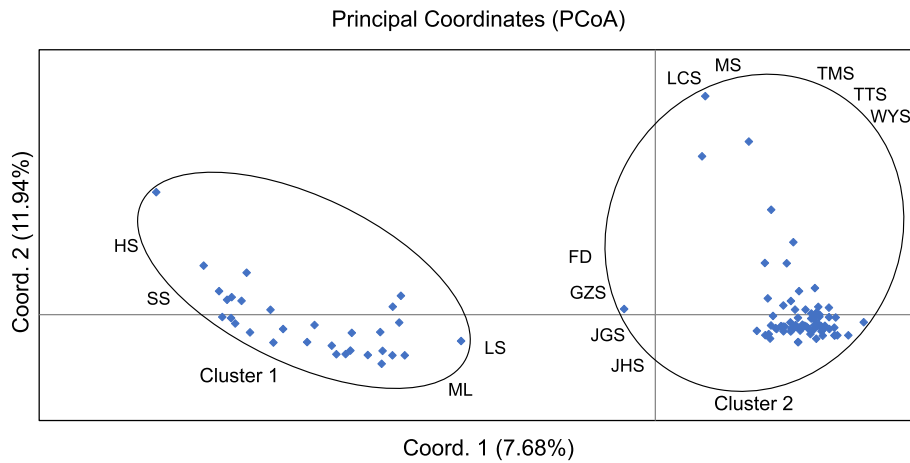


Fig. 2 Principal component analyses (PCA) among 13 populations. PCA axis 1 explained 7.68% of the variance, whereas PCA axis 2 explained 11.94%. Different colors represented 13 populations

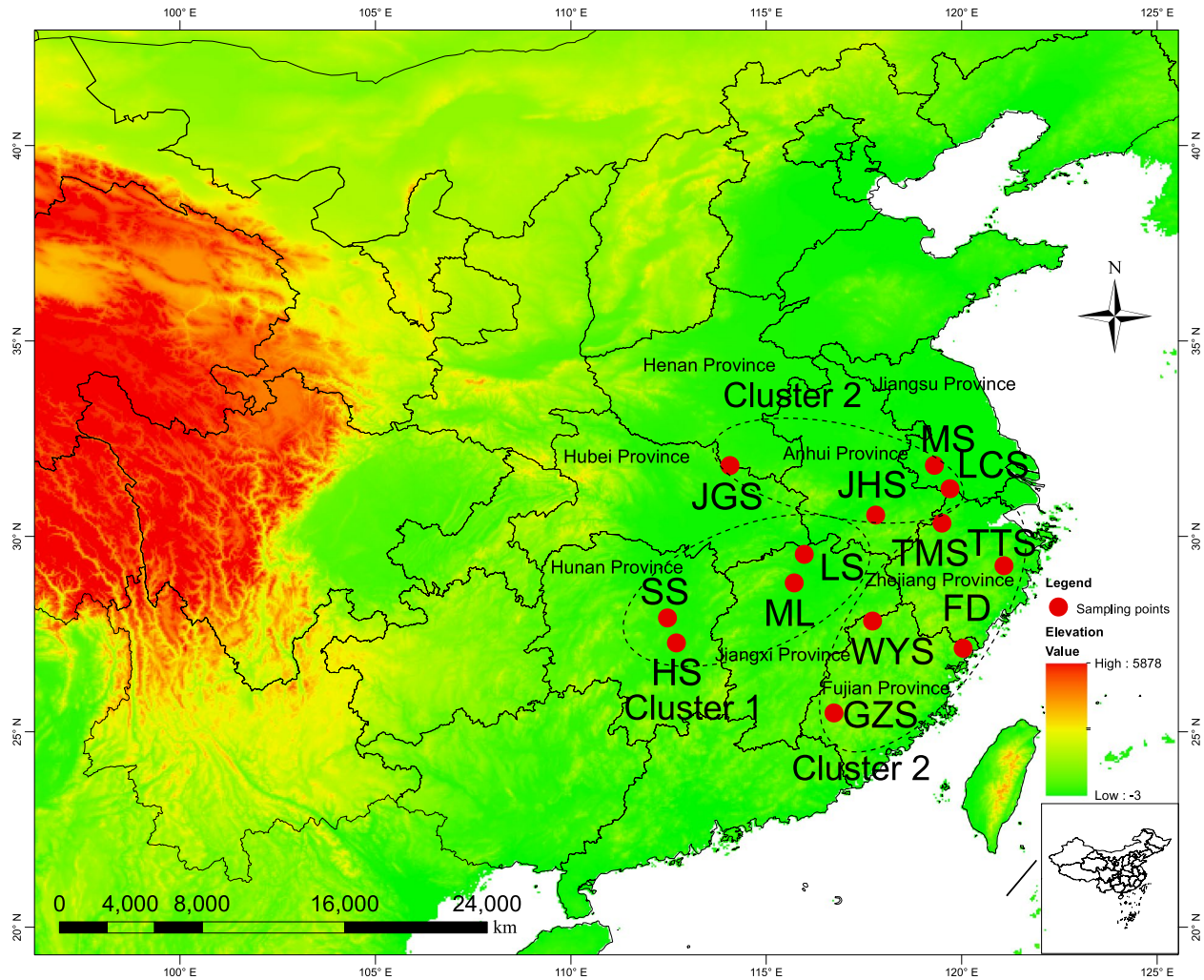


Fig. 3 Red points were sampling locations of *S. tzumu* used in this study. The map was obtained from the National Geomatics Center of China (NGCC, <https://www.ngcc.cn/ngcc/html/1/391/392/16114.html>)

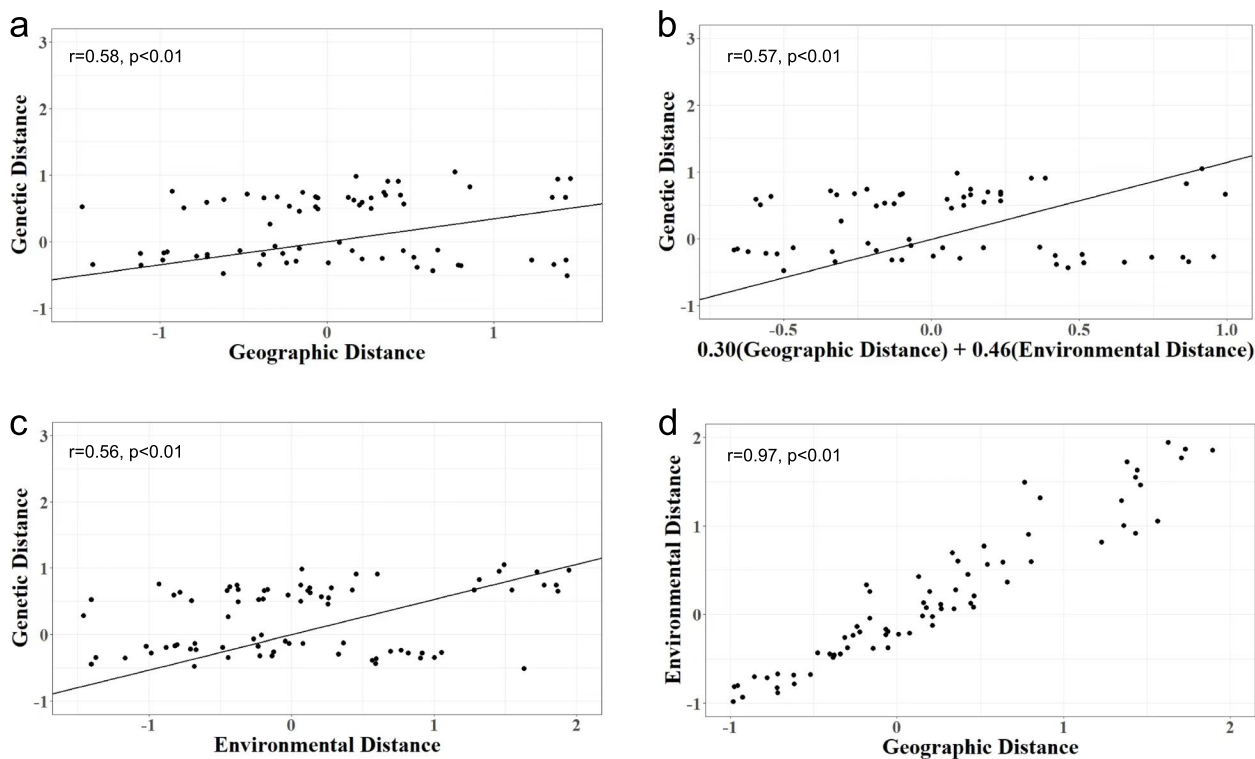


Fig. 4 **a** Correlations between genetic distance and geographical distance. **b** Geographical distance and environmental distance effects on genetic distance. **c** Correlations between genetic distance and environmental distance. **d** Correlations between environmental distance and geographical distance

Table 4 The results of multiple matrix regression with randomization (MMRR) analyses of among genetic distance and different factors

File1	File2	R. Squared	P-Value
Gen	Geo	0.58	$< 2.2e^{-16}$
Gen	Env	0.56	$9.831e^{-16}$
Gen	Geo and Env	0.57	$3.61e^{-16}$
Env	Geo	0.97	$< 2e^{-16}$

Abbreviations: Gen genetic distance, Geo geographical distance, Env environmental distance

variation of a species, the stronger its capacity to adapt to intricate and ever-changing environments [32]. We can quantify the genetic diversity of a species by calculating its expected heterozygosity (H_E) [33]. The mean H_E value of the 13 *S. tzumu* populations stood at 0.469, surpassing that of 39 *Cinnamomum camphora* populations distributed in China ($H_E=0.322$) [34]. Moreover, the H_E and H_O values of the 13 *S. tzumu* populations were lower than those of other arboreal species, such as *Xanthoceras sorbifolium* ($H_E=0.53$ and $H_O=0.72$) [35] and *Olea europaea* L. ($H_E=0.60$ and $H_O=0.75$) [36]. In contrast to the extensive genetic diversity observed in the aforementioned

studies, it can be inferred that the level of genetic diversity of the *S. tzumu* populations examined in this study is inherently high.

Wright [37] believed that genetic differentiation coefficient (F_{ST}) among populations was between 0 and 0.05, indicating that there was no genetic differentiation, and the F_{ST} value between 0.05 and 0.15 was moderately differentiated, then the F_{ST} value between 0.15 and 0.25 was highly differentiated. Furthermore, the mean F_{ST} value in our study was calculated to be 0.103, suggesting a moderate level of differentiation among 13 *S. tzumu* populations. Compared with other Lauraceae plants, it was much lower than and *Cryptocarya chinensis* ($F_{ST}=0.141$) [38] and *Litsea szemaois* ($G_{ST}=0.37$) [39], and had a similar value with *Cinnamomum camphora* ($F_{ST}=0.109$) [40]. The predominant distribution of genetic variation within 13 *S. tzumu* populations was observed to be within populations (71%, $p<0.01$), whereas the genetic variation among populations was minimal (10%, $p<0.01$). These findings aligned with the outcomes of previous studies on perennial plants [41] and further supported the notion that the genetic differentiation of *S. tzumu* primarily occurred within populations (93.6%, $p<0.001$)

[14]. Similar to Zhang's investigation, which focused on genus *Phoebe*, the primary source of variation in this species was within population (about 75.5%) rather than between populations (about 24.5%) [42]. Genetic diversity played a crucial role in adapting to extreme environments, and insufficient diversity could impede survival advantages in the face of environmental changes. Despite the current wide distribution of *S. tzumu*, it may lack sufficient genetic diversity to adapt to climate change. The inbreeding coefficient (F_{IS}) was calculated to be 0.215. It was worth noting that a high inbreeding coefficient might result in a decrease in genetic diversity and an increase in genetic structure within the 13 *S. tzumu* populations [43].

The reproductive strategy of *S. tzumu* entails cross-pollination, facilitated by its amphimerotic flowers. Insects act as the primary agents for pollen dispersal, while birds play a role in fruit dispersion [44]. These two mechanisms of transmission are the key drivers of gene flow among populations. The transfer of genetic information occurs through the fruits, thereby contributing to gene flow [16]. This study revealed a certain degree of gene exchange ($Nm=2.172$) among the 13 *S. tzumu* populations. This could be attributed to the prolonged presence of frugivorous birds within the same population, who consumed the fruit seeds of *S. tzumu* and facilitated gene exchange within populations. Additionally, it was determined that the genetic differentiation level among the 13 natural *S. tzumu* populations was a moderate significance.

Based on the genetic structure analysis results of 13 populations of *S. tzumu*, *S. tzumu* could be divided into four clusters: cluster 1 was distributed in Hunan Province (HS and SS populations), cluster 2 was distributed in Jiangxi Province (ML and LS populations), cluster 3 (FD, GZS, JGS, MS, TMS, TTS and WYS) was mostly located in the eastern part of the populations, and cluster 4 included JHS and LCS populations (Fig. 1c). However, the results of PCA supported the HS, SS, ML and LS groups were cluster 1 and other groups were cluster 2 (Figs. 2 and 3). And the results of IBD analysis showed that the spatial distribution pattern of gene clusters and geographical distance of *S. tzumu* populations were extremely significant correlation ($r=0.58$, $p<0.01$, see Fig. 4a). Based on the study of major peaks and mountains in China by Wang et al. [45], the divergence of cluster 1 and cluster 2 was related to the geographical boundaries of Wuyi Mountains. The four populations in this study, HS, SS, ML and LS formed one cluster, which was consistent with cluster 2 in Zhou's research results [46]. A total of 9 populations of *S. tzumu*, FD, GZS, JGS, JHS, LCS, MS, TMS, TTS and WYS constituted another cluster, which was consistent with cluster 1 in Zhou's study. Finally, we are more supportive of the grouping of genetic structures in PCA.

Woody plants have a similar life history. *Tapiscia sinensis* (Tapisciaceae), whose life history is defined as outcrossing, and its seeds and pollen also spread over short distances [47]. In addition, other species with dioecious or morphologically dioecious but functionally dioecious breeding systems tend to maintain high genetic structure [48, 49]. Meanwhile, some potential factors may affect genetic structure, such as habitat fragmentation, breeding system and life forms of plants [50, 51].

The Multiple Matrix Regression with Randomization (MMRR) analysis is a linear regression model that aims to quantify the impact of multiple explanatory variables. A portion of the genetic variation could be accounted for geographical distance ($r=0.58$, $p<0.01$; Fig. 4a). The *S. tzumu* populations in cluster 1 and cluster 2 were distinct due to geographical distance. This could explain the results of the previous principal component analysis. Moreover, the results of MMRR analyses demonstrated that genetic distance was influenced by both geographical distance and environmental distance ($r=0.57$, $p<0.01$; Fig. 4b). But environmental distance of *S. tzumu* in this study had a greater and profounder impact on its genetic diversity than geographical distance ($\beta_E=0.46$, $p<0.01$; $\beta_D=0.16$, $p<0.01$; Fig. 4b), which could be attributed to its preference for warm and humid habitats based on its biological characteristics [52]. IBD and IBE ($\beta_D=0.404$, $p<0.001$; $\beta_E=0.382$, $p<0.001$) were the major and equal contributor to genetic differentiation among 61 populations of *Neolitsea sericea* (Lauraceae) in the study of Cao et al. [53]. Notably, IBE ($r=0.56$, $p<0.01$; Fig. 4c) exerted a significant influence on genetic distance. In the event of unfavorable environmental conditions, such as cold frosts or inclement weather during the pollination period, the pollination success of *S. tzumu* could be easily compromised [16], resulting in blooming without fruiting. Meanwhile, the distribution pattern of *S. tzumu* populations was found to be associated with elevation, slope direction, slope position, and humus thickness [54]. Research indicated that relative humidity (accounting for 26.2% of permutation importance), mean temperature of coldest quarter (16.6%), annual precipitation (12.6%), and temperature annual range (10.3%) were the primary factors influencing the distribution of *S. tzumu* in China [15].

In the research, 13 populations of *S. tzumu* exhibited a moderate degree of genetic diversity, rendering them less resilient to diverse habitats and intricate climate fluctuations, thereby posing a potential threat to the species' survival [55]. Consequently, it is imperative to implement protective measures. Firstly, there is a need to enhance the promotion of plant preservation and cultivate public consciousness regarding this matter. This can be achieved by implementing localized protective measures

in economically underdeveloped and remote areas with limited transportation and inadequate awareness, as well as intensifying the dissemination of information regarding the conservation of *S. tzumu* to curb deforestation. Secondly, it is advisable to establish a nationwide network for ex situ conservation of plants, with National Botanical gardens playing a pivotal role. This network should encompass the creation of ecological corridors to facilitate gene flow among *S. tzumu* populations. Thirdly, in regions highly susceptible to climate change, it is recommended to introduce tree species that are known to adapt well to specific climatic conditions, instead of further expanding the cultivation of *S. tzumu*. Lastly, it is suggested to prioritize single plant selection and asexual propagation, while developing a comprehensive and scientifically guided breeding strategy for the development and utilization of *S. tzumu*.

In conclusion, we screened and developed 11,862 SNPs to effectively elucidate the genetic structure of *S. tzumu* populations. The findings revealed that *S. tzumu* populations exhibited a high level of genetic diversity. Simultaneously, it had come to our attention that the existing *S. tzumu* resources had suffered significant damage due to deforestation. In remote areas characterized by relatively underdeveloped economies, limited transportation, and insufficient protection and public awareness, it became imperative to implement in situ conservation measures and bolstered the dissemination of information regarding the protection of *S. tzumu* to deter indiscriminate deforestation. As for the development and utilization of *S. tzumu* resources, we propose placing emphasis on selective breeding of individual camphor trees, complemented by asexual reproduction. Additionally, it is crucial to formulate scientific and comprehensive breeding

strategies, manage *S. tzumu* genetic resources in a scientific manner, and enhance the sustainable utilization of these resources.

Methods

Collection and preservation of plant materials

During the summer of 2017, we gathered fresh leaves without pests or diseases as plant material samples from 13 naturally existing populations of *S. tzumu* across its range in China (Fig. 3). We made sure to collect individuals that were at least 50 m apart, with a collection of 2–10 individuals from each population. Ultimately, we carefully selected a total of 106 samples of *S. tzumu* for analysis (Table 5). To maintain the plant leaves' integrity, we carefully dried the leaves in sealed plastic bags with silica gel and stored them in a freezer at -20°C.

DNA extraction and genotyping-by-sequencing data set

For DNA extraction, dried floral leaves were subjected to the Plant Genomic DNA kit (Tiangen, Beijing, China) following the manufacturer's instructions. The extracted DNA from all samples was quantified using a Qubit spectrophotometer. To facilitate genotyping by sequencing, all samples were sent to the esteemed Institute of Shanghai OE Biotech. Co., Ltd. In brief, the DNAs from the 106 individuals were digested with *EcoRI*-HF and *MspI* to reduce the complexity of the genome. Subsequently, a 106-plex GBS library, comprising 105 DNA samples and a negative control (no DNA), was meticulously prepared by ligating the digested DNA to unique barcode nucleotide adapters, followed by standard PCR amplification. Finally, the resulting 106-plex library was sequenced on a single lane of an HiSeq X Ten platform. Sequence data obtained from the HiSeq X Ten platform were processed

Table 5 The sampling locations of *Sassafras tzumu*

Code	Longitude	Latitude	Number of Samples	Temperature	Precipitation	Altitude (m)
FD	120.0398	27.13686	10	16.48	1686	497
GZS	116.7378	25.4839	10	19.03	1708	157
HS	112.7029	27.27414	6	16.03	1763	598
JGS	114.0766	31.81327	2	14.08	1212	158
JHS	117.8004	30.54553	10	16.12	1477	91
LCS	119.702	31.22193	10	15.52	1168	59
LS	115.9744	29.54243	5	12.58	1626	98
ML	115.7197	28.80984	10	17.04	898	397
MS	119.3055	31.8222	10	15.5	1078	384
SS	112.4749	27.91357	10	17.24	1510	479
TMS	119.5007	30.3415	9	14.1	1548	882
TTS	121.086	29.2522	10	12.44	1708	902
WYS	117.7218	27.83386	4	12.75	2135	957

for SNPs identification using the GBS pipeline integrated within Stacks 3.0.166 [56]. In brief, variants were filtered using VcfTools 0.1.14 [57] based on criteria such as a minimum depth of coverage (>5), quality score (>30), and an initial maximum missingness of 50% [58]. Subsequently, the dataset was examined to ensure that no samples exhibited high levels of missingness (all samples were found to have less than 30% missing data). The final set of SNPs was further filtered to retain those with a maximum of 20% missing values and a minor allele frequency below 0.05.

Genetic diversity and genetic structure analysis

The resulting vcf file in our GBS data set was converted into compatible formats using PGDspider 2.0.9.0 [59]. Genetic differentiation (F_{ST}) was calculated for each SNP using GenAEx 6.501 [60], and these values were utilized for analyses of molecular variance (AMOVA) among populations. To determine the number of genetic clusters across populations, the STRUCTURE 2.3.4 [61] was employed. The genetic structure analysis was conducted with three independent replicates, ranging from $K=2$ to $K=13$. Each run consisted of 5.0×10^4 burn-in iterations and 3.0×10^4 MCMC replications under the admixture model. The resulting output was submitted to the Structure Harvester [62] website for further analysis. The best K value was determined to represent the best number of clusters. The Q-matrix data, including Kx.indfile (individuals) and Kx.popfile (populations), obtained from multiple runs were downloaded. These data were then analyzed using CLUMPP [63] and visualized using DISSTRUCT [64] and Adobe Illustrator 2021. To explore the clustering patterns between populations, principal components analysis (PCA) was performed using GenAEx 6.501 [60].

The multiple matrix regression with randomization analysis

Geographical distance and environmental distance affect gene flow and genetic differentiation between populations [65], and isolation by distance (IBD) and isolation by environment (IBE) are important ways to reflect that the complex landscapes affect the genetic structure of natural populations [66–70]. When gene flow may be affected by geographical and environmental variables, multiple matrix regression analysis (MMRR) [71] provides a valuable approach to quantify the impact of geographical and environmental isolation on genetic distance. In contrast to the many methods used to analyze distance matrices in landscape genetics [72–74], MMRR can ask not only whether variables are correlated but also how the dependent variable changes with respect to multiple independent variables, contributing to a

comprehensive understanding of how landscape influences gene flow patterns.

In the study, the geosphere package in R software [75] was utilized to compute the euclidean distance of the latitude and longitude information from 13 populations. Additionally, MEGA 11 [76] was employed to determine the genetic distance among these populations. In this study, we used climate variables, soil pH and elevation as environmental factors. The climate variables (1970–2000) and elevation layer were downloaded from WorldClim website (<https://worldclim.org>), and soil pH was downloaded from ISRIC — World Soil Information (<https://www.isric.org/>). Then, we used Maxent 3.3.1 [77] to simulate the suitable layer in the species distribution model. SDM Tools in ArcMap 10.8 was used for converting from environment suitability layer to migration resistance layer. The value of the environmental layer in sampling locations was extracted from the matrix format utilizing ArcMap 10.8 [78], and the specific operation was “Spatial Analyst Tools—Extract by points”.

The isolation by distance (IBD) and isolation by environment (IBE) were quantified using the “MMRR” function in R software [79]. We used genetic information as the response variable, with geographical and environmental distances serving as explanatory variables to describe the patterns of IBD and IBE, the correlation between the three matrices was computed to assess the impact of geographical distance and environmental distance on genetic distance.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12862-024-02264-9>.

Supplementary Material 1.

Acknowledgements

We thank the other members of our research group from Laboratory of Landscape Genetics, Nanchang University for their assistance.

Authors' contributions

Conceptualization, Bc.G., Q.L., X.L., and X.G.; Data curation, X.L.; Formal analysis, Q.L.; Funding acquisition, X.G.; Investigation, X.L. and X.G.; Methodology, Q.L. and Bc.G.; Resources, Bc.G.; Visualization, Q.L.; Writing-original draft, Q.L. and X.L.; Writing-review and editing, Bc. G., Q.L., X.L., and X.G.; Supervision, X.G. and Bc.G.; All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Natural Science Foundation of Jiangxi Province (Grant No.20212BAB205030).

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the Figshare repository under persistent web link “https://figshare.com/articles/dataset/Data_in_VCF_format_for_single_nucleotide_polymorphisms_SNPs_of_13_i_Sassafras_tzumu_i_Lauraceae_populations_with_106_individuals/24899328”.

Declarations

Ethics approval and consent to participate

We ensure the collection of the plants used in the study (*Sassafras tzumu*) complies with Chinese guidelines. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable to this study.

Competing interests

The authors declare no competing interests.

Received: 3 October 2023 Accepted: 24 May 2024

Published online: 13 June 2024

References

- Buzhdygan OY, Meyer ST, Weisser WWG, Eisenhauer N, Ebeling A, Borrett SR, et al. Biodiversity increases multitrophic energy use efficiency, flow and storage in grasslands. *Nat Ecol Evol.* 2020;4:393–405.
- Isbell F, Craven D, Connolly J, Loreau M, Schmid B, Beierkuhnlein C, et al. Biodiversity increases the resistance of ecosystem productivity to climate extremes. *Nature.* 2015;526:574–89.
- Agashe D, Falk JJ, Bolnick DI. Effects of founding genetic variation on adaptation to a novel resource. *Evolution.* 2011;65:2481–91.
- Hu XQ, Wu S, Wang YG, Zhao HB, Zhang YY. Genetic diversity and genetic structure of different types of natural populations in *Osmanthus fragrans Lour.* and the relationships with sex ratio, population structure, and geographic isolation. *Sci World J.* 2014;2014:817080.
- Hamrick JL, Godt MJW. Allozyme diversity in plant species. *Crop Sci.* 1997;37:26–30.
- Baker HG. Self-compatibility and establishment after “long-distance” dispersal. *Evolution.* 1955;9:347–9.
- Grossenbacher D, Runquist RB, Goldberg EE, Brandvain Y. Geographic range size is predicted by plant mating system. *Ecol Lett.* 2015;18:706–13.
- Pannell JR. Evolution of the mating system in colonizing plants. *Mol Ecol.* 2015;24:2018–37.
- Stebbins GL. Self-fertilization and population variability in the higher plants. *Am Nat.* 1957;91:337–54.
- Loveless MD, Hamrick JL. Ecological determinants of genetic structure in plant populations. *Annu Rev Ecol Evol Syst.* 1984;15:65–95.
- Yan JW, Bai WF, Li JH, Yu L, Nie DL, Xiong Y, Wu SZ. Genetic structure and diversity of *Cerasus conradinae* in Hunan Province. *J Plant Genet Resour.* 2022;23:376–84.
- Li XW, Li J, van der Werff H. *Sassafras* J. Presl in Berchtold & J. Presl Priir. *Roslin 2(2):* 30. 1825. Flora of China. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press; 2008:159–161.
- Chen HZ, Liu J, Jiang JM, Sun Y, Dong X, Fu AP. A study on flowering and fruit development of *Sassafras tzumu*. *For Res.* 2020;33:148–54.
- Jiang AP, Jiang JM, Liu J. Genetic diversity and genetic structure in *Sassafras tzumu* populations along altitudinal gradients in Tianmushan Mountain. *China J Appl Ecol.* 2016;27:1829–36.
- Zhang KL, Zhang Y, Jia DW, Tao J. Species distribution modeling of *Sassafras tzumu* and implications for forest management. *Sustainability.* 2020;12:4132.
- Chung KF, van der Werff H, Peng CI. Observations on the floral morphology of *Sassafras randaiense* (Lauraceae). *Ann Mo Bot Gard.* 2010;97:1–10.
- Wang S, Wang Y, Zhou JB, Li P, Lin HW, Peng Y, et al. Genetic diversity and population structure of an arctic tertiary relict tree endemic to China (*Sassafras tzumu*) revealed by novel nuclear microsatellite (nSSR) markers. *Plants.* 2022;11:2706.18.
- Ding XF, Chen HL, Cao J, Koozo K, Masatoshi U, Masanori O. Preliminary study on genetic structure of three natural populations in *Sassafras Tzumu*. *Hubei Forestry Science and Technology.* 2006;5:1–2.
- Liu X, Gong X, Chen SS, Guan BC. Simulation of the distribution pattern of *Sassafras tzumu* and changes in habitat based on ArcGIS and MaxEnt. *Plant Sci J.* 2018;36:320–6.
- Wu ZY, Zong D, Gan PH, Liu ZT, Zhu MC, He CZ. Genetic analysis on different stem forms of *Pinus yunnanensis* by reduced-representation genome sequencing. *J Yunnan Agric Univ (Nat Sci).* 2022;37:285–93.
- Poland JA, Rife TW. Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome.* 2012;5:92–102.
- Rowe HC, Renaut S, Guggisberg A. RAD in the realm of next-generation sequencing technologies. *Mol Ecol.* 2011;20:3499–502.
- Alipour H, Bihamta MR, Mohammadi V, Peyghambari SA, Bai GH, Zhang GR. Genotyping-by-sequencing (GBS) revealed molecular genetic diversity of Iranian wheat Landraces and cultivars. *Front Plant Sci.* 2017;8:1293.
- Wang J, Li ZJ, Guo QH, Ren GP, Wu YX. Genetic variation within and between populations of a desert poplar (*Populus euphratica*) revealed by SSR markers. *Ann For Sci.* 2011;68:1143–9.
- Zhao Y, Li GY, Yang J. Conservation and utilization of wild relatives of cultivated plants. *Biodiversity Science.* 2018;26:414–26.
- Khan MK, Pandey A, Thomas G, Akkaya MS, Kayis SA, Ozsensoy Y, et al. Genetic diversity and population structure of wheat in India and Turkey. *AoB Plants.* 2015;7:plv083.
- Levin DA, Kerster HW. Gene flow in seed plants. *Evol Biol.* 1974;7:139–220.
- Linhart YB, Grant MC. Evolutionary significance of local genetic differentiation in plants. *Annu Rev Ecol Evol Syst.* 1996;27:237–77.
- Gutiérrez-Ortega JS, Molina-Freaner F, Martínez JF, Pérez-Farrera MA, Vovides AP, Hernández-López A, et al. Speciation along a latitudinal gradient: The origin of the Neotropical cycad sister pair *Dioon sonorense*–*D. vovidesii* (Zamiaceae). *Ecol Evol.* 2021;11:6962–76.
- Peleg Z, Saranga Y, Krugman T, Abbo S, Nevo E, Fahima T. Allelic diversity associated with aridity gradient in wild emmer wheat populations. *Plant Cell Environ.* 2008;31:39–49.
- Castilla AR, Méndez-Vigo B, Marcer A, Martínez-Minaya J, Conesa D, Picó FX, et al. Ecological, genetic and evolutionary drivers of regional genetic differentiation in *Arabidopsis thaliana*. *BMC Evol Biol.* 2020;20:71.
- Du QZ, Wang BW, Wei ZZ, Zhang DQ, Li BL. Genetic diversity and population structure of Chinese white poplar (*Populus tomentosa*) revealed by SSR markers. *J Hered.* 2012;103:853–62.
- Li CH, Zheng YQ, Liu Y, Lin FR, Huang P. Development of genomic SSR for the subtropical hardwood tree *Dalbergia hupeana* and assessment of their transferability to other related species. *Forests.* 2021;12:804.
- Gong X, Yang A, Wu Z, Chen C, Li H, Liu Q, et al. Employing genome-wide SNP discovery to characterize the genetic diversity in *Cinnamomum camphora* using genotyping by sequencing. *Forests.* 2021;12(11):1511.
- Shen Z, Zhang KQ, Ma LY, Duan J, Ao Y. Analysis of the genetic relationships and diversity among 11 populations of *Xanthoceras sorbifolia* using phenotypic and microsatellite marker data. *Electron J Biotechnol.* 2017;26:33–9.
- Dervishi A, Jakše J, Ismaili H, Javornik B, Štajner N. Comparative assessment of genetic diversity in Albanian olive (*Olea europaea* L) using SSRs from anonymous and transcribed genomic regions. *Tree Genet Genomes.* 2018;14:53.
- Wright S. *Evolution and the genetics of population, variability within and among natural populations.* Chicago: University of Chicago Press; 1978;4:213–20.
- Gao SH, Wang ZF, Zhang JL, Tian SN. Genetic diversity of *Cryptocarya chinensis* life stages in heishiding Guangdong Province. *J Plant Genet Resour.* 2005;6:300–3.
- Chen JQ, Ci XQ, Li QM, Li J. Genetic diversity of *Litsea szemaosis*, an endangered species endemic to China, detected by inter-simple sequence repeat (ISSR). *Biodivers Sci.* 2006;14:410–20.
- Liu S. Population genetic structure of *Cinnamomum camphora* in China [D]. Master Dissertation, Jiangxi Agricultural University, Nanchang, China. 2019.
- Xing C, Tian Y, Guan FC, Meng FJ. Evaluation of genetic diversity in *Amygdalus mira* (Koehne) Ricker using SSR and ISSR markers. *Plant Syst Evol.* 2015;301:1055–64.
- Zhang R, Zhou ZC, Jin GQ, Wang SH, Wang XH. Genetic diversity and differentiation within three species of the family Lauraceae in southeast China. *Biochem Syst Ecol.* 2012;44:317–24.
- Cole CT. Genetic variation in rare and common plants. *Annu Rev Ecol Evol Syst.* 2003;34:213–37.
- Yang Z, Tan C, Wei YM, Rohwer JG, Liu B, Yang Y. Floral morphology and phenology of *Sassafras tzumu* (Lauraceae). *BMC Plant Biol.* 2022;22:327.

45. Wang XP, Wang ZH, Fang JY. Mountain ranges and peaks in China. *Biodiversity Science*. 2004;01:206–12.
46. Zhou JB. The population genetics of *Sassafras tzumu* based on nuclear microsatellite (nSSR) [D]. Master Dissertation, Nanjing University, Nanjing, China. 2021.
47. Zhang JJ, Li ZZ, Fritsch PW, Tian H, Yang AH, Yao XH. Phylogeography and genetic structure of a Tertiary relict tree species, *Tapiscia sinensis* (Tapisciaceae): implications for conservation. *Ann Bot*. 2015;116:727–37.
48. Rendell S, Ennos RA. Chloroplast DNA diversity of the dioecious European tree *Ilex aquifolium* L. (English holly). *Mol Ecol*. 2003;12:2681–8.
49. Wang J, Gao P, Kang M, Lowe AJ, Huang H. Refugia within refugia: the case study of a canopy tree *Eurycorymbus cavaleriei* in subtropical China. *J Biogeogr*. 2009;36:2156–64.
50. Wang HW, Ge S. Phylogeography of the endangered *Cathaya argyrophylla* (Pinaceae) inferred from sequence variation of mitochondrial and nuclear DNA. *Mol Ecol*. 2006;15:4109–22.
51. Vekemans X, Hardy OJ. New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol Ecol*. 2004;13(4):921–35.
52. Sun JJ, Jiang B, Wu CP, Yuan WG, Zhu JR, Huang YJ, et al. Study on the habitat and niche of *Sassafras tzumu* (Hemsl.) Hemsl. in Zhejiang Province. *Acta Ecologica Sinica*. 2019;39:884–94.
53. Cao YN, Wang JJ, Chen LY, Ding YQ, Liu LX, Qiu YX. Inferring spatial patterns and drivers of population divergence of *Neolitsea sericea* (Lauraceae), based on molecular phylogeography and landscape genomics. *Mol Phylogenet Evol*. 2018;126:162–72.
54. Sun JJ, Guo JH, Shen AH, Xu X, Feng HL, Zhang S, et al. Composition and environmental interpretation of the communities of *Sassafras tzumu*, a protected species, at Zhejiang Province in eastern China. *Glob Ecol Conserv*. 2020;24:e01218.
55. Zhou PY, Hui LX, Huang SJ, Ni ZX, Yu FX, Xu LA. Study on the genetic structure based on geographic populations of the endangered tree species: *Liriodendron chinense*. *Forests*. 2021;12:917.
56. Catchen JL, Hohenlohe PA, Bassham SS, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. *Mol Ecol*. 2013;22:3124–40.
57. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–8.
58. da Cunha NL, Xue HR, Wright SI, Barrett SCH. Genetic variation and clonal diversity in floating aquatic plants: Comparative genomic analysis of water hyacinth species in their native range. *Mol Ecol*. 2022;31(20):5307–25.
59. Lischer HEL, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*. 2012;28:298–9.
60. Smouse PE, Whitehead MR, Peakall R. An informational diversity framework, illustrated with sexually deceptive orchids in early stages of speciation. *Mol Ecol Resour*. 2015;15:1375–84.
61. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–59.
62. Earl DA, VonHoldt BM. Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Cons Genet Res*. 2012;4:359–61.
63. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 2007;23:1801–6.
64. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes*. 2003;4:137–8.
65. Wright S. Isolation by distance. *Genetics*. 1943;28:114–38.
66. Nosil P. Reproductive isolation caused by visual predation on migrants between divergent environments. *Proc Biol Sci*. 2004;271:1521–8.
67. Nosil P, Vines TH, Funk DJ. Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*. 2005;59:705–19.
68. Wang JJ, Summers K. Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Mol Ecol*. 2010;19:447–58.
69. Lee-Yaw JA, Davidson A, Mcrae BH, Green DM. Do landscape processes predict phylogeographic patterns in the wood frog? *Mol Ecol*. 2009;8:1863–74.
70. Zellmer AJ, Hanes MM, Hird SM, Carstens BC. Deep phylogeographic structure and environmental differentiation in the carnivorous plant *Sarracenia alata*. *Syst Biol*. 2012;61:763–77.
71. Wang JJ. Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*. 2013;67:3403–11.
72. Holderegger R, Wagner HH. Landscape genetics. *Bioscience*. 2008;58:199–207.
73. Balkenhol N, Waits LP, Dezzani RJ. Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*. 2009;32:818–30.
74. Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP. Landscape genetics: where are we now? *Mol Ecol*. 2010;19:3496–514.
75. Karney CFF. Algorithms for geodesics. *J Geod*. 2013;87:43–55.
76. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol*. 2021;38:3022–7.
77. Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecol Model*. 2006;190:231–59.
78. Brown JL. SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods Ecol Evol*. 2014;5:694–700.
79. Legendre P, Lapointe FJ, Casgrain P. Modeling brain evolution from behavior: a permutational regression approach. *Evolution*. 1994;48:1487–99.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.